HAL
open science

Une étude didactique sur l'introduction dans l'enseignement mathématique vietnamien de notions statistiques dans leurs liens avec les probabilités

Thu Huong Vu Nhu

To cite this version:

Thu Huong Vu Nhu. Une étude didactique sur l'introduction dans l'enseignement mathématique vietnamien de notions statistiques dans leurs liens avec les probabilités. Mathématiques [math]. Université Joseph-Fourier - Grenoble I, 2009. Français. NNT : . tel-00418306v2

HAL Id: tel-00418306

https://theses.hal.science/tel-00418306v2

Submitted on 27 Nov 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE EN CO-TUTELLE

Présentée par

VŨ NHƯ Thư Hương

Pour obtenir les titres de

Docteur de l'Université Joseph Fourier, Grenoble et Docteur du Vietnam

Spécialité : MATHÉMATIQUES ET INFORMATIQUE

UNE ETUDE DIDACTIQUE SUR L'INTRODUCTION DANS L'ENSEIGNEMENT MATHEMATIQUE VIETNAMIEN DE NOTIONS STATISTIQUES DANS LEURS LIENS AVEC LES PROBABILITES

Soutenue le 1er Septembre 2009

Devant le jury composé de :

Nicolas **BALACHEFF**	Directeur de thèse
Alain **BIREBENT**	Co-directeur de thèse
Jean-Luc **DORIER**	Rapporteur
Colette **LABORDE**	Examinatrice
LÊ THÁI BẢO Thiên Trung	Examinateur
LÊ THỊ Hoài Châu	Co-directeur de thèse
NGUYỄN Chí Thành	Examinateur
TRẦN LƯƠNG Công Khanh	Rapporteur
Annie **BESSOT**	Invitée

Thèse préparée au sein des équipes

Didactique et Informatique pour l'Apprentissage des Mathématiques
Laboratoire LIG, Université Joseph Fourier, Grenoble

Didactic Toán, Département de l'Informatique et des Mathématiques
Université de Pédagogie de Ho Chi Minh ville

Abstract

« A didactic study of introduction of statistic concepts in their relations with the probability in teaching of mathematics in Vietnam »

The study focuses on Vietnamese high school teaching of statistics in its potential links with concepts of probability. It derives from two observations, the first concerned with epistemology and the second with didactic:

- there is a body of knowledge in inferential statistic which mathematically characterizes the relationship between statistic and probabilities

- the two education systems, Vietnamese and French, differ, as in Vietnam statistics and probability are taught separately and these subjects are missing in the standard curriculum, while in France recent reforms have included some of these knowledges in teaching programs via such concepts as sampling statistical fluctuations or adequacy of a statistical distribution to a probability law.

Cross linking epistemological analysis and institutional analysis carries the study's main issue to develop into two successive questions: What could be the effect of the observed gap between probability and statistics on the institutional relationship to "objects-candidates for the statistics-probabilities links" in Vietnam from the teacher's side as well as from the learner's side? Is it possible to keep some of these objects-candidates going in the present conditions of the Vietnamese education? Each question is worked upon in an experimental way: the first through a questionnaire targeting Vietnamese students and teachers, and the second by mean of didactic engineering carried out in a Vietnamese high school class.

Keywords

Study of mathematic, descriptive and inferential statistic, analysis and didactical engineering

Remerciements

J’adresse mes remerciements sincères à Alain Birebent, qui a sacrifié beaucoup de son temps à la thèse, du point de départ au point d’arrivée : tu étais prêt, non seulement pour m’aider à compléter mes connaissances en statistique mais aussi pour m’apprendre des proverbes ; efficace et exigeant, tu m’as poussée et encouragée tout au long de ces trois années de thèse et cela m’a permis de mener cette recherche jusqu’au *« bout du tunnel »*.

Mes profonds remerciements vont aussi à Annie Bessot pour sa disponibilité : tu as largement participé à la progression de mon travail ; tes idées précieuses, notamment dans les rendez-vous à trois avec Alain et moi, ont toujours été pour moi des moments d’apprentissage privilégiés.

J’exprime une profonde gratitude à Lê Thị Hoài Châu avec qui j’ai commencé mon mastère : tu m’as accordé du temps pour réfléchir sur mon travail, tu m’as accompagnée pendant toute la phase d’exprérimentation, nous avons discuté du résumé de la thèse en vietnamien et de la présentation pendant la soutenance.

C’est pour moi un grand honneur que Jean-Luc Dorier et Tran Luong Cong Khanh aient rapporté sur cette thèse. Leurs remarques claires et encourageantes m’ont été utiles. Qu’ils veuillent bien recevoir ici mes chaleureux remerciements.

Merci à Nicolas Balacheff qui a accepté d’être le directeur officiel de la thèse, merci à Colette Laborde qui a bien voulu présider le jury, et merci à tous les membres vietnamiens du jury, Nguyen Chi Thanh , Le Thai bao Thien Trung.

J’ai eu le plaisir de partager avec mes collègues de l’équipe DDM puis l’équipe DIAM des moments agréables et je les remercie pour m’avoir toujours encouragé. Merci également pour leur soutien sans restriction, leur gentillesse. Je pense fortement à Claude, Jean-François, Khanh, Trung, Sophie-Sourry, Bernard, Zilora, Angela, Joris, Nga et aussi particulièrement à Henri, Jacky pour leurs aides en informatique.

Je n’oublierai jamais les moments agréables partagés avec la communauté catholique vietnamienne à Grenoble, avec l’équipe solidarité Vietnam, avec mes amis vietnamiens et français qui m’ont accompagné de près ou de loin, au cours de ces trois années. Même si je ne vous ai pas tous cités, sachez que je pense à vous.

Ma reconnaissance va, bien entendu, à la région Rhône-Alpes pour m’avoir octroyé une bourse de recherche qui m’a aidé financièrement à réaliser cette thèse.

Des pensées affectueuses à mon père qui aurait voulu être à mes côtés, à ma mère tant dévouée, à mes sœurs et mes frères, ma nièce Tu Dung et tous mes proches qui m’ont toujours encouragée.

Table des matières

Introduction **11**

I. Le contexte de l'EMS vietnamien 15

I.1. La formation des enseignants 15

I.2. L'interdisciplinarité 16

I.3. L'environnement technologique 17

I.4. Intérêt d'une analyse institutionnelle comparative France-Vietnam 18

II. Outils théoriques et méthodologiques 19

II.1. La praxéologie mathématique comme objet élémentaire d'une analyse institutionnelle 20

II.2. Le modèle épistémologique de référence 21

II.3. Contrat institutionnel et contrats didactiques 22

II.4. Ingénierie didactique 23

III. Organisation de la thèse 24

Partie A. Analyse institutionnelle comparative France-Vietnam **27**

I. Analyse institutionnelle française 29

I.1. Avant la réforme (avant l'an 2000) 29

I.2. Après la réforme (après l'an 2000) 36

I.3. Conclusion 48

II. Analyse institutionnelle vietnamienne 49

II.1. Les programmes actuels 49

II.2. Les manuels 50

III. Conclusions de cette première analyse institutionnelle 55

Partie B. Analyse épistémologique **57**

I. Présentations comparées des deux traités 59

I.1. Le traité Saporta 59

I.2. Le traité Wonnacott 61

II. Descriptions comparées des deux traités 61

II.1. Le traité Saporta 62

II.2. Le traité Wonnacott 66

III. Un modèle praxéologique de référence 71

III.1. Trois problèmes 71

III.2. Deux praxéologies 72

Partie C. Retour sur l'analyse institutionnelle pour le Vietnam 75

Chapitre C1. Contrat institutionnel à la statistique 77

I. Rappel des premiers résultats d'analyse institutionnelle 77

II. Incomplétude des praxéologies à enseigner 78

III. Le contrat institutionnel à la Statistique dans EMS Vietnam 84

Chapitre C2. Mises à l'épreuve d'hypothèses sur les contrats didactiques par un questionnaire aux élèves et aux enseignants..... 87

I. Exercice 1 87

I.1. Quelques éléments d'analyse institutionnelle vietnamienne à propos de l'objet histogramme 87

I.2. Présentation de l'exercice 98

I.3. Analyse a priori 100

I.4. Analyse a posteriori 111

I.5. Conclusion 118

II. Exercice 2 119

II.1. Quelques éléments d'analyse institutionnelle vietnamienne à propos de l'objet courbe polygonale 119

II.2. Présentation de l'exercice 127

II.3. Analyse a priori 128

II.4. Analyse a posteriori 132

II.5. Conclusion 135

III. Exercice 3 136

III.1. Quelques éléments d'analyse institutionnelle vietnamienne à propos de la notion de représentativité dans une étude statistique. 136

III.2. Présentation de l'exercice 149

III.3. Analyse a priori 149

III.4. Analyse a posteriori 160

III.5. Conclusion : 178

IV. Exercice 4 180

IV.1. Quelques éléments d'analyse institutionnelle vietnamienne à propos de la notion d'échantillon représentatif. 180

IV.2. Présentation de l'exercice 185

IV.3. Analyse a priori 186

IV.4. Analyse a posteriori 197

IV.3. Conclusion 211

Partie D. Ingénierie didactique 213

Chapitre D1. Introduction à l'ingénierie didactique.....215

I. Pourquoi une ingénierie didactique ? 215

II. Présentation des ingénieries de Brousseau et Briand.....216

II.1. La situation initiale de l'ingénierie de Brousseau, décrite par Brousseau lui-même : 217

II.2. La genèse du problème d'incertitude dans l'ingénierie de Brousseau, résumée par nous 219

II.3. Examen d'une possibilité de transposition hors du primaire 220

II.4. L'ingénierie de Briand, décrite par Briand lui-même 221

II.5. Examen de la transposition réalisée par Briand 223

III. Rapports institutionnels aux notions d'expérience et d'incertitude aléatoires 225

III.1. Notion d'expérience aléatoire : 226

III.2. Notion d'incertitude aléatoire 230

IV. Organisation de l'ingénierie 232

Chapitre D2. L'ingénierie didactique 233

I. Séance 1 233

I.1. Introduction.....233

I.2. Protocole et analyse a priori de la première phase (20 minutes) 233

I.3. Déroulement et analyse a posteriori de la première phase.....236

I.4. Protocole et analyse a priori de la deuxième phase (25 minutes).....239

I.5. Déroulement et analyse a posteriori de la deuxième phase 252

I.6. Conclusion 263

II. Séance 2 264

II.1. Introduction 264

II.2. La première phase (25 minutes) 268

II.3. La deuxième phase (20 minutes).....283

II.4. Conclusion.....289

III. Séance 3 291

III.1. Introduction.....291

III.2. La première phase (20 minutes).....296

III.3. La deuxième phase (25 minutes).....303

III.4. Conclusion 326

Conclusions et perspectives 327
I. Ce qui a motivé l'étude didactique de la thèse 329
II. Principaux apports du travail de la thèse 329
II.1. L'existence de trois problèmes-candidats 329
II.2. Une carte praxéologique de l'EMS VN en statistique 330
II.3. Une compréhension des rapports institutionnels aux objets-candidats 331
II.4. Une ingénierie didactique 332
III. Limites de l'étude et nouvelles directions de recherche possibles 333
III.1. Le choix de l'institution 333
III.2. Les choix de l'ingénierie 334
III.3. Quelques directions de recherche possibles 336
Bibliographie 337
I. Didactique des mathématiques - réflexions sur l'enseignement des mathématiques... 338
II. Didactique de la Statistique - réflexions sur l'enseignement des statistiques et des probabilités 339
III. Traités et manuels universitaires en statistique et probabilités 341
IV. Manuels et programmes scolaires du secondaire 342
V. Autres ouvrages 343
Résumé en vietnamien Tóm tắt luận án bằng tiếng Việt 345
I. Bối cảnh dạy toán ở trung học phổ thông Việt nam 348
I.1. Việc đào tạo giáo viên 348
I.2. Vấn đề liên môn 349
I.3. Môi trường công nghệ 350
II. Công cụ lý thuyết và phương pháp luận nghiên cứu 352
II.1. Lợi ích của một nghiên cứu so sánh thể chế Pháp - Việt 353
II.2. Tổ chức toán học - một công cụ cơ bản của phân tích thể chế 353
II.3. Mô hình khoa học luận tham chiếu 355
II.4. Hợp đồng thể chế và các hợp đồng sư phạm 355
II.5. Đồ án sư phạm 356
III. Cấu trúc luận án 357
IV. Những kết quả chính của luận án 358
IV.1. Ba bài toán - ứng viên 358
IV.2. Các tổ chức toán học trong dạy học toán ở trung học Việt nam 359
IV.3. Quan hệ thể chế đối với các đối tượng-ứng viên 360
IV.4. Đồ án sư phạm 361

V. Giới hạn và hướng nghiên cứu mới 363
V.1. Lựa chọn thể chế 363
V.2. Lựa chọn của đồ án sư phạm.....364
V.3. Một số hướng nghiên cứu mới 365
Annexes.....367

Introduction

Des questions d’enseignement aux questions de recherche

Il y a entre les probabilités et la statistique à la fois une solidarité lisible dans des concepts fondamentaux communs et une séparation que certains attribuent à des niveaux de mathématisation différents en renvoyant les probabilités aux mathématiques pures et la statistique aux mathématiques appliquées. Cette séparation a été vécue longtemps dans l'enseignement mathématique secondaire vietnamien sous une forme radicale : la statistique descriptive au collège, les probabilités dans l'enseignement supérieur.

La décision prise tout récemment par la noosphère vietnamienne d'introduire les probabilités dans l'enseignement mathématique secondaire, précédée d'un déplacement d'une partie de la statistique vers le lycée[1], fait surgir de multiples questions qui pour la plupart d'entre elles ont été longuement débattues en France mais qui deviennent actuelles pour notre pays et ses enseignants. Pourquoi cette introduction maintenant et pourquoi seulement en classe 11[2] ? Pourquoi circonscrire la statistique à la statistique descriptive et pourquoi la séparer des probabilités ? Quels apprentissages doit-on viser ? A quelles difficultés d'enseignement doit-on s'attendre ? Comment faire vivre la solidarité épistémologique entre statistique et probabilités alors qu'elles vont se côtoyer sans se confronter ?

Ces questions sont au cœur de l'étude didactique que nous avons décidé de conduire. La dernière d'entre elles nous est apparue comme primordiale car dans les liens que l'enseignement mathématique pourrait tisser entre ces deux domaines se jouent la nature et la portée des significations pour les différents concepts que les élèves et les enseignants construiront ensemble tant dans le domaine de la statistique que dans celui des probabilités.

En effet de nombreux travaux français de didactique des mathématiques[3], ceux de la Commission inter-IREM « Statistique et Probabilités »[4], ceux de la Commission de réflexion sur l'enseignement des mathématiques[5] qui avance l'expression globalisante de « formation à l'aléatoire », affirment la pertinence d'approches croisées entre statistique et probabilités dans l'enseignement des mathématiques. Beaucoup de ces travaux ont soit devancé soit accompagné les changements successifs dans les programmes français dont les derniers en date ont introduit la statistique inférentielle au lycée et les probabilités au collège[6]. En consultant ces travaux, nous avons pu prendre connaissance des débats qui ont secoué et qui secouent encore la noosphère française. Les questions épistémologiques sur la statistique et les probabilités y prennent une grande place. Nous les développerons dans la partie B de l'étude.

---

[1] Les programmes des classes 6 et 9 passèrent respectivement en classe 7 et 10 ; l'introduction des probabilités commença en 2006-2007.

[2] Classe de première en France.

[3] Les premiers en date sont certainement ceux de Brousseau [1974] exposés sous le titre : *Généralités sur l'enseignement des probabilités au niveau élémentaire*. On se réfèrera aussi aux travaux de Michel Henry cités dans la bibliographie.

[4] Citons les ouvrages suivants : Enseigner les probabilités au lycée (1997), Autour de la modélisation en probabilités (2001), Statistique au lycée (2007).

[5] Le rapport d'étape (2001) consacré aux statistiques-probabilités de cette commission dite commission Kahane, est consultable sur http://smf.emath.fr/Enseignement/CommissionKahane/

[6] Les programmes sont consultables sur le site http://eduscol.education.fr/

Cependant aucune des décisions prises quant aux programmes et aux méthodes d'enseignement ne s'explique uniquement par les choix épistémologiques retenus. Interviennent très fortement d'autres raisons dont beaucoup tiennent au contexte institutionnel dans lequel ses décisions vont se répercuter. Les débats ne manquent pas, notamment, de souligner l'importance, pour réussir les changements, de plusieurs facteurs que nous qualifions pour l'instant d'institutionnels : la formation des enseignants, l'interdisciplinarité et l'environnement technologique[7].

La décision noosphérienne vietnamienne d'introduire les probabilités rompt avec des décennies de renvoi des probabilités au-delà de l'enseignement secondaire obligatoire pour tous. La présence de la statistique dans les programmes de collège est elle-même récente, depuis 1990 seulement. Seules les classes bilingues français-vietnamien ont depuis longtemps à la fois des enseignements de statistique au collège et de probabilités au lycée[8]. On peut dès lors penser que la réussite de l'introduction des probabilités et plus encore de son articulation avec la statistique nécessitent de créer des conditions institutionnelles favorables. Certes la décision noosphérienne a été précédée d'une phase d'expérimentation avec des manuels spéciaux, qui a duré deux ans. On trouvera dans Vu Nhu (2005)[9] une description de cette expérimentation noosphérienne qui n'a concerné que certains lycées et une analyse didactique des programmes et des manuels complétée par une expérimentation en classe. A propos des raisons qui ont conduit à cette décision, nous remarquions :

> Aujourd'hui, les éducateurs vietnamiens souhaitent l'engagement de la statistique et de la probabilité dans l'enseignement au secondaire afin de suivre les tendances internationales et d'améliorer et renouveler les méthodes d'enseignement et l'éducation au Vietnam. […]
>
> En introduisant les probabilités dans l'enseignement secondaire beaucoup plus tard que les autres pays, les auteurs des manuels vietnamiens ont pu prendre connaissance des expériences de ces pays, à partir de documents, de manuels, de programmes à consulter et ont eu à faire des choix pour faire approcher par les élèves la notion de probabilité.
>
> (Vu Nhu [2005], p.1)

Cette décision de donner plus d'importance à l'aléatoire dans l'enseignement mathématique doit cependant composer avec des contraintes fortes concernant justement la formation des enseignants, l'interdisciplinarité et l'environnement technologique.

Il nous paraît donc utile, avant même d'exposer les questions et la méthodologie de recherche, de procéder à une brève description du contexte institutionnel vietnamien relativement à ces trois éléments : la formation des enseignants, l'interdisciplinarité et celle de l'environnement technologique, pour lesquels nous redonnons certaines réflexions de la commission Kahane comme base de comparaison avec le contexte français.

---

[7] Nous reprenons la terminologie des institutions d'enseignement.

[8] On pourra se référer au mémoire de Thac Si de TRAN Tuy An [2007] : Étude de pratiques d'enseignement de la notion de probabilité dans les classes bilingues et les classes ordinaires au Vietnam, Université Pédagogique d'Ho Chi Minh Ville

[9] VU NHU Thu Huong, 2005, *La notion de probabilité dans l'enseignement-apprentissage des mathématiques au lycée*, mémoire de Thac Si, Université de Pédagogie de Ho Chi Minh Ville

I. Le contexte de l'EMS [10] vietnamien

I.1. La formation des enseignants

Voici les recommandations de la commission Kahane pour la France de l'an 2000 :

> La formation des enseignants des lycées et collèges actuellement en poste est aujourd'hui un problème clé en ce qui concerne la formation citoyenne à l'aléatoire. Quelles sont les spécificités d'une telle formation ?
>
> - elle concerne pour les mathématiques environ 40 000 personnes (professeurs de collège et de lycée) ; étant donné le caractère nécessairement réparti d'un enseignement de la statistique, il concerne aussi les autres disciplines. Il n'y a aucune raison de cloisonner les formations ; il conviendrait donc de former ensemble les enseignants de mathématiques, de physique, de biologie et de sciences économiques et sociales, soit un public d'environ 60 000 personnes.
> - cette formation est à créer ; elle doit être construite en collaboration avec les enseignants, comme a été créé l'enseignement de la statistique aux médecins, en s'appuyant pour une très large part sur les médecins eux-mêmes.
> - une telle formation doit comporter des éléments théoriques, et elle doit aussi confronter les enseignants aux traitements de données réelles. Elle doit intégrer et utiliser la simulation aléatoire. Comme toutes les formations, sa qualité dépend de la pertinence des exemples traités, des questions posées, et bien que ce soit une formation professionnelle, les questions les plus motivantes ou éclairantes ne sont pas pour la plupart issues de la pratique professionnelle d'enseignant.
> - pour monter des dispositifs de formation, il convient de faire participer les professeurs de l'enseignement secondaire connaissant bien la statistique, des universitaires de différentes disciplines, mais aussi des professionnels d'institutions tels l'INSEE, l'INRA, l'INED, des enseignants des écoles d'ingénieurs, des médecins etc.

Au Vietnam, la formation initiale des professeurs de collège est séparée de celle des professeurs de lycée ; les durées et les contenus de la formation diffèrent. Celle des enseignants collèges est plus courte que celle des enseignants de lycée. Il y a tout lieu de penser que les rapports personnels à la statistique des enseignants de collège ignorent la statistique inférentielle.

Nous n'avons pas l'intention de faire du thème de la formation des enseignants l'objectif central de notre étude. Cependant, nous avons réalisé un questionnaire à l'adresse des élèves et des enseignants de la classe 10. Certaines des analyses conduites à propos de ce questionnaire (cf. chapitre C2) nous permettront de revenir sur ce thème de la formation des enseignants. Nous le ferons dans la conclusion.

[10] Enseignement mathématique secondaire.

I.2. L'interdisciplinarité

Le rapport de la commission Kahane développe longuement l'utilisation actuelle de la statistique et des probabilités dans plusieurs disciplines qui sont enseignées dans le secondaire, collège et lycée : la physique, la biologie, l'économie, la géographie. Pour plusieurs d'entre elles il insiste sur la concurrence entre des modélisations déterministes et des modélisations stochastiques :

> L'aléatoire se retrouve dans de nombreux champs disciplinaires et même dans la description de phénomènes pour lesquels on connaît des équations déterministes.

Il va jusqu'à avancer des recommandations pour l'enseignement d'autres disciplines que les mathématiques :

> Il appartient à tout enseignement de biologie, dans l'enseignement secondaire en particulier, de rendre familier l'aléatoire, notamment pour des phénomènes continus, de faire prendre conscience de la fluctuation d'échantillonnage, d'initier à la problématique de la prévision, de travailler sur la différence entre prédictibilité et causalité, etc.
>
> […] Dans un premier temps, au niveau de l'enseignement secondaire, l'enseignement de l'économie[11] pourrait prendre en charge la réflexion sur la nature de données qu'elle emploie, la difficulté à définir des indicateurs pertinents pour étudier la réalité économique ; par exemple une réflexion peut être menée en même temps en mathématiques et en économie sur l'élaboration des tables de mortalité et les calculs d'espérance de vie à tous les âges (pourquoi faut-il définir une cohorte fictive pour dresser les tables de mortalité, pourquoi l'espérance de vie d'un professeur est-elle supérieure à celle d'un individu pris au hasard dans la population ?). Ce travail sur la définition d'indices pertinents et leurs modes de calculs possibles est un élément fondamental de la formation en statistique.

Au Vietnam, plusieurs disciplines enseignées, autres que les mathématiques - la géographie, la biologie, la physique – pratiquent la statistique descriptive mais aucune n'aborde la modélisation stochastique nécessaire au passage de la description à la prévision. L'idée de prévision est certes présente dans les motivations à établir des statistiques. Mais cette idée ne rencontre pas la notion de probabilité, du moins explicitement.

Dans un lycée vietnamien, quelque soit la discipline, l'ordre et la durée des séances d'enseignement sont arrêtés par le programme, respectés par les manuels et ils s'imposent à tous les enseignants. Il n'y a donc pas place à un travail concerté sur les contenus mathématiques qui servent d'outils aux autres disciplines. Notre expérience d'enseignant au Vietnam confirme qu'il n'y a aucune concertation interdisciplinaire[12] sur les enseignements de la statistique.

Quant à la fluctuation d'échantillonnage évoquée par la commission, elle ne peut pas exister formellement au Vietnam puisque la notion d'échantillonnage est absente. Par contre, dans

---

[11] L'économie n'est pas enseignée au secondaire vietnamien.

[12] Nous reprenons ici l'adjectif usuel dans l'institution. Dans la conclusion, nbous parlerons plutôt de codisciplinarité en défendant l'idée d'une construction mixte de la mathématique et d'une autre discipline.

les disciplines comme la physique, la chimie et la biologie, on pratique en classe le recueil de mesures d'une même grandeur par plusieurs groupes d'élèves. Cette activité conduit à résoudre la question du choix de la valeur exacte à retenir à partir d'un ensemble de mesures. Aussi peut-on penser que la variabilité fait partie au moins implicitement des notions de statistique rencontrées dans l'enseignement d'autres disciplines que les mathématiques. Dans quelle mesure cette initiation à la variabilité peut-elle servir de référence pour les élèves dans une rencontre avec la fluctuation d'échantillonnage en mathématiques ?

I.3. L'environnement technologique

Partons une nouvelle fois du rapport de la commission Kahane :

> On pourrait dire […] que le matériau brut travaillé par la statistique est constitué de données expérimentales, les outils théoriques utilisés sont essentiellement la géométrie et l'algèbre linéaire pour la statistique exploratoire et les probabilités pour la statistique inférentielle et l'outil matériel est l'ordinateur. La mise en œuvre des méthodes de la statistique demande de gros moyens de calculs ; la statistique a pu se développer à grande échelle parce que les ordinateurs mettent à la disposition d'un large public des possibilités dépassant celles des centres de calculs d'il y a 30 ans.
>
> L'utilisation des outils logiciels dans les milieux professionnels ou de recherche est double :
>
> - mise en œuvre de calculs longs ou complexes (calculs de moyenne, de variance, inversions de matrices, calculs d'extrema sous contraintes, etc.), calculs des valeurs approchées (fonctions de répartition de lois de probabilités par exemple), calculs à distance finie, mise en œuvre de méthodes non paramétriques, de ré-échantillonnage.
> - estimation des résultats à partir de simulations : par exemple, le volume de certains ensembles, l'estimation de probabilités d'événements pour lesquels on n'a pas établi de formules qui en permettrait une approximation numérique directe.
>
> Dans une perspective de formation, l'usage de l'informatique est aussi double :
>
> - accès à des données diverses et de qualité et possibilité de mettre en œuvre sur ces données les traitements statistiques pertinents.
> - la simulation est un outil privilégié pour acquérir une expérience des phénomènes aléatoires, comprendre des théorèmes de convergence, voir où se situent les questions et appréhender la nature des preuves statistiques.

On constate au Vietnam un développement formidable des moyens informatiques dans la société que relaient des volontés politiques. Concernant la politique éducative, nous nous appuyons sur les références citées par Nguyen Chi Thanh [2009] [13] :

> Il faut rénover les méthodes d'enseignements. Eviter la transmission de connaissance dans un sens unique. Il faut stimuler la capacité de penser créativement. Il faut appliquer au fur et à mesure les nouvelles technologies dans l'enseignement.

[13] Nguyen C.T. 2009, L'usage du logiciel Cabri 3D dans l'enseignement de la géométrie dans l'espace au lycée du Viêt-nam.

(Décision du comité central du parti communiste en 1997)

> […] utiliser des nouvelles technologies dans l'enseignement afin de rendre les méthodes d'enseignement plus rénovatrices. Ces méthodes doivent stimuler le rôle actif de l'élève, la capacité de travailler en groupe […]

(Article 8.2 dans la loi éducative de l'année 2005)

> Il faut utiliser les nouvelles technologies dans l'enseignement à tous les niveaux comme un outil qui supporte efficacement la rénovation de la méthode d'enseignement de toutes les matières.

(Le décret ministériel 29/2001/BGD&DT)

Les enseignants sont encouragés mais encore faiblement épaulés. Leurs initiatives restent isolées. En particulier la simulation n'est pas mentionnée dans les programmes et sa pratique, même avec une calculatrice – par utilisation de la touche Random - est très marginale. Comment pourrait-elle venir dans l'enseignement de la statistique et des probabilités ?

L'exemple d'une réponse possible est donné dans le mémoire de Thac Si de Tran Tuy An [2007] [14] : dans l'une de ses expérimentations, un modèle multinomial est implémenté dans un tableur et l'élève demande au tableur les résultats de n répétitions de l'expérience aléatoire. L'objectif assigné à cette simulation est d'asseoir la méthode fréquentiste de calcul d'une probabilité dans le cas où la méthode laplacienne est remise en cause : répéter un grand nombre de fois une même expérience aléatoire, calculer la fréquence d'apparition d'un certain événement et proposer une valeur approchée « raisonnable » de la probabilité ; cette valeur approchée est appelée probabilité expérimentale.

La pratique de la simulation dépend de la place donnée à l'approche fréquentiste dans la définition et le calcul de la probabilité d'un événement. Or, nous aurons l'occasion de montrer que même si cette approche n'est pas ignorée des programmes et des manuels actuels vietnamiens, elle ne donne lieu à aucune activité des élèves en classe.

I.4. Intérêt d'une analyse institutionnelle comparative France-Vietnam

Ainsi le contexte institutionnel vietnamien impose des contraintes très fortes qui expliquent les résistances des enseignants lors de la mise en route des changements. Pour alléger temporairement ces contraintes, il a été décidé de ne pas mettre de probabilités dans les sujets d'examen ni dans les sujets de concours ce qui explique le choix de la classe 11 à la place de la classe 12 (classe terminale). De plus le programme retenu après son expérimentation abandonne une partie du programme expérimental qui est jugée trop difficile par les enseignants : ainsi a-t-on enlevé les probabilités conditionnelles dans les deux séries de manuels[15] ainsi que la notion de variable aléatoire dans la série élémentaire.

---

[14] Tran T. A. 2007, Etude des pratiques d'enseignement sur la notion de probabilité dans les classes bilingues et les classes ordinaires au Vietnam, Mémoire de Thac Si Université Pédagogique de Ho Chi Minh Ville.

[15] Il existe deux séries de manuels dans le lycée vietnamien : une série dite « avancé » et l'autre dite « élémentaire ».

L'étude que nous entendons conduire ne peut donc pas consister à valoriser les choix de l'enseignement français par rapport aux choix vietnamiens et à imaginer la possibilité de les reprendre au Vietnam en faisant fi du contexte institutionnel. Il s'agit pour nous de comprendre comment dans un certain contexte institutionnel, celui du Vietnam, certains choix sont pertinents et d'autres non quand on veut que les enseignements de la statistique et des probabilités se nourrissent l'un de l'autre pour construire des significations plus fortes aux différents concepts de l'un et l'autre des deux domaines. Mais nous n'abandonnons pas la comparaison avec la France car une telle comparaison peut apporter, selon Bessot & Comiti [2008], trois points d'appuis solides dans une étude didactique :

- une dénaturalisation du regard sur le fonctionnement scolaire d'une institution didactique ;
- la prise en compte de niveaux de détermination plus élevés que celui du domaine ;
- la constitution d'un répertoire de praxéologies à enseigner et l'évaluation des praxéologies mathématiques et didactiques.

(Bessot & Comiti [2008], p.191)

La méthodologie de comparaison sera mise en œuvre dans la partie A afin de dégager les praxéologies à enseigner, des objets de la Théorie Anthropologique du Didactique (TAD) que nous présentons tout de suite.

II. Outils théoriques et méthodologiques

Nous reformulons l'ensemble des questions initiales d'enseignement en trois questions de recherche que nous travaillerons avec les outils de la Théorie Anthropologique du Didactique (TAD) et de la théorie des Situations Didactiques (TSD).

QR1. Autour de quels problèmes et avec quels objets mathématiques – que nous appellerons problèmes et objets candidats – construire des liens statistique-probabilités dans l'enseignement mathématique secondaire ?

Pour répondre à cette question nous nous appuierons sur le concept de praxéologie mathématique tel que défini dans la TAD et dégagerons, à partir de l'examen de manuels universitaires français, des praxéologies de référence en nous restreignant aux praxéologies les plus significatives de la solidarité épistémologique entre statistique et probabilités. C'est l'objet de la partie B.

QR2. Quelles sont les conditions de vie actuelles de ces problèmes-candidats et de ces objets-candidats dans EMS Vietnam ?

Toujours grâce aux concepts de la TAD, il s'agit de procéder à ce que nous appelons une analyse institutionnelle et, comme nous l'avons annoncé, en nous aidant d'une comparaison avec la France. Nous dégagerons des praxéologies à enseigner et modéliserons l'écart aux praxéologies de référence dans des contrats institutionnels relativement aux objets candidats. C'est l'objet du chapitre C1.

Nous pourrons alors former des hypothèses sur les praxéologies enseignées et les contrats didactiques dans l'institution vietnamienne et procéder à une mise à l'épreuve de ces

hypothèses grâce à un questionnaire aux élèves et aux enseignants. C'est l'objet du chapitre C2.

QR3. Comment faire vivre dans les conditions actuelles de l'enseignement vietnamien des liens statistique-probabilités autour de quelques-uns des problèmes et des objets-candidats ?

A partir d'hypothèses sur la possibilité de faire vivre dans les conditions actuelles de l'enseignement vietnamien des liens statistique-probabilités autour de quelques-uns des problèmes et des objets-candidats, nous concevrons une ingénierie didactique que nous expérimenterons dans une classe vietnamienne. La confrontation des analyses a priori et a posteriori des situations didactiques de cette ingénierie servira de base aux conclusions de l'étude. C'est l'objet de la partie D.

Nous présentons maintenant brièvement les outils théoriques bien connus de l'anthropologie du didactique, que nous utilisons dans l'étude des questions QR1 et QR2, celui de praxéologie mathématique (Chevallard [1992]), celle de modèle épistémologique de référence (Bosch & alii, 2000) et celle de contrat institutionnel.

II.1. La praxéologie mathématique comme objet élémentaire d'une analyse institutionnelle

Nous laissons à Le Thai Bao Thien Trung [2007] le soin de redéfinir les objets de base de la TAD qui nous serviront pour ce que nous appelons une analyse institutionnelle :

> Pour analyser le rapport institutionnel à un objet de savoir, Chevallard (1998) introduit la notion de praxéologie mathématique ou d'organisation mathématique (OM) autour d'un savoir. Une OM est constituée du quadruplet : types de tâches T présents dans une institution donnée, techniques τ permettant d'accomplir les tâches de T, technologies θ justifiant les techniques τ, théories Θ justifiant les technologies θ.
>
> Chevallard (1998) distingue différents niveaux pour les OM.
>
> - $[T/\tau/\theta/\Theta]$: OM ponctuelle relative à un unique type de tâches T.
>
> Le tout, noté $[T/\tau/\theta/\Theta]$, constitue une praxéologie ponctuelle. Ce qualificatif signifiant qu'il s'agit d'une praxéologie relative à un unique type de tâches, T. (Op. cité, p. 95)
>
> - $[Ti/\tau i/\theta/\Theta]$: OM locale centrée sur une unique technologie θ.
>
> - $[Tij/\tau ij/\theta j/\Theta]$: OM régionale formée autour d'une unique théorie Θ.
>
> - $[Tijk/\tau ijk/\theta jk/\Theta k]$: OM globale, agrégation de plusieurs OM régionales correspondant à plusieurs théories Θk.
>
> (Chevallard, [1998])

Le concept de praxéologie est l'un des outils de la TAD dans une approche du didactique dont Bosch et Chevallard [2001] redisent ici la finalité :

> L'approche anthropologique répond à la question de la modélisation des pratiques sociales, de leurs composants, de leur évolution, de leurs produits. Pour ce qui est des mathématiques, considérée comme une activité humaine structurée en organisations

> praxéologiques, on pourra dire qu'elles naissent de la problématisation de certains types de tâches, dès lors regardées comme des types de problèmes dont l'étude donne lieu à la construction d'organisations praxéologiques locales. L'articulation de certaines de ces praxéologies autour d'une technologie commune permet de former des organisations régionales qui, à leur tour, s'articulent en organisations plus larges jusqu'à constituer globalement "le savoir mathématique". La description de ces organisations et l'étude de leur écologie institutionnelle sont au cœur du programme d'étude de la didactique des mathématiques.
>
> (Bosch & Chevallard [1999], p. 8)

Procéder à une analyse institutionnelle consiste alors en premier lieu à dégager un ensemble de praxéologies mathématiques vivantes dans l'institution pour en faire un modèle structurant l'ensemble des pratiques mathématiques systématiques partagées dans l'institution. La méthode est ainsi décrite par Chevallard [1998] :

> On ne rencontre en fait que rarement des praxéologies ponctuelles. Généralement, en une institution *I* donnée une théorie Θ répond de *plusieurs technologies* θ_j, dont chacune à son tour justifie et rend intelligibles *plusieurs techniques* τ_{ij} correspondant à *autant de types de tâches* T_{ij}. Les organisations ponctuelles vont ainsi s'agréger, d'abord en organisations *locales*, $[T_i/\tau_i/\theta/\Theta]$, centrées sur une technologie θ déterminée, ensuite en organisation *régionales*, $[T_{ij}/\tau_{ij}/\theta_j/\Theta]$, formées autour d'une théorie Θ. (Au-delà, on nommera organisation globale le complexe praxéologique $[T_{ijk}/\tau_{ijk}/\theta_{jk}/\Theta_k]$ obtenu, dans une institution donnée, par l'agrégation de plusieurs organisations régionales correspondant à plusieurs théories Θ_k). […].
>
> (Chevallard, [1998], p. 95)

La fabrication des praxéologies nécessite une plongée dans les programmes et les manuels actuels mais aussi dans les programmes et les manuels passés ou dans ceux d'autres institutions pour définir soigneusement les types de tâche présents et absents, les techniques et les éléments technologiques énoncés souvent sous forme de règles ou de théorèmes. Pour notre part, nous avons presque systématiquement mis en regard les types de tâche « vietnamiens » et les types de tâche « français ».

II.2. Le modèle épistémologique de référence

Référons-nous au schéma classique de la transposition didactique avec ses trois chaînons principaux que nous redonnons ci-dessous :

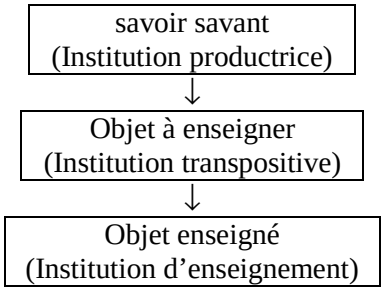


Schéma 1. Schéma de la transposition didactique

Nous comprenons que la phase de l'analyse institutionnelle que nous venons de présenter nous fournit ce que nous appelons des *praxéologies à enseigner* puisque programmes et manuels sont des produits d'une institution transpositive communément appelée *la noosphère*. Ce modèle de praxéologies à enseigner s'avère encore très descriptif, pas assez explicatif ni d'un point de vue épistémologique ni d'un point de vue didactique. Il doit être confronté d'une part à une vision épistémologique des objets de savoir que nous irons chercher du côté de l'institution « savante », d'autre part à une vision didactique qui ne peut ignorer les contraintes de l'institution d'enseignement elle-même et que nous irons chercher dans la classe.

La confrontation épistémologique débouche sur la fabrication de *praxéologies de référence* :

> Etant donné un processus didactique sur un thème mathématique déterminé, la première étape de notre technique d'analyse suppose l'adoption de ce que nous appelons un modèle épistémologique de référence sur le contenu mathématique en jeu. Dans le cadre de la TAD, ce modèle se formule en termes d'organisations ou praxéologies mathématiques.
>
> (Bolea, Bosch, Garcia, Gascon, Ruiz, Sierra, 2005, Colloque Hommage à Guy Brousseau).

Fabriqué empiriquement à partir des praxéologies à enseigner et des textes « savants » comme des manuels ou des traités universitaires[16], ce modèle doit permettre de rendre compte des rapports institutionnels aux objets de savoir impliqués dans les praxéologies à enseigner. Il doit aussi permettre de comprendre les contraintes qui pèsent sur l'institution d'enseignement et le processus de contractualisation entre les sujets-acteurs de l'institution relativement aux objets de savoir des domaines « statistique » et « probabilités ».

Cette contractualisation peut être décrite à deux niveaux : le contrat institutionnel et le contrat didactique.

II.3. Contrat institutionnel et contrats didactiques

Le contrat institutionnel relatif à un objet de savoir O – disons ici la statistique - peut être défini comme le système des rapports institutionnels à O et à o_i, o_i institutionnellement pertinents pour O. Un tel système que nous appelons ici *contrat institutionnel relatif à la Statistique s*'entend comme la mise en jeu didactique, de façon stable et prégnante dans l'institution, d'objets mathématiques qui appartiennent à la Statistique, ou qui entretiennent des liens étroits avec elle. Dans Birebent (2008), nous trouvons cette présentation plus complète du contrat institutionnel :

> Le contrat institutionnel émerge de la distance entre ce système des praxéologies de référence et celui des praxéologies à enseigner qu'affichent programmes et instructions et que reconstituent d'une certaine manière manuels et sujets d'examen ; il pilote dans l'institution le processus didactique de reconstitution des praxéologies de référence dans les praxéologies à enseigner et conditionne les praxéologies effectivement enseignées. A la fois processus et état, le contrat institutionnel :

[16] Nous avons choisi quelques manuels universitaires français.

- énonce la manière de se comporter par rapport à O, dans I, pour les élèves et pour l'enseignant (les attentes de I relativement à O) ;
- fabrique un clivage topogénique entre l'élève et l'enseignant relativement à O dans I ;
- assure, au moyen de ruptures / transitions ou de reprises / oublis, l'avancée du temps didactique pour O (chronogenèse et changements d'états du contrat) ;
- détermine un champ de contraintes et de conditions pour les praxéologies enseignées ;
- constitue une matrice des contrats didactiques pour les objets o_i institutionnellement associés au calcul numérique.

(Birebent [2007], p. 316)

Mais l'objet de savoir O qu'est ici la statistique est un « gros objet ». Le contrat institutionnel offre alors une lecture écologique « globale » des relations didactiques entre les différents sujets-acteurs de l'institution, relativement à O et tous les o_i, qui permet de décrire l'institution dans sa dynamique temporelle et spatiale. Cette lecture doit être affinée par des analyses plus « locales ». Par exemple dans le découpage praxéologique, on ne peut pas arrêter la définition de la tâche à la consigne verbale. Ce sont les contextes d'enseignement et donc les contrats didactiques locaux qui rendent la tâche repérable au-delà de la forme langagière de la consigne.

Le concept de contrat didactique[17] relève alors plutôt d'une modélisation économique des relations didactiques, relativement à un objet mathématique, entre des élèves et un enseignant, dans leur courte histoire commune.

Dans cette modélisation de la contractualisation des objets de savoir entre les différents sujets-acteurs de l'institution, le contrat didactique est l'instanciation spatiale et temporelle du contrat institutionnel ; c'est la forme que prend le contrat institutionnel dans la classe. Cette vision fait écho à la présentation formulée par Chevallard (1989) : « le contrat didactique est l'état de la chronogenèse et de la topogenèse à un instant donné ».

Pratiquement dans notre étude, nous proposerons au chapitre C1 une description du contrat institutionnel et présenterons dans le chapitre C2 certaines hypothèses sur des contrats didactiques immédiatement mises à l'épreuve dans une classe 10 vietnamienne.

II.4. Ingénierie didactique

Pour répondre à QR3, nous suivons une méthodologie propre à la didactique des mathématiques qui est de concevoir, expérimenter et analyser une ingénierie didactique.

> La notion d'ingénierie didactique a émergé en didactique des mathématiques au début des années 1980. Il s'agissait d'étiqueter par ce terme une forme du travail didactique : celle comparable au travail de l'ingénieur qui, pour réaliser un projet précis, s'appuie

[17] Pour une présentation de ce concept plus ancien que celui de contrat institutionnel, se référer à Bessot A. (1995), Le contrat didactique : un outil central pour l'analyse des situations d'enseignement in Comiti C. Ngo Anh T., Bessot A. et alii (eds), *Didactique des disciplines scientifiques et formation des enseignants*, Ho Chi Minh Ville, Université Pédagogique.

> sur les connaissances scientifiques de son domaine, accepte de se soumettre à un contrôle de type scientifique mais, dans le même temps, se trouve obligé de travailler sur des objets beaucoup plus complexes que les objets épurés de la science et donc de s'attaquer pratiquement, avec tous les moyens dont il dispose, à des problèmes que la science ne veut ou ne peut encore prendre en charge.

(Artigue [1990], p.283)

> L'ingénierie didactique, vue comme méthodologie de recherche, se caractérise en premier lieu par un schéma expérimental basé sur des « réalisations didactiques » en classe, c'est-à-dire sur la conception, la réalisation, l'observation et l'analyse de séquences d'enseignement. […]

(Ibid., p.286)

Reprenant les termes de Le Thai Bao (2007), nous dirons que :

> L'ingénierie didactique est une situation d'enseignement élaborée par le chercheur, une forme du travail didactique semblable à celui de l'ingénieur : elle s'appuie sur les connaissances scientifiques de son domaine pour travailler sur des objets beaucoup plus complexes que les objets épurés de la science. Elle permet
>
> - une action sur le système d'enseignement, basée sur des études didactiques préalables.
>
> - une mise à l'épreuve d'hypothèses ou de questions élaborées par la recherche par leur réalisation dans un système d'enseignement.
>
> Elle a donc une double fonction : méthodologique, et productrice de questions didactiques nouvelles.

(Le Thai Bao [2007], p.286)

Dans notre cas, prenant en compte les résultats des analyses épistémologique et institutionnelle – respectivement dans les parties B, A-C – nous avons conçu l'ingénierie sur un problème d'incertitude aléatoire à partir d'une hypothèse qui postule la possibilité de construire la résolution d'un tel problème dans suite de situations didactiques favorisant l'enrichissement épistémologique de liens entre la statistique et probabilités. Cette ingénierie est découpée en trois séances correspondant à autant de valeurs d'une *variable méso didactique*[18]. Cette ingénierie a été expérimentée en classe 11 d'un lycée vietnamien.

III. Organisation de la thèse

Après cette introduction, la thèse s'organise en quatre parties notées A, B, C et D ainsi constituées :

Partie A : analyse institutionnelle comparative France-Vietnam

Repérage des praxéologies à enseigner en France et au Vietnam.

[18] Le Thai Bao [2007] reprenant les termes d'un cours donné en DEA par Bessot & Birebent (2003).

Partie B : analyse épistémologique

Une modélisation des liens statistique-probabilités dans deux praxéologies de référence :

- la comparaison des distributions statistiques

- l'échantillonnage

Deux hypothèses de recherche :

- l'existence d'un point de vue fonctionnel sur les graphiques statistiques est une condition favorable à la construction de liens statistique-probabilités au sein d'une praxéologie de comparaison des distributions statistiques.

- le problème d'obtenir un échantillon représentatif d'une population mère inaccessible oblige à rencontrer et à construire à partir de statistiques des éléments de mathématisation du hasard au sein d'une praxéologie d'échantillonnage.

Partie C : retour sur l'analyse institutionnelle pour le Vietnam

Chapitre C1 : contrat institutionnel à la statistique

Incomplétude des praxéologies à enseigner et nouvelles questions de recherche

Chapitre C2 : mises à l'épreuve d'hypothèses sur les contrats didactiques par un questionnaire aux élèves et aux enseignants

I. Exercice 1, II. Exercice 2, III. Exercice 3, IV. Exercice 4

Partie D : ingénierie didactique

Chapitre D1 : introduction à l'ingénierie didactique

I. Pourquoi une ingénierie didactique ?

II. Les ingénieries de Brousseau et de Briand

III. Rapports institutionnels aux notions d'expérience et d'incertitude aléatoires

Chapitre D2 : expérimentation au Vietnam et analyses des séances

I. Séance 1, II. Séance 2, III. Séance 3

Partie A

Analyse institutionnelle comparative France-Vietnam

L'objet de cette partie est de repérer les praxéologies statistiques à enseigner dans l'EMS vietnamien. Compte tenu de l'importance que nous voulons donner aux types de tâche qui s'inscriraient dans la construction de liens entre statistique et probabilités, la méthode a consisté en fait à mener d'abord une étude praxéologique sur la France en considérant deux périodes, celle qui a précédé la réforme de l'an 2000 et celle qui l'a suivie. En effet ce que nous appelons ici la réforme de l'an 2000 est l'introduction d'éléments de statistique inférentielle au lycée dès la classe de seconde. Notre hypothèse de travail est que la structure praxéologique de l'EMS vietnamien actuel dans le domaine de la statistique se rapproche de celle de la France d'avant la réforme.

I. Analyse institutionnelle française

Nous analysons les programmes et les manuels de la classe de sixième à la classe Terminale (série S) en deux temps : avant et après la réforme. Les changements de programmes du lycée de l'an 2000 ont été précédés en collège entre les années 1996 et 1999 par de nouveaux programmes. Voici la structure de l'analyse institutionnelle française que nous présentons ci-après, les cases grisées renvoyant pour partie à l'annexe A[19] :

Avant la réforme	Collège	Programmes
		Manuels
	Lycée	Programmes
		Manuels
Après la réforme	Collège	Programmes
		Manuels
	Lycée	Programmes
		Manuels

Tableau 1. Présentation de la structure d'analyse institutionnelle française

I.1. Avant la réforme (avant l'an 2000)

Les programmes de collège

Nous entrons dans les programmes du collège pendant la période 1986-1999. A partir de l'année scolaire 1986, la partie de *Représentation et Organisation de données* prend place dans le programme mathématique de la sixième à la troisième. Dans le tableau qui suit nous distinguons deux sous-périodes, 1986-1989 et 1996-1999, correspondant à deux programmes distincts :

[19] L'annexe de la partie A.

		Classe 6è	Classe 5è	Classe 4è	Classe 3è
1986-1989[20]	Programme	1986 Lecture, interprétation et réalisation de tableaux et de graphiques.	1987	1988 Fréquences, expression en pourcentage. Effectifs cumulés, fréquences cumulées.	1989 Moyenne, moyennes pondérées. Médiane.
	Compétences exigibles	Appliquer un tableau de pourcentage. Effectuer, éventuellement avec une calculatrice, des calculs sur les mesures de grandeurs figurant au programme.	Lire des données statistiques présentées sous la forme de tableaux ou de représentations graphiques Traduire des données statistiques sous la forme d'un diagramme en bâton.	Savoir lire des données statistiques présentées sous la forme de tableaux ou de diagramme d'effectifs ou de fréquences. A partir de données statistiques, présenter les effectifs ou les fréquences dans des tableaux et tracer les diagrammes correspondants.	Savoir lire et exploiter des données statistiques mises sous forme de tableaux ou de diagrammes d'effectifs ou de fréquences, savoir calculer une moyenne. A partir de données statistiques, calculer les effectifs ou les fréquences, les présenter dans les tableaux et tracer les diagrammes correspondants.
1996-1999[21]	Programme	1996 Exemple conduisant à lire, à établir des tableaux, des graphiques.	1997 Classes, effectifs d'une distribution statistique. Fréquences. Diagrammes à barres, diagrammes circulaires.	1998 Effectifs cumulés. Fréquences cumulées. Moyennes. Initiation à l'usage de tableurs-grapheurs.	1999 <u>Approche de la comparaison de séries statistiques</u>[22].

[20] D'après le *Tableau synoptique pour le collège* fourni par M.E.N. Mathématiques 6è, 5è, 4è, 3è (1989), Horaires, Objectifs, Programmes, Instructions.

[21] Extrait du *Tableau synoptique pour le collège*, B.O. Hors-série n°10 Octobre1998.

[22] Soulignés par nous.

	Compétences exigibles		Lire et interpréter un tableau, un diagramme à barres, un diagramme circulaire ou semi-circulaire Regrouper des données statistiques en classes, calculer des effectifs. Présenter une série statistique sous la forme d'un tableau, la présenter sous la forme d'un diagramme à barres ou d'un graphique. Calculer des fréquences.	Calculer des effectifs cumulés, des fréquences cumulées. Calculer la moyenne d'une série statistique. Calculer une valeur approchée de la moyenne d'une série statistique regroupée en classes d'intervalles.	Une série statistique étant donnée (sous forme de liste ou de tableau, ou par une représentation graphique) proposer une valeur médiane de cette série et en donner la signification. Une série statistique étant donnée, déterminer son étendue ou celle d'une partie donnée de cette série.

Tableau 2. Tableau de programmes et compétences exigibles du collège de 1986 à 1999

On constate de la première à la seconde sous-période un glissement en amont de plusieurs notions ainsi qu'un renforcement avec l'arrivée du regroupement des données en classes (dès la 5ième) et celle d'un indicateur de dispersion en classe de troisième :

	5è	4è	3è
1986-1989		Effectif Fréquence Effectif cumulé Fréquence cumulée	Moyenne Médiane
1996-1999	Effectif, Fréquence Regroupement en classes	Moyenne Effectif cumulé Fréquence cumulée	Médiane Etendue

Tableau 3. Les notions de statistique dans les programmes du collège avant la réforme

Nous voyons dans ces changements une volonté noosphérienne de renforcer la place de la Statistique dans les mathématiques, tout en la cantonnant à la Statistique descriptive.

Pour nous faire comprendre, voici comment un manuel universitaire définit la Statistique descriptive en l'opposant à la Statistique inférentielle :

> On distingue deux grands types d'études statistiques. Il y a d'abord les études purement descriptives dans lesquelles on déduit certaines caractéristiques d'un ensemble de données. Ces données peuvent avoir été obtenues à partir d'observations partielles sur une population sans qu'on ait eu aucun contrôle ni aucun choix quant à la façon de procéder aux observations. Il peut aussi s'agir de résultats d'observations systématiques sur une population entière. Quoi qu'il en soit, on peut toujours effectuer une étude sommaire de résultats d'observations.
>
> La partie de la statistique qui s'occupe exclusivement de la description de données obtenues à partir d'observations quelconques est appelée **statistique descriptive** ou

> **statistique déductive**. La statistique descriptive s'intéresse par exemple aux pourcentages de données qui prennent certaines valeurs, aux données les plus fréquentes, aux variations des données les unes par rapport aux autres, aux comparaisons avec d'autres données, etc., en somme au traitement de données. La plupart du temps, il s'agit de classer des données selon des critères déterminés et de faire des calculs pour mesurer, décomposer et comparer les variations.
>
> Dans certains cas, les observations, quoique partielles, sont faites suivant des procédures bien définies et contrôlées : ou bien les individus sont sciemment pris au hasard, ou bien ils sont observés de façon indépendante et dans les même conditions, le choix au hasard étant laissé à la nature ou aux circonstances, ou bien ils sont observés d'après des critères spécifiques qui permettent de les catégoriser. Le but des procédures est généralement d'assurer que l'on obtient des échantillons représentatifs de populations bien définies, ce qui permet d'induire une information sur les populations à partir d'échantillon. La partie de statistique qui s'occupe de tirer des conclusions générales à partir d'enquêtes ou d'expérience statistiques contrôlées est appelée **statistique inductive** ou **inférence statistique**. C'est l'inférence statistique qui permet les généralisations et les prévisions, d'où son grand intérêt théorique et pratique.
>
> (Lessard S. & Monga E., 1993, MASSON éditeur, p. 16)

Le traitement des données observées tel que le définit ce manuel universitaire se retrouve dans les compétences exigibles des programmes scolaires. Nous traduisons et rangeons ces compétences dans quatre groupes de types de tâche et cette catégorisation nous sert de pré-découpage praxéologique :

Groupes	Types de tâche
Groupe 1 : regrouper et représenter	- Regrouper les données brutes - Calculer effectifs, pourcentages, fréquences → distributions des effectifs et des fréquences - Représenter des données statistiques fournies : - par tableaux - par graphiques
Groupe 2 : résumer numériquement	Calculer les indicateurs : - de position - de dispersion
Groupe 3 : comparer des distributions	Comparer : - les tableaux - les graphiques - les résumés numériques
Groupe 4 : exploiter les données traitées	- Interpréter - Commenter

Tableau 4. Les quatre groupes de types de tâche

Les programmes de lycée

Nous donnons le programme qui précède la réforme de 2000 :

	2de	1erS	Ter C, E
Connaissances et savoir-faire	1990 Série statistique à une variable : - Répartition d'une population en classes - Effectifs, fréquences. Séries statistiques à une variable quantitative : - Effectifs cumulés, fréquences cumulées. - Caractéristiques de position et de dispersion : moyenne, écart-type.	1991 Evénements, événements élémentaires ; la probabilité d'un événement est définie par addition de probabilités des événements élémentaires. Evénements disjoints (ou incompatibles), événement contraire, réunion et intersection de deux événements. Cas où les événements élémentaires sont équiprobables.	1992 Probabilité a) Variable aléatoire (réelle) prenant un nombre fini de valeurs et loi de probabilité associée ; fonction de répartition, espérance mathématique, variance, écart-type. b) Probabilité conditionnelle d'un événement par rapport à un événement de probabilité non nulle ; relation $p(A\cap B)=p(A\mid B)p(B)$ Indépendance de deux événements Formule des probabilités totales : étant donné des événements B_1, B_2,…B_n constituant une partition de E, pour tout événement A, $p(A)= p(A\cap B_1)+p(A\cap B_2)+\ldots p(A\cap B_n)$ et pour tout k, $p(A\cap B_k)=p(A\mid_{Bk})p(B_k)$
Champ de problèmes et des techniques		Exemples simples d'emplois de partitions et représentations (arbres, tableaux…) pour organiser et dénombrer des données relatives à la description d'une expérience aléatoire. Exemples simples d'étude de situations de probabilités issues d'expériences aléatoires (modèle d'urne, jeux...)	Exemples simples d'emploi de partitions et représentations (arbres, tableaux…) pour organiser et dénombrer des données. Exemples d'étude de situations de probabilités issues d'expériences aléatoires (modèle d'urne, jeux, schéma de Bernoulli...)

Tableau 5. Tableau des programmes du lycée de 1990 à 1999

- La classe de seconde reprend les notions de collège en leur adjoignant seulement un nouvel indicateur de dispersion, l'écart-type. Il y a là comme un frein au développement de la Statistique dans l'enseignement secondaire. Après la classe de seconde la Statistique disparaît laissant place aux probabilités[23].

[23] Nous avons choisi de ne regarder que la filière S (anciennement C-D-E). On remarquera que l'enseignement des probabilités venait après la séparation des filières et ne touchait donc pas tous les élèves de la même manière, au contraire des statistiques.

- La notion d'échantillonnage est absente : les données statistiques ne sont jamais considérées comme celles d'un échantillon d'une population mère mais toujours comme celles d'une population toute entière, fermée. L'idée d'élargir les résultats obtenus sur une sous-population à la population entière étudiée et l'idée de comparer à d'autres populations ne sont pas exploitées.

- Dans ses commentaires adressés aux enseignants, le programme stipule : *« Pour introduire la notion de probabilité, on s'appuie sur l'étude de séries statistiques obtenues par répétition d'une expérience aléatoire, en soulignant les propriétés des fréquences et la relative stabilité de la fréquence d'un événement donné lorsque cette expérience est répétée un grand nombre de fois »* Ainsi, même en l'absence d'échantillonnage, la loi des grand nombre est un élément théorique de référence. La construction d'un lien entre statistique et probabilités est ouverte.

- Toujours dans les commentaires adressés aux enseignants, on trouve cette phrase : « *La description d'expériences aléatoires amène aussi à organiser des données : on se limitera à quelques exemples permettant de mettre en valeur les idées […] ».* Le problème de l'organisation des données n'était pas visible dans les programmes de statistique en collège et en seconde. Ce problème vient en classe de première sous la forme de la description d'expériences aléatoires et de la modélisation de certaines d'entre elles.

Les manuels

Nous avons défini et recensé les types de tâche présents et pour certains absents de plusieurs manuels de collège et de lycée. Ce travail est consultable dans l'Annexe A (chapitre « Institution française avant la réforme »). Nous nous contentons ici de présenter certains résultats de ce travail à partir du tableau 9 :

- Commençons par remarquer l'absence de question à l'origine des types de tâche présents. Nous rejoignons ici les analyses de Wozniak [2005] :

> On observera qu'aucun des types de tâches envisagés n'est motivé par une question à laquelle on se soucierait d'apporter réponse, en dépit même d'une définition préalable de la statistique énonçant que celle-ci vise à « obtenir des rapports numériques sensiblement indépendants des anomalies du hasard et qui dénotent l'existence des causes régulières ».
>
> (Wozniak [2005], p. 54)

- Les types de tâche les plus fréquents concernent les représentations graphiques, les calculs d'effectifs et de fréquences pour obtenir les tableaux de distribution ainsi que les calculs des indicateurs de position. Ce constat avait conduit Wozniak [2005] à pointer dans l'enseignement le phénomène qu'elle qualifiait de « réduction arithmétique de la statistique » et dont elle cherchait l'explication dans des raisons d'économie didactique :

> Aucune question préalable n'engendre l'étude ; il n'y a pas de recherche des données utiles, ni de discussion sur les informations à extraire de ces données. Le travail « statistique » devient ici une affaire de petite arithmétique. Cette réduction arithmétique de la statistique est plus visible encore dans les cas – nombreux – où, selon une pente dont nous avons déjà observé l'effet, les auteurs se réfèrent à des situations statistiques

> complexes, pour lesquelles les outils appropriés sont hors de portée, et qui ne demandent alors aux élèves qu'un travail très guidé de calcul arithmétique

(Wozniak [2005], p. 319)

> On retrouve ici des considérations rencontrées déjà chez les professeurs dont nous avons parlé dans notre chapitre 3 : la statistique apparaît comme un domaine mathématiquement plus « facile ». Cette constance dans le jugement peut surprendre le statisticien s'il ne sait pas que, en réalité, comme on le verra plus loin, nombre de professeurs enseignent moins la statistique que la réduction arithmétique qu'ils en font, laquelle est alors, de fait, sans grande difficulté pour les élèves.

(Wozniak [2005], p. 370)

- Même après le changement de programme en collège qui introduit l'étendue d'une distribution, la notion de dispersion ne prend sa place qu'à partir de la classe de seconde, limitant ainsi la comparaison entre distributions en collège. Notons le type de tâche T49 : calculer le pourcentage de données situées dans l'intervalle [x-σ ; x+σ] / [x-2σ ; x+2σ] / [x-3σ ; x+3σ]) qui ne trouve sa raison d'être qu'en comparaison à une distribution normale absente du programme mais qui donne une occasion supplémentaire de faire des calculs.

- La comparaison des distributions statistiques n'occupe qu'une place très modeste : la comparaison des moyennes est présente en classe 5è (tâche T62) et on trouve des demandes plus vagues de comparaison de deux séries statistiques (tâche T64 et T65) dans les classes de 3ième et 2nde.

- L'interprétation statistique est réduite à la formulation de la signification contextuelle d'un effectif, d'une fréquence ou d'un indicateur comme la moyenne, la médiane ou le mode dans les classes de 4ième et 3ième. Le groupe 4 englobant ce type de tâches est très peu représenté et même absent de la classe de 2nde.

Conclusion

Avant la réforme, statistique et probabilités vivent séparément ; au collège de la statistique sans probabilités ; au lycée après la classe de seconde, des probabilités sans statistique.

Nous dégageons de cette étude le découpage praxéologique suivant :

- un ensemble de praxéologies locales de description graphique d'une série statistique s'appuyant sur des praxéologies ponctuelles de regroupement d'individus (selon les valeurs ou les classes de valeurs) et de fabrication de tableaux d'effectifs ou de fréquences ;

- un ensemble de praxéologies locales de description numérique d'une série statistique s'appuyant sur les mêmes praxéologies ponctuelles que les praxéologies locales de description graphique ;

- un ensemble de praxéologies locales de comparaison graphique de séries statistiques ;

- un ensemble de praxéologies locales de comparaison numérique de séries statistiques.

Par fusion et imbrication des praxéologies locales, nous trouvons deux praxéologies régionales :

- une praxéologie régionale de description d'une série statistique imbriquant les praxéologies locales de description graphique et numériques ; cette praxéologie développe les tâches et les techniques pour simplifier, résumer et synthétiser une série statistique ; elle vise la comparaison de chaque individu à l'ensemble des individus ; les éléments technologiques sont les tableaux, les graphiques et les vecteurs d'indicateurs numériques.

- une praxéologie régionale de comparaison de séries statistiques imbriquant les praxéologies locales de comparaison graphiques ou numériques ; cette praxéologie vise la comparaison entre sous-populations d'une même population ou entre une sous-population et sa population-mère.

La technologie de cette dernière est basée sur des règles implicites d'égalité des statistiques dont voici les deux qui sont présentes dans l'institution :

- deux statistiques sont égales si leurs graphiques de fréquences sont superposables

- deux statistiques sont égales si leurs vecteurs d'indicateurs numériques sont égaux.

I.2. Après la réforme (après l'an 2000)

Le collège a connu de nouveaux programmes à partir de 2005 qui ont apporté très peu de modifications dans le domaine de la statistique. Par contre à partir de l'année 2009, on introduira les probabilités en classe de 3ième. Seuls les programmes de la période 2005-2009 sont considérés ici.

Les programmes de collège

	Classe 6è	Classe 5è	Classe 4è	Classe 3è
Programme	Organisation et représentation de données.	Lecture. Interprétation, représentations graphiques de séries statistiques. Classes, effectifs. Fréquences.	Effectifs cumulés, Fréquences cumulées. Moyennes pondérées. Initiation à l'utilisation de tableurs-grapheurs.	Caractéristiques de position d'une série statistique. Approche de caractéristiques de dispersion d'une série statistique. Initiation à l'utilisation de tableurs-grapheurs en statistique.

Compétences exigibles	- Organiser des données en choisissant un mode de présentation adapté : + tableaux en deux ou plusieurs colonnes ; + tableaux à double entrée. - Lire et compléter une graduation sur une demi-droite graduée, à l'aide d'entiers naturels, de décimaux ou de quotients (placement exact ou approché). - Lire et interpréter des informations à partir d'une représentation graphique (diagrammes en bâtons, diagrammes circulaires ou demi-circulaires, graphiques cartésiens).	Lire et interpréter un tableau, un diagramme à barres, un diagramme circulaire ou semi-circulaire. Regrouper des données statistiques en classes, calculer des effectifs. Présenter une série statistique sous la forme d'un tableau, la représenter sous la forme d'un diagramme ou d'un graphique. Calculer des fréquences.	Calculer des effectifs cumulés, des fréquences cumulées. Calculer la moyenne d'une série statistique. Calculer une valeur approchée de la moyenne d'une série statistique regroupée en classes d'intervalles	Une série statistique étant donnée (sous forme de liste ou de tableau, ou par une représentation graphique), proposer une valeur médiane de cette série et en donner la signification. Une série statistique étant donnée, déterminer son étendue ou celle d'une partie donnée de cette série.

Tableau 6. Tableau de programmes et compétences exigibles du collège entre 2005 et 2009

Les seuls changements notables concernent l'organisation des données puisque les tableaux à double entrée sont présentés dès la classe de 6ième et que le tableur informatique fait son entrée en classe – sans qu'aucune compétence le concernant ne soit énoncée. On notera cependant aussi la disparition dans le programme de la comparaison des distributions statistiques qui était mentionnée dans le programme précédent en classe de 3ième.

Une étude des types de tâche est disponible dans l'annexe A (chapitre Institution française, après la réforme, collège). Contre toute attente on y découvre un renforcement des tâches de comparaison entre séries statistiques dès la classe de 6ième. D'autre part les techniques de calcul de la moyenne sont diversifiées : moyenne pondérée et moyenne approchée donnent lieu à de nombreux exercices pour les élèves.

6è	5è	4è	3è
Diagramme	Classes Effectifs Fréquences	Effectifs Fréquences Moyenne	Moyenne Médiane Étendue

Tableau 7. Les notions de statistique dans les programmes du collège après la réforme

Les changements importants vont en fait avoir lieu en lycée.

Les programmes de lycée

Nous rappelons que la classe de seconde est dite classe de détermination ce qui signifie que tous les élèves des lycées d'enseignement général ont le même programme de mathématiques dans cette classe. Pour les classes de première et de terminale nous nous restreignons à la filière S là où l'enseignement des mathématiques tient la plus grande place dans l'ensemble des disciplines enseignées. Voici le programme :

	2nde	1ère S	Te S
Contenus	**Statistique** Résumé numérique par une ou plusieurs mesures de tendance centrale (moyenne, médiane, classe modale, moyenne élaguée) et une mesure de dispersion (on se restreindra en classe de seconde à l'étendue). Définition de la distribution des fréquences d'une série prenant un petit nombre de valeurs et de la fréquence d'un événement. Simulation et fluctuation d'échantillonnage.	**Statistique** Variance et écart-type. Diagramme en boîtes ; intervalle interquartile. Influence sur l'écart type et l'intervalle interquartile d'une transformation affine des données. **Probabilité** Définition d'une loi de probabilité sur un ensemble fini. Espérance, variance, écart-type d'une loi de probabilité. Probabilité d'un événement, de la réunion et de l'intersection d'événements. Cas de l'équiprobabilité. Variable aléatoire, loi d'une variable aléatoire, espérance, variance, écart type. Modélisation d'expériences aléatoires de référence (lancers d'un ou plusieurs dés ou pièces discernables ou non, tirage au hasard dans une urne, choix de chiffres au hasard, etc.).	Conditionnement et indépendance Conditionnement par un événement de probabilité non nulle puis indépendance de deux événements. Indépendance de deux variables aléatoires. Formule des probabilités totales. **Statistique et modélisation.** Expériences indépendantes. Cas de la répétition d'expériences identiques et indépendantes. **Lois de probabilité** Exemples de lois discrètes Introduction des combinaisons, notées $\binom{n}{p}$ Formule du binôme. Loi de Bernoulli, loi binomiale ; espérance et variance de ces lois. Exemples de lois connues Lois continues à densité : – loi uniforme sur [0,1] ; – loi de durée de vie sans vieillissement.

Modalités de mise en œuvre	Utiliser les propriétés de linéarité de la moyenne d'une série statistique. Calculer la moyenne d'une série à partir des moyennes de sous-groupes. Calcul de la moyenne à partir de la distribution des fréquences. Concevoir et mettre en œuvre des simulations simples à partir d'échantillons de chiffres au hasard.	**Statistique** On cherchera des résumés pertinents et on commentera les diagrammes en boîtes de quantités numériques associées à des séries simulées ou non. On observera l'influence des valeurs extrêmes d'une série sur l'écart type ainsi que la fluctuation de l'écart type entre séries de même taille. L'usage d'un tableur ou d'une calculatrice permet d'observer, dynamiquement et en temps réel, les effets des modifications des données. **Probabilité** Le lien entre loi de probabilité et distributions de fréquences sera éclairé par un énoncé vulgarisé de la loi des grands nombres. On expliquera ainsi la convergence des moyennes vers l'espérance et des variances empiriques vers les variances théoriques ; on illustrera ceci par des simulations dans des cas simples. On pourra aussi illustrer cette loi avec les diagrammes en boîtes obtenus en simulant par exemple 100 sondages de taille n, pour n = 10 ; 100 ; 1000. On simulera des lois de probabilités simples obtenues comme images d'une loi équirépartie par une variable aléatoire (sondage, somme des faces de deux dés, etc.)	On justifiera la définition de la probabilité de B sachant A, notée PA(B), par des calculs fréquentiels. On utilisera à bon escient les représentations telles que tableaux, arbres, diagrammes, etc., efficaces pour résoudre des problèmes de probabilités. Application à la problématique des tests de dépistage en médecine et à la loi de l'équilibre génétique lors d'appariements au hasard. Application aux expériences de références vues en seconde et première (dés, pièces, urnes, etc.). On introduira la notation n !. L'élève devra savoir retrouver les formules : $\binom{n}{p} = \binom{n-1}{p-1} + \binom{n-1}{p}$ $\binom{n}{p} = \binom{n}{n-p}$ On appliquera ces résultats à des situations variées. Application à la désintégration radioactive: loi exponentielle de désintégration des noyaux. Étude d'un exemple traitant de l'adéquation de données expérimentales à une loi équirépartie.

Tableau 8. Tableau des programmes du lycée, filière S, après la réforme (après l'an 2000)

Mentionnons une différence notable entre la filière S et la filière ES (sciences économiques et sociales) : la présence dans le programme de la terminale ES de statistique bidimensionnelle. Le contenu annoncé est : « Nuage de points associé à une série statistique à deux variables numériques. Point moyen. Ajustement affine par moindres carrés ».Dans les modalités de mise en œuvre, le programme précise :

> On proposera aussi des exemples où la représentation directe en (x ; y) n'est pas possible et où il convient, par exemple, de représenter (x ; ln y) ou (ln x ; y) et on fera le lien avec des repères semi-logarithmiques. On fera percevoir le sens de l'expression « moindres carrés » par le calcul sur tableur, pour un exemple simple, de la somme $\sum (y_i - ax_i - b)^2$. On évoquera sur des exemples l'intérêt éventuel et l'effet d'une transformation affine des données sur les paramètres a et b. On étudiera avec des

> simulations la sensibilité des paramètres aux valeurs extrêmes. On proposera des exemples où une transformation des données conduit à proposer un ajustement affine sur les données transformées. On proposera un ou deux exemples où les points (xi ; yi) du nuage sont « presque » alignés et où cet alignement peut s'expliquer par la dépendance « presque » affine à une troisième variable.

On pourra s'étonner de l'absence de statistique bidimensionnelle dans la filière S alors que les notions de corrélation et de régression ne sont pas réservées aux sciences économiques et sociales. De plus le prix à payer pour ES n'est pas négligeable car il s'agit de l'abandon de la notion de variable aléatoire. Les commentaires du programme n'avancent pas d'explication à ce choix :

> L'objectif est de faire des interpolations ou des extrapolations. On admettra les formules donnant les paramètres de la droite des moindres carrés : coefficient directeur et ordonnée à l'origine. On traitera essentiellement des cas où, pour une valeur de x, on observe une seule valeur de y (par exemple, les séries chronologiques). Le coefficient de corrélation linéaire est hors programme (son interprétation est délicate, notamment pour juger de la qualité d'un ajustement affine). On verra ainsi que pouvoir prédire y à partir de x ne prouve pas qu'il y ait un lien de causalité entre x et y.

En réduisant les activités des élèves autour de techniques de calcul, l'ambition semble modeste. S'agit-il de singulariser la filière ES en lui évitant d'aborder la notion de variable aléatoire jugée plus difficile ?

Synthétisons dans un tableau les notions statistiques et de probabilités proposées à enseigner au lycée, filière S, après la réforme :

Notions	Seconde	Première S	Terminale S
Statistiques	Moyenne Moyenne élaguée Médiane Classe modale Étendue Simulation Fluctuation d'échantillonnage	Simulation Variance et écart-type Diagramme en boîte Intervalle interquartile	Simulation
Probabilités		Loi de probabilité Probabilité d'un événement Espérance, variance, écart-type d'une loi de probabilité Modélisation d'expériences aléatoires de référence	Modélisation Conditionnellement Indépendance de deux événements Indépendance de deux variables aléatoires Loi de Bernoulli, loi binomiale Espérance et variance de ces lois

Tableau 9. Les notions statistiques et probabilistes dans les programmes du lycée après la réforme

Par rapport à la période précédente, on constate à la fois un étalement et un renforcement des enseignements de statistique.

- étalement puisque certaines notions comme l'écart-type sont renvoyées en classe de première. A-t-on cherché ainsi un rapprochement avec la même notion présentée pour une variable aléatoire ce qui mettrait en parallèle, la même année, les variables statistiques et les variables aléatoires ?

- renforcement puisqu'apparaissent les quartiles, l'intervalle interquartile et la box-plot. La description d'une série statistique s'enrichit. Apparaît aussi dès la classe de seconde la notion de fluctuation d'échantillonnage. Le choix de la classe de seconde pour introduire cette notion toute nouvelle peut s'expliquer par la volonté de faire venir l'approche fréquentiste des probabilités avant son approche laplacienne qui se fera en première. En effet la fluctuation d'échantillonnage nécessite une répétition d'une même expérience aléatoire et même si la notion de probabilité n'est pas encore un objet d'enseignement à ce niveau elle pourrait émerger d'une comparaison des fréquences empiriques entre elles. La notion de fluctuation d'échantillonnage

La loi des grands nombres fait son entrée officielle et même si elle n'est pas un objet d'enseignement en elle-même, sa présence en tant qu'outil dès la classe de seconde devrait permettre de conjuguer ensemble les objets de la Statistique déjà installés et les objets des probabilités à venir. D'ailleurs cette imbrication des concepts de statistique et de probabilités se poursuit en classe terminale avec l'adéquation de données expérimentales à une loi équirépartie, même si ce concept d'adéquation semble marginalisé par le choix de ne le traiter que sur un exemple.

Il y a volonté de donner à la loi des grands nombres un rôle de pivot dans l'articulation statistique-probabilités. Dans le document d'accompagnement, on peut d'ailleurs lire cette explication à l'adresse des enseignants :

> Pour valider un modèle probabiliste, le premier outil dont on dispose est **un théorème de mathématiques** [24] appelé « loi (forte) des grands nombres », dont un énoncé intuitif est : Dans le monde théorique défini par une loi de probabilité P sur un ensemble fini, les fréquences des éléments de cet ensemble dans une suite de n expériences identiques et indépendantes « tendent» vers leur probabilité quand n augmente indéfiniment. Ou encore : Si on choisit n éléments selon une loi de probabilité P, indépendamment les uns des autres, alors la distribution des fréquences est voisine de P lorsque n est grand.
>
> Il s'agit là d'un énoncé vulgarisé. Pour être un peu plus précis, le théorème appelé loi forte des grands nombres dit que, dans l'ensemble, des suites infinies d'éléments choisis selon P, le sous-ensemble des suites pour lesquelles la distribution des fréquences ne convergent pas vers P est «négligeable». Par exemple, dans une suite finie de n lancers d'une pièce équilibrée, on peut n'obtenir que des faces (code 0) ou que des piles (code 1), ou 001001…001, etc. ; ces trois suites finies ont chacune une probabilité 2-n ; si on «prolonge » ces suites, les fréquences de 1 et de 0 ne tendent pas vers 1/2, mais la probabilité de l'ensemble de ces trois suites tend vers 0. Le théorème ci-dessus indique que l'ensemble de toutes les suites imaginables pour lesquelles les fréquences ne tendent pas vers 0,5 est de «probabilité nulle» – pour une loi de probabilité construite sur l'ensemble des suites infinies de 0 et de 1 à partir de l'équiprobabilité sur {0 ; 1}. Le mathématicien dira qu'il y a convergence presque sûre des fréquences des éléments vers leur probabilité. Le statisticien, s'il observe dans une longue série d'expériences des distributions de fréquences qui fluctuent de moins en moins, choisira comme modèle, en vertu de cette loi des grands nombres, une loi de probabilité «proche» de la dernière

[24] En gras dans le texte.

> distribution observée. Une validation du modèle de la loi équirépartie pour le lancer d'un dé consistera à vérifier que la distribution des fréquences est «proche» de (1/6,…, 1/6) sur {1,…, 6} quand le nombre de lancers est grand : cela sera traité en classe terminale.
>
> (Document d'accompagnement des programmes de classe première S, 2002, p. 71)

C'est aussi ce que pense Henry [2008] lorsqu'il écrit :

> Les programmes des années 2000 font largement appel à la simulation informatique permettant de traiter de véritables séries statistiques de grandes dimensions et de visualiser concrètement les effets de la loi des grands nombres.
>
> (Henry [2008], p. 7)

Il n'est pas besoin de simuler pour répéter une expérience aléatoire et faire venir la loi des grands nombres mais la simulation et notamment la simulation informatique a une légitimité épistémologique encore plus forte, celle de la modélisation, ce que montre Henry [2008] en citant l'Encyclopédie Universalis :

> La simulation est l'expérimentation sur un modèle. C'est une procédure de recherche scientifique qui consiste à réaliser une reproduction artificielle (modèle) du phénomène que l'on désire étudier, à observer le comportement de cette reproduction lorsque l'on fait varier expérimentalement les actions que l'on peut exercer sur celle-ci, et à en induire ce qui se passerait dans la réalité sous l'influence d'actions analogues.

Le document d'accompagnement des programmes de première précise pour sa part :

> Modéliser consiste à associer un modèle à des données expérimentales, alors que simuler consiste à produire des données à partir d'un modèle prédéfini. Pour simuler une expérience, on associe d'abord un modèle à l'expérience en cours, puis on simule la loi du modèle.
>
> (Document d'accompagnement des programmes de classe première S, 2002, p. 71)

Et Henry [2008] de commenter, en mettant en évidence des choix implicites :

> Pour effectuer une simulation, il convient donc de faire d'abord le choix d'un modèle. Remarquons que les situations aléatoires considérées en classe sont en principe toutes à base d'équiprobabilité. Les modèles utilisés pour les simulations proposées sont donc souvent implicites quand ils se réduisent à la loi uniforme discrète, produite par le générateur aléatoire de l'ordinateur ou de la calculette qui fournissent des chiffres pseudo-aléatoires supposés équirépartis.
>
> (Henry [2008], p. 7)

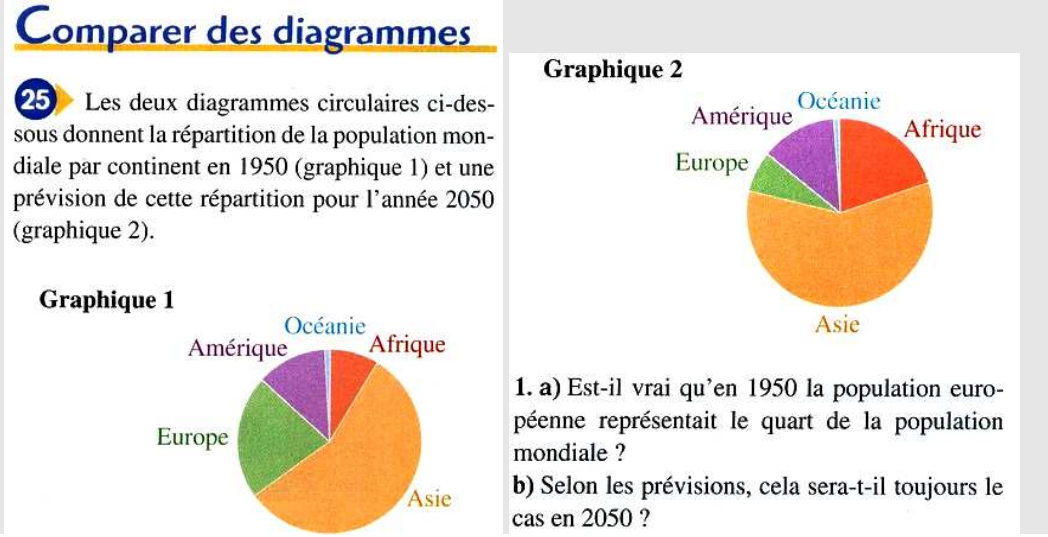
Ainsi, les nouveaux programmes de lycée ont pour intention manifeste de construire des liens entre la statistique et les probabilités. Le moteur est la loi des grands nombres et l'âme en est la modélisation. Comment ces liens vont-ils prendre forme dans les manuels ?

Les manuels

On se reportera à l'annexe A (chapitre « Institution française, après la réforme) pour les tableaux de types de tâche et les extraits de manuels. Nous ne présentons ici que les résultats qui illustrent les changements que nous venons de repérer dans les programmes. Nous nous référons pour les groupes de type de tâche au tableau 4 de cette partie.

Renforcement du groupe 3

- le problème de comparer deux distributions statistiques apparaît dès la classe de 6ième sous forme de comparaison de diagrammes circulaires. Voici un exemple illustratif d'une telle comparaison graphique :



Comparer des diagrammes

25 Les deux diagrammes circulaires ci-dessous donnent la répartition de la population mondiale par continent en 1950 (graphique 1) et une prévision de cette répartition pour l'année 2050 (graphique 2).

Graphique 1

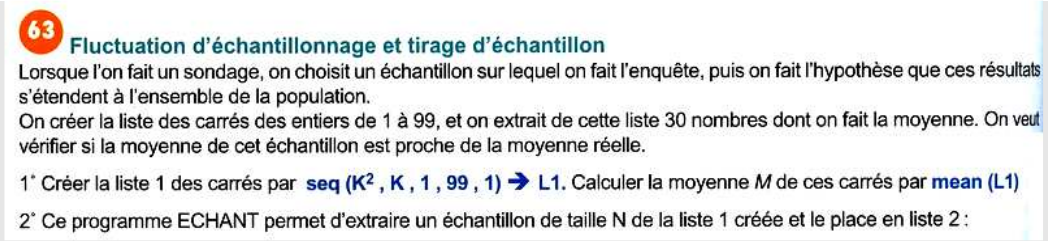
Graphique 2

1. a) Est-il vrai qu'en 1950 la population européenne représentait le quart de la population mondiale ?
b) Selon les prévisions, cela sera-t-il toujours le cas en 2050 ?

(Manuel 6è, collection DIABOLO, HACHETTE, 2005, p. 115)

Le type de tâche : comparer deux diagrammes circulaires ; la technique : évaluer les aires/angles de deux secteurs correspondants (sous-tâche) puis comparer les aires/angles entre eux ; la technologie : comparaison de deux fractions.

- l'observation de la fluctuation d'échantillonnage est prétexte à la comparaison entre un échantillon aléatoire et la population-mère. Le plus souvent on compare des moyennes. Voici un exemple illustratif d'une telle comparaison numérique :



63 **Fluctuation d'échantillonnage et tirage d'échantillon**

Lorsque l'on fait un sondage, on choisit un échantillon sur lequel on fait l'enquête, puis on fait l'hypothèse que ces résultats s'étendent à l'ensemble de la population.
On créer la liste des carrés des entiers de 1 à 99, et on extrait de cette liste 30 nombres dont on fait la moyenne. On veut vérifier si la moyenne de cet échantillon est proche de la moyenne réelle.

1° Créer la liste 1 des carrés par **seq (K^2 , K , 1 , 99 , 1) → L1.** Calculer la moyenne *M* de ces carrés par **mean (L1)**

2° Ce programme ECHANT permet d'extraire un échantillon de taille N de la liste 1 créée et le place en liste 2 :

T.I. 82 stats ou 83	Casio 35+	Programme ECHANT
dim (L1) → N	Dim List 1→ N ↵	Longueur de la liste L1 dans N
Input "TAILLE ECH. ?" , P	"TAILLE ECH." ? → P ↵	Demande la taille de l'échantillon…
P→ dim (L2)	Seq (0 , K , 1 , P , 1) → List 2 ↵	…qui sera la longueur P de la liste L2
L1→ L3	List 1→ List 3 ↵	Copie de L1 sur L3 pour préserver L1
For (K , 1, P)	For 1→ K To P ↵	Début de la boucle
randInt (1 , N) → J	Int (N × Ran#) + 1→ J ↵	Choix au hasard de la place J d'un nombre de
L3 (J) → L2 (K)	List 3 [J] → List 2 [K] ↵	L3 : L3 (J) rangé dans L2 et le dernier nombre
L3 (N) → L3 (J)	List 3 [N] → List 3 [J] ↵	de L3 vient prendre la place du nombre choisi
N – 1→ N	N – 1→ N ↵	N diminue de 1
End	Next	Fin de boucle

a) Entrer ce programme et l'exécuter pour $P = 30$; calculer la moyenne de l'échantillon par **mean (L2)** et noter cette moyenne arrondie à l'unité ; on pourra la rentrer en cellule 1 de la liste 4, c'est-à-dire L4 (1).

b) Recommencer 20 fois au total et noter les moyennes obtenues.

c) On obtient une liste 4 des moyennes des 20 échantillons.

Donner l'étendue de cette série et la moyenne m de cette série. Comparer m et M.

Calculer l'écart entre m et M ; à quel pourcentage de M cet écart correspond-t-il ?

Combien de valeurs de la liste 4 ont un écart avec la moyenne réelle M inférieur à 10 % de M ?

3° Recommencer l'étude avec des échantillons de taille 60.

(Manuel de seconde, collection Déclic, 2004, p. 214)

Le type de tâche : comparer les moyennes d'un échantillon et d'une population mère ; la technique : exécuter un programme de simulation d'un modèle de la loi mère (sous-tâche), calculer les moyennes des deux distributions (la mère et la fille) et apprécier la distance entre les deux nombres.

Ici pas de technologie statistique à proprement parler autre que calculatoire. En fait il faut chercher les motivations de cette tâche dans les deux dernières questions. La question « Combien de valeurs de la liste 4 ont un écart avec la moyenne réelle M inférieur à 10 % de M ? » est censée provoquer une réflexion sur le risque de se tromper de plus de 10 % en choisissant m à la place de M. La question « Recommencer l'étude avec des échantillons de taille 60 » pourrait permettre d'induire (si on a de la chance) une réduction de la fluctuation d'échantillonnage avec l'augmentation de la taille de l'échantillon. L'avantage de la simulation informatique est ici évident puisqu'elle rend négligeable le coût de production de l'échantillon même avec des tailles de plus en plus grandes. Par contre, puisqu'il cela ne coûte rien d'augmenter la taille, pourquoi s'intéresser au risque pour une certaine taille ? Quels sont les apprentissages visés ? Cette rencontre de la distribution d'échantillonnage de la moyenne sous la forme d'observations empiriques pilotées entièrement par l'enseignant peut-elle produire des apprentissages d'un savoir sur le risque à prendre quand on veut tirer la valeur d'un paramètre de la population-mère connue de celle d'un échantillon ? Nous reviendrons sur cette question dans la partie D car nous en ferons la question centrale d'une des séances de l'ingénierie.

- la comparaison de la distribution d'un échantillon avec un modèle de celle de sa population-mère est l'objet du test d'adéquation à une loi. Voici un exemple illustratif d'une telle comparaison entre distributions :

PROBLÈME

Une expérience aléatoire conduisant à k issues $x_1, \ldots, x_k$, dont les probabilités respectives sont $p_1, \ldots, p_k$, est réalisée n fois.

1) a) Pour $1 \leqslant i \leqslant k$, on note N_i la variable aléatoire égale au nombre de réalisations de l'issue x_i, sur les n expériences.
Donner la loi de N_i et son espérance mathématique.

b) Pour mesurer les « écarts aux moyennes », on considère la variable aléatoire :

$$D^2 = \sum_{i=1}^{k} \frac{(N_i - np_i)^2}{np_i}.$$

Montrer que l'espérance mathématique de D^2 : $E(D^2)$ est indépendante de n et des p_i.

2) Dans le cas de 120 lancers d'un dé équilibré, conjecturer, à l'aide de simulations, les vingtiles v_{18} et v_{19} de la loi de D^2 (définis par $P(D^2 > v_{18}) = 0{,}1$ et $P(D^2 > v_{19}) = 0{,}05$).

3) On convient, pour une expérience analogue, de rejeter la régularité d'un dé *de façon significative au seuil de 10 % (ou de 5 %)* lorsque la valeur observée d^2 de D^2 dépasse v_{18} (ou v_{19}).

Pour les 120 lancers, les six numéros sont sortis respectivement 26, 13, 22, 27, 12 et 20 fois.
Le dé est-il régulier ?

1) a) Sans ambiguïté, N_i suit la **loi binomiale** $\mathcal{B}(n, p_i)$.

On en déduit :

$$E(N_i) = np_i \text{ et } V(X) = np_i(1 - p_i).$$

b) D'après la linéarité de l'espérance :

$$E(D^2) = \sum_{i=1}^{k} \frac{1}{np_i} E[(N_i - np_i)^2] = \sum_{i=1}^{k} \frac{1}{np_i} V(N_i).$$

On déduit de ce qui précède :

$$E(D^2) = \sum_{i=1}^{k} (1 - p_i) = \sum_{i=1}^{k} 1 - \sum_{i=1}^{k} p_i.$$

Finalement :

$$E(D^2) = k - 1.$$

2) Nous avons simulé sur *Excel* 1 000 échantillons de 120 lancers d'un dé. Voici deux résumés visuels des résultats obtenus (*Fig.* 11 et *Fig.* 12).

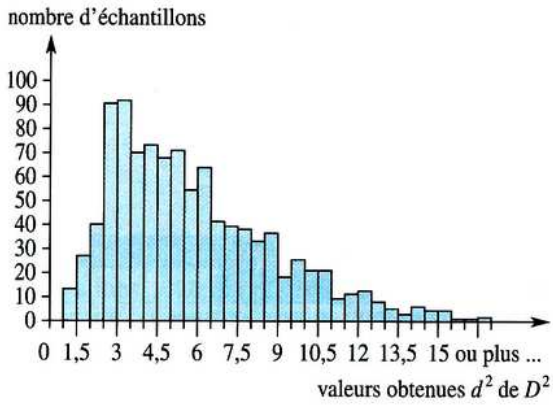


Fig. 11 : Histogramme

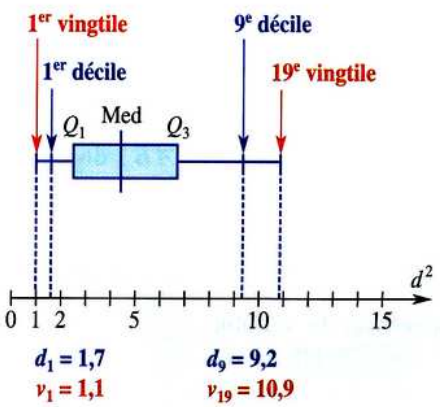


Fig. 12 : Boîte à moustaches

Commentaire : **d^2 dépasse la valeur 9,2 dans 10 % des cas, et dépasse 10,9 dans 5 % des cas.**

3) Pour le dé observé, $d^2 = 10,1$.

• $d^2 < v_{19}$: au seuil de 5 %, on ne peut pas rejeter l'hypothèse de régularité du dé, c'est-à-dire le modèle équiréparti ;

• $d_9 < d^2$: au seuil de 10 %, on décide de rejeter l'hypothèse de régularité du dé.

(Manuel Ter S, TERRACHER, HACHETTE, 2002, p. 314-315)

Cet exercice corrigé respecte la demande du programme qui incite les enseignants à conduire l'étude d'un exemple traitant de l'adéquation de données expérimentales à une loi équirépartie. Il s'agit en fait d'une comparaison entre deux distributions empiriques, l'une d'elles étant obtenue par simulation informatique d'une loi du Chi2 dont la présentation théorique est absente du programme. Le type de tâche : valider ou invalider le modèle théorique d'équirépartition ; la technique : calculer d^2 et D^2 (sous-tâches calculatoire) et comparer les deux nombres d^2 et D^2.

Ici pas de technologie statistique à proprement parler autre que calculatoire. En fait il faut chercher les motivations de cette tâche dans la possibilité offerte à l'enseignant de justifier l'utilité de la fluctuation d'échantillonnage puisque c'est la réalisation d'un échantillon de la population mère théorique qui produit une mesure du risque (sous forme de probabilité évaluée par la fréquence empirique) de se tromper si on accepte le modèle. Nous nous posons les mêmes questions que précédemment : quels sont les apprentissages visés ? Cette rencontre de la distribution d'échantillonnage de la distance du Chi2 sous la forme d'observations empiriques pilotées entièrement par l'enseignant peut-elle produire des apprentissages d'un savoir sur le risque à prendre quand on veut mesurer la distance d'une distribution empirique à une distribution modèle ?

Voici la nouvelle configuration du groupe 3 des types de tâche de comparaison :

Groupe 3 : comparer des distributions	- Comparer deux distributions empiriques - les tableaux - les graphiques - les résumés numériques - comparer la distribution d'un échantillon à celle de sa population-mère - Comparer une distribution empirique à une distribution théorique

Création d'un groupe 5 : une praxéologie d'échantillonnage ?

De l'analyse des activités et des exercices proposés aux élèves dans les manuels, nous concevons un nouveau groupe de types de tâche qui apparaissent dès la classe de seconde :

Groupe 5 : répéter une expérience aléatoire, produire et analyser des séries d'échantillons, modéliser une expérience aléatoire	observer la fluctuation d'échantillonnage modéliser et simuler une expérience aléatoire évaluer les indicateurs d'une population à partir d'un échantillon

Dans les manuels la production d'échantillons obtenus par la répétition d'une expérience aléatoire est réalisée sur calculatrice ou tableur et répond à l'incitation forte à la simulation[25] formulée dans le programme et le document d'accompagnement. On demande à l'élève de constater la fluctuation d'échantillonnage soit sur une fréquence particulière soit sur un indicateur comme la moyenne. Le plus souvent ce constat vise à mettre en évidence la réduction de la fluctuation par augmentation de la taille de l'échantillon.

Le seul type de tâche qui vise à mesurer et exploiter la fluctuation d'échantillonnage est celui que nous venons de présenter via l'exemple que fournit chaque manuel d'une adéquation de distribution empirique à une loi équirépartie. Ce type de tâche est isolé.

Tout ceci nous fait dire qu'il existe bien une *praxéologie d'échantillonnage* mais sa technologie se limite à la convergence des fréquences empiriques vers la probabilité dans le cas d'une répétition d'expérience aléatoire. C'est en quelque sorte une praxéologie de *stabilisation* de l'échantillonnage plutôt qu'une praxéologie de *fluctuation* d'échantillonnage.

[25] Nous n'avons pas repéré d'utilisation de tables de hasard comme le suggère le document d'accompagnement. Au-delà des manuels, nous nous demandons : comment cette simulation est-elle pratiquée en classe ?

Emergence d'un groupe 0 ?

L'insistance de la noosphère à faire valoir l'idée de sondage aléatoire conduit les manuels à mettre l'accent sur le recueil des données. En fait très souvent ce recueil est finalisé par une question relative à la fluctuation d'échantillonnage et n'est pas organisé en vue de répondre à une question « motivante » d'une étude didactique. Nous nous interrogeons sur l'émergence possible d'un groupe de type de tâches qui serait le suivant :

Groupe 0 : recueillir des données [26] dans une population, exhaustivement ou par sondage.	- Poser une question qui motive l'étude statistique - Procéder à une enquête - Organiser le recueil de données brutes

Un tel groupe se comprendrait en interrelation avec le groupe 4 et donnerait une raison d'être aux praxéologies de description, de comparaison et d'échantillonnage que nous avons mises en évidence.

I.3. Conclusion

Le renforcement de la Statistique dans l'EMS français s'est réalisé par l'introduction d'éléments de Statistique inférentielle axés sur la fluctuation d'échantillonnage que nous considérons comme une création didactique. En effet cette expression ne recouvre que partiellement ce que les statisticiens appellent communément la *distribution d'échantillonnage* et qui fait appel au concept de variable aléatoire. Cette création répondait pour ses promoteurs dans l'enseignement secondaire à la nécessité de créer des liens entre statistique et probabilités à la fois pour enrichir le concept de probabilité (l'approche fréquentiste en plus de l'approche laplacienne) et pour faire venir des problèmes dans lesquels un modèle probabiliste s'avère la clef de la résolution (comme celui de l'adéquation de données empiriques à une loi de probabilité qui fait appel à la loi du Chi2).

L'analyse des programmes et des manuels que nous venons de produire montre les effets praxéologiques de cette création dans le domaine de la statistique. On trouvera aussi en annexe A un complément d'étude portant sur le Brevet et le Baccalauréat français, avant et après la réforme. Cette analyse ne peut cependant suffire pour évaluer la portée didactique de la réforme de l'an 2000 ; nous renvoyons le lecteur intéressé à la thèse de Wozniak (2005). Par contre nous pouvons exploiter le repérage praxéologique que nous avons réalisé en France pour analyser maintenant l'institution vietnamienne.

[26] On recueille les données avant de les traiter. Cela explique le numéro 0 donné à ce groupe.

II. Analyse institutionnelle vietnamienne

Nous ne nous intéressons ici qu'à l'EMS vietnamien actuel. On trouvera dans l'annexe A, les programmes sous forme de tableaux.

II.1. Les programmes actuels

Au Vietnam, même si l'enseignement de la statistique est apparu plus tard qu'en France (il y a moins de dix ans) on le rencontre dès les classes primaires. A la classe 3 (équivalent à classe CE2 en France) se trouve une rubrique s'appelant « Eléments de statistique ». Cette première approche de la statistique consiste essentiellement dans la lecture de tableaux de distributions des effectifs, de diagrammes en rectangles et de diagrammes circulaires.

Mais l'enseignement de la statistique, s'il débute très tôt, ne se prolonge pas en continu dans le programme secondaire ce que montre le tableau suivant :

	Niveau de collège				Niveau de lycée		
Classe	6 (6è)	7 (5è)	8 (4è)	9 (3è)	10 (2de)	11 (1er)	12 (Ter)
Programme en vigueur		Statistique			Statistique	Combinatoire Probabilité	
Programme précédent	Statistique			Statistique			Combinatoire

Tableau 2. L'apparition de statistique et probabilité dans les classes secondaires au Vietnam

Au collège : la statistique ne se trouve qu'en classe 7 (5è) parmi les quatre classes du collège (10 séances sur 140 séances). Les données recueillies proviennent d'enquêtes au près de populations réelles supposées exhaustives. Ces populations sont généralement peu nombreuses et les valeurs prises par la variable statistique sont elles aussi peu nombreuses, et on n'effectue pas de regroupement de ces valeurs en classes. Les notions comme effectif, tableau des effectifs, moyenne, mode sont enseignées. La présentation des données statistiques est demandée sous la forme soit d'un tableau de distribution des effectifs soit sous celle d'un diagramme (diagramme en segment ou diagramme en rectangle ou diagramme circulaire). Par contre, la notion de fréquence n'est pas officiellement désignée.

Au lycée : la statistique occupe 7 séances sur 105 et son enseignement est découpé en quatre leçons intitulées : tableau de distribution des effectifs et des fréquences, diagrammes, moyenne-médiane-mode-variance et écart type. Les nouveautés de classe 10 par rapport à la classe 7 sont les suivantes :

- Le regroupement par classe : c'est la transformation de données observées en données regroupées par classes même si, notons le tout de suite, ce regroupement est entièrement à la charge de l'enseignant. Ceci conduit à la fabrication de diagrammes en colonnes (nous dirions histogrammes), aux calculs des indicateurs pour résumer les données.

- La fréquence en pourcentage : pour normaliser les données i.e. ramener les populations à une même taille égale à 100 ;

- La médiane (pour les données observées), la variance, l'écart type.

- Des résumés appelés « représentants » dans le programme.

Après la classe 10, les programmes de mathématiques ne comportent plus de chapitres de statistique. En classe 11, on trouve la combinatoire et les probabilités avec des notions d'expérience aléatoire, d'univers des possibles, et avec les théorèmes d'adition et de multiplication des probabilités.

Synthèse

	Collège	Lycée	
	7è	10è	11è
Statistique	Moyenne Diagramme en bâtons/rectangles Effectif - Tableau des effectifs Mode	Moyenne Diagramme en colonnes (histogramme) Mode	
		Médiane Fréquence - Tableau des fréquences Courbe polygonale des fréquences/effectifs Variance et écart type Regroupement en classes	
Probabilités			Expérience aléatoire Echantillon Evénement Probabilité d'un événement

Tableau 3. Tableau des notions de statistique-probabilité à enseigner au Vietnam

On retrouve dans ce tableau toutes les notions présentes dans les programmes français d'avant la réforme, certes à des moments différents de la scolarité et sur des durées moindres mais la structure praxéologique semble identique. Notons cependant l'absence des effectifs et fréquences cumulées et la présence d'un objet non repéré dans l'enseignement français : la courbe polygonale des effectifs/fréquences. Nous y deviendrons.

II.2. Les manuels

Nous plaçons en annexe A (chapitre « Institution vietnamienne ») les tableaux de type de tâche en comparaison avec ceux de la France. Nous ne pointons ici que quelques différences entre les enseignements vietnamiens actuels et les enseignements français avant réforme.

- Il n'existe qu'un seul ensemble de manuels[27] par niveau au collège et deux ensembles de manuels au lycée qui sont les seules références pour le travail des enseignants et pour les épreuves des examens et concours. En classe 10, les deux ensembles de manuels ont le même programme. L'élève utilise l'ensemble correspondant à son choix de section (sciences naturelles, sciences sociales, langue étrangère vivante, etc. ...). Généralement, celui qui choisit la section de science naturelle, utilise le manuel avancé et celui qui choisit la section de science sociale utilise le manuel élémentaire. Chaque ensemble se compose de trois livres :

- un livre des cours (manuel) pour les élèves : dans ce livre, chaque leçon est suivie par un nombre assez suffisant des exercices pour que les élèves s'entrainent et à la fin de chaque chapitre il y a un ensemble des exercices de révision.

- un livre des exercices (qui ne sont pas lesquels dans le livre des cours) pour les élèves : dans ce livre, pour chaque leçon, il y a toujours un résumé du cours, puis un exercice exemplaire avec sa solution, ensuite des exercices supplémentaires dont les solutions sont donnés à la fin de chaque chapitre.

- un livre des enseignants : ce livre présente la structure de chaque chapitre avec sa distribution d'emploi du temps. On y trouve des objectifs, des compétences et le contenu détaillé de chaque leçon et les corrigés des exercices correspondant du livre des cours, … Particulièrement, il y a une partie fournissant des connaissances supplémentaires sur une nouveauté aux enseignants à la fin du chapitre.

Venons-en à l'enseignement de la statistique :

- Les deux manuels de classe 10 ne diffèrent pas sensiblement ni sur l'ensemble des notions présentées ni sur l'ensemble des types de tâches. On retrouve dans les deux manuels une structure binaire : la représentation par tableaux et diagrammes, les indicateurs de position et de dispersion.

- Toutes les données sont tirées d'enquêtes sociétales, c'est-à-dire qu'il n'y a pas de simulation d'un modèle mathématique : le nombre des plantes plantées, la population des villes, les notes des élèves d'une classe, le nombre des élèves absents dans un école pendant une journée, le nombre d'enfants des familles, la durée du temps de déplacement entre la maison et l'école d'un élève, la durée du temps à courir un parcours de 100m des élèves d'une classe, les poids des boites de thé, le nombre d'élève d'une classe ayant le même mois de la date de naissance, le nombre d'années de travail des ouvriers dans une entreprise, les points gagnés notés d'un tireur d'arme à feu, la durée du temps à résoudre un problème mathématique des élèves d'une classe, … Ainsi, nous ne trouvons aucune situation où les données sont recueillies par des expériences aléatoires répétées. Le mot sondage n'est pas prononcé.

- On ne retrouve pas la distinction souvent présente en France entre données à caractère qualitatif et données à caractère quantitatif ni la distinction, elle aussi très fréquente en France, entre données quantitatives discrètes et données quantitatives continues. Ainsi, le diagramme en rectangles peut être formé soit par des rectangles isolés soit par des

[27] Un ensemble se compose d'un manuel pour élève, d'un livre d'exercices corrigés pour élève et d'un livre pour enseignant.

rectangles jointifs, sans lien avec la notion de données discrètes ou continues regroupées en classe. Autrement dit, il n'y a pas de notion d'histogramme distincte de celle de diagramme en rectangles.

- Les fréquences sont systématiquement écrites sous la forme fractionnaire ou en pourcentages.

- Un *échantillon* est présenté comme une partie de population sans aucune mention relative à un tirage au hasard :

> Un sous ensemble défini des individus enquêté est appelé **échantillon**. Le nombre des individus est appelé **la taille d'échantillon**. Les valeurs des caractères recueillis sur un échantillon est appelé **échantillon des données**[28].

(Manuel de classe 10, l'ensemble avancé, p. 160)

Tandis que le terme de *population* est expliqué comme suit :

> Si l'on fait une enquête sur tous les individus, dans ce cas, c'est une *enquête exhaustive*. Si l'on fait une enquête sur un échantillon, alors c'est une *enquête d'échantillon*[29].

(Manuel de classe 10, l'ensemble avancé, p. 160)

Et aussi :

> […] l'enquête sur tous les individus est impossible, il faut donc faire une enquête d'échantillon. Une des missions les plus importantes de la statistique est de construire les méthodes qui nous permettent de tirer des conclusions, d'établir des prévisions sur *tous les individus enquêtables*[30] en s'appuyant sur les informations acquis sur un échantillon […] L'ensemble de tous les individus enquêtés s'appelle *l'ensemble principal* […] .

(Livre d'enseignants, l'ensemble avancé, p. 217)

Nous découvrons ici un discours technologique pour aider les enseignants dans la justification des groupes de tâches 0 et 4 mais il est clair que le groupe 5 est totalement absent.

- pour comprendre le rôle dévolu à la courbe polygonale des fréquences, voici un exercice tiré du livre des exercices corrigés destinés aux élèves :

[28] En gras dans le document.

[29] En italique dans le document.

[30] En italique dans le document.

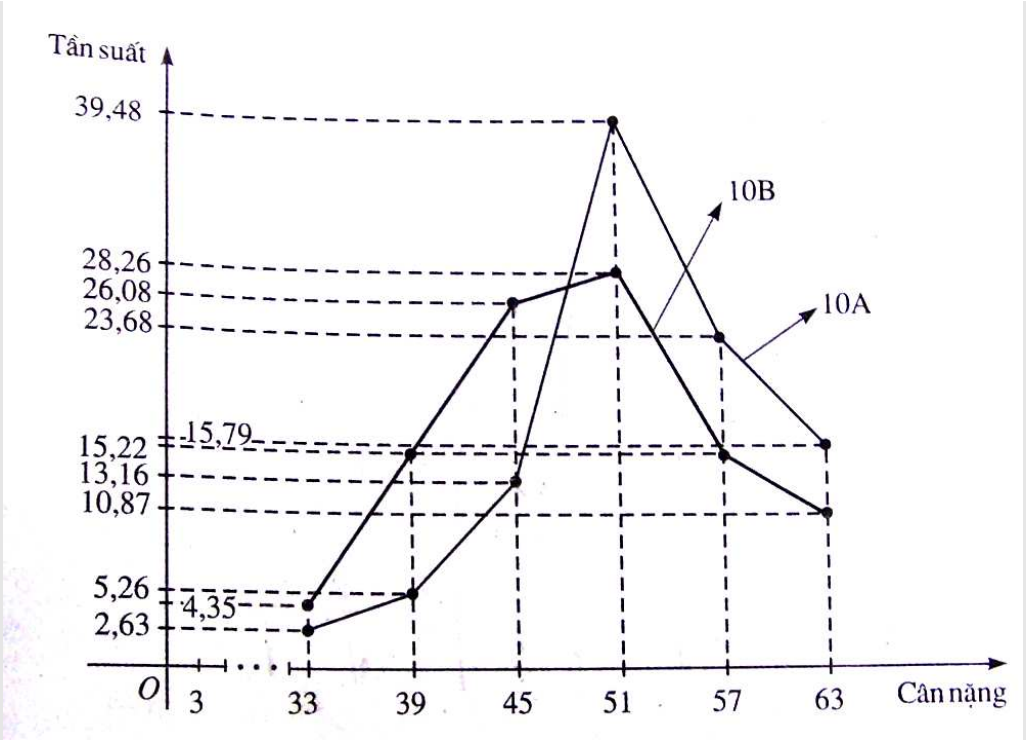
Exercice 19
Soit le tableau de distribution des effectifs regroupés en classes :
Le poids des élèves de classe 10A et 10B du lycée L.

Classe de poids	Effectif	
	10A	10B
[30 ; 36)	1	2
[36 ; 42)	2	7
[42 ; 48)	5	12
[48 ; 54)	15	13
[54 ; 60)	9	7
[60 ; 66)	6	5
Total	38	46

a) Etablir le tableau de distribution des effectifs regroupés en classes avec des classes données dans le tableau précédant.
b) Tracer dans le même plan de repère les courbes polygonales des effectifs regroupés en classes du poids des élèves de deux classes 10A et 10B. Puis comparer les poids des élèves de classe 10A et ceux des élèves de classe 10B.
c) Quel est le pourcentage des élèves de classe 10A pesant au moins de 42kg ? Même question pour les élèves de classe 10B.
d) Calculer la moyenne, l'écart type des données de classe 10A et 10B. Dans quelle classe que les élèves pèsent plus lourd ?

(Livre d'exercices corrigés de classe 10, l'ensemble avancé, p. 162)

En guise de solution à la question b), le livre donne la représentation graphique suivante :



Les courbes polygonales des fréquences regroupées en classes
du poids des élèves des classes 10A et 10B

(Livre d'exercices corrigés de classe 10, l'ensemble élémentaire, p. 173)

qu'il complète par cette interprétation :

> En observant deux courbes polygonales, on voit que : parmi les élèves pesant au plus 45kg, les élèves de classe 10A occupe un pourcentage plus grand que ceux de classe 10B. Et parmi les élèves pesant au moins de 51kg, les élèves de classe 10B occupe un pourcentage plus grand que ceux de classe 10A.

(Livre d'exercices corrigés de classe 10, l'ensemble élémentaire, p. 174)

Ainsi la courbe polygonale des fréquences sert d'élément technologique à la comparaison de deux distributions statistiques et s'inscrit dans une praxéologie de comparaison graphique. Une des questions qui nous intéressera dans la partie C sera celle de savoir de quel point de vue fonctionnel cette comparaison est porteuse.

- La praxéologie de comparaison numérique est également représentée. Voici un exercice qui demande de calculer les moyennes, les médianes, les variances et les écarts types de deux séries statistiques puis de fournir un commentaire :

Exercice 15 On a noté la vitesse (km/h) de 30 voitures passant chacun de deux parcours A et B : Parcours A :														
60	65	70	68	62	75	80	83	82	69	73	75	85	72	67
88	90	85	72	63	75	76	85	84	70	61	60	65	73	76
Parcours B :														
76	64	58	82	72	70	68	75	63	67	74	70	79	80	73
75	71	68	72	73	79	80	63	62	71	70	74	69	60	63
Calculer la moyenne, la médiane, la variance et l'écart type de vitesse des voitures roulant dans chaque parcours. D'après vous, dans quel parcours, les voitures roulent avec la plus sécurité ?														

(Manuel de classe 10, l'ensemble avancé, p. 179)

En fait la praxéologie de comparaison est très présente au Vietnam. Nous renvoyons à l'annexe A pour y trouver des exemples et des analyses.

- La répétition d'une même expérience aléatoire est introduite en classe 11 avec l'approche fréquentiste des probabilités. Mais cette notion ne donne pas lieu à des exercices pour les élèves et nous confirmons l'absence de toute praxéologie d'échantillonnage.

III. Conclusions de cette première analyse institutionnelle

On trouvera dans l'annexe A un tableau comparatif des types de tâches en France et au Vietnam. Les différences repérées entre l'enseignement français avant la réforme et l'enseignement vietnamien actuel n'empêchent de retrouver le même découpage praxéologique avec cependant une présence marquée des praxéologies locales de comparaison au Vietnam.

Le glissement d'une partie de la statistique du collège au lycée en classe 10 rapproche la statistique des probabilités maintenant enseignées en classe 11. On peut voir dans ce rapprochement un début d'évolution de l'enseignement de la statistique vers la construction de liens avec les probabilités, d'autant plus que l'approche fréquentiste n'est pas ignorée dans les manuels. Cependant cette évolution ne va pas jusqu'à imbriquer comme en France les deux enseignements et encore moins à faire des liens statistique-probabilités des objets d'enseignement[31].

Nous reprendrons l'analyse institutionnelle pour l'approfondir notamment sur les liens potentiels entre statistique et probabilités après un travail d'analyse épistémologique présenté dans la partie B.

[31] Certains de ces liens comme par exemple la ressemblance entre fréquence et probabilité, entre variable statistique et variable aléatoire, la moyenne et l'espérance, donne lieu à une lecture complémentaire dans le manuel avancé (pages 91-92).

Partie B

Analyse épistémologique

L'objet de cette partie est d'apporter des éléments de réponse à la question (QR1) :

QR1. Autour de quels problèmes et avec quels objets mathématiques – que nous appellerons problèmes et objets candidats – construire des liens statistique-probabilités dans l'enseignement mathématique secondaire ?

Pour cela nous nous interrogeons sur l'origine « savante » des praxéologies que nous avons dégagées dans la partie A quand nous avons exploré les savoirs à enseigner des EMS français et vietnamien. En nous appuyant sur les travaux de Chevallard [1985], nous partons en effet de l'hypothèse que ces praxéologies à enseigner résultent d'un processus de transposition didactique à partir de savoirs qualifiés socialement de Statistique ou de Probabilités et produits par une institution « savante », source de légitimation et de contrôle pour EMS des savoirs à enseigner. Certes les expressions de savoir savant et d'institution savante peuvent prêter à confusion et surtout à débat. Cela justifie que nous mettions des guillemets mais il nous paraît incontestable que les praxéologies à enseigner portent les traces de pratiques et de théories élaborées ailleurs que dans l'institution d'enseignement que nous considérons. Ces traces sont faites de reprises, de transformations et aussi de mises à l'écart qui sont toutes significatives des contraintes institutionnelles.

Il n'est pas possible dans le cadre de notre étude d'explorer toutes les institutions productrices des savoirs et savoir-faire auxquelles pourrait se référer EMS. Il nous apparaît que la principale d'entre elles, car la plus influente sur EMS, est l'enseignement supérieur. C'est pourquoi nous choisissons de mener notre enquête épistémologique en nous restreignant à l'étude de manuels universitaires. Nous choisissons, pour leur notoriété, deux traités[32] universitaires contemporains à partir desquels nous entreprenons de construire le modèle épistémologique de référence relatif aux liens statistique-probabilités unidimensionnelles.

Le titre du premier traité est « Probabilités, analyse des données et Statistique » ; son auteur est un universitaire français, Gilbert Saporta ; par la suite, nous le désignerons par « Traité Saporta ».

Le titre du second traité est « Statistique » ; ses auteurs sont deux universitaires américains Thomas H. Wonnacott et Ronald J. Wonnacott ; par la suite, nous le désignerons par « Traité Wonnacott ».

Nous allons expliquer nos choix en présentant les deux ouvrages.

I. Présentations comparées des deux traités

I.1. Le traité Saporta

Voici la préface de l'éditeur de la 2e édition (2006) à laquelle nous nous référons :

[32] Nous employons le mot *traité* au sens de Neyret : « Nous appelons « traité » le livre produit par un auteur (considéré comme une institution transpositive), appartenant à la sphère de production du savoir […], ou très proche de celle-ci, dans le but d'élémenter les savoirs à enseigner », Neyret [1995] p. 57.

> La démarche statistique n'est pas seulement une auxiliaire des sciences destinée à valider ou non des modèles préétablis, c'est aussi une méthodologie indispensable pour extraire des connaissances à partir de données et un élément essentiel pour la prise de décision. La très large diffusion d'outils informatiques peut donner l'illusion de la facilité à ceux qui n'en connaissent pas les limites, alors que la statistique est plus que jamais un mode de pensée fondamental pour maîtriser la complexité, l'aléatoire et les risques, en donnant la prudence scientifique nécessaire.
>
> Ce manuel présente l'ensemble des connaissances utiles pour pouvoir pratiquer la statistique. Il est destiné à un vaste public (étudiants, chercheurs, praticiens de toutes disciplines) possédant le niveau d'algèbre et d'analyse d'un premier cycle universitaire scientifique ou économique.
>
> Cette nouvelle édition est une révision complète, avec des ajouts, de l'édition de 1990 et comporte de nombreux développements sur des méthodes récentes. Les 21 chapitres sont structurés en cinq parties : outils probabilistes, analyse exploratoire, statistique inférentielle, modèles prédictifs et recueil de données. On y trouve l'essentiel de la théorie des probabilités, les différentes méthodes d'analyse exploratoire des données (analyses factorielles et classification), la statistique « classique » avec l'estimation et les tests mais aussi les méthodes basées sur la simulation, la régression linéaire et logistique ainsi que des techniques non linéaires, la théorie des sondages et la construction de plans d'expériences.
>
> (Saporta [2006])

On le voit, il s'agit d'un ouvrage qui ne se veut pas une initiation à la statistique mais a pour ambition de dresser un panorama le plus complet possible des « connaissances utiles pour pouvoir pratiquer la statistique » au-delà de l'utilisation de logiciels dont il critique la superficialité. Pour acquérir ces connaissances le minimum mathématique requis se situe à la licence de mathématiques. Nous nous attendons donc à un *régime épistémologique*[33] très mathématisé où les concepts et les méthodes exposés seraient libérés des diverses et multiples applications scientifiques.

Ce régime épistémologique correspond à l'une des deux traditions de l'enseignement de la statistique.qui se sont combattues dans les milieux universitaires français. Pour l'une de ces traditions, l'enseignement de la statistique relève de l'enseignement de la discipline qui en a besoin dans son investigation scientifique. Pour l'autre, « le fait même que la statistique serve à des utilisateurs aussi étrangers les uns aux autres que le naturaliste, le psychologue, l'économiste, le physicien, justifie la nécessité d'une synthèse et l'intervention du mathématicien »[34]. On pourra trouver dans Wozniak [2005][35] un panorama et les grandes lignes de force de l'histoire de l'enseignement statistique en France. Elle y souligne la « faible pénétration de la statistique dans la culture française » à l'inverse de la culture anglo-saxonne plus encline à développer des mathématiques mixtes que Wozniak [2005] présente ainsi pour la statistique :

> Toute statistique mêle nécessairement des objets mathématiques et des objets non mathématiques. La mise en œuvre de la « méthode statistique, c'est-à-dire de la technologie statistique, enclenche ainsi, dans le meilleur des cas, de véritables synergies

[33] Chevallard [1985]

[34] Wozniak (2005, p. 43) citant Hennequin P.L. dans *Initiation à la statistique* (1967)

[35] Plus particulièrement dans le chapitre 1

> co-disciplinaires, en articulant les énergies de deux disciplines au moins, l'une mathématique, l'autre non. Sur ce patron générique, on pourra faire de la statistique en médecine, ou en démographie, ou en lexicologie, ou en psychologie, ou en didactique, etc., ce qui appellera chaque fois, le cas échéant, des adjonctions spécifiques à « la » méthode statistique.

(Wozniak [2005] p. 29)

I.2. Le traité Wonnacott

Voici justement un traité renommé dans les universités anglo-saxonnes. Le titre complet de l'édition française (1991) de la 4ième édition à laquelle nous nous référons, est « Statistique : économie, gestion, sciences, médecine ». Dans la préface, Cohendet P., professeur de sciences économiques à l'université de Strasbourg, insiste sur les qualités pédagogiques de l'ouvrage destiné à un public d'étudiants non mathématiciens :

> Les auteurs ont cherché avant tout à expliquer le plus clairement possible les principaux concepts statistiques. L'usage des mathématiques est limité autant que possible aux principes de calcul les plus connus. Toutefois, certaines sections ou certains chapitres apparaissent un peu plus difficiles, et sont précédés d'un astérisque.
>
> Les quinze premiers chapitres présentent les principes statistiques fondamentaux. Les dix derniers chapitres traitent de sujets plus spécifiques mais tout aussi importants, comme les tests non paramétriques, les séries chronologiques, les arbres de décision, la statistique bayésienne, ou les systèmes à équations simultanées.
>
> Cette 4ème édition du « Wonnacott and Wonnacott » accorde encore davantage de place aux exercices d'application, insiste sur la discussion et l'interprétation des résultats, et invite le lecteur à un apprentissage particulièrement stimulant de la méthodologie statistique.

(Wonnacott [1991], p. IX)

On notera le parti pris de minorer les pré-requis mathématiques dans l'étude de la statistique tout en conservant « les principaux concepts statistiques ». On relèvera a contrario la volonté de donner aux applications un rôle dans l'apprentissage des concepts. Nous nous attendons donc à un régime épistémologique plus « littéraire » et plus attaché à l'usage professionnel des savoirs statistiques.

La question qui nous préoccupe maintenant est : par delà ces différences de régime épistémologique, les structures et les contenus dégagent-ils un même système praxéologique mathématique, lequel nous servirait de référence ?

II. Descriptions comparées des deux traités

Nous présentons les structures de chacun des deux traités et mettons en lumière certains choix épistémologiques et didactiques. Nous entrons dans certains chapitres pour montrer les différences entre les deux traités, plus particulièrement dans le chapitre introductif avec lequel les auteurs annoncent leurs intentions et certaines raisons de leurs choix.

II.1. Le traité Saporta

Il se décompose en cinq parties dans une structure linéaire :

La première partie

Appelée « Outils probabilistes », elle comporte quatre chapitres : le modèle probabiliste, variables aléatoires, couples de variables aléatoires et conditionnement, vecteurs aléatoires. Il s'agit pour l'auteur d'installer la théorie des probabilités en soubassement de tout son traité, dans ses aspects uni et multi dimensionnels, théorie dont il dit :

> En tant que théorie mathématique, la théorie des probabilités n'a pas à être justifiée : une fois ses axiomes posés, elle se développe de façon autonome par rapport à la réalité concrète.
>
> (Saporta [2006], p. 3)

Le modèle probabiliste est donc construit à partir de l'axiomatique de Kolmogorov et présente dans ce cadre les notions d'expérience aléatoire et d'événements dont on calcule les probabilités.

Le chapitre « Variables aléatoires » est le plus fourni et fournit un catalogue impressionnant de distributions discrètes et continues toutes introduites par leur loi de probabilité, leur fonction de répartition ou leur fonction de densité. La loi normale appelée ici loi de Laplace-Gauss n'est que l'une de ces multiples lois malgré le petit paragraphe d'introduction qui lui est réservé :

> Cette loi joue un rôle fondamental en probabilités et statistique mathématique. Elle constitue un modèle fréquemment utilisé dans divers domaines : variation du diamètre d'une pièce dans une fabrication industrielle, répartition des erreurs de mesure autour de la « vraie valeur », etc.
>
> […] Elle est cependant loin décrire tous les phénomènes physiques et il faut se garder de considérer comme anormale une variable ne suivant pas la loi de Laplace-Gauss. Son rôle principal en statistique provient en réalité de ce qu'elle apparaît comme loi limite de caractéristiques liées à un échantillon de grande taille.
>
> (Saporta [2006], p. 43)

Le chapitre développe donc logiquement et longuement les notions de convergence des suites de variables aléatoires suivi du théorème central-limite.

La deuxième partie

Appelée « Statistique exploratoire », elle se décline en sept chapitres dont les deux premiers intitulés respectivement « description unidimensionnelle de données numériques » et « description bidimensionnelle de données numériques et mesures de liaison entre variables » correspondent à ce qui est communément appelé Statistique descriptive. Les autres chapitres font partie de ce que d'aucuns appellent Analyse des données. D'emblée l'auteur se place dans la perspective multidimensionnelle :

> La plupart du temps les données se présentent sous la forme suivante : on a relevé sur n unités appelés « individus » p variables numériques. Lorsque n et p sont grands on cherche à synthétiser cette masse d'informations sous une forme exploitable et compréhensible. Une première étape consiste à décrire séparément les résultats obtenus pour chaque variable : c'est la description unidimensionnelle, phase indispensable mais insuffisante dans toutes étude statistique.
>
> (Saporta [2006], p. 109)

Le chapitre de statistique descriptive unidimensionnelle est très succinct puisque les paramètres des distributions ont été présentés avec les variables aléatoires : quelques considérations rapides sur les graphiques et les résumés numériques. Voici ce qui concerne les histogrammes :

> Analogue à la courbe de densité d'une variable aléatoire, un histogramme est un graphique à barres verticales accolées, obtenu après découpage en classes des observations d'une variable continue. La surface de chaque barre, encore appelée tuyau d'orgue, doit être proportionnelle à la fréquence de la classe.
>
> (Saporta [2006], p. 114)

L'histogramme est placé en filiation de la courbe de densité ce qui met en évidence le rôle de l'aire dans son usage mais dévalorise son intérêt théorique. Après un exemple, l'auteur poursuit :

> On peut d'ailleurs critiquer le fait de représenter par une fonction en escalier la distribution d'une variable continue : l'histogramme est une approximation assez pauvre d'une fonction de densité et il serait plus logique de chercher une fonction plus régulière. La théorie de l'estimation de l'estimation d'une densité permet de proposer des solutions à ce problème (voir chapitre 13).
>
> (Saporta [2006], p. 114)

On peut s'étonner du faible intérêt de l'auteur pour les techniques de description statistique usuelles dans la communication scientifique et sociale. C'est là, à notre avis, une conséquence de son choix d'une statistique très mathématisée.

La troisième partie

Appelée « Statistique inférentielle », elle est constituée de quatre chapitres dont voici les intitulés : distributions des caractéristiques d'un échantillon, l'estimation, les tests statistiques, méthodes de Monte-Carlo et de rééchantillonnage. Elle débute par la formulation du problème central de l'inférence statistique :

> Le problème central de l'inférence statistique est, rappelons-le, le suivant : disposant d'observations sur un échantillon de taille n, on désire en déduire les propriétés de la population dont il est issu. Ainsi on cherchera à estimer, par exemple, la moyenne m de la population à partir de la moyenne $\overline{x}$ d'un échantillon. Ceci n'est possible que si l'échantillon a été tiré selon des règles rigoureuses destinées à en assurer la « représentativité ». Le mode de tirage le plus simple et aussi le plus important est l'échantillonnage aléatoire simple correspondant à des tirages équiprobables et indépendants les uns des autre. Dans ces conditions les observations deviennent des

> variables aléatoires ainsi que les résumés numériques usuels : il convient donc d'en chercher les lois de probabilité avant de tenter d'extrapoler à la population.

(Saporta [2006], p. 272)

C'est seulement après s'être muni des distributions d'échantillonnage établies d'après le modèle probabiliste que l'auteur engage la résolution du problème et la présentation des différents tests paramétriques et non-paramétriques comme les tests d'ajustements dont le test du Khi 2.

Attardons-nous sur le premier chapitre qui traite des distributions d'échantillonnage. La base mathématique est ce que l'auteur appelle l'hypothèse fondamentale de la théorie de l'échantillonnage :

> Les valeurs observées x_i sont des réalisations d'une même variable aléatoire X, appelée variable parente. En associant à chaque individu i tiré une variable aléatoire X_i dont on observe une seule réalisation x_i, on peut dire que les X_i ont la même distribution, celle de X.

(Saporta [2006], p. 271)

Ainsi l'auteur décide-t-il de définir un échantillon sous la forme du n-uple (X_1, X_2, …, X_n) de *variables aléatoires indépendantes et identiquement distribuées* ce qui place les statistiques obtenues par répétition d'une même expérience aléatoire dans le cadre probabiliste. D'ailleurs, l'auteur annonce cette définition qui lie définitivement statistique et probabilités :

> Une statistique T est une variable aléatoire fonction mesurable de X_1, X_2, …, X_n : $T = f(X_1, X_2, \ldots, X_n)$

(Saporta [2006], p. 272)

La quatrième partie

Appelée « Modèles prédictifs », elle comporte quatre chapitres qui portent sur la régression (simple, multiple, logistique et algorithmes). L'auteur commence par la régression simple avec une seule variable explicative, en approfondissant ce qu'il avait présenté dans le deuxième chapitre de la deuxième partie (description bidimensionnelle de données numériques et mesures de liaison entre variables). Son projet est cependant plus général et multidimensionnel ; il le mène avec les objets de l'Algèbre linéaire qu'il qualifie souvent de géométriques.

La cinquième partie

Dernière partie, appelée « Recueil de données », elle présente deux chapitres, l'un sur les sondages et l'autre sur les plans d'expérience. L'auteur reconnaît que « le recueil de données constitue un préalable à l'analyse » mais c'est pour justifier immédiatement que cette partie plus appliquée soit la dernière. En effet, dit-il :

> […] la collecte des données ne peut se concevoir sans en connaître l'usage ultérieur, ce qui nécessite la compréhension de l'estimation et de la modélisation.

Mais il accepte de s'approcher de la pratique des usagers de la statistique. Il débute ainsi le chapitre sur les sondages :

> Les méthodes de sondage ont pour objectif de tirer dans une population concrète des échantillons destinés à estimer avec la meilleure précision possible des paramètres d'intérêt. Le tirage équiprobable avec remise qui conduit à des échantillons de variables aléatoires indépendantes et identiquement distribuées est la base des développements des chapitres précédents et est le modèle de la statistique mathématique ; ce mode de tirage ne correspond pas en fait à la pratique et n'est au mieux qu'une approximation commode. Les sondages réels portent sur des populations finies et sont effectués par tirage sans remise, pour ne pas risquer d'interroger deux fois le même individu. […] Ce chapitre a pour objectif de donner une initiation à la théorie des sondages aléatoires, et ne prétend nullement couvrir le sujet. […]
>
> (Saporta [2006], p. 511)

Le chapitre d'introduction

Dans ce chapitre l'auteur veut « préciser les notions fondamentales de la statistique et les rapports qu'elle entretient avec la théorie des probabilités ainsi que ce qu'on entend par démarche statistique ». C'est le deuxième point qui retient ici notre attention ; l'auteur voit trois liens que nous citons.

Premièrement, c'est *l'incertitude des mesures.*

> […] les données observées sont souvent imprécises, entachées d'erreurs. Le modèle probabiliste permet de représenter comme des variables aléatoires les déviations entre « vraies » valeurs et valeurs observées.
>
> (Saporta [2006], p. xxviii)

Deuxièmement, c'est *la comparaison des distributions* :

> […] on constate souvent que la répartition statistique d'une variable au sein d'une population est voisine de modèles mathématiques proposés par le calcul des probabilités.
>
> (Saporta [2006], p. xxviii)

Dernièrement, c'est *la représentativité d'un échantillon* :

> […] les échantillons d'individus observés sont la plupart du temps tirés au hasard dans la population, ceci pour assurer mathématiquement leur représentativité : si le tirage est fait de manière équiprobable chaque individu a une probabilité constante et bien définie d'appartenir à l'échantillon. Les caractéristiques observées sur l'échantillon deviennent, grâce à ce tirage au sort, des variables aléatoires et le calcul des probabilités permet d'étudier leurs répartitions […].
>
> (Saporta [2006], p. xxviii)

Pour l'auteur, il faut bien distinguer ce dernier rôle des probabilités des deux premiers :

> Dans les premiers cas le calcul des probabilités propose des modèles mathématiques simplificateurs, éventuellement contestables, du comportement d'un phénomène ; dans le dernier cas, le calcul des probabilités fournit des théorèmes si le processus d'échantillonnage est respecté […]
>
> (Saporta [2006], p. xxviii)

Ainsi l'association statistique-probabilité serait épistémologiquement double :

- une association réalité-modèle dans laquelle la statistique observe et les probabilités modélisent

- une association méthode-théorie dans laquelle la statistique fabrique et exploite les données dans des conditions que les probabilités justifient.

Ces associations ne sont pas exclusives puisque les probabilités peuvent modéliser des phénomènes pour lesquelles on n'a pas de statistiques et, comme le rappelle l'auteur, certaines méthodes statistiques ne sont pas aléatoires.

II.2. Le traité Wonnacott

Il se décompose lui aussi en cinq parties mais dans une structure arborescente. Le tronc commun comporte trois parties que nous présentons sommairement :

(I) Eléments de probabilité et de statistique

Après un premier chapitre intitulé « Nature de la statistique » sur lequel nous reviendrons, on y présente, dans cet ordre, Statistique descriptive, Probabilités, Distributions de probabilité et Couple de variables aléatoires.

Entrons dans les chapitres dont les titres s'apparentent à ceux de l'enseignement secondaire : Statistique descriptive, Probabilités, Distributions de probabilité.

Celui de statistique descriptive commence ainsi :

> Jusqu'ici, on a mis en évidence l'objet fondamental de la statistique : dégager d'un échantillon des résultats valables pour l'ensemble de la population. Dans une étape préliminaire, il faudra simplifier l'échantillon et le réduire à quelques nombres qui le décrivent, nombres appelés « statistiques » de l'échantillon.
>
> (Wonnacott [1991], p. 29)

Dans une telle démarche, échantillon et population ont le même statut et sont soumis au même traitement fait de tableaux, de diagrammes et de résumés numériques. L'auteur prend soin de présenter sur des exemples la pertinence de tel ou tel graphe et de tel ou tel résumé numérique. Les probabilités n'ont pas été encore abordées si bien que ce chapitre ressemble dans son organisation et son contenu à tous les chapitres de Statistique descriptive que l'on rencontre dans l'enseignement secondaire français avent la réforme.

Celui des probabilités ne s'éloigne pas non plus des chapitres « Probabilités » de certains manuels de l'enseignement secondaire français. La probabilité est définie d'abord de manière fréquentiste comme la limite des fréquences dans une infinité de répétitions d'une même expérience aléatoire. Puis l'auteur monte pas à pas un modèle fait d'événements sans recourir explicitement à une axiomatique.

Celui des distributions de probabilité installe les concepts de variable aléatoire, de lois de probabilité et de fonction de densité, ce dernier en lissant un histogramme par augmentation de la taille de l'échantillon. Une seule loi discrète : la loi binomiale. Une seule loi continue : la loi normale. Cette dernière est introduite par le graphe de sa fonction de densité, sans formule et sans justification autre que celle-ci :

> De nombreuses variables aléatoires ont pour fonction de densité une courbe en forme de cloche, appelée courbe normale […]. Il s'agit de la loi statistique la plus répandue et la plus utile. En effet les erreurs de mesure portant sur des phénomènes physiques ou économiques, par exemple, sont très souvent distribuées normalement. De plus, de nombreuses lois de probabilité (comme la loi binomiale) peuvent souvent être approchées par la loi normale.
>
> (Wonnacott [1991], p. 142)

Le chapitre est illustré de nombreux exemples tirés de la « vie quotidienne » et l'outillage mathématique ne comporte aucune notion d'Analyse autre que celle de fonction d'une variable réelle : on ne trouve ni dérivée ni intégrale.

(II) L'inférence statistique

Cinq chapitres pour cette partie : l'échantillonnage, l'estimation ponctuelle, l'estimation par intervalle, tests d'hypothèses, analyse de variance. La distribution d'échantillonnage en le cœur. Notons que le chapitre sur les tests ne présente que quelques tests paramétriques que l'auteur appelle tests classiques et ne comporte donc pas les tests d'ajustement.

Examinons de plus près le chapitre dédié à l'échantillonnage. Le but affiché est « de répondre à la question fondamentale de la statistique déductive : que peut-on attendre d'un échantillon aléatoire issu d'une population connue ? » :

> Mais avant de développer cette question, il faut mettre à nouveau en garde le lecteur sur un point important : la manière de recueillir les données de l'échantillon est au moins aussi importante que la façon de l'analyser. En particulier il faut qu'un échantillon soit représentatif de la population, et l'échantillon aléatoire constitue le meilleur moyen d'y parvenir. Un échantillon qui n'est pas aléatoire, c'est bien pire qu'un échantillon inutile.
>
> (Wonnacott [1991], p. 211)

La représentativité de l'échantillon fabriqué à partir de la population, d'ailleurs non définie, est donc la motivation de l'étude. Le tirage sans remise d'une urne sert de modèle à la définition d'un échantillon et aboutit à la définition d'un échantillon aléatoire simple comme étant un échantillon dont les observations sont indépendantes donc de même loi, de même moyenne et de même écart-type que la population. On notera que le langage bien que probabiliste délaisse la notion de variable aléatoire contrairement au choix du traité de Saporta.

L'auteur s'intéresse à la distribution d'échantillonnage de la moyenne. Prenant appui uniquement sur des graphiques, il fournit ce qu'il appelle la Règle de l'Approximation Normale qui tient lieu de théorème central-limite :

> Dans les échantillons aléatoires simples de taille n, la moyenne de l'échantillon varie autour de la moyenne μ de la population avec un écart-type égal à $\sigma/\sqrt{n}$ (où σ est l'écart-type de la population).
>
> Donc quand n s'accroît, la distribution d'échantillonnage de la moyenne est de plus en plus concentrée autour de l'objectif μ. Elle devient de plus en plus proche de la distribution normale (forme en cloche).

(Wonnacott [1991], p. 225)

Voici les graphiques :

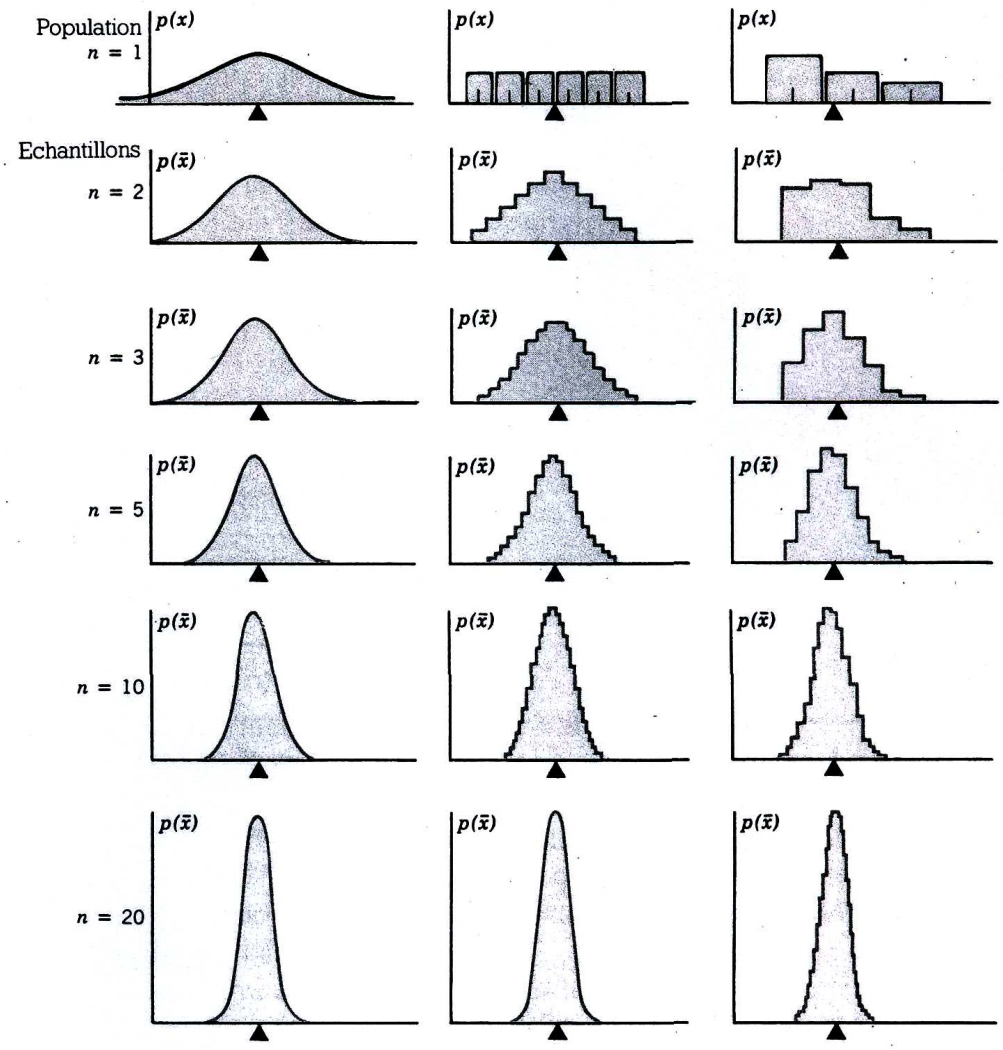


FIGURE 6-3
La distribution d'échantillonnage de $\bar{X}$ (en couleur) diffère de la distribution de la population-mère (en gris). La colonne de gauche représente l'échantillonnage à partir d'une population normale. A mesure que la taille de l'échantillon s'accroît, l'écart type de $\bar{X}$ diminue. Les deux colonnes suivantes montrent comment, en dépit du caractère non normal de la population, la distribution d'échantillonnage devient approximativement normale.

(Wonnacott [1991], p. 224)

Les graphiques servent de technologie[36] sans avoir recours aux théorèmes de convergence sur les suites de variables aléatoires comme le fait le traité Saporta. On quitte le cadre probabiliste pour retourner dans celui de la Statistique. Ce chapitre est très illustratif de la volonté de l'auteur de limiter la mathématisation de la Statistique inférentielle.

(III) Régression

L'ajustement linéaire, la régression simple, la régression multiple, extensions de la régression, corrélation constituent cette partie. Au-delà de la méthode des moindres carrés ordinaires, le modèle linéaire généralisé y est développé mais sans recourir aux objets de l'Algèbre linéaire.

Les deux autres parties assemblent, pour la première quelques thèmes de l'inférence classique (statistiques non paramétriques, statistiques robustes, tests du Chi 2, estimation par le maximum de vraisemblance) ou bayésienne, pour la seconde quelques sujets particuliers à l'économie et la gestion (arbres de décision, indices, méthodes de sondage, séries chronologiques). Chaque chapitre précise le pré-requis nécessaire dans le tronc commun. Ainsi le test du Khi 2 demande-t-il d'avoir pris connaissance du chapitre « Tests d'hypothèses » de la partie III. On remarquera l'absence de l'Analyse des données considérée certainement par l'auteur comme extérieure à la Statistique.

Le chapitre « Nature de la statistique »

Le premier chapitre intitulé « Nature de la statistique » donne le ton du traité. Trois concepts sont immédiatement mis sur le devant de la scène à partir d'exemples qui se veulent autant de problèmes à résoudre : échantillonnage au hasard, test et régression.

- Le concept d'échantillonnage au hasard est introduit à partir de l'exemple d'un sondage politique comme la réponse au problème d'obtenir une *approximation sans biais* de la proportion d'électeurs favorables à tel ou tel candidat. Notant π la proportion cherchée et P celle dans l'*échantillon aléatoire simple*, l'auteur annonce tout de suite que l'on peut construire un *intervalle de confiance* pour π : $\boxed{\pi = P \pm \text{marge d'échantillonnage}}$ et même que :

> Pour un échantillon aléatoire simple au hasard on peut affirmer avec environ 5 % de chances de se tromper que $\pi = P \pm 1{,}96\sqrt{\frac{P(1-P)}{n}}$ où n est la taille de l'échantillon.

(Wonnacott [1991], p. 6)

Cette introduction permet à l'auteur non seulement de proposer exemples et exercices mais aussi d'expliquer la démarche statistique en opposant induction et déduction :

[36] Au sens de la TAD

> L'un des objectifs essentiels de l'ouvrage est de construire des intervalles de confiance. Un autre objectif, lié au précédent est de tester des hypothèses. […] De façon générale, nous tiendrons le raisonnement suivant : une hypothèse sera rejetée si elle est en contradiction avec l'intervalle de confiance.
>
> Intervalles de confiance et tests d'hypothèses sont deux exemples d'inférence statistique ou raisonnement inductif : on utilise un échantillon observé pour en tirer des résultats s'appliquant à une population inconnue. Dans cet ouvrage, nous étudierons également la théorie des probabilités ou raisonnement déductif – en procédant en sens inverse, d'une population connue à un échantillon inconnu (en fait, c'est par cette étude que l'on commencera car c'est philosophiquement plus simple et cela fournit la base nécessaire à l'inférence statistique).

(Wonnacott [1991], p. 7)

Ainsi l'association statistique-probabilité serait épistémologiquement une association induction-déduction et c'est par souci d'économie didactique qu'il faudrait commencer par les probabilités. Selon l'auteur il y a obligation de passer par les probabilités si l'on veut étudier la Statistique et ce passage doit être fait tout de suite. L'auteur n'en prend pas moins le temps de présenter les problèmes de la Statistique.

- Le concept de test est introduit à propos l'influence supposée d'une habitude hospitalière, celle de « permettre à la jeune maman de jeter un coup d'œil sur son enfant après la délivrance puis à le tenir éloigné d'elle pendant près de 8 heures » alors que d'autres praticiens prônent de « permettre à la maman de passer une heure pleine avec le bébé juste après l'accouchement et de lui autoriser des contacts prolongés durant les trois premiers après-midi ». L'auteur justifie l'intérêt pour répondre à la question posée (y a-t-il influence ou pas ?) d'une comparaison entre deux groupes, le groupe traité et le groupe témoin, dont les individus ont été tirés au hasard – ce qu'il appelle la *randomisation*. L'auteur présente alors la notion d'*expérience aléatoire* et conclut sur cet encadré :

> Des expériences randomisées et aveugles garantissent qu'en moyenne les deux groupes sont initialement égaux et continuent à être traités de façon égale. Ainsi une comparaison valable est possible.

(Wonnacott [1991], p. 13)

Après la recherche d'un paramètre d'une population, la comparaison de populations est le second problème de la Statistique.

- Le concept de régression est introduit à partir de l'idée que – malheureusement ? – les expériences aléatoires sont souvent impossibles et qu'il n'y a pas d'autre moyen que de mener une *étude empirique*. Le problème est alors de réduire les biais dus aux facteurs exogènes non contrôlés. L'auteur annonce tout de suite qu'il existe – heureusement[37] – une réponse générale à ce problème :

> En même temps que les variables étudiées on observe les facteurs exogènes. Quand il est impossible de maintenir les facteurs exogènes constants par l'introduction du hasard, on analyse les données d'une façon qui, en compensation, donne à peu près et autant

[37] Mention de l'auteur ;

que possible les mêmes résultats. La technique utilisée est appelée *analyse de la régression multiple* ou plus simplement, *régression*.

(Wonnacott [1991], p. 18)

L'auteur dit attacher beaucoup d'importance à l'analyse de la régression qui consiste à regarder la façon dont une variable est reliée à une autre variable car, selon lui, elle fonde des règles de décision et autorise des prédictions, toutes contrôlées scientifiquement.

III. Un modèle praxéologique de référence

Au fil de la description des deux manuels, nous avons marqué les choix épistémologiques différents en les attribuant aux différences de niveau de mathématisation voulues par les deux auteurs compte-tenu de leurs propres projets. Il suffit de feuilleter les deux ouvrages pour être convaincu que le traité Saporta privilégie la précision du formalisme algébrique et le traité Wonnacott la force illustrative des exemples et des graphiques.

Mais ces différences se répercutent jusque dans les découpages praxéologiques et notre projet est de leur référer les praxéologies à enseigner de l'EMS vietnamien pour ce qui touche aux liens statistique-probabilités.

Constatons d'abord que les deux traités présentent un même ensemble de problèmes fondamentaux de la Statistique et une même structure générale en trois pôles : recueil, traitement et interprétation des données. C'est dans le pôle traitement que se situent les liens mathématiques entre statistique et probabilités. Ce pôle s'organise lui-même, dans les deux traités, autour de trois composantes : une composante descriptive (ou exploratoire), une composante inférentielle (ou inductive), une composante modélisatrice (ou prédictive).

III.1. Trois problèmes

Concernant la construction mathématique de liens statistique-probabilités, les deux traités mettent en lumière un modèle praxéologique possible à partir des trois problèmes que le traité Saporta énonce[38] :

Le problème de l'incertitude des mesures : comment choisir la vraie valeur d'une certaine grandeur (physique, économique, etc.) à partir des mesures de cette valeur, comment estimer les déviations entre les mesures (valeurs empiriques) et la vraie valeur, comment évaluer le degré de confiance à l'égard de la valeur choisie ? Ce problème est le lieu d'une rencontre possible avec la loi normale.

Le problème de la comparaison des distributions statistiques : comment reconnaître deux distributions statistiques égales, c'est-à-dire, pour reprendre les mots de Brousseau [2005], « deux statistiques qui définissent le même objet, l'acceptent toutes les deux comme modèle » ? Comment estimer la distance entre deux distributions statistiques ? Pour

[38] Nous appelons ces trois problèmes les problèmes candidats à fonder un modèle praxéologique de référence pour les liens mathématiques entre statistiques et probabilités.

Brousseau [2005] la résolution de ce problème « implique une stratégie explicite du test d'hypothèse » ; ce problème est donc le lieu d'une rencontre possible avec l'échantillonnage et le concept de distribution d'échantillonnage.

Le problème de la représentativité d'un échantillon : que peut-on extrapoler d'un échantillon vers la population si la population est inconnue, comment le faire et quel risque prend-on ? Ce problème qui est à l'origine de la théorie des tests est réciproque de celui qui se demande ce que l'on peut savoir d'un échantillon si la population est connue et qui motive la théorie de l'échantillonnage.

On notera que dans cette présentation, ces problèmes sont volontairement intégrés dans une vision unidimensionnelle de la statistique et des probabilités mais qu'une formulation multidimensionnelle est possible et même présente dans les traités.

III.2. Deux praxéologies

Nous avons montré dans les descriptions comparées des manuels que chacun des deux manuels s'appuie sur ces problèmes pour construire ses corpus d'enseignement. Ils le font à travers deux praxéologies :

- **une praxéologie de comparaison** dans laquelle les distributions statistiques sont comparées entre elles ou avec une distribution théorique. Les comparaisons entre elles sont soit graphiques (comparaison des formes) soit numériques (comparaison des résumés) ; les comparaisons avec une distribution théorique sont menées comme les précédentes ou bien par une technique d'ajustement après choix d'un modèle de loi dont les paramètres restent à déterminer.

Dans cette praxéologie, les manuels s'efforcent de rapprocher et d'articuler des objets auxquels la statistique et aux probabilités donnent des significations différentes. Citons : fréquence – probabilité ; histogramme – courbe de densité ; variable statistique – variable aléatoire ; tableau de distribution des fréquences – loi de probabilité d'une variable aléatoire ; densité de fréquence – densité de probabilité ; fonction cumulative des fréquences – fonction de répartition.

Cette praxéologie qu'on pourrait qualifier de régionale combine en fait trois praxéologies locales construites sur trois technologies distinctes :

- une technologie graphique : les distributions de fréquences sont représentées par des diagrammes (très souvent par des histogrammes même dans le cas de variables discrètes si les modalités du caractère étudié sont nombreuses) ; deux distributions sont égales si leurs diagrammes sont superposables ; nous avons remarqué cette technologie dans le traité Wonnacott quand il s'est agi de justifier la Règle de l'Approximation Normale ; quand les diagrammes sont des histogrammes cette superposition s'appuie sur une vision fonctionnelle où interviennent la fonction de répartition ou la courbe de densité et l'aire sous cette courbe..

Les objets de la Statistique candidats à construire des liens avec les probabilités sont dans le cadre de cette praxéologie les histogrammes, les courbes de densité de fréquences et les courbes de cumulation de fréquences.

Nous pouvons ici formuler une hypothèse de recherche que nous travaillerons dans l'institution EMS

HR1 : l'existence d'un point de vue fonctionnel sur les graphiques statistiques est une condition favorable à la construction de liens statistique-probabilités au sein d'une praxéologie de comparaison des distributions statistiques

- une technologie numérique : les distributions sont remplacées par leurs résumés ; cela peut-être une caractéristique de position mais le plus souvent le résumé comporte plusieurs caractéristiques – c'est le cas du box-plot, boîte à moustaches en français - ; deux distributions sont égales si leurs résumés sont les mêmes ; cette technologie est appelée lorsque les deux distributions sont modélisables – pour des raisons qui peuvent tenir au contexte réel que la statistique sont censées étudier – dans une même classe de loi (loi de Poisson à un paramètre, loi binomiale ou loi normale à deux paramètres, loi exponentielle à deux paramètres, etc.) ;

Les objets de la Statistique candidats à construire des liens avec les probabilités sont dans le cadre de cette praxéologie les résumés numériques.

- une technologie probabiliste : l'écart entre une distribution empirique et une distribution théorique qui la modélise est estimé grâce à la distribution d'échantillonnage ; l'acceptation du modèle se fait selon un niveau de confiance fixé a priori et justifié par le contexte de modélisation ; autrement dit la distribution empirique et la distribution théorique sont égales si l'écart entre les deux est attribuable à la fluctuation d'échantillonnage avec une probabilité jugée suffisante ; cette technologie est celle de nombreux tests dont le test du Khi2 ; les éléments théoriques sont ceux de la théorie de l'échantillonnage dont le théorème central-limite.

Les objets de la Statistique candidat à construire des liens avec les probabilités sont dans le cadre de cette praxéologie les distributions d'échantillonnage.

Bien qu'au niveau universitaire une praxéologie soit visible essentiellement par son logos (les éléments théoriques qui rythment le découpage en chapitres de cours), on trouvera en Annexe B une étude de cette praxéologie de comparaison menée sur quelques ouvrages universitaires français.

- **une praxéologie d'échantillonnage** dans laquelle est travaillé le rapport échantillon-population autour de l'idée de représentativité. Nous avons vu dans les deux traités comment cette praxéologie s'organise en deux praxéologies distinctes :

- l'une (qu'on pourrait qualifier de déductive) produit les lois et les caractéristiques des distributions d'échantillonnage lorsque la loi parente est connue ou partiellement connue[39] ; ses éléments technologiques sont produits par une praxéologie des

[39] Par exemple, lorsqu'on sait que la variable aléatoire X dans la population suit une loi normale $N(\mu, \sigma)$ d'écart-type inconnu σ, on montre que la variable aléatoire $T = \frac{\overline{X} - \mu}{s' / \sqrt{n}}$ suit une loi de Student à (n-1) degrés de liberté (noté aussi ddl) où s'est l'écart-type empirique.

probabilités qui peut-être plus ou moins mathématisée mais qui ne peut se passer de la loi normale ; cette praxéologie est un outil pour la praxéologie de comparaison probabiliste ;

- l'autre (qu'on pourrait qualifier d'inductive) produit les règles d'inférence sous la forme d'estimation des caractéristiques et des tests d'hypothèses ; ses éléments technologiques sont les lois des distributions d'échantillonnage produites par la première.

Les objets de la Statistique candidat à construire des liens avec les probabilités sont dans le cadre de cette praxéologie les distributions d'échantillonnage et la notion d'échantillon représentatif.

Nous pouvons ici formuler une deuxième hypothèse de recherche que nous travaillerons dans l'institution EMS :

HR2 : Le problème d'obtenir un échantillon représentatif d'une population mère inaccessible oblige à rencontrer et à construire à partir de statistiques des éléments de mathématisation du hasard au sein d'une praxéologie d'échantillonnage.

Nous entendons maintenant analyser les traces de ces praxéologies dans l'enseignement mathématique secondaire. Les écarts entre les praxéologies à enseigner et les praxéologies de référence qui leur sont « parentes » sont significatifs des contraintes institutionnelles. Cela nous conduira ainsi à transformer les deux hypothèses de recherche en nouvelles questions de recherche pour l'EMS vietnamien.

Partie C

Retour sur l’analyse institutionnelle pour le Vietnam

Chapitre C1
Contrat institutionnel à la statistique

L'objet de la partie C est d'apporter des éléments de réponse à la question (QR2) :

QR2. Quelles sont les conditions de vie actuelles de ces problèmes-candidats et de ces objets-candidats dans EMS Vietnam ?

Nous avons appelé problèmes et objets candidats, des problèmes et des objets mathématiques avec lesquels construire des liens statistique-probabilités dans l'enseignement mathématique secondaire.

Nous procédons en deux phases.

La première phase, qui constitue le chapitre C1, consiste à apprécier les praxéologies à enseigner repérées dans la partie A au regard des praxéologies de référence définies dans la partie B. Quel est le sort institutionnel des problèmes candidats énoncés dans la partie A ? Nous concluons par un essai de caractérisation du contrat institutionnel des liens statistique-probabilités dans EMS vietnamien.

La deuxième phase, qui constitue le chapitre C2, veut apprécier plus finement les rapports institutionnels à certains des objets-candidats. Pour cela nous analysons les réponses des élèves et des enseignants à un questionnaire conçu à partir d'hypothèses sur les contrats didactiques.

I. Rappel des premiers résultats d'analyse institutionnelle [40]

Nous avons mis en évidence le complexe praxéologique suivant dans EMS vietnamien, également valable pour EMS français d'avant la réforme :

- une praxéologie régionale de description d'une série statistique imbriquant les praxéologies locales de description graphique et numériques ; cette praxéologie développe les tâches et les techniques pour simplifier, résumer et synthétiser une série statistique ; elle vise la comparaison de chaque individu à l'ensemble des individus ; les éléments technologiques sont les tableaux, les graphiques et les vecteurs d'indicateurs numériques.

- une praxéologie régionale de comparaison de séries statistiques imbriquant les praxéologies locales de comparaison graphiques ou numériques ; cette praxéologie vise la comparaison entre sous-populations d'une même population ou entre une sous-population et sa population-mère.

Dans l'EMS français d'après la réforme de l'an 2000, nous avons élargi ce complexe en lui adjoignant une praxéologie d'échantillonnage dont les types de tâche sont la production d'échantillons par répétition d'une expérience aléatoire simulée sur ordinateur ou

[40] Ces résultats sont issus de la partie A.

calculatrice, l'observation de la fluctuation d'échantillonnage soit sur une fréquence particulière soit sur un indicateur comme la moyenne, la constatation de la réduction de la fluctuation par augmentation de la taille de l'échantillon.

La comparaison avec l'institution française a ainsi mis en évidence l'absence de praxéologie d'échantillonnage à enseigner au Vietnam mais nous savons que l'idée de répéter une expérience aléatoire est utilisée dans la présentation de l'approche fréquentiste en classe 11. Nous sommes donc conduits à nous poser une double question sur l'enseignement de la Statistique avant la classe 11 :

QR4 : Quel est le rapport institutionnel à la notion d'échantillon en fin de classe 10 ? Quel est le rapport institutionnel à la notion d'échantillon représentatif d'une population ?

Cette question sera étudiée dans le chapitre C2 en nous appuyant sur les exercices 3 et 4 du questionnaire que nous faisons passer à des élèves de classe 10 et à leurs enseignants. Cela nous permettra de reconsidérer l'hypothèse de recherche (HR2), ce que nous ferons dans le cadre de l'ingénierie didactique que nous expérimenterons en classe 11. Nous pensons en effet qu'en l'absence d'une problématisation de la représentativité d'un échantillon à partir d'un tirage aléatoire de cet échantillon, la vie de cette notion dans la statistique enseignée n'autorise pas la construction de liens avec les probabilités.

Intéressons-nous maintenant à la praxéologie de comparaison. En quoi la praxéologie à enseigner au Vietnam se rapproche-t-elle ou s'écarte-t-elle de la praxéologie de référence que nous avons dégagée dans la partie B ?

II. Incomplétude des praxéologies à enseigner

Pour cela nous entrons au cœur de la praxéologie dans l'institution vietnamienne.

Dans les ensembles de manuels, il y a des exercices concernant la comparaison des séries statistiques avec des données discrètes ou avec des données regroupées en classe. En fait nous trouvons que la comparaison des séries statistiques est souvent implicite dans ces exercices car les consignes de ces exercices sont en fait de formuler des remarques sur les données statistiques fournies.

Voici un choix d'exercices concernant la comparaison de données, extrait des ensembles des manuels, les numéros étant ceux des manuels :

Exercice 2

Les deux classes 10C et 10D d'un lycée ont passé un contrôle de littérature de même sujet. Les résultats sont présentés dans deux tableaux de distribution des effectifs suivants :

Notes de classe 10C

Note	5	6	7	8	9	10	Total
Effectifs	3	7	12	14	3	1	40

Notes de classe 10D

Note	6	7	8	9	Total
Effectifs	8	18	10	4	40

a) Calculer les moyennes, les variances, les écarts types des tableaux de distribution des effectifs donnés.

b) Quelle est la classe ayant des résultats les plus réguliers ?

(Manuel de classe 10, l'ensemble élémentaire, p. 128)

Les moyennes sont égales, la comparaison se fait sur les écarts-types. L'étendue paraît pourtant un bon critère de comparaison car peu coûteux et tout autant significatif [41].

Exercice 3

Soient deux tableaux de distribution des effectifs

Les poids des poissons du premier groupe :

Poids (en kg)	[0,6 ; 0,8)	[0,8 ; 1,0)	[1,0 ; 1,2)	[1,2 ; 1,4]	Total
Effectifs	4	6	6	4	20

Les poids des poissons du deuxième groupe :

Poids (en kg)	[0,5 ; 0,7)	[0,7 ; 0,9)	[0,9 ; 1,1)	[1,1 ; 1,3)	[1,3 ; 1,5]	Total
Effectifs	3	4	6	4	3	20

a) Calculer les moyennes des tableaux de distribution des effectifs regroupés en classe.

b) Calculer les variances des tableaux de distribution des effectifs regroupés en classe.

c) Quel est le groupe des poissons ayant des poids les plus réguliers ?

(Manuel de classe 10, l'ensemble élémentaire, p. 128)

Les moyennes sont égales, la comparaison se fait sur les écarts-types. Cette fois-ci les données sont regroupées en classes qui sont différentes.

Exercice 15

On a noté la vitesse (km/h) de 30 voitures passant chacun de deux parcours A et B :

Parcours A :

60 65 70 68 62 75 80 83 82 69 73 75 85 72 67

88 90 85 72 63 75 76 85 84 70 61 60 65 73 76

Parcours B :

76 64 58 82 72 70 68 75 63 67 74 70 79 80 73

75 71 68 72 73 79 80 63 62 71 70 74 69 60 63

[41] Nous rappelons que l'étendue n'est pas un indicateur de dispersion à enseigner au Vietnam.

a) Calculer la moyenne, la médiane, la variance et l'écart type de vitesse des voitures roulant dans chaque parcours.

b) D'après vous, dans quel parcours, les voitures roulent avec la plus sécurité ?

(Manuel de classe 10, l'ensemble avancé, p.179)

Chacun des indicateurs du premier groupe de données est inférieur à l'indicateur du second groupe correspondant. Pourtant, le corrigé ignore la médiane :

En général conduire une voiture sur la route B est plus sûr que sur la route A parce que la vitesse moyenne des voitures sur la route B est inférieure à celle sur la route A et l'écart-type des voitures sur la route B est aussi inférieur à celui sur la route A.

(Livre des enseignants, ensemble avancé, p.234)

La consigne du calcul de la médiane répond-elle à la seule incitation d'ordonner les données ? Dans ce cas on découvre que la vitesse la plus forte dans B est nettement inférieure à sa correspondante dans B. Encore une fois l'élève n'est pas conduit à choisir sa méthode de comparaison, celle-ci étant imposée par les consignes.

Nous repérons deux types de tâche de la praxéologie de comparaison numérique :

- Comparer les moyennes des séries statistiques

- Comparer les variances et écarts-types des séries statistiques si les moyennes sont égales

Les techniques sont les calculs et la comparaison de nombres réels et la technologie est celle de l'interprétation de ces nombres et de leur relation d'ordre dans le contexte.

Exercice 4

« Soient deux séries statistiques suivantes :

Le poids des poissons du premier groupe (en gramme)

645	650	645	644	650	635	650	654
650	650	650	643	650	630	647	650
645	650	645	642	652	635	647	652

Le poids des poissons du deuxième groupe (en gramme)

640	650	645	650	643	645	650	650
640	650	645	650	641	650	650	649
640	645	650	650	644	650	650	645

a) Etablir le tableau de distribution des effectifs et des fréquences de données du premier groupe de poisson en regroupant en classes suivantes : [630 ; 635), [635; 640), [640 ; 645), [645 ; 650), [650 ; 655]

b) Etablir le tableau de distribution des effectifs et des fréquences de données du deuxième groupe de poisson en regroupant en classes suivantes : [638 ; 642), [642; 646), [646 ; 650), [650 ; 654]

c) Décrire le tableau de distribution des fréquences regroupé en classe de la question a) en dressant le diagramme en bâton et la courbe représentative.

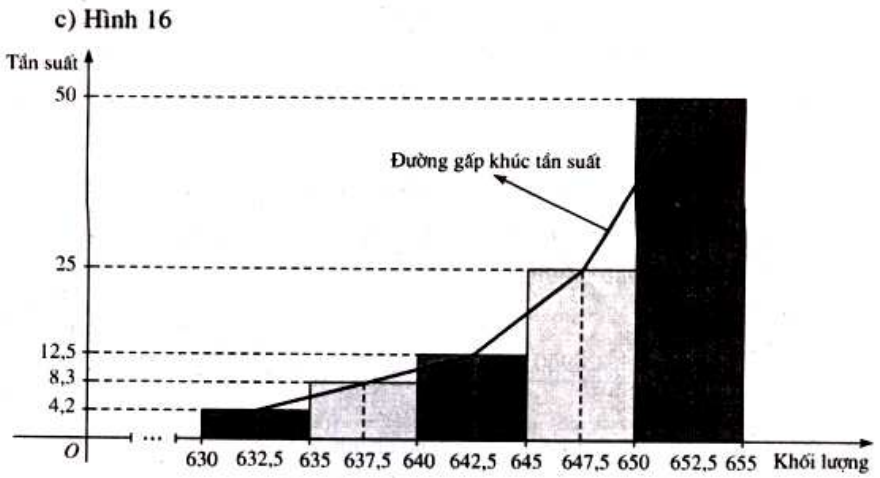
d) Décrire le tableau de distribution des effectifs regroupé en classe de la question b) en dressant le diagramme en bâton et la courbe représentative.

e) Calculer la moyenne, la variance, l'écart type des tableaux des distributions des effectifs et des fréquences effectués.

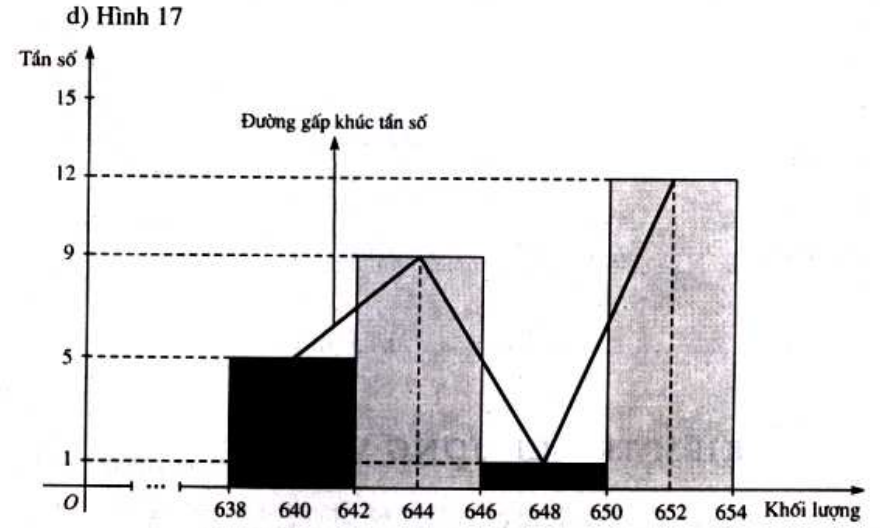
Quel est le groupe de poisson ayant des poids plus réguliers ?

(Manuel de classe 10, l'ensemble élémentaire, p. 129)

Lorsque les données exigent de regrouper en classes, c'est le manuel qui précise des classes et l'élève n'a donc pas la responsabilité de ce fait. Toutes les classes du regroupement de données sont choisies par le manuel de même taille. Il est demandé ici une représentation graphique sous la forme de la courbe polygonale des effectifs. Remarquons que les effectifs totaux sont égaux ce qui évite la nécessité de passer en fréquences pour la comparaison. Voici les graphiques fournis par le manuel dans son corrigé :



Hình 16. Biểu đồ tần suất hình cột và đường gấp khúc tần suất về khối lượng (g) của nhóm cá thứ 1



Hình 17. Biểu đồ tần số hình cột và đường gấp khúc tần số về khối lượng (g) của nhóm cá thứ 2

(Livre des enseignants, ensemble élémentaire, p.137)

Les graphiques peuvent-ils servir la comparaison ? Il nous semble possible de qualifier le second graphique de moins régulier que le premier, dans sa forme. Pourtant, dans sa réponse le manuel ignore le graphique. Il fournit les résultats des calculs de moyennes : elles sont égales. Et comme l'écart-type du groupe 2 est inférieur à celui du groupe 1, le manuel conclut que le groupe le plus régulier est le groupe 2.

Exercice 7

Les tailles de 120 élèves de classe 11 d'un lycée sont données dans le tableau suivant

Nam				Nữ			
175	163	146	150	172	141	155	150
176	162	147	151	172	142	156	154
176	161	149	152	172	142	157	152
177	165	148	153	175	150	158	152
176	169	152	155	175	154	159	153
170	144	168	160	170	150	144	160
170	143	167	160	170	152	144	165
170	142	166	160	170	152	143	159
165	141	174	161	170	160	143	165
166	144	173	162	170	160	140	159
175	156	161	172	175	160	145	168
175	157	162	171	176	161	146	159
176	160	158	170	176	162	147	168
176	164	159	170	175	164	148	159
175	163	160	170	176	165	149	168

1) Avec les classes [135 ; 145), [145 ; 155), [155 ; 165), [165 ; 175), [175 ; 185)

Effectuer le tableau de distribution des fréquences des tailles des garçons.

Effectuer le tableau de distribution des fréquences des tailles des filles.

2) Sur le même repère,

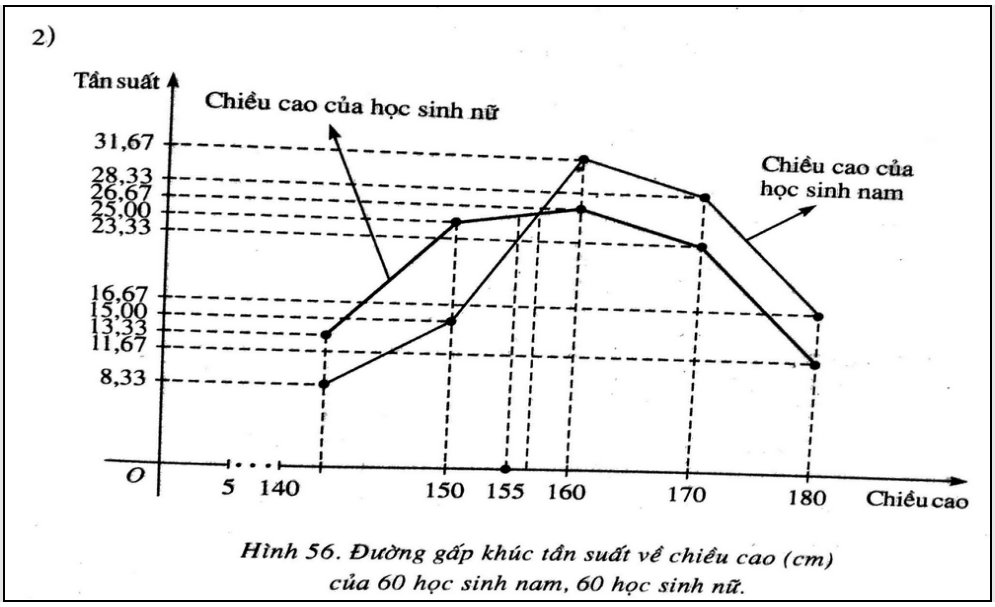
Tracer la courbe polygonale des fréquences pour décrire le tableau de distribution des fréquences regroupées en classes des tailles des garçons.

Tracer la courbe polygonale des fréquences pour décrire le tableau de distribution des fréquences regroupé en classes de tailles des filles.

3) En s'appuyant sur les courbes obtenues dans la question 2, comparer les distributions de tailles des garçons et des filles.

(Livre d'exercices corrigés de classe 10, l'ensemble élémentaire, p.145)

Comme solution, le livre donne les deux courbes polygonales sur le même graphique en marquant point d'intersection de deux courbes :



Hình 56. Đường gấp khúc tần suất về chiều cao (cm) của 60 học sinh nam, 60 học sinh nữ.

(Livre d'exercices corrigés de classe 10, l'ensemble élémentaire, p.167)

Puis il fournit l'explication suivante :

> En observant les deux courbes polygonales, on voit que : parmi les élèves de taille inférieure à 155 cm les filles occupent un pourcentage plus grand que celui des garçons. Et pour la taille supérieure à 160 cm, les garçons occupent un pourcentage plus grand que celui des filles.

(Livre d'exercices corrigés de classe 10, l'ensemble élémentaire, p. 167)

Nous repérons ici un nouveau type de tâche :

- Comparer les deux courbes polygonales

Donc, cet exercice appelle une technique de production graphique pour comparer des distributions. Mais, une question se pose : comment fait-on cette comparaison ? Doit-on s'appuyer sur l'aire sous la courbe ou bien sur la position relative des parties des courbes polygonales ? Ici la technologie mise en place par le manuel et qui justifie selon lui le résultat de la comparaison est celle de la position relative des deux courbes sur un intervalle de valeurs. Nous approfondirons cette analyse dans le chapitre C2 à l'occasion de l'exercice 2 du questionnaire.

C'est seulement à la fin du chapitre que le manuel propose un travail d'enquête – ce sera le seul - en groupes de 3 à 5 élèves qui doivent choisir une classe du lycée :

> 1. Faire une enquête et recueillir des données statistiques dans la classe choisie d'après un caractère choisi par chaque groupe (par exemple le nombre de frères de sœurs dans la famille, la durée du temps consacré pour l'étude de mathématiques à la maison de chaque élève, la taille des élèves de la classe, les notes de mathématiques de chaque élève d'un contrôle précédent,…).
>
> 2. Présenter, analyser, traiter des données statistiques recueillies.
>
> 3. Retirer une conclusion puis faire des propositions.

(Manuel de classe 10, l'ensemble élémentaire, p. 131)

C’est la première et unique fois qu’un type de tâche relevant du groupe 0 apparaît. L’activité proposée va-t-elle se poursuivre dans la comparaison des données des différents groupes ? Le mot comparaison n’est en tout cas pas prononcé.

Mais revenons aux liens statistique-probabilités et à l’hypothèse de recherche HR1 :

HR1 : l’existence d’un point de vue fonctionnel sur les graphiques statistiques est une condition favorable à la construction de liens statistique-probabilités au sein d’une praxéologie de comparaison des distributions statistiques.

Nous avons mieux décrit la praxéologie de comparaison et constaté la présence des technologies numérique et graphique mais l’absence sans surprise de la technologie probabiliste. A propos de la technologie graphique, l’analyse de l’exercice 7 nous conduit à une nouvelle question :

QR5 : au sein de la praxéologie de comparaison dans l’institution vietnamienne, existe-t-il un point de vue fonctionnel sur les graphiques statistiques ?

Nous traiterons cette question à partir des exercices 1 et 2 du questionnaire présenté dans le chapitre C2.

Il est manifeste que la praxéologie de comparaison à enseigner que nous venons de découvrir plus finement est incomplète au sens qu’elle ne dispose pas de modèles par rapport auxquels il faudrait comparer. La praxéologie est enfermée dans une vision descriptive interne où les distributions empiriques sont confrontées deux à deux sans référence à des modèles.

En l’absence de modèles que fourniraient par exemple les probabilités, le point de vue fonctionnel n’est pas absent de la production, de la lecture et de la comparaison des graphiques mais son usage est inadéquat pour construire des liens statistique-probabilités.

III. Le contrat institutionnel à la Statistique dans EMS Vietnam

A l’issue de cette analyse institutionnelle nous pouvons dire que l’enseignement de la Statistique ne rencontre significativement aucun des trois problèmes candidats à construire des liens entre statistique et probabilités. C’est évidemment le cas des problèmes de l’incertitude des mesures et de l’échantillonnage complètement étrangers à cet enseignement. Le seul que l’institution fait venir dans la Statistique est celui de la comparaison des distributions mais la vie institutionnelle des objets communs à la statistique et aux probabilités à travers ce problème est enfermée dans la description sans modélisation.

Ebauchons le contrat institutionnel à la Statistique dans EMS Vietnam :

- la prédominance pour ne pas dire l’exclusivité des praxéologies de description et de comparaison, la seconde s’appuyant sur les technologies de la première pour résoudre des tâches de comparaison entre distributions empiriques

- l’insuffisance pour ne pas dire l’absence complète des questionnements qui motiveraient les types de tâches de l’une et l’autre des deux praxéologies

- une chronogénèse très courte dont le seul moment fort est le passage de distributions discrètes aux distributions en classes

- une topogénèse très rigide construite sur l'application par l'élève de techniques de calculs ou de productions graphiques sur des données pré construites (l'élève n'a ni la responsabilité de la construction des classes ni celle de choisir le type de graphique)

- un isolement complet d'avec les probabilités

Redisons que le contrat institutionnel tel que nous venons de le décrire encore trop sommairement détermine un champ de contraintes et de conditions pour les praxéologies enseignées et constitue une matrice des contrats didactiques pour les objets institutionnellement associés à la Statistique. C'est une invitation à nous approcher encore plus des pratiques réelles d'enseignement ce que nous allons faire dans le chapitre C2.

Chapitre C2
Mises à l'épreuve d'hypothèses sur les contrats didactiques par un questionnaire aux élèves et aux enseignants

Le questionnaire s'adresse aux élèves de classe 10 après enseignement de la statistique et aussi aux enseignants qui enseignent la statistique. IL comporte quatre exercices que nous analysons dans l'ordre de présentation aux élèves. Nous avons placé dans l'annexe C tous les documents portant consignes et énoncés.

I. Exercice 1

Comme nous l'avons déjà dit, l'histogramme (appelé diagramme en colonnes au Vietnam) est un objet central des liens statistique-probabilités. Avant de procéder aux analyses relatives à cet exercice qui met en jeu de tels diagrammes, nous approfondissons l'analyse institutionnelle vietnamienne.

I.1. Quelques éléments d'analyse institutionnelle vietnamienne à propos de l'objet histogramme

Au Vietnam, la statistique est enseignée en classe 7 du collège et en classe 10 du lycée. Le diagramme en segment/rectangle est introduit dès la classe 7 pour décrire une distribution des effectifs sur des valeurs isolées sans regroupement en classes.

En reprenant le tableau ci-après de distribution des effectifs des plants plantés par les élèves de 20 classes d'un collège, le manuel de classe 7 guide ainsi l'élève dans la réalisation d'un diagramme en segments :

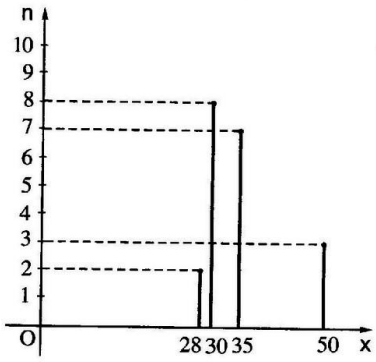
Valeurs (x)	28	30	35	50	
Effectif (n)	2	8	7	3	N = 20

Tracer le repère dont l'axe des abscisses représente les valeurs x, l'axe des ordonnés représente effectif n (la largeur des unités sur deux axes peuvent être différentes)

Préciser des points de coordonnés qui sont des couples des valeurs et son effectif : (28 ; 2) ; (30 ; 8) ;… (remarquer que la valeur est en première et l'effectif est écrit suivant).

Joindre chaque point avec le point sur l'axe des abscisses ayant la même abscisse. Par exemple le point (28 ; 2) est lié avec le point (28 ; 0) ; …

Le graphique venant d'être construit est un diagramme en segments.



(Manuel de classe 7, volume 2, p. 13)

Une particularité de ce graphique est la présence des traits pointillés horizontaux. Quel est le rôle de ces traits pointillés dans ce diagramme en segments ? Il s'agit certainement d'aider l'élève à lire les ordonnées des points en haut de chaque segment comme cela se fait dans les représentations graphiques de fonctions étudiées dans cette même classe. Le choix fait sur l'axe des abscisses (utilisé comme droite réelle avec une graduation arithmétique) de ne pas translater vers la gauche les valeurs comme cela se pratique dans les diagrammes statistiques confirme l'intention du manuel d'activer dans cette présentation du diagramme en segments des pratiques de construction et de lecture rencontrées dans le travail graphique des fonctions numériques.

Ce que nous appellerons dorénavant l'activité graphique fonctionnelle comprenant la construction et la lecture fonctionnelle naît de la définition de la représentation graphique d'une fonction comme un ensemble des points de coordonnés (x; y) dans un repère cartésien. Voici comment le manuel de cette même classe 7 l'introduit :

1. La représentation graphique d'une fonction, qu'est-ce que c'est ?

La fonction y=f(x) est donnée par le tableau suivant :

x	-2	-1	0	0,5	1,5
y	3	2	-1	1	-2

a) Ecrire l'ensemble {(x ; y)} des couples des valeurs correspondant de x et y qui déterminent la fonction précédente.

b) Tracer le repère Oxy et marquer des points de coordonnés qui sont des couples précédents.

L'ensemble des tels points qui représentent des couples de nombres est appelé la représentation graphique de la fonction y = f(x). Donc :

La représentation graphique de la fonction y = f(x) est l'ensemble de tous les points qui représentent des couples des valeurs correspondant (x ; y) sur le plan de repère.

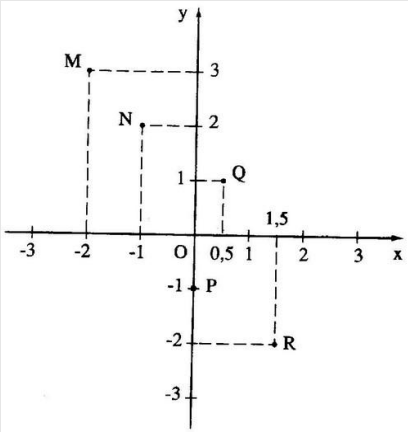
(Manuel de classe 7, volume 1, p. 69)

Ensuite, le manuel donne un exemple pour illustrer la représentation graphique d'une fonction :

Exercice : Tracer la représentation graphique de la fonction donnée dans un repère

Solution : D'abord, on trace le repère Oxy.

La représentation graphique de fonction y = f(x) se compose de 5 point M, N, P, Q comme dans la figure 23.



(Manuel de classe 7, volume 1, p. 70)

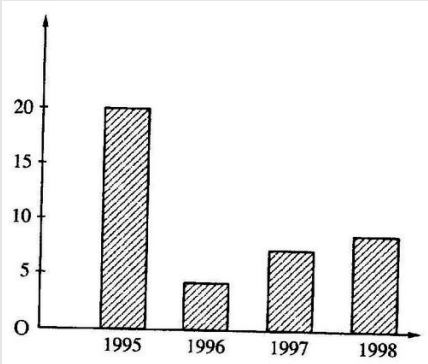
Mais, à la différence de ce qui vient d'être présenté pour les fonctions, le graphique statistique dit diagramme en segments, fabrique des segments verticaux en trait plein. Cette façon de mettre en évidence la hauteur sur la valeur isolée de x pour renforcer le rôle de la hauteur du segment dans l'appréciation visuelle de la distribution devrait s'opposer au tracé d'un trait plein qui rejoint les sommets de ces segments comme on le rencontre dès la classe 7 dans les graphiques fonctionnels où la variable x est considérée comme continue.

Puis, par une remarque, le manuel introduit le diagramme en rectangles comme une autre possibilité pour représenter une distribution des effectifs :

Remarque :

A côté des diagrammes en segments se trouvant dans les documents statistiques ou dans les livres, les journaux, on peut trouver un diagramme comme celui dans la figure 2 (les segments sont remplacés par les rectangles, et parfois, ces rectangles sont tracés de façon adjacente pour faciliter l'observation et la comparaison), c'est le diagramme en rectangles.

Figure 2 représente l'aire des forêts détruites dans notre pays et résume en fonction de l'année de 1995 à 1998 (unité de l'axe des ordonnés : mille hectare).



(Manuel de classe 7, volume 2, p. 14)

Ici les rectangles tracés ont la même largeur et leurs hauteurs sont égales aux effectifs correspondants.

Dans ces deux diagrammes en segments et en rectangles, les valeurs placées sur l'axe des abscisses sont des *valeurs isolées*. La lecture d'aire ne peut avoir lieu dans le diagramme en segments mais elle est possible dans le diagramme en rectangles[42] à la place de la hauteur. Ce passage de la hauteur à l'aire rompt lui aussi avec la lecture fonctionnelle (et ne sera retrouvé que dans le calcul intégral). Comment expliquer sa présence implicite dans le manuel ?

Une première explication est que le manuel de classe 7 veut mettre en avant les représentations graphiques présentes dans « la vie quotidienne » pour favoriser des pratiques pluridisciplinaires de la statistique.

Une deuxième explication est que le diagramme en rectangles sera un point d'appui pour l'histogramme (appelé diagramme en colonnes au Vietnam) que l'élève de classe 7 va retrouver en classe 10.

En classe 10, le regroupement en classes est l'une des nouveautés par rapport au programme de la classe 7. Avec le regroupement en classes, la notion d'histogramme est introduite comme un diagramme pour décrire une distribution des fréquences/effectifs regroupées en classes. Les deux manuels (élémentaire et avancé) ne présentent que des histogrammes à classes égales. Cependant les deux manuels se différencient dans la manière d'introduire la notion.

Dans le manuel élémentaire[43], l'histogramme appelé diagramme en *colonnes* est immédiatement introduit pour décrire une distribution des fréquences de valeurs regroupées en classes *égales* comme où les rectangles sont *adjacents*. Ces classes sont des intervalles adjacents fermés à gauche et ouverts à droite sauf le dernier intervalle qui est tout fermé.

Il faut savoir que la notion d'ensemble est introduite dès le premier chapitre du manuel algèbre de classe 10 avec toutes les codages et les définitions des ensembles numériques $\mathbb{N}$, $\mathbb{N}^*$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, ou des sous-ensembles de $\mathbb{R}$ comme suit :

Intervalle[44] :

$(a ; b) = \{x \in \mathbb{R} / a < x < b\}$

$(a ; + \infty) = \{x \in \mathbb{R} / a < x \}$

$(- \infty; b) = \{x \in \mathbb{R} / x < b\}$

Segment[45] :

$[a ; b] = \{x \in \mathbb{R} / a \leq x \leq b\}$

Demi-intervalle :

[42] Une unité d'aire est équivalente à une unité d'effectif.

[43] Il y a deux ensembles des manuels pour la classe 10 comme pour les classes 11 et 12 : un ensemble « élémentaire » et un ensemble « avancé ». Chaque ensemble se compose de trois livres : le manuel qui comporte cours et exercices non corrigés et un livre d'exercices corrigés tous deux destinés aux élèves, le livre de l'enseignant qui se présente comme un document d'accompagnement didactique et qui comporte des éléments de correction des exercices du manuel.

[44] Cette notion correspond à la notion d'intervalle ouvert en France.

[45] Cette notion correspond à la notion d'intervalle fermé en France.

$[a ; b) = \{x \in \mathbb{R} / a \leq x < b\}$

$(a ; b] = \{ x \in \mathbb{R} / a < x \leq b\}$

$[a ; +\infty) = \{ x \in \mathbb{R} / a \leq x \}$

$(-\infty ; b] = \{ x \in \mathbb{R} / x \leq b\}$

(Manuel de classe 10, l'ensemble élémentaire, p. 17)

L'extrait de l'exercice ci-après est donné pour introduire le diagramme en colonnes en prenant le tableau de distribution des effectifs (dans la deuxième colonne) et des fréquences (dans la troisième colonne) de tailles de 36 élèves regroupés en 4 classes égales :

Chiều cao của 36 học sinh

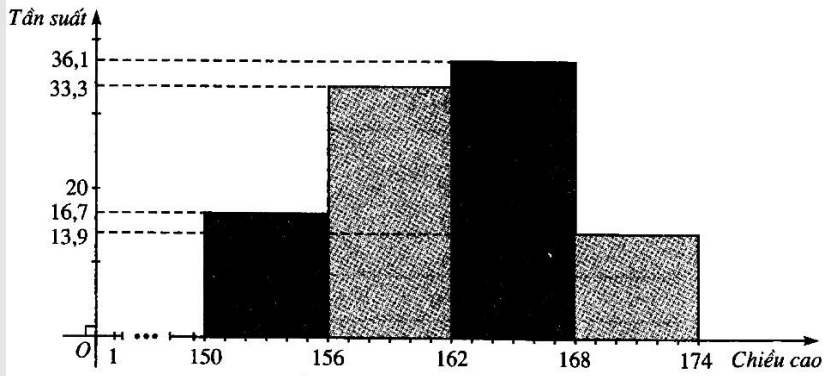
Lớp số đo chiều cao (cm)	Tần số	Tần suất (%)
[150 ; 156)	6	16,7
[156 ; 162)	12	33,3
[162 ; 168)	13	36,1
[168 ; 174]	5	13,9
Cộng	36	100 (%)

Bảng 4

(Distribution des effectifs et des fréquences regroupées en classes de 36 élèves)

(Manuel de classe 10, l'ensemble élémentaire, p. 112)

Ensuite, le diagramme suivant est donné :



Hình 34. Biểu đồ tần suất hình cột về chiều cao (cm) của 36 học sinh.

(Diagramme en colonnes de la taille (cm) de 36 élèves)

(Manuel de classe 10, l'ensemble élémentaire, p. 115)

Signalons ici quelques signes graphiques qui se retrouvent dans la plupart des graphiques statistiques (le diagramme en colonnes, la courbe polygonale) du manuel, du livre d'exercices corrigés, du livre d'enseignant de l'ensemble « élémentaire »[46].

[46] Étonnamment elles ne sont pas reprises dans l'ensemble « avancé ».

- les deux axes gradués portent des flèches, ils sont perpendiculaires en point d'origine O avec le codage de l'angle droit en O (conformité aux pratiques de graphique fonctionnel)

- sur l'axe des abscisses, sont marqués :

+ le point 1 qui est ajouté pour prendre la distance [0,1] comme la norme de longueur sur cet axe.

+ une discontinuité représentée par trois points laquelle est expliquée dans le livre d'enseignant comme un moyen pour réduire le segment [0 ; 150].

+ les petits traits verticaux entre les points 150 et 174 qui aident à compter des petits segments,

+ les extrémités de toutes les classes.

Ces derniers signes sont étrangers aux pratiques fonctionnelles dans l'institution.

Sur l'axe des ordonnés et sur l'axe des abscisses, le manuel écrit respectivement « Fréquence » et « Taille ». On voit aussi que l'axe des abscisses est gradué en unités ; par exemple, on peut lire 6 unités en comptant 6 petits segments entre valeur 150 et valeur 156. Comme dans la classe 7, les traits pointillés joignant les bases en haut des rectangles avec les fréquences sur l'axe des ordonnées activent une lecture fonctionnelle dans ce diagramme en colonnes. Dans ce cas, celui de classes égales, nous savons que l'aire d'une colonne/rectangle est égale à la fréquence de classe correspondante. On remarque aussi que les nombres écrits sur l'axe des ordonnées sont des décimaux supérieurs à 1 et donc la somme est égale à 100 donc il est sous-entendu qu'il faut lire en pourcentage.

Ce manuel propose une même façon pour tracer un diagramme en colonnes des effectifs en faisant cette remarque :

> On peut décrire le tableau de distribution des effectifs regroupés en classes par le diagramme en colonnes des effectifs …
>
> La méthode de tracé est comme celle du diagramme des fréquences … où l'on remplace l'axe des fréquences par l'axe des effectifs.
>
> (Manuel élémentaire classe 10, p. 117)

Mais le manuel n'explicite aucune propriété de ressemblance entre les diagrammes de fréquences et ceux des effectifs. Ce n'est pas le cas du manuel avancé puisque l'on y trouve cette remarque :

> Nous pouvons utiliser le diagramme en colonnes pour décrire une distribution des fréquences regroupées en classes et l'appeler diagramme des fréquences regroupées en classes. Dans ce cas, la hauteur du rectangle est égale à la fréquence [47] (en %).
>
> (Manuel de classe 10, l'ensemble avancé, p. 165)

Avec la même variable statistique (la taille), le manuel avancé de la classe 10 présente successivement deux méthodes de regroupements des données[48] en classes égales d'une même série statistique brute ordonnée :

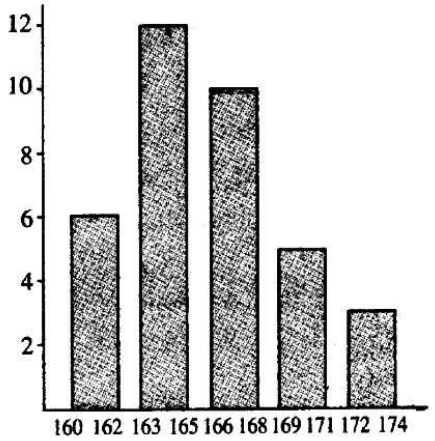
---

[47] Souligné par nous.

• La première méthode : c'est un regroupement en classes *égales*, *fermées* et *disjointes* dont les extrémités sont des valeurs entières. Voici un extrait du manuel :

Lớp	Tần số
[160 ; 162]	6
[163 ; 165]	12
[166 ; 168]	10
[169 ; 171]	5
[172 ; 174]	3
	$N = 36$

Bảng 4



Hình 5.1

(Tableau de distribution des effectifs regroupés en classes

& Diagramme en colonne des effectifs regroupés en classes)

(Manuel de classe 10, l'ensemble avancé, p. 163-165)

Le discours ci-après du manuel est pour construire le diagramme :

> « Tracer deux droites[49] perpendiculaires.
>
> Sur la droite horizontale (utilisée comme l'axe des réels), marquer des intervalles qui précisent des classes en commençant par l'intervalle [160 ; 162] jusqu'à l'intervalle [172 ; 174].
>
> Sur chaque intervalle, tracer une colonne rectangulaire ayant cet intervalle comme sa base et sa hauteur est égale à l'effectif de la classe correspondant.
>
> La figure obtenue est un diagramme en colonnes des effectifs »

(Manuel de classe 10, l'ensemble avancé, p. 165)

Le tableau 4 montre des classes égales qui sont des intervalles tous fermés avec des extrémités en nombres entiers. L'extrémité à droite de cet intervalle est l'entier précédent de l'extrémité à gauche de l'intervalle suivant. La conséquence est que les colonnes/rectangles du diagramme sont disjointes et leurs positions respectent la graduation arithmétique de l'axe des réels. Les rectangles sont de même largeur et leurs hauteurs sont égales aux

---

[48] Les données dans l'exemple est une série statistique ordonnée de mesures de tailles de 36 élèves comme suit :

160	161	161	162	162	162	163	163	163	164	164	164	164
165	165	165	165	165	166	166	166	166	167	167	168	168
168	168	169	169	170	171	171	172	172	174.			

[49] Soulignés par nous.

effectifs des classes. Rien n'est écrit comme légende sur chacun des deux axes. On n'y trouve pas non plus de flèches comme dans les graphiques fonctionnels[50].

En comparaison avec le diagramme en rectangles introduit en classe 7, les deux diagrammes en colonnes des effectifs obligent à des lectures sur l'axe des abscisses qui sont différentes :

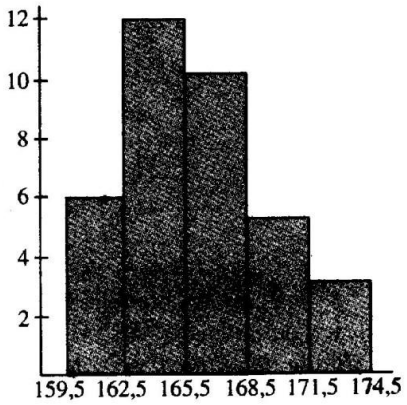
- dans le diagramme en rectangles de la classe 7, les valeurs écrites sur l'axe des abscisses sont des étiquettes données à toutes les valeurs (isolées) de la variable à la manière des valeurs d'une variable qualitative,

- par contre, dans ce diagramme en colonnes, seules les extrémités des classes sont écrites sur l'axe des abscisses et elles encadrent toutes les valeurs possibles de la variable.

• La deuxième méthode : c'est un regroupement en classes *égales*, *adjacentes* et *demi-fermées à gauche*. Voici ce qu'écrit le manuel :

Lớp	Tần số	Tần suất (%)
[159,5 ; 162,5)	6	16,7
[162,5 ; 165,5)	12	33,3
[165,5 ; 168,5)	10	27,8
[168,5 ; 171,5)	5	...
[171,5 ; 174,5)	3	...
	$N = 36$	

Bảng 6



Hình 5.2

(Tableau de distribution des effectifs regroupés en classes

& Diagramme en colonne des effectifs regroupés en classes)

(Manuel de classe 10, l'ensemble avancé, p. 164-165)

Comme dans le manuel élémentaire, les classes adjacentes sont présentées graphiquement par des rectangles jointifs de même largeur et les hauteurs des rectangles représentent les effectifs des classes.

En guise de justification pour différencier les deux diagrammes :

> Pour le regroupement en classes comme celui dans le tableau 6, ce tableau de distribution des effectifs est décrit par le diagramme en colonnes dans la figure 5.2. Dans ce cas, il n'y a pas d'« espace » entre des colonnes.

(Manuel de classe 10, l'ensemble avancé, p. 165)

[50] Dans l'institution vietnamienne, la représentation graphique d'une fonction est toujours présentée sur un plan rapporté à un repère cartésien formé par deux axes gradués perpendiculaires et se croisant en l'origine O. Chaque axe porte une flèche qui indique le sens de l'axe.

Le manuel avancé de classe 10 organise une concurrence entre deux méthodes de représentation graphiques d'une même série statistique de variable continue, elles-mêmes appuyées sur deux tableaux de distribution différents.

Dans la première méthode, la variable statistique continue est discrétisée par intervalles disjoints.

Dans la deuxième méthode, la variable statistique continue est présentée continûment par intervalles adjacents.

Quelle méthode l'enseignant proposera-t-il et quelle méthode l'élève devra-t-il utiliser pour décrire une distribution des effectifs ? En fait dans les exercices proposés, le regroupement en classes n'est pas à charge d'élève puisque le tableau de distribution est fourni avec l'énoncé de l'exercice. Ce tableau détermine le graphique attendu.

Parmi huit exercices proposés dans ce manuel, sept concernent le regroupement en classes égales disjointes dont les extrémités sont des nombres entiers et un seul concerne le regroupement en classes égales adjacentes dont les extrémités sont des nombres décimaux non entiers. Chaque fois, les classes sont données par l'auteur.

Le regroupement en classes inégales, absent du programme, n'est proposé ni dans le manuel ni dans le livre d'exercices. Pourtant, dans une partie du livre d'enseignants de l'ensemble avancé, intitulée « connaissance complémentaire », l'auteur introduit la méthode pour tracer l'histogramme dans le cas où les classes d'un regroupement sont inégales en donnant une formule pour calculer la hauteur du rectangle :

> Connaissance complémentaire
>
> En général, les intervalles qui déterminent les classes sont de même largeur. Mais dans certaines situations, il n'est pas obligé que les classes soient égales. Dans le cas où les classes sont inégales, pour tracer le diagramme en colonnes, la hauteur de la colonne qui représente la classe i-ième est comme suit :
>
> $$h_i = c\frac{n_i}{l_i}$$
>
> où ni est l'effectif de la ième classe, li est largeur de cette classe, c étant une constante choisie par nous-mêmes. Si c égale 1, l'aire d'une colonne qui représente l'effectif égale à l'effectif de cette classe.
>
> S'il n'y a pas d'espace entre les colonnes, le nombre des individus appartenant à l'intervalle [a ; b] (où a < b) est égal approximativement à la somme des aires des colonnes se trouvant entre deux droite x = a et x = b ».

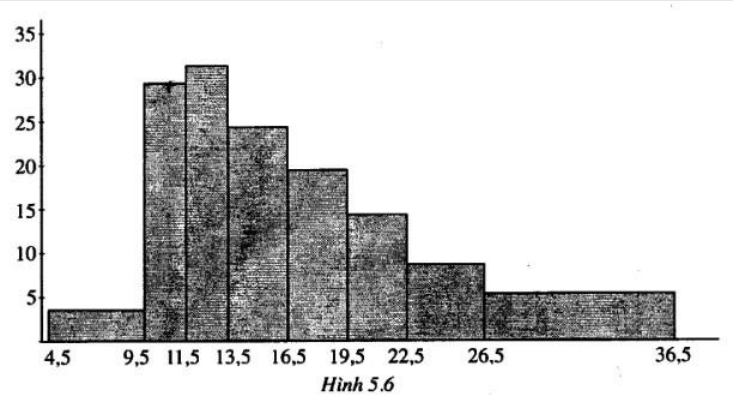
(Livre d'enseignant de classe 10, l'ensemble avancé, p. 224)

Il illustre son propos par un exercice corrigé :

> Exemple :
>
> Les sommes des notes de 4 matières (Maths, Physique, Chimie, Langue anglais) de 400 élèves sont données dans le tableau suivant (la note est sur 10) :

Lớp	Tần số	Độ dài nửa khoảng	Chiều cao cột
[4,5 ; 9,5)	18	5	3,6
[9,5 ; 11,5)	58	2	29
[11,5 ; 13,5)	62	2	31
[13,5 ; 16,5)	72	3	24
[16,5 ; 19,5)	57	3	19
[19,5 ; 22,5)	42	3	14
[22,5 ; 26,5)	36	4	9
[26,5 ; 36,5)	55	10	5,5
	$N = 400$		

Le graphique en colonnes correspondant est celui-ci :



Hình 5.6

(Livre d'enseignants de classe 10, l'ensemble avancé, p. 225)

Pour quelle raison le livre d'enseignants expose-t-il cette méthode de construction de l'histogramme avec des classes inégales ? Pourquoi ne justifie-t-il pas le caractère approximatif du résultat trouvé ? Et que peut faire l'enseignant de ce savoir sur les classes inégales ?

Dans le cas d'une série statistique où il y a des valeurs extrêmes, si on fait le regroupement en classes égales, cela peut causer l'apparition de classes vides ce que le livre d'enseignants recommande d'éviter :

« Méthode de regroupement en classes :

Variable qualitative (et aussi variable ordinale):

On groupe des individus ressemblants ou quasi ressemblants dans une même classe de façon que les classes soient disjointes[51].

Variable quantitative :

Dans ce cas, on suit les pas suivants :

Etape 1 : Préciser des intervalles

[51] Dans ce contexte, les classes disjointes doivent être comprises telle que l'intersection des classes est nulle.

> Chercher l'individu le plus grand noté x_{max} et le plus petit noté x_{min}. Tous les individus appartiennent au segment [a ; b] où $a \leq x_{min}$ et $x_{max} \leq b$; a et b sont choisis tel qu'il y ait aucune classe vide[52] (contenant aucun individu) lors de regroupement en classes.
>
> Etape 2 : Préciser le nombre des classes (noté k) et diviser le segment [a ; b] en k classes (en fait, c'est k-1 demi-intervalles et un segment) disjointes et non vides suivantes :
>
> $$[x_1 ; x_2),\ [x_2 ; x_3),\ldots,\ [x_{2k-1} ; x_k),[x_k ; x_{k+1}] \text{ où } x_1 = a,\ x_{k+1} = b$$
>
> Chaque individu appartient à une et seulement une classe de ces k classes.

(Livre d'enseignants de classe 10, l'ensemble élémentaire, p. 149)

On pourrait ainsi voir dans cette recommandation aux enseignants une raison de procéder à un regroupement en classes inégales à partir d'une série et le livre d'enseignants explique la méthode pour le faire, si nécessaire. Mais dans les tâches proposées aux élèves, on ne rencontre pas de série où un tel regroupement serait nécessaire et le regroupement en classes est systématiquement effectué par le manuel en classes égales.

Comme le dit le Livre d'enseignants, quand les classes sont inégales, l'effectif d'une classe ne se lit pas comme la hauteur sur l'axe des ordonnées mais peut se lire comme l'aire du rectangle dont la largeur est comptée dans l'unité qui a été choisie dans la construction de l'histogramme.

En conséquence, la présence des classes inégales apporte dans l'histogramme une lecture d'aire des rectangles qui se substitue avantageusement à la lecture fonctionnelle de la hauteur sur l'axe des ordonnées qui doit être suivie d'une multiplication par la largeur. Ce que l'on lit sur l'axe des ordonnées ce n'est pas l'effectif mais la densité des effectifs. Cette dernière notion n'est pas présentée par le manuel et n'est certainement pas un objet d'enseignement dans l'institution vietnamienne.

Quand le regroupement se fait à classes égales, l'effectif peut être lu aussi bien comme la hauteur sur l'axe des ordonnées que comme l'aire du rectangle dont la largeur est unitaire. Il est donc possible même dans ce cas de calculer les effectifs entre deux valeurs a et b quelconques par un calcul d'aire. Une telle tâche n'existe pas dans le manuel. Les enseignants la proposent-ils aux élèves ? Si oui, comment justifient-ils l'approximation dont parle le manuel.

Conclusion sous forme de deux hypothèses de recherche

Hypothèse 1 (H1)

Dans l'institution vietnamienne, il existe un contrat didactique à propos de l'objet d'histogramme que l'on peut énoncer ainsi :

> L'élève n'a pas à sa charge de faire le regroupement en classes d'une série statistique donnée. Le regroupement des données statistiques en classes est toujours fait en classes égales et il est à la charge de l'enseignant. Selon qu'il sera en classes disjointes

[52] Soulignés par nous.

ou non, le graphique attendu est un diagramme en colonnes disjointes ou un diagramme en colonnes adjacentes.

Hypothèse 2 (H2)

Dans l'institution vietnamienne actuelle, la notion d'aire d'un rectangle n'est pas mobilisable dans les tâches mathématiques attachées à l'histogramme.

Les deux exercices (1 et 2) du questionnaire sont conçus pour tester ces hypothèses de recherche.

I.2. Présentation de l'exercice

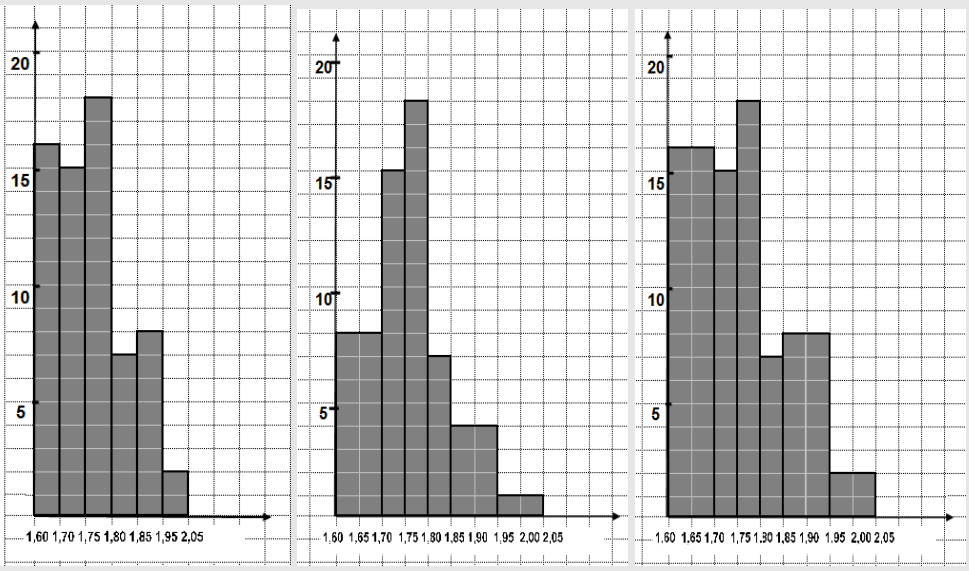
Le questionnaire s'adresse aux élèves de classe 10 après enseignement de la statistique et aussi aux enseignants qui enseignent la statistique.

Le sujet ci-après est la partie commune distribuée en feuille à chaque élève et à chaque enseignant :

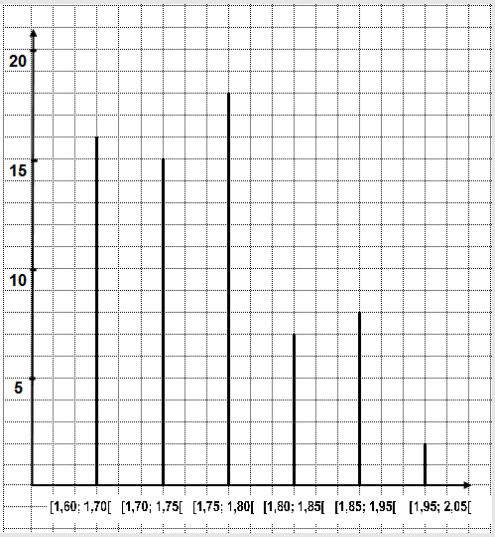
On a mesuré la taille des 66 joueurs d'un club de basket-ball. On a constitué le tableau suivant :

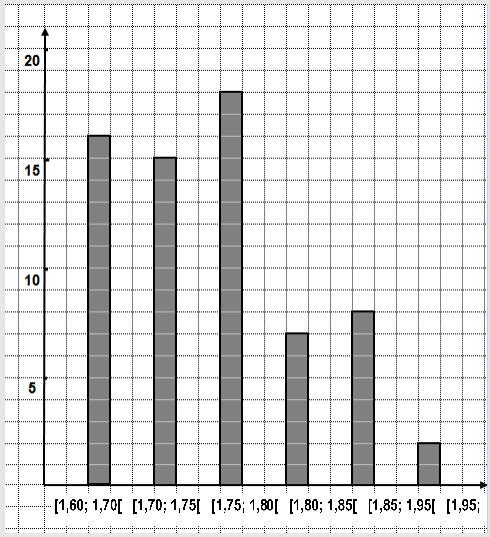
Classe de taille (mètre)	Nombre de joueurs
[1,60 ; 1,70[	16
[1,70 ; 1,75[	15
[1,75 ; 1,80[	18
[1,80 ; 1,85[	7
[1,85 ; 1,95[	8
[1,95 ; 2,05[	2

On a demandé à élèves d'un lycée vietnamien de représenter graphiquement les données du tableau. Voici cinq représentations graphiques qu'ils ont réalisées :



Graphique 1 Graphique 2 Graphique 3





Graphique 4 Graphique 5

1. Parmi les représentations graphiques proposées, lesquelles considérez-vous comme correctes ?
2. Parmi les représentations graphiques proposées, lesquelles considérez-vous comme incorrectes ? Justifiez vos réponses.
3. Écrire la légende sur l'axe horizontal et sur l'axe vertical pour chaque représentation graphique acceptée comme correcte.

- A l'élève il est proposé de résoudre individuellement l'exercice 1 dans une feuille de réponse avec la demande de lire attentivement l'exercice et de répondre aux questions proposées.

Voici un ensemble de 4 exercices. Lisez attentivement les questions et répondez aux questions, s'il vous plait.

Merci de votre participation.

Vos nom et prénom : .....

Classe : .....

Lycée : .....

Exercice 1 :

1) .....
.....
2) .....
.....
3) .....

- A l'enseignant il est demandé de donner ses réponses sur une autre feuille avec des questions préparées pour lui comme suit :

Voici 4 exercices à propos de la statistique enseignée en classe 10. Pour chacun d'eux, nous vous prions de répondre à la question suivante :

Proposeriez-vous un tel exercice en classe 10 ? Si oui, quelle réponse attendez-vous d'un élève de classe 10 ? Sinon, comment le modifieriez-vous ?

Merci de votre participation au questionnaire.

Fiche de réponses :

Votre nom et prénom : .....

Vous êtes enseignant(e) du lycée : .....

Avez-vous enseigné à l'aide du programme expérimental entre les années 2004 et 2006 ? .....

Exercice 1 :

Proposeriez-vous cet exercice en classe 10 ? Oui Non

La réponse que vous attendez d'un élève de classe 10 :.....
.....
Votre modification éventuelle de l'exercice :.....
.....

I.3. Analyse a priori

Quels sont les choix qui organisent le questionnaire ?

Les variables de situation

1) Le tableau de distribution est commun à tous les graphiques. Les données sont d'emblée regroupées en classes inégales (contre égales) qui sont représentées graphiquement sur l'axe des abscisses par des intervalles soit adjacents soit disjoints.

Classe de taille (mètre)	Nombre de joueurs
[1,60 ; 1,70[	16
[1,70 ; 1,75[	15
[1,75 ; 1,80[	18
[1,80 ; 1,85[	7
[1,85 ; 1,95[	8
[1,95 ; 2,05[	2

Tableau 10. Distribution commun

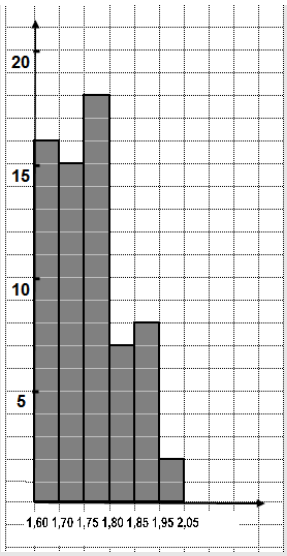
2) Tous les graphiques sont « fonctionnels » au sens qu'ils utilisent un repère cartésien (contre les graphiques circulaires).

3) Il n'y a pas de légende ni sur l'axe des abscisses ni sur l'axe des ordonnées (contre avec légende).

4) Le graphique est tracé sur un fond quadrillé pour favoriser la lecture d'aire mais sans précision d'unité d'aire (contre feuille vide)

Les différents graphiques

Le graphique 1 est un diagramme en rectangles adjacents dont les hauteurs sont égales aux effectifs des classes correspondantes mais dont les largeurs sont toutes égales (conformément à l'histogramme de l'institution vietnamienne). Les largeurs des rectangles ne sont donc pas proportionnelles aux largeurs des classes dans le tableau. Pourtant, l'aire[53] de chaque rectangle est égale à l'effectif de la classe. Ainsi ce diagramme respecte la lecture fonctionnelle d'un diagramme en rectangles et la lecture par l'aire d'un histogramme.

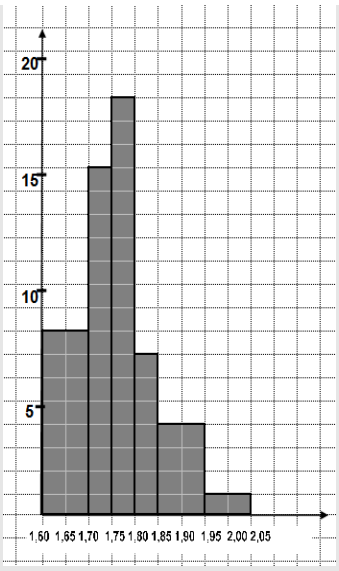


Graphique 1

- *Le graphique 2* est un diagramme en rectangles adjacents. Dans ce diagramme, l'aire de chaque rectangle est égale à l'effectif de classe correspondant et la proportionnalité des largeurs des classes inégales est respectée sur l'axe des abscisses. La rupture du contrat se trouve dans le fait que la hauteur du rectangle ne peut pas être lue comme l'effectif de classe à cause des classes inégales. C'est-à-dire il n'y a pas de lecture fonctionnelle sur l'axe des ordonnées sauf à attacher à l'axe des ordonnées la densité d'effectifs au lieu des effectifs eux-mêmes (la densité d'effectifs est le quotient de l'effectif par la largeur de classe exprimée en unités de classe[54]. Ce graphique est conforme à la définition de l'histogramme présente dans les manuels universitaires.

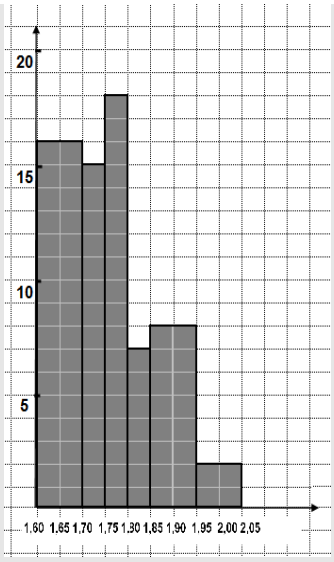
[53] Dans ce contexte, l'unité d'aire est l'aire d'un carré du quadrillage.

[54] Dans cet exercice, la largeur de classe-unité est égale à 0,5.



Graphique 2

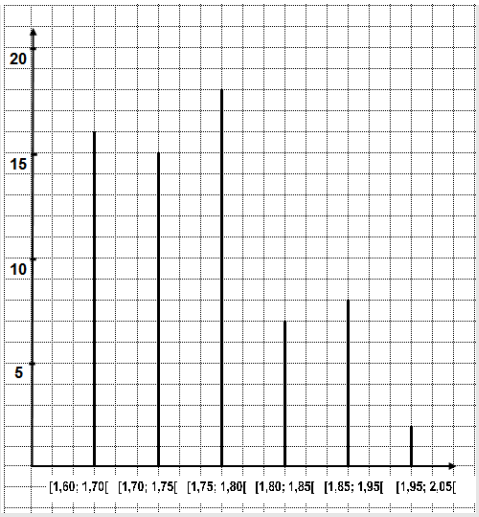
- *Le graphique 3* est aussi un diagramme en rectangles adjacents. Sur l’axe des ordonnées, on peut lire l’effectif d’une classe par la hauteur du rectangle et sur l’axe des abscisses, la proportionnalité des largeurs des classes est conservée graphiquement. C’est le graphique le plus conforme à l’institution vietnamienne. Pourtant, il ne respecte pas la propriété attribuée à l’histogramme de proportionnalité des aires et des effectifs.



Graphique 3

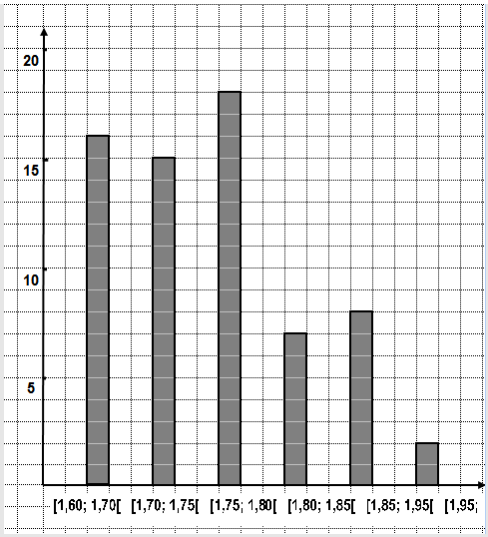
- *Le graphique 4* est un diagramme en segments qui conserve la lecture fonctionnelle des effectifs sur l’axe des ordonnées par les hauteurs des segments. Les valeurs sur l’axe des abscisses sont des valeurs isolées comme dans le cas d’un caractère qualitatif ou d’un caractère quantitatif discret. Il ressemble formellement au diagramme en segments introduit en classe 7 et devient donc un diagramme conforme à l’institution vietnamienne de classe 7.

Mais la substitution des rectangles par des segments fait disparaître l'aire comme quantité graphiquement évaluable.



Graphique 4

- *Le graphique 5* est un diagramme en colonnes/rectangles disjoints qui est proche le graphique 4 mais avec des « segments épais », c'est-à-dire des rectangles de même largeur que celle de classe-unité. L'aire existe donc dans ce diagramme et elle est proportionnelle aux effectifs. Les valeurs sur l'axe des abscisses sont des valeurs isolées du caractère qualitatif. Ce graphique 5 ressemble formellement au diagramme en rectangles introduit en classe 7 et au premier diagramme du manuel avancé de la classe 10.



Graphique 5

Ainsi les cinq diagrammes proposés varient sur les valeurs différentes de la variable didactique **V** qu'est le couple (**tableau, graphique**).

Paramètres du tableau :

En fait, le tableau des données peut être :

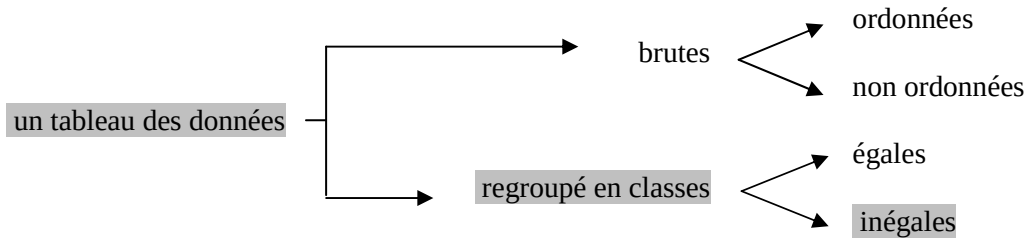


Schéma 2. Arbre des possibilités du tableau des données

Dans notre choix, ce paramètre de tableau est fixe. C'est que le tableau représente une distribution des effectifs regroupés en classes *inégales*.

Deux éléments importants dans ce choix sur le tableau sont :

- le regroupement en *classes* qui est un choix conforme à l'institution vietnamienne,

- l'*inégalité* des classes qui est une rupture du contrat didactique relatif au tableau de données.

Le tableau 1 suivant présente les possibilités des choix pour le tableau des données (les cases en fond gris font partie de notre choix) :

Tableau des données		Conformité à l'institution vietnamienne
Tableau des données brutes	ordonnées	oui
	non-ordonnées	oui
Tableau des données regroupées en classes	égales	oui
	inégales	non

Tableau 11. Tableau des possibilités des choix sur le tableau des données

Paramètres du graphique : nous distinguons la représentation graphique des classes du tableau des données sur l'axe des abscisses, l'interprétation des hauteurs des rectangles sur l'axe des ordonnés et la présence ou l'absence de légendes sur les axes. Concrètement, nous faisons le choix sur les valeurs suivantes :

- Sur l'axe des abscisses les classes sont représentées graphiquement par :

  + les intervalles ou les valeurs isolées Si intervalles, alors :

  + les intervalles sont égaux ou inégaux (c'est-à-dire leurs largeurs sont tous égales ou bien elles sont proportionnelles à celles des classes du tableau donné),

  + les intervalles sont adjacents ou disjoints

- Sur l'axe des ordonnées :

  + la hauteur d'un rectangle est égale à l'effectif de classe ou non,

  + la hauteur d'un rectangle est égale à la densité d'effectif de la classe ou non

+ la légende non écrite sur l'axe des ordonnées est « Effectif » ou « Densité d'effectif ».

- Pas de légendes

Chaque association de valeurs données aux deux premiers paramètres dicte l'égalité ou la non-égalité entre l'effectif de la classe et l'aire du rectangle (cette comparaison se fait seulement quand l'aire existe dans le graphique).

Le tableau 3 présente tous les graphiques possibles (les cases en fond gris font partie de notre choix) :

Axe des abscisses			Axe des ordonnées	Aire-Effectif	Légende sur l'axe des ordonnées	Graphique
Intervalles	égaux	adjacents	hauteur = effectif	aire = effectif	Effectif	G1
			hauteur = densité d'effectif	aire ≠ effectif		
		disjoints	hauteur = effectif	aire = effectif	Effectif	G5
			hauteur = densité d'effectif	aire ≠ effectif		
	inégaux	adjacents	hauteur = effectif	aire ≠ effectif	Effectif	G3
			hauteur = densité d'effectif	aire = effectif	Densité d'effectif	G2
		disjoints	hauteur = effectif	aire ≠ effectif		
			hauteur = densité d'effectif	aire = effectif		
Valeurs isolées		disjoints	hauteur = effectif	pas d'aire	Effectif	G4
			hauteur ≠ effectif	pas d'aire		

Tableau 12. Arbre des combinaisons possibles axe des abscisses-axe des ordonnées

Les cases vides de la dernière colonne expriment des graphiques possibles mais non utilisés dans le questionnaire.

Voici le tableau 4, réduction du tableau 3 aux graphiques choisis :

Axe des abscisses			Axe des ordonnées	Aire-Effectif	Légende sur l'axe des ordonnées	Graphique
Intervalles	égaux	adjacents	hauteur = effectif	aire = effectif	Effectif	G1
		disjoints	hauteur = effectif	aire = effectif	Effectif	G5
	inégaux	adjacents	hauteur = effectif	aire ≠ effectif	Effectif	G3
			hauteur = densité	aire = effectif	Densité d'effectif	G2
Valeurs isolées			hauteur = effectif	pas d'aire	Effectif	G4

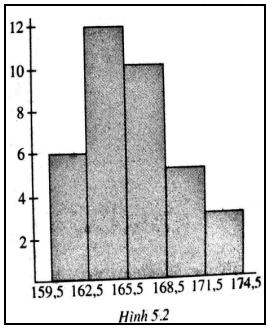
Tableau 13. Tableau réduit des choix sur l'axe des abscisses-axe des ordonnées

Le tableau 5 ci-après associe le tableau 1 et tableau 3 en ajoutant les possibilités induites par un regroupement en classes égales. Ce tableau résume nos choix pour la variable V1(tableau, graphique) et compare nos choix avec les attentes de l'institution vietnamienne (partie grisée représente nos choix).

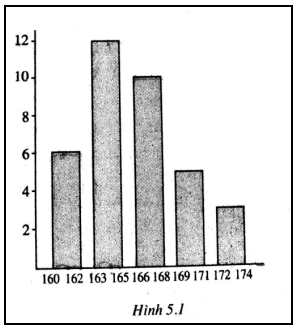
Graphique / Tableau	Axe des abscisses			Axe des ordonnées	Aire-Effectif	Légende sur l'axe des ordonnées	Graphique
Tableau des données regroupées en classes égales	Intervalles	égaux	adjacents	hauteur = densité d'effectif ou = effectif	aire = effectif	Densité d'effectif Ou effectif	g1
			disjoints	hauteur = densité d'effectif ou = effectif	aire = effectif	Densité d'effectif Ou effectif	g5
	Valeurs isolées		disjoints	hauteur = effectif	pas d'aire	Effectif	g4
Tableau des données regroupées en classes inégales	Intervalles	égaux	adjacents	hauteur = effectif	aire = effectif	Effectif	G1
			disjoints	hauteur = effectif	aire = effectif	Effectif	G5
		inégaux	adjacents	hauteur = effectif	aire ≠ effectif	Effectif	G3
				hauteur = densité d'effectif	aire = effectif	Densité d'effectif	G2
	Valeurs isolées		disjoints	hauteur = effectif	pas d'aire	Effectif	G4

Tableau 14. Tableau réduit des choix sur l'axe des abscisses-axe des ordonnées

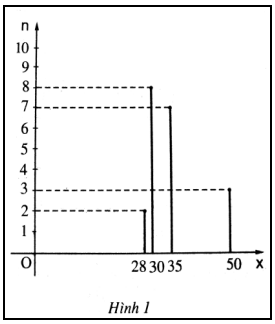
Dans ce tableau, gi est pour décrire le graphique utilisé dans l'institution vietnamienne qui ressemble Gi, c'est-à-dire :



g1
(méthode 2, Manuel de classe 10, l'ensemble avancé)



g5
(méthode 1, Manuel de classe 10, l'ensemble avancé)



g4
(manuel de classe 7)

Stratégies

$S_{fonction}$: Stratégie « *lecture fonctionnelle* » (stratégie analytique)

La lecture fonctionnelle dans le graphique fait comparer la hauteur d'un rectangle et l'effectif de la classe correspondant pour vérifier la proportionnalité entre eux. Pour les graphiques satisfaisant cette égalité, les trois réponses prévisibles sont comme suit :

- le graphique G3 est correct[55] car on voit en plus la proportionnalité des largeurs des intervalles (intervalles inégaux) et celles des classes (la réponse la plus conforme à l'institution vietnamienne).

- les graphiques G1 et/ou G5 seront pris comme corrects en considérant les intervalles égaux comme représentant des classes égales,

+ si on choisit en plus la représentation adjacente des intervalles, c'est le graphique G1 qui est correct,

+ par contre, si on accepte la représentation disjointe des intervalles, le graphique G5 qui est correct.

- le graphique G4 est correct si le segment est considéré comme un rectangle aplati (aire nulle)

Le schéma suivant résume les réponses possibles d'après la $S_{fonction}$:

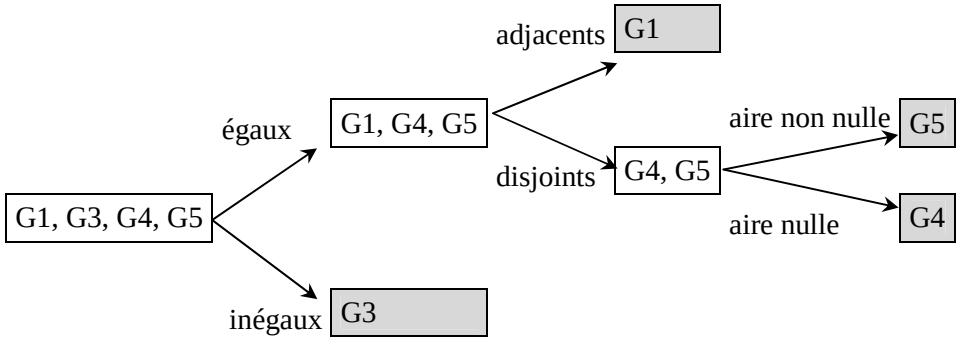


Schéma 3. Discrétisation des graphiques en utilisant $S_{fonction}$

S_{air} : Stratégie « *aire* » (stratégie géométrique)

La lecture d'aire dans le graphique fait comparer l'effectif d'une classe et l'aire du rectangle correspondant pour vérifier l'égalité ou la proportionnalité selon l'unité d'aire choisie.

Les graphiques G1, G2, G5 satisfont cette égalité mais les graphiques choisis pour les correctes dépendent à la représentation graphique des largeurs des intervalles :

[55] Il faut comprendre que ce graphique est correct dans le cadre de cette stratégie.

- la proportionnalité des largeurs des intervalles et celles des classes dans le graphique fait choisir le graphique G2 comme correct. D'après la statistique descriptive dans livre universitaire, le graphique G2 est l'histogramme correct,

- l'égalité des intervalles fait choisir le graphique G1 et G5,

+ si on considère la représentation adjacente des intervalles, le graphique G1 est correct,

+ inversement, si on considère la représentation disjointe des intervalles, le graphique G1 est correct.

Le schéma suivant résume les réponses possibles d'après la S_{air} :

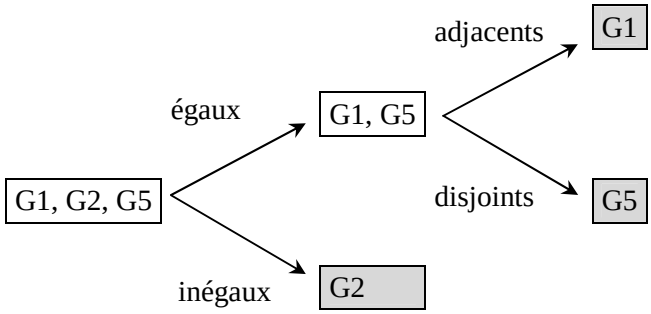
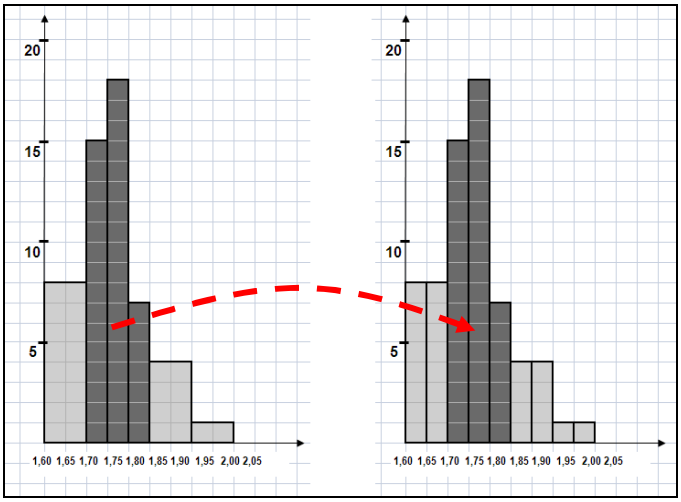


Schéma 4. Discrétisation des graphiques en utilisant S_{air}

S_{classe} : Stratégie « *classes égales* »

Le contrat didactique sur des classes égales fait accepter seulement les graphiques avec des classes égales ou fait ramener à classes égales par un découpage :

- *Découper des classes inégales en sous-classes égales* : accepter le graphique G2 et refuser le graphique G3 :



Graphique 2 Graphique 2'

Sur le graphique G2, par un découpage binaire, le premier rectangle de hauteur 8 qui représente la classe [1,60 ; 1,70[est divisé en deux rectangles égaux de même hauteur 8 qui

correspondent à deux sous-classes égales [1,60 ; 1,65[et [1,65 ; 1,70[. Maintenant, comme ces deux sous-classes sont égales, la lecture hauteur-effectif donne l'effectif 8 pour chacune. Puis, par une réunion de ces deux sous-classes en additionnant leurs effectifs, on obtient la classe initiale et la somme 16 qui est celui de classe initiale et il satisfait l'information dans le tableau donné. Ensuite, les deux derniers rectangles sont divisés de même façon, et finalement, le graphique G2 devient G2'. On voit que ce graphique G2' correspond au tableau de distribution des effectifs (Tableau 4) dont le regroupement est *transformé en 10 classes égales* (les lignes en fond gris et en lettres foncés sont pour les nouvelles sous-classes):

Classe de taille (mètre)	Nombre de joueurs
[1,60 ; 1,65[	8
[1,65 ; 1,70[	8
[1,70 ; 1,75[	15
[1,75 ; 1,80[	18
[1,80 ; 1,85[	7
[1,85 ; 1,90[	4
[1,90 ; 1,95[	4
[1,95 ; 2,00[	1
[2,00 ; 2,05[	1

Tableau 15. Distribution des effectifs en classes égales

Ce découpage conduit le lecteur à accepter le graphique G2.

La même technique s'applique au premier rectangle de hauteur 16 du graphique G3 et le transforme en deux rectangles égaux de même hauteur 16. Par conséquence, l'effectif de chacune des sous-classes [1,60 ; 1,65[et [1,65 ; 1,70[est 16. Comme la somme de ces deux effectifs fait 32 qui ne satisfait pas le tableau donné, on rejette le graphique G3.

- Utiliser des intervalles égaux pour représenter graphiquement des classes inégales : accepter les graphiques G1, G4 et G5

+ si on choisit en plus la représentation adjacente des intervalles, c'est le graphique G1 qui est correct,

+ par contre, si on accepte la représentation disjointe des intervalles, le graphique G5 qui est correct,

+ le graphique G4 est correct si le segment est considéré comme un rectangle aplati (aire nulle).

Le schéma suivant résume les réponses possibles d'après la S_{classe}

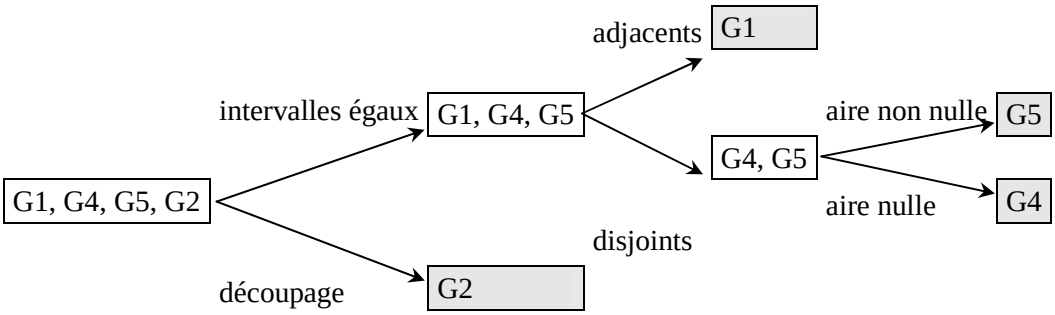


Schéma 5. Discrétisation des graphiques en utilisant S_{classe}

En résumé voici un diagramme de Venn qui présente la répartition des graphiques d'après les stratégies possibles :

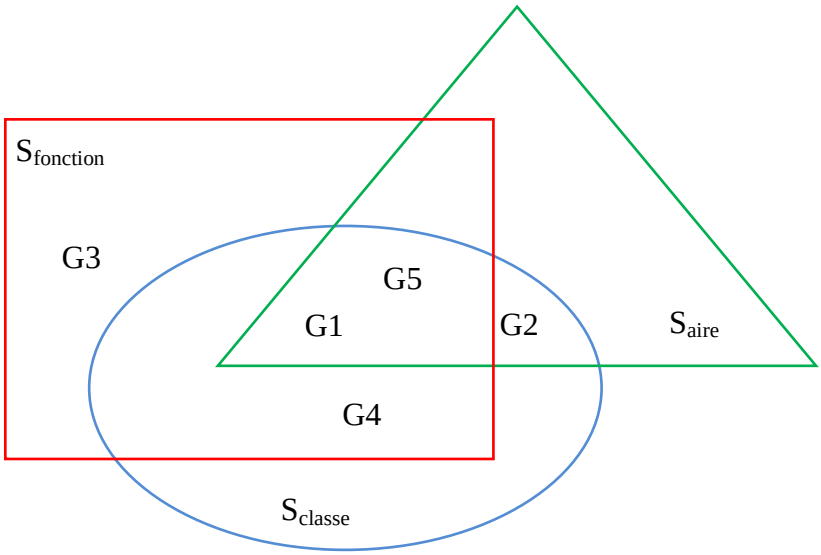


Figure 1. Diagramme de Venn des trois stratégies

Rectangle : $S_{fonction}$ Triangle : S_{air} Ovale : S_{classe}

Couples des graphiques à observer :

• Couple (G2, G3) :

G2 ne s'appuie pas sur une lecture fonctionnelle (par contre, il s'appuie sur une lecture d'aire) tandis que G3 s'appuie seulement sur une lecture fonctionnelle. Nous prévoyons que ce couple ne peut pas être choisi.

• Couple (G1, G3) :

G1 et G3 s'appuient sur une lecture fonctionnelle avec des intervalles adjacents mais diffèrent sur la proportionnalité des largeurs des intervalles et des classes.

• Couple (G1, G5) :

Ces deux graphiques sont acceptables par les trois stratégies et ne diffèrent que par la disjonction des rectangles.

I.4. Analyse a posteriori

Le questionnaire a été proposé aux élèves et aux enseignants, concrètement :

- les élèves de classes 10 à Ho Chi Minh ville entre mars et mais 2008 (Viêt-Nam). C'est juste après avoir appris la statistique qui a eu lieu au début de deuxième semestre dans l'année scolaire,

- les enseignants des lycées à Ho Chi Minh ville et en province qui ont donné des cours de statistique, c'est-à-dire qui ont enseigné dans des classes 10.

	Nom de l'établissement/Région	Effectif	Total
Élève	2 classes 10 (seconde) du lycée Le Hong Phong de HCM ville	57	208
	2 classes 10 (seconde) du lycée Vo Thi Sau de HCM ville	77	
	1 classe 10 (seconde) du lycée Mac Dinh Chi de HCM ville	42	
	1 classe 10 (seconde) du lycée Nguyen Huu Cau de HCM ville	32	
Enseignant	Ho Chi Minh ville	40	72
	Provinces (Dong Nai, Phan Thiet, Tien Giang)	32	

Tableau 16 : Distribution des effectifs des participants à l'expérimentation

Les conditions de passation ont été les suivantes :

• Elève :

Chacun reçoit un questionnaire[56] composant 4 exercices et un ensemble[57] de 5 feuilles de réponse ayant les versos vierges pour brouillon.

On a demandé à l'élève de résoudre individuellement les exercices, de donner des réponses écrites sur les feuilles fournies et d'effectuer si nécessaire des calculs sur les versos.

La passation a duré une heure et demi.

Chaque séance donne lieu à une observation et toutes les feuilles sont récupérées.

• Enseignant :

L'enseignant reçoit le même questionnaire que l'élève ainsi qu'un ensemble de 4 feuilles de réponses avec des questions qui lui sont réservées. On attend des réponses écrites de sa part

[56] Cette partie est commune pour le début dans l'analyse a posteriori de l'ensemble de 4 questions du questionnaire.

[57] Une feuille par question, sauf question 3, on distribue deux feuilles dont la deuxième est une feuille quadrillé millimétrique pour tracer le graphique.

les fiches nous étant retournées quelque temps après (il n'y pas, pour lui, de durée fixe pour répondre au questionnaire).

Réponses des enseignants

Dans le tableau à double entrée suivant, on lit la répartition du nombre des enseignants qui participent au questionnaire. En ligne, c'est le nombre des enseignants qui acceptent de proposer ou non cet exercice à l'élève ; en colonne, c'est le nombre de ceux qui donnent ou ne donnent pas leurs choix des graphiques comme corrects et incorrects.

Choix des graphiques / Proposer l'exercice 1 à l'élève	Oui	Non	Total
Oui	22	17	39
Non	12	18	30
Sans réponse		3	3
Total	34	38	72

Tableau 17. Répartition des enseignants

Le tableau des choix observés dont les cases marquées d'un « X » sont pour les graphiques déclarés corrects et dont les cases vides sont pour les graphiques déclarés incorrects :

G1	G2	G3	G4	G5	Nombre d'enseignants	Caractéristique du choix
					3	Aucun graphique n'est correct
			X		1	G4 est le seul correct
		X			23	G3 est le seul correct
		X		X	2	Couple (G3, G5) comme deux corrects
		X	X		1	Couple (G3, G4) comme deux corrects
X					3	G1 est le seul correct
X				X	1	Couple (G1, G5) comme deux corrects
X		X			2	Couple (G1, G3) comme deux corrects

Tableau 18. Combinaisons observées des graphiques choisis comme corrects de 34 enseignants

Le tableau ci-après est celui de distribution des effectifs des graphiques choisis comme corrects[58] :

Graphique	G1	G2	G3	G4	G5
Nombre de choix de l'enseignant	6	0	28	2	3

Tableau 19. Distribution des effectifs des graphiques corrects donnés par enseignant

Un autre tableau à double entrée[59] (tableau), symétrique par la diagonale, est pour résumer les nombres de choix des graphiques avec la relation relative des couples de graphiques. Les

[58] Comme l'enseignant peut choisir plusieurs graphiques comme corrects, la somme totale des graphiques choisis est supérieure au nombre de réponses.

cases diagonales sont pour les graphiques choisis comme le seul correct ; la dernière ligne est celle du tableau 8 :

	G1	G2	G3	G4	G5
G1	3	0	2	0	1
G2	0	0	0	0	0
G3	2	0	23	1	2
G4	0	0	1	1	0
G5	1	0	2	0	0
Total	6	0	28	2	3

Tableau 20: Nombres de couples des graphiques choisis comme corrects

Analyses

1) Couple (G2, G3)

Parmi les 28 choix du G3, 23 enseignants ont choisi G3 comme le seul correct, par contre aucun enseignant ne choisit G2 comme correct. De façon quasi-unanime G3 est choisi et G2 est rejeté. La stratégie $S_{fonction}$ est dominante et la notion d'aire ne prend aucune place dans les réponses des enseignants. On retrouve la stratégie $S_{fonction}$ dans leurs explications pour rejeter le graphique G2. Ainsi :

> Enseignant 5 : « Dans le graphique 2, les ordonnées sont fausses »
>
> Enseignant 42 : « Graphique 2 est incorrect car l'ordonnée de classe [1,60 ; 1,70[est fausse »

Ces enseignants ont lu les hauteurs des rectangles sur l'axe des ordonnées puis ils les comparent aux effectifs dans le tableau et trouvent que certaines sont fausses.

Par contre, même s'il n'y a aucun choix sur le graphique G2 comme correct, nous trouvons l'usage de la stratégie S_{classe} (stratégie de classes égales) à travers l'explication d'un enseignant pour refuser certains graphiques :

> Enseignant 1 : « Le fait de considérer que les deux classes [1,60 ; 1,65[et [1,65 ; 1,70[ont le même effectif qui est la moitié de l'effectif de classe [1,60 ; 1,70[est faux. »

C'est un « découpage-lecture » pour ramener le graphique au cas des classes égales mais cet enseignant n'accepte pas l'hypothèse de répartition uniforme dans la classe qui justifierait l'égalité des effectifs des deux sous-classes lors d'une division binaire de classe initiale.

Puis, pour refuser le graphique G3, cet enseignant utilise la même stratégie S_{classe} :

> Enseignant 1 : « Le fait de considérer que les deux classes [1,60 ; 1,65[et [1,65 ; 1,70[ont le même effectif que la classe [1,60 ; 1,70[est faux ».

---

[59] Les enseignants donnent au maximum deux graphiques comme corrects, c'est pour quoi nous pouvons établir ce tableau de deux entrées pour voir des couples choisis.

Une interprétation possible est qu’il a fait la somme de deux effectifs de deux sous-classes en trouvant que le résultat est beaucoup plus que l’effectif de classe initiale.

2) Couple (G1, G3)

Il n’est pratiquement pas présent ce qui peut s’interpréter par l’importance que donnent les enseignants à la proportionnalité de largeurs des classes à celles des intervalles.

3) La disjonction des rectangles est refusée par presque tous les enseignants.

> Enseignant 42 : « Graphique 5 est incorrect car c’est un graphique pour décrire des données continues regroupées en classes, donc il ne faut pas des espaces entre des intervalles».
>
> Enseignant 48 : « Les graphiques 4 et 5 sont incorrects car le manuel de classe 10 n’introduit pas ces formes »

4) Contrat didactique sur le regroupement en classes

Les 16 réponses de refus de proposer cet exercice à l’élève illustrent la force du contrat didactique sur le regroupement des données en classes égales. Elles proposent de modifier les classes du tableau de données en classes égales ou bien de remplacer le tableau par un autre tableau :

> Enseignant 19+41 : « Modifier le tableau des données (des classes égales) comme celui donné dans le manuel »
>
> Enseignant 44 : « L’élève n’apprend que des tableaux avec des classes de même tailles, donc il faut modifier des classes » …« il faut modifier des graphiques-solutions car il n’y a pas de graphique juste » … « Remplacer le tableau de distribution des effectifs regroupées en clases étant des intervalles égales ».

Pour répondre à la question « La réponse que vous attendez d’un élève de classe 10 », certains enseignants attendent que l’élève sache montrer que le tableau des données est erroné à cause des classes inégales :

> Enseignant 45 : « L’élève doit montrer que le tableau des effectifs regroupés en classe n’est pas juste car les classes n’ont pas de même largeurs (dans les cours, on n’enseigne que la division en classes égales) »
>
> Enseignant 49 : « L’élève répond que les graphiques donnés sont tous faux car le sujet a donné des classes de largeurs inégales ».

5) Trois enseignants qui ne trouvent aucun graphique correct ont refusé G3 en avançant comme raison le fait qu’il n’y a pas d’espace entre l’origine O et les intervalles sur l’axe des abscisses. Il est vrai que les manuels laissent toujours un tel espace lequel est quelquefois hachuré pour coder la réduction de la distance. Ces enseignants refusent l’identification par translation entre 0 et 1,60 sur l’axe des réels. Cette pratique courante dans les diagrammes statistiques ne l’est pas dans les graphiques fonctionnels.

Réponses des élèves

Nous avons reçu 206 réponses d’élèves pour l’exercice 1 alors que 207 élèves ont participé au questionnaire car il y a un élève qui ne résout pas cet exercice (mais il résout les trois autres du questionnaire).

Le tableau suivant présente une distribution des 5 graphiques déclarés corrects par les élèves :

Graphique	G1	G2	G3	G4	G5
Nombre de choix de l'élève	30	3	119	45	105

Tableau 21. Distribution des effectifs des graphiques corrects donnés par élève

Tableau 11 nous organisons les choix des élèves :

G1	G2	G3	G4	G5	Nombre d'élèves	Les choix
					17	Aucun graphique n'est correct
				x	26	G5 est le seul correct
			x	x	20	Couple (G4, G5) le seul correct
		x			68	G3 est le seul correct
		x		x	29	Couple (G3, G5) le seul correct
		x	x		4	Couple (G3, G4) le seul correct
		x	x	x	9	
	x		x	x	3	
x					3	G1 est le seul correct
x				x	10	Couple (G1, G5) le seul correct
x			x		2	Couple (G1, G4) le seul correct
x			x	x	6	
x		x			7	Couple (G1, G3) le seul correct
x		x		x	1	
x		x	x	x	1	

Tableau 22. Combinaisons observées des graphiques choisis comme corrects de 206 élèves

Analyses

1) Couple (G2 ; G3)

G3 est encore le graphique le plus choisi mais dans une proportion moindre que chez les enseignants. Le couple (G2 ; G3) n'apparaît pas : les deux stratégies $S_{fonction}$ et S_{classe} sont prises séparément.

> Elève L13 : « … Graphique 2 est incorrect car la classe [1,60 ; 1,70[contient 16 joueurs mais ce graphique ne représente qu'une valeur moins de 10 »

Par contre apparaît fortement le couple (G3 ; G5). Nous reviendrons plus loin sur l'analyse de ce fait.

2) Couple (G1, G3)

Ce couple est choisi par 7 élèves et rejeté par 3 élèves.

2a) Couple (G1, G3) seul accepté

Les élèves suivent la stratégie fonctionnelle sans tenir compte de la largeur des intervalles sur l'axe des abscisses et en rejetant la disjonction des rectangles.

2b) Couple (G1, G3) seul rejeté

> Elève V73 : « Les correctes sont 2, 4, 5.
>
> Le graphique G1 est incorrect car les largeurs des classes sur l'axe des abscisses doivent inégales.
>
> Le graphique 3 est incorrect car chaque carré dans le graphique représente une unité de valeur. La colonne [1,60 ; 1,70[représente la valeur de 16 doit être représentée par 16 carrés ».

C'est le seul élève qui déclare lire « aire-effectif» dans l'histogramme Il utilise la stratégie aire pour rejeter G3.

> Elève V72 : « Les graphiques corrects sont les graphiques 2, 4, 5 ».

Dans le graphique 1, les colonnes sont faussement tracées ; les largeurs des classes ne sont pas égales donc les largeurs des colonnes doivent être différentes.

Même raison que l'élève V73

Dans le graphique 3, si on prend l'exemple de demi-intervalle [1,60 ; 1,70[il correspond à l'effectif 16 mais le graphique se présente en deux colonnes égales, alors l'effectif doit être égale à 32 $\Rightarrow$ c'est faux ».

Quelle stratégie suit-il ? Il a procédé à une lecture « hauteur-effectif » après un découpage de classe. C'est-à-dire il coupe le rectangle qui représente l'effectif de classe [1,60 ; 1,70[en deux demi-rectangles égaux puis il lit la hauteur de chaque demi-rectangle comme l'effectif de chaque demi-classe. C'est la stratégie S_{classe} qu'il a suivie comme l'enseignant 1, mais cet élève accepte l'hypothèse de l'égalité des effectifs de deux sous-classes lors d'une bi-division de la classe initiale.

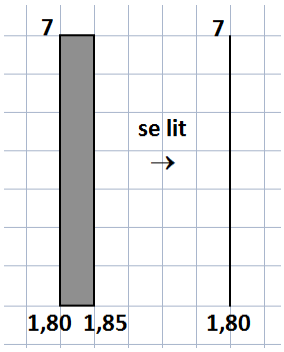
3) Les graphiques G4 et G5

A la différence des enseignants, ces graphiques sont beaucoup choisis par les élèves qui les ont rencontrés en classe 7. D'ailleurs G5 présent en classe 10 en tant qu'ensemble de rectangles disjoints est préféré à G4 qui n'apparaît jamais seul dans les réponses. La disjonction des rectangles est donc acceptée par la moitié des élèves alors qu'aucun enseignant ne l'acceptait. Pour ceux qui ne prennent que le couple (G4 ; G5), il y en a 20, c'est peut-être la façon d'accepter des classes inégales comme si ces classes étaient des valeurs isolées d'une variable qualitative.

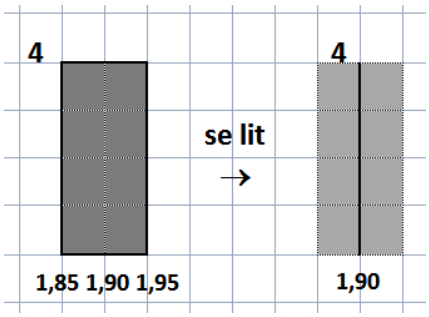
> Elève V58 : « Les trois graphiques 1, 2, 3 sont incorrects car les données sont regroupés en classes dans le tableau tandis que sur les graphiques, on ne prend qu'une seule valeur pour représenter toute la classe »

Cette réponse peut être interprétée comme suit :

- dans le graphique 1, en regardant le rectangle qui représente l'effectif de classe [1,80 ; 1,85[, cet élève ne lit qu'un côté à gauche[60] du rectangle comme un segment qui joint le point (1,80 ; 0) et le point (1,85 ; 7) et il considère qu'il y a 7 joueurs de même taille 1,80m. C'est pour quoi cet élève a dit qu'une seule valeur (1,80m) appartenant à l'intervalle (exemple [1,60 ; 1,70[ ) est prise pour représenter toute la classe (dans l'exemple : [1,60 ; 1,70[ ). De même façon, il peut lire qu'il y a 15 joueurs de taille 1,70m, 18 joueurs de taille 1,75,…



- de même dans les graphiques 2 et 3 : prenons un rectangle qui représente classe [1,85 ; 1,95[est remplacé par le segment qui joint le point (1,90 ; 0) et point (1,90 ; 4). C'est-à-dire, la valeur 1,90 est la seule qui est prise pour représenter toute la classe [1,85 ; 1,95[, donc cet élève a lit qu'il y a 4 joueurs de taille de 1,90m.



> Elève L35 : « Il y a 16 joueurs de tailles appartenant à [1,60 ; 1,70[mais c'est pas surement qu'il y ait un même nombre de 8 joueurs dont leurs tailles appartiennent à [1,60 ; 1,65[et à [1,65 ; 1,70[»

C'est un « coupement-lecture » en n'acceptant pas l'hypothèse sur la division d'une classe en deux sous-classes des effectifs égaux car cela fait perdre l'information.

Le fait que (G3 ; G5) soit largement préféré au (G1 ; G5) qui était plus attendu peut s'interpréter ainsi :

- on rejette les rectangles adjacents à largeurs égales (puisque les classes sont inégales)

- on accepte les rectangles disjoints comme si la variable était qualitative (non plus la taille mais la classe de taille)

[60] Cette classe est un intervalle fermé à gauche et ouvert à droite.

4) Contrat didactique sur le regroupement en classes

> Elève V54, V68 : « Graphique 2 et 3 sont incorrects car les largeurs des classes ne sont pas égales »

Ici, il y a sous-entendue l'idée sous-entendue de ne pas accepter un regroupement en classes inégales. Puisque l'élève n'a jamais rencontré de données regroupées en classes inégales, il ne sait donc pas dire que les classes doivent être égales comme l'enseignant (sachant que les savoir complémentaire sur la méthode pour tracer le diagramme en colonnes au cas des données regroupées en classes inégales ne se trouve que dans Livre d'enseignants). C'est pourquoi il trouve « quelque chose étrange » ici en voyant que les largeurs des rectangles sont inégales et la prend comme raison pour rejeter ces graphiques.

5) Notons que l'absence du point d'origine O sur l'axe des abscisses est invoquée plusieurs fois pour rejeter G1, G2 et G3.

I.5. Conclusion

L'obligation faite aux enseignants et aux élèves dans le questionnaire de choisir des graphiques associés à une distribution en classes inégales, construite comme une rupture du contrat didactique relatif au regroupement en classes, a permis non seulement de valider les hypothèses 1 et 2 mais a mis en lumière deux autres phénomènes didactiques relatifs à l'enseignement de la statistique au Vietnam :

- Le recours à l'analogie fonctionnelle pour construire et lire des diagrammes en segments n'est pas remise en cause avec les diagrammes en colonnes ;

- Le découpage des savoirs relatifs à la nature des variables statistiques (qualitatives ou quantitatives discrètes en classe 7 et quantitatives continues en classe 10) ne s'accompagne pas d'une rupture sur la nature de la représentation graphique : beaucoup d'élèves choisissent des diagrammes en segments ou en rectangles disjoints car ils reconnaissent à ces diagrammes la capacité de représenter correctement une distribution de variable continue. Le rejet des segments ou des rectangles disjoints chez les enseignants ne relève-t-il pas d'une soumission très forte au manuel reconnu comme le représentant de l'institution d'enseignement ? Les enseignants seraient garants des rapports institutionnels. Cette interprétation est confirmée par le grand nombre de refus de réponses de la part des enseignants.

L'absence de l'aire dans les rapports institutionnels relatifs à l'histogramme est corrélative à celle de la notion de densité d'effectifs dans le programme et les manuels. Quel est l'effet de cette double absence sur la vie institutionnelle de la courbe polygonale qui est introduite à la suite de l'histogramme ? Quelle signification cette courbe porte-elle ?

II. Exercice 2

II.1. Quelques éléments d'analyse institutionnelle vietnamienne à propos de l'objet courbe polygonale

Après l'introduction du diagramme en colonnes (histogramme), les deux manuels de classe 10, le manuel élémentaire et le manuel avancé, introduisent la courbe polygonale comme une autre façon de décrire une distribution des effectifs/fréquences regroupées en classes (égales).

Le manuel élémentaire

A partir du tableau de distribution des fréquences de la taille de 36 élèves,

Chiều cao của 36 học sinh

Lớp số đo chiều cao (cm)	Tần số	Tần suất (%)
[150 ; 156)	6	16,7
[156 ; 162)	12	33,3
[162 ; 168)	13	36,1
[168 ; 174]	5	13,9
Cộng	36	100 (%)

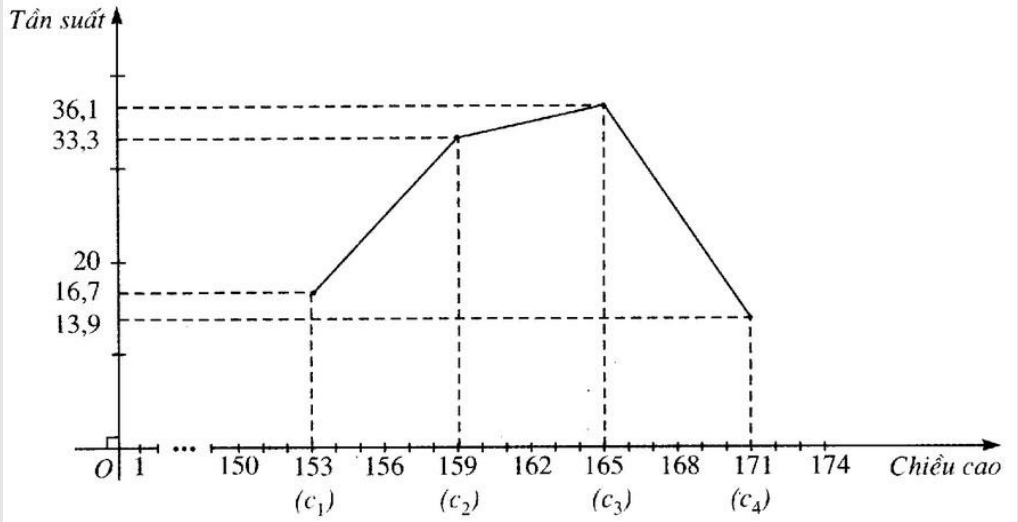
Bảng 4

(Manuel de classe 10, l'ensemble élémentaire, p. 116)

ce manuel propose une définition de la courbe et une technique pour la tracer :

> Courbe polygonale des fréquences :
>
> Le tableau de distribution des fréquences regroupées en classes (tableau 4) précédent peut être décrit par une courbe polygonale dont la méthode de tracer est comme suit :
>
> Dans le plan rapporté à un repère cartésien, préciser les points $(c_i ; f_i)$, i = 1, 2, 3, 4 où c_i est la moyenne numérique de deux extrémités de classe i (c_i est appelé la valeur représentante[61] de la classe i).
>
> Tracer les segments joignant le point $(c_i ; f_i)$ et le point $(c_{i+1} ; f_{i+1})$, i = 1, 2, 3, on obtient une courbe polygonale appelée courbe polygonale des fréquences (Figure 35)

[61] Ces mots sont soulignés par le manuel.



Hình 35. Đường gấp khúc tần suất về chiều cao (cm) của 36 học sinh.

(Figure 35. Courbe polygonale des fréquences de la taille (cm) de 36 élèves)

(Manuel de classe 10, l'ensemble élémentaire, p. 116)

La légende de l'axe des abscisses est « taille » et celle de l'axe des ordonnées est « fréquence ».

Les traits pointillés aident à lire les coordonnées des points $(c_i ; f_i)$, i = 1, 2, 3, 4, qui sont en nombre égal aux classes du tableau. Les traits pleins joignant ces points forment la courbe polygonale.

Analyses

- Cette présentation prépare un milieu fonctionnel pour l'objet courbe polygonale. La technique pour tracer la courbe ressemble totalement à celle de la représentation graphique d'une fonction numérique. Cette technique s'appuie seulement sur le tableau de données regroupé en classes (égales) sans avoir besoin de passer par l'histogramme. Comment l'institution vietnamienne relie-t-elle l'histogramme et la courbe polygonale ?

- Dans certains manuels universitaires cette courbe est présentée à partir de l'histogramme mais l'axe des ordonnées porte les densités d'effectifs ou de fréquences. Voici par exemple la présentation de la courbe polygonale intitulée « polygone des effectifs » :

> Pour donner de la distribution une image « plus continue », certains auteurs préconisent de tracer une ligne brisée reliant les points d'abscisse x_{cj} (centre de classe) et d'ordonnée n_j / h_j qui, sur l'histogramme des effectifs présentés dans la figure 2.18, sont situés au milieu des côtés supérieurs des rectangles. Deux points complètent le schéma aux extrémités de façon à ce que le graphique possède la même propriété que l'histogramme des effectifs : la surface délimitée par cette courbe et l'axe des abscisses doit valoir n. Cette représentation porte le nom de polygone des effectifs.

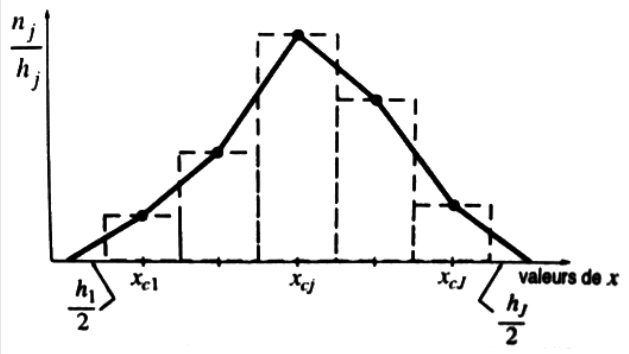
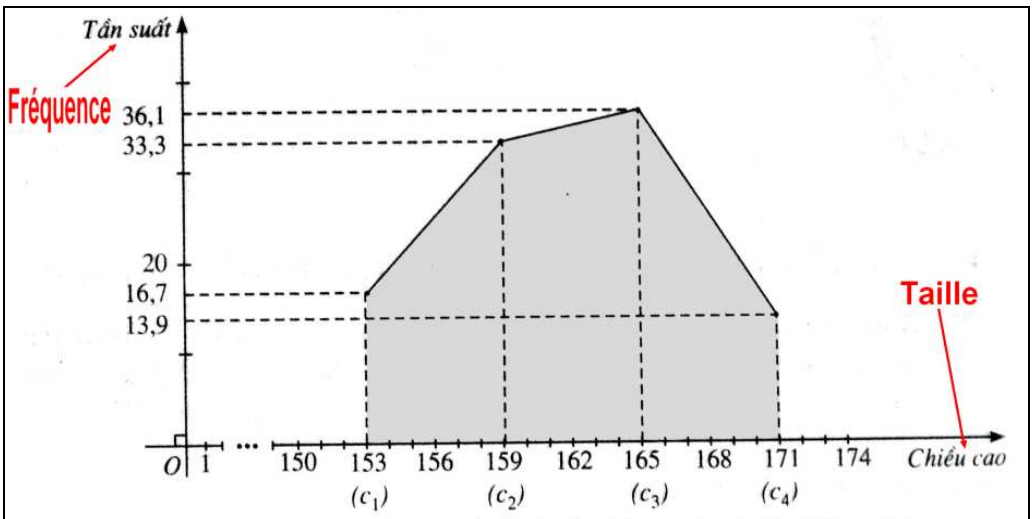


Figure 2.18 *Polygone des effectifs*

(DROESBEKE Jean-Jacques, 2001, Éléments de statistique, Editions de l'Université de Bruxelles, p. 46)

Ce manuel présente ensuite le polygone des fréquences dont il dit qu'il est parfois appelé « graphe de densité locale » lorsque les amplitudes des classes sont relativement petites.

En calculant la surface sous la courbe du manuel vietnamien, on trouvera que son aire, calculée avec une unité égale à celle d'un rectangle de hauteur 1% et de largeur égale à 6 (la largeur commune aux classes), est inférieure à la somme totale des fréquences exprimées en %, c'est-à-dire inférieure à 100 contrairement à la propriété mise en avant par les manuels universitaires quand ils font le parallèle entre une telle courbe et la courbe de densité d'une distribution théorique.



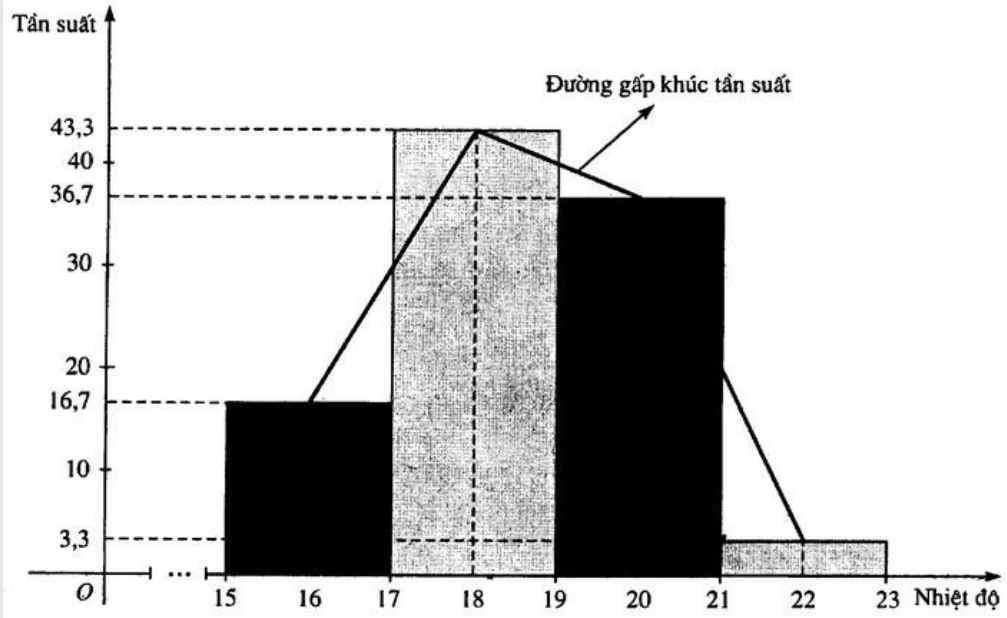
Hình 35. Đường gấp khúc tần suất về chiều cao (cm) của 36 học sinh.

En effet :

Aire sous courbe polygonale =

$$\frac{(16{,}7\% + 33{,}3\%)}{2} + \frac{(33{,}3\% + 36{,}1\%)}{2} + \frac{(36{,}1\% + 13{,}9\%)}{2} + 84{,}7\% < 100\%$$

Regardons maintenant le livre d'enseignant de l'ensemble élémentaire qui donne la solution pour l'Activité 1 proposée dans le manuel.



Hình 12. Biểu đồ tần suất hình cột và đường gấp khúc tần suất về nhiệt độ (°C) trung bình của tháng 12 tại thành phố Vinh từ 1961 đến hết 1990 (30 năm)

(Diagramme en colonnes et courbe polygonale des fréquences de la température (°C) moyenne en décembre de la ville de Vinh entre 1961 et 1990 (30 ans))

(Livre d'enseignant de l'ensemble élémentaire, p. 128)

On voit qu'il y a effectivement un passage du diagramme en colonnes des fréquences regroupées en classe à la courbe polygonale des fréquences. En traçant les deux diagrammes sur le même repère, ce livre fait un lien entre ces deux diagrammes : le diagramme en colonnes est un intermédiaire graphique entre le tableau de distribution des fréquences regroupées en classes et la courbe polygonale des fréquences. Mais la conservation de l'aire des rectangles du diagramme n'est pas prise en compte pour la construction de la courbe. Et aucune remarque n'invite à s'intéresser à la forme de cette courbe.

Un type de tâche concernant la courbe polygonale se trouve dans le livre d'exercices corrigés : il s'agit de comparer deux distributions d'effectifs par leur courbe polygonale. Voici l'un des exercices corrigés :

Exercice 19

Soit le tableau de distribution des effectifs regroupés en classes :

Le poids des élèves de classe 10A et 10B du lycée L.

Classe de poids	Effectif	
	10A	10B
[30 ; 36)	1	2
[36 ; 42)	2	7
[42 ; 48)	5	12
[48 ; 54)	15	13
[54 ; 60)	9	7
[60 ; 66)	6	5
Total	38	46

Etablir le tableau de distribution des fréquences regroupées en classes avec les classes données dans le tableau.

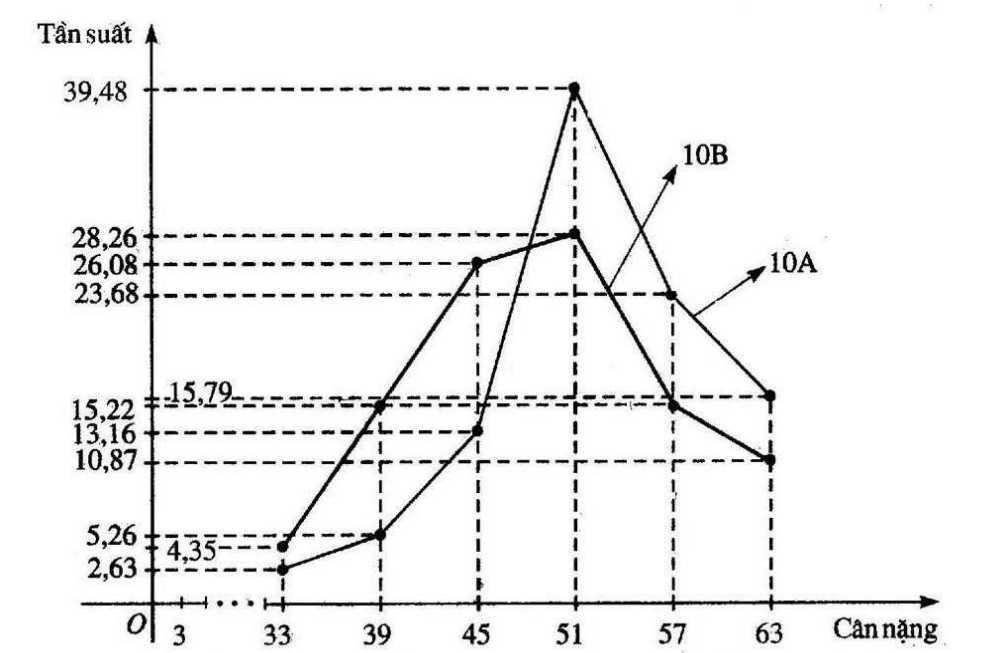
Tracer dans le même plan de repère les courbes polygonales des fréquences regroupées en classes du poids des élèves de deux classes 10A et 10B. Puis comparer les poids des élèves de classe 10A et ceux des élèves de classe 10B.

Quel est le pourcentage des élèves de classe 10A pesant au moins de 42kg ? Même question pour les élèves de classe 10B.

Calculer la moyenne, l'écart type des données de classe 10A et 10B. Dans quelle classe les élèves sont-ils les plus lourds ?

(Livre d'exercices corrigés de l'ensemble élémentaire, p. 162)

La solution de la question 19b :



Hình 59. Đường gấp khúc tần suất về cân nặng (kg) của học sinh lớp 10A, lớp 10B trường Trung học phổ thông L

(Figure 59. Les courbes polygonales des fréquences regroupées en classes du poids des élèves des classes 10A et 10B)

(Livre d'exercices corrigés de l'ensemble élémentaire, p. 173)

Deux courbes polygonales sont tracées sur le même repère cartésien. Chacune joint six points correspondant à six classes proposées dans l'énoncé. Les deux courbes se coupent en un seul point, donc une partie de la courbe 10A est au-dessus de la courbe 10B et une autre partie de la courbe 10A est au-dessous celle de 10B.

Ce graphique permet :

- de comparer visuellement les formes :
- d'étudier la position relative des deux courbes,
- de comparer les aires sous les deux courbes.

Quelle est la solution attendue par le manuel ? Et quelle technique utilise-t-il ?

Voici le corrigé du livre :

> En observant deux courbes polygonales, on voit que : parmi les élèves pesant au plus 45kg, les élèves de classe 10B occupent un pourcentage plus grand que ceux de classe 10A. Et parmi les élèves pesant au moins 51kg, les élèves de classe 10B occupent un pourcentage plus grand que ceux de classe 10A.
>
> (Livre d'exercices corrigés, l'ensemble élémentaire, p. 162)

Comment le manuel procède-t-il dans son observation des courbes ? Il s'appuie sur la position relative des courbes et les deux valeurs 45 et 51 qui sont choisies sont les extrémités des classes qui entourent l'abscisse du point d'intersection des deux courbes polygonales et qui sont marquées sur l'axe. Cela donne la conclusion : *« parmi les élèves pesant au plus de 45kg, les élèves de classe 10B occupent un pourcentage plus grand que ceux de classe 10A* ». De même manière, 51 est l'abscisse la plus petite et lisible sur l'axe des abscisses tel que la courbe polygonale 10B est au-dessus de la courbe polygonale 10B, cela donne donc : « *parmi les élèves pesant au moins de 51kg, les élèves de classe 10B occupe un pourcentage plus grand que ceux de classe 10A* ».

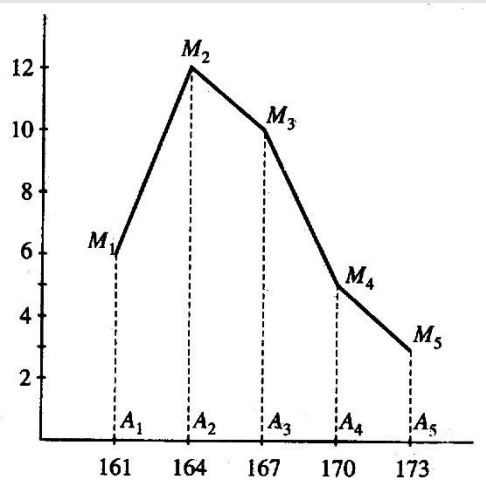
La position relative de deux courbes polygonales de fréquences est au cœur de la technique de comparaison des deux distributions de poids. Les aires sous les courbes ne sont pas mobilisées pour la comparaison alors qu'elles apporteraient (avec une autre technique) des réponses plus fortes. Par exemple, en reprenant les termes de comparaison utilisés par le manuel, on pourrait affirmer que « parmi les élèves qui pèsent au plus 51 kg, les élèves de la classe 10B occupent un pourcentage plus grand que ceux de la classe 10A ».

Le manuel avancé

Dans ce manuel, la courbe polygonale des effectifs est introduite pour décrire le tableau de distribution des effectifs de la manière suivante :

> Exemple 4 :
>
> Considérer le tableau de l'exemple 2 (tableau 4), la courbe polygonale pour décrire le tableau 4 est tracé de manière suivante :
>
> Tracer deux droites perpendiculaires. Sur la droite horizontale (utilisé comme l'axe numérique), marquer les points A_1, A_2, A_3, A_4, A_5 où A_i est le point milieu du

segment (ou du demi-intervalle) qui détermine la ième classe (i=1, 2, …, 5). A chaque point A_i, tracer le segment A_iM_i qui est perpendiculaire à la droite horizontale et sa longueur égale l'effectif de la ième classe : en concrète, $A_1M_1 = 6,...,A_5M_5 = 3$. Tracer les segments $M_1M_2, M_2M_3, M_3M_4, M_4M_5$, on obtient une courbe polygonale. C'est la courbe polygonale des effectifs qui décrit le tableau 4.



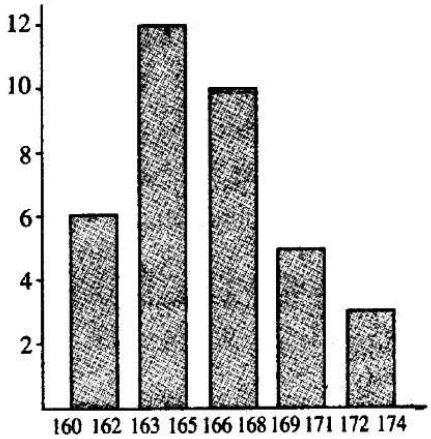
Hình 5.3

(Manuel de classe 10, l'ensemble avancé, p. 166)

Rappelons-nous que les classes données dans le tableau 4 sont disjointes et que le diagramme correspondant présenté dans le manuel est un diagramme en colonnes disjointes :

Lớp	Tần số
[160 ; 162]	6
[163 ; 165]	12
[166 ; 168]	10
[169 ; 171]	5
[172 ; 174]	3
	$N = 36$

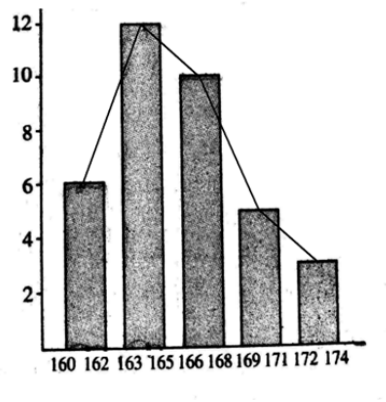
Bảng 4



Hình 5.1

(Manuel de classe 10, l'ensemble avancé, p. 165)

Comme dans le manuel élémentaire, la courbe polygonale est tracée à partir d'un tableau de distribution en classes égales. A la différence du manuel élémentaire, il s'agit d'une courbe d'effectifs. Il n'associe pas la courbe au diagramme en colonnes. Une telle association se traduirait par cette figure :



où l’aire sous la courbe polygonale ne prendrait aucun sens.

Mais, alors que cela se fait dans l’ensemble « élémentaire », nous ne retrouvons aucune association entre un diagramme et une courbe polygonale dans l’ensemble « avancé » même lorsqu’il est demandé de tracer la courbe polygonale à partir d’un tableau de distribution en classes adjacentes.

Conclusion

Ainsi dans l’institution vietnamienne les liens entre les diagrammes en colonnes, que les colonnes soient adjacentes ou disjointes, et la courbe polygonale, qu’elle soit une courbe d’effectifs ou de fréquences, sont très peu développés dans le programme et les manuels. Par contre des liens existent entre courbe et tableau d’une part, entre diagramme et tableau d’autre part. La construction de la courbe et son interprétation (quand elle est demandée aux élèves par le manuel) ne s’appuient que sur les points « centres des classes » dont l’abscisse est la valeur centrale de la classe et l’ordonnée est l’effectif (ou la fréquence) de la classe entière. Aucune signification n’est apportée à la forme de la courbe ni à l’aire sous cette courbe. Seule la position relative de deux courbes de fréquences est utilisée dans la comparaison de deux distributions.

Tout ceci dessine une réponse à l’une des question de recherche présentées dans la partie C1 et que nous rappelons :

QR5 : au sein de la praxéologie de comparaison dans l’institution vietnamienne, existe-t-il un point de vue fonctionnel sur les graphiques statistiques ?

Avant d’arrêter notre réponse nous mettons expérimentalement à l’épreuve deux hypothèses sur l’institution vietnamienne :

H1 : il existe une règle de contrat didactique à propos de l’objet courbe polygonale :

Pour construire la courbe polygonale des fréquences ou des effectifs, il faut joindre les points « centres des classes » présents dans le tableau de distribution par des segments de droite.

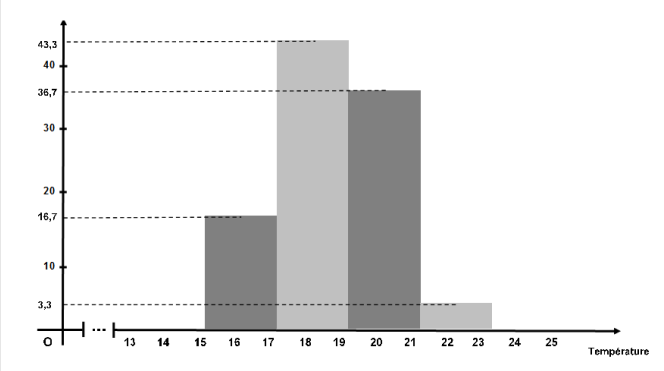
H2 : les rapports institutionnels à la courbe polygonale n’offrent aucun lien relatif à la notion d’aire entre la courbe et le diagramme en colonnes d’une même distribution.

Ces deux hypothèses sont mises à l’épreuve dans l’exercice 2 du questionnaire que voici.

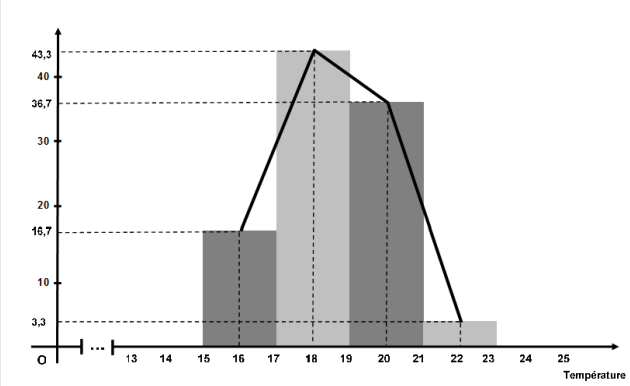
II.2. Présentation de l'exercice

Le sujet ci-après est la partie commune distribuée en feuille à chaque élève et à chaque enseignant :

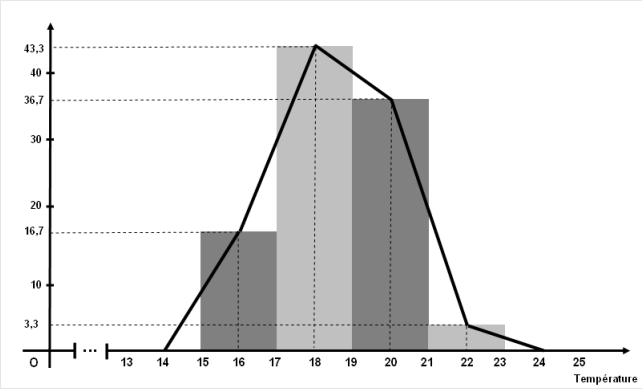
Un météorologue a relevé de 1961 à 1990 tous les jours de décembre à 10h la température dans un même lieu de la ville de Vinh. Voici l'histogramme de la distribution des fréquences :



Le météorologue a construit plusieurs courbes polygonales des fréquences pour décrire la distribution des fréquences :



Courbe polygonale 1



Courbe polygonale 2

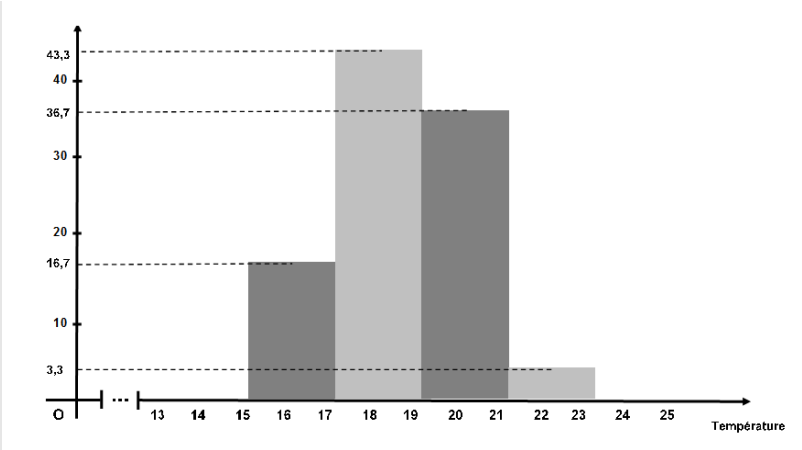
Question : parmi les deux courbes polygonales tracées, lesquelles considérez-vous comme correctes pour décrire la distribution des fréquences ? Justifiez votre réponse.

Les demandes pour l'élève et pour l'enseignant sont celles de l'exercice 1.

II.3. Analyse a priori

Les variables de situation

1) L’histogramme des fréquences des valeurs de la variable regroupées en classes égales est commun aux deux courbes polygonales 1 et 2. L’histogramme initial est donné sans tableau et il est présent en arrière-plan dans chaque graphique. Ce fait constitue une rupture avec l’institution vietnamienne qui ne propose pas aux élèves d’association entre le diagramme et la courbe.



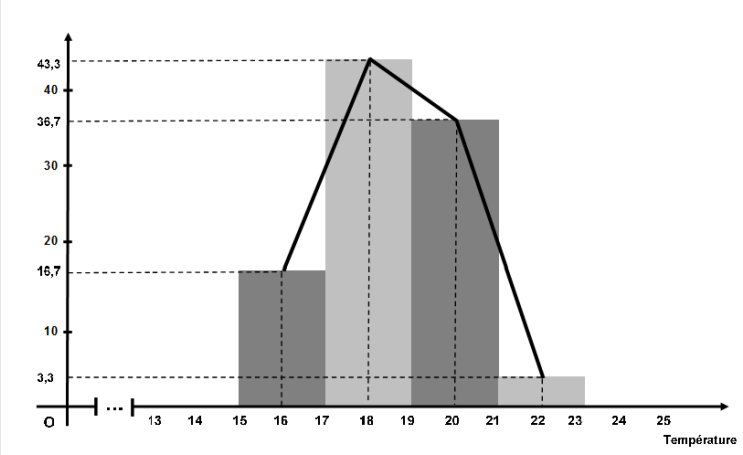
Histogramme

2) Une légende est écrite sur l’axe des abscisses tandis que la légende sur l'axe des ordonnées est non-écrite.

3) Les traits pointillés verticaux aux milieux des rectangles sont tracés dans les deux diagrammes des courbes.

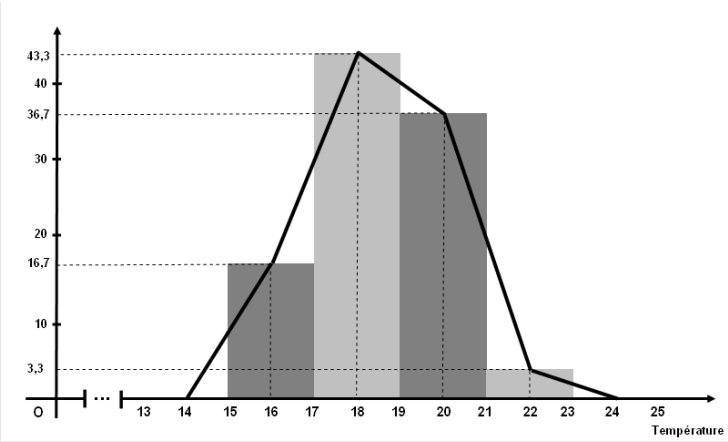
Les différentes courbes polygonales

La courbe 1 est conforme à la technique du manuel de classe 10 en assignant aux pointillés le rôle d’indicateur du centre de la classe. Elle est formée des segments qui joignent les points milieux des quatre segments horizontaux supérieurs des rectangles du diagramme initial. Les classes étant construites en largeur et en hauteurs à partir des rectangles du diagramme initial, les coordonnées de chacun de ces points milieux sont donc lisibles et assimilables aux couples (c_i, f_i) où c_i est la valeur centrale de la classe et f_i est la fréquence de la classe. L’aire sous la courbe est inférieure à 100 %.



Courbe polygonale 1

La courbe 2 est construite de la même manière que la courbe 1 en ajoutant aux extrémités de la distribution deux classes vides de même largeur que les autres : la classe [13; 15[avant la première classe et la classe [23; 25[après la dernière classe. Donc la courbe 2 est formée selon la même technique que la courbe 1 par six points au lieu de quatre pour la courbe 1. Cet ajout fait passer l'aire sous la courbe polygonale à l'aire totale des quatre rectangles de l'histogramme (100 %). Redisons que cette courbe est la seule présente dans le savoir savant.



Courbe polygonale 2

Variable didactique V(distribution, courbe polygonale)

Distribution : une même distribution des fréquences est donnée sous la forme d'un diagramme en colonnes (contre un tableau de distribution des fréquences).

Courbe polygonale : un élément différencie les deux courbes polygonales, la présence ou non de classes vides adjacentes.

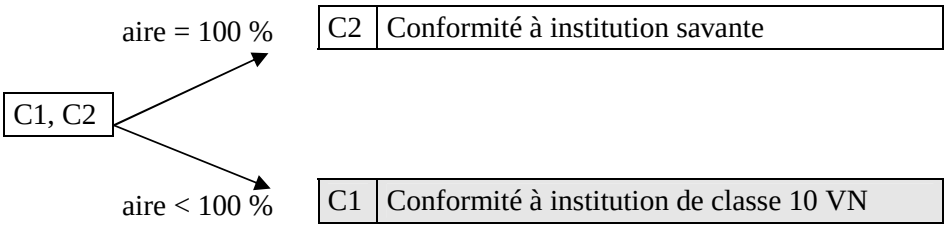


Schéma 6. Arbre des courbes polygonales choisies d'après les critères

Stratégies :

$S_{tableau}$: Stratégie « *tableau* »

Reconstituer un tableau de distribution des fréquences à partir du graphique (courbe + rectangles) à travers l'interprétation des signes graphiques suivants (trait vertical pointillé pour le centre de la classe, trait vertical plein pour une borne de la classe) puis le comparer avec le tableau de distribution initial (celui correspondant à l'histogramme donné).

Voici les tableaux reconstitués :

Classe	Fréquences
[15; 17[	16,7
[17; 19[	43,33
[19; 21[	36,7
[21; 23[	3,3

Tableau 23. Tableau de la distribution initiale

Classe	Fréquences
[15; 17[	16,7
[17; 19[	43,33
[19; 21[	36,7
[21; 23[	3,3

Tableau 24. Tableau de distribution correspondant à la courbe C1

Classe	Fréquences
[13; 15[	0
[15; 17[	16,7
[17; 19[	43,33
[19; 21[	36,7
[21; 23[	3,3
[23; 25[	0

Tableau 25. Tableau de distribution correspondant à la courbe C2

Cette stratégie conduit à accepter C1 et à rejeter C2.

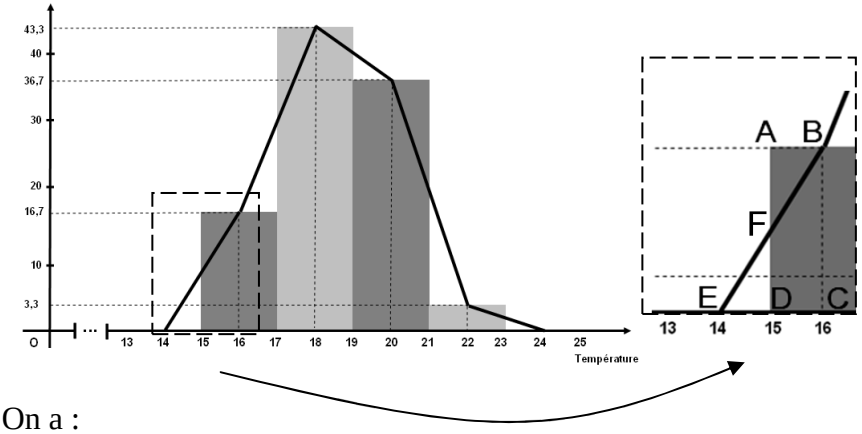
S_{air} : Stratégie « *aire* »

Comparer l'aire sous la courbe et accepter la courbe si et seulement cette aire est égale à l'aire totale des rectangles. Cette stratégie conduit à accepter C2 et à rejeter C1.

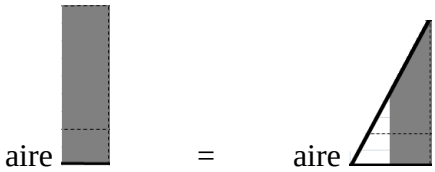
Conservation d'aire sous la courbe polygonale :

L'aire sous le premier segment de la courbe polygonale :

Comme deux triangles rectangulaires FAB et FDE sont égaux (opposés par le point F), l'aire du triangle BEC (l'aire sous le premier segment EB de la courbe polygonale) est égale à l'aire du rectangle ABCD :

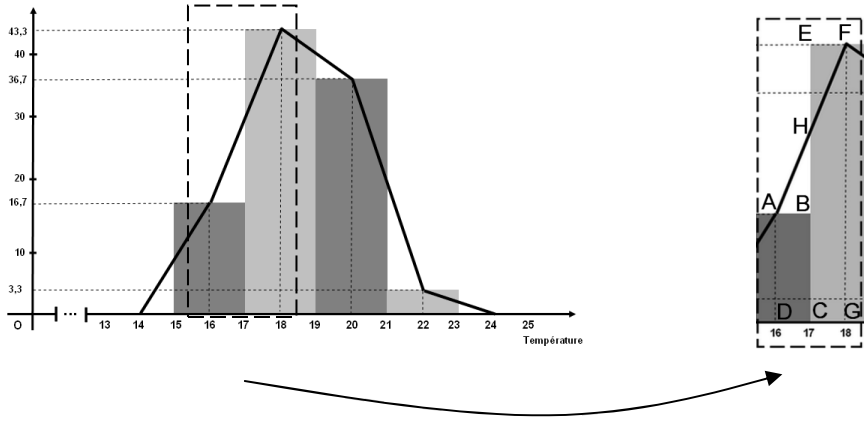


On a :

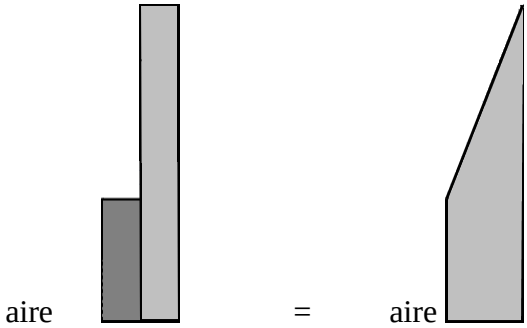


L'aire sous le deuxième segment de la courbe polygonale :

De même manière, deux triangles rectangulaires HEF et HBA sont égaux (opposés par le point H), l'aire du trapèze AFGD (l'aire sous le deuxième segment AF de la courbe polygonale) est égale à la somme des aires de deux demi-rectangles ABCD et EFGC.



On a :



Analoguement, l’aire sous chaque segment de la courbe polygonale C2 est égale à la somme des aires des demi-rectangles dont les bases sont superposables sur l’axe des abscisses.

Alors, l’aire sous la courbe polygonale est conservée.

Remarquons que le choix (C1 ; C2) exclut toute considération sur l’aire. Quelles peuvent être les raisons d’un tel choix, s’il existe ?

II.4. Analyse a posteriori

Enseignant :

Le tableau 5 à double entrée répartit la participation des enseignants d’après les réponses données ou non par eux et d’après leurs acceptations de proposer ou non l’exercice 2 aux élèves.

Choix des graphiques / Proposer l'exercice 2 à l’élève	Oui	Non	Total
Oui	30	9	39
Non	16	17	33
Total	46	26	72

Tableau 26. Répartition des enseignants

Nous nous intéressons aux réponses de 46 enseignants :

C1	C2	Nombre d’enseignants	Caractéristique du choix
X		44	C1 est la seule correcte
X	X	2	C1 et C2 sont correctes
	X	0	C2 est la seule correcte

Tableau 27. Combinaisons observées des graphiques choisis comme corrects de 46 enseignants

Graphique	C1	C2
Nombre de choix de l’enseignant	46	2

Tableau 28. Distribution des effectifs des graphiques corrects donnés par enseignant

La conformité de la courbe C1 au rapport institutionnel fait que cette courbe est quasi-unanimement choisie. Aucun enseignant ne choisit C2 seul et les deux enseignants qui choisissent le couple (C1 ; C2) ne fournissent aucune explication.

Elèves

Le tableau 8 présente les combinaisons observées de 188 élèves (207 élèves en total) :

	C1	C2	Total
C1	174	4	178
C2		10	10
Total	174	14	188

Tableau 29. Combinaisons observées des graphiques choisis comme corrects de 188 élèves

19 élèves ne donnent pas de réponse à cet exercice. On s'intéresse ci-après aux 14 élèves qui donnent C2 et particulièrement à ceux qui choisissent C2 contre C1.

Analyses

1) Courbe C2 seul

Parmi les 14 choix du C2, 10 élèves ont choisi C2 comme la seule correcte. Certains ont pris en compte le fait que, pour eux, la courbe polygonale doit toucher l'axe des abscisses.

> Elève V11 : « La courbe 2 est correcte car elle doit commencer par un point sur l'axe des abscisses puis se terminer par un autre point sur cet axe ».
>
> Elève V58 : « La courbe C2 est correcte car on doit joindre le point le plus bas, puis le points les plus haut de chaque rectangle et enfin le point le plus bas ».
>
> Elève V52 : « La courbe C2 est correcte car on doit tracer de bas en haut et aux milieux de chaque colonne ».
>
> Elève L53 : « Je pense que la courbe 2 est correcte car :
>
> - Ses points sont aux milieux des colonnes,
>
> - La courbe est prolongée jusqu'aux deux axes (cela permet de voir que les fréquences des températures au-delà des intervalles sont égales au zéro »
>
> Elève N8 : « Parmi les courbes polygonales, la deuxième est correcte, par ce que « mon maitre a enseigné comme cela », le point de départ est le milieu de classe [13; 15[, c'est à 0 %, et le dernier point est le milieu de classe [23; 25[est à 0 % aussi. C'est pourquoi on doit commencer par 0 % et puis terminer par 0 % ».

Les élèves font référence à la nécessité de placer le premier et le dernier point de la courbe sur l'axe des abscisses. Le dernier élève parle même d'un enseignement reçu dans ce sens. Les raison mathématiques avancées sont-elles de l'ordre du contrat didactique relatif au graphique d'une fonction numérique ? Ou bien certains enseignants justifient-ils cette pratique par des raisons spécifiques aux diagrammes statistiques par exemple en parlant de classes vides comme le fait l'élève N8 ? Un autre élève semble accepter l'idée des classes vides pour accepter C2 :

Elève L18 : « La courbe 2 est correcte car elle joint des points milieux »

Tandis qu’un autre le fait pour rejeter C2 :

Elève V50 : « La courbe C2 est incorrecte car elle passe par le point 14 qui est le milieu du segment [13 ; 15[» mais il n’y a pas de classe [13 ; 15[».

2) Couple (C1 ; C2)

4 élèves ont choisi C1 et C2 comme correctes à la fois. Dans leurs explications, la technique pour tracer la courbe polygonale est exprimée par la lecture des valeurs centrales des classes à travers les traits pointillés verticaux sans évoquer la notion d’aire.

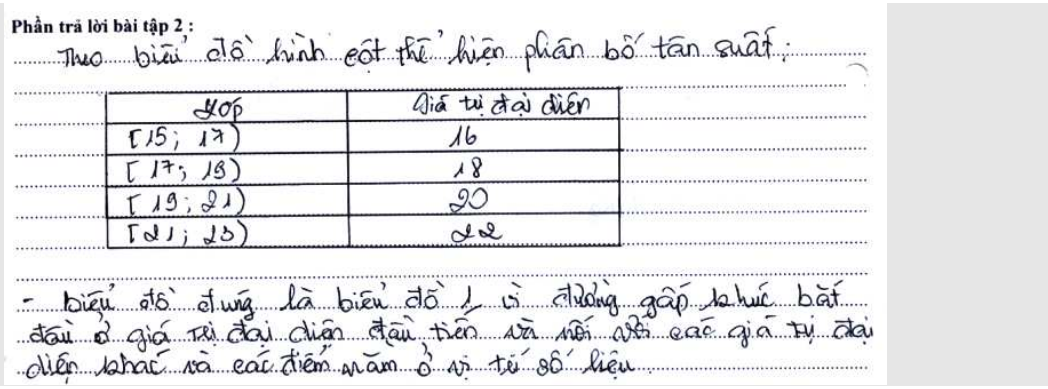
Elève V54 : « Les courbes C1 et C2 sont correctes car quand on trace la courbe polygonale, il faut prendre les points milieux des barres des rectangles ».

Elève N25 : « Les courbes C1 et C2 sont correctes parce qu’elles joignent les points qui appartiennent aux droites perpendiculaires[62] à l’axe des abscisses en les points qui présentent les données[63] ».

Elève V62 : « Les courbes correctes sont C1 et C2 »

Elève N11 : « Les courbes C1 et C2 sont correctes parce qu’elles représentent correctement les valeurs »

3) Le passage de l’histogramme à la courbe donne lieu à une reconstitution d’un tableau de distribution même si ce tableau est incorrectement construit comme pour l’élève V26.


Phần trả lời bài tập 2 :

Theo biểu đồ hình cột thể hiện phân bố tần suất ;

Nhóm	Giá trị đại diện
[15 ; 17)	16
[17 ; 19)	18
[19 ; 21)	20
[21 ; 23)	22

- biểu đồ đúng là biểu đồ 1 vì đường gấp khúc bắt đầu ở giá trị đại diện đầu tiên và nối với các giá trị đại diện khác và các điểm nằm ở vị trí số liệu

Cet élève a fabriqué un tableau à deux colonnes dont la première présente les quatre classes du tableau de la distribution initiale et la deuxième présente les valeurs centrales des classes dont les fréquences sont absentes. Puis il conclue :

Elève V26 :

Le premier diagramme est correct car la courbe commence par la première valeur[64] représentante de la classe puis elle joint les valeurs représentantes suivantes et les points se trouvent.

[62] Ce sont les traits pointillés verticaux visibles sur les rectangles du diagramme.

[63] Ce sont les valeurs centrales des classes lues sur l’axe des abscisses.

[64] Nous reviendrons sur l’usage du mot « représentant » pour désigner la valeur centrale de la classe.

II.5. Conclusion

Les résultats obtenus à l'exercice 2 du questionnaire confirment et complètent la première analyse institutionnelle.

L'aire sous la courbe polygonale et celle d'un histogramme ne jouent aucun rôle dans l'institution vietnamienne.

La courbe polygonale dans l'institution vietnamienne est construite et interprétée à partir d'un tableau de distribution présent ou reconstitué (via l'histogramme) de la même manière que la représentation graphique d'une fonction à partir d'un tableau de valeurs présent ou reconstitué (via la formule).

Variable statistique continue	Fonction numérique continue
- Tableau de distribution - Un nombre fini des couples (c_i, f_i) ou (c_i, n_i) - Représentation graphique = une ligne polygonale joignant les points (c_i, f_i) ou (c_i, n_i)	- Tableau des valeurs - Un nombre fini de couples (x ; y) - Représentation graphique = une ligne joignant les points (x ; y)
- Les coordonnées (x ; y) d'un point situé sur la ligne en dehors des extrémités des segments ne peuvent pas être interprétées sous la forme (x ; f(x)) où f(x) serait une fréquence. - L'aire sous la courbe polygonale entre deux droites x=a et x=b est interprétable en termes de fréquence.	- Les coordonnées (x ; y) de chaque point de la représentation graphique peuvent être interprétées sous la forme (x ; f(x)) - L'aire sous la représentation graphique encadrée entre deux droites x = a et x = b est l'intégrale définie de cette fonction de a à b.

Tableau 30. Ressemblances et différences entre une variable statistique et une fonction numérique

La partie grisée du tableau est objet d'enseignement dans l'institution vietnamienne. La nature statistique de la courbe polygonale n'est pas enseignée pas plus que l'aire sous cette courbe.

Nous sommes en mesure maintenant d'apporter une réponse argumentée à QR5 :

QR5 : au sein de la praxéologie de comparaison dans l'institution vietnamienne, existe-t-il un point de vue fonctionnel sur les graphiques statistiques ?

Notre réponse est très tranchée : le point de vue fonctionnel n'est pas absent de la production et la lecture des graphiques mais son usage est inadéquat pour construire des liens statistique-probabilités.

On peut même faire l'hypothèse que ce point de vue fonctionnel s'érige dans l'institution en obstacle didactique à la construction de liens statistique-probabilité dans la mesure où de tels liens font appel à la notion d'aire sous la courbe (de densité). Cet obstacle est renforcé par l'absence de cumulation des fréquences et de la fonction de répartition.

III. Exercice 3

III.1. Quelques éléments d'analyse institutionnelle vietnamienne à propos de la notion de représentativité dans une étude statistique.

La réduction des données brutes relatives à une population en vue de son analyse notamment par comparaison avec d'autres populations justifie pour les statisticiens la recherche d'indicateurs de position centrale et de « modèles réduits » des distributions dont le plus connu est le couple (moyenne, écart-type). La réduction des données s'accompagne d'un objectif de représentativité simplifiée de la population. C'est cette notion de représentativité que nous explorons dans l'institution d'enseignement

L'idée de disposer d'un représentant d'une série statistique apparaît la première fois en classe 7 avec la présentation des indicateurs de position centrale que sont la moyenne et le mode. Voici ce qu'écrit le manuel :

Quel nombre peut-il être le « représentant » des valeurs de caractère ?

La moyenne numérique du caractère

Les notes d'un contrôle des maths des élèves de classe 7C sont notées par le chef de la classe dans le tableau 19 suivant :

3	6	6	7	7	2	9	6
4	7	5	8	10	9	8	7
7	7	6	6	5	8	2	8
8	8	2	4	7	7	6	8
5	6	6	3	8	8	4	7

Bảng 19

(Tableau 19)

[…]

Si on considère la note du contrôle comme le caractère étudié, on peut donc établir le tableau des effectifs portant deux colonnes pour calculer la note moyenne (tableau 20) :

Điểm số (x)	Tần số (n)	Các tích (x.n)	
2	3	6	
3	2	6	
4	3	12	
5	3	15	
6	8	48	
7	9	63	
8	9	72	
9	2	18	
10	1	10	
	N = 40	Tổng : 250	$\bar{X} = \frac{250}{40} = 6,25$

Bảng 20

(Tableau 20)

[…]

Formule[65] :

En s'appuyant sur le tableau des effectifs, on peut calculer la moyenne d'un caractère (la moyenne arithmétique notée $\overline{X}$) :

Multiplier chaque valeur avec l'effectif correspondant,

Additionner tous ces produits,

Diviser cette somme par le nombre des valeurs (c'est-à-dire la somme de tous les effectifs)

$$\overline{X} = \frac{n_1 x_1 + n_2 x_2 + ... + n_k x_k}{N} \quad (1)$$

$x_1, x_2, ..., x_k$ est k valeurs différentes du caractère X

$n_1, n_2, ..., n_k$ est k effectifs correspondants

N est le nombre des valeurs

[…]

(Manuel de classe 7, volume 2, p. 17-18)

Ce manuel explique en quoi la moyenne joue un rôle de représentant de la série :

La signification de la moyenne arithmétique

La moyenne arithmétique est un « représentant » pour ce caractère quand il faut le présenter en abrégé ou quand il faut le comparer avec un autre caractère de même type (par exemple, on peut comparer la capacité d'étudier les maths de deux élèves après une année scolaire grâce à la moyenne des notes de maths à la fin d'année scolaire de chaque élève.

[...]

La moyenne arithmétique est utilisée souvent comme le « représentant » d'un caractère, surtout quand on veut comparer deux mêmes caractères.

[…].

(Manuel de classe 7, volume 2, p. 19)

En dehors de la comparaison entre séries, les significations à donner à la moyenne d'une série ne sont pas exposées ici. Dans le cas où la variable est quantitative, nous savons que la moyenne est la valeur du caractère qui serait donnée à chaque individu si tous les individus étaient égaux. Mais l'interprétation qui en découle souffre du défaut d'être difficilement acceptable si la moyenne n'est pas une valeur possible de la variable (exemple : le nombre moyen d'enfants par femme). Une autre signification présente la moyenne comme la valeur du caractère qui annule la somme des écarts. Elle justifie la comparaison, interne à la série, de chaque individu à la moyenne, en le positionnant soit en dessous soit au dessus de la moyenne. Ajoutons une signification qui ne sera pas présente en classe 7 : la moyenne

[65] Nous notons (1) cette formule pour la reprendre plus tard.

minimise la somme des carrés des écarts. Cette interprétation convient à une analyse de la série qui s'intéresserait à la dispersion des individus autour d'un représentant.

Bien que n'ayant pas vraiment suggéré d'interprétation de la moyenne, le manuel prend soin de prévenir que la moyenne de la série peut ne pas être portée par un individu contrairement au mode que nous regarderons par la suite. Mais cette remarque ne met pas en cause le choix de la moyenne comme le représentant de la série étudiée.

> Remarque :
>
> La moyenne arithmétique peut être n'appartient pas à la suite des valeurs du caractère.
>
> Exemple : 6,25 n'est pas une valeur du caractère présenté dans le tableau 20.

(Manuel de classe 7, volume 2, p. 19)

Par contre il rejette la moyenne lorsque la série a des valeurs extrêmes éloignées :

> Remarque
>
> Lorsqu'il y a une grande différence entre des valeurs du caractère, il ne faut pas prendre la moyenne arithmétique pour son « représentant ».
>
> Exemple : Soit un caractère X ayant la suite des valeurs : 4000 1000 500 100.
>
> Il est impossible de prendre la moyenne arithmétique $\overline{X} = 1400$ pour le représentant de X car il y a une très grande amplitude entre les valeurs (4000 et 100).

(Manuel de classe 7, volume 2, p. 19)

Cette série dépourvue de tout contexte peut-t-elle disposer d'un représentant ? La question vaut aussi pour cette autre série :

> Observer le tableau des « effectifs » (tableau 24) et répondre s'il faudrait utiliser la moyenne arithmétique pour le représentant du caractère ? Pourquoi ?

Giá trị (x)	2	3	4	90	100	
Tần số (n)	3	2	2	2	1	N = 10

Bảng 24

Tableau 24

(Manuel de classe 7, volume 2, p. 20)

D'ailleurs, dans la révision conduite à la fin du chapitre statistique, le manuel insiste encore une fois sur ce cas :

> Comment faire pour calculer la moyenne arithmétique d'un caractère ?
>
> Citer les pas pour la calculer. La signification de la moyenne arithmétique ? Quand est-ce que la moyenne arithmétique a peu de chance d'être le représentant du caractère ?

(Manuel de classe 7, volume 2, p. 20)

A partir d'un exemple tiré de la vie quotidienne le manuel introduit le mode et lui donne aussi un rôle de représentant :

Le mode d'un caractère

Exemple : Une boutique des chaussures a noté la taille de chaussures vendues pour les hommes pendant un trimestre dans le tableau 22 :

Cỡ dép (x)	36	37	38	39	40	41	42	
Số dép bán được (n)	13	45	110	184	126	40	5	N = 523

Bảng 22

(Tableau 22)

Ce que cette boutique s'intéresse le plus, c'est quelle taille de chaussure est vendue la plus, dans ce cas cette taille (taille 39) sera le « représentant » et ce n'est pas la moyenne arithmétique des tailles.

(Manuel de classe 7, volume 2, p. 19)

Le calcul du mode est appelé par la question à la quelle il faut répondre. Le rejet de la moyenne comme représentant est alors imposé à partir d'une unicité supposée du représentant. Le choix du représentant (mode, moyenne ou autre) serait-il donc prioritairement attaché à la question qui motive l'étude ? Il semble en fait que la moyenne est le représentant privilégié notamment dans le cas de comparaison de deux populations.

Dans le livre d'exercices corrigés, nous trouvons en effet deux exercices où la représentativité est assurée par la moyenne.

Première exercice :

Exercice 12

Les températures moyennes annuelles de 1956 à 1975 de deux villes A et B sont notées(en degré °C) dans les tableaux suivants :

• Đối với thành phố A

Nhiệt độ trung bình (x)	23	24	25	26	
Tần số (n)	5	12	2	1	N = 20

• Đối với thành phố B

Nhiệt độ trung bình (x)	23	24	25	
Tần số (n)	7	10	3	N = 20

Comparer la température moyenne annuelle de ces deux villes.

(Livre d'exercices corrigés de classe 7, p. 6)

Sa solution corrigée :

La température moyenne annuelle de chaque ville :

Ville A : $\overline{X} = 23{,}95^{o}\,C$

Ville B :

$\overline{X} = 23{,}8^{o}\,C$

En général, la ville A est un peu plus chaude que la ville B.

(Livre d'exercices corrigés de classe 7, p. 8-9)

Le corrigé fait le choix d'effectuer la comparaison par les moyennes et pas par les modes.

Deuxième exercice :

Exercice 13

Deux tireurs tirent 20 fois chacun, leurs résultats sont notés dans le tableau suivant :

A	8	10	10	10	8	9	9	9	10	8	10	10	8	8	9	9	9	10	10	10
B	10	10	9	10	9	9	9	10	10	10	10	10	7	10	6	6	10	9	10	10

Calculer la note moyenne de chaque tireur.

Faire des remarques du résultat et la capacité de chaque tireur.

(Livre d'exercices corrigés de classe 7, volume 2, p. 6)

La solution corrigée est :

Tireur A : $\overline{X} = 9{,}2$

Tireur B : $\overline{X} = 9{,}2$

Même si la note moyenne de deux tireurs sont les mêmes, le tireur A tire à la cible plus concentriquement que le tireur B.

(Livre d'exercices corrigés de classe 7, volume 2, p. 9)

Devant l'égalité (voulue) des moyennes qui arrête la comparaison, le corrigé introduit implicitement l'idée de dispersion alors qu'une comparaison par les modes av son intérêt : le tireur B fait 10 plus fréquemment que le tireur A.

Le terme de représentant d'une série brute de données est donc introduit en classe 7 à la fois pour réduire et synthétiser la série sous la forme d'un seul nombre, et pour représenter la série toute entière auprès d'autres séries. Ce terme est à distinguer de celui de représentation d'une série utilisé par le manuel pour parler des diagrammes et des graphiques. Même si le mode est introduit, c'est la moyenne qui est privilégiée au risque de choisir une valeur « représentante » qui n'appartient pas à l'ensemble des valeurs possibles de la variable et qui ne soit donc pas portée par l'un des individus de la série.

La classe 10 apporte un représentant de plus, la médiane ; elle introduit aussi la notion de dispersion pour augmenter la capacité de comparaison entre deux séries ; elle travaille aussi avec des séries dont les valeurs sont regroupées en classes.

Les manuels soulignent l'intérêt des indicateurs numériques :

> Pour avoir les informations importantes des données statistiques, on utilise les indicateurs comme la moyenne, le mode, la médiane, la variance et l'écart-type. Ces nombres spécifiques reflètent des côtés différents du caractère étudié.
>
> (Manuel de classe 10, l'ensemble élémentaire, p. 119)

Mais ils privilégient la moyenne dans le rôle de représentant dont ils réaffirment l'unicité :

> Signification de la moyenne
>
> La moyenne d'un échantillon est utilisée comme le représentant de tous les individus de l'échantillon. Elle est un indicateur très important de l'échantillon.
>
> (Manuel de classe 10, l'ensemble avancé, p. 172)

Le terme « représentativité » apparait dans le livre d'enseignants de l'ensemble élémentaire à propos de la moyenne :

A propos de représentativité de la moyenne arithmétique

> Si les données sont de même genre et de quantité n qui est suffisamment grande ($n \geq 30$) on peut choisir la moyenne arithmétique comme le représentant pour les données statistiques. Sa représentativité s'explique comme suit :
>
> Dans tous les individus d'une série statistique de même genre, les individualisations des individus (exprimées par la différence entre eux et par leur caractéristique casuelle) se compensent entre elles pour arriver à la moyenne arithmétique $\overline{X}$ qui a apodicticité et qui exprime le général, la nature commune à tous les individus. C'est pour quoi on peut prendre $\overline{X}$ comme le représentant pour tous les individus de données statistiques.
>
> (Livre d'enseignants de classe 10, l'ensemble élémentaire, p. 153)

A l'enseignant la moyenne est ainsi présentée avec deux significations différentes qui ne sont pas présentes dans le manuel destiné aux élèves :

- la moyenne est la valeur du caractère qui annule la somme des écarts à cette valeur

- la moyenne est la valeur du caractère qui serait la valeur commune à tous les individus si tous les individus étaient égaux, serait la valeur commune ; c'est à ce titre qu'elle endosse le rôle de représentant.

> Cependant il est prévu que la moyenne ne puisse pas être le représentant et son remplacement par un autre indicateur donne lieu à une méthode indépendante du contexte de l'étude statistique et qui est présentée dans deux livres différents, le livre d'exercices corrigés[66] et le livre d'enseignants :
>
> Choisir le représentant pour les données statistiques
>
> a) Dans le cas où l'on peut calculer les trois nombres : la moyenne, le mode, la médiane et tous les individus sont de même nature et le nombre total des individus est suffisamment grand ($n \geq 30$) on choisit de préférence la moyenne comme le représentant. Dans ce cas, la médiane ou le mode est utilisé pour compéter les informations nécessaires.

[66] Il est à la disposition des élèves.

b) Au cas où l'on ne peut pas calculer la moyenne alors on choisit la médiane ou le mode pour représenter les données statistiques.

c) Dans les cas suivants, il ne faut pas utiliser la moyenne pour le représentant des données statistiques (on peut utiliser la médiane ou le mode) :

- Il y a très peu d'individus

- Il n'y a pas de grande différence entre les individus

- La courbe polygonale n'est pas symétrique[67] et plusieurs d'autre cas.

(Livre d'exercices corrigés de classe 10, l'ensemble élémentaire, p. 153)

Remarque

[…]

La moyenne, la médiane et le mode donnent des informations très essentielles de l'échantillon et ils sont souvent utilisés pour le représentant de l'échantillon.

Quand l'échantillon a des valeurs extrêmes (la différence est trop grande ou trop petite par rapport à autres valeurs), il vaut mieux que la médiane soit le représentant. Généralement, quand un échantillon n'a pas de valeurs extrêmes comme cela, la médiane et la moyenne sont assez proches l'une de l'autre.

(Livre d'enseignants de classe 10, l'ensemble avancé, p. 228)

On peut parler d'un type de tâche « Choisir le représentant d'une population » présent dans l'institution vietnamienne dont la technique est algorithmisée. Nous présentons trois exercices de ce type de tâche, avec leurs corrigés.

Le premier exercice :

Soit le tableau de distribution des effectifs des revenus en l'an 2000 de 31 familles d'un village dans les régions montagneuses :

Revenu (en million dong VN)	Effectif
4	1
4,5	1
5	3
5,5	4
6	8
6,5	5
7,5	7
13	2
Total	31

a) Calculer la moyenne, la médiane, le mode des données.
b) Choisir la valeur représentante pour les données.

(Livre d'exercices corrigés de classe 10, ensemble élémentaire, p. 156)

[67] C'est la seule fois où la forme de la courbe polygonale est abordée.

Le corrigé de cet exercice :

> La moyenne $\overline{x} = 6{,}6$ millions. La médiane $M_e = 6$ millions. Le mode $M_o = 6$ millions.
>
> Dans les données, il y a une grande différence entre les individus, on ne choisit donc pas la moyenne mais on choisit la médiane pour le représentant de tous les revenus pendant l'an 2000 de chaque famille parmi 31 familles enquêtées.

(Livre d'exercices corrigés de classe 10, l'ensemble élémentaire, p. 169-170)

Le corrigé n'interprète pas le représentant choisi. Remplacer un revenu moyen par un revenu médian ne répond pas à la même question statistique, question qui n'est d'ailleurs pas présentée dans l'énoncé de l'exercice. Notons que dans cette série statistique, si on enlève les deux individus de la valeur extrême 13, la moyenne deviendrait 6,2 sans que changent la médiane et le mode. La pratique consistant à élaguer une série de ses valeurs extrêmes douteuses n'est pas exposée dans l'institution.

Voici le deuxième exercice (du même livre) où la moyenne n'est pas calculable car la variable est qualitative :

> Soit le tableau de rang en classes des activités des élèves d'une classe 10 dans l'année scolaire 2000-2001:

Rang	Effectif
A	10
B	16
C	16
D	7
Total	49

> a) Calculer la moyenne, la médiane, le mode du tableau 12 (s'il est possible).
> b) Choisir la valeur représentante pour les données.

(Livre d'exercices corrigés de classe 10, l'ensemble élémentaire, p. 156)

Le corrigé de cet exercice :

> On ne peut pas calculer la moyenne.
>
> Ce tableau de distribution comprend 49 individus, chaque individu a un rang d'activité. Il y a 4 rangs ordonnés en suite décroissante du rang le plus haut qui est le rang A vers plus bas qui est le rang D. En s'appuyant sur cette suite, on trouve la médiane $M_e = B$
>
> Il y a deux modes $M_o^1 = B$ et $M_o^2 = C$
>
> On choisit le rang B comme le représentant de tous les individus.

(Livre d'exercices corrigés de classe 10, l'ensemble élémentaire, p. 169-170)

Qu'est-ce qui motive le choix de B comme représentant unique ? Est-ce parce que B est à la fois médiane et mode ? Le choix n'est alors pas motivé par l'interprétation à avoir du représentant dans le contexte étudié. A quelle question le représentant répond-il ?

Voici le troisième exercice, de l'ensemble avancé, dont la série est peu nombreuse et très étendue :

> Les revenus de 8 boutiques pendant une journée dans la ville A (unité de million de dong VN) sont :
>
> 2 2 25 2 10 100 2 10
>
> Calculer la moyenne et la médiane. Qui est le meilleur représentant ?

(Livre d'exercices corrigés de classe 10, l'ensemble avancé, p. 175)

Le corrigé de cet exercice :

> La moyenne est 19,13. La médiane est 6. La médiane est le meilleur représentant parce qu'il y a une grande différence entre des individus dans l'échantillon.

(Livre d'exercices corrigés de classe 10, l'ensemble avancé, p. 187)

La présence d'une valeur extrême très éloignée des autres discrédite la moyenne sans pour autant donner une signification pertinente à la médiane ; l'information portée par le mode n'est-elle pas plus significative de la répartition des revenus des 8 boutiques ?

Ces trois exercices illustrent la méthode pour choisir un représentant numérique dans le cas d'une série discrète. Les corrigés ne fournissent pas d'interprétation du représentant élu alors que la signification dans un contexte ne peut être la même pour une moyenne, une médiane ou un mode. Tout se passe comme si le choix du représentant ne répondait pas à une question qui motiverait l'étude statistique.

La méthode exposée dans le cours et les exercices corrigés est appliquée dans le cas où la variable est qualitative ou quantitative discrète sans regroupement en classes. C'est donc une reprise des savoirs de la classe 7. Comment cette notion de représentant numérique se transforme-t-elle avec les séries regroupées en classe ?

Le manuel, le livre d'exercices corrigés de l'ensemble élémentaire présentent une formule pour calculer la moyenne dans le cas d'un regroupement en classes :

> Cas du tableau de distribution des effectifs, des fréquences regroupées en classes :
>
> $$\overline{X} = \frac{1}{n}(n_1c_1 + n_2c_2 + ... + n_kc_k) = f_1c_1 + f_2c_2 + ... + f_kc_k \qquad (2a)$$
>
> où c_i, n_i, f_i sont respectivement la valeur représentante, l'effectif, la fréquence de i-ième classe
>
> n est le nombre total des individus $n_1 + n_2 + ... + n_k = n$

(Manuel de classe 10, l'ensemble élémentaire, p. 120)

Le centre de la classe n'est pas désigné explicitement comme la valeur dite représentante de la classe et la formule ne se présente pas comme une approximation. La formule entretient la confusion entre un tableau d'effectifs et une distribution en classes.

La présentation dans l'ensemble avancé est différente :

> Supposons que l'échantillon de la taille N soit donné en tableau de distribution des effectifs regroupés en classes, que les données soient regroupées en m classes correspondant à m segments ou en m classes correspondant à m demi-intervalles. Le milieu x_i du segment correspondant à i-nième classe est appelé la valeur représentant de cette classe.
>
> Dans ce cas, la moyenne de cet échantillon est calculée approximativement par la formule : $\overline{X} \approx \frac{1}{n}\sum_{i=1}^{m} n_i x_i$ (2)

(Manuel de classe 10, l'ensemble avancé, p. 171)

Ainsi une moyenne qualifiée d'approximative est calculée à partir des centres des classes, appelés « valeurs représentantes des classes » et avec pondération par les effectifs. Il est dit dans le cours que cette moyenne pourrait être différente de celle que l'on calcule à partir de la série initiale. Certains exercices proposent même les deux calculs dont l'un est sur des classes données avec l'énoncé.

Exemple dans l'ensemble avancé :

> Une compagnie a 45 véhicules. Les consommations (unité litre) de ces véhicules pendant la semaine dernière sont notées suivant :

123	132	130	119	106	97	121	109	118
128	132	115	130	125	121	127	144	115
107	110	112	118	115	134	132	139	144
104	128	138	114	121	129	128	116	138
129	113	105	142	122	131	126	111	142

> Effectuer le tableau de distribution des effectifs regroupés en classes suivantes : [90 ; 100), [100 ; 110), …, [140 ; 150)
>
> Calculer la moyenne et la moyenne (approximative) en s'appuyant sur le tableau de distribution des effectifs regroupés en classes.
>
> Calculer la médiane.

(Livre d'exercices corrigés de classe 10, l'ensemble avancé, p. 181)

Voici le corrigé donné par ce livre :

> Tableau de distribution des effectifs regroupés en classes :

Lớp	Giá trị đại diện	Tần số
[90 ; 100)	95	1
[100 ; 110)	105	5
[110 ; 120)	115	12
[120 ; 130)	125	13
[130 ; 140)	135	10
[140 ; 150)	145	4
		$N = 45$

A partir de ce tableau, on calcule la moyenne qui est 123,44 litres. Si on calcule exactement sur les données (quand il n'y a pas de regroupement), alors que la moyenne est 123,11litres.

[…]

Pour calculer la médiane, on ordonne croissantement les données :

97	104	105	106	107	109	110	111	112
113	114	115	115	115	116	118	118	119
121	121	121	122	123	125	126	127	128
128	128	129	129	130	130	131	132	132
132	134	138	138	139	142	142	144	144

D'où, la médiane est $M_e = 12$

(Livre d'exercices corrigés de classe 10, l'ensemble avancé, p. 190-191)

Les trois indicateurs (moyenne, médiane et moyenne approximative) sont calculés sans être comparés ni interprétés. La médiane est calculée sur la série discrète même après le regroupement en classes effectué à la question précédente.

On peut retrouver le mode et la médiane (pour les séries regroupées en classes) mais seulement dans le livre d'enseignants de l'ensemble élémentaire :

b) Calculer approximativement la médiane dans le cas du tableau de distribution des effectifs regroupés en classes :

Etape 1 : Chercher la classe contenant la médiane

Calculer $\frac{n}{2}$ (n est le nombre des individus)

Additionner les effectifs des classes à partir de la première classe jusqu'à la classe où la somme des effectifs dépasse $\frac{n}{2}$

Etape 2 : Préciser approximativement la médiane d'après la formule : $M_e \approx x_t + \frac{h_t}{n_t}\left(\frac{n}{2} - S_t\right)$

où :

x_t est l'extrémité à gauche de la classe contenant la médiane,

h_t est la largeur de la classe contenant la médiane,

n_t est l'effectif de la classe contenant la médiane.

S_t est la somme des effectifs de toutes les classes avant la classe contenant la médiane.

c) Calculer approximativement le mode dans le cas du tableau de distribution des effectifs regroupés en classes :

Etape 1 : Chercher la classe contenant le mode

Si les classes sont égales, la classe qui porte l'effectif le plus grand est la classe contenant le mode.

Si les classes sont inégales, la classe dont la densité d'effectif la plus grande est celle contenant le mode, où la densité de classe i notée d_i et $d_i = \frac{n_i}{h_i}$, où n_i est l'effectif de classe i, h_i est la largeur de classe i.

Etape 2 : Préciser le mode d'après la formule : $M_o = x_o + h_o \frac{\varepsilon_1}{\delta_1 + \delta_2}$

où :

x_o est l'extrémité à gauche de la classe contenant le mode,

h_o est la largeur de la classe contenant le mode.

$\delta_1 = n_o - n_{o-1}$ (si les classes sont égales) ou bien $\delta_1 = d_o - d_{o-1}$ (si les classes sont inégales), où n_o et n_{o-1} sont l'effectif de la classe contenant le mode et de celle précédente de classe

(Livre d'enseignants de classe 10, l'ensemble élémentaire, p. 154-155)

C'est donc la moyenne qui est la seule habilitée à servir de représentant dans une comparaison entre deux distributions. C'est le point de départ de l'intérêt porté à l'écart-type qui est présenté comme capable de séparer deux distributions de même moyenne. L'écart-type est d'ailleurs le seul indicateur de dispersion présent dans le programme et le manuel.

Si les moyennes des deux distributions ne sont pas égales, le recours à l'écart-type est jugé inutile (pas de notion de coefficient de variation). Le couple (moyenne, écart-type) non seulement n'a pas de concurrent mais son utilisation est très limitée.

En résumé, voici le schéma de la technique de résolution de la tâche « choisir le représentant » :

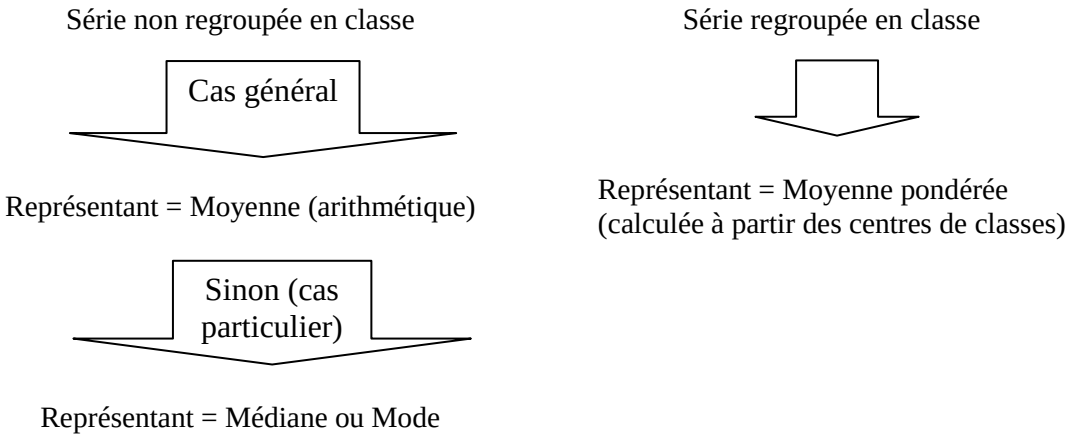


Schéma 7. Schéma de représentant d'après la série donnée

Et voici sa présentation algorithmisée :

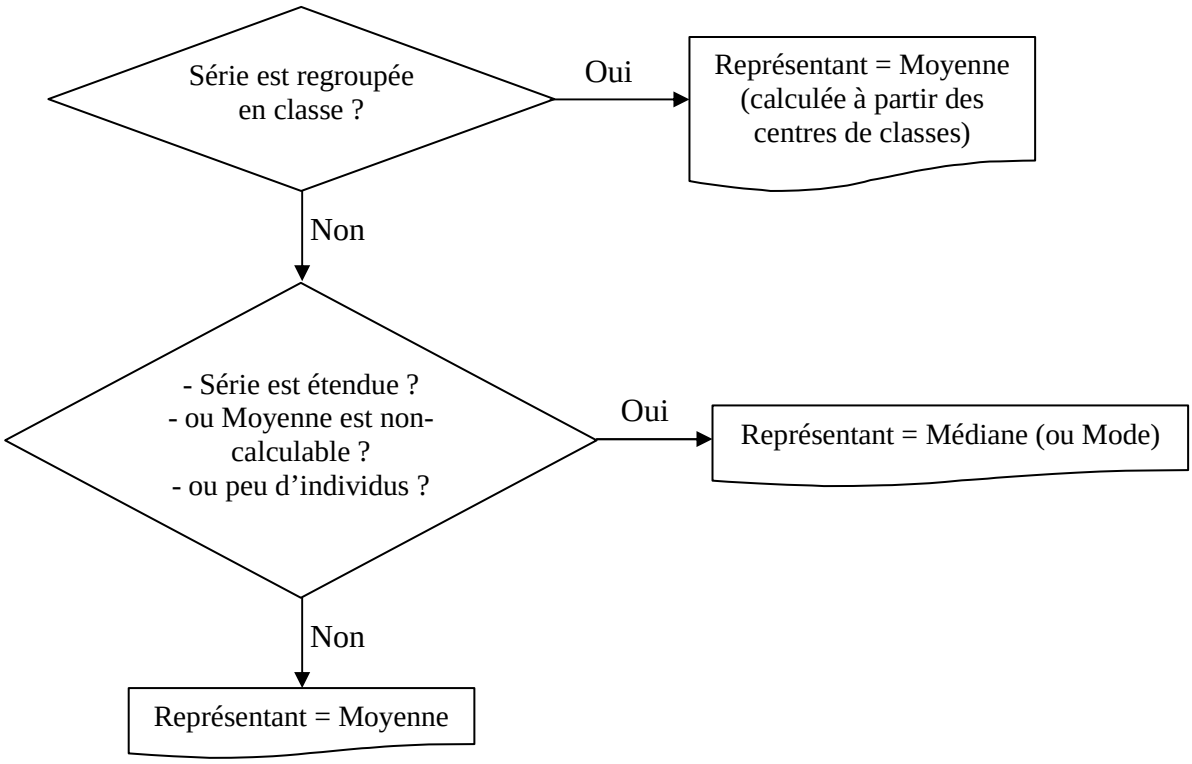


Schéma 8. Algorithme pour choisir le représentant d'après la série donnée

Conclusion

La notion de représentativité d'une série ou d'une distribution statistique par un ou plusieurs nombres est très présente dans l'enseignement vietnamien et sa raison d'être se trouve principalement dans la comparaison potentielle avec une autre distribution sans que cette comparaison soit toujours demandée et sans que l'interprétation du représentant soit partie prenante de la comparaison. Dans le cas de variables quantitatives discrètes, la représentativité est assumée par la moyenne sauf dans des conditions particulières où le mode et la médiane peuvent la remplacer. Dans le cas de variables continues cette notion de représentativité se réduit à celle de la moyenne.

En résumé, nous formulons deux hypothèses sur le contrat didactique relativement au choix d'un représentant dans l'institution vietnamienne :

H1 : pour toute série ou distribution il existe un représentant unique qui est prioritairement la moyenne.

H2 : dans une distribution discrète, ce représentant peut ne pas être la moyenne et dans ce cas il est remplacé par la médiane ou le mode après exécution d'un algorithme.

III.2. Présentation de l'exercice

Le sujet ci-après est la partie commune distribuée sur feuille à chaque élève et à chaque enseignant :

Pour contrôler la qualité de fonctionnement d'une machine remplissant des paquets de sucre, on a noté, en les ordonnant, les poids de 100 paquets remplis par la machine. Ils ont été choisis au hasard durant une journée de fonctionnement.
On obtient le tableau de recueil des poids en grammes

250,8	252,2	281,7	292,3	312,2	344,0	345,6	399,2	490,2	490,2
490,3	490,5	490,6	490,6	491,1	491,3	491,7	492,1	492,3	493,1
493,1	493,1	493,2	493,3	493,3	493,4	494,0	494,3	494,5	494,8
495,1	495,7	495,7	495,8	495,9	496,0	496,5	496,5	497,1	497,1
498,3	498,3	498,9	498,9	499,1	499,1	499,7	499,7	499,9	500,4
500,8	501,1	501,2	501,3	501,4	501,7	501,8	501,9	502,0	502,1
502,2	502,2	502,2	502,3	502,4	502,5	502,8	502,8	503,0	503,7
503,9	504,1	504,4	504,7	504,7	504,8	505,3	505,4	505,5	505,9
506,1	506,2	506,3	506,3	506,8	507,1	507,3	507,5	507,6	507,8
507,9	508,0	508,2	508,6	508,7	509,0	510,5	510,7	511,7	512,2

1. Donnez un indicateur qui est un bon représentant de ce recueil de données ? Justifiez votre choix.
2. Tracez la courbe polygonale des effectifs qui décrit ces données (sur la feuille quadrillée joint).
3. Que peut-on dire sur le fonctionnement de la machine ? Pour quel poids la machine est-elle, programmée ? En tant que contrôleur, que diriez-vous de la machine ? A-t-elle fonctionné normalement ?

Les demandes pour l'élève et pour l'enseignant sont les mêmes que dans les exercices 1 et 2.

L'exercice organise une concurrence entre la médiane et la moyenne pour représenter une série quantitative. Il veut éclairer les rapports institutionnels au choix du représentant, à son rôle et à sa signification dans une comparaison.

III.3. Analyse a priori

On fournit le tableau de données brutes ordonnées croissantes avec un grand nombre d'individus (100 individus) et un grand nombre de valeurs distinctes (82 valeurs distinctes) dont le regroupement en classes n'est pas effectué. Rappelons (cf. exercice 2) que l'élève dans l'institution n'a pas la responsabilité de regrouper les données en classes et quand il rencontre des classes c'est par l'intermédiaire d'un tableau de distribution avec des classes

égales, pas de classes vides, entre 6 et 8 classes[68] et pas de déséquilibre important dans les effectifs de classe.

Les questions et les variables didactiques

Question 1 :

> 1. Donnez un indicateur qui est un bon représentant de ce recueil de données ? Justifiez votre choix.

Cette question relève du type de tâche « choisir un représentant » présent dans l'institution et veut donc mettre à l'épreuve l'hypothèse sur le contrat didactique relatif à la résolution de cette tâche. Ici, le nombre d'individus dépasse largement 30, la moyenne est calculable mais son calcul est très couteux et la série des valeurs est très étendue et très nombreuse. Le choix d'un tableau en données brutes isolées autorise en fait les trois indicateurs. Mais le coût du calcul de la moyenne devrait induire des procédures de réduction de ce coût. On peut donc, dans ce conflit entre la prépondérance de la moyenne pour représenter une série et le coût de son calcul, s'attendre à la médiane comme premier choix du représentant sans écarter la possibilité de trouver la moyenne.

Les variables didactiques

V1 : Valeurs isolées (permet les trois indicateurs) / classes (favorise la moyenne)

Contrat didactique du discret établi depuis la classe 7 / contrat didactique du continu établi en classe 10

V2 : Grand nombre d'individus qui favorise le regroupement des individus /petit nb

V2bis : Fort pourcentage de valeurs distinctes (82 sur 100) qui favorise le calcul de la moyenne par un regroupement en classes / Faible pourcentage de valeurs distinctes qui favorise le calcul de la moyenne par un tableau d'effectifs.

V3 : Distribution numérique asymétrique (médiane $\neq$ moyenne et favorise la création de classes inégales) / symétrique (médiane = moyenne)

V4 : Etendue forte (favorise la formation de classes vides) / étendue faible

[68] Le nombre des classes par rapport au nombre des individus (Livre des enseignants de l'enseignant élémentaire, p. 150)

Nombre d'individus	Nombre de classes
De 25 à 60	De 4 à 8
De 60 à 100	De 7 à 10
Supérieur ou égale à 100	De 8 à 12

Stratégies

$S_{\text{discrète}}$: Stratégie « *discrète* »

Les calculs de chacun des trois indicateurs sont possibles mais ces calculs ne sont pas équivalents en coût.

La médiane ne nécessite qu'un dénombrement simple et de plus la série est étendue ce qui renforce son choix.

Le repérage d'un mode est long et de plus la série est bimodale ce qui limite son choix.

Le coût de calcul de la moyenne[69] est important et cela devrait faire provoquer des calculs intermédiaires structurés par le tableau pour diminuer ce coût. Quels sont ces calculs intermédiaires et sur quels regroupements se font-ils ?

Voici les trois indicateurs de la série :

Moyenne	Modes	Médiane
485,173	493,1 et 502,2	500,6

Tableau 31. Les indicateurs de position centrale

S_{continue} : Stratégie « *continue* »

Cette stratégie s'exprime dans la construction de classes qui utilisent l'ostensif « intervalle de R » sans que la valence sémantique de la continuité soit nécessairement présente.

Pour l'institution n'existe que le regroupement en classes égales toutes non vides dont la formation passe par la division de l'étendue par un nombre entier. Le tableau de données est alors transformé en tableau de distribution.

Dans ce cas, la formule institutionnelle pour calculer la moyenne est la formule (2).

- Sous-stratégie « *classes égales toutes non vides* »

Le refus de classes vides produit des classes très déséquilibrées en effectifs. Le nombre maximum de classes égales toutes non vides est 3.

Voici un exemple de regroupement en 3 classes égales sans classe vide :

[69] D'après la formule (1) de l'institution : $\overline{X} = \dfrac{n_1x_1 + n_2x_2 + \ldots + n_kx_k}{N}$

Classe	Effectif	Valeur centrale de classe
[250;350[	7	300
[350; 450[	1	400
[450; 550[	92	500
Total	100	

Tableau 32. Distribution regroupée en classes égales sans classes vides

Le représentant est la moyenne $\overline{X} = 485$

- Sous-stratégie « *classes égales dont certaines vides* »

La présence d'au moins une classe vide garantit un nombre « suffisant[70] » de classes.

Un exemple :

Classe	Effectif	Valeur centrale de classe
[250; 300[	4	275
[300; 350[	3	325
[350; 400[	1	375
[400; 450[	0	425
[450; 500[	41	475
[500; 550[	51	525
Total	100	

Tableau 33 : Distribution des effectifs regroupés en classes égales dont certaines sont vides

Le représentant est la moyenne $\overline{X} = 487$

- Sous-stratégie « *classes inégales* »

Cette stratégie permet d'équilibrer les effectifs, d'obtenir un nombre « suffisant » de classes et d'éviter des classes vides. Elle est en rupture de contrat.

Voici un exemple de regroupement en 5 classes inégales :

Classe	Valeur centrale de classe	Effectif
[250; 490[	370	8
[490; 495[	492,5	22
[495; 505[	500	46
[505; 510[	507,5	20
[510; 520[	515	4

Tableau 34. Distribution regroupée en classes inégales

Le représentant est la moyenne $\overline{X} = 490{,}05$

[70] Le nombre habituel dans le manuel est compris entre 5 et 8.

Le tableau suivant résume les stratégies avec le représentant :

<table>
<tr><th colspan="3">Stratégies</th><th>Représentant</th></tr>
<tr><td colspan="3">Stratégie discrète</td><td>Moyenne < Mode < Médiane[71]</td></tr>
<tr><td rowspan="4">Stratégie classe</td><td rowspan="2">égales</td><td>sans classes vides</td><td rowspan="4">Moyenne pondérée</td></tr>
<tr><td>avec classes vides</td></tr>
<tr><td rowspan="2">inégales</td><td>adjacentes</td></tr>
<tr><td>disjointes</td></tr>
</table>

Tableau 35 : Synthèse des stratégies et les représentants

Question 2 :

> 2. Tracez la courbe polygonale des effectifs qui décrit ces données (sur la feuille quadrillée jointe).

Conformément au contrat didactique (cf. exercice 2 du questionnaire) la courbe polygonale ne doit être tracée qu'à partir d'un tableau de distribution comportant des classes égales et non vides.

V3 : Distribution asymétrique numérique (favorise la formation de classes inégales) / symétrique (classes égales)

V4 : Etendue forte (favorise la formation de classes vides) / étendue faible

On s'attend donc :

- dans le cadre institutionnel à un conflit entre un regroupement en classes égales peu nombreuses (3 au plus) pour éviter des classes vides ou en classes plus nombreuses mais avec plusieurs classes vides

- à un non-respect du contrat didactique se manifestant par un regroupement en classes inégales pour éviter les classes vides tout en ayant suffisamment de classes.

La stratégie continue provoquée par le regroupement en classes attendu par l'institution pour tracer la courbe polygonale devrait être largement majoritaire, la poursuite d'une stratégie discrète mise en place dans la question 1 étant peu probable.

$S_{\text{discrète}}$: Stratégie « *discrète* »

- Fabriquer le tableau de distribution des effectifs :

Valeur	Eff.
250,8	1
252,2	1
281,7	1
292,3	1
312,2	1

493,4	1
494,0	1
494,3	1
494,5	1
494,8	1

501,2	1
501,3	1
501,4	1
501,7	1
501,8	1
501,9	1

505,4	1
505,5	1
505,9	1
506,1	1
506,2	1
506,3	2

[71] La préférence est dans l'ordre inverse du coût : la médiane est préférée au mode car son calcul est moins coûteux.

344,0	1	495,1	1	502,0	1	506,8	1
345,6	1	495,7	2	502,1	1	507,1	1
399,2	1	495,8	1	502,2	3	507,3	1
490,2	2	495,9	1	502,3	1	507,5	1
490,3	1	496,0	1	502,4	1	507,6	1
490,5	1	496,5	2	502,5	1	507,8	1
490,6	2	497,1	2	502,8	2	507,9	1
491,1	1	498,3	2	503,0	1	508,0	1
491,3	1	498,9	2	503,7	1	508,2	1
491,7	1	499,1	2	503,9	1	508,6	1
492,1	1	499,7	2	504,1	1	508,7	1
492,3	1	499,9	1	504,4	1	509,0	1
493,1	3	500,4	1	504,7	2	510,5	1
493,2	1	500,8	1	504,8	1	510,7	2
493,3	2	501,1	1	505,3	1	512,2	1

Tableau 36. Distribution des effectifs de poids des paquets

- placer tous les points de coordonnées (x_i, e_i) où x_i est la valeur du poids et e_i est l'effectif correspondant.

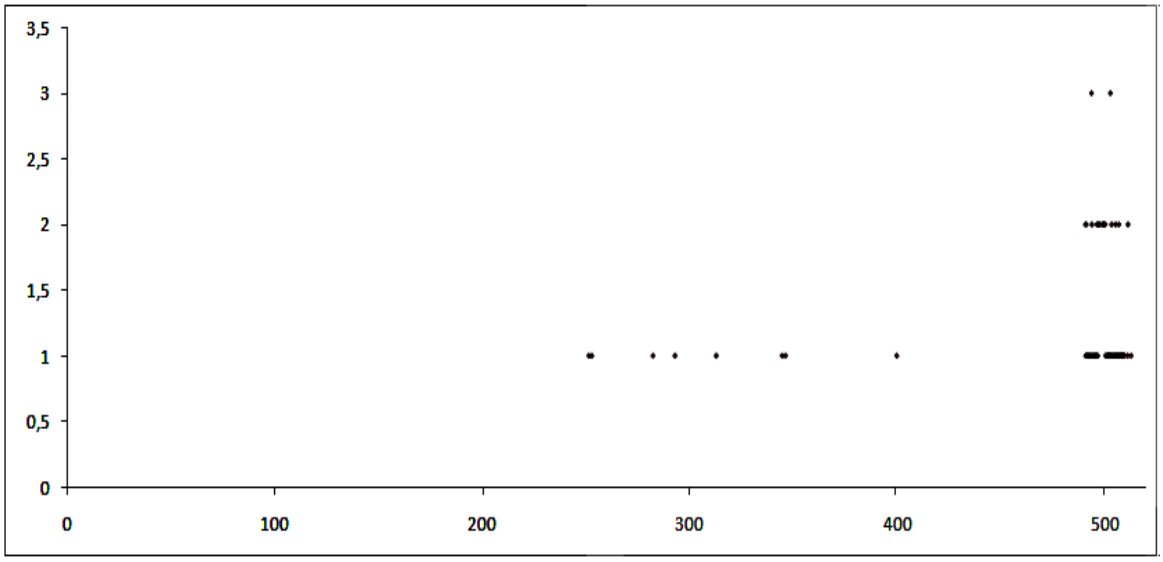


Figure 2. Courbe 1

- ou à partir du tableau 6, ajouter des valeurs intermédiaires entre deux valeurs avec effectifs zéro puis tracer la courbe 1bis :

Valeur	Eff.
250,8	1
250,9	0
251,0	0
….	0
252,1	0
252,2	1
…..	0

Tableau 37. Illustration des valeurs ajoutées correspondant à l'effectif zéro

Dans ce tableau, les lignes grisées sont les valeurs ajoutées en s'appuyant sur un seul chiffre de la partie décimale des valeurs pour une fausse continuité.

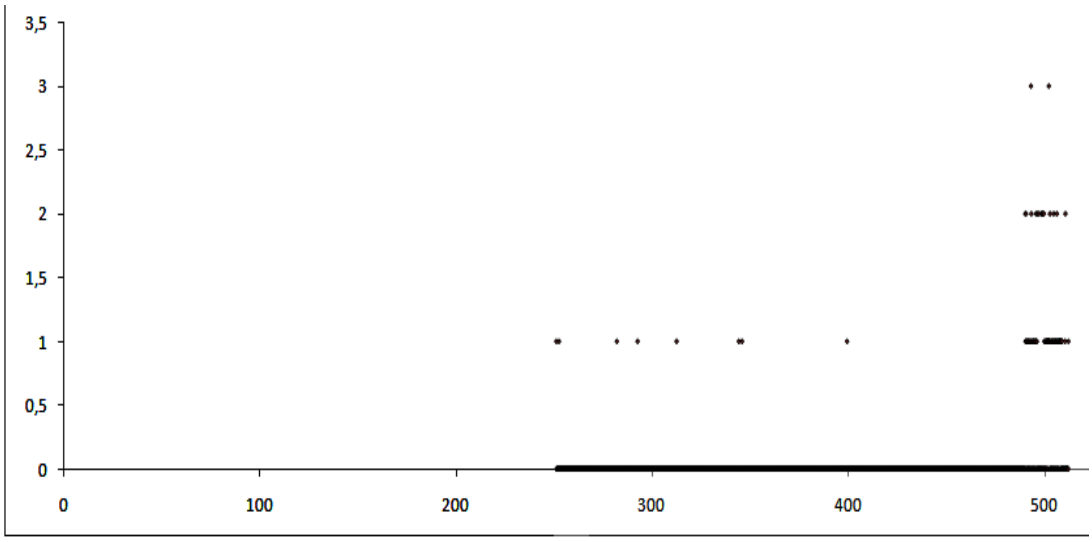


Figure 3. Courbe 1bis

- ou tracer la courbe 1ter en reprenant la courbe 1bis et en joignant tous les points de coordonnées (x_i, e_i) :

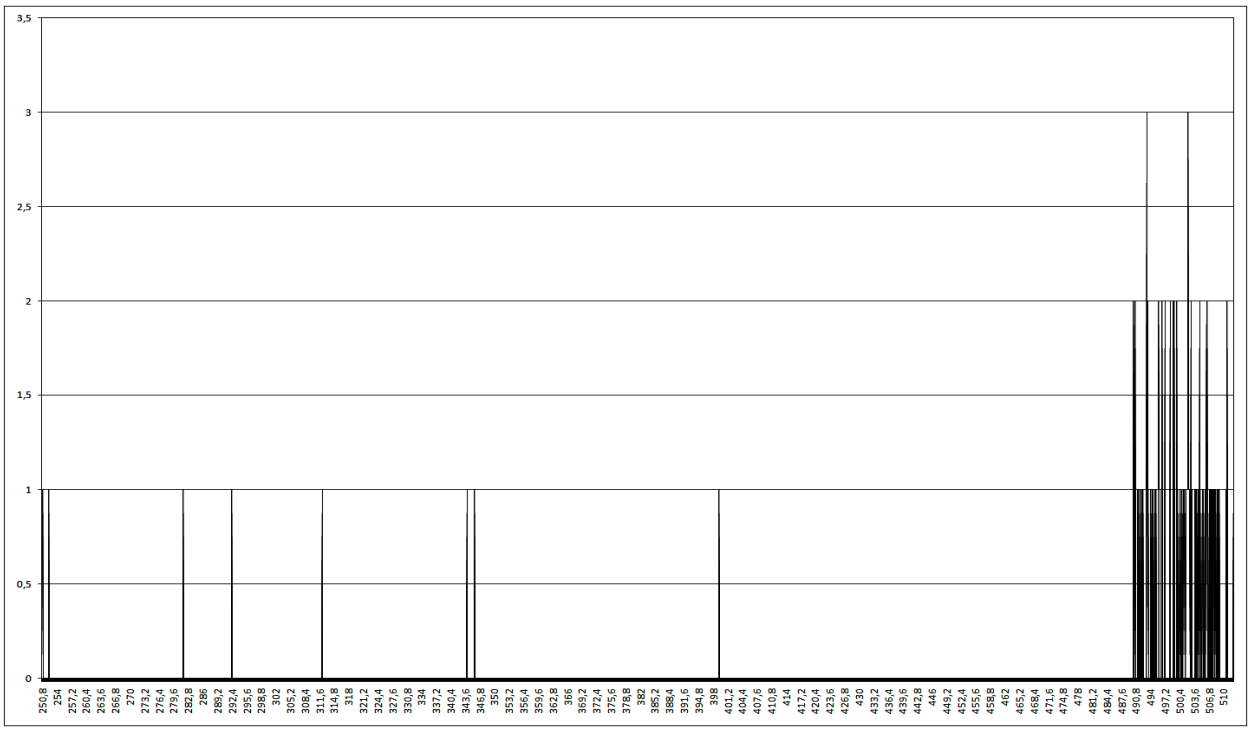


Figure 4. Courbe 1ter

S_{continue} : Stratégie « *continue* »

A partir du tableau de distribution des effectifs regroupés en classes, tracer la courbe polygonale.

- Sous-stratégie « *classes égales toutes non vides* »

Courbe polygonale correspondant au tableau 4 est très asymétrique (forme en J):

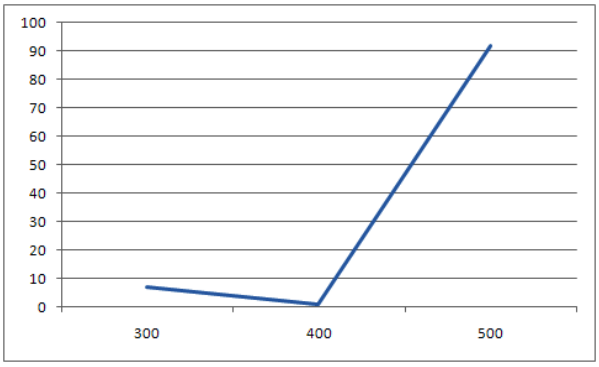


Figure 5. Courbe polygonale 4 : regroupement en classes égales sans classe vide

- Sous-stratégie « *classes égales dont certaines vides* »

Courbe polygonale à partir du tableau 5 :

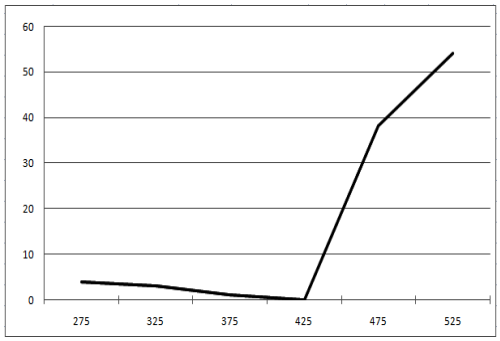


Figure 6. Courbe polygonale 5 : regroupement en classes égales dont une vide

Cette courbe est aussi très asymétrique et sa forme ressemble celle de la courbe 4.

- Sous-Stratégie « *classes inégales* »

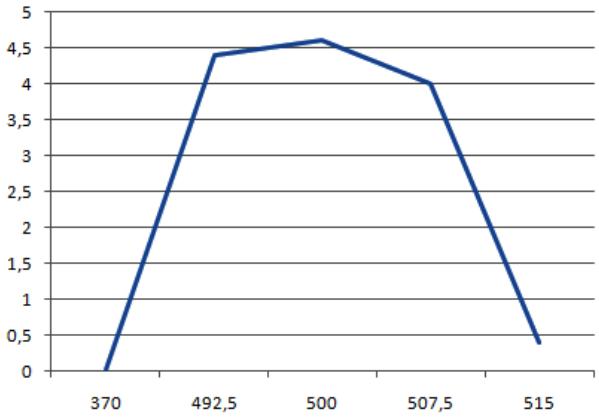


Figure 7. Courbe polygonale 6 : regroupement en classes inégales

La courbe polygonale 6 est quasi-symétrique comme toutes les courbes polygonales présentes dans les manuels.

Question 3

> 3. Que peut-on dire sur le fonctionnement de la machine ? Pour quel poids la machine est-elle, programmée ? En tant que contrôleur, que diriez-vous de la machine ? A-t-elle fonctionné normalement ?

La demande de commenter le fonctionnement de la machine vise à provoquer l'interprétation des données brutes de la série, de sa distribution, de la forme de la courbe polygonale ainsi que la signification donnée au représentant et à la dispersion de la série autour de ce représentant dans une comparaison à une série fictive qui caractériserait un fonctionnement normal de la machine. Le regroupement en classes pourrait conduire à choisir la moyenne comme nouveau représentant de la série car c'est le seul indicateur institutionnel couplé à un calcul de dispersion. Mais le calcul de l'écart-type est très couteux. Pour éviter ce calcul, comment la dispersion sera-t-elle évaluée, autour de quelle valeur ? Quel rôle pourrait-il être donné à la médiane ? La valeur proposée en réponse à la question « Pour quel poids la machine est-elle programmée ? » va-t-elle provoquer une autre interprétation ?

V5 : Faible aplatissement de la distribution (favorise la concentration de beaucoup de valeurs autour de la pointe de distribution → diminution de la dispersion) / Fort aplatissement de la distribution (→ augmentation de la dispersion)

Réponses quantitatives

Pour la question « Pour quel poids la machine est-elle programmée ? », voici les réponses observables :

- 500 g poids d'un paquet de sucre dans le commerce (référence culturelle) ;
- La reprise du représentant retenu à la question 1 (la moyenne ou la médiane ou une autre valeur) ;
- La moyenne calculée à partir du regroupement en classes réalisée dans la question 2.

Réponses qualitatives

Pour les questions :

> - Que peut-on dire sur le fonctionnement de la machine ?
>
> - En tant que contrôleur, que diriez-vous de la machine ? A-t-elle fonctionné normalement ?

Ce sont des stratégies de comparaison à une distribution considérée comme normale pour une machine qui fonctionne bien. Soit elles excluent le hasard soient elles le prennent en compte.

Stratégies « *sans hasard* »

Stratégie « *dispersion* »

Dispersion forte → la machine ne fonctionne pas bien

Dispersion faible (voire nulle : machine idéale) → la machine fonctionne bien

Par rapport à quel indicateur la dispersion est-elle évaluée ? Quels sont les critères pour juger de la dispersion (forte ou faible) ?

Exemple : la machine fonctionne bien puisque 92 % des paquets ont des poids proches les uns des autres (dans un petit intervalle).

-Sous-stratégie « *machine idéale* »

> Tous les poids sont égaux au poids programmé (écart-type nul), sinon la machine ne fonctionne pas bien

- Sous-stratégie « *loi normale* »

> L'intervalle $[\mu - 2\sigma ; \mu + 2\sigma]$ contient ou ne contient pas 95 % des individus → la machine fonctionne bien ou mal.

Stratégie « *forme* »

Symétrie → la machine fonctionne bien

Asymétrie → la machine ne fonctionne pas bien

Stratégies « *avec hasard* »

La machine étant programmée pour une certaine valeur et fonctionnant normalement, le tirage au hasard des paquets peut-il expliquer la distribution obtenue ? La prise en compte du hasard dans l'analyse de la série relativement à la question posée fait passer la statistique du statut de population exhaustive connue à celui d'échantillon issu d'une population inconnue. Une stratégie « avec hasard » doit formuler une hypothèse accompagnée d'un niveau de risque acceptable.

Stratégie « *test d'adéquation à une loi normale* »

Le fonctionnement d'une machine est généralement perturbé par des causes multiples aux effets petits et additifs qui répartissent <u>aléatoirement</u> selon une loi normale les valeurs prises (les poids P) par les individus (ici les paquets) autour de la valeur programmée. Un test d'adéquation à une loi normale consiste à définir une distance entre la distribution empirique et une distribution théorique selon une loi normale dont on a choisi les paramètres. Un test usuellement pratiqué est celui du Chi2 où la distance est la somme des $\frac{(N_i - N_{P_{i_i}})^2}{N_{P_i}}$ pour toutes les classes i, N_i étant l'effectif observé (effectif empirique), $N_{P_{i_i}}$ étant l'effectif

théorique d'une telle classe selon la loi normale choisie. Voici un calcul en partant de la répartition suivante dont les effectifs ne sont pas trop petits :

Classes	Effectifs observés
[250; 490[	8
[490; 495[	22
[495; 500[	19
[500; 505[	27
[505; 510[	20
[510; 515[	4

L'écart-type empirique est 36 (pour une moyenne empirique de 490). Faisons l'hypothèse que la variable aléatoire P suive une loi normale de moyenne 500 et d'écart-type 36 donc que la variable aléatoire $Z = \frac{P-500}{36}$ suive une loi normale centrée réduite. L'effectif de la première classe serait alors $100 \text{ x proba}(P \in [250;490[) = 100 \text{ x proba}(Z \in [-0{,}28;-6{,}94[)$ $\approx 100 \text{ x } 0{,}39 \approx 39$.

Ainsi la part du Khi 2 due seulement à la première classe serait $\frac{(4-39)^2}{39} \approx 31{,}4$.

Comme le Khi2 à 3 degrés de liberté[72] ne dépasse 31 qu'avec une probabilité très faible (inférieure à 1 %), il faut rejeter l'hypothèse sans grand risque.

Stratégie « *test d'hypothèse sur la moyenne* »

Quelle que soit la loi de la variable aléatoire poids (P) si la valeur programmée est 500, on s'attend pour un fonctionnement normal à ce que la moyenne empirique (m) ne s'éloigne pas trop de 500. L'effectif de l'échantillon étant grand ($n \geq 30$) la variable aléatoire $\frac{m-500}{\frac{s}{10}}$ (où s est l'écart-type empirique ici égal à 53) suit approximativement une loi normale centrée réduite. C'est sur base que se construit le test d'hypothèse de la moyenne. Faisons l'hypothèse de base (H_o) que la moyenne de P soit égale à 500 contre l'hypothèse alternative (H_1) que cette moyenne soit inférieure à 500. La zone d'acceptation[73] de l'hypothèse H0, au risque α de la rejeter au profit de H_1 alors qu'elle est vraie, est alors $[500 - 5{,}3 \text{ x } t_\alpha ; +\infty[$ où t_α est le quantile d'ordre 1 - α de la variable normale centrée réduite. Avec un risque de 5 % la zone d'acceptation est $[491{,}25 ; +\infty[$; dans cette zone ne se trouve pas m (m = 485,2). Les données fournies par l'échantillon de paquets autorisent de contester le bon fonctionnement de la machine programmée à 500 avec un risque inférieur à 0,05 de se tromper.

[72] Le nombre de degrés liberté est égale ici à 6 (le nombre de classes) – 3 (le nombre de relations imposées, l'effectif total et les deux paramètres).

[73] Zone dans laquelle doit se trouver la moyenne empirique m pour accepter l'hypothèse de base.

III.4. Analyse a posteriori

Le cas des enseignants :

Voici un tableau à double entrée qui répartit la participation des enseignants d'après les réponses données ou non par eux et d'après leurs acceptations de proposer ou non l'exercice 3 aux élèves.

Répondre aux questions / Proposer l'exercice 3 à l'élève	Oui	Non	Total
Oui	13	10	23
Non	15	29	44
Sans réponse	0	5	5
Total	28	44	72

Tableau 38. Répartition des enseignants

Analyses

Nous nous intéressons aux réponses des 28 enseignants.

1) Question 1

Stratégie discrète				Stratégie classes			
Moyenne	Médiane	Mode	Autres	Moyenne	Médiane	Mode	Autres
13	18	0	0	0	0	0	0

Tableau 39. Répartition des stratégies de l'enseignant pour question 1

Dix enseignants s'abstiennent de répondre à cette question. Ceux qui répondent expriment ce qu'ils attendent d'un élève de classe 10 en faisant mention de la moyenne calculée selon la formule (1) (mais sans fournir les calculs) et de la médiane (quelquefois calculée explicitement par la demi somme des valeurs de rang 50 et 51).

La plupart se conforment à l'algorithme du manuel :

> Enseignant 9 : « 1) La moyenne $\overline{X} =$; $M_e = 500{,}6$
>
> Le représentant est la médiane car les valeurs ne sont pas régulières (grande différence entre 10 premiers paquets et 90 paquets derniers) »
>
> Enseignant 4 : « 1) $\overline{X} = 485{,}173$; $M_e = \dfrac{500{,}4 + 500{,}8}{2} = 500{,}6$
>
> La médiane est le meilleur représentant car les valeurs ont de très grandes différences »

D'autres (moins nombreux) ne suivent pas cet algorithme. Parmi eux certains calculent et choisissent la médiane en faisant valoir d'autres arguments :

Enseignant 42 : « 1) Il faudrait choisir la médiane $M_e = 500{,}6$

Raison : Il y a 100 individus et seulement 8 individus portent les petites valeurs tandis que 92 autres individus portent les valeurs tout près de Me. Même si la différence entre max et min est ½ de Me mais avec 92% individus près de Me alors Me est proche de la constante $\overline{x}$. D'ailleurs, si on calcule la moyenne pour 100 individus, cela prend du temps fou et ce n'est pas nécessaire. »

D'autres encore calculent la moyenne puis la rejette car elle devrait être selon eux plus au centre de la série :

Enseignant 46, 59 : « 1) La moyenne $\overline{X} = \frac{1}{n}\sum_{i=1}^{m} m_i x_i = 485{,}173$;

On voit que la plus part des paquets (92 paquets) dépassent la moyenne. Cette moyenne ne représente pas bien le poids commun de ce groupe. On utilise donc la médiane : $M_e = \frac{500{,}4 + 500{,}8}{2} = 500{,}6$ »

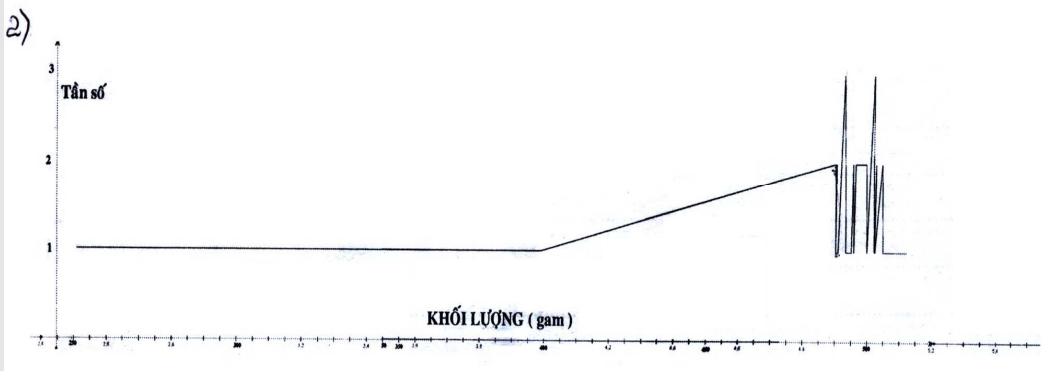
2) Question 2

Stratégie			Nombre
Stratégie « *discrète* »			1
Stratégie « *classe* »	Classes égales	avec classes vides	19
		sans classes vides	0
	Classes inégales	adjacentes	0
		disjointes	0

Tableau 40. Répartition des stratégies de l'enseignant pour question 2

Un seul enseignant emprunte la stratégie discrète. A partir du tableau des effectifs pour la variable considérée comme discrète, il relie tous les points sommitaux d'un diagramme en segments. Il obtient ainsi la courbe ci-dessous qu'il qualifie de courbe polygonale :

Enseignant 50 :



Cette courbe est construite en donnant au tableau des effectifs la signification d'un tableau de distribution en classes dont on « oublie » qu'elles sont disjointes. Elle a donc la continuité de la courbe polygonale définie par le manuel pour une variable continue dont les valeurs sont regroupées dans des classes.

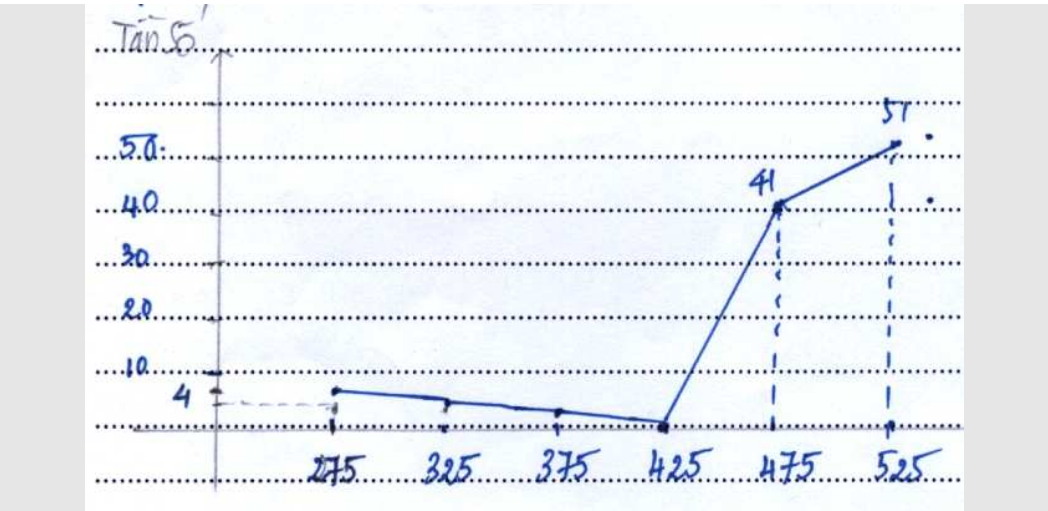
Tous les autres enseignants s’engagent dans le regroupement en classes égales, systématiquement avec la même distribution comportant six classes dont l’une est vide. La voici :

Enseignant 46 : « Tableau de distribution des effectifs regroupés en classes

STT	LỚP	TẦN SỐ
1	[250 ; 300]	4
2	[300 ; 350)	3
3	[350 ; 400)	1
4	[400 ; 450)	0
5	[450 ; 500)	41
6	[500 ; 550)	51
		N = 100

Quelles sont les règles qui dictent la formation de cette distribution ? La simplicité de calcul des valeurs centrales semble être un critère primordial au point de laisser la dernière classe déborder largement au-delà de l’intervalle des valeurs de la série.

Le tableau est suivi de la courbe :



L’absence de réponse de la part des enseignants provient du refus de rompre le contrat didactique qui donne à l’enseignant le charge du regroupement en classes et à l’élève la charge de la réponse correcte, réponse facilement contrôlable par l’enseignant car elle est unique. C’est le sens de la déclaration suivante :

Enseignant 10 : « L’élève ne peut pas tracer la courbe polygonale si l’enseignant ne donne pas les classes. Sinon, il y aurait plusieurs solutions « correctes » qui dépendent du regroupement ».

3) Question 3

Réponse quantitative			Réponse qualitative						
			Stratégies « *sans hasard* »				Stratégies « *avec hasard* »		
			Stratégie dispersion						
500	Autre	Sans réponse	dispersion	« machine idéale »	loi normale	Stratégie forme	Stratégie adéquation à une loi normale	Stratégie test d'hypothèse sur la moyenne	Sans réponse
14	6	8	8	7	0	0	0	2	11

Tableau 41. Répartition des réponses des enseignants pour la question 3

La plupart des enseignants (14 sur 20 réponses) se réfèrent à 500 g et plus encore abandonnent le représentant qu'ils avaient choisi. Leur évaluation de la dispersion est diverse. Voici deux formes fréquentes qui aboutissent à deux conclusions opposées :

> Enseignant 42 : « 3) Cette machine est programmée pour un poids de 500 gr. A partir du diagramme, l'élève trouve que les paquets mal remplis peuvent être enlevés, le nombre des ces paquets est acceptable. Le nombre des paquets bien remplis (proches de 500 gr) occupent un grand pourcentage (92%) et alors on peut dire que cette machine fonctionne normalement ».
>
> Enseignant 58 : « 3) […] Cette machine est programmée pour remplir un paquet du poids de 500gr. Il y a plusieurs paquets dont les poids sont presque la moitié du poids demandé, d'ailleurs, il y a des paquets trop légers (250,8gr) et d'autres dépassent le poids standard. Elle ne fonctionne pas normalement ».

Un seul enseignant fait valoir l'écart-type :

> Enseignant 4 : « 3) $s^2 \approx 22262{,}8$; $s \approx 47{,}56$
>
> Cette machine est programmée pour le poids de 500 gr. En tant que contrôleur, je dirais que cette machine ne fonctionne pas bien car l'écart-type est grand (la dispersion est grande) »

Après avoir choisi la médiane comme représentant cet enseignant juge la dispersion de la série sur l'écart-type[74] (écart standard à la moyenne et pas à la valeur 500 qu'il considère comme la valeur programmée).

Quelque soit l'indicateur de position centrale choisi (on se rappelle que la médiane a été majoritairement choisie), la valeur de dispersion est appréciée en absolu sans critère explicite pour juger si elle est forte ou faible :

> Enseignant 57 : « La machine ne fonctionne pas normalement car il y a une grande différence entre les individus et la médiane ».

[74] Comment a-t-il calculé cet écart-type ? En série discrète, nous avons trouvé 53,776 et en série continue avec les six classes proposées par les enseignants nous avons trouvé 59,21

La différence n'est-elle pas appréciée uniquement en tenant compte des huit paquets de poids les plus faibles ? Et que peut signifier le poids médian dans l'interprétation demandée sur le fonctionnement de la machine ?

Certains enseignants comparent les poids des individus avec le poids programmé et imaginent une machine idéale qui remplirait tous les paquets avec le même poids :

> Enseignant 18 : « La machine est programmée pour le poids de 500 gr. Cette machine ne fonctionne pas exactement par rapport au poids programmé → pas normalement. Il ne faut pas compter sur cette machine »
>
> Enseignant 46 : « La machine est programmée pour le poids de 500gr. Cette machine ne fonctionne pas exactement et non stablement. Les poids des paquets remplis ne sont pas exactement 500gr »

Seuls deux enseignants formulent des réponses qui prennent en compte le fait que la statistique fournie est celle d'un échantillon tiré au hasard :

L'un parle du test d'hypothèse (sur une moyenne égale à 500 ?) en demandant de connaître le risque accepté :

> Enseignant 22 : « La machine est programmée pour remplir un paquet du poids de 500gr. Pour voir si la machine fonctionne normalement ou non, il faut que l'enseignant donne le pourcentage acceptable des produits mal fabriqués pour tester l'hypothèse »

L'autre émet le souhait de connaître le taux de sondage considérant à tort que le prélèvement de 100 paquets pourrait être insuffisant :

> Enseignant 21 : « Fournissez, svp : 100 paquets occupent quel pourcentage de tous les paquets remplis pendant une journée »

Enfin, huit enseignants ne veulent proposer aux élèves cet exercice qu'à condition d'en modifier ainsi l'énoncé :

> 1) calculer la moyenne, la médiane et l'écart-type.
>
> D'après vous, quel indicateur est le bon représentant pour la série ?
>
> 2) Créer le tableau de distribution des effectifs regroupés en classes : [250; 300), [300; 350), [350; 400), [400; 450), [450; 500), [500; 550).
>
> Tracer la courbe polygonale des effectifs.
>
> 3) Commenter : cette machine est programmée pour remplir un paquet de quel poids ? En tant que le contrôleur, trouvez-vous que cette machine fonctionne normalement ?

On retrouve les points forts de notre analyse du contrat didactique relatif au traitement des séries statistiques : calculs des indicateurs à partir des données brutes, choix d'un bon représentant, tracé de la courbe polygonale à partir d'un regroupement donné en classes égales. Seule la présence de l'écart-type pour le choix du représentant sort de ce contrat.

Le cas des élèves :

Il y a au total 207 élèves mais ils ne donnent pas les réponses à toutes les questions posées :

Répondent à la question 1	Répondent à la question 2	Répondent à la question 3
174	146	170

Tableau 42. Nombre de réponses des élèves

Analyses

1) Question 1

Stratégie « *discrète* »				Stratégie « *continue* »			
Moyenne	Médiane	Mode	Autres	Moyenne	Médiane	Mode	Autres
31	75	16	18	29	0	3	3

Tableau 43. Répartition des réponses à la question 1 de 174 élèves

Les choix de représentant sont plus variés que chez les enseignants :

• la médiane

Elle est calculée par demi somme de 500,4 et 500,8 est de nouveau la plus choisie.

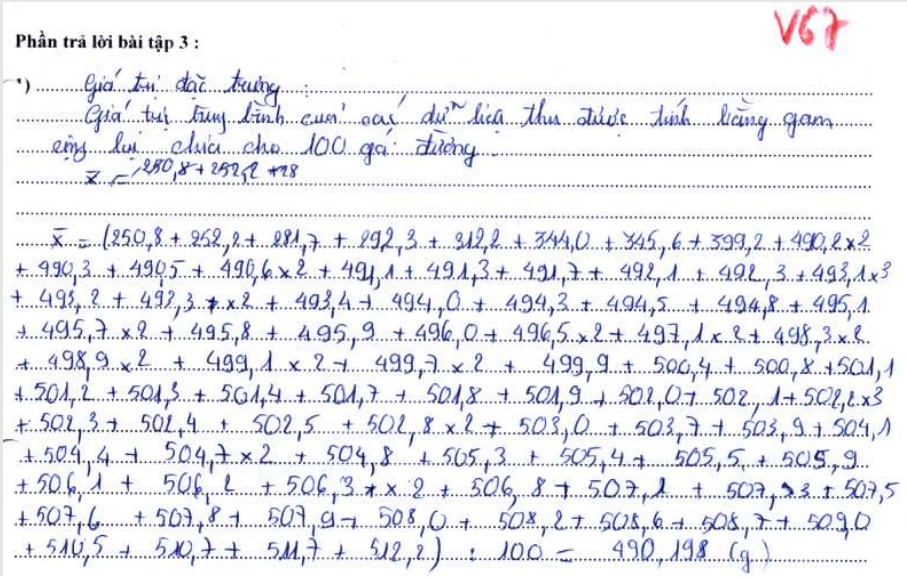
• la moyenne

Elle est calculée suivant les deux stratégies l’une discrète, l’autre continue :

- soit discrète à partir de toutes les données brutes selon la formule (1) ou d’une autre formule notée (1’) et présentée ci-après :

+ Formule (1)

Elève V67 :


V67

Phần trả lời bài tập 3 :

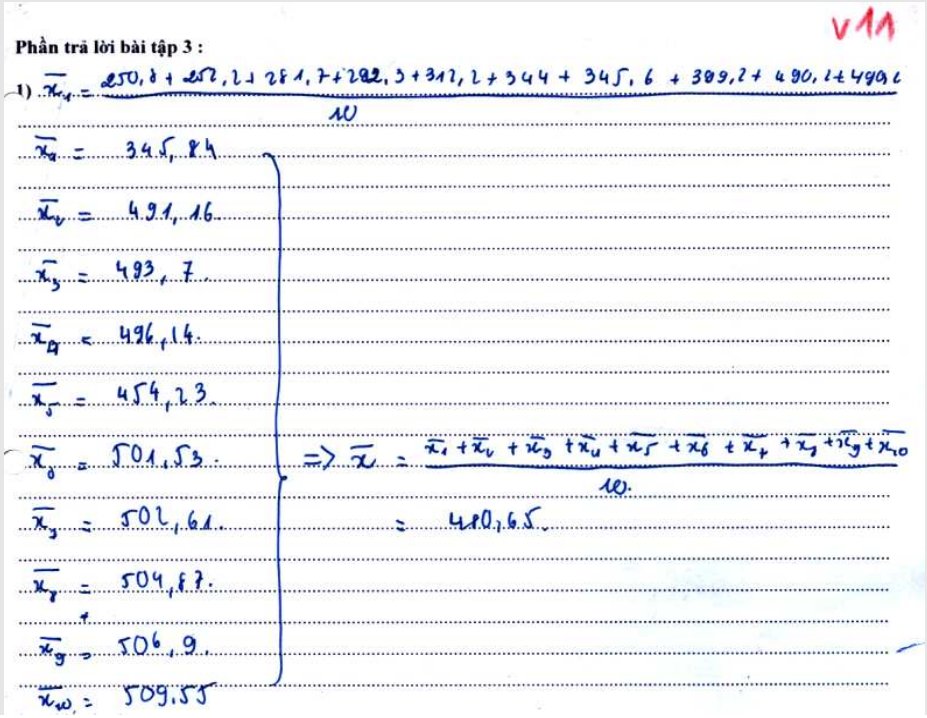
1) Giá trị đặc trưng :
Giá trị trung bình cuối cùng dù lấy thu được tính bằng gam cộng lại chia cho 100 quả trứng.
$\bar{x} = 250,8 + 252,2 + 281,7$

$\bar{x} = (250,8 + 252,2 + 281,7 + 292,3 + 312,2 + 344,0 + 345,6 + 399,2 + 490,2 \times 2 + 490,3 + 490,5 + 490,6 \times 2 + 491,1 + 491,3 + 491,7 + 492,1 + 492,3 + 493,1 \times 3 + 493,2 + 493,3 \times 2 + 493,4 + 494,0 + 494,3 + 494,5 + 494,8 + 495,1 + 495,7 \times 2 + 495,8 + 495,9 + 496,0 + 496,5 \times 2 + 497,1 \times 2 + 498,3 \times 2 + 498,9 \times 2 + 499,1 \times 2 + 499,7 \times 2 + 499,9 + 500,4 + 500,8 + 501,1 + 501,2 + 501,3 + 501,4 + 501,7 + 501,8 + 501,9 + 502,0 + 502,1 + 502,2 \times 3 + 502,3 + 502,4 + 502,5 + 502,8 \times 2 + 503,0 + 503,7 + 503,9 + 504,1 + 504,4 + 504,7 \times 2 + 504,8 + 505,3 + 505,4 + 505,5 + 505,9 + 506,1 + 506,2 + 506,3 \times 2 + 506,8 + 507,2 + 507,3 + 507,5 + 507,6 + 507,8 + 507,9 + 508,0 + 508,2 + 508,6 + 508,7 + 509,0 + 510,5 + 510,7 + 511,7 + 512,2) : 100 = 490,198$ (g)

Cet élève applique la formule (1) et trouve un résultat faux.

+ Formule (1’) par décomposition en sous-séries

Elève V11 :



Cet élève fait une répartition de la série en 10 sous-séries puis calcule la moyenne de chaque sous-série d’après la stratégie discrète, ensuit il arrondit les sous-moyennes obtenues et enfin, il fait la moyenne arithmétique simple de ces moyennes. La formule (1’) utilisée :

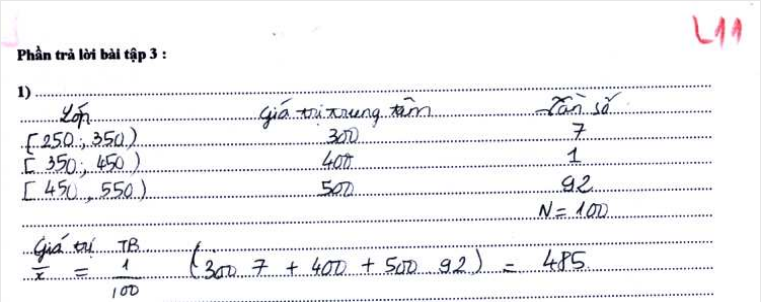
$$\overline{X} = \frac{\dfrac{n_{11}x_{11}+n_{12}x_{12}+\ldots n_{1p}x_{1p}}{N_1} + \dfrac{n_{21}x_{21}+n_{22}x_{22}+\ldots n_{2q}x_{2q}}{N_2} + \ldots + \dfrac{n_{m1}x_{m1}+n_{m2}x_{m2}+\ldots n_{mr}x_{mr}}{N_m}}{m}$$

La série est répartie en m sous-séries et l’intérêt est de diminuer le coût des calculs.

soit à partir des représentants de classes extraites du recueil. Ces représentants sont les valeurs centrales de classes (la formule (2)) ou sont choisis par l’élève (formule (2’) présentée ci-après) :

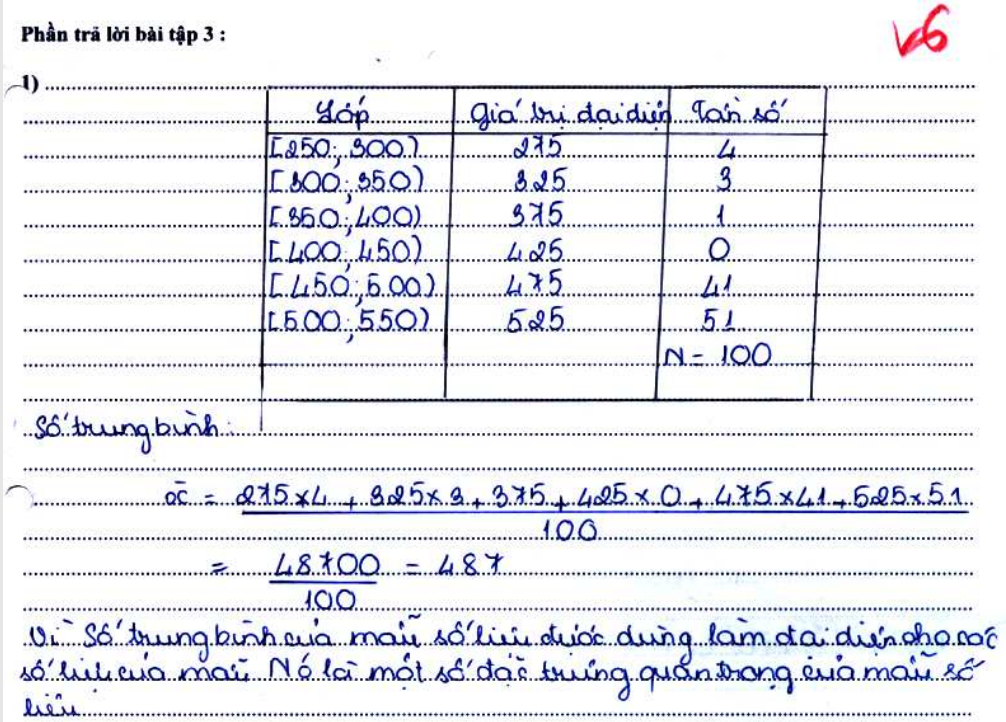
+ Classes égales sans classes vides :

Elève L11 :



+ Classes égales dont certaines sont vides :

Elève V6 : «



Phần trả lời bài tập 3 :

V6

1)

Lớp	Giá trị đại diện	Tần số
[250; 300)	275	4
[300; 350)	325	3
[350; 400)	375	1
[400; 450)	425	0
[450; 500)	475	41
[500; 550)	525	51
		N = 100

Số trung bình:

$$\bar{x} = \frac{275 \times 4 + 325 \times 3 + 375 + 425 \times 0 + 475 \times 41 + 525 \times 51}{100}$$

$$= \frac{48700}{100} = 487$$

Vì Số trung bình của mẫu số liệu được dùng làm đại diện cho các số liệu của mẫu. Nó là một số đặc trưng quan trọng của mẫu số liệu

La moyenne de la série est utilisée pour le représentant. Il est un indicateur important de données ».

Pour chacun d’eux, c’est la formule (2) qui permet le calcul.

- Classes inégales disjointes :

Un élève fait 10 classes disjointes dont est une ligne du tableau de données brutes puis il calcule la moyenne d’après la formule (2). Il termine son choix du représentant en choisissant dans la série la valeur la plus proche de la moyenne trouvée :

Elève M27 :

« D’après moi, la meilleure valeur est :

$$\frac{250{,}8 + 490{,}2}{2} = 370{,}5$$

$$\frac{490{,}3 + 493{,}1}{2} = 491{,}8$$

$$\frac{490{,}1 + 494{,}8}{2} = 493{,}9$$

……..

J’obtiens 487,2

Le meilleur nombre est 490,2 car il est le plus proche »

On trouve ici l'idée que le représentant doit être porté par un individu de la série. Cette idée va être systématisée dans la formule (2') ou on remplace la valeur centrale de chaque classe par une valeur de cette classe qui est proche de valeur centrale.

+ Formule (2')

$$\overline{X} = \frac{n_1\tilde{c}_1 + n_2\tilde{c}_2 + \ldots + n_p\tilde{c}_p}{N} \quad (2')$$

où $\tilde{c}_i$ est la valeur choisie pour remplacer la valeur centrale c_i de la classe par rapport à la formule (2).

Elève M19 :

« Le meilleur représentant :

$$\frac{252{,}2.4 + 345{,}6.4 + 495{,}7.41 + 506{,}3.51}{100} = 485{,}4$$

Nous expliquons ci-dessous les calculs intermédiaires :

- d'abord le regroupement en classes inégales et les positions des valeurs représentantes de chaque classe

250,8	252,2	281,7	292,3	312,2	344	345,6	399,2	490,2	490,2
490,3	490,5	490,6	490,6	491,1	491,3	491,7	492,1	492,3	493,1
493,1	493,1	493,2	493,3	493,3	493,4	494	494,3	494,5	494,8
495,1	495,7	495,7	495,8	495,9	496	496,5	496,5	497,1	497,1
498,3	498,3	498,9	498,9	499,1	499,1	499,7	499,7	499,9	500,4
500,8	501,1	501,2	501,3	501,4	501,7	501,8	501,9	502	502,1
502,2	502,2	502,2	502,3	502,4	502,5	502,8	502,8	503	503,7
503,9	504,1	504,4	504,7	504,7	504,8	505,3	505,4	505,5	505,9
506,1	506,2	506,3	506,3	506,8	507,1	507,3	507,5	507,6	507,8
507,9	508	508,2	508,6	508,7	509	510,5	510,7	511,7	512,2

Tableau 44. Positions des valeurs représentantes dans les classes

- ensuite le tableau de distribution avec la valeur représentante de chaque classe :

Classe	[250,8; 292,3]	[312,2; 399,2]	[490,2; 499,9]	[500,4; 512,2]
Valeur centrale c_i	271,55	355,7	495,05	506,3
Valeur choisie $\tilde{c}_i$	252,2	345,6	495,7	506,3
Effectif	4	4	41	51

Tableau 45. Regroupement en 4 classes inégales et disjointes

- enfin, le résultat obtenu en arrondissant au dixième la moyenne égale à 485,362[75].

Le tableau suivant résume les formules utilisées par l'élève en calculant la moyenne pour la choisir comme le représentant.

Stratégie « *discrète* »		Stratégie « *continue* »	
Moyenne (1)	Moyenne (1')	Moyenne (2)	Moyenne (2')
29	2	28	1

Tableau 46. Répartition des formules de la moyenne selon deux stratégies

Le coût du calcul de la moyenne a donc bien conduit nombre d'élèves à des calculs intermédiaires dont certains sont interprétables en termes de regroupement par classes. Cela démontre qu'il est possible de transférer la responsabilité de ce regroupement en lui fournissant d'autres éléments technologiques de contrôle de ce regroupement.

• Le mode

A la différence de l'enseignant, l'élève choisit le mode comme représentant des données.

- Stratégie « *discrète* »

L'élève trouve qu'il y a deux modes 493,1 et 502,2, il propose de prendre l'un des deux :

> Elève L2 :
>
> « L'effectif de la valeur 493,1 est 3
>
> L'effectif de la valeur 502,2 est aussi 3.
>
> Alor, ces deux valeurs ont le maximum effectif
>
> Choisir une de ces deux valeurs pour le représentant ».

Certains autres proposent le deuxième mode pour le représentant :

> Elève V54 :
>
> « La valeur qui est le meilleur représentant est 502,2 car elle apparait 3 fois »

Un élève propose la moyenne de deux modes :

> Elève L23 :
>
> « Le représentant est 497,7. Je prends la moyenne de deux valeurs dont l'effectif est le plus grand ».

- Stratégie « *continue* » :

Deux élèves (L5, L22) donnent la valeur centrale de classe modale[76] après avoir regroupé la série en 10 classes égales dont l'amplitude est le dixième de l'étendue de la série :

---

[75] Pourtant, quand cet élève tracera la courbe polygonale pour répondre à la deuxième question, il respectera les valeurs centrales des classes pour abscisse des points qui forment la courbe.

[76] Ce terme de classe modale n'est pas introduit dans l'institution vietnamienne.

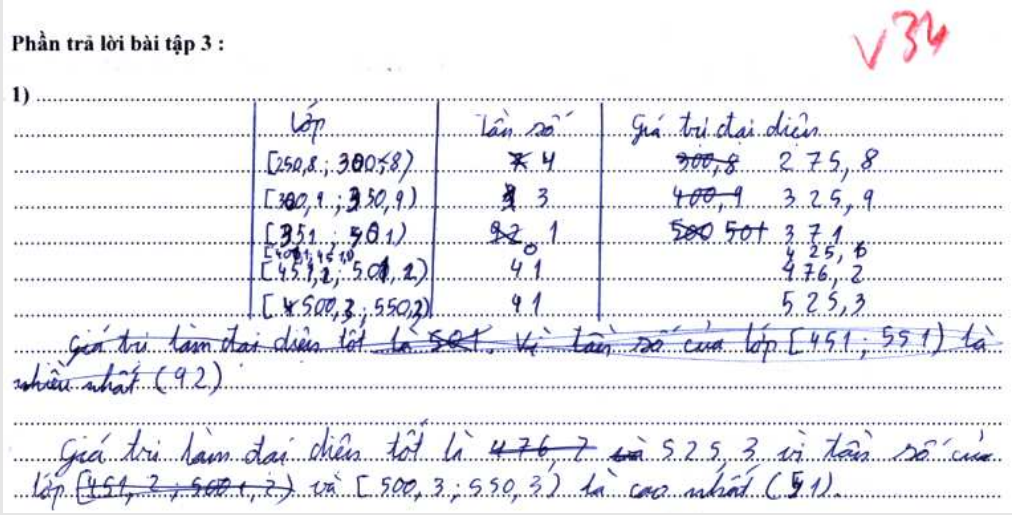
Elève L5 :

« 499,13 est le représentant car classe [486,06; 512,2) est la classe dont l’effectif est le plus grand. Je choisis la valeur centrale de cette classe ».

Un autre élève fait le tableau de distribution des effectifs regroupés en 6 classes égales puis il propose comme représentant la valeur centrale de la classe modale :

Elève V34 :

«



Phần trả lời bài tập 3 :

V34

1)

Lớp	Tần số	Giá trị đại diện
[250,8 ; 300,8)	4	275,8
[300,9 ; 350,9)	3	325,9
[351 ; 401)	1	371
[451,2 ; 501,2)	4	476,2
[500,3 ; 550,3)	4	525,3

~~Giá trị làm đại diện tốt là 501, vì tần số của lớp [451 ; 551) là nhiều nhất (92)~~

Giá trị làm đại diện tốt là ~~476,7~~ ~~và~~ 525,3 vì tần số của lớp ~~[451,2 ; 501,2)~~ vì [500,3 ; 550,3) là cao nhất (4).

La valeur qui est le meilleur représentant est 525,3 car l’effectif de la classe [500,3 ; 550,3) est le plus grand »

L’absence institutionnelle de technique pour choisir les classes conduit à un représentant hors de l’intervalle des valeurs de la série. Ce fait met crûment en lumière que l’interprétation du représentant dans le contexte de l’étude n’est pas une exigence intégrée au contrat didactique.

- autres valeurs :

Trois élèves calculent la moyenne de deux extrémités de la série pour le représentant comme si la série était regroupée en une seule classe (moyenne en stratégie continue, formule (2)) :

Elève M26 :

« Je vais calculer la moyenne de deux valeurs : la plus petite et la plus grande

[…]

Moyenne = $\dfrac{250{,}8+512{,}2}{2}=381{,}5$ »

A l’issue du même calcul, d’autres retiennent la valeur 399,2 de la série qui est la plus proche de cette moyenne (formule (2’).

Un élève choisit l’individu le plus proche à la valeur 500 qui est la réponse à la question 2 :

Elève V28 :

« La valeur 499,9 est le meilleur représentant parce qu’elle est la plus proche à 500 ».

Un élève calcule la moyenne de dix valeurs se trouvant dans la diagonale du tableau pour le représentant (moyenne en stratégie discrète sur un échantillon ou continue en formule (2’)) :

Elève M24 :

« Le représentant est 475,91 (g)

Je prends les valeurs dans la diagonale de gauche à droite, de la plus petite valeur 250,8g à la plus grande valeur 512,2g. J’additionne tous et je fait la moyenne qui est le représentant ».

2) Question 2

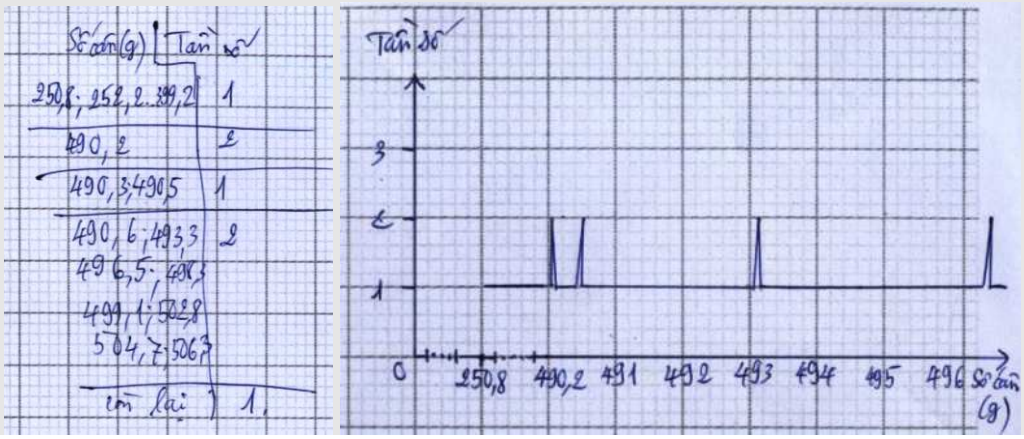
Stratégies			Nb de réponse d’élève
Stratégie discrète			1
Stratégie continue	Classes égales	sans classe vide	63
		avec classes vides	58
	Classes inégales	adjacentes	10
		disjointes	15

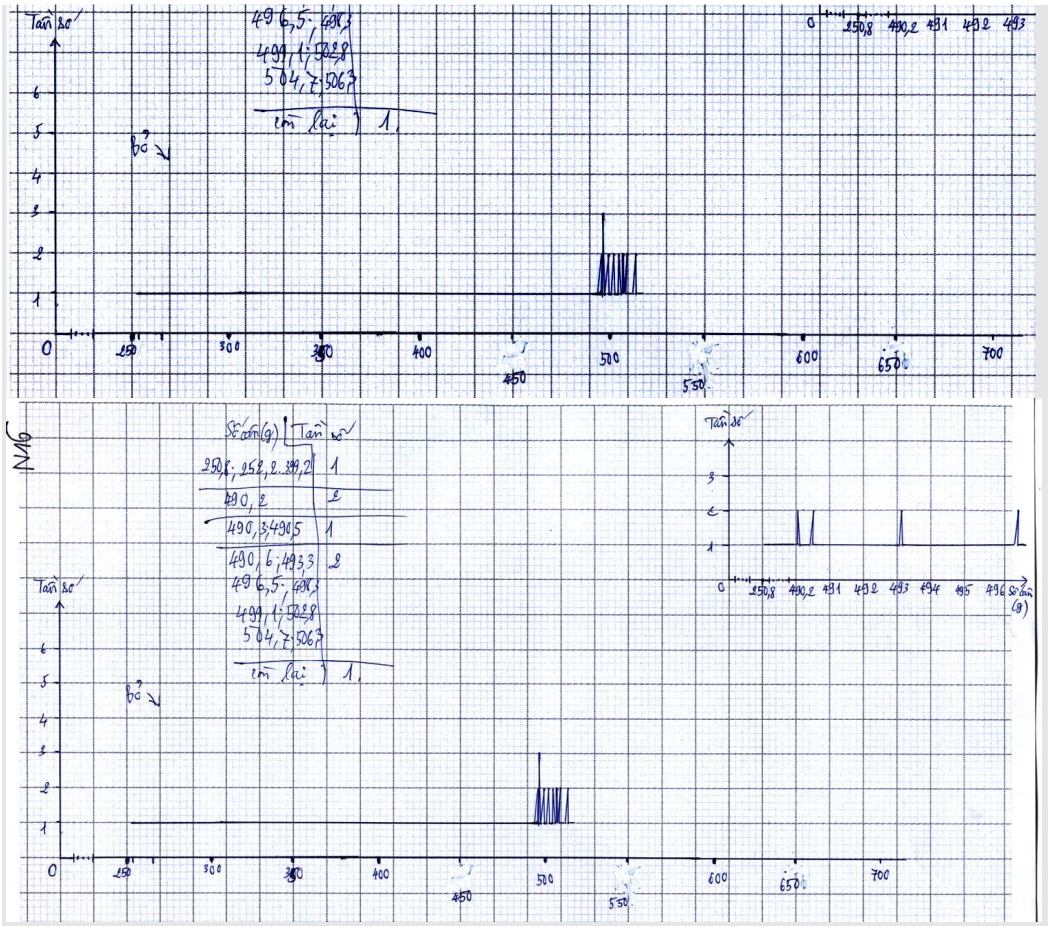
Tableau 47. Répartition des réponses à la question 2

Les cases grisées pour la stratégie conforme à l’l’institution vietnamienne.

Il y a un seul élève utilisant la stratégie discrète pour tracer la courbe polygonale :

Elève N16 :

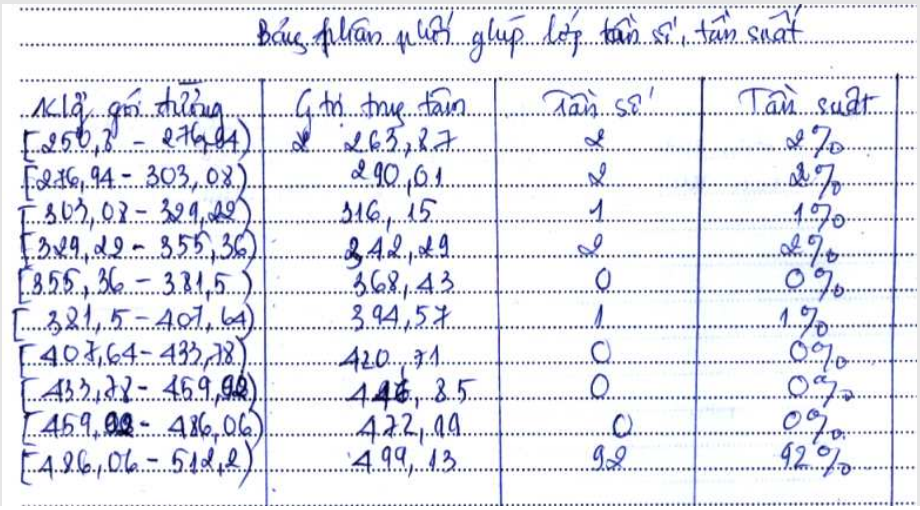




Les élèves font des regroupements soit en classes égales avec ou sans classe vide soit en classes inégales[77]. S’il y a eu le regroupement en classes dans la question 1, ce regroupement est généralement repris pour la question 2.

Contrairement à l’enseignant, l’élève ne limite pas à six le nombre de classes et n’hésite pas à choisir des bornes non entières. Certains acceptent des classes vides :

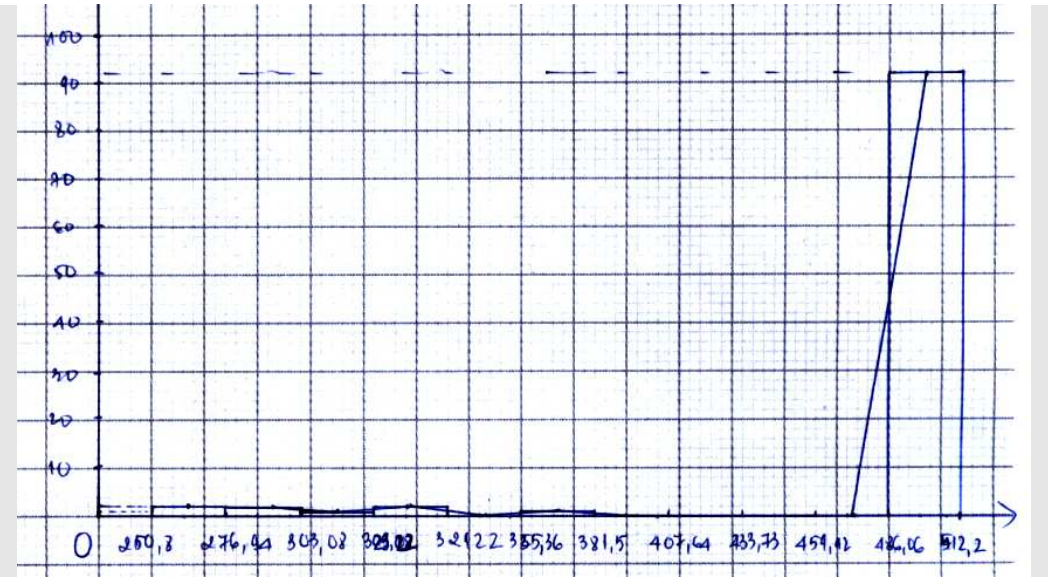
Elève L3 : (10 classes)



Bảng phân phối giúp lớp tần số, tần suất

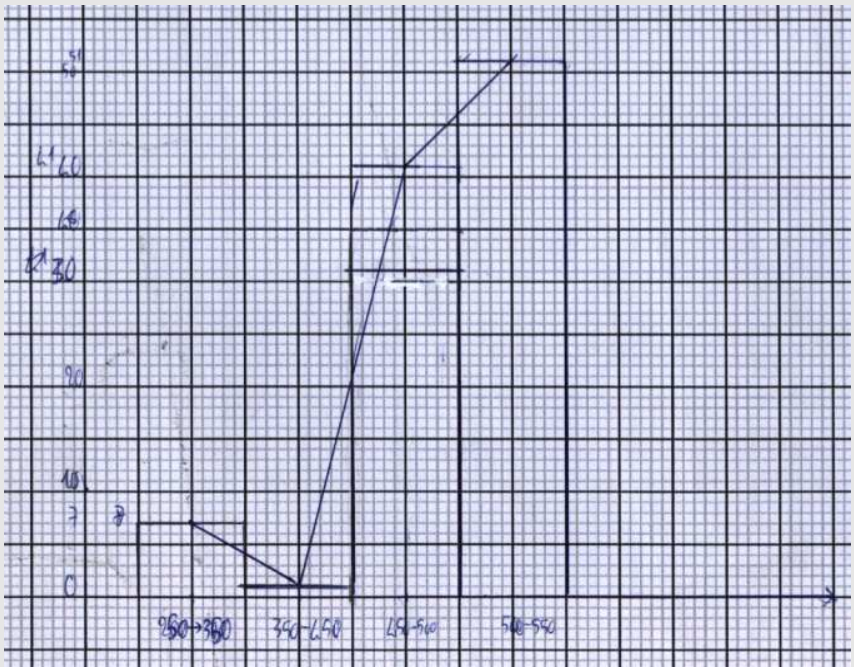
Khg giá trị	G trị trung tâm	Tần số	Tần suất
[250,3 - 276,94)	263,27	2	2%
[276,94 - 303,08)	290,01	2	2%
[303,08 - 329,22)	316,15	1	1%
[329,22 - 355,36)	342,29	2	2%
[355,36 - 381,5)	368,43	0	0%
[381,5 - 407,64)	394,57	1	1%
[407,64 - 433,78)	420,71	0	0%
[433,78 - 459,92)	446,85	0	0%
[459,92 - 486,06)	472,99	0	0%
[486,06 - 512,2)	499,13	92	92%

[77] Nous soulignons ici les différences avec les enseignants.



D’autres évitent les classes vides en diminuant le nombre de classes ou en créant des classes inégales dans le tableau sans pour autant respecter cette inégalité de largeur dans le graphique. Par exemple :

Elève M24 :



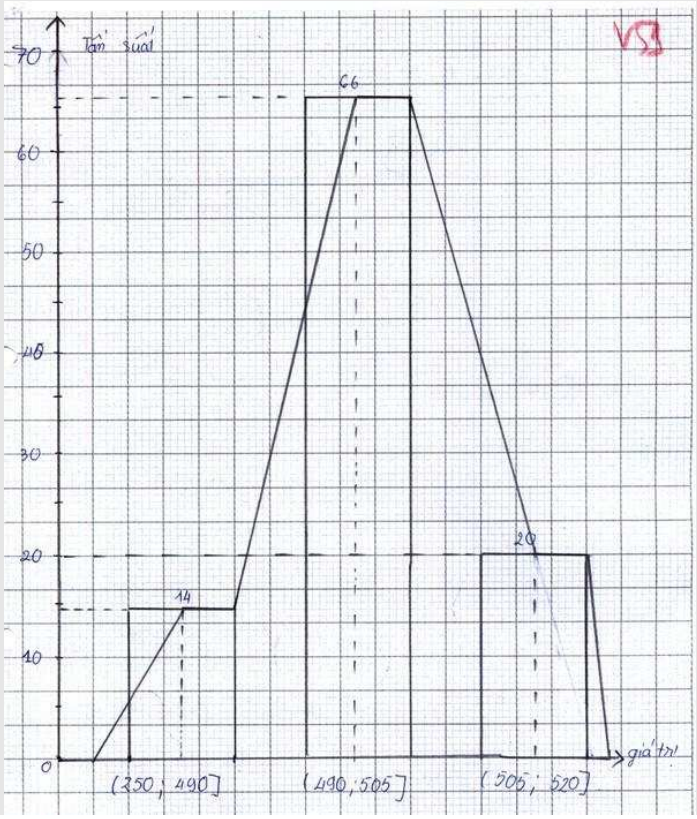
En effet le graphique est fait de classes égales tandis que le tableau de distribution correspondant comporte des classes inégales :

Classe	[250; 350)	[350; 450)	[450; 500)	[500; 550)
Effectif	7	1	41	51

Tableau 48. Regroupement en classes de l’élève M24

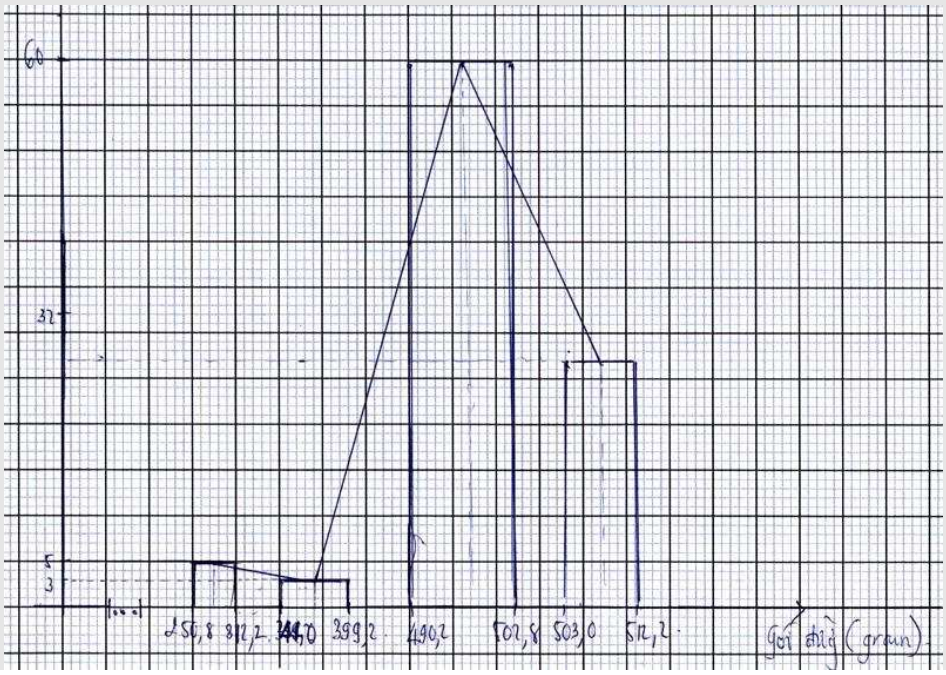
La conformité au graphique institutionnel se retrouve dans cette autre production où l’inégalité des classes est peut-être compensée par la disjonction des rectangles.

Elève V59 :



Apparaît aussi et fortement un regroupement en classes inégales disjointes. Ce regroupement permet d’ignorer les classes vides et s’accompagne d’incohérences entre le tableau et le graphique.

Elève M29 :

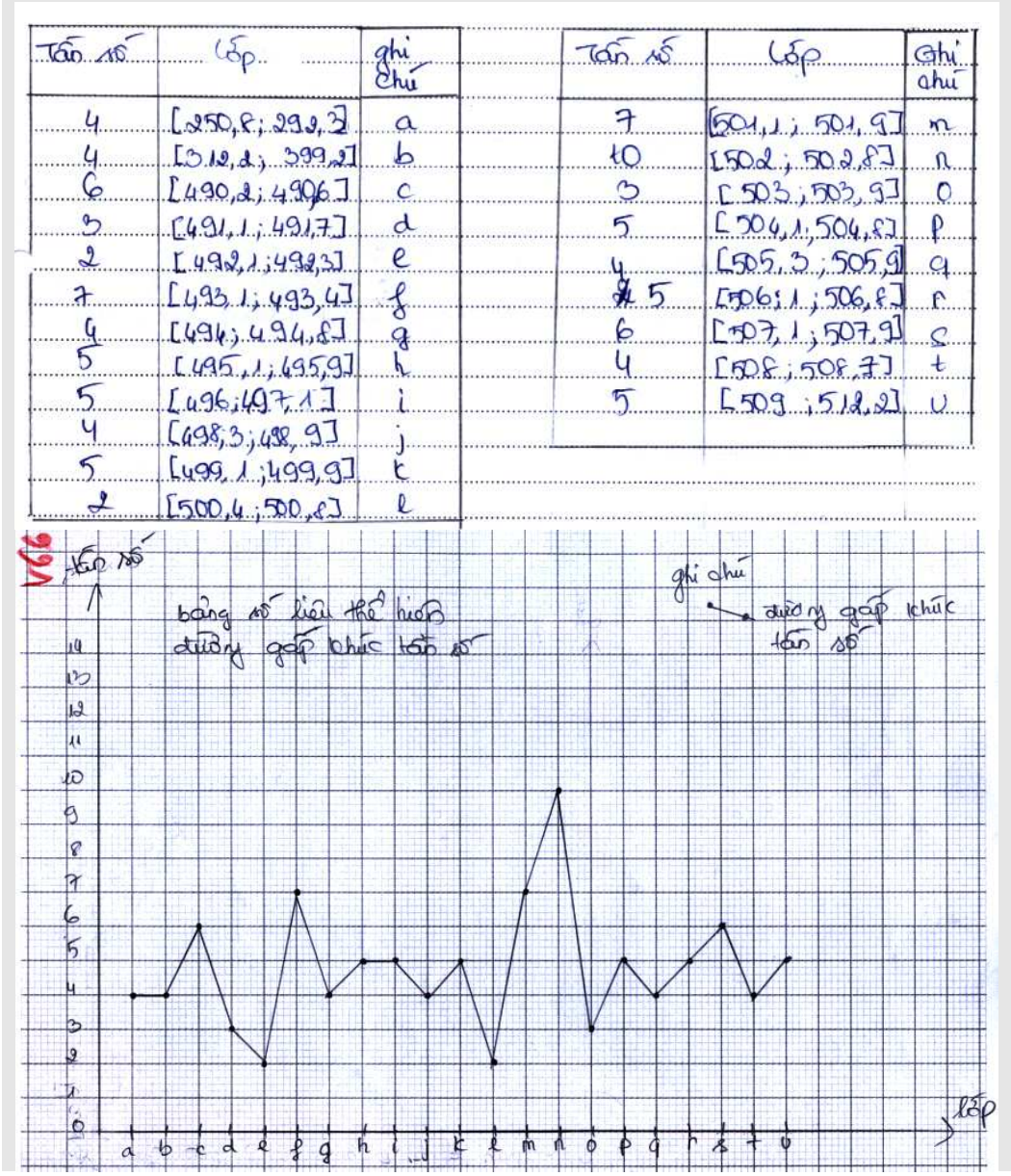


Ce graphique ne fait pas voir de classes vides comme si le tableau de distribution était :

Classe	[250,8; 312,2]	[344,0; 399,2]	[490,2; 502,8]	[503,0; 512,2]
Effectif	5	3	60	32

Tableau 49. Regroupement en classes de l’élève M29

Elève V66 :



Le graphique comporte des points successifs avec des abscisses régulièrement espacées (comme pour des classes égales) alors que le tableau est fait de classes inégales.

3) Question 3 :

Réponse quantitative			Réponse qualitative							
			Stratégies « sans hasard »				Stratégies « avec hasard »			
			Stratégie dispersion			Stratégie forme	Stratégie adéquation à une loi normale	Stratégie test d’hypothèse sur la moyenne	Autre	Sans réponse
500	Autre	Sans réponse	dispersion	« machine idéale »	loi normale					
68	79	55	47	45	0	0	0	0	14	64

Tableau 50. Répartition des réponses de 170 élèves pour la question 3

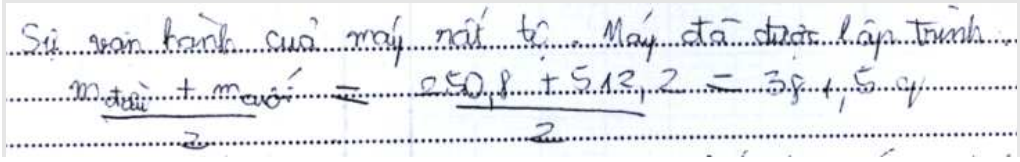
Réponse quantitative (valeur programmée =) :

Contrairement aux enseignants, c’est une minorité d’élèves qui choisissent 500 comme la valeur programmée de la machine et la proportion des « sans réponses » augmente. La plupart des 75 élèves qui avaient choisi la médiane décident de l’abandonner et certains prennent la moyenne. En dehors de 500, la médiane ou la moyenne on trouve aussi :

- soit une des deux valeurs extrêmes (le maximum ou le minimum),

> Elève V15 : « La machine est programmée pour remplir un poids de 512,2 gr »
>
> Elève M5 : « Le poids est programmé pour la machine est de 512,2 gr »

- soit la moyenne de deux valeurs extrêmes de la série,

> Elève M38 : « […] La machine est programmée pour
>
> $$\frac{m_{premier} + m_{dernier}}{2} = \frac{250,8 + 512,2}{2} = 381,5$$
>
> 
> Sử dụng hành của máy nất tố. Máy đã được lập trình
> $\frac{m_{đầu} + m_{cuối}}{2} = \frac{250,8 + 512,2}{2} = 381,5$ g

soit certaines valeurs choisies différentes du représentant sans critère apparent :

> Elève L10 : « La machine est programmée pour le poids de 450 gr »
>
> Elève L16 : « La machine est programmée pour le poids de 400 gr »
>
> Elève L24 : « La machine est programmée pour le poids de 425,8 gr »

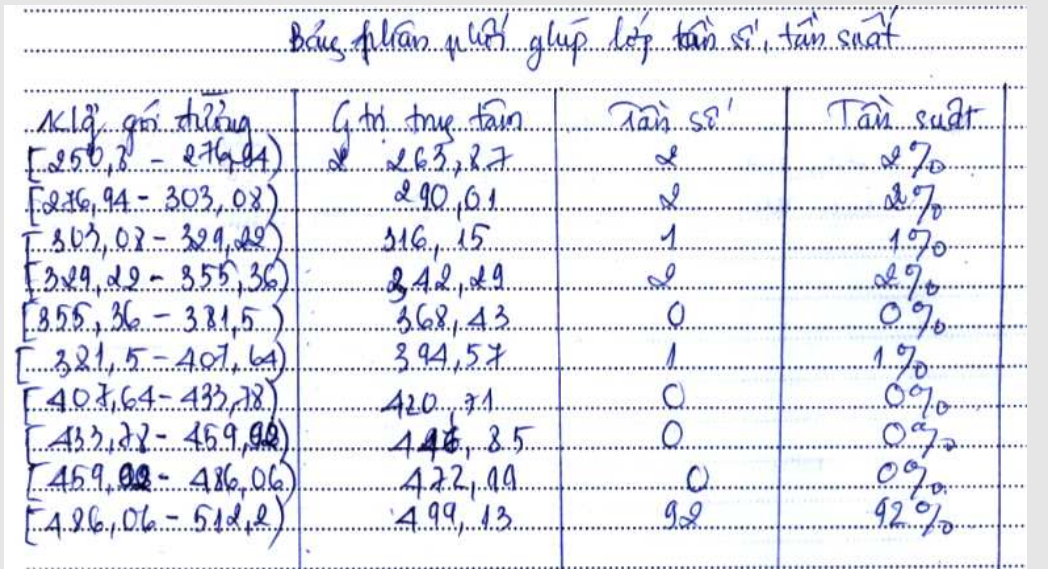
Réponse qualitative :

Aucun élève n’utilise une stratégie « avec hasard ».

La prise en compte de la dispersion de la série conduit comme chez les enseignants soit à accepter un fonctionnement normal soit à le dénier. Cette prise en compte s’exprime autour des idées suivantes :

- un grand nombre d’individus appartenant à un intervalle de centre 500 considérée comme la valeur programmée :

Elève L5 :



Bảng phân phối giúp lớp tần số, tần suất

Khoảng giá trị	Giá trị trung tâm	Tần số	Tần suất
[250,8 - 276,94)	263,87	2	2%
[276,94 - 303,08)	290,01	2	2%
[303,08 - 329,22)	316,15	1	1%
[329,22 - 355,36)	342,29	2	2%
[355,36 - 381,5)	368,43	0	0%
[381,5 - 407,64)	394,57	1	1%
[407,64 - 433,78)	420,71	0	0%
[433,78 - 459,92)	446,85	0	0%
[459,92 - 486,06)	472,99	0	0%
[486,06 - 512,2)	499,13	92	92%

3) La machine fonctionne assez bien, elle est programmée pour remplir un paquet du poids de 500gr car la plupart de poids appartiennent à cet intervalle »

A partir d’un regroupement en dix classes égales, cet élève constate qu’il y a un fort pourcentage de paquets dans l’intervalle]486,6 ; 512,2] qui contient la valeur 500 et pense alors que la machine fonctionne normalement.

Elève V59 : « D’après le tableau, on voit que la plupart des poids des paquets appartiennent à l’intervalle]490 ; 505] et parmi 100 paquets, il y a 90 paquets dont les poids satisfont le standard de production. Il y a environ 10% des paquets des poids inférieurs à 490gr. Donc cette machine fonctionne assez bien »

- un nombre de paquets (les huit premiers) de poids trop faible :

Elève L39 : « La machine est programmée pour un poids de 500gr. Si j’étais le contrôleur je dirais qu’elle ne fonctionne pas normalement car il y a 8 paquets à qui il manque presque 50% du poids tandis que les autres ont une différence de 2,5% »

- un grand nombre de paquets dont le poids dépasse la moyenne :

Elève M10 : « La machine ne fonctionne pas bien. Elle est programmée pour remplir un poids de 462,93gr. Je crois que la compagnie va perdre l’intérêt car l’effectif des paquets dont les poids sont supérieurs à la moyenne est très grand »

- la grande étendue de la série :

Elève L21 : « La machine fonctionne trop mal car il y a une très grande différence entre des valeurs (512,2 – 250,8 = 261,4). Elle est programmée pour remplir le poids de 500gr. En tant que contrôleur, je proposerais de jeter cette machine car il est difficile de réparer une telle mauvaise machine »

- l’écart-type (un seul élève) :

> Elève V17 : « L'écart-type étant environ 53,77 est assez grand → la machine ne fonctionne pas bien »

- la « machine idéale » (pour de très nombreux élèves) :

> Elève L3 : « La machine ne fonctionne pas normalement car les poids des paquets ne sont pas identiques, il y a une grande différence entre eux (le poids maximum est 512,2 et le minimum est 250,8), l'effectif de chaque valeur n'est pas grand (≤ 3)».
>
> Elève N26 : « La machine ne fonctionne pas normalement.
>
> Pour une machine normale, le poids de chacun des paquets doivent être les mêmes »

Des élèves (14) introduisent le temps dans leur analyse du fonctionnement de la machine en considérant que l'ordre des individus de la série est celui du processus de production. Ils commentent :

- soit sur l'augmentation des poids des paquets,

> Elève M2 : « La machine ne fonctionne pas bien. La preuve est que les poids sont croissants selon 100 fois de mesurer ».
>
> Elève M5 : « La machine ne fonctionne pas bien. Au début, elle remplit un paquet du poids de 250,8gr et au centième paquet, le poids augmente 261,4 graphique son poids est donc 512,2gr ».

- soit sur la productivité de la machine,

> Elève N3 : « La machine fait de plus en plus une productivité croissante […. D'après moi, la machine ne fonctionne pas normalement car la productivité n'est pas régulière ».

- soit sur la vitesse de fonctionnement de la machine :

> Elève V14 : « Le fonctionnement de la machine est de plus en plus vite […] Elle ne fonctionne pas normalement, tantôt lentement tantôt vite ».

III.5. Conclusion :

Malgré le coût du calcul de la moyenne comparativement à celui de la médiane, beaucoup d'élèves et d'enseignants choisissent la moyenne comme représentant (hypothèse (H1)). Le grand nombre de valeurs distinctes favorise la conduite d'une stratégie continue qui passe par un regroupement en classes. Les initiatives des élèves dans ce sens démontrent qu'il est possible de transférer la responsabilité de ce regroupement si l'institution lui fournit des éléments technologiques de contrôle de ce regroupement.

La signification possible du représentant dans le contexte étudié n'est pas à l'origine du choix de tel représentant ou de tel autre et elle n'accompagne jamais la réponse fournie à la question 1. A noter cependant que les élèves préfèrent que la moyenne soit portée par un individu de la série statistique[78]. En dehors de ce cas, si justification du choix du représentant il a elle revient entièrement à l'algorithme et aux calculs tant chez les élèves que chez les enseignants (hypothèse (H2)).

[78] Par construction de la médiane, ils acceptent que la médiane ne soit pas une valeur de la série.

La demande formulée par la question 2 de tracer la courbe polygonale conduit à une seule exception près à la fabrication de classes. Le regroupement effectué est le même (six classes égales) pour tous les enseignants comme s’il existait des règles de fabrication de ces classes connues des enseignants et d’eux seuls. En effet les regroupements présentés par les élèves sont très divers tant dans le choix du nombre de classes que dans celui des bornes des intervalles. Cependant même quand ils fabriquent des classes inégales, les élèves produisent des graphiques avec des intervalles égaux respectant ainsi le contrat didactique relatif à la fabrication de la courbe polygonale.

Comme nous nous y attendions, le fait mentionné dans l’énoncé que le sondage avait été réalisé par un échantillon aléatoire simple n’a pas conduit les élèves et les enseignants à questionner l’effet du hasard sur la distribution donnée par rapport à une distribution fictive de machine fonctionnant normalement. Le rôle du hasard dans la production de statistiques pour répondre à des questions sur une population inconnue ne fait pas l’objet d’un enseignement intégré au domaine de la statistique. Les données brutes sont systématiquement traitées comme si elles provenaient d’un recensement exhaustif et toutes les caractéristiques de la population sont alors considérées comme entièrement déductibles, sans risque, du tableau des données.

Nous avons ainsi ébauché une réponse à la question formulée dans le chapitre C1 :

QR4 : Quel est le rapport institutionnel à la notion d’échantillon en fin de classe 10 ? Quel est le rapport institutionnel à la notion d’échantillon représentatif d’une population ?

Dans l’exercice 4, nous élargissons notre étude de la notion de représentativité dans l’institution jusqu’à la notion d’échantillon représentatif issu d’une population connue par une série de données.

IV. Exercice 4

IV.1. Quelques éléments d'analyse institutionnelle vietnamienne à propos de la notion d'échantillon représentatif.

La notion d'échantillon

Dans l'institution vietnamienne, le terme « échantillon » n'apparaît qu'à partir de la classe 10 et seulement dans l'ensemble avancé.

L'ensemble élémentaire utilise l'idée d'enquête statistique pour introduire la notion de recueil des données sur une population et l'expression retenue est « données statistiques » :

> Une étude statistique d'un phénomène économique-sociologique commence par une « enquête statistique » d'un caractère quelconque.
>
> […]
>
> Les données recueillies après avoir enquêté selon un caractère sont appelées données statistiques.

(Livre d'enseignants de classe 10, l'ensemble élémentaire, p. 144)

Le manuel élémentaire reprend cette expression pour désigner les résultats d'une enquête :

> Données statistiques
>
> Lors d'un enquête statistique (selon un objectif prévu), il faut déterminer l'ensemble des individus enquêtés, le caractère étudié et recueillir des données.
>
> Exemple 1
>
> En enquêtant « les productions de riz durant la saison été-automne 1998 » de 31 provinces, on obtient les données écrites dans le tableau suivant :
>
> Productivités du riz de saison d'été-automne (quintal[79] / hectare) 1998 de 31 provinces

30	30	25	25	35	45	40	40	35	45	
25	45	30	30	30	40	30	25	45	45	
35	35	30	40	40	40	35	35	35	35	35

Tableau 1

> […]
>
> Les données dans le tableau 1 sont appelées les données statistiques et aussi les valeurs du caractère.

(Manuel de classe 10, l'ensemble élémentaire, p. 110)

[79] 1 quintal = 100 kilogrammes

Comme il y a 58 provinces, on comprend que ce manuel utilise l'expression « ensemble des individus enquêtés » en laissant entendre que l'enquête n'a pas couvert toutes les provinces.

L'ensemble avancé introduit explicitement la notion d'échantillon au début du chapitre « statistique » :

> Échantillon des données
>
> […]
>
> Pour enquêter le nombre des élèves de chaque classe d'un lycée à Hanoi, l'enquêteur vient dans chaque classe et note le nombre des élèves de cette classe. Voici un extrait de son cahier de notes :

STT	Lớp	Số học sinh
1	10*A*	47
2	10*B*	55
3	10*C*	48
4	10*D*	50
5	10*E*	50
6	11*A*	45
7	11*B*	53
8	11*C*	48
9	11*D*	54
10	11*E*	55

> […]
>
> Un sous-ensemble fini des individus enquêtés est appelé un <u>échantillon</u>. Le nombre des éléments de cet échantillon est appelé <u>la taille de l'échantillon</u>. Les valeurs du caractère obtenues de l'échantillon sont appelées <u>un échantillon des données</u>[80].
>
> Dans l'exemple, l'échantillon est l'ensemble des classes {10A, 10B, 10C, …, 11D, 11E} et l'échantillon des données est {47, 55, 48, …, 54, 55}, la taille de l'échantillon est 10.
>
> Si l'enquête est faite sur tous les individus, c'est une enquête entière. Si l'enquête est faite sur un échantillon, c'est une enquête d'échantillon.

(Manuel de classe 10, l'ensemble avancé, p. 159-160)

Le terme d'échantillon s'utilise ici avec deux significations liées entre elles : un échantillon d'individus qui est une sous-population (sous-ensemble de la population) et un échantillon de données qui est une sous-série des valeurs (et pas un sous-ensemble des valeurs possibles de la variable statistique). Dans l'enquête d'échantillon (on dirait habituellement sondage), comment l'échantillon est-il formé ? N'apparaît pas l'idée qu'un échantillon est le résultat de tirages aléatoires répétés d'individus dans la population-mère plus nombreuse que l'échantillon.

[80] Soulignés par nous.

La notion d'échantillon représentatif

A partir de l'exemple suivant, le manuel élémentaire aborde la notion de représentativité d'un échantillon :

> Exemple 2
>
> Pour préparer les uniformes des élèves, on mesure les tailles de 36 élèves d'une classe 10 et obtient des données statistiques notées dans le tableau suivant :
>
> Les tailles de 36 élèves (unité : cm)
>
> 158 152 156 158 168 160 170 166 161 160 172 173
>
> 150 167 165 163 158 162 169 159 163 164 161 160
>
> 164 159 163 155 163 165 154 161 164 151 164 152
>
> Tableau 2
>
> […]

	Classes de mesure de taille	Effectif	Fréquences (%)	
	[150 ; 156[	6	16,7	
	[156 ; 162[	12	33,3	
	[162 ; 168[	13	36,1	
	[168 ; 174[	5	13,9	
	Total	36	100%	

> Tableau 4
>
> […]
>
> Le tableau 4 nous fournit une base/ un appui pour préciser la quantité des uniformes nécessaires pour chaque taille (correspondant à chaque classe). Par exemple, comme le nombre des élèves dont la taille appartient à la première classe occupe 16,7% d'élèves, la quantité des uniformes de taille correspondant à cette classe occupe donc 16,7% de quantité des uniformes nécessaires. On conclut le même pour des autres classes.
>
> Si cette classe 10 pouvait représenter toute l'école on pourrait appliquer ce résultat pour préparer les uniformes pour tous les élèves de l'école[81].

(Manuel de classe 10, l'ensemble élémentaire, p. 111-113)

L'expression « échantillon représentatif » n'est pas employée mais l'idée formulée dans la dernière phrase est celle de la reproduction possible à l'identique de la distribution des valeurs de la classe sur celle de l'école.

Quand doit-on faire une enquête sur un échantillon au lieu de la faire entièrement sur la population ? Le manuel avancé cite les cas d'une population inaccessible en entier à cause du nombre important d'individus ou de la destruction des individus au cours de l'enquête. Voici l'activité proposée par le manuel avancé :

> H1[82]. Un enquêteur doit enquêter la qualité des boites du lait d'une usine qui traite du lait en ouvrant la boite pour l'examiner. Est-ce que l'on peut faire une enquête entière ?

[81] Soulignés par nous.

[82] Activité 1.

> L'enquête entière n'est pas toujours faisable à cause du trop grand nombre des individus ou parce que quand on veut enquêter, on doit détruire l'individu. On fait souvent une enquête d'échantillon et une analyse de l'échantillon recueilli.
>
> (Manuel de classe 10, l'ensemble avancé, p. 160)

Mais contrairement au manuel élémentaire, ce manuel n'aborde pas le fait que les résultats d'une analyse statistique sur un échantillon pourraient permettre, sous certaines conditions, de conclure sur la population mère toute entière. D'ailleurs le manuel ne dit rien sur la méthode à employer pour réaliser « l'enquête d'échantillon » et donc pour choisir les individus retenus dans l'échantillon.

Le terme « échantillon » est repris dans tous les exercices du manuel avancé proposés aux élèves sans qu'il soit précisé si cet échantillon est représentatif ou non d'une population plus nombreuse. Est-il sous-entendu que l'échantillon étudié est le résultat d'une « enquête d'échantillon » ?

Ainsi les deux ensembles, élémentaire et avancé, utilisent deux expressions différentes, « données statistiques » et « échantillon », en assignant à ces deux expressions la même signification qui est celle d'une extraction de la population mère. Cela correspond à un tirage sans remise dans la population. L'échantillon n'est donc pas conçu comme une série de résultats issus de la répétition d'une expérience aléatoire correspondant à un tirage avec remise dans la population.

Dans aucun des deux ensembles, élémentaire et avancé, on ne rencontre de types de tâche relatifs à la notion d'échantillon représentatif (par exemple « Déduire les caractéristiques de la population mère à partir de celles d'un échantillon », « Comparer les résumés d'un échantillon avec ceux de la population mère connue », « Comparer des échantillons entre eux », etc.).

Les résultats du traitement des données de l'enquête (manuel élémentaire) ou de l'échantillon (manuel avancé) comme les résumés numériques et graphiques ne débouchent jamais sur des questions relatives à la population toute entière.

La représentativité d'un échantillon et la manière d'obtenir un échantillon représentatif à partir du hasard font l'objet de compléments théoriques destinés seulement aux enseignants:

> **Echantillon par hasard**
>
> Dans plusieurs situations, l'enquête entière n'est pas faisable, une enquête d'échantillon est alors nécessaire. Dans la statistique, la question de choisir un échantillon est très importante et très riche. Les méthodes pour choisir un échantillon tel qu'il assure la représentativité dépendent des ensembles des individus et de la spécification du caractère. Un tel échantillon doit refléter le plus possible et le plus fidèlement.
>
> Une méthode très générale est de choisir un échantillon par <u>hasard</u>. Un choix de l'échantillon est appelé choix par hasard si chaque individu est choisi indépendamment de l'autre et qu'il a une même chance d'être choisi.
>
> Un échantillon choisi de telle façon est appelé <u>échantillon aléatoire</u>[83].

[83] Soulignés par nous.

> De nos jours, le fait de faire un échantillon dans une enquête sociologique est très universel. En s'appuyant sur l'analyse des informations obtenues à partir d'un échantillon on peut inférer un « tableau[84] » pour l'ensemble tout entier des individus avec une grande exactitude. Par exemple, avant une élection présidentielle, on fait un sondage sur un échantillon de 1000 personnes choisies au hasard. Alors le pourcentage d'électeurs pour le candidat A dans cet échantillon sera très proche du pourcentage des électeurs pour le candidat A lors de la vraie élection.

(Livre d'enseignants de classe 10, l'ensemble avancé, p. 218-219)

L'idée d'échantillonnage est développée ici et avec elle la production d'un échantillon représentatif d'une population par tirage aléatoire des individus. Mais cette idée est totalement absente des manuels avec lesquels travaillent les élèves.

Résumé et hypothèses sur le contrat didactique

En résumé, la seule qualité présente dans l'institution relativement à l'objet « échantillon représentatif d'une population » est celle d'être un modèle réduit avec moins d'individus de la population mère, fondé sur une distribution identique et la conservation des représentants. Ce rapport à la représentativité d'un échantillon prend appui dans l'institution d'une part sur la proportionnalité dans les tableaux de nombres d'autre part sur les résumés numériques formés à partir des indicateurs considérés comme représentants de la série. Revenant sur la question QR4, nous nous demandons comment un tel rapport à la notion de représentativité pourrait laisser une place à la variabilité résultant de l'échantillonnage. En particulier, comment la ressemblance entre une population connue et un des ses échantillons est-elle appréciée lorsque la représentativité idéale sous-jacente aux hypothèses n'est pas réalisable ? Comment la comparaison des échantillons et de leur population mère permet-elle de mobiliser les notions de proportionnalité et de représentant ? La notion de distance entre les distributions de fréquences peut-elle alors émerger d'une telle comparaison ?

Avec l'intention de construire des éléments de réponse à ces nouvelles questions plus précises que QR4 elle-même, nous mettons à l'épreuve expérimentalement dans la classe 10 deux hypothèses sur les qualités d'un échantillon représentatif par rapport à sa population mère :

H1 : un échantillon représentatif d'une population doit avoir des effectifs proportionnels à ceux de la population mère

H2 : un échantillon représentatif d'une population conserve tous les représentants de la population mère.

C'est le dernier exercice du questionnaire.

[84] Tableau au sens d'image.

IV.2. Présentation de l'exercice

L'exercice propose de classer les échantillons d'une population selon leur capacité à être représentatifs de cette population. C'est un problème de comparaison entre distributions.

Le sujet ci-après est la partie commune distribuée sur feuille à chaque élève et à chaque enseignant :

On a recueilli et ordonné les notes d'un examen de mathématiques des 100 élèves d'un lycée de Hô Chi Minh ville qui ont eu des notes supérieures ou égales à 5.

5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	6	6	6	6
6	7	7	7	7	7	7	7	7	7	7	7	7	7	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	9	9	9	9	9	9	9	9	9	9
9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
9	9	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10

Recueil de notes

Pour communiquer des informations restreintes à comparer avec celles d'autres lycées de Hô Chi Minh ville, on extrait du recueil, des échantillons de 10 notes. On voudrait des échantillons représentatifs de l'ensemble des notes recueillies. Voici cinq échantillons :

6	7	7	8	8	8	8	9	9	10

Échantillon 1

7	7	8	8	8	9	9	9	10	10

Échantillon 2

5	6	9	9	9	9	9	9	10	10

Échantillon 3

5	6	7	7	8	8	8	9	9	10

Échantillon 4

6	6	7	8	8	8	8	9	10	10

Échantillon 5

1. D'après vous, parmi les cinq échantillons, quel est le plus représentatif du recueil tout entier ? Pourquoi ?
2. D'après vous, parmi les cinq échantillons, quel est le moins représentatif du recueil tout entier ? Pourquoi ?
3. Si possible, proposez un sixième échantillon plus représentatif que les cinq échantillons déjà donnés.

Les consignes pour l'élève et pour l'enseignant sont les mêmes que dans les exercices 1, 2 et 3.

Nous appellerons « échantillon 6 » et noterons E6 l'échantillon à proposer dans la dernière question.

IV.3. Analyse a priori

Les variables de situation

1) Les tableaux de données brutes ordonnées se composent de valeurs entières supérieures ou égales à 5. Le choix de seulement six valeurs différentes dans la population mère favorise la fabrication de tableaux d'effectifs ou de fréquences et évite des regroupements en classes qui donneraient lieu à des distributions diverses (cf. exercice 3).

2) Tous les échantillons sont de même taille (10 individus) pour contrarier l'idée que l'échantillon le plus nombreux serait le plus représentatif.

3) Le couple (N ; n) où N est l'effectif total de la population mère et n l'effectif total commun à tous les échantillons est tel que :

- la stricte reproduction de la distribution par proportionnalité n'est pas possible

- la somme des arrondis entiers des Ni x n/N est égale à n et donc l'échantillon résultant de la proportionnalité « par arrondis » est unique[85].

En effet la distribution mère produit un seul échantillon de taille 10 dont chaque effectif est l'entier le plus proche de l'effectif correspondant dans la population mère ramenée à 10 notes :

Notes	5	6	7	8	9	10	Total
Effectifs de la population mère	9	12	13	16	32	18	100
« Effectifs » sur 10 notes (x (n / N))	0,9	1,2	1,3	1,6	3,2	1,8	10
Effectifs arrondis à l'unité	1	1	1	2	3	2	10

Tableau 51. Construction de l'échantillon par proportionnalité + arrondi entier

- les calculs de fréquences des échantillons et de la population mère peuvent se faire mentalement ce qui facilite leur comparaison

- les calculs des effectifs de l'échantillon par proportionnalité à partir ceux de la population-mère peuvent se faire mentalement (N/n = 10).

4) La population mère est asymétrique donc les trois indicateurs de position centrale sont distincts. Cela permet de proposer des échantillons différents dont chacun peut avoir un seul

---

[85] Un exemple :

Notes	**5**	**6**	**7**	**8**	**9**	**10**	**Total**
Fréquences x 10	0,9	**1,3**	**1,4**	**1,4**	3,2	1,8	10
Arrondir à unité près	1	1	1	**1**	3	2	**9**

Les échantillons possibles suivant la stratégie « proportionnalité + arrondi à l'entier » :

Notes	**5**	**6**	**7**	**8**	**9**	**10**	**Total**
Distribution des effectifs de l'échantillon E6-1	1	1	1	**2**	3	2	**10**
Distribution des effectifs de l'échantillon E6-2	1	1	**2**	1	3	2	**10**

indicateur égal ou très proche de celui de la population mère. Dans ce cas, la proximité entre la population-mère et un échantillon sera-t-elle mesurée par l'écart entre deux autres représentants ?

Les variables didactiques

Variable V1 « *représentant* » :

	Étendue	$\overline{X}$	M_e	Mode	σ_n	σ_{n-1}	Ressemblance avec la population mère
Population mère	5	8,04	8,5	9	1,57		
Échantillon 1	4	8,00	8,0	8	1,10	1,15	Moyenne Écart-type éloigné
Échantillon 2	3	8,50	8,5	8	1,02	1,08	Médiane
Échantillon 3	5	8,50	9,0	9	1,57	1,65	Écart-type, Mode
Échantillon 4	5	7,70	8,0	8	1,42	1,49	Etendue
Échantillon 5	4	8,00	8,0	8	1,34	1,41	Moyenne Écart-type proche

Tableau 52. Indicateurs de la population mère et des cinq échantillons

Remarque : dans l'institution vietnamienne, l'écart-type est toujours calculé par la formule:

$$\sigma_n = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

sans distinguer le cas de l'échantillon où l'écart-type (empirique) pourrait être calculé par la formule :

$$\sigma_{n-1} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

Parmi la multitude d'indicateurs susceptibles de servir de représentants, quels sont ceux qui sont privilégiés et comment le choix de ce représentant conduit-il au classement des échantillons et à la fabrication de l'échantillon E6 ?

Variable V2 « *ordre des effectifs* » (ordre total strict / ordre total large)

Dans la population mère, les effectifs sont tous différents et :

effectif de 5 < effectif de 6 < effectif de 7 < effectif de 8 < effectif de 10 < effectif 9

Cet ordre est strict. Dans certains échantillons l'ordre est large avec égalité entre certains effectifs.

Nous formons deux hypothèses :

- le passage d'un ordre strict de la population à un ordre large pour l'échantillon est acceptable

- le changement d'ordre strict n'est pas acceptable.

Par exemple : dans la population mère, effectif de 5 < effectif de 6, alors on accepte que effectif de 5 = effectif de 6 dans un échantillon mais on n'accepte pas l'inversion d'ordre que serait effectif de 5 > effectif de 6.

Nous classerons le cas « effectif de 5 = effectif de 6 » dans le groupe des relations acceptables et le cas « effectif de 5 > effectif de 6 » dans le groupe des relations inacceptables.

Comment l'ordre des effectifs est-il opérationnalisé dans le classement et la fabrication de E6 ?

Variable V3 « *absence de valeurs* »

Dans la population mère toutes les valeurs sont présentes (avec des effectifs non nuls). Dans certains échantillons, certaines valeurs sont absentes. Une absence dans un échantillon perturbe la mise en proportionnalité des effectifs de cet échantillon et de la population mère. Quelle propriété mathématique peut-elle alors soutenir le classement demandé sur la représentativité des échantillons ?

Variable V4 « *rapport d'effectifs* »

Dans la population mère : effectif de 9 = 2 fois d'effectif de 8 (rapports N9/N8 entier) et effectif de 10 = 2 fois d'effectif de 5 (N10/N5 entier). Ces rapports entiers ne se retrouvent pas tous dans les échantillons. Sont-ils considérés comme nécessaires dans l'échantillon E6 ?

Les différents échantillons

Échantillon 1

6	7	7	8	8	8	8	9	9	10

La moyenne de cet échantillon est quasi-égale à celle de la population mère mais avec l'absence de valeur 5. La distribution est symétrique (la moyenne = la médiane = le mode = 8). Cet échantillon ne respecte pas effectif de 8 ≤ effectif de 10.

Échantillon 2

7	7	8	8	8	9	9	9	10	10

Cet échantillon est le seul qui a la médiane de la population mère mais avec absence des deux valeurs 5 et 6. Il respecte cependant l'ordre des effectifs.

Échantillon 3

5	6	9	9	9	9	9	9	10	10

Le mode et l'étendue de cet échantillon sont les mêmes que ceux de la population mère mais l'absence de deux valeurs 7 et 8 modifie sensiblement l'ordre des effectifs. Cet échantillon a la même moyenne que l'échantillon 2 (la moyenne = 8,5) mais son écart-type est plus élevé

(1,65 contre 1,08) et le plus proche de celui de la population mère (1,58) parmi les cinq échantillons. Ainsi du point de vue de la dispersion dans l'institution vietnamienne cet échantillon peut être regardé comme le plus ressemblant à la population mère.

Échantillon 4

5	6	7	7	8	8	8	9	9	10

C'est le seul échantillon ayant toutes les valeurs de la population mère. Pourtant, il n'a aucun indicateur de même valeur que celui de la population mère. Il bouleverse deux fois l'ordre des effectifs puisque :

effectif de 9 < effectif de 8 et effectif de 8 > effectif de 10

Échantillon 5

6	6	7	8	8	8	8	9	10	10

Comme l'échantillon 1, cet échantillon est symétrique autour de 8. Ce qui est le distingue de l'échantillon 1, c'est son écart-type (1,41 contre 1,15) qui est alors plus proche de celui de la population mère (1,58). Il bouleverse 3 fois l'ordre.

Variables	Valeurs	Population mère	E1	E2	E3	E4	E5
Représentant	Moyenne, Mode, Médiane, écart-type, étendue		Moyenne	Médiane	Mode, Ecart-type	Etendue	Moyenne
Ordre	Nombre d'inversions		2	1	4	2	3
Rapports d'effectifs	Certains sont entiers, d'autres non	$\frac{N_9}{N_8} = 2$ $\frac{N_{10}}{N_5} = 2$			$\frac{N_{10}}{N_5} = 2$		

Tableau 53. Les variables didactiques et leurs valeurs

Les stratégies

$S_{représenant}$: Stratégie « *comparaison des représentants des distributions mère et filles* »

Cette stratégie s'inscrit dans le rapport institutionnel qui veut que des séries statistiques sont comparables grâce à leur représentant. On s'attend donc au choix d'un indicateur qui sera considéré comme le représentant ; Après le calcul de ce représentant (moyenne, médiane, mode, étendue, écart-type, etc.) pour la population mère et pour les cinq échantillons, le classement des échantillons du plus représentatif au moins représentatif peut se faire selon la distance de son représentant à celui de la population mère : plus/moins le représentant de l'échantillon est proche de celui de la population mère, plus/moins cet échantillon est représentatif de la population mère. En cas d'égalité, la comparer avec un deuxième indicateur.

Voici les indicateurs et couples d’indicateurs observables :

- Moyenne

  - Les plus représentatifs sont ceux dont la moyenne est la plus proche de celle de la population mère : échantillons 1 et 5.

  - Les moins représentatifs sont ceux dont la moyenne est la plus éloignée de celle de la population mère : échantillons 2 et 3.

- Médiane

  - L’échantillon 2 est le plus représentatif car il a la même médiane que la population mère.

  - Les autres échantillons sont les moins représentatifs.

- Mode

  - L’échantillon 3 est le plus représentatif car il a le même mode que la population mère.

  - Les autres échantillons sont les moins représentatifs.

- Ecart-type (indicateur de dispersion)

  - L’échantillon 3 est le plus représentatif car son écart-type est le même que celui de la population mère.

  - L’échantillon 2 est le moins représentatif car son écart-type est le plus éloigné de celui de la population mère.

- Couple (moyenne; écart-type)

La moyenne est le premier critère pour choisir l’échantillon le plus représentatif. Lorsque les échantillons ont la même moyenne, l’écart-type est le deuxième critère.

- Les plus représentatifs sont ceux dont la moyenne est la plus proche de celle de la population mère : l’échantillon 1 et l’échantillon 5 (ils ont la même moyenne de 8,0. Pour le deuxième critère d’écart-type, choisir l’échantillon 5 car son écart-type est plus proche de celui de la population mère que celui de l’échantillon 1.

- Les moins représentatifs sont ceux dont la moyenne est la plus éloignée de celle de la population mère : échantillon 2 et échantillon 3. Le deuxième critère d’écart-type fait choisir l’échantillon 3 car son écart-type est plus éloigné de celui de la population mère que celui de l’échantillon 3.

En résumé :

<table>
<tr><th>Indicateurs</th><th>Étendue</th><th>Moyenne</th><th>Médiane</th><th>Mode</th><th>Écart type</th><th>(Moyenne, Écart type)</th></tr>
<tr><td>Population mère</td><td>5</td><td>8,04</td><td>8,5</td><td>9</td><td>1,57</td><td></td></tr>
<tr><td rowspan="5">Echantillons « les plus proches »

Echantillons « les plus éloignés »</td><td>E4</td><td rowspan="2">E1, E5</td><td>E2</td><td>E3</td><td>E3</td><td>E5</td></tr>
<tr><td>E3</td><td rowspan="4">E1, E3, E4, E5</td><td rowspan="4">E1, E2, E4, E5</td><td>E4</td><td>E1</td></tr>
<tr><td rowspan="2">E1, E5</td><td>E4</td><td>E5</td><td>E4</td></tr>
<tr><td rowspan="2">E2, E3</td><td>E1</td><td>E2</td></tr>
<tr><td>E2</td><td>E2</td><td>E3</td></tr>
</table>

Tableau 54. La population mère et les échantillons selon les indicateurs

Cette stratégie peut construire un échantillon E6 par modification de l'échantillon considéré comme le plus représentatif ; par exemple si le représentant est le couple (moyenne, écart-type) très utilisé dans l'institution vietnamienne, on modifie l'échantillon E5 pour rapprocher respectivement la moyenne et l'écart-type de la moyenne et l'écart-type de la population mère. Ainsi l'échantillon E5 :

E5	6	6	7	8	8	8	8	9	10	10

devient :

E6r	5	6	7	8	8	8	9	9	10	10

Explication : dans l'échantillon E5, remplacer une note 6 par une note 5, c'est-à-dire diminuer d'un point, il faut donc ajouter un point en remplaçant une note 8 par une note 9 : cela pour conserver la moyenne et pour augmenter l'écart-type.

	Etendue	Moyenne	Médiane	Mode	Écart type
Population mère	5	8,04	8,5	9	1,57
échantillon E5	4	8,00	8,0	8	1,34
échantillon E6r	5	8,00	8,0	8	1,55

Tableau 55. Comparaison des représentants

E6r est plus proche que E5 de la population mère mais ne respecte pas l'ordre.

$S_{proportion}$ Stratégie « *proportionnalité des effectifs* » ou « *égalité des fréquences* »

Effectifs des notes	5	6	7	8	9	10
Population mère	9	12	13	16	32	18
E1	0	1	2	4	2	1
E2	0	0	2	3	3	2
E3	1	1	0	0	6	2
E4	1	1	2	3	2	1
E5	0	2	1	4	1	2

Tableau 56. Distribution des effectifs des notes de la population mère et des échantillons

Les effectifs de chaque échantillon sont comparables à ceux de la population mère :

soit par colonne : égalités deux à deux des rapports $\frac{N_i}{N_j}$ et $\frac{n_i}{n_j}$ (i, j = 5.. 10) où et N_j sont les effectifs des notes i et j dans la population mère ; n_i et n_j sont ceux dans l'échantillon,

Population mère	…	N_i	…	N_j	…	…
Echantillon	…	n_i	…	n_j	…	…

soit par ligne : égalité de tous les rapports $\frac{N_i}{n_i}$ (i = 5.. 9) où N_i est l'effectif de la note i dans la population mère et n_i celui dans l'échantillon.

Population mère	…	N_i	…	N_j	…	…
Echantillon	…	n_i	…	n_j	…	…

Cette stratégie permet de produire un échantillon E6 considéré comme le plus représentatif avec des effectifs quasi-proportionnels aux effectifs de la population mère.

Notes	5	6	7	8	9	10	Total
Effectifs de la population mère	9	12	13	16	32	18	100
« Effectifs » sur 10 notes (x $\frac{n}{N}$)	0,9	1,2	1,3	1,6	3,2	1,8	10
Effectifs arrondis à l'unité	1	1	1	2	3	2	10

Tableau 57. Construction de l'échantillon E6p par proportionnalité + arrondi entier

L'échantillon E6p obtenu :

E6p	5	6	7	8	8	9	9	9	10	10

Cet échantillon est différent de l'échantillon E6r calculé précédemment. Il respecte l'ordre mais sa moyenne est plus éloignée de celle de la population mère.

Cette stratégie ne permet pas de classer les échantillons entre eux. Par contre le critère de proportionnalité peut être remplacé par des critères moins forts basés sur des propriétés résultant de la proportionnalité : le respect de l'ordre des effectifs et celui de la présence des valeurs pondérée par leurs effectifs.

S_{ordre} : Stratégie « *ordre* » (sous-stratégie de la stratégie « *proportion* »)

Choisir l'échantillon qui respecte le mieux/le moins les relations comparatives des effectifs dans la série de population mère.

Rappel de l'ordre dans la population mère :

effectif de 5 < effectif de 6 < effectif de 7 < effectif de 8 < effectif de 10 < effectif 9

Le tableau suivant présente les relations acceptables et celles inacceptables de chaque échantillon :

Echantillon	Relations acceptables	Relations inacceptables
E1	effectif de 5 = 0	effectif de 8 > effectif de 9 effectif de 8 > effectif de 10
E2	effectif de 5 = 0 effectif de 5 = effectif de 6 effectif de 8 = effectif de 9	effectif de 8 > effectif de 10
E3	effectif de 5 = effectif de 6 effectif de 7 = effectif 8	effectif de 5 > effectif de 7 effectif de 5 > effectif de 8 effectif de 6 > effectif de 7 effectif de 6 > effectif de 8
E4	effectif de 5 = effectif de 6 effectif de 7 = effectif 9 effectif de 8 = effectif 9	effectif de 8 > effectif de 9 effectif de 8 > effectif de 10
E5	effectif de 5 = 0 effectif de 6 = effectif de 10 effectif de 7 = effectif de 9	effectif de 6 > effectif de 7 effectif de 8 > effectif de 9 effectif de 8 > effectif de 10

Tableau 58. Des relations acceptables et celles inacceptables des échantillons

Un classement possible des échantillons peut consister à compter le nombre de relations inacceptables. D'après ce tableau, l'échantillon E2 est le plus représentatif car il ne produit qu'une seule relation inacceptable et l'échantillon E3 est le moins représentatif à cause du plus grand nombre de relations inacceptables. Un autre classement pourrait prendre en compte l'importance de l'ordre effectif de 8 < effectif de 10 < effectif 9 qui concerne les valeurs les plus nombreuses, ce qui pourrait donner le couple (E2 ; E5). Ce classement serait un observable d'une stratégie que nous qualifierons d'ordre pondéré.

Une stratégie d'ordre permet de proposer plusieurs échantillons E6 sans relation inacceptable. Ils seront différents et ne pourront être départagés par l'ordre. Il y a l'obligation d'un ordre large à la place de l'ordre strict, avec plus ou moins d'égalités. Voici deux échantillons E6 possibles :

5	6	7	8	8	9	9	9	10	10

(effectif de 5 = effectif de 6 = effectif de 7 < effectif de 8 = effectif de 10 < effectif de 9)

6	7	8	8	9	9	9	10	10	10

(effectif de 5 < effectif de 6 = effectif de 7 < effectif de 8 < effectif de 10 = effectif de 9)

$S_{valeurs}$: Stratégie « *présence/absence de valeurs* » (sous-stratégie de la stratégie « *proportion* »)

Pour être représentatif un échantillon doit afficher toutes les valeurs de la population mère. L'échantillon 4 est le seul qui satisfait ce critère et il est donc le plus représentatif. Par

contre, deux échantillons, E2 et E3 sont candidats selon cette stratégie à être le moins représentatif car il leur manque à chacun deux valeurs par rapport à la population mère. Comment le choix se fera-t-il entre E2 et E3 ? S'agira-t-il de respect l'étendue ? Ou le poids des effectifs dans la population mère des valeurs absentes de l'échantillon interviendra-t-il comme moyen de «respecter au mieux » la proportionnalité des effectifs ?

Comme la sous-stratégie « ordre » cette stratégie ne permet pas de proposer un échantillon E6 unique.

$S_{\text{fréquence}}$: Stratégie « *distance entre les distributions de fréquences* »

Fréquences des notes	5	6	7	8	9	10
Population mère	0,09	0,12	0,13	0,16	0,32	0,18
E1	0,00	0,10	0,20	0,40	0,20	0,10
E2	0,00	0,00	0,20	0,30	0,30	0,20
E3	0,10	0,10	0,00	0,00	0,60	0,20
E4	0,10	0,10	0,20	0,30	0,20	0,10
E5	0,00	0,20	0,10	0,40	0,10	0,20

Tableau 59. Fréquences des notes de la population mère et des échantillons

Le passage des distributions d'effectifs aux distributions de fréquences est mis en avant dans les manuels dans des tâches de comparaison. Comment mener la comparaison entre les distributions de fréquences ? La réponse institutionnelle est, dans le cas de séries de valeurs regroupées en classes, l'analyse de la position relative des courbes polygonales.

- Sous-stratégie « *graphique* »

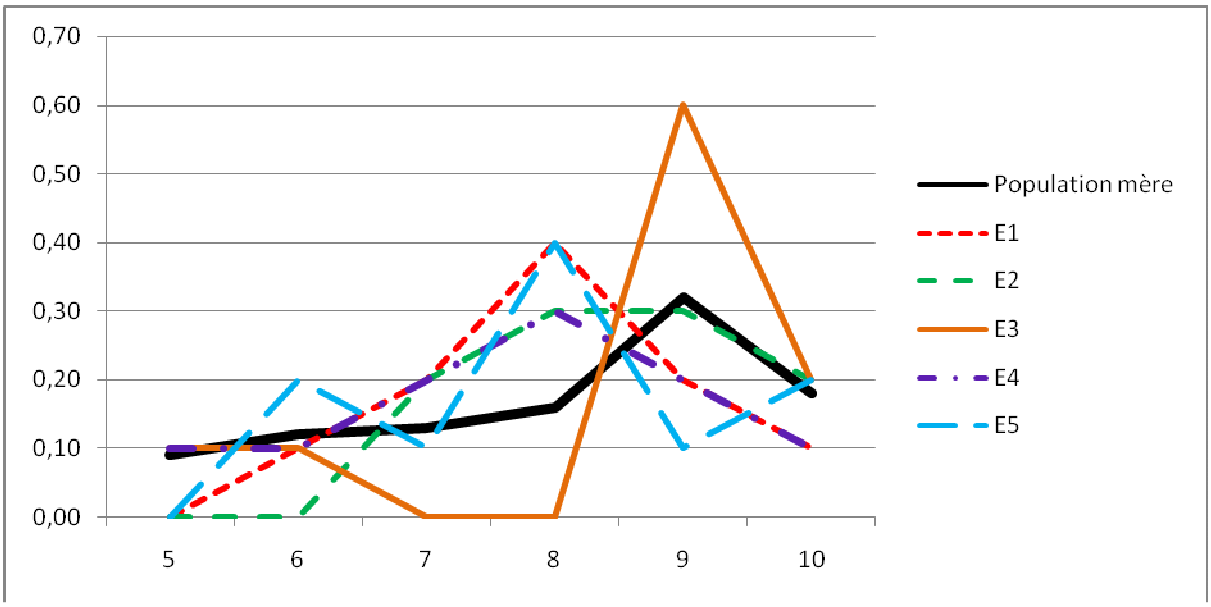


Figure 8. Courbes polygonales des fréquences de la population mère et de 5 échantillons

Tracer les courbes polygonales de fréquences de la population mère et des cinq échantillons. Chaque courbe échantillon est plus ou moins éloignée de celle de la population mère. L'éloignement s'apprécie soit par l'ensemble des distances entre les sommets correspondants soit par la surface qui sépare les deux courbes.

Cette stratégie n'est pas une stratégie attendue car dans l'institution vietnamienne, d'une part la courbe polygonale exige le regroupement en classes et d'autre part elle n'est utilisée à la demande de l'enseignant que pour comparer deux distributions.

- Sous-stratégie « *vectorielle* »

Entre distributions discrètes, la comparaison peut se mener sur les vecteurs-fréquences : calculer la « distance » entre le vecteur-fréquence de la population mère et celui de chaque échantillon selon une certaine norme. Puis proposer le sixième échantillon qui minimise la distance avec la population mère.

Utilisons par exemple les trois normes habituelles de R^n :

Norme 1 : $\| \|_1 = \sum_i^n \left| f_{i_{\text{échantillon}}} - f_{i_{\text{polulation mère}}} \right|$

Norme 2 : $\| \|_2 = \sqrt{\sum_i^n \left| f_{i_{\text{échantillon}}} - f_{i_{\text{polulation mère}}} \right|^2}$

Norme 3 : $\| \|_3 = \sup_i \left| f_{i_{\text{échantillon}}} - f_{i_{\text{polulation mère}}} \right|$

Voici le tableau des distances selon ces trois normes :

Echantillons	Norme 1	Norme 2	Norme 3
E1	0,62	0,30	0,24
E2	0,46	0,22	0,14
E3	0,62	0,35	0,28
E4	0,44	0,22	0,14
E5	0,68	0,35	0,24

Tableau 60. Distances des vecteurs fréquences entre échantillons et population mère

(Dans chaque colonne, les cases grisées correspondent aux plus petits écarts et les cases graves correspondent aux plus grands écarts)

Classement des échantillons	Norme 1	Norme 2	Norme 3
le plus proche de la population mère	E4	E4	E4
	E2	E2	E2
	E1	E1	E1
	E3	E3	E5
le plus éloigné de la population mère	E5	E5	E3

Tableau 61. Ordre croissant des échantillons d'après les distances

Le plus représentatif est l'échantillon 4 (le moindre écart des fréquences entre lui et la population mère, d'après les trois normes) ; les moins représentatifs sont l'échantillon 3 et l'échantillon 5.

On construit l'échantillon E6 en modifiant E4 pour réduire l'écart entre E4 et population mère selon la norme. Voici les résultats.

Selon les normes 1 et 2 :

5	6	7	8	8	9	9	9	10	10

On retrouve l'échantillon E6p.

Selon la norme 3 :

$E6_{f3}$	5	6	7	7	8	8	9	9	9	10

Cet échantillon ne respecte pas l'ordre.

Echantillons	Norme 1	Norme 2	Norme 3
E4	0,44	0,22	0,14
$E6_{f1}$	0,14	0,06	0,14
$E6_{f3}$	0,24	0,12	0,08

Tableau 62. Distances des échantillons E6f et E4 avec la population mère

Synthèse des stratégies

Représentants	Proportions entre effectifs ou égalité des fréquences		Distance entre distributions des fréquences
	Ordres	Valeurs	
Idéal = même représentant	Idéal = proportionnalité des effectifs ou égalité des fréquences		Idéal = égalité des vecteurs et des graphiques

Absence de distance ⟶ Présence de distance

E6											Nb. de bouleversements d'ordre	Distances par rapport à population mère				
												Moyenne	Écart type	Norme 1	Norme 2	Norme 3
E6r	5	6	7	8	8	8	9	9	10	10	2	0,04	0,07	0,34	0,19	0,28
E6p	5	6	7	8	8	9	9	9	10	10	0	0,06	0,10	0,14	0,06	0,08
E6f3	5	6	7	7	8	8	9	9	9	10	1	0,24	0,02	0,24	0,12	0,04

Tableau 63. Les différents échantillons E6 comparés à la population mère

Remarque : dans le tableau précédent, E6r est l'échantillon idéal conçu selon la stratégie représentant mais son écart-type est plus éloigné de celui de la population mère par rapport à celui de l'échantillon E6f3 issu de la stratégie vecteurs-fréquences.

Synthèse des stratégies attendues

Dans le tableau ci-après, nous présentons les couples $(E_i\,;E_j)$ où E_i est le plus représentatif et E_j est le moins représentatif d'après les stratégies attendues :

Stratégies	Représentant				Proportion	
	$\overline{M}$	σ	M_e	Mode	Ordre	Valeurs
Le plus représentatif	E1, E5	E3	E5	E3	E2	E4
Le moins représentatif	E2, E3	E2	E2	E1, E2, E4, E5	E3	E3/E2
Echantillon idéal E6			E6r		E6p	

Tableau 64. Les couples (le plus représentatif ; le moins représentatif)

selon les stratégies attendues

IV.4. Analyse a posteriori

Enseignant

Voici un tableau à double entrée qui répartit les enseignants d'après la présence ou l'absence de réponses et d'après leur acceptation ou non à proposer l'exercice 4 aux élèves.

Répondre aux questions / Proposer l'exercice 4 à l'élève	Oui	Non	Total
Oui	14	4	18
Non	16	32	48
Sans réponse	3	3	6
Total	33	39	72

Tableau 65. Répartition des enseignants

Ainsi les enseignants considèrent majoritairement que l'exercice 4 est très éloigné de leurs attentes relativement à la notion de représentativité. Ce fait traduit-il le refus d'attribuer une ordinalité à la représentativité d'un échantillon ? Cela confirmerait nos hypothèses (H1) et (H2) qui réduisaient la relation population-échantillon représentatif à la stricte proportionnalité des distributions d'effectifs ou au strict copiage de la distribution des fréquences. N'oublions pas non plus que les tâches de comparaison n'occupent pas une place importante dans la statistique enseignée et que cet exercice. Or cet exercice s'inscrit pleinement dans une problématique de comparaison entre distributions.

Nous nous intéressons aux réponses des 33 enseignants.

Le plus représentatif					
E1	E2	E3	E4	E5	Total
2	3	6	8	5	24

Le moins représentatif					
E1	E2	E3	E4	E5	Total
1	13	3	7	0	24

Proposition		
E6p	Autres	Total
19	5	24

Tableau 66. Les choix des échantillons représentatifs de 33 enseignants

(Ces 33 enseignants ne donnent pas toujours toutes les réponses pour chaque question posée.)

Aucun enseignant n'utilise la stratégie vecteurs-fréquences, aucun la stratégie graphique et aucun la stratégie d'ordre. Voici un tableau récapitulatif des couples choisis :

Le moins représentatif / Le plus représentatif	E1	E2	E3	E4	E5	Total
E1		1	0	1 (1 E6p)	0	2
E2	1 (1 E6p)		1 (1 E6p)	0	0	2
E3	0	0		6 (1 E6p)	0	6
E4	0	7 (4 E6p + 2 E6f3)	0		0	8
		1 (1 E6p)				
E5	0	4 (3 E6p + 1 E6r)	0	1 (1 E6p)		5
Total	1	13	1	8	0	23

Tableau 67. Les couples (le plus représentatif ; le moins représentatif) choisis par les enseignants

Des stratégies « représentant »

Elles sont visibles à travers les lignes E1, E3 et E5 : 13 enseignants sur 23 ce qui est une forte proportion.

- avec la moyenne

> Enseignant 41 :
>
> « L'élève calcule la moyenne des notes de 100 élèves et celles de chaque échantillon puis les compare. »

- avec le couple (moyenne ; écart-type) : (E5 ; E2)

> Enseignant 9 :
>
> 1) L'échantillon 5 car sa moyenne et sa variance sont les plus proches de celles du tableau des données.

2) L'échantillon 2 car sa moyenne et sa variance sont les plus éloignées de celles du tableau des données.

3) 5, 6, 7, 8, 8, 9, 9, 9, 10, 10 [86]

Enseignant 22 : « A partir du tableau des données de 100 élèves, on obtient : la moyenne $\overline{X} = 8{,}04$ (la médiane $M_e = 8{,}5$). L'écart-type : S ≈1,5. Parmi 5 échantillons, l'échantillon 5 représente mieux les données entières et l'échantillon 2 est le moins représentatif parce que :

+ L'échantillon 1 et l'échantillon 5 ont : $\overline{X} = M_e = 8$ qui est presque celui du tableau des données (n=100) mais l'écart-type de l'échantillon 5 est S ≈ 1,31 qui est plus grand que celui de l'échantillon 1 (S ≈ 11). Cela veut dire que l'échantillon 5 représente mieux la dispersion de données du tableau (n=100) dont l'écart-type est S ≈ 1,57

+ L'échantillon 2 est le moins représentatif car il a : $\overline{x} = 8{,}5 ; M_e = 8{,}5 ; S \approx 1{,}1$. Cet échantillon ne représente pas bien la situation réelle de l'étude des élèves (il n'y a que les grandes notes et la dispersion est faible).

On peut proposer un sixième échantillon qui représente au mieux que les cinq échantillons donnés : 5, 6, 7, 8, 8, 8, 9, 9, 10, 10[87] parce que l'échantillon 6 a : $\overline{X} = M_e = 8 ; S \approx 1{,}55$ »

- avec le triplet (écart-type ; moyenne ; médiane) : (E3 ; E4)

Enseignant 64 : «

Calculer les indicateurs des données :

: Dĩ lich chuẩn ; S = 1,57 ; Số trung bình : $\bar{x}$ = 8,04 điểm
Số trung vị Me = 8,5 điểm

Faire les mêmes pour les échantillons 1, 2, 3, 4, 5. On obtient :

+ Mẫu 1 : S = 1,1 ; $\bar{x}$ = 8 ; Me = 8 ;
+ Mẫu 2 : S = 1,02 ; $\bar{x}$ = 8,5 ; Me = 8,5
+ Mẫu 3 : S = 1,57 ; $\bar{x}$ = 8,5 ; Me = 9
+ Mẫu 4 : S = 1,42 ; $\bar{x}$ = 7,7 ; Me = 8
+ Mẫu 5 : S = 1,34 ; $\bar{x}$ = 8 ; Me = 8

d'où on conclut :

L'échantillon le plus représentatif : échantillon 3

L'échantillon le moins représentatif : échantillon 4

L'échantillon proposé avec : S = 1,58 ; $\bar{x}$ = 8,1 ; Me = 8,5

[86] C'est l'echantillon E6r

[87] C'est l'échantillon E6r

5	6	7	8	8	9	9	9	10	10

[88]

Le choix du couple (E3 ; E4) est conforme à une stratégie « représentant », E4 étant considéré comme le moins représentatif car il ne possède aucun des trois éléments du triplet (écart-type = 1,57 ; moyenne = 8 ; médiane = 8,5) de la population mère.

Deux enseignants donnent un sens différent à la notion de représentativité laquelle n’est pas basée sur la ressemblance (la proximité) avec la population mère. Pour eux le représentant doit être bien noté ou même avoir la note 10. Pour l’échantillon, ils suivent alors le critère : plus la moyenne d’un échantillon est grande mieux il représente la population.

> Enseignant 54 : Calculer les moyennes de cinq échantillons, puis choisir celui dont la moyenne est la plus grande pour l’échantillon le plus représentatif et faire réciproquement pour celui le moins représentatif. Je propose de choisir un échantillon de toutes les notes 10 pour le meilleur représentatif. E6*= que des 10
>
> Ainsi un échantillon représentatif idéal du lycée ne devrait comporter que les meilleurs élèves, dont la note serait 10.

Des stratégies « valeurs » pondérées

La stratégie « valeurs » est visible sur la ligne E4. Le couple (E4 ; E2) est le plus choisi (8 enseignants). Les commentaires qui accompagnent ce choix de (E4 ; E2) nous permettent de mieux le comprendre :

> Enseignant 8 :
>
> « J’attends que l’élève choisisse l’échantillon 4 comme le plus représentant car il comprend toutes les valeurs de 5 à 10 et les fréquences sont très proches des fréquences de données de 100 élèves dans le tableau. »
>
> Enseignant 67 :
>
> Parmi cinq échantillons, l’échantillon 4 est le plus représentatif car il est convenable aux données.
>
> Parmi cinq échantillons, l’échantillon 2 est le moins représentatif car il n’est pas convenable aux données, il n’y a pas de notes 5 et 6 comme dans les données.

Le critère de la présence de toutes les valeurs est accompagné d’une pondération par les fréquences. Cette pondération agit aussi pour rejeter E2 car il ne comporte pas toutes les valeurs de la population mère. On peut interpréter cette procédure de classement des échantillons à l’aide des deux théorèmes en acte suivant :

Théorème direct : pour qu’il y ait la proportionnalité des effectifs, il faut la présence de toutes les valeurs,

Théorème réciproque : s’il n’y a pas de présence de toutes les valeurs, alors il n’y a pas de la proportionnalité.

---

[88] C’est l’échantillon E6r

Le rejet de E3 relève de la même stratégie :

Enseignant 7 : l'échantillon 3 est le moins représentatif car il ne comprend pas la note 8 (dont le pourcentage dans les données est 16%) et il comprend 6 notes de 9 (qui occupent 32% dans les données).

L'importance de l'absence de la valeur 8 est pondérée par la fréquence de cette valeur dans la population ; il en est de même pour la présence de la valeur 9.

Des stratégies mixtes « proportion » et « représentant »

Enseignant 46 :

« Tableau de distribution des effectifs :

Notes	Effectifs
5	9
6	12
7	13
8	16
9	32
10	18
	N=100

Comme on doit retirer 10 individus parmi 100 individus pour avoir un bon représentatif, on fait comme suit :

Etape 1 : Diviser les effectifs par 10. On obtient les effectifs de l'échantillon de 10 individus

Etape 2 : Arrondir les effectifs obtenus par la division (car l'effectif ne doit pas être un nombre décimal)

On a :

Notes	Effectifs	Effectifs arrondis
5	0,9	1
6	1,2	1
7	1,3	1
8	1,6	2
9	3,2	3
10	1,8	2
	N = 10	N = 10

Alors le sixième échantillon à proposer est : 5, 6, 7, 8, 8, 9, 9, 9, 10, 10 E6p

Vérifier :

La moyenne des notes de 100 élèves : $\bar{x} = 8{,}04$

La moyenne des notes de 10 élèves qui sont les représentants $\overline{x} = 8{,}01$ [89]

On voit que $\overline{x} = 8{,}04 \approx 8{,}01$. Donc 1es 10 élèves représentent très bien

L'écart-type des notes de 100 élèves : $S = \sqrt{\overline{x^2} - (\overline{x})^2} = 1{,}568$

L'écart-type des notes de 10 élèves : $S = 1{,}578$

On voit que 1es 10 élèves de l'échantillon 6 représentent bien.

Conclusion : Ce sixième échantillon est très bon, il est l'échantillon représentatif.

La vérification que s'impose cet enseignant montre que son critère de représentativité est celui de l'égalité du couple (moyenne ; écart-type) entre l'échantillon et la population mère ; cette égalité est effectivement vérifiée lorsque les effectifs de l'échantillon et de la population sont proportionnels. Ce critère issu du rapport institutionnel qui met en avant le couple (moyenne ; écart-type) dans la comparaison des distributions ne permet cependant pas à lui seul d'être assuré de la proportionnalité des effectifs entre l'échantillon et la population. L'enseignant revient ensuite sur une stratégie « représentant » pour le classement des échantillons :

Comparaison avec la population mère :

A partir de 5 échantillons, on calcule la moyenne et l'écart-type :

Echantillon 1	8	1,095
Echantillon 2	8,5	1,025
Echantillon 3	8,5	1,565
Echantillon 4	7,7	1,418
Echantillon 5	8	1,342

- L'échantillon 2 est le moins représentatif car sa moyenne et son écart-type sont très différents de ceux de 100 élèves et ceux de l'échantillon 6.

- L'échantillon 3 et l'échantillon 4 ne sont pas les bons (seul un de deux indicateurs est proche de celui de 100 élèves.

- L'échantillon 4 est comme l'échantillon2 mais moins faible que l'échantillon 2.

- L'échantillon 5 est le meilleur car sa moyenne est très proche de celle de 100 élèves et son écart-type est très peu différent de celui de 100 élèves ».

Tableau synthétique sur E6

Résultats	E6*	E6r	E6p	E6f3	Total
Nb enseignants	2	1	19	2	24

Tableau 68. Distribution des choix des E6 selon les stratégies de 24 réponses d'enseignants

[89] Faux de calcul.

Même après avoir emprunté une stratégie « représentant » dans la comparaison des échantillons, les enseignants suivent la stratégie « proportion » pour produire l'échantillon E6 mais pas unanimement.

Quant aux deux enseignants qui produisent E6f3, ils sont partis de l'échantillon E4 considéré comme le plus représentatif à cause de la présence de toutes les valeurs et ont prolongé la stratégie « valeurs » en pondérant les présences respectives du 8 et du 9 ; il leur a suffi pour cela de mettre un 9 à la place d'un 8.

Commentaires d'enseignants :

Plusieurs enseignants suggèrent d'utiliser les échantillons pour comparer deux populations mères

> Enseignant 27 :
>
> + On ne discute pas la question de comparer deux échantillons pour trouver lequel est le meilleur mais du fait qu'en s'appuyant sur deux échantillons représentatifs de deux populations entières, on prévoit la comparaison de ces deux populations entières »
>
> Enseignant 4 : « L'idée de cet exercice est de faire une comparaison. Dans ce cas, il vaut mieux se donner deux échantillons (deux établissements) dont chacun de taille 30.
>
> Lycée A : 30 élèves
>
> Lycée B : 30 élèves
>
> Calculer la moyenne, la variance et l'écart-type de deux lycées.
>
> Dans quel lycée, les élèves étudient mieux ? »
>
> L'un d'entre eux va jusqu'à proposer un exercice à la place de l'exercice 4 :
>
> + Remplacer l'exercice par un exercice qui demande d'évaluer les copies d'un examen de deux groupes d'élèves d'une classe grâce à leurs indicateurs.
>
> + Evaluer la capacité d'étude des garçons et des filles d'une classe par <u>deux échantillons représentatifs</u>[90], un pour les garçons et un pour les filles ».

Ces enseignants se référent à des méthodes de statistique inférentielle et le second pense que la représentativité ne peut être assurée qu'à partir d'une taille 30 pour l'échantillon. Souhaitent-ils que les problèmes de comparaison de population à partir d'échantillons soient présents dans l'enseignement secondaire ?

Elève

Comme chez les enseignants, la stratégie graphique n'est pas utilisée. Par contre, certains élèves utilisent la stratégie « ordre » dans la comparaison des échantillons.

[90] Souligné par nous.

Le plus représentatif					
E1	E2	E3	E4	E5	Total
5	32	10	128	6	181

Le moins représentatif					
E1	E2	E3	E4	E5	Total
2	73	71	19	27	192

Proposition		
E6p	Autres	Total
102	48	150

Tableau 69. Les choix des échantillons représentatifs de 191 élèves

Le moins représentatif / Le plus représentatif	E1	E2	E3	E4	E5	Total
E1		1 (1 E6p)	2 (1 E6p)	2 (0 E6p)	0	5
E2	0		16 (9E6p)	8 (0 E6p)	8 (2 E6p + 1E6o1)	32
E3	0	2 (2 E6p)		4 (0 E6p)	4 (1 E6p)	10
E4	2 (1 E6p)	61 (43 E6p)	50 (30 E6p)		13 (6 E6p)	128
		5 (3 E6p)				
	2 (1 E6p)					
E5	0	2 (1 E6p)	1 (0 E6p)	3 (0 E6p)		6
Total	2	73	71	19	27	192/181

Tableau 70. Les couples (le plus représentatif ; le moins représentatif) choisis par les élèves

Des stratégies « représentant »

Ce sont 17 élèves sur 191 ce qui est en proportion nettement moindre que chez les enseignants. De plus le couple (moyenne ; écart-type) visible chez les enseignants est absent chez les élèves qui se contentent de la moyenne considérant peut-être que le calcul de l'écart-type est trop coûteux et pas assez significatif.

- avec la moyenne : le couple (E1/E5 ; E2/E3)

L'élève calcule la moyenne des données et celles des échantillons pour choisir le plus et le moins représentatif et aussi pour justifier que le sixième échantillon qu'il propose par proportions est le meilleur possible.

Elève V16 : «

1) La note moyenne des données est $\bar{x} = 8{,}04$

La note moyenne des échantillons 1 et 5 sont 8, celles des échantillons 2 et 3 sont 8,5; celle de l’échantillon 4 est 7,7.

Alors, les échantillons 1 et 5 sont les meilleurs représentatifs car leurs moyennes sont les plus proches de la moyenne des données.

2) Les échantillons 2 et 3 sont les moins représentatifs car leurs moyennes sont les plus éloignées de celle des données

3) Le sixième échantillon :

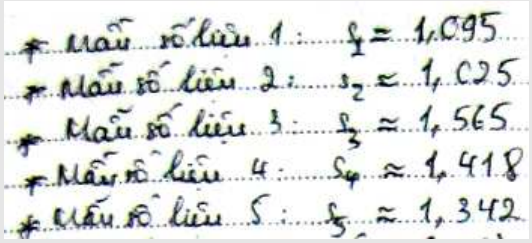
5 6 7 8 8 9 9 9 10 10 E6p

Parce que la moyenne est 8,1 ≈ 8 et les proportions des notes par rapport aux données sont relativement exacts, il est donc, pour moi, le plus représentatif »

- avec écart-type :

Quatre élèves font les comparaisons entre échantillons en utilisant l’écart-type sans se référer à la population mère. Ainsi choisissent-ils le couple (E2 ; E3) car E2 a le plus petit écart-type (1,1) et E3 le plus grand écart-type (1,57). Suivant la même stratégie ils proposent un échantillon E6 ayant un écart-type plus faible que E2.

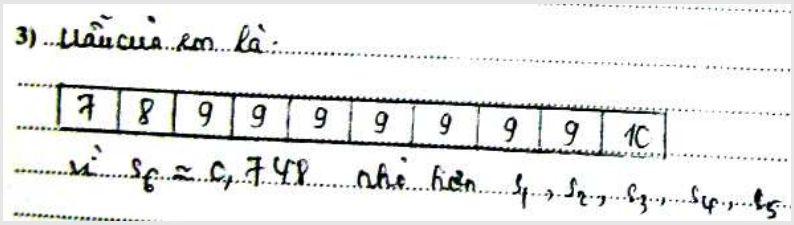
Elève N22 : «



L’échantillon 2 est le plus représentatif car S2 est le plus petit.

L’échantillon 3 est le moins représentatif car S3 est le plus grand.

Mon échantillon proposé est :



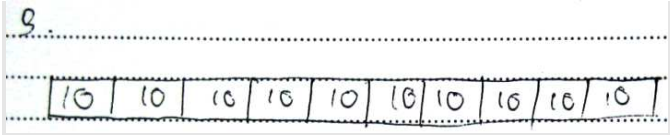
Car l’écart-type S6 = 0,748 est plus petit que S1, S2, S3, S4, S5.

Un élève suit les procédures attendues par l’enseignant 54 :

Elève V69 :

« Les échantillons 2 et 3 sont les plus représentatifs car leurs moyennes sont les plus grandes (8,5).

L’échantillon 4 est le moins représentatif car sa moyenne est la plus petite (7,7).

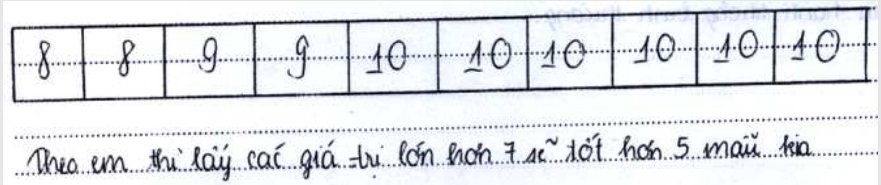


» [91]

Cette conception de la représentativité de l’échantillon qui n’est pas basée sur une « image » de la population mère mais sur la bonne qualité intrinsèque de l’échantillon se décline chez de nombreux élèves avec les choix des couples (E2 ; E4) comme par exemple l’élève V58 :

D’après moi, parmi les cinq échantillons, l’échantillon 2 est le plus représentant car ses valeurs sont très grandes, supérieurs à 7.

D’après moi, parmi les cinq échantillons, l’échantillon 4 est le moins représentant car il comprend des valeurs inférieures à 7.



Je pense qu’un échantillon ayant des valeurs supérieures de 7 serait mieux que les cinq échantillons donnés.

Des stratégies « ordre »

Cette stratégie n’est pas présente chez les enseignants. Chez les élèves, on la trouve avec des procédures diverses de pondération.

Elève L21 : «

1) Valeur	5	6	7	8	9	10
Effectif 9	12	15	16	32	18	

Diviser les effectifs par l’effectif le plus petit, on le tableau :

Valeur 5	6	7	8	9	10	
	1	1,33	1,67	1,78	3,36	2

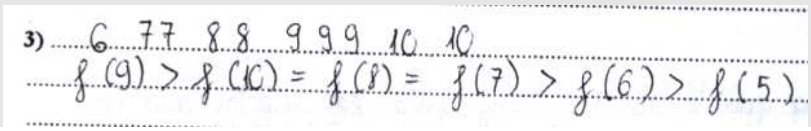
Dans l’échantillon représentatif, on doit avoir des effectifs comme suit :

$f(9) > f(10) \geq f(8) \geq f(7) > f(6) \geq f(5)$

Donc l’échantillon 2 est le meilleur : $f(9) = f(8) \geq f(10) = f(7)$

2) De même manière, l’échantillon 5 est le pire :

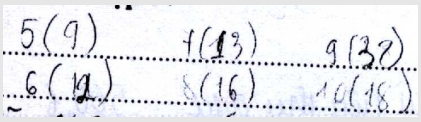
$f(8) > f(6) = f(10) > f(7) = f(9)$



[91] C’est l’échantillon E6*

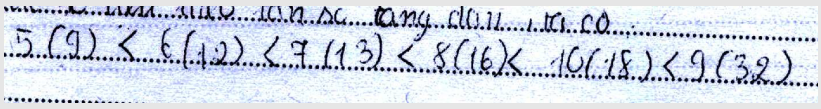
»

Elève L27 : «

1) 

L’échantillon 2 est le plus convenable car il y a un <u>équilibre</u>[92] des effectifs.

Ordonner les valeurs en ordre croissant des effectifs, on a :

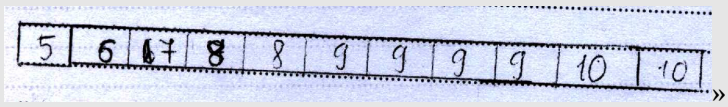


L’échantillon 5 est le moins représentatif car :

- effectif de 7 < effectif de 6

- effectif de 9 est le plus grand mais dans cet échantillon, son effectif n’est que 1.

3)



»

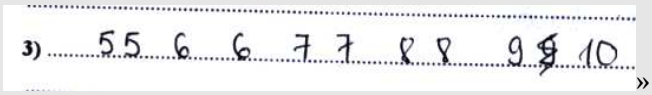
Des stratégies « valeurs »

C’est la stratégie la plus empruntée par les élèves. Elle se traduit principalement par les couples (E4 ; E2) et (E4 ; E3). On trouve diverses procédures de pondération par les effectifs.

Elève 25 : «

1) Echantillon 4 car il comprend toutes les valeurs.

2) Echantillon 2 car il lui manque des valeurs.

3) 

»

L’échantillon E6 choisi ne prend pas en compte les poids des effectifs.

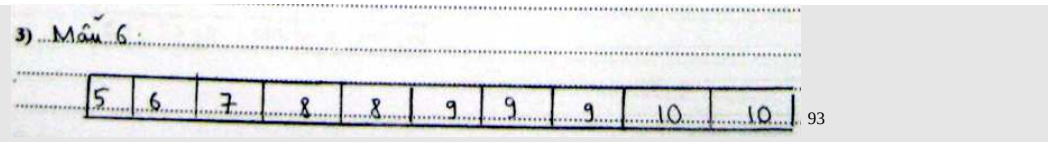
- 53 élèves sur 191 ont choisi le couple (E4 ; E3) par la présence de toutes les valeurs dans l’échantillon E4 et par l’absence des notes 7 et 8 qui occupent un grand pourcentage dans la population mère. La pondération des valeurs :

Elève L48 : «

1) Echantillon 4 est le plus représentant car il comprend toutes les valeurs.

2) Echantillon 2 est le moins représentant car il lui manque le plus des valeurs et d’ailleurs les notes ne sont pas bien distribués.

[92] Souligné par nous.


3) Mẫu 6

5	6	7	8	8	9	9	9	10	10

[93]

[93] C’est l’échantillon E6p.

Des stratégies « proportion »

- les rapports par colonne :

Elève M2 : «

1) A partir du tableau, on a :

$$\begin{cases} n_5 = 9 \\ n_6 = 13 \\ n_7 = 13 \\ n_8 = 16 \\ n_9 = 31 \end{cases}$$
$n_{10} = 18$

On a la formule : $f = \frac{n_i}{N}$

$n_5 = \frac{9}{100} = 0,09$

$n_6 = \frac{13}{100} = 0,13$

$n_7 = \frac{13}{100} = 0,13$

$n_8 = \frac{16}{100} = 0,16$

$n_9 = \frac{31}{100} = 0,31$

$n_{10} = \frac{18}{100} = 0,18$

$\Rightarrow n_8 = \frac{1}{2} n_9$

3) 6è échantillon :

Comme n6 = n7= 0,13 on peut donc choisir 6 ou 7 → on prend 1 note

$n_5/n_6 = \frac{0,13}{0,09} = 1,4$ fois ≈ 1 fois

(= 1,4 fois ≈ 1 fois)

n9 = 2 n8 et n8/n6 = 1 fois → n9 = 2 fois

n10 = 1 fois

Si on prend 6 individus, alors 5 6 8 9 9 10

Si on prend 10 individus : 5 6 8 8 9 9 9 9 10 10

Il est possible de remplacer 6 par 7 »

Les échantillons E6 proposés respectent exactement le rapport (égal à 2) des effectifs de 8 et de 9. Les autres rapports sont ramenés à l’entier le plus proche.

- les rapports par regroupements :

Un élève utilise le fait que la somme des effectifs de 9 et de 10 est moitié de l’effectif total :

Elève N19 : «

Phần trả lời bài tập 4 :

1) Số lượng học sinh đạt điểm 5 điểm là 9 học sinh ⇔ 9% về tỉ lệ
6 điểm là 12 học sinh ⇔ 12% về tỉ lệ
7 điểm là 13 học sinh ⇔ 13% về tỉ lệ
8 điểm là 16 học sinh ⇔ 16% về tỉ lệ
9 điểm là 32 học sinh ⇔ 32% về tỉ lệ
10 điểm là 18 học sinh ⇔ 18% về tỉ lệ
Với ra hệ số 10 ta có : 5 điểm = 0,9 ; 6 điểm = 1,2 ; 7 điểm = 1,3 ; 8 điểm = 1,6 ; 9 điểm = 3,2 ; 10 điểm = 1,8

[…]

Selon mon calcul, la somme des élèves ayant la note 9 et 10 est 5, en m’appuyant sur les proportions, je prends trois notes 9 et deux notes 10.

⇒ ? ? ? 7 ? 9 9 9 10 10

en m’appuyant sur les proportions des nombres d’élève ayant la note 5, 6, 7, 8, je prends la suite : 5, 6, 7, 8, 8

alors mon échantillon représentant est :

⇒ Mẫu số liệu của em là : 5 6 7 8 8 9 9 9 10 10 E6p

S’il est permis d’utiliser des nombres de R, je pourrais proposer le mieux tableau d’échantillon.

La dernière remarque nous confirme que la quasi-proportionnalité est acceptée par défaut.

Tableau synthétique sur E6

Résultats	E6r	E6*	E6p	E6f3	Autres	Total
Nombre d’élèves	1	1	102	4	42	150

Tableau 71. Distribution des choix des E6 selon les stratégies de 150 réponses d’élèves

Deux tiers des élèves (contre trois quarts pour les enseignants) proposent l’échantillon E6p. Contrairement aux enseignants qui prennent soin de justifier le recours à la proportionnalité, les élèves ne le font pas, sauf exception notable :

Elève N29 : «

D’après le tableau des données, on obtient le tableau des effectifs suivant :

Giá trị	5	6	7	8	9	10	
Tần số (%)	9	12	13	16	32	18	N = 100

En m'appuyant sur le tableau des effectifs précédent, je fabrique le tableau des effectifs de 10 élèves (arrondi en entier)

Giá trị	5	6	7	8	9	10	
Tần số	1	1	1	2	3	2	N=10

Et je vais proposer l'échantillon représentant :

Vậy em sẽ chọn ra mẫu số liệu sau

5	6	7	8	8	9	9	9	10	10

IV.3. Conclusion

L'exercice proposé, nous l'avons dit, renvoie à une praxéologie de comparaison de statistiques. Les réponses observées tant chez les enseignants que chez les élèves prouvent l'absence de technique institutionnelle pour le type de tâche proposé, à savoir comparer des distributions d'effectifs. La technique de proportionnalité des effectifs est certes disponible mais elle entre fortement en concurrence avec d'autres techniques dont les fondements mathématiques n'ont pas été éprouvés par l'institution. Ainsi en est-il de retenir dans un échantillon toutes les valeurs présentes dans la population pour le rendre représentatif de cette population. Ainsi en est-il également d'égaler deux à deux certains paramètres numériques pour s'assurer de l'égalité de deux distributions. On sait que cette dernière technique est valide pour certaines classes de distribution mais pas de façon générale pour toutes les distributions. La technique graphique pourtant reconnue dans l'institution dans d'autres circonstances n'est pas mobilisée dès lors qu'elle n'est pas explicitement demandée dans la consigne.

L'hypothèse (H1) est validée sans surprise, la quasi-proportionnalité est acceptée par défaut et sert d'expression de la proximité de deux distributions. La quantification de cette proximité est présente seulement à travers l'écart entre représentants.

L'hypothèse (H2) est validée par la forte présence des stratégies représentants et par l'emploi des procédures de vérification de la conservation des représentants dans le cas de l'utilisation de la proportionnalité.

Dans ce cadre, comment la notion de variabilité aléatoire de la distribution de l'échantillon par rapport à celle de la population peut-elle être construite dans l'institution vietnamienne ? Cette question est justifiée par notre projet d'ingénierie sur le problème d'obtenir un échantillon représentatif d'une population mère inaccessible. Nous rappelons en effet l'hypothèse de recherche formulée à la fin de l'analyse épistémologique (partie A) :

HR2 : le problème d'obtenir un échantillon représentatif d'une population mère inaccessible oblige à rencontrer et à construire à partir de statistiques des éléments de mathématisation du hasard au sein d'une praxéologie d'échantillonnage.

Nous prévoyons donc d'utiliser la notion d'échantillon représentatif dans le cas où l'échantillonnage est possible mais pour une population mère dont la distribution est

inconnue. Cela nous pose évidemment une nouvelle question **:** comment évolue le concept de représentativité quand on fait intervenir le hasard ? Et comment en faire un pont entre statistique et probabilités ?

Partie D

Ingénierie didactique

Chapitre D1
Introduction à l'ingénierie didactique

I. Pourquoi une ingénierie didactique ?

Revenons sur notre question de départ : est-il possible de faire vivre dans les conditions actuelles de l'enseignement vietnamien quelques uns des liens statistique-probabilités décrits par l'analyse épistémologique alors que statistique et probabilités sont enseignées séparément au Vietnam ? Les analyses institutionnelles que nous avons produites révèlent qu'on ne trouve pas de trace de savoirs caractéristiques des liens entre statistique et probabilité ni dans le cours ni dans les exercices des manuels secondaires vietnamiens, même si certains commentaires des livres pour les enseignants font référence à ces savoirs. Les réponses au questionnaire montrent cependant que certains graphiques et tableaux de distribution des fréquences pourraient potentiellement permettre de tisser quelques uns de ces liens notamment quand il s'agit de construire et d'utiliser un modèle probabiliste à partir de l'observation d'une distribution de fréquences.

C'est le cas dans le problème d'obtenir un échantillon représentatif d'une population mère dont la distribution n'est accessible autrement que par sondage qui est l'un des problèmes centraux de la statistique inférentielle et dont l'enjeu est une réduction de l'incertitude aléatoire pour prendre une décision. Cette réduction s'effectue à partir de distributions de fréquences et d'un modèle probabiliste. Les réponses des élèves dans l'exercice 4 montrent la disponibilité de la linéarité (proportionnalité) pour reconnaître un échantillon représentatif d'une population dont la distribution est connue. Comment organiser le passage au cas où la distribution de la population est inconnue ?

Pour apporter des réponses à cette question, nous avons décidé de concevoir, réaliser et observer une ingénierie fondée sur une hypothèse issue de nos analyses épistémologique et institutionnelle. Les raisons de cette décision ont été exposées dans l'introduction générale à notre étude lorsque nous avons présenté et justifié les choix théoriques et méthodologiques.

L'hypothèse de recherche que nous retenons est logiquement centrée sur les *problèmes d'incertitude aléatoire.* Nous appelons problème d'incertitude aléatoire un problème dont l'enjeu est de réduire l'incertitude sur une réalité observée lorsque cette incertitude est produite par la variabilité des résultats d'observation et que :

- la variabilité du résultat de chaque observation n'est ni accidentelle ni réductible mais est irréductible car inhérente au processus d'observation

- le résultat de l'observation est modélisable par un phénomène aléatoire[94] pour lequel l'observation est reproductible à l'infini

(HR3)[95]: l'étude coopérative d'une situation d'incertitude aléatoire permet le passage d'un point de vue déterministe à un point de vue probabiliste et par là à la construction de liens statistique-probabilités.

---

[94] Un phénomène aléatoire (ou stochastique) est un phénomène qui ne se prête qu'à une analyse statistique, par opposition à un phénomène déterministe. (Wikipédia : http://fr.wiktionary.org/wiki/stochastique)

.

En effet statistique et probabilités fonctionnent comme des objets de savoir complémentaires pour résoudre un problème d'incertitude aléatoire, les statistiques donnant une représentation de la réalité dont l'insuffisance n'est dépassable que dans la construction d'un modèle probabiliste.

Notre ingénierie est conçue à partir des ingénieries de Brousseau et de Briand, la seconde étant une reprise de la première. Ces deux ingénieries confrontent les élèves à *une machine de hasard* de structure inconnue (une bouteille opaque dans laquelle se trouvent des boules de deux couleurs différentes. Nous les présentons ci-après en montrant en quoi elles nous ont inspirés et en quoi nous avons innové.

II. Présentation des ingénieries de Brousseau et Briand

Précisons d'abord de quelles ingénieries nous parlons :

- la première dite « expérimentation Brousseau » fut menée à l'école élémentaire entre 1973 et 1974 ; elle est présentée par Brousseau lui-même comme « un curriculum de 31 leçons de Statistique et de Probabilités enseignées dans des classes de CM2 » ; on trouvera sa description complète, ses motivations et des analyses dans Brousseau [2005] et Brousseau [2009] ;

- la seconde dite « expérimentation Briand » fut menée au lycée en 2002 et 2003 ; elle est décrite par Briand lui-même comme une « transposition en classe de seconde du projet de Brousseau » ; on trouvera sa description complète, ses motivations et des analyses dans [Briand 2005]

Brousseau comme Briand insistent dans leurs motivations sur l'intérêt à conduire des apprentissages où Statistique et Probabilités fonctionnent de concert :

> Ce curriculum présente les probabilités et les statistiques ensemble, dans leur rapport épistémologique et fonctionnel, afin de montrer d'emblée leur complémentarité dans la résolution des problèmes d'incertitude.
>
> (Brousseau [2009])

> Dans une situation mettant en jeu des phénomènes aléatoires, l'élève est confronté à une incertitude : celle de la non reproductibilité d'une expérience pourtant construite à dessein […]
>
> La situation [didactique] doit donc permettre d'ébaucher un modèle prenant comme objet d'étude cette nouvelle incertitude : celle de la variabilité. […]. C'est donc l'incertitude qui est objet d'étude. Nous nous plaçons là, au delà de simples descriptions statistiques, dans une perspective de pensée probabiliste.
>
> (Briand [2005])

---

[95] Une telle hypothèse pourrait être qualifiée d'hypothèse fondatrice de l'ingénierie.

On comprend que la proximité de leurs intentions et des nôtres nous ait conduits à examiner les conditions d'une reprise dans le contexte institutionnel vietnamien des situations didactiques qui structurent les ingénieries. Cet examen cherchait à répondre à la question suivante : dans quelle mesure les apprentissages visés par des situations didactiques destinées à des élèves non initiés à la statistique et aux probabilités peuvent-ils conserver leur pertinence pour des élèves initiés à la statistique ou pour des élèves initiés à la statistique et aux probabilités ?

Pour comprendre comment nous avons répondu à cette question, commençons par entrer dans les ingénieries, d'abord dans celle de Brousseau puis dans celle de Briand.

II.1. La situation initiale de l'ingénierie de Brousseau, décrite par Brousseau lui-même[96] :

Le professeur :

> Votre camarade Pierre va mettre dans cette bouteille (opaque) et vide, 5 boules prises dans ce sac (opaque) qui en contient une trentaine. Venez vérifier qu'il n'y a dans le sac que des boules blanches et des boules noires.
>
> Pierre, mélange les boules dans le sac, et maintenant, sans regarder, isole 5 boules que tu tiens à travers le tissu ».
>
> Venez constater qu'il y en a bien 5. J'introduis alors la bouteille dans le sac.
>
> « Pierre, fais entrer les cinq boules dans la bouteille ! et maintenant bouche le goulot avec ce bouchon translucide.
>
> Vous êtes tous bien sûrs que dans cette bouteille il y a bien 5 boules et rien d'autre mais vous ne connaissez pas leur couleur. Nous allons essayer de savoir ce que contient cette bouteille sans jamais l'ouvrir.

Cette question heurte évidemment les modalités des raisonnements déterministes en usage dans les classes et les élèves ne comprennent pas quel calcul ils pourraient faire pour obtenir la solution de ce problème.

Ils commencent donc par essayer de regarder à travers le bouchon, puis l'un renverse la bouteille et là il peut voir la couleur de la boule qui se loge contre le bouchon.

> Il y a une boule blanche !

La question qui vient spontanément alors est « y en a-t-il aussi une noire ? » deux retournements de la bouteille… encore une blanche (la même ? une autre ? personne ne sait). L'idée que s'il y a une noire elle va finir par se montrer justifie des retournements jusqu'à ce que l'événement se produise (s'il se produit sinon ?). Ce ne sont évidement pas des tirages.

> Il y a une blanche et une noire…

C'est la fin de l'épisode déterministe et rien ne saurait justifier dans ce mode de pensée de continuer à retourner cette bouteille (est-ce que la lecture d'un deuxième exemplaire d'un même journal confirme les nouvelles du premier ?).

---

[96] Brousseau [2005]

Mais en général les enfants font des hypothèses qui vont démarrer le processus :

« Recommence cinq fois pour qu'on voie toutes les boules » dit un enfant qui a l'idée que les boules doivent se montrer successivement en bon ordre

> Voilà ! il y a trois boules blanches et deux noires !

L'histoire n'a pas de raison d'aller plus loin… sauf si cette idée que les boules se montrent sagement dans le même ordre est relevée, que ce soit par le professeur ou par un élève.

> Si ce que tu dis est vrai alors en recommençant on devrait voir encore les trois blanches et les deux noires…

Surtout ne pas s'arrêter aux hésitations des élèves qu'une prévision directe effarouche

> Cette fois on voit quatre fois une blanche et une noire.

Exit sans bruit l'hypothèse de la succession régulière … tout pourrait s'arrêter là mais…

« En tout cas il y plus de blanches que de noires… » dit un autre

> Alors on devrait continuer à voir plus de blanches que de noires si on continuait ?
>
> Non, si les blanches se sont montrées c'est au tour des noires (il y aura compensation).

De là une attention particulière pour les résultats des observations et pour leur enregistrement.

Il faudrait examiner toutes les évolutions possibles dans tous les cas (par exemple s'il n'y avait que des boules blanches) mais le démarrage est maintenant assuré et le professeur va pouvoir relancer des « expériences » (qui ne sont toujours pas des tirages puisqu'elles se font en situation supposée déterministe) à l'aide du moteur suivant :

Si ce qu'on voit donne des indications sur ce que contient la bouteille, alors, en reproduisant ce qu'on a fait on doit reproduire ce qu'on a vu. La question est évidemment : qu'y a-t-il de commun entre ces expériences ? Que peut-on voir ?

Ce qu'on a vu, va devenir l'image du passé de la machine de hasard, c'est-à-dire une *statistique*, ce qu'on prévoit va devenir des événements plus ou moins *probables* associés à la machine et les deux sont reliées par des hypothèses sur la structure de la machine que l'on peut facilement imaginer.

Il faut remarquer que cette situation est fondamentale à la fois pour la statistique et pour les probabilités. Autrement dit, seule l'invention de ces deux concepts (et les notions afférentes) et la compréhension de leurs rapports peuvent résoudre ce problème de façon satisfaisante. […] Nous le vérifierons en suivant la genèse que ce problème peut provoquer. Cela suppose évidemment que le professeur n'apportera et n'acceptera aucun apport extérieur sur les questions cruciales que nous allons rencontrer.

II.2. La genèse du problème d'incertitude dans l'ingénierie de Brousseau, résumée par nous[97]

- Les élèves font des groupes de cinq observations puis dénombrent les compositions obtenues :

> Nous serons sûrs, disent les enfants, lorsque l'effectif d'une des compositions possibles dépassera les autres de 3…
>
> Ainsi les suites de tirages s'allongent, nourries de doutes et de convictions des élèves et de leur désir de vérifier et de prouver ce qu'ils pensent. Les séances sont courtes : le plus souvent dix minutes ou un quart d'heure : le temps de discuter un résultat et de décider une action.

- Les élèves décident de fabriquer des bouteilles transparentes et l'enseignant leur demande de prévoir les résultats :

> Il y a des écarts, pensent-ils mais si on recommence on devrait voir les écarts se réduire !
>
> Espoir fallacieux pour l'instant puisqu'on considère des effectifs, mais l'important est que cette étape fournit des raisons de faire de longues suites d'observations et de noter les résultats dans le cadre de questions pertinentes à leur sujet.

- L'enseignant propose des tirages déjà faits par un ordinateur :

> Dès lors, tout est en place pour calculer le rapport vers lequel on s'attend à voir la fréquence observée des blancs ou des noirs se rapprocher : 3/5 pour (3b, 2n), 2/5 pour (2b, 3n) etc. Les élèves fixent eux-mêmes le nombre de tirages dont ils demandent les résultats à l'ordinateur (entré entre temps dans la classe). Les séries de fréquences observées sont reportées sur des *graphiques* et les élèves « observent » que les courbes après des oscillations assez importantes viennent s'approcher des valeurs prévues.

- L'enseignant propose un jeu de pari avec une mise donc un risque de perdre la mise :

> Le *jeu* consiste à deviner quel est le contenu d'une bouteille à la lecture d'une série de tirages qu'elle a produits. Pour cela les élèves peuvent « acheter » des séries de la longueur qu'ils veulent s'ils ont des jetons pour les payer. Cette situation simule exactement le problème du test d'hypothèse.

- L'enseignant change de machine de hasard et leur demande des prédictions :

> C'est seulement alors que l'on a pu introduire le vocabulaire des événements et des opérations sur les événements, puis le calcul sur des valeurs de « probabilité ». Tous les calculs élémentaires dits « de probabilités » ont alors pu être découverts par les élèves ou enseignés. Il faut remarquer qu'à aucun moment nous n'avons parlé formellement de probabilité, ni introduit des spéculations sur la couleur qui allait sortir au cours d'une expérience isolée. De ce fait, les grandeurs manipulées l'ont été dans le cadre d'une pensée déterministe en situation d'incertitude, tout à fait adaptée aux enfants de cet âge. L'expérience s'est étendue sur 32 séances (hors des cours de mathématiques) et dont seulement une dizaine ont duré plus d'une demi heure.

---

[97] Les citations sont tirées de Brousseau 2005.

II.3. Examen d'une possibilité de transposition hors du primaire

Brousseau envisage une telle possibilité et avance cette idée :

> Aujourd'hui, surtout avec des élèves plus âgés, la tension serait insoutenable et le professeur serait obligé de rendre des comptes, c'est-à-dire d'exposer les savoirs et les techniques dont il dispose… ce qui les mettrait sous la dépendance du jugement des élèves basé sur des connaissances superficielles et des préjugés.
>
> (Brousseau [2009])

Plus précisément, à propos des savoirs visés par l'ingénierie, il déclare :

> Les situations n'ont pour effet que de stimuler des questions, jamais elles ne fournissent de réponse décisive à ces questions. Ce qui a été découvert et enseigné est une démarche, pas une preuve.
>
> (Brousseau [2005])

Doit-on en conclure que la reprise de l'ingénierie pour des élèves instruits de statistique et de probabilités, doive être écartée ? Nous pensons que non, et ce pour deux raisons :

- la première est que les situations de l'ingénierie feront nécessairement entrer les élèves dans une démarche de résolution d'un problème d'incertitude aléatoire dès lors que l'enseignant se refusera à ouvrir la bouteille, ce que des élèves plus âgés accepteront volontiers, à notre avis ; cette raison est avancée par Briand qui la développe dans la présentation de son ingénierie en parlant d'initiation à la pensée stochastique ;

- la seconde est que les situations de l'ingénierie placent les élèves devant la confrontation de probabilités qu'ils ont appris à calculer et de distributions statistiques qu'ils ont appris à décrire ; les savoirs en jeu peuvent alors se porter sur la modélisation probabiliste de la fluctuation des distributions empiriques.

La transposition réalisée par Briand pour une classe de seconde française nous permettra d'étoffer notre argumentation. Auparavant, notons qu'elle a lieu dans le contexte des nouveaux programmes de statistique en lycée (cf. l'analyse que nous faisons par ailleurs de ces programmes). Briand précise en outre qu'il est important de prendre en compte la culture des élèves du lycée dans les domaines de la statistique et des probabilités, affirmant que « les significations acquises [dans d'autres lieux que la classe] constituent un foyer d'obstacles en tous genres ». Il rejoint en cela l'avis de Brousseau à propos « du jugement des élèves basé sur des connaissances superficielles et des préjugés »[98]. Poursuivant l'idée d'obstacles à faire franchir, il argumente sur la nécessité de construire des situations qui organisent une rupture entre le modèle déterministe et le modèle stochastique :

> Il s'agit de construire un milieu d'apprentissage dans lequel les élèves devront prendre des décisions qui seront validées ou invalidées par des expérimentations, mais par rapport à des situations adidactiques relatives à des familles de savoirs mathématiques à caractère déterministe, les rétroactions ne seront pas de même nature, puisqu'il s'agit maintenant d'un modèle stochastique.

---

[98] Cf. infra

(Briand [2005])

La machine de hasard au cœur de l'ingénierie de Brousseau lui paraît constituer un milieu capable de porter la rupture visée entre déterminisme et aléatoire :

> Il s'agit de construire un milieu « adéquat » et des situations en sachant que celles-ci ne recouvrent que rarement les savoirs tels que ceux répertoriés dans l'institution et que les élèves construisent des savoirs à partir de leurs propres connaissances, de leurs conceptions, de ce qu'ils pensent être des vérités, en commettant des erreurs, en ayant à vivre avec des obstacles.
>
> Dans le cas présent, le dispositif utilisé : la machine de hasard, permet aux élèves de constater des faits. Le travail sur les expérimentations à conduire apparaît alors comme résultant des premières affirmations, des premières conjectures, déclarées par les élèves dans les moments de formulation. Vont se construire un ensemble de conjectures dont le professeur aura à se saisir pour faire vivre la curiosité et dépasser ainsi la demande d'ouverture de la bouteille.

(Briand [2005])

II.4. L'ingénierie de Briand, décrite par Briand lui-même [99]

« Nous avons travaillé de la façon suivante : lors de réunions préparatoires, le professeur a pris connaissance de l'activité initiale de l'école primaire. Nous avons ensuite construit un protocole *a priori.* Notre collègue a proposé de modifier la consigne de travail. C'est sur la base de sa proposition que nous avons commencé le travail.

[…]

Les deux premières séances ont consisté à poser un problème dans le milieu matériel bouteilles-billes. La troisième séance a mis en jeu des outils de simulation informatique afin que les élèves puissent disposer d'informations statistiques leur permettant de mettre à l'épreuve des conjectures qui sont apparues, surtout lors de la deuxième séance.

Première séance

Les élèves sont répartis par groupe de quatre et disposent de bouteilles ayant 5 boules : des rouges et des vertes. Les bouteilles n'ont pas toutes la même composition. Le professeur ne connaît pas les contenus des bouteilles, ce qui permet à celui-ci de maintenir plus aisément l'aspect adidactique de la situation. La consigne prévue est : *« On va considérer que vous êtes en quelque sorte des prévisionnistes. Un client s'adresse à vous et vous demande un sondage. Qu'est-ce que c'est qu'un sondage dans ce dispositif ? Par exemple, un sondage de taille 10, cela veut dire que l'on interroge 10 fois la bouteille. Il a besoin que vous lui disiez le résultat de ce sondage de 10 tirages et en plus, il a besoin de savoir si vous êtes moyennement sûrs, assez sûrs ou au contraire tout à fait sûrs ».*

Lors de cette première séance, les élèves étaient répartis par groupes de 4. Nous avons pu observer une grande similarité avec les comportements des élèves de fin d'école primaire. Les premiers constats se font : « il y a des rouges et des vertes », puis l'obéissance à la

[99] Briand [2005]

consigne fait que les élèves font des séries de 10 tirages. La fluctuation des résultats obtenus constitue le moteur qui va permettre d'augmenter le nombre de tirages. Le va-et-vient entre le passé (les faits observés) et ce que l'on veut vérifier (l'expérience construite, c'est-à-dire l'expérimentation) montre à la fois l'incertitude et le début de régularités qui incitent à recommencer et à augmenter le nombre de tirages. Les élèves cherchent en fait à connaître la composition de leur bouteille.

Deuxième séance

Le dispositif de classe est le même. Le professeur fait le point de la séance précédente et propose alors : *« Je garde en tête ma question initiale, mais il me semble que l'on pourrait aujourd'hui essayer de voir si l'on pouvait avancer sur le contenu des bouteilles »*. Cette nouvelle consigne clarifie le travail. A la différence des élèves de fin d'école primaire, les élèves de seconde sont vite conscients de la possibilité de mettre en place des raisonnements. L'ouverture de la bouteille (demandée par les élèves d'école primaire) ne fait plus l'objet de demandes pressantes. Les premières conjectures apparaissent.

Troisième séance

A la fin de la deuxième séance les élèves sont convaincus de certaines régularités et font l'hypothèse que *« plus on fera de tirages, plus on pourra affirmer des compositions de bouteilles avec certitude »*. Ils ont fait des déclarations à propos des intervalles de confiance. Il est plus facile de décider de la composition de 4R, 1V puisque l'intervalle [0,7 ; 1] est plus grand que l'intervalle [0,5 ; 0,7] qui permettrait de décider 3R, 2V.

Afin de mieux expliciter ces nouveaux savoirs d'expériences, le professeur propose un travail à l'aide d'un programme effectué sur tableur : l'élève entre une composition de bouteille et demande 15, 25, 100, 500 ou 1000 tirages. Le programme propose le résultat de ces tirages avec effectif et fréquence (figure 5).

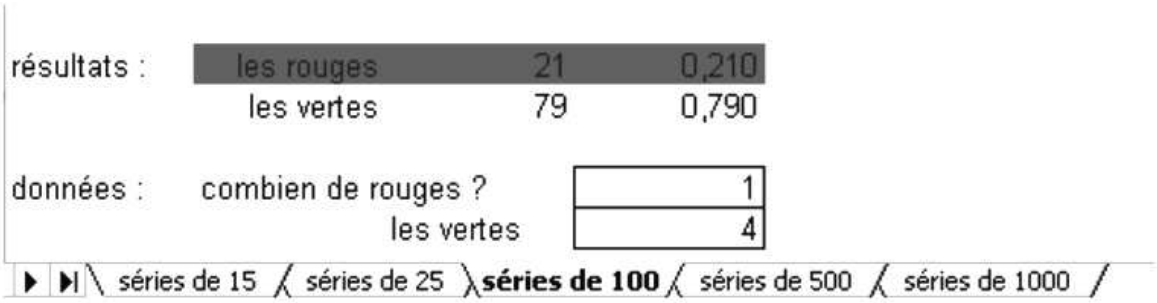


Figure 5. – Écran de travail du tableur

Afin de mieux tester l'hypothèse : « plus on fera de tirages, plus on pourra affirmer des compositions de bouteilles avec certitude », le professeur propose de mettre sur un graphique en abscisse le nombre de tirages, en ordonnée le résultat d'une expérience (statistique) et ce qui constitue de leur point de vue la meilleure hypothèse à avancer. C'est ainsi que les élèves proposent de matérialiser par une droite parallèle à l'« axe des x » ce qui, pour l'observateur, constitue un début de travail probabiliste. Le sentiment de la probabilité porte sur la sûreté de la convergence.

À ce moment de l'expérimentation, les élèves ont donc compris l'existence de régularités, malgré les alea de l'expérience statistique. En effectuant un « bon nombre » d'expériences,

les élèves vont disposer d'un outil leur permettant de commencer à formuler des hypothèses prenant en compte le risque dans une prise de décision[100] en fonction du nombre de tirages. »

(Briand [2005], p. 263-265)

II.5. Examen de la transposition réalisée par Briand

Par rapport à l'ingénierie de Brousseau, celle de Briand introduit deux nouveautés dans la première séance : l'utilisation de la notion de sondage en plus de celle de tirage appelée « interrogation de la bouteille » et l'évaluation par l'élève de sa certitude en trois degrés « moyennement sûrs, assez sûrs ou au contraire tout à fait sûrs ».

La première, dont Briand nous dit qu'elle est dictée par la volonté professorale de se conformer aux recommandations décrites dans le document d'accompagnement du nouveau programme, présente l'intérêt de mettre en scène la fluctuation d'échantillonnage qui est la manifestation de la variabilité aléatoire sur laquelle on bâtit la notion de risque dans la séance suivante. Il nous semble pertinent de la retenir mais sans la mettre en scène dans la première situation laquelle doit permettre aux élèves d'entrer dans le problème d'incertitude.

La seconde nous semble elle aussi pertinente car elle permet à l'enseignant de tisser implicitement un lien entre un résultat statistique et une inférence.

En dehors de ces deux nouveautés, l'ingénierie de Briand reprend en les condensant les situations successives de celle de Brousseau.

Dans les commentaires que nous avons soulignés ci-dessus, Briand désigne un savoir à viser dans l'ingénierie, celui de risque, ce qu'il réaffirme un peu plus loin dans une référence à la théorie des situations didactiques :

> Le professeur entretient une relation privilégiée avec le savoir (par exemple, ici, savoir que le risque est objet d'étude).
>
> (Briand [2005])

Nous sommes en accord avec le choix du savoir visé et faisons nôtres les raisons qu'il en donne :

> Rappelons quelques entrées possibles pour l'enseignement des statistiques et des probabilités :
>
> - Une voie est celle de la combinatoire et du calcul des probabilités.
>
> On suppose que l'élève comprend ce qu'est d'avoir 3 chances sur 4 de tirer une carte donnée. À partir de ce type de constat, un travail de recensement des cas favorables et des cas possibles permet une construction élémentaire des probabilités.
>
> - Une autre voie consiste à observer des phénomènes aléatoires (passage de voitures par exemple). Le recueil des données et leur organisation aboutit à l'élaboration de tableaux, de diagrammes, etc. À partir de cela, l'analyse des données permet de chercher ce qui serait intéressant.

[100] C'est nous qui soulignons.

> La recherche de 1975 [celle de Brousseau] proposait une troisième voie, celle qui permet d'aborder les statistiques et les probabilités dans leurs relations réciproques. Il ne s'agit plus seulement de travailler à partir de descriptions, mais (De Finetti 1974) *« de définir la probabilité d'un événement donné E (bien déterminé), pour un individu donné, comme le degré de confiance qu'il a dans son arrivée »*. Comme définition opérationnelle, la probabilité est alors *« le prix p (pour cet individu) conditionné au fait que E soit vrai »*. Et l'auteur d'ajouter : *« c'est, avec les précisions que l'on tire aisément des conditions de "cohérence", la notion sur laquelle s'appuient, d'une part, les paris (sur des faits de sports, d'élections politiques, etc.) et en général toutes les décisions que nous prenons sans cesse, plus ou moins consciemment, au cours de la vie »*.
>
> Pour donner un sens à la probabilité associée à un événement, il paraît alors nécessaire que l'élève la matérialise par un gain attaché à une décision. Pour cela, nous pensons que les données doivent être des informations recherchées intentionnellement par l'élève dans des expériences faciles à mettre en œuvre dès qu'il le souhaite.
>
> (Briand [2005])

Brousseau a lui aussi la notion de risque en vue lorsqu'il propose aux élèves le pari sur la composition de la bouteille :

> Devoir payer les informations les amène à s'interroger sur les rapports entre la longueur de la série et le risque de se tromper. L'expérience est « décisive » pour eux.
>
> (Brousseau [2005])

Comment le jeu est-il organisé ?

> Lors de notre expérience, ce jeu a dû être écourté car il nous manquait du temps pour arriver au terme du processus. D'ailleurs nous n'avions pas pu calculer les paramètres de la situation qui pouvaient assurer un déroulement intéressant et vif : combien de jetons donner au départ ? Quels nombres de tirages proposer ? A quels tarifs ? etc. Nous avons pu observer qu'il fallait un temps pour que les élèves se réfèrent aux fréquences observées pour les grands nombres de tirages et non pas aux fluctuations du départ.
>
> (Brousseau [2009])

Dans son article de 2003, Brousseau parle seulement de l'activité graphique que l'on retrouve aussi dans la séance 3 de Briand :

> Autour des valeurs de convergence 0,2 (1 blanc) ; 0,4 (2b) ; 0,6 (3b) ; 0,8 (4b) ils repèrent des bandes à l'intérieur desquelles les séries finissent par entrer et à rester. En demandant 15 séries de 15 tirages 8 séries seulement sont dans la bande et 7 à l'extérieur, avec 15 séries de 20, on réussit 9 fois et on échoue 6 fois, avec des séries de 100, 15 réussites sur 15 mais avec 160 seulement 14 réussites sur 15. Les séries de 1000 tirages se resserrent beaucoup autour de la valeur théorique… Au vocabulaire près, les élèves utilisent la notion de seuil de signification et d'intervalle de décision.
>
> (Brousseau [2005])

La mise y est donc symbolique et l'enjeu de la situation est, sous la conduite de l'enseignant, la formulation du risque de se tromper sur la base d'un certain nombre de tirages ainsi que son évaluation empirique sous la forme d'une proportion de tirages produisant un résultat extérieur à un intervalle fixé par la composition qui est elle-même connue.

Le risque étant une probabilité[101], aucune validation empirique n'est possible, pas plus en ouvrant la bouteille qu'en la maintenant fermée ; la seule validation est de nature théorique sous la forme d'une probabilité a priori calculée à partir d'un modèle mathématique pour la distribution d'échantillonnage consistant et « crédible ».

Les conditions proposées dans l'ingénierie de Brousseau et de Briand sont-elles suffisantes pour initier une modélisation probabiliste de la distribution d'échantillonnage ? L'absence de pari effectif ne rend pas nécessaire pour l'élève la mesure du risque qu'il prend. La formulation et sa mesure du risque dans le cas d'une population dont le nombre de compositions est petit se fait sur la base d'un intervalle de décision construit théoriquement. Pour rendre plus prégnante l'idée de risque, nous pensons qu'il est nécessaire, dans la situation de formulation, que les élèves soient effectivement confrontés à un nombre de compositions possibles très grand qui oblige à construire un intervalle de confiance plutôt que de se référer à un intervalle de décision.

Auparavant nous complétons notre analyse institutionnelle relativement aux deux notions d'incertitude aléatoire et d'expérience aléatoire. Nous pensons en effet que ces deux notions sont cruciales dans les enjeux didactiques de l'ingénierie.

III. Rapports institutionnels aux notions d'expérience et d'incertitude aléatoires

Contrairement aux élèves des expérimentations de Brousseau et de Briand, les élèves du lycée vietnamien où sera expérimentée l'ingénierie auront reçu un enseignement des probabilités. Cet enseignement n'existe que depuis l'année scolaire 2006-2007. La noosphère vietnamienne a mis en place cet enseignement après avoir procédé à l'expérimentation d'un nouveau curriculum mathématique secondaire comportant pour la première fois une introduction aux probabilités, en classe 11. Cet enseignement expérimental a fait l'objet d'un mémoire de recherche de notre part[102] dont nous extrayons cette conclusion :

> L'organisation mathématique de la notion de probabilité repose essentiellement sur l'approche classique de Laplace. L'approche fréquentiste n'est pratiquement présente qu'au travers une définition fréquentiste de la probabilité : elle est rudimentaire dans M1 (dirigé par Doan Quynh) et absente de la batterie d'exemples et d'exercices dans M2 (dirigé par Tran Van Hao)[103]. La définition axiomatique de la probabilité n'est pas retenue mais l'introduction à la notion de probabilité s'appuie implicitement sur la mise en relation des opérations sur les ensembles finis et des opérations sur les cardinaux. Nous pouvons dire que l'enseignement de la probabilité au lycée est centré essentiellement sur les types de tâche "Calculer la probabilité" et ne se préoccupe guère de l'établissement du « sens réel » de la notion de probabilité (« sens » que l'approche fréquentiste pourrait apporter plus facilement), comme de l'élaboration des techniques caractéristiques de ce sens. Ainsi, la probabilité calculée a un caractère théorique […]

[101] En fait une probabilité conditionnée par la composition de la bouteille.

[102] Vu Nhu Thu Huong, 2005, La notion de probabilité dans l'enseignement-apprentissage des mathématiques au lycée.

[103] Il existait deux manuels, celui de la série 1 considérée comme série « avancée » et celui de la série 2 considérée comme série « élémentaire ».

> bâti à partir des règles de la Combinatoire sans réalisation effective d'expériences aléatoires. […]
>
> La centration du manuel sur la définition classique risque de faire amalgamer la notion de probabilité à un nombre exact représentant la possibilité d'apparition d'un événement dans le cas très restrictif d'expériences aléatoires dont les issues sont équiprobables.
>
> (Vu Nhu, [2005])

A la suite des deux années scolaires de cette expérimentation noosphérienne - les années 2004-2005 et 2005-2006 - les probabilités seront effectivement présentes en classe 11 selon un programme largement inspiré du programme expérimental. Les seules modifications porteront sur le retrait des probabilités conditionnelles et des variables aléatoires. Ces modifications ne sont pas susceptibles de produire des effets qui infirmeraient les analyses et les conclusions de notre recherche précédente.

En ajoutant les programmes et les manuels actuels du lycée vietnamien, nous nous appuyons sur ces analyses pour caractériser plus précisément les rapports institutionnels actuels aux notions d'expérience et d'incertitude aléatoire.

III.1. Notion d'expérience aléatoire :

Les manuels expérimentaux de classe 11

Le manuel de la série 1 présente un modèle connu de l'expérience aléatoire : le lancement d'un dé.

> Lorsque nous lançons un dé, le nombre des points de la face apparue est considéré comme le résultat du lancement. Nous constatons qu'il est très difficile de prédire[104] le résultat de chaque lancement. Ce résultat peut être n'importe quel nombre dans l'ensemble {1, 2, 3, 4, 5, 6}. Nous appelons ce lancement de dé une expérience aléatoire.
>
> (Manuel expérimental de classe 11, 2004, p. 79)

Puis il donne une définition de l'expérience aléatoire avec trois conditions :

> **Une expérience aléatoire**[105] (ou expérience en abréviation) est une expérience ou une action
>
> - qui peut être répétée plusieurs fois dans les mêmes conditions ;
>
> - dont les résultats ne sont pas prévisibles ;
>
> - où l'on peut déterminer l'ensemble de tous les résultats attendus de l'expérience.

[104] C'est nous qui soulignons.

[105] Dans cette citation comme dans les suivantes nous respectons la mise en forme du manuel.

> Cette expérience est souvent désignée par le signe T.
>
> (Manuel expérimental de classe 11, 2004, p. 80)

Notons les contradictions internes entre la partie introductive et sa définition puisque la prédiction est d'abord qualifiée de « très difficile » avant d'être annoncée comme impossible.

Les manuels actuels de classe 11

Le manuel avancé reprend le contenu du manuel expérimental de classe 11 (2004-2005) mais dans la définition d'une expérience aléatoire la première condition qui concerne la répétition est enlevée :

> **Une expérience aléatoire** (ou expérience en abréviation) est une expérience ou une action :
>
> - dont les résultats ne sont pas prévisibles;
>
> - où l'on peut déterminer l'ensemble de tous les résultats possibles de cette expérience.
>
> L'expérience est souvent désignée par le signe T.
>
> (Manuel de classe 11, l'ensemble avancé, p. 70)

Cette omission reflète-t-elle un changement épistémologique sur la notion d'expérience aléatoire ou préfigure-telle seulement l'effacement du point de vue fréquentiste lequel nécessite la reproductibilité d'une expérience ?

Le Manuel de classe 11, l'ensemble élémentaire (2006) quant à lui s'attache à distinguer une expérience aléatoire d'une expérience quelconque pour mettre en avant l'idée qu'une expérience est aléatoire quand on ne peut pas en prévoir le résultat :

> Une des notions importantes de la théorie de probabilité est l'expérience : une expérience[106], un fait de mesure ou une observation d'un phénomène est considérée comme une expérience.
>
> Par exemple, le lancement d'une monnaie métal (abrévié monnaie), tirage d'une carte de l'ensemble des cartes (52 cartes) ou le fait de tirer d'une balle à la cible, … sont des exemples de l'expérience.
>
> Quand on lance une monnaie, on ne peut pas prévoir si face ou pile apparait. C'est un exemple de l'expérience aléatoire.
>
> **L'expérience aléatoire** est une expérience dont on ne peut pas prévoir le résultat même si on en connaît l'ensemble des résultats possibles.
>
> Pour abréger, on l'appelle désormais une expérience. Dans les mathématiques secondaires, on n'étudie que les expériences ayant un nombre fini de résultats.
>
> (Manuel de classe 11, l'ensemble élémentaire, p. 59)

[106] Sens physique, chimique ou biologique.

Il est clair que l'institution vietnamienne entend se limiter à des expériences aléatoires ayant un nombre fini de résultats que l'on peut connaître a priori sans avoir à répéter les expériences ni même à les réaliser et donc pour lesquelles on peut « déterminer l'ensemble de tous les résultats possibles ». Cela rejoint l'analyse précédente de la domination du point de vue laplacien (combinatoire) sur le point de vue fréquentiste.

Ce point de vue fréquentiste existe cependant dans l'institution et ce de deux manières :

- à partir de la répétition d'une même expérience aléatoire supposée reproductible :

> Considérons l'expérience T et l'événement A en relation à cette expérience. Nous répétons N fois l'expérience T et notons les fois où se produit l'événement A.
>
> Le nombre d'apparitions de l'événement A est appelé l'effectif de A dans l'expérience T effectuée N fois.
>
> Le rapport entre l'effectif de A et le nombre n est appelé la fréquence de A dans l'expérience T effectuée N fois. On a pu démontrer que plus le nombre N d'expériences est grand plus la fréquence de A se rapproche d'un nombre déterminé, appelé la probabilité de A dans le sens fréquentiste (c'est aussi le nombre P(A) dans la définition classique de la probabilité.

(Manuel de classe 11, l'ensemble avancé, p. 70)

Cette première manière donne lieu à la « définition fréquentiste de la probabilité » comme limite des fréquences empiriques et que certains manuels français appellent fréquence théorique.

> Exemple 7
>
> Si on lance une pièce d'argent équilibrée, la probabilité d'apparition de la face est de 0,5. Buffon, mathématicien français du XVIIIe siècle, a expérimenté plusieurs fois le lancement de pièce et il a obtenu les issues suivantes:

Nombre des lancements	Effectif d'apparition des faces	Fréquence d'apparition des faces
4040	2048	0,5070
12000	6019	0,5016
24000	12012	0,5005

(Manuel de classe 11, l'ensemble avancé, p. 74)

A travers la notion de probabilité expérimentale définie comme la fréquence de réalisation d'un événement, obtenue par une statistique dont l'effectif total est jugé suffisamment grand. L'exemple suivant illustre cette définition en montrant son usage possible à la place d'une probabilité a priori :

> Une compagnie d'assurances-vie a établi les statistiques suivantes : parmi 100 000 hommes âgés de 50 ans il y a 568 hommes qui décèdent avant qu'ils passent à l'âge de 51 ans, et parmi 100 000 femmes de 50 ans, 284 meurent quand elles ont 51 ans. La probabilité expérimentale pour qu'un homme de 50 ans décède avant de passer à l'âge de 51 ans est $\frac{568}{100\ 000} = 0{,}00568$ et la probabilité expérimentale pour qu'une femme de 50 ans décède avant de passer à l'âge de 51 ans est $\frac{284}{100\ 000} = 0{,}00284$

(Manuel de classe 11, l'ensemble avancé, p. 75)

Une probabilité expérimentale est donc une fréquence empirique et dans cet exemple il n'y a pas explicitement d'expérience répétée. Une seule activité de répétition d'une expérience aléatoire est proposée :

> Lancer un dé 50 fois. Noter les résultats de ces lancements et calculer les fréquences d'apparition des chaque face de 1, 2, 3, 4, 5, 6 points :

Nombre des points apparus	Effectifs	Fréquences
1		
2		
3		
4		
5		
6		

(Manuel de classe 11, l'ensemble avancé, p. 75)

avec le conseil d'organiser cette activité dans le livre des enseignants :

> Conseil de réalisation. L'enseignant prépare 5 dés identiques. Appeler 5 élèves et demander chacun de lancer le dé 10 fois et de noter le nombre de fois que la face de k points apparait pour son 10-lancements (k = 1, 2, 3, 4, 5, 6). En additionnant les résultats de ces 5 élèves, l'enseignant obtient l'effectif de la face de k points pour 50 lancements d'un dé.

(Livre d'enseignants de classe 11, l'ensemble avancé, p. 95)

L'objectif semble être de constater que les fréquences empiriques sont différentes (et peut-être « proches ») des probabilités a priori calculables du point de vue laplacien en supposant que le dé n'est pas pipé. La fluctuation des fréquences autour des probabilités a priori n'est cependant pas travaillée par le manuel. Et comment l'enseignant peut-il articuler cette activité de répétition du lancer de dés avec l'exemple précédent de l'assurance sans évoquer la notion de variable aléatoire ? La notion de variable aléatoire est certes présente mais seulement dans le manuel avancé et elle ne donne pas lieu à une reprise des activités de répétition des expériences aléatoires. Le manuel avancé se contente d'une lecture complémentaire :

> LIEN ENTRE LA VARIABLE ALÉATOIRE DISCRETE ET LA STATISTIQUE
>
> Soit le caractère X avec l'ensemble des valeurs finies $\{x_1, x_2, ... x_n\}$. Suppose qu'un échantillon de taille N du caractère X, il y a des valeurs x_i (i = 1, 2, …, n), c'est-à-dire la valeur x_i porte l'effectif n_i. La fréquence de la valeur x_i est $f_i = \frac{n_i}{N}$.
>
> Alors le tableau de distribution des fréquences de l'échantillon est :

Valeur x	x_1	x_2	…	x_n
Fréquence f_i	f_1	f_2	…	f_n

> On peut considérer le caractère X comme une variable aléatoire discrète avec l'ensemble des valeurs $\{x_1, x_2, ... x_n\}$.
>
> Suppose que la distribution des probabilités de X soit :

X	x_1	x_2	$\cdots$	x_n
P	p_1	p_2	$\cdots$	p_n

On sait que la fréquence f_i est une valeur approchée de la probabilité p_i. Le tableau de distribution des fréquences d'un échantillon nous donne donc une « image » approchée du tableau de distribution des probabilités de X.

La moyenne de l'échantillon est : $\bar{x} = \dfrac{\sum_{i=1}^{n} n_i x_i}{N} = \sum_{i=1}^{n} x_i \dfrac{n_i}{N} = \sum_{i=1}^{n} x_i f_i$

Comme $f_i \approx p_i$ alors : $\bar{x} = \sum_{i=1}^{n} x_i f_i \approx \sum_{i=1}^{n} x_i p_i \approx E(X)$

Ainsi, la moyenne de l'échantillon est une valeur approchée de l'espérance de X.

De même façon, la variance de l'échantillon est

$$s^2 = \frac{\sum_{i=1}^{n} n_i (x_i - \bar{x})^2}{N} = \sum_{i=1}^{n} (x_i - \mu)^2 p_i \approx \sum_{i=1}^{n} (x_i - \bar{x})^2 f_i \approx V(X)$$

où $\mu \approx E(X)$.

Enfin, la variance et l'écart-type de l'échantillon sont les valeurs approchées de la variance et l'écart-type de X.

(Manuel de classe 11, l'ensemble avancé, p. 91-92)

III.2. Notion d'incertitude aléatoire

Nous distinguons deux significations différentes de la notion d'incertitude aléatoire :

Il y a d'une part l'incertitude qui porte, dans une expérience aléatoire unique, sur le résultat à venir de l'expérience parmi tous les résultats possibles dont on connaît l'ensemble. C'est le cas lorsqu'on construit un échantillon d'une population connue – échantillon qui peut-être de taille 1 dans le cas d'une expérience aléatoire unique. Dans ce cas l'incertitude porte sur le résultat de l'échantillonnage.

Il y a d'autre part l'incertitude qui porte sur les paramètres d'une population inconnue à laquelle seule des expériences aléatoires répétées donnent accès. C'est le cas lorsqu'on a extrait un échantillon d'une population inconnue et que l'on infère sur la population. Dans ce cas l'incertitude porte sur le résultat de l'inférence.

Abordée à travers les définitions d'une expérience aléatoire (« les résultats ne sont pas prévisibles »), la première signification est développée dans l'enseignement vietnamien sur la notion de probabilité laquelle quantifie la possibilité d'un résultat de l'expérience aléatoire parmi un ensemble de résultats possibles :

On peut lire dans les manuels actuels de la classe 11 :

Dans la vie quotidienne, quand on parle d'un événement, on dit souvent que cet

événement a une grande possibilité d'apparition et que l'autre événement a une faible possibilité d'apparition. Les mathématiques quantifient ces possibilités en assignant à chaque événement un nombre non-négatif, inférieur ou égal à 1 ; ce nombre est appelé la probabilité de l'événement. La probabilité de l'événement A est noté P(A). Elle mesure la possibilité objective d'apparition de l'événement A.

(Manuel de classe 11, l'ensemble avancé, p. 71)

Une caractéristique qualitative importante d'un événement concernant une expérience est si cet événement se réalise ou non lors de la réalisation d'une expérience. Une question est posée : est-ce que cet événement peut se réaliser ? Quelle est la possibilité de son apparition ? Le problème est qu'il faut assigner un nombre convenable à cet événement pour évaluer sa possibilité d'apparition. On appelle ce nombre la *probabilité de l'événement*.

(Manuel de classe 11, l'ensemble élémentaire, p. 65)

L'assignation de cette probabilité à l'événement est réalisée dans l'institution selon la méthode laplacienne sous forme de probabilité a priori supposée unique et objectivable (car calculable dans l'univers des résultats possibles). L'absence de l'approche fréquentiste dans les activités des élèves conduit à privilégier le calcul de la probabilité a priori dans la résolution de l'incertitude aléatoire pour une expérience unique et à se passer de la réalisation effective de l'expérience aléatoire.

La deuxième signification qui s'appuie sur la notion de sondage dans une population n'est pas un objet d'enseignement au Vietnam. L'incertitude sur le résultat d'une suite d'expériences aléatoires « semblables » n'y relève donc pas d'un calcul de probabilité (comme l'incertitude dans le résultat de la réalisation d'une seule expérience aléatoire) et seule est présente l'idée de convergence de la suite des fréquences empiriques vers la probabilité.

Le rapport institutionnel à la notion d'incertitude aléatoire prend trois formes :

- une forme non mathématisée issue du langage commun avec la notion d'imprévisibilité

- une forme mathématisée issue du langage des probabilités avec le calcul de la probabilité d'un événement pour une expérience aléatoire unique bien définie

- une forme mathématisée issue du langage de la statistique descriptive avec l'approximation d'une probabilité par des fréquences empiriques dans une répétition d'expériences aléatoires identiques.

Comment ces trois formes usent-elles des différents objets mathématiques apportés par les enseignements de la statistique et des probabilités ? En particulier, comment l'incertitude aléatoire s'exprime-t-elle dans l'enseignement vietnamien en dehors des situations où la technique laplacienne n'est pas exploitable ? L'approche fréquentiste à peine évoquée dans l'institution est-elle disponible ? Sous quelles formes ? Dans quelles conditions ? Comment, dans ce cas, les notions de fréquences et de probabilités s'articulent-elles ? Comment la notion de risque complètement absente de l'institution peut-elle être construite ? L'ingénierie didactique cherche à répondre à ces questions en mettant en scène la loi des grands nombres et en interrogeant son caractère probabiliste.

IV. Organisation de l'ingénierie

Notre ingénierie s'organise en trois séances de classe, chacune répondant à un objectif :

- la dévolution du problème d'incertitude aléatoire et la mise au point d'une loi empirique des grands nombres par confrontation des fréquences empiriques et de fréquences théoriques (les proportions possibles de chaque couleur) ;

- l'exploration du caractère probabiliste de cette loi des grands nombres pour décider d'une fréquence théorique inconnue en appréciant le risque de se tromper ;

- l'élargissement du fonctionnement de cette loi des grands nombres au problème de la recherche d'un échantillon représentatif d'une population en l'absence de fréquences théoriques accessibles.

Dans les trois séances, les liens entre statistique et probabilités se tissent au fil de la confrontation entre les observations de ce que produit la machine de hasard et les prévisions sur sa structure :

> Ce qu'on a vu, va devenir l'image du passé de la machine de hasard, c'est-à-dire une statistique, ce qu'on prévoit va devenir des événements plus ou moins probables associés à la machine et les deux sont reliées par des hypothèses sur la structure de la machine que l'on peut facilement imaginer.
>
> (Brousseau [2005])

Dans la deuxième séance, l'hypothèse consistant à retenir telle ou telle structure de la machine est conditionnée par le résultat d'un échantillon et le risque de se tromper se présente comme une probabilité conditionnelle expérimentale[107].

Dans la troisième séance, la structure de la machine de hasard est complexifiée pour passer d'une estimation ponctuelle à une estimation par intervalle d'une fréquence théorique.

La suite des trois séances est bâtie sur les changements successifs des valeurs de la variable didactique (nombre des compositions ; nombre de tirages possibles) :

Nombre des compositions / Nombre des tirages possibles	petit	grand
petit	Séance 1	
grand	Séance 2	Séance 3

Tableau 72. Les choix des valeurs des variables dans trois séances

Le passage de la valeur (petit ; petit) à la valeur (petit ; grand) permet d'explorer la loi des grands nombres en acte et le passage de la valeur (petit ; grand) à la valeur (grand ; grand) vise à installer le caractère probabiliste de la loi des grands nombres.

[107] L'expression « probabilité expérimentale » sera commentée plus loin.

Chapitre D2
L'ingénierie didactique

I. Séance 1

I.1. Introduction

Objectifs de la séance : installer la répétition d'une expérience aléatoire et révéler une loi pragmatique des grands nombres.

L'enseignant dispose de bouteilles opaques avec bouchon translucide, contenant 5 boules, des rouges et des vertes : on ne peut voir qu'une boule à la fois par le bouchon. Les bouteilles n'ont pas toutes la même composition.

Figure 9. La bouteille fabriquée pour l'expérimentation

I.2. Protocole et analyse a priori de la première phase (20 minutes)

Objectif de la phase : passage de l'observation à la répétition de l'observation.

Les élèves sont répartis en 8 groupes de 4 ou 5.

Chaque groupe reçoit l'une des bouteilles et une feuille (feuille 1) et le professeur ne mentionne ni que les boules sont de deux couleurs ni le nombre de boules.

L'enseignant a fabriqué des bouteilles de 4 types (4R-1V, 3R-2V, 2R-3V et 1R-4V) mais il ne connaît pas le contenu de la bouteille qu'il donne au groupe[108] (distribution au hasard de bouteilles indiscernables entre elles). Nous nous accordons avec Briand :

> « Le professeur ne connaît pas les contenus des bouteilles, ce qui permet à celui-ci de maintenir plus aisément l'aspect adidactique de la situation »
>
> (Briand [2005], p. 263)

[108] L'enseignant dispose de 10 bouteilles et chaque composition est représentée par au moins bouteille.

Consigne :

Chaque groupe dispose d'une bouteille contenant des boules. On ne peut voir qu'une boule à la fois : on dira que l'on interroge la bouteille. Que pouvez-vous dire sur le contenu de votre bouteille ? Vous avez 5 minutes pour noter toutes vos déclarations sur la feuille (1) que nous vous distribuons.

La feuille 1 :

Feuille 1 (noms, prénoms, lycée, classe, n° groupe)

Durée : 5 minutes

Que pouvez-vous dire sur le contenu de votre bouteille ? Noter ci-dessous toutes les déclarations que vous jugez pertinentes à propos de ce contenu ainsi que vos remarques.

Au bout de 5-10 minutes, l'enseignant ramasse les feuilles 1 de tous les groupes puis organise une synthèse portant sur les déclarations et les remarques.

Analyse a priori

La consigne demande à l'élève une activité d'observation sur un objet étrange, une bouteille opaque contenant des boules et l'élève peut conduire cette observation soit en « écoutant » la bouteille quand il la remue et seul est accessible (difficilement) le nombre de boules soit en la « regardant » et seule est accessible (facilement) la couleur de la boule qui est dans le goulot. On peut donc s'attendre à des constats auditifs mais surtout visuels, ces derniers étant fortement suggérés par la consigne.

A la différence de l'expérimentation de Briand où le nombre de boules est donné, la consigne pourrait être comprise par le groupe comme visant le nombre de boules dans sa bouteille. Cette interprétation de la consigne peut conduire au recours de l'observation auditive même après une série d'observations visuelles mais nous faisons l'hypothèse que les élèves de la classe 11 confrontés à l'art difficile « d'écouter » la bouteille abandonneront observation auditive au profit de l'observation visuelle.

Pour Briand, l'incitation à recommencer l'observation provient d'une anticipation du résultat d'une observation contrariée par le constat issu de cette observation :

> « Le va-et-vient entre le passé (les faits observés) et ce que l'on peut vérifier (l'expérience construite, c'est-à-dire l'expérimentation) montre à la fois l'incertitude et le début des régularités qui incitent à recommencer […] »
>
> (Briand [2005], p. 264)

De ce point de vue, les observations auditives et visuelles diffèrent.

L'observation (« expérience immédiate » Briand 2002) auditive faite une fois donne une information sur l'ensemble des boules (j'entends trois boules, les boules sont en bois, etc.) et sa répétition pour un même observateur peut effectivement vouloir confirmer le résultat de la première observation. Mais ce résultat n'est pas objectivable par le groupe dans les

conditions de l'observation et devrait être controversé. Cette expérience sera-t-elle répétée par le groupe en faisant passer la bouteille d'élève à élève et en relevant le nombre de boules annoncé ? Les élèves identifieront-ils alors cette répétition d'observations visuelles à une expérience aléatoire ?

L'observation visuelle faite une fois ne donne d'information que sur l'une des boules sans que l'observateur puisse maîtriser le choix de la boule. Ce constat débouche sur un événement élémentaire facilement objectivable et non controversable (la boule que l'on voit est rouge, ou bien elle est verte, etc.). L'assurance qu'il y a plusieurs boules incite à recommencer l'observation pour connaître la couleur des autres boules. L'action peut-être recommencée dans les mêmes conditions. Le résultat obtenu les fois suivantes est-il considéré comme ne dépendant pas des résultats précédents ? Si oui, nous dirons que l'élève a construit implicitement une expérience aléatoire. Nous faisons l'hypothèse que les élèves de la classe 11 déjà instruits en probabilités pourront identifier la répétition d'observations visuelles à une expérience aléatoire telle qu'elle vit dans l'institution d'enseignement.

Résumons : l'enseignant n'explicite pas ce qu'il appelle « observer la bouteille » ; quelle expérience les élèves construiront-ils et lui donneront-ils le statut d'expérience aléatoire ?

Les variables didactiques

a) Le nombre de couleurs : 2 couleurs, un choix qui entretient la répétition de l'expérience visuelle alors qu'une seule couleur aurait conduit à vite l'abandonner et que plus de deux couleurs aurait augmenté fortement le nombre et la complexité des constats possibles.

b) Le nombre de boules : il n'est pas donné ce qui augmente le champ des incertitudes sur le contenu de la bouteille. Comment vont se combiner les stratégies de recherche du nombre de boules et celles de recherche du rapport des nombres de boules de chaque couleur.

Stratégies

Nous n'examinons ici que celles qui s'intéressent aux nombres de boules ou au rapport entre ces nombres.

S_{auditive} : agiter la bouteille pour l'écouter.

Elle permet de s'assurer qu'il y a bien plusieurs boules et d'évaluer le nombre de boules présentes dans la bouteille. Sa répétition peut conduire à relever plusieurs nombres. Que faire de ces nombres ? Les élèves en feront-ils une statistique ? Calculeront-ils la moyenne ?

S_{visuelle} : renverser la bouteille pour voir la bouteille à travers le bouchon.

Cette stratégie permet de relever des couleurs. Sa répétition peut conduire à calculer le rapport ou la différence des nombres de couleurs relevés et à des annonces sur le nombre de boules.

Voici les réponses possibles sur le contenu quantitatif de la bouteille :

Nb. de boules	Ce qui est observé	= 2 (le nombre des couleurs vues)
		= 5 (j'entends 5 boules)
	Au-delà de ce qui est observé	= n (il y a deux couleurs vues et l'une des couleurs apparaît n –1 fois plus souvent que l'autre)
		> 2 (il y a deux couleurs vues et l'une des couleurs apparaît plus souvent que l'autre)
		= n (il y a deux couleurs vues et le rapport des effectifs des couleurs vaut p/q où p + q = n)
Comparaison des nombres de boules par couleur		plus de rouges que des vertes
		plus de vertes que des rouges
		autant de vertes que des rouges
		rapport exact des effectifs R/V
		rapport arrondi des effectifs R/V

Tableau 73. Les réponses possibles sur le contenu quantitatif de la bouteille

I.3. Déroulement et analyse a posteriori de la première phase

Dans la salle d'étude, les tables sont regroupées pour 8 groupes de 5 à 6 élèves, notés Groupe 1, Groupe 2, … Groupe 8 (la classe 11A1 se compose 44 élèves).

À la fin de la première phase, l'enseignant a résumé publiquement toutes les annonces proposées par 8 groupes.

Rappelons que les élèves ne connaissent ni le nombre de boules ni le nombre de couleurs.

Au début, les élèves ne savent pas renverser la bouteille et les groupes n'observent que l'apparence de la bouteille ou l'agitent sans la renverser. Aux sons que fait la bouteille les groupes constatent qu'il y a plusieurs objets dans la bouteille (l'un des groupes affirme qu'il y en a exactement 5) ; certains décrivent la nature de ces objets (des boules en verre, des pièces de monnaie (groupe 7), des morceaux de métaux, des grains de sable (groupe 4), etc.). Ils cherchent à comprendre ce que l'enseignant attend comme résultat :

El	Observe-t-on la couleur ?
Ens	(*en rigolant*) Oui, si vos yeux sont … bons …
El	Je vois la couleur rouge et verte, et la blanche aussi
Ens	D'accord, observez bien !
	(*Elève ne sait pas encore renverser la bouteille*)
	Pourquoi y-a-t il cette fente ?
El1	Ah, je ne savais pas qu'il y a une fente.
	(*cet élève commence à renverser la bouteille*)
El2	Une boule rouge, une verte… Il y a deux rouges.

(Phase 1, Script de la séance 1)

Ils découvrent qu'il y a des boules de couleurs différentes et décident de continuer l'expérience visuelle. Le constat du changement de couleur les conduit à répéter l'expérience. Mais la déclaration « il y a deux rouges » prouve que les deux boules rouges

apparues successivement sont nécessairement des boules différentes donc que l'élève n'a pas encore construit la répétition de l'observation visuelle comme une expérience aléatoire.

Tous les groupes finissent par observer les boules à travers le bouchon et abandonnent la stratégie auditive. Voici une synthèse des fiches rendues par les groupes :

Groupe	Description des objets dans la bouteille	Nb. de boules	Composition	Justifications et commentaires
7	- boules R-V - grains de sables - monnaies	3	2R-1V ou 1R-2V	Aucune justification
3	- boules R-V	> 2	R = V	
4	- boules R-V - monnaies		R > V	
1	- boules R-V - mêmes masses	~ 5	3R-2V	« il est possible qu'il y ait 3R et 2V »
5	- boules R-V	≥ 2	R > V	« Parmi 10 expériences, on a vu 3 fois une boule verte et 7 fois une boule rouge. S'il y a plus de 2 boules, la *probabilité* de voir une R est plus grande »
6	- boules R-V	3	1R-1V + 1	« V-V-V-R-R-V-V-R-R-V la troisième boule est peut-être verte »
2	- boules R-V	3	2R-1V	« 32 essais ; *proportion* R/V ~ 2 ; c'est la *variable aléatoire discrète* ; *il est possible* qu'il y ait 2R-1V »
8	- boules R-V	4	3V-1R puis 2V-2R	« 30 essais ; la *probabilité* d'apparition d'une boule verte est $\frac{2}{3}$ ($\frac{20}{30}$ *de fois d'essai*) »

Tableau 74. Résumé des réponses des groupes - phase 1

Les trois premiers groupes ne justifient pas leurs déclarations par le recours à la répétition. Le groupe 1 fait allusion à une répétition mais sans dresser de statistique. Tous les groupes ont réalisé des observations visuelles répétées mais seulement les quatre derniers groupes dressent des statistiques (nombre d'essais, nombre de rouges, nombre de vertes) pour justifier leurs nombres de boules. Ils présentent des raisonnements qui justifient à leurs yeux leurs déclarations :

> « À la différence des élèves de fin d'école primaire, les élèves de seconde sont vite conscients de la possibilité de mettre en place des raisonnements. »
>
> (Briand [2005], p. 264)

Pour les quatre derniers groupes du tableau, il y a utilisation de l'observation visuelle répétée et le nombre d'observations est supérieur au nombre de boules annoncés qui est calculé à partir du rapport des effectifs pour chaque couleur. Ces groupes utilisent dans un contexte statistique des expressions apportées par l'enseignement des probabilités (probabilité, variable aléatoire, expérience). Par exemple, dans la fiche 1 du groupe 8, on voit une technique statistique :

Première fois : nous essayons 10 fois, il apparait 5 vertes / 5 rouges (fois)

Deuxième fois : nous essayons 10 fois, il apparait 8 vertes / 2 rouges (fois)

Troisième fois : nous essayons 10 fois, il apparait 7 vertes / 3 rouges (fois)

$\Rightarrow$ La <u>probabilité</u>[109] d'apparition d'une boule verte est $\frac{2}{3}\left(\frac{20}{30}\text{ fois d'essai}\right)$

La <u>probabilité</u> d'apparition d'une boule rouge est $\frac{1}{3}\left(\frac{10}{30}\text{ fois d'essai}\right)$

$\Rightarrow$ Dans la bouteille, il y a des boules de deux couleurs verte et rouge.

En concrète, $\begin{cases} Verte: 3\ 2 \\ Rouge: 1\ 2 \end{cases}$

(La fiche 1 du groupe 8)

Ici, le mot *probabilité* remplace le mot *fréquence*. Peut-être faut-il y voir la référence à la notion de *probabilité expérimentale* présentée dans la partie cours du manuel avancé de classe 11 ?

L'idée d'incertitude aléatoire, présente à travers le vocabulaire (expressions du type « peut-être », etc.) employé par ces groupes[110] ne se distingue pas nettement encore de l'idée d'imprécision réductible. Même si certains expriment un doute sur le nombre de boules qu'ils annoncent ce doute n'est pas radical au sens d'une impossibilité à connaître exactement le nombre de boules par l'expérience visuelle construite sans risquer de se tromper. Elle ne va pas par exemple jusqu'à émettre la possibilité d'une troisième couleur non vue dans les observations. Le déterminisme sous-jacent aux raisonnements implicitement ou explicitement à l'origine de certaines réponses, côtoie donc une première appréhension du caractère aléatoire du résultat de l'observation visuelle.

Pour terminer la phase 1, l'enseignant présente et commente les réponses des 8 groupes :

Ens Après la première activité, ce que vous avez « vu », ce sont :
- dans la bouteille, il y a des boules
- dans la bouteille, il y a tant de boules
- dans la bouteille il y a deux couleurs : rouge et verte.

Certains groupes disent que telle couleur est plus nombreuse que l'autre couleur. Par exemple les rouges sont plus nombreuses que les vertes ; ils ont noté leurs tirages et ont vu que telle couleur apparaît plus fréquemment que l'autre couleur.

(Phase 1, Script de la séance 1)

Ce bilan a publiquement mis en avant deux questions que tous les groupes se sont posées :

- combien de boules dans la bouteille ?

- combien de boules rouges et combien de boules vertes ?

---

[109] Souligné par nous.

[110] Elle est complètement absente des réponses apportées par les trois premiers groupes du tableau.

La phase 1 de l'ingénierie a non seulement permis l'émergence dans la classe de ces deux questions mais aussi montré la disponibilité des calculs de fréquences des couleurs (cf. groupe 8) ou de proportions rouge/vert (cf. groupe 2) sur des statistiques pour y répondre. Pour permettre l'apparition généralisée de statistiques et l'utilisation de ces calculs dans la prise de décision relative à la proportion de chaque couleur, l'enseignant donne dans la phase suivante le nombre de boules. Suivant en cela l'ingénierie de Brousseau, ce choix permet de donner du souffle à ces stratégies calculatoires et peut conduire à une explicitation de la loi des grands nombres. En effet les rapports (ou les fréquences) théoriques autour desquels doivent osciller les rapports (ou les fréquences) empiriques sont alors calculables et ils sont peu nombreux car le nombre de boules dans la bouteille est faible. La séance 3 prolongera ce choix en imaginant un nombre de boules tellement grand que les rapports (ou les fréquences) empiriques ne pourront pas s'appuyer directement sur les rapports (ou les fréquences) théoriques.

I.4. Protocole et analyse a priori de la deuxième phase (25 minutes)

Objectif de la phase : formulation de conjectures sur la composition ; loi des grands nombres en actes.

Consigne :

On cherche toujours à connaître la bouteille en l'interrogeant. Je vous donne une information supplémentaire : chacune de vos bouteilles contient 5 boules mais pas plus que vous je ne connais la composition en rouges et en vertes de votre bouteille.

Cette fois-ci on cherche à connaître la composition de votre bouteille.

Vous avez 10 minutes pour noter sur la feuille 2 votre proposition et les arguments qui la soutiennent. Utilisez la feuille 2 bis comme brouillon.

Voici la fiche 2 :

Fiche 2 (noms, prénoms, lycée, classe, n° groupe)

Durée : 10 minutes

Que pouvez-vous dire sur la composition de votre bouteille ? Noter ci-dessous votre proposition et les arguments qui la soutiennent.

La composition de votre bouteille :

Donnez vos arguments :

Etes-vous moyennement sûrs ? Etes-vous tout à fait sûrs ?

Pourquoi ?

L'enseignant circule entre les groupes pour les inciter à répondre aux deux questions : la composition de la bouteille et le degré de certitude. Au bout de 10-15 minutes, l'enseignant ramasse les feuilles 2 de tous les groupes. Il relève publiquement toutes les compositions proposées et les synthétise au tableau.

Voici la feuille utilisée par l'enseignant pour regrouper au tableau les numéros des groupes qui annoncent la même composition :

N° groupe	1R4V		2R3V		3R2V		4R1V	
	moyennement sûr	tout à fait sûr	moyennement sûr	tout à fait sûr	moyennement sûr	tout à fait sûr	moyennement sûr	tout à fait sûr
1	X							
2				X				
3								
4								
…								

Pour un groupe qui se déclare « tout à fait sûr » de la composition de sa bouteille, l'enseignant demande aux élèves de ce groupe de faire voir avec leur bouteille aux autres élèves comment ils sont arrivés à ce degré de certitude.

Analyse a priori

Le fait d'annoncer qu'il y a 5 boules fait passer les élèves dans une nouvelle situation : chercher la composition par la stratégie visuelle en abandonnant la stratégie auditive qui a surtout servi précédemment à évaluer le nombre de boules dans la bouteille. La répétition des observations est ainsi favorisée ainsi que la comparaison des fréquences ou des rapports observés avec des fréquences ou des rapports théoriques issus du calcul des probabilités. Cette comparaison devrait être à l'origine de raisonnements sur les compositions :

> « À la différence des élèves de fin d'école primaire, les élèves de seconde sont vite conscients de la possibilité de mettre en place des raisonnements. »
>
> (Briand [2005], p. 264)

Brousseau parle d'hypothèses sur la machine de hasard qu'est la bouteille :

> « Ce qu'on a vu, va devenir l'image du passé de la machine de hasard, c'est-à-dire une statistique, ce qu'on prévoit va devenir des événements plus ou moins probables associés à la machine et les deux sont reliées par des hypothèses sur la structure de la machine que l'on peut facilement imaginer. »
>
> (Brousseau [2001], p. 4)

Ces hypothèses sont les compositions possibles de la bouteille (il y en a quatre).

Les variables didactiques

Nous faisons l'hypothèse que la préférence des élèves va à des séries de 10 tirages ou d'un multiple de 10 tirages, habituellement utilisées dans les opérations de comptage.

a) Le nombre de boules égal à 5 favorise l'obtention de fréquences empiriques exactement égales aux fréquences théoriques et peut conduire les élèves à annoncer une composition dont ils se déclareraient être « totalement sûrs » ; avec un nombre de boules égal à 7, une telle certitude basée sur la stricte égalité de la fréquence empirique et de la fréquence théorique ne pourrait provenir que d'un nombre de tirages multiple de 70 ce qui ne serait pas réalisable dans le temps imparti à cette phase.

b) Le nombre n de boules est « petit » : le nombre de fréquences et de rapports théoriques est n – 1. Donc, pour n petit (disons n < 10), le nombre de compositions possibles est lui-même petit et le choix de la composition à partir de la fréquence ou du rapport empirique peut se faire sans calcul de proportionnalité, sur la base d'un intervalle de décision non explicité. Si n est plus grand, la liste des compositions possibles est trop importante et le choix de la composition appelle un calcul de proportionnalité sur les effectifs car la comparaison des rapports (ou des fréquences) empiriques et des rapports (ou des fréquences) théoriques est trop coûteuse.

Stratégies de décision de la composition

Toujours selon Briand :

> « Les élèves sont confrontés à au moins deux problèmes :
>
> - tout d'abord, concevoir les intervalles que nous nommons intervalles de confiance[111] et qui se construisent, dans la classe, non pas à partir de la probabilité, mais de résultats statistiques et d'une compréhension, en acte, de la loi des grands nombres : les fréquences de l'événement (3V, 2R) se répartissent de part et d'autre de la fréquence la plus probable[112].
>
> - ensuite, commencer à évaluer le risque encouru à faire une prévision fausse à l'issue de n tirages […]»
>
> (Briand [2005], p. 255)

En effet la notion d'intervalle de décision pourrait venir de la confrontation du calcul des fréquences observées et des fréquences théoriques[113] ou bien des rapports observés et des

---

[111] Qu'il nommera plus tard, comme nous le ferons dorénavant, intervalles de décision. Briand dénombre quatre intervalles de décision :]0 ; 0,3], [0,3 ; 0,5], [0,5 ; 0,7], [0,7 ; 1[qui contiennent chacun une fréquence théorique et une seule, respectivement 0.2, 0.4, 0.6 et 0.8, et qui permettent de discriminer toute fréquence observée par sa proximité à une fréquence théorique.

[112] L'appartenance d'une fréquence observée à un intervalle de décision rend la fréquence observée plus proche de la fréquence théorique attachée à l'intervalle que de toutes les autres fréquences théoriques ; cette fréquence théorique est donc la plus probable des fréquences théoriques.

[113] Fréquence théorique et non probabilité car on connaît le nombre (5) de boules et le nombre (2) de couleurs : on peut donc envisager toutes les compositions (en nombre fini = 4). Les fréquences théoriques sont donc 0.2, 0.4, 0.6 et 0.8 (on a exclu la possibilité d'une seule couleur).

rapports théoriques[114], les fréquences et les rapports théoriques étant calculés à partir des compositions possibles : s'il y avait 4R-1V, la fréquence théorique des boules rouges serait 4/5 et le rapport théorique serait 4/1; alors que s'il y avait 3R-2V la fréquence théorique des boules rouges serait de 3/5 et le rapport théorique serait 3/2. Cependant le recours aux fréquences et aux rapports doit être précédé d'un recueil des données et on peut s'attendre à ce que chaque groupe procède à un dénombrement (écrit ou non) des boules apparues. Ces dénombrements peuvent alors conduire à l'utilisation de la *proportionnalité entre les effectifs respectifs des échantillons et de la bouteille*. Par exemple, supposons 15 boules dans la bouteille et qu'un échantillon de 50 tirages donne 28 vertes et 22 rouges : le nombre de vertes dans la bouteille se calcule par 28 x 15 / 50 = 8,4 qu'on arrondit à l'entier le plus proche ce qui aboutit à la composition 8V-7R. Cette stratégie n'a pas besoin de la liste des compositions possibles et le calcul de l'arrondi à l'entier le plus proche permet le choix entre les deux compositions 8V-7R et 9V-6R sans avoir à comparer la fréquence empirique et les fréquences théoriques voisines ou le rapport empirique et les rapports théoriques voisins (stratégie « *proportionnalité* »)

Certes on peut s'attendre aussi à la confrontation en acte des fréquences observées et des fréquences théoriques[115] ou à celle des rapports observés et des rapports théoriques. Cette confrontation devrait utiliser des distances entre valeurs empiriques et valeurs théoriques voisines (stratégie « *distances* »).

Nombre		Fréquences théoriques		Rapports théoriques	
Rouge	Verte	Rouge	Verte	Fraction	Décimal
1	4	1/5	4/5	1/4	0,25
2	3	2/5	3/5	2/3	0,66…
3	2	3/5	2/5	3/2	1,50
4	1	4/5	1/5	4	4,00

Tableau 75 : Les rapports et les fréquences théoriques possibles

Mais pourquoi choisir les fréquences plutôt que les rapports ? La *comparaison de distances à deux fréquences théoriques conduisant à la création d'intervalles de décision* comme le suggère Briand comporte un pas supplémentaire vers une vision plus globale de l'ensemble des possibles qui ne devrait vraiment être explicitée que pour répondre du degré de certitude auquel le groupe est arrivé. C'est pour viser une telle explicitation que nous posons la deuxième question, à savoir « quel est votre degré de certitude ? ». Cette question oblige à confronter le point de vue déterministe déclenché par la décision à prendre sur la composition de la bouteille à partir de calculs sur les statistiques et le point de vue probabiliste crée par les aléas issus des interrogations de la bouteille. Voici une petite étude théorique qui permet de souligner les enjeux mathématiques de cette confrontation.

[114] Les intervalles de décision sur les rapports ne sont pas les mêmes que ceux sur les fréquences.

[115] Fréquence théorique et non probabilité car on connaît le nombre (5) de boules et le nombre (2) de couleurs : on peut donc envisager toutes les compositions (en nombre fini = 4). Les fréquences théoriques sont donc 0.2, 0.4, 0.6 et 0.8 (on a exclu la possibilité d'une seule couleur).

Les interrogations de la bouteille : étude théorique

On fait N interrogations de la bouteille dont les **n** boules sont soit vertes soit rouges avec au moins une verte ou au moins une rouge. A l'issue des N interrogations on peut calculer :

- la fréquence empirique du nombre de vertes, notée f_e

- le rapport théorique du nombre de vertes au nombre de rouges, noté r_e

- la différence entre les nombres de vertes et de rouges, notée d_e .

Parmi les n–1 compositions possibles, la composition de la bouteille sur la base des N interrogations peut se décider à partir :

- de la fréquence empirique, en comparant cette fréquence aux différentes fréquences théoriques associées aux différentes compositions possibles ;

- du rapport empirique, en comparant ce rapport aux différents rapports théoriques associés aux différentes compositions possibles (il y en a n–1) ;

- de la différence empirique, en comparant cette différence aux différentes différences moyennes (différence des fréquences théoriques x N)

1. Les intervalles de décision des fréquences

Si la décision se prend en choisissant la fréquence théorique la plus proche, cela revient à placer la fréquence empirique dans un intervalle qu'on qualifie d'intervalle de décision.

Les n – 1 intervalles de décision sont $[0;\frac{3}{2n}[$ noté I_1 ; $]\frac{2i-1}{2n};\frac{2i+1}{2n}[$ notés I_i et $]\frac{2n-3}{2n};1[$ noté I_{n-1} où i est un entier $2 \leq i \leq n-2$.

Point de vue déterministe

Si la fréquence empirique f_e appartient à I_i alors on décide la composition (i, n – i) pour la bouteille. Remarquons que si f_e est l'une des bornes autre que 0 et 1, il y a indécision. Cette règle de décision s'accompagne implicitement d'un choix sur le nombre de tirages à partir duquel on l'applique mais, ce choix étant fait, cette règle n'est assortie d'aucune probabilité et nous pouvons la qualifier de déterministe.

Point de vue probabiliste

Si la fréquence empirique f_e appartient à I_i alors on décide la composition (i, n – i) pour la bouteille en annonçant le risque de se tromper. Le risque R de se tromper est ici la

probabilité de l'événement $f_e \notin I_k$ alors que la composition de la bouteille est (k, n – k)[116].

Choisissons une bouteille de composition (k vertes, (n – k) rouges) avec $1 \leq k \leq n - 1$ et procédons à N interrogations. Le nombre d'interrogations dont l'issue est une boule verte est une variable aléatoire notée X_N qui suit une loi binomiale de paramètres N et $\frac{k}{h}$ et la fréquence empirique f_e est la variable aléatoire $\frac{X_N}{N}$. La probabilité que f_e appartienne à I_i (i entier entre 1 et n – 1) est donc égale à $p(X_N \in N.I_i)$ c'est-à-dire à la somme des $p(X_N = h)$ où $h \in N.I_i$; plus formellement, cette probabilité s'écrit $\sum_{h=1}^{n}\binom{n}{h}\left(\frac{k}{h}\right)^h\left(\frac{n-k}{h}\right)^{n-h}$ où h est un nombre entier de $N.I_i$.. Le risque R est égal à $p(f_e \notin I_k)$ qui vaut $1 - p(f_e \in I_k)$ car on supposera qu'en cas d'indécision possible, il est procédé à au moins un tirage supplémentaire faisant passer le nombre d'interrogations de N à N + 1. Ce risque est évalué en % qui se comprend ainsi : dans R % des cas, la décision prise sur la base de N interrogations est erronée.

Exemples pour n « petit (n ≤10) » et N = 10

- **n = 4** ; k = 2 ; N = 10 ; I_2 =]3/8 ; 5/8 [; h = 4, 5 ou 6

$p(f_e \in I_2) \approx 0{,}42$; $R \approx 58$ %

- **n = 5 ; k = 3** ; N = 10 ; I_3 =]0,5 ; 0,7[; h = 6

$p(f_e \in I_3) = \binom{10}{6}.0{,}6^6.0{,}4^4 \approx 0{,}25$; $p(f_e = 0{,}5 \text{ ou } 0{,}7) \approx 0{,}415$ et $R \approx 33{,}5$ %

- **n = 5 ; k = 4** ; N = 10 ; I_4 =]0,7 ; 1] ; h = 8, 9 ou 10

$p(f_e \in I_4) = \binom{10}{8}.0{,}8^8.0{,}2^2 + \binom{10}{9}.0{,}8^9.0{,}2^1 + \binom{10}{10}.0{,}8^{10}.0{,}2^0 \approx 0{,}678$;

$p(f_e = I_7) \approx 0{,}215$ et $R \approx 10{,}7$ %

- **n = 7** ; k = 4 ; N = 10 ; I_4 =]0,5 ; 9/14 [; h = 6

$p(f_e \in I_4) \approx 0{,}25$; $p(f_e = 0{,}5) \approx 0{,}22$ et $R \approx 53$ %

- **n = 9 ; k = 5** ; N = 10 ; I_k =]0,5 ; 0,5 + 1/9 [; h = 6

$p(f_e \in I_k) \approx 0{,}24$; $p(f_e = 0{,}5) \approx 0{,}23$ et $R \approx 53$ %.

---

[116] En se référant à la théorie des tests d'hypothèse paramétriques (ici, un test de proportion), c'est le risque de deuxième espèce puisque c'est la probabilité d'accepter l'hypothèse nulle ($f = f_i \neq f_k$) alors que cette hypothèse est fausse. On le calcule comme une probabilité conditionnelle.

Exemples pour n « petit (n ≤10) » et N > 30

Pour X_N on remplace (approximation) la loi binomiale par la loi normale de paramètres N. $\frac{k}{n}$ et $\sqrt{N.\frac{k}{n}.\frac{n-k}{n}}$ et pour f_e on utilise la loi normale de paramètres $\frac{k}{n}$ et $\sqrt{\frac{k}{n}.\frac{n-k}{nN}}$ si bien que la variable aléatoire $\frac{f_e - \frac{k}{n}}{\sqrt{\frac{k}{n}.\frac{n-k}{nN}}}$ notée U suit la loi normale centrée réduite.

$$p(f_e \in I_k) = p\left(\frac{\frac{2k-1}{2n} - \frac{k}{n}}{\sqrt{\frac{k}{n}.\frac{n-k}{nN}}} < U < \frac{\frac{2k+1}{2n} - \frac{k}{n}}{\sqrt{\frac{k}{n}.\frac{n-k}{nN}}}\right) = 2.p\left(U < \frac{\frac{1}{2n}}{\sqrt{\frac{k}{n}.\frac{n-k}{nN}}}\right) - 1$$

$= 2p\left(U \langle \frac{1}{2}\sqrt{\frac{N}{k(n-k)}}\right) - 1$ qui, pour N fixé, diminue[117] lorsque k se rapproche de $\frac{n}{2}$.

- **n = 5 ; k = 3** ; N = 50 ; $p(f_e \in I_3) \approx 0{,}85$; R ≤ 15 %
- **n = 7 ; k = 4** ; N = 50 ; $p(f_e \in I_3) \approx 0{,}69$; R ≤ 31 %

Pour $\frac{k}{n}$ « à peu près fixe » la probabilité diminue quand n augmente et augmente quand N augmente.

Exemples pour n « grand (n > 10) » et N > 30

- **n = 100 ; k = 60** ; N = 50 ; $p(f_e \in I_{60}) \approx 0{,}056$
- **n = 1000 ; k = 600** ; N = 50 ; $p(f_e \in I_{600}) \approx 0{,}005$

pour n et N fixé, la probabilité augmente si k diminue

- **n = 1000 ; k = 60** ; N = 50 ; $p(f_e \in I_{60}) \approx 0{,}012$

[117] Le produit k.(n – k) est max pour k = $\frac{n}{2}$.

2. Les intervalles de décision des rapports

Si la décision se prend en choisissant le rapport théorique la plus proche, cela revient à placer le rapport empirique dans un intervalle qu'on qualifie d'intervalle de décision.

Les différents rapports théoriques sont $\frac{1}{n-1}, \frac{2}{n-2}, ..., \frac{n-1}{1}$ et les n – 1 intervalles de décision sont $\left[0\,;\frac{3n-4}{2(n-1)(n-2)}\right[$ noté J_1 ; $\left]\frac{1}{2}\left(\frac{i-1}{n-i+1}+\frac{i}{n-i}\right);\frac{1}{2}\left(\frac{i}{n-i}+\frac{i+1}{n-i-1}\right)\right[$ notés J_i et $\left]\frac{3n-4}{4}\,;+\infty\right[$ noté J_{n-1} où i est un entier $2 \le i \le n-2$.

Point de vue déterministe

Si le rapport empirique r_e appartient à J_i alors on décide la composition (i, n – i) pour la bouteille. Remarquons que si r_e est l'une des bornes autre que 0, il y a indécision. Cette règle de décision s'accompagne implicitement d'un choix sur le nombre de tirages à partir duquel on l'applique mais, ce choix étant fait, cette règle n'est assortie d'aucune probabilité et nous pouvons la qualifier de déterministe.

Notons que les décisions à partir des rapports peuvent ne pas être les mêmes qu'à partir des fréquences. Pour s'en convaincre, imaginons 25 interrogations dont l'issue est (18V-7R) alors $f_e = 0{,}72$ et on choisit la composition 4V-1R tandis que $r_e = \frac{18}{7}$ et on choisit 3V-2R.

Point de vue probabiliste

Choisissons une bouteille de composition (k vertes, (n – k) rouges) avec $1 \le k \le n-1$ et procédons à N interrogations. Le nombre d'interrogations dont l'issue est une boule verte est une variable aléatoire notée X_N donc le rapport r_e est la variable aléatoire $\frac{X_N}{N-X_N}$. Notons φ la fonction monotone croissante $X \to \frac{X}{N-X}$, on a $\varphi(f_e) = r_e$ donc $p(r_e \in J_k) = p(f_e \in \varphi^{-1}(J_k))$. Comme, en général, $\varphi^{-1}(J_k) \neq I_k$ on peut s'attendre que les probabilités $p(r_e \in J_k)$ et $p(f_e \in I_k)$ ne soient pas égales et que le risque R de se tromper soit différent de celui qui a été pris avec l'intervalle de décision des fréquences.

Exemple pour n « petit (≤10) » et N = 10

- **n = 5 ; k = 3** ; N = 10 ; $J_3 = \left]\frac{13}{12};\frac{14}{4}\right[$ (remarquer que $\varphi^{-1}(J_3) = \left]\frac{13}{25};\frac{11}{15}\right[\neq I_3 =$]0,5 ; 0,7[) ; comme $\frac{6}{4}$ et $\frac{7}{3}$ appartiennent à J_3, on a $p(r_e \in J_3) = \binom{10}{6}.0{,}6^6.0{,}4^4 +$

$\binom{10}{7}.0{,}6^7.0{,}4^3 \approx 0{,}25 + 0{,}215 \approx 0{,}465$; il n'y a pas d'indécision possible et donc R ≈ 53,5 % plus fort que le risque pris avec l'intervalle de décision des fréquences.

- **n = 5 ; k = 4** ; N = 10 ; $J_4 = \left]\frac{11}{15} ; +\infty\right[$; comme $\frac{8}{2}, \frac{9}{1}$ et $\frac{10}{0}$ appartiennent à J_4, on a $p(r_e \in J_4) = \binom{10}{8}.0{,}8^8.0{,}2^2 + \binom{10}{9}.0{,}8^9.0{,}2^1 + \binom{10}{10}.0{,}8^{10}.0{,}2^0 \approx 0{,}678$; il n'y a pas d'indécision possible et donc R ≈ 32,2 % nettement plus fort que le risque pris avec l'intervalle de décision des fréquences.

Exemple pour n « petit (≤10) » et N > 30

- **n = 5 ; k = 3** ; N = 50 ;

$$p(r_e \in J_3) = p\left(f_e \in \left]\frac{13}{25} ; \frac{11}{15}\right[\right) \approx p\left(\frac{\frac{13}{25} - \frac{3}{5}}{\sqrt{\frac{3.2}{5.5.50}}} < U < \frac{\frac{13}{15} - \frac{3}{5}}{\sqrt{\frac{3.2}{5.5.50}}}\right)^3 = p\left(U < \frac{10}{3\sqrt{3}}\right) - p\left(U < -\frac{2}{\sqrt{3}}\right)$$

$$= p\left(U < \frac{10}{3\sqrt{3}}\right) + p\left(U < \frac{2}{\sqrt{3}}\right) - 1 \approx 0{,}876 \text{ ; R = 12,4 \%}$$[118]

3. Les intervalles de décision des différences

A l'issue des N interrogations de la bouteille on calcule la différence entre le nombre de boules vertes et celui des boules rouges. Si la décision se prend en positionnant la différence empirique par rapport aux différences théoriques (N fois la différence des fréquences théoriques), cela revient à placer la différence empirique dans un intervalle qu'on peut encore qualifier d'intervalle de décision. Contrairement aux fréquences et aux rapports, un tel intervalle dépend de N et se conçoit comme un intervalle de $\mathbb{N}$.

Exemple avec n = 5 et N = 40 : supposons qu'on obtienne 27 vertes et 13 rouges donc 14 vertes de plus que de rouges. Sur 40 interrogations de la bouteille, la composition 3V-2R, fournit en moyenne 8 boules vertes de plus que de rouges tandis que la composition 4V-1R en fournit 24. Que choisir ?

Les raisonnements déterministe et probabiliste sont calqués sur ceux des fréquences.

Point de vue déterministe : 14 étant plus proche de 8 que de 24, on choisit 3V-2R.

Point de vue probabiliste : 14 étant plus proche de 8 que de 24, on choisit 3V-2R mais il y a un risque de se tromper à évaluer.

[118] Ce résultat ne diffère guère de celui des fréquences.

Si N n'est pas un multiple de n, les différences théoriques peuvent ne pas être des nombres entiers ce qui complique le choix.

Exemple avec n = 5 et N = 46 ; supposons qu'on obtienne 37 vertes et 9 rouges donc 18 vertes de plus que de rouges ; or, sur 46 interrogations de la bouteille, la composition 3V-2R, fournit en moyenne 9,2 boules vertes de plus que de rouges tandis que la composition 4V-1R en fournit 27,6. Que choisir ? Sur quel intervalle de décision s'appuyer, [9 ; 28] , [10 ; 27], etc. ?

Point de vue probabiliste pour N > 30

La variable aléatoire d_e est égale à $N(2f_e - 1)$ et suit approximativement la loi normale de paramètres $2N\frac{k}{n} - N$ (sa moyenne) et $\frac{2}{n}\sqrt{N.k.(n-k)}$ (son écart-type). On peut donc utiliser les valeurs de la fonction de répartition de la loi normale centrée réduite pour évaluer le risque de se tromper dans une décision basée sur un intervalle de différences.

Exemples pour n petit

- **n = 5 ; k = 3 ; N = 50** ; d_e suit $N(10\ ;\ 4\sqrt{3})$. L'intervalle de décision des différences conduisant à choisir la composition 3V-2R est]0 ; 20[.

La probabilité que d_e soit négative (plus de rouges que de vertes) est égale à $p\left(U < -\frac{10}{4\sqrt{3}}\right)$ où $U = \frac{d_e - 10}{4\sqrt{3}}$ suit la loi normale centrée réduite[119] ; on trouve 0 ,075. La probabilité que d_e soit plus que 20 est de même égale à $p\left(U > \frac{20-10}{4\sqrt{3}}\right)$; par symétrie on trouve également 0 ,075. Le risque de se tromper est d'environ 15 %.

4. Le maximum de vraisemblance

Le choix de la composition de la bouteille relève d'un problème plus général appelé estimation d'un paramètre d'une population à partir d'un échantillon aléatoire simple issu de cette population. Ce paramètre peut être ici soit la proportion d'une des couleurs, soit le rapport d'une couleur à l'autre, soit la différence entre les couleurs. Parmi les multiples méthodes probabilistes développées en statistique inférentielle, celle dite du maximum de vraisemblance a pour principe de choisir la valeur du paramètre qui a la plus forte probabilité de provoquer l'apparition de la valeur empirique. Elle fait donc appel aux probabilités conditionnelles calculables ici dans un univers fini d'événements[120].

Nous la présentons en prenant pour paramètre la proportion de boules vertes (que nous avons appelé fréquence théorique et noté f) et à partir de l'exemple déjà traité où n = 5, N = 40.

---

[119] Pour une meilleure précision il faudrait effectuer une correction de continuité puisqu'on remplace une loi discrète par une loi continue.

[120] Enseignées en France mais pas au Vietnam.

Soit un échantillon de 27 vertes et 13 rouges donc une proportion empirique de $\frac{27}{40}$ sise entre $\frac{3}{5}$ et $\frac{4}{5}$: on compare les probabilités conditionnelles $p\left(f_e = \frac{27}{40} \middle/ f = \frac{3}{5}\right)$ et $p\left(f_e = \frac{27}{40} \middle/ f = \frac{4}{5}\right)$. La première vaut $\binom{40}{27}.0{,}6^{27}.0{,}4^{13}$ et la seconde vaut $\binom{40}{27}.0{,}8^{27}.0{,}2^{13}$. La première étant plus forte que la seconde, la composition la plus vraisemblable est 3V-2R. L'inégalité serait le même pour 28 boules vertes (12 rouges) mais serait inversée pour 29 boules vertes (11 rouges)[121]. Dans ce dernier cas (n = 5, N = 40, 29 vertes) il est à noter que la méthode de l'intervalle de décision des fréquences fournirait aussi la composition 4V-1R (puisque $\frac{27}{40} = 0{,}725 > 0{,}7$) tandis que celle de l'intervalle de décision des rapports fournirait la composition 3V-2R puisque $\frac{29}{11} \in J_3$.

Degrés de certitude

Comment vont-ils fonder leur degré de certitude sachant qu'ils ont affaire à des aléas ? L'étude précédente met en évidence que deux conjectures distinctes peuvent fonder ce degré de certitude après un certain nombre de tirages :

- l'appréciation quantitative de la distance de la fréquence empirique à la fréquence théorique la plus proche ou du rapport empirique au rapport théorique le plus proche : par exemple, plus on est proche en fréquence de 0,4 plus on est sûr de la composition 2V/3R ;

- l'appréciation quantitative du nombre de tirages : plus il est grand plus on est sûr de la fréquence théorique la plus proche même si elle n'est pas « très proche » de la fréquence empirique.

La combinaison de ces deux appréciations fonde la notion de niveau de confiance attachée à un intervalle (et donc à la distance à la fréquence théorique ou au rapport théorique) et qui dépend effectivement du nombre de tirages. Pour favoriser cette combinaison il faudrait confronter des statistiques différentes sur une même bouteille et c'est ce que mettra en place la séance 2. Nous faisons cependant ici l'hypothèse que cette phase 2 de la séance 1 favorise la formulation implicite ou explicite de l'une ou l'autre de ces conjectures. Briand rapporte :

> « … les élèves sont convaincus de certaines régularités et font l'hypothèse que « plus on fera de tirages, plus on pourra affirmer des compositions de bouteilles avec certitude ».
>
> (Briand [2005], p. 264)

[121] Derrière ces calculs se profile un problème d'optimisation de la fonction dite de vraisemblance.

Cette formulation empirique de la loi des grands nombres pourrait donc faire l'objet d'une institutionnalisation par l'enseignant à la fin de la séance 1 après la réalisation publique par un groupe de la répétition des tirages de sa bouteille.

Comment les élèves vont-ils organiser les tirages pour en tirer des fréquences ou des rapports empiriques entre couleurs ? Vont-ils réaliser des séries de tirages en nombre égal au nombre de boules ou à l'un de ses multiples avec l'intention d'obtenir des rapports (fréquences) empiriques égaux aux rapports (fréquences) théoriques ? Vont-ils observer la succession des résultats des séries ou bien l'ignorer ?

Les stratégies d'organisation des tirages

Quant à l'organisation des calculs pour décider la composition, nous avons dégagé précédemment deux stratégies :

- celle de la proportionnalité entre les effectifs de l'échantillon et ceux la bouteille

- celle de la distance entre fréquences théoriques et fréquences empiriques ou bien entre rapports théoriques et rapports empiriques ou bien encore entre différences empiriques et différences théoriques.

Intéressons-nous maintenant à l'organisation des tirages, en particulier au nombre de tirages à effectuer.

a) Stratégie « *nombre fixe de tirages* »

On arrête les tirages sur un nombre que l'on juge a priori suffisant, indépendamment des résultats produits par les tirages.

- faire des calculs sur l'ensemble des tirages : effectifs / fréquences / rapport des effectifs / rapport des fréquences, …

- calculer le rapport / les fréquences obtenu(es) avec les rapports / fréquences théoriques pour choisir celui le plus proche du rapport / de la fréquence obtenu(e)

- ou bien procéder par proportionnalité

- proposer la composition correspondant au rapport théorique choisi.

- Sous-stratégie « *sans séries* » : on ne fait les calculs que sur le nombre total de tirages.

- Sous-stratégie « *avec séries* » : on effectue des relevés partiels mais leurs résultats n'interviennent pas dans la décision finale.

Un exemple

+ Décider de réaliser 30 expériences et obtenir 17 rouges et 13 vertes.

+ Le rapport des effectifs/fréquences des boules rouges et boules vertes est 17/13 qui vaut environ 1,31.

+ Comparer les distances entre ce rapport et les rapports théoriques :

Composition possibles	Rapports théoriques R/V	Rapport obtenu	Distances
1R-4V	0,25	1,31	1,06
2R-3V	0,66	1,31	0,65
3R-2V	1,5	1,31	**0,19**
4R-1V	4	1,31	2,69

Tableau 76. Comparaison des distances entre les rapports des effectifs/fréquences

avec les rapports théoriques

+ Comme la distance la plus petite est 0,19, choisir le rapport théorique 1,5 qui correspond à la composition 3R-2V.

Ou bien, le nombre des boules rouges dans la bouteille est l'entier le plus proche de 17 x 5 /30 soit 3 donc la composition de la bouteille est 3R-2V.

Cette stratégie, avec ou sans séries, ignore la fluctuation des séries s'il y en a. Que se passe-t-il si on obtient un rapport tel que 1/1 qui ne permet pas aux procédures calculs de décider de la composition ? La loi pragmatique des grands nombres prend la signification suivante : « si l'échantillon est suffisamment grand, la fréquence (respectivement le rapport) observée est suffisamment proche de la fréquence (respectivement le rapport) théorique pour que la composition de la bouteille soit conforme à cette fréquence (ce rapport) théorique ».

b) Stratégie « *nombre variable de tirages* »

On réalise un nombre minimum de tirages (par exemple une série) puis on répète l'expérience (expérience par expérience ou série d'expériences par série d'expériences) en calculant les fréquences (ou rapports) cumulés de boules rouges et de boules vertes jusqu'à obtenir un rapport exact des boules rouges et boules vertes, par exemple : 1/4, 2/3, 3/2, 4/1 ou un rapport très proche.

La loi pragmatique des grands nombres prend une première signification : « si on n'a pas assez de tirages, on ne peut savoir »

- Sous-stratégie « *séries* » : série par série on analyse le rapport de la série ; si la « tendance » est sur un rapport, on choisit ce rapport (les séries ont une couleur).

La loi pragmatique des grands nombres prend une seconde signification : « chaque série reflète plus ou moins la population».

- Sous-stratégie « *total* » : on cumule les séries les unes après les autres et on arrête si bon rapport.

La loi pragmatique des grands nombres prend une deuxième signification : « seul le total des séries reflète la population ».

Comme le nombre total des boules est 5, alors un rapport exact n'est qu'obtenu au bout d'un nombre multiple de 5 des expériences, c'est-à-dire au bout de 5, 10, 15, 20, … expériences.

Deux exemples :

- répéter expérience par expérience en calculant les effectifs des boules de chaque couleur :

Ordinal de l'expérience	Effectifs de boule rouge	Effectifs de boule verte
1	1	0
2	2	0
3	3	0
4	4	0
5	5	0
6	5	1
7	6	1
8	6	2
9	6	3
10	7	3
11	7	4
12	7	5
13	8	5
14	9	5
15	9	6

Dès l'obtention le bon rapport : 9R/6V (au bout de 15 tirages), arrêter la répétition des tirages et proposer la composition 3R-2V.

- répéter série d'expériences par série d'expériences :

Première série (9 expériences)		
Ord	Rouge	Verte
1	X	
2	X	
3	X	
4	X	
5	X	
6		X
7	X	
8	X	
9	X	
Total 1	**8**	**1**

Deuxième série (6 expériences)		
Ord	Rouge	Verte
1	X	
2	X	
3		X
4	X	
5		X
6	X	
Total 2	**4**	**2**

S'arrêter après deux séries grâce au bon rapport 12 rouges / 3 vertes et proposer la composition 4R-1V.

I.5. Déroulement et analyse a posteriori de la deuxième phase

Pour passer à la deuxième phase, l'enseignant a posé une question :

> « Êtes-vous sûr de vos annonces sur le contenu des bouteilles ? Maintenant, je vous précise que je sais que dans chaque bouteille, il y a 5 boules et qu'il y a deux couleurs : verte et rouge. C'est tout. Il n'y a pas de sable ni de pièces de monnaie.
>
> Je veux que vous essayiez de me dire combien de boules rouges et combien de boules vertes il y a dans la bouteille de votre groupe».

(Phase 1, Script de la séance 1)

Lors de l'affirmation de 5 boules de l'enseignant, le groupe 1 criait très fort comme s'il avait gagné avec son annonce : « *Il y a des boules de couleurs rouge et verte, environ 5 boules de même masse, il est possible qu'il y ait 3 rouges et 2 vertes.* » Ensuite, tous les groupes se mettent à répéter des tirages.

Tous les groupes ont fait de 20 à 50 tirages, deux d'entre eux sans séries.

Nombre de tirages	20 fois	30 fois	50 fois	
Nombre de groupe	1	3	4	Total : 8
Nombre de groupe	1 totalement sûr	1 totalement sûr	2 totalement sûr	Total : 4

Tableau 77. Nombres des tirages et déclaration d'incertitude des groupes

Les nombres totaux de tirages sont des multiples de 10 ce à quoi nous nous attendions et que nous expliquons essentiellement par la recherche d'une diminution des coûts de comptage et de calculs intermédiaires et finaux.

Distribution des stratégies d'organisation des tirages selon les groupes :

Stratégie / **Groupe**	Nombre fixe de tirages		Nombre variable de tirages		**Nombre de tirages**
	sans série	**avec séries**	**total**	**séries**	
Gr1		30 tirages (5 exps/série)			30
Gr2	30 tirages				30
Gr3		4 séries instables (21, 11, 8, 10 exps)			50
Gr4				5 séries stables (10 exps/série)	50
Gr5 (4,1)			5 séries stables (10 exps/série)		50
Gr6			2 séries instables (20 exps, 30exps)		50
Gr7	20 tirages				20
Gr8 (4,1)			2 séries instables (20exps, 10exps)		30
Total	**4**		**4**		

Tableau 78 : Distribution des groupes selon les stratégies

(Les cases grisées correspondent aux réponses « totalement sûr »)

Calculs pour la composition / Organisation tirages	Proportionnalité	Distances	Autres ou mixtes
Nombre de tirages fixe	Groupe 1	Groupe 3 Groupe 2	Groupe 7
Nombre de tirages variable	Groupe 5	Groupe 6 Groupe 8	Groupe 4

Tableau 79 : distribution croisée des groupes selon l'organisation des tirages et la réalisation des calculs pour prendre la décision

Analyses de certains groupes

- La stratégie tirages sans série à nombre fixe est celle des groupes 2 et 7.

Fiche 2bis (brouillon)

Fiche 2

Renverser la bouteille 30 fois, les probabilités d'apparition des boules :

$$\text{Rouge} = \frac{16}{30}, \text{Verte} = \frac{14}{30}$$

On voit que le nombre d'apparition des rouges est plus grande que celle des vertes

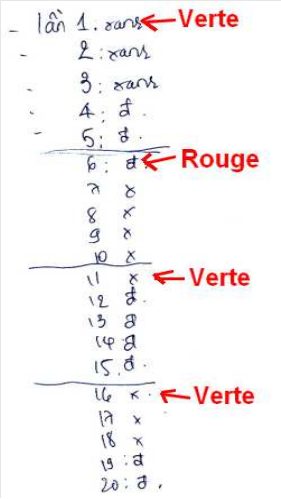
Comme il y a 5 boules dans la bouteille, alors : 3 rouge et 2 verte.

(Fiche 2 et fiche 2bis du groupe 2)

Ce groupe n'explicite pas la raison de son choix 3R-2V. On peut penser qu'il utilise le fait que le nombre de rouges est plus grand que le nombre de vertes mais pas trop, même s'il ne mentionne pas d'argument quantitatif.

- Dans la même situation, (11 vertes / 9 rouges) le groupe 7 hésite, invoque les aléas des tirages pour refuser de donner une composition puis opte en faveur de 3V/2R en prenant en compte le nombre d'apparitions de vertes en première position des séries.

Fiche 2bis (brouillon)



Fiche 2

La probabilité d'apparition d'une boule verte est plus grande que celle d'une boule rouge après 20 tirages.

La probabilité ne peut pas dire de la certitude mais possède un caractère relatif. Les résultats des tirages ne sont pas les mêmes.

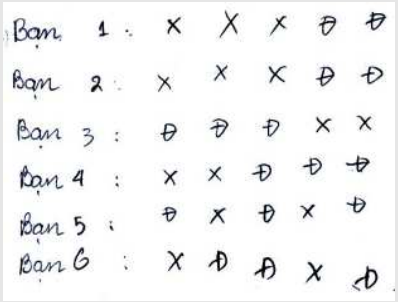
Par l'observation, on voit que la probabilité d'une boule verte est plus grande que celle d'une boule rouge. Il est impossible d'avoir le cas de 4 vertes, 1 rouge → Il y a 3 vertes, 2 rouges.

(Fiche 2 et fiche 2bis du groupe 7)

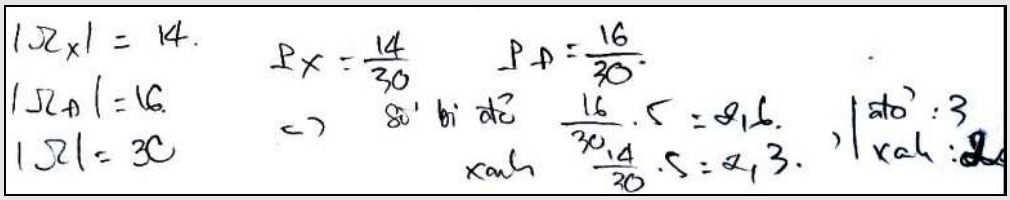
- La stratégie « nombre fixe de tirages » avec séries est empruntée par les groupes 1 et 3.

Remarquons que le groupe 1 décide de fabriquer 6 séries, une par chacun des 6 élèves, chaque élève faisant à son tour une série de 5 tirages. Ce groupe calcule des fréquences totales, qu'il appelle probabilités. Pour trouver la composition, ce groupe suit la technique de proportion pour calculer le nombre des boules rouges et celui des boules vertes.

Fiche 2bis (brouillon)



Fiche 2



Traduction en français :

$|\Omega_V| = 14$

$|\Omega_R| = 16$

$|\Omega_V| = 30$

$P_V = \frac{14}{30}$; $P_R = \frac{16}{30}$

Le nombre de boules rouges : $\frac{16}{30}.5 = 2{,}6$

Le nombre de boules vertes : $\frac{14}{30}.5 = 2{,}3$

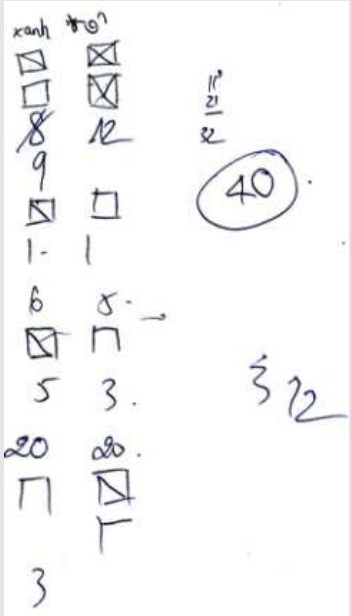
Rouge : 3

Verte : 2

(Fiche 2 et fiche 2bis du groupe 1)

- Le groupe 3 décide de faire 40 tirages en plusieurs séries mais tombe sur le rapport 1 (20R/20V). Cela provoque une nouvelle série uniquement de 10 tirages qui aboutit au résultat (7R/3V) que le groupe exploite pour annoncer la composition 3R/2V de la bouteille.

Fiche 2bis (brouillon)



Fiche 2 :

Après avoir réalisé 50 tirages, on obtient le résultat :

23 fois d'apparition des vertes

27 fois d'apparition des rouges

La différence entre le nombre d'apparition des vertes et des rouges n'est pas si grande ⇒ il existe des cas suivants : 3R–2V, 3V–2R

Comme l'apparition des rouges est plus grande que celle des vertes ⇒ Choisir 3R–2V

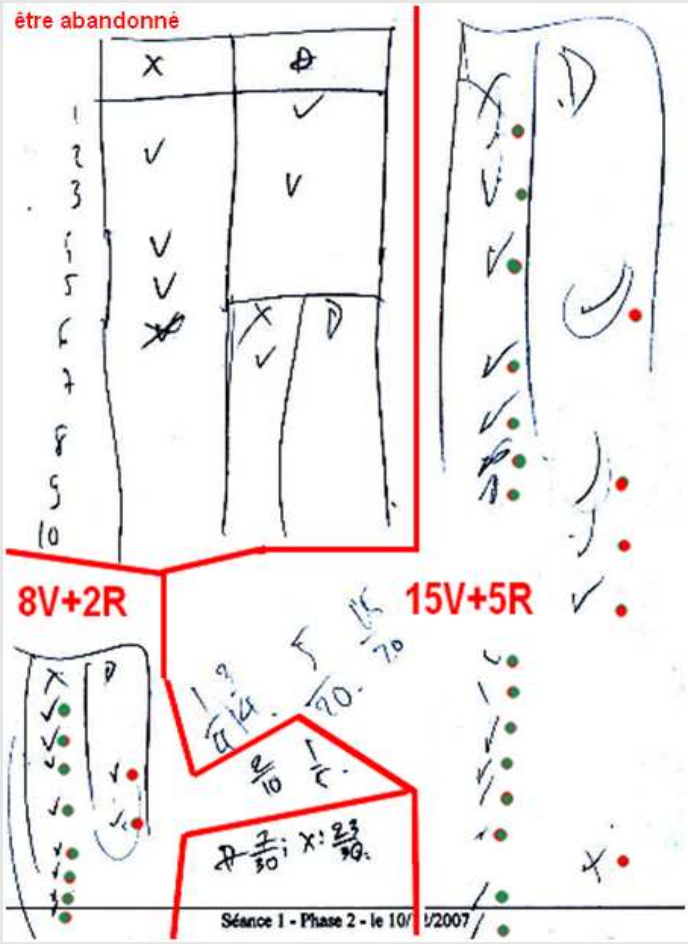
(Fiche 2 et fiche 2bis du groupe 3)

Remarquons le recours à la différence entre les nombres d'apparition pour séparer les compositions 4-1 de 3-2.

- Illustrons la stratégie « nombre variable de tirages » par le groupe 8 : après avoir abandonné une première série, le groupe réalise une série de 20 tirages en obtenant le rapport 15 vertes / 5 rouges. Les fréquences 15/20 et 5/20 sont réduites à 3/4 et 1/4 qui ne lui donnent pas une « belle » composition. Ce groupe décide de faire une autre série de 10 tirages et obtient le rapport 8 vertes / 2 rouges.

La décision de la composition de la bouteille est prise après constat de la proximité du rapport des effectifs obtenus et d'un rapport théorique.

Fiche 2bis[122] (brouillon)



Fiche 2

Après plusieurs tirages, on voit que le rapport Verte/Rouge $\approx \frac{4}{1}$

Faire 30 tirages, on voit qu'il y a 23 fois de verte et 7 fois de rouge

Fréquence de verte : $\frac{23}{30}$

Fréquence de rouge : $\frac{7}{30}$

Il y a en total 5 boules

Donc : Nombre de boules vertes ≈ 4 ; Nombre de boules rouges ≈ 1

(Fiche 2 et fiche 2bis du groupe 8)

Bien qu'il ait deux « bons rapports » dans les séries partielles et que les fréquences empiriques partielles soient dans le même intervalle de décision, le groupe additionne les effectifs des deux séries et justifie la composition 4R-1V par les fréquences totales 23/30 et 7/30. Ainsi le groupe 8 débute par la sous-stratégie « séries » et termine par la sous-stratégie « total ».

[122] La fiche est marquée en couleur par nous.

- Le groupe 4 est le seul groupe à ne pas faire apparaître de calculs de fréquences ou d'un rapport sur la totalité des tirages. A la place il fait un relevé des rapports par séries dont il note la variation. En effet ce groupe fait 5 séries de 10 tirages avec leurs rapports. Il mentionne que les « séries-rouges[123] » sont majoritaires mais hésite sur la composition de sa bouteille en refusant de choisir entre 4R-1V et 3R-2V.

Fiche 2bis (brouillon)

Fiche 2

Par 50 tirages, on trouve que le rapport d'apparition des rouges est plus grand que celui des vertes

Les rapports des tours des tirages :

1ère fois : 50/50 [124]

2è fois : 60/40

3è fois : 80/20

4è fois : 50/50

5è fois : 90/10

⇒ il est possible qu'il y ait plus ou égale à 3 rouges car il y a des fois que la boule rouge apparait plus que la boule verte.

(Fiche 2 et fiche 2bis du groupe 4)

Voici un tableau synthétique des groupes selon les trois blocs analysés dans la phase 1. Cela permet d'apprécier l'évolution de leurs comportements.

[123] Nous appelons une série-rouge si dans cette série, les boules rouges apparaissent plus nombreuses que les boules vertes.

[124] L'écriture de 50/50 devrait être lu 50% / 50%.

Groupes	Phase 1		Phase 2		Commentaire
	Composition	Nombre de tirages	Composition	Nombre de tirages	
7	2R-1V ou 1R-2V	Pas mention de la répétition	2R-3V	20	Passage à la répétition des tirages
3	R = V		3R-2V	50	Passage à une forte répétition des tirages
4	R > V		3R-2V	50	
1	3R-2V	il est possible qu'il y ait 3R et 2V	3R-2V	30	Passage à la répétition des tirages
5	R > V	10	1R-4V	50	Forte augmentation du nombre des tirages
6	1R-1V + 1V	10	2R-3V	50	
2	2R-1V	32	3R-2V	30	Conservation du nombre des tirages
8	3V-1R ; 2V-2R	30	1R-4V	30	

Tableau 80: Comparaison des nombres des tirages faites par les groupes pour deux phases

(Les cases grisées correspondent aux réponses « totalement sûr »)

Le nombre de tirages n'est pas uniforme et certains groupes qui avaient déjà pratiqué la répétition lors de la première phase ne jugent pas nécessaire d'augmenter le nombre. Parmi les groupes du premier bloc, le groupe 3 affiche sa totale certitude de la composition de la bouteille et est appelé au tableau par l'enseignante.

Evolution du groupe 3

Phase 1

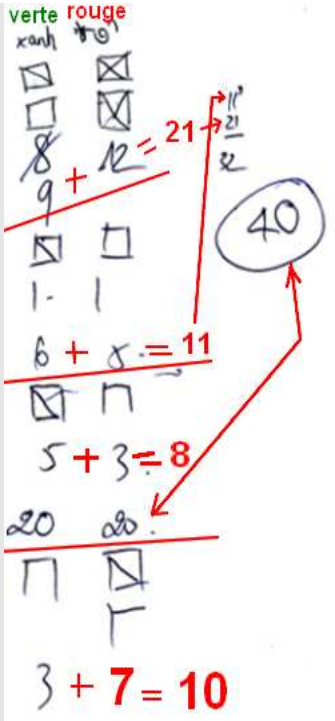
Le groupe 3 ne mentionne pas une répétition des tirages pour annoncer ses découvertes sur le contenu de la boule. Sa réponse à la fin de cette phase est :

> Dans la bouteille, il y a des boules qui sont des rouges et des vertes
>
> Le nombre des boules est supérieur à 2
>
> Le nombre des rouges est égal à celui des vertes.

(Fiche 1 du groupe 3)

Phase 2

Ce groupe décide de faire 40 tirages et le résultat à l'issue de ces 40 tirages, qu'il note sur la feuille, est 20 vertes contre 20 rouges. Confronté comme en phase 1 à l'égalité du nombre de vertes et de rouges, le groupe augmente le nombre des tirages et fait confiance à la dernière série pour départager les compositions 3V-2R et 3R-2V. Il obtient 27 rouges contre 23 vertes. Voici ses fiches :



(Fiche 2bis du groupe 3)

Après avoir réalisé 50 tirages, on obtient le résultat :

23 fois d'apparition des vertes

27 fois d'apparition des rouges

Comme la différence entre le nombre d'apparition des boules vertes et des boules rouges n'est pas si grande ⇒ il existe des cas suivants : 3R–2V, 3V–2R

Comme l'apparition des rouges est plus grande que celle des vertes

⇒ Choisir le cas 3R–2V

(Fiche 2 du groupe 3)

Tout se passe comme si la différence de 4 boules entre vertes et rouges était significative après 50 tirages alors que celle de 3 boules constatée sur les 21 premiers tirages ne l'était pas. Il semble donc que sa « totale certitude » est fondée sur l'appréciation quantitative du nombre de tirages, jugé suffisant à partir de 50.

Le passage au tableau

L'enseignant choisit ce groupe pour expliquer aux autres élèves comment ils sont arrivés à ce degré de certitude ; elle les invite à renouveler publiquement l'expérience statistique pour la comparer à leur expérience privée.

L'élève représentant du groupe 3 propose d'abord de faire 10 tirages. La proportion des boules de deux couleurs à la fin de cette scène est 1:1 (5 rouges contre 5 vertes).

Pour convaincre toute la classe, le groupe 3 propose de refaire 10 tirages de plus. Résultat, à la fin de 20 tirages en total, il y a 10 rouges contre 10 vertes :

Sans abandonner, il répète encore 10 tirages et surprise, obtient 15 rouges contre 15 vertes.

Malgré les rires et les contestations de la classe, l'élève représentant le groupe continue à soutenir sa proposition de la composition de bouteille : 3R-2V.

Question de l'enseignante :

> « Que penses-tu faire pour convaincre toute la classe d'accepter la proposition de la bouteille de ton groupe ? ».

Réponse de l'élève :

> « Il faut faire beaucoup plus des tirages, par exemple environ 1000 tirages ».
>
> (Phase 2, Script de la séance 1)

La fuite vers un « très grand nombre de tirages » jugé seul capable de justifier la composition retenue est l'indice de la tension entre le point de vue déterministe et le point de vue probabiliste. L'idée *« plus on fera de tirages, plus on pourra affirmer des compositions de bouteilles avec certitude »* est certes présente chez l'élève mais elle ne n'est pas portée par une appréciation en probabilité des variations d'échantillonnage.

À la fin de phase 2, les degrés de certitude déclarés (« tout à fait sûr » sur la première colonne et « moyennement sûr » sur la deuxième colonne) sont résumés dans le tableau suivant duquel ne se dégage aucun lien entre le degré de certitude et la composition annoncée :

Nhóm N°	1 Đỏ - 4 Xanh		2 Đỏ - 3 Xanh		3 Đỏ - 2 Xanh		4 Đỏ - 1 Xanh	
	Rất chắc	*Tương đối chắc*	*Rất chắc*	*Tương đối chắc*	*Rất chắc*	*Tương đối chắc*	*Rất chắc*	*Tương đối chắc*
1						X		
2						X		
3					X			
4						X		
5								X
6			X					
7			X					
8	X							

Tableau 81 : Distribution des réponses des groupes selon 4 compositions avec degré de certitude

En particulier nous n'avons rencontré aucun indice d'appréciation de l'incertitude selon la composition alors que selon Briand, les élèves ont fait des déclarations à propos des intervalles de décision :

« Il est plus facile de déclarer la composition 4R-1V puisque l'intervalle [0,7 ; 1] est plus grand que l'intervalle [0,5 ; 0,7] qui permettrait de décider 3V-2R. »

I.6. Conclusion

La loi des grands nombres en acte dans les réponses des élèves s'articule bien autour d'une appréciation du nombre de tirages et de la proximité entre la fréquence empirique et la fréquence théorique mais cette appréciation est globalisante et non probabiliste. Globalisante car ne se dégage pas l'idée que l'on peut relier le degré de certitude à un seul des facteurs (le nombre de tirages ou l'écart des fréquences) en fixant l'autre. Non probabiliste car le degré de certitude (ici moyennement ou totalement sûr) s'exprime généralement sans recourir aux ostensifs langagiers ou symboliques des probabilités.

Le groupe 5 fait exception : « Comme c'est une expérience aléatoire discrète et il existe d'autres conditions objectives qui peuvent affecter l'expérience, donc la probabilité est correcte jusqu'à environ 95% et non 100%. ». Cette déclaration peut s'interpréter comme l'affirmation d'une probabilité égale à 0,95 mise sur la validité de la décision de la composition 4R-1V. C'est le seul groupe qui utilise le langage des probabilités pour mesurer l'incertitude.

La mobilisation pragmatique de cette loi passe par la constitution de séries agrégées sans évaluation de leur variation. La fluctuation d'échantillonnage est un phénomène reçu comme normal mais qui ne peut pas participer à l'interprétation des aléas donc sans signification pour se forger un degré de certitude. Seul compte au final le jugement porté sur la proximité du rapport empirique (ou de la fréquence empirique) avec le rapport théorique (ou la fréquence théorique) pour l'ensemble des tirages.

Ainsi les groupes 5 et 8 qui n'ont pas une grande fluctuation d'échantillonnage autour de la composition 4-1 ne s'appuient pas sur cette faible fluctuation pour justifier leur degré de certitude dans la composition de la bouteille : le groupe 5 se dit moyennement sûr avec une fréquence empirique de 0,8 sur 50 tirages et le groupe 8 se dit totalement sûr avec une fréquence empirique de 0,77 sur 30 tirages.

La séance 2 se donne pour objectif de travailler les significations probabilistes de la fluctuation d'échantillonnage.

II. Séance 2

II.1. Introduction

Objectifs de la séance : notion d'expérience statistique, mise à l'épreuve expérimentale d'une loi « pragmatique » des grands nombres, notion de fluctuation d'échantillonnage, notion de risque dans une décision.

La connaissance ou non de la composition de la bouteille est une variable didactique qui change de valeur de la séance 1 à la séance 2 et à l'intérieur de la séance 2. En effet, dans la séance 1, la composition étant inconnue, les stratégies visent essentiellement la connaissance de cette composition et s'appuient sur une loi des grands nombres en acte. Dans la première phase de la séance 2, la composition étant connue, on met à l'épreuve pragmatiquement cette loi des grands nombres. Dans la deuxième phase de la séance 2, en rendant de nouveau inconnue la composition de la bouteille, on place les élèves dans une situation semblable à celle de la séance 1 mais avec des savoirs nouveaux sur la loi des grands nombres et sur la notion d'incertitude qui viennent d'être institutionnalisés. La similitude entre les deux situations porte sur l'enjeu de décider de la composition inconnue mais la machine de hasard est différente : elle est multiple et ses rétroactions sont quantitatives dans la séance 2 alors qu'elle est unique et ses rétroactions sont qualitatives dans la séance 1.

Matériel :

La classe dispose d'ordinateurs (dont un public) sur lesquels sont installés deux programmes.

- un programme « bouteille connue » : l'utilisateur entre la composition de son choix et il peut demander 25, 100 ou 500 tirages ; le programme propose alors les résultats de ces tirages : effectif et fréquence des deux couleurs. Plus précisément :

Nous avons écrit le programme d'un logiciel qui permet :

+ d'entrer la composition de 5 boules de deux couleurs rouges et vertes en entrant le nombre des boules rouges,

```
Turbo Pascal

        * * * * * * * * * * *      GROUPE N°1      * * * * * * * * * * * *

La bouteille contient 5 boules rouges et vertes, choisissez le nombre des boules

rouges en entrant un nombre entier de 1 à 4
                          Le nombre des boules ROUGE  :
```

+ de demander 25 ou 100 ou 500 tirages, seulement ces trois possibilités, sans limiter le nombre de demandes,

```
Turbo Pascal

        * * * * * * * * * * *      GROUPE N°1      * * * * * * * * * * * *

Selon votre choix, la bouteille contient 3 boule(s) ROUGE et 2 boule(s) VERTE
Combien de fois voulez-vous "interroger" la bouteille ?
Choisissez un parmi trois nombres de tirages suivants :  25      100     500
puis tapez le nombre choisi :
```

+ de donner des résultats sous la forme d'un tableau de distribution des effectifs et des fréquences des boules de chaque couleur.

```
Turbo Pascal

Selon votre choix, la bouteille contient 3 boule(s) ROUGE et 2 boule(s) VERTE
Combien de fois voulez-vous "interroger" la bouteille ?
Choisissez un parmi trois nombres de tirages suivants :  25      100     500
puis tapez le nombre choisi : 25

Vous avez "interrogé" la bouteille 25 fois, vous obtenez le tableau suivant :
                    Boule              ROUGE          VERT
                    ------------------------------------------
                    Effectif             16             9
                    Fréquence         0.6400        0.3600

Voulez-vous "interroger" la bouteille encore ?
Tapez lettre O pour Oui ou N pour Non : O
```

```
Turbo Pascal

Selon votre choix, la bouteille contient 3 boule(s) ROUGE et 2 boule(s) VERTE
Combien de fois voulez-vous "interroger" la bouteille ?
Choisissez un parmi trois nombres de tirages suivants :  25      100     500
puis tapez le nombre choisi : 100

Vous avez "interrogé" la bouteille 100 fois, vous obtenez le tableau suivant :
                    Boule              ROUGE          VERT
                    ------------------------------------------
                    Effectif             50             50
                    Fréquence         0.5000         0.5000

Voulez-vous "interroger" la bouteille encore ?
Tapez lettre O pour Oui ou N pour Non :
```

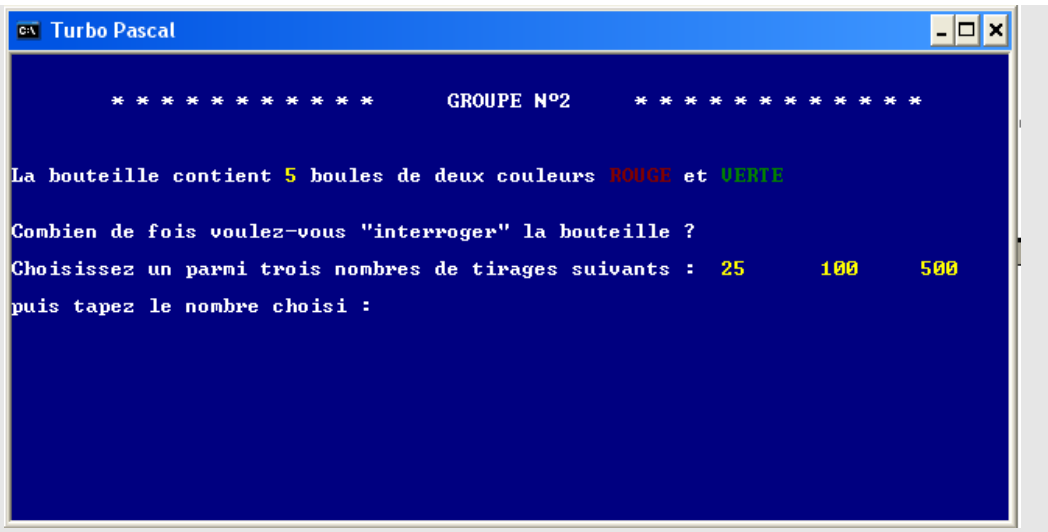
```
Turbo Pascal

Selon votre choix, la bouteille contient 3 boule(s) ROUGE et 2 boule(s) VERTE
Combien de fois voulez-vous "interroger" la bouteille ?
Choisissez un parmi trois nombres de tirages suivants :  25      100     500
puis tapez le nombre choisi : 500

Vous avez "interrogé" la bouteille 500 fois, vous obtenez le tableau suivant :
                    Boule              ROUGE          VERT
                    ------------------------------------------
                    Effectif            279            221
                    Fréquence         0.5580         0.4420

Voulez-vous "interroger" la bouteille encore ?
Tapez lettre O pour Oui ou N pour Non : N
```

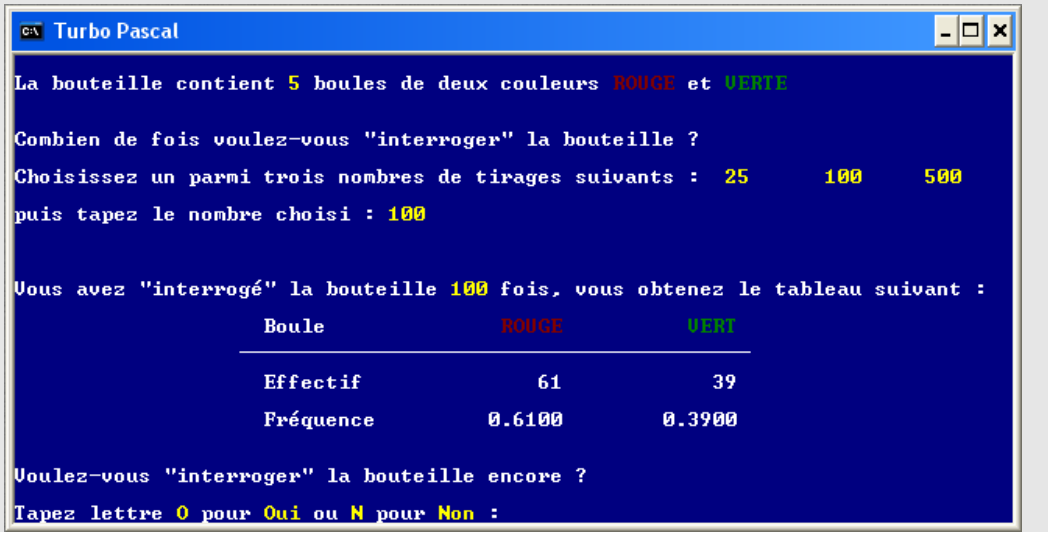
+ de récupérer, sous la forme d'un fichier et grâce à une fonction cachée, l'historique de toutes les traces du travail de l'élève et les fréquences totales de boule de chaque couleur

```
BC124625 - Bloc-notes
Fichier  Edition  Format  Affichage  ?
Séance2 : Bouteille_CONNUE
Fiche des traces du Groupe_1
Heure: 12:46:25
Date :  1/ 5/2009

NOMBRE DE BOULES ROUGES :
  3

NOMBRE DE TIRAGE                    ROUGE                          VERT
      25                   Eff =   16 ; Freq =   0.640      Eff =    9 ; Freq =   0.360
     100                   Eff =   50 ; Freq =   0.500      Eff =   50 ; Freq =   0.500
     500                   Eff =  279 ; Freq =   0.558      Eff =  221 ; Freq =   0.442

EFFECTIF TOTAL  =  625
Eff_ROUGE total =  345 ; Freq_ROUGE total =   0.552
Eff_VERT  total =  280 ; Freq_VERT  total =   0.448

Ln 1, Col 1
```

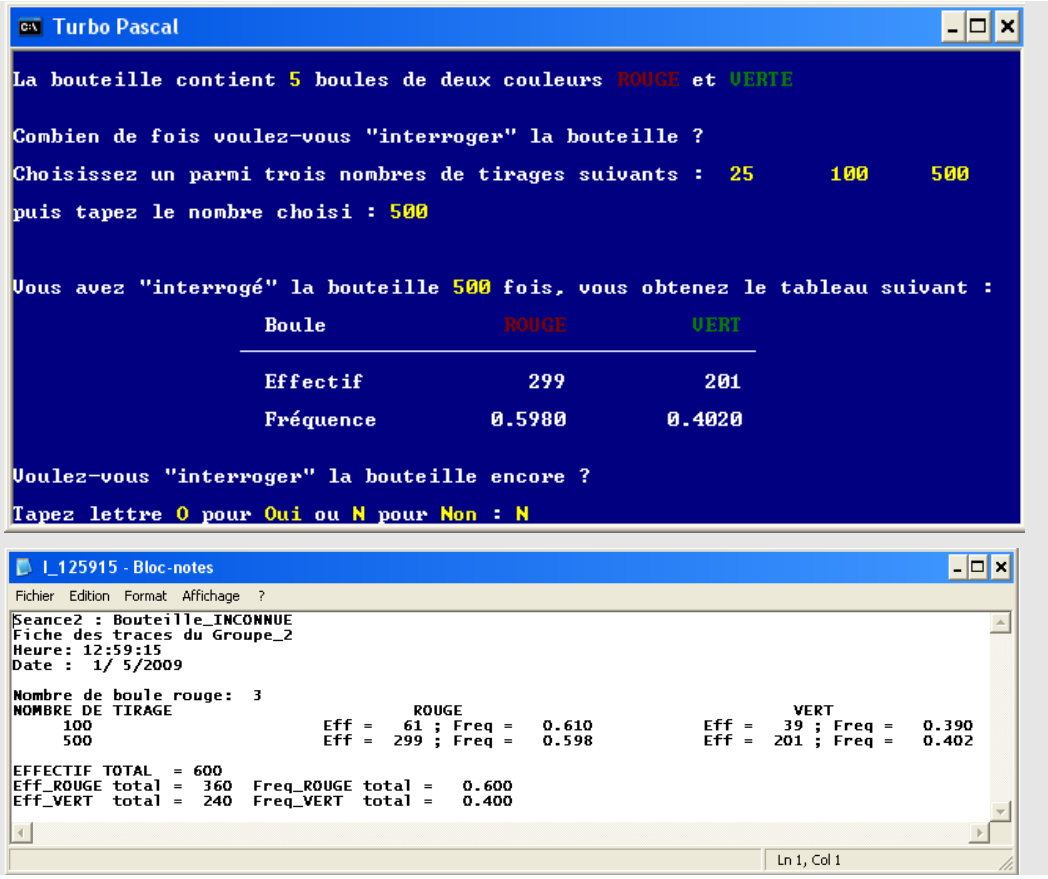
Figure 10. Illustration du fichier de trace

- un programme « bouteille inconnue » : la composition de bouteille a été pré remplie (3R-2V et 4R-1V) et cachée, et on ne peut demander que 50 tirages[125] ; le programme propose alors les résultats de ces tirages : effectif et fréquence des deux couleurs. Les ordinateurs n'ont pas tous la même composition.





[125] Du fait d'une erreur de programmation, les élèves auront en fait le choix entre 25, 100 et 500 tirages. Nous reviendrons dans l'analyse de la deuxième phase sur les conséquences de cette modification du milieu.



II.2. La première phase (25 minutes)

Objectif de la phase : réalisation d'expériences statistiques, mise à l'épreuve expérimentale d'une loi « pragmatique » des grands nombres, notion de fluctuation d'échantillonnage.

Le premier temps

Protocole

L'enseignant fait d'abord un résumé de ce qui s'est passé à la fin de la première séance. Il donne le statut de conjecture à la déclaration suivante : « plus on fera de tirages, plus on pourra affirmer des compositions de bouteilles avec certitude ». Il annonce qu'il s'intéresse à la validité de cette conjecture et que pour faire plus de tirages et plus vite on va faire appel à l'ordinateur et à ses programmes « bouteille ».

L'enseignant propose alors un travail avec le programme « bouteille connue ».

> Vous pouvez entrer une composition de bouteille. Le programme de l'ordinateur peut interroger la bouteille 25, 100 ou 500 fois : on dira qu'il fait l'expérience de 25, 100 ou 500 tirages.
>
> Ce programme calcule alors des résultats (des statistiques) de cette expérience qui sont l'effectif et la fréquence des deux couleurs (les tableaux de distribution des effectifs et des fréquences).

Demandez des expériences à l'ordinateur et écrivez vos remarques sur la fiche 3.

Il distribue la fiche 3 :

Fiche 3 (noms, prénoms, lycée, classe, n° groupe)

Vos demandes de tirages et leurs résultats :

La classe est organisée selon les mêmes groupes que dans la séance 1, sans leur bouteille. Chaque groupe dispose d'un ordinateur avec le programme « bouteille connue ».

Au bout de 10 minutes durant lesquelles les élèves se familiarisent avec le programme, l'enseignant ramasse toutes les fiches 3 puis organise une synthèse des remarques en introduisant la notion de *fluctuation des fréquences* : des expériences semblables[126] (même machine de hasard et même nombre de tirages) donnent des distributions de fréquences différentes.

Analyse a priori

Le dispositif « bouteille-ordinateur » permet de réaliser effectivement un grand nombre de tirages et affiche les fréquences empiriques sans calcul de la part de l'utilisateur. Par contre les rapports empiriques nécessitent des calculs. Le dispositif favorise donc fortement la comparaison des fréquences empiriques avec les fréquences théoriques qui sont déterminées par la composition qui est connue, plutôt que la comparaison des rapports qui n'apparaissent pas à l'écran.

La restriction à seulement 3 possibilités (25, 100, 500) des nombres de tirages demandés à la machine permet d'envisager sur toute la classe un ensemble d'expériences statistiques de même nombre de tirages et donc la comparaison des fréquences dans cet ensemble d'expériences. On peut espérer constater sur l'ensemble des 25 tirages une plus forte fluctuation des fréquences empiriques que sur l'ensemble des 500 tirages et faire ainsi émerger un lien entre le nombre de tirages et la possibilité, faible pour 25 tirages et forte pour 500 tirages, d'avoir une fréquence qui soit conforme à la fréquence théorique attendue, la conformité se décrétant par exemple par l'appartenance à l'intervalle de décision associé à la fréquence théorique (cf. étude pour la séance 1).

Comment les élèves vont-ils organiser les expériences statistiques, comment vont-ils présenter les données recueillies ? Comment l'objectif annoncé par l'enseignant d'étudier la conjecture « plus on fera de tirages, plus on pourra affirmer des compositions de bouteilles avec certitude » guidera-t-il les élèves dans la conduite des expériences statistiques et des recueils des données fournies par l'ordinateur ?

On peut s'attendre à ce que les élèves récupèrent les résultats fournis par l'ordinateur et les organisent en tableaux d'effectifs ou de fréquences. Le recueil de données est un type de tâche apparu en classe 7 et classe 10 au Vietnam (cf. analyse institutionnelle).

[126] L'expression consacrée est plutôt « expériences identiques et indépendantes ».

Les différentes organisations des statistiques et les différents recueils de données

Après le choix d'une composition, les différentes expériences peuvent être organisées de façon « cumulative » ou « non-cumulative ».

• Organisation « *non cumulative* » : réaliser une suite plus ou moins aléatoire des nombres des tirages choisis parmi 25, 100, 500 et noter la suite des effectifs, des fréquences de boule rouge ou celles de boule verte.

- Organisation « *suite consécutive* » : faire des suites dans l'ordre 25, 100, 500. Cette organisation si elle est suivie d'un recueil des fréquences, permet de constater plusieurs fois de suite le rapprochement de la fréquence empirique et de la fréquence théorique quand on passe de 25 à 500 tirages (la stabilisation des fréquences empiriques). Cette organisation est un observable d'une stratégie de vérification de la stabilisation croissante des fréquences avec l'augmentation du nombre des tirages.

Exemple :

Ordre des interrogations
25 tirages
100 tirages
500 tirages
25 tirages
100 tirages
500 tirages
…

- Organisation « *suite constante* (ou suite constante partielle) » : choisir un seul des nombres des tirages parmi 25, 100 et 500 puis demander plusieurs fois ce nombre des tirages. Cette organisation est un observable de l'intention de mettre en avant la fluctuation d'échantillonnage (des effectifs, des fréquences et des rapports) pour un même nombre de tirages.

Un exemple : une suite constante de 500 tirages

Ordre des interrogations
500 tirages
500 tirages
500 tirages
….

et une suite comprenant des sous-suites constantes

Ordre des interrogations
25 tirages
25 tirages
25 tirages
100 tirages
100 tirages
500 tirages
500 tirages
500 tirages
…

• Organisation « *cumulative* » : réaliser un certain nombre de fois des expériences en choisissant le nombre de tirages parmi les possibilités offertes (25, 100, 500) puis calculer les fréquences sur le total des tirages déjà effectués. Cette organisation permet, dans le cas des calculs cumulés des fréquences, d'observer la convergence de la suite des fréquences vers la fréquence théorique.

Exemple de calcul total *sans calculs cumulés* :

	Tirages
	25 tirages
	100 tirages
	500 tirages
	100 tirages
	500 tirages
	…
Total	(Nombre total des tirages)

Exemple de calcul total *avec calculs cumulés* :

	Tirages	Nombre de tirages cumulés
	25 tirages	25 tirages
	100 tirages	125 tirages
	500 tirages	625 tirages
	100 tirages	725 tirages
	500 tirages	1225 tirages
	…	…
Total		(Nombre total des tirages)

L'organisation « *cumulative* » oblige l'utilisateur à prendre à sa charge des calculs au-delà de ceux que lui fournit l'ordinateur. Nous faisons l'hypothèse que ce coût défavorise cette organisation des statistiques. De plus l'enseignement de la statistique au Vietnam ne comporte pas la notion de cumulation tant sur les effectifs que sur les fréquences (cf. analyse institutionnelle).

Pour chacune de ces organisations, le choix peut se porter sur une seule couleur ou sur les deux couleurs. Dans le cas des deux couleurs, cela ouvre la possibilité après calculs d'observer les variations du rapport.

A ces différentes organisations d'expériences et de recueil de données s'ajoute la possibilité d'essayer plusieurs compositions.

Déroulement et analyse a posteriori

La classe 11A1 se compose des mêmes groupes que dans la séance 1. Tous ces groupes explorent le programme sans relever les données sur la fiche. Certains le font en entrant successivement des compositions différentes et des nombres de tirages différents : nous assimilons ce moment à une phase exploratoire qui permet un apprentissage de l'instrument qu'est le logiciel. Puis les groupes écrivent sur la fiche 3 les données relativement à une unique composition sans avancer de remarque sur l'influence possible du choix de la composition.

Voici le tableau croisé de distribution des organisations selon les 8 groupes :

Organisation «*non cumulative* »	Organisation « suite consécutive »	Gr1, Gr2, Gr5
	Organisation « suite constante (partielle) »	- Suite constante : Gr8 - Suite constante partielle : Gr3, Gr6, Gr7.
Organisation « *cumulative* »		Aucun groupe
Autre		Gr4

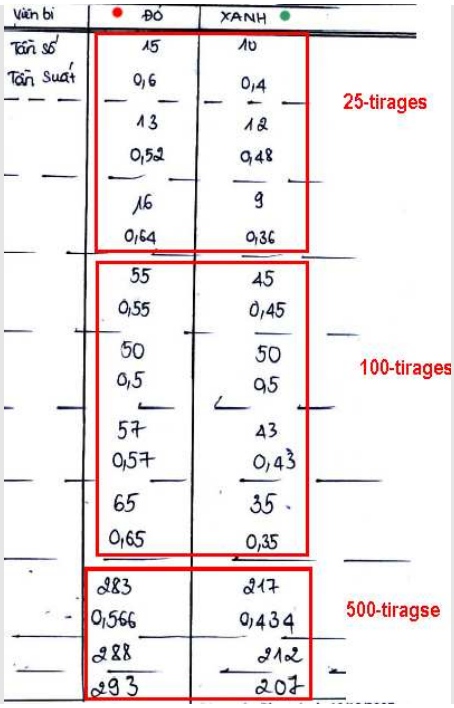
Tableau 82. Distribution des organisations des 8 groupes – Phase 1

L'organisation « *cumulative* » n'est effectivement pas apparue.

Dans cette première étape de la phase 1, les statistiques sont apparues sous forme de tableaux de distribution des effectifs et des fréquences lus sur l'écran d'ordinateur de chaque groupe. Dans les fiches 3 recueillies, la plupart des groupes (7 sur 8 groupes) relèvent une partie des résultats obtenus sur l'ordinateur sauf le groupe 4 qui ne fait pas de statistiques[127].

- La fiche du groupe 3 illustre l'organisation « *suite constante partielle* » ; le groupe fait la suite 25, 25, 25, 100, 100, 100, 100, 500, 500, 500 tirages avec la composition 3R-2V. Ses résultats pourraient lui permettre d'observer la fluctuation d'échantillonnage et de comparer les fluctuations pour 25-tirages, 100-tirages et 500-tirages mais aucune remarque n'est écrite sur la fiche.

[127] Ce groupe lit les fréquences, les multiplie par 10 puis en tire une composition fictive de la bouteille.



Viên bi	ĐỎ	XANH
Tần số	15	10
Tần suất	0,6	0,4
	13	18
	0,52	0,48
	16	9
	0,64	0,36
	55	45
	0,55	0,45
	50	50
	0,5	0,5
	57	43
	0,57	0,43
	65	35
	0,65	0,35
	283	217
	0,566	0,434
	288	212
	293	207

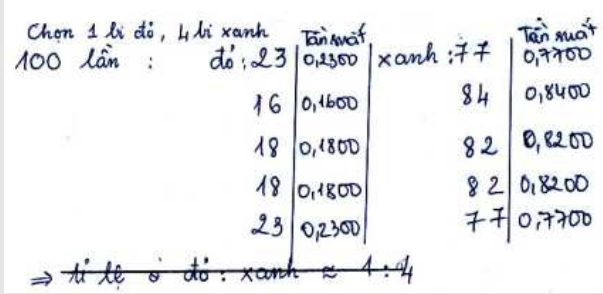
25-tirages

100-tirages

500-tiragse

(Fiche 3 du groupe 3)

- Le groupe 8 expose sur sa fiche une organisation « *suite constante* »



Chọn 1 bi đỏ, 4 bi xanh		Tần suất		Tần suất
100 lần : đỏ :	23	0,2300	xanh : 77	0,7700
	16	0,1600	84	0,8400
	18	0,1800	82	0,8200
	18	0,1800	82	0,8200
	23	0,2300	77	0,7700

⇒ tỉ lệ ở đỏ : xanh ≈ 1 : 4

Traduction en français

Choisir 1 boule rouge, 4 boules vertes

		Fréquences		Fréquences
100 fois :	Rouge :	23	0,2300	Verte : 77 0,7700
	16	0,1600	84	0,8400
	18	0,1800	82	0,8200
	18	0,1800	82	0,8200
	23	0,2300	77	0,7700

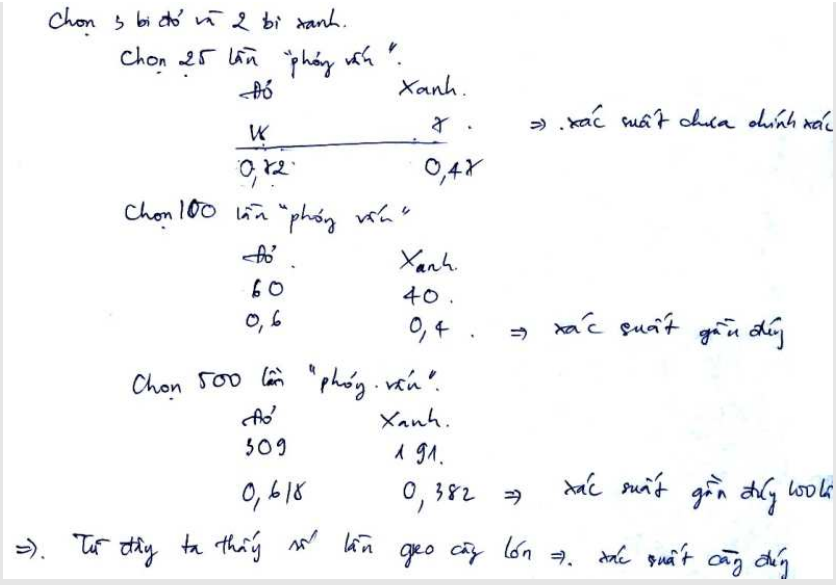
⇒ la proportion rouge : verte ≈ 1 : 4

Les fréquences obtenues de l’interrogation de l’ordinateur est convenable au résultat de l’activité 2.

(Fiche 3 du groupe 8)

C’est le seul groupe qui note une suite constante de 100 tirages (cinq fois la même expérience de la bouteille-ordinateur) avec la composition de 1R-4V. Comme le nombre des tirages est 100, il a obtenu des fréquences qui oscillent autour des fréquences théoriques ce qui aurait pu le conduire à des remarques sur la fluctuation des fréquences empiriques autour de la fréquence théorique égale à 0,2. Mais sur sa fiche seule est présente l’idée d’approximation à travers l’usage de l’ostensif ≈ dans la phrase barrée.

- Le groupe 2 expose quant à lui une organisation « *suite consécutive* ». Auparavant, durant la phase exploratoire, ce groupe a changé plusieurs fois de composition et d’expérience de manière erratique.



Traduction en français :

Choix de 3R et 2V :

Choisir 25 fois « d’interroger » la bouteille :

Rouge	Verte	
18	7	⇒ la probabilité n’est pas encore exacte
0,72	0,47	

Choisir 100 fois « d’interroger » la bouteille :

Rouge	Verte	
60	40	⇒ la probabilité est approximative
0,60	0,40	

Choisir 500 fois « d’interroger » la bouteille :

Rouge	Verte	
309	191	⇒ la probabilité est approximative comme
0,618	0,382	dans le cas de 100 fois

On voit que plus le nombre des tirages est grand plus la probabilité est précise.

(Fiche 3 du groupe 2)

Finalement ce groupe décide de noter sur la fiche les résultats d'une suite croissante des nombres des tirages (25 tirages, 100 tirages, 500 tirages) qu'ils ont réalisée en dernier comme le prouve l'historique des dernières expériences (les résultats correspondant à ce qui est noté dans la fiche 3 sont grisés ; rappelons que les calculs finaux sont faits par une fonction cachée du programme bouteille-connue et ne sont pas affichés à l'écran de l'ordinateur du groupe) :

BC163603.G2 - Bloc-notes

Fichier Edition Format Affichage ?

```
Seance2 : Bouteille_CONNUE
Fiche des traces du Groupe_2
Heure: 16:36:03
Date : 10/12/2007

NOMBRE DE BOULES ROUGES :
  3

NOMBRE DE TIRAGE                          ROUGE                                    VERT
     25                          Eff =   18 ; Freq =   0.720          Eff =    7 ; Freq =   0.280
     25                          Eff =   13 ; Freq =   0.520          Eff =   12 ; Freq =   0.480
     50  Valeur invalide
    100                          Eff =   60 ; Freq =   0.600          Eff =   40 ; Freq =   0.400
    500                          Eff =  309 ; Freq =   0.618          Eff =  191 ; Freq =   0.382

EFFECTIF TOTAL  =  650
Eff_ROUGE total =  400 ; Freq_ROUGE total =   0.615
Eff_VERT  total =  250 ; Freq_VERT  total =   0.385
```

Figure 11. Le fichier des traces du groupe 2 pour le programme « Bouteille Connue »

enregistré à 16:36:03

Les remarques pour chaque expérience proviennent de la comparaison entre des fréquences empiriques (appelées probabilités[128]) avec les fréquences théoriques qui sont respectivement 0,6 et 0,4. La dernière remarque « *plus le nombre des tirages est grand plus la probabilité est précise* » fait écho à la conjecture énoncée par l'enseignant. Mais comment ce groupe comprend-il le fait qu'à 500 tirages la fréquence empirique soit plus éloignée de la fréquence théorique qu'à 100 tirages où elle est exactement égale à la fréquence théorique ?

En résumé deux connaissances sont exprimées ici :

• La fréquence empirique est nécessairement un approché de la fréquence théorique

• La fréquence empirique s'approche avec le nombre de tirages de la fréquence théorique et pour s'en assurer il suffit d'une suite « croissante » d'expériences.

Cette dernière connaissance est une interprétation de la conjecture énoncée par l'enseignant – interprétation que nous qualifions de fonctionnelle. Nous faisons l'hypothèse qu'elle s'érige en obstacle à une lecture probabiliste de la loi des grands nombres qui énonce une *convergence en probabilité* des fréquences empiriques vers la fréquence théorique et qui est présente dans la notion de fluctuation d'échantillonnage et de risque dans la décision à prendre. Pour conduire l'observation de la fluctuation d'échantillonnage et de la réduction

[128] Le mot « probabilité » est certainement utilisé à la place du mot fréquence et reprendrait ainsi l'expression « probabilité expérimentale » enseignée en classe 11.

de l'intervalle « min-max » des fréquences empiriques, l'enseignant décide d'imposer à tous une unique composition et de rassembler de très nombreuses expériences dans un graphique.

Le deuxième temps

Protocole

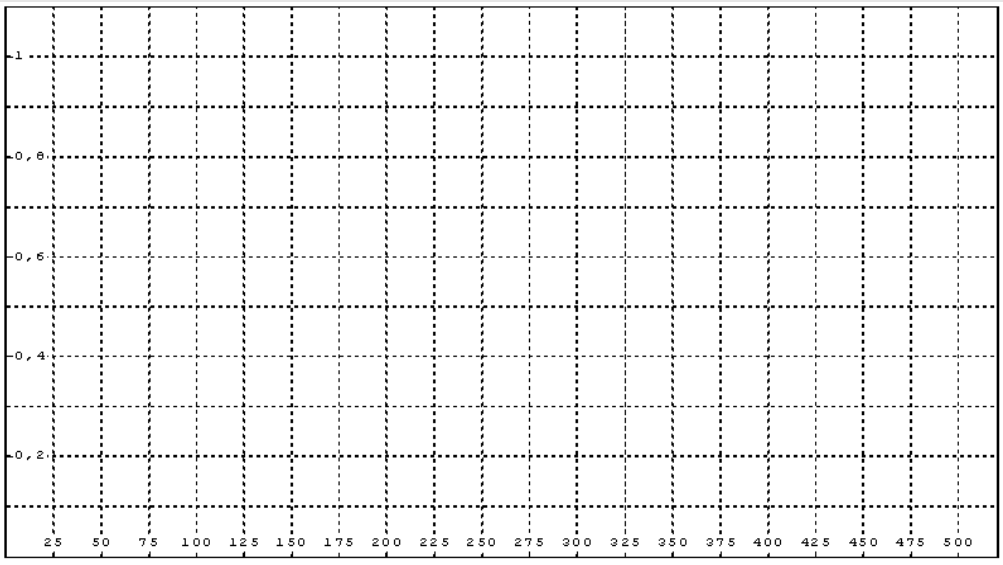
L'enseignant donne alors une nouvelle consigne :

> Je vous propose maintenant de décider entre nous d'une composition de la bouteille (la même pour toute la classe, par exemple 3R-2V) et de faire un graphique sur la fiche quadrillée (fiche 4) en mettant, pour chaque expérience statistique, en abscisse, le nombre de tirages de l'expérience et en ordonnée la fréquence des rouges calculée par l'ordinateur.

Voici la fiche 4 :

Fiche 4 (noms, prénoms, lycée, classe, n° groupe)

Le graphique (en abscisse le nombre de tirages et en ordonnée la fréquence de boules rouges)



Au bout de 10 minutes, l'enseignant ramasse les fiches 4 de tous les groupes puis regroupe[129] publiquement sur un graphique tous les points obtenus par les différents groupes.

[129] Ce regroupement de tous les points donne implicitement une réponse positive au problème suivant : les expériences des uns et des autres sont-elles « semblables ? » sans fournir de modèle.

Analyse a priori

Son intention est alors d'expliciter publiquement la réduction de la fluctuation des fréquences quand le nombre de tirages augmente.

Sur la fiche 4 les élèves doivent associer un point du repère cartésien à chaque expérience et représenter toutes leurs expériences.

Selon Briand :

> « C'est ainsi que les élèves proposent de matérialiser par une droite parallèle à « l'axe des x » ce qui, pour l'observateur, constitue un début de travail probabiliste. Le sentiment de la probabilité porte sur la sûreté de la convergence.
>
> […]
>
> À ce moment de l'expérimentation, les élèves ont donc compris l'existence de régularités, malgré les aléas de l'expérience statistique. En effectuant un « bon nombre » d'expériences, les élèves vont disposer d'un outil leur permettant de commencer à formuler les hypothèses prenant en compte le risque dans une prise de décision, en fonction du nombre de tirages. »

(Briand [2005], p. 265)

Les conditions mises en place sont-elles suffisantes pour faire émerger chez les élèves la notion de fluctuation d'échantillonnage ? Nous en doutons pour deux raisons :

• dans le temps imparti, les expériences des élèves sont peu nombreuses ; aussi l'enseignant prévoit-il de rassembler toutes les expériences semblables pour mener l'explicitation de la notion.

• l'outil graphique mis en place donne lieu à la mobilisation de connaissances fonctionnelles qui peuvent renforcer l'obstacle fonctionnel dont nous avons fait l'hypothèse ci-dessus, obstacle qui s'exprimera dans des tracés de segments joignant les différents points considérés comme des éléments de la courbe représentative d'une fonction (cf. analyse institutionnelle page …)

Les stratégies graphiques prolongent les organisations de statistiques dégagées dans le premier temps.

• Soit point par point : réaliser plusieurs expériences de même nombre de tirages et placer les points de coordonnés (nombre de tirages ; fréquence empirique) dans le repère ; cette stratégie met en valeur la fluctuation d'échantillonnage et la réduction de fluctuation d'échantillonnage selon le nombre de tirages.

Exemple :

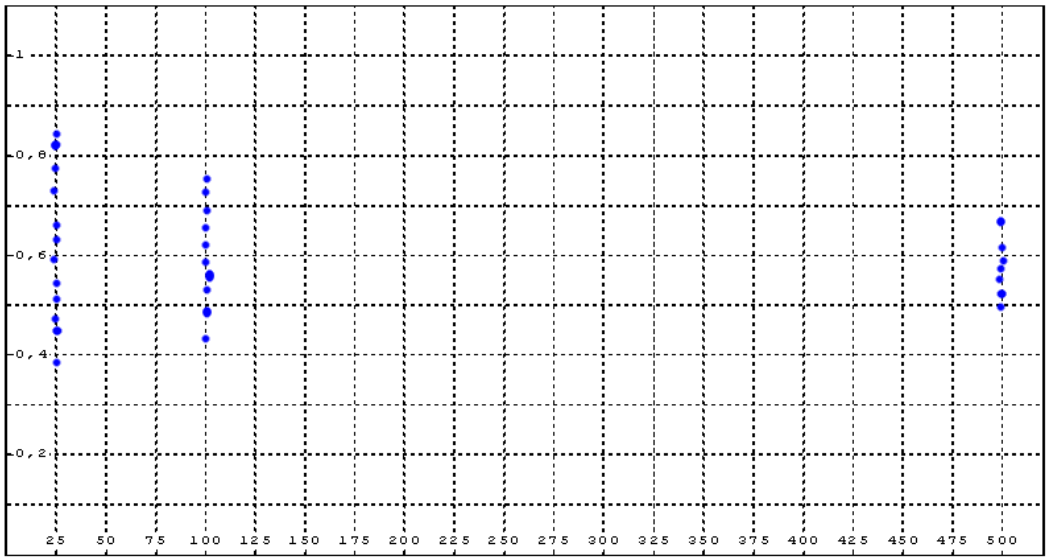


Figure 12. Exemple de réponse en fiche graphique

• soit par segments : réaliser successivement et plusieurs fois trois expériences de 25, 100 et 500 tirages, placer les 3 points de coordonnées (nombre de tirages ; fréquence empirique) puis les joindre par des segments ; cette stratégie conduit à une lecture fonctionnelle de la variation de la fréquence empirique en fonction du nombre de tirages conduisant à la convergence des fréquences empiriques vers la fréquence théorique. Cette stratégie est favorisée par le dispositif graphique où :

- l'axe des abscisses est gradué régulièrement par les nombres visibles des tirages qui sont 25, 50, 75, …, 500,

- l'axe des ordonnés est gradué avec les valeurs visibles de fréquences qui sont 0,2 ; 0,4 ; 0,6 ; 0,8 et 1

Exemple :

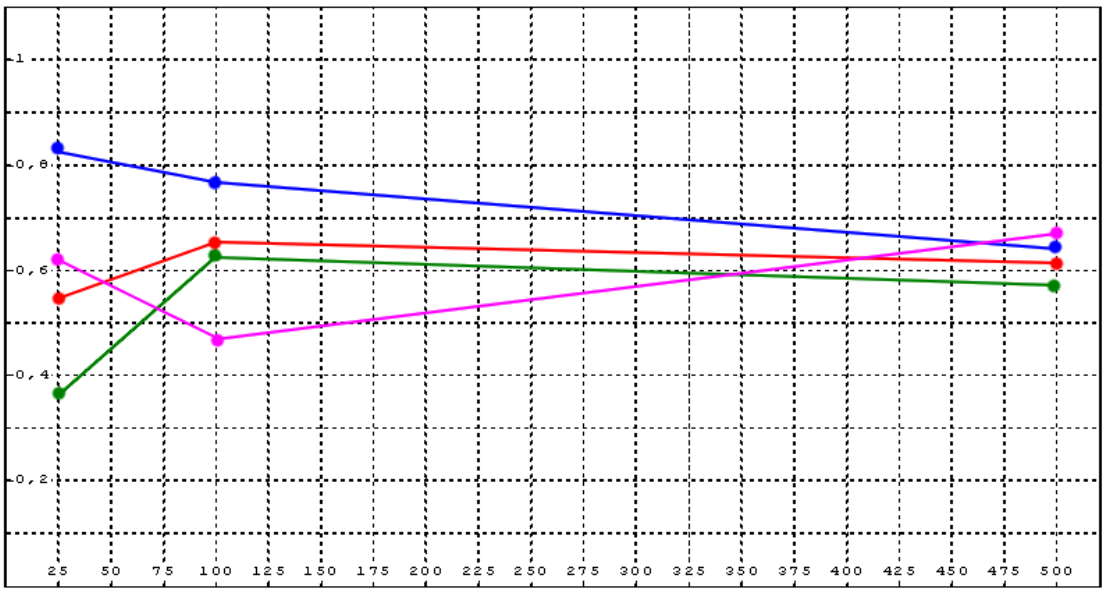


Figure 13. Exemple de réponse en fiche graphique

Déroulement et analyse a posteriori

L'enseignant propose :

> Bon, on est d'accord pour choisir 3R. Maintenant, tous les groupes refont fonctionner le même programme avec le même nombre des rouges. Tous les groupes font pareillement. Puis vous allez noter les résultats sur vos fiches 4. Après, on va faire ensemble un bilan des résultats.

(Script de la séance 2)

Rapidement un groupe l'interpelle ; ce groupe a remarqué les résultats différents issus d'expériences semblables et hésite sur la manière de répondre à la consigne :

El	Madame, chaque fois de 25 tirages, j'ai un résultat différent[130] que le précédent, qu'est-ce que je dois alors faire ?
Ens	Ah, tu fais un point pour chaque fois. (*l'enseignant parle à toute la classe*) Un groupe a fait plusieurs fois 25 tirages et obtenu des résultats différents. Il m'a demandé comment faire. Alors, je vous propose de placer autant de points différents correspondant au nombre de vos résultats.

(Script de la séance 2)

Malgré l'explication fournie par l'enseignant deux groupes (les groupes 4 et 7) ont joint les points par des segments. Ainsi le groupe 7 transforme graphiquement des statistiques élaborées dans une organisation « suite consécutive » :

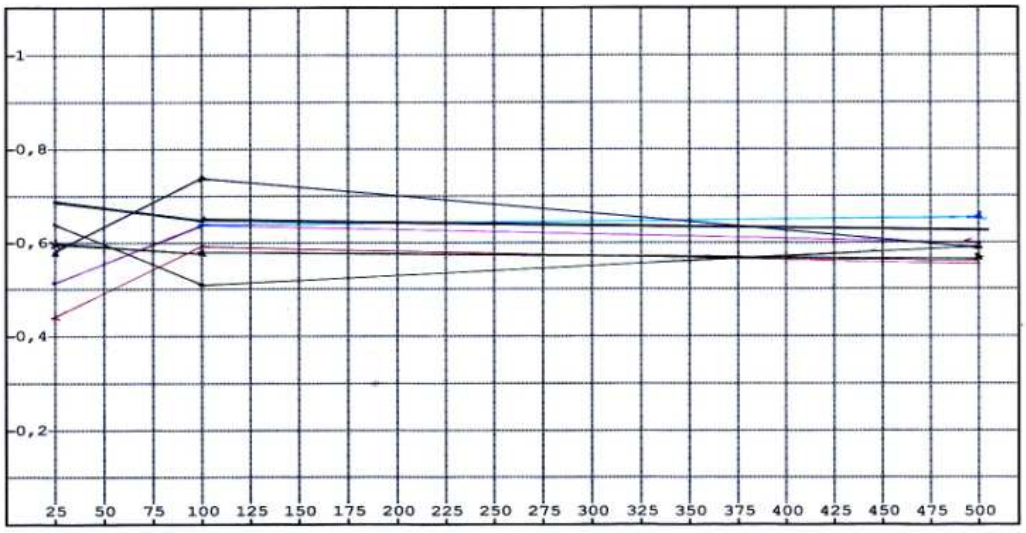


Figure 14. Fiche de réponse graphique du groupe 7

Nous reconnaissons le point de vue fonctionnel qui fait disparaître la fluctuation d'échantillonnage. Cependant un autre groupe s'essaye à une formulation de la fluctuation d'échantillonnage :

[130] Souligné par nous pour mettre en valeur la rupture de la règle du contrat relatif au graphique d'une fonction numérique qui stipule qu'il ne peut y avoir plusieurs ordonnées pour une même abscisse.

El	[…] Madame, est-ce que quand le nombre des tirages est faible, la fréquence n'est pas exacte ; et quand le nombre des tirages est plus grand, la certitude est grande aussi, n'est-ce pas ?
Ens	Vous observez et vous le trouvez ? Très bien,
El	Continuons…, ceci doit être 0,6 ; mais on a 0,63, ça veut dire il est plus 0,03 ; 25 fois, on a 0,54,…, 0,76 ; …

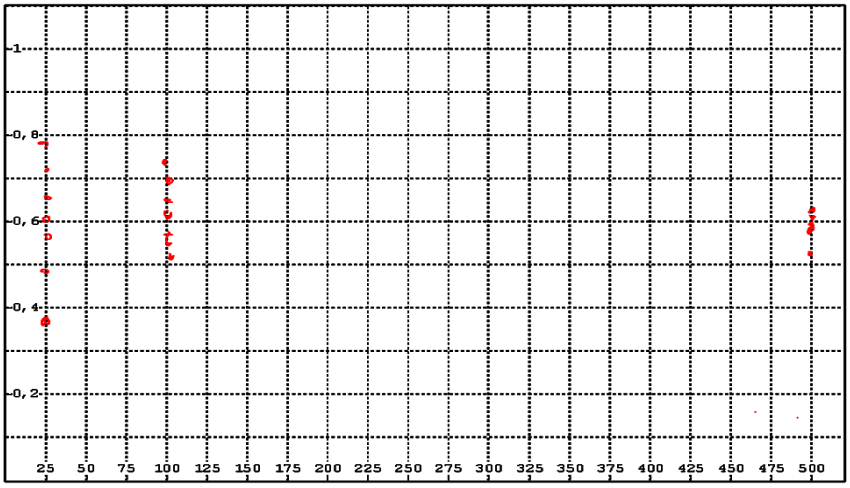
(Script de la séance 2)

Dans son intervention publique, l'enseignant va séparer les trois ensembles d'expériences, celles de 25 tirages, celles de 100 tirages et celles de 500 tirages :

Ens	Ah, d'abord, je vais vous poser la question suivante : pourquoi avec la même machine, le même programme, le même choix de 3 boules rouges, pour la première fois où on demande 25 tirages, on obtient un certain résultat, puis la deuxième demande de 25-tirages, on obtient au autre résultat ; ces résultats sont différents… Pourquoi ?
El	Parce qu'un 25-tirages n'est pas exact
Ens	Ah, pourquoi ? pourquoi un 25-tirages n'est pas exact[131] ?
El	C'est parce que le nombre de tirages est petit.
Ens	Ah, votre ami dit que c'est à cause de petit nombre des tirages. Bon, vous voyez, le groupe utilise une machine commune, fait fonctionner le même programme, avec le choix de 3 boules rouges, quand ce groupe fait 2 ou 3 fois un 25-tirages, il obtient des résultats différents, ce phénomène est appelé la fluctuation d'échantillonnage. Pour les 100-tirages, ou les 500-tirages, il y a aussi une fluctuation d'échantillonnage […]

(Script de la séance 2)

L'enseignant reprend les résultats de tous les groupes puis trace devant toute la classe les points sur une fiche. Voici cette fiche :



Séance 2 - Phase 1 - le 10/12/2007

Figure 15. Affichage des résultats des groupes

[131] L'enseignant ne s'intéresse pas aux significations possibles de cette idée d'exactitude parce qu'il vise la notion de « réduction de la fluctuation » selon le nombre de tirages.

L'affichage public pouvait permettre d'analyser empiriquement la décision selon l'expérience :

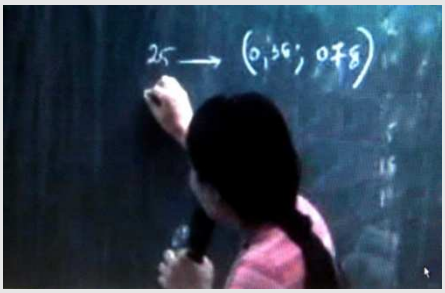
- pour 7 expériences de 25-tirages, il y 4 fréquences hors de l'intervalle de décision]0,5 ; 0,7[→ la probabilité de prendre une bonne décision est 3/7,

- pour 7 fois de 100-tirages, il y a 1 fréquences étant hors de l'intervalle de décision]0,5 ; 0,7[→ la probabilité de prendre une bonne décision est 5/7,

- pour 5 fois de 500-tirages, il n'y a aucune fréquence étant hors de l'intervalle de décision]0,5 ; 0,7[→ la probabilité de prendre une bonne décision est 5/5,

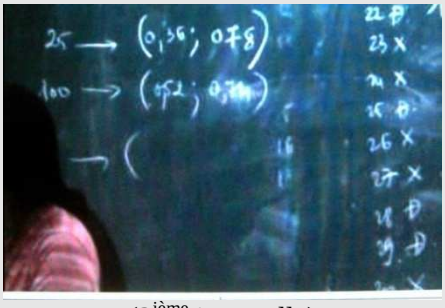
En fait, l'enseignant va s'attacher d'abord à mettre en lumière la réduction de l'amplitude des intervalles [min ; max] lors de l'augmentation du nombre des tirages :

Ens	Regardez au tableau, SVP. Si on regarde 25 fois, le groupe 1 a remarqué que cet intervalle est large, rappelons que tous les groupes ont choisi 3 boules rouges, donc, la probabilité d'une boule rouge calculée par la formule vous donne combien ?
El	soixante
Ens	0,6. Vous voyez, c'est un intervalle de 0,36 à … zéro virgule … combien ? … 0,78, je vais noter sur le tableau… Quand nous faisons un tel nombre des 25-tirages, nous avons l'intervalle]0,36 ; 0,78[;



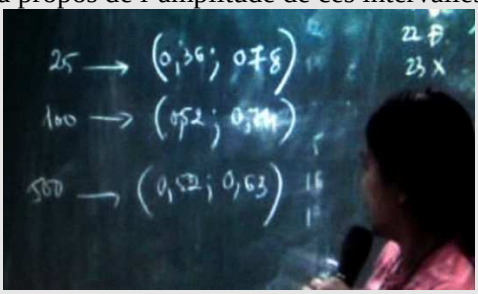
(1er intervalle)

	pour 100 fois, nous avons des points …. comment ?
El	Plus proches…
Ens	[…] Mais pour les 100-tirages, qu'est-ce que l'on a pour l'intervalle de valeur la plus basse et la plus haute ? (les fréquences sont sous-entendues ici)
El	Entre 0,52 …
Ens	Et combien ?
El	et 0,74
Ens	D'accord, 0,74.



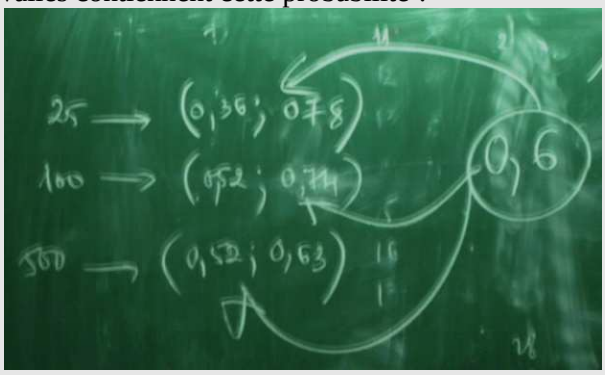
(2ième intervalle)

Ens	Puis pour des 500-tirages ?
El	0,52 et 0,63

Ens	Oui, environ 0,63 ou 0,64, n'est-ce pas ? Que trouvez-vous à propos de l'amplitude de ces intervalles ? 
	(3ième intervalle)
Els	L'amplitude est grande puis elle devient petite,
Ens	Ah, ça veut dire que … plus le nombre des tirages est grand alors… plus l'amplitude est …
Els	elle est petite…

(Script de la séance 2)

Puis il conduit l'analyse de cette réduction de l'intervalle en termes de risque pour la décision à prendre en cas de composition inconnue :

Ens	Et qu'est-ce que vous remarquez ?
Els	Plus on fait des tirages, plus il est exact …
Ens	Hm… c'est la certitude de quoi ?
El1	… plus la certitude est grande.
El2	De la probabilité[132] …
Ens	Est-ce que vous avez la probabilité ici ?
Els	Oui, 0,6…
Ens	D'accord, vous savez déjà le rapport des boules rouges et vertes. Alors, d'après la formule apprise, vous trouvez que la probabilité est 0,6. Est-ce que tous ces intervalles contiennent cette probabilité ? 
	(l'enseignant écrit au tableau les intervalles contenant la valeur 0,6)
Els	Oui,

[132] Souligné par nous.

Ens	Ah, on voit que tous ces intervalles contiennent 0,6 qui appartient à cet intervalle, à cet intervalle et à cet intervalle aussi. Et on voit que les intervalles se réduisent de plus en plus … Bon, supposons que l'on ne sache pas quelle est la valeur de la probabilité d'apparition d'une boule rouge, mais on sait qu'elle appartient à cet intervalle, est-ce que vous pouvez préciser la probabilité ? Est-il facile à donner la probabilité ? (*l'enseignant montre l'intervalle formé par des 25-tirages*),
Els	non
Ens	… à cet intervalle ? (*l'enseignant montre l'intervalle formé par des 100-tirages*),
Els	Un peu…
Ens	et à cet intervalle (*l'enseignant montre l'intervalle formé par des 500-tirages*),
Els	Plus facile …
Ens	Ah, les intervalles deviennent plus étroits… et le dernier est le plus étroit, et il semble que les valeurs à choisir dans ce dernier intervalle soient moins nombreux que dans les précédents. On voit que la probabilité se trouve dans les intervalles qui sont de plus en plus étroits. Et on voit aussi que plus on fait un grand nombre des tirages, plus la fréquence est approchée la probabilité.

(Script de la séance 2)

Cette intervention publique de l'enseignant pour introduire le concept de fluctuation d'échantillonnage insiste sur la réduction de l'intervalle [min ; max] des fréquences empiriques. L'enseignant rebondit sur la remarque d'un élève pour faire venir les notions de probabilités déjà connues de tous. Mais la probabilité à laquelle il fait référence est celle d'obtenir une boule rouge lors d'une seule interrogation de la bouteille de composition 3R-2V – ce que nous avons appelé la fréquence théorique. Et pour définir le degré de certitude dans la décision issue d'une expérience de n tirages, l'enseignant ne traduit pas les fréquences obtenues en probabilités préférant utiliser les termes qualitatifs de « facile »et « plus facile ». La notion de risque n'est donc pas institutionnalisée en termes probabilistes. Comment le concept de fluctuation d'échantillonnage ainsi institutionnalisé produira-t-il des connaissances nouvelles aux élèves pour prendre une décision avec risque lorsque la composition de la bouteille est inconnue ?

II.3. La deuxième phase (20 minutes)

Objectif de la phase : faire formuler une hypothèse sur la composition d'une bouteille de composition inconnue et le risque que cette hypothèse soit fausse.

Le protocole envisagé

L'enseignant propose un travail à partir du programme « bouteille inconnue ».

> Consigne :
>
> Dans la bouteille-ordinateur il y a cinq boules, des vertes et des rouges mais vous ignorez la composition. Il va falloir décider entre vous de la composition de votre bouteille.

Quelle hypothèse faites-vous sur la composition ? Comment testez-vous cette hypothèse ? Ecrivez sur la fiche (5) votre hypothèse et comment vous la testez.

Voici la fiche (5) :

Fiche 5 (noms, prénoms, lycée, classe, n° groupe) Votre hypothèse sur la composition de votre bouteille : Comment testez-vous cette hypothèse ?

La contrainte imposée aux élèves de ne pouvoir répéter que des expériences de 50 tirages avait pour finalité de répondre aux questions suivantes : que feront-ils d'une telle répétition des expériences ? Ajouteront-ils les effectifs de tous les tirages de toutes les expériences pour en tirer de nouvelles fréquences ? (utilisation implicite de l'indépendance des tirages).

Rappelons que pour 50 tirages (c'est ce qu'aurait du permettre le programme « bouteille inconnue ») avec la composition 3R/2V, la probabilité que la fréquence observée soit dans l'intervalle de décision associé est égale à 0,806 ce qui laisse « prévoir » *qu'en moyenne 1 expérience sur 5 aboutira à une prévision fausse*. Avec la composition 4R/1V, la probabilité que la fréquence observée soit dans l'intervalle de décision associé est égale à 0,939 ce qui laisse « prévoir » *qu'en moyenne 1 seule expérience sur 16 aboutira à une prévision fausse.*

L'enseignant conclut la séance sur la notion de décision avec risque après expériences statistiques sur une population *de composition inconnue et inaccessible autrement que par sondage*.

Analyse a priori

Lors de l'expérimentation, l'enseignant découvre que le programme est différent de celui qui était envisagé. Les élèves ont en fait le choix entre 25, 100 et 500 tirages. La possibilité de faire 500 tirages devrait être privilégiée par les élèves à qui l'enseignante vient d'expliquer la diminution du risque en augmentant le nombre de tirages. Vont-ils réaliser plusieurs expériences de 500, comparer leurs résultats ou/et cumuler les nombres de tirages des différentes expériences ?

• Stratégie « *intervalle de fluctuation des fréquences* » :

- réaliser plusieurs expériences

- puis fabriquer l'intervalle de fluctuation d'échantillonnage [f_{min} ; f_{max}],

- plus l'intervalle est étroit et centré sur une fréquence théorique, plus la certitude est grande sur la composition associée à cette fréquence théorique

Cette stratégie vise à observer la distribution des fréquences empiriques et s'appuie sur les savoirs qui viennent d'être institutionnalisés par l'enseignant : on augmente la certitude en demandant des expériences semblables à grand nombre de tirages.

Exemple :

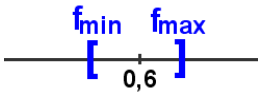


Figure 16. Intervalle centré par la fréquence théorique 0,6

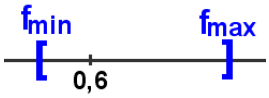


Figure 17. Intervalle « moins centré » par la fréquence théorique 0,6

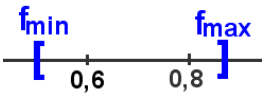


Figure 18. Intervalle contenant deux fréquences théoriques 0,6 et 0,8

Le dernier exemple montre comment les rétroactions du milieu déterminent la poursuite ou l'arrêt de cette stratégie.

Une variante de cette stratégie, où l'intervalle de fluctuation des fréquences n'est pas formé explicitement, prend la décision à partir de la stabilité des fréquences empiriques à proximité d'une fréquence théorique.

Exemple :

Une suite des expériences de 500-tirages donne une suite des fréquences empiriques qui se stabilise autour de la fréquence théorique 0,8 permet de proposer la composition 4R-1V :

0,778
0,83
0,814
0,824
0,83
0,791
0,812
0,809

Pour cette stratégie comme pour sa variante, la certitude s'installe dès que l'intervalle de fluctuation est inclus dans l'intervalle de décision. Et cette éventualité est favorisée par la possibilité de choisir des expériences de 500 tirages. On peut donc s'attendre à ce que la quantification et même l'idée de risque dans la décision soit absente des réponses des élèves.

• Stratégie « *fréquence moyenne* » :

- réaliser plusieurs expériences

- calculer la fréquence moyenne des toutes les fréquences empiriques reçues,

- comparer la fréquence moyenne avec les fréquences théoriques et choisir la fréquence théorique la plus proche de la fréquence moyenne pour proposer la composition de la bouteille-ordinateur.

Cette stratégie ignore la fluctuation d'échantillonnage mais cherche à obtenir un nombre de tirages jugé suffisamment grand pour garantir sans risque la composition choisie. Si les expériences sont semblables le coût du calcul de la fréquence moyenne est moins coûteux que si les expériences ne le sont pas. Dans le premier cas, la fréquence moyenne s'obtient par moyenne arithmétique simple des fréquences empiriques affichées à l'écran.

Exemple :

	Nombre des tirages	Fréquences de couleur rouge
	100	0,77
	25	0,88
	25	0,84
	500	0,824
	100	0,83
	25	0,84
Total	775	

Cette stratégie est une reprise des stratégies utilisées dans la phase 2 de la séance 1 en remplaçant les tirages effectués manuellement un par un par des blocs de tirages de 25, 100 ou 500 effectués par l'ordinateur.

• Stratégie « *convergence de fréquences* » :

- réaliser une ou plusieurs suites d'expériences de tirages en nombre croissant,
- constater une convergence des fréquences empiriques vers l'une des fréquences théoriques,
- choisir la composition associée à cette fréquence théorique.

Exemple :

Suite d'expériences	Fréquences de boule rouge
25 tirages	0,520
100 tirages	0,650
500 tirages	0,592

La suite des fréquences empiriques 0,520 ; 0,650 ; 0,592 tend vers la fréquence théorique 0,6 ; la composition est donc 3R-2V

Cette stratégie s'appuie sur l'interprétation fonctionnelle de la loi des grands nombres. Elle ignore donc la notion de risque.

Déroulement et analyse a posteriori

La moitié des groupes commencent l'interrogation de la bouteille-ordinateur par une expérience de 500-tirages puis tous multiplient les expériences sans privilégier celles de 500 tirages, sauf le groupe 5, et sans cumuler les effectifs ou effectuer la moyenne des fréquences, sauf le groupe 6.

Organisation « *non cumulative* »	Organisation « suite consécutive »	Gr1, Gr2, Gr6, Gr8
	Organisation « suite constante (ou constante partielle) »	- Suite constante : Gr5, Gr7 - Suite constante partielle : Gr3
	Organisation « erratique »	Gr4
Organisation « *cumulative* »		Gr7

Tableau 83. Distribution des organisations des 8 groupes – Phase 2

Un seul groupe (le groupe 6) se trompe sur la composition de sa bouteille. Il semble qu'il n'ait pris en compte que 50 tirages. Les autres groupes prennent leur décision selon le tableau suivant :

Stratégie « convergence de fréquences »	Gr2, Gr8 et Gr1
Stratégie « intervalle de fluctuation des fréquences »	Gr5 et Gr4
Stratégie « fréquence moyenne »	Gr 7
Autres	Gr 3 et Gr6

Tableau 84. Distribution des stratégies des 8 groupes

- Le groupe 5 emprunte la stratégie « *intervalle de fluctuation des fréquences* » et expose sur sa fiche les résultats de 5 expériences de 500 tirages sur les 17 expériences réalisées dont 13 de 500 tirages ; ce groupe recueille les fréquences de la seule couleur rouge et établit l'intervalle dont les bornes sont la fréquence la plus faible et la fréquence la plus forte :

"Phỏng vấn" : 500 lần
- Lần 1 : tần suất của bi đỏ : 0,598
- Lần 2 : tần suất của bi đỏ : 0,614
- Lần 3 : tần suất của bi đỏ : 0,632
- Lần 4 : tần suất của bi đỏ : 0,566
- Lần 5 : tần suất của bi đỏ : 0,614.

⇒ tần suất ∈ [0,566 ; 0,632]

⇒ tần suất của bi đỏ ≈ 0,6 ⇒ có 3 bi đỏ, 2 bi xanh

Traduction en français

Interrogation : 500 tirages

1ère fois : la fréquence de boule rouge : 0,598

2ième fois : la fréquence de boule rouge : 0,614

3ième fois : la fréquence de boule rouge : 0,632

4ième fois : la fréquence de boule rouge : 0,566

5ième fois : la fréquence de boule rouge : 0,614

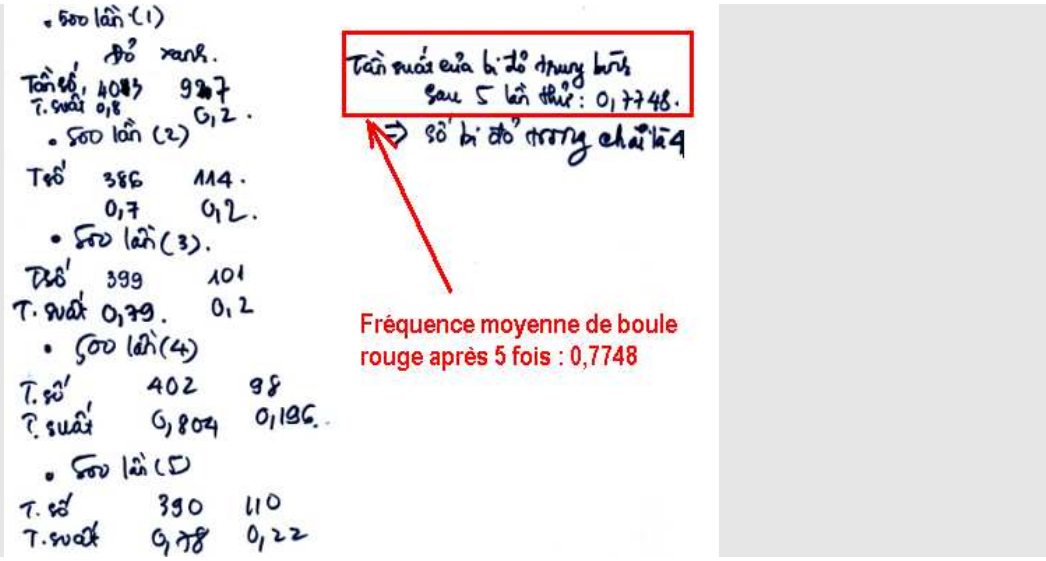
⇒ fréquence ∈ [0,566 ; 0,632]

⇒ fréquence de boule rouge ≈ 0,6 ⇒ il y a 3R-2V

(Fiche 5 du groupe 5)

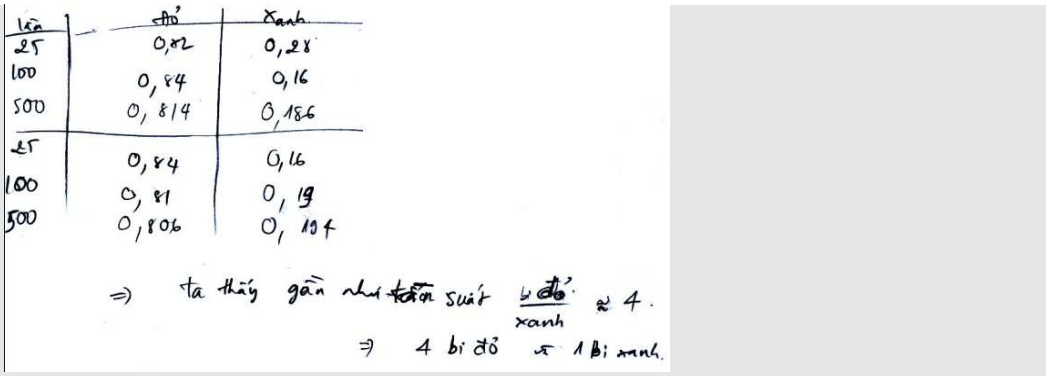
L'intervalle [0,566 ; 0,632] est d'amplitude faible (0,066) et est bien centré sur 0,6 ce qui justifie pour ce groupe le choix de la composition.

- Le groupe 7 emprunte la stratégie « *fréquence moyenne* » ; c'est le seul qui calcule la fréquence moyenne (qui est effectivement la fréquence pour le total des tirages) sur 5 expériences de 500-tirages parmi les 8 expériences réalisées :



(Fiche 5 du groupe 7)

- Le groupe 2 emprunte la stratégie « *convergence des fréquences*» ; il répète en cela l'organisation qu'il avait choisie dans le premier temps de la phase 1. Deux suites consécutives lui suffisent pour ne pas douter de la composition.



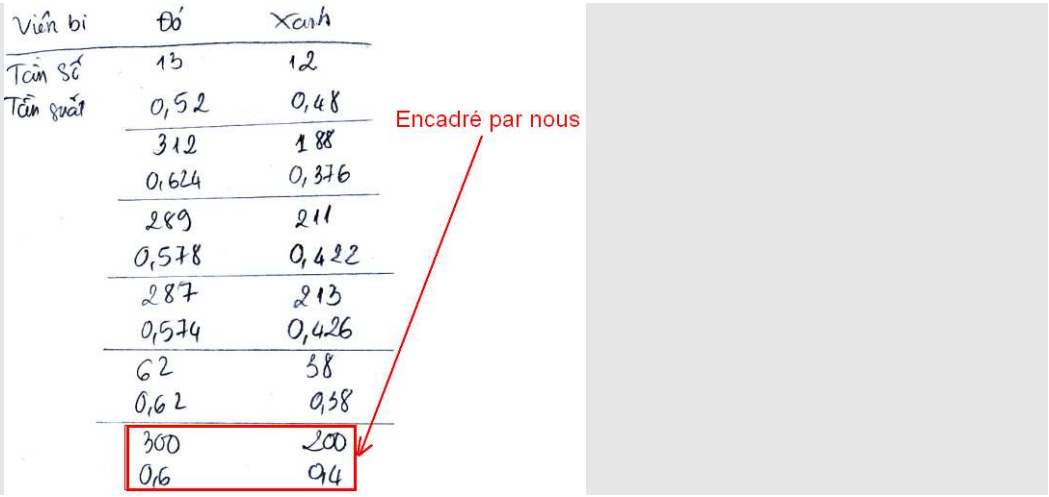
Traduction en français

⇒ on voit presque les fréquences Rouge/Verte ≈ 4

⇒ 4 boules rouges et 1 boule verte

(Fiche 5 du groupe 2)

- Le groupe 3 arrête la suite des expériences lorsque les fréquences affichées sont exactement les fréquences théoriques (c'est le groupe qui a fait des tirages devant toute la classe à la fin de la séance 1) :



Viên bi	Đỏ	Xanh
Tần số	13	12
Tần suất	0,52	0,48
	312	188
	0,624	0,376
	289	211
	0,578	0,422
	287	213
	0,574	0,426
	62	38
	0,62	0,38
	300	200
	0,6	0,4

(Fiche 5 du groupe 3)

Terminons cette analyse par le constat qu'aucun groupe n'accompagne sa décision d'un commentaire sur le risque de s'être trompé. Il est vrai que la consigne n'exprime pas explicitement une demande dans ce sens. Le recueil de l'ensemble des expériences pour chaque groupe et ce qui en est extrait par le groupe pour servir de justification à la composition nous a permis cependant d'établir ce qui fonde leur certitude essentiellement la proximité des fréquences empiriques et de la fréquence théorique. Ce constat pose la question des conditions qui permettent de déstabiliser le point de vue déterministe qui fait que la décision n'est accompagnée d'aucune annonce sur le risque pris.

II.4. Conclusion

Sur quelles significations de la notion de fluctuation d'échantillonnage la séance 2 débouche-t-elle ? Nous en voyons deux :

1. la convergence des fréquences empiriques vers la fréquence théorique (la probabilité) quand le nombre de tirages augmente

2. la réduction de la largeur de l'intervalle de fluctuation des fréquences empiriques quand le nombre de tirages augmente

La loi des grands nombres prend alors deux formes graphiques :

- l'une est analogue au phénomène d'asymptote horizontale pour la courbe représentative de la fréquence considérée comme fonction du nombre de tirages

- l'autre est analogue à l'emboîtement d'intervalles numériques sur la droite réelle

Mais le point de vue probabiliste est absent de ces deux formes. Rappelons ce point de vue : quand le nombre de tirages augmente, la probabilité pour que la fréquence empirique appartienne à un intervalle de largeur fixée et centré sur la fréquence théorique augmente (faisant diminuer dans notre exemple le risque de se tromper en décidant de la composition sur la base de la fréquence théorique proche). Cette probabilité ne peut être égale à 1 et donc le risque n'est jamais nul.

Nous faisons l'hypothèse que le point de vue déterministe n'a pas pu être déstabilisé parce que le nombre de compositions possibles étant faible, tout intervalle de fluctuation des fréquences est presque sûrement inclus strictement dans un intervalle de décision pour les expériences à portée des élèves. L'inclusion de l'intervalle de fluctuation dans l'intervalle de décision bloque l'idée de doute que l'enseignant a tenté d'installer à la fin de la phase 1. La séance 3 est construite sur l'idée que si on augmente le nombre de compositions possibles, la diminution consécutive des intervalles de décision va faire ressurgir la nécessité du doute donc de la quantification du risque.

III. Séance 3

III.1. Introduction

Objectif de la séance : rendre opératoire la loi des grands nombres dans l'évaluation d'un risque pour une décision sur une population « inaccessible ».

La recherche d'un échantillon représentatif d'une population mère « inaccessible » car de distribution inconnue, et de taille grande et indéterminée repose la question du risque pris à décider un tel échantillon car le grand nombre des compositions possibles empêche la comparaison des fréquences empiriques avec des fréquences théoriques qui avait été possible dans la situation-bouteille. L'estimation ponctuelle de telle ou telle fréquence théorique doit alors être remplacée par l'estimation par un intervalle dans lequel on met une certaine confiance. La loi des grands nombres fournit un appui théorique, via la fluctuation d'échantillonnage, à la fabrication d'un tel intervalle à condition de disposer de statistiques sur des expériences répétées de façon identique et indépendante.

La phase 1 est organisée à partir de l'exercice 4 du questionnaire pour installer dans la classe une première signification de la notion d'échantillon représentatif d'une population qui est celle de l'institution vietnamienne : un échantillon représentatif d'une population est un sous-ensemble de la population-mère dont les effectifs par valeurs sont *le mieux proportionnels* à ceux de la population mère. Dans cette signification, la quasi-égalité des fréquences qui résulte de la quasi-proportionnalité des effectifs fait office implicitement de distance entre les deux distributions de fréquences, celle de l'échantillon et celle de la population toute entière.

La phase 2 redonne l'occasion d'exploiter la loi des grands nombres comme outil d'exploration d'une population inconnue. Pour cela la classe, enseignant et élèves, peuvent s'appuyer sur une histoire commune qui leur permet de se référer à la situation-bouteille comme situation typique d'expériences aléatoires identiques et indépendantes sous la forme de tirages répétés avec remise (mémoire didactique). Cette référence commune est susceptible dans de nouvelles conditions de favoriser la construction, indépendamment de la notion de représentativité, de la notion d'échantillon dans le sens utilisé en Statistique inférentielle[133].

La situation bouteille n'a pas pu déboucher sur la notion d'échantillon comme résultat de la répétition des tirages. En effet même avec la bouteille-ordinateur les groupes de tirages ne prennent pas le statut d'échantillon car l'institution réserve ce terme à l'extraction d'un groupe d'individus de la population entière et cela n'est pas possible quand on tire 25 fois une boule parmi 5 boules. Pour construire la notion d'échantillon en s'appuyant sur la situation-bouteille, il faut donc augmenter sensiblement la taille de la population-mère (de 5 à très grande). Les conditions ainsi mises en place dans cette phase 2 déstabilisent la première signification d'échantillon représentatif et permettent d'en construire une nouvelle

[133] Nous appellerons échantillon de taille *n* d'une expérience la série des résultats obtenus en réalisant *n* fois cette expérience ; on dira aussi qu'un échantillon est une liste de résultats de *n* expériences identiques et indépendantes » (document d'accompagnement du programme de Seconde, 2000)

: un échantillon représentatif est un échantillon dans lequel on place une confiance fondée sur la loi des grands nombres et mesurée empiriquement grâce à la fluctuation d'échantillonnage[134]. La distance entre deux distributions de fréquences – celle de l'échantillon et celle de la population – mise en place implicitement dans un cadre déterministe pour la première signification est alors élargie dans un cadre probabiliste aux notions conjointes de risque et de confiance.

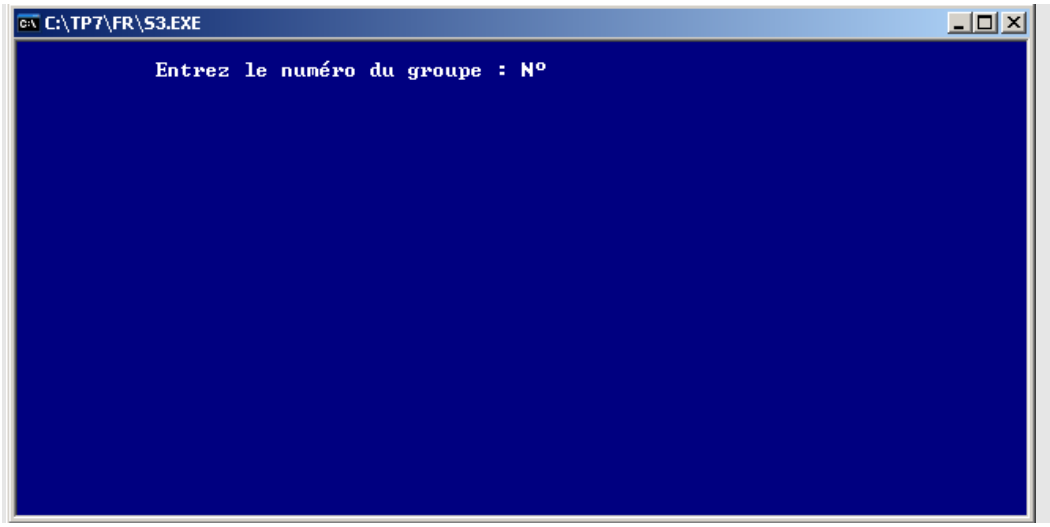
Matériel :

- l'exercice 4 du questionnaire

- le transparent des différentes réponses aux questions 1 et 2 de l'exercice 4 du questionnaire ;

- la fiche 6 pour écrire un sixième échantillon plus représentatif que les cinq échantillons déjà donnés.

- un ordinateur public sur lequel est installé un programme « notes de HCMV » (1000 notes[135]) : l'utilisateur peut lui demander de fabriquer des échantillons de 10 notes qu'il fournit sous la même forme que dans l'exercice 4 (nombres décimaux à 3 chiffres décimaux).



[134] La justification théorique par le théorème de Moivre-Laplace et la mesure grâce à la loi normale ne sont pas intégrés à l'ingénierie.

[135] La distribution est : 90 cinq, 110 six, 140 sept, 400 huit, 160 neuf et 100 dix ; elle est ignorée des élèves.

```
C:\TP7\FR\S3.EXE

        * * * * * * * * * * *      GROUPE N°1      * * * * * * * * * * * *

 Dans l ordinateur, il y a les notes suppérieurs ou égales à  5 de l  élève  de
 tous les lycées de Ho Chi Minh ville. L ordinateur peut fournir un échantillon
 de 10 notes pris par hasard dans l ensemble de toutes les notes

Voulez-vous demander un échantillon de 10 notes ? (Tapez O (Oui)/N (Non)) : _
```

```
C:\TP7\BIN\TURBO.EXE
        * * * * * * * * * * *      GROUPE N°1      * * * * * * * * * * * *

 Dans l ordinateur, il y a les notes suppérieurs ou égales à  5 de l  élève  de
 tous les lycées de Ho Chi Minh ville. L ordinateur peut fournir un échantillon
 de 10 notes pris par hasard dans l ensemble de toutes les notes

Voulez-vous demander un échantillon de 10 notes ? (Tapez O (Oui)/N (Non)) : Oui
Voici un échantillon de 10 notes :    6  6  6  7  8  8  8  8 10 10 10

Voulez-vous demander un échantillon de 10 notes ? (Tapez O (Oui)/N (Non)) :
```

Dans ce programme chaque note est remise dans la population après tirage (« tirage avec remise » au lieu du « tirage sans remise » que laisse entendre l'idée d'extraire un échantillon de la population). Le tirage programmé permet ainsi de fabriquer des échantillons au sens de la statistique inférentielle. Remarquons que ce programme, appelons-le programme-notes, diffère du programme-bouteille en cela qu'il ne fournit ni l'effectif ni la fréquence de chacune des 6 notes.

Le programme a deux commandes pour calculer les résultats de n échantillons de 10 notes :

- commande n°1 : demander n échantillons de 10 notes puis donner les fréquences des 6 valeurs sur les n x 10 notes (cf. partie 1 de la phase 2 où l'on prend n = 50),

```
C:\TP7\FR\S3.EXE

La commande N°1 :

Tapez touche Entrer pour avoir 50 échantillons de 10 notes
```

```
C:\TP7\FR\S3.EXE
Voici un chantillon de 10 notes :    5  6  6  8  8  8  8  9 10 10 10
Voici un chantillon de 10 notes :    7  7  8  8  8  8  9  9 10 10 10
Voici un chantillon de 10 notes :    6  8  8  8  8  8  8  8  9  9  9
Voici un chantillon de 10 notes :    5  6  7  7  7  8  8  8 10 10 10
Voici un chantillon de 10 notes :    5  5  7  8  8  9  9  9 10 10 10
Voici un chantillon de 10 notes :    6  8  8  8  8  8  8  8  9  9  9
Voici un chantillon de 10 notes :    6  6  6  7  8  8  8  8  8  9  9
Voici un chantillon de 10 notes :    6  7  7  8  8  8  8  8  9  9  9
Voici un chantillon de 10 notes :    5  5  7  7  7  7  8  8  8 10 10
Voici un chantillon de 10 notes :    5  5  6  6  8  8  8 10 10 10 10
Voici un chantillon de 10 notes :    5  7  8  8  8  8  9  9 10 10 10
Voici un chantillon de 10 notes :    5  7  7  7  7  8  8  8  8  9  9
Voici un chantillon de 10 notes :    5  5  6  7  8  8  8  8  8  9  9
Voici un chantillon de 10 notes :    5  6  6  8  8  8  8  9  9 10 10
Voici un chantillon de 10 notes :    5  7  8  8  8  8  9  9  9 10 10
Voici un chantillon de 10 notes :    5  5  6  8  8  8  8  9  9  9  9
Voici un chantillon de 10 notes :    5  5  7  7  7  8  8  8 10 10 10
Voici un chantillon de 10 notes :    6  6  7  8  8  8  9  9 10 10 10
Voici un chantillon de 10 notes :    5  5  6  7  7  8  8  9 10 10 10
Voici un chantillon de 10 notes :    5  6  7  7  8  8  8  8  8  9  9
Voici un chantillon de 10 notes :    5  5  8  8  9  9  9 10 10 10 10
Voici un chantillon de 10 notes :    6  7  8  8  8  8  9  9  9  9  9
Voici un chantillon de 10 notes :    5  6  6  7  7  7  8  9  9 10 10

Tapez touche Entrer_
```

```
C:\TP7\FR\S3.EXE

Après 50 fois de demander un échantillon de 10 notes, on obtient le tableau

de distribution des fréquences de 500 notes :

Note              5         6         7         8         9        10

Fréquence    0.0920    0.0920    0.1480    0.3960    0.1740    0.0980

Tapez touche Entrer_
```

Notons que l'affichage des résultats d'une commande n°1 disparaît à chaque nouvelle exécution de la commande.

- commande n°2 : les p fréquences d'une certaine valeur[136] pour p utilisations de la commande n°1 (cf. partie 2 de la deuxième phase et l'on prend p = 20) :

```
C:\TP7\FR\S3.EXE

La commande N°2 :

Choisissez la note que vous voulez observer ses fréquences ? Note :
```

```
C:\TP7\FR\S3_FR.EXE

La commande N°2 :

Choisissez la note que vous voulez observer ses fréquences ? Note : 7

Voici 20 fréquences de note 7 obtenues par 20 demandes de la commande N°2

  0.134  0.162  0.156  0.132  0.162  0.158  0.154  0.160  0.130  0.136
  0.142  0.118  0.132  0.130  0.172  0.138  0.162  0.138  0.156  0.144

Fréquence moyenne de note 7 après 20 fois de demander la commande N°1 :  0.146

Tapez touche Entrer pour ordonner les 20 fréquences
_
```

```
C:\TP7\FR\S3_FR.EXE

Voici 20 fréquences ordonnées croissantes :

  0.118  0.130  0.130  0.132  0.132  0.134  0.136  0.138  0.138  0.142
  0.144  0.154  0.156  0.156  0.158  0.160  0.162  0.162  0.162  0.172

Fréquence moyenne de note 7 après 20 fois de demander la commande N°1 :  0.146

Tapez touche Entrer
```

[136] Soit la valeur 7 soit la valeur 9 (cf. analyse du choix de l'échantillon représentatif par les élèves)

III.2. La première phase (20 minutes)

Objectif de la phase : institutionnaliser la notion d'échantillon représentatif d'une population mère finie et accessible.

Protocole

Cet exercice a été proposé avant l'ingénierie aux élèves de la classe 11 où s'est faite l'expérimentation. C'est une reprise de l'exercice 4 du questionnaire passé en classe 10. Une différence notable est que les deux premières questions de l'exercice sont à faire individuellement à la maison (une semaine avant de réaliser l'expérimentation de la séance 3) ; seule la question 3 est travaillée en groupes dans la classe.

Le secrétariat d'un lycée de Hô Chi Minh ville a recueilli et ordonné les notes obtenues par les 100 élèves du lycée qui ont eu des notes supérieures ou égales à 5 à un examen de mathématiques.

5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	6	6	6	6
6	7	7	7	7	7	7	7	7	7	7	7	7	7	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	9	9	9	9	9	9	9	9	9	9
9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
9	9	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10

Ensemble de notes

Pour communiquer des informations restreintes à comparer avec celles d'autres lycées de Hô Chi Minh ville, l'administration du lycée doit fournir un échantillon de 10 notes qui soit représentatif de l'ensemble des notes obtenues. Pour cela elle extrait du recueil, des échantillons de 10 notes. Voici cinq échantillons possibles :

6	7	7	8	8	8	8	9	9	10

Échantillon 1

7	7	8	8	8	9	9	9	10	10

Échantillon 2

5	6	9	9	9	9	9	9	10	10

Échantillon 3

5	6	7	7	8	8	8	9	9	10

Échantillon 4

6	6	7	8	8	8	8	9	10	10

Échantillon 5

1. D'après vous, parmi les cinq échantillons, quel est le plus représentatif du recueil tout entier ? Pourquoi ?
2. D'après vous, parmi les cinq échantillons, quel est le moins représentatif du recueil tout entier ? Pourquoi ?

L'enseignant prend soin d'analyser auparavant les réponses de ses élèves aux questions 1 et 2 et prépare sur un transparent les différentes réponses apparues. Il projette ce transparent pour rendre publiques ces réponses et organise un débat.

L'enseignant repose la question 3 via la consigne suivante :

> Si possible, proposer un sixième échantillon plus représentatif que les cinq échantillons déjà donnés. L'écrire sur la fiche (6) en disant comment vous l'avez construit.

L'enseignant conclut sur la quasi-proportionnalité des effectifs et la quasi-égalité des fréquences comme critères de bonne représentativité d'un échantillon par rapport à une population dont il est extrait.

La fiche 6 :

> **Fiche 6** (noms, prénoms, lycée, classe, n° groupe)
>
> Votre échantillon de 10 notes : | | | | | | | | | | |
>
> Comment l'avez-vous construit ?

Analyse a priori

On la trouvera dans le chapitre B2 dont nous extrayons le tableau de synthèse des stratégies :

Stratégie Représentants	Stratégie Proportions entre effectifs		Stratégie Distance entre fréquences
	Ordres	Valeurs	
Idéal = même représentant	Idéal = proportionnalité		Idéal = égalité des vecteurs ou des graphiques

Tableau 85. Synthèses des stratégies (exercice 4 du questionnaire)

On se rappelle que la stratégie « représentants » est proche du rapport institutionnel de la classe 10 au traitement numérique d'une série statistique mais qu'elle est coûteuse en calculs. Les élèves ayant le temps chez eux de répondre aux questions 1 et 2, cette stratégie pourrait être empruntée plus franchement que dans la classe 10 choisie pour la passation du questionnaire.

Déroulement et analyse a posteriori

Tableau des résultats des réponses individuelles à deux premières questions :

	E1	E2	E3	E4	E5	Total
Echantillon le plus représentatif	5	5	2	26	5	43
Echantillon le moins représentatif	1	22	12	3	3	42

Tableau 86. Résultat des réponses individuelles

Les couples des échantillons (le plus représentatif ; le moins représentatif) :

Le moins représentatif / Le plus représentatif	E1	E2	E3	E4	E5	Total
E1		0	3	2	0	5
E2	1		1	1	1	4
E3	0	1		0	1	2
E4	0	20	5		1	26
E5	0	2	2	0		4
Total	1	23	11	3	3	41

Tableau 87. Les couples des échantillons

Rappelons que :

- les lignes E1, E3, E5 correspondent à la stratégie « représentant »

- les lignes E2 et E4 correspondent à la stratégie « proportion des effectifs ».

Comme en classe 10 la grande majorité des réponses relèvent de la stratégie « proportion des effectifs ».

Représentants	Proportions entre effectifs	Distance entre fréquences	Total
11	30	0	41
26,8 %	73,2 %	0 %	100 %

Tableau 88. Distribution des stratégies des réponses de classe 11

Rappel des résultats de la classe 10 :

Représentants	Proportions entre effectifs	Distance entre fréquences	Total
21	160	0	181
11,6 %	88,4 %	0 %	100 %

Tableau 89. Distribution des stratégies des réponses de classe 10

L'enseignant projette publiquement un relevé des calculs et des critères de choix utilisés par les élèves :

Les calculs :

Sur la population mère ou sur les échantillons

- Effectif total N, effectifs et fréquences par note, moyenne, mode, médiane, variance V(X), écart-type σ,…,

- Pourcentages des notes.

- Probabilités des notes.

Les critères :

- la présence de toutes les notes de la population mère,

- comparaison des moyennes,

- comparaison des moyennes et des écarts-types,

- comparaison des moyennes, des médianes et des écarts-types,

- comparaison des modes,

- proportionnalité des notes / fréquences des notes

(Relevé des résumés préparé et rendu public par l'enseignant)

Les réponses des élèves sont synthétisées. La troisième question est alors proposée pour un travail en groupes.

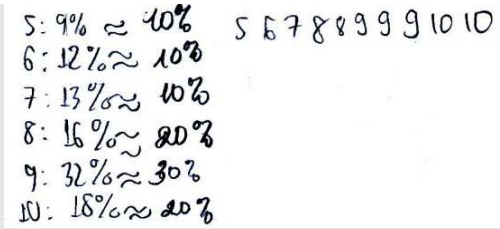
L'enseignant distribue la fiche 6 à chaque groupe.

Les 8 groupes donnent le même échantillon E6p[137] en empruntant tous la stratégie de quasi-proportionnalité des effectifs (ou de quasi-égalité des fréquences) et en complétant leur justification par deux critères :

- l'apparition de toutes les valeurs de notes dans l'échantillon

- la proximité des représentants de l'échantillon et de la population

C'est le cas des groupes 3 et 8 :



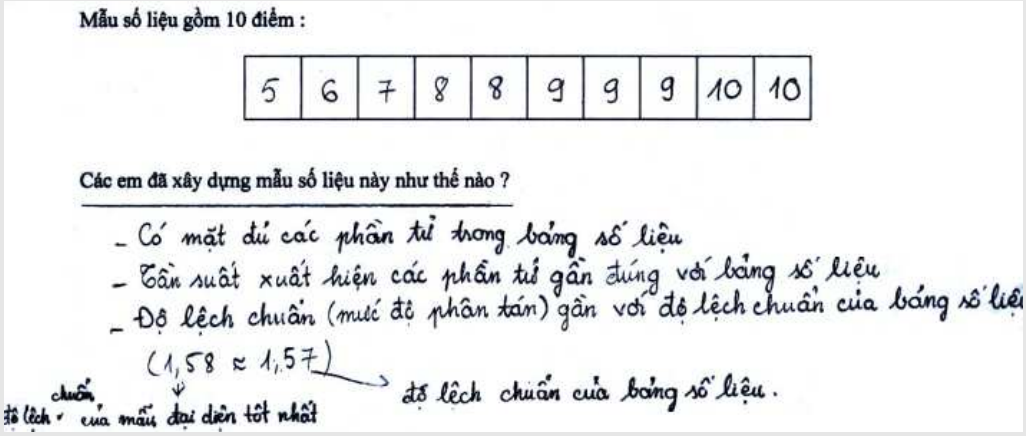
Traduction en français :

La note moyenne de l'échantillon est 8,1, donc la différence entre elle et celle du tableau n'est pas grande.

[137] Échantillon E6p : 5, 6, 7, 8, 8, 9, 9, 9, 10, 10

(Fiche 6 du groupe 3)

«



Mẫu số liệu gồm 10 điểm :

5	6	7	8	8	9	9	9	10	10

Các em đã xây dựng mẫu số liệu này như thế nào ?

- Có mặt đủ các phần tử trong bảng số liệu
- Tần suất xuất hiện các phần tử gần đúng với bảng số liệu
- Độ lệch chuẩn (mức độ phân tán) gần với độ lệch chuẩn của bảng số liệu

(1,58 ≈ 1,57) → độ lệch chuẩn của bảng số liệu.

độ lệch chuẩn của mẫu đại diện tốt nhất

Traduction en français :

- Apparition des toutes les notes du tableau,

- Les fréquences des notes de l'échantillon 6 doivent être proches des fréquences des notes dans le tableau,

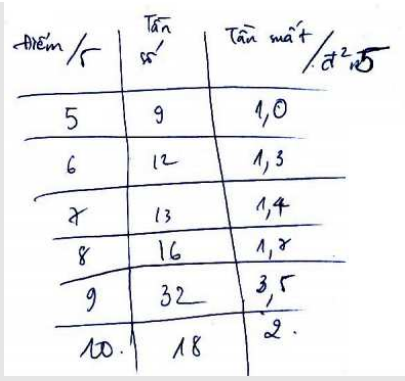
- L'écart-type (le niveau de dispersion) est proche de celui du tableau

(1,58 ≈ 1,57)

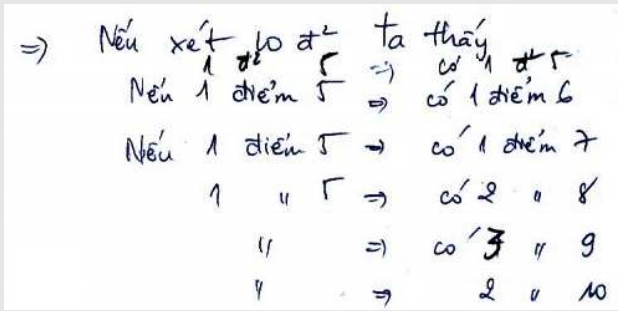
écart-type de l'échantillon le plus représentatif écart-type du tableau

(Fiche 6 du groupe 8)

Le groupe 2, quant à lui, rapporte tous les effectifs de notes à celui de la note 5 pris comme unité :



Điểm / 5	Tần số	Tần suất / đ² 5
5	9	1,0
6	12	1,3
7	13	1,4
8	16	1,7
9	32	3,5
10	18	2



⇒ Nếu xét 10 đ² ta thấy
1 đ² 5 ⇒ có 1 đ² 5
Nếu 1 điểm 5 ⇒ có 1 điểm 6
Nếu 1 điểm 5 → có 1 điểm 7
1 " 5 ⇒ có 2 " 8
" ⇒ có 3 " 9
" ⇒ 2 " 10

Traduction en français :

On voit que :

Pour 1 note 5 ⇒ il y a 1 note 5

Pour 1 note 5 ⇒ il y a 1 note 6

Pour 1 note 5 ⇒ il y a 1 note 7

Pour 1 note 5 ⇒ il y a 2 notes 8

Pour 1 note 5 ⇒ il y a 3 notes 9

Pour 1 note 5 ⇒ il y a 2 notes 10

(Fiche 6 du groupe 2)

Notes	Effectifs	Rapport d'effectif d'une note sur effectif de note 5	Rapports « arrondis » en gardant le 1er chiffre de la partie décimale
5	9	9/9 = 1	1
6	12	12/9 ≈ 1,33	1,3
7	13	13/9 ≈ 1,44	1,4
8	16	16/9 ≈ 1,78	1,7
9	32	32/9 ≈ 3,56	3,5
10	18	18/9 = 2	2

(La colonne grise est notre explication pour détailler les calculs)

Le retour aux représentants numériques pour vérifier le caractère représentatif de l'échantillon E6p chez les élèves de classe 11 montre encore une fois la force du rapport institutionnel à la notion de représentant numérique pour résumer les données statistiques. Cette vérification exploite une condition nécessaire (mais pas suffisante) de la similitude des distributions de fréquences.

L'insistance sur « l'apparition de toutes les valeurs » révèle une fois de plus la *normalité* des tableaux de distribution proposés aux élèves : ces tableaux ne comportent pas de valeurs rares qui pourraient être exclues dans un échantillon même représentatif.

A la fin de la phase 1, l'enseignant institue la quasi-proportionnalité des effectifs des notes comme seul critère de choix de l'échantillon le plus représentatif.

Ens Je vois l'échantillon représentatif du groupe 1 : 5, 6, 7, 8, 8, 9, 9, 9, 10, 10.

Quels sont les groupes qui proposent le même échantillon que celui-ci ?

(*Tous les groupes lèvent le doigt pour dire qu'ils ont le même résultat*)
Bon, quel groupe peut aller au tableau pour présenter comment il trouve cet échantillon ?
(*Un élève du groupe 3 vient au tableau*)

El D'abord, j'ai calculé les fréquences des notes et j'ai obtenu les résultats : la note 5 occupe 99%...

Ens 99 ???

El Ah non, 9%,
(*Un élève se trouvant près de l'appareil de vidéo dit : « 0,09 »*)
La note 6, 12% ; la note 7, 13%, la note 8 occupe 16%, la note 9 occupe 32%, et la note 1à occupe 18%. Puis notre groupe a arrondi ces pourcentages.

Ens Ah, ton groupe a arrondi ces nombres, mais comment le ais-tu ? Je vois ici 9%, je ne comprends pas bien comment les arrondir ? Peux-tu l'expliquer clairement ?

El Je fais comme ça : pour les nombres dont l'unité est de 5 à 9, je les augmente à 10, par exemple, le 9%, je l'arrondis à 10% ; le 12%, je l'arrondis à 10% aussi,…

Ens D'accord, tu les écris au-dessus des pourcentages, et avec % car tu dis pourcentage.
(*cet élève continue à écrire : 10%, 10%, 10%, 20%, 30%, 20%*)
La classe, vous comprenez ce qu'il fait pour arrondir ces pourcentages ?

Els Comme je vois que les trois notes 5, 6, 7 occupent 10% des notes, donc les nombre des ces notes doivent le même, c'est-à-dire chacun de ces trois notes doit prendre une place.

Ens Pourquoi une place pour chacun de ces trois note ?

Els Car on a 10 cases pour 10 notes, cette note occupe 10% c'est-à-dire il occupe une place et chacun de ces trois notes prend une place.

Ens Ah, tu veux dire que 10% est égale à 1/10.

Els Les notes 8 et 10 occupent 20%, ils prennent donc deux places et enfin la note 9 occupant 30% doit prendre 3 places. Ça fait en total 10 places.

Ens Bon, merci.
Et les autres groupes ? vous faites pareillement avec les pourcentages ?

Els Oui …

Ens Est-ce qu'il y a une autre méthode pour construire un échantillon qui est mieux que celui-ci ? Non ?
Vous voyez, cette méthode nous donne un échantillon qui est <u>très proportionnel</u>[138] à l'ensemble de toutes les notes. Bon, je vous remercie de m'apprendre cette très bonne méthode pour construire un échantillon représentatif par la proportionnalité des effectifs entre cet échantillon et l'ensemble de 100 notes.

(Script de la séance 3)

[138] Souligné par nous.

La signification d'un échantillon représentatif d'une population a encore le sens que lui donne l'institution, à savoir celle d'un extrait quasi-proportionnel de la population. Avec la deuxième phase, il s'agit de mettre en place les conditions d'une nouvelle signification où l'échantillon serait le résultat aléatoire et donc fluctuant d'expériences aléatoires identiques et indépendantes.

III.3. La deuxième phase (25 minutes)

Objectif de la phase 2 : produire un échantillon représentatif d'une population « inaccessible » en s'appuyant sur la loi des grands nombres.

III.3.1 Le début de la deuxième phase : engagement de la loi des grands nombres

Protocole

Les élèves restent répartis selon les mêmes groupes que précédemment. L'enseignant introduit la situation « notes » :

> Le rectorat de la ville d'Hô Chi Minh ville a recueilli toutes les notes supérieures ou égales à 5 de tous les lycées de la ville. Toutes ces notes sont dans l'ordinateur. Pour communiquer des informations restreintes à comparer avec celles des autres villes du pays, le rectorat doit fournir au ministère un échantillon de 10 notes qui soit représentatif de l'ensemble des notes obtenues. Mais cet ensemble est si grand que le rectorat se demande comment faire. Pourriez-vous lui proposer une méthode ?

L'enseignant sollicite des réponses orales durant quelques minutes ; même si l'enseignant insiste sur le coût du traitement exhaustif à cause de la taille de la population à étudier il est possible que la méthode de sondage aléatoire simple ne soit pas formulée (pourquoi tirer au hasard alors que toutes les données sont accessibles dans l'ordinateur et qu'on peut faire faire ce qu'on veut à un ordinateur ?).

Puis l'enseignant présente le programme disponible sur l'ordinateur public :

> L'ordinateur peut fabriquer pour vous un échantillon de 10 notes en faisant une expérience de 10 tirages au hasard de notes et il peut présenter pour vous cet échantillon dans un tableau en ordonnant les notes tirées, comme dans l'exercice 4 du questionnaire. Je fais quelques essais devant vous.

L'enseignant explique la méthode de fabrication d'un échantillon en se référant à la situation-bouteille.

L'enseignant donne alors la consigne suivante :

> Comment faire avec l'aide de l'ordinateur pour obtenir un échantillon de 10 notes représentatif de l'ensemble des notes (celles de tous les élèves de tous les lycées de Hô Chi Minh Ville qui ont eu des notes supérieures ou égales à 5) ?

Pouvez-vous rédiger sur la fiche (7) ce que vous auriez envie de demander à l'ordinateur en donnant vos raisons. En particulier quels calculs lui demanderiez-vous de faire à votre place ?

Voici la fiche 7 :

Feuille 7 (noms, prénoms, lycée, classe, n° groupe) Ce que vous auriez envie de demander à l'ordinateur en donnant vos raisons. En particulier quels calculs lui demanderiez-vous de faire à votre place ?

Après 10 minutes, l'enseignant fait la synthèse rapide des demandes.

Analyse a priori

Dans le tout début de la phase 2, les élèves ignorent le nombre de notes mais ils savent que ces notes sont dans l'ordinateur. Vont-ils proposer une méthode de recherche de toutes les notes ou vont-ils faire le rapprochement avec la situation bouteille-ordinateur même si le nombre de boules est maintenant inconnu ? Dans ce cas on s'attend à ce qu'ils demandent que l'ordinateur puisse effectuer des tirages répétés note par note ?

Pour la question qui suit, les élèves n'ont pas accès à l'ordinateur que seul l'enseignant manipule. L'élève devra donc anticiper individuellement puis en se concertant avec les élèves de son groupe une procédure générale intégrant des calculs particuliers de production d'un échantillon de 10 notes le plus représentatif.

Le problème de trouver un échantillon représentatif de 10 notes avec 6 valeurs à partir d'échantillons de 10 notes fournis par l'ordinateur est en continuité avec le problème de l'exercice 4 du questionnaire lui-même repris dans la première phase. Les expériences demandées aux ordinateurs ressemblent elles-aussi à celles de la séance 2. Par contre en annonçant que la population est très nombreuse et inaccessible dans son ensemble on change la signification du résultat de l'expérience réalisée par l'ordinateur puisque les échantillons de taille 10 sont plus petits que la population elle-même et peuvent être regardés comme des sous-ensembles de la population à l'instar de l'exercice 4 mais à la différence de la séance 2.

Il est possible de demander un seul échantillon et de déclarer qu'il est représentatif ou bien de se référer à la loi des grands nombres en répétant cette expérience n fois.

Dans ce deuxième cas, on peut envisager au moins quatre types de calculs :

- calcul d'un échantillon « moyen » de 10 notes où l'effectif (non nécessairement entier) de chaque note est la moyenne arithmétique des effectifs des n échantillons de 10 notes demandés ; cela revient à considérer l'échantillon moyen ainsi calculé comme plus proche de l'échantillon représentatif que chacun des échantillons demandés (la moyenne est un bon représentant de la population)

- calcul d'un échantillon modal de 10 notes, l'échantillon modal étant l'échantillon le plus fréquent dans la population de tous les échantillons demandés ((le mode est un bon représentant de la population)

Dans ces deux premiers calculs, on fait des statistiques sur une population d'échantillons de 10 notes.

- calcul d'une distribution d'effectifs de la population de n x 10 notes par sommation des effectifs des n échantillons de 10 notes[139]
    - soit on regarde cette distribution empirique comme proche de la distribution de la population toute entière (celle des notes du rectorat) et on utilise la méthode institutionnalisée à la fin de la phase 1 pour trouver l'échantillon E6
    - soit on considère que l'on a fait 10n tirages et comme dans l'approche fréquentiste on déduit les probabilités expérimentales

Déroulement et analyse a posteriori :

- Au début de la phase 2, l'enseignant introduit la nouvelle situation en la comparant avec celle de l'exercice 4 :

Ens	Dans l'exercice fait à la maison on a 100 notes de 100 élèves d'un lycée, un seul lycée, et ce lycée doit prendre un échantillon de 10 notes pour comparer avec ceux des autres lycées. Maintenant, la nouvelle situation est plus large : après un examen commun d'une séance d'une durée de 45 minutes concernant tous les élèves de notre ville de HCMV, toutes les notes sont envoyées au directeur du service d'éducation et de formation de Ho Chi Minh ville. Cet ensemble des notes est très grand. Je ne sais combien il y a de notes mais je sais qu'il y en a un très grand nombre. […] Le directeur du service d'éducation et de formation de Ho Chi Minh ville a reçu les notes supérieures ou égales à 5 de tous les élèves de tous les lycées de la ville. Il veut choisir un échantillon de 10 notes pour communiquer avec les autres villes du Vietnam. Toutes les notes sont dans cet ordinateur. (l'enseignant désigne du doigt l'ordinateur public). Comme il y a trop des notes, il ne peut pas les compter. Alors, qu'est-ce que vous pouvez lui conseiller de faire pour obtenir un échantillon le plus représentatif possible ?

(Script de la séance 3)

Un élève du groupe 2 propose oralement devant la classe deux solutions pour « aider » le directeur. La première d'entre elles s'appuie sur la loi de grands nombres comme nous nous y attendions. La seconde est une plaisanterie. Dans la première solution, les fréquences des notes doivent être calculées après un grand nombre des tirages d'une note. Voici l'échange public entre l'enseignant et le groupe 2 :

El	J'ai deux solutions : - première solution : le directeur doit écrire un programme qui permet de faire un grand nombre des tirages pour avoir des fréquences très exactes des notes

[139] Cette sommation fait appel à l'indépendance des n x 10 tirages considérés comme expériences identiques et indépendantes.

	5, 6, … etc. A partir de ces fréquences, on peut prévoir combien il y a de notes 5, de notes 6, … etc. - deuxième solution : comme il est le directeur du service d'éducation et de formation, il peut ordonner aux autres personnes de compter les effectifs de toutes les notes pour lui. (la classe rit)
Ens	Ah, tu veux un programme qui peut demander un très grand nombre des tirages. Mais je veux savoir combien, tu va prendre de notes par tirage ?
El	… une note… c'est comme on prend une boule dans une urne
Ens	A chaque fois que tu vas demander une note de l'ensemble qu'est-ce que tu vas observer ?
El	Je vais voir la note, comme la couleur d'une boule…
Ens	Tu veux faire combien de demande ?
El	Plus le nombre des tirages est grand, plus ce sera mieux… environ 500 ou 1000 fois.
Ens	Et qu'est-ce que tu feras avec ces notes ?
El	Je vais calculer les fréquences de toutes ces notes…

(Script de la séance 3)

Cet élève formule donc la nécessité de la répétition de l'expérience. De plus il s'appuie sur la situation-bouteille pour justifier sa solution : « … une note… c'est comme on prend une boule dans une urne » qui fait référence au mode du tirage et « Je vais voir la note, comme la couleur d'une boule… » qui fait référence à l'expérience et à son observation. Cet échange public devrait induire dans la classe une procédure générale de construction d'un échantillon représentatif par répétition des tirages justifié par la loi des grands nombres. La question des calculs demeure. Quels calculs les groupes vont-ils proposer pour fabriquer l'échantillon.

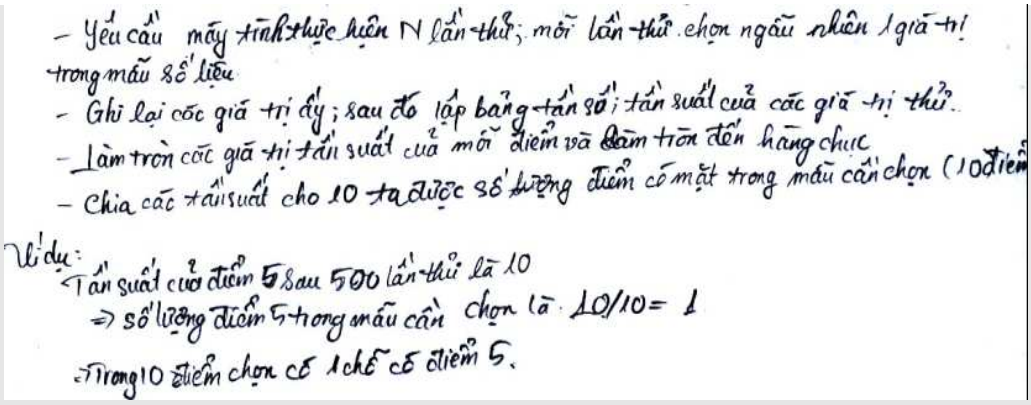
- L'enseignant se saisit de la réponse de l'élève du groupe 2 et présente les commandes d'exécution d'un tel programme en expliquant les choix faits dans sa conception mathématique :

Ens	Dans cet ordinateur, il y a les notes supérieures ou égales à 5 de tous les élèves. Un programme intégré dans l'ordinateur dispose d'une commande particulière. Chaque fois que l'on utilise cette commande, la machine nous donnera un sous-ensemble de 10 notes. Ces 10 notes sont prises de la façon suivante : on prend au hasard une note dans l'ensemble de toutes les notes, puis on la remet, c'est-à-dire prendre une note avec remise. Imaginez qu'il y a beaucoup, beaucoup,… de boules et que l'on prenne une boule au hasard puis qu'on la remette. On fait 10 fois comme ça et on obtient un sous-ensemble de 10 boules. Vous comprenez ? (L'enseignant fait certains essais sur l'écran public) Bon, maintenant, je vous propose de bien réfléchir pour me dire quelles fonctions vous voulez pour cet ordinateur ou quels calculs vous voulez lui demander de faire pour vous, pour que vous obteniez le meilleur échantillon, c'est-à-dire l'échantillon le plus représentatif ?

(Script de la séance 3)

Remarquons que l'enseignant ne reprend que partiellement la solution du groupe 2 puisqu'il ne mentionne pas l'idée de répétition de la commande. Il distribue alors la fiche 7.

- Comme nous l'avions prévu, toutes les réponses à cette nouvelle question contiennent des demandes de répétitions de tirages mais aucune n'envisage d'être confrontée à une fluctuation d'échantillonnage des échantillons de taille 10 que fournit l'ordinateur. Certains groupes réclament à l'ordinateur ce qu'ils obtenaient de la bouteille sans tenir compte de la commande disponible qui impose la taille de l'échantillon (égale à 10). Par exemple le groupe 8 réclame des échantillons d'une seule note comme dans la situation-bouteille.


- Yêu cầu máy tính thực hiện N lần thử; mỗi lần thử chọn ngẫu nhiên 1 giá trị trong mẫu số liệu
- Ghi lại các giá trị đó; sau đó lập bảng tần số; tần suất của các giá trị thử.
- Làm tròn các giá trị tần suất của mỗi điểm và làm tròn đến hàng chục
- Chia các tần suất cho 10 ta được số lượng điểm có mặt trong mẫu cần chọn (10 điểm

Ví dụ:
Tần suất của điểm 5 sau 500 lần thử là 10
=> số lượng điểm 5 trong mẫu cần chọn là 10/10 = 1
=> Trong 10 điểm chọn có 1 chế có điểm 5.

Traduction en français :

- Demander la machine de réaliser N essais ; pour chaque essai, choisir au hasard 1 valeur dans les données

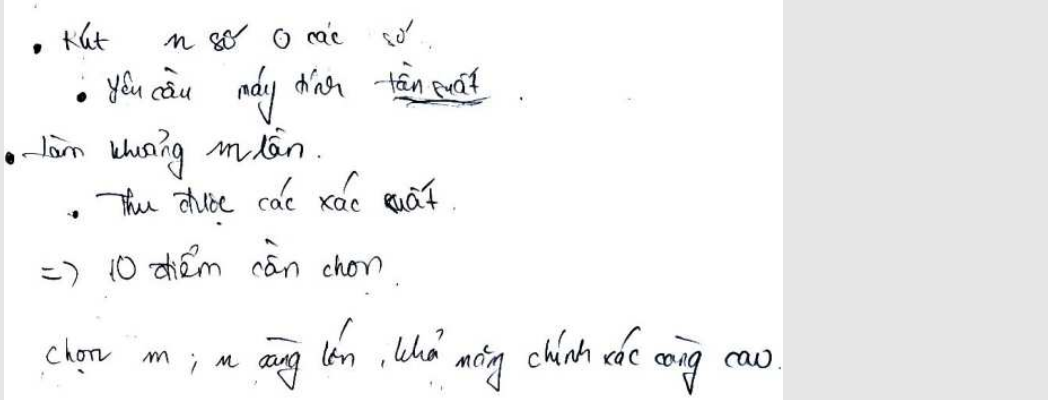
- Noter ces valeurs puis établir le tableau des effectifs, des fréquences des valeurs essayées.

- Arrondir les fréquences des notes à peu près de dixième.

- Diviser[140] les fréquences en 10 pour obtenir les notes qui se présentent dans l'échantillon choisi (10 notes).

(Fiche 7 du groupe 8)

Quant au groupe 1, il entre tout de suite dans un processus plus général avec les paramètres m, n où n est la taille d'un échantillon et m est le nombre des échantillons de n notes. Mais il ne mentionne pas le traitement mathématique des différentes fréquences pour chaque note qu'il obtiendra de l'ordinateur à la fin des m fois n-tirages. Comment obtiendra-t-il les « probabilités » dont il parle ?


• Kết n số 0 các số.
• Yêu cầu máy tính tần suất.
• Làm khoảng m lần.
• Thu được các xác suất.
=> 10 điểm cần chọn.
Chọn m; n càng lớn, khả năng chính xác càng cao.

[140] Ce groupe propose un calcul faux car il « doit » multiplier les fréquences par 10 au lieu de les diviser par 10.

Traduction en français :

- Tirer n notes dans toutes les notes

  Demander à la machine de calculer les fréquences[141]

- Refaire environ m fois

Obtenir les probabilités ⇒ Obtenir 10 notes que l'on doit choisir.

Plus on choisit la valeur grande pour m et n, plus la possibilité de d'avoir l'exactitude est grande.

(Fiche 7 du groupe 1)

En fait la loi des grands nombres continue à être mise en avant sans son corollaire de la fluctuation des effectifs entre échantillons de même taille (ici 10 notes).

III.3.2 Suite de la deuxième phase : choix d'une fréquence empirique comme fréquence théorique.

Protocole

Avec l'intention de déboucher sur la nécessité de prendre en compte cette fluctuation pour décider d'un échantillon représentatif, l'enseignant va demander aux élèves de fournir effectivement un échantillon représentatif. Pour cela il commence par expliquer le fonctionnement de la commande n°1 et donne la consigne suivante :

Utiliser la commande n°1 et en déduire un échantillon représentatif de la population. Ecrire les résultats sur la fiche 8.

Voici la fiche 8 :

Fiche 8 (noms, prénoms, lycée, classe, n° groupe) Ecrivez les résultats de la commande n°1 :						
Valeurs	5	6	7	8	9	10
Fréquences						
Décidez de l'échantillon représentatif de l'ensemble des notes de la ville d'Ho Chi Minh						

Cette fiche a été distribuée en double exemplaire (fiche 8a et 8b) avec l'idée que certains groupes rencontreraient des distributions de fréquences produisant des échantillons représentatifs différents. En effet, examinons la distribution des données des 1000 notes qui a été implémentée et qui est cachée à l'élève :

[141] Souligné par le groupe 1

Notes	5	6	7	8	9	10
Effectifs	90	110	140	400	160	100
Fréquences	0,09	0,11	0,14	0,40	0,16	0,10

Tableau 90. Distribution des données de 1000 notes

Analyse a priori

La commande n°1 présente l'ensemble des 50 suites de 10 notes ordonnées puis fournit, à la demande, la distribution des fréquences d'un échantillon de 500 notes. Les élèves peuvent constater la fluctuation d'échantillonnage comme dans le programme-bouteille (séance 2). L'idée de retenir un seul de ces échantillons comme échantillon représentatif devrait être écartée. On s'attend à ce que les élèves soient immédiatement intéressés par la touche T qui donne accès au tableau de distribution des fréquences des 500 notes et qu'ils soient ainsi confrontés, après plusieurs essais à la fluctuation des fréquences des notes pour ces échantillons de 500 notes.

Les caractéristiques de la distribution des proportions de notes - faible différence entre la proportion de notes 7 et celle de notes 9 et toutes deux proches de 0,15 ; autres proportions proches de dixièmes d'entiers - produisent une distribution d'échantillonnage des fréquences où les probabilités « un 7 deux 9 » et « deux 7 un 9 » sont suffisamment fortes pour que les deux choix puissent apparaître dans les fiches 8a et 8b d'un même groupe et surtout dans l'ensemble des fiches de tous les groupes[142]. Les deux échantillons attendus car les plus probables[143] et concurrents sont donc :

Notes	5	6	7	8	9	10
Effectifs dans l'échantillon représentatif	1	1	1	4	2	1

(1)

et

Notes	5	6	7	8	9	10
Effectifs dans l'échantillon représentatif	1	1	2	4	1	1

(2)

Les élèves ont pratiqué auparavant l'arrondi du produit 10 x fréquence de la note pour obtenir l'effectif de la note. Par contre, comment vont-ils ajuster à 10 l'effectif de l'échantillon dans les cas où les deux fréquences empiriques, celle de la note 7 et celle de la note 9 sont toutes les deux du même côté de 0,15 ?

[142] Le plus fort reste pourrait s'appliquer à d'autres notes que 7 ou 9 mais cette éventualité est très peu probable.

[143] Nous avons procédé à une simulation par un modèle multinomial et les probabilités estimées sont : 0.8 pour le type (1), 0.19 pour le type (2) et 0.01 pour l'ensemble des autres types.

Stratégies

Stratégie « *échantillon majoritaire* » : exécuter plusieurs fois la commande n°1 et choisir l'échantillon du type (1) ou (2) qui apparait majoritairement.

Exemple : exécuter 6 fois de la commande n°1 ; on obtient ici quatre fois un échantillon représentatif de type (1) et deux fois un échantillon représentatif de type (2) ; on propose l'échantillon de type (1) car c'est le plus fréquemment apparu :

Notes	5	6	7	8	9	10	
Effectifs	38	59	60	205	86	52	
Frequences	0.0760	0.1180	0.1200	0.4100	0.1720	0.1040	(1)

Notes	5	6	7	8	9	10	
Effectifs	39	55	83	191	81	51	
Frequences	0.0780	0.1100	0.1660	0.3820	0.1620	0.1020	(2)

Notes	5	6	7	8	9	10	
Effectifs	37	68	55	202	85	53	
Frequences	0.0740	0.1360	0.1100	0.4040	0.1700	0.1060	(1)

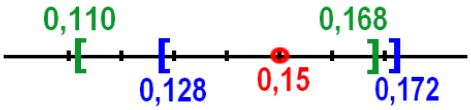
Notes	5	6	7	8	9	10	
Effectifs	44	61	65	212	73	45	
Frequences	0.0880	0.1220	0.1300	0.4240	0.1460	0.0900	(1)

Notes	5	6	7	8	9	10	
Effectifs	50	70	72	191	64	53	
Frequences	0.1000	0.1400	0.1440	0.3820	0.1280	0.1060	(2)

Notes	5	6	7	8	9	10	
Effectifs	43	45	71	197	84	60	
Frequences	0.0860	0.0900	0.1420	0.3940	0.1680	0.1200	(1)

Stratégie « *fluctuation* » : exécuter plusieurs fois la commande n°1, construire l'intervalle de fluctuation des fréquences de la note 7 et celui de la note 9 puis choisir l'effectif égal à 2 contre l'effectif égal à 1 pour la note dont l'intervalle de fluctuation des fréquences se présente « plus à droite » que l'autre[144].

Exemple :



L'intervalle de fluctuation des fréquences de la note 7 (l'intervalle [0,128 ; 0,172]) est plus à droite de celui de la note 9 (l'intervalle [0,110 ; 0,168]). Choisir donc deux notes 7 et une note 9, c'est-à-dire l'échantillon représentatif de type (2).

Stratégie « *moyenne des fréquences* » : réaliser plusieurs fois la commande n°1, calculer la fréquence moyenne des fréquences pour toutes les notes et proposer l'échantillon « moyen » comme l'échantillon représentatif.

Exemple :

[144] L'ordre des intervalles décide de l'ordre des fréquences. L'intervalle le plus à droite est celui dont le centre est le plus à droite. De légalité des centres on déduit l'égalité des fréquences.

Notes	5	6	7	8	9	10
Fréquences	0,094	0,112	0,134	0,392	0,160	0,108
Fréquences	0,084	0,102	0,133	0,434	0,146	0,111
Fréquences	0,106	0,126	0,140	0,408	0,132	0,088
Fréquences	0,092	0,102	0,136	0,414	0,138	0,118
Fréquences	0,078	0,110	0,166	0,382	0,162	0,102
Moyennes des fréquences	0,091	0,110	0,142	0,406	0,148	0,105
Effectifs	1	1	1	4	2	1

On choisit l'échantillon représentatif de type (1). Remarquons que la stratégie « *fluctuation* » aurait conduit à choisir le type (2).

Les deux dernières stratégies s'inspirent de la loi des grands nombres qui est une loi sur les fréquences mais nécessitent des relevés écrits de toutes les fréquences de toutes les notes et sont donc très coûteuses en écritures et calculs. La première stratégie ne prend pas directement en compte les fréquences puisqu'elle n'observe que les échantillons dits représentatifs mais elle est moins coûteuse puisqu'elle peut se passer du relevé écrit de tous les tableaux de fréquences.

Déroulement et analyse a posteriori

Tous les groupes multiplient les exécutions de la commande n°1. Ils ne relèvent pas par écrit les tableaux de fréquences successifs sauf deux des tableaux qu'ils mettent sur les fiches 8a et 8b comme la consigne les incitait à le faire.

Une moitié des échantillons représentatifs produits par les groupes sont de type (1) et l'autre moitié de type (2). Plus précisément

- 7 groupes sur 8 ont affichent un échantillon du type (2) sur les fiches 8a et 8b et pourtant l'historique informatique de leurs actions révèlent qu'ils ont rencontré les deux types (1) et (2).

Par exemple le groupe 1 a réalisé effectivement 4 fois la commande n°1 en obtenant trois fois le type (1) et une fois le type (2) :

Notes	5	6	7	8	9	10
Effectifs	40	51	85	188	74	62
Frequences	0.0800	0.1020	0.1700	0.3760	0.1480	0.1240

(Enregistré à 16:50:27)

Notes	5	6	7	8	9	10
Effectifs	39	43	72	215	82	49
Frequences	0.0780	0.0860	0.1440	0.4300	0.1640	0.0980

(Enregistré à 16:51:27)

Notes	5	6	7	8	9	10
Effectifs	53	58	74	207	62	46
Frequences	0.1060	0.1160	0.1480	0.4140	0.1240	0.0920

(Enregistré à 16:54:06)

Notes	5	6	7	8	9	10
Effectifs	42	67	68	201	62	60
Frequences	0.0840	0.1340	0.1360	0.4020	0.1240	0.1200

(Enregistré à 16:55:26)

Ce groupe décide d'afficher les deux dernières distributions qui conduisent à des échantillons de type (2). Le fait que le type (2) apparaît plus souvent que le type (1) explique-t-il sa décision ? La procédure relèverait alors d'une stratégie « échantillon majoritaire ».

Hãy ghi các kết quả của Lệnh N°1 :

Điểm	5	6	7	8	9	10
Tần suất	0,166	0,116	0,148	0,414	0,124	0,092
	1	1	2	4	1	1

Hãy ghi mẫu số liệu chọn làm đại diện cho tập hợp toàn bộ các điểm của học sinh TP. Hồ Chí Minh

5	6	7	7	8	8	8	8	9	10

(Fiche 8a du groupe 1)

Hãy ghi các kết quả của Lệnh N°1 :

Điểm	5	6	7	8	9	10
Tần suất	0,084	0,134	0,136	0,402	0,124	[illegible]
	1	1	2	4	1	1

Hãy ghi mẫu số liệu chọn làm đại diện cho tập hợp toàn bộ các điểm của học sinh TP. Hồ Chí Minh

5	6	7	7	8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

(Fiche 8b du groupe 1)

- un seul groupe (le groupe 4) a hésité entre les deux types à la suite de ses tirages. Ce groupe a réalisé 4 fois la commande n°1 et a obtenu deux fois le type (1) et deux fois le type (2). Il décide de copier dans la fiche 8a l'échantillon représentatif de type (2) ; puis pour la deuxième fiche 8b, il note une distribution de fréquences correspondant à un échantillon de type (1) en laissant vide sa proposition de l'échantillon représentatif :

Notes	5	6	7	8	9	10
Effectifs	37	56	70	202	75	60
Frequences	0.0740	0.1120	0.1400	0.4040	0.1500	0.1200

(Enregistré à 16:45:16)

Notes	5	6	7	8	9	10
Effectifs	47	51	69	211	68	54
Frequences	0.0940	0.1020	0.1380	0.4220	0.1360	0.1080

(Enregistré à 16:46:55)

Notes	5	6	7	8	9	10
Effectifs	49	71	67	200	63	50
Frequences	0.0980	0.1420	0.1340	0.4000	0.1260	0.1000

(Enregistré à 16:49:57)

Notes	5	6	7	8	9	10
Effectifs	43	47	68	220	82	40
Frequences	0.0860	0.0940	0.1360	0.4400	0.1640	0.0800

(Enregistré à 16:51:43)

Hãy ghi các kết quả của Lệnh N°1 :

Điểm	5	6	7	8	9	10
Tần suất	0,0940	0,1020	0,1380	0,4220	0,1360	0,1080

Hãy ghi mẫu số liệu chọn làm đại diện cho tập hợp toàn bộ các điểm của học sinh TP. Hồ Chí Minh

5	6	7	7	8	8	8	8	9	10

(Fiche 8a du groupe 4)

Hãy ghi các kết quả của Lệnh N°1 :

Điểm	5	6	7	8	9	10
Tần suất	0,086	0,094	0,136	0,44	0,164	0,08

Hãy ghi mẫu số liệu chọn làm đại diện cho tập hợp toàn bộ các điểm của học sinh TP. Hồ Chí Minh

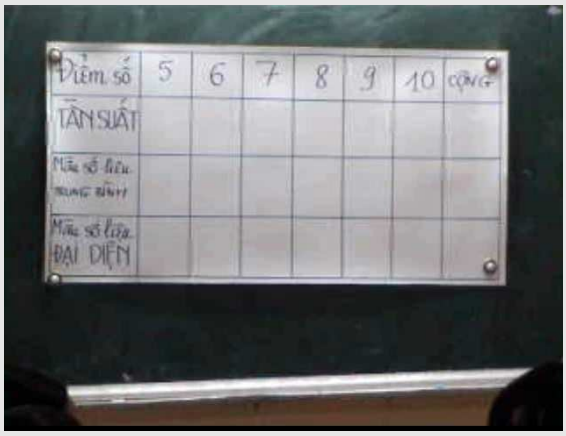
(Fiche 8b du groupe 4)

L'unicité implicite de l'échantillon représentatif est un obstacle à l'affichage de deux réponses différentes.

Intervention publique de l'enseignant

Cette intervention a pour objectif de rendre public le problème du choix de l'échantillon représentatif entre les deux types (1) et (2) et de préparer ainsi la situation didactique qui va suivre. Pour cela l'enseignant demande à la machine 50 (n = 50) échantillons de 10 notes, exécute la commande n°1, affiche la distribution de fréquences des 500 notes, la recopie dans un tableau préparé à l'avance :

Ens	Bon, regardez au tableau, SVP. Donnez- moi le résultat obtenu sur l'écran pour que je l'écrive sur cette fiche, SVP.



	(*l'affiche commune est collée au tableau*)
El	0,092 … 0,114 … 0,156 … 0,39 … 0,15 … et 0,098.
Ens	Merci. Bon, j'ajoute un zéro pour que la partie décimale de tous ces nombres ait 3 chiffres. (*l'enseignant ajoute 0 pour avoir 0,390 et 0,150*) Nous avons maintenant cette ligne des fréquences de 10 notes. Si nous additionnons ces nombres, il fait surement …
Els	1
Ens	Oui, la somme est égale à 1. Je vais multiplier cette fréquence par 10. (*l'enseignant montre la fréquence 0,092*). Alors, combien j'obtiendrai ?
Els	0,92
Ens	Bon, 0,92 ; puis …
Els	1,14 ; 1,56, 3,9 ; 1,5, 0,98
Ens	Merci. Et si nous additionnons tous ces nombres, alors … ?
Els	Cela fait 10…
Ens	Oui, la somme est égale à 10. Si j'arrondis ce nombre ? (*l'enseignant montre 1,14*)
Els	1 ; 1 ; 2 ; 4 ; <u>1 … mais 2</u> …[145] (Certains élèves disent 2, et les autres disent 1)
Ens	1 ou 2 ? Dites-moi, combien vous voulez ? 1 ou 2 ?

	(*l'enseignant s'arrête dans le cas d'arrondi de la fréquence 1,5 de la note 9 pour une discussion publique*)

[145] Souligné par nous.

Els	1 (*les élèves parlent très fort*)
Autres	2 (*ces élèves parlent plus faiblement*)

(Script de la séance 3)

Les élèves défendent-ils un point de vue sur l'arrondi entier d'un nombre décimal ou le résultat de l'échantillon qu'ils ont écrit sur leur fiche 8 ?

Ens	Attendez. Pourquoi voulez-vous 1 ? Pourquoi l'arrondi de 1,5 vous donne 1 ? Selon le principe d'arrondi, il doit être combien ?
Els	Mais la somme fait 11.
Ens	Ah, d'accord. Si je suis le principe d'arrondi d'un nombre réel au dixième, je dois augmenter ce nombre en 2 ici, n'est-ce pas ? Mais la somme fait 11.

	Alors, que dois-je faire ? Ce nombre 1,5 devient 2 après arrondi et 1,56 aussi, il est devenu 2 par l'arrondi. Bon, quel est le nombre[146] que l'on va garder ?
Els	1,56 !
Ens	Ah, vous décidez de garder 1,56 et vous voulez que ce nombre soit réduit à 1 ? (*l'enseignant montre le nombre 2 et fait une flèche de déduction à 1*)

[146] Soit 1,56 soit 1,5.

	Et grâce à ça, cette somme recule à …
Els	10
Ens	Oui, à 10. Je vois que pour ces deux notes 7 et 9, leurs fréquences 1,56 et 1,5 sont très proches. Comme on a 50 fois un échantillon de 10 notes, c'est-à-dire 500 notes. Ces 500 notes nous donnent deux pour la note 7 et un pour la note 9. On ne sait pas si en refaisant la commande n°1 avec 500 autres notes, on obtiendrait de nouveau « deux notes 7 - une note 9 » ou peut être « une note 7 - deux notes 9 » dans l'échantillon représentatif ? Il semble qu'il y a une hésitation entre le choix de « réduire le nombre de notes 9 » et celui de « réduire le nombre de notes 7 ». Par exemple, dans ce tableau, si on redonne 1 pour la note 7 et 2 pour la note 9, est-ce que l'on prendra un risque de se tromper ? C'est le problème posé.

(Script de la séance 3)

Ainsi, à la fin de ce travail en public, l'enseignant ne conclue pas et introduit la notion de risque à prendre pour décider d'un échantillon représentatif via l'incertitude entre les types (2) et (1). L'évaluation de ce risque est l'enjeu de la situation qui suit.

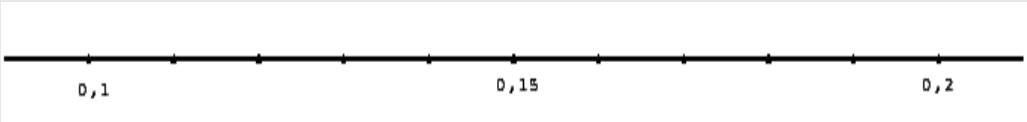
III.3.3 Situation-notes : les intervalles de confiance sur une fréquence et le calcul du risque pris à choisir une fréquence empirique comme fréquence théorique.

Protocole

L'enseignant présente la commande n°2 :

> Utiliser la commande n°2 pour obtenir 20 fréquences de la valeur choisie (7 ou 9) ; compléter le graphique de la fiche 9 et décider du nombre de notes de cette valeur dans l'échantillon représentatif.

Voici la fiche 9 :

> **Fiche 9** (noms, prénoms, lycée, classe, n° groupe)
>
> Placez sur la droite graduée les 20 fréquences de la valeur choisie (7 ou 9) :
>
> 
>
>
> Combien de notes de cette valeur mettez-vous dans l'échantillon représentatif ? ☐

L'enseignant ramasse les fiches et fait lui-même publiquement le travail en dégageant la notion d'intervalle de confiance au niveau 95 % (il faut écarter 5 fréquences sur 100 ou 1 fréquence sur 20).

Analyse a priori

Dans la situation-notes, contrairement à la situation-bouteille on ne peut pas calculer d'intervalle de décision ni, donc, de risque attaché à un tel intervalle attribué après p expériences de n tirages. En effet, par comparaison avec la situation-bouteille des séances 1 et 2, le nombre de notes est fini mais inaccessible, ce qui empêche toute construction a priori des distributions possibles.

Même si le nombre de boules[147] était connu, quand le nombre de boules (de notes) augmente même avec deux couleurs (valeurs), le nombre de distributions théoriques possibles augmente ainsi que le nombre d'intervalles de décision, chacun de ces intervalles diminuant en largeur. La conséquence est une augmentation du risque[148].

Dans la situation-bouteille il est possible de comparer les fréquences empiriques aux *fréquences théoriques connues* (connues parce qu'elles correspondent aux différentes compositions possibles de bouteille qui sont calculables) donc de faire, pour chacune des deux couleurs, *une prévision de fréquence théorique* ; on utilise l'intervalle de décision attaché à cette fréquence théorique pour évaluer empiriquement le risque inhérent à une telle prévision c'est-à-dire la proportion d'expériences de n tirages qui conduiraient à une autre fréquence théorique car la fréquence observée sortirait de l'intervalle de décision. Ce risque n'est pas décidé par l'expérimentateur ; il est issu du résultat de l'expérimentation.

Dans la situation-notes *les fréquences théorique étant inconnues* (inconnues car on ne peut établir une liste des différentes distributions possibles), la comparaison des fréquences entre elles pour les p expériences de 500 tirages permet, pour chacune des notes, *de faire une prévision de fréquence théorique seulement à partir de l'intervalle de fluctuation et de construire un intervalle de confiance*. Cet intervalle de confiance est construit empiriquement sur une statistique : par exemple on ne retient, pour une valeur de note, qu'un intervalle strictement inclus dans l'intervalle de fluctuation et qui contient un certain pourcentage des p fréquences de cette valeur ; à cet intervalle est donc attaché, pour des expériences de 500 tirages, un risque de se tromper qui est la proportion d'expériences de 500 tirages dont la fréquence observée sort de cet intervalle. Contrairement à la situation-bouteille le seuil de risque est choisi par l'expérimentateur.

L'intervalle de confiance est calculable théoriquement à partir d'une seule expérience de 500 tirages sans avoir à la répéter : avec la distribution choisie, notons f la fréquence empirique de 7 (ou 9) obtenue après 500 tirages au hasard, la théorie de l'échantillonnage basée sur le

[147] Dans la séance 3, ce ne sont plus des boules de certaines couleurs mais des notes de certaines valeurs. Cependant quelquefois nous emploierons métaphoriquement le mot boules à la place du mot notes.

[148] Par exemple avec 10 boules l'intervalle de décision attaché à la composition 6R/4V est [0,55 ; 0,65] et le risque sur 50 tirages de se trouver hors de cet intervalle est égal à 0,542 au lieu de 0,194 pour l'intervalle de décision [0,5 ; 0,7] attaché à la composition 3R-2V sur 5 boules.

théorème « central-limite » fournit l'intervalle $\left[f - 1{,}96\sqrt{\frac{f(1-f)}{500}} \; ; \; f + 1{,}96\sqrt{\frac{f(1-f)}{500}} \right]$ comme intervalle de confiance possible à 95 % pour la proportion de 7 dans la population (pour f proche de 0,15 cela donne 0,15 ± 0,03).

Soit p la proportion de notes égales à 7 dans la population des notes du rectorat. Faisons 500 tirages avec remise et notons f la fréquence de la valeur 7 dans cette expérience ; 500 x f est une variable aléatoire qui suit une loi binomiale de paramètres 500 et p ; on est dans les conditions où cette loi peut être approchée par la loi normale de paramètres 500p et $\sqrt{500p(1-p)}$; donc f suit une loi normale de paramètre p et $\sqrt{\frac{p(1-p)}{500}}$; et $\frac{f-p}{\sqrt{\frac{p(1-p)}{500}}}$ est une variable aléatoire qui suit la loi normale centrée réduite.

Nous avons choisi p = 0,14 : la probabilité que f soit inférieure à 0,15 est donc égale à la probabilité que la variable normale centrée réduite soit inférieure à $\frac{0{,}01}{\sqrt{0{,}14 \text{ x } \frac{0{,}86}{500}}} \approx 0{,}6444$ et par calcul ou lecture de table on trouve la probabilité égale à 0,74. On comprend donc que en faisant 500 tirages, la fréquence de la valeur 7 apparaisse 3 fois sur 4 inférieure à 0,15 et 1 fois sur 4 supérieure à 0,15. Sur 20 expériences de 500 tirages, en moyenne, 15 d'entre elles devraient fournir une fréquence de 7 inférieure à 0,15 et 5 d'entre elles une fréquence de 7 supérieure à 0,15.

Pourquoi avoir choisi 0,14 ? Rappelons que nous attendons une décision sur le nombre de notes 7 dans l'échantillon représentatif. L'avantage d'éloigner p de 0,15 est de rendre la décision moins risquée ; le désavantage est de rencontrer moins souvent d'expériences de 500 tirages qui donnent une fréquence empirique plus grande que 0,15. Or on voudrait en rencontrer plusieurs pour justifier la répétition de 500 tirages par la commande n° 2 (partie 2 de la deuxième phase). Nous avons choisi 0,14 comme un bon compromis.

Les stratégies

Stratégie « *évaluation de la fréquence théorique* »

- Sous-stratégie « *fréquence la plus fréquente* » : retenir comme fréquence de la note choisie celle qui, au centième près, est la plus fréquente.

Cette stratégie s'appuie sur le regroupement en classes égales d'amplitude 0,01 et la fréquence théorique est égale la fréquence empirique modale.

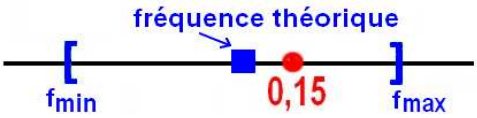
Classes des fréquences
[0,095; 0,105 [
[0,105; 0,115 [
[0,115; 0,125 [
[0,125; 0,135 [
[0,135; 0,145[
[0,145; 0,155 [
[0,155; 0,165 [
[0,165; 0,175 [
[0,175; 0,185 [
[0,185; 0,195 [
[0,195; 0,205 [

Exemple :

Après avoir arrondi au centième près, 0,14 est la plus fréquente alors choisir 0,14 pour la fréquence théorique de la note choisie et proposer une note 7 dans l'échantillon représentatif.

Fréquences	Fréquences arrondies
0,130	0,13
0,136	**0,14**
0,138	**0,14**
0,140	**0,14**
0,140	**0,14**
0,141	**0,14**
0,141	**0,14**
0,142	**0,14**
0,143	**0,14**
0,144	**0,14**
0,144	**0,14**
0,151	0,15
0,152	0,15
0,153	0,15
0,162	0,16
0,182	0,18

- Sous-stratégie « *fréquence centrale* » : construire l'intervalle de fluctuation des fréquences de la note choisie, considérer la valeur centrale de cet intervalle comme la fréquence théorique puis la comparer à 0,15. Dans le cas où la fréquence centrale est inférieure à 0,15 (resp. supérieure à 0,15), choisir l'effectif de la note égal à 1 (resp. 2).



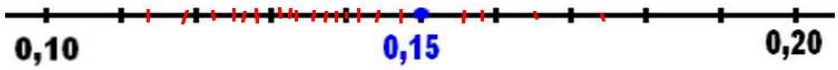
- Sous-stratégie « *fréquence moyenne* » : arrondir la fréquence moyenne des 20 fréquences de la note choisie et qui est affichée à l'écran puis proposer l'effectif de cette note. La fréquence moyenne est implicitement assimilée à la fréquence de la note choisie pour un échantillon de taille égale à 10 000.

La stratégie « *évaluation de la fréquence théorique* » est conforme au rapport institutionnel du traitement de données statistiques avec les objets classes, centre de classe, mode et moyenne. De plus elle n'est pas très coûteuse en calculs surtout la sous-stratégie « fréquence moyenne » puisque le résultat est affiché à l'écran[149].

Stratégie « *majorité des points* »

Cette stratégie est provoquée par le travail graphique proposé dans la fiche 9. Ce travail graphique favorise le comptage des points et l'appréciation de la densité des points à gauche et à droite de 0,15. Elle consiste à réaliser une seule fois de la commande n°2 et à choisir l'effectif égal à 1 pour la note s'il y a plus de points à gauche de 0,15 qu'à droite ; dans le cas opposé, choisir l'effectif égal à 2.

Exemple :



Il y a 16 points à gauche de 0,15 (contre 4 points à droite), choisir alors un pour la note choisie.

Cette stratégie est également une stratégie de statistiques puisqu'il s'agit de placer la médiane par rapport à 0,15 [150]. Elle peut déboucher sur une appréciation probabiliste (probabilité expérimentale) du risque pris dans la décision. Dans l'exemple présenté ci-dessus, il y a 4 points à droite de 0,15 sur les 20 points, donc on peut dire qu'il y un risque de un sur cinq de se tromper en décidant un pour la note choisie.

Déroulement et analyse a posteriori :

Tous les groupes prennent répondent 1 note 7 alors que seuls quatre groupes avaient l'échantillon de type (1) dans la situation précédente où seule la commande n°1 était disponible.

Tableau de distribution des stratégies empruntées par les groupes :

---

[149] Le programme est écrit de façon que les 20 fréquences et la fréquence moyenne sont affichées.

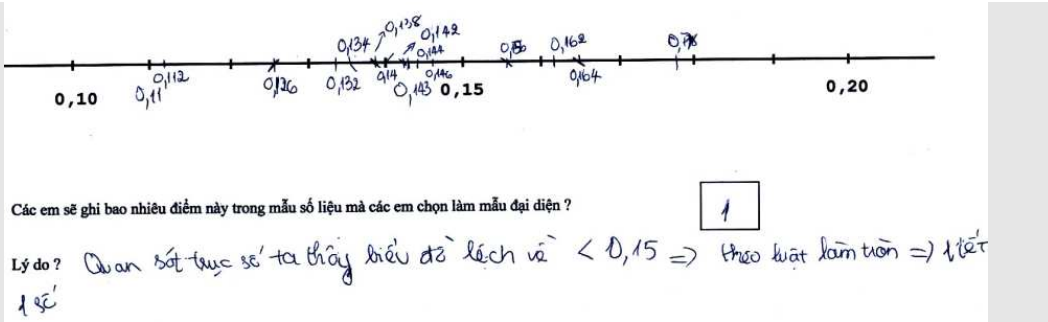
[150] Dans l'institution vietnamienne, la médiane est effectivement à enseigner mais la signification que l'institution lui donne est celle d'être le représentant seulement quand il y a des valeurs extrêmes dans les données.

Stratégie « *évaluation de la fréquence théorique* »			Stratégie « *majorité des points* »
Sous-stratégie « *fréquence la plus fréquente* »	Sous-stratégie « *fréquence centrale* »	Sous-stratégie « *fréquence moyenne* »	
Gr1	0	Gr3, Gr5	Gr1, Gr2, Gr4, **Gr6, Gr7**, Gr8

Tableau 91. Distribution des stratégies

Les groupes 6 et 7 n'exécutent qu'une fois la commande n°2 et s'appuient sur le graphique pour conclure (stratégie « majorité de points »). Les autres groupes exécutent plusieurs fois la commande n°2 et ne retiennent pour le graphique qu'une seule des distributions obtenues.

- le groupe 7 :



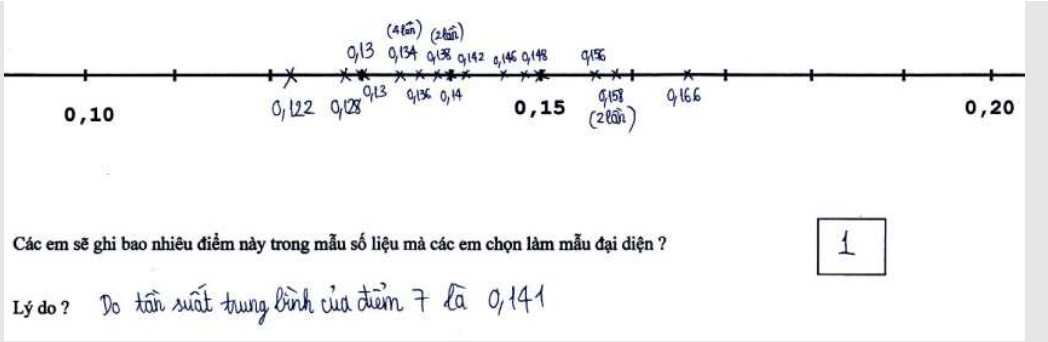
Traduction en français

Raison : En observant l'axe numérique, nous constatons que le graphique est scalène[151] à < 0,15

⇒ selon le principe d'arrondir ⇒ écrire 1 note.

(Fiche 9 du groupe 7)

- Les groupes 5 et 3 prennent la moyenne des fréquences comme fréquence théorique :



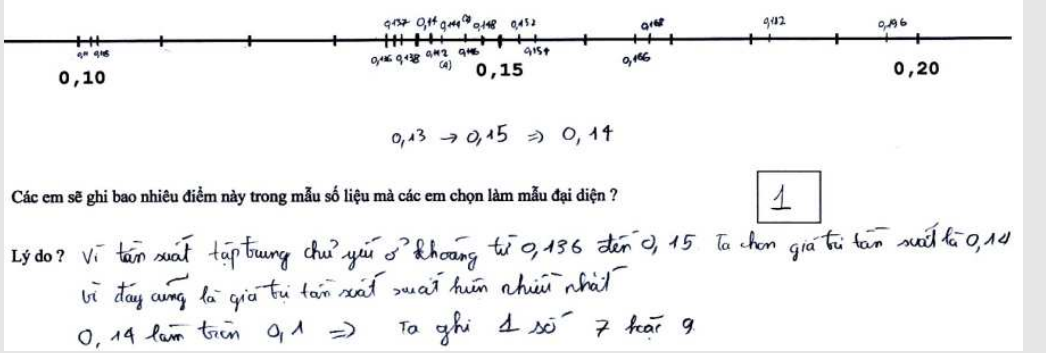
Traduction en français

Raison : Car la moyenne des fréquences est 0,141

(Fiche 9 du groupe 5)

[151] Le dictionnaire franco-vietnamien suggère deux traductions pour le mot « lêch » : de travers et scalène qui toutes les deux renvoient à l'image d'une dissymétrie.

- Le groupe 1 utilise un mixte des deux stratégies en choisissant la valeur 0,14 comme la fréquence la plus fréquente, en argumentant sur la concentration des points représentant les fréquences à gauche de 0,15.



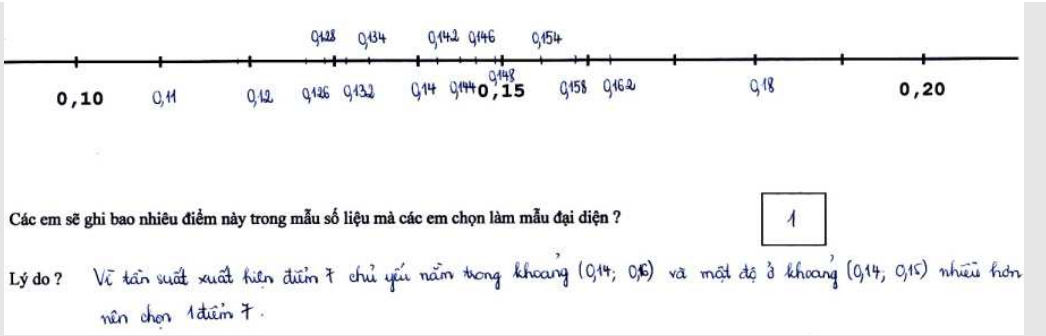
Traduction en français

Raison : Car les fréquences se concentrent <u>essentiellement</u>[152] dans l'intervalle de 0,136 à 0,15, nous choisissons la valeur de 0,14 pour la fréquence comme elle apparait le plus souvent

0,14 est arrondi à 0,1 ⇒ Nous écrivons 1 note 7 ou 9.

(Fiche 9 du groupe 1)

- Le groupe 2 emploie l'expression « densité forte » dans l'interprétation graphique :



Traduction en français

Raison : Car les fréquences de note 7 sont <u>essentiellement</u>[153] dans l'intervalle (0,14 ; 0,16) et la densité dans l'intervalle (0,14 ; 0,15) est plus grande, c'est pourquoi nous choisissons 1 note 7.

(Fiche 9 du groupe 2)

L'utilisation du mot « essentiellement » par les deux groupes montre qu'ils ont repéré que la fréquence empirique pour une expérience peut ne pas être dans l'intervalle qu'ils retiennent pour justifier leur décision. Cet intervalle, [0,136 ; 0,15] pour le groupe 1 et [0,14 ; 0,15] pour le groupe 2, n'est pas conçu comme un intervalle de confiance mais comme un intervalle de décision. Ainsi même si la fluctuation des fréquences pour les échantillons de taille 500 est prise en compte grâce à l'élaboration du graphique, elle ne conduit ni à

[152] C'est nous qui soulignons.

[153] C'est nous qui soulignons

formuler ni à mesurer une incertitude dans la décision prise. Nous faisons l'hypothèse que les groupes répondent à la question posée sans se reporter en arrière lorsque seule la commande n°1 était disponible et lorsque la décision pouvait être prise à l'issue d'une seule expérience. Ainsi la proportion de notes hors de l'intervalle qu'ils choisissent ne leur apparaît pas comme la probabilité d'une possible mauvaise décision[154] à l'issue d'une seule expérience.

L'inférence pratiquée reste prisonnière du cadre déterministe compatible avec la loi des grands nombres qui est comprise comme une convergence fonctionnelle des fréquences empiriques vers la fréquence théorique.

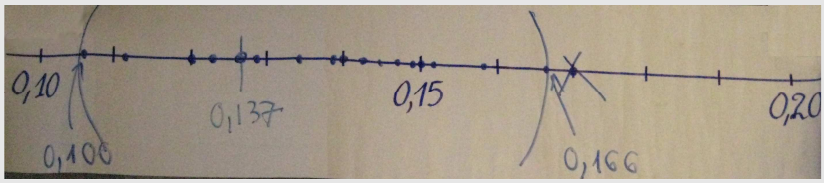
Intervention publique de l'enseignant

Cette intervention a pour objectif d'introduire la notion d'intervalle de confiance et celle de risque dans la décision. Pour cela il propose à la classe de choisir une note dont on observera les fréquences en exécutant la commande n°2. La note choisie est la note 7. L'enseignant affiche 20 fréquences sur un graphique préparé à l'avance.

Ens Maintenant, je vais placer 20 points pour 20 fréquences obtenues sur cette affiche. Pouvez-vous me lire les fréquences affichées à l'écran ?

Els 0,106… 0,112… 0,117… 0,120… 0,122… 0,126… 0,128… 0,128… 0,134… 0,138… 0,140… 0,142… 0,144… 0,146… 0,148… 0,150… 0,150… 0,152… 0,158… 0,166.

Ens Bon, regardez ce graphique, d'après vous, on va prendre combien pour la note 7 ?



Els U…n…
(*La classe crie très fort*)

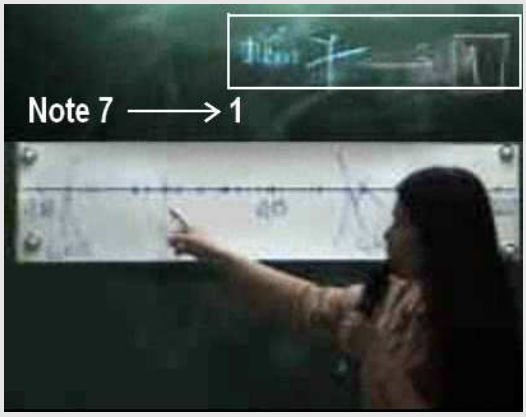
Ens Pourquoi vous choisissez 1 ?

Els Parce qu'il y a beaucoup de points à gauche de 0,15.

(Script de la séance 3)

[154] Rappelons que tous les groupes affichent le type (1) dans leur réponse alors que la moitié d'entre eux avaient choisi le type (2) pour leur réponse précédente.

La classe exprime ici son accord avec la procédure « majorité de points » mais l'enseignant ne valide pas cette procédure bien qu'il pouvait s'en servir pour estimer empiriquement le risque pris à choisir 1 en exploitant immédiatement la possibilité de compter le nombre de points à droite de 0,15. En fait l'enseignant n'a pas les moyens de présenter à la classe une justification mathématique. Une justification aurait-elle pu se passer d'un modèle théorique de la distribution des fréquences empiriques emportant la symétrie par rapport à la fréquence théorique[155] ? Pourtant, l'enseignant décide de mettre l'accent sur la fréquence moyenne :

Ens	Et si je regarde la moyenne des fréquences, elle est combien ?
Els	0,137
Ens	Bon, je vais placer un point sur cet axe pour la moyenne. (*l'enseignant marque un point pour 0,137*) On peut constater que les « points-fréquences » se trouvent autour de cette moyenne et ils forment un intervalle…, la valeur la plus petite des fréquences, combien ? 0,106 … et la valeur la plus grande ? 0,166. Bon, on a l'intervalle]0,106 ; 0,166[. Et cet intervalle contient la valeur moyenne 0,137. La plupart des points se trouvent où ? 
Els	A gauche [156]…

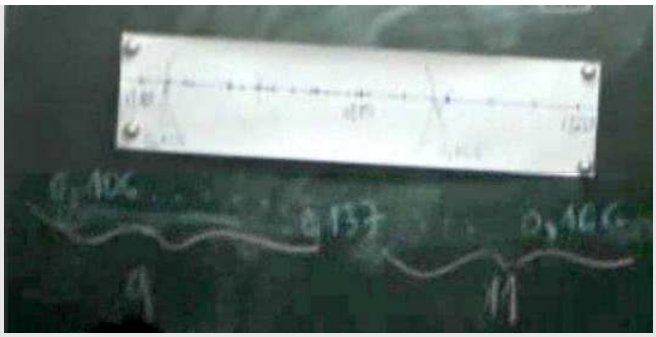
(Script de la séance 3)

L'enseignant réemploie les intervalles de fluctuation de fréquences mis en place dans la séance 2.

Ens	Maintenant, si je veux enlever certaines fréquences, par exemple 5 fréquences, alors, lesquelles dois-je enlever ? Bon, si je veux comparer ces fréquences à enlever avec la fréquence moyenne 0,137, où se trouvent-elle ? on va compter le nombre des fréquences à gauche et à droite de la fréquence moyenne. Combien à gauche ? (*l'enseignant fait compter la classe grâce à l'affichage sur l'écran*)
Els	… un, deux, trois, … neuf … Il y a neuf fréquences à gauche de 0,137.
Ens	Bon, il reste donc … 11 fréquences à droite.

[155] Le modèle théorique habituellement retenu est celui de la loi normale.

[156] Les élèves veulent dire à gauche de 0,15.



	Alors, les fréquences que vous voulez enlever où se trouvent-elles ?
Els	A droite…
Ens	Pourquoi ?
Els	Car il y a plus de points à droite qu'à gauche…

(Script de la séance 3)

De cet ensemble, après l'avoir ordonné et transformé en intervalle, l'enseignant propose d'enlever 5 fréquences avec l'intention de former un nouvel intervalle plus petit qui contienne la fréquence théorique avec une certaine chance égale à 0,75. Mais l'enseignant cherche à construire cet intervalle de confiance en le centrant sur la fréquence empirique moyenne.

Les élèves, ignorants de l'intention didactique, veulent alors écarter les fréquences à droite de 0,15 puisque l'enjeu était pour eux de choisir entre une fréquence théorique inférieure à 0,15 ou supérieure à 0,15. Influencé par les interventions des élèves, l'enseignant abandonne la centration de l'intervalle sur la fréquence moyenne et exhibe un intervalle de confiance au niveau 75 %.

Ens	Bon, le nouvel intervalle [0,106 ; 0,148] est plus concentré sur 0,137. Comme on a enlevé 5 fréquences parmi les 20 fréquences, on peut dire que l'on a une certaine erreur. Si on enlève 5 valeurs parmi 100 valeurs, on dit que l'on a une confiance de 95% dans cet intervalle.

(Script de la séance 3)

L'enseignant maintient publiquement la référence à 0,137 qui est la fréquence empirique moyenne alors que cette valeur n'est pas intervenue dans la confection de l'intervalle retenu.

Notre interprétation de cet épisode de classe est que l'enseignant a négocié à la baisse pour imposer le savoir visé : il a remplacé l'intervalle de confiance théorique[157] par un intervalle de confiance empirique dépendant des données de l'expérience réalisée publiquement. Cet intervalle a en fait la signification d'un intervalle de décision. Mais l'enseignant peut-il faire autrement en l'absence institutionnelle de toute technologie probabiliste ?

L'intervalle de confiance accepté par la classe permet à l'enseignant de revenir sur la notion de risque dans la décision à prendre :

[157] Partiellement théorique puisque la largeur de l'intervalle n'aurait pas été choisie en utilisant la distribution normale justifiée par le théorème central-limit. Seul le centrage sur la fréquence moyenne nécessitait une justification théorique.

Ens	[…] Les cinq fréquences enlevées sont celles à partir de 0,150. Je dis que je prends un certain risque en décidant de choisir un pour la note 7.

(Script de la séance 3)

Mais il ne peut pas associer la valeur du risque à la probabilité expérimentale que fournit l'échantillon de 20 fois 500 tirages. Il revient par contre sur le gain de confiance que donne une augmentation du nombre de tirages :

	[…] Vous voyez, dans ce travail, la commande n°2 exécute 20 fois la commande n°1 qui prend 500 notes, c'est-à-dire nous avons un ensemble de 20 fois 500 notes c'est-à-dire 10 000 notes. Et nous avons choisi une seule note 7 au lieu de deux notes 7 comme dans le travail tout à l'heure. Est-ce que plus on prend un grand nombre des notes, plus on a une fréquence plus exacte ?

(Script de la séance 3)

De nouveau le point de vue fonctionnel sur la convergence des fréquences empiriques vers la fréquence théorique prend le pas sur le point de vue probabiliste de la fluctuation d'échantillonnage.

III.4. Conclusion

La situation a rendu nécessaire l'examen de l'ensemble des fréquences et a permis à l'intervalle de fluctuation des fréquences de servir implicitement ou explicitement d'intervalle de décision pour la grande majorité des groupes. Ainsi la fluctuation des fréquences peut servir d'outil à la prise de décision. Mais l'utilisation de cet outil ne peut à lui seul conduire à faire germer la notion de risque car l'intervalle de confiance ne peut se construire seulement à partir de l'intervalle de décision.

L'enseignant ne parvient pas à s'appuyer sur la notion de fluctuation d'échantillonnage autrement qu'empiriquement : du coup la notion de risque qu'il s'efforce de présenter apparaît comme un coup de force. Il faut dire que non seulement la notion de fluctuation d'échantillonnage est absente du rapport institutionnel à la statistique mais également la notion de modélisation probabiliste par une loi de probabilités si bien qu'aucune modélisation probabiliste de cette fluctuation ne pouvait être mobilisée.

Il était attendu que la notion apparaisse dans la dernière situation avec la signification d'une probabilité expérimentale grâce à l'utilisation de la commande n°2 qui exécute la répétition de la commande n°1. Mais l'absence de retour à l'expérience unique (celle de la commande n°1) n'a pas permis que les résultats de la commande n°2 soient réinterprétés comme une probabilité (expérimentale) dans le cas de l'usage unique de la commande n°1. Nous faisons l'hypothèse qu'un tel retour est nécessaire pour toute utilisation de l'approche fréquentiste dès lors qu'il s'agit de faire le lien entre les fréquences empiriques obtenues dans la répétition d'une expérience et la probabilité que l'on veut induire.

Conclusions et perspectives

Conclusion générale, portée et limites de l'étude didactique

I. Ce qui a motivé l'étude didactique de la thèse

Une première étude didactique[158] sur l'introduction des probabilités dans l'enseignement secondaire vietnamien nous a conduits vers de nouvelles questions suscitées par l'intérêt que nous portions à la solidarité épistémologique entre la statistique et les probabilités. Postulant que « dans les liens que l'enseignement mathématique pourrait tisser entre les deux domaines, celui de la statistique et celui des probabilités, se jouent la nature et la portée des significations pour les différents concepts que les élèves et les enseignants construiront ensemble tant dans l'un des domaines que dans l'autre »[159] et constatant que l'introduction les probabilités se faisait en classe 11 alors que l'enseignement de la statistique prenait fin en classe 10, constatant aussi que l'enseignement de la statistique était circonscrit à la statistique descriptive, la question qui se posait pour nous était alors :

Comment la solidarité épistémologique entre statistique et probabilités peut-elle vivre dans EMS vietnamien alors que statistique et probabilités vont se côtoyer sans se confronter ? Cette question se décline dans le travail de la thèse sous deux formes :

- comment la solidarité épistémologique entre statistique et probabilités fonctionne-t-elle dans EMS vietnamien actuel ?

- à quelles conditions l'EMS vietnamien actuel peut-il renforcer la solidarité épistémologique entre statistique et probabilités ?

Nous avons travaillé ce double questionnement dans un cadre méthodologique qui fait appel à la fois à la Théorie Anthropologique du Didactique et à la Théorie des Situations Didactiques. Ce cadre nous a permis de croiser deux types d'analyses qui s'appuient l'une sur l'autre : l'analyse épistémologique et l'analyse institutionnelle. Nous avons mené ces analyses dans différents dispositifs d'étude : un repérage des praxéologies dans les programmes et les manuels de l'EMS puis dans des manuels universitaires, un questionnaire aux élèves de la classe 10 et à leurs enseignants, une ingénierie didactique en classe 11. Chacun de ces dispositifs a contribué à étoffer les deux analyses et à les articuler entre elles pour fournir des éléments de réponse au questionnement initial.

II. Principaux apports du travail de la thèse

II.1. L'existence de trois problèmes-candidats

Il s'agit de problèmes susceptibles de faire vivre des liens statistique-probabilités dans une institution d'enseignement des mathématiques. Ces problèmes-candidats ont été mis en évidence par le premier dispositif d'étude.

---

[158] Vu Nhu [2005]

[159] Ce postulat a été présenté dans notre introduction.

Le problème de l'incertitude des mesures : ce problème qui est à l'origine historique de la loi normale est le lieu d'une rencontre possible avec la loi des grands nombres et à la distribution d'échantillonnage de la moyenne ;

Le problème de la comparaison des distributions statistiques empiriques entre elles ou avec des distributions théoriques : ce problème est le lieu d'un rapprochement possible des objets communs à la statistique descriptive et à la théorie des probabilités telles les notions de variable (statistique ou aléatoire), les notions de distributions (statistiques ou de probabilités), etc.

Le problème de la représentativité d'un échantillon issu d'une population : ce problème qui est à l'origine de l'inférence statistique et de la théorie des tests est le lieu d'une rencontre possible avec la loi des grands nombres et la distribution d'échantillonnage.

Pour tisser les liens statistique-probabilités, chacun de ces problèmes a besoin d'objets qui sont autant d'objets-candidats : les distributions de fréquences ou de probabilités avec leurs ostensifs graphiques, tabulaires, numériques et algébriques, la loi des grands nombres, l'échantillonnage.

II.2. Une carte praxéologique de l'EMS VN en statistique

- **une praxéologie dominante** : celle de description d'une série statistique imbriquant les praxéologies locales de description graphique et numériques ; cette praxéologie développe les tâches et les techniques pour simplifier, résumer et synthétiser une série statistique ; elle vise la comparaison de chaque individu à l'ensemble des individus ; les éléments technologiques sont les tableaux, les graphiques et les vecteurs d'indicateurs numériques.

- **une praxéologie incomplète :** celle de comparaison de séries statistiques imbriquant les praxéologies locales de comparaison graphiques ou numériques ; cette praxéologie vise la comparaison entre sous-populations d'une même population ou entre une sous-population et sa population-mère

- **une praxéologie absente :** celle d'échantillonnage dont les types de tâche seraient la production d'échantillons par répétition d'une expérience aléatoire, l'étude de la fluctuation d'échantillonnage soit sur une fréquence particulière soit sur un indicateur comme la moyenne, l'étude de la réduction de la fluctuation par augmentation de la taille de l'échantillon.

Ce complexe praxéologique permet une vie de la statistique dans EMS VN séparée de celle des probabilités, le seul lien tissé étant la relation fréquence-probabilité dans la définition théorique mais non opératoire de la probabilité d'un événement que l'on trouve dans l'approche fréquentiste de la probabilité (en classe 11). Ce lien est très ténu car il ne donne lieu à aucune construction praxéologique sur la loi des grands nombres ni sur la fluctuation d'échantillonnage. Son renforcement sera l'objet principal de l'ingénierie didactique.

L'EMS VN ne prend en charge qu'un seul des trois problèmes-candidats, celui de la comparaison mais l'absence de modèles probabilistes discrets ou continus empêche de pousser cette comparaison au-delà d'une comparaison graphique ou d'une comparaison numérique entre distributions empiriques.

II.3. Une compréhension des rapports institutionnels aux objets-candidats

Les seuls objets-candidats présents dans l'institution sont les distributions statistiques avec leurs ostensifs graphiques et numériques.

- **Les graphiques**

Nous sommes partis de l'hypothèse de recherche HR1 formulée à l'issue de la comparaison des praxéologies à enseigner et des praxéologies de référence :

HR1 : l'existence d'un point de vue fonctionnel sur les graphiques statistiques est une condition favorable à la construction de liens statistique-probabilités au sein d'une praxéologie de comparaison des distributions statistiques

Et avons travaillé la question QR5 :

QR5 : au sein de la praxéologie de comparaison dans l'institution vietnamienne, existe-t-il un point de vue fonctionnel sur les graphiques statistiques ?

Nous sommes arrivés à la conclusion suivante : le point de vue fonctionnel n'est pas absent de la production et la lecture des graphiques mais il s'érige dans l'institution en *obstacle didactique à la construction de liens statistique-probabilité* dans la mesure où de tels liens font appel aux notions d'aire de l'histogramme et d'aire sous la courbe (de densité) alors que cette notion est ignorée de l'institution. Cet obstacle est renforcé par l'absence institutionnelle de cumulation des fréquences et donc de la fonction de répartition.

- **Les résumés numériques**

En tant que représentant d'une série statistique, un résumé numérique est un outil de comparaison entre un échantillon et la population-mère. Le dispositif expérimental des exercices 3 et 4 du questionnaire a permis de renforcer les résultats de l'étude des manuels : le représentant dominant est la moyenne ou, à défaut, le couple moyenne/écart-type et il est un critère très fort de la conservation de la distribution entre échantillon et population. L'expérimentation a cependant mis en lumière la disponibilité de critères de quasi-proportionnalité pour juger de la représentativité d'un échantillon par rapport à la population.

- **Une praxéologie d'échantillonnage à construire**

Le recours à l'échantillonnage aléatoire simple dans la production de statistiques sur une population inconnue pour répondre à des questions sur cette population ne fait pas l'objet d'un enseignement intégré au domaine de la statistique. Seule est présente, en classe 11 donc dans l'enseignement des probabilités, la notion de probabilité dite expérimentale obtenue par répétition d'une expérience aléatoire et considérée comme valeur approchée d'une probabilité inconnue. Cette restriction de l'inférence statistique à l'estimation ponctuelle d'un nombre (une probabilité) par un autre (une fréquence) enferme l'inférence dans le cadre de l'approximation numérique et d'une vision fonctionnelle de la loi des grands nombres - la probabilité expérimentale est une fonction du nombre de tirages dont la limite, à l'infini, est la probabilité vraie[160]. L'ingénierie didactique va montrer que cette *vision*

[160] Que nous appelons fréquence théorique dans le cas où l'expérience aléatoire est de type Bernoulli.

fonctionnelle de la loi des grands nombres se constitue en obstacle épistémologique de la vision probabiliste[161] parce qu'elle ne rend pas nécessaire la mesure, même expérimentale, du risque pris dans la décision de retenir telle ou telle probabilité.

II.4. Une ingénierie didactique

L'ingénierie, point d'orgue du dispositif expérimental, a permis de revenir sur les analyses épistémologique et institutionnelle et de les compléter. Son fil conducteur est une transformation progressive de la loi des grands nombres pour en faire un outil de décision avec conscience et mesure expérimentale du risque dans un problème d'incertitude aléatoire. Cette transformation s'appuie sur la comparaison d'échantillons en mettant en scène la fluctuation d'échantillonnage ; elle touche donc, outre la loi elle-même, la notion d'échantillon et celle de degré de certitude dans la décision :

Objets / Séance	La loi des grands nombres	La fluctuation d'échantillonnage	Le risque
Séance 1	La loi des grands nombres est en acte dans le choix de la fréquence théorique.	La fluctuation d'échantillonnage est un phénomène naturel incontrôlable et non mesuré.	Le degré de certitude : - n'est relié à aucun facteur (le nombre de tirages ou l'écart des fréquences empiriques à la fréquence théorique par exemple) - n'est pas exprimé dans le langage probabiliste
Séance 2	La loi des grands nombres est un objet avec deux attributs graphiques : - une asymptote horizontale pour la courbe représentative de la fréquence considérée comme fonction du nombre de tirages - un emboîtement d'intervalles numériques sur la droite réelle.	La fluctuation d'échantillonnage est mesurée : l'intervalle de fluctuation des fréquences empiriques est réduit jusqu'à un intervalle de décision théorique (avec certitude).	Le risque devient nul à partir d'un nombre de tirages fourni par l'expérience
Séance 3	La loi des grands nombres est un outil d'évaluation d'un risque pour une décision.	La fluctuation d'échantillonnage est mesurée : l'intervalle de fluctuation des fréquences empiriques devient un intervalle de confiance empirique (avec risque).	Le risque est la probabilité expérimentale de prendre une mauvaise décision

Tableau 92. Evolution de trois objets au cours des trois séances de l'ingénierie didactique

[161] Nous avons parlé d'obstacle « fonctionnel »

La séance 3 construit la transformation de l'intervalle de décision (théorique) en un intervalle de confiance (empirique). Cette transformation est nécessaire pour faire émerger la notion de risque dans l'inférence statistique.

L'expérimentation de l'ingénierie dans une classe d'un lycée vietnamien a porté une modification du rapport personnel des élèves à l'objet échantillon et ainsi montré une *modification possible du rapport institutionnel à la notion d'échantillon* :

- d'une part, un échantillon est le produit de la répétition d'une expérience aléatoire et n'est pas seulement un extrait d'une population mère avec le sens d'un sous-ensemble de la population ;

- d'autre part, un échantillon est un outil de connaissance d'une population inconnue grâce à la variabilité des échantillons de même taille.

L'expérimentation a aussi confirmé l'existence de l'obstacle « fonctionnel » dans la construction d'une signification probabiliste de la loi des grands nombres.

En quelque sorte cette ingénierie se voulait une amorce de praxéologie d'échantillonnage à construire dans l'EMS vietnamien. Son expérimentation met en lumière certaines conditions de viabilité dans l'institution d'une telle praxéologie. Le passage de la seule stabilisation des fréquences au couple variabilité/régularité des échantillons de même taille est l'une de ces conditions et elle oblige, dans la méthode fréquentiste, à concevoir la répétition d'une même expérience aléatoire non pas sous la forme d'une accumulation de tirages unitaires mais sous celle d'un ensemble d'échantillons d'un même nombre de tirages.

III. Limites de l'étude et nouvelles directions de recherche possibles

Toute étude arrête des choix qui l'exposent à des critiques et bornent la portée de ses résultats mais conséquemment initient de nouvelles questions et ouvrent des perspectives à explorer en s'appuyant sur elle. À partir d'une reformulation des choix principaux, nous proposons de nouvelles directions de recherche dans le contexte institutionnel vietnamien.

III.1. Le choix de l'institution

Notre étude a porté sur les notions de statistique désignées par la noosphère vietnamienne comme à enseigner au collège et au lycée. L'institution que nous avons ciblée est donc l'EMS vietnamien pour la statistique en marginalisant l'enseignement des probabilités. En dehors du fait que nous disposions des résultats d'une étude précédente[162] sur l'enseignement des probabilités, nous avions deux raisons pour cela :

- l'enseignement de la statistique devance au Vietnam celui des probabilités et il s'agissait pour nous de réfléchir aux conditions d'une modification des rapports personnels au concept de probabilité pour des élèves vietnamiens ayant déjà reçu un enseignement de statistique ;

[162] Vu Nhu [2005]

- l'EMS français présente des choix curriculaires différents de l'EMS vietnamien en statistique et nous pouvions ainsi nous appuyer sur une analyse comparative des institutions.

Il serait judicieux de prolonger au niveau universitaire ce travail réalisé avec l'EMS en initiant une problématique de l'enseignement de la statistique et des probabilités à ce niveau. Une question de départ pourrait concerner la dynamique de continuité/rupture avec l'EMS. Nous pourrions nous appuyer sur l'analyse des manuels universitaires exposée dans la partie B ainsi que sur l'analyse de manuels universitaires vietnamiens[163].

III.2. Les choix de l'ingénierie

Deux hypothèses de recherche formulées à la fin de l'analyse épistémologique, HR1 et HR2, pouvaient être travaillées dans le cadre de cette méthodologie. Nous les rappelons ici :

HR1 : l'existence d'un point de vue fonctionnel sur les graphiques statistiques est une condition favorable à la construction de liens statistique-probabilités au sein d'une praxéologie de comparaison des distributions statistiques

HR2 : le problème d'obtenir un échantillon représentatif d'une population mère inaccessible oblige à rencontrer et à construire à partir de statistiques des éléments de mathématisation du hasard au sein d'une praxéologie d'échantillonnage.

Pour continuer à travailler HR1, l'ingénierie aurait pu se fixer comme objectif la recherche de conditions pour une complétude de la praxéologie de comparaison. Cela aurait nécessité l'introduction de modèles probabilistes dont celui de la loi normale. Une telle ingénierie nous a paru ne pas pouvoir s'inscrire dans les contraintes institutionnelles de l'EMS vietnamien. Pour travailler HR2, nous disposions de travaux d'ingénierie déjà expérimentés et diffusés dans la recherche en didactique, ceux de Brousseau et Briand. Nous avons donc emprunté les idées-forces de ces deux ingénieries. Nous revenons maintenant sur trois décisions qui, dans cette reprise, permettent d'apprécier la portée des enseignements de notre ingénierie :

- nous avons expérimenté l'ingénierie auprès d'élèves qui étaient non seulement instruits en statistique mais aussi en probabilités ; en faisant ce choix (contre celui d'expérimenter auprès d'élèves n'ayant pas reçu l'enseignement des probabilités) nous favorisions l'utilisation du concept de probabilité et du langage probabiliste dans l'expression de l'incertitude sur ce qui pouvait être inféré des résultats des expériences aléatoires répétées ; ce que nous apprend la séance 1 de l'ingénierie telle qu'elle a été expérimentée, c'est que le concept de probabilité n'est pas, pour ces élèves, un concept opératoire dans ce problème d'incertitude sur la proportion inconnue (dans la bouteille)[164] ;

[163] Nous avons consulté l'ouvrage Nguyen Cao Van et Tran Thai Ninh (2005)

[164] Supposons que nous obtenions 7 rouges et 3 vertes après 10 tirages de la bouteille ; le choix de la proportion R/V dans la bouteille peut être réglé par la comparaison des probabilités a posteriori d'obtenir l'événement 7R/3V pour chacune des compositions possibles.

- nous avons placé la notion d'échantillon représentatif d'une population en amont et en aval de l'ingénierie, en amont par un exercice du questionnaire et en aval en créant la séance 3 qui se démarque des ingénieries de Brousseau et Briand en ce sens que le modèle probabiliste est multinomial avec une population tellement nombreuse que les intervalles de décision ne peuvent être calculés ; cette décision fait suite à celle de renoncer à organiser la situation du pari telle que Brousseau l'a décrite dans Brousseau (2003)[165] ; en effet nous pensions qu'une telle situation ne débouchait pas sur l'apprentissage de la fluctuation d'échantillonnage comme outil de mesure de la probabilité visée ;

- nous avons utilisé une simulation informatique des modèles binomial et multinomial des situations concrètes présentées (la bouteille puis les notes) pour que les élèves aient accès, au moindre coût, à des échantillons aléatoires simples ; une telle simulation n'a pas cours dans l'enseignement mathématique secondaire vietnamien ; cette décision a permis de mettre la fluctuation d'échantillonnage sur le devant de la scène didactique mais elle ne préjuge pas de la viabilité de la simulation informatique dans l'EMS vietnamien. En remplaçant la bouteille par la bouteille-ordinateur un enjeu didactique se dessinait, que nous n'avons pas étudié. Pour le cerner, citons Girard & Henry (2005)[166] à propos de la réforme de l'an 2000 en France :

> Ne négligeons pas l'intérêt de l'outil informatique. Il permet une approche expérimentale des situations de sondage, et si la proportion p est cachée aux élèves, nous avons un outil de résolution de problèmes jouant le même rôle que les calculettes graphiques quand elles tracent des courbes représentatives de fonctions données.
>
> Au moyen d'un habillage éventuellement évocateur d'une situation aléatoire concrète, l'ordinateur est donc implicitement considéré comme un générateur de hasard qui, concrètement programmé, fournit des résultats parfaitement représentatifs des issues que l'expérience concrète évoquée fournirait. Cette assimilation ne va pas de soi pour tous les élèves. L'ordinateur donne au *hasard* des chiffres de 1 à 6 équirépartis, mais *ce n'est pas un vrai dé*, même non pipé. L'équivalence pour être admise suppose implicitement l'idée que la probabilité de sortie de l'as sur le dé est égale à celle de sortie du 1 de l'ordinateur, c'est-à-dire que le concept de probabilité soit installé. L'utilisation de l'ordinateur pour introduire la notion de probabilité, comme représentant une expérience concrète, repose donc sur un cercle vicieux didactique, ou bien sur un contrat didactique préalable impossible à expliciter : c'est bien le hasard du dé qui est représenté par l'ordinateur. C'est à cette condition que la réalisation d'expériences aléatoires exploitant le générateur aléatoire de l'ordinateur peut être assimilée à une *simulation*, terme omniprésent dans les programmes [français]. Il convient donc de cerner plus précisément le statut de la simulation informatique telle qu'est présentée en classe.
>
> (Girard&Henry [2005], p. 151-152)

On comprend que la viabilité du recours à la simulation informatique dans l'EMS vietnamien mérite d'être étudiée. Cela représente une direction de recherche que nous discutons ci-après.

---

[165] La probabilité visée par notre ingénierie comme par celle de Brousseau est la probabilité d'un écart |fréquence empirique - fréquence théorique| inférieur à un certain seuil de risque

[166] In Statistique au lycée (2005) de la Commission Inter-Irem

III.3. Quelques directions de recherche possibles

Les directions de recherche que nous ébauchons ici sous forme de questions se référent aux éléments du contexte institutionnel vietnamien que nous avons mentionnés dans l'introduction.

La formation des enseignants de mathématiques

Redisons que la formation des enseignants sépare collège et lycée. Avec le questionnaire distribué aux enseignants de la classe 10, nous avons pu faire l'hypothèse d'une soumission très forte des enseignants aux manuels reconnus selon nous comme les représentants de l'institution d'enseignement. Au-delà de cette hypothèse qui positionne sans surprise les enseignants comme les garants des rapports institutionnels aux objets de la statistique, nous nous posons la question des rapports personnels de ces enseignants à la statistique et aux probabilités. Pour les étudier il serait judicieux de s'intéresser aux instituts de formation de ces enseignants, à leurs enseignements disciplinaires et à leurs dispositifs d'étude de la méthodologie d'enseignement[167]. Les méthodes d'enseignement de la statistique proposées donnent-elles une place aux liens entre statistique et probabilités ? Cette place est-elle la même pour les futurs enseignants de collège et les futurs enseignants de lycée ?

L'environnement technologique

Nous avons abordé ce thème avec la question de la simulation informatique. Quelques soient les environnements technologiques dédiés à l'apprentissage d'une notion mathématique la question de la genèse instrumentale ne peut être éludée. Au Vietnam seule la calculatrice avec touche « random » est actuellement disponible pour tous les élèves et l'enseignant. Mais les équipements en salles informatiques et même en salles informatisées se multiplient très vite. Comment les enseignants vont-ils se saisir des possibilités offertes par l'informatique pour enseigner la statistique et les probabilités ? Quelles sont les ressources didactiques qui devraient être mises à la disposition des enseignants[168] ?

L'interdisciplinarité

Ce thème n'a pas été abordé dans notre étude pour les raisons que nous avons avancées dans l'introduction. Nous savons que d'autres enseignants, autres que de mathématiques, reçoivent une formation en statistique et probabilités. Comment cette formation prend-elle en compte les questionnements mathématiques à l'origine de l'échantillonnage et des sondages ? Quels outils mathématiques ces enseignants sont-ils appelés à mobiliser pour conduire des études statistiques dans leur propre enseignement disciplinaire ? Comment cette co-disciplinarité entre mathématique et une autre discipline peut-elle se développer dans l'institution vietnamienne ?

---

[167] C'est l'expression utilisée au Vietnam pour la formation des futurs enseignants.

[168] Peut-on envisager au Vietnam un site de simulation en ligne tel que http://homeomath.imingo.net/simulations.htm ?

Bibliographie

I. Didactique des mathématiques - réflexions sur l'enseignement des mathématiques

ARTIGUE M. (1989), Ingénierie didactique, *Recherches en didactique des mathématiques,* Vol. 9/3, p. 281-308, Editions La Pensée Sauvage, Grenoble.

BESSOT A. (1995), Contrat didactique : outil central pour l'analyse des situations d'enseignement, in Comiti C., Ngo Anh T., Bessot A. al. (eds), *Didactique des disciplines scientifiques et formation des enseignants, Actes du Premier Colloque Francophone du Sud-est Asiatique*, Université Pédagogique d'Ho Chi Minh Ville, Vietnam.

BESSOT A. & COMITI C. (2002), *Cours de Thac si Didactique des mathématiques*, U.P.H.C.M – U.J.F. Grenoble 1.

BESSOT A. & COMITI C. (2008), Apport des études comparatives aux recherches en didactique des mathématiques : le cas Viêt-Nam / France, in L. Coulange & C. Hache (édits) *Actes du séminaire national de didactique des mathématiques, Année 2008*, p.171-194 Editions ARDM et IREM de Paris 7.

BESSOT A. & BIREBENT A. (2003), Modèles didactiques des situations d'apprentissage, *cours de DEA EIAHD module D3*, U.J.F. Grenoble 1.

BESSOT A. & NGUYEN C.T. (2005), L'ingénierie didactique, l'exemple d'une recherche coopérative franco-vietnamienne », *Actes du premier séminaire franco – vietnamien en Didactique des Mathématiques*, Juin 2005, Université de Pédagogie de Ho Chi Minh Ville.

BIREBENT A. (2007), Le calcul numérique dans l'enseignement mathématique secondaire : reprise des travaux de notre thèse, in G. Gueudet et Y. Matheron (eds) *Actes du séminaire national de didactique des mathématiques, Année 2007,* p. 308-324, Editions ARDM et IREM de Paris 7.

BIREBENT A. (2007), Le calcul numérique à l'épreuve de l'intégration de la calculatrice : le problème didactique de l'approximation décimale et son analyse à l'aide du concept de contrat institutionnel, in R. Floris et F. Conne (eds) *Environnements informatiques, enjeux pour l'enseignement des mathématiques*, De Boeck éditeurs, 75-94.

BOLEA P., BOSCH M., GARCIA J., GASCON J., RUIZ L. & SIERRA T. (2005), Analyse de « La mesure en CM1 » d'après la Théorie Anthropologique du Didactique, in *Sur la théorie des* situations *didactiques. Questions, réponses, ouvertures. Hommage à Guy Brousseau*, Editions La pensée sauvage, Grenoble.

BOSCH M., CHEVALLARD Y. (1999), La sensibilité de l'activité mathématique aux ostensifs. Objet d'étude et problématique, *Recherches en didactique des mathématiques,* Vol.19, n°1, p. 77–124, Editions la Pensée Sauvage, Grenoble.

BOSCH M., FONSECO C., GASCÓN J. (2004), Incompletud de las organizaciones matemáticas locales en las instituciones escolares, *Recherches en didactique des mathématiques*, Vol. 24/2.3, p. 205–250, Editions La Pensée Sauvage, Grenoble.

BROUSSEAU G. (1986), Fondements et méthodes de la didactique des mathématiques, *Recherches en didactique des mathématiques,* Vol. 7/2, p. 33–116, Editions La Pensée Sauvage, Grenoble.

BROUSSEAU G. (1990), Le contrat didactique : le milieu, *Recherches en didactique des mathématiques,* Vol. 9/3, p. 309–336, Editions La Pensée Sauvage, Grenoble.

BROUSSEAU G. (1998), *Théorie des situations didactiques*, Pensée Sauvage, Grenoble.

CHEVALLARD Y. (1985), *La transposition didactique – du savoir savant au savoir enseigné*, Editions La Pensée Sauvage, Grenoble.

CHEVALLARD Y. (1989), Le concept de rapport au savoir, rapport personnel, rapport institutionnel, rapport officiel, *Séminaire de didactique des mathématiques et de l'informatique,* LSD-IMAG Grenoble, p. 211–235.

CHEVALLARD Y. (1992) Concepts fondamentaux de la didactique : perspectives apportées par une approche anthropologique, *Recherches en didactique des mathématiques,* Vol. 12/1, p. 73–111, Editions La Pensée Sauvage, Grenoble.

CHEVALLARD Y. (1994), Les processus de transposition didactique et leur théorisation, in Arsac et al. (eds.), *La transposition didactique à l'épreuve*, La Pensée Sauvage, Grenoble.

CHEVALLARD Y. (1998), Analyse des pratiques enseignantes et didactique des mathématiques : l'approche anthropologique, *Actes de l'U.E. de la Rochelle*.

LE THAI BAO T.T. (2007), *Étude didactique des relations entre enseignement de la notion de limite au lycée et décimalisation des nombres réels dans un environnement « calculatrice »*. Thèse en cotutelle France – Viêt-Nam, Université Joseph Fourier et Université Pédagogique d'Ho Chi Minh Ville.

NGUYEN B.K. & VU D.T. (1992), *Phương pháp dạy học môn Toán* (Méthodologie de l'enseignement des mathématiques), Edition « Nhà xuất bản Giáo dục », Hanoï, Vietnam.

NGUYEN C.T. (2009), L'usage du logiciel Cabri 3D dans l'enseignement de la géométrie dans l'espace au lycée du Vietnam, présentation au séminaire de l'équipe DIAM, Laboratoire LIG.

II. Didactique de la Statistique - réflexions sur l'enseignement des statistiques et des probabilités

ACADEMIE DES SCIENCES (2000), *La Statistique,* Rapport sur la science et la technologie n°8, Editeur Tec & Doc.

BRIAND J., (2005), Une expérience statistique et une première approche des lois du hasard au lycée par une confrontation avec une machine simple, *Recherches en Didactique des Mathématiques*, Vol. 25/2, p. 247-282, Editions La Pensée Sauvage, Grenoble.

BROUSSEAU G., BROUSSEAU N. & WARFIELD G. (2002), An experiment on the teaching of statistics and probability, *Journal of Mathematical Behaviour of children*, Vol. 20, Number 3, p. 363-411 (49), Editions Elsevier Science.

BROUSSEAU G. (2005), Situations fondamentales et processus génétiques de la statistique, in A. Mercier & C. Margolinas (Ed.), *Balises en Didactique des Mathématiques*, p. 165-194, Editions La pensée sauvage, Grenoble.

BROUSSEAU G. (2009), Alternatives en didactique de la statistique, in *Actes de la 41ième Journée de Statistique, Société Française de Statistique*, Bordeaux.

CHEVALLARD Y. & WOZNIAK F. (2005), Enseigner la statistique au secondaire, entre genre prochain et différence spécifique, in A. Mercier & C. Margolinas (Ed.), *Balises en Didactique des Mathématiques*, p. 195-218, Editions La pensée sauvage, Grenoble

COMMISSION DE REFLEXION SUR L'ENSEIGNEMENT DES MATHEMATIQUES (2001), *Statistique et probabilités*, Rapport d'étape.

COMMISSION INTER-IREM STATISTIQUE ET PROBABILITES (1997), *Enseigner les probabilités au lycée: Ouvertures statistiques, enjeux épistémologiques, questions didactiques et idées d'activités*, ouvrage collectif publié par le réseau des IREM, avec le soutien de la Direction des Lycées et Collèges.

COMMISSION INTER-IREM STATISTIQUE ET PROBABILITES (2001), *Autour de la modélisation en probabilités*, ouvrage collectif coordonné par HENRY M. et publié au Presses Universitaires Franc-Comtoises.

COMMISSION INTER-IREM STATISTIQUE ET PROBABILITES (2005 et 2007), *Statistique au lycée*, vol. 1 et 2, ouvrages collectifs publiés par l'APMEP.

COMMISSION INTER-IREM STATISTIQUE ET PROBABILITES (2003), P*robabilités au lycée*, ouvrage collectif publié par l'APMEP.

COUTINHO C. (2001), *Introduction aux situations aléatoires dès le collège: de la modélisation à la simulation d'expériences de Bernoulli dans l'environnement informatique Cabri-géomètre 2*. Thèse de doctorat, Université Joseph Fourier, Grenoble.

HENRY M. (1994), *L'enseignement du calcul des probabilités dans le second degré. Perspectives historiques, épistémologiques et didactiques*. Editions IREM de Besançon.

HENRY M. (2008), Évolution des programmes de l'enseignement secondaire français en statistique et probabilités, contributions des IREM à la formation continue des enseignants, *Actes du 1er Colloque international sur l'enseignement de la statistique*, (à paraitre).

IREM de Lyon (1998), *Enseigner la statistique du CM à la seconde. Pourquoi, comment ?*, brochure collective publiée par l'Irem de Lyon.

IREM de Montpellier (2001), *Des statistiques à la pensée statistique*, brochure collective publiée par l'Irem de Montpellier.

LE KE D. (2000), *Evidence dans le hasard*, Maison d'édition éducative du Vietnam.

PARSYSZ B., *L'enseignement des probabilités et de la statistique en France depuis* 1965, http://www.caib.es/ibae/esdeveniment/jornades_10_01/doc/prostat-F.doc

ROBERT C. (1995), *L'empereur et la girafe. Initiation à la statistique*, Diderot éditeur, Paris.

TRAN T.A. (2007), *Etude des pratiques d'enseignement sur la notion de probabilité dans les classes bilingues et les classes ordinaires au Vietnam*, Mémoire de Thac Si de l'Université Pédagogique de Ho Chi Minh Ville.

VU NHU T.H. (2005), *La notion de probabilité dans l'enseignement-apprentissage des mathématiques au lycée*, Mémoire de Thac Si de l'Université Pédagogique de Ho Chi Minh Ville.

WOZNIAK, F**.** (2005), *Conditions et contraintes de l'enseignement de la statistique en classe de seconde générale. Un repérage didactique*, Thèse de l'Université Claude Bernard. Lyon 1.

III. Traités et manuels universitaires en statistique et probabilités

CALOT G. (1973), *Cours de statistique descriptive*, Editions Dunod, Paris.

DROESBEKE J.J. (2001), *Éléments de statistique*, Editions de l'Université de Bruxelles

DODGE Y. (2005), *Premiers pas en statistique*. Editions Springer.

FOURASTIÉ J., LASLIER J.-F. (1997), *Probabilité et statistique.* Dunod, Paris.

LESSARD S. & MONGA E. (1993), *Statistique, Concepts et Méthodes, avec exercices et corigés.* (M. éditeur, Ed.), Les presses de l'université de Montréal, Canada.

MÉOT A. (2003), *Introduction aux Statistiques inférentielles - De la logique à la Pratique, avec exercices et corrigés*, Editions Boeck Université, Bruxelles.

MURRAY R. SPIEGEL (1972), *Théorie et applications de la statistique*. Série Schaum.

NGUYEN C.V. & TRAN T.N. (2005), Giáo trình Lý thuyết Xác suất và Thống kê Toán (Cours de Théorie des Probabilités et de la Statistique mathématique), Edition « Nhà xuất bản Thống kê », Hanoï, Vietnam.

SAPORTA G. (2006), *Probabilités, Analyse des données et Statistique*, Editions Technip, Paris.

SPIEGEL M.R. (1981), *Probabilités et Statistiques, Cours et problèmes* (Theory and Problems of Probability and Statistics) transl. JACOUD R., Editions McGraw-Hill, Paris.

WONNACOTT T.H. & R.J. (1991), *Statistique : économie, gestion, sciences, médecine.* Traduction française, Economica, Paris.

IV. Manuels et programmes scolaires du secondaire

Manuels et programmes scolaires français

Ministère de l'Éducation nationale (1999), Programme de mathématiques de la classe de seconde générale et technologique. *BOEN*, *HS6*. CNDP Paris.

[http://www.education.gouv.fr/bo/1999/hs6/default.htm]

Ministère de l'Éducation nationale (2000a). Document d'accompagnement des programmes de Mathématiques - classe de seconde. CNDP Paris.

[http://www.cndp.fr/archivage/valid/14963/14963-8208-9261.pdf]

Ministère de l'Éducation nationale (2000c). Programme de mathématiques des classes de première série économique et sociale. *BOEN*, *HS*8. CNDP Paris.

[http://www.education.gouv.fr/bo/2000/hs8/math.htm]

Ministère de l'Éducation nationale (2001a). Document d'accompagnement de programme - Mathématiques - classe de première des séries générales. CNDP Paris.

[http://www.cndp.fr/archivage/valid/35165/35165-6099-5919.pdf]

Ministère de l'Éducation nationale (2001c). Programme de l'enseignement des mathématiques en classe terminale des séries économique et sociale & scientifique. *BOEN*, *HS4*. Paris : CNDP. [http://www.education.gouv.fr/bo/2001/hs4/default.htm]

Ministère de l'Éducation nationale (2003). Enseigner au collège. Mathématiques. Programmes et accompagnement. CNDP Paris.

[http://www.cndp.fr/archivage/valid/54412/54412-9698-12059.pdf]

Annabrevet 1996, Editions Hatier Paris.

Annabrevet Sujets, 2007, Mathématiques, Editions Hatier Paris.

Annabrevet Corrigés, 2007, Mathématiques, Editions Hatier Paris.

Manuel 6ième, Collection Nathan, 1986 (p. 181-194)

Manuel 5ième, Collection Nathan, 1987 (p. 164-169)

Manuel 4ième, Collection Terracher, Editions Hachette, 1988 (p. 156-173)

Manuel 3ième, Collection Terracher, Editions Hachette, 1989 (p. 261-272)

Manuel 6ième, Collection Diabolo, Editions Hachette, 2005 (p. 107-120)

Manuel 5ième, Collection Diabolo, Editions Hachette, 2006 (p. 121-140)

Manuel 4ième, Collection Triangle, Editions Hatier, 2002 (p. 71-84)

Manuel 3ième, Collection Diabolo, Editions Hachette, 2005 (p. 137-154)

Manuel Seconde, Collection Terracher, Editions Hachette, 1994 (p. 109-132)

Manuel Première S, Collection Terracher, Editions Hachette, 1995 (p. 91-116)

Manuel Terminale S, Collection Terracher, Editions Hachette, 1994 (p. 283-312)

Manuel Seconde, Collection Déclic, Editions Hachette, 2004 (p. 193-216),

Manuel Première S, Collection Terracher, Editions Hachette, 2001 (p. 227-280)

Manuel Terminale S, Collection Terracher, Editions Hachette, 2002 (p. 265-294)

Manuels et programmes scolaires vietnamiens (édition du Ministère d'Education et de Formation)

Nouveau programme du Collège (à partir de 2002).

L'ensemble de manuel de classe 7, (manuel, livre d'exercices corrigés, livre des enseignants), 2003, Maison d'édition Education et Formation.

Nouveau programme du Lycée (à partir de 2006).

L'ensemble expérimental 1 de classe 11 (manuel, livre d'exercices corrigés, livre des enseignants), 2003, Maison d'édition Education et Formation.

L'ensemble expérimental 2 de classe 11 (manuel, livre d'exercices corrigés, livre des enseignants), 2003, Maison d'édition Education et Formation.

L'ensemble élémentaire de classe 10 (manuel, livre d'exercices corrigés, livre des enseignants), 2006, Maison d'édition Education et Formation.

L'ensemble avancé de classe 10 (manuel, livre d'exercices corrigés, livre des enseignants), 2006, Maison d'édition Education et Formation

L'ensemble élémentaire de classe 11 (manuel, livre d'exercices corrigés, livre des enseignants), 2007, Maison d'édition Education et Formation.

L'ensemble avancé de classe 11 (manuel, livre d'exercices corrigés, livre des enseignants), 2007, Maison d'édition Education et Formation.

V. Autres ouvrages

NGUYEN B.K. & LE KHAC T. (1993), *Dạy học một số yếu tố của Toán học tính toán và Tin học* (Enseignement de quelques éléments de l'Analyse numérique et de l'Informatique), Editions « Nhà xuất bản Giáo dục », Hanoï, Vietnnam.

Résumé en vietnamien

Tóm tắt luận án bằng tiếng Việt

Giữa xác suất và thống kê vừa có những liên hệ - thể hiện rõ nét ở các khái niệm cơ bản chung, vừa có sự tách biệt - chẳng hạn như xác suất được gắn với toán học thuần tuý trong khi thống kê lại được gắn với toán ứng dụng. Sự tách biệt này tồn tại khá lâu trong dạy toán ở phổ thông Việt nam, cụ thể là thống kê mô tả được dạy ở bậc phổ thông và xác suất được dạy ở bậc đại học.

Về điều này, chương trình năm 2006 dành cho bậc trung học phổ thông Việt Nam đã mang lại một sự thay đổi quan trọng : cùng với việc mở rộng một số kiến thức về *thống kê* đã được đề cập ở trung học cơ sở, người ta còn đưa vào cả *Xác suất*. Sự thay đổi này đặt ra cho giáo viên và nhà nghiên cứu nhiều vấn đề để xem xét. Cần phải nói rằng phần lớn trong số đó đã là chủ đề của nhiều cuộc tranh luận lâu dài ở một số nước, chẳng hạn như Pháp, nơi mà từ lâu cả *xác suất* lẫn *thống kê* đã có mặt trong chương trình phổ thông. Đứng trước sự thay đổi đó, chúng tôi tự hỏi : Vì sao lại hạn chế chỉ dạy Thống kê mô tả ? Vì sao bây giờ xác suất mới được đưa vào dạy ở trung học phổ thông Việt Nam và vì sao lại được chọn dạy ở lớp 11 ? Hiện tượng *xác suất* và *thống kê* được dạy ở hai lớp khác nhau dẫu không thể lấy làm cơ sở để khẳng định rằng hai nội dung này được đưa vào theo cách hoàn toàn tách rời, vẫn có thể khiến người ta đặt ra dấu hỏi về sự gắn kết giữa chúng trong dạy học. Việc học tập nào được nhắm đến ? Những khó khăn nào có thể tác động đến việc dạy học ? Làm thế nào để cho những liên hệ về mặt khoa học luận giữa *thống kê* và *xác suất* « sống được » gần kề nhau mà không đối lập nhau ?

Tìm kiếm các yếu tố trả lời cho những câu hỏi trên chính là trọng tâm của nghiên cứu didactic mà chúng tôi đặt ra cho luận án này. Đối với nghiên cứu của chúng tôi, câu hỏi cuối cùng trong các câu hỏi nêu trên là cốt lõi nhất, vì trong những mối liên hệ mà việc dạy toán có thể gầy dựng nên giữa hai lĩnh vực *xác suất* và *thống kê*, thì bản chất tự nhiên và việc đem lại nghĩa (do giáo viên và học sinh cùng xây dựng trong lĩnh vực này hay lĩnh vực kia) cho các khái niệm khác nhau sẽ cùng đồng hành với nhau.

Hiện nay, nhiều công trình nghiên cứu ở Pháp về didactic toán, chẳng hạn các công trình của Ủy ban Liêp hiệp các Viện nghiện cứu dạy toán « Thống kê và Xác suất » (Commission inter-IREM « Statistiques et Probabilités »[169]), của Ủy ban nhận xét việc dạy toán (Commision de réflexion sur l'enseignement des mathématiques), đã khẳng định sự thích đáng của những tiếp cận giao nhau giữa *thống kê* và *xác suất* trong dạy toán. Nhiều trong số những công trình nghiên cứu này hoặc đã đi trước hoặc đã đi kèm với những thay đổi liên tiếp của chương trình giảng dạy ở Pháp, mà thay đổi gần đây nhất là đưa *thống kê suy diễn* vào dạy ở trung học phổ thông và đưa *xác suất* vào bậc trung học cơ sở [170]. Khi tham khảo những tài liệu này, chúng tôi thấy rằng các cuộc tranh luận đã và còn đang tác động đến các nhà soạn thảo sách giáo khoa của Pháp.

Tuy nhiên, không phải việc giữ lại những lựa chọn về mặt khoa học luận là lý do duy nhất tác động đến các quyết định về phần chương trình cũng như về phần phương pháp giảng dạy. Thực ra, còn có sự can thiệp rất mạnh của nhiều lý do khác mà đa phần là gắn liền với bối cảnh của thể chế. Đó cũng là nơi mà các quyết định sẽ được phản hồi.

---

[169] Trích dẫn các tài liệu sau : Dạy xác suất ở phổ thông (1997), Thống kê ở trung học phổ thông (2007).

[170] Có thể tham khảo các chương trình này ở trang web : http://eduscol.education.fr/

Tại Việt Nam, trong những chương trình trước đây, chỉ các lớp song ngữ Pháp-Việt mới được học thống kê ở trung học cơ sở và xác suất ở trung học phổ thông[171]. Đối với hệ thống giáo dục chung, thống kê mới hiện diện trong chương trình trung học cơ sở từ những năm 1990 trở lại đây. Quyết định đưa xác suất vào dạy ở bậc trung học phổ thông vào chương trình mới (áp dụng trên toàn quốc từ năm 2006) đã là một bước ngoặt lớn so với việc xác suất chỉ được dạy ở bậc đại học trong nhiều thập niên qua. Quyết định ấy của các nhà soạn thảo chương trình và sách giáo khoa Việt nam đã được tính trước thông qua pha thực nghiệm kéo dài trong ba năm với hai bộ sách giáo khoa đặc biệt viết theo chương trình mới. Mục đích nhắm đến của việc thay đổi chương trình nói chung, của việc đưa vào *xác suất* và *thống kê* nói riêng là làm cho toán học dạy ở nhà trường gần với cuộc sống hơn, đồng thời cũng để tạo điều kiện cho sự hòa nhập của học sinh Việt Nam vào thế giới bên ngoài khi họ bước lên bậc đại học.

Do đi sau các nước khác trong việc đưa khái niệm xác suất vào giảng dạy ở phổ thông trung học nên các tác giả sách giáo khoa Việt nam có nhiều tư liệu, sách giáo khoa, chương trình,… của các nước khác để tham khảo cũng như có nhiều chọn lựa trong cách giới thiệu khái niệm xác suất. Tuy nhiên, để việc thay đổi chương trình được thành công thì cần phải tính đến nhiều yếu tố. Liên quan đến *xác suất – thống kê*, Ủy ban nhận xét việc dạy toán của Pháp đã đưa ra thuật ngữ « đào tạo về ngẫu nhiên ». Thế nhưng, để quyết định làm nổi bật tầm quan trọng của sự ngẫu nhiên trong dạy học toán đi được vào thực tế thì phải tính đến những ràng buộc liên quan đến việc đào tạo giáo viên, đến vấn đề liên môn và đến môi trường công nghệ.

Trước khi trình bày câu hỏi và phương pháp nghiên cứu, chúng tôi xin mô tả ngắn gọn bối cảnh của thể chế Việt nam liên quan đến ba yếu tố này. Với mỗi yếu tố, chúng tôi sẽ đưa ra một số nhận xét phản hồi của ủy ban Kahane, xem đó như một cơ sở để so sánh với thể chế của Pháp.

I. Bối cảnh dạy toán ở trung học phổ thông Việt nam

I.1. Việc đào tạo giáo viên

Sau đây là những nhận xét của ủy ban Kahane dành cho Pháp và được phổ biến vào năm 2000 :

> Ngày nay, việc đào tạo giáo viên trung học phổ thông và trung học cơ sở đang đương nhiệm là vấn đề cốt lõi của cái gọi là việc đào tạo phổ thông về sự ngẫu nhiên. Đâu là những đặc thù của một kiểu đào tạo như vậy ?
>
> - việc đào tạo này liên quan đến khoảng 40.000 người (giáo viên trung học cơ sở và trung học phổ thông) ; trong đó, việc dạy thống kê được phân bổ như một điều cần thiết và nó liên quan đến nhiều môn học khác. Không có lý do gì để tạo vách ngăn giữa các đào tạo này. Nên đào tạo chung các giáo viên toán, lý, sinh và khoa học tự nhiên cũng như khoa học xã hội, như vậy, trên một cộng đồng khoảng 60.000 người.

[171] Có thể tham khảo luận văn thạc sĩ của tác giả Trần Tuý An (2007) : *Nghiên cứu thực hành dạy khái niệm xác suất ở lớp song ngữ và lớp thường ở Việt nam*, ĐHSP TP. .HCM.

- việc đào tạo này là cần thực hiện và phải được xây dựng trên sự cộng tác với các giáo viên, như đã dạy thống kê cho các bác sĩ bằng cách dựa trên một số đông chính các bác sĩ.
- một đào tạo như vậy đòi hỏi phải có các yếu tố lý thuyết và phải làm cho các giáo viên đối diện với việc xử lý dữ liệu thực. Việc đào tạo này cũng cần phải được tích hợp và sử dụng việc giả lập ngẫu nhiên. Như tất cả các loại đào tạo khác, chất lượng của việc đào tạo phụ thuộc vào tính thích đáng của các ví dụ được xử lý, của các vấn đề được đặt ra và mặc dù đây là một đào tạo chuyên môn, các vấn đề mang tính thúc đẩy nhất hoặc soi rõ nhất lại không phải là dành cho phần lớn các thực hành chuyên môn của giáo viên.
- để tăng cường lực lượng cho việc đào tạo, phải làm cho các giáo viên phổ thông giỏi về thống kê tham cùng gia vào việc đào tạo, cũng như các giảng viên đại học các môn khác, các trường chuyên như INSEE, INRA, INED, các giảng viên các trường đại học kỹ thuật, các bác sĩ, v.v…

Luận án của chúng tôi không đi sâu nghiên cứu việc đào tạo giáo viên. Dẫu vậy, một số phân tích đã dẫn chúng tôi đến chỗ thực hiện một nghiên cứu thực nghiệm liên quan đến giáo viên và học sinh lớp 10. Điều đó cho phép chúng tôi quay trở lại bàn luận chút ít về chủ đề đào tạo giáo viên.

I.2. Vấn đề liên môn

Báo cáo của ủy ban Kahane phát triển việc sử dụng thống kê và xác suất hiện nay trong nhiều môn học ở phổ thông như vật lý, sinh vật, kinh tế, địa lý. Đối với nhiều môn trong số này, báo cáo còn nhấn mạnh đến sự cạnh tranh giữa mô hình hóa xác định và mô hình hóa ngẫu nhiên :

> Sự ngẫu nhiên có mặt trong nhiều môn học và ngay trong việc mô tả các hiện tượng mà ta biết các phương trình xác định.

Báo cáo của ủy ban này còn nêu lên các nhận xét dành cho việc dạy các bộ môn khác :

> Trong dạy học môn sinh vật, nói riêng về dạy học ở phổ thông, cần cho học sinh làm quen với sự ngẫu nhiên, nhất là các hiện tượng tiếp diễn, làm cho học sinh ý thức về sự biến động của việc chọn mẫu, khởi động vấn đề dự đoán trước và làm việc trên sự khác nhau của cái dự đoán được và cái có tính nhân quả. v.v…
>
> […] Lúc đầu, ở mức độ dạy học phổ thông, việc dạy môn kinh tế có thể đảm nhận việc nhận xét bản chất của dữ liệu, cân nhắc về khó khăn của việc định nghĩa các chỉ số thích đáng để nghiên cứu thực tế của kinh tế ; chẳng hạn như một nhận xét có thể cùng lúc được dẫn đến toán học và đến kinh tế khi xây dựng các bảng tử vong và các phép tính về kỳ vọng sống của mọi lứa tuổi (vì sao phải định nghĩa một loạt giả định để thiết lập bảng tử vong, vì sao kỳ vọng sống của một giáo sư lại cao hơn kỳ vọng sống của một cá thể được lấy ngẫu nhiên trong dữ liệu tổng thể ? Công việc định nghĩa các chỉ số thích đáng này và các kiểu tính toán có thể có, là một yếu tố nền tảng của việc đào tạo thống kê.

Ở Việt nam, ngoài môn toán, nhiều môn học khác cũng có phần thực hành thống kê mô tả như địa lý, sinh vật, vật lý. Nhưng, qua trao đổi với một số đồng nghiệp dạy các môn này, chúng tôi thấy dường như vấn đề mô hình hóa ngẫu nhiên không được họ lưu tâm đến. Ấy thế mà đây lại là một việc rất cần thiết cho bước chuyển từ mô tả sang dự đoán. Ý tưởng dự

đoán trước chắc chắn là một trong những động cơ của việc làm thống kê. Nhưng ý định này ít ra là đã không thấy một cách tường minh đối với khái niệm xác suất được trình bày trong các sách giáo khoa toán viết theo chương trình mới.

Đối với mọi môn học ở trường trung học phổ thông việt nam, thứ tự và phân bố thời gian của các tiết dạy đều được ấn định trong sách giáo khoa, chương trình và giáo viên phải tuân thủ chặt chẽ. Như thế, khó mà đạt được một thỏa thuận giữa các giáo viên dạy những bộ môn khác nhau để đưa những vấn đề của môn này vào việc giảng dạy môn kia, một cách làm khiến cho học sinh thấy được lợi ích của toán học. Kinh nghiệm giảng dạy của bản thân chúng tôi cho thấy là không hề có cuộc bàn thảo nào về việc dạy môn thống kê giữa các liên môn.

Sự biến động của quá trình chọn mẫu mà ủy ban Kahane gợi ý không thể tồn tại một cách chính thức ở Việt nam vì khái niệm chọn mẫu không có mặt trong chương trình toán. Ngược lại, trong khi thực hành thí nghiệm của các môn học như vật lý, hóa học, sinh vật, các nhóm học sinh thường thu thập các số liệu đo lường của cùng một đại lượng. Hoạt động này dẫn đến việc phải chọn một giá trị chính xác từ tập hợp nhiều kết quả đo lường. Như vậy, ta cũng có thể thấy được là tính biến đổi này (là một phần trong các khái niệm thống kê) ngầm xuất hiện trong việc dạy học các môn học khác với môn toán. Như vậy, bằng cách nào mà bước khởi đầu của tính biến đổi này có thể góp phần làm tham chiếu cho khởi đầu của sự biến động trong quá trình chọn mẫu ? Câu hỏi này dường như không được đặt ra trong thể chế Việt Nam, dù liên môn vấn là một trong những định hướng cho việc đổi mới phương pháp dạy học.

I.3. Môi trường công nghệ

Chúng ta cùng trở lại với báo cáo của ủy ban Kahane :

> Có thể nói là […] ngành thống kê làm việc chủ yếu là trên dữ liệu thực nghiệm; đối với ngành thống kê khai thác, các công cụ lý thuyết được sử dụng chủ yếu là hình học và đại số tuyến tính, còn đối với thống kê suy diễn thì xác suất là công cụ lý thuyết và cơ sở vật chất chính là máy vi tính. Việc thực hiện các phương pháp thống kê đòi hỏi rất nhiều phương tiện tính toán ; ngày nay, ngành thống kê đã có thể phát triển rất mạnh là nhờ đã khả năng của máy tính vượt trội so với máy tính ở các trung tâm tính toán cách đây 30 năm.
>
> Việc sử dụng các công cụ phần mềm trong lĩnh vực chuyên môn hay trong các nghiên cứu ~~là~~ có hai mặt :
>
> - thực hiện các phép tính dài và phức tạp (tính giá trị trung bình, phương sai, nghịch đảo ma trận, tính toán tối ưu có ràng buộc, v.v…), tính các giá trị gần đúng (các hàm phân bố theo luật xác suất chẳng hạn), tính khoảng cách hữu hạn, thực hiện các phương pháp phi tham số, thực hiện tái chọn mẫu,
> - ước lượng các kết quả từ các giả lập, chẳng hạn thể tích của một số tổng thể nào đó, ước lượng xác suất của một biến cố mà ta không thể thiết lập công thức để tìm giá trị số gần đúng trực tiếp.
>
> Trong triển vọng của việc đào tạo, việc sử dụng công nghệ thông tin cũng có hai mặt :
>
> - truy cập được nhiều dữ liệu khác nhau về chất và có khả năng xử lý thống kê một cách thích đáng các dữ liệu này,

- việc giả lập là một công cụ được ưu tiên để đạt được một thí nghiệm của các hiện tượng ngẫu nhiên, hiểu được các lý thuyết về hội tụ, thấy được các vấn đề được đặt ở đâu, hiểu được bản chất của các bằng chứng thống kê.

Có thể thấy là hiện nay ở xã hội Việt nam, cùng với chính sách của nhà nước, các phương tiện công nghệ thông tin phát triển rất nhanh và tuyệt vời. Liên quan đến chính sách giáo dục, chúng tôi dựa trên tham khảo của tác giả Nguyễn Chí Thành (2009)[172] :

> Cần đổi mới phương pháp dạy học. Tránh truyền đạt kiến thức một chiều. Phải khuyến khích suy nghĩ sáng tạo. Phải áp dụng dần các công nghệ hiện đại vào việc dạy học.
>
> (Quyết định của Ủy ban trung ương Đảng cộng sản năm 1997)
>
> […] sử dụng các công nghệ hiện đại vào giảng dạy để đổi mới phương pháp dạy học. Các phương pháp này phải khuyến khích vai trò chủ động của học sinh, khả năng làm việc theo nhóm […]
>
> (Điều 8.2, Luật giáo dục năm 2005)
>
> Phải sử dụng các công nghệ hiện đại trong giảng dạy ở mọi cấp học như một công cụ hỗ trợ hiệu quả cho việc đổi mới phương pháp dạy tất cả các môn học.
>
> (Nghị định của thủ tướng chính phủ, 29/2001/BGD&DT)

Giáo viên được động viên nhưng thực ra vẫn còn ít được hỗ trợ. Việc khởi động của họ vẫn còn rất đơn độc. Nói riêng, việc giả lập không được chỉ định trong chương trình lẫn trong thực hành, dù là với máy tính bỏ túi – bằng cách sử dụng phím « Random ». Vậy làm thế nào để có thể đưa việc giả lập vào việc dạy thống kê và xác suất ?

Luận văn Thạc sĩ của tác giả Trần Tuý An [2007][173] có thể trả lời cho câu hỏi này bằng một trong các thực nghiệm, đó là đã cài đặt vào bảng tính (Excel) một mô hình bội và học sinh sử dụng bảng tính này để nhận được các kết quả của n lần thực hiện một phép thử ngẫu nhiên. Mục tiêu nhắm đến sự giả lập này là đặt cơ sở phương pháp tần suất cho việc tính xác suất trong trường hợp phương pháp cổ điển của Laplace không thể thực hiện được : lặp lại cùng một phép thử ngẫu nhiên thật nhiều lần và tính xác suất xuất hiện của một biến cố nào đó rồi đưa ra một giá trị gần đúng « hợp lý »cho xác suất. ; giá trị gần đúng này được gọi là xác suất thực nghiệm.

Việc thực hiện giả lập dữ liệu phụ thuộc vào vị trí mà thể chế dành cho tiếp cận tần suất trong định nghĩa và trong phép tính xác suất của một biến cố. Nghiên cứu của chúng tôi (Vũ Như Thư Hương, 2005) và của tác giả Trần Túy An đã chỉ ra rằng cách tiếp cận này không phải là không được biết đến trong chương trình và trong sách giáo khoa hiện hành của Việt nam[174], nhưng dường như « môi trường sống » cho nó đã không được tạo ra.

---

[172] Nguyễn Chí Thành (2009), *Sử dụng phần mềm cabri 3D trong dạy học hình học không gian ở trung học phổ thông Việt nam*, báo cáo trước êkíp DIAM, labo LIG.

[173] Trần Tuý An (2007), *Nghiên cứu thực hành dạy học khái niệm xác suất ở các lớp song ngữ và lớp thường của Việt nam*, Luận văn Thạc sĩ, ĐHSP TP.HCM.

[174] Vũ Như Thư Hương (2005), *Khái niệm xác suất trong dạy học toán ở trung học phổ thông*, Luận văn Thạc sĩ, ĐHSP TP.HCM.

II. Công cụ lý thuyết và phương pháp luận nghiên cứu

Chúng tôi đã thiết lập lại các câu hỏi ban đầu về giảng dạy thành ba câu hỏi nghiên cứu mà Lý thuyết Nhân học và Lý thuyết tình huống sẽ được sử dụng làm công cụ để tìm kiếm câu trả lời.

Câu hỏi 1. Trong giảng dạy toán phổ thông, các mối liên hệ giữa thống kê-xác suất được xây dựng qua những bài toán nào và những đối tượng toán học nào ?

Để trả lời cho câu hỏi này, chúng tôi sẽ phân tích hai cuốn giáo trình sách được sử dụng ở đại học của Pháp. Phân tích này được thực hiện theo cách tiếp cận của Thuyết nhân học, đặc biệt là với khái niệm tổ chức toán học. Việc phân tích các giáo trình đại học nhằm mục đích đưa ra các tổ chức mà chúng tôi sẽ lấy làm tham chiếu cho những phần nghiên cứu tiếp theo. Lưu ý rằng ở đây chúng tôi chỉ hạn chế nghiên cứu trên các tổ chức có ý nghĩa nhất đối với mối liên hệ khoa học luận giữa thống kê và xác suất. Chúng tôi sẽ chỉ ra những bài toán mà việc dạy học có thể khai thác để thiết lập mối quan hệ giữa xác suất với thống kê. Chúng tôi gọi đó là những *bài toán - ứng viên*. Để tạo nên được các mối liên hệ thống kê-xác suất, mỗi một trong các bài toán này cần đến những đối tượng mà chúng tôi gọi là các *đối tượng-ứng viên*.

Câu hỏi 2. Đâu là những điều kiện sống hiện hành của các *bài toán-ứng viên* và của các *đối tượng-ứng viên* trong giảng dạy toán ở phổ thông Việt nam ?

Cũng vẫn với các khái niệm của Thuyết nhân học, chúng tôi thực hiện một phân tích thể chế dạy học ở Việt nam và so sánh với thể chế dạy học ở Pháp. Chúng tôi đưa ra những tổ chức cần giảng dạy và mô hình hóa sự cách biệt của tổ chức này với tổ chức tham chiếu trong những hợp đồng thể chế liên quan đến đối tượng-ứng viên.

Chúng tôi cũng thiết lập các giả thuyết trên những tổ chức được giảng dạy và những hợp đồng sư phạm trong thể chế Việt nam rồi kiểm chứng các giả thuyết này bằng một thực nghiệm dưới dạng bộ câu hỏi điều tra giáo viên và học sinh.

Câu hỏi 3. Làm thế nào để trong những điều kiện giảng dạy hiện nay ở Việt nam, có thể làm « sống » được các mối liên hệ thống kê-xác suất qua một số trong các *bài toán-ứng viên* và *đối tượng-ứng viên* ?

Từ các giả thuyết đưa ra về khả năng có thể làm « sống » được các mối liên hệ thống kê-xác suất qua một số trong các *bài toán-ứng viên* và *đối tượng-ứng viên* trong điều kiện hiện nay, chúng tôi dự kiến một đồ án didactic và thực nghiệm tại 1 lớp học ở Việt nam. Sự đối chiếu các phân tích tiên nghiệm và phân tích hậu nghiệm các tình huống didactic của đồ án này sẽ tạo cơ sở cho kết luận của nghiên cứu.

Dưới đây, trước khi trình bày những kết quả chính của luận án, chúng tôi sẽ giới thiệu vắn tắt lợi ích của các công cụ lấy từ Thuyết nhân học trong didactic rất được biết đến (mà chúng tôi sử dụng trong nghiên cứu dành cho các câu hỏi 1 và 2). Đặc biệt, chúng tôi sẽ nói đến lợi ích của các khái niệm tổ chức toán học (Chevallard 1992, 1998), phân tích khoa học luận (Bosch & alii, 2000) và hợp đồng thể chế. Giới thiệu này cũng cho phép làm rõ tính hợp lý của phương pháp luận nghiên cứu chỉ đạo các nghiên cứu thực hiện trong phạm vi luận án.

II.1. Lợi ích của một nghiên cứu so sánh thể chế Pháp - Việt

Như chúng tôi đã nói trên, bối cảnh thể chế Việt nam áp đặt nhiều ràng buộc chặt chẽ và điều này giải thích vì sao có sự khựng lại của giáo viên trước việc thay đổi chương trình. Có lẽ là để giảm nhẹ tạm thời những khó khăn sinh ra từ nội dung mới của chương trình, phần xác suất được quyết định là không đưa vào trong đề thi Tú tài và đề thi tuyển sinh đại học. Hơn nữa, chương trình được giữ lại sau pha thí điểm đã bỏ đi một phần bị cho là khó, đó là *xác suất có điều kiện*. Thêm vào đó, bộ sách cơ bản (2007) còn bỏ đi phần *Biến ngẫu nhiên*.

Nghiên cứu thể chế trong luận án này không nhằm vào việc đánh giá những chọn lựa trong giảng dạy của Pháp so với những chọn lựa của Việt nam, lại càng cũng không hình dung đến khả năng lấy lại những chọn lựa của Pháp mà coi thường bối cảnh thể chế Việt nam. Mục đích của chúng tôi là tìm hiểu xem trong một bối cảnh thể chế như Việt nam, làm sao mà một số chọn lựa nào đó được xem là thích đáng còn số khác thì không, với mong muốn làm cho việc dạy thống kê và xác suất, cái này duy trì cái kia để xây dựng rõ nghĩa hơn cho một số khái niệm khác nhau của hai lĩnh vực. Tuy vậy, chúng tôi cũng không bỏ qua việc so sánh với nước Pháp vì theo Bessot và Comiti, một nghiên cứu so sánh như vậy có thể đem lại ba điểm tựa chắc chắn cho một nghiên cứu didactique :

- một sự phi đồng hóa cách nhìn vấn đề trên hoạt động học đường của một thể chế sư phạm;
- tính đến các cấp độ quyết định cao cấp hơn so với cấp độ quyết định của phạm vi ;
- tổ chức một kho tư liệu các praxéologie cần dạy đánh giá các tổ chức toán học và tổ chức sư phạm.

(Bessot&Comiti [2008], trang 191)

Phương pháp học về so sánh này được thực hiện để nêu lên các tố chức cần dạy, các đối tượng của Lý thuyết Nhân chủng học didactic.

II.2. Tổ chức toán học - một công cụ cơ bản của phân tích thể chế

Nhằm phân tích mối quan hệ thể chế với một đối tượng tri thức (ở đây là tri thức toán học), Chevallard (1998) giới thiệu khái niệm praxéologie toán học, còn gọi là *tổ chức toán học* (OM). Một OM được lập thành bởi bộ tứ : kiểu nhiệm vụ $\boldsymbol{T}$ xuất hiện trong thể chế, các kỹ thuật $\boldsymbol{\tau}$ cho phép thực hiện các nhiệm vụ của $\boldsymbol{T}$, các công nghệ $\boldsymbol{\theta}$ giải thích các kỹ thuật $\boldsymbol{\tau}$, các lý thuyết $\boldsymbol{\Theta}$ giải thích cho các công nghệ $\boldsymbol{\theta}$.

Chevallard phân biệt các cấp độ khác nhau của các OM.

- $[T/\tau/\theta/\Theta]$: OM *bộ phận* liên quan đến một kiểu nhiệm vụ duy nhất

Theo nguyên tắc, quanh một kiểu nhiệm vụ T, chúng ta sẽ tìm thấy một bộ ba tạo bởi một kỹ thuật (tối thiểu), τ, một công nghệ của τ, θ, và một lý thuyết của θ, Θ. Bộ tứ $[T/\tau/\theta/\Theta]$ tạo thành một tổ chức toán học *điểm*, nghĩa là một tổ chức toán học liên quan đến một kiểu nhiệm vụ duy nhất, T. (Chevallard 1998, trang 95)

- $[T_i/\tau_i/\theta/\Theta]$: OM *địa phương* tập trung vào một công nghệ θ duy nhất.

- $[T_{ij}/\tau_{ij}/\theta_j/\Theta]$: OM *vùng* được hợp thành quanh một lý thuyết Θ duy nhất.

- $[T_{ijk}/\tau_{ijk}/\theta_{jk}/\Theta_k]$: OM *toàn phần* được tích hợp từ nhiều OM vùng ứng với nhiều lý thuyết Θ_k.

(Chevallard 1998, trang 95)

Khái niệm praxéologie là một trong các công cụ của Thuyết nhân học theo cách tiếp cận mà Bosch và Chevallard [2001] trình bày như sau:

Tiếp cận nhân chủng học trả lời cho câu hỏi về việc mô hình hóa các hoạt động xã hội, mô hình hóa các thành phần, sự tiến triển và các thành quả của chúng. Đối với toán học, cái được xem là một hoạt động nhân văn được cấu trúc thành tổ chức praxéologie – tổ chức, ta có thể nói là chúng nảy sinh từ việc đặt vấn đề của một số kiểu nhiệm vụ nào đó và được xem là như kiểu bài toán mà việc nghiên cứu tạo điều kiện cho việc xây dựng các praxéologie địa phương. Sự nối khớp một số trong các praxéologie quanh một công nghệ chung cho phép thiết lập các praxéologie vùng và đến phiên chúng nối khớp với nhau toàn bộ để tạo nên « tri thức toán học ». Việc mô tả các tổ chức và nghiên cứu sinh thái thể chế chính là trọng tâm của chương trình nghiên cứu didactic toán.

(Bosch & Chevallard [2001], trang 8)

Khởi đầu của một phân tích thể chế là nêu lên tập hợp các praxéologie toán học đang sống trong thể chế, từ đó thực hiện một mô hình cải thiện cơ cấu tập hợp các thực hành toán học được hệ thống và được chia sẻ trong thể chế. Phương pháp này được Chevallard [1998] mô tả như sau :

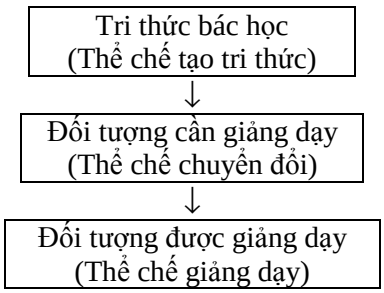
Thật ra, chúng ta hiếm khi bắt gặp các OM điểm. Một cách tổng quát, trong một thể chế đang xét, một lý thuyết Θ đáp ứng *nhiều công nghệ* θ_j. Đến lượt mình, mỗi công nghệ giải thích và giúp hiểu được *nhiều kỹ thuật* τ_{ij} tương ứng với *chừng ấy kiểu nhiệm vụ* T_{ij}. Vậy là, các tổ chức toán học điểm được tích hợp lại, đầu tiên thành các tổ chức toán học *địa phương*, $[T_i/\tau_i/\theta/\Theta]$, tập trung vào một công nghệ θ duy nhất, kế đến thành các tổ chức toán học *vùng*, $[T_{ij}/\tau_{ij}/\theta_j/\Theta]$, được hợp thành quanh một lý thuyết Θ duy nhất. (Trên giới hạn này, chúng ta sẽ gọi sự phức hợp $[T_{ijk}/\tau_{ijk}/\theta_{jk}/\Theta_k]$ là tổ chức toán học *toàn phần*. Tổ chức toán học toàn phần là sự tích hợp nhiều OM vùng ứng với nhiều lý thuyết Θ_k).

(Chevallard [1998], trang 95)

Việc thiết lập các praxéologie đòi hỏi phải đi sâu không chỉ vào chương trình, sách giáo khoa hiện hành mà còn cả vào chương trình, sách giáo khoa cũ, hay vào các thể chế khác nữa, để định nghĩa một cách đầy đủ các kiểu nhiệm vụ có mặt và vắng mặt, các kỹ thuật và các yếu tố công nghệ được trình bày dưới dạng các quy tắc hay định lý. Trong phần nghiên cứu của mình, chúng tôi quan tâm ngay đến các kiểu nhiệm vụ « Việt nam » và các kiểu nhiệm vụ « Pháp ».

II.3. Mô hình khoa học luận tham chiếu

Sơ đồ cổ điển mô tả sự chuyển đổi sư phạm gồm ba mắt xích sau :



Sơ đồ chuyển đổi didactic cổ điển

Sự đối chiếu khoa học luận khai thông việc thiết lập các tổ chức tham chiếu :

> Cho một tiến trình sư phạm trên một chủ đề toán học xác định. Giai đoạn đầu của kỹ thuật nghiên cứu giả định thừa nhận cái được gọi là một kiểu mẫu khoa học luận về sự tham chiếu trên nội dung toán học đang bàn đến. Trong khuôn khổ của ~~lý~~ thuyết nhân chủng học, kiểu mẫu này tạo nên các từ ngữ tổ chức toán học hay praxéologie toán học.
>
> (Bolea, Bosch, Garcia, Gascon, Ruiz, Sierra, 2000, Hội thảo Hommage à Guy Brousseau).

Được tạo thành một cách ngầm ẩn từ các tổ chức cần giảng dạy và các văn bản « bác học » như sách giáo khoa hay các sách đại học[175], kiểu mẫu này phải cho phép chỉ ra các mối quan hệ thể chế của các đối tượng tri thức được bao gồm trong các tổ chức cần giảng dạy. Nó cũng phải cho phép chúng ta hiểu được là các ràng buộc đè nặng trên thể chế dạy học và trên tiến trình thiết lập hợp đồng giữa các chủ thể hành động của thể chế liên quan đến tri thức

Hợp đồng này có thể được mô tả theo hai mức độ : hợp đồng thể chế và hợp đồng sư phạm.

II.4. Hợp đồng thể chế và các hợp đồng sư phạm

Hợp đồng thể chế liên quan đến đối tượng tri thức O - ở đây, chúng tôi muốn nói đến đối tượng thống kê – có thể được định nghĩa như một hệ thống các mối quan hệ thể chế với O và với o_i, o_i thích đáng đối với O nếu xét trong thể chế này. Một hệ thống như vậy được gọi là *hợp đồng thể chế liên quan đến Thống kê*. Cụ thể hơn, hợp đồng ấy liên quan đến các đối tượng toán học thuộc lĩnh vực thống kê hay có mối liên hệ hẹp với thống kê.

Để giải thích thêm khái niệm này, chúng tôi trích dưới đây một trình bày chi tiết hơn về hợp đồng thể chế liên quan đến phép tính số do tác giả Alain Birebent (2008) nêu ra :

> Hợp đồng thể chế xuất hiện từ khoảng cách giữa hệ thống các tổ chức tham chiếu và hệ thống các tổ chức cần giảng dạy mà chương trình và hướng dẫn giảng dạy thể hiện và sách giáo khoa cũng như các đề thi, đề kiểm tra phản ánh ; trong thể chế, hợp đồng này chỉ đạo tiến trình sư phạm nhằm khôi phục lại các tổ chức tham chiếu trong các tổ chức

[175] Chúng tôi chọn một số sách đại học của Pháp.

cần giảng dạy và quyết định các tổ chức thực sự được giảng dạy. Vừa là tiến trình, vừa là trạng thái, hợp đồng thể chế :

- trình bày cách ứng xử của học sinh và của giáo viên đối với O, trong I (những mong đợi của I liên quan đến O) ;
- thiết lập một phân cách hẹp giữa học sinh và giáo viên trong quan hệ với O trong I;
- bảo đảm sự hoàn thiện các thời didactic đối với O (chronogenèse) và thay đổi trạng thái của hợp đồng bằng các ngắt quãng / chuyển tiếp hay lấy lại / quên đi ;
- xác định một trường các ràng buộc và các điều kiện cho các tổ chức cần giảng dạy ;
- lập thành một ma trận các hợp đồng đối với đối tượng o_i có liên kết với phép tính số trong thể chế.

(Alain Birebent [2008], trang 316)

Ở đây, đối tượng tri thức O là thống kê và là một « đối tượng lớn ». Hợp đồng thể chế đưa ra cái nhìn « toàn cục » theo cách tiếp cận sinh thái cho các quan hệ didactic giữa các chủ thể hành động khác nhau liên quan đến O và tất cả các o_i. Điều này cho phép mô tả thể chế trong thời gian và không gian « sống » của O. Khái niệm hợp đồng didactic[176] gợi lại một sự mô hình hóa các mối quan hệ sư phạm, liên quan đến một đối tượng toán học, giữa các học sinh và một giáo viên trong thời gian làm việc chung ngắn ngủi của họ.

Trong sự mô hình hóa bằng hai kiểu hợp đồng này, hợp đồng sư phạm là sự thiết lập một ví dụ cụ thể về một đối tượng, được thực hiện từ lớp các đối tượng theo không gian và thời gian của hợp đồng thể chế. Cách nhìn này đã được Chevallard [1989] diễn đạt như sau : « hợp đồng didactic là một trạng thái của sự hình thành chronogenèse và topogenèse ở một thời điểm » của hợp đồng thể chế.

II.5. Đồ án sư phạm

Để trả lời câu hỏi 3, chúng tôi dự kiến, tiến hành thực nghiệm và phân tích một đồ án sư phạm.

Khái niệm đồ án sư phạm xuất hiện trong khoa học Didactic Toán từ đầu những năm 1980. Công việc sư phạm này mang tên đồ án sư phạm vì nó có thể so sánh với công việc của một kỹ sư.

Để thực hiện một dự án cụ thể nào đó, người kỹ sư dựa vào các kiến thức khoa học trong lĩnh vực liên quan và chịu sự kiểm soát của khoa học. Nhưng anh ta phải làm việc với những đối tượng phức tạp hơn nhiều so với những đối tượng khoa học thuần túy thuộc lĩnh vực của mình. Với tất cả các phương tiện hiện có, anh ta phải tiến hành giải quyết các bài toán vượt ngoài tầm lĩnh vực khoa học liên quan.

(Artigue [1990], trang 283)

[176] Có thể tham khảo một trình bày có từ trước về khái niệm hợp đồng thể chế trong Bessot A. (1995), Hợp đồng thể chế : một công cụ trung tâm để phân tích các tình huống dạy học, trích trong Comiti C. Ngo Anh T., Bessot A. và các tác giả, *Didactic các môn khoa học và đào tạo giáo viên*, ĐHSP TP. HCM.

> Đồ án sư phạm là một phương pháp nghiên cứu. Phương pháp này được đặc trưng bằng một sơ đồ thực nghiệm « thực hiện dạy học » trong lớp. Nói cách khác, sơ đồ thực nghiệm này bao gồm xây dựng, thực hiện, quan sát và phân tích các công đoạn dạy học.

(Tài liệu đã dẫn, trang 286)

Theo trích dẫn của tác giả Lê Thái Bảo Thiên Trung (2007) :

> Đồ án sư phạm là một tình huống dạy học (hay tập hợp các tình huống dạy học) do nhà nghiên cứu didactic thiết kế, dựa trên các khái niệm cơ bản trong lý thuyết tình huống của Brousseau, như tình huống cơ bản, biến didactic, môi trường… Mục tiêu của một đồ án sư phạm là:
>
> - thực hiện việc giảng dạy trong thể chế dạy học sau khi đã nghiên cứu thể chế và khoa học luận,
>
> - và kiểm chứng các giả thuyết hay trả lời các câu hỏi đề ra (bằng cách đối chiếu giữa các phân tích tiên nghiệm và phân tích hậu nghiệm).
>
> Đồ án sư phạm có hai chức năng : phương pháp luận ~~học~~ và sản sinh ra những câu hỏi sư phạm mới.

(Lê Thái Bảo Thiên Trung [2007], trang 286)

Trong khuôn khổ luận án này, chúng tôi dựa vào các phân tích khoa học luận và phân tích thể chế ở các phần B, A, C để dự kiến một đồ án đối với vấn đề « sự không chắc chắn của ngẫu nhiên ». Việc dự kiến này còn dựa trên giả thuyết cho phép xây dựng một lời giải cho vấn đề không chắc chắn của ngẫu nhiên bằng một chuỗi các tình huống sư phạm tạo thuận lợi cho việc làm phong phú trên phương diện khoa học luận mối liên hệ giữa *thống kê* với *xác suất*. Đồ án sư phạm này được chia thành ba buổi thực nghiệm tương ứng với các giá trị của một biến méso-didactic và được thực hiện ở một lớp 11 tại Việt nam.

III. Cấu trúc luận án

Luận án bao gồm bốn phần chính : A, B, C và D

Phần A

> Trong phần này, chúng tôi thực hiện một phân tích thể chế so sánh Pháp-Việt và định vị các tổ chức cần giảng dạy ở Pháp và ở Việt nam

Phần B

> Đây là phần nghiên cứu về khoa học luận, nhằm chỉ ra một mô hình các mối liên hệ thống kê-xác suất trong hai tổ chức tham chiếu, đó là :

- so sánh các phân bố thống kê
- chọn mẫu dữ liệu

Phần C

> Ở đây chúng tôi quay trở lại phân tích thể chế Việt nam. Kết quả nghiên cứu được trình bày trong hai chương :

Chương C1 : Dành cho việc làm rõ các hợp đồng thể chế liên quan đến các đối tượng của thống kê và liên quan đến các tổ chức dạy học chưa hoàn chỉnh. Nghiên cứu này dẫn chúng tôi đến với những câu hỏi nghiên cứu mới.

Chương C2 : Để kiểm chức các giả thuyết nghiên cứu về hợp đồng sư phạm ~~didactic~~, chúng tôi tiến hành một bộ câu hỏi điều tra dành cho học sinh và giáo viên. Bộ câu hỏi được xây dựng trên bốn bài tập dành cho học sinh lớp 10, sau khi đã học thống kê và dành cho giáo viên toán đã dạy thống kê.

Phần D

Đây là phần dành cho đồ án sư phạm, bao gồm hai chương :

Chương D1 : Trong chương này, trước hết chúng tôi giải thích vì sao cần đưa ra một đồ án sư phạm. Ở đây, chúng tôi còn giới thiệu hai đồ án của Brousseau và Briand mà chúng tôi đã sử dụng để xây dựng các tình huống cơ bản cho đồ án của chúng tôi ở chương D2. Chương này cũng nêu lên các mối quan hệ thể chế đối với các khái niệm về phép thử và sự không chắc chắn của ngẫu nhiên.

Chương D2 : Chương này dành cho việc trình bày thực nghiệm đã thực hiện ở Việt nam và phân tích ba buổi thực nghiệm thuộc đồ án đã được xây dựng.

IV. Những kết quả chính của luận án

IV.1. Ba bài toán - ứng viên

Phân tích các sách đại học cho phép chúng tôi chỉ ra ba *bài toán-ứng viên*. Chúng tôi nhắc lại rằng thuật ngữ *bài toán-ứng viên* được sử dụng để chỉ những bài toán có thể làm cho các mối quan hệ thống kê-xác suất « sống » được trong một thể chế dạy học toán.

Bài toán về sự không chắc chắn của các độ đo : bài toán nguồn gốc lịch sử của luật phân bố thường này là nơi có thể « gặp » luật các số lớn và phân bố chọn mẫu của giá trị trung bình ;

Bài toán so sánh giữa các phân bố thống kê thực nghiệm hay so sánh các phân bố thống kê thực nghiệm với các phân bố thống kê lý thuyết : bài toán này là nơi có thể tiếp cận gần của các đối tượng chung của thống kê mô tả và lý thuyết xác suất, như các khái niệm biến (thống kê hay ngẫu nhiên), phân bố (thống kê hay xác suất), v.v…

Bài toán về tính đại diện của một mẫu dữ liệu : bài toán gốc của thống kê suy diễn và của lý thuyết kiểm định này là nơi có thể « gặp gỡ » luật số lớn và phân bố của việc chọn mẫu.

Để tạo nên được các mối liên hệ thống kê-xác suất, mỗi bài toán trong các bài toán này cần đến các đối tượng mà chúng tôi gọi là các *đối tượng-ứng viên* : các phân bố tần suất hay các phân bố xác suất với các ostensif như biểu đồ, bảng, số và đại số, luật các số lớn, việc chọn mẫu.

Tuy nhiên, trong thể chế dạy học trung học phổ thông Việt nam cũng như Pháp, chúng tôi nhận thấy bài toán ứng viên về *sự không chắc chắn của độ đo* hoàn toàn vắng mặt. Do đó, chúng tôi sẽ chỉ đi sâu nghiên cứu hai tổ chức tham chiếu sau đây cho dạy học ở trung học phổ thông Việt nam. Đó là hai tổ chức tương ứng với hai bài toán cuối :

- Tổ chức so sánh

- Tổ chức chọn mẫu dữ liệu

IV.2. Các tổ chức toán học trong dạy học toán ở trung học Việt nam

- **Một tổ chức thống trị** : đó là tổ chức mô tả một dãy số liệu thống kê. Tổ chức này bao gồm các tổ chức địa phương như mô tả dãy số liệu thống kê bằng biểu đồ hay bằng các chỉ số đặc trưng ; tổ chức này phát triển các kiểu nhiệm vụ và các kỹ thuật để thu gọn, đặc trưng và tổng hợp một dãy số liệu ; nó cũng hướng đến việc so sánh mỗi số liệu với tập hợp các số liệu. Những yếu tố kỹ thuật chính là bảng, biểu đồ và véctơ có thành phần là các chỉ số đặc trưng.

- **Một tổ chức chưa hoàn thiện :** đây là một tổ chức so sánh các dãy số liệu thống kê, được hình thành từ các tổ chức địa phương như so sánh biểu đồ và so sánh số ; nó cũng hướng đến việc so sánh giữa các tập số liệu con của cùng một tổng thể dữ liệu hay giữa một tập số liệu con với tổng thể dữ liệu.

Đặc biệt, phân tích thể chế dạy học toán lớp 10 của Việt nam cho thấy đường gấp khúc tần suất được sử dụng như một công cụ để so sánh hai phân bố thống kê trong khi nó hoàn toàn vắng mặt trong chế dạy học toán của Pháp. Tuy nhiên, việc so sánh hai đường gấp khúc tần suất thực ra không phải được thực hiện trên việc so sánh diện tích miền phẳng nằm dưới đường gấp khúc tần suất mà là so sánh vị trí tương đối của hai đường gấp khúc. Cũng cần nói rõ rằng diện tích miền phẳng này chính là mật độ tần suất.

- **Một tổ chức vắng mặt :** đây là một tổ chức có các kiểu nhiệm vụ là : lấy ra một mẫu dữ liệu bằng cách lặp lại một phép thử ngẫu nhiên, nghiên cứu sự biến động của quá trình chọn mẫu, tức là hoặc nghiên cứu trên một tần suất riêng, hoặc nghiên cứu trên một chỉ số đặc trưng như giá trị trung bình, nghiên cứu việc giảm độ biến động bằng cách tăng kích thước của mẫu dữ liệu.

Tuy nhiên, một số khái niệm cần dạy như mẫu dữ liệu, đại diện của mẫu dữ liệu và mẫu dữ liệu đại diện cho tổng thể (riêng khái niệm này được giới thiệu một cách ngầm ẩn) là những yếu tố thuận lợi cho việc xây dựng tổ chức chọn mẫu dữ liệu.

Sự phức tạp của tổ chức này là nguyên nhân khiến cho thống kê tồn tại theo kiểu tách rời với xác suất trong dạy học toán ở trung học tại Việt nam. Mối liên hệ duy nhất có được là giữa tần suất và xác suất trong định nghĩa thống kê của xác suất (ở lớp 11) nhưng lại không được học sinh thực hành. Mối liên hệ này rất tinh tế vì nó không tạo điều kiện để xây dựng tổ chức về luật các số lớn và cũng như sự biến động của việc chọn mẫu. Mục tiêu chính của đồ án didactic của chúng tôi là làm tăng cường mối quan hệ này.

Chỉ có bài toán so sánh là *bài toán-ứng viên* duy nhất trong các *bài toán-ứng viên* được giảng dạy trong dạy học toán ở phổ thông Việt nam nhưng lại thiếu đi các mô hình xác suất rời rạc hay liên tục. Điều này làm cản trở việc so sánh biểu đồ hay so sánh những đặc trưng số giữa các phân bố thực nghiệm.

IV.3. Quan hệ thể chế đối với các đối tượng-ứng viên

Trong thể chế dạy học toán trung học phổ thông Việt nam chỉ có hai *đối tượng-ứng viên* hiện diện, đó là các phân bố thống kê được phô bày qua biểu đồ và chỉ số đặc trưng.

- **Các biểu đồ**

Dựa trên việc so sánh các tổ chức cần giảng dạy với tổ chức tham chiếu, chúng tôi đã thiết lập nên giả thuyết nghiên cứu sau :

> Sự tồn tại của quan điểm hàm số trên các biểu đồ thống kê là một điều kiện thuận lợi cho việc xây dựng các mối liên hệ thống kê-xác suất về tổ chức so sánh các phân bố tần suất.

Từ giả thuyết này chúng tôi đã đưa ra câu hỏi :

Câu hỏi 5 : Trong tổ chức so sánh của thể chế việt nam, có tồn tại quan điểm hàm số trên các biểu đồ thống kê không ?

Để kiểm chứng giả thuyết, chúng tôi thực hiện một nghiên cứu thực nghiệm với bộ câu hỏi điều tra gồm 4 bài tập, dành cho học sinh và giáo viên.

Kết quả phân tích câu trả lời cho các bài tập 1 và 2 cho phép chúng tôi đi đến kết luận: ở thể chế Việt Nam, quan điểm hàm số không hề vắng mặt trong quá trình hình thành cách đọc biểu đồ, nhưng nó lại gây ra một *chướng ngại sư phạm cho việc hình thành các mối quan hệ thống kê-xác suất*. Cụ thể là việc hình thành các mối quan hệ như vậy cần đến khái niệm diện tích biểu đồ hình chữ nhật và diện tích miền nằm dưới đường biểu diễn (mật độ), trong khi khái niệm này lại bị thể chế « lờ đi ». Chướng ngại này càng thể hiện rõ ở sự vắng mặt khái niệm tần suất tích lũy và vì vậy kéo theo sự vắng mặt của hàm phân bố trong thể chế dạy học Việt nam.

- **Các chỉ số đặc trưng**

Đóng vai trò đại diện cho một dữ liệu thống kê, việc thu gọn dữ liệu thống kê dưới dạng các chỉ số đặc trưng là một công cụ của việc so sánh một mẫu dữ liệu với dữ liệu tổng thể. Kết quả phân tích câu trả lời dành cho câu hỏi 3 và 4 trong bộ câu hỏi điều tra thực nghiệm đã cho phép làm nổi rõ các kết quả nghiên cứu sách giáo khoa : chỉ số đại diện giữ vai trò thống trị được mặc định trong thể chế là giá trị trung bình, hoặc là cặp *giá trị trung bình/độ lệch chuẩn*. Đây là một tiêu chuẩn rất mạnh nhằm bảo toàn phân bố từ dữ liệu tổng thể trên một mẫu dữ liệu đại diện. Thực nghiệm cũng làm sáng tỏ một vấn đề là tiêu chuẩn « *gần như tỉ lệ* chiếm một vị trí quan trọng trong quan niệm của học sinh và giáo viên. Họ thường sử dụng tiêu chuẩn này để đánh giá tính đại diện của một mẫu dữ liệu so với dữ liệu tổng thể.

- **Một tổ chức chọn mẫu cần được xây dựng**

Việc dùng đến chọn mẫu ngẫu nhiên đơn giản để làm thống kê trên một dữ liệu tổng thể chưa biết rõ nhằm trả lời cho các câu hỏi trên tổng thể này không phải là đối tượng giảng dạy được tích hợp trong lĩnh vực thống kê. Chỉ trong giảng dạy xác suất ở lớp 11, mới thấy nói đến xác suất thực nghiệm. Ở đây, xác suất thực nghiệm nhận được bằng cách lặp lại một phép thử ngẫu nhiên nhiều lần và được xem là một giá trị gần đúng của một xác suất chưa biết. Sự thu hẹp của thống kê suy diễn vào việc ước lượng điểm một con số (xác suất) bởi một số khác (tần suất) đã bó hẹp sự suy diễn trong khuôn khổ xấp xỉ số và một quan điểm hàm số cho luật các số lớn - ở đây, xác suất thực nghiệm là một hàm số theo số phép thử mà giới hạn ở vô hạn của hàm số này là xác suất thực[177].

Để xây dựng tổ chức chọn mẫu, chúng tôi thiết kế và tiến hành thực nghiệm thứ hai, đó là một đồ án didactic. Chứng tỏ rằng *cách nhìn theo hàm số của luật các số lớn gây ra một chướng ngại khoa học luận cho cách nhìn theo xác suất* [178] vì nó không thể hiện một độ đo cần thiết cho rủi ro mắc phải khi quyết định lấy một giá trị này hay giá trị kia cho xác suất, dù là trong thực nghiệm.

IV.4. Đồ án sư phạm

Đồ án sư phạm đã cho phép quay lại với các phân tích khoa học luận, phân tích thể chế và bổ sung cho chúng. Sợi dây chỉ đạo xuyên suốt đồ án sư phạm này là quá trình biến đổi từng bước của luật các số lớn, nhằm tạo ra công cụ để thực hiện một quyết định với ý thức về sự có mặt của rủi ro với một độ đo thực nghiệm (của rủi ro này) trong bài toán về sự không chắc chắn ngẫu nhiên. Sự biến đổi này được xây dựng dựa trên việc so sánh các mẫu dữ liệu bằng cách dàn dựng nên một biến động của quá trình chọn mẫu ; ngoài luật các số lớn, sự biến đổi ấy còn liên quan đến khái niệm mẫu dữ liệu và khái niệm mức độ chắc chắn của việc ra quyết định.

---

[177] Chúng tôi gọi xác suất thực này là tần suất lý thuyết trong trường hợp phép thử ngẫu nhiên là thuộc dạng Bernoulli.

[178] Chúng tôi đã nói về chướng ngại « hàm số ».

Đối tượng / Buổi	Luật số lớn	Sự biến động của việc chọn mẫu	Rủi ro
Buổi thứ nhất	Luật số lớn là một hành động trong việc chọn lựa tần suất lý thuyết.	Sự biến động của việc chọn mẫu là một hiện tượng tự nhiên, không kiểm soát được và không đo được.	Mức độ của rủi ro : - không gắn liền với bất kỳ hệ số nào (số lần thực hiện phép thử hay biên độ của tần suất thực nghiệm với tần suất lý thuyết chẳng hạn) - không được biểu thị trong ngôn ngữ xác suất.
Buổi thứ hai	Luật số lớn là một đối tượng với hai thuộc tính đồ thị : - một tiệm cận ngang đối với đường biểu diễn tần suất được xem như một hàm số của số phép thử, - hình ảnh các khoảng chứa lẫn nhau trên trục số thực.	Sự biến động của việc chọn mẫu là đo được : khoảng biến động của tần suất được thu hẹp thành khoảng quyết định lý thuyết (với sự chắc chắn).	Rủi ro bị triệt tiêu kể từ một số lần thực hiện phép thử được cung cấp bởi thực nghiệm.
Buổi thứ ba	Luật số lớn là một công cụ đánh giá rủi ro khi lấy quyết định.	Sự biến động của việc chọn mẫu là đo được : khoảng biến động của tần suất thực nghiệm trở thành một khoảng tin cậy thực nghiệm (có rủi ro).	Rủi ro chính là xác suất thực nghiệm của việc lấy quyết định sai.

Tiến triển của ba đối tượng trong ba buổi thực nghiệm của đồ án sư phạm

Buổi thứ ba của thực nghiệm là nhằm xây dựng sự chuyển đổi *khoảng quyết định* (lý thuyết) thành *khoảng tin cậy* (thực nghiệm). Sự biến đổi này là cần thiết để làm xuất hiện khái niệm rủi ro của thống kê suy diễn.

Đồ án sư phạm đã được thực hiện tại một lớp 11 ở Việt nam. Thực nghiệm này đã bổ sung cho mối quan hệ cá nhân của học sinh với đối tượng mẫu dữ liệu và cũng chứng tỏ một *sự bổ sung có thể cho mối quan hệ thể chế với đối tượng này* :

- một mặt, một mẫu dữ liệu là sản phẩm của việc lặp lại một phép thử ngẫu nhiên và không chỉ là một trích xuất từ dữ liệu tổng thể theo nghĩa tập con của tổng thể ấy ;

- mặt khác, mẫu dữ liệu là một công cụ để « biết được[179] » một tổng thể chưa biết, nhờ vào tính biến đổi của các mẫu dữ liệu có cùng kích cỡ.

[179] Theo nghĩa suy diễn ra các đặc trưng của dữ liệu tổng thể.

Thực nghiệm cũng khẳng định sự tồn tại của chướng ngại « hàm số » trong việc xây dựng nghĩa xác suất cho luật các số lớn.

Trong một chừng mực nào đó, thực nghiệm này cũng đưa ra một bước khởi đầu cho tổ chức chọn mẫu để xây dựng trong thể chế dạy học toán ở trung học của Việt nam. Thực nghiệm đã làm sáng tỏ một số điều kiện sống cho một tổ chức như vậy trong thể chế dạy học Việt nam. Bước chuyển từ chỉ một sự ổn định tần suất sang cặp *tính biến đổi/tính đều đặn* của các mẫu dữ liệu cùng kích cỡ là một trong những điều kiện này và trong phương pháp tần suất, ta buộc phải dự kiến một sự lặp lại cùng một phép thử, không phải dưới dạng luỹ tích các phép thử đơn vị mà phải là dưới dạng tập hợp các mẫu dữ liệu với cùng một số lần thực hiện phép thử.

V. Giới hạn và hướng nghiên cứu mới

Từ việc thiết lập lại các lựa chọn chính, chúng tôi sẽ đề nghị những hướng nghiên cứu mới trong bối cảnh thể chế Việt nam.

V.1. Lựa chọn thể chế

Nghiên cứu của chúng tôi thực hiện trên các khái niệm trong thống kê đã được các nhà soạn thảo chương trình và sách giáo khoa Việt nam chỉ định. Thể chế mà chúng tôi nhắm đến là thể chế giảng dạy thống kê trong dạy học toán ở trung học Việt nam và để ngoài lề việc giảng dạy xác suất. Ngoài lý do là chúng tôi đã có một số kết quả trong một nghiên cứu gần đây về giảng dạy xác suất[180], chúng tôi còn có hai lý do khác cho chọn lựa này, đó là :

- ở Việt nam, thống kê được dạy trước xác suất và điều này làm chúng tôi nghĩ đến các điều kiện để bổ sung cho các mối quan hệ cá nhân của học sinh việt nam đối với khái niệm xác suất một khi học sinh đã được học thống kê ;

- dạy học toán ở trung học của Pháp thể hiện những chọn lựa về giờ giấc, chương trình, thực hành giảng dạy, … trong giảng dạy thống kê khác với Việt nam và vì vậy, chúng tôi có thể dựa trên phân tích so sánh giữa hai thể chế.

Bằng cách đi từ vấn đề giảng dạy thống kê và xác suất, việc mở rộng nghiên cứu đã thực hiện ở mức độ dạy học toán ở trung học lên nghiên cứu trên mức độ đại học là một điều cần được thực hiện. Một trong những câu hỏi khởi đầu có thể là câu hỏi liên quan đến sự sinh động của việc *tiếp tục/ngắt quãng* với dạy học toán ở trung học. Chúng tôi cũng có thể dựa vào phần phân tích sách giáo khoa đại học cũng như trên việc phân tích các tài liệu giảng dạy đại học của Việt nam[181].

---

[180] Luận văn thạc sĩ của Vũ Như Thư Hương [2005]

[181] Chúng tôi đã tham khảo sách của tác giả Nguyễn Cao Văn và Trần Thái Ninh (2005)

V.2. Lựa chọn của đồ án sư phạm

Chúng tôi xin nói lại là chúng tôi đã mượn những ý tưởng trong đồ án của Brousseau và Briand. Chúng tôi xin nêu lại ba quyết định khi lấy lại các ý tưởng này. Những quyết định ấy cho phép đánh giá cao những gì mà đồ án « của chúng tôi » có thể đem lại cho việc giảng dạy :

- chúng tôi đã thực hiện thực nghiệm trên những học sinh không chỉ được học thống kê mà còn đã được học xác suất ; quyết định chọn như vậy (đối lập với việc chỉ thực hiện thực nghiệm trên các học sinh chưa được học xác suất) là để tạo thuận lợi cho việc sử dụng các khái niệm xác suất và thuận lợi cho ngôn ngữ xác suất xuất hiện trong thành ngữ của sự *không chắc chắn* khi suy diễn từ các kết quả của phép thử ngẫu nhiên được lặp lại ; điều mà chúng tôi thu nhận được từ thực nghiệm buổi thứ nhất của đồ án didactic, đó là đối với học sinh, khái niệm xác suất không phải là khái niệm được hàm số thao tác trong bài toán về sự không chắc chắn về tỉ lệ bbi đỏ và bi xanh chưa biết (của cái chai)[182] ;

- chúng tôi đưa ra khái niệm mẫu dữ liệu đại diện cho một tổng thể trong một bài tập thuộc bộ câu hỏi điều tra trước khi thực hiện đồ án didactic và một lần thứ hai khi thiết kế buổi thực nghiệm thứ 3 của đồ án didactic. Buổi thực nghiệm thứ 3 này thoát ra khỏi đồ án của Brousseau và Briand theo hướng là mô hình xác suất được sử dụng trong buổi thứ 3 một mô hình siêu bội với một tổng thể cực lớn mà các khoảng quyết định không thể tính toán trước được ; quyết định này của chúng tôi đi tiếp quyết định đã được thông báo là tổ chức tình huống « cá cược » như tác giả Brousseau đã mô tả trong Brousseau [2005] [183] ; thực ra, chúng tôi đã nghĩ rằng một tình huống như vậy không thể khai thông được trong việc học sự biến động của quá trình chọn mẫu như là một công cụ để đo xác suất mà chúng tôi nhắm đến ;

Chúng tôi sử dụng một giả lập bằng tin học theo các mô hình nhị phân và mô hình siêu bội để giả lập các tình huống cụ thể đã được trình bày (tình huống cái chai rồi đến tình huống điểm) để học sinh truy cập được các mẫu dữ liệu ngẫu nhiên đơn giản mà ít phải trả giá ; một sự giả lập như vậy vốn không có mặt trong dạy toán ở trung học Việt nam ; quyết định này cho phép đưa sự biến động của quá trình chọn mẫu lên trước màn nhưng quyết định này không vội đoán định tính biến đổi của việc giả lập tin học trong dạy toán ở trung học Việt nam. Một trò chơi (mà chúng tôi không nghiên cứu) được tổ chức bằng cách thay thế cái chai thật bởi cái chai–máy tính. Để nói một chút về vấn đề này, chúng tôi xin trích dẫn điều mà Girard và Henry (2005)[184] đã nói về cải cách chương trình ở Pháp năm 2000 :

> Không được coi thường lợi ích của công cụ tin học. Nó cho phép tiếp cận các tình huống thăm dò một cách thực nghiệm, và nếu giấu học sinh tỉ lệ p, chúng ta có được một công cụ giải quyết vấn đề cũng giống như vai trò của máy tính bỏ túi có đồ thị khi máy tính vẽ các đường cong biểu diễn cho các hàm số được nhập vào.

---

[182] Giả định rằng ta nhận được 7 bi đỏ và 3 bi xanh sau 10 lần lắc chai ; việc chọn một tỉ lệ Đỏ/Xanh cho các vien bi chứa trong chai có thể được giải quyết bằng cách so sánh các xác suất tiên nghiệm để nhận được biến cố 7 đỏ/3 xanh với xác suất của mỗi tỉ lệ có thể.

[183] Xác suất được nhắm đến trong đồ án didactic của chúng tôi cũng như của Brousseau là xác suất của biên độ | tần suất thực nghiệm - tần suất lý thuyết| nhỏ hơn một ngưỡng rủi ro nào đó.

[184] Trích trong « Thống kê ở trung học phổ thông » (2005) của Commission Inter-Irem.

> Nhờ cách sắp đặt tuỳ theo tình hình gợi lại một tình huống ngẫu nhiên cụ thể, máy tính lại được xem một cách ngầm ẩn như là một chiếc máy sản sinh ra cái ngẫu nhiên mà sự ngẫu nhiên này được lập trình cụ thể, nó cung cấp các kết quả đại diện một cách hoàn hảo cho các kết xuất mà một phép thử cụ thể có thể cung cấp. Sự đồng hóa này không phải là tất nhiên đối với tất cả các học sinh. Máy tính cho ngẫu nhiên các chữ số từ 1 đến 6 đồng khả năng nhưng *đó không phải là một con xúc sắc thật*, mà cũng không phải là xúc sắc gian lận. Sự tương đương để được chấp nhận đòi hỏi một cách ngầm ẩn một ý tưởng là xác suất xuất hiện mặt của con xúc sắc là bằng với xác suất xuất hiện số 1 của máy tính, nghĩa là khái niệm xác suất đã được cài đặt. Việc sử dụng máy tính để sản sinh ra khái niệm xác suất, như đại diện cho một phép thử cụ thể, đặt lại một vấn đề trên một cái vòng luẩn quẩn didactic, hoặc là trên một hợp đồng didactic trước không thể nói rõ : chính là cái ngẫu nhiên của con xúc sắc được biểu thị bởi máy tính. Chính vì điều kiện này mà việc thực hiện phép thử ngẫu nhiên khái thác chiếc máy sản sinh ra ngẫu nhiên của máy tính có thể được đồng hóa với một sự giả lập, từ ngữ có mặt khắp nơi trong chương trình [chương trình của Pháp]. Nên rào quanh một cách chính xác hơn cương vị của giả lập tin học sao cho nó được trình bày trong lớp học.

(Girard&Henry [2005], trang 151-152)

Chúng tôi cho là khả năng « sống » của việc sử dụng tin học để mô phỏng trong dạy học ở phổ thông Việt nam là rất đáng để nghiên cứu. Điều này là một trong các hướng nghiên cứu mới mà chúng tôi sẽ đề cập ngay sau đây.

V.3. Một số hướng nghiên cứu mới

V.3.1. Việc đào tạo giáo viên toán

Chúng tôi nhắc lại là việc đào tạo giáo viên ở trung học cơ sở và trung học phổ thông tại Việt nam là tách biệt nhau. Với bộ câu hỏi điều tra dành cho giáo viên dạy lớp 10, chúng tôi có thể đưa ra một giả thuyết là giáo viên tuân thủ sách giáo khoa rất nghiêm ngặt và coi đó như đại diện của thể chế dạy học. Không có gì là ngạc nhiên khi giả thuyết này chỉ ra rằng giáo viên chính là những người bảo đảm cho các mối quan hệ thể chế đối với các đối tượng của thống kê. Từ đó, chúng tôi tự đặt câu hỏi về các quan hệ cá nhân của giáo viên với thống kê và xác suất. Để nghiên cứu các mối quan hệ cá nhân này, chúng tôi cho rằng cần quan tâm đến các việc đào tạo các giáo viên, không chỉ về trong chương trình toán mà còn cả trong chương trình môn phương pháp giảng dạy. Kiến thức về thống kê - xác suất và về phương pháp dạy học mà người ta trang bị cho giáo viên có tạo ra được chỗ đứng cho các mối liên hệ giữa thống kê và xác suất không ? Chỗ đứng cho những mối liên hệ này có giống nhau trong đào tạo giáo viên trung học cơ sở và giáo viên trung học phổ thông hay không ?

V.3.2. Môi trường công nghệ

Chủ đề này được nói đến với bài toán về việc giả lập bằng tin học. Đôi khi, trong môi trường công nghệ dành riêng cho việc học một khái niệm toán học, vấn đề hình thành công cụ ban đầu là không thể tránh khỏi. Ở Việt nam, học sinh và giáo viên chỉ có sẵn máy tính bỏ túi với nút chức năng « random ». Tuy nhiên, việc trang bị các phòng tin học và trang bị phương tiện tin học ngay tại từng lớp học đang phát triển mạnh. Các giáo viên sẽ tận dụng

các phương tiện tin học có sẵn này như thế nào để dạy học thống kê và xác suất ? Những nguồn tư liệu didactic nào cần được chuẩn bị sẵn cho giáo viên sử dụng [185] ?

V.3.3. Vấn đề liên môn

Chủ đề này không được đề cập đến trong nghiên cứu của chúng tôi như đã nêu lý do ở phần giới thiệu đề tài nghiên cứu. Chúng tôi biết là giáo viên các bộ môn khác cũng có được đào tạo về xác suất và thống kê. Làm thế nào để việc đào tạo này dựa trên các vấn đề toán học đặt ra cho nguồn gốc của việc chọn mẫu dữ liệu và việc thăm dò ? Những công cụ toán học nào được các giáo viên này sử dụng để thực hiện các nghiên cứu thống kê trong việc dạy môn học của mình ? Làm thế nào để việc liên môn giữa toán học và một môn học khác có thể phát triển trong thể chế Việt nam ?

[185] Có thể nào xây dựng một website « online » để giả lập phép thử ngẫu nhiên như http://homeomath.imingo.net/simulations.htm không ?

Annexes

Table des matières

Annexe A **371**
I. Institution française 371
II. Institution vietnamienne.....422
III. Analyse des sujets de Brevet et de Baccalauréat français 445
Annexe B **473**
I. Type de tâche T1 : 474
II. Type de tâche T2 :.....493
Annexe C **501**
I. Questions pour l'enseignant 501
II. Questions pour l'élève.....502
III. Quatre exercices.....503
V. Câu hỏi dành cho giáo viên.....508
VI. Câu hỏi dành cho học sinh.....508
VII. Bốn bài tập 510
Annexe D **515**
I. Programmes informatiques.....515
I.1. Séance 2 : Bouteille_Connue.....515
I.2. Séance 2 : Bouteille_Inconnue 518
I.3. Séance 3 : Situation_Note.....521
II. Fiches de l'ingénierie didactique 527
III. Scripts des séances.....531
III.1. Script de la séance 1.....531
III.2. Script de la séance 2.....537
III.3. Script de la séance 3.....547

Annexe A

I. Institution française

Dans cette étude, nous faisons une analyse des programmes et des manuels de la classe de sixième à la classe Terminale (série S) sur les notions concernant la statistique et les probabilités pendant des années juste avant et après la réforme en rangeant tous les types de tâches trouvés dans les 6 groupes présentés dans la partie A.

Groupes	Types de tâche
Groupe 0 : recueillir des données [186] dans une population, exhaustivement ou par sondage.	- Poser une question qui motive l'étude statistique - Procéder à une enquête - Organiser le recueil de données brutes
Groupe 1 : regrouper et représenter	- Regrouper les données brutes - Calculer effectifs, pourcentages, fréquences → distributions des effectifs et des fréquences - Représenter des données statistiques fournies : - par tableaux - par graphiques
Groupe 2 : résumer numériquement	Calculer les indicateurs : - de position - de dispersion
Groupe 3 : comparer des distributions	Comparer : - les tableaux - les graphiques - les résumés numériques
Groupe 4 : exploiter les données traitées	- Interpréter - Commenter
Groupe 5 : répéter une expérience aléatoire, produire et analyser des séries d'échantillons, modéliser une expérience aléatoire	- Observer la fluctuation d'échantillonnage - Modéliser et simuler une expérience aléatoire - Evaluer les indicateurs d'une population à partir d'un échantillon

Tableau 93. Les six groupes des types de tâches

Les cases grisées sont pour les groupes qui apparaissent au fur et à mesure des changements de programme.

Pour voir l'évolution des types de tâches dans le manuel, nous numérotons les types de tâches en gardant ces numéros pour tous niveaux analysés.

[186] On recueille les données avant de les traiter. Cela explique le numéro 0 donné à ce groupe.

I.1. Avant la réforme (période 1986-1999)

I.1.1. Collège :

Programme :

Cette analyse est visée le programme du collège pendant la période 1986-1999. A partir de l'année scolaire 1986, la partie de *Représentation et Organisation de données* occupent un chapitre dans le programme mathématique de sixième à troisième. Comme il y a deux programmes mis en place, nous faisons un résumé en tableau des connaissances essentielles et des compétences exigibles retirés des programmes d'après deux épisodes de quatre ans : 1986-1989 et 1995-1998.

		Classe 6è	Classe 5è	Classe 4è	Classe 3è
1986-1989[187]	Programme	1986 Lecture, interprétation et réalisation de tableaux et de graphiques.	1987	1988 Fréquences, expression en pourcentage. Effectifs cumulés, fréquences cumulées.	1989 Moyenne, moyennes pondérées. Médiane.
	Compétences exigibles	Appliquer un tableau de pourcentage. Effectuer, éventuellement avec une calculatrice, des calculs sur les mesures de grandeurs figurant au programme.	Lire des données statistiques présentées sous la forme de tableaux ou de représentations graphiques Traduire des données statistiques sous la forme d'un diagramme en bâton.	Savoir lire des données statistiques présentées sous la forme de tableaux ou de diagramme d'effectifs ou de fréquences. A partir de données statistiques, présenter les effectifs ou les fréquences dans des tableaux et tracer les diagrammes correspondants.	Savoir lire et exploiter des données statistiques mises sous forme de tableaux ou de diagrammes d'effectifs ou de fréquences, savoir calculer une moyenne. A partir de données statistiques, calculer les effectifs ou les fréquences, les présenter dans les tableaux et tracer les diagrammes correspondants.

[187] D'après le *Tableau synoptique pour le collège* fourni par M.E.N. Mathématiques 6è, 5è, 4è, 3è (1989), Horaires, Objectifs, Programmes, Instructions.

1995-1998[188]	Programme	1995 Exemple conduisant à lire, à établir des tableaux, des graphiques.	1996 Classes, effectifs d'une distribution statistique. Fréquences. Diagrammes à barres, diagrammes circulaires.	1997 Effectifs cumulés. Fréquences cumulées. Moyennes. Initiation à l'usage de tableurs-grapheurs.	1998 Approche de la comparaison de séries statistiques.
	Compétences exigibles		Lire et interpréter un tableau, un diagramme à barres, un diagramme circulaire ou semi-circulaire Regrouper des données statistiques en classes, calculer des effectifs. Présenter une série statistique sous la forme d'un tableau, la présenter sous la forme d'un diagramme à barres ou d'un graphique. Calculer des fréquences.	Calculer des effectifs cumulés, des fréquences cumulées. Calculer la moyenne d'une série statistique. Calculer une valeur approchée de la moyenne d'une série statistique regroupée en classes d'intervalles.	Une série statistique étant donnée (sous forme de liste ou de tableau, ou par une représentation graphique) proposer une valeur médiane de cette série et en donner la signification. Une série statistique étant donnée, déterminer son étendue ou celle d'une partie donnée de cette série.

Tableau 94. Les programmes et les compétences exigibles du collège de 1986 à 1999

Les premières connaissances de la statistique enseignées au collège sont pour représenter et organiser les données :

Dans le premier épisode :

Les connaissances exigibles restent simplement dans la lecture des tableaux, des représentations graphiques et dans leurs réalisations aux niveaux de 6è et de 5è, puis les calculs des indicateurs comme effectif, effectif cumulé, fréquence, fréquence cumulée sont présentée à la classe 4è. Moyenne, moyenne pondérée et médiane sont pour la classe 3è. Ces notions se présentent discrètement dans le programme de collège. On ne demande pas l'élève d'exploiter les données statistiques. Il est difficile à trouver la raison d'être de la notion de fréquence, de moyenne et de médiane dans le processus d'enseignement de statistique au collège.

Dans l'épisode suivant :

- Le programme ne change pas beaucoup pour la classe 6. Cependant, dès la classe 5è, nous voyons l'apparition de nouvelles notions comme classes, effectif et particulièrement, la fréquence. Ces notions ne sont qu'engagées à la classe de 4è pendant l'épisode d'avance. De même cas, à la classe 4è on trouve la présence de moyenne.

[188] Extrait du *Tableau synoptique pour le collège* écrit dans Le B.O. N°10 Oct 1998 HORS-SERIE.

Ainsi, le programme « a poussé » la statistique en avant un an. Par contre, la notion de médian est disparue pendant cet épisode.

	5è	4è	3è
1986-1989		Effectif Fréquence Effectif cumulé Fréquence cumulée	Moyenne Médiane
1995-1998	Effectif, Fréquence	Moyenne Effectif cumulé Fréquence cumulée	Médiane Etendue

Tableau 95. Les notions déplacées vers le niveau immédiatement inférieur

- Avec le progrès de la technique moderne comme l'ordinateur, le programme demande les élèves de classe 4è d'utiliser les tableurs-grapheurs.

- Au niveau de 3è, l'élève s'approche à la « comparaison des séries statistiques ».

Manuel :

Pour analyser les manuels du collège, nous choisissons une série des années consécutives : 1986, 1987, 1988, 1989, en concret, ils sont :

- Manuel 6è, NATHAN, 1986

- Manuel 5è, NATHAN, 1987

- Manuel 4è, Terracher, HACHETTE, 1988

- Manuel 3è, Terracher, HACHETTE, 1989

Pour chaque classe, nous cherchons les types de tâches et les mettons en groupe.

Manuel 6ième : manuel 6ième, NATHAN, 1986 (p. 181-194)

Chapitre 15 : **Proportionnalité, pourcentage**

Groupe	T	Types de tâches
0	T1	Simuler une expérience aléatoire n fois / Etablir une série statistique
1	T5	Lire des informations d'un tableau
	T7	Représenter en graphique des données statistiques
	T11	Chercher un exemple de « grandeurs » proportionnelles / non proportionnelles
2	T25	Compléter un tableau (des « grandeurs » proportionnelles)
	T26	Chercher le coefficient de proportionnalité d'un tableau (des « grandeurs » proportionnelles)
	T27	Vérifier si un couple de deux entiers est proportionnel à un autre couple

Tableau 96. Les types de tâches du manuel 6ième, NATHAN, 1986

- Les types de tâches dans ce manuel de 6è concentrent à la représentation et organisation des données comme ce qui sont indiqués dans le programme. Au niveau de la classe 6è, on n'introduit que la notion de proportionnalité et celle de pourcentage.

- Dans le manuel, on trouve le mot la suite pour des valeurs de même ligne du tableau des valeurs proportionnelles. Ainsi, des suites proportionnelles sont pour dire des « grandeurs » proportionnelles.

- Comme le chapitre « Proportionnalité, pourcentage » est le suivant du chapitre « Représenter une fonction », nous trouvons que la représentation graphique des données est pour faire un retour à la représentation graphique d'une fonction.

Manuel 5ième : manuel 5ième, NATHAN, 1987 (p. 164-169)

Chapitre 26 : **Relevés statistiques**

Groupe	T	Types de tâches
0	T1	Simuler une expérience aléatoire n fois / Etablir une série statistique
1	T7	Représenter en graphique des données statistiques
	T12	Calculer la mesure d'angle d'un diagramme qui représente des données statistiques
2	T33	Transformer des répartitions de données en pourcentages
	T36	Calculer la moyenne
	T38	Calculer une valeur pour ajouter à une liste des valeurs statistiques telles que la moyenne ne change pas
	T39	Expliquer comment on peut calculer la moyenne pour l'ensemble de population
	T60	Prévoir une évolution des moyennes lors de changement de population
3	T62	Comparer les moyennes des séries statistiques

Tableau 97. Les types de tâches du manuel 5ième, NATHAN, 1987

Dans ce manuel de 5è, on trouve T21 concernant la moyenne qui n'est pas demandé dans le programme. En plus, T4 peut être compris comme une demande implicite sur la fréquence qui n'existe pas dans le programme. Particulièrement, nous nous intéressons à l'exercice 3.15

3.15 Dans un pays comme la France, explique comment on peut calculer, pour l'ensemble de la population :

a) L'âge moyen.

b) Le salaire moyen.

c) Le nombre moyen d'enfants par famille.

d) La durée de vie moyenne.

Note bien que ces moyennes n'ont de sens que si l'on dit à quel moment on fait le calcul. Comment penses-tu que ces moyennes évolueront dans l'avenir. Augmenteront-elles ou diminueront-elles ?

(Manuel 5ième, NATHAN, 1987, p. 168)

car :

- Par la question : « *Dans un pays comme la France, explique comment on peut calculer, pour l'ensemble de la population…* », nous pensons qu'il aborde implicitement la question de choisir une échantillon, c'est-à-dire l'échantillonnage.

- La question « *Comment penses-tu que ces moyennes évolueront dans l'avenir. Augmenteront-elles ou diminueront-elles ?* » demande d'une **estimation d'évolution** des moyennes lorsque la population change au fur et à mesure de temps. Alors, aborde-elle à une fluctuation d'échantillonnage ?

Ce manuel propose effectivement des travaux beaucoup avancés par rapport au programme.

Manuel 4ième : manuel 4ième, collection Terracher, HACHETTE, 1988 (p. 156-173)

Chapitre 7 : **Exploitation de données statistiques**

Groupe	T	Types de tâches
0	T1	Simuler une expérience aléatoire n fois / Etablir une série statistique
1	T7	Représenter en graphique des données statistiques
2	T28	Calculer l'effectif
	T32	Calculer le pourcentage
	T34	Calculer la fréquence
	T46	Calculer la taille de données statistiques
4	T73	Interpréter des indicateurs statistiques

Tableau 98. Les types de tâches du manuel 4ième, collection Terracher, HACHETTE, 1988

Par les types de tâches T28 (Calculer l'effectif), T34 (Calculer la fréquence) appartenant aux groupes 2, nous voyons que ce manuel suit le programme. Cependant, il existe particulièrement et explicitement le mot « ***simulation*** » pour la première fois dans l'exercice 1 pour expliquer comment fait-on pour avoir une liste des valeurs. En plus, on voit que « hasard » apparaît comme une conception indisponible et que l'on doit « *simuler* » :

1 Avec un dé

On lance un dé et on relève à chaque jet le nombre obtenu.

Voici la liste :

5 ; 3 ; 2 ; 3 ; 1 ; 3 ; 3 ; 3 ;
6 ; 1 ; 4 ; 1 ; 5 ; 2 ; 5 ; 6 ;
5 ; 4 ; 5 ; 4 ; 2 ; 5 ; 4 ; 1 ;
1 ; 3 ; 3 ; 5 ; 1 ; 5 ; 5 ; 6 .

Organiser ces résultats en présentant une série statistique faisant apparaître les **effectifs** et les **fréquences relatives** (en %) de chaque valeur du caractère.

Remarque : La liste ci-dessus a été obtenue par simulation sur micro-ordinateur à l'aide du programme ci-dessous (écrit en LOGO) en appelant JETDE 32.

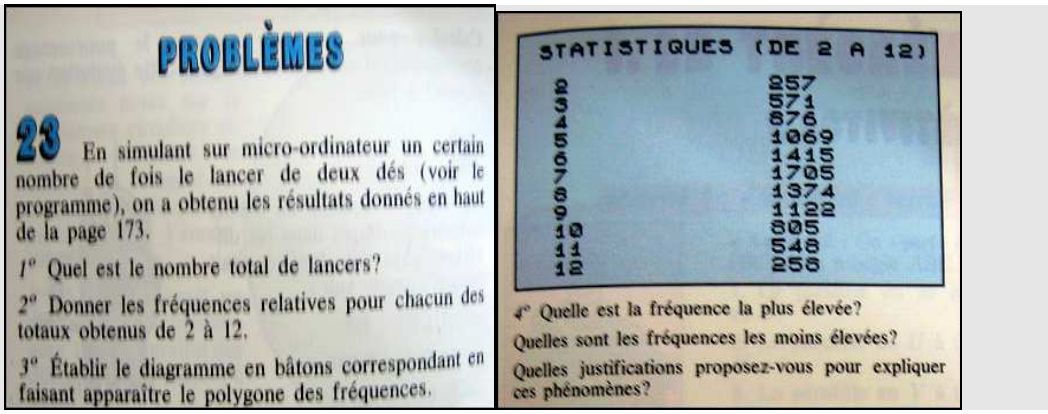
```
POUR JETDE :N
REPETE : N [ECRIS 1 + HASARD 6]
FIN
```

Programme de simulation du jet d'un dé (N fois).

(Manuel 4ième, collection Terracher, HACHETTE, 1988, p. 166)

On écrit effectivement un programme pour obtenir une simulation sur micro-ordinateur. Ainsi, la simulation est abordée mais est-elle pour proposer une pratique[189] ? Est-ce que cet exercice intéresse l'enseignant ?

Ensuite, le problème 23 à la page 172 donne des résultats d'une simulation de lancer de deux dés et on obtient les effectifs des sommes des faces de deux dés. Par ce tableau, l'élève peut-il trouver une grande« possibilité » d'apparition de la somme 7 ?



PROBLÈMES

23 En simulant sur micro-ordinateur un certain nombre de fois le lancer de deux dés (voir le programme), on a obtenu les résultats donnés en haut de la page 173.

1° Quel est le nombre total de lancers?

2° Donner les fréquences relatives pour chacun des totaux obtenus de 2 à 12.

3° Établir le diagramme en bâtons correspondant en faisant apparaître le polygone des fréquences.

STATISTIQUES (DE 2 A 12)

2	257
3	571
4	876
5	1069
6	1415
7	1705
8	1374
9	1122
10	805
11	548
12	256

4° Quelle est la fréquence la plus élevée?

Quelles sont les fréquences les moins élevées?

Quelles justifications proposez-vous pour expliquer ces phénomènes?

(Manuel 4ième, collection Terracher, HACHETTE, 1988, p. 172-173)

On trouve aussi une question intéressante dans ce problème, c'est la demande « *Quelles justifications proposez-vous pour expliquer ces phénomènes ?* ». Les phénomènes dits ici sont la somme des faces de deux dés la plus fréquente est 7 et la moine fréquente est 2 ou 12. Les justifications visées dans cette question peuvent être les liste des cas où deux dés ayant la somme de 7 et de 12 (ou bien de 2). D'autre côté, on peut s'appuyer sur le nombre total de lancers qui est assez grand (10.000 lancers).

Manuel 3ième : manuel 3ième, collection Terracher, HACHETTE, 1989 (p. 261-272)

Chapitre 13 : **Exploitation de données statistiques**

Groupe	T	Types de tâches
0	T1	Simuler une expérience aléatoire n fois / Etablir une série statistique
1	T7	Représenter en graphique des données statistiques
2	T28	Calculer l'effectif
	T34	Calculer la fréquence
	T36	Calculer la moyenne
	T47	Calculer la médiane
	T48	Calculer la mode
	T59	Construire une série statistique connaissant la moyenne et/ou la médiane
3	T64	Commenter des séries statistiques
4	T74	Interpréter les résultats statistiques

Tableau 99. Les types de tâches du manuel 3ième, collection Terracher, HACHETTE, 1989

[189] Souligné par nous.

Dans ce manuel du dernier niveau de collège, on voit presque tous les types de tâches principaux du programme en 1989, surtout le groupe 2 des calculs. C'est comme un « bilan » de la statistique descriptive.

A propos de la notion de moyenne, ce manuel a donné la définition « calculatoire » suivante :

> La moyenne d'une série est le quotient de la somme de toutes les valeurs du caractère par l'effectif total
>
> (Manuel 3ième, collection Terracher, HACHETTE, 1989, p. 261-272)

Ensuite, dans un exemple illustrant la méthode du calcul la moyenne avec une série statistique des valeurs discrètes, on voit la moyenne est 10,04. Puis, le manuel donne une remarque importante concernant au cas où la série statistique se présente sous forme de classes :

> Remarques :
>
> Quand le caractère de la série statistique se présente sous forme de classes, on admet que toutes les notes se regroupent au « centre » de la classe.
>
> Moyenne des notes : […] = 9,92
>
> Il est normal que l'on ne trouve pas la même moyenne dans deux cas. En effet, dans la deuxième démarche, il s'agit d'une approximation liées au centre des classes.
>
> (Manuel 3ième, collection Terracher, HACHETTE, 1989, p.)

Là, nous voyons qu'il y a deux méthodes du calcul de moyenne. Ceux-ci sont des transformations de la théorie universitaire[190] en l'appliquant au cas des valeurs statistiques discrètes.

---

[190] La définition formelle de la moyenne dans le livre « Statistique » (des auteurs Sabin Lessard – Monga, page 56-57) présenté au niveau universitaire :

> « Étant donné un échantillon d'une variable quantitative constitué de n données brutes
>
> $$x_1, x_2, \ldots, x_n$$
>
> la **moyenne de ces données**, notée $\bar{x}$, est la quantité : $\frac{x_1 + x_2 + \ldots + x_n}{n}$
>
> Si ces données sont regroupées en r classes dont les valeurs ou les valeurs centrales, selon le cas, sont :
>
> $$v_1, v_2, \ldots, v_r$$
>
> et dont les fréquences sont :
>
> $$f_1, f_2, \ldots, f_r$$
>
> alors **la moyenne de ces valeurs**, notée $\bar{v}$, est donnée par : $v_1 f_1 + v_2 f_2 + \ldots + v_r f_r$
>
> une formule équivalente pour $\bar{v}$ est : $\frac{n_1 v_1 + n_2 v_2 + \ldots + n_r v_r}{n}$
>
> où $n_1, n_2, \ldots, n_r$ sont les effectifs des r classes et n la somme des effectifs. En général, $\bar{v}$ n'est pas égale à $\bar{x}$, à moins que les classes n correspondent à des valeurs isolées. Cependant, $\bar{v}$ est toujours près de $\bar{x}$ puisque, même dans le cas des classes correspondant à des intervalles de valeurs, les centres des

Ensuite, le manuel (Mathématiques 3è, Terracher, HACHETTE, 1989) donne un commentaire sur les trois notions : la mode, la médiane et la moyenne à partir d'un exemple comme suit :

> Les trois valeurs « centrales » : MODE (8), MÉDIANE (9) et MOYENNE (10) donnent une idée de l'ordre de grandeur des notes recueillies.

(Mathématiques 3ième, collection Terracher, HACHETTE, 1989)

Parmi des exercices, on trouve que la simulation du problème 23 du manuel 4è, Terracher, HACHETTE, 1988 est reprise dans le problème 32 à la p. 272.

```
NOMBRE D'EXPERIENCES? 10000

STATISTIQUES (DE 2 A 12)

 2            257
 3            571
 4            876
 5            1069
 6            1415
 7            1705
 8            1374
 9            1122
 10           805
 11           548
 12           258
OK
SCREENPRINT
Device I/O Error
OK
```

(Mathématiques 3ième, collection Terracher, HACHETTE, 1989, p. 272)

Cette fois-ci, le problème demande de calculer la mode, la médiane et la moyenne puis une question : « Parmi les trois résultats précédents, quel résultat est le plus significatif ? ». Nous ne voyons pas la raison d'être de cette question. (Signification de la moyenne en probabilité : espérance de gain dans un jeu de hasard)

---

classes sont près de valeurs données que celles-ci contiennent. On considère donc que $\bar{v}$ est une approximation de $\bar{x}$ et on note $\bar{x} \approx \bar{v}$. Si on rapetisse de plus en plus la longueur des classes, $\bar{v}$ se rapproche de plus en plus de $\bar{x}$ jusqu'à devenir pratiquement indiscernable de $\bar{x}$. On peut donc, obtenir une approximation aussi précise que l'on veut de la moyenne exacte des données brutes en regroupant les données selon des classes de longueur suffisamment petite, puis en calculant la moyenne des données groupées à l'aide des valeurs centrales des classes. »

D'après ce document, on distingue deux calculs de la moyenne :

- La moyenne de données brutes : $\bar{x}$
- La moyenne de données regroupées : $\bar{v}$

I.1.2. Lycée :

Programme :

	2de	1erS	Ter C, E
Connaissances et savoir-faire	1990 Série statistique à une variable : - Répartition d'une population en classes - Effectifs, fréquences. Séries statistiques à une variable quantitative : - Effectifs cumulés, fréquences cumulées. - Caractéristiques de position et de dispersion : moyenne, écart-type.	1991 Evénements, événements élémentaires ; la probabilité d'un événement est définie par addition de probabilités des événements élémentaires. Evénements disjoints (ou incompatibles), événement contraire, réunion et intersection de deux événements. Cas où les événements élémentaires sont équiprobables.	1992 Probabilité a) Variable aléatoire (réelle) prenant un nombre fini de valeurs et loi de probabilité associée ; fonction de répartition, espérance mathématique, variance, écart-type. b) Probabilité conditionnelle d'un événement par rapport à un événement de probabilité non nulle ; relation $p(A\cap B)=p(A \mid B)p(B)$ Indépendance de deux événements Formule des probabilités totales : étant donné des événements B_1, B_2,…B_n constituant une partition de E, pour tout événement A, $p(A)=$ $p(A\cap B_1)+p(A\cap B_2)+\ldots p(A\cap B_n)$ et pour tout k, $p(A\cap B_k)=p(A \mid_{Bk})p(B_k)$
Champ de problèmes et des techniques		Exemples simples d'emplois de partitions et représentations (arbres, tableaux…) pour organiser et dénombrer des données relatives à la description d'une expérience aléatoire. <u>Exemples simples d'étude de situations de probabilités issues d'expériences aléatoires (modèle d'urne, jeux...)</u>	Exemples simples d'emploi de partitions et représentations (arbres, tableaux…) pour organiser et dénombrer des données. <u>Exemples d'étude de situations de probabilités issues d'expériences aléatoires (modèle d'urne, jeux, schéma de Bernoulli...)</u>[191]

Tableau 100. Les programmes du lycée de 1990 à 1992

<u>Commentaire :</u>

Seconde :

- Le programme fait un rappel sur les deus notions caractéristiques enseignées au collège comme effectif, fréquence et la répartition d'une population en classes, puis effectifs cumulés, fréquences cumulées et moyenne.

- Une nouvelle notion est l'écart type qui sert à interpréter les données statistiques.

[191] Soulignés par nous.

- L'usage des fonctions intégrées dans la calculatrice est autorisé au cas de la moyenne et de l'écart type.

1er S :

- Le programme dit : *« Pour introduire la notion de probabilité, on s'appuie sur l'étude de séries statistiques obtenues par répétition d'une expérience aléatoire, en soulignant les propriétés des fréquences et la relative stabilité de la fréquence d'un événement donné lorsque cette expérience est répétée un grand nombre de fois »* Ainsi, la loi de grand nombre est un élément théorique de référence.

- Dans la phrase : « *La description d'expériences aléatoires amène aussi à organiser des données »*, pourrait-on trouver une idée de la simulation des expériences aléatoires ? L'organisation de données est un travail de plusieurs faits qui se composent sa simulation. On voit que ce fait n'est qu'abordé théoriquement et qu'en Première comme l'intérêt de *la description d'expériences aléatoires*. Mais aucun mot se trouve dans le programme pour demander l'élève à faire la simulation des expériences aléatoires, par contre, le fait de « décrire une expérience aléatoire » est un de deux objectifs qui doivent être entraîné chez l'élève à propos de cette partie de probabilité du programme. Ceci existe dans les travaux pratiques comme « *Exemples simples d'étude de situations de probabilités issues d'expériences aléatoires (modèle d'urne, jeux...) »*.

Terminal C, E (équivalent à Terminal S) :

- L'objectif du programme est de « *poursuivre de l'étude de phénomène aléatoire* » engagée en Première. Pour cet objectif, le programme dit des « *Exemples d'étude de situations de probabilités issues d'expériences aléatoires (modèle d'urne, jeux, schéma de Bernoulli...)* » Ainsi, on voit que les expériences aléatoires sont considérées comme des phénomènes donnant des issues dans les exemples, sans pratique.

En résumé, l'idée de faire un lien entre la statistique et la probabilité par l'étude de phénomène aléatoire est précisée dans le programme.

Manuel :

Manuel seconde : manuel Seconde, collection Terracher, HACHETTE, 1994 (p. 109-132)

Chapitre 2 : **Statistique**

1. Séries statistiques :
   a. Vocabulaire statistique
   b. Représentations graphiques
   c. Effectifs et fréquences cumulées
2. Indicateurs :
   a. Un indicateur de position
   b. Un indicateur de dispersion

Groupe	T	Types de tâches
0	T1	Simuler une expérience aléatoire n fois / Etablir une série statistique
1	T7	Représenter en graphique des données statistiques
	T13	Décrire la statistique (population, caractère étudié, type de caractère)
	T14	Effectuer un regroupement des données
	T15	Donner l'allure de la courbe des fréquences cumulées croissantes
	T17	Indiquer la nature des caractères
2	T31	Constituer un tableau des fréquences cumulées
	T32	Calculer le pourcentage
	T34	Calculer la fréquence
	T36	Calculer la moyenne
	T40	Calculer la valeur d'un terme manquant d'une suite connaissant la moyenne
	T43	Préciser si la moyenne existe dans un énoncé et elle a une signification
	T44	Calculer la taille d'une série statistique connaissant des hypothèses sur les moyennes
	T48	Calculer la mode
	T49	Calculer l'étendue
	T51	Calculer l'écart type
	T56	Calculer le pourcentage de données situées dans l'intervalle [x-s; x+s] / [x-2s; x+2s] / [x-3s; x+3s]
3	T65	Comparer la répartition de deux séries statistiques

Tableau 101. Les types de tâches du manuel de Seconde, collection Terracher, HACHETTE, 1994

Sur les types de tâches, nous voyons que :

- La statistique enseignée en seconde pendant cette période concentre sur la description de données, comme : lecture de données, résumé des données, présentation graphique des données, puis calculer les indicateurs de position et de dispersion comme la moyenne, l'écart type ; et enfin interpréter des données statistiques.

- La notion de moyenne est définie de même que celle présentée au collège et appliquée à deux cas : celui d'un caractère discret et celui d'un regroupement par classes.

- Les deux notions de mode et de médiane ne sont pas dites officiellement dans la partie de théorie mais dans les travaux pratiques (Travaux A, p. 121)

- Il y a un seul type de tâche T55 (comparer la répartition de deux séries statistiques) du groupe 3 et aucun type de tâche des groupes 4 et 5 dans le manuel.

Manuel première : manuel Première S, collection Terracher, HACHETTE, 1995 (p. 91-116)

Chapitre 2 : Probabilités

1. Expérience aléatoire - Evénements :
    a) Expérience aléatoire, issues, univers
    b) Evénements associés à une expérience aléatoire
2. Indicateurs :
    a) Loi de probabilité
    b) Probabilité d'un événement

Remarque :

- Dans la partie d'introduction (avant le cours théorique), le manuel fait des « rappels » sur la statistique enseignée en seconde comme la notion de fréquence et une propriété concernant à celle de la probabilité introduite plus tard.

> La notion de fréquence introduite en statistique intervient dans ce chapitre sous une forme telle que : « Si en lançant n fois une pièce de monnaie, le résultat Pile apparaît p fois, la fréquence relative du Pile est $\frac{p}{n}$ »
>
> Rappelons également que, dans une série statistique, la somme des fréquences est égale à 1.

(Manuel 1ière S, collection Terracher, HACHETTE, 1995, p. 94)

- Ensuite, sous le titre « Vers la notion de probabilité », le manuel précise le rôle de la statistique en s'écrivant :

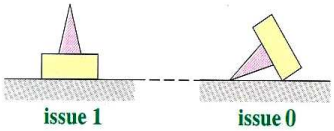
> La statistique permet une première approche de la notion de probabilité, connue sous le nom de Principe des fréquences, que nous explorons dans l'activité 2.

(Manuel 1ière S, collection Terracher, HACHETTE, 1995, p. 96)

L'activité 2 est un travail sur l'expérience aléatoire réalisé par 200 fois de lancers d'une punaise dont le résultat est donné par une série statistique. La demande de faire un tableau de répartition des fréquences en fonction de l'augmentation du nombre de lancer est pour engager à l'estimation d'une fréquence stabilisée.

ACTIVITÉ 2

1° *L'expérience aléatoire*
Elle est réalisée avec une punaise (*figure* ci-contre) que l'on jette en l'air et dont on observe la position lorsqu'elle s'est immobilisée sur le sol.
Cette expérience a deux *issues*, schématisées et codées comme indiqué ci-dessous.



2° *Les données statistiques*
En lançant la punaise 200 fois, on obtient la série statistique suivante :

10001	11101	11101	11111	11100	01110
10101	01000	01001	11111	01110	01001
10111	10011	11101	10010	01010	01111
11010	10010	10001	10011	00101	11100
01101	10111	01110	01101	11010	11000
11001	10100	11110	11111	11101	11000
	10100	11111	11101	11111	

3° *La gestion des données statistiques*
Compléter le tableau suivant :

nombre de chiffres	5	10	15	20	25	30	.....	180	200
fréquence relative du chiffre 1	0,4								

Après 40, on utilisera les données de 20 en 20 : 60, 80, 100, ..., 180, 200.
Effectuer une représentation graphique de ces résultats.

4° *L'interprétation*
Vers quelle valeur la fréquence relative du chiffre 1 tend-t-elle à se stabiliser... à la longue ? (On essaiera de donner une réponse sous forme fractionnaire.)

(Manuel 1ière S, collection Terracher, HACHETTE, 1995, p. 96)

Après ça, on ne trouve plus ce type d'exercice ni dans les exemples du cours ni dans les exercices à s'entraîner.

Par la suite de l'activité 2, la notion de probabilité est introduite comme « limite[192] » d'une suite des fréquences lors de répétition un grand nombre de fois d'une expérience aléatoire. Ainsi, la loi des grands nombres est mise en place pour approcher l'élève à la notion de probabilité :

> « La valeur précise autour de laquelle la fréquence relative du chiffre 1 tend à se concentrer quand on répète l'expérience un grand nombre de fois sera appelé probabilité de l'issue 1 »
>
> (Manuel 1ière S, collection Terracher, HACHETTE, 1995, p. 97)

- La fréquence est la seule notion statistique enseignée en seconde et puis reprise en 1erS pour introduire la probabilité.

- Dans la partie du cours, le manuel définit la loi des probabilités et la probabilité d'un événement comme la somme des probabilités des éléments élémentaires.

- Deux définitions concernant la probabilité sont présentées par des formules : la probabilité d'un issu (événement élémentaire) du cas d'équiprobabilité et celle d'un événement général.

Manuel Terminale : manuel Ter S, collection Terracher, HACHETTE, 1994 (p. 283-312)

Chapitre 12 : Probabilités

1. Probabilités
    a) Définitions
    b) Propriétés des probabilités
    c) Équiprobabilité
2. Probabilité conditionnelle
    a) Définitions
    b) Formules des « Probabilités totales »
3. Variable aléatoire
    a) Notion de variable aléatoire réelle
    b) Fonction de répartition
    c) Espérance mathématique
    d) Linéarité de l'espérance
    e) Variance et écart type

Remarque :

- Pour la partie « Les connaissances mises en jeu », le manuel propose de reprendre des notions concernant la probabilité depuis classe première comme notion d'issue, d'expérience aléatoire, univers, événement, probabilité, équiprobable (p. 284).

- La définition de probabilité est comme celle en classe première : La statistique et la probabilité sont séparées dans le manuel de terminal S. On ne trouve aucun lien entre ces deux notions sauf la notion d'expérience aléatoire.

[192] Le mot « limite » est utilisé par nous.

- Les expériences aléatoires usuelles (modèle d'urne, jeux, schéma de Bernoulli…) sont reprises dans les exercices pour la demande de calculer les probabilités.

Pour résumer l'analyse des programmes et des manuels, nous synthétisons un tableau des types de tâches suivant :

Synthèse des types de tâches dans les manuels de collège-lycée avant la réforme

Groupe	T	Types de Tâches	Avant 2000				
			6è	5è	4è	3è	2de
0	T1	Simuler une expérience aléatoire n fois / Etablir une série statistique					
	T2	Exécuter un programme pour simuler une expérience aléatoire (pour une machine à calcul)					
	T3	Faire un compte rendu de l'expérience					
	T4	Recueillir des données dans la vie réelle puis les écrire en tableau					
1	T5	Lire des informations d'un tableau	**X**				
	T6	Lire des informations à partir d'une représentation graphique					
	T7	Représenter en graphique des données statistiques	**X**	**X**	**X**	**X**	**X**
	T8	Représenter/Regrouper des données statistiques en tableau					
	T9	Représenter une série statistique par la boite moustaches					
	T10	Représenter la boîte de dispersion associée à une série statistique					
	T11	Chercher un exemple de « grandeurs » proportionnelles / non proportionnelles	**X**				
	T12	Calculer la mesure d'angle d'un diagramme qui représente des données statistiques		**X**			
	T13	Décrire la statistique (population, caractère étudié, type de caractère)					x
	T14	Effectuer un regroupement des données					**X**
	T15	Donner l'allure de la courbe des fréquences cumulées croissantes					x
	T16	Lire des indicateurs (l'effectif, la moyenne, la médian) dans une représentation graphique					
	T17	Indiquer la nature des caractères					x
	T18	Dresser le tableau des effectifs (à partir des graphiques)					
	T19	Dresser le tableau des fréquences (à partir des graphiques)					
	T20	Trouver la classe modale					
	T21	Établir la distribution des fréquences (à partir des données)					
	T22	Chercher la valeur la plus grande/petite d'une série statistique					
	T23	Chercher la valeur la plus fréquente					
	T24	Construire un diagramme avec un tableur					
2	T25	Compléter un tableau (des « grandeurs » proportionnelles)	**X**				
	T26	Chercher le coefficient de proportionnalité d'un tableau (des « grandeurs » proportionnelles)	**X**				
	T27	Vérifier si un couple de deux entiers est proportionnel à un autre couple	x				

	T28	Calculer l'effectif			**X**	**X**	
	T29	Calculer des effectifs cumulés					
	T30	Constituer un tableau des effectifs cumulés (croissant)					
	T31	Constituer un tableau des fréquences cumulées					**X**
	T32	Calculer le pourcentage			**X**		x
	T33	Transformer des répartitions de données en pourcentages		x			
	T34	Calculer la fréquence			**X**	**X**	**X**
	T35	Calculer les fréquences cumulées					
	T36	Calculer la moyenne		x		**X**	**X**
	T37	Estimer la moyenne des données regroupées					
	T38	Calculer une valeur pour ajouter à une liste des valeurs statistiques telles que la moyenne ne change pas		**X**			
	T39	Expliquer comment on peut calculer la moyenne pour l'ensemble de population		x			
	T40	Calculer la valeur d'un terme manquant d'une suite connaissant la moyenne					**X**
	T41	Calculer la valeur d'un terme manquant d'une suite connaissant la médiane					
	T42	Calculer la valeur d'un terme manquant d'une suite connaissant l'étendue					
	T43	Préciser si la moyenne existe dans un énoncé et elle a une signification					x
	T44	Calculer la taille d'une série statistique connaissant des hypothèses sur les moyennes					x
	T45	Préciser la position relative de la moyenne par rapport aux valeurs de série statistique					
	T46	Calculer la taille de données statistiques			x		
	T47	Calculer la médiane				**X**	
	T48	Calculer la mode				**X**	**X**
	T49	Calculer l'étendue					x
	T50	Calculer la variance					
	T51	Calculer l'écart type					**X**
	T52	Calculer une valeur à partir le pourcentage et le nombre total des valeurs statistiques					
	T53	Compéter un tableau des effectifs sachant certains effectifs et pourcentages					
	T54	Déterminer les quartiles					
	T55	Déterminer l'intervalle interquartile					
	T56	Calculer le pourcentage de données situées dans l'intervalle [x-s; x+s] / [x-2s; x+2s] / [x-3s; x+3s]					x
	T57	Calculer les bornes de l'intervalle théorique					
	T58	Calculer les bornes de l'intervalle de confiance pour la fréquence observée					
	T59	Construire une série statistique connaissant la moyenne et/ou la médiane				**X**	
	T60	Prévoir une évolution des moyennes lors de changement de population		x			

	T61	Chercher deux valeurs extrêmes où se trouve le nombre d'expérience aléatoire pour que la distribution des fréquences soit équirépartie					
3	T62	Comparer les moyennes des séries statistiques		x			
	T63	Comparer les variances des séries statistiques					
	T64	Commenter des séries statistiques				x	
	T65	Comparer la répartition de deux séries statistiques					**X**
	T66	Comparer la distribution des fréquences des échantillons différents					
	T67	Comparer les indicateurs statistiques des échantillons différents					
	T68	Comparer des diagrammes différents de mêmes données statistiques					
	T69	Comparer deux diagrammes des données statistiques différentes					
	T70	Comparer deux séries statistiques					
	T71	Comparer les fluctuations d'échantillonnage des séries statistiques					
	T72	Comparer une distribution de fréquence à une loi de probabilité					
4	T73	Interpréter des indicateurs statistiques			**X**		
	T74	Interpréter les résultats statistiques				x	
	T75	Donner la pensé à la lecture des résultats de fréquence et d'effectif					
	T76	En observant le tableau de données, répondre si la moyenne pourrait être utilisée comme un représentant pour le caractère étudié					
	T77	Présenter des remarques sur le caractère étudié à partir du tableau de distribution des fréquences de données regroupées en classe					
	T78	Présenter des remarques sur le caractère étudié à partir de la représentation graphique des fréquences de données regroupées en classe					
5	T79	Observer la fluctuation d'échantillonnage					
	T80	Modéliser et simuler une expérience aléatoire					
	T81	Évaluer les indicateurs d'une population à partir d'un échantillon					

Tableau 102. Les types de tâches dans les manuels de collège - lycée avant l'an 2000

Note : x rare **X** moyennement fréquent **X** très fréquent

I.1.3. Conclusion (avant la réforme) :

- Avant la réforme, la statistique enseignée au secondaire est la statistique descriptive. La statistique descriptive, qu'est-ce que c'est ? Quelle statistique descriptive ? Dans le livre « Statistique – Concepts et méthodes » (Sabin Lessard, Monga, 1993, MASSON éditeur, p. 16), la statistique descriptive est définie comme suit :

« On distingue deux grands types d'étude statistiques. Il y a d'abord les études purement descriptives dans lesquelles on déduit certaines caractéristiques d'un ensemble de données. Ces données peuvent avoir été obtenues à partir d'observations partielles sur une population sans qu'on ait eu aucun contrôle ni aucun choix quant à la façon de procéder aux observations. Il peut aussi s'agir de résultats d'observations systématiques sur une population entière. Quoi qu'il en soit, on peut toujours effectuer une étude sommaire de résultats d'observations.

La partie de la statistique qui s'occupe exclusivement de la description de données obtenues à partir d'observations quelconques est appelée **statistique descriptive** ou **statistique déductive**. La statistique descriptive s'intéresse par exemple aux pourcentages de données qui prennent certaines valeurs, aux données les plus fréquentes, aux variations des données les unes par rapport aux autres, aux comparaisons avec d'autres données, etc., en somme au traitement de données. La plupart du temps, il s'agit de classer des données selon des critères déterminés et de faire des calculs pour mesurer, décomposer et comparer les variations.

Dans certains cas, les observations, quoique partielles, sont faites suivant des procédures bien définies et contrôlées : ou bien les individus sont sciemment pris au hasard, ou bien ils sont observés de façon indépendante et dans les même conditions, le choix au hasard étant laissé à la nature ou aux circonstances, ou bien ils sont observés d'après des critères spécifiques qui permettent de les catégoriser. Le but des procédures est généralement d'assurer que l'on obtient des échantillons représentatifs de populations bien définies, ce qui permet d'induire une information sur les populations à partir d'échantillon. La partie de statistique qui s'occupe de tirer des conclusions générales à partir d'enquêtes ou d'expérience statistiques contrôlées est appelée **statistique inductive** ou **inférence statistique**. C'est l'inférence statistique qui permet les généralisations et les prévisions, d'où son grand intérêt théorique et pratique. »

(Sabin Lessard, Monga, 1993, MASSON éditeur, p. 16)

- Dans la statistique descriptive présentée au collège, la notion d'échantillonnage n'est pas abordée. Les données statistiques ne sont jamais considérées comme un échantillon d'une population mère mais elles sont considérées comme une population exhaustive, d'où sans hasard. De plus, même la **construction d'une série statistique** peu recueil de données pour répondre à ces questions est absente.

- Les premières connaissances de la statistique descriptive présentées dans les programmes et les manuels du collège à la seconde se trouvent dans tous les cinq groupes des types de tâches.

- Les indicateurs de dispersion ne sont pas présentés au collège pendant ces temps.

- L'observation des fréquences d'un événement lors de répétition d'un grand nombre d'expérience aléatoire pouvant faire le lien entre la statistique et la probabilité dans le programme avant la réforme (avant l'an 2000). Il n'y a que des traces infinies de telles observations. On peut en conclure qu'il y n'existe pas au collège et en seconde avant 2000 un tel lien.

- Sur le tableau 10 des types de tâches, on voit que:

+ T7 apparaît dans tous les niveaux de collège à la seconde.

+ Les types de tâches concernant les calculs des notions principales comme T28 (effectif), T34 (fréquence), T37 (moyenne), T48 (mode) sont présents fréquemment et dominants au niveau de collège.

+ Les calculs des indicateurs de dispersion comme T49 (étendue), T51 (écart type), T56 (pourcentage de données situées dans l'intervalle [x-σ; x+σ] / [x-2σ; x+2σ] / [x-3σ; x+3σ]) n'existent qu'en classe seconde.

+ La comparaison des séries statistiques n'occupe qu'une place très modeste en commençant rarement par la comparaison des moyennes en classe 5è (T62), puis commentaire des séries statistiques en classe 3è (T64) et enfin comparaison des répartitions de deux séries statistiques en seconde (T65).

+ L'interprétation statistique reste pour des indicateurs simples comme effectif, fréquence, moyenne, médiane, mode pour des classes 4ième et 3ième. On voit l'absence de ce travail en classe seconde et par conséquence, il n'y a pas de comparaison des indicateurs de dispersion.

- En seconde, on voit un résumé des travaux sur la statistique comme un bilan à la fin du collège. Puis on ne trouve plus sa présence dans les classes suivantes (la première et la terminale) sauf une « trace » dite précédemment.

- Avant la réforme, on peut voir que la statistique au collège et en seconde est la statistique sans probabilité. Par contre, on trouve la probabilité sans statistique en classe première et en classe terminal. Même s'il existe des tableaux statistiques, on laisse à l'élève la responsabilité ou non construire lui même.

I.2. Après la réforme (à partir de 2000)

I.2.1. Collège :

Programme :

	Classe 6è	Classe 5è	Classe 4è	Classe 3è
Programme	Organisation et représentation de données.	Lecture. Interprétation, représentations graphiques de séries statistiques. Classes, effectifs. Fréquences.	Effectifs cumulés, Fréquences cumulées. Moyennes pondérées. Initiation à l'utilisation de tableurs-grapheurs.	Caractéristiques de position d'une série statistique. Approche de caractéristiques de dispersion d'une série statistique. Initiation à l'utilisation de tableurs-grapheurs en statistique.

Compétences exigibles	- Organiser des données en choisissant un mode de présentation adapté : + tableaux en deux ou plusieurs colonnes ; + tableaux à double entrée. - Lire et compléter une graduation sur une demi-droite graduée, à l'aide d'entiers naturels, de décimaux ou de quotients (placement exact ou approché). - Lire et interpréter des informations à partir d'une représentation graphique (diagrammes en bâtons, diagrammes circulaires ou demi-circulaires, graphiques cartésiens).	Lire et interpréter un tableau, un diagramme à barres, un diagramme circulaire ou semi-circulaire. Regrouper des données statistiques en classes, calculer des effectifs. Présenter une série statistique sous la forme d'un tableau, la représenter sous la forme d'un diagramme ou d'un graphique. Calculer des fréquences.	Calculer des effectifs cumulés, des fréquences cumulées. Calculer la moyenne d'une série statistique. Calculer une valeur approchée de la moyenne d'une série statistique regroupée en classes d'intervalles	Une série statistique étant donnée (sous forme de liste ou de tableau, ou par une représentation graphique), proposer une valeur médiane de cette série et en donner la signification. Une série statistique étant donnée, déterminer son étendue ou celle d'une partie donnée de cette série.

Tableau 103. Les programmes et compétences exigibles du collège après la réforme

Les notions statistiques enseignées au collège après la réforme :

6è	5è	4è	3è
Diagramme	Classe Effectif Fréquence	Effectif Fréquence Moyenne	Moyenne Médiane Étendue

Tableau 104. Les notions statistiques dans les programmes du collège après la réforme

- Nous voyons l'apparition de notion d'étendue qui n'existe pas au niveau collège avant la réforme.

- Au niveau de collège, le programme de la statistique enseignée après la réforme ne change pas beaucoup par rapport à l'avance.

Manuel :

Pour analyser les manuels du collège, nous choisissons les suivants :

- Manuel 6è, collection DIABOLO, HACHETTE, 2005
- Manuel 5è, collection DIABOLO, HACHETTE, 2006
- Manuel 4è, collection Triangle, HATIER, 2002
- Manuel 3è, collection DIABOLO, HACHETTE, 2005

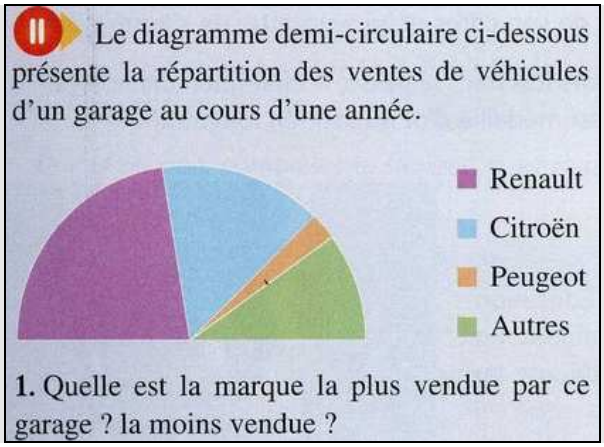
Manuel 6ième : manuel 6ième, collection Diabolo, HACHETTE, 2005 (p. 107-120)

Chapitre 6 : **Gestion de données**

Groupe	T	Types de tâches
0	T1	Simuler une expérience aléatoire n fois / Etablir une série statistique
1	T5	Lire des informations d'un tableau
	T7	Représenter en graphique des données statistiques
	T8	Représenter/Regrouper des données statistiques en tableau
	T22	Chercher la valeur la plus grande/petite d'une série statistique
	T23	Chercher la valeur la plus fréquente
	T24	Construire un diagramme avec un tableur
2	T32	Calculer le pourcentage
	T52	Calculer une valeur à partir le pourcentage et le nombre total des valeurs statistiques
3	T68	Comparer des diagrammes différents de mêmes données statistiques
	T69	Comparer deux diagrammes des données statistiques différentes

Tableau 105. Les types de tâches du manuel 6ième, NATHAN, 1986

- Parmi les types de tâches, on trouve certains nouveaux par rapport au manuel analysé avant : il existe des exercices qui demande de préciser « la mode » (question 1, exercice 11 - p. 120 ou bien question 3, exercice 21 – p. 114), « l'effectif cumulé » (question 4, exercice 21 – p. 114),… même si ces notions ne sont pas présentées dans ce niveau de sixième.



(Manuel 6ième, collection Diabolo, HACHETTE, 2005, p. 120)

21 La liste de nombres suivante représente les notes obtenues par une classe de sixième au devoir commun de mathématiques.

5 – 7 – 7 – 9 – 9 – 10 – 10 – 15 – 15 – 11 – 11
10 – 10 – 11 – 13 – 13 – 13 – 13 – 15
15 – 15 – 15 – 7 – 11 – 11

1. Regrouper ces données dans un tableau.

2. Combien d'élèves ont fait ce devoir ?
3. Quelle est la note obtenue le plus fréquemment ?
4. Quel est le nombre d'élèves ayant obtenu 10 sur 20 ou plus ?
5. Construire un diagramme en bâtons pour représenter ces données.

(Manuel 6ième, collection Diabolo, HACHETTE, 2005, p. 114)

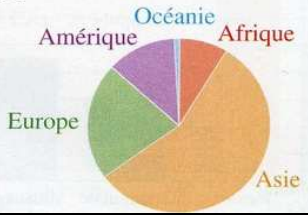
- Le pratique le plus insisté est celui sur les tableaux et les diagrammes.

- Comparaison deux diagrammes de deux données statistiques est un travail sur une comparaison des données statistique (exercice 25, p. 115). Cela veut dire que les types de tâches évoluent.

Comparer des diagrammes

25 Les deux diagrammes circulaires ci-dessous donnent la répartition de la population mondiale par continent en 1950 (graphique 1) et une prévision de cette répartition pour l'année 2050 (graphique 2).

Graphique 1



Graphique 2

1. a) Est-il vrai qu'en 1950 la population européenne représentait le quart de la population mondiale ?
b) Selon les prévisions, cela sera-t-il toujours le cas en 2050 ?

(Manuel 6ième, collection Diabolo, HACHETTE, 2005, p. 115)

Manuel 5ième: manuel 5ième, collection Diabolo, HACHETTE, 2006 (p. 121-140)

Chapitre 7 : **Statistique**

Groupe	T	Types de tâches
0	T1	Simuler une expérience aléatoire n fois / Etablir une série statistique
1	T7	Représenter en graphique des données statistiques
	T8	Représenter/Regrouper des données statistiques en tableau

	T12	Calculer la mesure d'angle d'un diagramme qui représente des données statistiques
2	T26	Chercher le coefficient de proportionnalité d'un tableau (des « grandeurs » proportionnelles)
	T28	Calculer l'effectif
	T30	Constituer un tableau des effectifs cumulés (croissant)
	T34	Calculer la fréquence
	T52	Calculer une valeur à partir le pourcentage et le nombre total des valeurs statistiques
	T53	Compéter un tableau des effectifs sachant certains effectifs et pourcentages
3	T68	Comparer des diagrammes différents de mêmes données statistiques

Tableau 106. Les types de tâches du manuel 5ième, collection Diabolo, HACHETTE, 2006

- La comparaison des diagrammes se répète en classe 5è mais cette fois-ci, elle est une comparaison de forme des diagrammes (exercice 55, p. 136), question d) : « *Comparer la forme de ce diagramme avec la forme obtenue à la question 1.a)* »

55 Comparaison de diagrammes

Dans un village, on dénombre 84 familles avec un enfant, 42 familles avec 2 enfants, 28 familles avec 3 enfants et 21 familles avec 4 enfants.

1. a) Représenter cette situation par un diagramme en barres.

b) Dresser un tableau des fréquences des familles selon le nombre d'enfants qui la composent.

2. a) Combien d'enfants vivent dans des familles d'exactement 2 enfants ? de 4 enfants ?

b) Recopier et compléter le tableau suivant :

enfant vivant dans une famille de …	effectif	fréquence (en %)
1 enfant		
2 enfants		
3 enfants		
4 enfants		
Total		

c) Représenter cette répartition par un diagramme en barres.

d) Comparer la forme de ce diagramme avec la forme obtenue à la question **1.a)**.

(Manuel 5ième, collection Diabolo, HACHETTE, 2006, p. 136)

- Le regroupement des données en classe (T8) est le nouveau travail par rapport au niveau 6ième.

40 Sécurité routière

Le tableau ci-dessous donne les vitesses, en kilomètre par heure, enregistrées par un radar sur une autoroute pendant une minute.

129	131	129	125	95	110
108	148	118	135	119	140
98	110	124	112	124	130
130	107	112	115	123	129
133	130	128	131	121	127

1. Regrouper ces données dans des classes telles que : $90 \leqslant v < 100$, $100 \leqslant v < 110$, etc.

(Manuel 5ième, collection Diabolo, HACHETTE, 2006, p. 133)

Manuel 4ième : manuel 4ième, collection Triangle, HATIER, 2002 (p. 71-84)

Chapitre 5 : **Statistique**

Groupe	T	Types de tâches
0	T1	Simuler une expérience aléatoire n fois / Etablir une série statistique
1	T5	Lire des informations d'un tableau
	T7	Représenter en graphique des données statistiques
	T14	Effectuer un regroupement des données
	T16	Lire des indicateurs (l'effectif, la moyenne, la médian) dans une représentation graphique
2	T28	Calculer l'effectif
	T29	Calculer des effectifs cumulés
	T30	Constituer un tableau des effectifs cumulés (croissant)
	T31	Constituer un tableau des fréquences cumulées
	T32	Calculer le pourcentage
	T34	Calculer la fréquence
	T35	Calculer les fréquences cumulées
	T36	Calculer la moyenne
	T40	Calculer la valeur d'un terme manquant d'une suite connaissant la moyenne
	T45	Préciser la position relative de la moyenne par rapport aux valeurs de série statistique
	T52	Calculer une valeur à partir le pourcentage et le nombre total des valeurs statistiques
	T59	Construire une série statistique connaissant la moyenne et/ou la médiane

Tableau 107. Les types de tâches du manuel 4ième, collection Triangle, HATIER, 2002

- Les nouvelles notions enseignées officiellement dans la partie de cours en 4è sont : effectif cumulé, fréquence cumulée, et moyenne. Particulièrement, les calculs des moyennes sont présentés en trois cas : moyenne, moyenne pondérée, moyenne approchée d'une série regroupée en classe.

- Un exercice concernant une propriété de la moyenne est Exercice 31 : (Lors d'ajout d'un même nombre à toutes les valeurs d'une série statistique, la moyenne ne change pas)

31. a/ Calculer la moyenne des notes :
4 ; 7 ; 15 ; 12 ; 2 ; 16 ; 9.
b/ Des élèves demandent à leur professeur d'augmenter chaque note de 2 points. Que deviendrait la moyenne ?

(Manuel 4ième, collection Triangle, HATIER, 2002, p. 83)

- Dans le manuel, on trouve un seul exercice qui demande l'élève de faire une simulation en jetant deux dés pour obtenir une série statistique (exercice 35, p. 84) :

35. Voici un jeu : on jette deux dés et on note la somme des valeurs obtenues sur chaque dé.

a/ Faire 50 jets et noter les sommes obtenues.

b/ Regrouper ces données dans un tableau.

c/ Calculer les fréquences de chaque somme.

d/ Les résultats trouvés au **c/** étaient-ils prévisibles ?

(Manuel 4è, collection Triangle, HATIER, 2002, p. 84)

Manuel 3ième: manuel 3ième, collection Diabolo, HACHETTE, 2005 (p. 137-154)

Chapitre 7 : **Statistique**

Groupe	T	Types de tâches
0	T1	Simuler une expérience aléatoire n fois / Etablir une série statistique
1	T8	Représenter/Regrouper des données statistiques en tableau
	T16	Lire des indicateurs (l'effectif, la moyenne, la médian) dans une représentation graphique
	T22	Chercher la valeur la plus grande/petite d'une série statistique
2	T29	Calculer des effectifs cumulés
	T30	Constituer un tableau des effectifs cumulés (croissant)
	T32	Calculer le pourcentage
	T36	Calculer la moyenne
	T40	Calculer la valeur d'un terme manquant d'une suite connaissant la moyenne
	T41	Calculer la valeur d'un terme manquant d'une suite connaissant la médiane
	T42	Calculer la valeur d'un terme manquant d'une suite connaissant l'étendue
	T47	Calculer la médiane
	T49	Calculer l'étendue
	T59	Construire une série statistique connaissant la moyenne et/ou la médiane
3	T62	Comparer les moyennes des séries statistiques

Tableau 108. Les types de tâches du Manuel 3è, collection Diabolo, HACHETTE, 2005

- Les nouvelles notions sont : la médiane, l'étendue. Particulièrement, les calculs des effectifs cumulés sont demandés pour déduire la médiane.

- Deux séries statistiques à comparer se forment de l'une à l'autre. C'est-à-dire, on supprime certaines valeurs d'une série statistique pour obtenir une nouvelle puis les compare (exercice 34, p. 149).

34 On considère la série statistique suivante :

27 ; 29 ; 37 ; 43 ; 28 ; 37 ; 25 ; 26 ; 45.

1° Calculer sa moyenne, sa médiane et son étendue.

2° Que deviennent ces trois caractéristiques si l'on supprime les valeurs extrêmes de cette série ?

(Manuel 3ième, collection Diabolo, HACHETTE, 2005, p. 149)

- L'exercice 51 est le seul qui aborde le hasard par le lancement de deux dés. En plus, la question 2° demande de répondre l'existence d'une somme possible ne jamais apparaîtrant.

51 **Au hasard !**

Julien lance deux dés cubiques, vingt fois consécutives. Sur chaque face d'un dé figure l'un des nombres :

1, 2, 3, 4, 5, 6.

Chaque résultat obtenu est représenté par le couple $(a\ ;\ b)$, où a et b désignent deux entiers compris entre 1 et 6.

Voici les résultats obtenus par Julien lors des vingt lancers :

(4 ; 1) ; (2 ; 3) ; (3 ; 6) ; (5 ; 5) ; (1 ; 6) ;
(4 ; 4) ; (2 ; 5) ; (6 ; 6) ; (1 ; 1) ; (2 ; 4) ;
(3 ; 4) ; (2 ; 5) ; (6 ; 3) ; (1 ; 5) ; (2 ; 1) ;
(2 ; 6) ; (4 ; 3) ; (3 ; 5) ; (1 ; 1) ; (5 ; 6).

Julien s'intéresse à la somme S des nombres obtenus à chaque lancer.

1° Écrire les vingt sommes obtenues par Julien (on regroupera les résultats dans un tableau).

2° Existe-t-il une somme possible, mais jamais obtenue lors des vingt lancers ? Si oui, laquelle ?

3° Quelle est la moyenne des sommes obtenues ?

4° Quelle est la médiane des sommes obtenues ?

5° Quelle est l'étendue des sommes obtenues ?

6° Quelle est la fréquence de la valeur 7 ?

7° Tracer un histogramme des sommes S obtenues.

(Manuel 3ième, collection Diabolo, HACHETTE, 2005, p. 149)

I.2.2. Lycée :

Programme :

	2nde	1ère S	Te S
Contenus	**Statistique** Résumé numérique par une ou plusieurs mesures de tendance centrale (moyenne, médiane, classe modale, moyenne élaguée) et une mesure de dispersion (on se restreindra en classe de seconde à l'étendue). Définition de la distribution des fréquences d'une série prenant un petit nombre de valeurs et de la fréquence d'un événement. Simulation et fluctuation d'échantillonnage.	**Statistique** Variance et écart-type. Diagramme en boîtes ; intervalle interquartile. Influence sur l'écart type et l'intervalle interquartile d'une transformation affine des données. **Probabilité** Définition d'une loi de probabilité sur un ensemble fini. Espérance, variance, écart-type d'une loi de probabilité. Probabilité d'un événement, de la réunion et de l'intersection d'événements. Cas de l'équiprobabilité. Variable aléatoire, loi d'une variable aléatoire, espérance, variance, écart type. Modélisation d'expériences aléatoires de référence (lancers d'un ou plusieurs dés ou pièces discernables ou non, tirage au hasard dans une urne, choix de chiffres au hasard, etc.).	Conditionnement et indépendance Conditionnement par un événement de probabilité non nulle puis indépendance de deux événements. Indépendance de deux variables aléatoires. Formule des probabilités totales. **Statistique et modélisation.** Expériences indépendantes. Cas de la répétition d'expériences identiques et indépendantes. **Lois de probabilité** Exemples de lois discrètes Introduction des combinaisons, notées $\binom{n}{p}$ Formule du binôme. Loi de Bernoulli, loi binomiale ; espérance et variance de ces lois. Exemples de lois connues Lois continues à densité : – loi uniforme sur [0,1] ; – loi de durée de vie sans vieillissement.

Modalités de mise en œuvre	Utiliser les propriétés de linéarité de la moyenne d'une série statistique. Calculer la moyenne d'une série à partir des moyennes de sous-groupes. Calcul de la moyenne à partir de la distribution des fréquences. Concevoir et mettre en œuvre des simulations simples à partir d'échantillons de chiffres au hasard.	**Statistique** On cherchera des résumés pertinents et on commentera les diagrammes en boîtes de quantités numériques associées à des séries simulées ou non. On observera l'influence des valeurs extrêmes d'une série sur l'écart type ainsi que la fluctuation de l'écart type entre séries de même taille. L'usage d'un tableur ou d'une calculatrice permet d'observer, dynamiquement et en temps réel, les effets des modifications des données. **Probabilité** Le lien entre loi de probabilité et distributions de fréquences sera éclairé par un énoncé vulgarisé de la loi des grands nombres. On expliquera ainsi la convergence des moyennes vers l'espérance et des variances empiriques vers les variances théoriques ; on illustrera ceci par des simulations dans des cas simples. On pourra aussi illustrer cette loi avec les diagrammes en boîtes obtenus en simulant par exemple 100 sondages de taille n, pour n = 10 ; 100 ; 1000. On simulera des lois de probabilités simples obtenues comme images d'une loi équirépartie par une variable aléatoire (sondage, somme des faces de deux dés, etc.)	On justifiera la définition de la probabilité de B sachant A, notée PA(B), par des calculs fréquentiels. On utilisera à bon escient les représentations telles que tableaux, arbres, diagrammes, etc., efficaces pour résoudre des problèmes de probabilités. Application à la problématique des tests de dépistage en médecine et à la loi de l'équilibre génétique lors d'appariements au hasard. Application aux expériences de références vues en seconde et première (dés, pièces, urnes, etc.). On introduira la notation n !. L'élève devra savoir retrouver les formules : $\binom{n}{p} = \binom{n-1}{p-1} + \binom{n-1}{p}$ $\binom{n}{p} = \binom{n}{n-p}$ On appliquera ces résultats à des situations variées. Application à la désintégration radioactive: loi exponentielle de désintégration des noyaux. Étude d'un exemple traitant de l'adéquation de données expérimentales à une loi équirépartie.

Tableau 109. Les programmes du lycée après la réforme

Les notions statistiques et probabilistes proposées à enseigner au lycée après la réforme :

Notion	2de	Première S	Terminal
Statistique	Moyenne Moyenne élaguée Médiane Classe modale Étendue Simulation Fluctuation d'échantillonnage	Simulation Variance et écart-type Diagramme en boîtes Intervalle interquartile	Simulation

Probabilité		Loi de probabilité Probabilité d'un événement Espérance, variance, écart-type d'une loi de probabilité Modélisation d'expériences aléatoires de référence	Modélisation Conditionnellement Indépendance de deux événements Indépendance de deux variables aléatoires Loi de Bernoulli, loi binomiale Espérance et variance de ces lois

Tableau 110. Les notions statistiques et probabilistes dans les programmes du lycée après la réforme

- En Seconde :

+ Moyenne, médiane, classe modale, moyenne élaguée, étendue sont des notions enseignées en Seconde.

+ Pour la notion de moyenne habituelle, le programme dit du « *calcul la moyenne à partir des moyennes de sous-groupes* ».

+ La statistique est concentrée sur la simulation et la fluctuation. C'est pourquoi, on ne voit pas les notions d'indicateurs de dispersion habituelles auparavant comme l'écart type, variance, … dans la classe seconde.

- En Première :

+ Par contre, la statistique est continuée à enseigner et occupe un chapitre dans le programme de classe première avec les notions statistiques restant : variance et écart-type ; diagramme en boîtes; intervalle interquartile.

+ Le programme veut renforcer le travail sur le résumé et le commentaire des séries statistiques par les diagrammes en boites.

+ Il semble que l'intension de présenter les notions de variance et d'écart type dans la partie de statistique et pour faciliter et lier avec les notions variance et d'écart type d'une loi de probabilité et puis celles de variable aléatoire dans la partie de probabilité suivante.

+ Et enfin, il y a des liens entre les connaissances de statistique en Seconde et celles en première comme « *lien entre loi de probabilité et distributions de fréquences* », « *la convergence des moyennes vers l'espérance et des variances empiriques vers les variances théoriques* » en illustrant par la simulation des cas simples.

- En Terminal :

+ Même si la statistique n'occupe pas une place explicite comme celle dans la classe de première, mais « *Statistique et modélisation* » et « *Statistique et simulation* » sont présentes alternativement avec des autres parties de la probabilité dans le programme de Terminal. Elles se présentent dans « *La répétition des expériences aléatoire* », « *Application aux expériences de références vues en seconde et première (dés, pièces, urnes, etc.)* » et sont des éléments faisant le lien entre la statistique et la probabilité dans le programme de Terminal.

+ « *L'adéquation de données expérimentales à une loi équirépartie* » est proposée d'être étudié par un exemple. Alors, est-ce qu'il existe dans la partie d'exercice ?

+ Le programme veut prendre « *des calculs fréquentiels* » pour justifier la définition de la probabilité conditionnelle. Alors, comment cela se fait dans le manuel ?

Ainsi, le nouveau dans le programme de lycée est de garder un lien entre la statistique et la probabilité.

Manuel :

Manuel seconde : manuel Seconde, collection Déclic, HACHETTE, 2004 (p. 193-216),

Chapitre 2 : **Statistique**

1. Vocabulaire des séries statistiques
    a) Statistique descriptive
    b) Effectif et fréquence
2. Mesure d'une série quantitative
    a) Étendue et moyenne
    b) Médiane
    c) Mode et Classe modale
3. Fluctuation d'échantillonnage et simulations
    a) Fluctuation d'échantillonnage
    b) Simulation

Groupe	T	Types de tâches
0	T1	Simuler une expérience aléatoire n fois / Etablir une série statistique
	T2	Exécuter un programme pour simuler une expérience aléatoire (pour une machine à calcul)
	T3	Faire un compte rendu de l'expérience
1	T7	Représenter en graphique des données statistiques
	T17	Indiquer la nature des caractères
	T18	Dresser le tableau des effectifs (à partir des graphiques)
	T20	Trouver la classe modale
	T21	Établir la distribution des fréquences (à partir des données)
2	T32	Calculer le pourcentage
	T34	Calculer la fréquence
	T35	Calculer les fréquences cumulées
	T36	Calculer la moyenne
	T47	Calculer la médiane
	T48	Calculer la mode
	T49	Calculer l'étendue
	T57	Calculer les bornes de l'intervalle théorique
	T58	Calculer les bornes de l'intervalle de confiance pour la fréquence observée
3	T66	Comparer la distribution des fréquences des échantillons différents
	T67	Comparer les indicateurs statistiques des échantillons différents
5	T79	Observer la fluctuation d'échantillonnage

Tableau 111. Les types de tâches du manuel Seconde, collection Déclic, HACHETTE, 2004

Dans le manuel, nous voyons que :

- Les notions usuelles comme effectif, fréquence, moyenne, mode et médiane sont présentées, et particulièrement, la moyenne pondérée est dite. Par contre, la variance et l'écart type ne sont pas abordés dans ce manuel comme celles dans la classe seconde avant la réforme.

- Les deux cas de série quantitative discrète et celle continue sont remarqués pour le choix du tableau des effectifs.

- L'observation de la fluctuation d'échantillonnage et la simulation sont insistées dans la partie théorique et même dans les exercices avec 4 types de tâches des groupes 4 et 5 s'y trouvant.

Manuel première : manuel Première S, collection Terracher, HACHETTE, 2001 (p. 227-280)

Chapitre 9 : **Statistique**
1. Moyenne et l'écart type
   a) Moyenne
   b) L'écart type et variance
   c) Propriétés de la moyenne et l'écart type
2. Médiane et interquartile

Médiane
   a) Quartiles et intervalle interquartile
   b) Les boîtes de dispersion

Chapitre 10 : Probabilités
1. Modélisation d'une expérience aléatoire
   a) Loi de probabilité
   b) Loi équirépartie
   c) Paramètres associés à des issues numériques
2. Evénement
   a) Evénements associés à une expérience aléatoire
   b) Probabilité d'un événement
   c) Propriétés des probabilités
3. Variable aléatoire
   a) Notion de variable aléatoire
   b) Espérance, variance, l'écart type d'une variable aléatoire

Groupe	T	Types de tâches
0	T1	Simuler une expérience aléatoire n fois / Etablir une série statistique
1	T7	Représenter en graphique des données statistiques
	T9	Représenter une série statistique par la boite moustaches
	T10	Représenter la boîte de dispersion associée à une série statistique
	T14	Effectuer un regroupement des données
2	T32	Calculer le pourcentage
	T36	Calculer la moyenne
	T43	Préciser si la moyenne existe dans un énoncé et elle a une signification
	T47	Calculer la médiane

	T49	Calculer l'étendue
	T50	Calculer la variance
	T51	Calculer l'écart type
	T54	Déterminer les quartiles
	T55	Déterminer l'intervalle interquartile
	T56	Calculer le pourcentage de données situées dans l'intervalle [x-s; x+s] / [x-2s; x+2s] / [x-3s; x+3s]
3	T63	Comparer les variances des séries statistiques
	T70	Comparer deux séries statistiques
	T71	Comparer les fluctuations d'échantillonnage des séries statistiques

Tableau 112. Les types de tâches du manuel Première S, collection Terracher, HACHETTE, 2001

- Dans le manuel, la statistique avec la présence des notions de moyenne, médiane, l'écart type, variance, quartile, boite de dispersion est pour garder le lien entre la statistique en seconde avec la probabilité en première.

- La moyenne est calculée à partir des moyennes de sous-groupes.

- Il y a un exercice à la fin de ce chapitre (exercice 35, p. 250) avec la question 4b) intéressante qui pose un problème du dé « truqué » ou « honnête » par l'observation d'une distribution des effectifs. C'est un exercice concernant une connaissance de la loi « Khi-deux » qui est hors de programme.

35 Distance statistique

Lorsqu'on lance n fois un dé, chacun sait que les distributions $(f_1, f_2, f_3, f_4, f_5, f_6)$ des fréquences obtenues des six numéros fluctuent autour de la distribution théorique

$$\left(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\right).$$

Nous nommerons *distance statistique* associée à la distribution $(f_1, f_2, f_3, f_4, f_5, f_6)$ correspondant à n lancers, le nombre :

$$d^2 = n \sum_{i=1}^{6} \left(f_i - \frac{1}{6}\right)^2.$$

Nous avons réalisé 1 000 simulations de 120 lancers de dé et nous avons obtenu les valeurs suivantes de d^2 (correspondant à $n = 120$).

Intervalle	Effectif
[0 ; 0,2[	67
[0,2 ; 0,4[	153
[0,4 ; 0,6[	167
[0,6 ; 0,8[	160
[0,8 ; 1[	135
[1 ; 1,2[	94
[1,2 ; 1,4[	73
[1,4 ; 1,6[	48
[1,6 ; 1,8[	40

Intervalle	Effectif
[1,8 ; 2[	23
[2 ; 2,2[	13
[2,2 ; 2,4[	6
[2,4 ; 2,6[	5
[2,6 ; 2,8[	6
[2,8 ; 3[	3
[3 ; 3,2[	0
[3,2 ; 3,4[	4
[3,4 ; 3,6[	3

1) Représenter l'histogramme de la série.

2) Représenter la courbe cumulative des fréquences.
En déduire la médiane de la série, la valeur des déciles d_1 et d_9, la valeur des « vingtiles » v_1 et v_{19}.

Si F est la fonction de répartition, v_1 et v_{19} sont définis par : $F(v_1) = 0{,}05$ et $F(v_{19}) = 0{,}95$.

3) Représenter la boîte à moustaches ayant pour extrémités v_1 et v_{19}.

4) Nous admettons qu'un dé donné est *irrégulier de façon significative au seuil de* 5 % lorsque, pour 120 lancers, sa distance statistique d^2 dépasse v_{19}.

a) Commenter et justifier cette définition.

b) En lançant 120 fois un dé, nous avons obtenu la distribution des effectifs des six numéros : (13, 18, 11, 21, 27, 30).
Ce dé est-il régulier ?

c) Recommencer avec un dé personnel et conclure.

(Manuel 1ière S, collection Terracher, HACHETTE, 2001, p. 250)

- La probabilité occupe un chapitre suivant de celui de statistique.

Manuel terminale : manuel Terminal S, collection Terracher, HACHETTE, 2002 (p. 265-294)

Chapitre 9 : Probabilités, conditionnellement et indépendance
- 1. Le point sur les probabilités
  - 1.1. Modélisation d'une expérience aléatoire
  - 1.2. Probabilité d'un événement
  - 1.3. Propriétés des probabilités
  - 1.4. Variable aléatoire
  - 1.5. Propriétés de l'espérance, de la variance…
- 2. Probabilités conditionnelles
  - 2.1. Définition
  - 2.2. Formule des probabilités totales
  - 2.3. Mise en œuvre dans le cas d'épreuves répétées
- 3. Indépendance
  - 3.1. Evénements indépendants
  - 3.2. Variables aléatoires indépendantes
  - 3.3. Expériences aléatoires indépendantes

Chapitre 10 : Loi de probabilité (p. 295-330)
- 1. Dénombrement
  - 1.1 Liste d'événement d'un ensemble fini
  - 1.2 Permutation
- 2. Combinaisons
  - 2.1 Définition, notation
  - 2.2 Dénombrement des combinaisons
  - 2.3 Coefficients binomiaux et formule des binômes

3. Exemple de loi discrète
   3.1 La loi de Bernoulli
   3.2 La loi binomiale
   3.3 Caractéristique de la loi binomiale
4. Lois continues (généralités), lois uniformes
   4.1 Loi de probabilité à densité
   4.2 La loi uniforme
   4.3 Variable aléatoire, Fonction de répartition
5. La loi exponentielle
5.1 Définition, premières propriétés
   5.2 Loi de durée de vie sans vieillissement

Groupe	T	Types de tâches
0	T1	Simuler une expérience aléatoire n fois / Etablir une série statistique
2	T61	Chercher deux valeurs extrêmes où se trouve le nombre d'expérience aléatoire pour que la distribution des fréquences soit équirépartie
3	T72	Comparer une distribution de fréquence à une loi de probabilité
4	T76	Donner la pensé à la lecture des résultats de fréquence et d'effectif

Tableau 113. Les types de tâches du manuel Ter. S, collection Terracher, HACHETTE, 2002

Remarque :

- Pour introduire à la notion de probabilité conditionnelle, un tableau de deux entrées est montré pour « une approche fréquentiste » où nous voyons que l'on a montré une « formule statistique » pour calculer la probabilité conditionnelle en s'appuyant sur les effectifs, puis les fréquences.

Séquence 3 **Une approche fréquentiste**

Les 400 élèves de Terminale d'un lycée sont répartis en deux groupes : A et D selon leur langue vivante 1 (LV1) : anglais ou allemand. Le tableau suivant précise cette répartition pour les garçons (G) et les filles (F).

LV1	anglais (A)	allemand (D)
garçons (G)	130	50
filles (F)	140	80

Un élève est choisi au hasard dans le fichier des Terminales.

1) Calculer la probabilité :

a) que l'élève soit une fille (notée $P(F)$) ;

b) que l'élève ait pour LV1 l'allemand.

2) Sachant que l'élève est une fille, quelle est la probabilité que sa LV1 soit l'allemand ?

3) Comparer ce résultat avec $\frac{P(D \cap F)}{P(F)}$.

Nous avons calculé à la *question* **2) la probabilité de D sachant que F est réalisé.**
Ce nombre, que nous noterons $P(D/F)$) en cours, est égal à :

$$\frac{\text{effectif de } D \cap F}{\text{effectif de } F} = \frac{\text{fréquence de } D \cap F \text{ dans } E}{\text{fréquence de } F \text{ dans } E}.$$

Cela explique le résultat obtenu en **3)**.
Nous nous appuierons là-dessus pour la définition de la probabilité conditionnelle.

(Manuel Ter. S, collection Terracher, HACHETTE, 2002, p. 267)

- Pour faciliter l'étude des notions comme l'espérance mathématique, la variance et l'écart type d'une variable aléatoire, le manuel fait une liaison des notions équivalentes de la statistique et même des notations :

• $E(X)$ est **la moyenne** des valeurs x_i, **pondérées** par les valeurs p_i ; l'analogie avec la moyenne statistique permet de noter $E(X) = \bar{X}$.

(Manuel Ter. S, collection Terracher, HACHETTE, 2002, p. 271)

Et ensuite, le manuel dit « *Tout comme en statistique…* » pour une imitation :

• Tout comme en statistique, variance et écart-type sont des paramètres de dispersion. Plus qu'à leur valeur intrinsèque, il faut accorder une signification à leur comparaison, pour deux variables aléatoires définies sur le même ensemble.

(Manuel Ter. S, collection Terracher, HACHETTE, 2002, p. 271)

- Dans le manuel, « *statistique et modélisation* » se trouve dans la définition de la loi de probabilité. Un modèle pour modéliser une expérience aléatoire est dit bon lorsque « *les fréquences se rapprochent des probabilités* » théoriques. Là, nous voyons l'apparition de la notion de *fréquence* et la loi des grands nombres comme une théorie de référence :

Loi de probabilité

Lorsqu'une expérience aléatoire comporte un nombre fini d'issues[1], on définit sur l'ensemble $E = \{x_1, \ldots, x_r\}$ – E est appelé **univers** – des issues une **loi de probabilité** en se donnant une suite de nombres $(p_1, \ldots, p_r)$ vérifiant :

$$\textbf{pour tout } i \ (1 \leq i \leq r) : p_i \geq 0 \quad \text{et} \quad \sum_{i=1}^{r} p_i = 1.$$

Exemple : lorsqu'on prélève au hasard une boule dans l'urne représentée sur la *figure* 2, si l'on considère les trois issues R (rouge), B (bleue) et J (jaune), plusieurs choix des p_i remplissent les conditions précédentes, mais le modèle choisi n'est bon que lorsque **les fréquences statistiques f_i se rapprochent des p_i**, quand le nombre d'expériences devient grand.

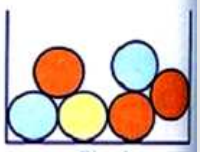


Fig. 2

Notre intuition nous conduit au modèle ci-contre :

issue x_i	R	B	J
probabilité p_i	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

(Manuel Ter. S, collection Terracher, HACHETTE, 2002, p. 268)

- Comme le programme dit de l'« Application aux expériences de références vues en seconde et première (dés, pièces, urnes, etc.) », dans les travaux pratiques du manuel, nous voyons la partie de l'« Utilisation des modèles de référence » avec l'objectif de reprendre des modèles classiques des expériences aléatoires :

1. Utilisation de modèles de référence

OBJECTIF

Utiliser des modèles – désormais bien connus – d'expériences aléatoires classiques (telles que **lancer de dés, de pièces, tirages dans une urne**) dans le traitement d'autres expériences aléatoires.

(Manuel Ter. S, collection Terracher, HACHETTE, 2002, p. 268)

Et ensuite, la plus part des exercices concerne des expériences aléatoires de tels modèles.

- La « Statistique et simulation » dit dans le programme est apparue comme la simulation d'expérience aléatoire dans deux chapitres concernant la probabilité :

+ à la fin du chapitre 9, il y a un seul exercice (problème 70, p. 293) où on demande de simuler une expérience aléatoire non équiprobable à l'aide d'une calculatrice ou d'un tableur :

70 Les deux pièces

Deux pièces de monnaie A et B, indiscernables à la vue et au toucher, sont truquées de telle sorte que « Face » sort avec la probabilité a (pièce A) ou b (pièce B).

On suppose qu'un joueur choisit une pièce au hasard et la lance ; s'il obtient « Face », il gagne et relance la même pièce ; s'il obtient « Pile », il perd et change de pièce avant de rejouer ; ainsi de suite.

On note respectivement, pour $n \geqslant 1$, A_n, B_n et S_n les événements :

• A_n : « la pièce A est lancée au n-ième coup » ;

• B_n : « la pièce B est lancée au n-ième coup » ;

• S_n : « le joueur gagne au n-ième coup ».

On pose enfin $\alpha_n = P(A_n)$ et $p_n = P(S_n)$.

1) *Simulation de l'expérience*

Simuler, à l'aide d'une calculatrice ou d'un tableur, cette expérience avec 100 lancers :

a) lorsque $a = 0{,}3$ et $b = 0{,}5$;

b) lorsque $a = 0{,}4$ et $b = 0{,}6$;

c) lorsque $a = b = 0{,}4$.

Dans chaque cas, on observera une stabilisation des valeurs de p_n vers un nombre à préciser.

(Manuel Ter. S, collection Terracher, HACHETTE, 2002, p. 293)

+ par contre, dans le chapitre 10, la simulation est proposée explicitement dans la partie théorique pour simuler la loi d'uniforme mais non dans la partie des exercices :

+ Partie d' « Avant-première » : le manuel montre une activité de préparation sous la forme d'une « séquence » pour une approche à la loi uniforme :

Séquence 2 « La loi uniforme »

1 Simulation

a) Choisir un intervalle $[a, b]$ de $[0, 1]$, simuler 1 000 tirages d'un réel de $[0, 1]$ (avec ALEA() ou rand()) et déterminer la fréquence des nombres « tombant » dans $[a, b]$.

b) Quelle conjecture peut-on émettre entre cette fréquence et la longueur de $[a, b]$ (diversifier dans la classe le choix de $[a, b]$) ?

2 Étude de la loi

La loi uniforme sur $[0, 1]$ « attribue » à chaque intervalle $[a, b]$ de $[0, 1]$ la probabilité :

$$P([a, b]) = b - a.$$

a) Représenter graphiquement la fonction F qui à $x (0 \leqslant x \leqslant 1)$ associe $F(x) = P([0, x])$.

b) Quelle est la densité de cette loi de probabilité ?

c) Interpréter par une aire, sur le graphique représentatif de f, chacune des probabilités suivantes :

- $P([0{,}6 ; 1])$, $P([0{,}2 ; 0{,}3] \cup [0{,}5 ; 1])$;
- $P\left(\left\{\frac{1}{3}\right\}\right)$ plus généralement $P(\{x_0\})$ où x_0 est donné dans $[0, 1]$;
- $P(A)$, $P(B)$, $P(A \cap B)$, $P(A \cup B)$ où : $A = [0{,}1 ; 0{,}6]$ et $B = [0{,}3 ; 0{,}7]$.

(Manuel Ter. S, collection Terracher, HACHETTE, 2002, p. 297)

+ Partie d' « Au cœur des problèmes » : un des travaux pratiques tourne « autour de la loi uniforme » avec l'objectif bien précisé de simuler pour relever la loi d'uniforme :

2. Autour de la loi uniforme

OBJECTIF

Choisir au hasard un réel dans $[0, 1]$, un point sur un segment ou sur un cercle, un instant entre midi et quatorze heures, etc. : autant de situations qui peuvent relever de la *loi uniforme*, autant de questions à examiner de près : **simulation, calcul de probabilité, recherche de densité**, etc. C'est l'idée.

Objectif de travail autour de la loi uniforme, p. 310, Manuel Ter S, TERRACHER, 2002

+ Puis, un autre est la « Simulations de « référence » » où il y a 4 thèmes à étudier avec l'aide de tableur d'Excel pour avoir des résultats de simulation d'une expérience aléatoire. Là, la fonction ALEA () est précisée comme « *une variable expérience aléatoire* ». L'extrait du thème 1 montre la « Mise en place de la simulation » et l'« Évolution de la fréquence » :

THÈME 1

« Calcul de probabilités » (voir *problème* 1, page 310)

Il s'agit de simuler à l'aide du logiciel *Excel* le choix d'un point M pris au hasard sur un demi-cercle Γ de diamètre $[AB]$ de centre O et de rayon 1 et d'évaluer la probabilité que aire$(OAM) \leqslant 0{,}25$.

• Étape 1 : Mise en place de la simulation

- *Colonne* 1 : numéro des tirages.
- *Colonne* 2 : valeurs prises par la variable aléatoire θ uniforme sur l'intervalle $[0, \pi[$, ce qui revient au même que sur $[0, \pi]$.

Dans la formule, **Pi()** est pour *Excel* une valeur approchée du nombre π et **ALEA()** est une variable aléatoire uniforme sur l'intervalle $[0, 1[$.

	1	2	3	4	5	6
1	N° tirage	θ	aire(AOM)	Fréquence		
2	1	1,284577932	0,47965919	0,00000000		
3	2	3,009271448	0,065967705	0,50000000		
4	3	1,756699457	0,491384861	0,33333333		
5		1,240005652	0,202777993	0,50000000		
6		2,461962001	0,314552802	0,50000000		
7	6					
8	7	0,754544072				

=PI()*ALEA()

=0,5*SIN(LC(-1))

=NB.SI(L2C3:LC(-1);"<=0,25")/LC(-3)

• Étape 2 : Évolution de la fréquence

Effectuer 1 000 tirages environ, sélectionner la colonne 4 et activer l'assistant graphique d'*Excel* pour obtenir le graphique (*Fig.* 13).
On prendra soin d'activer l'axe des ordonnées (par un double clic de la souris) afin d'afficher les ordonnées avec seulement deux chiffres après la virgule (pour une meilleure lisibilité) et de fixer le minimum de y à 0 et le maximum à 1.
Appuyer sur la touche F9 pour obtenir une autre série de tirages.
Conjecturer la valeur à laquelle se stabilise la fréquence. Pour affiner cette conjecture, fixer le minimum de y à 0,3 et le maximum à 0,4 et augmenter le nombre de tirages effectués.

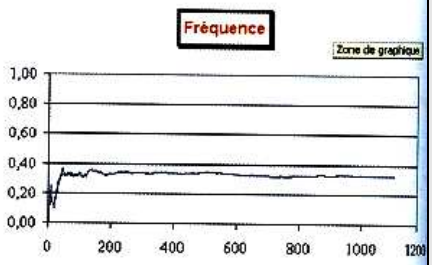


Fig. 13

(Manuel Ter. S, collection Terracher, HACHETTE, 2002, p. 316)

- Enfin, à propos d'« *adéquation de données expérimentales à une loi équirépartie* », on le trouve dans le rubrique « *Adéquation à une loi équirépartie* » avec l'objectif précisé suivant :

4. Adéquation à une loi équirépartie

OBJECTIF

Sensibiliser à la notion de preuve statistique.
À partir de résultats théoriques, simplement abordés à l'aide d'une simulation, il va s'agir de réfléchir sur des données (plutôt que d'appliquer des recettes), et décider le rejet d'un modèle théorique si une certaine fonction des données expérimentales a une valeur jugée *a priori* exceptionnelle.

PROBLÈME

Une expérience aléatoire conduisant à k issues $x_1, \ldots, x_k$, dont les probabilités respectives sont $p_1, \ldots, p_k$, est réalisée n fois.

1) a) Pour $1 \leqslant i \leqslant k$, on note N_i la variable aléatoire égale au nombre de réalisations de l'issue x_i, sur les n expériences.
Donner la loi de N_i et son espérance mathématique.

b) Pour mesurer les « écarts aux moyennes », on considère la variable aléatoire :

$$D^2 = \sum_{i=1}^{k} \frac{(N_i - np_i)^2}{np_i}.$$

Montrer que l'espérance mathématique de D^2 : $E(D^2)$ est indépendante de n et des p_i.

2) Dans le cas de 120 lancers d'un dé équilibré, conjecturer, à l'aide de simulations, les vingtiles v_{18} et v_{19} de la loi de D^2 (définis par $P(D^2 > v_{18}) = 0{,}1$ et $P(D^2 > v_{19}) = 0{,}05$).

3) On convient, pour une expérience analogue, de rejeter la régularité d'un dé *de façon significative au seuil de 10 % (ou de 5 %)* lorsque la valeur observée d^2 de D^2 dépasse v_{18} (ou v_{19}).

Pour les 120 lancers, les six numéros sont sortis respectivement 26, 13, 22, 27, 12 et 20 fois.
Le dé est-il régulier ?

1) a) Sans ambiguïté, N_i suit la **loi binomiale** $\mathcal{B}(n, p_i)$.

On en déduit :

$$\boldsymbol{E(N_i) = np_i} \text{ et } \boldsymbol{V(X) = np_i(1-p_i)}.$$

b) D'après la linéarité de l'espérance :

$$E(D^2) = \sum_{i=1}^{k} \frac{1}{np_i} E[(N_i - np_i)^2] = \sum_{i=1}^{k} \frac{1}{np_i} V(N_i).$$

On déduit de ce qui précède :

$$E(D^2) = \sum_{i=1}^{k} (1-p_i) = \sum_{i=1}^{k} 1 - \sum_{i=1}^{k} p_i.$$

Finalement :

$$\boldsymbol{E(D^2) = k-1}.$$

2) Nous avons simulé sur *Excel* 1 000 échantillons de 120 lancers d'un dé. Voici deux résumés visuels des résultats obtenus (*Fig.* 11 et *Fig.* 12).

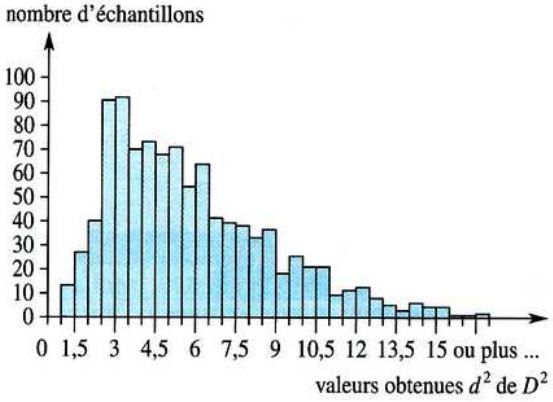


Fig. 11 : Histogramme

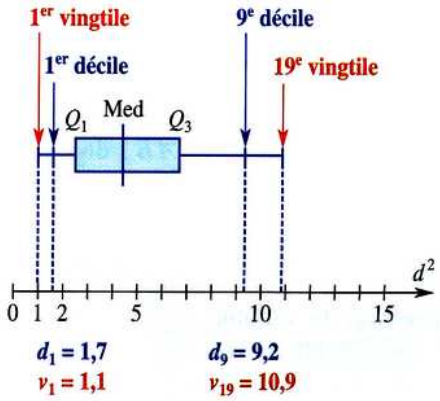


Fig. 12 : Boîte à moustaches

Commentaire : **d^2 dépasse la valeur 9,2 dans 10 % des cas, et dépasse 10,9 dans 5 % des cas.**

3) Pour le dé observé, $d^2 = 10,1$.

• $\boldsymbol{d^2 < v_{19}}$: au seuil de 5 %, on ne peut pas rejeter l'hypothèse de régularité du dé, c'est-à-dire le modèle équiréparti ;

• $\boldsymbol{d_9 < d^2}$: au seuil de 10 %, on décide de rejeter l'hypothèse de régularité du dé.

(Manuel Ter. S, collection Terracher, HACHETTE, 2002, p. 314)

Et la loi « khi-deux » se présente comme « petit infos » à la fin de ce rubrique. Dans la partie d'exercice, il y a en total 6 exercices concernant l'adéquation à une loi équirépartie.

Synthèse des types de tâches dans les manuels de 6è à Terminal après la réforme

Groupe	T	Types de Tâches	Arès 2000						
			6è	5è	4è	3è	2de	1er	Ter
0	T1	Simuler une expérience aléatoire n fois / Etablir une série statistique					**X**	x	x
	T2	Exécuter un programme pour simuler une expérience aléatoire (pour une machine à calcul)					x		
	T3	Faire un compte rendu de l'expérience					x		
	T4	Recueillir des données dans la vie réelle puis les écrire en tableau							
1	T5	Lire des informations d'un tableau	**X**		x				
	T6	Lire des informations à partir d'une représentation graphique							
	T7	Représenter en graphique des données statistiques	**X**	**X**	**X**		x	x	
	T8	Représenter/Regrouper des données statistiques en tableau	**X**	**X**		**X**			
	T9	Représenter une série statistique par la boite moustaches						**X**	
	T10	Représenter la boîte de dispersion associée à une série statistique						x	
	T11	Chercher un exemple de « grandeurs » proportionnelles / non proportionnelles							
	T12	Calculer la mesure d'angle d'un diagramme qui représente des données statistiques		**X**					
	T13	Décrire la statistique (population, caractère étudié, type de caractère)							
	T14	Effectuer un regroupement des données			**X**			x	
	T15	Donner l'allure de la courbe des fréquences cumulées croissantes							
	T16	Lire des indicateurs (l'effectif, la moyenne, la médian) dans une représentation graphique			x	x			
	T17	Indiquer la nature des caractères					x		
	T18	Dresser le tableau des effectifs (à partir des graphiques)					**X**		
	T19	Dresser le tableau des fréquences (à partir des graphiques)							
	T20	Trouver la classe modale					**X**		
	T21	Établir la distribution des fréquences (à partir des données)					**X**		
	T22	Chercher la valeur la plus grande/petite d'une série statistique	**X**			x			
	T23	Chercher la valeur la plus fréquente	x						
	T24	Construire un diagramme avec un tableur	x						
2	T25	Compléter un tableau (des « grandeurs » proportionnelles)							
	T26	Chercher le coefficient de proportionnalité d'un tableau (des « grandeurs » proportionnelles)		**X**					
	T27	Vérifier si un couple de deux entiers est proportionnel à un autre couple							
	T28	Calculer l'effectif		**X**	**X**				
	T29	Calculer des effectifs cumulés			**X**	**X**			
	T30	Constituer un tableau des effectifs cumulés (croissant)		**X**	x	**X**			

	T31	Constituer un tableau des fréquences cumulées			x				
	T32	Calculer le pourcentage	**X**		**X**	**X**	x	x	
	T33	Transformer des répartitions de données en pourcentages							
	T34	Calculer la fréquence		**X**	**X**		**X**		
	T35	Calculer les fréquences cumulées			**X**		**X**		
	T36	Calculer la moyenne			**X**	**X**	**X**	**X**	
	T37	Estimer la moyenne des données regroupées							
	T38	Calculer une valeur pour ajouter à une liste des valeurs statistiques telles que la moyenne ne change pas							
	T39	Expliquer comment on peut calculer la moyenne pour l'ensemble de population							
	T40	Calculer la valeur d'un terme manquant d'une suite connaissant la moyenne			**X**	**X**			
	T41	Calculer la valeur d'un terme manquant d'une suite connaissant la médiane				**X**			
	T42	Calculer la valeur d'un terme manquant d'une suite connaissant l'étendue				**X**			
	T43	Préciser si la moyenne existe dans un énoncé et elle a une signification						x	
	T44	Calculer la taille d'une série statistique connaissant des hypothèses sur les moyennes							
	T45	Préciser la position relative de la moyenne par rapport aux valeurs de série statistique			x				
	T46	Calculer la taille de données statistiques							
	T47	Calculer la médiane				**X**	**X**	x	
	T48	Calculer la mode					**X**		
	T49	Calculer l'étendue				**X**	**X**	x	
	T50	Calculer la variance						**X**	
	T51	Calculer l'écart type						**X**	
	T52	Calculer une valeur à partir le pourcentage et le nombre total des valeurs statistiques	x	x	x				
	T53	Compéter un tableau des effectifs sachant certains effectifs et pourcentages		**X**					
	T54	Déterminer les quartiles						**X**	
	T55	Déterminer l'intervalle interquartile						**X**	
	T56	Calculer le pourcentage de données situées dans l'intervalle [x-s; x+s] / [x-2s; x+2s] / [x-3s; x+3s]						x	
	T57	Calculer les bornes de l'intervalle théorique					**X**		
	T58	Calculer les bornes de l'intervalle de confiance pour la fréquence observée					**X**		
	T59	Construire une série statistique connaissant la moyenne et/ou la médiane			**X**	**X**			
	T60	Prévoir une évolution des moyennes lors de changement de population							
	T61	Chercher deux valeurs extrêmes où se trouve le nombre d'expérience aléatoire pour que la distribution des fréquences soit équirépartie							x
3	T62	Comparer les moyennes des séries statistiques				x			
	T63	Comparer les variances des séries statistiques						x	
	T64	Commenter des séries statistiques							

	T65	Comparer la répartition de deux séries statistiques							
	T66	Comparer la distribution des fréquences des échantillons différents					**X**		
	T67	Comparer les indicateurs statistiques des échantillons différents					**X**		
	T68	Comparer des diagrammes différents de mêmes données statistiques	x	x					
	T69	Comparer deux diagrammes des données statistiques différentes	x						
	T70	Comparer deux séries statistiques						x	
	T71	Comparer les fluctuations d'échantillonnage des séries statistiques						x	
	T72	Comparer une distribution de fréquence à une loi de probabilité							x
4	T73	Interpréter des indicateurs statistiques							
	T74	Interpréter les résultats statistiques							
	T75	Donner la pensé à la lecture des résultats de fréquence et d'effectif							x
	T76	En observant le tableau de données, répondre si la moyenne pourrait être utilisée comme un représentant pour le caractère étudié							
	T77	Présenter des remarques sur le caractère étudié à partir du tableau de distribution des fréquences de données regroupées en classe							
	T78	Présenter des remarques sur le caractère étudié à partir de la représentation graphique des fréquences de données regroupées en classe							
5	T79	Observer la fluctuation d'échantillonnage					**X**		
	T80	Modéliser et simuler une expérience aléatoire							
	T81	Évaluer les indicateurs d'une population à partir d'un échantillon							

Tableau 114. Les types de tâches dans les manuels secondaires après la réforme

I.2.3. Conclusion (après la réforme) :

- A partir du tableau des notions statistiques et probabilistes proposées à enseigner dans les programmes du lycée après la réforme (Tableau 18) et tableau 22 précédent, la statistique dans classe seconde se concentre sur la simulation et la fluctuation d'échantillonnage tandis que les notions enseignées en seconde avant la réforme comme la variance, l'écart type sont déplacés à classe première. En plus, dans la classe première, il apparaît de nouvelles notions comme diagramme en boite (boite moustache), interquartile, intervalle interquartile qui servent à interpréter et à comparer des séries statistiques. Ce fait pourrait être pour que la statistique reste plus longtemps au lycée. La simulation d'une expérience aléatoire se trouve dans le programme de tous les trois niveaux de lycée et la modélisation d'une expérience aléatoire n'apparairent qu'en classe première et classe terminal. Le lien entre la statistique et la probabilité est bien mis en accent avec la présence de la statistique dans tout au long du programme secondaire, c'est-à-dire de la classe sixième à la classe terminale. Cela permet de dire qu'il y a un « fil statistique » dans le programme du collège au lycée.

- Les types de tâches se trouvent dans tous les six groupes et particulièrement, les types des groupes 4 et 5 concentrent au niveau de classe seconde.

I.3. Conclusion

1. Au niveau de collège, les notions proposées à enseigner dans le programme ne changent pas beaucoup sauf que l'étendue est la seule nouvelle notion enseignée au collège après la réforme par rapport à l'avance. Mais, on peut voir que dans le programme après la réforme, les notions statistiques se répètent dans les niveaux consécutifs pour renforcer des connaissances de statistique au collège tandis qu'avant la réforme, ces notion se présentent séparément dans chaque niveau (voir Tableau 23). Par contre, au niveau de lycée, les notions de la statistique sont enseignées prolongement jusqu'à la classe Première et la présence de la statistique se trouve jusqu'à la classe terminale.

	6è	5è	4è	3è	2de	1er S	Ter. S
Avant la réforme	Diagramme	Effectif, Fréquence	Moyenne	Médiane	Effectifs, Effectifs cumulés, Fréquences, Fréquence cumulées. Moyenne Ecart type		
Après la réforme	Diagramme	Classe Effectif Fréquence	Effectif Fréquence Moyenne	Moyenne Médiane Étendue	Moyenne, Moyenne élaguée Médiane Étendue Classe modale Simulation Fluctuation d'échantillonnage Modélisation	Moyenne Médiane Variance Ecart type Quartile Intervalle interquartile Boite de dispersion Modélisation	Modélisation

Tableau 115. Les notions statistiques dans les programmes du collège-lycée avant et après la réforme

2. À propos du niveau de classe seconde, voyons le tableau 26 où se trouvent des types de tâches communs, des présents avant la réforme mais disparus après la réforme, et des absents avant la réforme mais apparus après la réforme, concrètement, comme suivant :

- Pour groupe 1 : T8 et T17 sont les deux communs, tandis que T13, T14, T15 ont disparus après la réforme et par contre T18, T20, T21 ne se présentent qu'après la réforme.

- Pour groupe 2 : le travail sur la statistique ne concentre pas trop sur des calculs divers mais sur les calculs des effectifs, des fréquences, de moyenne, de mode, de médiane et d'étendue

sauf T52 (calculer écart type) est particulièrement un travail principal avant la réforme. Par contre, T47, T49, T57, T58 n'existent qu'après la réforme.

- Pour groupes 3, 4, 5 : seul T65 du groupe 3 apparaît avant la réforme et aucun de groupe 4, 5 se trouvant avant la réforme tandis que les T66, T67 concernant la comparaison, puis T79 (Observer la fluctuation d'échantillonnage), T1, T2, T3 ne se présentent qu'après la réforme. Ainsi, le travail est beaucoup concentré sur la simulation et l'observation la fluctuation d'échantillonnage.

Alors, pourrait-on dire qu'il y a une modification du rôle de la classe de seconde dans le projet statistique ?

3. À propos de la comparaison des séries statistiques, regardons sur l'extrait des types de tâche du groupe 3 du tableau 24, on pourrait voir qu'il y un nombreux de types de tâches de comparaison des séries statistiques et ils se trouvent de classe 6è à la classe terminale après l'2000.

Synthèse des types de tâches du groupe 3 dans les manuels secondaire avant et après la réforme

Groupe	T	Types de Tâches	AVANT 2000					APRES 2000						
			6è	5è	4è	3è	2de	6è	5è	4è	3è	2de	1er	Ter
3	T62	Comparer les moyennes des séries statistiques		x							x			
	T63	Comparer les variances des séries statistiques											x	
	T64	Commenter des séries statistiques				x								
	T65	Comparer la répartition de deux séries statistiques					**X**							
	T66	Comparer la distribution des fréquences des échantillons différents										**X**		
	T67	Comparer les indicateurs statistiques des échantillons différents										**X**		
	T68	Comparer des diagrammes différents de mêmes données statistiques						x	x					
	T69	Comparer deux diagrammes des données statistiques différentes						x						
	T70	Comparer deux séries statistiques											x	
	T71	Comparer les fluctuations d'échantillonnage des séries statistiques											x	
	T72	Comparer une distribution de fréquence à une loi de probabilité												x

Tableau 116. Les types de tâches du groupe 3 avant et après la réforme

Alors, est-ce que l'intension du programme et des auteurs des manuels est de compléter la praxéologie de comparaison ?

4. Cette analyse nous provoque de nous poser des hypothèses à étudier suivants :

H1. Il y a un renforcement de la praxéologie de comparaison.

H2. Il y a un renforcement des liens entre statistique et probabilité.

5. Des questions à répondre :

Q1. Est-ce que l'intension de faire changer d'enseignement des enseignants en France ?

Q2. Est-ce que le changement du contenue de la statistique en lien avec la probabilité modifie les connaissances de statistique chez élèves ?

Synthèse des types de tâches dans les manuels secondaire avant et après la réforme

Groupe	T	Types de Tâches	Avant 2000	Après 2000
0	T1	Simuler une expérience aléatoire n fois / Etablir une série statistique		x
	T2	Exécuter un programme pour simuler une expérience aléatoire (pour une machine à calcul)		x
	T3	Faire un compte rendu de l'expérience		x
	T4	Recueillir des données dans la vie réelle puis les écrire en tableau		
1	T5	Lire des informations d'un tableau	x	x
	T6	Lire des informations à partir d'une représentation graphique		
	T7	Représenter en graphique des données statistiques	x	x
	T8	Représenter/Regrouper des données statistiques en tableau		x
	T9	Représenter une série statistique par la boite moustaches		x
	T10	Représenter la boîte de dispersion associée à une série statistique		x
	T11	Chercher un exemple de « grandeurs » proportionnelles / non proportionnelles	x	
	T12	Calculer la mesure d'angle d'un diagramme qui représente des données statistiques	x	x
	T13	Décrire la statistique (population, caractère étudié, type de caractère)	x	
	T14	Effectuer un regroupement des données	x	x
	T15	Donner l'allure de la courbe des fréquences cumulées croissantes	x	
	T16	Lire des indicateurs (l'effectif, la moyenne, la médian) dans une représentation graphique		x
	T17	Indiquer la nature des caractères	x	x
	T18	Dresser le tableau des effectifs (à partir des graphiques)		x
	T19	Dresser le tableau des fréquences (à partir des graphiques)		
	T20	Trouver la classe modale		x
	T21	Établir la distribution des fréquences (à partir des données)		x
	T22	Chercher la valeur la plus grande/petite d'une série statistique		x
	T23	Chercher la valeur la plus fréquente		x
	T24	Construire un diagramme avec un tableur		x
2	T25	Compléter un tableau (des « grandeurs » proportionnelles)	x	
	T26	Chercher le coefficient de proportionnalité d'un tableau (des « grandeurs » proportionnelles)	x	x
	T27	Vérifier si un couple de deux entiers est proportionnel à un autre couple	x	
	T28	Calculer l'effectif	x	x
	T29	Calculer des effectifs cumulés		x
	T30	Constituer un tableau des effectifs cumulés (croissant)		x
	T31	Constituer un tableau des fréquences cumulées	x	x

	T32	Calculer le pourcentage	x	x
	T33	Transformer des répartitions de données en pourcentages	x	
	T34	Calculer la fréquence	x	x
	T35	Calculer les fréquences cumulées		x
	T36	Calculer la moyenne	x	x
	T37	Estimer la moyenne des données regroupées		
	T38	Calculer une valeur pour ajouter à une liste des valeurs statistiques telles que la moyenne ne change pas	x	
	T39	Expliquer comment on peut calculer la moyenne pour l'ensemble de population	x	
	T40	Calculer la valeur d'un terme manquant d'une suite connaissant la moyenne	x	x
	T41	Calculer la valeur d'un terme manquant d'une suite connaissant la médiane		x
	T42	Calculer la valeur d'un terme manquant d'une suite connaissant l'étendue		x
	T43	Préciser si la moyenne existe dans un énoncé et elle a une signification	x	x
	T44	Calculer la taille d'une série statistique connaissant des hypothèses sur les moyennes	x	
	T45	Préciser la position relative de la moyenne par rapport aux valeurs de série statistique		x
	T46	Calculer la taille de données statistiques	x	
	T47	Calculer la médiane	x	x
	T48	Calculer la mode	x	x
	T49	Calculer l'étendue	x	x
	T50	Calculer la variance		x
	T51	Calculer l'écart type	x	x
	T52	Calculer une valeur à partir le pourcentage et le nombre total des valeurs statistiques		x
	T53	Compéter un tableau des effectifs sachant certains effectifs et pourcentages		x
	T54	Déterminer les quartiles		x
	T55	Déterminer l'intervalle interquartile		x
	T56	Calculer le pourcentage de données situées dans l'intervalle [x-s; x+s] / [x-2s; x+2s] / [x-3s; x+3s]	x	x
	T57	Calculer les bornes de l'intervalle théorique		x
	T58	Calculer les bornes de l'intervalle de confiance pour la fréquence observée		x
	T59	Construire une série statistique connaissant la moyenne et/ou la médiane	x	x
	T60	Prévoir une évolution des moyennes lors de changement de population	x	
	T61	Chercher deux valeurs extrêmes où se trouve le nombre d'expérience aléatoire pour que la distribution des fréquences soit équirépartie		x
3	T62	Comparer les moyennes des séries statistiques	x	x
	T63	Comparer les variances des séries statistiques		x
	T64	Commenter des séries statistiques	x	
	T65	Comparer la répartition de deux séries statistiques	x	
	T66	Comparer la distribution des fréquences des échantillons différents		x
	T67	Comparer les indicateurs statistiques des échantillons différents		x
	T68	Comparer des diagrammes différents de mêmes données statistiques		x
	T69	Comparer deux diagrammes des données statistiques différentes		x
	T70	Comparer deux séries statistiques		x
	T71	Comparer les fluctuations d'échantillonnage des séries statistiques		x
	T72	Comparer une distribution de fréquence à une loi de probabilité		
4	T73	Interpréter des indicateurs statistiques	x	
	T74	Interpréter les résultats statistiques	x	
	T75	Donner la pensé à la lecture des résultats de fréquence et d'effectif		x

	T76	En observant le tableau de données, répondre si la moyenne pourrait être utilisée comme un représentant pour le caractère étudié		
	T77	Présenter des remarques sur le caractère étudié à partir du tableau de distribution des fréquences de données regroupées en classe		
	T78	Présenter des remarques sur le caractère étudié à partir de la représentation graphique des fréquences de données regroupées en classe		
5	T79	Observer la fluctuation d'échantillonnage		x
	T80	Modéliser et simuler une expérience aléatoire		
	T81	Évaluer les indicateurs d'une population à partir d'un échantillon		

Tableau 117. Les types de tâches dans les manuels secondaires avant et après la réforme

Synthèse des types de tâches dans les manuels français avant et après la réforme

Groupe	T	Types de Tâches	Avant 2000	Après 2000
0	T1	Simuler une expérience aléatoire n fois / Etablir une série statistique		x
	T2	Exécuter un programme pour simuler une expérience aléatoire (pour une machine à calcul)		x
	T3	Faire un compte rendu de l'expérience		x
	T4	Recueillir des données dans la vie réelle puis les écrire en tableau		
1	T5	Lire des informations d'un tableau	x	x
	T6	Lire des informations à partir d'une représentation graphique		
	T7	Représenter en graphique des données statistiques	x	x
	T8	Représenter/Regrouper des données statistiques en tableau		x
	T9	Représenter une série statistique par la boite moustaches		x
	T10	Représenter la boîte de dispersion associée à une série statistique		x
	T11	Chercher un exemple de « grandeurs » proportionnelles / non proportionnelles	x	
	T12	Calculer la mesure d'angle d'un diagramme qui représente des données statistiques	x	x
	T13	Décrire la statistique (population, caractère étudié, type de caractère)	x	
	T14	Effectuer un regroupement des données	x	x
	T15	Donner l'allure de la courbe des fréquences cumulées croissantes	x	
	T16	Lire des indicateurs (l'effectif, la moyenne, la médian) dans une représentation graphique		x
	T17	Indiquer la nature des caractères	x	x
	T18	Dresser le tableau des effectifs (à partir des graphiques)		x
	T19	Dresser le tableau des fréquences (à partir des graphiques)		
	T20	Trouver la classe modale		x
	T21	Établir la distribution des fréquences (à partir des données)		x
	T22	Chercher la valeur la plus grande/petite d'une série statistique		x
	T23	Chercher la valeur la plus fréquente		x
	T24	Construire un diagramme avec un tableur		x
2	T25	Compléter un tableau (des « grandeurs » proportionnelles)	x	
	T26	Chercher le coefficient de proportionnalité d'un tableau (des « grandeurs » proportionnelles)	x	x
	T27	Vérifier si un couple de deux entiers est proportionnel à un autre couple	x	
	T28	Calculer l'effectif	x	x
	T29	Calculer des effectifs cumulés		x
	T30	Constituer un tableau des effectifs cumulés (croissant)		x
	T31	Constituer un tableau des fréquences cumulées	x	x

	T32	Calculer le pourcentage	x	x
	T33	Transformer des répartitions de données en pourcentages	x	
	T34	Calculer la fréquence	x	x
	T35	Calculer les fréquences cumulées		x
	T36	Calculer la moyenne	x	x
	T37	Estimer la moyenne des données regroupées		
	T38	Calculer une valeur pour ajouter à une liste des valeurs statistiques telles que la moyenne ne change pas	x	
	T39	Expliquer comment on peut calculer la moyenne pour l'ensemble de population	x	
	T40	Calculer la valeur d'un terme manquant d'une suite connaissant la moyenne	x	x
	T41	Calculer la valeur d'un terme manquant d'une suite connaissant la médiane		x
	T42	Calculer la valeur d'un terme manquant d'une suite connaissant l'étendue		x
	T43	Préciser si la moyenne existe dans un énoncé et elle a une signification	x	x
	T44	Calculer la taille d'une série statistique connaissant des hypothèses sur les moyennes	x	
	T45	Préciser la position relative de la moyenne par rapport aux valeurs de série statistique		x
	T46	Calculer la taille de données statistiques	x	
	T47	Calculer la médiane	x	x
	T48	Calculer la mode	x	x
	T49	Calculer l'étendue	x	x
	T50	Calculer la variance		x
	T51	Calculer l'écart type	x	x
	T52	Calculer une valeur à partir le pourcentage et le nombre total des valeurs statistiques		x
	T53	Compéter un tableau des effectifs sachant certains effectifs et pourcentages		x
	T54	Déterminer les quartiles		x
	T55	Déterminer l'intervalle interquartile		x
	T56	Calculer le pourcentage de données situées dans l'intervalle [x-s; x+s] / [x-2s; x+2s] / [x-3s; x+3s]	x	x
	T57	Calculer les bornes de l'intervalle théorique		x
	T58	Calculer les bornes de l'intervalle de confiance pour la fréquence observée		x
	T59	Construire une série statistique connaissant la moyenne et/ou la médiane	x	x
	T60	Prévoir une évolution des moyennes lors de changement de population	x	
	T61	Chercher deux valeurs extrêmes où se trouve le nombre d'expérience aléatoire pour que la distribution des fréquences soit équirépartie		x
3	T62	Comparer les moyennes des séries statistiques	x	x
	T63	Comparer les variances des séries statistiques		x
	T64	Commenter des séries statistiques	x	
	T65	Comparer la répartition de deux séries statistiques	x	
	T66	Comparer la distribution des fréquences des échantillons différents		x
	T67	Comparer les indicateurs statistiques des échantillons différents		x
	T68	Comparer des diagrammes différents de mêmes données statistiques		x
	T69	Comparer deux diagrammes des données statistiques différentes		x
	T70	Comparer deux séries statistiques		x
	T71	Comparer les fluctuations d'échantillonnage des séries statistiques		x
	T72	Comparer une distribution de fréquence à une loi de probabilité		
4	T73	Interpréter des indicateurs statistiques	x	

	T74	Interpréter les résultats statistiques	x	
	T75	Donner la pensé à la lecture des résultats de fréquence et d'effectif		x
	T76	En observant le tableau de données, répondre si la moyenne pourrait être utilisée comme un représentant pour le caractère étudié		
	T77	Présenter des remarques sur le caractère étudié à partir du tableau de distribution des fréquences de données regroupées en classe		
	T78	Présenter des remarques sur le caractère étudié à partir de la représentation graphique des fréquences de données regroupées en classe		
5	T79	Observer la fluctuation d'échantillonnage		x
	T80	Modéliser et simuler une expérience aléatoire		
	T81	Évaluer les indicateurs d'une population à partir d'un échantillon		

Tableau 118. Les types de tâches dans les manuels français avant et après la réforme

Synthèse des types de tâches dans les manuels de classe 2de avant et après la réforme

Groupe	T	Types de Tâches	Avant 2000	Après 2000
0	T1	Simuler une expérience aléatoire n fois / Etablir une série statistique		**X**
	T2	Exécuter un programme pour simuler une expérience aléatoire (pour une machine à calcul)		x
	T3	Faire un compte rendu de l'expérience		x
	T4	Recueillir des données dans la vie réelle puis les écrire en tableau		
1	T5	Lire des informations d'un tableau		
	T6	Lire des informations à partir d'une représentation graphique		
	T7	Représenter en graphique des données statistiques	**X**	x
	T8	Représenter/Regrouper des données statistiques en tableau		
	T9	Représenter une série statistique par la boite moustaches		
	T10	Représenter la boîte de dispersion associée à une série statistique		
	T11	Chercher un exemple de « grandeurs » proportionnelles / non proportionnelles		
	T12	Calculer la mesure d'angle d'un diagramme qui représente des données statistiques		
	T13	Décrire la statistique (population, caractère étudié, type de caractère)	x	
	T14	Effectuer un regroupement des données	**X**	
	T15	Donner l'allure de la courbe des fréquences cumulées croissantes	x	
	T16	Lire des indicateurs (l'effectif, la moyenne, la médian) dans une représentation graphique		
	T17	Indiquer la nature des caractères	x	x
	T18	Dresser le tableau des effectifs (à partir des graphiques)		**X**
	T19	Dresser le tableau des fréquences (à partir des graphiques)		
	T20	Trouver la classe modale		**X**
	T21	Établir la distribution des fréquences (à partir des données)		**X**
	T22	Chercher la valeur la plus grande/petite d'une série statistique		
	T23	Chercher la valeur la plus fréquente		

	T24	Construire un diagramme avec un tableur		
2	T25	Compléter un tableau (des « grandeurs » proportionnelles)		
	T26	Chercher le coefficient de proportionnalité d'un tableau (des « grandeurs » proportionnelles)		
	T27	Vérifier si un couple de deux entiers est proportionnel à un autre couple		
	T28	Calculer l'effectif		
	T29	Calculer des effectifs cumulés		
	T30	Constituer un tableau des effectifs cumulés (croissant)		
	T31	Constituer un tableau des fréquences cumulées	**X**	
	T32	Calculer le pourcentage	x	x
	T33	Transformer des répartitions de données en pourcentages		
	T34	Calculer la fréquence	**X**	**X**
	T35	Calculer les fréquences cumulées		**X**
	T36	Calculer la moyenne	**X**	**X**
	T37	Estimer la moyenne des données regroupées		
	T38	Calculer une valeur pour ajouter à une liste des valeurs statistiques telles que la moyenne ne change pas		
	T39	Expliquer comment on peut calculer la moyenne pour l'ensemble de population		
	T40	Calculer la valeur d'un terme manquant d'une suite connaissant la moyenne	**X**	
	T41	Calculer la valeur d'un terme manquant d'une suite connaissant la médiane		
	T42	Calculer la valeur d'un terme manquant d'une suite connaissant l'étendue		
	T43	Préciser si la moyenne existe dans un énoncé et elle a une signification	x	
	T44	Calculer la taille d'une série statistique connaissant des hypothèses sur les moyennes	x	
	T45	Préciser la position relative de la moyenne par rapport aux valeurs de série statistique		
	T46	Calculer la taille de données statistiques		
	T47	Calculer la médiane		**X**
	T48	Calculer la mode	**X**	**X**
	T49	Calculer l'étendue	x	**X**
	T50	Calculer la variance		
	T51	Calculer l'écart type	**X**	
	T52	Calculer une valeur à partir le pourcentage et le nombre total des valeurs statistiques		
	T53	Compéter un tableau des effectifs sachant certains effectifs et pourcentages		
	T54	Déterminer les quartiles		
	T55	Déterminer l'intervalle interquartile		
	T56	Calculer le pourcentage de données situées dans l'intervalle [x-s; x+s] / [x-2s; x+2s] / [x-3s; x+3s]	x	
	T57	Calculer les bornes de l'intervalle théorique		**X**
	T58	Calculer les bornes de l'intervalle de confiance pour la fréquence observée		**X**
	T59	Construire une série statistique connaissant la moyenne et/ou la médiane		
	T60	Prévoir une évolution des moyennes lors de changement de population		
	T61	Chercher deux valeurs extrêmes où se trouve le nombre d'expérience aléatoire pour que la distribution des fréquences soit équirépartie		
3	T62	Comparer les moyennes des séries statistiques		

	T63	Comparer les variances des séries statistiques		
	T64	Commenter des séries statistiques		
	T65	Comparer la répartition de deux séries statistiques	**X**	
	T66	Comparer la distribution des fréquences des échantillons différents		**X**
	T67	Comparer les indicateurs statistiques des échantillons différents		**X**
	T68	Comparer des diagrammes différents de mêmes données statistiques		
	T69	Comparer deux diagrammes des données statistiques différentes		
	T70	Comparer deux séries statistiques		
	T71	Comparer les fluctuations d'échantillonnage des séries statistiques		
	T72	Comparer une distribution de fréquence à une loi de probabilité		
4	T73	Interpréter des indicateurs statistiques		
	T74	Interpréter les résultats statistiques		
	T75	Donner la pensé à la lecture des résultats de fréquence et d'effectif		
	T76	En observant le tableau de données, répondre si la moyenne pourrait être utilisée comme un représentant pour le caractère étudié		
	T77	Présenter des remarques sur le caractère étudié à partir du tableau de distribution des fréquences de données regroupées en classe		
	T78	Présenter des remarques sur le caractère étudié à partir de la représentation graphique des fréquences de données regroupées en classe		
5	T79	Observer la fluctuation d'échantillonnage		**X**
	T80	Modéliser et simuler une expérience aléatoire		
	T81	Évaluer les indicateurs d’une population à partir d’un échantillon		

Tableau 119. Les types de tâches dans la classe **2de** avant et après la réforme

II. Institution vietnamienne

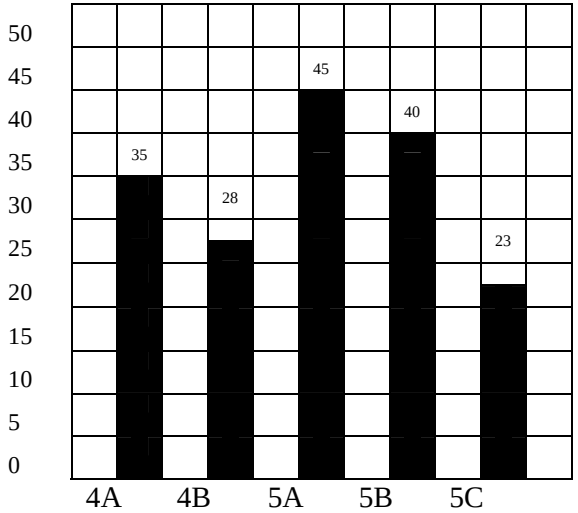
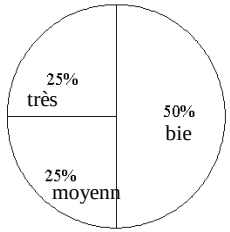
II.1. Analyse du programme

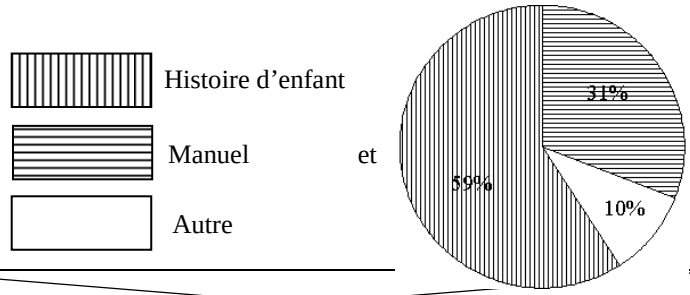
Les programmes des mathématiques actuels du Vietnam (2006)

<table>
<tr><td>Classe</td><td>Contenus</td><td>Compétences exigibles</td><td>Note</td></tr>
<tr><td>1 (CP)</td><td></td><td></td><td></td></tr>
<tr><td>2 (CE1)</td><td></td><td></td><td></td></tr>
<tr><td>3
(CE2)</td><td>Eléments de statistique</td><td>1) Première approche d’suite des données. Savoir ordonner des données pour former une suite de données.

2) Premiers contacts avec le tableau de données. Comprendre les données dans un tableau statistique simple, savoir lire un tableau statistique et donner des observations sur le tableau statistique.</td><td>1) Ex1 : Les quatre amis Dung, Ha, Hung, Quan mesurent respectivement : 129cm ; 132 cm ; 125cm ; 135cm.
S’appuyer sur les données précédentes pour répondre aux questions suivantes :
- Quel est la taille de Hung ?
- Qui est le plus grand, le plus petit ?
- En combien de cm, Dung est plus grand que Hung ?
Ex2 : Les poids des packs du riz sont notés comme suivant :
Ecrire les poids en kg de 5 packs du riz :
a) en ordre croissant
b) en ordre décroissant

2) Ex : Soit le tableau statistique des plants que les classes 3 ont plantés :
<table><tr><td>Classe</td><td>3A</td><td>3B</td><td>3C</td><td>3D</td></tr><tr><td>Nombre de plants</td><td>40</td><td>25</td><td>45</td><td>28</td></tr></table>Observer le tableau précédent pour répondre aux questions suivantes :
a) Classe 3C, combien de plants a-elle planté ?
b) Quelle est la classe ayant planté le plus de plants ? Le moins ?
c) Classe 3A et 3C, combien de plants ont-elles plantés ?</td></tr>
<tr><td>4</td><td>Eléments de statistique</td><td>1) Savoir calculer la moyenne de</td><td>1) Ex : Calculer la moyenne des nombres suivants : 36, 42 et 57.</td></tr>
</table>

(CM1)		plusieurs nombres. 2) Savoir relever simplement certaines informations sur un diagramme en bâtons.	2) Ex : Le diagramme suivant présente le nombre de plants plantés par des classes 4 et classes 5.  Observer le diagramme pour répondre aux questions suivantes : a) Quelles sont les classes qui ont planté des plants ? b) Classe 4A, combien de plants a-elle planté ? Classe 5B ? Classe 5C
5 (CM2)	Diagramme circulaire	1) Savoir reconnaître un diagramme circulaire et sa signification réelle. 2) Savoir recueillir des données et traiter simplement certaines informations à partir d'un diagramme circulaire	1) Ex : La figure suivante est un diagramme qui représente les résultats finaux d'une classe 5A. En regardant ce diagramme, on peut lire certains résultat d'une classe 5A comme : 25% d'élèves très bien 50% d'élèves bien 25% d'élèves moyens.  2) Ex. Diagramme circulaire relatif au pourcentage des livres dans une bibliothèque d'une école primaire :

			
6 (6è)			
7 (5è)	Recueil de données. Effectif. Tableau des effectifs et diagrammes des effectifs (diagramme en segments ou diagramme en rectangles). Moyenne, mode du tableau de données.	Connaissances : - Savoir des notions: données statistiques, effectif. - Savoir tableau des effectifs, diagramme en segments, diagramme en rectangles. Savoir-faire: - Comprendre et savoir appliquer la moyenne, mode du tableau des données dans les situations réelles. - Savoir recueillir des données statistiques - Savoir présenter des données statistiques par un tableau des effectifs, par un diagramme en segments ou en rectangles.	Ex: Effectuer des faits suivants: a) Copier des notes du contrôle des mathématiques à la fin du premier semestre de chaque élève de la classe. b) Dresser le tableau des effectifs et le diagramme en bâtons correspondant. c) Donner des observations en utilisant le tableau des effectifs effectué (nombre des valeurs des caractères ; nombre des valeurs différentes ; la valeur la plus grande, la plus petite ; la valeur ayant le plus grand effectif ; l'intervalle de la plus part des valeurs) d) Calculer la moyenne des valeurs statistiques.
8 (4è)			
9 (3è)			
10 (seconde)	1. Tableau de distribution des effectifs – Effectif. Tableau de distribution des fréquences – Fréquences des classes regroupées.	Connaissances : - Savoir des notions : effectif, fréquence de chaque valeur de la série statistique (échantillon), tableau de distribution des effectifs–fréquences,	- Ne pas demander (aux élèves) : savoir regrouper en classe et savoir dans quels cas on doit effectuer un tableau de distribution des effectifs - fréquences des données regroupées. - La présentation du contenu doit être effectué à la fois avec l'étude des problèmes réels.

<table>
<tr>
<td></td>
<td></td>
<td>tableau des effectifs–fréquences de données regroupées en classes.
Savoir-faire :
- Préciser l’effectif, la fréquence de chaque valeur de la série statistique.
- Dresser le tableau des effectifs–fréquences de données regroupées en classe</td>
<td>- Remarquer la valeur centrale de chaque classe.
Ex : Les tailles de 30 élèves d’une classe 10 sont listées dans le tableau suivant (unité mètre):
<table>
<tr><td>1,45</td><td>1,58</td><td>1,61</td><td>1,52</td><td>1,52</td><td>1,67</td></tr>
<tr><td>1,50</td><td>1,60</td><td>1,65</td><td>1,55</td><td>1,55</td><td>1,64</td></tr>
<tr><td>1,47</td><td>1,70</td><td>1,73</td><td>1,59</td><td>1,62</td><td>1,56</td></tr>
<tr><td>1,48</td><td>1,48</td><td>1,58</td><td>1,55</td><td>1,49</td><td>1,52</td></tr>
<tr><td>1,52</td><td>1,50</td><td>1,60</td><td>1,50</td><td>1,63</td><td>1,71</td></tr>
</table>
a) Dresser le tableau de distribution des effectifs – fréquences d’après modèle suivant :
<table>
<tr><td>Taille xi (m)</td><td>Effectif</td><td>Fréquence</td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td>Total</td><td></td><td></td></tr>
</table>
b) Effectuer le tableau de distribution des effectifs - fréquences en regroupant des données en classes suivantes : [1,45 ; 1,55) ; [1,55 ; 1,65) ; [1,65 ; 1,75)</td>
</tr>
<tr>
<td></td>
<td>2. Diagramme
Diagramme en colonnes des effectifs, des fréquences.
La courbe des effectifs, des fréquences
Diagramme circulaire</td>
<td>Connaissances :
- Savoir le diagramme en colonnes des effectifs, des fréquences ; la courbe des effectifs ; le diagramme circulaire.
Savoir-faire :
- Lire des diagrammes en colonnes, diagrammes circulaires.
- Dresser des diagrammes en colonnes, diagrammes circulaires.
- Dresser des courbes polygonales des effectifs ; des fréquences.</td>
<td>Ex : Dresser le diagramme en colonnes des effectifs, des fréquences ; la courbe des effectifs correspondant au exemple précédent.
Ex : Soit le tableau de distribution des fréquences de données regroupées en classe suivant : La température moyenne en décembre à la ville de Vinh de 1961 à 1990.
<table>
<tr><td>Classes de température X (°C)</td><td>Valeur représentant x_i°</td><td>Fréquence f_i (%)</td></tr>
<tr><td>[15; 17)
[17; 19)
[19; 21)
[21; 23)</td><td>16
18
20
22</td><td>16,7
43,3
36,7
3,3</td></tr>
<tr><td>Total</td><td></td><td>100%</td></tr>
</table>
Décrire le tableau précédent en dressant :
a) Le diagramme en colonnes des effectifs
b) La courbe des effectifs.</td>
</tr>
<tr>
<td></td>
<td>3. Moyenne. Médiane, mode</td>
<td>Connaissances :</td>
<td>Ex : Les notes de mathématiques du deuxième semestre d’un groupe</td>
</tr>
</table>

		- Savoir certaines caractéristiques d'une série statistique : la moyenne, la médiane, la mode et leur signification. Savoir-faire : - Savoir calculer la moyenne, la médiane, la mode d'une série statistique (dans les situations étudiées)	d'élève de classe 10A sont listées suivantes : (avec convention qu'elles soient arrondies à 0,5) 2 ; 5 ; 7,5 ; 8 ; 5 ; 7 ; 6,5 ; 9 ; 4,5 ; 10 a) Calculer la note moyenne de ces 10 élèves (arrondie en dixième) b) Calculer la médiane de cette série.
	4. Variance et écart type d'une série statistique.	Connaissances: - Savoir la notion de variance, d'écart type d'une série statistique et leur signification. Savoir-faire : - Savoir calculer la variance, l'écart type d'une série statistique.	
11 (1er)	1. Combinatoire Règle d'adition et règle de multiplication. Arrangement. Permutation. Combinaison. Binôme de Newton.	Connaissances : - Savoir : Règle d'adition et règle de multiplication ; permutation, arrangement, combinaison k de n éléments ; formule de binôme de Newton $(a+b)^n$. Savoir-faire : - Savoir appliquer les règles d'addition et de multiplication. - Savoir calculer le nombre des permutations, des arrangements, des combinaisons - Savoir développer binôme de Newton d'une puissance donnée. - Trouver le coefficient de x^k dans le développement en polynôme de $(ax+b)^n$.	Ex : Un groupe de joueurs de ping-pong se compose 8 garçons et 7 filles. Y-a-il combien de propositions de : a) Un garçon simple, une fille simple. b) Couple garçon - fille. Ex : Soit des chiffres : 1, 2, 3, 4, 5. Y-a-il combien de naturels de 5 chiffres (ces 5 chiffres sont différents deux à deux) formés des chiffres donnés ?
	2. Probabilité Expérience aléatoire et événement. Probabilité d'un événement et	Connaissances : - Savoir : expérience aléatoire, échantillon, événement concernant à une expérience aléatoire. Définition de	Ex : Lancer un dé (identique) a) Décrire l'échantillon. b) Calculer la probabilité de l'événement « apparition d'un impaire de points sur la face de dé»

	propriétés élémentaires des la probabilité.	la probabilité d'un événement. - Savoir des propriétés : $P(\emptyset) = 0$; $P(\Omega) = 1$; $0 \leq P(A) \leq 1$ - Savoir (sans prouver) la théorème d'addition de probabilité et celle de multiplication de probabilité. Savoir-faire : - Savoir précise une expérience aléatoire, un échantillon, un événement concernant une expérience aléatoire. - Savoir utiliser la calculatrice pour calculer la probabilité.	Ex : Lancer deux dés. Calculer la probabilité de l'événement : « La somme des points sur les faces de deux dés est 8 » Ex : Y-a-il combien de façon à partager une classe de 40 élèves en groupe de 8 élèves ? Ex : a) Développer $(2x+1)^5$ en polynôme. b) Trouver le coefficient de x^3 dans ce polynôme. Ex : Pour tout $n \in N$, prouver que : $C_n^0 + C_n^1 + C_n^2 + ... + C_n^n = 2^n$
12 (Ter)			

Tableau 120. La synthèse de la statistique dans les programmes en vigueur au Vietnam à partir de l'an 2006

Au Vietnam, même si l'enseignement de la statistique est apparu plus tard qu'en France (il y a moins de dix ans) il vient dès les classes primaires. C'est à la classe 3 (équivalent à classe CE2 en France) on voit une rubrique s'appelant « Eléments de statistique ». Cette première approche de la statistique consiste essentiellement dans la lecture de tableaux de distributions des effectifs, diagrammes en rectangles et diagrammes circulaires.

Mais cet enseignement ne se fait pas en continu dans le programme secondaire ce que montre le tableau suivant

	Niveau de collège				Niveau de lycée		
Classe	6 (6è)	7 (5è)	8 (4è)	9 (3è)	10 (2de)	11 (1er)	12 (Ter)
Programme en vigueur		Statistique			Statistique	Combinatoire Probabilité	
Programme précédant	Statistique			Statistique			Combinatoire

Tableau 121. L'apparition de statistique et probabilité dans les classes secondaires au Vietnam

Au collège : la statistique ne se trouve qu'en classe 7 (5è) parmi les quatre classes du collège (10 séances sur 140 séances). Les données recueillies proviennent d'enquêtes au près de populations réelles supposées exhaustives et jamais d'expériences aléatoires. Ces populations sont généralement peu nombreuses et les valeurs prises par la variable statistique sont elles aussi peu nombreuses, et on n'effectue pas de regroupement de ces valeurs en classes. Les notions comme effectif, tableau des effectifs, moyenne, mode sont enseignées. La présentation des données statistiques est demandée sous la forme soit d'un tableau de distribution des effectifs soit sous celle d'un diagramme (diagramme en segment ou diagramme en rectangle ou diagramme circulaire). Par contre, la notion de fréquence n'est pas officiellement enseignée.

Au lycée : la statistique occupe 7 séances sur 105 et son enseignement est découpé un quatre leçons intitulées : tableau de distribution des effectifs et des fréquences ; diagramme ; moyenne, médiane, mode ; variance et écart type. Les nouveautés de classe 10 par rapport à la classe 7 sont les suivantes :

- Le regroupement par classe : c'est la transformation de données observées en données regroupées par classes même si ce regroupement est entièrement à la charge de l'enseignant. Ceci conduit à la fabrication de diagrammes en colonnes (histogrammes), calculs des résumés.

- La fréquence : pour normaliser les données i.e. ramener les populations à une même « taille » 100%

- La médiane (pour les données observées), la variance, l'écart type.

- Des résumés appelés « représentants » dans le programme.

Après la classe 10, les programmes de mathématiques ne comportent plus de chapitres de statistique. En classe 11, on trouve la combinatoire et les probabilités avec des notions d'expérience aléatoire, d'univers des possibles, et avec les théorèmes d'adition et de multiplication des probabilités.

	Collège	Lycée	
	7è	10è	11è
Statistique	Moyenne Diagramme en bâtons/rectangles Effectif - Tableau des effectifs Mode	Moyenne Diagramme en colonnes (histogramme) Mode	
		Médiane Fréquence - Tableau des fréquences Courbe polygonale des fréquences/effectifs Variance et écart type Regroupement en classes	
Probabilités			Expérience aléatoire Echantillon Evénement Probabilité d'un événement

Tableau 122. Les notions de statistique-probabilité enseignées au Vietnam

II.2. Analyse du manuel

Nous dégageons les types de tâches en utilisant la même catégorisation que pour les manuels français.

II.2. 1. Manuel mathématiques de classe 7 (volume 2, 2003, Maison d'édition Education et Formation, p. 3-23)

Chapitre 3 : **Statistique**

1§. Recueil de données statistique, effectif
2§. Tableau des « effectifs » des valeurs des caractères
3§. Diagramme
4§. Moyenne

Groupe	T	Types de tâches
0	T1	Simuler une expérience aléatoire n fois / Etablir une série statistique
	T4	Recueillir des données dans la vie réelle puis les écrire en tableau
1	T5	Lire des informations d'un tableau
	T6	Lire des informations à partir d'une représentation graphique
	T7	Représenter en graphique des données statistiques
	T8	Représenter/Regrouper des données statistiques en tableau
	T13	Décrire la statistique (population, caractère étudié, type de caractère)
	T22	Chercher la valeur la plus grande/petite d'une série statistique
2	T36	Calculer la moyenne
	T37	Estimer la moyenne des données regroupées
	T46	Calculer la taille de données statistiques
	T48	Calculer la mode

4	T76	En observant le tableau de données, répondre si la moyenne pourrait être utilisée comme un représentant pour le caractère étudié

Tableau 123. Tableau des types de tâches du manuel de classe 7, volume 2, 2003

Toutes les données sont tirées d'enquêtes sociétales (c'est-à-dire qu'il n'y a pas de simulation) : le nombre des plantes plantées, la population des villes, les notes des élèves d'une classe, le nombre des élèves absents dans un école pendant une journée, le nombre d'enfants des familles, la durée du temps de déplacement entre la maison et l'école d'un élève, la durée du temps à courir un parcours de 100m des élèves d'une classe, les poids des boites de thé, le nombre d'élève d'une classe ayant le même mois de la date de naissance, le nombre d'années de travail des ouvriers dans une entreprise, les points gagnés notés d'un tireur d'arme à feu, la durée du temps à résoudre un problème mathématique des élèves d'une classe, … Ainsi, nous ne trouvons aucune situation où les données sont recueillies par des expériences aléatoires répétées. Le mot hasard n'est jamais prononcé.

Le manuel expose la représentation graphique des données statistiques sous deux formes : le diagramme en segments (en bâtons) et le diagramme en rectangles (en barres). Après avoir présenté le diagramme en segment, il fait la remarque suivante : « *A côté des diagrammes en segment se trouvant dans les documents statistiques ou dans les livres, les journaux, on peut trouver un diagramme comme celui dans la figure 2 (les segments sont remplacés par les rectangles, et parfois, ces rectangles sont tracés de façon que l'un est contre l'autre pour faciliter l'observation et la comparaison), c'est le diagramme en rectangles* ». Cette remarque ne s'accompagne pas d'une distinction usuelle entre données à caractère qualitatif et données à caractère quantitatif ni d'une distinction, elle aussi très fréquente entre données quantitatives discrètes et données quantitatives continues. Ainsi, le diagramme en rectangles peut être formé soit par des rectangles isolés soit par des rectangles jointifs, sans lien avec la notion de données discrètes ou continues regroupées en classe. Autrement dit, il n'y a pas de notion d'histogramme dans ce manuel.

Dans une partie facultative (p. 15-16, après la leçon « 3§. Diagramme » et avant la leçon « 4§. La moyenne »), on trouve la notion de fréquence avec la formule de calcul $f = n/N$ (où n est l'effectif d'une valeur, N est l'effectif total, et f est la fréquence de cette valeur). La fréquence est présentée sous la forme fractionnaire et puis en pourcentage. Cette notion donne l'occasion de présenter le diagramme circulaire ayant des angles proportionnels aux fréquences mais elle n'est pas développée pour d'autres usages.

A propos de la moyenne, le manuel présente sa signification comme suit :

> La moyenne est utilisée pratiquement comme « le représentant » pour des caractères étudiés, particulièrement au cas où l'on veut comparer des caractères de même genre.
>
> (Manuel de classe 7, volume 2, p. 19)

Puis il insiste sur les limites de son usage par cette remarque :

> Lorsque les valeurs du caractère étudié ont de très grands écarts l'un de l'autre, il ne faudrait pas prendre la moyenne comme leur représentant.
>
> (Manuel de classe 7, volume 2, p. 19)

Il prépare ainsi l'idée de disposer d'autres indicateurs de position et présente le mode comme un représentant possible de la série statistique. La médiane sera présentée en classe 10.

La moyenne est le premier indicateur que le manuel aborde au service d'une comparaison entre séries statistiques mais aucun exercice exigeant une comparaison ne se trouve dans ce manuel.

Nous trouvons en fin de leçon une préparation du passage des variables discrètes aux variables continues à travers un tableau de distribution des effectifs regroupés en classe se trouvant dans l'exercice 18.

On mesure la taille de 100 élèves de classe 6 (unité cm) et obtient les résultats dans le tableau 26 suivant :

La taille (ordonnée croissant en intervalle)	Effectif (n)
105	1
110-120	7
121-131	35
132-142	45
143-153	11
155	1
	N = 100

a) Quelle est la différence entre ce tableau et les autres tableaux des effectifs connus au préalable ?
b) Estimez la moyenne dans ce cas.

(Manuel de classe 7, volume 2, p. 21)

Puis, le manuel fournit une technique de calcul de la moyenne :

- Calculer la moyenne de chaque intervalle. Ce nombre est juste la moyenne de la valeur la plus grande et de valeur la plus petite de l'intervalle. Par exemple : la moyenne de l'intervalle 110-120 est 115.

- Multipliez les moyennes obtenues avec leurs effectifs correspondant.

- Continuez les étapes d'après les règles étudiées.

(Manuel de classe 7, volume 2, p. 21)

Non seulement le manuel ne fournit aucune justification sur la validité de la technique mais il utilise ici un étrange mélange de valeurs isolées (classe d'une valeur unique à largeur nulle) et de classes dans ce tableau de regroupement. Quel diagramme choisir pour représenter le tableau de distribution ?

II.2. 2. Manuels mathématiques de classe 10 (2006, Maison d'édition Education et Formation)

Au niveau de classe 10, il existe deux ensembles de manuels de même programme, l'un s'appelle l'ensemble élémentaire et l'autre est l'ensemble avancé. L'élève utilise l'ensemble correspondant à son choix de section (sciences naturelles, sciences sociales, langue étrangère vivante, etc. ...). Généralement, ce qui choisit la section de science naturelle, il utilise le

manuel avancé, ce qui choisit la section de science sociale, il utilise le manuel élémentaire. Chaque ensemble se compose de trois livres :

- un livre de cours (manuel) pour les élèves : dans ce livre, chaque leçon est suivie par un nombre assez suffisant des exercices pour que les élèves s'entrainent et à la fin de chaque chapitre il y a un ensemble des exercices de révision.

- un livre d'exercices corrigés (qui ne sont pas lesquels dans le livre des cours) pour les élèves : dans ce livre, pour chaque leçon, il y a toujours un résumé du cours, puis un exercice exemplaire avec sa solution, ensuite des exercices supplémentaires dont les solutions sont donnés à la fin de chaque chapitre.

- un livre d'enseignants : ce livre présente la structure de chaque chapitre avec sa distribution d'emploi du temps. On y trouve des objectifs, des compétences et le contenu détaillé de chaque leçon et les corrigés des exercices correspondant du livre des cours, … Particulièrement, il y a une partie fournissant des connaissances supplémentaires sur une nouveauté aux enseignants à la fin du chapitre.

Dans le domaine des statistiques, ces manuels ne diffèrent pas sensiblement ni sur l'ensemble des notions présentées ni sur l'ensemble des types de tâches. On retrouve dans les deux manuels une structure binaire : la représentation par tableaux et diagrammes, les indicateurs de position et de dispersion. Sauf dans des cas que nous expliciterons « le manuel » signifie l'un et l'autre des deux manuels. Nous utiliserons aussi les livres des exercices pour accéder au rapport institutionnel et les livres des enseignants dans notre analyse pour accéder aux intentions de la noosphère.

Groupe	T	Types de tâches
0	T1	Simuler une expérience aléatoire n fois / Etablir une série statistique
1	T7	Représenter en graphique des données statistiques
	T8	Représenter/Regrouper des données statistiques en tableau
	T14	Effectuer un regroupement des données
	T19	Dresser le tableau des fréquences (à partir des graphiques)
	T21	Établir la distribution des fréquences (à partir des données)
2	T32	Calculer le pourcentage
	T36	Calculer la moyenne
	T47	Calculer la médiane
	T48	Calculer la mode
	T50	Calculer la variance
	T51	Calculer l'écart type
3	T62	Comparer les moyennes des séries statistiques
	T64	Commenter des séries statistiques
	T65	Comparer la répartition de deux séries statistiques
	T66	Comparer la distribution des fréquences des échantillons différents
	T67	Comparer les indicateurs statistiques des échantillons différents
	T69	Comparer deux diagrammes des données statistiques différentes
4	T73	Interpréter des indicateurs statistiques
	T74	Interpréter les résultats statistiques
	T75	Donner la pensé à la lecture des résultats de fréquence et d'effectif

	T77	Présenter des remarques sur le caractère étudié à partir du tableau de distribution des fréquences de données regroupées en classe
	T78	Présenter des remarques sur le caractère étudié à partir de la représentation graphique des fréquences de données regroupées en classe

Tableau 124. Les types de tâches du manuel de classe 10, l'an 2006

Après une disparition de la statistique en classes 8 et 9, le manuel de classe 10 fait un bref résumé des notions élémentaires de la statistique abordées en classe 7 comme données statistiques, individu, caractère, valeur du caractère, recueil de données. Le tableau des distributions des effectifs comporte souvent maintenant une colonne fréquences en pourcentage. Et le manuel donne le nom « tableau de distribution des effectifs et des fréquences » à ce tableau.

La notion d'*échantillon* est présentée pour la première fois comme une partie de population sans aucune mention à un tirage du hasard :

> Un sous ensemble défini des individus enquêté est appelé **échantillon**. Le nombre des individus est appelé **la taille d'échantillon**. Les valeurs des caractères recueillis sur un échantillon est appelé **échantillon des données**.
>
> (Manuel de classe 10, l'ensemble avancé, p. 160)

Tandis que le terme de *population* est dite sous-entendu comme suit :

> Si l'on fait une enquête sur tous les individus, dans ce cas, c'est une **enquête exhaustive**. Si l'on fait une enquête sur un échantillon, alors c'est une **enquête d'échantillon**.
>
> (Manuel de classe 10, l'ensemble avancé, p. 160)

Et aussi :

> […] l'enquête sur tous les individus est impossible, il faut donc faire une **enquête d'échantillon**. Une des missions les plus importantes de la statistique est de construire les méthodes qui nous permettent de tirer des conclusions, d'établir des prévisions sur **tous les individus enquêtables** en s'appuyant sur les informations acquis sur un échantillon […] Ensemble de tous les individus enquêtés s'appelle **l'ensemble principal** […]
>
> (Livre d'enseignants, l'ensemble avancé, p. 217)

→ Discours technologique pour aider les enseignants dans la justification des groupes de tâches 0 et 4

Ainsi, nous voyons que la population (ou l'ensemble principal) est l'ensemble de tous les individus enquêtables.

- Le terme d'échantillon ne fait pas référence à un extracteur au hasard.

A propos du regroupement de données qui est une nouveauté importante de la classe 10, le manuel utilise aussi bien des variables quantitatives discrètes que des variables continues[193]. Même si les variables sont continues (la taille, le poids, la durée du temps, …), les données utilisées sont généralement discrètes. La justification avancée pour regrouper les valeurs en classes est la réduction du nombre de valeurs. Dans les exercices proposés l'élève n'a pas la charge ni de la décision de regrouper en classes ni de la fabrication des classes.

Est-ce que le choix sur des données dans le manuel est pertinent pour faire un passage de variable statistique discrète à la variable statistique continue ? Prenons la représentation des données statistiques du manuel avancé. D'abord, pour représenter des données en tableau de taille en centimètre ce manuel donne un exemple sur des données brutes comme suit :

160	161	161	162	162	162	163	163	163	164	164	164	164
165	165	165	165	165	166	166	166	166	167	167	168	168
168	168	169	169	170	171	171	172	172	174.			

Remarquons que ces données prennent des valeurs entières et discrètes. Puis le manuel présente deux tableaux des distributions des fréquences regroupées en classes (Tableau 5 et Tableau 6)

Lớp	Tần số	Tần suất (%)
[160 ; 162]	6	16,7
[163 ; 165]	12	33,3
[166 ; 168]	10	...
[169 ; 171]	5	...
[172 ; 174]	3	...
	$N = 36$	

Bảng 5

Lớp	Tần số	Tần suất (%)
[159,5 ; 162,5)	6	16,7
[162,5 ; 165,5)	12	33,3
[165,5 ; 168,5)	10	27,8
[168,5 ; 171,5)	5	...
[171,5 ; 174,5)	3	...
	$N = 36$	

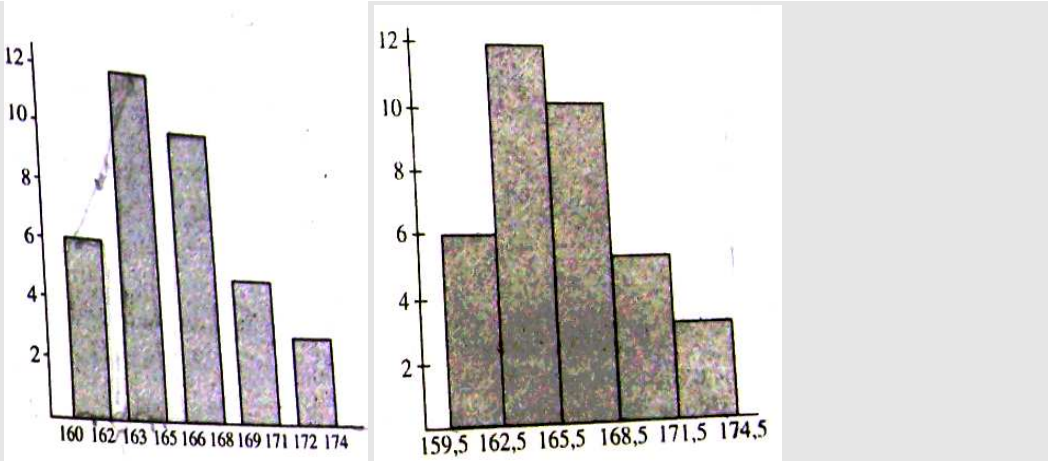
Bảng 6

(Manuel de classe 10, l'ensemble avancé, p. 164)

Dans le tableau 5, non seulement les classes sont des intervalles fermés et disjoints mais il y a aussi des « trous » entre ces classes (c'est à dire que l'on ne peut pas remplacer l'union de ces intervalles par un seul !). Par contre, dans le tableau 6, les classes sont des intervalles semi-fermés, disjoint sans trous entre des classes consécutifs. Même si les deux tableaux sont pour représenter le regroupement en classes, mais pourquoi ces choix ? Continuons avec les deux représentations graphiques du manuel :

[193] Le livre des enseignants présente trois types de variables :

- Variable quantitative
- Variable ordinale
- Variable qualitative



(Manuel de classe 10, l'ensemble avancé, p. 161)

On voit clairement que le diagramme en rectangles isolés (diagramme en bâtons) correspond au tableau 5 et le diagramme en rectangles jointifs (histogramme) correspond au tableau 6. Ainsi, le passage du tableau 5 au tableau 6 peut être celui de diagramme en bâton à l'histogramme.

Le diagramme appelé *diagramme en rectangles* dans le manuel de classe 7 disparaît et apparaît en classe 10 le *diagramme en colonnes (histogramme).* Il est utilisé pour décrire le tableau des distributions des effectifs ou des fréquences regroupés en classe. Apparaît ensuite, la courbe polygonale des fréquences présentée aussi comme un autre moyen pour décrire graphiquement le tableau des distributions des effectifs ou des fréquences regroupés en classe. A propos de la technique pour tracer la courbe polygonale des fréquences, le manuel propose de préciser d'abord des points de coordonnés (c_i ; f_i) où c_i est la moyenne de deux extrémités de l'intervalle nommé valeur représentant de la classe (la valeur centrale de la classe) ; et f_i est la fréquence de la classe puis de joindre ces points. Nous trouvons que dans toutes les questions concernant la demande de tracer la courbe polygonale des fréquences, on demande toujours de tracer à la fois le diagramme en colonnes puis la courbe polygonale des fréquences. Quel est le lien entre ces deux diagrammes que le manuel veut apporter aux élèves ?

Voici deux diagrammes de même données.

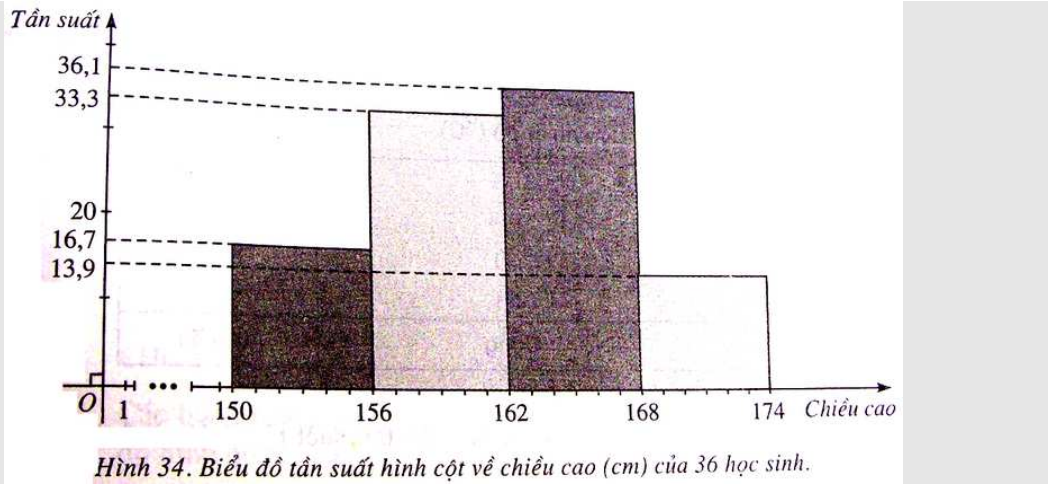


Diagramme en colonnes de la taille (cm) de 36 élèves

(Manuel de classe 10, l'ensemble élémentaire, p. 115)

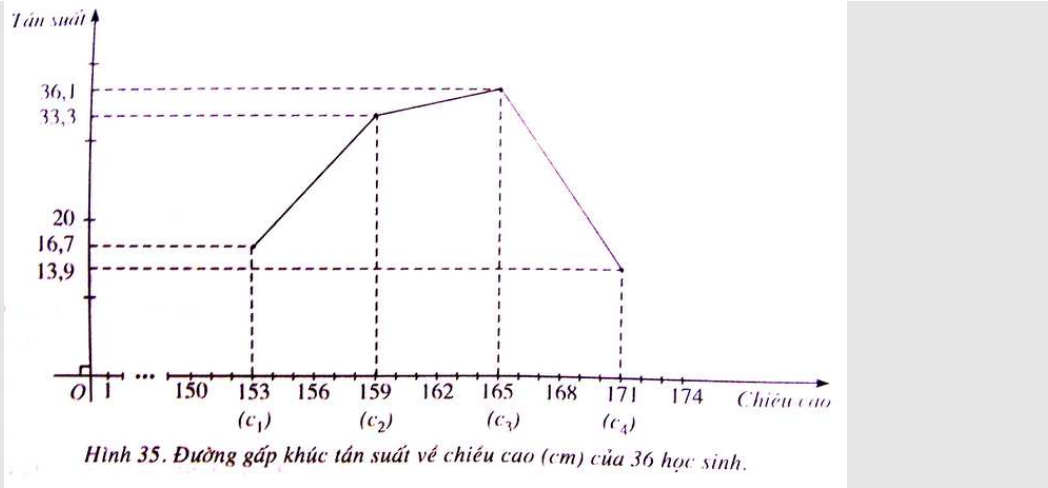


Figure 35. Courbe polygonale des fréquences de la taille (cm) de 36 élèves

(Manuel de classe 10, l'ensemble élémentaire, p. 116)

Regardons maintenant le livre réservé pour les enseignants, on voit qu'il y a un passage de diagramme en colonnes des fréquences regroupés en classe à la courbe polygonale des fréquences. En traçant les deux diagrammes sur le même repère, ce livre veut peut-être faire un lien entre ces deux diagrammes : le diagramme en colonnes est un intermédiaire du passage du tableau de distribution des fréquences regroupées en classes à la courbe polygonale des fréquences ?

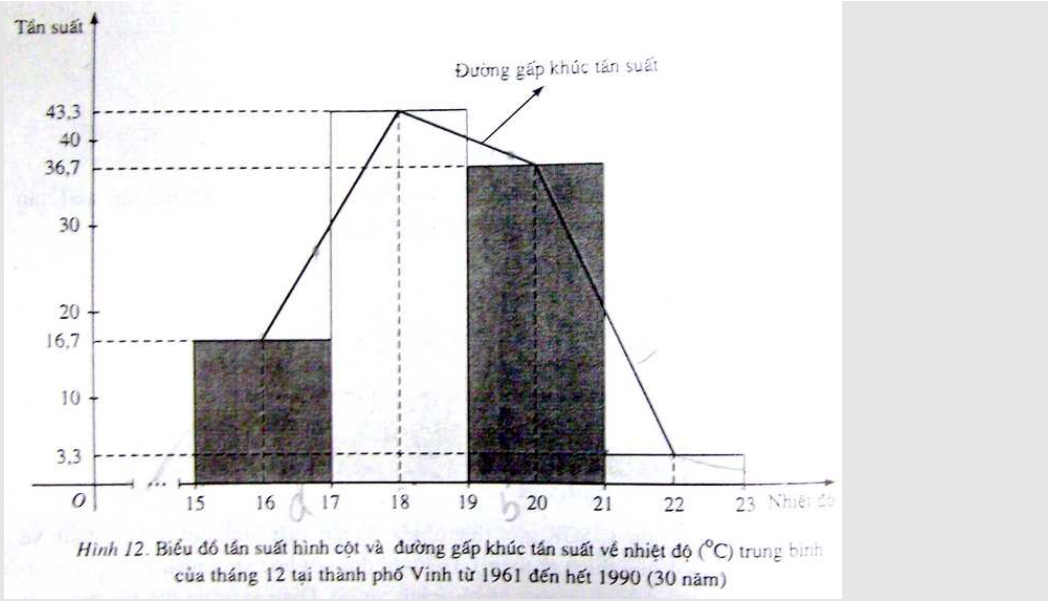


Figure 12. Diagramme en colonnes et courbe polygonale des fréquences de la température (°C) moyenne en décembre de la ville de Vinh entre 1961 et 1990 (30 ans)

(Livre d'enseignants de classe 10, l'ensemble élémentaire, p. 128)

Une remarque très forte se trouve ici, c'est que la courbe polygonale des fréquences ne conserve pas l'aire sous la courbe même si elle joint des milieux des classes, elle assure la forme de la densité des fréquences,…

On peut poser aussi une question : de la part de l'institution, quel est le rôle de la courbe polygonale des fréquences ? Dans un exercice du livre des exercices, nous trouvons deux exercices dans le manuel élémentaire qui nous permet de trouver un des éléments pour répondre à la question.

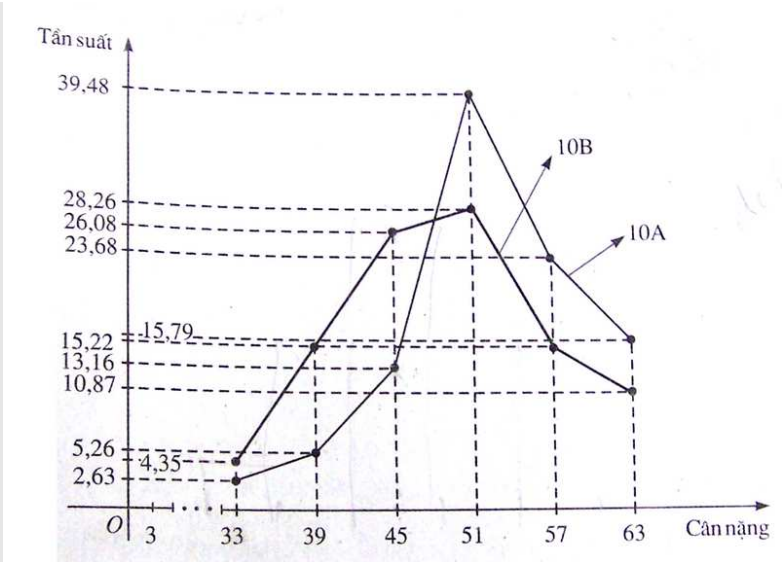
Exercice 19
Soit le tableau de distribution des effectifs regroupés en classes :
Le poids des élèves de classe 10A et 10B du lycée L.

Classe de poids	Effectif	
	10A	10B
[30 ; 36)	1	2
[36 ; 42)	2	7
[42 ; 48)	5	12
[48 ; 54)	15	13
[54 ; 60)	9	7
[60 ; 66)	6	5
Total	38	46

a) Etablir le tableau de distribution des effectifs regroupés en classes avec des classes données dans le tableau précédant.
b) Tracer dans le même plan de repère les courbes polygonales des effectifs regroupés en classes du poids des élèves de deux classes 10A et 10B. Puis comparer les poids des élèves de classe 10A et ceux des élèves de classe 10B.
c) Quel est le pourcentage des élèves de classe 10A pesant au moins de 42kg ? Même question pour les élèves de classe 10B.
d) Calculer la moyenne, l'écart type des données de classe 10A et 10B. Dans quelle classe que les élèves pèsent plus lourd ?

(Livre d'exercices corrigés de classe 10, l'ensemble élémentaire, p. 162)

Puis comme la solution de la question 19b), le livre donne la représentation graphiques :



Les courbes polygonales des fréquences regroupées en classes
du poids des élèves des classes 10A et 10B

(Livre d'exercices corrigés de classe 10, l'ensemble élémentaire, p. 173)

Puis il fait son explication :

> En observant deux courbes polygonales, on voit que : parmi les élèves pesant au plus 45kg, les élèves de classe 10A occupe un pourcentage plus grand que ceux de classe 10B. Et parmi les élèves pesant au moins de 51kg, les élèves de classe 10B occupe un pourcentage plus grand que ceux de classe 10A.

(Livre d'exercices corrigés de classe 10, l'ensemble élémentaire, p. 174)

La question se pose est comment fait-on cette observation ? On s'appuie sur l'aire sous la courbe, ou la hauteur de segments représentants le diagramme en segments, ou bien la position relative des parties des courbes polygonales ? Comme la notion d'aire sous la courbe n'existe pas dans le manuel,

Comparaison :

- A propose des types de tâche concernant la comparaison, il n'y a qu'un seul exercice (contre 21 exercices en total) qui demande de calculer les moyennes, les médianes, les variances et les écarts types de deux séries statistiques puis demander un commentaire.

> Exercice 15
> On a noté la vitesse (km/h) de 30 voitures passant chacun de deux parcours A et B :
>
> Parcours A :

60	65	70	68	62	75	80	83	82	69	73	75	85	72	67
88	90	85	72	63	75	76	85	84	70	61	60	65	73	76

> Parcours B :

76	64	58	82	72	70	68	75	63	67	74	70	79	80	73
75	71	68	72	73	79	80	63	62	71	70	74	69	60	63

> 1. Calculer la moyenne, la médiane, la variance et l'écart type de vitesse des voitures roulant dans chaque parcours.
> 2. D'après vous, dans quel parcours, les voitures roulent avec la plus sécurité ?

(Manuel de classe 10, l'ensemble avancé, p. 179)

La moyenne est présentée avec deux techniques de calcul et le choix de la technique dépend de la présentation des données : soit le tableau des données brutes (ensemble des valeurs observées pour tous les individus) ou tableau de distribution des données regroupées en classes (page ….). Pourtant, l'exemple l'illustrant ne donne aucun commentaire sur ces deux techniques de calcul car les résultats obtenus sont pareils en arrondissant en unité.

Pour présenter la notion de médiane, le manuel a dit :

> Lorsque les données statistiques ont de grande différence entre leurs valeurs statistiques, la moyenne ne peut pas être leur représentant. Dans ce cas, on choisit un autre indicateur caractéristique plus convenable, c'est la médiane ;

(Manuel de classe 10, l'ensemble élémentaire, p. 121)

Il y a trois exercices (contre 27 en total) concernant la comparaison des séries statistiques : un avec des données discrètes, deux avec des données regroupées en classe. Mais nous

trouvons que la comparaison des séries statistiques est sous-entendu dans ces exercices car on le consigne de ces exercices ne demande pas directement une comparaison, mais sous la forme d'une demande de remarquer sur deux données statistiques fournies.

Voici l'extrait de ces trois exercices concernant la comparaison de données dans le manuel élémentaire :

Exercice 2
Les deux classe 10C et 10D d'un lycée ont passé un contrôle de littérature de même sujet. Les résultats sont présentés dans deux tableaux de distribution des effectifs suivants :

Notes de classe 10C :

Note	5	6	7	8	9	10	Total
Effectifs	3	7	12	14	3	1	40

Notes de classe 10D :

Note	6	7	8	9	Total
Effectifs	8	18	10	4	40

a) Calculer les moyennes, les variances, les écarts types des tableaux de distribution des effectifs donnés.
b) Quelle est la classe ayant des résultats les plus réguliers ?

(Manuel de classe 10, l'ensemble élémentaire, p. 128)

Exercice 3
Soient deux tableaux de distribution des effectifs

Les poids des poissons du premier groupe :

Poids (en kg)	[0,6 ; 0,8)	[0,8 ; 1,0)	[1,0 ; 1,2)	[1,2 ; 1,4]	Total
Effectifs	4	6	6	4	20

Les poids des poissons du deuxième groupe :

Poids (en kg)	[0,5 ; 0,7)	[0,7 ; 0,9)	[0,9 ; 1,1)	[1,1 ; 1,3)	[1,3 ; 1,5]	Total
Effectifs	3	4	6	4	3	20

a) Calculer les moyennes des tableaux de distribution des effectifs regroupés en classe.
b) Calculer les variances des tableaux de distribution des effectifs regroupés en classe.
c) Quel est le groupe des poissons ayant des poids les plus réguliers ?

(Manuel de classe 10, l'ensemble élémentaire, p. 128)

Exercice 4

Soient deux séries statistiques suivantes :

Le poids des poissons du premier groupe (en gramme) :

645	650	645	644	650	635	650	654
650	650	650	643	650	630	647	650
645	650	645	642	652	635	647	652

Le poids des poissons du deuxième groupe (en gramme) :

640	650	645	650	643	645	650	650	642
640	650	645	650	641	650	650	649	645
640	645	650	650	644	650	650	645	640

a) Etablir le tableau de distribution des effectifs et des fréquences de données du premier groupe de poisson en regroupant en classes suivantes :
[630 ; 635), [635; 640), [640 ; 645), [645 ; 650), [650 ; 655]

b) Etablir le tableau de distribution des effectifs et des fréquences de données du deuxième groupe de poisson en regroupant en classes suivantes :
[638 ; 642), [642; 646), [646 ; 650), [650 ; 654]

c) Décrire le tableau de distribution des fréquences regroupé en classe de la question a) en dressant le diagramme en bâton et la courbe représentative.

d) Décrire le tableau de distribution des effectifs regroupé en classe de la question b) en dressant le diagramme en bâton et la courbe représentative.

e) Calculer la moyenne, la variance, l'écart type des tableaux des distributions des effectifs et des fréquences effectués. Quel est le groupe de poisson ayant des poids plus réguliers ?

(Manuel de classe 10, l'ensemble élémentaire, p. 129)

Dans le dernier de ces trois exercices, nous voyons que le nombre de classes, la taille des classes proposés dans deux séries statistiques sont différentes : 5 groupes avec taille de 5 pour la première série statistique, et 4 groupes avec taille de 4 pour la deuxième série statistique.

- Lorsque les données exigeant de regrouper en classe, c'est le manuel qui précise des classes et l'élève donc n'a pas de responsabilité de ce fait. En plus, toutes les classes du regroupement de données ont la même taille, c'est-à-dire elles sont régulières. Et enfin, la suite des classes d'une répartition est celle des intervalles demi-fermées sauf le dernier intervalle est une fermée.

- Enfin du chapitre, le manuel propose un seul travail pratique en groupe de 3 à 5 élèves en proposant de choir une classe du lycée pour ce travail. C'est une synthèse des travaux sur la statistique.

1. Faire un sondage et recueillir des données statistiques dans la classe choisie d'après un caractère choisi par chaque groupe (par exemple le nombre de frères de sœurs dans la famille, la durée du temps consacré pour l'étude de mathématiques à la maison de chaque élève, la taille des élèves de la classe, les notes de mathématiques de chaque élève d'un contrôle précédent,…).

2. Présenter, analyser, traiter des données statistiques recueillies.

3. Retirer une conclusion puis proposer des propositions.

(Manuel de classe 10, l'ensemble élémentaire, p. 131)

II.2. 3. Manuel mathématiques de classe 11 (2007, Maison d'édition Education et Formation)

Comme le cas de classe 10, il existe aussi deux ensembles des manuels pour la classe 11.

Dans le programme de classe 11, le chapitre « Combinatoire et probabilité » est consacré pour l'étude des probabilités. Les deux manuels de deux ensembles comprennent trois premières leçons communes :

1. Principes de comptage
2. Permutation, arrangement, combinaison
3. Binôme Newton

Puis les deux manuels se diffèrent par la partie de probabilité :

Manuel de l'ensemble élémentaire :

4. Expérience aléatoire et événement
5. Probabilité d'un événement

Manuel de l'ensemble avancé :

4. Evénement et expérience aléatoire
5. Règles des calculs des probabilités
6. Variable aléatoire

Dans le tableau suivant, nous présentons les définitions de notions principales avec son apparition dans les deux manuels :

Notions	L'ensemble avancé	L'ensemble élémentaire
Expérience aléatoire	Une expérience aléatoire est une expérience dont les résultats ne sont pas prévisibles même si on connait l'ensemble de tous les résultats attendus de l'expérience.[194]	Une expérience aléatoire (ou expérience en abréviation) est une expérience ou une action : - dont les résultats ne sont pas prévisibles; - où l'on peut déterminer l'ensemble de tous les résultats possibles de cette expérience.[195]
Définition laplacienne de la notion de probabilité	Soit A un événement concernant une expérience ayant un nombre fini des résultats équiprobables. On appelle le	Supposons que l'expérience T a un ensemble des issues Ω qui est un ensemble fini et les issues de T sont équiprobables. Si A est un événement concernant l'expérience T

[194] Manuel de classe 11, l'ensemble élémentaire, p. 59.

[195] Manuel de classe 11, l'ensemble avancé, p. 70.

[196] Manuel de classe 11, l'ensemble élémentaire, p. 66.

	quotient $\frac{n(A)}{n(\Omega)}$ la probabilité de l'événement A, noté P(A). $$P(A) = \frac{n(A)}{n(\Omega)}$$ n(A) est le nombre des éléments de l'ensemble A ou le nombre des résultats favorables pour l'événement A, n(Ω) est le nombre des résultats possibles de l'expérience.[196]	et si Ω_A est l'ensemble des issues décrivant A, la probabilité de A est un nombre dont le signe est P(A) est un nombre déterminé par la formule suivante: $$P(A) = \frac{\lvert \Omega_A \rvert}{\lvert \Omega \rvert}$$ où $\lvert \Omega_A \rvert$, $\lvert \Omega \rvert$ sont successivement le nombre des éléments des ensembles Ω_A et Ω.[197]
Définition fréquentiste de la notion de probabilité	Soit **n(A)**, le nombre d'apparition de l'événement A dans une suite de **n** expériences répétées. Le rapport $\frac{n(A)}{n}$ est appelé **fréquence d'apparition** de l'événement A. Lorsque n augmente, $\frac{n(A)}{n}$ se rapproche de plus en plus un nombre P(A) qui est déterminé. On appelle ce nombre la **probabilité de l'événement A selon le point de vue fréquentiste.** Dans le cas d'une expérience ayant un nombre fini de résultats équiprobables, le nombre P(A) dans cette définition est coïncidé avec le nombre P(A) dans la définition classique de la probabilité.[198]	Considérons l'expérience T et l'événement A en relation à cette expérience. Nous répétons N fois l'expérience T et notons les fois où se produit l'événement A. Le nombre d'apparitions de l'événement A est appelé l'effectif de A dans l'expérience T effectuée N fois. Le rapport entre l'effectif de A et le nombre n est appelé la fréquence de A dans l'expérience T effectuée N fois. On a pu démontrer que plus le nombre N d'expériences est grand plus la fréquence de A se rapproche d'un nombre déterminé, appelé la probabilité de A dans le sens fréquentiste (c'est aussi le nombre P(A) dans la définition classique de la probabilité. [199]
Probabilité expérimentale		Une compagnie d'assurances-vie a établi les statistiques suivantes : parmi 100 000 hommes âgés de 50 ans il y a 568 hommes qui décèdent avant qu'ils passent à l'âge de 51 ans, et parmi 100 000 femmes de 50 ans, 284 meurent quand elles ont 51 ans. La probabilité expérimentale[200] pour qu'un homme de 50 ans décède avant de passer à l'âge de 51 ans est

[197] Manuel de classe 11, l'ensemble avancé, p. 72.

[198] Manuel de classe 11, l'ensemble élémentaire, p. 76.

[199] Manuel de classe 11, l'ensemble avancé, p. 74.

[200] Soulignés par nous.

		$\frac{568}{100\ 000} = 0{,}00568$ et la probabilité expérimentale pour qu'une femme de 50 ans décède avant de passer à l'âge de 51 ans est $\frac{284}{100\ 000} = 0{,}00284$[201]
Variable aléatoire		Une grandeur X est appelée une variable aléatoire discrète si elle prend des valeurs en nombres qui appartiennent à un ensemble fini et ces valeurs sont au hasard et on ne peut pas les deviner à l'avance.
Lien entre les statistiques et les probabilités		- La fréquence f_i est une valeur approchée de la probabilité p_i. - Le tableau de distribution des fréquences d'un échantillon nous donne donc une « image » approchée du tableau de distribution des probabilités de X. - La moyenne de l'échantillon est une valeur approchée de l'espérance de X. - la variance et l'écart-type de l'échantillon sont les valeurs approchées de la variance et l'écart-type de X.[202]

Tableau 125. Les définitions des notions avec leurs apparitions dans les deux manuels de classe 11

(La case grisée est pour une notion présentée dans une lecture complémentaire du manuel.)

II.3. Conclusion

- La statistique est présentée discrètement dans le programme de mathématiques au Vietnam avec une grande discontinuité : elle ne se trouve que dans classe 7 (5è) au collège et puis en classe 10 (seconde) au lycée. Et l'année prochaine, la probabilité sera engagée en classe 11 (Première). En comparaison avec le programme de mathématiques dans le passé, la statistique a été enseignée en classe 6 (6è) et classe 9 (3è) au collège et rien au lycée ; ensuit, la combinatoire (considérée comme une préparation pour la probabilité) a été enseignée en classe 12 (terminal). Cela montre une évolution de statistique-probabilité dans le programme vietnamien. Autrement dit, on pourrait voir qu'il y a un projet de faire évoluer la statistique de la part de la noosphère vietnamienne même s'il est encore faible et s'arrête au niveau de la statistique descriptive sans reprise dans la probabilité.

---

[201] Manuel de classe 11, l'ensemble avancé, p. 75.

[202] Manuel de classe 11, l'ensemble avancé, p. 91-92.

- Une détaille remarquable est que dans le passé, la statistique et la combinatoire sont enseignées en classe à la fin de collège (classe 9) et à la fin du lycée (classe 12). Au Vietnam, il y a une contrainte que les savoirs enseignés à la fin d'un niveau doivent probablement apparaître dans les épreuves de brevet ou de bac. Donc, on peut considérer que la noosphère vietnamienne veut éviter l'existence de sujets de statistique et de probabilité dans ces épreuves. Et est-ce que le concours influence évidement dans l'enseignement des enseignants vietnamiens ?

- Il existe la comparaison des données même sous-entendu et sur les critères de moyenne, variance et l'écart type avec très peu d'exercice et deux types de tâches T62 et T67 du groupe 3.

III. Analyse des sujets de Brevet et de Baccalauréat français

Pour notre analyse, nous choisissions les ouvrages l'Annabrevet pour le niveau de collège et l'Annabac pour le niveau de lycée.

En France, ces ouvrages sont des très importantes bases de documents car l'enseignant et l'élève de classe fin de collège et fin de lycée le référencient pour préparer l'enseignement-apprentissage et surtout la révision pour le brevet ou le bac national à la fin de collège ou de lycée.

III.1. Brevet :

III.1.1. Avant la réforme :

Un petit rappel :

Les notions enseignées pendant cette période sont : Effectif, Effectif cumulé, Fréquence, Fréquence cumulée, Moyenne.

a) Analyse des sujets d'épreuve :

Il y a 17 sujets de brevet en Juin 1995 dans l'ouvrage Annabrevet 1996. Chaque sujet de brevet se compose 3 parties[203] :

– Travaux numérique et gestion des données : cette partie comprend de 2 à 5 exercices.

– Travaux géométriques : cette partie comprend de 2 à 3 exercices.

– Problème.

En total, ces 17 sujets[204] comprennent 118 exercices dont 6 concernent la statistique (ça occupe environ 5,08%).

Dans la partie suivante, nous allons faire une analyse sur ces 6 exercices pour voir la place de la statistique enseignée au collège dans l'épreuve de brevet.

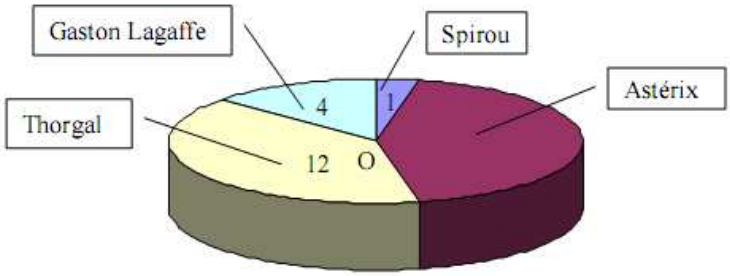
[203] Voir le détail du programme de brevet dans l'Annexe 3.

[204] Voir l'Annexe 1.

1.1. Sujet 2 : Sujet du brevet Amiens, Juin 1995

Exercice 4

Une enquête sur les bandes dessinées lues par les 30 élèves d'une classe a donné le diagramme suivant :



1 élève lit Spirou, 4 lisent Gaston Lagaffe....

1. Combien d'élèves lisent Astérix ?
2. Quel est le pourcentage d'élèves de cette classe lisant Thorgal ?
3. Calculer l'angle AOB correspondant, sur le diagramme, aux lecteurs de Gaston Lagaffe.

(Sujet 2, Annabrevet 1996, p. 15-16)

Analyse :

Les données statistiques sont fournies en forme d'un diagramme demi-circulaire qui est une répartition des effectifs des bandes dessinées lues par 30 élèves (ce n'est pas un diagramme en pourcentage familial). Cet exercice est pour demander de calculer l'effectif total, le pourcentage et la mesure de l'angle précisé.

On peut voit que cet exercice n'est que pour des calculs et n'est pas pour la statistique.

1.2. Sujet 5 : Sujet du brevet Caen, Juin 1995

Exercice 3

En Suisse, il y a quatre groupes d'habitants qui parlent chacun une langue différente :

- 4 150 000 parlent allemand ;
- 1 200 000 parlent ;
- 600 000 parlent italien ;
- 50 000 parlent romanche.

On veut présenter cette situation par un diagramme circulaire.

1. Reproduire et compléter le tableau :

Suisses parlant	allemand	français	italien	romanche	Total
Effectif	4 150 000				
Pourcentage					100
Angle					

2. Construire un diagramme circulaire (prendre un cercle de rayon 5 cm).

(Sujet 5, Annabrevet 1996, p. 18-19)

Analyse :

Dans ce sujet, les données statistiques sont fournies en forme d'une liste des effectifs de quatre langues parlées en Suisse. Le sujet dit clairement que l'on veut « représenter cette situation par un diagramme circulaire » en posant un tableau des effectifs, des pourcentages, et des mesures des angles que le candidat doit remplir. La dernière question est la demande de tracer le diagramme circulaire.

On peut voir qu'il est un exercice de calcul numérique et de représentation graphique de données statistiques.

1.3. Sujet 9 : Sujet du brevet Lilles, Juin 1995

Exercice 3

Lors de fête de fin d'année, un grand magasin fait une enquête sur la nature du cadeau le plus souvent offert aux enfants.
Pour cela, les enquêteurs interrogent les clients à la sortie du magasin.
Les nombres de réponses obtenues ont été regroupés dans le tableau suivant

	Jouets	Livres	Autres
Enquêteur A	243	158	78
Enquêteur B	201	187	52
Enquêteur C	190	55	36

1. Combien de réponses l'enquêteur A a-t-il obtenues ?
2. On veut représenter l'ensemble des réponses obtenues par les trois enquêteurs à l'aide d'un diagramme ayant la forme d'un demi-disque de rayon 4cm.
Pour préparer ce diagramme, reproduire et compléter le tableau suivant où les mesures des angles seront arrondies au degré :

	Jouets	**Livres**	**Autres**	
Effectif	634			1200
Angle		60°		180°

Dessiner le diagramme.

(Sujet 9, Annabrevet 1996, p. 27-28)

Analyse :

Les données statistiques sont fournies par un tableau de deux entrées qui est le résultat d'un enquête de trois enquêteurs sur la nature du cadeau le plus souvent offert aux enfants avec une description brève de la façon à faire l'enquête. La demande de l'exercice est de calculer l'effectif total, de remplir un nouveau tableau de distribution des effectifs et de mesures des angles en catégorisant suivant la nature du cadeau, puis dessiner un diagramme. Même si le consigne du sujet ne précise pas le genre de diagramme, mais le candidat doit s'appuyer sur la dernière colonne du tableau où la mesure total des angles est 180° pour reconnaître qu'il doit être un diagramme demi-circulaire.

On peut voit que cet exercice est pour but de présenter le recueil des données, de représenter des résultats d'un enquête de manière statistique (en tableau et en diagramme).

1.4. Sujet 11 : Sujet du brevet Nantes, Juin 1995

Exercice 3
Voici un tableau (incomplet) indiquant la production de voitures particulières, en 1993, de trois constructeurs français :

Constructeur	**Renault**	**Peugeot**	**Citroën**	**Production total**
Effectif	1 264 628	946 988		2 836 280
Fréquence en %				

1. Recopier, en le complétant, le tableau ci-dessus.
2. Quel nombre de voitures aurait produit le constructeur Renault s'il avait eu la possibilité d'augmenter sa production de 3% ?

(Sujet 2, Annabrevet 1996, p. 31)

Analyse :

Les données sont présentées par un tableau de distribution des effectifs concernant la production de voitures de trois constructeurs français en 1993 (Renault, Peugeot, Citroën). Pour remplir ce tableau, le candidat doit calculer les fréquences mais les résultats des calculs doivent être en pourcentage d'après l'indication du sujet « Fréquence en % ». La deuxième question est une demande d'un calcul du nombre de voitures produit par le constructeur Renault avec l'hypothèse qu'il a une possibilité d'augmenter sa production de 3%.

Quoique le mot « fréquence » est apparu dans le sujet, mais on peut voir clairement que cet exercice est totalement pour la raison des calculs pourcentage.

1.5. Sujet 12 : Sujet du brevet Orléans-Tours, Juin 1995

Exercice 6
Le tableau suivant représente la répartition des notes obtenues par les élèves d'une classe lors d'un contrôle.

Note n	$0 \leq n < 5$	$5 \leq n < 10$	$10 \leq n < 15$	$15 \leq n < 20$
Effectif	2	8	11	7

1. Représenter sur la copie cette répartition par un diagramme en barres.
On prendra :
- horizontalement : 2 cm pour 5 points ;
- verticalement : 0,5 cm pour 1 élève.
2. Calculer le pourcentage des élèves de la classe qui ont une note supérieure ou égale à 10, arrondir à 0,1 % près.

(Sujet 12, Annabrevet 1996, p. 32)

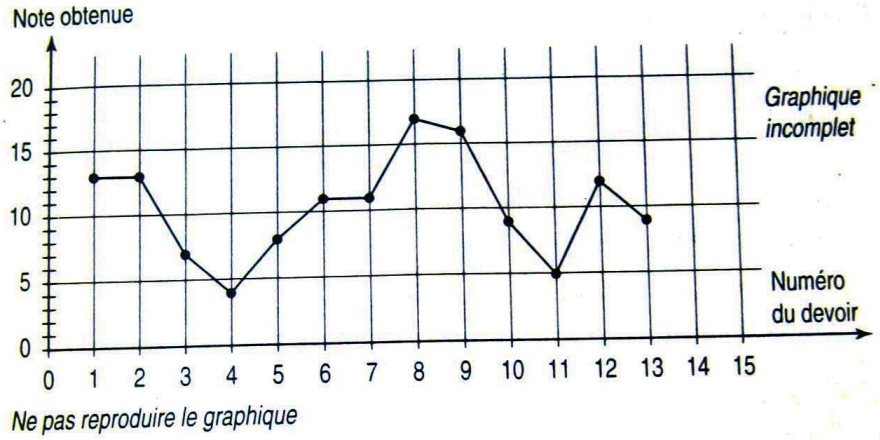
Analyse :

Dans ce sujet, on donne un tableau de répartition des notes obtenues par les élèves à un contrôle en regroupement en 4 classes des notes. Le sujet demande de représenter cette répartition par un diagramme en barres avec la précision des mesures horizontale et verticale du rectangle représentant les données. Un pourcentage demandé de calculer est effectivement une fréquence cumulée.

Ainsi, on peut voir que cet exercice est pour la raison de calcul même s'il y a une demande de représenter les données en diagramme.

1.6. Sujet 15 : Sujet du brevet Rouen, Juin 1995

EXERCICE 3



Un élève a reporté sur le graphique ci-dessus les notes de ses devoirs. Il a oublié d'y inscrire ses deux dernières notes : 12 et 16.

Soit n la note obtenue à un devoir.

1. Reproduire et compléter le tableau suivant :

Note obtenue	$0 \leqslant n \leqslant 5$	$5 < n \leqslant 10$	$10 < n \leqslant 15$	$15 < n \leqslant 20$	Nombre total de devoirs
Nombre de devoirs			6		15

2. Calculer le pourcentage de devoirs ayant obtenu la note n, telle que $10 < n \leqslant 15$

(Sujet 15, Annabrevet 1996, p. 38-39)

Analyse :

Le sujet donne un graphique « inhabituel » de deux axes : l'axe abscisse est pour *Numéro de devoir* (15 devoirs en total que le candidat peut lire soit sur le graphique soit dans le tableau au-dessous de ce graphique) et l'axe ordonné est pour *Note obtenue* d'un élève. Il y un manque des informations dans ce graphique. Ce sont les notes 12 et 16 qui ne sont pas marquées sur le graphique, mais on ne sait pas que ces deux notes correspondent aux quels devoirs qui restent (le numéro14 et le 15). Le candidat doit compléter le tableau de répartition des « Nombre de devoirs » (des effectifs des notes regroupées en classe) en comptant des « points » marquant les notes des devoirs sur le graphique et en faisant des calculs.

A travers de ce sujet, on peut voir qu'il est pour la lecture graphique faisant partie de Travaux numériques et n'est pas pour la statistique. Nous le disons car dans ce sujet, nous voyons que le sujet évite d'utiliser le mot « effectif » pris habituellement dans tous les autres sujets restant lors d'avoir un tableau de distribution des effectifs. En plus, nous n'y trouvons aucun mot concernant la statistique, il n'y a que des mots comme : graphique, nombre de devoirs, pourcentage.

b) Conclusion :

- Par l'analyse précédente, nous voyons que la statistique n'a pas effectivement de place dans les sujets du brevet à la fin de classe troisième en juin 1995. En effet, l'effectif existe dans ces 6 sujets pour le calcul des pourcentages plus tard qui ne sert qu'à tracer un diagramme circulaire ou diagramme demi-circulaire. En plus, le mot « fréquence » n'apparaît qu'une seule fois dans un seul sujet mais comme nous avons déjà analysé précédemment, il est écrit « Fréquence en % »

- La lecture graphique, la représentation graphique et le calcul sont dominants dans la partie statistique très modeste de « Travaux numériques » ces sujets de brevet.

- Le tableau suivant est une synthèse de l'apparition des notions dans les questions du sujet de brevet :

Sujet N°	Nombre de questions demandant à calculer					
	Effectif	Fréquence en %	Effectif / Fréquence cumulé(e)	Moyenne	Médiane	Graphique
2	1	1 (%)				1
5	1	1 (%)				1
9						
11	1	1				
12		1 (%)	1			1
15	1	1 (%)				
Total	4	5	1	0	0	3

Tableau 126. Tableau de nombre des questions dans les 6 sujets de brevet en Juin 1995

En observant ce tableau, on peut trouver une grande surprise, c'est que la moyenne n'existe pas dans ces 6 sujets. Par contre, ce n'est pas le cas de la médiane car elle n'est pas enseignée avant la réforme.

- Et tous les types de tâches se trouvent dans le tableau 1 suivant le montrent aussi :

Groupe	T	Types de Tâches
1	T5	Lire des informations d'un tableau
	T6	Lire des informations à partir d'une représentation graphique
	T7	Représenter en graphique des données statistiques
	T8	Représenter/Regrouper des données statistiques en tableau
	T12	Calculer la mesure d'angle d'un diagramme qui représente des données statistiques

	T16	Lire des indicateurs (l'effectif, la moyenne, la médian) dans une représentation graphique
	T18	Dresser le tableau des effectifs (à partir des graphiques)
2	T28	Calculer l'effectif
	T29	Calculer des effectifs cumulés
	T26	Calculer le pourcentage
	T32	Calculer la fréquence
	T35	Calculer les fréquences cumulées
	T46	Calculer la taille de données statistiques

Tableau 127. Tableau des types tâches dans les sujets de Brevet en Juin 1995[205]

III.1.2. Après la réforme :

Un petit rappel :

Les notions enseignées pendant cette période sont : Classe, Effectif, Fréquence, Moyenne, Médiane, Étendue.

a) Analyse des sujets d'épreuve :

Il y a 18 sujets [206] de brevet en 2005 et en 2006 présentés dans l'ouvrage Annabrevet 2007 (l'ouvrage le plus récent). Chaque sujet de brevet se compose 3 parties :

– Travaux numérique et gestion des données : comprend de 3 à 5 exercices

– Travaux géométriques : comprend de 2 à 3 exercices

– Problème : ne comprend pas des questions concernant la statistique.

En total, ces 18 sujets comprennent 128 exercices dont 6 concernent la statistique (ça occupe environ 4,69%).

Dans la partie suivante, nous allons faire une analyse sur ces 6 exercices pour voir la place de la statistique enseignée au collège dans l'épreuve de brevet.

[205] L'ouvrage Annabrevet 1996

[206] Voir l'Annexe 2.

2.1. Sujet 2 : Sujet du brevet Nants, Bordeaux, Caen, Clermont-Ferrand, Limoges, Poitiers, Rennes, Juin 2006

Exercice 3

Le tableau ci-dessous donne la répartition des notes obtenues à un contrôle de mathématiques par les 27 élèves d'une classe de troisième.

Notes	6	8	10	13	14	17
Effectifs	3	5	6	7	5	1

1. Calculer la note moyenne de la classe à ce contrôle. Arrondir le résultat à l'unité.
2. Calculer le pourcentage d'élèves ayant eu une note supérieure ou égale à 10. Arrondir le résultat au dixième.

(Sujet 2, Annabrevet 2007, p. 24)

Analyse :

Les données statistiques sont fournies en forme d'un tableau de deux lignes concernant les notes du contrôle de maths de 27 élèves avec ses effectifs. Dans ce cas, la moyenne demandée est celle pondérée. Et le pourcentage demandé de calculer est la fréquence cumulée.

2.2. Sujet 6 : Sujet du brevet Centres étrangers, Juin 2006

Exercice 3

Le tableau ci-dessous présente la série des notes obtenues par les éléves de 3e B lors du dernier devoir en classe :

Note sur 20	5	6	8	9	11	12	13	15	18	19
Effectif	1	2	6	2	1	4	2	3	1	1

1. Quel est l'effectif de la classe de 3e B ?
2. Calculer la note moyenne de ce devoir. En donner la valeur arrondie au dixiéme de point.
3. Quel est le pourcentage, arrondi à l'unité, de l'effectif total représentent les éléves ayant obtenu une note inférieure ou égale à 8.
4. Déterminer la note médiane de cette séne. Que représente cette note ?

(Sujet 6, Annabrevet 2007, p. 45)

Analyse :

Comme sujet 2, dans ce sujet, les données sont fournies en forme d'un tableau de deux lignes concernant les notes d'un devoir de classe 3e avec ses effectifs (sans savoir le nombre total d'élève). Là, on voit que :

- L'effectif demandé de calculer est l'effectif total.

- La moyenne demandée de calculer est celle pondérée.

- Le pourcentage demandé de calculer est la fréquence cumulée.

- A part de la demande de calculer la note médiane, question 4 demande de la représentation de cette note. Dans le Corrigé, l'ouvrage donne la réponse à cette question comme suit :

> « Nous savons que la médiane M est la valeur qui partage la série statistique, rangée par ordre croissant (ou décroissant), en deux parties de même effectif.
> Écrivons toutes les notes obtenues parles élèves de la classe 3ᵉB.
> 5-6-6-8-8-8-8-8-8-9-9-**11**-12-12-12-12-13-13-15-15-15-18-19.
> Avant la note **11** il existe onze notes et, après cette même note **11**, il existe également onze notes. Nous pouvons affirmer que la médiane est :
> **M = 11**
> Il existe autant d'élève ayant une note inférieure à 11 que d'élève ayant une note supérieure à 11. »

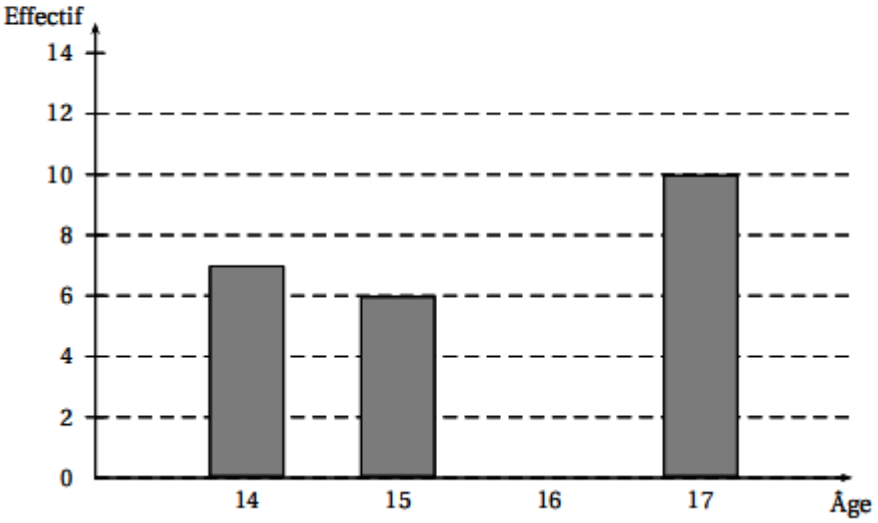
(Annabrevet 2007, p. 49)

Nous voyons que l'ouvrage donne la représentation de la médiane en reprenant sa définition sans autre remarque par rapport à la moyenne calculée qui est « m = 10,9 ». A partir de ces deux valeurs de moyenne et de médiane, il est possible de faire une comparaison entre eux.

2.3. Sujet 7 : Sujet du Brevet Madagascar, Océan Indien, Juin 2006

Exercice 4

L'histogramme ci-dessous illustre une enquête faire sur l'âge des 30 adhérents d'un club de badminton mais le rectangle correspondant aux adhérents de 16 ans a été effacé.



1. Calculer le nombre d'adhérents ayant 16 ans.
2. Quel est le pourcentage du nombre d'adhérents ayant 15 ans?
3. Quel est l'âge moyen des adhérents du club? Donner la valeur arrondie au dixième.
4. Reproduire et compléter le tableau ci-dessous pour réaliser un diagramme semi-circulaire représentant la répartition des adhérents selon leur âge (on prendra un rayon de 4 cm).

Âge	14 ans	15 ans	16 ans	17 ans	Total
Nombre d'adhérents	7	6		10	30
Mesure de l'angle (en degrés)					180

(Sujet 7, Annabrevet 2007, p. 50)

Analyse :

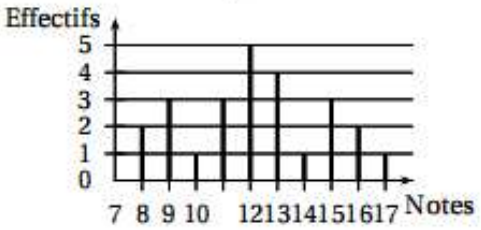
Dans ce sujet, les données sont fournies en forme d'un histogramme concernant l'âge de 30 adhérents d'un club de badminton mais avec un manque d'information. Le manque est le rectangle représentant le nombre d'adhérents ayant 16 ans. En observant cet histogramme, le candidat doit trouver ce nombre (qui est l'effectif des adhérents de 16 ans). Puis, à partir de l'effectif lisible sur l'histogramme, le candidat doit calculer le pourcentage (qui est la fréquence) des adhérents de 15 ans. Ensuite, un calcul de l'âge moyen des adhérents est demandé. Et enfin, les données sont reproduites sous la forme d'un tableau des effectifs dont la troisième ligne est vierge pour que le candidat remplisse les mesures des angles correspondant à un diagramme demi-circulaire à tracer.

On peut voir que cet exercice est une illustration riche d'une répartition des effectifs de données en plusieurs formes : histogramme, tableau, diagramme demi-circulaire. En plus, le candidat doit effectuer plusieurs de tâches comme : lire un graphique, calculer l'effectif, calculer la moyenne, calculer les mesures des angles, représenter les données en diagramme.

2.4. Sujet 9 : Sujet du brevet Pondichéry, Avril 2006

Exercice 4

Voici le diagramme en bâtons des notes obtenues par une classe de Troisième de 25 élèves au dernier devoir de mathématiques



1. Calculer la moyenne des notes.
2. Déterminer la médiane des notes.
3. Calculer le pourcentage des élèves ayant obtenu une note strictement supérieure 'a 13.

(Sujet 9, Annabrevet 2007, p. 60)

Analyse :

Dans ce sujet, les données sont en forme d'un graphique concernant les notes d'un devoir de mathématiques de 25 élèves d'une classe Troisième. En lisant ce graphique, le candidat doit trouver :

- Les effectifs de chaque note pour calculer la moyenne des notes.

- Puis trouver la médiane des notes.

- Enfin, le pourcentage à calculer est la fréquence cumulée.

On peut voir que cet exercice du sujet 9 ressemble celui du sujet 6 précédent sauf la forme de données (ici, on a le diagramme en bâtons des effectifs au lieu de tableau des effectifs). L'ouvrage donne la réponse de même manière trouvée dans le sujet 6. Et nous remarquons que la moyenne m = 12,24 tandis que la médiane M = 12. Encore une fois, il n'y a pas de comparaison entre ces deux indicateurs de position.

2.5. Sujet 12 : Sujet du brevet Nancy-Metz, Besançon, Dijon, Grenoble, Lyon, Reims, Strasbourg, Septembre 2005

Exercice 5

Voici les notes de 13 élèves à un devoir de mathématiques :

6 ; 8 ; 8 ; 9 ; 10 ; 11 ; 12 ; 14 ; 17 ; 18 ; 18 ; 19.

1. Calculer la moyenne arrondie au centième de cette série de
2. Déterminer la médiane de cette série de notes.

(Sujet 12, p. 77)

Analyse :

Dans ce sujet, les données une liste de notes d'un devoir de mathématiques de 13 élèves. Là, on trouve des mêmes questions que les sujets précédents. Pour la réponse, la moyenne m = 12,23 et la médiane M = 11.

2.6. Sujet 14 : Sujet du brevet Nouvelle-Calédonie, Novembre 2005

EXERCICE 3

Voici les résultats d'un sondage effectué dans une classe de troisième concernant les moyens de transport utilisés par ces élèves pour venir au collège.

Recopier et compléter le tableau suivant puis construire un diagramme circulaire de 3 cm de rayon :

	Voiture	Bus	À pied	Booster	total
Fréquence	45%	25%	20%	10%	
Angle					

(Sujet 14, Annabrevet 2007, p. 86)

Analyse :

Dans ce sujet, les données statistiques sont données par un tableau des fréquences. Cet exercice est le seul où se trouve le mot « fréquence ». Pourtant, ces fréquences sont en écriture de pourcentage et arrondisses à unité. Il semble que ce fait est pour qu'il soit familial pour les élèves. Nous le disons car parmi touts les 6 exercices,

Il semble que cet exercice ne comprend que des calculs.

b) Conclusion :

- Une synthèse de l'apparition des notions dans les questions du sujet de brevet se trouve dans le tableau suivant :

Sujet N°	Nombre de questions demandant à calculer / tracer					
	Effectif	Fréquence	Effectif / Fréquence cumulé(e)	Moyenne	Médiane	Graphique
2		1	1	1		
6	1	1	1	1	1	
7	1	1		1		1
9		1	1	1	1	
12		1 (%)		1	1	
14		1 (%)				1
Total	2	6	3	5	3	2

Tableau 128. Tableau de nombre des questions dans les 6 sujets de brevet 2005-2006

En l'observant, on voit que les données sont très souvent des notes d'un contrôle des élèves. La moyenne est la demande la plus souvent de ces 6 exercices concernant la statistique. La fréquence n'existe qu'une seule fois en forme de pourcentage. C'est le même cas de 6 sujets du brevet en juin 1995 que nous avons analysé dans la partie précédente.

Le tableau suivant présente des types de tâches se trouvant à travers des questions des exercices analysés précédents :

Groupe	T	Types de Tâches
1	T7	Représenter en graphique des données statistiques
	T12	Calculer la mesure d'angle d'un diagramme qui représente des données statistiques
2	T28	Calculer l'effectif
	T32	Calculer le pourcentage
	T35	Calculer les fréquences cumulées
	T36	Calculer la moyenne
	T47	Calculer la médiane

Tableau 129. Tableau des types de tâches dans les sujets de Brevet en 2005 et en 2006

Une remarque est qu'il n'y a aucune question demandant de calculer l'étendue même si elle est la seule nouvelle notion enseignée après la réforme. D'ailleurs, même si le programme ne propose pas l'enseignement de la notion *d'écart moyen*[207], mais dans la partie « Formulaire » à la fin de l'ouvrage Annabrevet 2007, on définit cette notion comme une des indicateurs de dispersion sans trouver aucune question concernant à cette notion.

- Les exercices de statistique dans ces 6 sujets apparient pour la raison de calcul et non pour la statistique.

- Il semble que la partie du problème d'un sujet de brevet n'soit jamais réservée pour la statistique. En effet, parmi 18 problèmes de 18 sujets présentés dans l'ouvrage Annabrevet 2007, le sujet du Brevet Amérique du Nord, Juin 2006, est le seul ayant un problème dont la troisième partie nous fait penser qu'il est un exercice statistique au début :

Partie 3

Le responsable du supermarché a relevé le nombre de saladiers vendus par chacune de ses quatre vendeuses et l'a inscrit dans le tableau suivant :

Nom de la vendeuse	Sofia	Natacha	Lorie	Magali
Nombre de saladiers vendus	220	200	290	250

1. Combien de saladiers ont été vendus ?
2. Calculer le pourcentage de saladiers vendus par Natacha. Arrondir au dixième.
3. Le responsable du supermarché affirme qu'il a vendu 80 % de son stock. Combien avait-il acheté de saladiers ?

(Sujet du Brevet Amérique du Nord, Juin 2006)

Quoique la première question demande de calculer le nombre total des saladiers vendus (qui est également l'effectif total), puis de calculer un pourcentage (qui est la fréquence), mais aucun mot concernant la statistique s'y trouve. La dernière question est effectivement un calcul d'un nombre en sachant son pourcentage.

III.1.3. Conclusion sur l'analyse des sujets du brevet :

Au niveau de collège, comme l'analyse des programmes et des manuels, la partie de statistique ne change pas beaucoup et même dans les sujets de brevet avant et après la réforme.

La statistique enseignée au collège est vraiment la statistique descriptive avec des travaux sur la description des données, sur des calculs des indicateurs, et rien du travail sur l'interprétation des indicateurs statistiques (de position) ni de comparaison (entre la moyenne et la médiane, ou bien entre des séries statistiques,…).

[207] La définition dans l'Annabrevet : « Ecart moyen : c'est la moyenne de la série obtenue enseignant prenant les valeurs positives des différences entre chaque valeur de la série statistique et la valeur moyenne de la série » (Annabrevet 2007, p. 286)

Avant la réforme, exercice concernant la statistique ne demande que de décrire la répartition de données en graphique, de calculer le pourcentage, … mais après la réforme, à part de ces travaux, les exercices du brevet demandent de calculer des caractéristiques de position comme la moyenne, la médiane,... En conclusion, on peut dire que les questions des sujets concentrent au cœur des notions statistiques plus qu'à l'avance.

La statistique occupe toujours une place très modeste dans les sujets d'épreuve de mathématiques au brevet. Ce ne sont pas les cas de Calcul dans R, de Factorisation, de Racine carrée, d'Équation, …. dans la partie des Travaux numériques et gestion de données du sujet d'épreuve.

Même si la notion de fréquence est enseignée au collège, mais nous trouvons que le sujet de brevet évite d'utiliser le mot « fréquence » pour la demande de calcul, sinon, le mot « fréquence » doit être lié avec des mots « en % ». Par contre, on « aime » utiliser le mot « pourcentage ». Est-ce que ce fait est pour que le sujet satisfasse à la fois deux buts : la proportionnalité (par le mot « pourcentage ») et la statistique (la signification statistique de pourcentage dans ces cas) ?

III.2. Baccalauréat :

III.2.1. Avant la réforme :

Un petit rappel :

La statistique se trouve dans la classe de seconde avec des notions statistiques comme : effectifs, fréquences, effectifs cumulés, fréquences cumulées, moyenne, écart-type.

Classe première est réservée pour la probabilité, puis en Terminal, des notions concernant la statistique comme variable aléatoire, fonction de répartition, espérance mathématique, variance, écart-type.

Remarque :

Parmi 55 sujets d'épreuve du BAC en 1997 et 1998 se trouvant dans l'ouvrage Annabac 1999 (Hatier, août 1998), il n'y a que 10 sujets ayant des questions concernant la statistique au niveau de la variable aléatoire. Concrètement, ces questions demandent de déterminer la loi de probabilité et de calculer l'espérance mathématique. Par contre, il y a 14 sujets ayant des questions concernant la probabilité élémentaire où certaines ont des questions demandant de calculer des probabilités d'un événement qui forment une suite à étudier sa limite lorsque n tend vers l'infini,….

III.2.2. Après la réforme :

Un petit rappel :

La statistique dans la classe de seconde se présente avec des notions statistiques comme : moyenne, médiane, classe modale, moyenne élaguée, étendue, distribution des fréquences, fréquence d'un événement, simulation et fluctuation d'échantillonnage.

Puis en classe première, les notions enseignées sont variance, écart-type, variable aléatoire, loi d'une variable aléatoire, espérance, variance, écart type et particulièrement, intervalle interquartile, diagrammes en boîtes, modélisation d'expériences aléatoires sont enseignés.

Enfin, en Terminal, à part de la probabilité conditionnelle, les lois de probabilité comme loi de Bernoulli, loi binomiale ; espérance et variance de ces lois sont présentés. En plus, l'idée de simulation et de modélisation concernant la statistique est prise comme un lien entre la statistique et la probabilité. En particulier, l'adéquation de données expérimentales à une loi équirépartie est demandée à étudier par un exemple.

Remarque :

Nous prenons l'ouvrage Annabac 2006 (Nathan, 2005) comme la référence. Dans cet ouvrage, il y a 9 sujets d'épreuve du BAC en Juin 2005 mais la variable aléatoire ne se trouve que dans deux sujets : sujet 7 (Polynésie 2005) et sujet 8 (Liban 2005) tandis que les calculs de la probabilité élémentaire sont demandés dans 7 sujets (France métropolitaine, Amérique du Nord, Antilles-Guyane, Centres étrangers I, La réunion, Polynésie et Liban en 2005).

Comme la demande concernant la variable aléatoire du sujet 7, l'ouvrage le prend pour définir une variable aléatoire puis demande de calculer la probabilité sans autre demande concernant l'espérance,… Par contre, sujet 8 est le seul ayant des questions demandant de déterminer la loi de probabilité puis de calculer l'espérance de la variable aléatoire étudiée.

Pourtant, dans cet ouvrage, à part de 9 sujets du Bac, pour donner plus de sujets exemplaires aux candidats, on a présenté aussi :

- 2 sujets de concours (un ESIEE en mai 2005 et GEPEI en mai 2005)

- 41 exercices qui sont retirés des sujets de bac entre 2002 et 2004, numérotés de 1 à 41, et classés par 6 thèmes principaux :

- 6 exercices de thème Analyse ;

- 4 exercices de Nombres complexes ;

- 12 exercices de thème Probabilité ;

- 7 exercices de thème Géométrie dans l'espace ;

- 6 exercices de thème Arithmétique ;

- 6 exercices de thème Similitude).

Parmi 12 exercices de thème Probabilité (numérotés de 11 à 22), nous nous intéressons au dernier exercice (l'exercice 22, p. 338) qui est extrait du sujet de Bac *d'Antilles-Guyane en septembre 2004* [208]. Cet exercice se compose trois questions indépendantes dont la troisième (question 3) est la seule qui concerne l'adéquation à une loi de probabilité.

[208] Voir Annexe 4

Voici l'extrait de cette question :

3. Les 1 000 premières décimales de π sont données ici par un ordinateur :

1415926535 8979323846 2643383279 5028841971 6939937510
5820974944 5923078164 0628620899 8628034825 3421170679
8214808651 3233066470 9384460959 0582235725 3594085234
8111745028 4102701930 5211055596 4462294895 4930301964
4288109756 6593344612 8475648233 7867831652 7120190914
5648566923 4603486534 5432664825 3393607260 2491412737
2450700660 6315580574 8815209209 6282925409 1715364367
8925903600 1133053054 8820466525 3841469519 4151160943
3057270365 7595919530 9218611738 1932611793 1051185480
7446297996 2749567355 8857527240 9122793318 3011949129
8336733624 4065664308 6025394946 3952247371 9070217986
0943702770 5392171762 9317675238 4674818467 6691051320
0056812714 5263560827 7857753427 9778900917 3637178721
4684409012 2495343054 6549585371 0507922796 8925892354
2019956112 1290219608 6403441815 9813629774 7713099605
1870721134 9999998372 9780499510 5973173281 6096318599
0244594553 4690830264 2522300253 3446850352 6193110017
1010003137 8387528865 8753320830 1420617177 6691473035
9825349042 8755460731 1595620633 8235378759 3751957781
8577805321 7122600661 3001927876 6111959092 1642019894

En groupant par valeurs entre 0 et 9 ces décimales, on obtient le tableau suivant :

Valeurs	0	1	2	3	4	5	6	7	8	9
Occurrences	93	116	102	102	94	97	94	95	101	106

Avec un tableur, on a simulé 1 000 expériences de 1 000 tirages aléatoires d'un chiffre compris entre 0 et 9.

Pour chaque expérience, on a calculé $d^2 = \sum_{k=0}^{k=9} (f_k - 0,1)^2$ où f_k représente, pour l'expérience, la fréquence observée du chiffre k.

On a alors obtenu une série statistique pour laquelle on a calculé le premier et neuvième décile (d_1 et d_9), le premier et troisième quartile (Q_1 et Q_3) et la médiane (Me) :

$d_1 = 0,000422$; $Q_1 = 0,000582$; Me $= 0,000822$; $Q_3 = 0,001136$; $d_9 = 0,00145$.

En effectuant le calcul de d_2 sur la série des 1 000 premières décimales de π, on obtient :

☐ 0,000 456 ☐ 0,004 56 ☐ 0,000 314

Un statisticien découvrant le tableau et ignorant qu'il s'agit des décimales de π, fait l'hypothèse que la série est issue de tirages aléatoires indépendants suivant une loi équirépartie. Il prend un risque de 10 % de rejeter cette hypothèse quand elle est vraie. Accepte-t-il cette hypothèse ?

☐ Oui ☐ Non ☐ Il ne peut pas conclure.

(Sujet de bac d'Antilles-Guyane, septembre 2004)

Analyse d'exercice :

Cet exercice est un travail concernant l'adéquation d'une distribution observée e loi équirépartie. Pour le résoudre, on applique la méthode du test Khi-deux[209].

Sous tâches :

ST1 : Calcul des fréquences observées.

ST2: Calcul de la probabilité de la loi uniforme associée à l'expérience aléatoire.

ST3: Calcul des déciles, d1, d9 ; des interquartiles : Q1, Q3, et médiane Me.

ST4 : Calcul de la distance d^2 entre la distribution des fréquences observée et la distribution des fréquences théorique.

ST5 : Simuler m fois d'une expérience aléatoire pour avoir une série des fréquences observées

$$(f_k^i)\ ,0 \leq i \leq m$$

sT6 : Vérifier l'hypothèse d'adéquation à une loi de probabilité pour jeter ou non.

Technique :

C'est la composition des sous tâches :

1. Construire le tableau des fréquences à partir d'un tableau des données ;

2. Calculer la probabilité théorique de la loi uniforme (équiprobabilité) correspondant : 0,1

3. Calculer la distance entre la fréquence observée et la fréquence théorique :

$$f_k - 0{,}1 \qquad où\ \ 0 \leq k \leq 9$$

4. Calculer la distance entre la distribution des fréquences observée et la distribution des fréquences théoriques (probabilité théorique) :

$$d_{obs}^2 = \sum_{k=0}^{9} (\ f_k - 0{,}1)^2$$

5. Simulation 1000 fois de avec un tableur de la loi équirépartie

[209] Voir Annexe 5

$$d_i^2 = \sum_{k=0}^{9} (f_k^i - 0{,}1)^2 \qquad , \ i = 1, 2, 3, 4, ...$$

Puis on peut calculer d_1 (premier décile), d_9 (neuvième décile), Q_1 (première interquartile), Q_3 (première interquartile), Me (médiane) : dans ce sujet, ces valeurs sont toutes données.

6. Hypothèse d'adéquation à une loi de probabilité via les distances (outils pour mesurer l'adéquation)

7. Rejet ou non de cette hypothèse en comparaissant d^2_{obs} et d_9 (90%).

Remarque :

Il y a une rupture entre les étapes 4 et 5. Là, on introduit le hasard ! Est-ce qu'un échantillon d'une loi équirépartie ?

III.2.3. Conclusion :

Au niveau du lycée, d'près l'analyse de programme et de manuel, on voit que la statistique est bien insistée à renforcée après la réforme avec plusieurs nouvelles comme la simulation, la fluctuation d'échantillonnage, puis l'adéquation de données expérimentales à une loi équirépartie… Mais dans ces sujets de BAC, il semble que cette idée de changement ne se réalise pas. Aucune nouveauté ne s'y trouve.

Annexe A1 :

Sujet	Travaux numérique et gestion des données	Travaux géométriques	Problème
1	2	2	1
2	4	3	1
3	3	3	1
4	3	2	1
5	4	3	1
6	3	2	1
7	3	3	1
8	3	2	1
9	5	3	1
10	3	2	1
11	4	3	1
12	6	3	1
13	3	2	1
14	3	2	1
15	3	3	1
16	3	2	1
17	3	3	1
TOTAL : 118	58	43	17

Tableau 130. Le nombre des exercices et des problèmes dans les 17 sujets de brevet en Juin 1995

Annexe A2 :

Sujet	Travaux numérique et gestion des données	Travaux géométriques	Problème
1	3	2	1
2	3	3	1
3	3	3	1
4	4	3	1
5	3	2	1
6	3	3	1
7	4	3	1
8	3	3	1
9	4	3	1
10	4	3	1
11	3	2	1
12	5	2	1
13	3	2	1
14	4	2	1
15	4	2	1
16	3	3	1
17	3	3	1
18	4	3	1
TOTAL : 128	63	47	18

Tableau 131. Le nombre des exercices et des problèmes dans les 18 sujets de brevet 2005-2006

Annexe A3 :

Programme de mathématiques de 3ème pour le brevet

Travaux géométriques

- Géométrie dans l'espace
  - Sphère
  - Problèmes de sections planes de solides
- Triangle rectangle : relations trigonométriques, distance de deux points dans un repère orthonormé du plan
- Propriété de Thalès
- Vecteurs et translations
  - Egalité vectorielle
  - Composition de deux translations ; somme de deux vecteurs
  - Coordonnées d'un vecteur dans le plan muni d'un repère
  - Composition de deux symétries centrales
- Rotation, angles, polygones réguliers
  - Images de figures par une rotation
  - Polygones réguliers
  - Angle inscrit

Travaux numériques

- Ecritures littérales ; identités remarquables
- Calculs élémentaires sur les radicaux (racines carrées)
  - Racine carrée d'un nombre positif
  - Produit et quotient de deux radicaux
- Equations et inéquations du premier degré
  - Ordre et multiplication
  - Inéquation du premier degré à une inconnue
  - Système de deux équations à deux inconnues
  - Résolution de problèmes du premier degré ou s'y ramenant
- Nombres entiers et rationnels
  - Diviseurs communs à deux entiers
  - Fractions irréductibles

Organisation et gestion de données - Fonctions

- Fonction linéaire et fonction affine
  - Fonction linéaire
  - Fonction affine
  - Fonction affine et fonction linéaire associée
- Proportionnalité et traitements usuels sur les grandeurs
  - Applications de la proportionnalité
  - Grandeurs composées
  - Changement d'unités
  - Calculs d'aires et de volumes
  - Effets d'une réduction ou d'un agrandissement sur des aires ou des volumes
- Statistique
  - Caractéristiques de position d'une série statistique
  - Approche de caractéristiques de dispersion d'une série statistique
  - Initiation à l'utilisation de tableurs-grapheurs en statistique

Annexe A4 :

EXERCICE 4 **5 points**

Pour chacune des trois questions, la totalité des points sera donnée si la réponse est correctement justifiée.

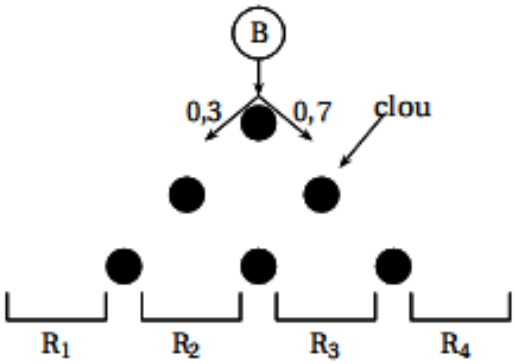
Les trois questions sont indépendantes.

1. La probabilité pour un individu d'une population d'être atteint d'une maladie M est égale à 0,003. Un test de dépistage, pour cette maladie, a été réalisé ; avec ce test, on peut dire que
   - si une personne est atteinte de la maladie M, le test est positif dans 50 % des cas ;
   - le test est positif pour 3 % des personnes saines.

   Quelle est à 0,01 près la probabilité d'avoir la maladie M lorsque le test est positif ?

   ☐ 0,95 ☐ 0,9 ☐ 0,15 ☐ 0,05

2. On considère une planche à clous de ce type :



On lance une boule B du haut de la planche, elle tombe alors dans l'un des quatre récipients notés R_1, R_2, R_3 et R_4. À chaque étape, la bille a une probabilité de 0,3 d'aller vers la gauche et 0,7 d'aller vers la droite (gauche et droite relatives à l'observateur).

On note p_1 la probabilité que la bille tombe dans le bac R_1 ou dans le bac R_3 et p_2 la probabilité que la bille tombe dans le bac R_2 ou dans le bac R_4.

Que valent p_1 et p_2 ?

- ☐ $p_1 = p_2 = 0{,}5$
- ☐ $p_1 = 0{,}216$ et $p_2 = 0{,}784$
- ☐ $p_1 = 0{,}468$ et $p_2 = 0{,}532$
- ☐ $p_1 = 0{,}468$ et $p_2 = 0{,}432$.

3. Les 1 000 premières décimales de π sont données ici par un ordinateur :

1415926535 8979323846 2643383279 5028841971 6939937510
5820974944 5923078164 0628620899 8628034825 3421170679
8214808651 3233066470 9384460959 0582235725 3594085234
8111745028 4102701930 5211055596 4462294895 4930301964
4288109756 6593344612 8475648233 7867831652 7120190914
5648566923 4603486534 5432664825 3393607260 2491412737
2450700660 6315580574 8815209209 6282925409 1715364367
8925903600 1133053054 8820466525 3841469519 4151160943
3057270365 7595919530 9218611738 1932611793 1051185480
7446297996 2749567355 8857527240 9122793318 3011949129
8336733624 4065664308 6025394946 3952247371 9070217986
0943702770 5392171762 9317675238 4674818467 6691051320
0056812714 5263560827 7857753427 9778900917 3637178721
4684409012 2495343054 6549585371 0507922796 8925892354
2019956112 1290219608 6403441815 9813629774 7713099605
1870721134 9999998372 9780499510 5973173281 6096318599
0244594553 4690830264 2522300253 3446850352 6193110017
1010003137 8387528865 8753320830 1420617177 6691473035
9825349042 8755460731 1595620633 8235378759 3751957781
8577805321 7122600661 3001927876 6111959092 1642019894

En groupant par valeurs entre 0 et 9 ces décimales, on obtient le tableau suivant :

Valeurs	0	1	2	3	4	5	6	7	8	9
Occurrences	93	116	102	102	94	97	94	95	101	106

Avec un tableur, on a simulé 1 000 expériences de 1 000 tirages aléatoires d'un chiffre compris entre 0 et 9.

Pour chaque expérience, on a calculé $d^2 = \sum_{k=0}^{k=9} (f_k - 0,1)^2$ où f_k représente, pour l'expérience, la fréquence observée du chiffre k.

On a alors obtenu une série statistique pour laquelle on a calculé le premier et neuvième décile (d_1 et d_9), le premier et troisième quartile (Q_1 et Q_3) et la médiane (Me) :
$d_1 = 0,000422$; $Q_1 = 0,000582$; Me $= 0,000822$; $Q_3 = 0,001136$; $d_9 = 0,00145$.

En effectuant le calcul de d_2 sur la série des 1 000 premières décimales de π, on obtient :

☐ 0,000 456 ☐ 0,004 56 ☐ 0,000 314

Un statisticien découvrant le tableau et ignorant qu'il s'agit des décimales de π, fait l'hypothèse que la série est issue de tirages aléatoires indépendants suivant une loi équirépartie. Il prend un risque de 10 % de rejeter cette hypothèse quand elle est vraie. Accepte-t-il cette hypothèse ?

☐ Oui ☐ Non ☐ Il ne peut pas conclure.

Annexe A5 :

Adéquation à une loi

Les problèmes de validation de modèles sont complexes. Pour y sensibiliser les élèves, on peut commencer par des expériences de référence (tirages de boules dans des urnes, de choix au hasard, etc.) pour comprendre de quoi il s'agit, puis traiter des exemples.

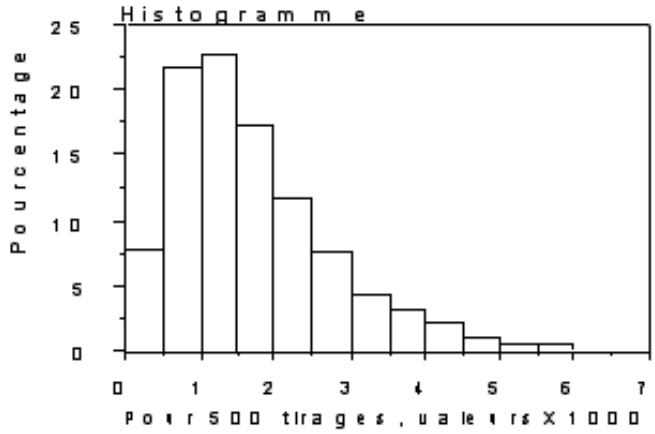
Un joueur veut vérifier si le dé qu'il utilise est équilibré. Comment peut-il faire ?

Il pourrait étudier la symétrie du dé, mais ce n'est pas exactement de cela qu'il s'agit. Il convient plutôt de vérifier que dans des conditions normales d'utilisation du dé, les résultats sont *compatibles* avec un modèle d'équiprobabilité sur {1,2,3,4,5,6}.
Essayons donc de définir sur cet exemple un critère de compatibilité de données expérimentales avec un modèle.

On peut par exemple s'intéresser à la distance entre la distribution des fréquences $(f_1, \ldots, f_6)$ obtenues en lançant n fois un dé et la loi de probabilité $\{1/6, \ldots, 1/6\}$ et regarder si cette distance est *petite*. En prenant la définition classique de la distance, on peut fonder la notion de compatibilité sur l'étude du carré de cette distance, à savoir :

$$d^2 = (f_1 - 1/6)^2 + (f_2 - 1/6)^2 + \ldots + (f_6 - 1/6)^2.$$

La quantité d^2 est soumise à la fluctuation d'échantillonnage, *i.e.* sa valeur varie d'une série de lancers à l'autre. C'est précisément l'étude de la fluctuation d'échantillonnage qui va permettre de convenir d'un seuil entre valeur *petite* et valeur *non petite* de d^2.
Imaginons que le joueur ait lancé 500 fois le dé et ait obtenu une distribution de fréquence $(f_1, \ldots, f_6)$, d'où une valeur observée d^2_{obs} qu'on va comparer à d'autres. Pour cela, on simule des séries de $n = 500$ chiffres au hasard dans $\{1, \ldots, 6\}$. Ci-dessous, on voit un histogramme de 2 000 valeurs de d^2 obtenues par des simulations de séries de 500 chiffres au hasard dans $\{1, \ldots, 6\}$.



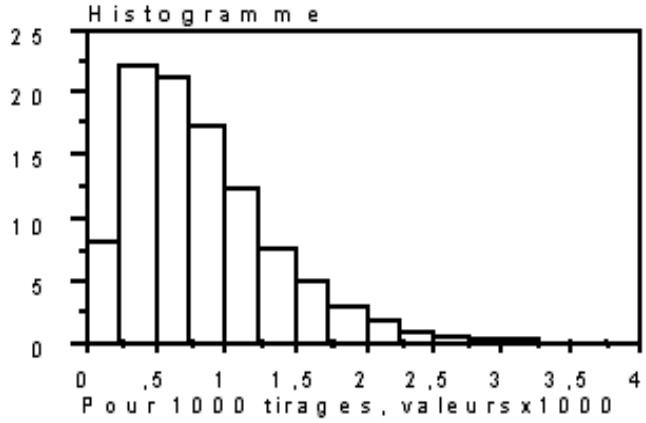
Le 9e décile de la série des valeurs simulées de d^2 est 0,003 (soit 90 % des valeurs simulées de d^2 sont dans l'intervalle [0 ; 0,003]).

Convenons de la décision suivante :

– si $d^2_{obs} \leqslant 0{,}003$, alors le dé sera déclaré équilibré ;

– si $d^2_{obs} > 0{,}003$, alors le dé sera déclaré non équilibré.

On associera à cette conclusion le risque $\alpha = 0{,}1$ correspondant au fait suivant : en utilisant cette règle de décision sur les données simulées, on se serait « trompé » dans 10 % des cas.
Deux éléments semblent arbitraires dans cette façon de conclure ou non à l'équilibre du dé : que se passerait-il pour une autre simulation, de taille égale ou non, et que se passerait-il si au lieu de lancer 500 fois le dé, on le lançait par exemple 1 000 fois ?
Qu'à cela ne tienne : étudions les résultats de 5 000 simulations de séries de 1 000 tirages de chiffres au hasard dans $\{1, \ldots, 6\}$; le neuvième décile des valeurs d^2 est 0,0015, *i.e.* 90 % des valeurs de d^2 simulées sont dans l'intervalle [0 ; 0,0015]. Comparons les résumés graphiques dans les deux séries simulées : même allure, à l'échelle près. L'existence d'un changement d'échelle est conforme au théorème des grands nombres vu en première : plus le nombre de tirages est grand, plus la distribution des fréquences est proche de la loi de probabilité, donc plus les valeurs de d^2 sont petites.



Des résultats théoriques expliquent la ressemblance de forme entre les histogrammes : on sait démontrer que la loi de probabilité de nd^2, où n est le nombre de tirages (n vaut 500 puis 1 000 dans notre étude), ne bouge sensiblement plus avec n lorsque n est suffisamment grand ; c'est là un des nombreux résultats de la théorie des probabilités. En pratique, on pourra considérer que la loi de probabilité de nd^2 dépend peu de n, pour tout $n > 100$.

On peut adapter ce qui vient d'être fait en changeant le risque a, ou en testant l'adéquation à une loi équirépartie sur un ensemble à k éléments pour diverses valeurs de k. La répartition des valeurs de loi de probabilité de nd_k^2 où :

$$d_k^2 = \left(f_1 - 1/k\right)^2 + \left(f_2 - 1/k\right)^2 + \ldots + \left(f_k - 1/k\right)^2$$

ne dépend quasiment plus de n pour n grand mais dépend cependant de k.

Le choix du risque α dépend du contexte : il est fonction de l'enjeu lié à la question et parfois on conclut non par rapport à un risque, mais par rapport à une plage de risques ; ce risque n'est en général pas choisi par le statisticien. On prend souvent par défaut $\alpha = 0{,}05$: il est important que les élèves sachent qu'il s'agit d'un consensus et non d'une constante immuable. L'enjeu est de comprendre sur quoi porte le risque (refuser à tort le modèle) et que plus le risque est petit, plus on aura tendance à accepter le modèle de l'équiprobabilité.

Remarques

1) Une idée naturelle serait ici d'étudier les fluctuations de $\delta = |f_1 - 1/2|$ et de les comparer aux fluctuations de la même quantité lorsqu'on simule un grand nombre de séries de taille 53 041 de nombres (0,1) tirés au hasard. Mais on peut écrire : $d_2^2 = 2\delta^2$ et les deux études sont donc identiques. En particulier, comme on a vu que les variations en n de la répartition des valeurs de nd_2^2 pouvaient être négligées dès que n est grand, il en est de même pour $\sqrt{n\delta}$. La règle de décision adoptée ici est d'accepter le modèle équiréparti au niveau 0,95 si $\sqrt{n\delta} \leqslant 1$, soit si 1/2 est dans l'intervalle de confiance au niveau 0,95 de la fréquence f_1 observée (voir sur le cédérom le complément théorique de la fiche sur les sondages du document d'accompagnement de seconde).

2) Il existe en fait un théorème plus général, à savoir : dans le monde théorique défini par une loi $P = (p_1, \ldots, p_k)$, alors la loi de probabilité de la quantité :

$$\chi^2 = n\sum_{i=1}^{k}\frac{\left(f_i - p_i\right)^2}{p_i}$$

est, pour n grand, distribuée selon une loi qui ne dépend que de k. Cette loi s'appelle loi du khi deux à $k - 1$ degrés de liberté et on trouve dans des tables numériques la liste des 9e déciles de ces lois (voir tableau ci-dessous). On peut ainsi généraliser la méthode ci-dessus et définir un critère de compatibilité d'une série de données avec une loi quelconque sur un ensemble fini.

$k - 1$	1	2	3	4	5	6	10	20	30
$\alpha = 0{,}1$	2,71	4,61	6,25	7,78	9,24	10,64	15,99	28,41	40,26
$\alpha = 0{,}05$	3,84	5,99	7,81	9,49	11,07	12,59	18,31	31,41	43,77

Pour $k = 2$, on part d'une loi binomiale et les calculs peuvent se retrouver autrement (voir sur le cédérom, « Compléments aux documents d'accompagnement »).
L'objectif ici n'est pas que les élèves fassent eux-mêmes la simulation, mais qu'ils soient capables de définir une règle de décision et d'exploiter les résultats de simulations.

(Document accompagnement des programmes, Probabilités et statistique Séries ES et S, p. 145-148)

Synthèse des types de tâches dans les manuels français et vietnamien

Groupe	T	Types de Tâches	France	Vietnam
0	T1	Simuler une expérience aléatoire n fois / Etablir une série statistique	x	
	T2	Exécuter un programme pour simuler une expérience aléatoire (pour une machine à calcul)	x	
	T3	Faire un compte rendu de l'expérience	x	
	T4	Recueillir des données dans la vie réelle puis les écrire en tableau		x
1	T5	Lire des informations d'un tableau	x	x
	T6	Lire des informations à partir d'une représentation graphique		x
	T7	Représenter en graphique des données statistiques	x	x
	T8	Représenter/Regrouper des données statistiques en tableau	x	x
	T9	Représenter une série statistique par la boite moustaches	x	
	T10	Représenter la boîte de dispersion associée à une série statistique	x	
	T11	Chercher un exemple de « grandeurs » proportionnelles / non proportionnelles	x	
	T12	Calculer la mesure d'angle d'un diagramme qui représente des données statistiques	x	
	T13	Décrire la statistique (population, caractère étudié, type de caractère)	x	x
	T14	Effectuer un regroupement des données	x	x
	T15	Donner l'allure de la courbe des fréquences cumulées croissantes	x	
	T16	Lire des indicateurs (l'effectif, la moyenne, la médian) dans une représentation graphique	x	
	T17	Indiquer la nature des caractères	x	
	T18	Dresser le tableau des effectifs (à partir des graphiques)	x	
	T19	Dresser le tableau des fréquences (à partir des graphiques)		x
	T20	Trouver la classe modale	x	x
	T21	Établir la distribution des fréquences (à partir des données)	x	x
	T22	Chercher la valeur la plus grande/petite d'une série statistique	x	x
	T23	Chercher la valeur la plus fréquente	x	x
	T24	Construire un diagramme avec un tableur	x	
2	T25	Compléter un tableau (des « grandeurs » proportionnelles)	x	
	T26	Chercher le coefficient de proportionnalité d'un tableau (des « grandeurs » proportionnelles)	x	
	T27	Vérifier si un couple de deux entiers est proportionnel à un autre couple	x	
	T28	Calculer l'effectif	x	
	T29	Calculer des effectifs cumulés	x	
	T30	Constituer un tableau des effectifs cumulés (croissant)	x	
	T31	Constituer un tableau des fréquences cumulées	x	
	T32	Calculer le pourcentage	x	x
	T33	Transformer des répartitions de données en pourcentages	x	
	T34	Calculer la fréquence	x	
	T35	Calculer les fréquences cumulées	x	
	T36	Calculer la moyenne	x	x
	T37	Estimer la moyenne des données regroupées		x
	T38	Calculer une valeur pour ajouter à une liste des valeurs statistiques telles que la moyenne ne change pas	x	
	T39	Expliquer comment on peut calculer la moyenne pour l'ensemble de population	x	
	T40	Calculer la valeur d'un terme manquant d'une suite connaissant la moyenne	x	
	T41	Calculer la valeur d'un terme manquant d'une suite connaissant la médiane	x	

	T42	Calculer la valeur d'un terme manquant d'une suite connaissant l'étendue	x	
	T43	Préciser si la moyenne existe dans un énoncé et elle a une signification	x	
	T44	Calculer la taille d'une série statistique connaissant des hypothèses sur les moyennes	x	
	T45	Préciser la position relative de la moyenne par rapport aux valeurs de série statistique	x	
	T46	Calculer la taille de données statistiques	x	x
	T47	Calculer la médiane	x	x
	T48	Calculer la mode	x	x
	T49	Calculer l'étendue	x	
	T50	Calculer la variance	x	x
	T51	Calculer l'écart type	x	x
	T52	Calculer une valeur à partir le pourcentage et le nombre total des valeurs statistiques	x	
	T53	Compéter un tableau des effectifs sachant certains effectifs et pourcentages	x	
	T54	Déterminer les quartiles	x	
	T55	Déterminer l'intervalle interquartile	x	
	T56	Calculer le pourcentage de données situées dans l'intervalle [x-s; x+s] / [x-2s; x+2s] / [x-3s; x+3s]	x	
	T57	Calculer les bornes de l'intervalle théorique	x	
	T58	Calculer les bornes de l'intervalle de confiance pour la fréquence observée	x	
	T59	Construire une série statistique connaissant la moyenne et/ou la médiane	x	
	T60	Prévoir une évolution des moyennes lors de changement de population	x	
	T61	Chercher deux valeurs extrêmes où se trouve le nombre d'expérience aléatoire pour que la distribution des fréquences soit équirépartie	x	
3	T62	Comparer les moyennes des séries statistiques	x	x
	T63	Comparer les variances des séries statistiques	x	
	T64	Commenter des séries statistiques	x	x
	T65	Comparer la répartition de deux séries statistiques	x	x
	T66	Comparer la distribution des fréquences des échantillons différents	x	
	T67	Comparer les indicateurs statistiques des échantillons différents	x	x
	T68	Comparer des diagrammes différents de mêmes données statistiques	x	
	T69	Comparer deux diagrammes des données statistiques différentes	x	
	T70	Comparer deux séries statistiques	x	
	T71	Comparer les fluctuations d'échantillonnage des séries statistiques	x	
	T72	Comparer une distribution de fréquence à une loi de probabilité		
4	T73	Interpréter des indicateurs statistiques	x	x
	T74	Interpréter les résultats statistiques	x	x
	T75	Donner la pensé à la lecture des résultats de fréquence et d'effectif	x	
	T76	En observant le tableau de données, répondre si la moyenne pourrait être utilisée comme un représentant pour le caractère étudié		x
	T77	Présenter des remarques sur le caractère étudié à partir du tableau de distribution des fréquences de données regroupées en classe		x
	T78	Présenter des remarques sur le caractère étudié à partir de la représentation graphique des fréquences de données regroupées en classe		x
5	T79	Observer la fluctuation d'échantillonnage	x	
	T80	Modéliser et simuler une expérience aléatoire		
	T81	Évaluer les indicateurs d'une population à partir d'un échantillon		

Tableau 132. Les types de tâches dans les manuels français et vietnamiens

Annexe B

Voici les questions que nous nous posons :

- La comparaison des séries statistiques, est-elle un objet d'enseignement au niveau universitaire ?

- Est-ce que ce travail de comparaison prend son sens dans des interprétations et appelle des comparaisons ? Dans quelle mesure, la comparaison entre séries statistiques fait-elle partie des savoirs, des pratiques présentés dans l'enseignement statistique ?

Cela nous conduit à étudier les pratiques de comparaison des séries statistiques dans ce travail. Nous allons analyser des manuels universitaires de première année pour répondre à ces questions. Pour les références, nous prenons les ouvrages universitaires suivants :

CALOT G. [1965]. *Cours de statistique descriptive* (éd. 9è [1992]). Dunod.

DODGE Y. [1999]. Premiers pas en statistique. Springer.

FOURASTIÉ J., LASLIER J.-F. [1997]. *Probabilité et statistique.* Dunod, Paris.

Ce manuel est utilisé dans le cadre l'étude d'économie. Il se compose trois grande parties : Statistique descriptive, Calcul des probabilités, Probabilité et Statistique. À travers la présentation de ce manuel, on voit que la Probabilité est le passage de la Statistique descriptive à la Statistique inférentielle.

LESSARD S., MONGA. [1993]. *Statistique, Concepts et Méthodes avec exercices et corigés.* (M. éditeur, Éd.) Paris-Barcelone-Milan: Les presses de l'université de Montréal, Canada.

MÉOT A. [2003]. Introduction aux Statistiques inférentielles - De la logique à la Pratique - Avec exercices et corrigés. Bruxelles: De Boeck Université.

SPIEGEL M.R. [1981]. *Probabilités et Statistiques, Cours et problèmes* (éd. Theory and Problems of Probability and Statistics). (R. JACOUD, Trad.) McGraw-Hill, Paris.

Nous dégageons trois types de tâches :

T_1. Comparaison de n distributions statistiques à une variable (des distributions statistiques étant données) où les n distributions statistiques sont n sous-populations d'une même population mère exhaustive (une sous-population peut être aussi la population totale ou réduite à un individu).

T_2. Comparaison d'une série statistique à une distribution théorique : exemple d'adéquation à une loi équirépartie.

Pour chaque type de tâche, nous cherchons les techniques utilisables, des exemples pour l'illustrer.

I. Type de tâche T1

Comparaison de n distributions statistiques à une variable (des distributions statistiques étant données) où n distributions statistiques sont n sous-populations d'une même population mère exhaustive (les sous-populations peuvent être lues population total ou réduite à l'individu).

Pour résoudre le type de tâche T_1, les techniques (τ) des solutions possibles sont les suivantes :

$\tau_{1.1}$: Dresser un seul tableau des distributions de n sous-populations avec n colonnes des effectifs et/ou des fréquences, puis comparer ces n colonnes pour comparer ces n sous-populations.

$\tau_{1.2}$: Tracer des représentations graphiques (diagramme, histogramme en assurant la proportionnalité entre des effectifs et des aires ou les courbes de représentation, diagramme en boîte, courbe des fréquences cumulées, courbes de concentration) des distributions de n sous-populations puis comparer visuellement ces représentations graphiques.

$\tau_{1.3}$: Calculer les caractéristiques de tendance centrale comme : moyenne, médiane, mode de n sous-populations puis les comparer.

$\tau_{1.4}$: Calculer les caractéristiques de dispersion comme : étendue, variance, écart-type, coefficient de variation de n sous-populations puis les comparer.

$\tau_{1.5}$: Calculer les moyennes, les écarts types des n sous-populations puis les comparer.

Technique	Avec graphique	Avec calcul
$\tau_{1.1}$		x
$\tau_{1.2}$	x	x
$\tau_{1.3}$		x
$\tau_{1.4}$		x
$\tau_{1.5}$		x

Tableau 133. Techniques pour T1

Nous prenons exercice 3 de l'ouvrage *Probabilité et statistique* (FOURASTIÉ & LASLIER [1997]), comme un exemple du cas de la variable continue pour ce type de tâche :

> On observe deux groupes de 100 personnes atteintes d'une maladie cancéreuse. Les uns ont subi le traitement A et les autres le traitement B. Le temps d'hospitalisation, avant guérison, se répartissent comme suit :

Tableau 7. *Temps d'hospitalisation.*

Temps d'hospitalisation	Effectif de malades	
	Traitement A	Traitement B
0 à 3 mois	2	12
3 à 5 mois	3	18
5 à 6 mois	6	21
6 à 7 mois	10	14
7 à 8 mois	18	12
8 à 9 mois	23	8
9 à 10 mois	18	5
10 à 12 mois	12	6
12 à 18 mois	8	4
Total	100	100

Représenter ces répartitions par le diagramme qui convient. Que peut-on conclure ?

(FOURASTIÉ J. & LASLIER J.-F. [1997], Exercice 3, P. 17)

Remarque sur le sujet donné :

- Les deux sous-populations sont extraites de la population mère des patients de la maladie cancéreuse. Elles ne sont pas choisies par hasard mais suivant les deux traitements A et B. Elles ont la même taille (effectif total) de 100 patients. Dans ce cas, on lit facilement les pourcentages (fréquences) ; par exemple, pour traitement A, l'effectif de classe [0; 3[est 2, donc sa fréquence en pourcentage est 2%. Dans le tableau de distribution des effectifs, les données sont regroupées en classe qui sont les mêmes pour deux sous-populations mais ne sont pas régulières.

Là, on voit les trois choix de l'énoncé : même taille, même classe et les classes irrégulières. Premièrement, nous voyons la discrétisation[211] des données recueillies. Puis, ces remarques sur le sujet nous font penser à poser des questions suivantes :

- Ces deux sous-populations, pourquoi ont-elles la même taille ?
- Pourquoi sont-elles regroupées en mêmes classes ?
- Ces classes, pourquoi ne sont-elles pas régulières ?
- Quel est effet de ces trois choix sur le choix de solution ?
- Le regroupement, pourquoi est-il fait dans l'énoncé au lieu de laisser à faire par le lecteur (l'étudiant) ?
- Quand doit-on regrouper des données en classe ?
- Comment fait-on pour construire des classes ?

Regardons la première classe de sous-population A, pendant trois mois, il y a 2 patients qui ont subit le traitement A. Les données sont regroupées en classe en fonction du temps d'hospitalisation. On peut imaginer qu'un patient a subit le traitement A pendant j jours, h heures, m minute, s seconde …, et un autre pour un temps de j' jours, h' heures, m' minute,

[211] Discrétisation signifie regroupement par classes.

s' seconde (où le temps total de chacun est inférieur à 3 mois) puis le même processus pour les classes restantes. Ainsi, les données dans ce cas sont discrètes. Après le regroupement en durée (classe), on voit que les classes sont consécutives. Les valeurs de la variable statistique deviennent des intervalles de P. Dans ce cas là, les statisticiens parlent de variable continue (ou plutôt les classes sont continues).

Donc, une hypothèse formée est que :

H1 : Le regroupement est un passage du caractère discret de variable statistique à la continuité de variable statistique.

- Le regroupement dans cet exercice nous permet de retirer une règle de construction des classes[212] :

- Quand les effectifs sont trop faibles, on doit regrouper des effectifs en classe.

- Le nombre de classes n'est pas trop grand.

- L'unité de classe est le plus petit (dans cet exercice, l'unité de classe est le mois et n'est pas le jour).

- Les classes ne sont pas toujours régulières et le regroupement en classe dépend au but de résumer des données.

- La question principale est de choisir *le diagramme qui convient, a*lors, en quoi est-ce qu'il « convient » ?

Une recherche sur internet nous fournit des principes pour la discrétisation des données :

Quantile ou Effectifs égaux : chaque classe a le même nombre de valeurs (Nombre total de valeurs / Nombre de classes)

Équivalence ou même amplitude ou égale étendue : Les étendues des classes sont d'égales valeurs (valeur maxi - valeur mini) / Nombre de classes

Progression arithmétique : classes découpées selon une progression arithmétique

Progression géométrique : classes découpées selon une progression géométrique

Standard selon moyenne et écart type : Les classes sont découpées en fonction d'écarts types par rapport à la moyenne.

Manuelle : Seuils observés ou Seuils naturels

Moyennes emboîtées : classes découpées selon leurs "moyennes".

Équiprobabilités : classes découpées selon loi de probabilité.

Méthode JENKS : classes découpées selon des "indices de pertinence.

---

[212] Au secondaire, les classements sont déjà précis.

I.1. Pré-Solution (sans graphique):

En comparant les deux colonnes des effectifs de Traitement A et de Traitement B du tableau au-dessus, on peut voir que la majorité des patients subis traitement A doit être traité pendant longtemps (à partir de 6 mois jusqu'à un an), par contre, la plupart des patients subis traitement B ne doit qu'être traité pendant une durée de moins de 8 mois.

Commentaire :

- Seuls indicateurs lisibles non graphiquement sont mode, quartiles.

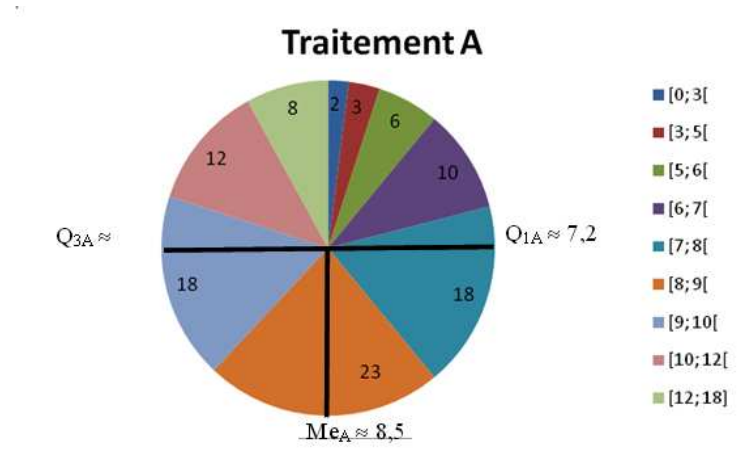
- Même si cette solution ne répond pas à la question de l'énoncé de l'exercice (choisir un graphique convenable), c'est-à-dire elle n'est pas vraiment une solution mais elle joue plus ou moins un rôle intermédiaire ou elle est une étape comme « pré-solution ».

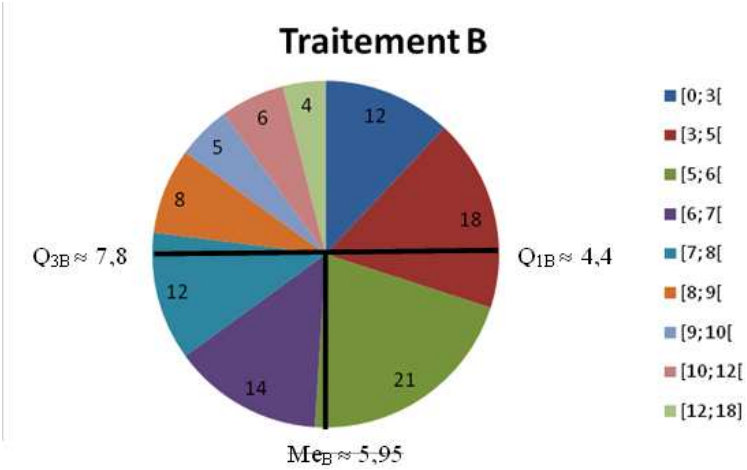
I.2. Solution 1 (avec diagramme figuratif):

Choisir une forme qui porte la proportionnalité entre les effectifs et les surfaces/longueurs (diagramme, rectangle,…) puis tracer les diagrammes figuratifs de deux distributions et les comparer.

Solution 1a : diagramme circulaire

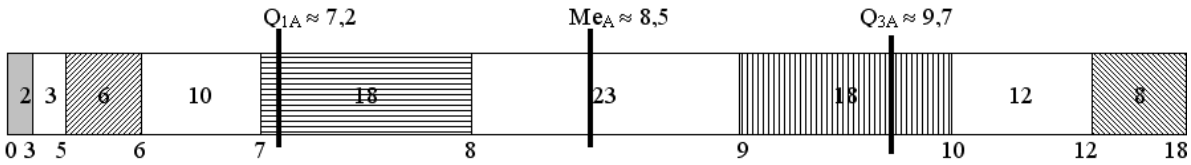
Tracer deux diagrammes circulaires des distributions des effectifs des patients subits traitements A et B



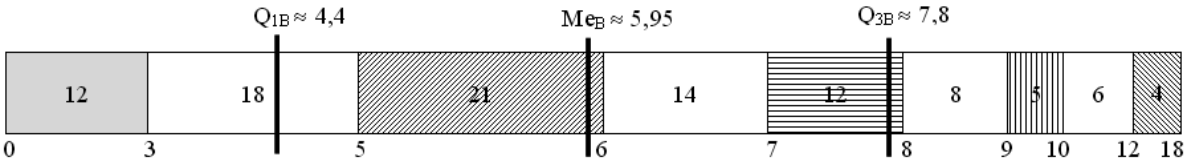


Solution 1b : diagramme en bande

Tracer deux diagrammes en bande des distributions des effectifs des patients subits traitements A et B (hauteur constante)



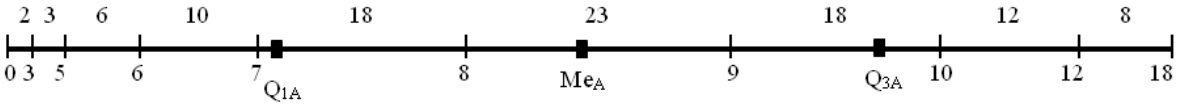
Traitement A



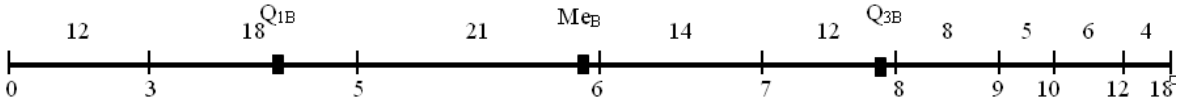
Traitement B

Solution 1c : diagramme en « fil »

Tracer deux « fils » de même longueur puis les partager en segments dont la longueur représente l'effectif de la classe correspondant.



Traitement A



Traitement B

- Classe modale est celle correspond à effectif le plus grand (dans les diagrammes, elle correspond au secteur le plus grand ou bien le rectangle/segment le plus long)

Classe modale A = [8; 9[; Classe modale B = [5; 6[

- Les quartiles sont obtenus par le partage de diagramme en quart :

$Q1_A = 7 + (8-7)/18 * 4 \approx 7,2$;	$Q1_B = 3 + (5-3)/18 * 13 \approx 4,4$;
$Me_A = 8 + (9-8)/23 * 11 \approx 8,5$;	$Me_B = 5 + (6-5)/21 * 20 \approx 5,95$;
$Q3_A = 9 + (10-9)/18 * 13 \approx 9,7$	$Q3_B = 7 + (8-7)/12 * 10 \approx 7,8$

- Médiane différentes : $Me_A \approx 8,5 > 5,95 \approx Me_B$ (médiane est obtenue en prenant les moitiés)

Conclusion :

Un grand nombre de patient qui subit le traitement A se trouve à partir de classe [6; 7[tandis qu'un grand nombre de patient qui subit le traitement B se trouve dans 5 premières classes regroupées (de classe [0 ; 3[à classe [7; 8[ ). C'est-à-dire si on choisit le traitement A, il est probable que l'on devrait le subir pour une plus grande durée de temps par rapport au traitement B.

Commentaire :

Même si les classes du regroupement des données ne sont pas régulières, mais par le calcul des quartiles et de la médiane avec le partage en quart ou en moitié, nous faisons une deuxième hypothèse :

H2 : La répartition dans chaque classe est uniforme.

- Avec les deux diagrammes figuratifs de Solution 1a et Solution 1b (diagramme circulaire et diagramme rectangulaire), même si on peut comparer les surfaces des secteurs/rectangles représentant des pourcentages de l'effectif, préciser les médianes, les quartiles, les classes modales mais le diagramme ne nous permet pas de comparer les variances, les écarts types, … car ces notions n'ont pas de lien avec ce genre de diagramme. C'est-à-dire on les calcule puis on compare ces résultats en nombre mais ils ne prennent pas de signification sur le diagramme.

- A partir du tableau de données, on trace :

  - Solution 1a : le diagramme circulaire en calculant la mesure des angles correspondant (effectifs sont proportionnelles aux angles/surfaces/longueurs des secteurs).

  - Solution 1b : le diagramme en bande en calculant la largeur des rectangles correspondant (les effectifs sont proportionnels aux largeurs/surfaces des rectangles).

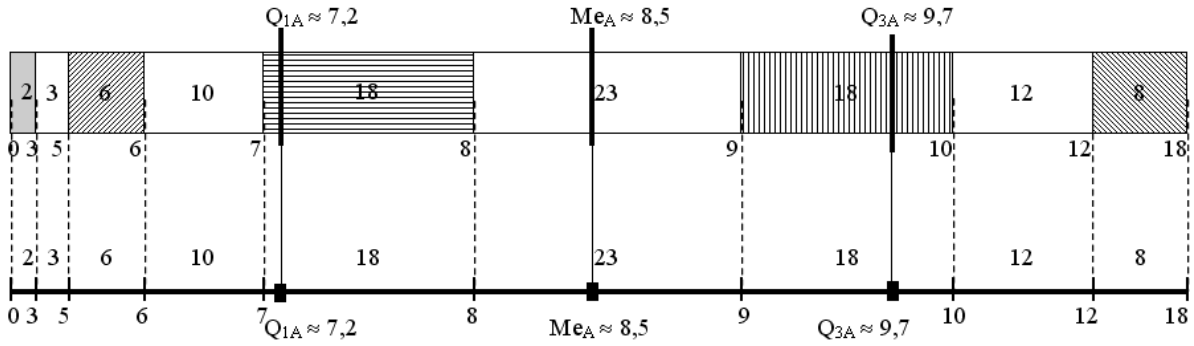
  - Solution 1c : le diagramme en fil en calculant la largeur des segments correspondant (effectifs sont proportionnelles aux largeurs des segments).

Inversement, sans besoin de masse de chaque classe, on peut récupérer les effectifs de chaque classe par la mesure des secteurs/longueurs et par la proportion s'appuyant sur la répartition uniforme dans une classe. Autrement dit, on peut retrouver le tableau de données à partir de ces diagrammes en sachant l'effectif total.

Ainsi, le diagramme figuratif est lié avec le tableau de données de façon réversible.

- Entre le diagramme de Solution 1b et Solution 1c, il y a une relation d'équivalence qui est la longueur du rectangle correspondant à une classe dans Solution 1b est égale à celle du segment représentant la même classe dans Solution 1c.

- En prenant le diagramme en bande de Solution 1b, on réduit sa hauteur à zéro (rectangle devient segment) pour obtenir le diagramme en fil de Solution 1c. Une autre façon à obtenir le diagramme en fil de Solution 1c est de prendre la projection du diagramme en bande dans Solution 1b sur l'axe vertical :



Traitement A

Inversement, en donnant une hauteur comme l'on veut, on peut retracer le diagramme en bande à partir de diagramme en fil. Donc, entre ces deux diagrammes, il y a une relation « aller-retour ».

Le type de diagramme circulaire est le plus habituel pour les élèves car le diagramme est présenté très tôt dans le programme de maths. Même si on doit calculer la mesure des angles, le diagramme peut être le premier choix parmi des graphiques.

À partir de graphique, on peut retrouver facilement la classe modale ; puis la médiane (en traçant le rayon opposé du premier rayon qui marque « 12h » sur le diagramme circulaire ou bien en partageant le diagramme rectangulaire en deux demi-rectangles égaux), et enfin le premier quartile et le troisième quartile (qui peuvent être trouvés en traçant le diamètre perpendiculaire au premier rayon du diagramme circulaire ou bien en partageant chaque demi-rectangle en deux).

Les deux premiers types de diagramme figuratif (circulaire et rectangulaire) s'appuient sur la proportionnalité de la surface et les deux derniers types s'appuient sur la proportionnalité de la longueur. Quoique la taille des classes ne soit pas régulière mais pour les calculs des quartiles, on partage régulièrement la surface de chaque pièce (secteur/rectangle) représentant la classe.

Ces diagrammes ne permettent pas de récupérer la moyenne.

Ici, il y a un problème de retour entre quantitatif et qualitatif : à partir du tableau quantitatif de effectifs, on représente des données en diagramme qualitatif. Mais pour le retour, comment fait-on pour récupérer les effectifs et les fréquences à partir de ce diagramme ? Pour la médiane, les quartiles, c'est facile mais pas pour les autres effectifs, fréquences. Cela coute fort car il faut calculer par la formule :

- Le cas de diagramme circulaire :

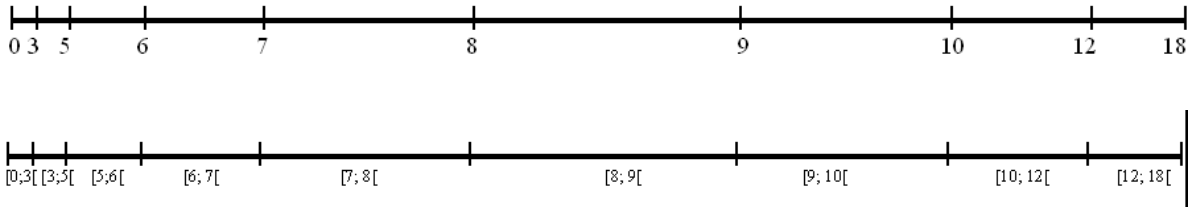
100	→	360°
$n = \alpha \cdot 100 / 360°$	→	α (mesure d'angle du secteur lite grâce à un rapporteur)

- Le cas de diagramme en bande et en fil :

100	→	L (largeur du diagramme rectangulaire)

n = ℓ. 100 / L → ℓ (largeur du rectangle/segment représentant une classe lu grâce à une règle graduée)

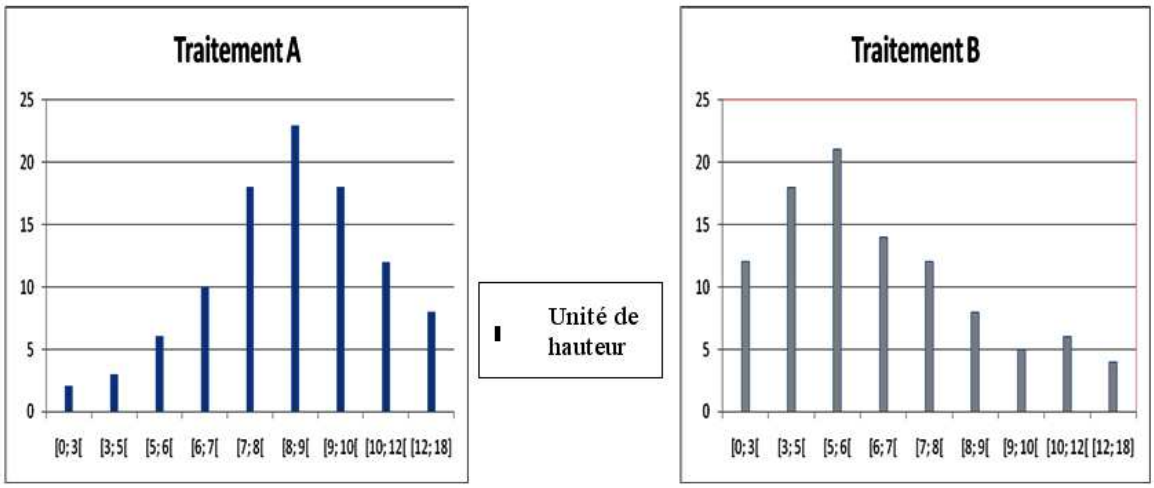
Dans deux diagrammes en bande et en fil, si on place des entiers pour marquer des classes comme des intervalles consécutives sur l'axe de R, cela nous donne une « vision continue ». Par contre, si on code chaque classe par légende d'intervalle semi-ouvert de R, une « vision discrète » est vue ici.



I.3. Solution 2 (avec diagramme) :

Solution 2a (avec diagramme en bâtons) :

- Tracer les diagrammes en bâtons dont les hauteurs représentent les effectifs des patients subits traitements A et B.



Conclusion :

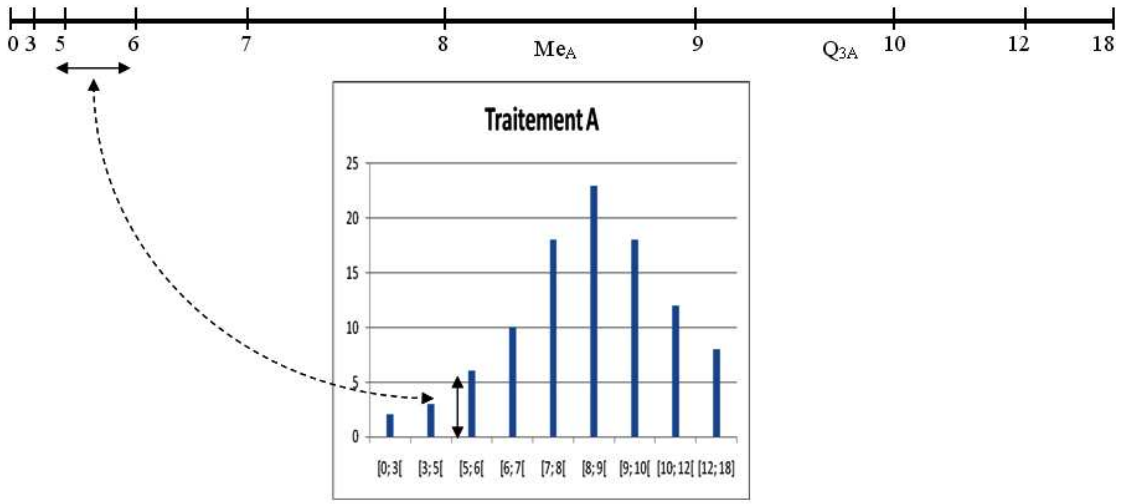
Un grand nombre de patient qui subit le traitement A se trouve à partir de classe [6; 7[tandis qu'un grand nombre de patient qui subit le traitement B se trouve dans 5 premières classes regroupées (de classe [0 ; 3[à classe [7; 8[ ). C'est-à-dire si on choisit le traitement A, il est probable que l'on devrait le subir pour une plus grande durée de temps par rapport au traitement B.

Commentaire :

- Dans ce type de diagramme, les classes sont considérées comme des « étiquettes » pour marquer des périodes du temps et ces étiquettes sont présentées régulièrement et dans l'ordre sur l'axe horizontal. Ainsi, on ne peut que lire les effectifs correspondant à la classe intéressée par la hauteur du bâton.

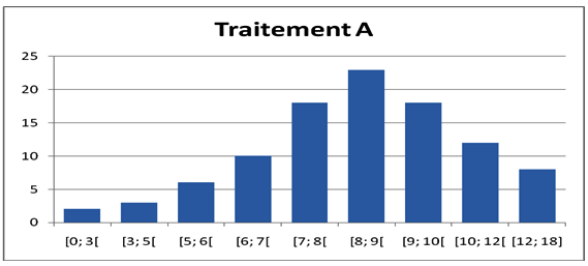
- On peut trouver les quartiles, la médiane à travers ce diagramme en comptant la somme totale des hauteurs puis on les partage en deux pour la médiane ou en quart pour les quartiles. Donc, ce travail est un comptage grâce à l'unité de longueur. Pourtant, on ne peut pas récupérer la moyenne.

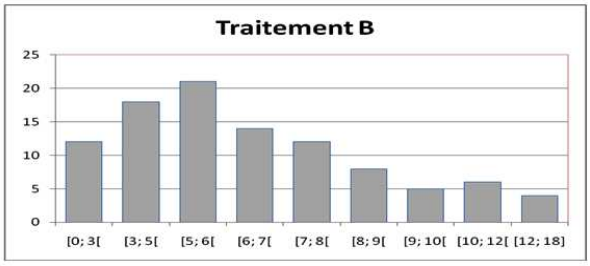
- Si on « coupe » le fil dans la Solution 1c en segment en fonction des classes, ensuit on les met verticalement (rotation d'angle droit dans le sens trigonométrique (i.e. inverse des aiguilles d'une montre)), enfin on pose ces segments de gauche à droit régulièrement, on obtiendra le diagramme en bâton. Inversement, quand on range respectivement et consécutivement tous les bâtons sur une droite verticale, on peut reformer le diagramme en fil à partir du diagramme en bâtons.



Solution 2b (diagramme en rectangles isolés)

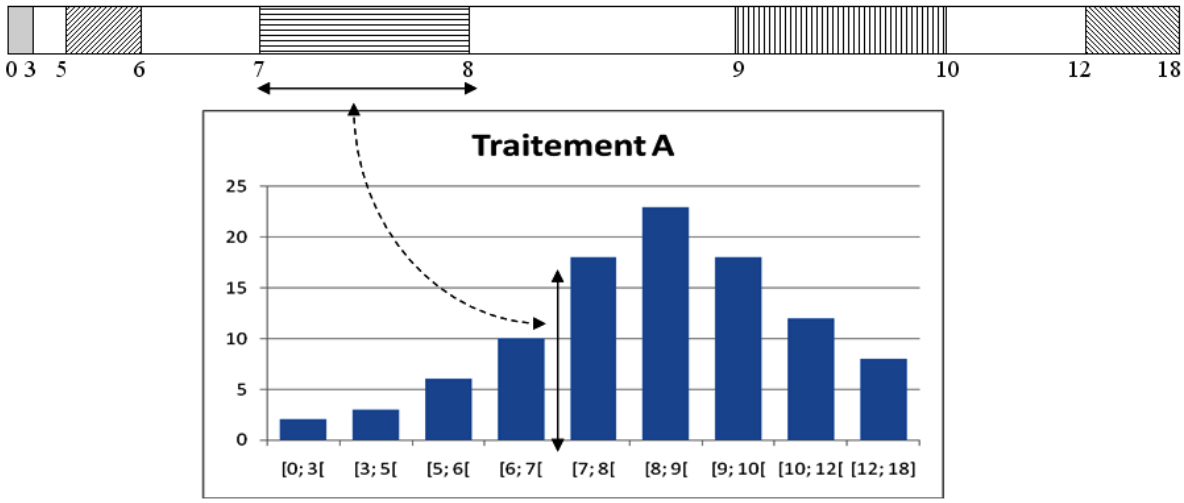
En remplaçant des bâtons dans Solution 2a par des rectangles de même largeur, on obtiendra diagramme en rectangles isolés.





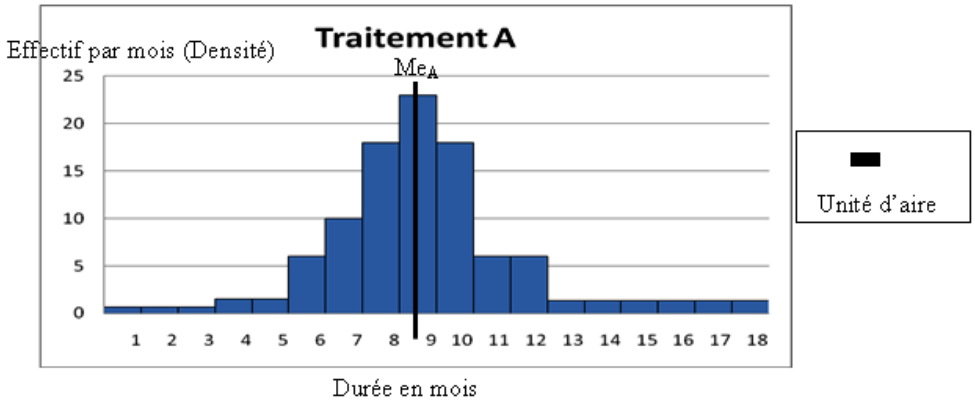
Commentaire :

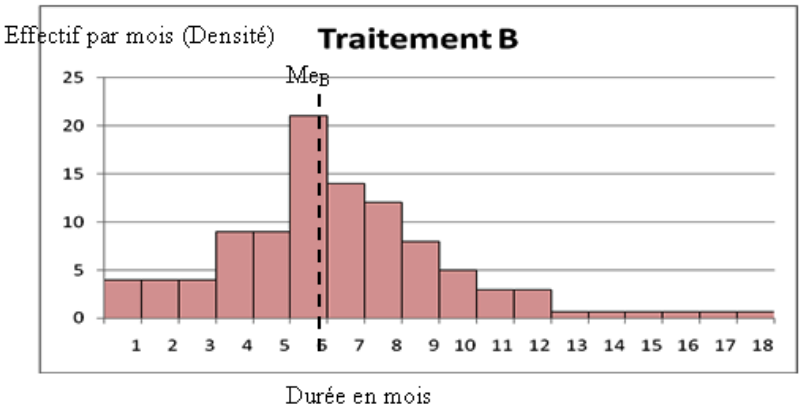
Ce diagramme en rectangles isolés peut être obtenu par le même passage du diagramme en fil au diagramme en bâtons : on refait en prenant le diagramme en bande pour passer au diagramme en rectangle isolés. Et inversement, à partir du diagramme en rectangles isolés, on retrouve le diagramme en bande en tournant et en mettant les rectangles en ligne consécutivement et respectivement.



I.4. Solution 3 (avec histogramme):

Tracer l'histogramme représentant des distributions des effectifs des patients subits traitements A et B.





Calculer la classe modale, médiane, étendue de deux sous-populations :

- Classe modale est celle ayant la hauteur la plus grande :

  Classe modale A = [8; 9[; Classe modale B = [5; 6[

- Médiane est trouvée grâce à l'unité d'aire : $Me_A \approx 8{,}5 > Me_B \approx 5{,}95$

- Étendu est la différence de la valeur la plus grande et la plus petite des durées en mois : Etendu A = Etendu B = 18 – 0 = 18

- MA ≈ 8,655 > MB ≈ 6,28

- σA ≈ 1,36 < σB ≈ 1.44

Conclusion :

Les patients subirent le traitement A pour une durée plus longue que ceux subits le traitement B. La sous-population A est un peu moins dispersée que celle B.

Commentaire :

Dans cet histogramme, l'aire est égale à l'effectif (aire totale est 100); la hauteur est la densité d'une classe et n'est pas constante comme dans le cas de diagramme en bande. Ce type de graphique cache en arrière la fonction de densité de la distribution.

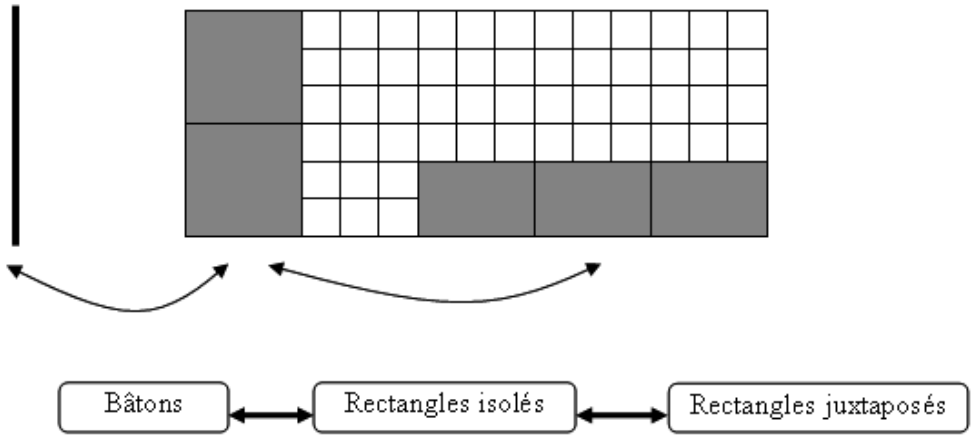
Histogramme peut être un bon candidat pour le choix de graphique car on peut comparer la forme de l'histogramme, placer la médiane par le comptage d'unité d'aire, les quartiles, … puis les comparer graphiquement.

Comme le diagramme, l'histogramme est lié directement au tableau de données.

Par estimation, on peut retrouver sur l'histogramme la médiane, les quartiles sans calculer grâce à l'unité d'aire. Mais l'histogramme ne nous permet pas de récupérer la moyenne.

À partir du diagramme en rectangles isolés, en prenant un mois comme l'unité, on transforme le rectangle représentant une classe à un nouveau rectangle de même aire mais sa nouvel largueur obtenue par le produit de l'amplitude de classe et la taille d'une unité ; et sa nouvel hauteur obtenu par le quotient l'ancien hauteur par sa nouvel largueur. Ce fait respecte la proportion des classes. Enfin, tous les nouveaux rectangles sont posés côté à côté pour former l'histogramme. Ainsi, les rectangles « isolés » deviennent les rectangles « juxtaposés ». Le retour est possible en faisant le calcul contraire.

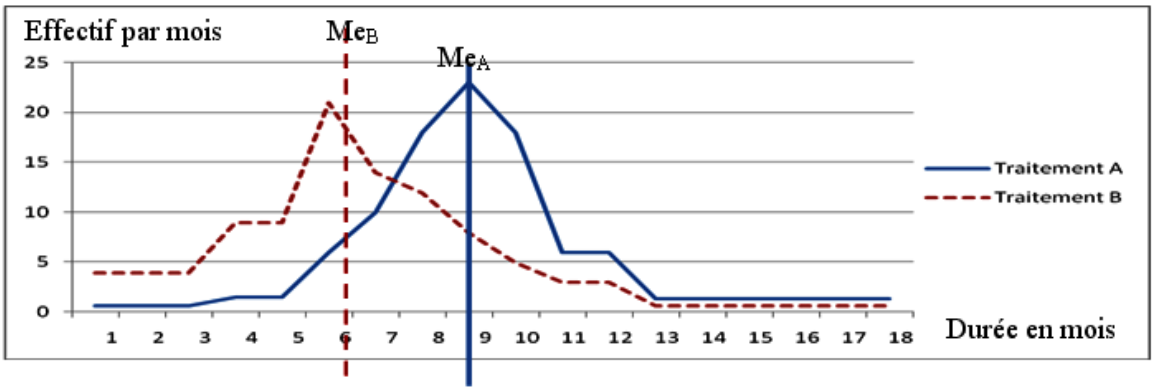
Par exemple : le rectangle de dimension 2 x 1 qui représente classe [0 ; 3[(dont l'amplitude est 3-0=3) avec l'effectif 2 dans le diagramme en bâtons est transformé au rectangle de dimension 2/3 x 3 dans l'histogramme.



Un passage possible entre Solution 1b et Solution 3 :

I.5. Solution 4 (avec courbe représentative de la densité):

- Tracer la courbe représentative des distributions des effectifs (la courbe représentative de la fonction de densité) des patients ayant subit traitements A et B



- Calculer moyenne, mode, médiane de deux sous-populations :

  - MA ≈ 8,655 > MB ≈ 6,28

  - Médiane différentes : $Me_A \approx 8,5 > Me_B \approx 5,95$

  - Classe modale A = [8; 9[; Classe modale B = [5; 6[

Conclusion :

La courbe de B est étalée à droite tandis que celle de A est assez symétrique (les trois indicateurs de A sont presque égaux). Les patients subirent le traitement A pour une durée plus longue que ceux subits le traitement B. La sous-population A est un peu moins dispersée que B.

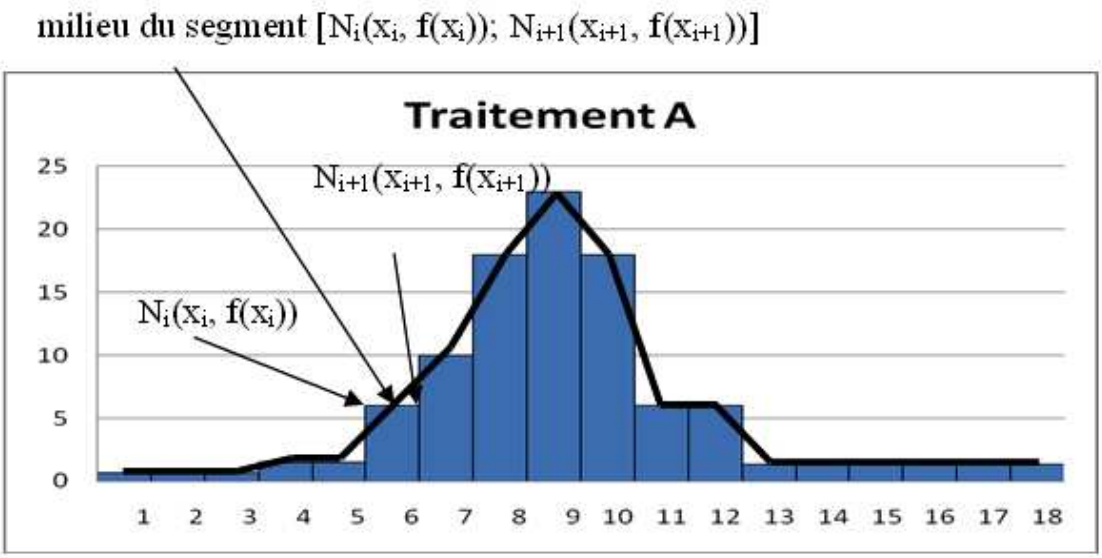
Commentaire :

- La courbe représente un modèle continu de la fonction de densité (cas de données chronologies ou considérées comme continues mais discrétisées).

- Une troisième hypothèse est posée :

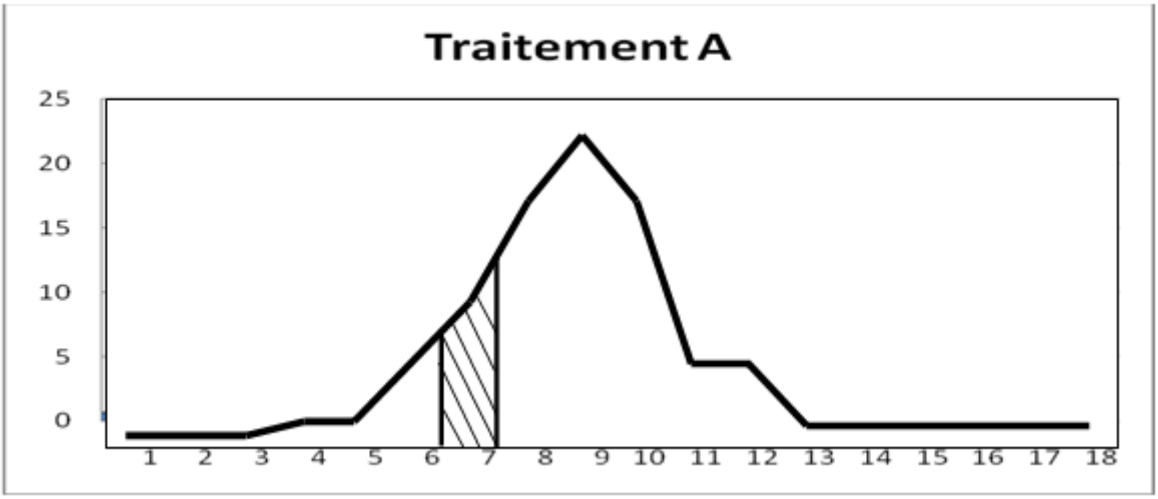
- On choisit le mois comme unité parce qu'il est l'unité le plus petit des classes.

- A partir du graphique de Solution 3, on peut obtenir celui de *Solution 4* en joignant les points milieux des segments $[N_i(x_i, f(x_i)); N_{i+1}(x_{i+1}, f(x_{i+1}))]$ qui forment des rectangles dans l'histogramme dans Solution 2.



- Pourquoi doit-on prendre les milieux à joindre ? C'est que cette façon assure graphiquement la conservation d'aire. En plus, … ?

- À partir de ce graphique, il faut lire les aires sous la courbe qui sont les effectifs.



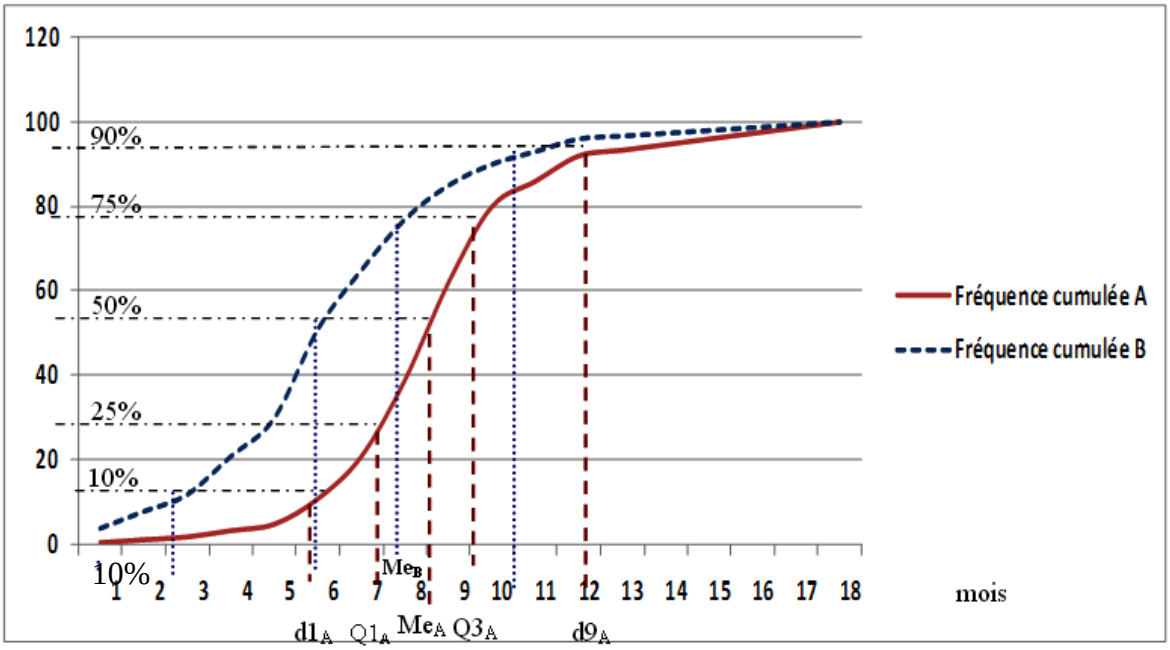
La plus petite unité correspond au plus grand effectif.

I.6. Solution 5 (avec la courbe cumulative) :

- Calculer les fréquences cumulées.

- Tracer la courbe cumulative (la courbe de la fonction de répartition) représentant des distributions des effectifs des patients subits des traitements A et B.

-



- Par la lecture graphique, on peut préciser approximativement les médianes, les quartiles, les intervalles portant 90% des effectifs, … de deux distributions des effectifs, puis les comparer.

- Min A = Min B = 0

- Max A = Max B = 18

- $Q1_A \approx 7{,}2 > Q1_B \approx 4{,}4$;

- $Me_A \approx 8{,}5 > Me_B \approx 5{,}95$;

- $Q3_A \approx 9{,}7 > Q3_B \approx 7{,}8$

Conclusion :

Les patients subirent le traitement A pour une durée plus longue que ceux subirent le traitement B.

Commentaire :

- Ce graphique porte des propriétés de la fonction analytique dont un des intérêts est la lecture des coordonnés d'un point appartenant à la courbe : en sachant son ordonné (fréquence) on peut retrouver son abscisse (mois) et contrairement, on obtient son ordonné en sachant son abscisse. C'est pourquoi ce graphique le plus facile à récupérer des fréquences cumulées, des quartiles (mais pas pour la mode, moyenne,…).

Le point d'inflexion de la courbe est la mode.

I.7. Solution 6 (avec boite moustache):

- Calculer Min, Max, Q1, Médiane, Q3, d1, d9 (reprendre les calculs des solutions précédents):

- Min A = Min B = 0

- Max A = Max B = 18

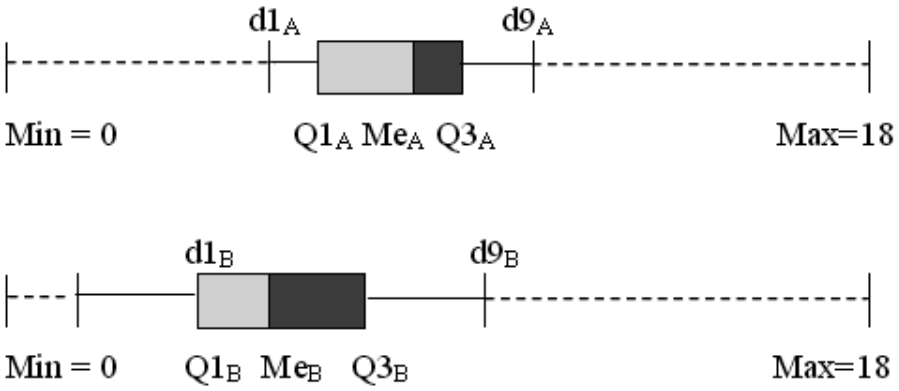
- $Q1_A \approx 7,2$; $Me_A \approx 8,5$; $Q3_A \approx 9,7$

- $Q1_B \approx 4,4$; $Me_B \approx 5,95$; $Q3_B \approx 7,8$

- $d1_A = 5 + (6-5)/6 * 5 \approx 5,8$; $d9_A = 10 + (12-10)/12 * 10 \approx 11,7$

- $d1_B = 0 + (3-0)/12 * 10 \approx 2,5$; $d9_B = 10$

- Tracer la boite moustache représentant des distributions des effectifs des patients subits traitements A et B.



Conclusion :

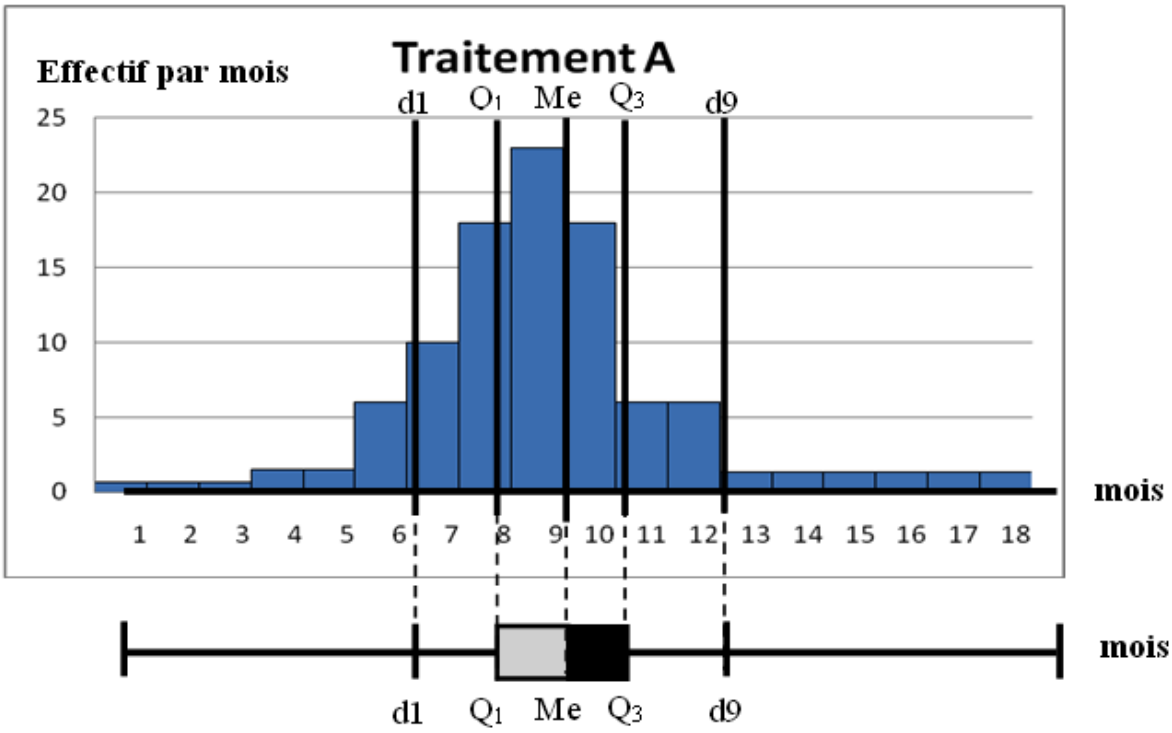
Les patients subirent le traitement A pour une durée plus longue que ceux subirent le traitement B.

Commentaire :

- Le travail avec le diagramme en boite n'est pas directement relié au tableau de données et on doit calculer les quartiles, les valeurs extrêmes, les deux déciles d1 et d9

- Ce fait est une représentation graphique d'un résumé de données.

- A partir de la projection des indicateurs comme Max, Min, les quartiles Q1, Me, Q3, les déciles d1 et d9 du histogramme dans *Solution 3* sur l'axe horizontal (mois), on peut retrouver le diagramme en boite.



Solution 7 (avec la courbe de concentration):

- Calculer les fréquences cumulées, les masses, pi et qi; puis tracer la courbe de concentration.

Temps	Centre	Effectif A	Effectif cumulé A	Masse	Masse cumulé	pi	qi
	ck	nk	nk cumulé	nk*ck	nk*ck cumulé	0	0
[0; 3[	1,5	2	2	3	3	0,02	0,00346
[3; 5[	4	3	5	12	15	0,05	0,01731
[5; 6[	5,5	6	11	33	48	0,11	0,0554
[6; 7[	6,5	10	21	65	113	0,21	0,13041
[7; 8[	7,5	18	39	135	248	0,39	0,28621
[8; 9[	8,5	23	62	195,5	443,5	0,62	0,51183
[9; 10[	9,5	18	80	171	614,5	0,8	0,70917
[10; 12[	11	12	92	132	746,5	0,92	0,86151
[12; 18]	15	8	100	120	866,5	1	1
		100		866,5			

Tableau 134. Les calculs du traitement A

Temps	Centre	Effectif B	Effectif cumulé B	Masse	Masse cumulé	pi	qi
	ck	nk	nk cumulé	nk*ck	nk*ck cumulé	0	0
[0; 3[	1,5	12	12	18	18	0,12	0,02866
[3; 5[	4	18	30	72	90	0,3	0,14331
[5; 6[	5,5	21	51	115,5	205,5	0,51	0,32723
[6; 7[	6,5	14	65	91	296,5	0,65	0,47213

[7; 8[	7,5	12	77	90	386,5	0,77	0,61545
[8; 9[	8,5	8	85	68	454,5	0,85	0,72373
[9; 10[	9,5	5	90	47,5	502	0,9	0,79936
[10; 12[	11	6	96	66	568	0,96	0,90446
[12; 18]	15	4	100	60	628	1	1
		100		628			

Tableau 135. Les calculs du traitement B

ck : centre de classe

nk : effectif de la classe

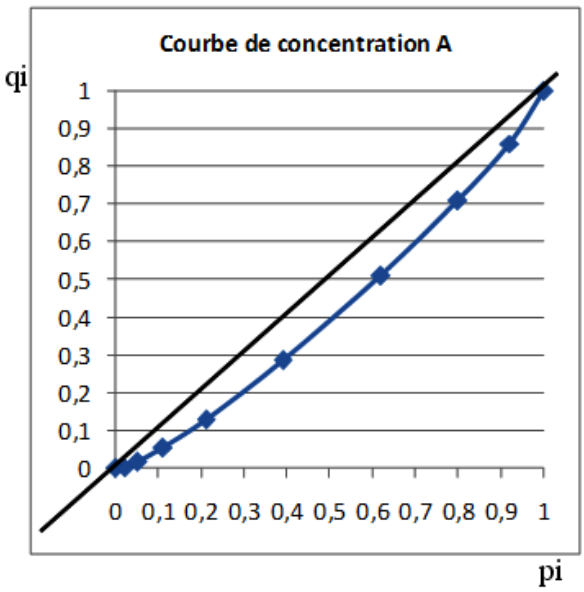
nk cumulé : effectif cumulé des classes

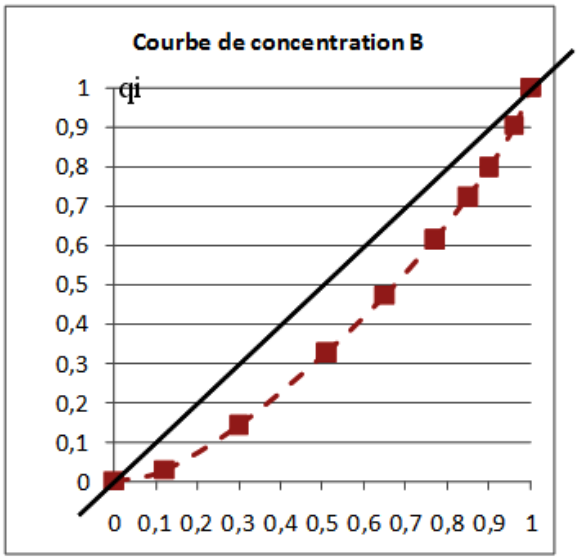
nk * ck : masse

nk * ck cumulé : masse cumulée

pi : fréquence cumulée

qi : fréquence des masses cumulées





Conclusion :

L'observation de ces courbes nous permet de dire que la distribution des effectifs des patients subits le traitement A est à faible concentration par rapport à celle B.

Conclusion :

1. Pour comparer des sous-populations regroupées en classe, il faut les mêmes classes pour des données. Il y a des règles cachées pour la construction des classes :

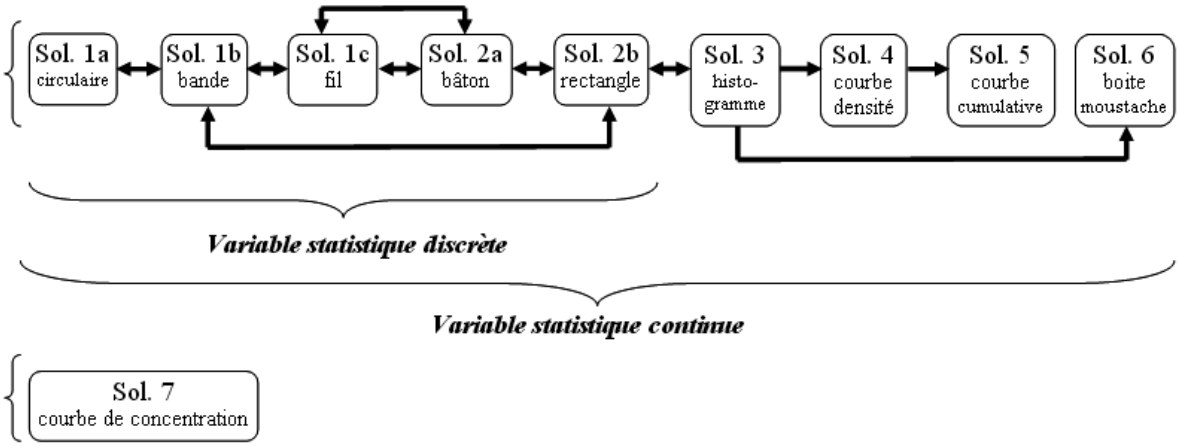
- Quand l'effectif n'est pas très grand, on doit regrouper des effectifs en classe.

- Le nombre de classes n'est pas trop grand.
- L'unité de classe est la plus petite classe en général.
- La plus petite unité correspond à d'effectif le plus grand.
- Les classes ne sont pas toujours régulières et le regroupement en classe dépend à

2. Le diagramme figuratif ne permet pas de comparer complètement des distributions statistiques parce qu'en observant les diagrammes, on ne voit pas la dispersion, … Et les autres, non plus.

3. Les solutions 2, 3, 4 avec l'histogramme, la courbe représentative, la courbe densité bénéficient des outils d'analyse de l'étude des fonctions. Par exemple, on peut récupérer les valeurs de la médiane, des quartiles, … comme des abscisses des points sur la courbe sachant leurs ordonnés (50%, 25%, 75%,…) ; par contre, le diagramme circulaire n'ayant pas de cet intérêt.

4. Le schéma partageant les solutions en 3 groupes suivant leurs implications :



Solution 1a, 1b, 1c, 2a, 2b, 3, 4, 5, 6 : ont des relations (un sens unique ou sens aller-retour).

Solution 7 : courbe de concentration

Nous faisons un tableau de synthèse de présence d'absence des graphiques dans les ouvrages universitaires référenciés :

Solutions / **Manuels universitaires**	S.1a diagramme circulaire	S.1b diagramme en bande	S.1c diagramme en fil	S.2a diagramme en bâton	S.2b diagramme en rectangles isolés	S.3 histogramme	S.4 courbe représentative de densité	S.5 courbe cumulative	S.6 boite de moustache	S.7 courbe de concentration
Calot G. [1965]	x			x		x	x	x	x	x
DODGE Y. [1999]	x				x	x	x		x	
FOURASTIÉ J., LASLIER J.-F. [1997]	x	x		x		x	x	x		
LESSARD S., MONGA. [1993]	x	x			x	x			x	
MÉOT A. [2003]				x		x	x	x		
SPIEGEL M.R. [1981]				x	x	x	x	x		

Tableau 136. Synthèse des solutions de comparaison deux traitements

D'après ce tableau de synthèse, Solution 2c n'existe pas dans les manuels référenciés.

5. Mais aucune représentation graphique ne nous permet de récupérer la moyenne.

Voici la réponse donnée par le livre :

> Le diagramme qui convient est un histogramme. Il s'agit de comparer deux distributions d'un caractère quantitatif continu (temps d'hospitalisation). On peut tracer les deux histogrammes sur un même graphique et les comparer.
>
> La distribution B est représentée à gauche de la distribution A : elle correspond à des temps d'hospitalisation plus courts. Le traitement B est le plus rapide que le traitement A (voir fig. 9, p. 174)

(FOURASTIÉ J., LASLIER J.-F. [1997], p. 173)

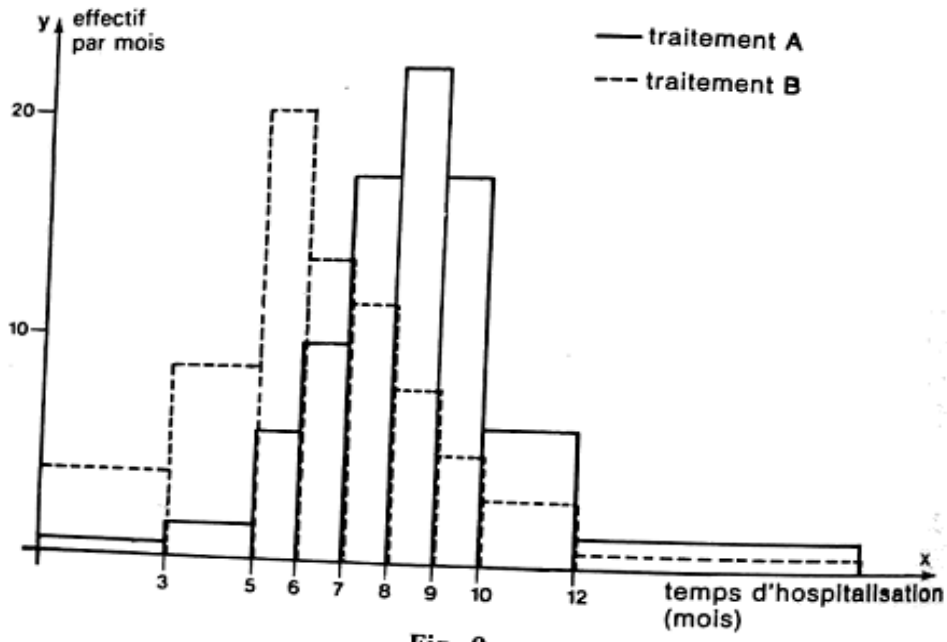


Fig. 9.

(FOURASTIÉ J., LASLIER J.-F. [1997], p. 174)

• Les objets de statistique qui ont un avenir dans les probabilités :

- fréquence-probabilité : par la loi des grands nombres, la probabilité est considérée comme la limite des fréquences.

- graphique (solution 3, 4, 5)

- variable statistique - variable aléatoire

- tableau de distribution (valeur des variables f_i) – la loi de probabilité

- densité de fréquence – densité de probabilité

- fonction cumulative – fonction de répartition

• Lien entre statistique et probabilité :

La praxéologie de comparaison est un bon candidat pour le lien entre les statistiques et les probabilités.

II. Type de tâche T2

Comparaison d'une série statistique à une distribution théorique : adéquation à une loi équirépartie.

Pour les deux lois concrètes : loi uniforme et loi normale, les techniques correspondantes sont présentées comme suivant :

$\tau_{2.1}$: Test du Khi-deux (pour la loi uniforme) :

Alain MÉOT [2003][213] propose la technique générale du Test du X^2 d'adéquation à une distribution théorique :

Le test dans le cas général

Étape 1

H_0 : Une variable aléatoire X suit une distribution donnée $\Leftrightarrow$ la probabilité π_j que X prenne le résultat x_j ($j = 1, \ldots, k$; où k est le nombre de résultats possibles considérés) est connue et fixée.

H_1 : La variable aléatoire ne suit pas la distribution spécifiée en H_0 $\Leftrightarrow$ Au moins une des probabilités spécifiées en H_0 est incorrecte.

Étape 2

$$t = \sum \frac{(\text{effectifs observés} - \text{effectifs attendus})^2}{\text{effectifs attendus}} = \sum_{j=1}^{k} \frac{(n_j - N\pi_j)^2}{N\pi_j}$$

Calcul de t sur un échantillon.

Étape 3

$p = p_{H_0}(T \geq t)$ où T suit une loi du χ^2 à $k - k' - 1$ ddl. k' est ici le nombre de paramètres à estimer pour pouvoir spécifier complètement la distribution supposée en H_0

Étape 4

Rejet de H_0 si $p \leq \alpha$.

(MÉOT [2003], p. 117-118)

$\tau_{2.2}$: Droite de Henry[214] (méthode graphique pour la loi normale)

1. Tracer la courbe des fréquences cumulées d'un *n*-échantillon au hasard extrait de la population dans laquelle on veut vérifier si la distribution de Y est Gaussienne, tracée sur un papier graphique spécial dit "papier Gausso-arithmétique".

2. Tracer des points ayant pour abscisse une valeur *y* observée de la variable d'intérêt, pour ordonnée la fréquence *cumulée* correspondante observée dans l'échantillon, $\frac{k}{n}$, où *k* est l'effectif des individus de l'échantillon qui ont une valeur inférieure ou égale à la valeur *y* (autant de points que l'on souhaite)

3. Si les points obtenus sont *sensiblement alignés*, on conclut qu'il n'y a aucune raison de rejeter l'hypothèse de normalité dans la population.

[213] MÉOT A. (2003). Introduction aux Statistiques inférentielles - De la logique à la Pratique - Avec exercices et corrigés. Bruxelles: De Boeck Université.

[214] Source : http://www.inrialpes.fr/is2/people/girard/Enseign/STA230/tp7/STA230TP7.pdf

4. Si les points obtenus ne sont *pas sensiblement alignés*, on conclut qu'on rejette l'hypothèse de normalité dans la population.

Particulièrement, lors de la normalité de la distribution, on peut estimer graphiquement :

5. La moyenne par sa médiane, à l'abscisse de l'intersection de la droite de Henry avec l'horizontale à la fréquence de 50%.

6. L'écart-type par la demi-amplitude de l'intervalle des abscisses de.ses intersections avec les horizontales aux fréquences de 0,1587 et 0,8413 (en pointillés sur le graphique).

Un exemple[215] :

Lors d'un examen noté sur 20, on obtient les résultats suivants :

- 10% des candidats ont obtenu moins de 4
- 30% des candidats ont obtenu moins de 8
- 60% des candidats ont obtenu moins de 12
- 80% des candidats ont obtenu moins de 16

Déterminer si la distribution des notes est gaussienne, et, si oui, ce que valent son espérance et son écart type ?

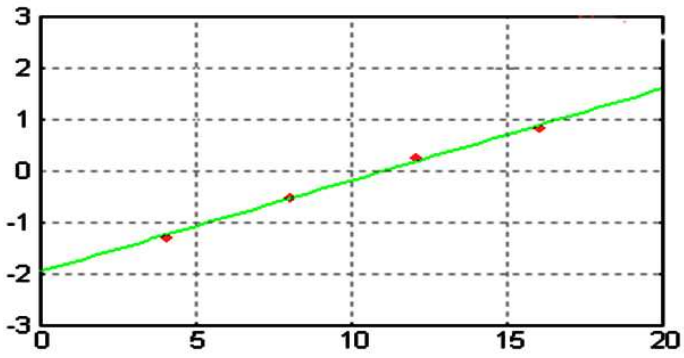
Solution :

On connaît donc 4 valeurs x_i, et, pour ces 4 valeurs, on connaît $P(X < x_i)$.
En utilisant la table de la fonction de répartition de la loi normale centrée réduite, on détermine les t_i correspondants :

x_i	$P(X<x_i) = \Phi(t_i)$	t_i
4	0,10	-1,28
8	0,30	-0,525
12	0,60	0,255
16	0,80	0,84

On trace les points de coordonnées (x_i ; t_i).

[215] Source : http://www.techno-science.net/?onglet=glossaire&definition=5951



Les points paraissent alignés ; la droite coupe l'axe des abscisses au point d'abscisse 11 et le coefficient directeur est (0,84 +1,28)/12 environ, ce qui donnerait un écart type de 12/2,12 = 5,7.

Cela laisse penser que la distribution est gaussienne de paramètres *m*, σ^2, où *m* = 11 et σ = 5,7.

$\tau_{2..3}$: Test de Shapiro-Wilk[216] (pour la loi normale, basé sur la comparaison la distribution cumulative à une distribution normale ajustée aux observations)

1. Classer les n observations par ordre de grandeur croissante : $x_1 \leq x_2 \leq \ldots \leq x_j \leq \ldots \leq x_{n-1} \leq x_n$

2. Calculer la Somme des Carrés des Écarts: $SCE = \sum (x_i - \bar{x})^2 = (n-1)S^2$

3. Calculer les différences :

$$d_1 = x_n - x_1$$
$$d_2 = x_{n-1} - x_2$$
$$d_3 = x_{n-2} - x_3$$
$$\ldots$$

Si n est pair il y aura alors n/2 différences.

Si n est impair il y aura alors (n-1)/2 différences, l'observation médiane ne sera pas utilisée.

4. Calculer : $b = \sum a_i d_i$ (les coefficients a_i sont donnés dans une table en fonction de n et i)

5. Calculer : $W = \dfrac{b^2}{SCE}$

6. Comparer W à $W_{1-\alpha,\, n}$

[216] Source http://www.cjonquiere.qc.ca/begin/ZEAA2005/TestShapiroWilk.htm

$W_{1-\alpha,\, n}$ est trouvé dans la table de Shapiro-Wilk en fonction du risque d'erreur α et de la taille de l'échantillon (le nombre d'observations) n

si $W < W_{1-\alpha,\, n}$ la distribution NE SUIT PAS UNE LOI NORMALE

si $W \geq W_{1-\alpha,\, n}$ la distribution SUIT UNE LOI NORMALE

Un exemple :

Exemple : On a fait des essais de fatigue sur un certain matériau (nombre de cycles avant rupture) et on a obtenu la série suivante : 31, 39, 62, 89, 115, 125, 140, 225, 251, 270, 342, 400, 442, 580, 850. Peut-on conclure avec un risque d'erreur de 5% (niveau de confiance 95%) que ces données proviennent d'une distribution suivant une loi normale ?

Solution :

1. On place les données en ordre croissant :

31, 39, 62, 89, 115, 125, 140, 225, 251, 270, 342, 400, 442, 580, 850

2. On calcule la somme des carrés des écarts : (NB : $\bar{x} = 264{,}0\overline{6}$)

$$
\begin{aligned}
SCE &= \sum (x_i - \bar{x})^2 \\
&= (31 - 264{,}0\overline{6})^2 + (39 - 264{,}0\overline{6})^2 + (62 - 264{,}0\overline{6})^2 + (89 - 264{,}0\overline{6})^2 \\
&+ (115 - 264{,}0\overline{6})^2 + (125 - 264{,}0\overline{6})^2 + (140 - 264{,}0\overline{6})^2 + (225 - 264{,}0\overline{6})^2 \\
&+ (251 - 264{,}0\overline{6})^2 + (270 - 264{,}0\overline{6})^2 + (342 - 264{,}0\overline{6})^2 + (400 - 264{,}0\overline{6})^2 \\
&+ (442 - 264{,}0\overline{6})^2 + (580 - 264{,}0\overline{6})^2 + (850 - 264{,}0\overline{6})^2 \\
&= 734482{,}93
\end{aligned}
$$

3. On calcule les différences d_i

Différences	di
$x_{15} - x_1 = 850 - 31$	d1=819
$x_{14} - x_2 = 580 - 39$	d2=541
$x_{13} - x_3 = 442 - 62$	d3=380
$x_{12} - x_4 = 400 - 89$	d4=311
$x_{11} - x_5 = 342 - 115$	d5=227
$x_{10} - x_6 = 270 - 125$	d6=145
$x_9 - x_7 = 251 - 140$	d7=111
$x_8 = 225$	Aucune : puisque n est impair la valeur médiane n'est pas utilisée

4. On calcule la valeur de b.

Pour ce calcul nous avons besoins des coefficients a_i de la table Shapiro-Wilk pour n = 15.

ai si n =15	di	
a1=0,5150	d1= a3=0,2495	d3=380
a3=0,2495	d3=380	
a4=0,1878	d4=311	
a5=0,1353	d5=227	
a6=0,0880	d6=145	
a7=0,0433	d7=111	

$$b = \sum a_i d_i = 0{,}5150x819 + 0{,}3306x541 + 0{,}2495x380 + 0{,}1878x311$$
$$+ 0{,}1353x227 + 0{,}0880x145 + 0{,}0433x111 = 802{,}1348$$

$$W = \frac{b^2}{SCE} = \frac{(802{,}1348)^2}{734482{,}93} = 0{,}876$$

5. On calcule

6. On compare W à $W_{1-\alpha,n}$

Avec α=5% et n = 15 on trouvera dans la table le $W_{95\%,15}$= 0,881

Puisque 0,876 < 0,881 on a donc W < $W_{1-\alpha,n}$ et par le fait même, la distribution ne suit pas une loi normale avec un risque d'erreur de 5%.

$\tau_{2.4}$: Test de Kolmogorov-Smirnov (pour la loi normale (loi continue) →résulté d'une adaptation de test Kolmogorov-Lilliefors)

1. Classement des données brutes par ordre croissant.

2. Centrage-réduction des données brutes (on obtient les z_i).

3. Recherche, dans la table III de Scherrer (p. 765) de l'aire de la courbe normale centrée-réduite correspondant aux valeurs z_i ci-dessus.

4. Calcul des fréquences observées cumulées.

5. Calcul des différences absolues en + (col. 5) et en – (col. 6).

6. Comparaison de la plus grande valeur trouvée dans les colonnes 5 et 6 avec le D critique D_a (formule donnée en p. 2 de ce document).

Annexe

Pratique d'un test d'ajustement du Khi-deux [217]

Décrivons maintenant les différentes étapes à réaliser pour ajuster des données expérimentales à une loi continue de probabilité par la méthode du Khi-deux.

On suppose que ces données sont les valeurs prises par un caractère *C* continu sur un échantillon aléatoire de taille n prélevé dans une population *P*.

6 Mise en place d'un modèle probabiliste

1) Considérer une variable aléatoire X à valeurs dans un intervalle réel I, modélisant les variations de C dans la population P. Une interprétation probabiliste du caractère C permet de conjecturer le type de la loi inconnue P_X de X.

2) Expliciter le paramètre q (éventuellement multidimensionnel) déterminant la loi P_X. Les valeurs numériques fixant θ et déterminant une loi P_θ sont :

   - soit obtenues par r estimations à partir de l'échantillon donné,

   - soit choisies a priori (r = 0).

3) Partager l'intervalle I en k intervalles I_i :

   - soit d'amplitudes égales (sauf éventuellement pour les extrémités) et calculer les probabilités $p_i = P_\theta(X \in I_i)$,

   - soit de probabilités $p_i = 1/k$ égales et déterminer les quantiles ai successifs ($P_\theta(X < a_1) = 1/k$, $P_\theta(a_1 < X < a_2) = 1/k$, …) de la loi P_θ délimitant les intervalles I_i.

4) Vérifier que pour tous les i de 1 à k, on a np_i (1 - p_i) ³ 5 (éventuellement regrouper des intervalles Ii contigus).

   Les p_i constituent alors la loi discrète finie (p), sous-modèle probabiliste pour représenter la distribution de fréquences des valeurs de C dans les différentes modalités Mi définies dans la population P par les intervalles I_i.

- Hypothèses à tester et mise en œuvre du test

5) H_0 : « La loi P_θ est un modèle satisfaisant et la distance entre (p) et la distribution (f) des fréquences des modalités Mi au sein de l'échantillon observé, n'est dû qu'aux fluctuations d'échantillonnage ».

   H_1 : « Le hasard du prélèvement de l'échantillon ne peut expliquer à lui seul l'écart observé entre (p) et (f) ».

---

[217] L'article « Ajustement de données expérimentales à une loi de probabilités» de Michel HENRY présenté pendant *Journées nationales APMEP* à Caen (le 21 - 24 octobre 2005)

http://www.apmep. asso.fr/IMG/pdf/ASm12.pdf

http://www.mission-laique.com/pedagogie/pdf/math56/am56_p59_72.pdf

6) Un seuil α étant donné, trouver dans la table du χ^2_v à k–r–1 degrés de liberté la valeur critique $\chi^2_c(\alpha)$ telle que $P(\chi^2_v > \chi^2_c(\alpha)) = \alpha$. La fonction KHIDEUX.INVERSE d'Excel donne aussi ces valeurs.

7) Expliciter la distribution (f) des fréquences f_i des modalités M_i dans l'échantillon observé. On note $n_i = nf_i$ les effectifs des M_i dans l'échantillon.

8) Calculer la *« distance »* du Khi-deux : $d_n^{\ 2} = n\sum_{i=1}^{k}\frac{(f_i - p_i)^2}{p_i} = \sum_{i=1}^{k}\frac{(n_i - np_i)^2}{np_i} = (\sum_{i=1}^{k}\frac{n_i^{\ 2}}{np_i}) - n$

9) Si $d_n^2 > \chi^2_c(\alpha)$, conclure que la loi P_θ n'est pas un modèle suffisamment adéquat pour la loi P_X représentant les variations de *C* dans *P*.

Rejeter H_0 en prenant un risque de se tromper $P_{H0}(d_n^2 > \chi^2_c(\alpha))$ proche de α.

10) Si $d_n^2 \leq \chi^2_c(\alpha)$, décider de conserver l'hypothèse H_0 : admettre que la loi P_θ est un modèle suffisamment adéquat pour représenter les variations de *C* dans *P*.

Annexe C

I. Questions pour l'enseignant

Voici 4 exercices à propos de la statistique enseignée en classe 10. Pour chacun d'eux, nous vous prions de répondre à la question suivante :

> *Proposeriez-vous un tel exercice en classe 10 ? Si oui, quelle réponse attendez-vous d'un élève de classe 10 ? Sinon, comment le modifieriez-vous ?*

Merci de votre participation au questionnaire.

Votre nom et prénom :.....
Vous êtes enseignant(e) du lycée : .....
Avez-vous enseigné à l'aide du programme expérimental entre les années 2004 et 2006 ?.....

Fiche de réponses :

Exercice 1 :

Proposeriez-vous cet exercice en classe 10 ? Oui Non

La réponse que vous attendez d'un élève de classe 10 :

.....

Votre modification éventuelle de l'exercice :

.....

Exercice 2 :

Proposeriez-vous cet exercice en classe 10 ? Oui Non

La réponse que vous attendez d'un élève de classe 10 :

.....

Votre modification éventuelle de l'exercice :

.....

Exercice 3 :

Proposeriez-vous cet exercice en classe 10 ? Oui Non

La réponse que vous attendez d'un élève de classe 10 :

.....

Votre modification éventuelle de l'exercice :

.....

Exercice 4 :

Proposeriez-vous cet exercice en classe 10 ? Oui Non

La réponse que vous attendez d'un élève de classe 10 :

.....

Votre modification éventuelle de l'exercice :

.....

II. Questions pour l'élève

Voici un ensemble de 4 exercices. Lisez attentivement les questions et répondez aux questions, s'il vous plait.

Merci de votre participation.

Vos nom et prénom : .....
Classe : .....
Lycée : .....

Fiche des réponses

Exercice 1 :

.....

.....

.....

Exercice 2 :

.....

.....

.....

Exercice 3 :

.....

.....

.....

Exercice 4 :

.....

.....

.....

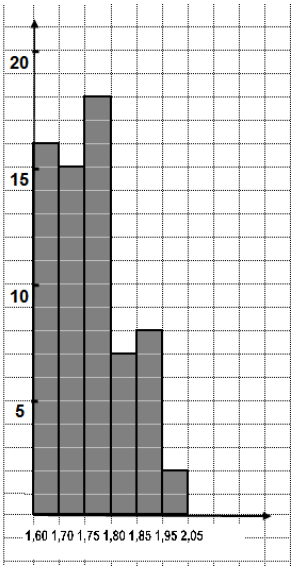
III. Quatre exercices

III.1. Exercice 1

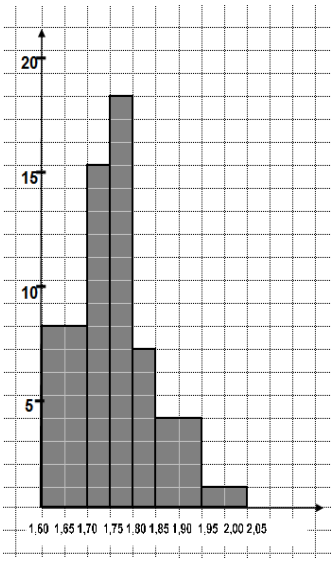
On a mesuré la taille des 66 joueurs d'un club de basket-ball. On a constitué le tableau suivant :

Classe de taille (mètre)	Nombre de joueurs
[1,60 ; 1,70[	16
[1,70 ; 1,75[	15
[1,75 ; 1,80[	18
[1,80 ; 1,85[	7
[1,85 ; 1,95[	8
[1,95 ; 2,05[	2

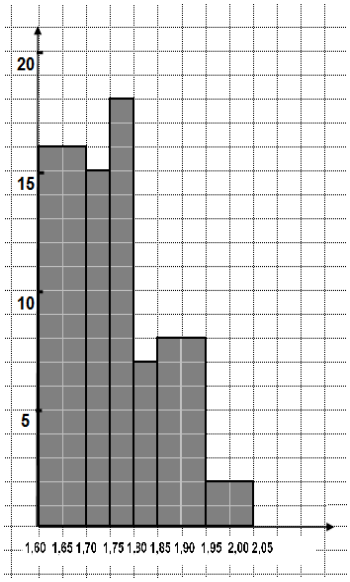
On a demandé à élèves d'un lycée vietnamien de représenter graphiquement les données du tableau. Voici cinq représentations graphiques qu'ils ont réalisées :



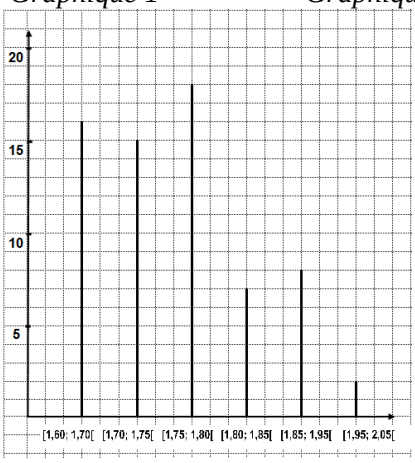
Graphique 1



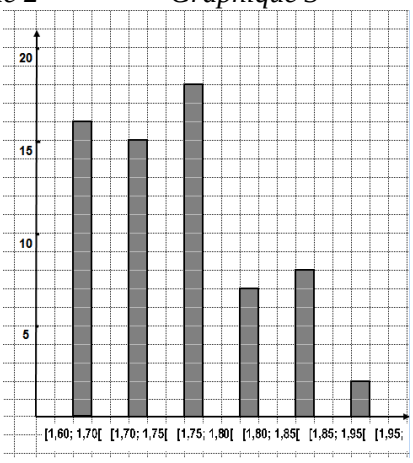
Graphique 2



Graphique 3



Graphique 4

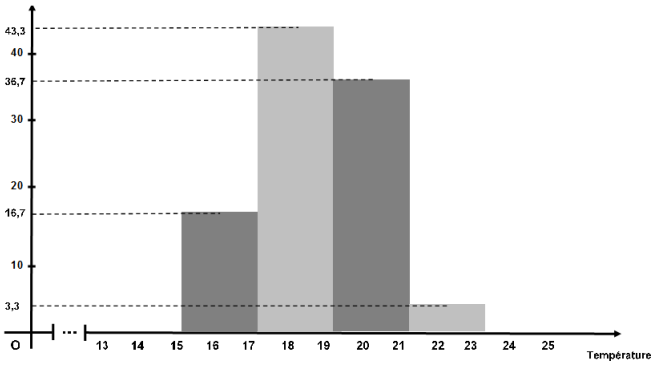


Graphique 5

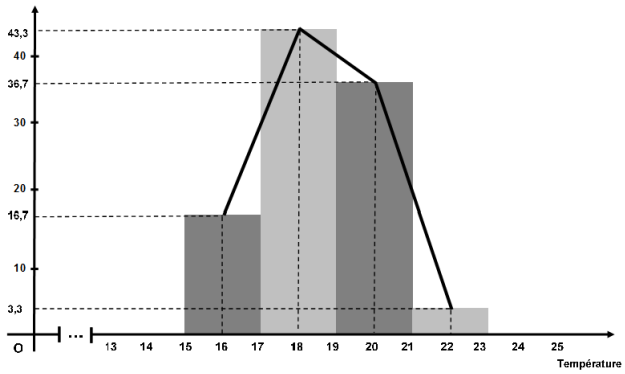
1. Parmi les représentations graphiques proposées, lesquelles considérez-vous comme correctes ?
2. Parmi les représentations graphiques proposées, lesquelles considérez-vous comme incorrectes ? Justifiez vos réponses.
3. Écrire la légende sur l'axe horizontal et sur l'axe vertical pour chaque représentation graphique acceptée comme correcte.

III.2. Exercice 2

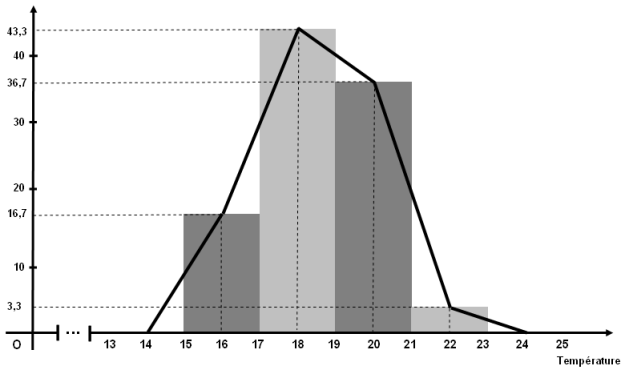
Un météorologue a relevé de 1961 à 1990 tous les jours de décembre à 10h la température dans un même lieu de la ville de Vinh. Voici l'histogramme de la distribution des fréquences :



Le météorologue a construit plusieurs courbes polygonales des fréquences pour décrire la distribution des fréquences :



Courbe polygonale 1



Courbe polygonale 2

Question : parmi les deux courbes polygonales tracées, lesquelles considérez-vous comme correctes pour décrire la distribution des fréquences ? Justifiez votre réponse.

III.3. Exercice 3

Pour contrôler la qualité de fonctionnement d'une machine remplissant des paquets de sucre, on a noté, en les ordonnant, les poids de 100 paquets remplis par la machine. Ils ont été choisis au hasard durant une journée de fonctionnement.

On obtient le tableau de recueil des poids en grammes :

250,8	252,2	281,7	292,3	312,2	344,0	345,6	399,2	490,2	490,2
490,3	490,5	490,6	490,6	491,1	491,3	491,7	492,1	492,3	493,1
493,1	493,1	493,2	493,3	493,3	493,4	494,0	494,3	494,5	494,8
495,1	495,7	495,7	495,8	495,9	496,0	496,5	496,5	497,1	497,1
498,3	498,3	498,9	498,9	499,1	499,1	499,7	499,7	499,9	500,4
500,8	501,1	501,2	501,3	501,4	501,7	501,8	501,9	502,0	502,1
502,2	502,2	502,2	502,3	502,4	502,5	502,8	502,8	503,0	503,7
503,9	504,1	504,4	504,7	504,7	504,8	505,3	505,4	505,5	505,9
506,1	506,2	506,3	506,3	506,8	507,1	507,3	507,5	507,6	507,8
507,9	508,0	508,2	508,6	508,7	509,0	510,5	510,7	511,7	512,2

1. Donnez un indicateur qui est un bon représentant de ce recueil de données ? Justifiez votre choix.

2. Tracez la courbe polygonale des effectifs qui décrit ces données (sur la feuille quadrillée joint).

3. Que peut-on dire sur le fonctionnement de la machine ? Pour quel poids la machine est-elle, programmée ? En tant que contrôleur, que diriez-vous de la machine ? A-t-elle fonctionné normalement ?

(Feuille quadrillée joint)

III.4. Exercice 4

On a recueilli et ordonné les notes d'un examen de mathématiques des 100 élèves d'un lycée de Hô Chi Minh ville qui ont eu des notes supérieures ou égales à 5.

5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	6	6	6	6
6	7	7	7	7	7	7	7	7	7	7	7	7	7	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	9	9	9	9	9	9	9	9	9	9
9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
9	9	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10

Recueil de notes

Pour communiquer des informations restreintes à comparer avec celles d'autres lycées de Hô Chi Minh ville, on extrait du recueil, des échantillons de 10 notes. On voudrait des échantillons représentatifs de l'ensemble des notes recueillies. Voici cinq échantillons :

6	7	7	8	8	8	8	9	9	10

Échantillon 1

7	7	8	8	8	9	9	9	10	10

Échantillon 2

5	6	9	9	9	9	9	9	10	10

Échantillon 3

5	6	7	7	8	8	8	9	9	10

Échantillon 4

6	6	7	8	8	8	8	9	10	10

Échantillon 5

1. D'après vous, parmi les cinq échantillons, quel est le plus représentatif du recueil tout entier ? Pourquoi ?

2. D'après vous, parmi les cinq échantillons, quel est le moins représentatif du recueil tout entier ? Pourquoi ?

3. Si possible, proposez un sixième échantillon plus représentatif que les cinq échantillons déjà donnés.

IV. Câu hỏi dành cho giáo viên

Thưa quý Thầy/Cô, sau đây là 4 bài tập về thống kê ở lớp 10. Với mỗi bài tập, mong quý Thầy/Cô vui lòng trả lời cho câu hỏi sau :

Quý Thầy/Cô có cho học sinh làm bài tập như loại bài tập này không ? Nếu có, quý Thầy/Cô mong muốn học sinh lớp 10 giải bài tập này thế nào ? Nếu không, quý Thầy/Cô sẽ điều chỉnh lại bài tập này như thế nào để cho học sinh làm ?

Chúng tôi xin chân thành cám ơn sự tham gia của quý Thầy Cô.

Phiếu trả lời

Họ và tên của quý Thầy Cô : ……

Quý Thầy/Cô là giáo viên trường THPT :……

Quý Thầy/Cô có tham gia giảng dạy chương trình thí điểm phân ban giai đoạn 2004-2006 không ?.....

Bài tập 1 :

Quý Thầy Cô có cho học sinh làm bài tập này ở lớp 10 không ? Có Không

a) Câu trả lời mà quý Thầy/Cô mong đợi nơi học sinh lớp 10 :

.....

b) Dự kiến điều chỉnh lại bài tập này của quý Thầy/Cô :

.....

Bài tập 2 :

Quý Thầy Cô có cho học sinh làm bài tập này ở lớp 10 không ? Có Không

a) Câu trả lời mà quý Thầy/Cô mong đợi nơi học sinh lớp 10 :

.....

b) Dự kiến điều chỉnh lại bài tập này của quý Thầy/Cô :

.....

Bài tập 3 :

Quý Thầy Cô có cho học sinh làm bài tập này ở lớp 10 không ? Có Không

a) Câu trả lời mà quý Thầy/Cô mong đợi nơi học sinh lớp 10 :

.....

b) Dự kiến điều chỉnh lại bài tập này của quý Thầy/Cô :

.....

Bài tập 4 :

Quý Thầy Cô có cho học sinh làm bài tập này ở lớp 10 không ? Có Không

a) Câu trả lời mà quý Thầy/Cô mong đợi nơi học sinh lớp 10 :

.....

b) Dự kiến điều chỉnh lại bài tập này của quý Thầy/Cô :

.....

V. Câu hỏi dành cho học sinh

Các em hãy đọc thật kỹ các câu hỏi trong bộ gồm 4 bài tập và cố gắng trả lời từng câu hỏi. Xin cám ơn sự tham gia của các em.

Họ và tên :

Lớp : ..... Trường THPT :

Phiếu trả lời

Phần trả lời bài tập 1 :

1) .....

.....

.....

2) .....

.....

3) Biểu đồ ..... Trục tung : .....Trục hoành :.....
Biểu đồ ..... Trục tung : .....Trục hoành :.....
Biểu đồ ..... Trục tung : .....Trục hoành :.....
Biểu đồ ..... Trục tung : .....Trục hoành :.....
Biểu đồ ..... Trục tung : .....Trục hoành :.....

Phần trả lời bài tập 2 :

.....

.....

Phần trả lời bài tập 3 :

1) .....

.....

.....

2) (Giấy kẻ ô)
3) .....

.....

Phần trả lời bài tập 4 :

1) .....

.....

2) .....

.....

3) .....

.....

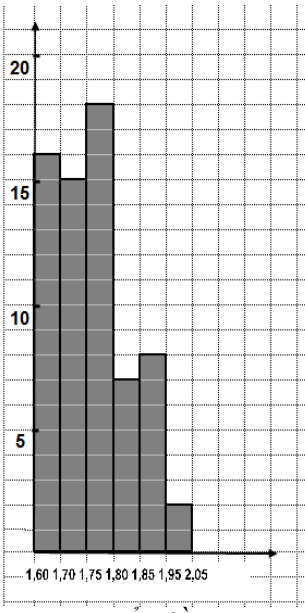
VI. Bốn bài tập

Bài tập 1

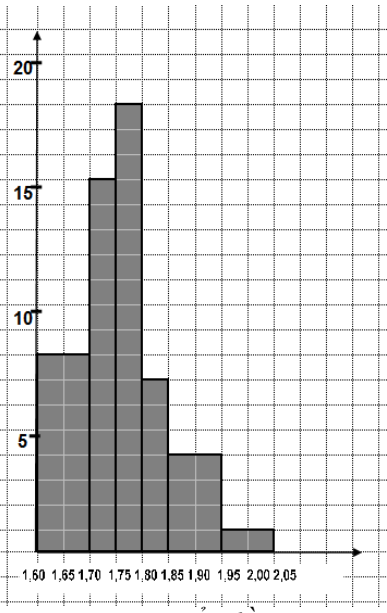
Sau khi đo chiều cao của 66 vận động viên bóng rổ của một câu lạc bộ, người ta thu được bảng sau :

Chiều cao (mét)	Số vận động viên
[1,60 ; 1,70)	16
[1,70 ; 1,75)	15
[1,75 ; 1,80)	18
[1,80 ; 1,85)	7
[1,85 ; 1,95)	8
[1,95 ; 2,05)	2

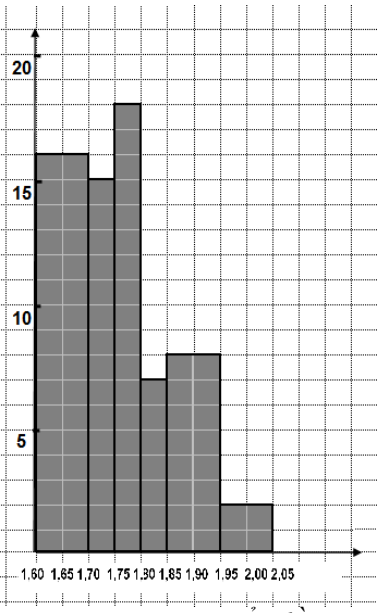
Người ta đã đề nghị các học sinh một trường trung học phổ thông vẽ biểu đồ mô tả bảng dữ liệu trên. Sau đây là 5 biểu đồ đã được thực hiện :



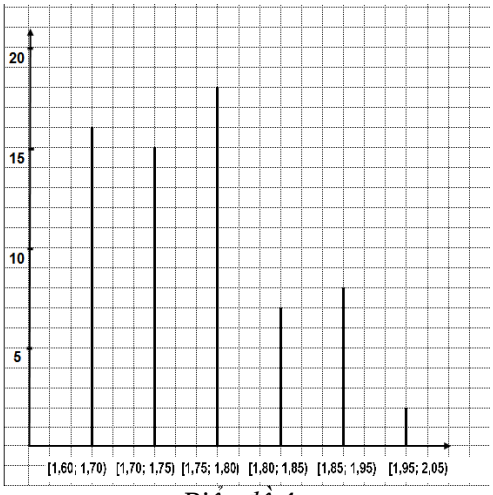
Biểu đồ 1



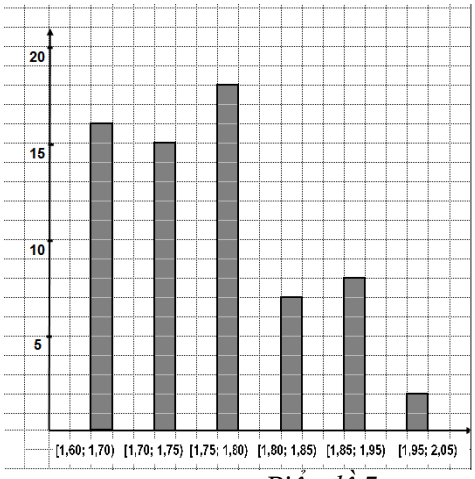
Biểu đồ 2



Biểu đồ 3



Biểu đồ 4



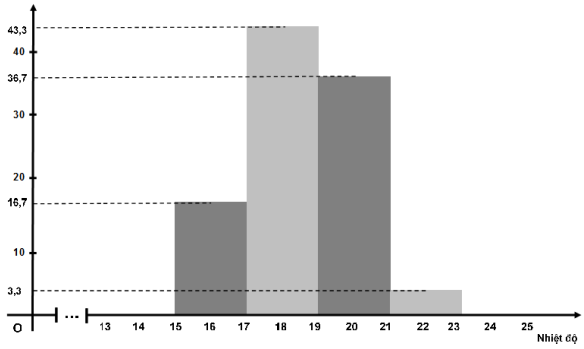
Biểu đồ 5

<u>Câu hỏi :</u>

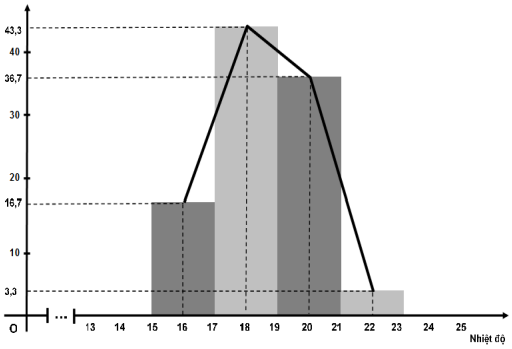
1) Trong số các biểu đồ trên, những biểu đồ nào được xem là đúng ?
2) Trong số các biểu đồ trên, những biểu đồ nào được xem là sai ? Giải thích câu trả lời của em.
3) Hãy viết chú thích trên trục tung và trục hoành cho mỗi biểu đồ mà em cho là đúng.

Bài tập 2

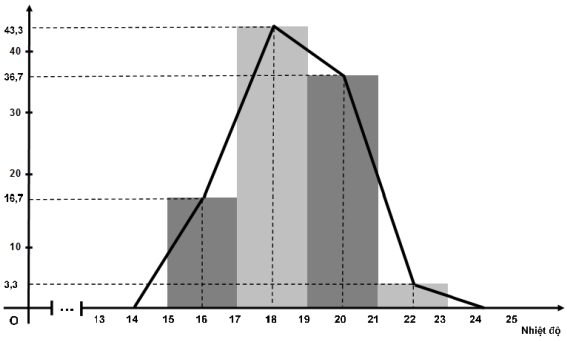
Từ năm 1961 đến 1990, một nhà khí tượng học đã ghi nhận nhiệt độ của thành phố Vinh vào lúc 10 giờ sáng mỗi ngày trong suốt tháng 12 tại cùng một địa điểm. Đây là biểu đồ hình cột thể hiện phân bố tần suất :



Nhà khí tượng học này đã vẽ nhiều đường biểu diễn gấp khúc tần suất để mô tả phân bố tần suất như sau :



Đường gấp khúc 1



Đường gấp khúc 2

<u>Câu hỏi :</u>

Theo em, trong số 4 đường gấp khúc tần suất vẽ ở trên, những đường gấp khúc tần suất nào được xem là đúng để mô tả phân bố tần suất ? Hãy giải thích câu trả lời của em.

Bài tập 3

Để kiểm tra chất lượng vận hành của một máy đóng bao bì, người ta đã cân 100 gói đường được đóng bao từ máy này và ghi lại kết quả theo thứ tự tăng dần. Những gói đường này được chọn ngẫu nhiên trong một ngày làm việc của máy.

Dưới đây là bảng dữ liệu thu được (tính bằng gam):

250,8	252,2	281,7	292,3	312,2	344,0	345,6	399,2	490,2	490,2
490,3	490,5	490,6	490,6	491,1	491,3	491,7	492,1	492,3	493,1
493,1	493,1	493,2	493,3	493,3	493,4	494,0	494,3	494,5	494,8
495,1	495,7	495,7	495,8	495,9	496,0	496,5	496,5	497,1	497,1
498,3	498,3	498,9	498,9	499,1	499,1	499,7	499,7	499,9	500,4
500,8	501,1	501,2	501,3	501,4	501,7	501,8	501,9	502,0	502,1
502,2	502,2	502,2	502,3	502,4	502,5	502,8	502,8	503,0	503,7
503,9	504,1	504,4	504,7	504,7	504,8	505,3	505,4	505,5	505,9
506,1	506,2	506,3	506,3	506,8	507,1	507,3	507,5	507,6	507,8
507,9	508,0	508,2	508,6	508,7	509,0	510,5	510,7	511,7	512,2

Câu hỏi :

1. Hãy đưa ra một giá trị đặc trưng làm đại diện tốt cho dữ liệu thu được. Đề nghị giải thích câu trả lời của em.
2. Vẽ (trên giấy kẻ ô đính kèm) đường gấp khúc tần số mô tả dữ liệu.
3. Ta có thể nói gì về sự vận hành của máy ? Máy này đã được lập trình để đóng bao với khối lượng bao nhiêu cho mỗi bao ? Với vai trò là người kiểm tra, em sẽ nói gì về chiếc máy này ? Nó vận hành có bình thường không ?

Bài tập 4

Sau khi chấm điểm bài kiểm tra toán theo đề chung dành cho tất cả học sinh lớp 10 trong thành phố, người ta thu thập điểm của 100 học sinh thuộc cùng một trường Trung học phổ thông. Bảng dưới đây trình bày theo thứ tự tăng dần điểm của những học sinh.

5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	6	6	6	6
6	7	7	7	7	7	7	7	7	7	7	7	7	7	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	9	9	9	9	9	9	9	9	9	9
9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
9	9	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10

Dữ liệu điểm thu thập được

Cần phải trích ra từ bảng trên một mẫu số liệu chỉ gồm điểm của 10 học sinh để trao đổi và so sánh với thông tin (về bài kiểm tra chung này) thu được từ các trường bạn trong thành phố. Vấn đề là phải có những mẫu số liệu đại diện cho tập hợp toàn thể số liệu điểm thu thập được. Sau đây là 5 mẫu số liệu được trích ra :

6	7	7	8	8	8	8	9	9	10

Mẫu số liệu 1

7	7	8	8	8	9	9	9	10	10

Mẫu số liệu 2

5	6	9	9	9	9	9	9	10	10

Mẫu số liệu 3

5	6	7	7	8	8	8	9	9	10

Mẫu số liệu 4

6	6	7	8	8	8	8	9	10	10

Mẫu số liệu 5

1) Theo em, trong số 5 mẫu số liệu trên, mẫu nào đại diện tốt nhất cho số liệu tổng thể ? Tại sao ?
2) Theo em, trong số 5 mẫu số liệu trên, mẫu nào đại diện tệ nhất cho số liệu tổng thể ? Tại sao ?
3) Nếu có thể, em hãy đưa ra một mẫu số liệu thứ 6, mà theo em, nó đại diện tốt hơn cả 5 mẫu số liệu đã cho ở trên.

Annexe D

I. Programmes informatiques

(Ces programmes sont écrits selon le langage Pascal et ils fonctionnent dans l'environnement du compilateur de Turbo Pascal 7.0)

I.1. Séance 2 : Bouteille_Connue

```
PROGRAM Seance_2_Bouteille_Connue;

USES crt,dos;

CONST f1=25;
      f2=100;
      f3=500;

VAR
  Rouge                                                    : byte;
  fois, fois_total, i, j, e_rouge, e_rouge_total, e_vert   : integer;
  c,reponse                                                : char;
  t                                                        : boolean;
  f                                                        : text;
  c_composition, c_fois, groupe, eff_rouge_s, eff_vert_s,
  fr_rouge_s, fr_vert_s                                    : string;
  f_rouge, f_vert                                          : real;
  h,m,s,cc, jj,mm,aa,jds                                   : word;
  hs,ms,ss, jjs,mms,aas                                    : string;


Procedure r; Begin textcolor(red); End;
Procedure g; Begin textcolor(green); End;
Procedure w; Begin textcolor(white); End;
Procedure y; Begin textcolor(yellow); End;
Procedure b; Begin textcolor(blue); End;

Function reel_en_text(r: real):string;
Var d     : longint;
    ds    : string;
Begin
  d:=round(r*1000);
  str(d,ds);
  reel_en_text:='O.'+ds;
End;

Procedure saut(n: byte);
Var i: byte;
Begin
  for i:=1 to n do writeln;
End;

Procedure annonce_1;
Begin
  Repeat

  clrscr;
  fois_total:=0;
  e_rouge_total:=0;

  saut(2);
  writeln('        * * * * * * * * * *      GROUPE N°',groupe,'     * * * * * * * * * * * *
');
  saut(2);

  w;
  writeln('La bouteille contient 5 boules rouges et vertes, choisissez le nombre des boules
');
  saut(1);
```

```
  writeln('rouges en entrant un nombre entier de 1 … 4');
  saut(1);
  y; write('                        Le nombre des boules ');
  r; write('ROUGE ');
  y; write(' : ');
  r; readln(rouge); w;
  str(rouge,c_composition);
  write(f,c_composition:3);
  if (rouge < 1) or (rouge > 4)
  then
    begin
      writeln(f,'Valeur invalid,');
      saut(5);
      textcolor(yellow+blink);
      writeln('Valeur entrée est invalide ! Tapez touche Entrer et entrez une autre valeur
');
      w;
      readln;
    end;
  Until (rouge >0) and (rouge < 5);
End;

Procedure annonce_2;
Begin
  clrscr;

  saut(2);
  writeln('        * * * * * * * * * * *      GROUPE N°',groupe,'     * * * * * * * * * * * *
');
  saut(2);

  write('Selon votre choix, la bouteille contient ');
  r; write(rouge);
  w; write(' boule(s) ');
  r; write('ROUGE');
  w; write(' et ');
  g; write(5-rouge);
  w; write(' boule(s) ');
  g; write('VERTE');
  w;
  saut(2);
  writeln('Combien de fois voulez-vous "interroger" la bouteille ? ');
  saut(1);
  write('Choisissez un parmi trois nombres de tirages suivants : ');
  y; write(' 25      100     500'); w;
  saut(2);
  w;
  write('puis tapez le nombre choisi : ');
  y; readln(fois); w;
  t:= (fois = f1) or (fois = f2) or (fois = f3);
  str(fois,c_fois);
  write(f,c_fois:8);
  If t then write(f,'              ');
End;

Procedure tableau;
Begin
    e_vert      :=fois-e_rouge;
    f_rouge     :=e_rouge/fois;
    f_vert      :=e_vert/fois;
    saut(1);
    write('                    Boule          ');
    textcolor(red);
    write('    ROUGE');
    textcolor(green);
    writeln('          VERT');
    textcolor(white);
    writeln('                  ___________________________________________');
    saut(1);
    writeln('                    Effectif  ',e_rouge:14,e_vert:14);
    saut(1);
    write  ('                    Fréquence ',f_rouge:14:4,f_vert:14:4);

    str(e_rouge,eff_rouge_s);
    str(e_vert,eff_vert_s);
    write(f,'         Eff = ',eff_rouge_s:4,' ; Freq = ',reel_en_text(f_rouge):7);
    writeln(f,'           Eff = ',eff_vert_s:4,' ; Freq = ',reel_en_text(f_vert):7);
```

```
End;

BEGIN                  {programme principal}
  clrscr;
  saut(2);
  write('                      Numéro du groupe : N°'); {Demander N° du groupe pour le nom du
fichier plus tard}
  readln(groupe);
  saut(1);
  writeln('       * * * * * * * * * * * *      GROUPE N°',groupe,'     * * * * * * * * * * *
* * ');
  saut(1);
  getdate(aa,mm,jj,jds);          {annee, mois, jour, jour de la semaine}
  str(aa,aas); str(mm,mms);  str(jj,jjs);
  gettime(h,m,s,cc);
  str(h,hs);  str(m,ms);  str(s,ss);
  IF h<10 then hs:='0'+hs;
  IF m<10 then ms:='0'+ms;
  IF s<10 then ss:='0'+ss;

  assign(f,'C:\VNTH\BC'+hs+ms+ss+'.G'+groupe);              {Créer fichier}
  rewrite(f);
  writeln(f,'Séance2 : Bouteille_CONNUE');                  {Trace du titre dans fichier}
  writeln(f,'Fiche des traces du Groupe_'+groupe);
  writeln(f,'Heure: ',hs,':',ms,':',ss);
  writeln(f,'Date : ',jjs:2,'/',mms:2,'/',aas);
  writeln(f,'');
  saut(3);
  writeln(f,'NOMBRE DE BOULES ROUGES : ');   {Trace du nombre de boule rouge}
  annonce_1;    {Entrer la composition dans la bouteille}
  writeln(f,'');
  writeln(f,'');
  writeln(f,'NOMBRE  DE  TIRAGE                                                          ROUGE
VERT');

  Repeat
    saut(1);
    annonce_2;  {Entrer le nombre de tirage}
    if not t {Limiter le nombre des tirages parmi : 25, 100 et 500}
    then
      repeat
        saut(2);
        textcolor(yellow+blink); {Attirer l'attention de l'utilisateur}
        writeln(' Valeur entrée est invalide ! Tapez touche Entrer puis entrez une autre
valeur  ');
        w;
        readln;
        writeln(f,'  Valeur invalide');
        annonce_2; {Refaire entrer le nombre de tirages}
      until t;
    randomize;     {Simuler les tirage}
    e_rouge:=0;
    for i:= 1 to fois do
    begin
      j:=random(5)+1;
      if j mod 5 < rouge then e_rouge:=e_rouge+1;
    end;
    saut(3);  {Ecrire tableau de distribution des résultats}
    write('Vous avez "interrogé" la bouteille ');
    y; write(fois); w;
    writeln(' fois, vous obtenez le tableau suivant : ');
    tableau;
    e_rouge_total:=e_rouge_total+e_rouge;
    fois_total:=fois_total+fois;
    saut(3);  {Demander si l'élève veut refaire des tirages}
    writeln('Voulez-vous "interroger" la bouteille encore ? ');
    saut(1);
    write('Tapez lettre ');
    y; write('O'); w;
    write(' pour ');
    y; write('Oui'); w;
    write(' ou ');
    y; write('N'); w;
    write(' pour ');
    y; write('Non'); w;
    write(' : ');
```

```
    repeat    {Restriction de la réponse pour seulement OUI et NON}
      c:=upcase(readkey);
    until (c in ['O','N']);
    y; write(c); w;
    readln;
  Until upcase(c)='N';
  writeln(f,'');
  writeln(f,'EFFECTIF TOTAL  = ',fois_total:4);
  writeln(f,'Eff_ROUGE    total    =    ',e_rouge_total:4,'    ;    Freq_ROUGE    total    =
',reel_en_text(e_rouge_total/fois_total):7);
  write(f,'Eff_VERT  total = ',fois_total-e_rouge_total:4);
  write(f,' ; Freq_VERT  total = ',reel_en_text(1-e_rouge_total/fois_total):7);
  close(f);
  readln;
END.
```

I.2. Séance 2 : Bouteille_Inconnue

```
PROGRAM Seance_2_Bouteille_Inconnue_3R;      {Ce programme est pour 3 Rouges -2 Vertes}

USES crt,dos;

CONST f0     = 50;
      f1     = 25;
      f2     = 100;
      f3     = 500;
      nr     = 3;      {nr = NOMBRE DES BOULES ROUGES}

VAR
  Rouge                                                    : byte;
  fois, fois_total, i,j, e_rouge, e_rouge_total, e_vert    : integer;
  c,reponse                                                : char;
  t                                                        : boolean;
  f                                                        : text;
  c_composition, c_fois, groupe, eff_rouge_s, eff_vert_s,
  fr_rouge_s, fr_vert_s                                    : string;
  f_rouge, f_vert                                          : real;
  h,m,s,cc, jj,mm,aa,jds                                   : word;
  hs,ms,ss, jjs,mms,aas                                    : string;

Procedure r; Begin textcolor(red); End;
Procedure g; Begin textcolor(green); End;
Procedure w; Begin textcolor(white); End;
Procedure y; Begin textcolor(yellow); End;
Procedure b; Begin textcolor(blue); End;

Function reel_en_text(r: real):string;
Var d     : longint;
    ds    : string;
Begin
  d:=round(r*1000);
  str(d,ds);
  reel_en_text:='O.'+ds;
End;

Procedure saut(n: byte);
Var i: byte;
Begin
  for i:=1 to n do writeln;
End;

Procedure annonce_2;
Begin
  clrscr;
  saut(2);
  w;
  writeln('        * * * * * * * * * * *      GROUPE N°',groupe,'     * * * * * * * * * * * *
');
  saut(3);
  write('La bouteille contient ');
  y; write('5');
  w; write(' boules de deux couleurs ');
```

```
  r; write('ROUGE');
  w; write(' et ');
  g; write('VERTE');
  w;
  saut(3);

  writeln('Combien de fois voulez-vous "interroger" la bouteille ? ');
  saut(1);
  write('Choisissez un parmi trois nombres de tirages suivants : ');
  y; write(' 25      100     500'); w;
  saut(2);
  w;
  write('puis tapez le nombre choisi : ');

  y; readln(fois); w;
  t:= (fois = f1) or (fois = f2) or (fois = f3);
  str(fois,c_fois);
  write(f,c_fois:8);
  If t then write(f,'              ');
End;

Procedure tableau;
Begin
    e_vert      :=fois-e_rouge;
    f_rouge     :=e_rouge/fois;
    f_vert      :=e_vert/fois;
    saut(1);
    write('                    Boule          ');
    textcolor(red);
    write('    ROUGE');
    textcolor(green);
    writeln('          VERT');
    textcolor(white);
    writeln('                  ________________________________________');
    saut(1);
    writeln('                    Effectif  ',e_rouge:14,e_vert:14);
    saut(1);
    write  ('                    Fréquence ',f_rouge:14:4,f_vert:14:4);

    str(e_rouge,eff_rouge_s);
    str(e_vert,eff_vert_s);
    write(f,'         Eff = ',eff_rouge_s:4,' ; Freq = ',reel_en_text(f_rouge):7);
    writeln(f,'           Eff = ',eff_vert_s:4,' ; Freq = ',reel_en_text(f_vert):7);

End;


BEGIN                  {programme principal}

  clrscr;
  saut(2);
  rouge:=nr;
  write('                      Numéro du groupe : N°'); {Demander N° du groupe pour le nom du
fichier plus tard}
  readln(groupe);

  getdate(aa,mm,jj,jds);          {année, mois, jour, jour de la semaine}
  str(aa,aas); str(mm,mms);  str(jj,jjs);

  gettime(h,m,s,cc);
  str(h,hs);  str(m,ms);  str(s,ss);
  IF h<10 then hs:='0'+hs;
  IF m<10 then ms:='0'+ms;
  IF s<10 then ss:='0'+ss;
  assign(f,'C:\I_'+hs+ms+ss+'.G'+groupe); {Créer fichier}
  rewrite(f);
  writeln(f,'Seance2 : Bouteille_INCONNUE'); {Trace du titre dans fichier}
  writeln(f,'Fiche des traces du Groupe_'+groupe);
  writeln(f,'Heure: ',hs,':',ms,':',ss);
  writeln(f,'Date : ',jjs:2,'/',mms:2,'/',aas);
  writeln(f,'');

  saut(3);
  writeln(f,'Nombre de boule rouge: ', nr:2);
  writeln(f,'NOMBRE  DE  TIRAGE                                                          ROUGE
VERT');
```

```
  Repeat
    saut(1);
    annonce_2;  {Entrer le nombre de tirage}
    if not t {Limiter le nombre des tirages parmi : 25, 100 et 500}
    then
      repeat
        gotoxy(1,20);
        textcolor(yellow+blink); {Attirer l'attention de l'utilisateur}
        writeln('Valeur  entrée  est  invalide  !  Tapez  touche  Entrer  puis  entrez  une  autre
valeur  ');
        w;
        readln;
        writeln(f,'  Valeur invalide');
        annonce_2; {Refaire l'Entrée du nombre de tirage}
      until t;

    randomize;     {Simuler les tirage}
    e_rouge:=0;
    for i:= 1 to fois do
    begin
      j:=random(5)+1;
      if j <= rouge then e_rouge:=e_rouge+1;
    end;

    saut(3);  {Ecrire tableau de distribution des résultats}
    write('Vous avez "interrogé" la bouteille ');
    y; write(fois); w;
    writeln(' fois, vous obtenez le tableau suivant : ');
    tableau;

    e_rouge_total:=e_rouge_total+e_rouge;
    fois_total:=fois_total+fois;

    saut(3);  {Demander si l'utilisateur veut refaire des tirages}
    writeln('Voulez-vous "interroger" la bouteille encore ? ');
    saut(1);
    write('Tapez lettre ');
    y; write('O'); w;
    write(' pour ');
    y; write('Oui'); w;
    write(' ou ');
    y; write('N'); w;
    write(' pour ');
    y; write('Non'); w;
    write(' : ');

    repeat    {Limiter la réponse OUI ou NON seulement}
      c:=upcase(readkey);
    until (c in ['O','N']);
    y; write(c); w;

    readln;
  Until upcase(c)='N';
  writeln(f,'');
  writeln(f,'EFFECTIF TOTAL  = ',fois_total);
  writeln(f,'Eff_ROUGE    total    =    ',e_rouge_total:4,'         Freq_ROUGE    total    =
',reel_en_text(e_rouge_total/fois_total):7);
  write(f,'Eff_VERT     total  =  ',fois_total-e_rouge_total:4,'     Freq_VERT     total  =
',reel_en_text(1-e_rouge_total/fois_total):7);
  close(f);
  readln;
END.
```

I.3. Séance 3 : Situation_Note

```
PROGRAM Seance_3_Avec_Remise;

USES crt,dos,graph;

CONST
     P         = 1000;   {Taille de la population mère}
     E         = 10;     {Taille de l'échantillon}
     v5        = 90;
     v6        = 110;
     v7        = 140;
     v8        = 400;
     v9        = 160;
     v10       = 100;
     fois_2    = 20;     {Command n°2 appelle 20 fois de n°1}

TYPE
     tab_e     = array[1..e]  of byte;   {array de 10 casses pour un échantillon}
     tab_eff   = array[5..10] of longint;{6 notes, indice de 5 . 10}

VAR
     i,j, n_tirage, choisi, note_choisie     : integer; {commande n°2}
     freq_note                               : array[1..500] of real;
                                                          {tableau de 500 notes-commande n°1}
     C                                       : char; {pour la réponse Oui ou Non}
     freq_note_choisie_totale,        {fréquence totales de 20 fréquences}
     tempo                                   : real;
     echantillon                             : tab_e;
     tab_eff_note,                                  {sauver garder 6 effectifs de 6 valeurs}
     tab_eff_note_total                      : tab_eff; {cumuler 6 effs de 6 valeurs}
     nombre, entier, decimal, n_choisi_total,
     freq_choisie_moyenne                    : real; {fréquence moyenne de 10000 notes}
     f                                       : text;  {fichier}
     ch, groupe, statistique                 : string;
     h,m,s, cc,jj,mm,aa,jds                  : word;
                                             {heure, minutes, seconde, cc, jour, mois, année}
     hs,ms,ss, jjs,mms,aas                   : string;
     pilote, mode, errcode,
     d, mx, my, x1, y1, x2, y2               : integer;
     tab_note_choisie                        : array[1..fois_2] of real;
                                               {pour sauver 20 fréquences de la note choisie}
     type_e                                  : array[1..5] of integer;   {commande n°3}

Procedure r; Begin textcolor(red); End;       {couleur rouge}
Procedure w; Begin textcolor(white); End;     {couleur blanche}
Procedure y; Begin textcolor(yellow); End;    {couleur jaune}

Procedure saut(n: byte);     {sauter n ligne}
Var i: byte;
Begin
  for i:=1 to n do writeln;
End;

Procedure ordonner_ech(var a: tab_e;k:integer);  {Ordonner l'échantillon en croissant}
var i,j,temp:integer;
Begin
  for i:=1 to k-1 do
    for j:=1 to k-i do
      if a[j]>a[j+1] then
        begin
          temp:= a[j];
          a[j]:=a[j+1];
          a[j+1]:=temp;
        end;
End;

Procedure vide_tab;      {vider un tableau effectif d'échantillon}
var i: byte;
Begin
  For i:= 5 to 10 do
  tab_eff_note[i]:=0;
End;
```

```
Procedure ecrire_ech(a:tab_e);  {exporter un échantillon en ligne sur l'écran}
var i: integer;
Begin
  y;
  For i:=1 To e Do write(a[i]:3);      {e : taille de l'échantillon}
    write(a[i]:3);
  w;
End;

Procedure ecrire_ech_f(a:tab_e);       {Exporter un échantillon dans un fichier}
var i: integer;
Begin
  writeln(f,'');
  write(f,'Echantillon ',n_tirage:3,' :    ');
  For i:=1 To e Do write(f,a[i]:4);    {e : taille de l'échantillon}
    write(f,a[i]:4);
End;

Function note_alea:integer;        {Prendre une note au hasard avec remise}
Var r, i: integer;
Begin
  r:=random(p)+1;
  If r<=v5 then note_alea:=5
  else if r<=(v5+v6) then note_alea:= 6
       else if r<=(v5+v6+v7) then note_alea:= 7
            else if r<=(v5+v6+v7+v8) then note_alea:= 8
                 else if r<=(v5+v6+v7+v8+v9) then note_alea:= 9
                      else note_alea:=10;
End;

Procedure creer_ech(var a:tab_e);   {Fabriquer un échantillon de 10 notes}
Var i,n: integer;
Begin
  n_tirage:=n_tirage+1;          {compter le nombre de tirage}
  For i:=1 To e Do
  begin
    n:=note_alea;
    a[i]:=n;
    tab_eff_note[n]:=tab_eff_note[n]+1; {compter effectif de valeur n}
  end;

  ordonner_ech(a,e);            {ordonner échantillon reçu}
  if upcase(c) <> 'H' then
  begin
    ecrire_ech(a);              {exporter l'échantillon obtenu}
    ecrire_ech_f(a);            {exporter l'échantillon obtenu dans un fichier}
  end;

End;

Procedure creer_ech_bis(var a:tab_e);   {Fabriquer un échantillon de 10 notes pour n°3}
Var i,n: integer;
Begin
  For i:=1 To e Do
  begin
    n:=note_alea;
    a[i]:=n;
    tab_eff_note[n]:=tab_eff_note[n]+1; {compter effectif de valeur n}
  end;
End;

Procedure Commande_1;            {50 échantillons de 10 notes = 500 notes}
Var i: integer;
Begin
  vide_tab;                      {vider array de l'échantillon}
  if upcase(c) <> 'H' then       {commande n°2 = Hai = H}
  begin
    clrscr;
    saut(2);
    y;
    writeln('La commande n°1 :');
    w;
    saut(1);
    writeln('Tapez touche Entrer pour avoir 50 échantillons de 10 notes');;
    readln;
    writeln(f,'Voici 50 tirages de 10 notes');
  end;
```

```
  For i:=1 To 50 Do
    begin
      if upcase(c) <> 'H' then           {commande n°2 = Hai = H}
      begin
        saut(1);
        write('Voici un échantillon de ',e,' notes :  ');
      end;
      creer_ech(echantillon);
    end;
  if upcase(c)='H' then
    for i:= 5 to 10 do
      tab_eff_note_total[i]:=tab_eff_note_total[i]+tab_eff_note[i];
      {compter l'effectif de 6 valeurs pour récupérer celui de la note choisie pour n°2}
End;

Procedure Commande_1bis;       {50 échantillons de 10 notes = 500 notes}
Var i: integer;
Begin
  vide_tab;                      {vider array de l'échantillon}
  For i:=1 To 50 Do
    creer_ech_bis(echantillon);
End;

Procedure Commande_2;
{fois_2 =20 fois de n°1 pour avoir une suite de 20 fréquences}
Var i,j,n: integer;
Begin
  clrscr;
  saut(1);
  {w;}
  y;
  writeln('La commande n°2 : ');
  saut(1);
  w;
  write('Choisissez la note que vous voulez observer ses fréquences ? Note : ');
  y;
  readln(choisi);
  writeln;
  freq_note_choisie_totale:= 0;   {vider la variable fréquence totale de note choisie}
  y;
  writeln('Voici ',fois_2,' fréquences de note ',choisi,' obtenues par ',fois_2,' demandes de
la commande n°2');
  w;
  saut(2);
  writeln(f,'Voici ',fois_2, ' fréquence de note ',choisi);

  For i:=1 to fois_2 do       {20 fréquences}
  begin
    {vide_tab;                 vider array de l'échantillon}
    commande_1;
    tab_note_choisie[i]:=tab_eff_note[choisi]/50/e;
    write(tab_note_choisie[i]:7:3);
    freq_note_choisie_totale:=freq_note_choisie_totale+tab_note_choisie[i];
    if i mod 10 = 0 then begin saut(1); writeln(f,'');end;
  end;

  freq_choisie_moyenne:=freq_note_choisie_totale/fois_2;
  saut(2);
  write('Fréquence moyenne de note ',choisi,' après ',fois_2,' fois de demander la commande
n°1 : ');
  y;
  writeln(freq_choisie_moyenne:6:3);
  w;
  writeln(f,'Fréquence moyenne de note ',choisi,' après ',fois_2,' fois de Command n°1 :
',freq_choisie_moyenne:6:3);

  for i:=1 to fois_2-1 do      {METTRE EN ORDRE LES 100 FREQUENCES}
    for j:=1 to fois_2-i do
      if tab_note_choisie[j]>tab_note_choisie[j+1] then
        begin
          tempo:= tab_note_choisie[j];
          tab_note_choisie[j]:=tab_note_choisie[j+1];
          tab_note_choisie[j+1]:=tempo;
        end;
  saut(1);
  y;
  writeln('Tapez touche Entrer pour ordonner les ',fois_2, ' fréquences');
```

```
  readln;
  clrscr;
  saut(2);
  writeln('Voici ', fois_2, ' fréquences ordonnées croissantes :');
  w;
  writeln;
  for i:=1 to fois_2 do
  begin
    write(tab_note_choisie[i]:7:3);
    write(f,tab_note_choisie[i]:8:3);
    if i mod 10 = 0 then
    begin
      writeln;
      writeln(f,'');
    end;
  end;
  saut(2);
  write('Fréquence moyenne de note ',choisi,' après ',fois_2,' fois de demander la commande
n°1 : ');
  y;
  writeln(freq_choisie_moyenne:6:3);
  w;
End;
```

Procedure Commande_3;
{Simulation de n°1 environ n fois pour compter le nombre d'échantillon de type 1, 2, autre}

```
Var i,j,max,type_1,type_2,autre, fois_3 : integer;
Begin
  type_1:=0;                  {compter les échantillons représentatifs de type 1, 2 et autre}
  type_2:=0;
  autre:=0;
  For i:=1 To 5 Do type_e[i]:=0;      {vider array type_e}

  y;
  writeln('La commande n°3 : ');
  saut(1);
  w;
  saut(3);
  writeln('Combien d échantillons représentatifs obtenus par la commande n°1');
  saut(1);
  write('que vous voulez ? ');
  readln(fois_3);
  saut(2);
  For i:=1 To fois_3 Do
    begin
      commande_1bis;             {réduction de n°1 sans garder les traces}
      type_e[1]:=tab_eff_note[5];
      type_e[2]:=tab_eff_note[6];
      type_e[3]:=tab_eff_note[7];
      type_e[4]:=tab_eff_note[9];
      type_e[5]:=tab_eff_note[10];
      max:=type_e[1];
      For j:=2 To 5 Do
        if max < type_e[j] then max:=type_e[j];
      if max = type_e[4] then type_1:=type_1+1         {fr. de la note 9 est la + grande}
        else if max = type_e[3] then type_2:=type_2+1  {fr. de la note 7 est la + grande}
          else autre:=autre+1;                         {fr. d'une autre note est la + grande}
    end;
  writeln('Parmi ',fois_3, ' échantillons représentatifs, il y a :');
  saut(1);
  writeln(type_1:6,  '  échantillons  représentatifs  de  type  (1)  avec  la  fréquence    ',
type_1/fois_3:8:5);
  saut(1);
  writeln(type_2:6,  '  échantillons  représentatifs  de  type  (2)  avec  la  fréquence    ',
type_2/fois_3:8:5);
  saut(1);
  writeln(autre:6,  '  échantillons  représentatifs  d"autre  type  avec  la  fréquence  ',
autre/fois_3:8:5);

End;


{Partie cachée pour calculer les effectifs et les fréquences de l'échantillon moyen}
```

```
Procedure Tableau_Secret;      {créer le tableau de distribution des fréquences}
Begin
    clrscr;
    gotoxy(10,1);
    writeln;
    w;
    writeln('Après ',n_tirage,' fois de demander un échantillon de 10 notes, on obtient le
tableau ');
    writeln;
    writeln('de distribution des fréquences de ',n_tirage*10,' notes : ');
    writeln;
    writeln;
    y;
    writeln('Note               5         6         7         8         9        10');
    writeln;
    write  ('Fréquence  ');
    For i:=5 To 10 Do
    if  n_tirage>0
    then
      write(tab_eff_note[i]/(n_tirage*10):10:4)
    else
    write(n_tirage:10);

End;

Procedure Tableau_Secret_f;    {Écrire le tableau de distribution des fréquences comme trace}
Begin
  writeln(f,'');
  writeln(f,'    Après ',n_tirage,' fois de tirage de 10 notes, on obtient une distribution
');
  writeln(f,'        des effectifs et des fréquences de ',n_tirage*10,' notes : ');
  writeln(f,'');
  writeln(f,'Notes             5         6         7         8         9        10');
  write(f,'Effectifs');

  For i:=5 To 10 Do write(f,tab_eff_note[i]:10);
  writeln(f,'');
  write(f,'Fréquences');

  For i:=5 To 10 Do
  if  n_tirage>0
  then
    write(f,tab_eff_note[i]/(n_tirage*10):10:4)
  else
    write(f,n_tirage:10);

End;

BEGIN       {Programme principal}
  clrscr;
  randomize;
  n_choisi_total:=0;

  w;
  saut(1);
  write('           Entrez le numéro du groupe : N°'); {Demander N° du groupe pour le nom du
fichier plus tard}
  readln(groupe);
  clrscr;
  saut(2);
  writeln('       * * * * * * * * * * *      GROUPE N°',groupe,'     * * * * * * * * * * * *
');
  saut(5);

  getdate(aa,mm,jj,jds);                  {année, mois, jour, jour de la semaine}
  str(aa,aas); str(mm,mms);  str(jj,jjs);

  gettime(h,m,s,cc);
  str(h,hs);  str(m,ms);  str(s,ss);

  IF h<10 then hs:='0'+hs;
  IF m<10 then ms:='0'+ms;
  IF s<10 then ss:='0'+ss;
  assign(f,'C:\'+hs+ms+ss+'S3.G'+groupe); {Créer nom du fichier des traces}

  rewrite(f);
  writeln(f,'Seance3: Echantillon representatif');          {Trace du titre dans fichier}
```

```
  writeln(f,'Fiche des traces du Groupe_'+groupe);
  writeln(f,'Heure: ',hs,':',ms,':',ss);
  writeln(f,'Date : ',jjs:2,'/',mms:2,'/',aas);
  writeln(f,'');

  writeln(f,'');
  writeln(f,'Population mere :');

  writeln(f,'Note');
  writeln(f,'        5         6         7         8         9        10');
  writeln(f,v5:10,v6:10,v7:10,v8:10,v9:10,v10:10);
  y;

  writeln(' Dans l ordinateur, il y a les notes suppérieurs ou égales à 5 de l  élève  de');
  writeln;
  writeln(' tous les lycées de Ho Chi Minh ville. L ordinateur peut fournir un échantillon');
  writeln;
  writeln(' de 10 notes pris par hasard dans l ensemble de toutes les notes ');
  writeln;

  For i:=6 To 10 Do tab_eff_note[i]:=0;

{Créer un échantillon de taille 10 de la population en tirant 10 fois d'une note avec remise}

  n_tirage:=0;
  Repeat
    saut(3);
    w;
    write('Voulez-vous demander un échantillon de ',e,' notes ? (Tapez O (Oui)/N (Non)) : ');
    c:=readkey;
    if upcase(c)='O'
    then
      begin
        y; writeln('Oui'); w;
        saut(1);
        write('Voici un échantillon de ',e,' notes :  ');
        creer_ech(echantillon);
      end
    else if upcase(c)='N' then
           begin y; write('Non'); w; end;
  Until upcase(c)='N';
  Tableau_Secret_f;

  saut(5);

  repeat                                                {Autres commandes}
    saut(2);
    y;
    write('Tapez touche Entrer');
    readln;
    w;
    clrscr;
    saut(2);
    {textcolor(yellow+blink);}
    y;
    writeln('                 Qu est-ce que vous voulez faire encore ?');
    c:=readkey;
    w;
    case upcase(c) of
    'M' : begin n_tirage:=0; vide_tab; Commande_1; Tableau_Secret_f; end;  {Command n°1= M}
    'H' : begin n_tirage:=0; Commande_2; end;                              {Command n°2= H}
    'B' : commande_3;
    'G' : Graphique_Secret;                                                {Graphique  = G}
    'T' : Tableau_Secret;                                                  {Tableau Ef = T}
    end;
  until upcase(c) = 'N';

  close(f);
  readln;
END.
```

II. Fiches de l'ingénierie didactique

Fiche 1 (noms, prénoms, lycée, classe, n° groupe)

Durée : 5 minutes

Que pouvez-vous dire sur le *contenu* de votre bouteille ? Noter ci-dessous toutes les déclarations que vous jugez pertinentes à propos de ce contenu ainsi que vos remarques.

Fiche 2 (noms, prénoms, lycée, classe, n° groupe)

Que pouvez-vous dire sur la *composition* de votre bouteille ? Noter ci-dessous votre proposition et les arguments qui la soutiennent.

Rappel : Briand présente pages 254 et 255 les probabilités de tirages favorables sur n (= 15, 25, 50 et 500) tirages pour les deux compositions (3/2 et 4/1).

La composition de votre bouteille :

Donnez vos arguments :

Etes-vous moyennement sûrs ? Etes-vous tout à fait sûrs ?

Pourquoi ?

Fiche enseignant (exemple)

N° groupe	1R4V	2R3V	3R2V	4R1V
1		X « Tout à fait sûr »		
2				
3			X « moyennement sûr	
4				

Fiche 2 bis (noms, prénoms, lycée, classe, n° groupe)

Durée : 10 minutes

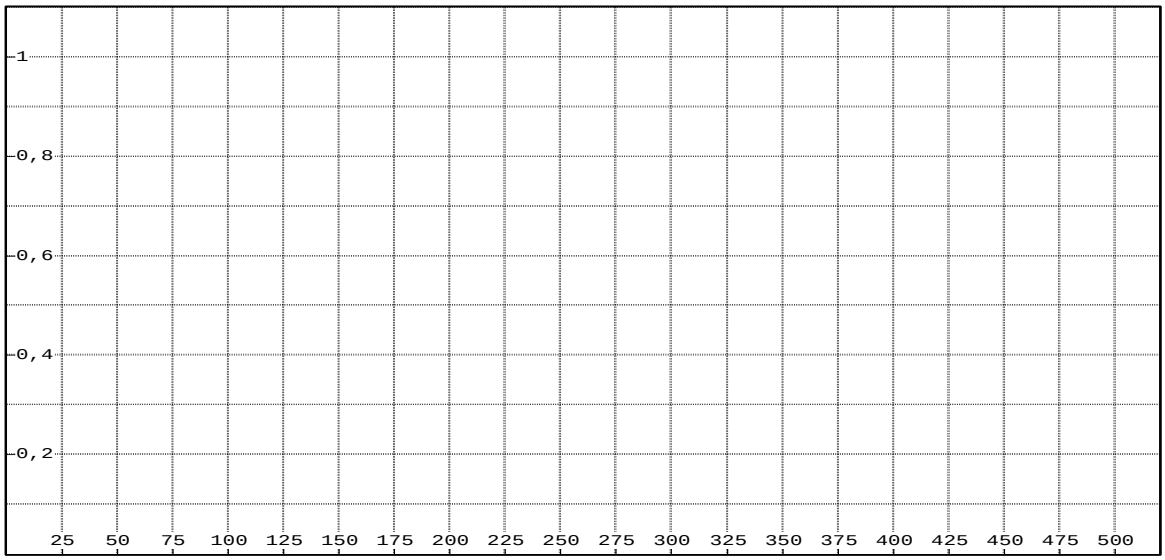
Brouillon de la fiche 2 :

Fiche 3 (noms, prénoms, lycée, classe, n° groupe)

Vos demandes de tirages et leurs résultats :

Fiche 4 (noms, prénoms, lycée, classe, n° groupe)

Le graphique (en abscisse le nombre de tirages et en ordonnée la fréquence de boules rouges)



Fiche 5 (noms, prénoms, lycée, classe, n° groupe)

Votre hypothèse sur la composition de votre bouteille :

Comment testez-vous cette hypothèse ?

Fiche 6 (noms, prénoms, lycée, classe, n° groupe)

Votre échantillon de 10 notes :

Comment l'avez-vous construit ?

Fiche 7 (noms, prénoms, lycée, classe, n° groupe)

Ce que vous auriez envie de demander à l'ordinateur en donnant vos raisons. En particulier quels calculs lui demanderiez-vous de faire à votre place ?

Fiche 8 (noms, prénoms, lycée, classe, n° groupe)

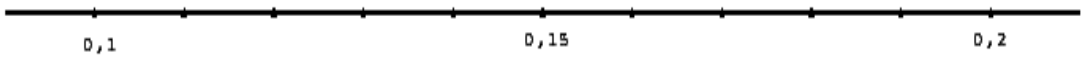
Ecrivez les résultats de la commande n°1 :Votre échantillon de 10 notes :

Valeurs	5	6	7	8	9	10
Fréquences						

Décidez de l'échantillon représentatif de l'ensemble des notes de la ville d'Hô Chi Minh

Fiche 9 (noms, prénoms, lycée, classe, n° groupe)

Placez sur la droite graduée les 20 fréquences de la valeur choisie (7 ou 9) :

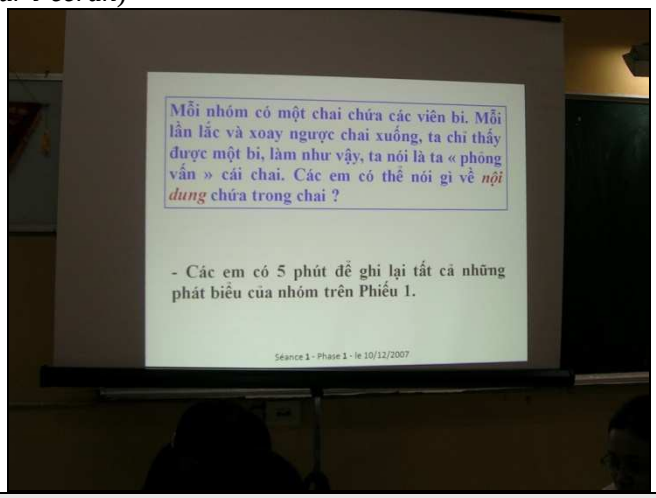


Combien de notes de cette valeur mettez-vous dans l'échantillon représentatif ? ☐

Justifiez votre réponse :

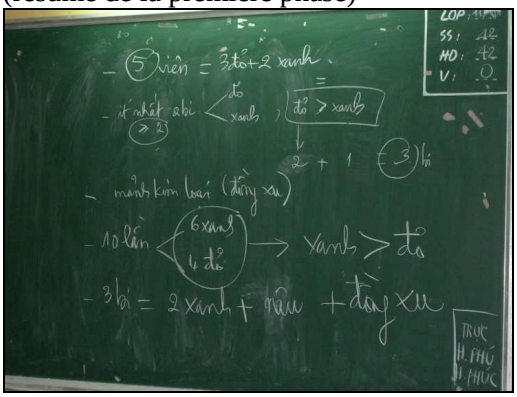
III. Scripts des séances

III.1. Script de la séance 1

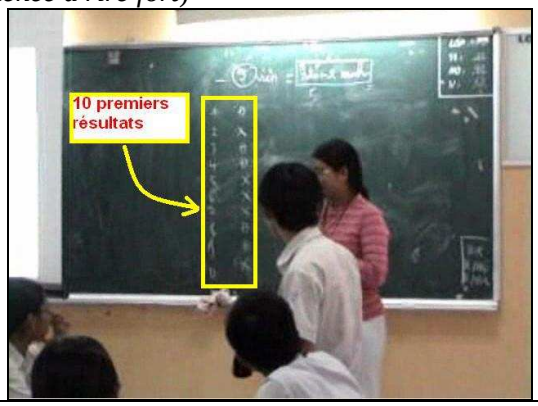
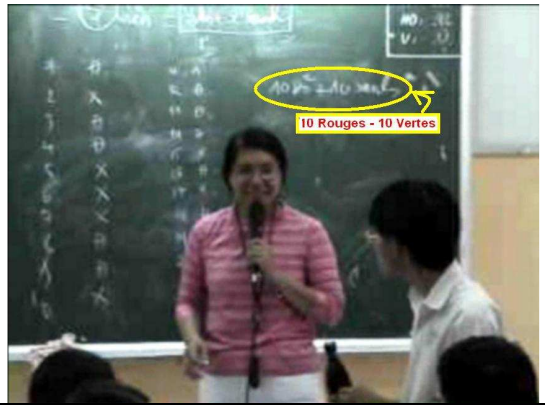
Phase 1		
	Ens	Aujourd'hui, nous allons travailler en suivant des activités. La première activité que je vous propose est celle-ci : (*l'enseignant prend une bouteille dans un sac plastique puis la montre en position debout devant la classe*) Je vais distribuer une bouteille comme celle-ci pour chaque groupe. Je voudrais que vous l'observiez bien puis vous me disiez ce qu'il y en a dans la bouteille. Dites-moi du contenu de votre bouteille. Allez-y ! Observez et notez bien ! Ne devez pas ouvrir le bouchon. Je l'ai bien fermé. Cherchez à tout faire pour me dire qu'est-ce qu'il y en a dans votre bouteille. Vous avez 5 minutes pour observer la bouteille. Essayez de dire qu'est-ce qu'il y a dans vos bouteilles. (*Après avoir distribué des bouteilles à chaque groupe, l'enseignant projette la consigne*[218] *sur l'écran*) 
Un groupe	El	Observe-t-on la couleur ?
	Ens	(*en rigolant*) Oui, si vos yeux sont … bons …
	El	Je vois la couleur rouge et verte, et la blanche aussi

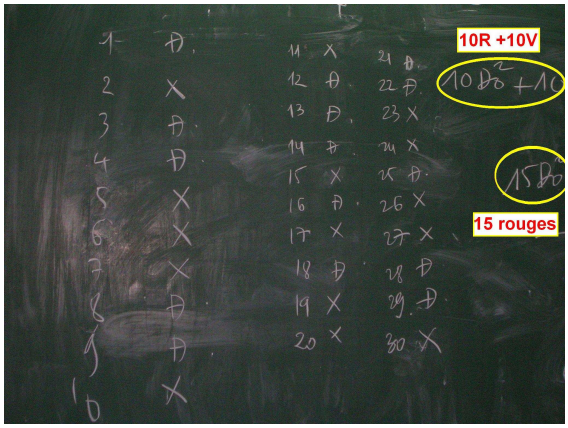
[218] « Chaque groupe dispose d'une bouteille contenant des boules. On ne peut voir qu'une boule à la fois : on dira que l'on interroge la bouteille. Que pouvez-vous dire sur le *contenu* de votre bouteille ? Vous avez 5 minutes pour noter toutes vos déclarations sur la feuille (1) que nous vous distribuons ».

	Ens	D'accord, observez bien ! (*Elève ne sait pas encore renverser la bouteille*) Pourquoi y-a-t il cette fente ?	
	El1	Ah, je ne savais pas qu'il y a une fente. (*Cet élève commence à renverser la bouteille*)	
	El2	Une boule rouge, une verte… Il y a deux rouges.	
Gr1	El	Il y a 5 boules.	
	Ens	Pourquoi ? Vous êtes sures ?	
	El	J'entends qu'il tombe 5 fois, je devine, …	
	Ens	5 minutes s'est passé !	
	Els	(*Un groupe se discute*) - Refaisons encore 10 fois ! - On a fait 5-5, puis 8-2. - Continuons ! - Alors, 3 fois ? - Non, pas de 3 fois, mais 10 fois	
	Ens	(*Conversation de l'enseignant avec un autre groupe*) Ens : Vous ne faites que 5 fois ? El : Non, plus que ça, 10 fois Ens : Pourquoi vous ne notez que 5 fois El1 : Il lit si vite, je ne peux pas noter tout El2 : Rouge, rouge, … il n'y a que des rouges	
Classe	Ens	Avez-vous fini ? Donnez-moi vos feuilles ! Bien notez le numéro du groupe. Je vais faire très vite un bilan des réponses :	
Classe	Ens	- *1er groupe*	« Dans la bouteille, il y a au moins deux boules de deux couleurs : rouge et verte ». (*un élève répond : c'est juste !)* S'il y a plus de deux boules alors la probabilité de voir une boule rouge est plus grande car parmi 10 expériences, nous avons vu 3 fois de boule verte et 7 fois de boule rouge ». Alors, ce groupe a conclut qu'il y a plus de rouges que de vertes.
		- *2ième groupe*	« dans la bouteille, il y a deux sortes de boules : rouges et vertes ». Alors, je n'ajoute pas au tableau. « Chaque fois de renverser la bouteille, nous voyons une seule boules, c'est la variable aléatoire discrète. Nous avons fait 20 rouges et 12 vertes. C'est possible qu'il y a deux boules rouges et une boule verte ». Donc, ce groupe a dit qu'il y a 3 boules tandis que le premier groupe a dit qu'il y a au moins deux boules.
		- *un autre groupe*	« dans la bouteille il y a des boules rouges et vertes. Le nombre des boules est plus grand ou égale à 2 ». Bon, c'est comme le premier groupe. « Le nombre des boules rouges est égale au nombre des boules vertes ». Alors, ce groupe dit que le nombre des rouges et celui des vertes sont égaux.
		- *encore un autre groupe*	« Il y a des boules rouges et vertes et plus de rouges que des vertes » Ce groupe *écoute* le son et dit qu'il y a des morceaux de métaux. (*la classe rit*) Et il devine que c'est des monnaies.

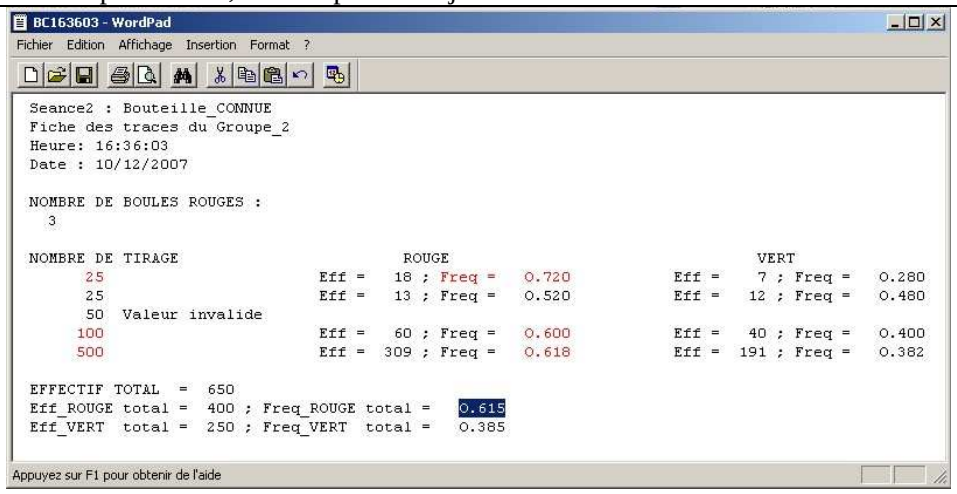
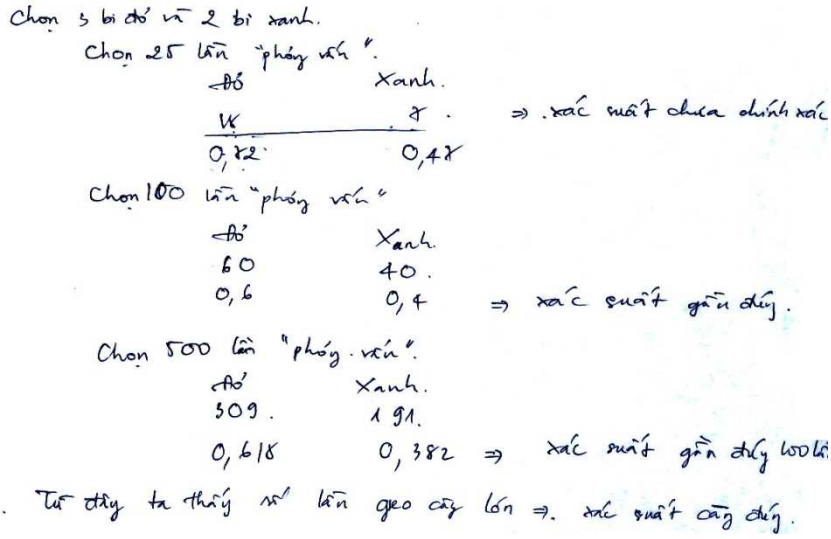
		- *un autre groupe*	« Il y a 3 boules ». Ce groupe a fait 10 fois, alors je vais compter, un, deux, trois, quatre, cinq, six. Il y a six fois de boule verte et 4 fois de boule rouge. Donc, ce groupe trouve qu'il y a plus de vertes que des rouges.
		- *un autre groupe* :	« Il y a 3 boules, deux vertes et une brune ». Ah, ce groupe a vu la couleur rouge mais il pense que c'est la couleur brune. Donc, je vais écrire rouge ? Non, je vais écrire brune comme ce que ce groupe a vu. En suite, je vois que ce groupe a écris : « il y a des sables » mais il est barré. Ça veut dire ce groupe a écouté le son de sable. Il dit aussi qu'il y a des monnaies.
		- *Le dernier groupe*	« Il y a des boules en pierre, de couleur rouge et verte, environ 5 boules » Ce groupe dit qu'il peut voir qu'il y a 3 rouges et 2 vertes. Bon, il y a un groupe qui est presque sur qu'il y a 5 boules.
		(résumé de la première phase)  	- 5 boulles = 3 rouges + 2 vertes - au moins 2 boulles : rouge ; verte ; rouge > verte 2 + 1 = 3 boules - morceau en métal (monnaie) - 10 fois : 6 vertes, 4 rouges → verte > rouge - 3 boules = 2 vertes + brune + monnaies
		Après la première activité, ce que vous avez « vu », ce sont : - Dans la bouteille, il y a des boules - La plus part des groupes disent qu'il y a deux couleurs : rouge et verte. - Certains groupes disent que telle couleur est plus que l'autre couleur. Par exemple les rouges sont plus nombreux que les vertes et vous avez noté vos tirages et comme vous avez vu que telle couleur apparaissent plus fréquentement que l'autre couleur. - Certains groupes écoutent le son des métaux et disent qu'il y a des monnaies. - Particulièrement, un seul groupe précise qu'il y a 5 boules : 3 rouges et 2 vertes. Ce groupe a montré ces tels chiffres. J'ai déjà interrogé ce groupe « comment faite-vous pour dire qu'il y a 5 boules ? ». Ce groupe m'a répondu qu'il a incliné la bouteille un peu puis chaque fois qu'une boule tombe au cou de la bouteille, il écoute et compte le son. Il a écouté cinq fois tomber la boule et il a dit qu'il y a 5 boules. Je l'ai demandé s'il est sur totalement de ce 5 boules et ce groupe m'a répondu qu'il est sur mais pas totalement 100%.		
Phase 2			
	Ens	(*Tenir une bouteille et la montrer devant la classe*) Êtes-vous sûr de vos annonces sur le contenu des bouteilles ? Maintenant, je vous précise que je sais que dans chaque bouteille…	
	Classe	Oh ! Oh !!!!!!!!!!!!!!! (*le groupe a gagné avec sa précision de 5 boules criait le plus fort*)	
	Ens	il y a 5 boules et qu'il y a deux couleurs : verte et rouge. C'est tout. Il n'y a pas de sable ni de pièces de monnaie. Je veux que vous essayiez de me dire combien de boules rouges et combien de boules vertes il y a dans la bouteille de votre groupe.	
	1 élève	3 rouges et 2 vertes	

	Ens	Combien de rouges, combien de vertes. Chaque bouteille porte une composition que je ne sais pas exactement. Essayez de chercher à dire cette composition. Je vous distribue la fiche 2 et sur laquelle vous allez écrire les résultats que vous avez faits pour chaque expérience. Et pour brouillon, vous avez une autre fiche. N'oubliez pas de noter le numéro de votre groupe sur les fiches. Vous avez 10 minutes pour cette activité. (*Tous les groupes commencent les tirages*)
	Un groupe	(*Un élève a fait 5 fois et demande un autre membre du groupe*) Fais un autre 5 fois…
	Ens	Ah, votre groupe fait de façon que chacun à son tour pour qu'il soit égalité ? C'est bien.
		(*enseignant fait le bilan des résultats des groupes à la fin de phase 2*) Groupe 1 : il y en a 3 rouges et 2 vertes, et vous êtes assez surement de ce que vous proposez. Groupe 2 : il y en a 3 rouges et 2 vertes, et vous êtes assez surement Groupe 3 : il y en a 3 rouges et 2 vertes, et vous êtes totalement sûrs. Groupe 5 : il y en a 4 rouges et 1 verte, et vous êtes assez sure. Groupe 4 : il y en a 3 rouges et 2 vertes, et vous êtes assez surement Groupe 6 : il y en a 3 vertes et 2 rouges, et vous êtes totalement sûrs. Groupe 7 : il y en a 3 vertes et 2 rouges, et vous êtes totalement sûrs. Groupe 8 : il y en a 4 vertes et 1 rouge, et totalement sûrs. Ce sont des résultats des groupes. Il y a certains groupes qui sont totalement surs. Lesquels ? Levez vos mains pour que la classe vous voie. Bon, ce groupe est totalement sur, ce groupe est totalement sur, ce groupe aussi. Je vais donc demander un de ces groupes de montrer devant la classe ce que vous avez fait. Comment faite-vous pour convaincre la classe ? (*L'enseignant choisit le groupe 3 qui dit fort qu'il est totalement sur*). Vous gardez encore votre bouteille, n'est-ce pas ? Vous avez confirmez que vous êtes surs de votre proposition de la composition de la bouteille, n'est-ce pas ? Pouvez-vous refaire des expériences pour convaincre la classe de votre réponse sure ? Alors, maintenant le groupe 3, je vous propose de redire votre affirmation de la composition dans votre bouteille.
	El	(cet *élève est représentant du groupe 3*) Notre groupe a proposé 3 rouges et 2 vertes et nous somme totalement surs.
	Ens	Bon, toute la classe, observez bien !
	El	Maintenant, je vais faire 10 fois de renverser la bouteille et je vais préciser la couleur de la boule qui est au cou de cette bouteille.
	Ens	Bon, ce groupe décide de faire 10 fois. Allez-y ! R – V – R (*troisième fois*) – R – V – V (*la classe commence à rire fort par ce résultat provisoire : moitié R et moitié V*) – V – R – R – V.

		(*la classe recommence à rire fort*) 
	Classe	(*crie*) Cinquante – cinquante
	Ens	Ah, après 10 fois, vous avez combien de rouges ? Un, deux, trois, quatre, cinq.
	El	Cinq rouges et cinq vertes.
	Ens	Ce cas, est-il possible de réalisation ?
	El	Non,
	Ens	Pourquoi ?
	El	Car il y en a 5 boules.
	Ens	(*Enseignant fait un geste de douter et étonner et parle au élève représentant du groupe*) Donc, qu'est-ce que vous décidez ?
	El	Faisons encore dix fois.
	Ens	D'accord, 11ième fois : V – R – R – R – V – R – V – R – V – V. Bon, cinq, six, sept, huit, neuf, dix. Dix rouges sur vingt. Ce groupe fait 20 fois et le résultat est pareil ! (*l'enseignant écrit sur le tableau le résultat final de 20 essais : 10 rouges – 10 vertes*) 
	Classe	C'est drôle !
	El	(*cet élève essaie de protéger la proposition de son groupe*)

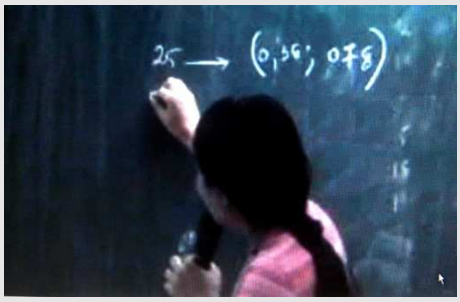
		Madame, si je ferais 50 fois, et je pourrais avoir plus de rouge que de verte, mais la différence ne serait pas très grande.
	Ens	Ah, tu pourras avoir plus de boules rouges que des boules vertes ? Toute la classe, on a fait 20 fois et on a 10 rouges 10 vertes. Voulez-vous faire encore ?
	Classe	Faisons encore !
	Ens	Ok, continue à faire expérience.
	Ens	R-R-V-V-R-V-V-R-R-V (*Toute la classe rit en applaudissant*)  Il semble que ce groupe n’est pas content … (*La classe rit encore*) Quand vous dîtes que vous êtes surs totalement, vous devez avoir quelque chose pour votre confiance. Vous avez fait un certain nombre de fois et compté des rouges et des vertes. Vos résultats vous conduit à conclure que le nombre des boules de telle couleur est plus que celui de l’autre couleur. Mais maintenant, vous refaites devant la classe 30 fois de façon : 10 premières expériences, vous avez le même rapport de deux couleurs, puis pour 20 expériences, alors le même rapport, et ensuite, 30 expériences, encore le même rapport. Bon, réfléchissez-vous maintenant ?
	El	Non, toujours 3 rouges - 2 vertes.
	Ens	Que penses-tu faire pour convaincre toute la classe d’accepter la proposition de la bouteille de ton groupe ?
	El	Il faut faire beaucoup plus des tirages, par exemple environ 1000 tirages
	Ens	Bon, d’accord. On va prendre une pause de quelques minutes puis on va passer à une autre activité.

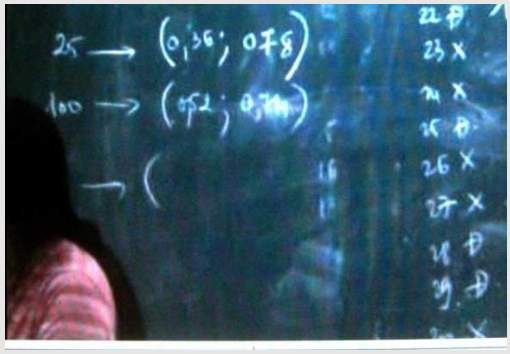
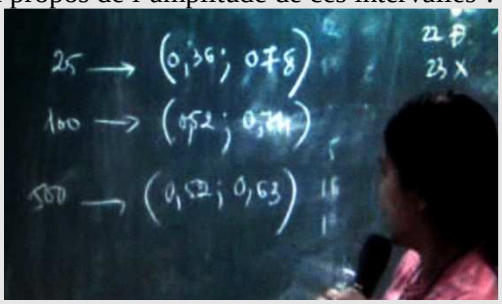
III.2. Script de la séance 2

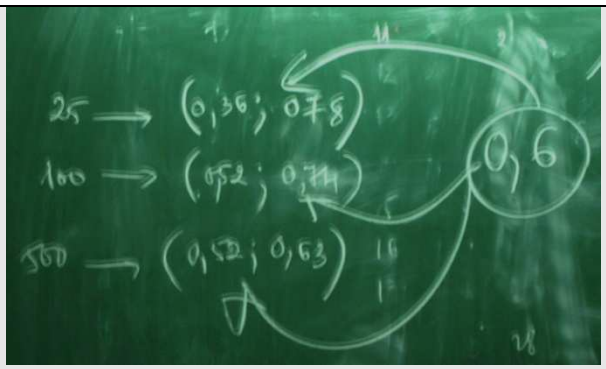
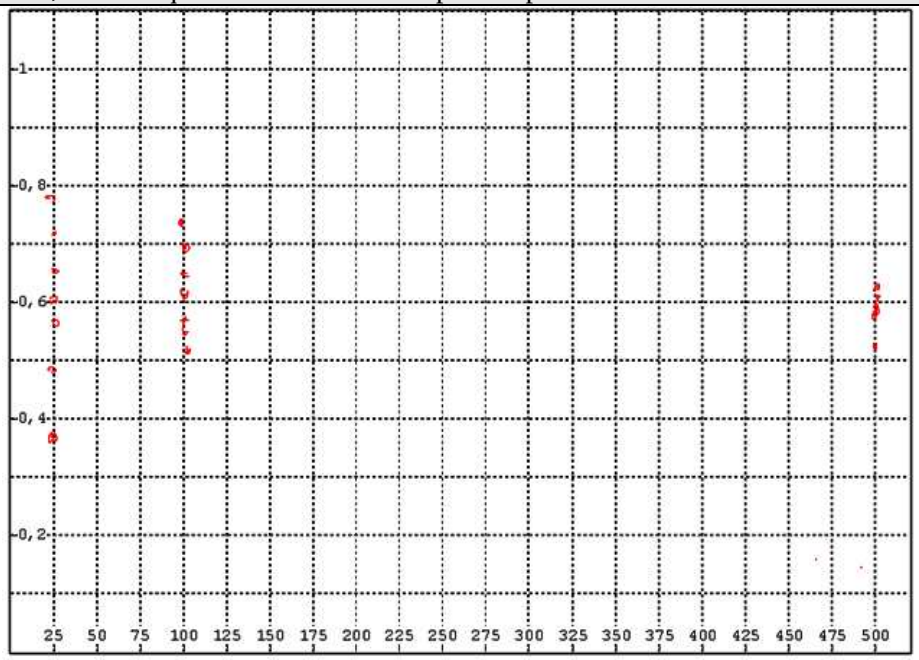
Phase 1		
Gr2	El1	Est-ce que je dois dire de l'exactitude ? Je vois que plus le nombre des tirages augmente plus l'incertitude est exact.
	Ens	Observe les résultats, puis tu donne tes remarques. Je ne veux pas que vous écrivez seulement « exact » ou « inexact ». Qu'est-ce que vous êtes en train d'observer ? (*L'enseignant regarde les notes dans la fiche 3 du groupe*) Ah, vous dites que : « plus le nombre des tirages est grand, plus la certitude est grande ». Je veux savoir avec quoi vous avez comparé pour dire qu'il est exact ?
	El2	… avec la fréquence …
	Ens	(l'enseignant regarde l'écran d'ordinateur et demande) - ok, je vois 0,618, comment fais-tu pour dire qu'il est plus exact qu'auparavant ? par exemple, 25 fois, rouge 0,72 ; 100 fois 0,6 ; 500 fois, 0,618. Avec quoi vous avez comparé ?
	El2	Il est proche de ce qui est au-dessus.
	El1	(cet élève ajoute un argument) Notre groupe a choisi 3 rouges
	Ens	Quelle est la différence entre le cas de 100 fois et celui de 500 fois ?
	El2	Ils sont pareils. Ah, notre réponse est juste…
		 BC163603 - WordPad Fichier Edition Affichage Insertion Format ? Seance2 : Bouteille_CONNUE Fiche des traces du Groupe_2 Heure: 16:36:03 Date : 10/12/2007 NOMBRE DE BOULES ROUGES : 3 NOMBRE DE TIRAGE — ROUGE — VERT 25 — Eff = 18 ; Freq = 0.720 — Eff = 7 ; Freq = 0.280 25 — Eff = 13 ; Freq = 0.520 — Eff = 12 ; Freq = 0.480 50 Valeur invalide 100 — Eff = 60 ; Freq = 0.600 — Eff = 40 ; Freq = 0.400 500 — Eff = 309 ; Freq = 0.618 — Eff = 191 ; Freq = 0.382 EFFECTIF TOTAL = 650 Eff_ROUGE total = 400 ; Freq_ROUGE total = 0.615 Eff_VERT total = 250 ; Freq_VERT total = 0.385 Appuyez sur F1 pour obtenir de l'aide (*Fichier de trace du groupe 2 de Bouteille-Connue*)  Chọn 3 bi đỏ và 2 bi xanh. Chọn 25 lần "phóng vén". Đỏ Xanh. 18 7 ⇒ xác suất chưa chính xác 0,72 0,28 Chọn 100 lần "phóng vén" Đỏ Xanh. 60 40. 0,6 0,4 ⇒ xác suất gần đúng. Chọn 500 lần "phóng vén". Đỏ Xanh. 309. 191. 0,618 0,382 ⇒ xác suất gần đúng hơn ⇒. Từ đây ta thấy số lần gieo càng lớn ⇒. xác suất càng đúng. (*Fichier 3 du groupe 2*)

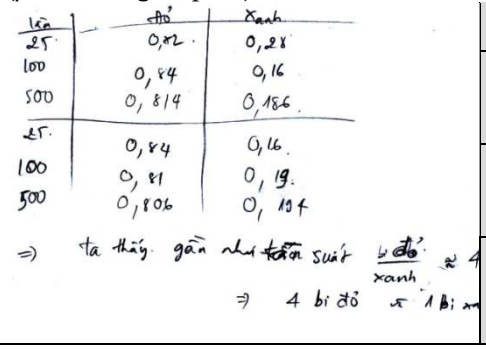
		Choisir 3R et 2V : 25 fois « d'interroger » la bouteille : Rouge Verte 18 7 ⇒ la probabilité n'est pas encore exacte 0,72 0,47 100 fois « d'interroger » la bouteille : Rouge Verte 60 40 ⇒ la probabilité est approximative 0,60 0,40 500 fois « d'interroger » la bouteille : Rouge Verte 309 191 0,618 0,382 ⇒ la probabilité est approximative comme dans le cas de 100 fois (*Traduction en français de fiche 3 du groupe 2*)
Classe	Ens	Bon, on s'arrête la première activité ici. Vous savez maintenant comment ça marche le programme et vous avez noté les résultats quand le programme fonctionne. Vous avez demandé la machine, et elle vous a répondu. Cela dépend du nombre de fois que vous voulez. Par exemple vous voulez 25 fois, 100 fois, 500 fois. Et vous avez vos résultats. Donnez-moi vos fiches, SVP. Arrêtez tous vos programmes. Maintenant, on va travailler ensemble. Alors, je vais interviewer chaque groupe. Groupe 1, combien de boules rouges avez-vous choisi ?
	Gr1	3R
	Ens	3R… Groupe 2 ?
	Gr2	3R
	Ens	3R… Groupe 3 ?
	Gr3	3R
	Ens	Oh mon dieu… Groupe 4 ?
	Gr4	3R
	Ens	Groupe 5 ?
	Gr5	4R
	Ens	Groupe 6 ?
	Gr6	2R
	Ens	Groupe 7 ?
	Gr7	2R
	Ens	Groupe 8 ?
	Gr8	1R
	Ens	Bon, vous avez choisi des …. des rapports différents des boules entre des groupes différents. Il est comme si vous avez des bouteilles différentes mais cette fois-ci, vous avez déjà connu le nombre de boules rouges et celui de boules vertes dans votre bouteille, la bouteille-ordinateur, et vous avez observé les fréquences d'apparitions des boules rouges et vertes correspondant au nombre de réalisation des expériences aléatoires. Maintenant, je propose que toute la classe choisisse un même nombre de boules rouges. Par exemple …
	El1	3R, 3R…
	Ens	3R ???
	El2	4R…
	Ens	(*comme il y a beaucoup d'élève voulant 3R, l'enseignant décide de choisir 3R*)

		Bon, on est d'accord pour choisir 3R. Maintenant, tous les groupes refont fonctionner le même programme avec le même nombre des rouges. Tous les groupes font pareillement. Puis vous allez noter les résultats sur vos fiches 4. Après, on va faire ensemble un bilan des résultats.
	1élève	3R, n'est-ce pas, madame ?
	Ens	Oui, 3R … Bon, toute la classe, écoutez moi, la nouvelle fiche que je vous distribue n'est pas pour noter les fréquences d'apparition. Écoutez moi, je vous propose comme suit : premièrement, on est d'accord de choisir 3 rouges ; deuxièmement, chaque fois où votre programme fonction, vous choisissez un nombre de tirages comme vous voulez, soit 25 fois, soit 100 fois, soit 500 fois. Pour les résultats, on… ne s'intéresse… que…. la fréquence… de boule rouge…, seulement de boule rouge. On choisit 3 R et on s'intéresse seulement la fréquence d'apparition des boules rouges. Puis vous allez tracer un point sur la fiche de graphique, ce point a des coordonnés correspondant au nombre de tirages, c'est-à-dire, … regardez l'axe des abscisses, vous voyez 25, 100, 500 ; et l'ordonnée sera la fréquence donnée par la machine. Ça va ? Vous pointerez un point chaque fois le programme fonctionne. Tous les groupes, vous comprenez ce que vous allez travailler ? avez-vous besoin que je répète encore une fois ? Oui ? Bon, premièrement, 3 rouges ; deuxièment, tracer un graphique en plaçant des points de coordonnées : nombre de fois, fréquence des rouges.
	Un élève	(*cet élève veut choisir un nombre quelconque des tirages*) Madame, mais il n'y a que 25, 100 et 500 fois.
	Ens	Oui, on fait le choix parmi 25, 100 ou 500. Pas d'autre choix. Encore des questions ? Non ? Ok, allez-y.
Un groupe	El	Madame, chaque fois de 25 tirages, j'ai un résultat différent que le précédent, qu'est-ce que je dois alors faire ?
	Ens	Ah, tu fais un point pour chaque fois. (*l'enseignant parle à toute la classe*) Un groupe a fait plusieurs fois de 25 tirages et obtenu des résultats différents. Il m'a demandé comment faire. Alors, je vous propose de placer autant de points différents correspondant au nombre de vos résultats.
Un groupe	El	Est-ce que l'on doit joindre des points ?
	Ens	Non, tu ne dois que tracer des points.
Gr4	El1	Madame, est-ce que quand le nombre des tirages est faible, la fréquence n'est pas exacte ; et quand le nombre des tirages est plus grand, la certitude est grande aussi, n'est-ce pas ?
	Ens	Vous observez et vous le trouvez ? Très bien,
	El2	Continuons…, ceci doit être 0,6 ; mais on a 0,63, ça veut dire il est plus 0,03 ; 25 fois, on a 0,54,…, 0,76 ; …
15:42		
Gr4	El1	La distance entre les points…
	El2	… est-ce que je dois joindre ces points, madame ?
	El1	Non, ne les joins pas
	Ens	Vous pouvez faire ce que vous voulez. Si vous voulez les joindre, vous le faites comment ?
	El2	C'est impossible de joindre ces points…
	El1	Si, ceci est plus haut, joins le plus haut avec le plus haut, le plus bas avec le plus bas…
	El2	D'accord, joins les deux pour voir la distance, pour voir que plus le nombre des tirages est grand, plus la distance est petite…
	Ens	Un intervalle de 0,48 à 0,76. Puis, l'intervalle, il devient plus étroit, mais il est proche de quelle valeur ?

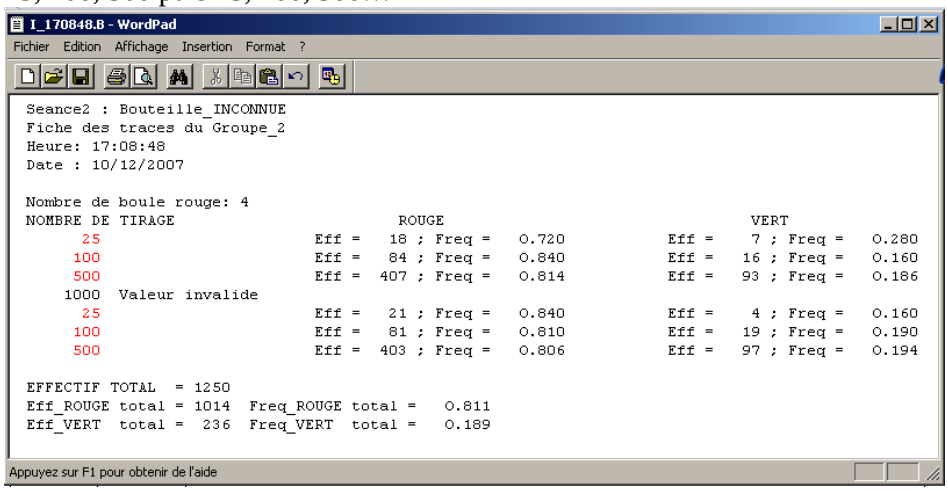
	Els	0,6
	Ens	L'autre est plus petit, nous avons combien de rouges, 3 n'est-ce pas ? Quelle est la probabilité pour voir une boule rouge ? 0,6 ? Vous voyez, tous ces intervalles contiennent toujours cette valeur 0,6 ; mais l'un est grand, l'autre est petit. Si on nous demande de trouver cette valeur 0,6 dans un grand intervalle, ce n'est pas facile à la trouver.
	Els	0,48,…. 0,6
Classe	Ens	Maintenant, tous les groupes, avez-vous fini ?
	Els	O…u…i …
	Ens	Bon, Pouvez-vous me donner vos fiches 4 ?
	Els	O…u…i …
19 :48	Ens	Le groupe 7 a dit que sa fiche est si b…e…l…l…e… (*L'enseignant ramasse toutes les fiches 4*) Regardez au tableau, je vais tracer tous les points obtenus par tous les groupes. D'abord, je vais prends certains points pour le cas de 25 tirages, dans la fiche de groupe 4, je vois des points ; je vais tracer comme ce que je vois.
20 :26	Ens	- Il y a un point ici, (*l'enseignant trace un point comme ce qu'on voit sur l'affiche au tableau*) ; un point est exactement au 0,6 ; un point plus bas, environ zéro virgule quarante… hum,…, quarante… combien ? Je vois ces trois points. - Un autre groupe, je vois un point très proche de 0,8 ; il y a trois points : un est au-dessus et un est au-dessous de 0,6 ; un point près de 0,7 ; 0,68 ; un point ici, je vois que ce groupe a écrit 0,36 à côté… Je vais remarquer le point au plus haut et celui est au plus bas. Maintenant je passe aux points correspondant à 100 fois, je vois un point environ 0,65 ; un autre plus proche, encore un autre proche,… le plus bas est à 0,5…, un point 0,7… Enfin, je vois 0,594 (*écrit*), un point tout près de 0,6 ;
Classe (24:15 → 28:29)	Ens	Regardez au tableau, SVP. Si on regarde 25 fois, le groupe 1 a remarqué que cet intervalle est large, rappelons que tous les groupes ont choisi 3 boules rouges, donc, la probabilité d'une boule rouge calculée par la formule vous donne combien ?
	El	soixante
	Ens	0,6. Vous voyez, c'est un intervalle de 0,36 à … zéro virgule … combien ? … 0,78, je vais noter sur le tableau… Quand nous faisons un tel nombre des 25-tirages, nous avons l'intervalle]0,36 ; 0,78[;  (*premier intervalle*) pour 100 fois, nous avons des points …. comment ?
	Els	Plus proches…
	Ens	Ah, d'abord, je vais vous poser la question suivante : pourquoi avec la même machine, le même programme, le même choix de 3 boules rouges, pour la première fois où on demande 25 tirages, on obtient un certain résultat, puis la deuxième demande de 25-tirages, on obtient au autre résultat ; ces résultats sont différents… Pourquoi ?
	El	Parce qu'un 25-tirages n'est pas exacte

		Ah, pourquoi ? pourquoi un 25-tirages n’est pas exacte ?
	El	C’est parce que le nombre de tirages est petit.
	Ens	Ah, votre ami dit que c’est à cause de petit nombre des tirages. Bon, vous voyez, le groupe utilise une machine commune, fait fonctionner le même programme, avec le choix de 3 boules rouges, quand ce groupe fait 2 ou 3 fois un 25-tirages, il obtient des résultats différents, ce phénomène est appelé la fluctuation d’échantillonnage. Pour les 100-tirages, ou les 500-tirages, il y a aussi une fluctuation d’échantillonnage. Mais pour les 100-tirages, qu’est-ce que l’on a pour l’intervalle de valeur le plus base et la plus haute ? (*les fréquences sont sous-entendues ici*)
	El	Entre 0,52 …
	Ens	Et combien ?
	El	et 0,74
	Ens	D’accord, 0,74.  Puis pour des 500-tirages ?
	El	0,52 et 0,63
	Ens	Oui, environ 0,63 ou 0,64, n’est-ce pas ? Que trouvez-vous à propos de l’amplitude de ces intervalles ?  (*l’enseignant écrit au tableau les intervalles*)
	Els	L’amplitude est grande puis elle devient petite,
	Ens	Ah, ça veut dire que … plus le nombre des tirages est grand alors… plus l’amplitude est …
	Els	elle est petite…
	Ens	Et qu’est-ce que vous remarquez ?
	Els	Plus on fait des tirages, plus il est exact …
	Ens	Hm… c’est la certitude de quoi ?
	El1	… plus la certitude est grande.
	El2	De la probabilité …
	Ens	Est-ce que vous avez la probabilité ici ?
	Els	Oui, 0,6…
	Ens	D’accord, vous savez déjà le rapport des boules rouges et vertes. Alors, d’après la formule apprise, vous trouvez que la probabilité est 0,6. Est-ce que tous ces intervalles contiennent cette probabilité ?

		 (*l’enseignant écrit au tableau les intervalles contenant la valeur 0,6*)
	Els	Oui,
	Ens	Ah, on voit que tous ces intervalles contiennent 0,6 qui appartient à cet intervalle, à cet intervalle et à cet intervalle aussi. Et on voit que les intervalles se réduisent de plus en plus … Bon, suppose que l’on ne sache pas quelle est la valeur de la probabilité d’apparition d’une boule rouge, mais on sait qu’elle appartient à cet intervalle, est-ce que vous pouvez préciser la probabilité ? Est-il facile à donner la probabilité ? (*l’enseignant montre l’intervalle formé par des 25-tirages*),
	Els	non
	Ens	… à cet intervalle ? (*l’enseignant montre l’intervalle formé par des 100-tirages*),
	Els	Un peu…
	Ens	et à cet intervalle (*l’enseignant montre l’intervalle formé par des 500-tirages*),
	Els	Plus facile …
	Ens	Ah, les intervalles deviennent plus étroits… et le dernier est le plus étroit, et il semble que les valeurs à choisir dans ce dernier intervalle est moins nombreux que dans les précédents. On voit que la probabilité se trouve dans les intervalles qui sont de plus en plus étroit. Et on voit aussi que plus on fait un grand nombre des tirages, plus la fréquence est approchée la probabilité. Bon, c’est ce que vous avez obtenu après ce premier travail de cette séance.
		 (*Fiche 4 du travail public devant la classe*)

		Phase 2
Classe 28:30	Ens	Nous avons encore un deuxième programme dans l’ordinateur. Dans ce programme, je ne sais pas qu’il y a combien de boules rouges, de boules vertes. Donc, cette fois-ci, nous n’avons pas de droit de choisir le nombre des boules rouges. La situation est comme suit : imaginez que vous avez une bouteille, comme la bouteille dans la séance 1, contenant 5 boules de deux couleurs rouge et verte, et vous ne savez ni le nombre de boules rouges et ni celui de boules vertes. Tout à l’heure, on a renversé la boule puis on a compté le nombre de fois de boules rouges et celui de boules vertes, maintenant vous utilisez le programme pour interroger la bouteille-ordinateur, comme vous avez fait avec le programme en sachant le nombre de boules rouges, mais cette fois-ci, on nen sait pas le nombre de boules rouges. Vous allez faire un tel nombre de fois comme vous voulez puis vous essayerez de me donner votre proposition de la composition des boules. D’accord ?
	Els	Oui
	Ens	Vous devez discuter dans votre groupe pour me dire qu’il y a combien de boules rouges. Vous allez me dire que le programme dans votre ordinateur est correspondant à une bouteille ayant combien de boules rouges. Vous devrez écrire dans la fiche 5 distribuée comment vous faites pour proposer votre hypothèse d’un tel nombre de boules rouges et d’un tel autre nombre de boules vertes. Vous comprenez ce que vous allez faire ?
	Els	O…u…i…
31:10	Ens	Bon, allez-y (*Tous les groupes travaillent*) Notez tous ce que vous avez. N’oublie pas de donner la raison pourquoi vous pouvez proposer votre hypothèse sur la composition.
Classe 36:29	Ens	Bon, toute la classe, ça y est-ce ?
	Els	Oui,
	Ens	Bon, groupe 2, vous avez dit que vous êtes sûr, alors quel est votre proposition de la composition des boules ?
	Gr2	4R et 1V (*fiche 5 du groupe 2*)  Nb. de fois | Rouge | Verte 25 | 0,72 | 0,28 100 | 0,84 | 0,16 500 | 0,814 | 0,186 25 | 0,84 | 0,16 100 | 0,81 | 0 ,19 500 | 0,806 | 0,197 Il est presque fréquence Rouge/Verte ≈4 ⇒ 4 boules rouges et 1 boule verte
	Ens	Comment faites-vous pour l’assurer ?
	Gr2	Le nombre d’apparitions d’une boule rouge divise par le nombre d’apparitions d’une boule verte, ça fait presque 4.
	Ens	Ah, vous avez divisé l’effectif des rouges par l’effectif des vertes. Et vous avez fait combien de *fois* ? (*l’enseignant veut demander le nombre total des tirages*)
	Gr2	4 ou 5 *fois* *(l’élève parle de l’interrogation de la machine)*
	Ens	Oh, 4 ou 5 fois ?
	Gr2	6 *fois*
	Ens	Qu’est-ce que ça veut dire ?

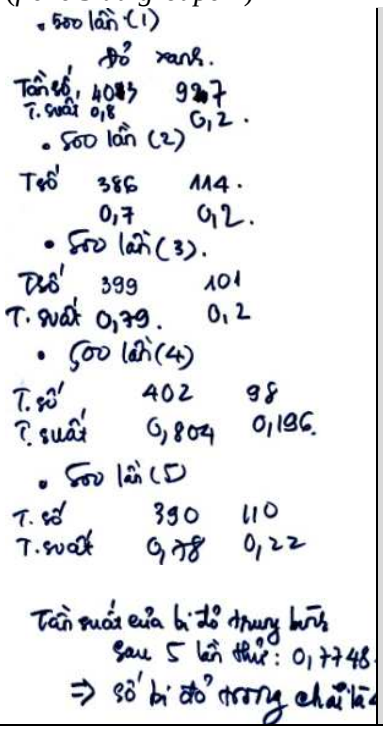
	Gr2	Ah, 6 fois d’interroger l’ordinateur…
	Ens	Bon, chaque fois que vous interrogez la machine, vous avez choisi combien ? 25 ou 100 ou 500 ?
	Gr2	25, 100, 500 puis 25, 100, 500…



```
I_170848.B - WordPad
Fichier  Edition  Affichage  Insertion  Format  ?

Seance2 : Bouteille_INCONNUE
Fiche des traces du Groupe_2
Heure: 17:08:48
Date : 10/12/2007

Nombre de boule rouge: 4
NOMBRE DE TIRAGE                    ROUGE                          VERT
      25                  Eff =   18 ; Freq =   0.720      Eff =    7 ; Freq =   0.280
     100                  Eff =   84 ; Freq =   0.840      Eff =   16 ; Freq =   0.160
     500                  Eff =  407 ; Freq =   0.814      Eff =   93 ; Freq =   0.186
    1000  Valeur invalide
      25                  Eff =   21 ; Freq =   0.840      Eff =    4 ; Freq =   0.160
     100                  Eff =   81 ; Freq =   0.810      Eff =   19 ; Freq =   0.190
     500                  Eff =  403 ; Freq =   0.806      Eff =   97 ; Freq =   0.194

EFFECTIF TOTAL  = 1250
Eff_ROUGE total = 1014  Freq_ROUGE total =   0.811
Eff_VERT  total =  236  Freq_VERT  total =   0.189

Appuyez sur F1 pour obtenir de l'aide
```

(*Fichier de trace du groupe 2 de Bouteille-Inconnue*)

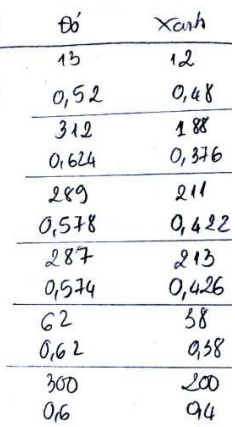
	Ens	Ah, vous faites 25, 100, 500 puis vous refaites 25, 100, 500, et vous le faites plusieurs fois comme ça ? Et enfin, vous avez un rapport 4 contre 1 ? Donc, 4 rouges et 1 verte.
	Gr2	Oui…
	Ens	Bon, groupe … 7, combien ?
	Gr7	4R et1V (*fiche 5 du groupe 7*)

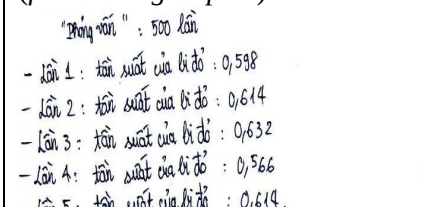


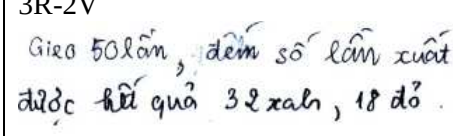
500 (1è fois)		
	Rouge	Verte
Eff	403	97
Fr	0,8	0,2
500 (2è fois)		
Eff	386	114
Fr	0,7	0,2
500 (3è fois)		
Eff	399	101
Fr	0,79	0,2
500 (4è fois)		
Eff	402	98
Fr	0,804	0,196
500 (5è fois)		
Eff	390	110
Fr	0,78	0,22

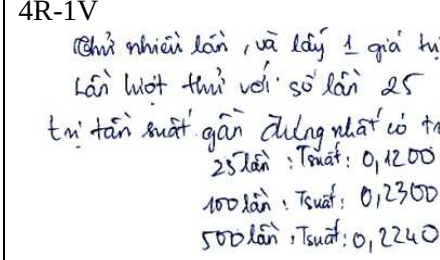
Moyenne des fréquences après 5 fois : 0,7748
⇒ Le nombre des boules rouges dans la bouteille est 4

	Ens	Oh, 4R-1V aussi ? Quel est votre dernier rapport ? Vous vous souvenez ?
	Gr7	0,7748
	Ens	(*un élève du groupe 3 dit avec les autres membres :* « la nôtre est sûrement 3R-2V ») Groupe 3 ?
	Gr3	3R-2V ! (*fiche 5 du groupe 3*)

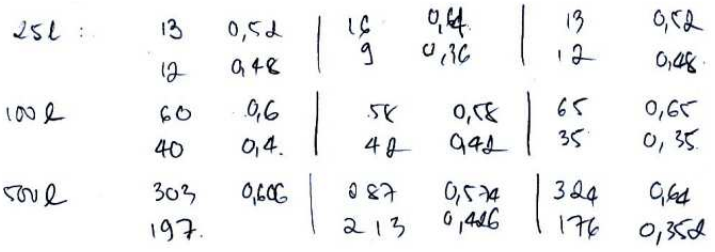
<table>
<tr><td></td><td></td><td>


<table>
<tr><td>Boule</td><td>Rouge</td><td>Verte</td></tr>
<tr><td>Effectif</td><td>13</td><td>12</td></tr>
<tr><td>Fréquence</td><td>0,52</td><td>0,48</td></tr>
<tr><td></td><td>312</td><td>188</td></tr>
<tr><td></td><td>0,624</td><td>0,376</td></tr>
<tr><td></td><td>289</td><td>211</td></tr>
<tr><td></td><td>0,578</td><td>0,422</td></tr>
<tr><td></td><td>287</td><td>213</td></tr>
<tr><td></td><td>0,574</td><td>0,426</td></tr>
<tr><td></td><td>62</td><td>38</td></tr>
<tr><td></td><td>0,62</td><td>0,38</td></tr>
<tr><td></td><td>300</td><td>200</td></tr>
<tr><td></td><td>0,6</td><td>0,4</td></tr>
</table>
</td></tr>
<tr><td></td><td>Ens</td><td>Ah encore fidèle à votre proposition 3R-2V ?</td></tr>
<tr><td></td><td></td><td>(l’élève du groupe 3 rit, puis une courte conversation en voix faible du groupe)
El2 : - mais je ne suis pas sûre ...
(cette élève dit faiblement)
El3 : - quoi ?
(El3 n’accepte pas hésitation de l’E2)
El1 : - non, ayons confiance sur nous-mêmes, il faut être calme
(El1 essaye d’assurer leur proposition)</td></tr>
<tr><td></td><td>Ens</td><td>Et alors, groupe 5 ?</td></tr>
<tr><td></td><td>Gr5</td><td>4R-1V
(fiche 5 du groupe 5)


« Interroger » : 500 fois
1è fois : fréquence de boule rouge : 0,598
2è fois : fréquence de boule rouge : 0,614
3è fois : fréquence de boule rouge : 0,632
4è fois : fréquence de boule rouge : 0,566
5è fois : fréquence de boule rouge : 0,614
⇒ fréquence ∈ [0,566 ; 0,632]
⇒fréquence de boule rouge ≈ 0,6
⇒ il y a 3 boules rouges et 2 boules vertes.</td></tr>
<tr><td></td><td>Ens</td><td>Groupe 6 ?</td></tr>
<tr><td></td><td>Gr6</td><td>3R-2V


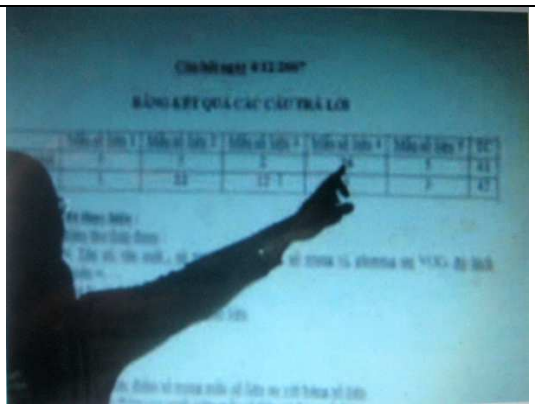
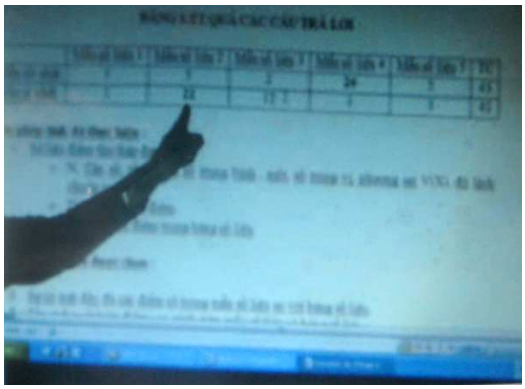
Lancer 50 fois, compter le nombre d’apparition des boules vertes, et celui de boules rouges, on obtient le résultat de 32 vertes et 18 rouges.
Le rapport verte/rouge ≈ 3/2</td></tr>
<tr><td></td><td>Ens</td><td>Groupe 8 ?</td></tr>
<tr><td></td><td>Gr8</td><td>4R-1V


Nous essayons plusieurs fois en prenant la fréquence de boule verte.
Nous faisons 25 fois, 100 fois, 500 fois ; puis les comparons et en fin, nous prenons la fréquence la plus approchée
25 fois : fréquence 0,1200</td></tr>
</table>

		100 fois : fréquence 0,2300 *500 fois : fréquence 0,2240*
	Ens	Groupe 1 ?
	Gr1	3R-2V  25 l : 13 0,52 | 16 0,64 | 13 0,52 12 0,48 | 9 0,36 | 12 0,48 100 l 60 0,6 | 58 0,58 | 65 0,65 40 0,4 | 42 0,42 | 35 0,35 500 l 303 0,606 | 287 0,574 | 324 0,64 197 | 213 0,426 | 176 0,352
	Ens	Bon, quand vous dites que, par exemple, vous avez 4R-1V, c'est-à-dire vous décidez de dire que votre programme est correspondant à la composition de 4R-1V. Bon, je veux savoir si vous êtes totalement sûr de vos propositions ?
	certains groupes	- Totalement sûr … - Assez sûr, - Moyennement,…
Gr3	El1	(*une conversation du groupe 3 enregistré*) Sûrement … 100% …
	Ens	Comment vous êtes sûr 100% ?
	El2 :	Pas sûr … (*rit*)
Classe 39 :45	Els	On fait et on voit …
	Ens	Ah, on fait et on voit …, et avec la machine ? Non ? Effectivement, quand on décide quelque chose, par exemple, on dit qu’on a 4R-1V. C'est-à-dire on va annoncer : « je dis qu’il y a 4R et 1V ». C’est qu’on a fait une décision. Mais quand on fait une décision, il y a toujours un tel risque. C'est-à-dire il y a une part incertaine, un tel pourcentage incertain, qui n’est pas vraiment exact. Donc, le risque combien vaut-il ?... Bon, je vous remercie de participer aux séances aujourd’hui. Ecoutez-moi, mémorisez ce que vous avez fait, gardez les mêmes groupes pour demain, gardez les tables regroupés. N’oubliez pas d’apporter vos calculatrices pour les utiliser s’il est nécessaire demain.

III.3. Script de la séance 3

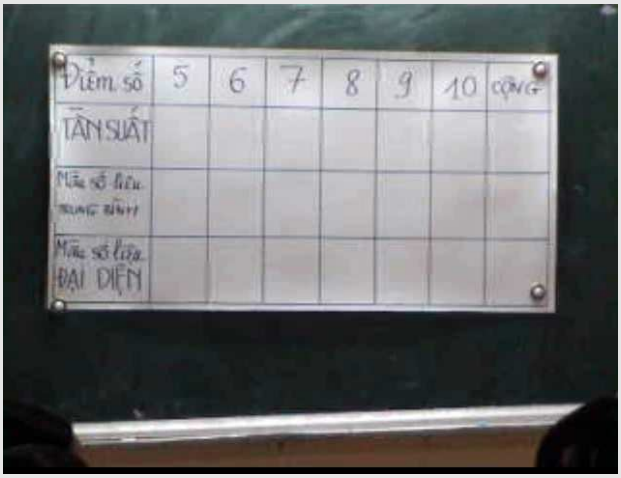
Phase 1		
	Ens	Qui peut me dire de ce que nous avons fait hier ?
	El	On a travaillé avec des bouteilles et on a cherché à savoir le nombre des boules rouges et celui des boules vertes.
	Ens	Bon, comment avez-vous fait pour proposer la composition ?
	El	On a calculé la probabilité,
	Ens	Comment avez-vous calculé la probabilité ?
	El	On a calculé des fréquences,
	Ens	Oui, et pour pouvoir calculer les fréquences, qu'a-on du faire ?
	El	On fait des expériences …
	Ens	Bon, très bien, on a fait des expériences. Notre expérience est de renverser la bouteille pour voir la boule à son goulot. Combien d'expériences a-on fait ?
	El	Beaucoup …
	Ens	D'accord, on a fait beaucoup d'expériences. Bon, pouvez-vous me dire ce que vous avez marqué lorsque vous faites plusieurs d'expériences par rapport à peu d'expériences ?
	El	C'est que le niveau d'exactitude de la proposition de la composition est plus grand quand on fait plusieurs d'expériences.
	Ens	Bon, hier, vous avez observé la bouteille. Au début, vous n'êtes rien su de son contenu, puis en sachant que la bouteille contient 5 boules, vous avez renversé la bouteille plusieurs fois pour compter le nombre d'apparition des boules rouges par rapport au nombre total des répétitions, puis donné vos premières conclusions. Comme vous avez trouvé qu'il faut un grand nombre des expériences, Ensuite, avec quoi vous avez fait la répétition ?
	El	Ordinateur
	Ens	Oui, comme vous trouvez qu'il fait du temps pour la répétition et quand vous voulez un grand nombre de répétition, vous avez refait grâce à une machine et à un programme de simulation, la bouteille-ordinateur, en suivant deux étapes : la première étape avec une composition connue ; et la deuxième étape, vous n'avez plus la bouteille et ni les boules, et grâce à la machine, vous avez des fréquences sans savoir le nombre des boules rouges et celui des boules vertes. la composition.
[…]		
05:00	Ens	Qui peut résumer le sujet de l'exercice que je vous ai demandé de faire à la maison la semaine dernière ? (*Une élève résume l'exercice 4 en parole*) (Enseignant projette sur l'écran le sujet de l'exercice 4 avec deux premières questions)
07:00		Je projette maintenant les résultats de vos réponses. La plupart parmi vous a choisi l'échantillon 4 comme le plus représentatif, il y a 26 réponses choisissant cet échantillon.

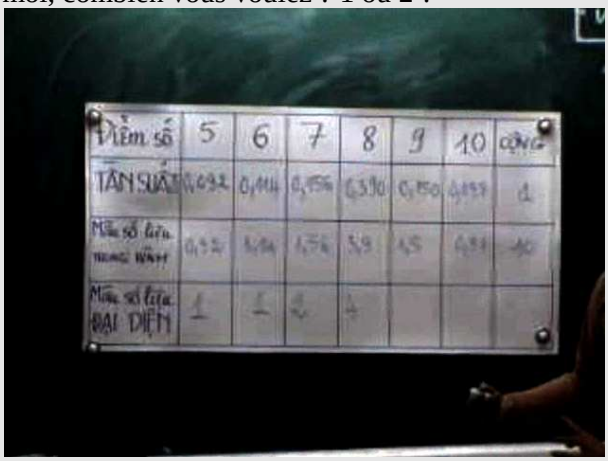
		 Pour l’échantillon le moins représentatif, vous avez choisi l’échantillon 2 très nombreux, 22 réponses. 
[…]		
	Ens	(Après avoir ramassé les fiches 6, l’enseignant fait un résume publique) Je vois l’échantillon représentatif du groupe 1 est comme suit : 5, 6, 7, 8, 8, 9, 9, 9, 10, 10. Quels sont les groupes qui proposent le même échantillon comme celui-ci ? (*Tous les groupes montrent leurs doigts pour dire qu’ils ont le même résultat*) Bon, quel groupe peut aller au tableau pour présenter comment vous trouvez cet échantillon ? (*Un élève du groupe 3 va au tableau*)
	El	D’abord, j’ai calculé les fréquences des notes et j’ai obtenu les résultats : la note 5 occupe 99%...

	Ens	99 ???
	El	Ah non, 9%, (*Un élève se trouvant près de l'appareil de vidéo dit : « 0,09 »*) La note 6, 12% ; la note 7, 13%, la note 8 occupe 16%, la note 9 occupe 32%, et la note 1à occupe 18%. Puis notre groupe a arrondi ces pourcentages.
	Ens	Ah, ton groupe a arrondi ces nombres, mais comment fais-tu-le ? Je vois ici 9%, je ne comprends pas bien comment les arrondir ? Peux-tu l'expliquer clairement ?
	El	Je fais comme ça : pour les nombres dont l'unité est de 5 à 9, je les augmente à 10, par exemple, le 9%, je l'arrondis à 10% ; le 12%, je l'arrondis à 10% aussi,…
	Ens	D'accord, tu les écris au-dessus des pourcentages, et avec % car tu dis pourcentage. (*cet élève continue à écrire : 10%, 10%, 10%, 20%, 30%, 20%*) La classe, vous comprenez ce que votre fait pour arrondir ces pourcentages ?
	Els	Comme je vois que les trois notes 5, 6, 7 occupent 10% des notes, donc les nombre des ces notes doivent le même, c'est-à-dire chacun de ces trois notes doit prendre une place.
	Ens	Pourquoi une place pour chacun de ces trois note ?
	Els	Car on a 10 cases pour 10 notes, cette note occupe 10% c'est-à-dire il occupe une place et chacun de ces trois notes prend une place.
	Ens	Ah, tu veux dire que 10% est égale à 1/10.
	Els	Les notes 8 et 10 occupent 20%, ils prennent donc deux places et enfin la note 9 occupant 30% doit prendre 3 places. Ça fait en total 10 places.
	Ens	Bon, merci. Et les autres groupes ? vous faites pareillement avec les pourcentages ?
	Els	Oui …
	Ens	Est-ce qu'il y a une autre méthode pour construire un échantillon qui est mieux que celui-ci ? Non ? Vous voyez, cette méthode nous donne un échantillon qui est <u>très proportionnel</u>[219] à l'ensemble de toutes les notes. Bon, je vous remercie de m'apprendre cette très bonne méthode pour construire un échantillon représentatif par la proportionnalité des effectifs entre cet échantillon et l'ensemble de 100 notes.
Phase 2		
	Ens	Dans l'exercice fait à la maison on a 100 notes de 100 élèves d'un lycée, un seul lycée, et ce lycée doit prendre un échantillon de 10 notes pour comparer avec ceux des autres lycées. Maintenant, la nouvelle situation est plus large : après un examen commun d'une séance d'une durée de 45 minutes concernant tous les élèves de notre ville de HCMV, toutes les notes sont envoyées au directeur du service d'éducation et de formation de Ho Chi Minh ville. Cet ensemble des notes est très grand. Je ne sais combien il y a de notes mais je sais qu'il y en a un très grand nombre. […] Le directeur du service d'éducation et de formation de Ho Chi Minh ville a reçu les notes supérieures ou égales à 5 de tous les élèves de tous les lycées de la ville. Il veut choisir un échantillon de 10 notes pour communiquer avec les autres villes du Vietnam. Toutes les notes sont dans cet ordinateur. (*l'enseignant désigne du doigt l'ordinateur public*). Comme il y a trop des notes, il ne peut pas les compter. Alors, qu'est-ce que

[219] Souligné par nous.

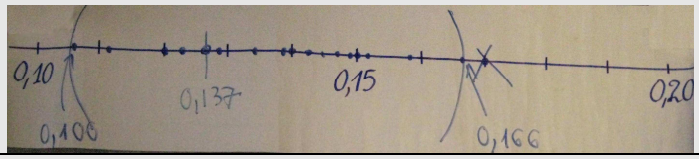
		vous pouvez lui conseiller de faire pour obtenir un échantillon le plus représentatif possible ?
[…]		
Gr.2	El	J'ai deux solutions : - première solution : le directeur doit écrire un programme qui permet de faire un grand nombre des tirages pour avoir des fréquences très exactes des notes 5, 6, … etc. A partir de ces fréquences, on peut prévoir combien il y a de notes 5, de notes 6, … etc. - deuxième solution : comme il est le directeur du service d'éducation et de formation, il peut ordonner aux autres personnes de compter les effectifs de toutes les notes pour lui. (*la classe rit*)
	Ens	Ah, tu veux un programme qui peut demander un très grand nombre des tirages. Mais je veux savoir combien, tu va prendre de notes par tirage ?
	El	… une note… c'est comme on prend une boule dans une urne
	Ens	A chaque fois que tu vas demander une note de l'ensemble qu'est-ce que tu vas observer ?
	El	Je vais voir la note, comme la couleur d'une boule…
	Ens	Tu veux faire combien de demande ?
	El	Plus le nombre des tirages est grand, plus ce sera mieux… environ 500 ou 1000 fois.
	Ens	Et qu'est-ce que tu feras avec ces notes ?
	El	Je vais calculer les fréquences de toutes ces notes…
[…]		
	Ens	Dans cet ordinateur, il y a les notes supérieures ou égales à 5 de tous les élèves. Un programme intégré dans l'ordinateur dispose d'une commande particulière. Chaque fois que l'on utilise cette commande, la machine nous donnera un sous-ensemble de 10 notes. Ces 10 notes sont prises de la façon suivante : on prend au hasard une note dans l'ensemble de toutes les notes, puis on la remet, c'est-à-dire prendre une note avec remise. Imaginez qu'il y a beaucoup, beaucoup,… de boules et que l'on prenne une boule au hasard puis qu'on la remette. On fait 10 fois comme ça et on obtient un sous-ensemble de 10 boules. Vous comprenez ? (*L'enseignant fait certains essais sur l'écran public*) Bon, maintenant, je vous propose de bien réfléchir pour me dire quelles fonctions vous voulez pour cet ordinateur ou quels calculs vous voulez lui demander de faire pour vous, pour que vous obteniez le meilleur échantillon, c'est-à-dire l'échantillon le plus représentatif ?
Gr.3	El	Le directeur doit entrer toutes les notes dans un tableur puis utiliser une fonction de tableur pour compter les effectifs des notes. Puis il peut calculer les fréquences des notes…
	Ens	Bon, l'idée du groupe 3 est d'entrer toutes les notes à l'ordinateur, c'est-à-dire il a un grand besoin de beaucoup de personnes. Mais, malheureusement, comme je vous ai dit, seul le directeur a le droit de voir toutes les notes et comme les notes sont trop nombreux, le directeur ne peut pas passer jour à jour pour entrer toutes les notes. En fait, votre solution est une méthode si on a le droit, le temps et des personnels. Pourtant, l'essentiel est qu'il faut avoir les fréquences des notes 5, 6, 7, 8, 9, 10, n'est-ce pas ? Bon, ce sont vos solutions.
[…]		

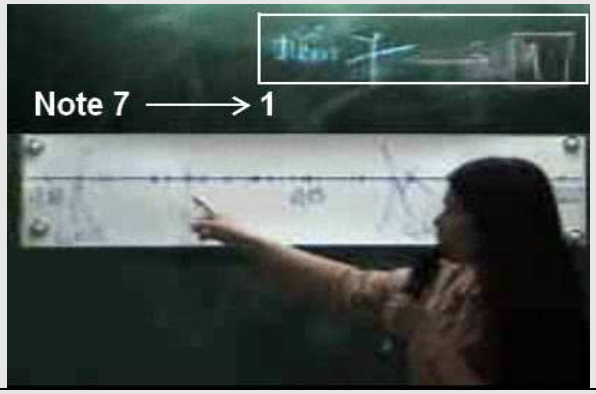
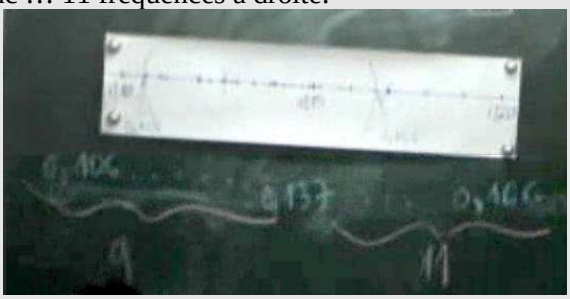
[…]		
58:13 Phase 2	Ens	Bon, regardez au tableau, SVP. Donnez- moi le résultat obtenu sur l’écran pour que je l’écrive sur cette fiche, SVP.  (*l’affiche commune est attachée sur le tableau*)
		0,092 ; 0,114 ; 0,156 ; 0,390 ; 0,150, 0,098
	Ens	Merci. Bon, j’ajoute un zéro pour que la partie décimale de tous ces nombres ait 3 chiffres. (*l’enseignant ajoute 0 pour avoir 0,390 et 0,150*) Nous avons maintenant cette ligne des fréquences de 10 notes. Si nous additionnons ces nombres, il fait surement …
	Els	1
	Ens	Oui, la somme est égale à 1. Je vais multiplier cette fréquence par 10. (*l’enseignant montre la fréquence 0,092*). Alors, combien j’obtiendrai ?
	Els	0,92
	Ens	Bon, 0,92 ; puis …
	Els	1,14 ; 1,56, 3,9 ; 1,5, 0,98
	Ens	Merci. Et si j’additions tous ces nombres, alors … ?
	Els	Cela fait 10…
	Ens	Oui, la somme est égale à 10. Si j’arrondis ce nombre ? (*l’enseignant montre 1,14*)

	Els	1 ; 1 ; 2 ; 4 ; 1 … mais 2 …[220] (Certains élèves disent 2, et les autres disent 1) 1 ou 2 ? Dites-moi, combien vous voulez ? 1 ou 2 ?  (*l’enseignant s’arrête dans le cas d’arrondi de la fréquence 1,5 de la note 9 pour une discussion publique*)
	Els	1 (*les élèves parlent très fort*)
	Autres els	2 (*ces élèves parlent plus faiblement*)
	Ens	Attendez. Pourquoi voulez-vous 1 ? Pourquoi l’arrondi de 1,5 vous donne 1 ? Selon le principe d’arrondi, il doit être combien ?
		Mais la somme fait 11.
		Ah, d’accord. Si je suis le principe d’arrondi d’un nombre réel au dixième, je dois augmenter ce nombre en 2 ici, n’est-ce pas ? Mais la somme fait 11. Alors, que dois-je faire ? Ce nombre 1,5 devient 2 après arrondi et 1,56 aussi, il est devenu 2 par l’arrondi. Bon, quel est le nombre[221] que l’on va garder ?
		1,56
	Ens	Ah, vous décidez de garder 1,56 et vous voulez que ce nombre soit réduit à 1 ? (*l’enseignant montre le nombre 2 et fait une flèche de déduction à 1*)

[220] Souligné par nous.

[221] Soit 1,56 soit 1,5.

		 Et grâce à ça, cette somme recule à …
	Els	10
	Ens	Oui, à 10. Je vois que pour ces deux notes 7 et 9, leurs fréquences 1,56 et 1,5 sont très proches. Comme on a 50 fois un échantillon de 10 notes, c'est-à-dire 500 notes. Ces 500 notes nous donnent deux pour la note 7 et un pour la note 9. On ne sait pas si en refaisant la commande n°1 avec 500 autres notes, on obtiendrait de nouveau « deux notes 7 - une note 9 » ou peut être « une note 7 - deux notes 9 » dans l’échantillon représentatif ? Il semble qu’il y a une hésitation entre le choix de « réduire le nombre de notes 9 » et celui de « réduire le nombre de notes 7 ». Par exemple, dans ce tableau, si on redonne 1 pour la note 7 et 2 pour la note 9, est-ce que l’on prendra un risque de se tromper ? C’est le problème posé.
	Ens	Maintenant, je vais placer 20 points pour 20 fréquences obtenues sur cette affiche. Pouvez-vous me lire les fréquences affichées à l’écran ? 0,106… 0,112… 0,170… 0,120… 0,122… 0,126… 0,128… 0,128… 0,134… 0,138… 0,140… 0,142… 0,144… 0,146… 0,148… 0,150… 0,150… 0,152… 0,158… 0,166. Bon, regardez ce graphique, d’après vous, on va prendre combien pour la note 7 ? 
	Els	U…n… (*La classe crie très fort*)
	Ens	Pour quoi vous choisissez 1 ?
	Els	Parce qu’il y beaucoup de points à gauche de 0,15.
	Ens	Et si je regarde la moyenne des fréquences, elle est combien ?
	Els	0,137
	Ens	Bon, je vais placer un point sur cet axe pour la moyenne. (*enseignant marque un point pour 0,137*) On peut constater que les « points-fréquences » se trouvent autour de cette

		moyenne et ils forment un intervalle…, la valeur la plus petite des fréquences, combien ? 0,106 … et la valeur la plus grande ? 0,166. Bon, on a l’intervalle]0,106 ; 0,166[. Et cet intervalle contient la valeur moyenne 0,137. La plupart des points se trouvent où ? 
	Els	A gauche…
	Ens	[…] Maintenant, si je veux enlever certaines fréquences, par exemple 5 fréquences, alors, lesquelles dois-je enlever ? Bon, si je veux comparer ces fréquences à enlever avec la fréquence moyenne 0,137, où se trouvent-elle ? on va compter le nombre des fréquences à gauche et à droite de la fréquence moyenne. Combien à gauche ? (*l’enseignant fait compter la classe grâce à l’affichage sur l’écran*)
	Els	… un , deux, trois, … neuf … Il y a neuf fréquences à gauche de 0,137.
	Ens	Bon, il reste donc … 11 fréquences à droite.  Alors, les fréquences que vous voulez enlever où se trouvent-elles ?
	Els	A droite de 0,137.
	Ens	Pourquoi ?
	Els	Car il y a plus de points à droite qu’à gauche…
	Ens	Bon, le nouvel intervalle [0,106 ; 0,148] est plus concentré sur 0,137. Comme on a enlevé 5 fréquences parmi les 20 fréquences, on peut dire que l’on a une certaine erreur. Si on enlève 5 valeurs parmi 100 valeurs, on dit que l’on a une confiance de 95% dans cet intervalle. […] Les cinq fréquences enlevées sont celles à partir de 0,150. Je dis que je prends un certain risque en décidant de choisir un pour la note 7.
		Vous voyez, dans ce travail, la commande n°2 exécute 20 fois la commande n°1 qui prend 500 notes, c'est-à-dire nous avons un ensemble de 20 fois 500 notes c’est-à-dire 10 000 notes. Et nous avons choisi une seule note 7 au lieu de deux notes 7 comme dans le travail tout à l’heure. Est-ce que plus on prend un grand nombre des notes, plus on a une fréquence plus exacte ?

Tóm tắt

« Một nghiên cứu didactic về việc đưa các khái niệm thống kê trong mối liên hệ với xác suất vào dạy học toán ở Việt nam »

Nghiên cứu didactic về dạy học thống kê trong dạy học toán ở Việt nam trong các mối liên hệ với các khái niệm của xác suất. Nghiên cứu này xuất phát từ hai ghi nhận :

- ghi nhận về phương diện khoa học luận : các tri thức của thống kê suy diễn nêu lên đặc tính toán học của mối liên hệ giữa thống kê và xác suất,

- ghi nhận về phương diện didactic : hai hệ thống dạy học Pháp và Việt nam khác nhau rõ nét. Thống kê và xác suất được dạy tách biệt nhau và các tri thức của thống kê suy diễn hoàn toàn vắng mặt trong chương trình toán của Việt nam. Hoàn toàn khác với quan điểm này, chương trình toán ở Pháp gần đây đã đưa vào một số tri thức của thống kê suy diễn như sự biến động của việc chọn mẫu hay sự tương thích của một phân bố thống kê với một luật xác suất.

Sự tách rời này đặt ra vấn đề về phương diện khoa học luận của tri thức được dạy. Sự nối khớp của một phân tích khoa học luận với một phân tích thể chế đã dẫn đến việc nghiên cứu hai câu hỏi có tính kế thừa nhau : Sự gián đoạn mà ta nhận thấy được giữa thống kê và xác suất đã ảnh hưởng thế nào đến các quan hệ thể chế của giáo viên cũng như của học sinh Việt nam đối với « các đối tượng-ứng viên của các mối liên hệ giữa thống kê và xác suất » ? Có thể nào làm cho một số trong các đối tượng-ứng viên này « sống được » trong các điều kiện dạy học hiện nay tại Việt nam không ?

Mỗi câu hỏi trong hai câu hỏi trên được nghiên cứu trong một thực nghiệm : thực nghiệm thứ nhất là bộ câu hỏi điều tra học sinh và giáo viên; thực nghiệm thứ hai là một đồ án didactic được thực hiện ở một lớp ở trung học phổ thông tại Việt nam.

Từ khoá

Dạy học toán, thống kê mô tả và thống kê suy diễn, phân tích và đồ án sư phạm.

Résumé

L'étude didactique porte sur l'enseignement secondaire vietnamien des statistiques dans ses liens potentiels avec les concepts de probabilité. Elle part de deux constats, le premier de nature épistémologique et le second de nature didactique :

- il existe un corpus des savoirs en statistique inférentielle qui caractérise mathématiquement les liens entre statistiques et probabilités

- les deux systèmes d'enseignement, vietnamien et français diffèrent profondément puisque statistiques et probabilités sont enseignées séparément au Vietnam et ces savoirs sont absents des programmes alors qu'en France plusieurs réformes récentes ont inscrit dans les programmes certains de ces savoirs au travers de notions comme la fluctuation d'échantillonnage ou l'adéquation d'une distribution statistique à une loi probabiliste.

L'articulation d'une analyse épistémologique et d'une analyse institutionnelle développe la problématique de l'étude dans deux questions successives : quel effet a la coupure constatée entre statistiques et probabilités sur le rapport institutionnel à des « objets-candidats pour les liens statistiques-probabilités » au Vietnam aussi bien du côté enseignant que du côté élève ? Est-il possible de faire vivre dans les conditions actuelles de l'enseignement vietnamien quelques uns de ces objets-candidats ? Chacune des questions est travaillée via un dispositif expérimental : la première à travers un questionnaire aux élèves et aux enseignants vietnamiens et la seconde dans une ingénierie didactique réalisée dans une classe de lycée vietnamien.

Mots-clés

Enseignement mathématique, statistique descriptive et inférentielle, analyses et ingénierie didactique.