



N° d'ordre NNT : 2019LYSE2064

THESE de DOCTORAT DE L'UNIVERSITÉ DE LYON  
en COTUTELLE avec LE CENTRE UNIVERSITAIRE FRANCISCAIN

Opérée au sein de

L'UNIVERSITÉ LUMIÈRE LYON 2

**École Doctorale : ED 485**

**Education Psychologie Information Communication**

Discipline : Sciences de l'Éducation

Soutenue publiquement le 23 Août 2019, par :

**Leonardo DALLA PORTA**

---

**Formação do raciocínio estatístico na  
conceptualização da estimação estatística.**

*estudo exploratório de um dispositivo pedagógico no ensino  
superior.*

---

Devant le jury composé de :

Nadja ACIOLY-REGNIER, Professeure des universités, Université Claude Bernard Lyon 1, Présidente

Saddo AG ALMOULOU, Professeur d'université, Universidade São Paulo, Examineur

Eleni BISOGNIN, Professeure, Centre universitaire Franciscain, Examinatrice

Luciane FLORES JACOBI, Professeure, Universidade Santa Maria, Examinatrice

Vladimir LIRA VERAS XAVIER DE ANDRADE, Professeur d'université, Université Federale rural do  
Pernambuco, Examineur

Adriano MENDOÇA SOUZA, Professeur Universidade Santa Maria, Examineur

André D. ROBERT, Professeur des universités, Université Lumière Lyon 2, Co-Directeur de thèse

Silvia Maria DE AGUIAR ISAIA, Professeure, Centre universitaire Franciscain, Co-Directrice de thèse

## Contrat de diffusion

Ce document est diffusé sous le contrat *Creative Commons* « [Paternité – pas d'utilisation commerciale – pas de modification](#) » : vous êtes libre de le reproduire, de le distribuer et de le communiquer au public à condition d'en mentionner le nom de l'auteur et de ne pas le modifier, le transformer, l'adapter ni l'utiliser à des fins commerciales.

**École Doctorale: 485 EPIC**  
Laboratoire de recherche UMR 5191 ICAR  
**Programa de Pós-graduação em Ensino de Ciências e Matemática –**  
PPGECIMAT

## **La formation du raisonnement statistique à travers la conceptualisation de l'estimation statistique.**

**Étude exploratoire à partir d'un dispositif pédagogique dans  
l'enseignement supérieur**

**Résumé de la Thèse de Doctorat en Sciences de l'Éducation  
en régime de cotutelle**

Présentée par Leonardo DALLA PORTA

Sous la Direction de:

Jean-Claude RÉGNIER et Silvia Maria DE AGUIAR ISAIA

Soutenue publiquement le 23 août de 2019

devant le jury composé de :

Jean-Claude RÉGNIER, Directeur de thèse, Université Lumière Lyon 2  
Silvia Maria DE AGUIAR ISAIA, Directrice de thèse, Universidade Franciscana  
Nadja Maria ACIOLY-RÉGNIER, Professeure des Universités, Université Lumière  
Lyon 1  
Eleni BISOGNIN, Professeure des Universités, Universidade Franciscana  
Saddo AG ALMOULOUD, Professeur des Universités, Pontifícia Universidade Católica  
de São Paulo  
Vladimir LIRA VERAS XAVIER DE ANDRADE, Professeur des Universités,  
Universidade Federal Rural de Pernambuco  
Adriano MENDONÇA SOUZA, Professeur des Universités, Universidade Federal de  
Santa Maria  
Luciane FLORES JACOBI, Professeure des Universités, Universidade Federal de Santa  
Maria

## Introduction

Concepts et procédures statistiques sont utilisés pour collaborer aux traitements et aux interprétations d'innombrables données qui sont diffusées au travers de moyens de communication les plus divers. Au cours des dernières années, comme le fait remarquer Ignácio (2010), la statistique poursuit sa dissémination dans les universités ainsi que dans les diverses branches de la société moderne. Carzola (2002) affirme que les concepts et techniques statistiques, auparavant à usage limité à quelques domaines, n'ont jamais été tant présents dans la vie quotidienne du citoyen. Face à cela, une discussion importante s'ouvre sur le développement du raisonnement statistique qui est une des compétences statistiques fondamentales pour affronter des situations qui requiert la statistique.

Observer le développement d'une compétence n'est pas une tâche facile, toutefois, dans notre recherche, nous visons à analyser la formation même du raisonnement statistique dans un contexte d'enseignement. Dans cette perspective, nous avons conduit notre recherche auprès d'un échantillon de 81 étudiants inscrits dans des cours de statistique dispensés au niveau de l'Enseignement supérieur d'une université communautaire de la ville de Santa Maria, dans l'état du Rio Grande do Sul. Ces étudiants sont originaires de divers cours de graduation, où la statistique est identifiée, selon Wada (1996), comme une discipline de service qui objective l'instrumentalisation de futurs professionnels pour un usage adéquat des outils statistiques.

Pour notre recherche sur la formation du raisonnement statistique, nous avons choisi le domaine de la statistique inférentielle, plus spécifiquement, à ce qui touche à l'estimation statistique, et la mise en œuvre pour estimer une moyenne ou une proportion. Ce choix est dû au fait que tous les étudiants de l'échantillon avaient suivi, antérieurement, une formation en statistique descriptive, c'est à dire, qui possédaient des connaissances sur des concepts et procédures de base en statistique. Nous avons utilisé la modalité de construction des données par questionnaire appliqué avant et après les cours portant sur l'estimation statistique. Ces questionnaires ont été construits autour de situations-problèmes qui renvoient aux contenus de l'estimation et des questions sur les relations des étudiants à la statistique. Gal & Ginsburg (1994) soutiennent que les enseignants de statistique devraient se préoccuper davantage des aspects affectifs des processus d'enseignement et d'apprentissage, en cherchant à identifier des attitudes et des frustrations des étudiants et en proposant des stratégies qui visent à réduire les aspects

négatifs. En ce sens, nous avons tenté d'observer au travers des réponses fournies dans les questionnaires dans quelle mesure les relations des étudiants à la statistique influencent la formation du raisonnement statistique.

Les informations générées par les questionnaires ont été traitées et exploitées dans le cadre théorique de l'Analyse Statistique Implicative (ASI), en utilisant le logiciel CHIC.7<sup>1</sup>. Dans le cadre théorique de la statistique classique, pour le traitement descriptif et inférentiel des données, nous avons utilisé SPAD.9.1<sup>2</sup>. Au-delà, nous avons eu recours à la théorie des Champs Conceptuels pour analyser le champ conceptuel de l'estimation statistique comme aide à l'interprétation des résultats.

Par ailleurs notons que les résultats de la recherche deviennent une production pour l'enseignement spécifiquement tournée vers les expériences en salle de cours, visant à contribuer à l'amélioration de celui-ci.

## 1.1 Objectifs de la recherche

Considérant ce qui a été écrit, nous présentons l'objectif général en lien avec les objectifs spécifiques de la recherche, au-delà des hypothèses soulevées pour la construction des résultats.

### 1.2.1 Objectif général

Mieux comprendre le processus de formation du raisonnement statistique dans l'apprentissage de l'estimation statistique dans l'Enseignement supérieur.

### 1.2.2 Objectifs spécifiques.

- I) Construire des données par une enquête par questionnaire intégrant des situations problèmes qui permettent de procéder à des observations sur la formation du raisonnement statistique;
- II) Mieux comprendre comment procède la formation du raisonnement statistique à la lumière de la Théorie des champs conceptuels ;
- III) Traiter les données construites avec des outils statistiques dans le cadre de l'Analyse Statistique Implicative – ASI avec le logiciel CHIC et dans le cadre de la statistique descriptive et inférentielle avec le logiciel SPAD

#### 1.2.1. Hypothèses

H1 – Il existe des limitations dans le processus de formation du raisonnement statistique, dans l'estimation statistique, liées aux procédures et aux concepts de base de la Statistique descriptive ;

---

<sup>1</sup> Classification Hierarchique Implicative Cohésitive <https://ardm.eu/partenaires/logiciel-danalyse-de-donnees-c-h-i-c/>

<sup>2</sup> SPAD = Système Portable d'Analyse des Données. COHERIS – [www.coheris.com](http://www.coheris.com) - 4, Rue du Port aux Vins – 92150 Suresnes – France

H2 – La relation affective de l'étudiant à la Statistique a une influence sur la formation du raisonnement statistique ;

H3 – La relation affective de l'étudiant aux Mathématiques a une influence sur la formation du raisonnement statistique ;

## **PARTIE 1 – Fondements théoriques et éléments de la recherche**

Les fondements théoriques et les éléments de la recherche apportent des discussions pertinentes pour les bases de notre travail. Nous avons traité la Statistique en abordant une vision éducationnelle de cette discipline dans l'Enseignement Supérieur, en nous centrant sur les compétences statistiques, la littéracie, la pensée et le raisonnement statistiques. Au-delà de cela, dans le cadre de la Théorie des champs conceptuels nous avons considéré l'estimation statistique comme un champ conceptuel.

### **1. Éducation statistique dans l'Enseignement supérieur : la statistique comme une discipline de service**

Nous commençons nos réflexions sur l'Éducation statistique en observant maintenant quelques points significatifs dans les processus d'enseignement et d'apprentissage d'une discipline, celle de la Statistique, dans l'Enseignement supérieur. Avant de commencer, nous allons situer la formation statistique qui a été l'objet de nos observations dans nos analyses, ceci en considérant deux types de public dans l'Enseignement supérieur. Le premier type est constitué par les étudiants qui viennent se former en Statistique pour devenir statisticien et qui possèdent des connaissances spécifiques couvrant de larges aspects de la discipline. Toutefois ce n'est pas sur cette catégorie qu'a porté notre investigation. Pour nous, il s'agit du second type constitué par les étudiants issus de divers cursus de *graduação*, c'est-à-dire les quatre premières années de l'enseignement supérieur universitaire au Brésil, dans lesquels une ou plusieurs disciplines du domaine de la Statistique sont obligatoires. Wada (1996) désignent ces disciplines comme disciplines de service pour mettre en évidence l'instrumentation de futurs professionnels qui feront un usage adéquat des outils statistiques. Cette formation vise à atteindre ce que Carzola (2002) appelle usagers de la statistique et qui sont des consommateurs de d'informations statistiques, mais nous considérons aussi qu'ils pourront produire en raison de la formation acquise. L'Éducation statistique nécessite de porter un regard plus attentif pour ce public, ayant en vue que le processus d'enseignement n'est pas toujours satisfaisant ainsi que le pointent Barceló, Capilla et Zúnica (2000), Carzola (2002), Batanero (2001), Grácio et Oliveira (2005), Mantovani et Viana (2004), entre autres.

Le faible profit tiré de ces disciplines de service soulève des discussions à propos de l'Éducation statistique, en pointant les Mathématiques comme l'un des principaux facteurs qui contribuent à générer des attitudes négatives des étudiants à l'égard de la

statistique. Dans l'Éducation de base (élèves âgés de 6 ans à 14ans correspond à l'École élémentaire et au Collège en France), la Statistique fait partie du programme d'enseignement des mathématiques et, donc, assuré par des enseignants *licenciés* en mathématiques qui, selon Carzola (2002) ne sont pas toujours suffisamment compétents pour enseigner la Statistique, ce qui affecte l'apprentissage tout au long du cursus. Barbetta (2006) déclare que nombre d'étudiants arrivent dans l'Enseignement supérieur avec peu ou pas de connaissances en statistique. Cette représentation de la statistique insérée dans les mathématiques à l'École élémentaire et au Collège pousse les étudiants qui arrivent à l'université à comprendre la statistique comme une branche des mathématiques. Cette représentation est la base du rejet pour beaucoup de ceux qui n'ont pas eu de bonnes performances en mathématiques durant leur scolarité. La déconstruction de ces représentations n'est pas une tâche facile, en amenant un débat sur comment les mathématiques doivent être employées dans l'Enseignement supérieur dans le cadre de la statistique. De façon générale, les mathématiques exigent des techniques de résolution qui conduisent à une solution unique pour un problème déterminé, en éloignant l'étudiant, de la nature de la statistique qui exige une posture plus critique face à la confrontation à l'incertitude et aux problèmes avec des possibilités variées pour des prises de décision. Silva (2007) fait ressortir que le renforcement des notions fondamentales de mathématiques est important quand l'enseignement vise à développer les capacités statistiques, mais discutable pour des étudiants qui seront des usagers de la statistique. Hand (1998) reconnaît l'importance des mathématiques pour l'enseignant de la statistique, en soulignant que la propension à commettre des erreurs dans la résolution des problèmes est d'autant moindre que les étudiants ont d'autant plus de connaissances mathématiques. Toutefois, l'auteur défend l'idée que l'usage de logiciels de statistique pour le développement des calculs privilégie le développement des habiletés en analyse statistique, en diminuant les efforts liés au fondement mathématique de l'analyse. L'éducation statistique doit donner priorité au développement d'une formation statistique qui stimule l'esprit critique capable de reconnaître l'usage des formules de mathématiques insérées dans les concepts de statistique afin de promouvoir des réflexions en face des résultats trouvés.

Les fondements théoriques de la Statistique sont à chercher dans les Mathématiques, delà, par le fait que de nombreux étudiants entrant dans l'Enseignement supérieur rencontrent des difficultés même avec des opérations mathématiques de base, ceux-ci ont tendance à transférer le même sentiment à l'égard de la Statistique. La nature

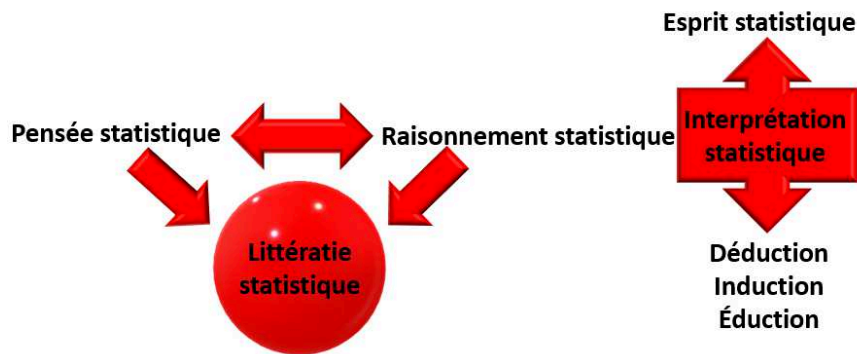


de la Statistique, comme nous l'avons déjà mentionné, requiert un travail avec des expériences aléatoires et des concepts de nature probabiliste, soumettant l'étudiant à la réalisation de procédures pour des prises de décision en situations incertaines. Un autre point saillant cité par l'auteur est la charge horaire trop réduite pour de la Statistique. Ainsi malgré son importance reconnue dans la plupart des formations professionnelles, la Statistique, pour être une discipline de service, est insérée comme coadjuvant dans divers domaines de connaissance sur la base d'une charge horaire réduite dans les cours de *graduation* universitaire. Ceci étant, les enseignants finissent par remédier aux insuffisances issues de la formation scolaire en privilégiant les procédures mathématiques et laissant de côté le développement du raisonnement statistique.

Face aux questions soulevées, nous entendons qu'il est nécessaire de réfléchir sur la principale fonction de la Statistique dans l'Enseignement supérieur en observant les compétences nécessaires pour une formation en Statistique qui réponde aux demandes de la société. Rumsey (2002), Gal et Garfield (1997), Campos (2007), delMas (2002) et autres, défendent, dans leurs recherches, l'importance du développement de trois compétences : littéracie, pensée, raisonnement statistique

## **2. Réflexions sur littéracie, pensée et raisonnement statistique**

delMas (2002) perçoit la littéracie statistique comme une compétence qui englobe la finalité générale de l'enseignement de la statistique, tenant compte des aspects de la pensée et du raisonnement statistique. Campos (2007) déclare comprendre qu'il n'est pas productif de penser une formation statistique en considérant les compétences statistiques de façon indépendantes, car elles sont complémentaires et c'est seulement ensemble qu'elles permettent de couvrir la compréhension globale de la Statistique. Pour Rumsey (2002) le développement de la littéracie statistique rend nécessaire de toujours penser en termes de capacités corrélées qui vont s'ajouter pour promouvoir une meilleure compréhension des concepts statistiques. Régnier (2003) insiste sur un raisonnement statistique intégrateur avec une centration sur les interprétations statistiques, au-delà de la formation de ce que l'auteur appelle esprit statistique à la manière de Bachelard qui parle d'esprit scientifique. Face à cela, nous avons élaboré un schéma pour visualiser et définir le rôle de chaque compétence comme le montre la Figure 1 ci-dessous :



Source: L'auteur

**Figure 1: Compétences statistiques**

Nous nous accordons avec Chance (2002) qui conçoit la littéracie statistique comme la compétence qui exerce la compréhension et l'interprétation statistique ainsi qu'avec Silva (2007) qui perçoit cette compétence comme dépendante du raisonnement et de la pensée statistique. Nous avons construit le Tableau 1 ci-dessous pour mieux visualiser les actions et les demandes pour la littéracie, la pensée et le raisonnement statistique.

	<b>Littératie</b>	<b>Pensée</b>	<b>Raisonnement</b>
<b>Actions connexes</b>	Lire Écrire Décrire Interpréter	Généraliser Observer le processus	Expliquer interpréter les résultats
<b>Demande</b>	Connaissance des concepts et des méthodes statistiques	Observer le processus statistique, reconnaissance de la variation.	Induction Dédution Éducation Esprit statistique

Source: L'auteur

**Tableau 1: actions et demandes de compétences statistiques**

La littéracie statistique est la compétence des individus à lire, écrire, décrire et interpréter des informations statistiques, dont le niveau de perception sera plus affiné en concordance avec l'élévation des niveaux de pensée et de raisonnement. Nous partageons le point de vue de Silva (2007) qui observe l'avancement bilatéral entre les niveaux de pensée et de raisonnement conduisant à la compréhension de ce que ces compétences sont mutuellement connectées. À la pensée statistique, il revient le soin de la planification, le contrôle et la généralisation tout au long de l'application d'une procédure statistique. Le raisonnement statistique, au-delà de l'habileté à travailler avec des outils et avec des concepts de statistique assume l'interprétation des résultats obtenus par le moyen des

raisonnements déductifs, inductifs et éductifs<sup>3</sup> sur lesquels se fonde un esprit critique et scientifiques que nous appelons aussi esprit statistique.

### **3. Estimation statistique**

La Statistique moderne peut être comprise au travers de deux grands domaines : la statistique descriptive et la statistique inférentielle. Pour Freund (2006), la statistique descriptive propose des outils qui permettent, de forme systématique, le traitement dans le recueil, l'organisation, la classification, la représentation et la description des données dans quelque domaine de connaissance que ce soit. Bien que la statistique descriptive soit une branche extrêmement importante de la statistique, les données relevées dans une recherche sont presque toujours extraites d'échantillons, ce qui requiert une exigence particulière dans les analyses au moyen de généralisations qui dépasse le niveau des descriptions des données. Dans ce sens, nous trouvons dans la statistique inférentielle, des méthodes qui encadrent ces procédures de généralisation. Traditionnellement, les problèmes d'inférence statistique sont alors traités soit dans le cadre des tests d'hypothèses statistiques, soit dans celui de l'estimation statistique.

Notre recherche porte sur la notion d'estimation statistique qui concerne la question d'évaluation de paramètres à partir d'opérations réalisées sur les données échantillonales. Il s'agit d'un raisonnement typiquement inductif par lequel les résultats obtenus sur l'échantillon sont généralisés sur la population. De cette manière, nous pouvons prendre l'estimation comme un outil statistique qui consiste à déterminer une valeur échantillonnale en lieu et place de la valeur respective inconnue sur la population. Précisons que nous nommons aussi statistique, la variable qui associe à un ensemble de données échantillonales une valeur d'un paramètre ou d'une caractéristique, par exemple la variable moyenne échantillonnale dont la réalisation est la moyenne des données sur l'échantillon. Quand cet échantillon est aléatoire, une statistique est une variable aléatoire. Dans le contexte de l'estimation statistique, une telle statistique est appelée un estimateur et une réalisation, une estimation.

Il existe deux types d'estimation : l'estimation ponctuelle et l'estimation par intervalle de confiance. Dans le cas de l'estimation ponctuelle, il s'agit de fournir une valeur

---

<sup>3</sup> La notion d'éduction a été utilisée par Spearman (1927), avec comme signifié l'idée de faire émerger des informations face à un ensemble de données sans structure apparemment saillante, c'est-à-dire sollicitant la capacité de l'individu à détecter des relations entre deux ou plusieurs idées perçues dans un problème ou évoquées de mémoire.

unique pour estimer une valeur inconnue sur la population. Un estimateur ponctuel d'un paramètre  $\theta$  sur la population est généralement noté par  $\hat{\theta}$ . Dans notre recherche, nous nous sommes limité à un estimateur  $\bar{X}$  d'une moyenne arithmétique  $\mu$  et d'une proportion  $\hat{p}$  sur une population  $p$ .

Nom de l'estimateur	Estimateur	Nom du paramètre	Paramètre
Moyenne	$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$	Moyenne	$\mu = \frac{\sum_{i=1}^N X_i}{N}$
Proportion	$\hat{p} = \frac{\sum_{i=1}^n X_i}{n}$	Proportion	$p = \frac{\sum_{i=1}^N X_i}{N}$

Source: L'auteur

**Tableau 2: exemple estimateurs**

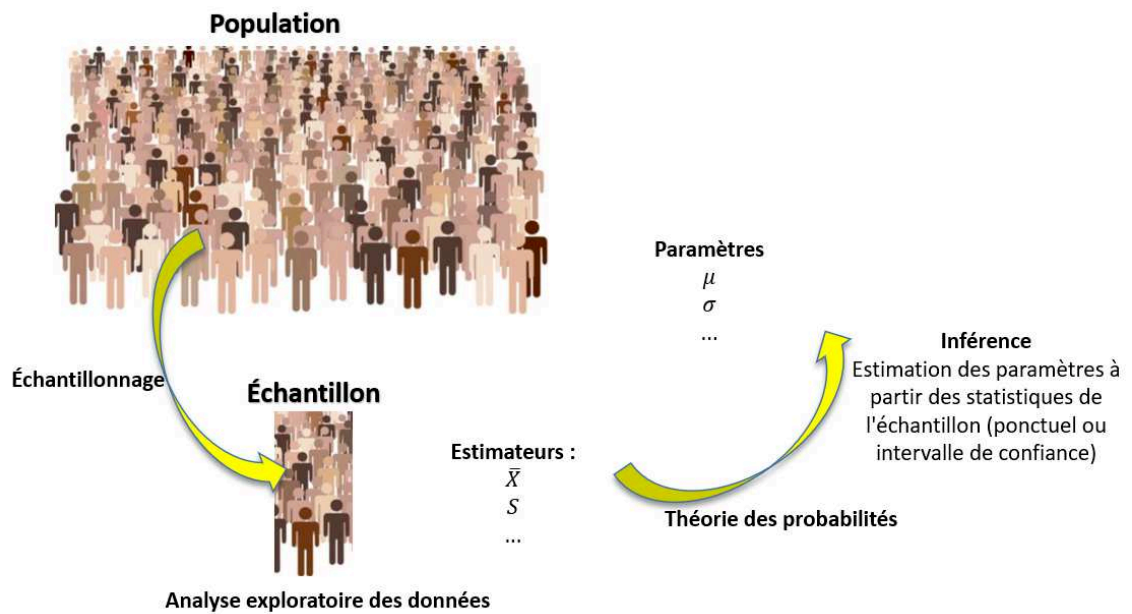
Quand nous obtenons une estimation ponctuelle, elle a peu de chance de coïncider avec la valeur inconnue du paramètre. Un des inconvénients de l'usage des estimations ponctuelles est que, si nous ne disposons d'aucune information additionnelle, nous n'avons aucun moyen d'évaluer la précision de l'estimation. De cette forme, l'estimation ponctuelle ne permet pas de lui attribuer en général un niveau de confiance. C'est alors tout l'intérêt de recourir à l'estimation par intervalle de confiance qui prend en compte la distribution de probabilité des estimateurs sur les échantillons. Le tableau ci-dessous donne les bornes des intervalles de confiance d'une moyenne.

	Moyenne	
	$\sigma$ connu	$\sigma$ inconnu
Populations infinies	$\bar{X} \mp z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$	$\bar{X} \mp z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$
Populations finies	$\bar{X} \mp t_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$	$\bar{X} \mp t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$

Source: L'auteur

**Tableau 3: estimation de la moyenne**

D'une manière concise, nous présentons quelques structures basiques sur l'estimation de paramètres pour lesquelles la théorie des probabilités et des concepts statistiques comme celui de population, échantillon, moyenne, écart type et autres, sont parmi ceux des fondements pour la prise de décision.



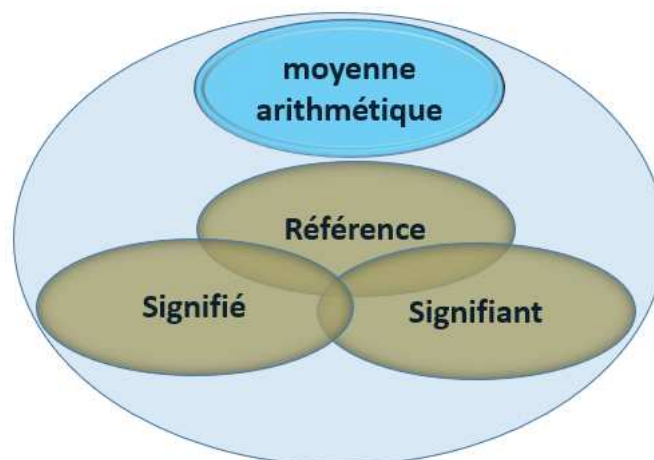
Source: L'auteur

**Figure 2: schéma d'estimation**

Dans ce cadre, nous pouvons aussi définir l'estimation d'un paramètre comme une méthode de décision qui permet d'estimer des caractéristiques de la population à partir de données échantillonnales extraites d'une population déterminée, en considérant divers facteurs qui caractérisent les cas d'observation.

#### 4. L'estimation statistique comme champ conceptuel

Nous reprenons le point de vue de Oriol (2007) et Diaz (2016) qui ont adopté le cadre de la théorie des champs conceptuels dans leurs travaux qui relèvent de la didactique de la statistique. Ces auteurs recourent à la notion de moyenne arithmétique d'une série de données pour tenter d'observer la formation du concept.



Source: L'auteur

**Figure 3 :le concept de moyenne**

La moyenne est incontestablement un concept qui renvoie à un ensemble de situations (réfèrent), à une ensemble d'invariants sur lesquels se fonde son opérationnalité (signifié) et recourt à un ensemble de formes langagières qui permettent de représenter symboliquement ses propriétés, situations et procédures de traitement (signifiant)

Par cette entrée, nous considérons que la théorie des champs conceptuels de l'estimation statistique est à même de nous fournir un cadre théorique pour mieux comprendre la conceptualisation de l'estimation statistique.

Concept: estimation ponctuelle de la moyenne		
Des situations (Réfèrent)	Invariants opératoires (Signifié)	Représentations symboliques (Signifiant)
Estimation ponctuelle de la moyenne de la population $\mu$ à partir de de l'échantillon n.	Mesures de tendance centrale	$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n}$
Decouvrir à travers d'un échantillon le temps moyen qu'un enfant prend pour assembler un jouet.	Somme de tous les éléments d'un échantillon divisée par le nombre total d'éléments.	$\frac{1}{n} \sum_{i=1}^n x_i$
	Estimateur $\bar{X}$ du paramètre $\mu$ .	

Source: L'auteur

**Tableau 4: estimation ponctuelle de la moyenne**

Concept: estimation par intervalle de confiance de la moyenne		
Des situations (Réfèrent)	Invariants opératoires (Signifié)	Représentations symboliques (Signifiant)
Estimation par intervalle de confiance de la moyenne de la population $\mu$ à partir des données de l'échantillon n.	Mesures de tendance centrale.	$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n}$
Decouvrir à travers d'un échantillon le temps moyen qu'un enfant prend pour assembler un jouet.	Mesures de dispersion.	$\frac{1}{n} \sum_{i=1}^n x_i$
	Estimateur $\bar{X}$ du paramètre $\mu$ .	$\left[ \bar{X} \mp z_{\alpha} \frac{\sigma}{\sqrt{n}} \right]$
	Intervalle de confiance.	
	Distributions de probabilités.	
	Erreur d'échantillonnage.	

Source: L'auteur

**Tableau 5: estimation par intervalle de confiance de la moyenne**

Rappelons que les cadres présentés visent à décrire des concepts scientifiques mis en forme pour la médiation didactique en vue de l'appropriation par une partie des étudiants, de tels concepts ne peuvent être considérés comme des concepts-en-acte (articulés par des théorèmes-en-actes). En effet ce ne sont pas des catégories de l'entendement d'un sujet en action face à une situation donnée. Notre intention a été de contextualiser la théorie des champs conceptuels en utilisant quelques concepts possibles que nous avons jugé nécessaires à un sujet donné pour rendre compte des situations.

Nous avons considéré alors qu'un concept, comme ici celui de l'estimation statistique, n'est pas une structure isolée mais est un élément d'un champ conceptuel qui met en jeu d'autres concepts comme celui d'échantillon, de population, de variable statistique, de moyenne, d'écart-type, pour ne citer que les plus communs. Vergnaud (1990a) affirme que, pour l'étude d'un concept, il est nécessaire d'être en présence d'un champ de problèmes distincts mais qui possèdent une base identique. Notre recherche a porté sur des étudiants en situations de résolutions de problèmes qui convoquent les mêmes concepts. Ainsi nous assumons l'idée que les concepts restent en tension entre eux et forment un champ conceptuel.

## **PARTIE 2 – MÉTHODOLOGIE DE CONSTRUCTION ET TRAITEMENT DES DONNÉES**

Nous avons opté pour une recherche qui prend appui sur une approche méthodologique mixte pour explorer des données construites selon des méthodes dites qualitatives et quantitatives, que certains auteurs nomment quali-quantitatif<sup>4</sup>. En accord avec Gonsalves (2003), durant de nombreuses années les recherches dites qualitatives ou quantitatives furent traitées comme deux paradigmes. Celles dites quantitatives transformaient les phénomènes explorés en nombres, quantités, et faisaient usage d'outils statistiques pour mettre à plat leurs causes et tester les hypothèses associées. D'un autre côté, les recherches dites qualitatives se préoccupaient davantage de la compréhension et de l'interprétation des phénomènes. Nous avons compris la nécessité de dépasser cette dichotomie en articulant des techniques dites quantitatives et celles dites qualitatives dans notre propre recherche accompagnant ainsi une tendance actuelle des recherches dans divers domaines. Régnier (2007) fait observer que nous ne devons pas couper le qualitatif du quantitatif tout au long d'une recherche. Pour lui, il y a une complémentarité fondamentale entre le qualitatif et le quantitatif et cela s'articule avec la nécessité de la variété et diversités des approches pour expliquer et comprendre un phénomène.

Pour construire les données valides, pertinentes et fiables pour notre recherche, nous avons eu recours à une enquête par questionnaires par lesquels nous avons apporté des situations en relation avec des connaissances sur l'estimation statistique et avec lesquels nous avons tenté d'identifier les rapports des étudiants avec la Statistique. Notre visée était d'explorer le processus de construction du raisonnement statistique au cours de l'apprentissage de l'estimation statistique en exploitant les réponses données dans les questionnaires par des étudiants inscrits dans les disciplines de Statistique d'un cursus universitaire. Le protocole de passation a consisté à réaliser l'enquête en deux étapes, l'une avant les cours sur l'estimation statistique et l'autre, après en soumettant le même questionnaire.

Ayant comme finalité l'obtention d'éléments pour pouvoir conduire nos inférences à partir des données construites par l'enquête par questionnaire puis traitées avec des logiciels de statistique, nous avons retenu, dans notre référentiel théorique, les

---

<sup>4</sup> Exemple d'approche quali-quantitative : Dominique Cardon (2015) *Méthodes qualitatives et données numériques : analyse topologique des territoires du web*, <https://iet.hypotheses.org/44>



composantes les plus importantes issues des réflexions portant sur l'éducation statistique comme la littéracie, la pensée et le raisonnement statistique. Au-delà de cela, nous avons cherché dans la théorie des champs conceptuels de Gérard Vergnaud, les arguments nécessaires pour explorer le champ conceptuel de l'estimation statistique et mieux comprendre le processus de conceptualisation.

En ce qui concerne notre recherche, la construction des données a été réalisée durant les deux semestres de l'année universitaire 2016 dans une université de la ville de Santa Maria dans l'état du Rio Grande do Sul au Brésil. Il s'agit de l'UFN – Université franciscaine. L'enquête a été menée auprès d'un échantillon non aléatoire constitué par des étudiants issus des cours de *graduation* de Sciences sociales et de Sciences technologiques. La taille de cet échantillon est de 81 individus. Ces étudiants devaient suivre les cours de Statistique au 3<sup>ème</sup> semestre du cursus de 8 semestres.

## **1. Outils informatiques**

Étant en possession des questionnaires complétés par les étudiants, nous avons rassemblé toutes les données dans un tableau du logiciel Excel afin de ne négliger aucune information recueillie. Ce traitement des données brutes a permis de pouvoir les consulter rapidement à quelque moment que ce soit afin de vérifier quelque information que nous souhaitions analyser. Ajoutons à cela que le recours à un tableur tel que le logiciel Excel permet d'avoir une base de données qui peut être exploitable avec d'autres logiciels de statistique comme SPAD 9.1 ou CHIC 7

### **1.1 Logiciel CHIC – Classification Hiérarchique Implicative Cohésitive de l'Analyse statistique implicative**

Pour analyser une partie des données construites par l'enquête par questionnaires, nous avons utilisé le cadre de l'Analyse Statistique Implicative – ASI instrumenté par le logiciel CHIC 7.0 - Classification Hiérarchique Implicative Cohésitive basé sur des algorithmes construits à l'origine par Régis Gras dans les années 80 et perfectionné au fil des années par Saddo Ag Almouloud, Harrison Ratsimba-Rajon et, actuellement sous la responsabilité de Raphaël Couturier.

En accord avec Régnier et al. (2013), Gras et al. (2009, 2013, 2017) le cadre théorique de l'ASI – Analyse statistique implicative a comme objectif principal la structuration des données dans les croisements entre un ensemble d'individus et un ensemble de variables. Pour reprendre les propos de Régnier (2019) sur le site du X<sup>ème</sup>

colloque sur l'Analyse statistique implicative<sup>5</sup> « À ce jour, l'analyse statistique implicative désigne un champ théorique centré sur le concept d'implication statistique ou plus précisément sur le concept de quasi-implication pour le distinguer de celui d'implication logique des domaines de la logique et des mathématiques. L'étude de ce concept de quasi-implication en tant qu'objet mathématique, dans les champs des probabilités et de la statistique, a permis de construire des outils théoriques qui instrumentent une méthode d'analyse de données. Force est de constater que les racines épistémologiques de ce concept se sont nourries de questions qui ont surgi principalement d'un autre champ : celui de la didactique des mathématiques. Historiquement, une des questions abordées concernait la mise en évidence des niveaux de complexité des exercices de mathématiques posés à de jeunes élèves s'énonçant ainsi :

R " si un exercice est plus complexe qu'un autre, alors tout élève qui résout le premier devrait réussir à résoudre le second ".

Plus précisément, Régis Gras, (Gras, 1979) avait conçu a priori en 1976 une taxonomie d'objectifs cognitifs, c'est-à-dire un préordre partiel entre des compétences attendues de l'élève au cours de l'apprentissage et du fonctionnement opératoire des concepts mathématiques. Par exemple, " Choix et ordonnancement d'arguments " y précéderait " Critique d'argumentation et construction de contre-exemples " et y suivrait " Effectuation d'algorithmes simples ". Il attendait de tests divers, constitués de variantes d'exercices présentés à des élèves de collège (13 à 15 ans), la validation ou la remise en question de cette taxonomie a priori. Sous forme de graphe orienté sans cycle, l'organisation des performances observées devrait permettre d'étudier l'adéquation de la taxonomie au préordre restitué par le graphe et, accessoirement d'étudier des distorsions liées à deux méthodes d'enseignement différentes. Tout enseignant, comme tout chercheur en didactique des mathématiques, sait par la pratique pédagogique ou par l'observation que des contre-exemples surgissent dans les situations observées par rapport aux hypothèses émises sur la performance. Un outil statistique s'avérait alors nécessaire pour évaluer et représenter les quasi-règles dégagées de la contingence sur la base des résultats obtenus<sup>2</sup>.

Revenant à l'énoncé R ci-dessus, il exprime une règle qui n'est que rarement strictement observée. Il ne peut donc pas avoir le statut de théorème au sens défini dans le champ des mathématiques. Cependant, celui-ci s'inscrit pleinement dans le cadre

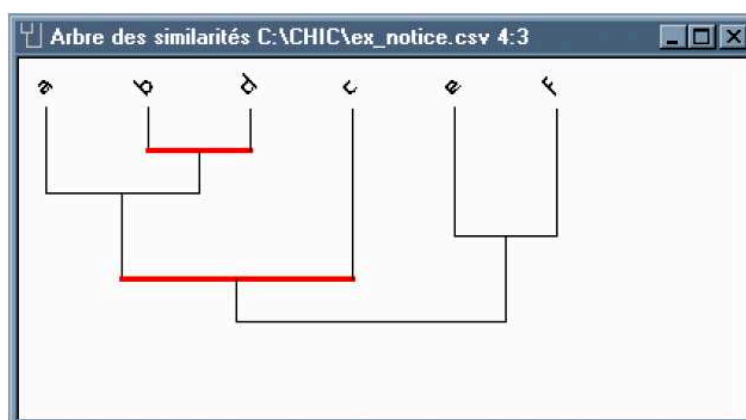
---

<sup>5</sup> <http://sites.univ-lyon2.fr/asi/10/index.php?page=0>

paradigmatique de la relation d'implication statistique qui fait l'objet central de cet ouvrage et où les règles s'expriment sous la forme : " Si on observe a, alors on observe généralement b ". »

Selon Gras et Aga Almouloud (2005) l'ASI rend possible non seulement la classification des éléments d'une banque de données comme cela est fait avec les méthodes d'Analyse multivariée de données mais aussi la détermination d'un niveau d'implication au moyen d'un indice, c'est-à-dire dans quelle mesure une variable a suit une variable b, ou en d'autres termes, en accord avec Régnier et al (2012) quand une variable a est choisie, nous observons une tendance à choisir une variable b. La relation binaire non symétrique nommée implication est une relation logique par laquelle une chose se conclut à partir d'une autre. Toutefois dans le cadre de l'ASI les relations établies entre les variables ne sont la plupart du temps que des règles de quasi-implication. Pour des raisons historiques, le cadre de l'ASI inclut aussi une procédure de classification fondée sur la relation symétrique de similarité basée sur l'approche développée par Lerman (1981).

Selon l'article de Couturier (2001) le logiciel CHIC dans ses versions successives possède une série de fonction, parmi elles, celle de calculer des paramètres de la statistique de base, tel que la moyenne, l'écart-type, le coefficient de corrélation. Une autre fonction est la construction de représentations graphiques. Pour donner un premier exemple nous présentons par la Figure 4 ci-dessous un arbre de similarité construit à partir d'indices de similarité fondés sur une approche probabiliste.



Source: Couturier, 2001

**Figure 4: arbre des similarités**

Couturier (2001) explique que les variables b et d sont, en principe, les variables les plus semblables. Ensuite l'algorithme conduit à associer pour des raisons de proximité la variable a à la classe déjà formée (b, d) pour obtenir au niveau suivant la nouvelle classe

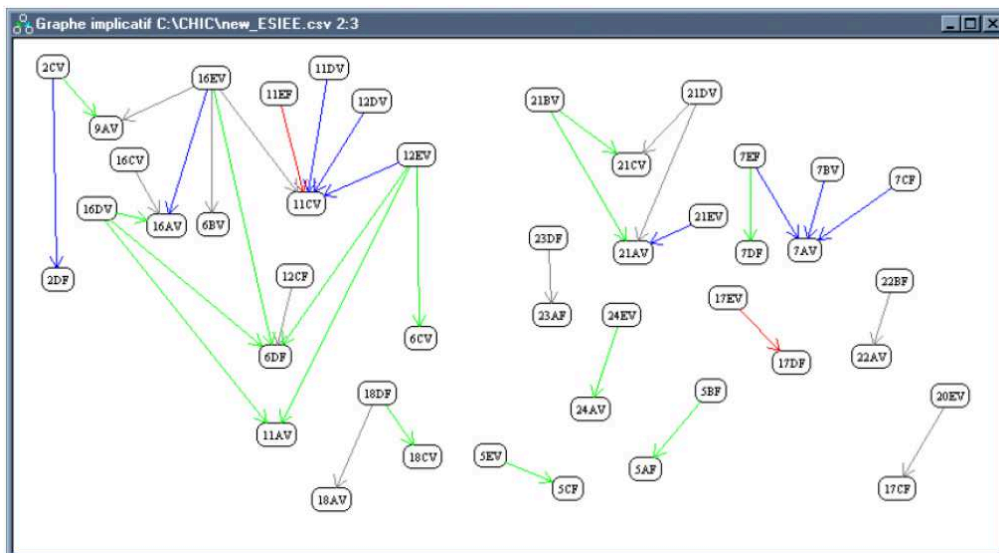
(a, (b, d)). Au troisième niveau c'est la classe (e, f) qui est formée, puis au quatrième niveau la variable c vient s'agréger à la classe (a, (b, d)). Finalement à l'ultime niveau nous obtenons une classe unique qui constitue le tout par agrégation de (a, b, d, c) et (e, f). Nous avons ainsi une succession de partition où les classes sont de moins en moins homogènes. Un critère mis en œuvre par le logiciel permet de mettre en évidence les liens qui sont dits les plus significatifs dans un sens que nous ne développons pas ici mais ce marquage indique à l'utilisateur de porter davantage son attention sur l'interprétation de ces liaisons.

Par le logiciel CHIC, nous pouvons réaliser le calcul de ce qui est appelé l'intensité d'implication entre deux variables a et b qui est une valeur probabiliste, et non une fréquence, qui fonde la décision de retenir ou non une relation de quasi-implication entre ces deux variables. Plus précisément dans le cas des variables binaires, nous appuyons sur les deux définitions suivantes :

Définition 1: La quasi-règle  $a \Rightarrow b$  est admissible au niveau de confiance  $1-\alpha$  si et seulement si  $\Pr[\text{Card}(X \cap Y) \leq \text{card}(A \cap B)] \leq \alpha$

Définition 2 : On appelle intensité d'implication de la quasi-règle  $a \Rightarrow b$ , le nombre  $\phi(a,b) = 1 - \Pr[\text{Card}(X \cap Y) \leq \text{Card}(A \cap B)]$  si  $nb \neq n$  et  $\phi(a,b) = 0$  si  $nb = n$

De là en faisant choix de divers niveaux d'intensité d'implication identifiés par les couleurs des flèches le logiciel CHIC produit la représentation graphique ci-dessous appelé graphe implicatif.

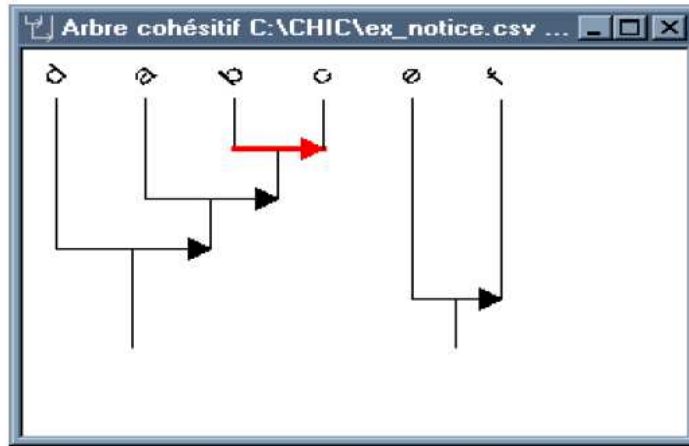


Source: Couturier, 2001

**Figure 5: graphe implicatif**

Il s'agit alors d'interpréter ces liaisons ordonnées et les chemins ainsi construits.

Nous pouvons aussi nous intéresser aux relations entre classes de variables. C'est alors le but de la classification hiérarchique orientée basée sur une relation nommée cohésion. La classification est réalisée à partir d'une mesure de cette cohésion qui intègre dans sa formule de calcul les informations apportées par l'intensité d'implication. Le logiciel produit alors une représentation graphique.



Source: Couturier, 2001

**Figure 6: arbre cohésitif**

Le graphique ci-dessus désigné sous le nom arbre cohésitif représente une classification hiérarchique orientée fournissant à chaque niveau une partition de l'ensemble des variables. Gras (2004) explique que la construction de l'arbre cohésitif comporte des marqueurs (flèche en rouge) pour indiquer les liaisons sur lesquelles l'attention du chercheur peut être portée en priorité pour l'interprétation. Par ailleurs le critère dit de cohésion de la classe à un niveau  $r$  pour réaliser la classification peut conduire à un arrêt qui conduit à ce que la partition ne soit pas nécessairement la partition triviale  $\{\emptyset, \mathbb{E}\}$ . Pour plus de précision nous rappelons ici la définition de la cohésion d'une classe à un niveau  $k$  de la classification (Gras et al., 2009 p.62).

**Définition 1:** La **cohésion d'une classe** de degré 2, correspondant au couple  $(a,b)$  est définie, à partir de l'entropie au sens de Shannon, par la formule,

$$coh(a,b) = \left(1 - \left[-p(\log_2(p)) - (1-p)(\log_2(1-p))\right]^2\right)^{\frac{1}{2}} \text{ où } p = \varphi(a,b) \geq 0,50 \text{ et } coh(a,b) = 0 \text{ si } p = \varphi(a,b) < 0,50.$$

**Définition 2:** La cohésion d'une classe  $C = (a_1, a_2, \dots, a_r)$  de degré  $r$ , est définie par la formule

$$coh(C) = \left[ \prod_{i=1, \dots, r-1}^{j=2, \dots, r; j>i} coh(a_i, a_j) \right]^{\frac{2}{r(r-1)}}$$

**Définition 3:** La cohésion d'une classe  $C = (a)$  de degré 1, est définie par  $coh(a) = 1$

Revenant au marquage des flèches sur l'arbre cohésitif, Couturier (2001) en explique la raison selon laquelle à ce niveau la valeur de la cohésion est la plus élevée entre toutes les valeurs correspondant à tous les couples de variables. Acioly-Régnier et Régnier (2005) ont étudié la question de la lecture-interprétation de ce genre de graphique par des utilisateurs non spécialistes du domaine. Malgré l'apparente simplicité de la représentation, ces auteurs ont identifié plusieurs difficultés parmi lesquelles l'orientation de l'arbre et le marquage des flèches en rouge peuvent constituer des obstacles à l'interprétation.

## **1.2 Logiciel SPAD 9.1 Système Portable d'Analyse des Données**

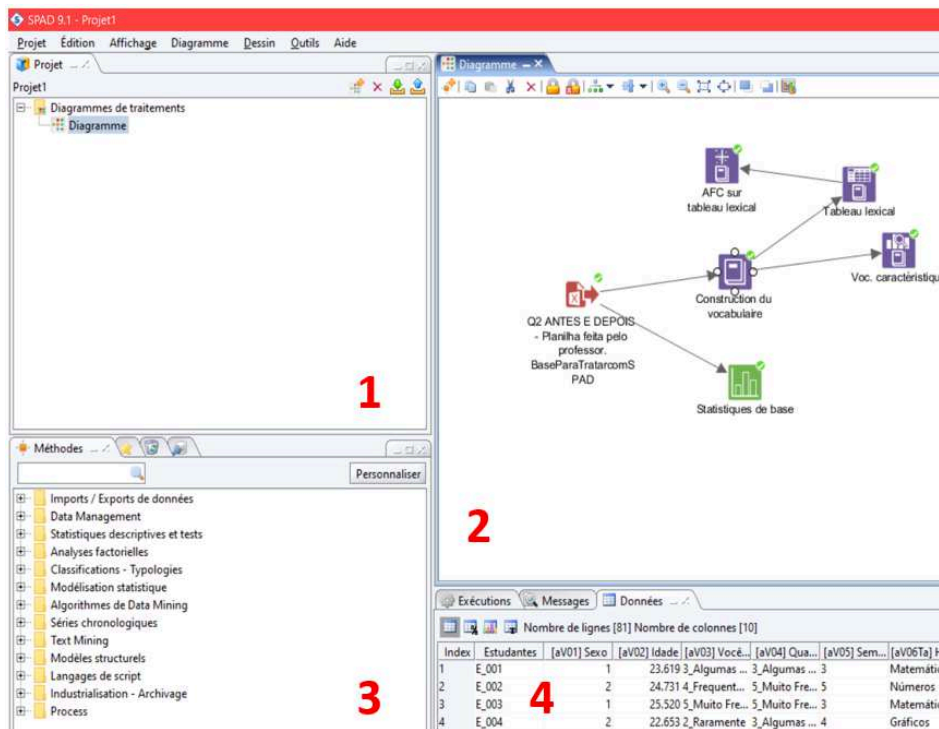
Le logiciel SPAD – Système Portable pour l'Analyse des Données a été développé par l'entreprise Coheris<sup>6</sup> et classé parmi la famille des logiciels de traitements statistiques comme SAS développé par Addinsoft<sup>7</sup>, Tanagra ou R qui est un langage dont la diffusion est soutenue par R Foundation for Statistical Computing. R fait partie de la liste des paquets GNU et est écrit en C (langage), Fortran et R. Cet inventaire ne recouvre pas toute la famille des logiciels de traitements statistiques, il permet seulement de pointer quelques-uns qui permettent l'exploration de grands volumes de données. Tufféry (2010) fait remarquer que, face à la profusion de logiciels de statistique, SPAD se détache de ses concurrents par son importante diffusion dans le monde universitaire français, en raison de la qualité des algorithmes mis en œuvre dans les procédures de traitements des données.

Ce logiciel SPAD dans sa version 9.1 implémente une série de méthodes, comme l'analyse descriptive et exploratoire des données avec la classification selon plusieurs modèles, les traitements de tableaux de croisements de variables, les analyses bivariées, etc.. Il permet de réaliser des analyses factorielles comme l'analyse en composantes principales ACP, analyse des correspondances simples AFC, analyse des correspondances multiples AFCM, etc. Il fournit aussi des méthodes de Text Mining pour réaliser des analyses de données textuelles.

---

<sup>6</sup> [www.coheris.com/produits/analytics/logiciel-data-mining/](http://www.coheris.com/produits/analytics/logiciel-data-mining/)

<sup>7</sup> [www.xlstat.com/en/](http://www.xlstat.com/en/)



Source: L'auteur

**Figure 7: interface logicielle SPAD 9.1**

Nous avons donc utilisé la version SPAD 9.1 qui présente une interface intuitive telle que la figure ci-dessus le montre. Nous avons bénéficié de l'accès à ce logiciel grâce au Professeur Jean-Claude Régnier qui assure un enseignement de statistique basé l'utilisation de ce logiciel mis à disposition par l'Université Lumière Lyon2 sur la base d'une licence de site. Dans ce cadre les doctorants encadrés par Jean-Claude Régnier bénéficient d'une licence annuelle gratuite pour l'usage du logiciel installé sur leur propre ordinateur. L'interface est divisée en 4 espaces comme le montre la figure. Les résultats des traitements algébriques sont fournis par l'intermédiaire d'un tableur comme Excel et les traitements graphiques par des interfaces fournissant les figures.

## 2. Enquête par questionnaires de recherche

Pour construire les données nous avons procédé par une enquête par questionnaire. Nous avons construit deux questionnaires que nous présentons dans les sections suivantes.

### 2.1 Premier questionnaire Q1

Le premier questionnaire Q1 (voir annexe 1) est une adaptation de celui construit par Jean-Claude Régnier pour nombreuses recherches en France qu'il a effectuées sur le thème de la formation à la statistique pour des utilisateurs non spécialistes. Régnier (2003) rappelle que ce questionnaire peut servir d'appui aux études portant sur les difficultés et

facilités dans l'apprentissage de la statistique. Adapté au contexte de notre recherche, le questionnaire en plus des données démographiques est articulé autour de trois axes :

1. Vous et votre parcours antérieur de formation
2. Vous et la statistique
3. Vous et votre sentiment envers la statistique

Nous avons réalisé des analyses des données construites par cette enquête par questionnaire Q1 à partir d'approches statistiques descriptives et inférentielles avec l'aide du logiciel SPAD 9.1. En ce qui concerne la structure des questions qui portent sur le sentiment à l'égard de la statistique (question V15 constituant un vecteur-variable à 24 composantes), nous nous inspirons des formats des questionnaires d'attitude basés sur 24 items énonçant un sentiment positif ou négatif à l'égard de la statistique associés à une échelle de Lickert à quatre niveaux : 1\_Tout à fait en désaccord, 2\_Putôt en désaccord, 3\_Plutôt en accord, 4\_Tout à fait d'accord. Plusieurs études parmi lesquelles nous citons Gal e Ginsburg (1994), Sgoutas-Emch e Johnson (1998), Gonzalez (2002) mettent en avant l'importance de l'analyse des composantes affectives pour l'étude de quelque discipline que ce soit car elles peuvent indiquer les causes de difficultés dans le développement cognitif des étudiants. Nous entendons par affectivité, un processus interne qui englobe d'autres processus psychologiques telles que la motivation ou les attitudes. Ces processus exercent un rôle important dans la cognition des étudiants dans la mesure où les attitudes comportent une lourde charge affective et conditionnent les actions à conduire sur les objets d'apprentissage constitués par les concepts et techniques de la statistique. Les réponses apportées en face de chaque item par les étudiants fournissent une information précieuse susceptible de mieux comprendre les difficultés et les facilités dans l'apprentissage de la Statistique.

Parmi les 24 items de la question V15, 13 exprime une relation négative et 11 à une relation positive. Delà être en accord avec un item positif traduit une attitude positive à l'égard de l'objet contenu dans celui, tandis être en désaccord traduit une attitude non positive que nous interprétons comme négative. De même être en accord avec un item négatif traduit une attitude négative à l'égard de l'objet contenu dans celui, tandis qu'être en désaccord traduit une attitude non négative que nous interprétons comme positive. Nous avons alors, de cette façon, ramené à un indice allant de 0 à 1 traduisant un degré de positivité à l'égard de la statistique que nous nommons *indice de positivité* et calculé ainsi :



$$\text{indice de positivité} = \frac{65 - \sum \text{items}_{\text{négatifs}} + \sum \text{items}_{\text{positifs}} - 24}{72}$$

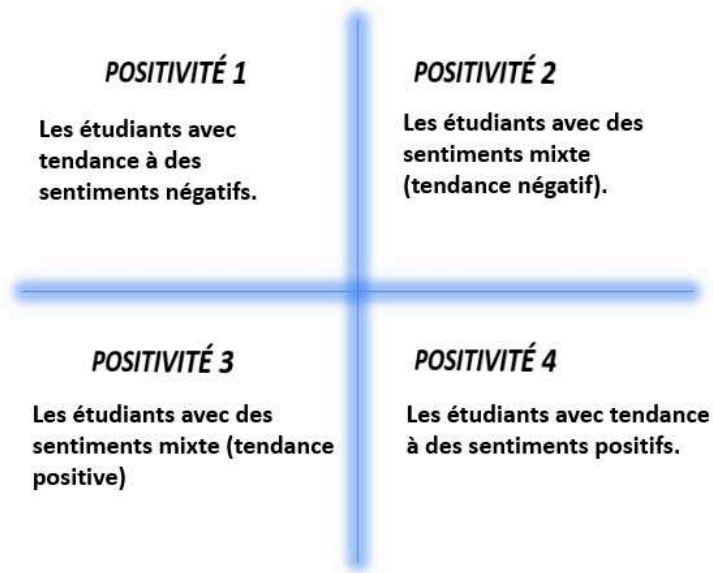
$$\text{indice de positivité} = \frac{41 + \sum \text{items}_{\text{positifs}} - \sum \text{items}_{\text{négatifs}}}{72}$$

Source: L'auteur

**Figure 8: indice de positivité**

Après avoir calculé l'ensemble des indices de positivité individuels avant et après la réalisation de la formation en statistique centrée sur l'estimation statistique, nous avons procédé au quartilage de la série des données *avant* et de celle *après*.

Ainsi la première zone, appelée Positivité 1, regroupe les individus dont les indices de positivité respectifs *avant*(a) et *après*(d) sont inférieurs ou égaux à la valeur du 1<sup>er</sup> quartile. Ce sont les étudiants (a\_Positividade\_1; d\_Positividade\_1) qui manifestent une tendance à une représentation plutôt négative à l'égard de la statistique. La deuxième zone, appelée Positivité 2, regroupe les individus dont les indices de positivité *avant* et *après* sont supérieurs à la valeur du 1<sup>er</sup> quartile et inférieurs à la médiane. Ce sont les étudiants (a\_Positividade\_2; d\_Positividade\_2) qui manifestent un sentiment mixte mais plutôt orientée négativement. La troisième zone, appelée Positivité 3, regroupe les individus dont les indices de positivité *avant* et *après* sont entre la médiane et le 3<sup>ème</sup> quartile. Ce sont les étudiants (a\_Positividade\_3; d\_Positividade\_3) qui manifestent un sentiment mixte mais plutôt orientée positivement. Enfin la quatrième zone, appelée Positivité 4, regroupe les individus dont les indices de positivité *avant* et *après* sont supérieurs au 3<sup>ème</sup> quartile. Ce sont les étudiants (a\_Positividade\_4; d\_Positividade\_4) qui manifestent une tendance à une représentation plutôt positive à l'égard de la statistique. Cette classification des étudiants déterminées par l'orientation de leurs attitudes à l'égard de la statistique nous a servi dans nos interprétations ultérieures des liens possibles entre l'attitude et nos hypothèses de recherche.



Source: L'auteur

**Figure 9: zones de sentiments à l'égard de la statistique**

## 2.2 Second questionnaire Q2

Le second questionnaire (Q2) vise à soumettre les étudiants à des situations-problème que nous considérons comme propices pour identifier les connaissances déjà acquises par relatives à l'estimation statistique. Ces situations ne sont pas associées aux connaissances spécifiques visées dans le cursus de *graduation*, terrain de notre recherche. Elles possèdent une des caractéristiques principales de la Statistique comme le souligne Gal (2002) à savoir son ancrage dans des situations du quotidien.

Avec comme finalité, celle de construire de données en ayant en toile de fond le champ conceptuel de l'estimation statistique, nous avons proposé de situations qui sollicitent les mêmes concepts statistiques dans le sens de Vergnaud (1996b) qui soutient que l'étude d'un concept doit s'appuyer sur un champ de problèmes distincts mais avec une base identique. Les deux premières situations-problèmes proposée dans le questionnaire visent identifier chez les étudiants, les composantes cognitives à travers la mobilisation des invariants opératoires (concept-en-acte, théorème-en-acte) qui adviennent en statistique descriptive, comme les notions de population, échantillon, variable statistique, moyenne, écart-type, et les relations mathématiques associées. La troisième situation cherche à permettre d'observer les composantes de disposition en analysant la posture critique à travers la capacité à générer des questionnements par la confrontation à informations de nature quantitative. Dans ce sens, les variables ont été déterminées selon la proposition de Gal (2002) qui pointe la nécessité d'établir un lien

entre les composantes cognitives et celles de la disposition pour le développement du raisonnement statistique.

Les énoncés de problèmes proposés dans le questionnaire Q2 sont les suivants.

Commençons par la première situation

Situation S1_1	Supposons que vous êtes intéressé par une recherche sur le nombre d'occupants des automobiles personnelles qui passent chaque jour dans une avenue mouvementée de votre ville. De quelle façon procéderiez-vous pour initier cette recherche ?  <i>Vamos supor que você esteja interessado em investigar a quantidade de ocupantes dentro dos automóveis de passeio que transitam por dia em uma avenida movimentada de sua cidade. Como você procederia para dar início a essa investigação?</i>
Situation S1_2	En face des informations recueillies, quel type de mesures statistiques utiliseriez-vous pour établir un type de modèle. Faites une simulation.  <i>Frente as informações levantadas, que tipo de medidas estatísticas você utilizaria para estabelecer algum tipo de padrão ? Faça uma simulação.</i>

Cette première situation vise à ce que les étudiants perçoivent l'usage de quelques procédures de la statistique descriptive. Ainsi ce que nous cherchions à identifier sont les notions suivantes

Avant la formation sur l'estimation statistique	Après ...
aS1_1 échantillon	dS1_1 échantillon
aS1_2A moyenne	dS1_2A moyenne
aS1_2B écart-type	dS1_2B écart-type
aS1_2C calcul	dS1_2C calcul
	dS1_2D notion d'estimation

Passons à la seconde situation.

Situation S2_1	Dans un supermarché, le responsable du secteur des fruits et légumes a été chargé de mettre les pommes en paquets de deux kilogrammes. Qu'est-ce que cet employé devra prendre en compte ?  <i>Em um supermercado o funcionário responsável pelo hortifrutigranjeiro recebeu a incumbência de empacotar as maçãs em pacotes de dois quilos. O que o funcionário deverá levar em consideração ?</i>
Situation S2_2	Après avoir mis en paquets toutes les pommes, au total 150 paquets, l'employé est resté curieux de connaître la distribution des fruits dans les sacs. N'ayant pas de temps pour faire un comptage dans tous les sacs, quel procédure pourrait-il réaliser pour obtenir cette information ?  <i>Após empacotar todas as maçãs, num total de 150 pacotes, o funcionário ficou curioso em saber sobre a distribuição das frutas nos pacotes. Não possuindo tempo para fazer uma</i>

	<i>contagem sobre todos os pacotes, que procedimento ele poderia realizar para fazer essa descoberta? .</i>
Situation S2_3	Quels types de mesures statistiques pouvons-nous utiliser pour essayer d'établir un modèle relatif aux paquets de pommes ? Faites une simulation.  <i>Que tipos de medidas estatísticas podemos proceder para tentarmos estabelecer algum padrão em relação aos pacotes de maçãs? Faça uma simulação.</i>

Cette seconde situation vise à ce que les étudiants perçoivent l'usage des procédures suivantes

Avant la formation sur l'estimation statistique	Après ...
aS2_1 variabilité	dS2_1 variabilité
aS2_2 échantillon	dS2_2 échantillon
aS2_3A moyenne	dS2_3A moyenne
aS2_3B écart-type	dS2_3B écart-type
aS2_3C calcul	dS2_3C calcul
	dS2_2D notion d'estimation

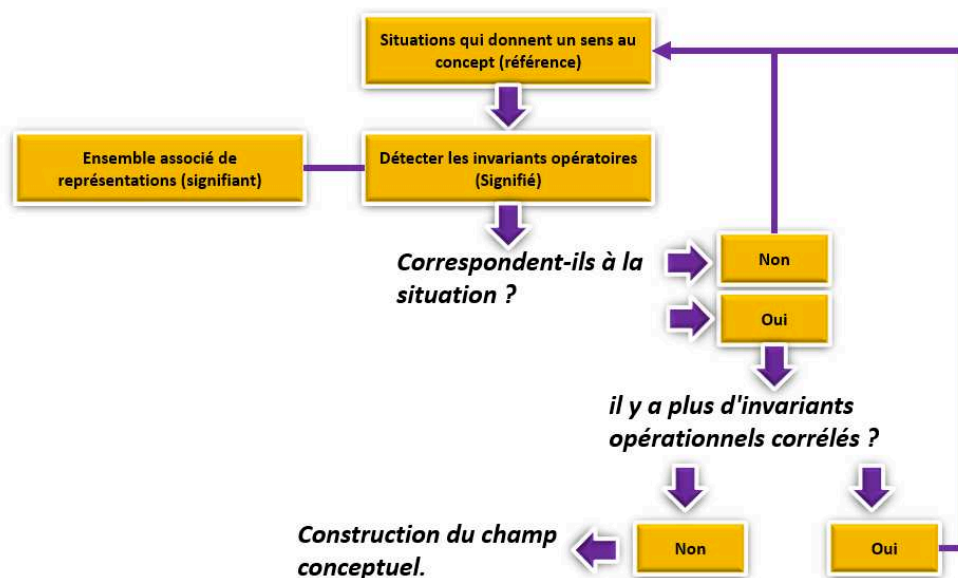
Enfin nous abordons la troisième situation

Situation S3_1	Dans les deux situations, croyez-vous qu'il puisse y avoir des erreurs en relation aux modèles établis ? Justifiez votre réponse.  <i>Em ambas as situações, você acredita que possam haver erros em relação aos padrões estabelecidos ? Justifique sua respostas.</i>
-------------------	--

Cette troisième situation vise à ce que les étudiants perçoivent les éléments suivants

Avant la formation sur l'estimation statistique	Après ...
aS3_1 notion d'erreur	dS3_1 notion d'erreur échantillonnale
aS3_2 justification	dS3_2 justification en lien avec la conception de l'estimation statistique

Pour explorer les réponses à la lumière de la théorie de champs conceptuels et pour les analyser dans le cadre théorique de l'Analyse statistique implicite avec l'aide du logiciel CHIC 7.0, nous avons schématisé un algorithme que nous utilisons comme critère et que nous présentons dans la figure ci-dessous (Figure 10)



Source: L'auteur

Figure 10: algorithme pour analyser les données construites avec le questionnaire Q2

Ce schéma montre de façon simplifiée les étapes fondamentales appuyées sur l'approche de Vergnaud (1990a) pour la construction d'un champ conceptuel. Les situations proposées en Q2 cherchent à donner un sens aux concepts visés. Sous cette forme, nous cherchons à identifier les invariants opératoires correspondance avec les variables analysées dans chaque question. Nous avons aussi considéré l'ensemble des représentations qui sont associées aux variables et liées horizontalement aux invariants opératoires. En ce sens, pour chaque réponse cohérente avec la visée des variables analysées, nous avons inséré le nombre 1 dans le tableau Excel sinon 0.

A	B	C	D	E	F	G	H	I	J	K	L
	aS1_1	aS1_2A	aS1_2B	aS1_2C	aS2_1	aS2_2	aS2_3A	aS2_3B	aS2_3C	aS3_1	aS3_2
E01_001	0	1	0	0	0	0	1	0	0	1	0
E01_002	1	0	0	1	0	0	1	0	1	0	0
E01_003	1	0	0	1	1	1	0	0	0	0	0
E01_004	1	1	0	0	1	1	0	0	1	0	0
E01_005	0	1	1	0	0	1	1	0	1	0	0
E01_006	0	0	0	1	0	0	0	0	0	0	0
E01_007	1	0	0	0	0	1	0	0	0	1	0
E01_008	1	0	0	0	0	1	1	1	1	1	0
E01_009	1	0	0	0	0	1	0	0	0	0	0
E01_010	1	0	0	1	0	1	0	0	1	1	0
E01_011	1	1	1	0	1	1	1	0	0	1	0
E01_012	0	0	0	0	0	1	0	0	0	0	0
E01_013	1	0	0	0	0	1	1	0	1	1	0
E01_014	1	0	0	1	0	0	0	0	0	1	0
E01_015	1	0	0	0	1	0	1	0	1	0	0
E01_016	1	0	0	0	0	1	0	0	0	1	0
E01_017	1	0	0	0	0	1	1	1	0	0	0
E01_018	1	0	0	0	0	0	1	0	0	1	0

Source: L'auteur

Figure 11: variables binaires

Ainsi nous avons construit un tableau croisant les 81 individus et l'ensemble des variables binaires qui traduisent la présence ou l'absence de l'attribut analysé, qui sert de base pour le traitement de données avec le logiciels CHIC 7.0 et construire les représentations graphiques tels que le graphe implicatif, l'arbre cohésitif et même l'arbre des similarités, qui constituent les outils de base pour les interprétations dans le cadre de l'analyse statistique implicative.

### **PARTIE 3 – RÉSULTATS, DISCUSSIONS ET PROLONGEMENTS**

Après avoir présenté nos choix méthodologiques en indiquant le contexte de l'étude et en spécifiant les instruments utilisés pour la construction et les traitements des données, nous abordons maintenant les résultats obtenus. Pour le premier questionnaire Q1, nous donnons ici dans ce résumé une synthèse des principales informations relatives aux variables analysées. Concernant l'échantillon des individus, nous obtenons les caractéristiques suivantes

V01 : parmi les individus dans l'échantillon, 51% sont de sexe féminin et 49% du sexe masculin. Il n'apparaît aucune différence significative entre le genre selon les zones de sentiments à l'égard de la statistique.

V02 L'âge moyen est environ de 24,2 ans avec un écart-type de 5,6 ans

V03 les étudiants reconnaissent les situations dans lesquelles la statistique est nécessaire, indépendamment du niveau de positivité donné par l'indice pris en référence.

V04 les étudiants ayant un indice de positivité bas sont plus sujets à rencontrer des difficultés dans des situations qui exigent des connaissances en statistique. Toutefois ce fait se réduit après la reprise des concepts et procédures opératoires de base de la statistique descriptive.

V05 Il apparaît une tendance parmi les étudiants à laisser la discipline de statistique pour être étudiée avec plus de temps, indépendamment du niveau de positivité donné par l'indice pris en référence

V06 la statistique est rappelée, principalement, par les mathématiques ou par des éléments mathématiques, indépendamment du niveau de positivité donné par l'indice pris en référence

V07 les étudiants ont une tendance à attribuer un bon niveau d'intérêt personnel pour la statistique dans toutes les zones de sentiments

V08 les étudiants manifestent une tendance à attribuer un bon niveau d'utilité à la statistique dans toutes les zones de sentiments, et il ressort une corrélation positive entre le niveau d'intérêt et le niveau d'utilité.

V09 les étudiants reconnaissent les termes statistiques, toutefois pour les étudiants ayant un plus fort indice de positivité il apparaît une tendance à utiliser des termes plus techniques dans les définitions.

V10 les étudiants ont une tendance à reconnaître les termes de statistique dans toutes les zones de sentiments

V11 les étudiants manifestent une tendance à reconnaître l'utilité de la statistique dans les relations de recherches et les interprétations des informations, dans toutes les zones de sentiments

V12 la principale difficulté évoquée par les étudiants en relation à la statistique est liée aux mathématiques et ceci dans toutes les zones de sentiments

V13 les étudiants expriment une forte tendance à associer l'utilité de la statistique, à l'interprétation des informations et à l'élaboration de recherche, dans toutes les zones de sentiments

V14 les étudiants ont une tendance à déclarer avoir suivi un enseignement de statistique entre un niveau superficiel et un niveau basique

V15 les indices de positivité ont eu une tendance à augmenter après la formation portant sur l'estimation statistique, qui peut avoir une cause dans la reprise des contenus de la statistique descriptive et des opérations mathématiques.

V16 les étudiants ont une tendance à reconnaître la statistique comme une science qui collecte, analyse et interprète des données pour une recherche.

Nous avons analysé les données construites avec le second questionnaire au moyen du logiciel CHIC 7.0 dans le cadre de l'analyse statistique implicite en considérant les réponses données par les étudiants avant et après la formation sur l'estimation statistique. L'analyse a eu comme base la recherche des invariants opératoires utilisés par les étudiants dans la résolution des situations-problèmes, en suivant le schéma présenté en Figure 10. Cette analyse des réponses a mis en évidence la possible formation d'un champ conceptuel de l'estimation statistique. Vergnaud (1990) insiste sur le fait que dans la théorie des champs conceptuels il convient de considérer les individus comme des sujets-en-situation. Cette caractéristique rend la théorie utile dans la planification et dans l'analyse des situations d'enseignement, dès lors que nous devons accompagner les étudiants dans leur apprentissage, en recherchant dans les concepts et les théorèmes en acte, l'évolution temporelle de leur niveau de connaissance. Dans ce

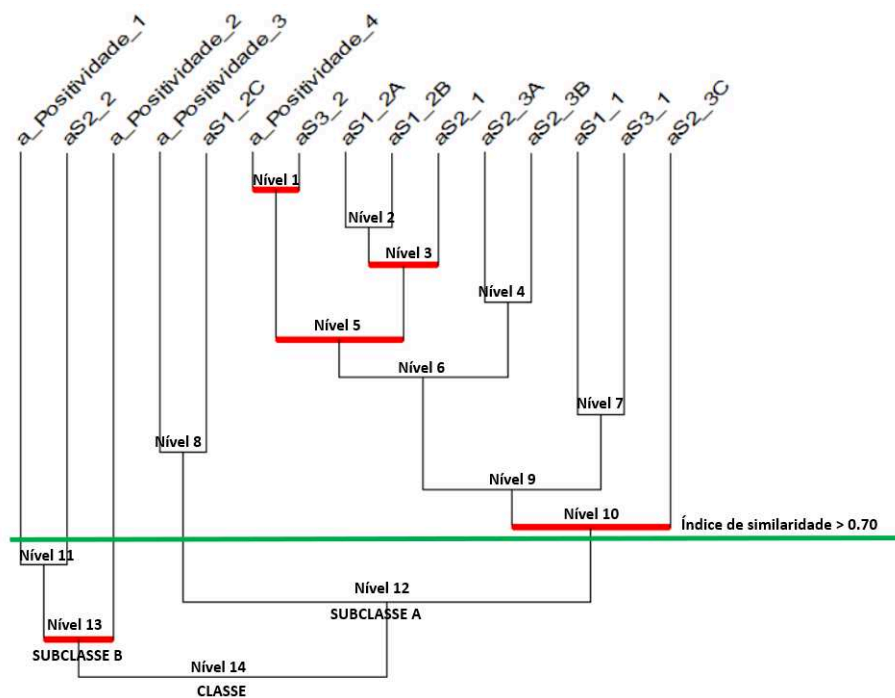


sens nous mettons en relation les variables impliquées dans le second questionnaire avec les variables de V15 *A ce jour votre sentiment envers la statistique ?* du premier questionnaire à partir de laquelle nous avons construit ce que nous avons appelé les zones de sentiment à l'égard de la statistique. De cette manière, nous avons cherché à observer si l'évolution des connaissances des sujets est reliée aux attitudes à l'égard de la statistique.

Les variables ont été modélisées par des variables binaires {0 ; 1} où la présence de l'occurrence est notée 1 et son absence 0. Comme nous l'avons déjà dans les sections précédentes nous avons construit sous Excel un tableau constitué de 0 et de 1 qui traduit le croisement de l'ensemble des sujets avec l'ensemble des variables. Ce type d'analyse permet comme le rappelle Gras et al (2009, 2013, 2017) d'étudier et d'interpréter en termes de typologie et relation de ressemblance, des classes de variables constituées à certains niveaux de l'arbre de classification ou s'opposant à d'autres niveaux. Nous avons aussi analysé le graphe implicatif construit à partir de l'indice d'implication selon le modèle binomial et en adoptant la valeur 0.70 comme seuil minimal de l'indice.

### **1. Résultats issus de la résolution des situations-problèmes avant la formation sur l'estimation statistique**

Nous commençons nos analyses avec l'arbre des similarités (Figure 12). Pour avoir une idée sur la formation de cette classification, nous pouvons déjà observer qu'en prenant comme repère le seuil critique de 0.70 de l'indice de similarité, nous obtenons alors une partition constituée par la classe du niveau 10 et celle de niveau 8 auxquelles s'adjoignent deux singletons {a\_Positividade\_1} et {a\_S2\_2}.



Source: L'auteur

**Figure 12: arbre des similarités "avant"**

Toutefois, si, à nouveau, nous prenons en considération l'arbre dans sa totalité, nous trouvons dans la sous-classe B au niveau 13, seulement un nœud au niveau 11 avec un indice de similarité moindre que 0.70. Cette sous-classe B est partitionnée deux éléments {a\_Positividade\_1, a\_S2\_2}.et {a\_Positividade\_2} qui regroupent les étudiants qui manifestent un indice de positivité plutôt bas. La variable aS2\_2 correspond à la reconnaissance du concept d'échantillon à travers des représentations sémiotiques du registre graphique comme nous le rapportons en Figure 13.



Source: L'auteur

**Figure 13 : réponses des étudiants aS1\_1 et aS2\_2**

La sous-classe B indique la tendance des étudiants avec un niveau de positivité plutôt bas à ne pas réussir explicitement dans les autres solutions que les situations-problème ont proposées ou mobilisées implicitement sous forme de concepts-en-acte comme celui d'échantillon. Il ne nous a pas été possible d'identifier des indices de

théorème-en-acte, dès lors qu'ils n'ont pas utilisés de relations mathématiques pour justifier leurs réponses.

La sous-classe A formée au-delà du seuil de 0.70 regroupe les dix variables {a\_Positividade\_4, aS3\_2, aS1\_2A, aS1\_2B, aS2\_1, aS2\_3A, aS2\_3B, aS1\_1, aS3\_1, aS2\_3C}. Cette sous-classe est construite à partir des regroupements successifs des nœuds 1, 2, 3, 4, 5, 6, 7, 9, 10 parmi lesquels les niveaux 1, 3,5 et 10 sont considérés comme les plus significatifs au sens de la théorie ASI et méritent une attention particulière dans l'interprétation des résultats. Il ressort que ces regroupements ont été produits par les réponses fournies par les étudiants qui ont manifesté un niveau de positivité plutôt fort.

<i>Nós com índice superior a 0,70</i>	<i>Variável com maior contribuição</i>
Nível 1 (índice 1)	a_Positividade_4
Nível 2 (índice 0,99)	a_Positividade_4
Nível 3 (índice 0,99)	a_Positividade_4
Nível 4 (índice 0,99)	a_Positividade_4
Nível 5 (índice 0,99)	a_Positividade_4
Nível 6 (índice 0,98)	a_Positividade_4
Nível 7 (índice 0,89)	a_Positividade_4
Nível 8 (índice 0,81)	a_Positividade_3
Nível 9 (índice 0,79)	a_Positividade_4
Nível 10 (índice 0,72)	a_Positividade_4

Source: L'auteur

**Tableau 6 : contributions des variables**

Nous intéressons maintenant à l'interprétation des classes formées aux niveau 1, 3, 5 et 10 identifiées comme les plus significatives ainsi que nous l'avons signalé plus haut.

Niveau 1 {a\_Positividade\_4, aS3\_2} indique une forte similarité entre le fait d'avoir manifesté un niveau de positivité dans la relation à la statistique situé dans le 4<sup>ème</sup> quartile et le fait de fournir des justifications cohérentes pour de possibles erreurs dans les modèles établis pour les situations S1 et S2. Nous obtenons des arguments de ce type:

*“Peguei uma amostra, então não posso afirmar que todos os pacotes de maçãs sigam o padrão que elaborei”* [J'ai pris un échantillon, alors je ne peux pas affirmer que tous les paquets de pommes suivent le modèle que j'ai élaboré.] ;

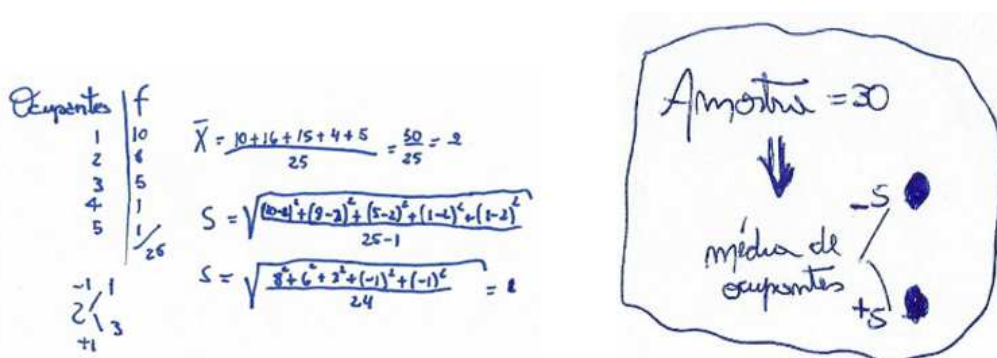
“No caso dos carros, depende muito do tamanho da amostra e do tempo que ficamos observando” [Dans le cas des automobiles, cela dépend beaucoup de la taille de l'échantillon et du temps d'observation.]

Niveau 3 {{aS1\_2A, aS1\_2B}, aS2\_1} indique une forte similarité entre le fait d'identifier la moyenne et l'écart-type dans la situation S1 et celui d'identifier la variabilité de l'élément dans la situation S2.

Niveau 5 {{a\_Positividade\_4 aS3\_2}, {{aS1\_2A aS1\_2B} aS2\_1}} – indique une forte similarité entre le fait que les étudiants qui ont manifesté un niveau élevé de positivité et qui ont apporté des arguments cohérents et pertinents dans la situation S3, et celui d'identifier les concepts des situations S1 et S2.

Niveau 10 {{{{{a\_Positividade\_4, aS3\_2} {{aS1\_2A, aS1\_2B} aS2\_1}} {aS2\_3A, aS2\_3B}} {aS1\_1, aS3\_1}} aS2\_3C} correspond au regroupement des classes de niveaux précédents et intègre globalement les interprétations formulées sur ses composantes.

Nous pouvons remarquer, pour la classe formée au niveau 10, l'importance du niveau le plus élevé de la variable positivité du sentiment à l'égard de la statistique. Les étudiants appartenant à cette catégorie ont interprété les situations S1, S2 et S3 en les plaçant dans un contexte plus scientifique en faisant usage d'éléments conceptuels adéquats comme nous l'exemplifions ci-dessous.



Source: L'auteur

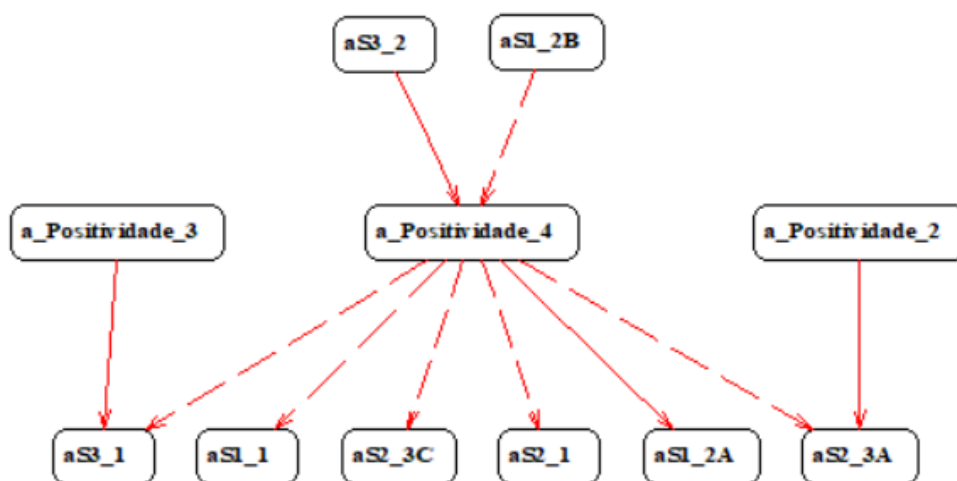
**Figure 14: Réponses fournies à la situation S1 parmi les étudiants ayant un indice de positivité dans la catégorie la plus élevée.**

De ce point de vue, nous pouvons ainsi observer le fait que les étudiants ayant un indice de positivité dans la catégorie la plus élevée ont tendance à manifester une meilleure compréhension du langage statistique utilisé dans les situations-problèmes en opérant avec des théorèmes et des concepts en acte adéquats au contexte. De façon

générale, les deux types d'invariants opératoires sont imbriqués dialectiquement dans l'action de ces étudiants et pour lesquels, selon Vergnaud (1990) les concepts-en-acte ont permis d'extraire du milieu les informations pertinentes et de sélectionner les théorèmes-en-acte nécessaires à la réalisation des calculs.

Il est aussi intéressant de remarquer la structure de l'arbre de similarité qui sépare nettement la nature des réponses fournies par les étudiants ayant un indice de positivité dans la catégorie la plus élevée, de celle des réponses fournies par les étudiants ayant un indice de positivité dans la catégorie la plus basse. Ce fait nous conduit à retenir l'idée de l'influence de la relation à l'objet statistique sur les réponses apportées aux questions du questionnaire Q2 et nous renforce dans l'idée de la validation de la seconde hypothèse de notre recherche.

Nous avons aussi construit le graphe implicatif tel que défini dans le cadre théorique de l'Analyse statistique implicative. Nous avons fait choix du seuil de 0.70 pour l'indice d'implication dont la définition a été rappelée dans nos propos de la Partie 2.



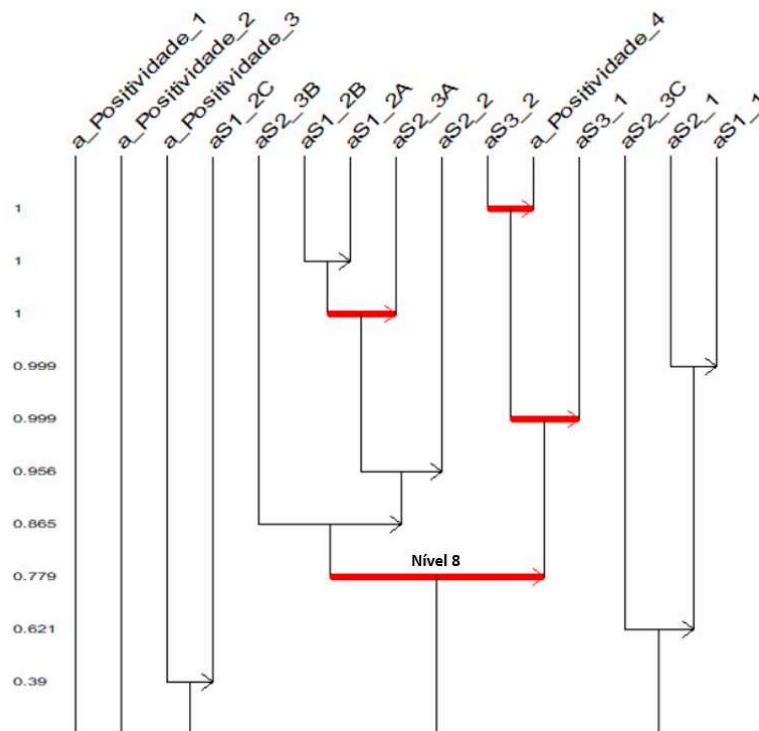
Source: L'auteur

**Figure 15 : graphe implicatif "avant"**

Il est possible de percevoir que la variable a\_Positividade\_1 qui regroupe les étudiants ayant un indice de positivité dans la catégorie la plus basse, n'apparaît pas dans ce graphe. Ceci peut s'interpréter par le fait que les étudiants de cette catégorie n'ont pas obtenus un niveau de performance suffisant dans la résolution des situations-problème. En revanche nous pouvons remarquer l'importance de la variable a\_Positividade\_4, qui regroupe les étudiants ayant un indice de positivité dans la catégorie la plus élevée, dans le réseau formé et mis en évidence par la représentation graphique. Cette variable peut être

interprétée comme une *cause* pour le recours aux concepts-en-acte et aux théorèmes-en-acte adéquats à la résolution des situations-problèmes. Dans ce sens, en présence de ces informations jusqu'ici contrôlées, nous identifions des liens pertinents de type *cause à effet*, entre les composantes de disposition (aS3\_1 e aS3\_2) et les composantes cognitives (aS1\_1, aS1\_2A, aS1\_2B, aS2\_3C, aS2\_1) qui, selon Gal (2002), doivent être en corrélation pour assurer le développement de la littéracie statistique.

Nous avons enfin construit l'arbre cohésitif tel que défini dans la cadre théorique de l'Analyse statistique implicite. Cette représentation graphique est basée sur l'indice de cohésion dont la définition a été rappelée dans nos propos de la Partie 2.



Source: L'auteur

**Figure 16 : arbre cohésitif "avant"**

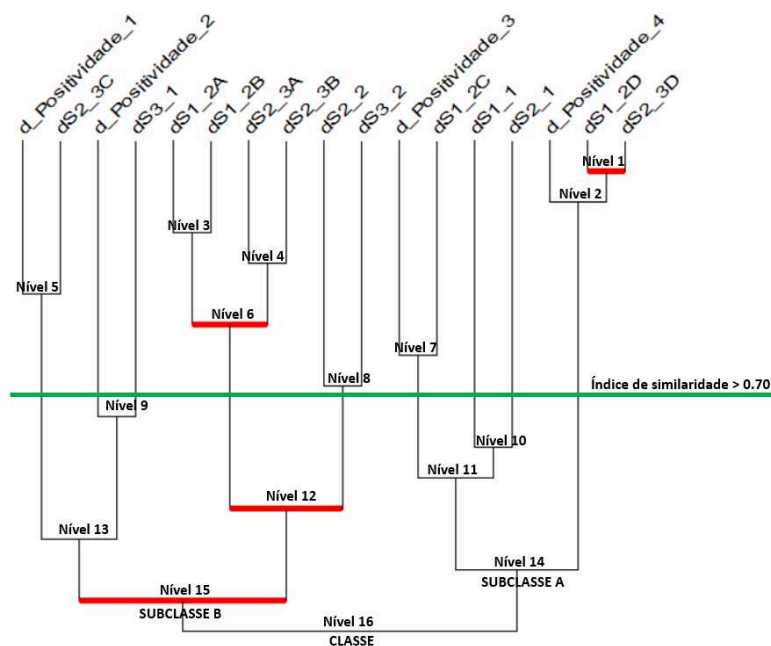
Selon Couturier et al. (2003) la représentation graphique de la classification hiérarchique orientée, se présente sous la forme d'un arbre, appelé arbre cohésitif, et d'un axe vertical sur lequel sont marqués de façon décroissante, les indices de cohésion qui déterminent les niveaux de regroupement des variables. Chaque classe est constituée à partir des indices de cohésion eux-mêmes intégrant les indices d'implication et maintient

une relation de type implication entre ses éléments qui la constituent. Il s'agit de règles qui portent sur des règles.

Pour commencer, nous pouvons remarquer que les variables qui n'ont pas été regroupées et restent isolées sont a\_Positividade\_1 e a\_Positividade\_2 qui ont été interprétées dans les analyses précédentes comme liées à l'absence de mobilisation expressive des invariants opératoires. Ensuite nous identifions la classe formée au niveau 1 de la hiérarchie qui regroupe les variables aS3\_2 et a\_Positividade\_4 et qui peut être interprétée comme l'expression d'une posture critique en relation aux résultats issus des résolutions des situations-problèmes S1 et S2 associée à un indice de positivité dans la catégorie la plus élevée. La classe formée au niveau 2 par le regroupement des variables aS1\_2B et aS1\_2A peut être interprétée par le fait d'identifier l'écart-type est associé à celui d'identifier la moyenne. Ce qui semble cohérent dans la mesure où le calcul de l'écart-type mobilise aussi le concept de moyenne. Enfin si nous considérons la classe formée au niveau 8 et marquée comme significative au sens de l'ASI avec un indice de cohésion de l'ordre de 0.779, nous pouvons remarquer qu'elle s'est constituée autour de la variable a\_Positividade\_4, ce qui nous conforte encore dans le rôle du rapport à la statistique que nous avons mesuré avec l'indice de positivité. Nous retrouvons par là le fait que les étudiants ayant un indice de positivité dans la catégorie la plus élevée ont une nette tendance à employer correctement les concepts de variabilité, d'échantillon, de moyenne et d'écart-type, en plus de mobiliser les relations mathématiques pour justifier leurs réponses.

## **2. Résultats issus de la résolution des situations-problèmes avant la formation sur l'estimation statistique**

Nous procédons comme nous l'avons fait précédemment en réalisant une analyse des similarités. Le traitement avec CHIC 7.0 aboutit à la construction de la classification hiérarchique représentée par l'arbre des similarités (Figure 17). La procédure de regroupement se poursuit en 16 niveaux jusqu'à la partition triviale constituée par l'ensemble  $E$  de toutes les variables et  $\emptyset$ . Le nœud 16 est issu du regroupement de deux classes obtenues respectivement aux niveaux 14 et 15.



Source: L'auteur

Figure 17: arbre des similarités "après"

Une fois de plus, nous pouvons observer le rôle de la variable traduisant ce que nous avons appelé le niveau de positivité du sentiment à l'égard de la statistique. Les variables qui sont regroupées dans la classe B sont associées aux niveaux de positivité plutôt bas alors que dans la classe A, elles sont associées aux niveaux de positivité plutôt élevés. Notons cependant que la classe regroupe un plus grand nombre de variables analysées antérieurement. Cela semble inciter à penser qu'il s'est produit une amélioration du développement de la littéracie statistique chez les étudiants, dès lors que les variables qui contribuèrent le plus dans la formation des nœuds, furent mieux distribuées comme le montre le Tableau 7.

<i>Nós com índice superior a 0,70</i>	<i>Variável com maior contribuição</i>
Nível 1 (índice 0,99)	d_Positividade_4
Nível 2 (índice 0,99)	d_Positividade_4
Nível 3 (índice 0,96)	d_Positividade_4
Nível 4 (índice 0,93)	d_Positividade_1
Nível 5 (índice 0,88)	d_Positividade_1
Nível 6 (índice 0,85)	d_Positividade_2
Nível 7 (índice 0,83)	d_Positividade_3
Nível 8 (índice 0,71)	d_Positividade_4

Source: L'auteur

Tableau 7 : contributions des variables "après"



Analysons maintenant la partition obtenue en fixant le seuil critique de l'indice de similarité au niveau 0.70. Cette partition est constituée par neuf classes dont quatre singletons :  $\{\{d\_Positividade\_1, dS2\_3C\}_1, \{d\_Positividade\_2\}_2, \{dS3\_1\}_3, \{\{dS1\_2A, dS1\_2B\} \{dS2\_3A, dS2\_3B\}\}_4, \{dS2\_2, dS3\_2\}_5, \{d\_Positividade\_3, dS1\_2C\}_6, \{dS1\_1\}_7, \{dS2\_1\}_8, \{d\_Positividade\_4 \{dS1\_2D, dS2\_3D\}\}_9$ .

Niveau 1  $\{dS1\_2D, dS2\_3D\}$  – indique la tendance à ce que les étudiants qui utilisent des éléments de l'estimation dans la situation S1, le font aussi dans la situation S2

Niveau 2  $\{d\_Positividade\_4 \{dS1\_2D, dS2\_3D\}\}$  indique la tendance à ce que les étudiants qui présentent les indices de positivité les plus élevés, ont identifié des éléments de l'estimation dans la situation S1 et la situation S2.

Niveau 3  $\{dS1\_2A, dS1\_2B\}$  indique la tendance à ce que les étudiants qui ont mobilisé le concept de moyenne, ont aussi mobilisé le concept d'écart-type dans la situation S1. Ce qui semble logique de fait de l'utilisation de la moyenne dans la définition de l'écart-type.

Niveau 4  $\{dS2\_3A, dS2\_3B\}$  indique la tendance à ce que les étudiants qui ont mobilisé le concept de moyenne, ont aussi mobilisé le concept d'écart-type dans la situation S2.

Niveau 5  $\{d\_Positividade\_1, dS2\_3C\}$  indique la tendance à ce que les étudiants qui ont pourtant manifesté un bas niveau de positivité, ont toutefois mobilisé les concepts-en-acte et les théorèmes-en-acte adéquats dans la situation S2

Niveau 6  $\{\{dS1\_2A, dS1\_2B\} \{dS2\_3A, dS2\_3B\}\}$  indique la tendance à ce que les étudiants qui ont mobilisé le concept de moyenne, et le concept d'écart-type dans la situation S1, l'ont aussi fait dans la situation S2.

Niveau 7  $\{d\_Positividade\_3, dS1\_2C\}$  indique la tendance à ce que les étudiants qui ont manifesté un haut niveau de positivité, ont mobilisé les concepts-en-acte et les théorèmes-en-acte adéquats dans la situation S1

Niveau 8  $\{dS2\_2, dS3\_2\}$  indique la tendance à ce que les étudiants qui ont mobilisé le concept d'échantillonnage, reconnaissent aussi de possibles erreurs dans les modèles établis.

Nous avons constaté que les invariants opératoires ont été mieux mobilisés quel que soit le niveau de positivité, alors dans la phase antérieure il s'agissait plutôt d'une

caractéristique du groupe de niveau de positivité élevé. De façon générale, les étudiants employèrent les concepts-en-acte et les théorèmes-en-acte adéquats, en justifiant les réponses au moyen de l'explicitation des concepts de moyenne et d'écart-type, par exemple, au-delà de l'utilisation des relations mathématiques existantes. Cependant nous soulignons, dans la classe A, l'élément de niveau 2 {d\_Positividade\_4, dS1\_2D, dS2\_3D} par lequel nous voyons que les étudiants avec les indices de positivité les plus élevés sont ceux qui mobilisèrent les invariants opératoires relatifs à l'estimation statistique comme le montre la figure 18

amostra = 30 carros  
 média = 2 por carro  
 $s = 1$   
 nível de confiança 95%  
 $1,96 \cdot \frac{s}{\sqrt{n}} = 1$   
 $\sqrt{n} = 30$

Source: L'auteur

**Figure 18: invariants opératoires identifiés en relation à l'estimation statistique**

Cette constatation a révélé que les étudiants situés dans la catégorie de ceux qui ont manifesté un sentiment à l'égard de la statistique d'un niveau plutôt élevé ont eu plutôt tendance à utiliser de nouvelles connaissances de forme plus satisfaisante que les autres. Nous continuons ainsi à penser que la relation à la statistique influence le développement du raisonnement statistique.

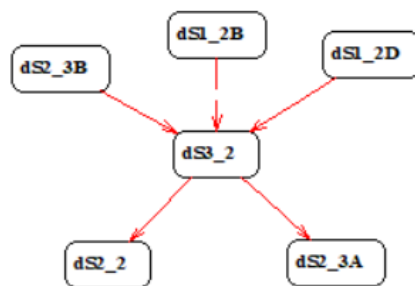
Pour poursuivre notre exploration, nous avons eu recours à la relation de quasi-implication telle que définie dans l'ASI et dont nous avons déjà fait état dans cet écrit. Pour ce faire nous avons construit le graphe implicatif (Figure 19) en fixant le seuil de l'indice d'implication au niveau 0.70



Source: L'auteur

**Figure 19: graphe implicatif "après"**

Pour cette fois, à la différence du graphe implicatif « avant » rapporté à la figure 15, toutes les variables « catégorie de positivité du sentiment à l'égard de la statistique » apparaissent comme des *causes*. De façon générale, la résolution des situations proposées S1, S2, S3 a été mieux développée pour les quatre catégories de positivité. Si nous regardons le quatrième bloc situé à droite du graphe, nous pourrions interpréter ainsi le fait qu'une relation plutôt positive conduit les étudiants à recourir aux invariants opératoires issus de l'estimation statistique. Ceci renforce notre interprétation de la relation à la statistique comme variable d'influence du développement de la littéracie statistique. Le graphe implicatif (Figure 20) a été construit en centrant les relations de quasi-implication sur la variable dS3\_2 « justification en lien avec la conception de l'estimation statistique » qui cherche à vérifier dans les réponses des étudiants des arguments qui renvoient à la conceptualisation de l'estimation statistique.



Source: L'auteur

**Figure 20 : graphe implicatif sur la variable dS3\_2**

De cette manière, ce graphe nous permet de voir que les étudiants qui ont identifié la moyenne, l'écart-type, l'échantillon et ont intégré dans leurs réponses, des éléments d'estimation statistique parviennent à justifier les erreurs possibles dans des simulations réalisées dans les situations-problèmes, évoquant une conceptualisation de l'estimation statistique :

« Pour vérifier la quantité de pommes, j'ai pris un échantillon, fait la moyenne, l'écart-type et un intervalle de confiance pour la moyenne. *(Para verificar a quantidade de maçãs nos pacotes, ocupei uma amostra, fiz a média, o desvio padrão e um intervalo de confiança para a média.)* »

« J'ai réalisé une estimation pour la moyenne en calculant l'erreur échantillonnale pour les automobiles et pour les paquets de pommes ( *Realizei uma estimativa para a média, calculando o erro amostral para os carros e para os pacotes de maçãs* ) »

« L'estimation ponctuelle peut ne pas être fiable, alors j'ai trouvé l'erreur échantillonnale avec un niveau de confiance pour établir un intervalle de confiance ( *A estimativa pontual pode não ser confiável, então encontrei o erro amostral com um nível de confiança para estabelecer um intervalo de confiança* ) »

Toutefois, l'utilisation du logiciel CHIC ne nous fournit qu'une vision globale sur l'influence des attitudes des étudiants à l'égard de la statistique en relation aux réponses apportées par le questionnaire Q2. Avec l'aide de la théorie des champs conceptuels, nous avons pu mieux préciser le fait que la mobilisation des invariants opératoires plus adéquats pour la résolution des situations-problèmes se manifestait principalement chez les étudiants qui avaient manifesté d'un indice de positivité plutôt élevé.

### **3. Considérations finales sur les hypothèses soulevées dans la recherche**

À la suite des analyses réalisées, nous avons commencé à percevoir qu'elles nous aidèrent dans la démarche de mise à l'épreuve de nos hypothèses de recherche. Nous les examinons tour à tour.

H1 – Il existe des limitations dans le processus de formation du raisonnement statistique, dans l'estimation statistique, liées aux procédures et aux concepts de base de la Statistique descriptive ;

Nous considérons cette première hypothèse comme validée. En effet, les données construites par l'enquête par questionnaires Q1 et Q2, auprès du groupe d'étudiants impliqués par notre recherche mettent en évidence des arguments qui soutiennent la validation. Ainsi les étudiants qui ont exprimé être plutôt positivement (modalité : souvent ou plus) à la question V03 « Vous êtes-vous déjà trouvé(e) confronté(e) à des situations qui demandent des connaissances en statistique ? » ont une nette tendance à reconnaître les éléments de nature statistique dans les trois situations auxquelles ils ont été confrontés. Toutefois une majorité indique des difficultés à utiliser les connaissances statistiques dans la résolution des situations-problème rencontrées au travers des réponses à la question V04 « Quand des connaissances en statistique sont requises, cela vous pose-t-il des problèmes ? ». Lors de la confrontation aux situations-problèmes S1, S2 et S2 qui constitue le questionnaire Q2, dans son application AVANT le cours sur l'estimation statistique, nous avons pu observer que seul le groupe d'étudiants de la catégorie a\_Positivité\_4, c'est-à-dire ayant un rapport à la statistique nettement positif a manifesté une propension à fournir des solutions valides en mobilisant des invariants opératoires adéquats. Ces solutions englobaient des concepts basiques de la statistique descriptive

comme celui d'échantillon, de moyenne et d'écart-type, au-delà des processus opératoires. De cette manière, il ressort que, sans une reprise des concepts et procédures opératoires, les étudiants rencontreraient des obstacles cognitifs pour l'avancement du raisonnement statistique avec le nouveau contenu de l'estimation statistique. Dans ce sens, malgré une amélioration importante observée lors de la seconde phase de l'application du questionnaire Q2 APRES, de nombreux étudiants portent les stigmates découlant d'une base statistique affaiblie, issue peut-être de l'éducation de base.

H2 – La relation affective de l'étudiant à la Statistique a une influence sur la formation du raisonnement statistique.

Nous tenons cette seconde hypothèse comme aussi validée. En effet, les données construites par l'enquête par questionnaires Q1 et Q2, auprès du groupe d'étudiants impliqués par notre recherche mettent en évidence des arguments qui soutiennent la validation. Ainsi en considérant les variables V07 « intérêt pour une formation en statistique » et V08 « utilité d'une formation en statistique », nous pouvons identifier une tendance positive à attribuer un bon niveau d'intérêt personnel en relation aux connaissances statistiques. Ils semblent prendre en compte le fait que cette formation peut être importante pour leur carrière professionnelle. Cependant l'information apportée par la variable V05 qui indique le semestre choisi pour suivre les cours de statistique montre la tendance à reporter « à plus tard » cette formation. Considérons maintenant les informations apportées par le vecteur- variable V15 « sentiment à l'égard de la statistique » dont nous rappelons ici son format :

[V15] A ce jour votre sentiment envers la statistique ?

**Consignes :** Chacune des 24 phrases suivantes exprime un sentiment envers la statistique. Il vous est demandé ici de comparer votre sentiment personnel avec celui exprimé en cochant l'une des 4 modalités selon votre degré d'adhésion.

1	2	3	4
Je ne suis pas du tout d'accord	Je ne suis plutôt pas d'accord	Je suis plutôt d'accord	Je suis tout à fait d'accord

Comme nous l'avons présenté succinctement dans cet écrit, nous avons construit une variable prenant ses valeurs dans l'intervalle [0 ;1] que nous avons nommée indice de positivité du sentiment envers la statistique dont les réalisations ont été produites dans les deux phases AVANT et APRES. A partir de cette variable, nous avons construit quatre catégories ordonnées par le niveau de positivité croissant que nous avons codée ainsi :

AVANT	a_Positivité_1	a_Positivité_2	a_Positivité_3	a_Positivité_4
APRES	d_Positivité_1	d_Positivité_2	d_Positivité_3	d_Positivité_4

Nous avons inféré que la nature du rapport affectif à la statistique pouvait influencer l'attitude à l'égard de la statistique, de ses usages et de son apprentissage. Nous avons pu alors que les étudiants des catégories de niveaux 1 et 2 sont ceux qui ont exprimé un intérêt plutôt moindre pour la formation à la statistique ou se sont montrés avec beaucoup de difficultés pour résoudre les situations-problèmes S1, S2 et S3 du questionnaire Q2. Notons alors que la reprise des concepts et procédures de base de la statistique descriptive semble avoir contribué à une amélioration en relation aux attitudes et même que les étudiants qui ont d'autant plus accru leur propension à se rapprocher de la statistique, ont eu aussi tendance à positiver leur attitude. Après avoir suivi le cours sur l'estimation statistique, malgré une amélioration globale, ce sont toutefois les étudiants des catégories de positivité les plus élevées, qui firent usage du nouveau contenu dans la résolution des situations-problèmes S1, S2 et S3, progressant dans le développement du raisonnement statistique.

H3 – La relation affective de l'étudiant aux Mathématiques a une influence sur la formation du raisonnement statistique.

Nous tenons cette troisième hypothèse comme aussi validée. À partir des informations issues de la variable textuelle V10 « A ce jour que pensez-vous des sondages d'opinion ? » qui touche à la signification des sondages d'opinion et de la variable textuelle V11 « Dans la vie sociale actuelle, quelle utilité attribuez-vous à la statistique ? » qui mobilise les représentations sociales ou individuelles relatives à la statistique, nous observons que les étudiants formulent leurs réponses en recourant à des termes du langage de la statistique et en reconnaissant une utilité sociale certaine de la statistique. Toutefois, il ne suffit pas de comprendre et d'appliquer des concepts de la statistique pour développer des attitudes positives envers la statistique, il faut aussi considérer les processus opératoires. Les mathématiques restent une référence dominante dans les réponses fournies à la question portant sur les évocations de la statistique, vecteur-variable V06 « Aujourd'hui qu'évoque pour vous la statistique ? » ou encore sur les difficultés rencontrées dans l'apprentissage ou dans l'usage de la statistique, variable V12 « Selon vous, quelles sont les trois **difficultés majeures** que vous estimez devoir affronter dans le cours de statistique ». Les étudiants qui comprennent la statistique comme un élément des mathématiques et, si leur vécu mathématique est quelque chose qui provoque des frustrations, ils vont se laisser aller à démontrer ces désenchantements par rapport à la statistique, entraînant des attitudes négatives. À l'exception de ceux qui ont manifesté,

avant et après le cours sur l'estimation statistique un niveau élevé de positivité , l'investissement de la part des étudiants dans la résolution des situations-problème S1, S2 et S3 fut plutôt timide. Nous constatons ainsi une forte tendance à ce que les attitudes des étudiants vis-à-vis de la statistique puissent se manifester par des attitudes vis-à-vis des mathématiques et influencer le développement du raisonnement statistique.

#### **4. Prolongements de la recherche**

Notre objectif général était de parvenir à mieux comprendre le processus de formation du raisonnement statistique à travers l'apprentissage de l'estimation scientifique au niveau de l'Enseignement supérieur. Comme nous l'avons présenté, nous avons construit les données au moyen d'une enquête par questionnaires auprès d'un échantillon de convenance d'étudiants de *graduation* ( 1<sup>er</sup> cycle universitaire de 4 années) et d'un dispositif pédagogique centré sur l'estimation statistique. Nous avons traité les données avec des outils statistiques dans le cadre de l'Analyse statistique implicite – ASI avec le logiciel CHIC 7.0 et dans le cadre de la statistique descriptive et inférentielle avec le logiciel SPAD 9.1. Nous avons identifié un champ conceptuel pour l'estimation statistique et recouru à la théorie des champs conceptuels pour analyser les données construites à partir des réponses aux questionnaires.

À partir de là, suivant une démarche habituelle dans le contexte d'une recherche à caractère scientifique, nous avons mis à l'épreuve nos hypothèses et mis aussi en lumière quelques limitations dans la formation du raisonnement statistique dans le cadre de l'estimation statistique.

Dans cette recherche, nous avons construit une grande quantité de données dont une partie seulement a été traitée et exploitée ici. De cette façon, nous soulignons quelques limitations auxquelles nous prétendons remédier. Nous explicitons les ébauches de projets de quelques travaux que nous voulons poursuivre et qui feront partie de la continuité de notre vie scientifique académique :

1 – Notre recherche a traité des attitudes des étudiants en relation à la statistique, cependant, elle s'est limitée à une exploration ponctuelle, sans rechercher les appuis dans la littérature qui puissent mieux fonder ces occurrences. Ainsi souhaitons-nous poursuivre l'exploration, dans le champ de l'affectivité, des composantes pédagogiques et anthropologiques des attitudes dans l'enseignement de la statistique

2 – Dans le cadre théorique de l’A.S.I., nous avons limité le traitement aux données d’un questionnaire. Nous souhaitons alors explorer les autres variables tout en recherchant de nouvelles approches qui pourraient contribuer au développement de l’A.S.I.

3 - Le logiciel SPAD 9.1 est un puissant outil statistique informatique, tout à fait approprié pour mettre en œuvre une série de méthodes dans l’analyse des données. Ici notre utilisation a été modeste, ce qui nous incite à effectuer une exploitation plus avancée de ses potentialités. Cette perspective place ce logiciel dans le cadre d’un dispositif pédagogique et d’une ingénierie pédagogique pour l’enseignement et l’apprentissage de la statistique, que nous projetons de mettre sur pied

4 – Nous avons limité notre investigation au domaine de l’estimation statistique. Nous souhaitons nous investir dans d’autres contenus statistiques. Nous continuerons à utiliser nos questionnaires pour construire des données sur d’autres échantillons d’étudiants.

Nous sommes d’ailleurs qu’une recherche scientifique conduit dans le cadre d’une formation ne peut en aucune façon épuiser l’objet d’étude. Notre intérêt est bien de promouvoir de nouvelles discussions sur la formation statistique et contribuer à son amélioration au niveau de l’Enseignement universitaire.



## Conclusion générale

Le raisonnement statistique associé à la pensée statistique tient une part de responsabilité dans l'extension de la littéracie statistique. Compte tenu des spécificités du développement du raisonnement statistique et pour avancer dans la compréhension d'une science, l'étudiant doit posséder des outils basiques de la statistique issues de sa formation scolaire. L'évolution des compétences statistiques dépend d'un ensemble de processus mentaux qui articulent les outils statistiques pour exécuter des calculs et interpréter les résultats. Dans ce sens, les mathématiques et le raisonnement mathématique visent à fournir des attributs opérationnels et des déductions logiques, alors que la statistique et le raisonnement statistique portent sur les interprétations subjectives liées, souvent, à un concept mathématique qui est dépassé par des analyses et des relations abstraites du langage statistique. De cette façon, la relation frontalière entre les mathématiques et la statistique empêche les étudiants de les dissocier, ce qui rend la statistique dépendante de la qualité de la relation entre les étudiants et les mathématiques. Par conséquent, si l'on considère les analyses réalisées, nombre d'attitudes négatives vis-à-vis de la statistique ont été promues en fonction des attitudes à l'égard des mathématiques, introduisant des perturbations dans le développement du raisonnement statistique.

Dans cette perspective, nous entendons que pour promouvoir le développement du raisonnement statistique, les étudiants doivent être munis de quelques références comme nous le montrons dans la Figure 21.



Source: L'auteur

Figure 21 : Références pour le développement du raisonnement statistique

Les données sont des nombres qui sont contextualisés pour situation déterminée et servent à orienter les interprétations des résultats. En face de cela, on s'assure de la meilleure procédure statistique qui doit être employée pour la manipulation des données. Les opérations mathématiques sont insérées dans l'effectuation des procédures. Un bon renforcement de ces opérations est propice à une avancée cognitive par rapport à la statistique, en diminuant les attitudes négatives vis-à-vis des mathématiques, même dans

le contexte des procédures statistiques spécifiques, ce qui permet de rapprocher l'étudiant de la statistique. En ce qui concerne l'estimation des paramètres statistiques, le renforcement a constitué une aide dans le domaine des concepts et des procédures opérationnelles liées à la moyenne et à l'écart type, par exemple, qui ont été fondamentales pour observer une avancée pour les étudiants de toutes les catégories de positivité construites dans la recherche.

Les résultats obtenus dans la résolution d'une situation-problème statistique ne doivent pas être confondus avec une solution mathématique. En général, les problèmes de statistique présentent un questionnement dont le résultat doit être articulé à la manipulation intervenue sur les données et par l'interprétation particulière du contexte de la situation-problème. Cette subjectivité du résultat n'est pas facile à être conceptualisée chez les étudiants, mais la prise de conscience de ce que les réponses ne sont pas absolues, promeut le concept fondateur de la statistique qu'est la variabilité.

Bien que de nombreux chercheurs comme Hand (1998), Snee (1988), Stuart (1995) et d'autres, plaident en faveur de la simplification des mathématiques dans la formation des utilisateurs de la statistique, en offrant des ressources informatiques pour l'amplification des calculs, ce que nous utilisons également dans notre pratique enseignante, nous ne pouvons pas ne pas reconnaître la présence de ces obstacles opérationnels découlant des mathématiques, générant ainsi des utilisateurs de statistique inaccomplis. Nous devons donc encourager les initiatives et les dispositifs pédagogiques visant à l'atténuation de ces difficultés liées aux mathématiques, en donnant accès aux conditions favorables au développement du raisonnement statistique dans le cadre de l'enseignement de la statistique à tous les niveaux de la scolarité.

## Références

- Acioly-Régner, N. M., & Régner, J-C. (2005). Repérage d'obstacles didactiques et socioculturels au travers de l'A.S.I. des données issues d'un questionnaire. In R. Gras, F. Spagnolo, & J. David (Coords.). *Proceedings Third International Conference A.S.I* (pp. 63-87). Palerme: Implicative Statistic Analysis.
- Barbetta, P. A., Reis, M. M., & Bornia, A. C. (2010). *Estatística para cursos de engenharia e informática* (3a ed.). São Paulo: Atlas.
- Barceló, S., Capilla, C., & Zúnica, L. (2000). Estatística e probabilidades no ensino superior: diagnóstico de competências e propostas curriculares globais. In C. Loureiro, F. Oliveira, & L. Brunheira (Eds.). *Ensino e aprendizagem da estatística* (pp. 83-92). Lisboa: SPE e APM.
- Batanero, C. (2001). *Didáctica de la estadística*. Granada: Grupo de investigación em Educación Estadística do Departamento de Didáctica de la Matemática da Universidad de Granada.
- Campos, C. R. (2007). *A educação estatística: uma investigação acerca dos aspectos relevantes à didática da estatística em cursos de graduação*. Tese de doutorado, Universidade Estadual Paulista, Rio Claro, SP, Brasil.
- Carzola, I. M. (2002). *A relação entre a habilidade viso-pictórica e o domínio de conceitos estatísticos na leitura de gráficos*. Tese de doutorado, Universidade Estadual de Campinas, Campinas, SP, Brasil.
- Chance, B. L. (2002). Components of statistical thinking and implications for instruction and assessment. *Journal of Statistics Education*, 10(3), 1-14.
- Couturier, R. (2001). Traitement de l'analyse statistique dans CHIC. In R. Couturier (Ed.). *Actes des journées: la fouille dans les données par la méthode d'analyse statistique implicative. Applications et traitement par CHIC* (pp. 33-50). Caen: IUFM.
- Couturier, R., Bodin, A., Gras, R. (2003), *A Classificação Hierárquica Implicativa e Coesitiva*. [http://math.unipa.it/~grim/asi/asi\\_03\\_gras\\_bodin\\_cout.pdf](http://math.unipa.it/~grim/asi/asi_03_gras_bodin_cout.pdf). Acesso em 15 de março de 2019.
- Delmas, B. C. (2001). What makes the standard deviation larger or smaller?. In B. C. Delmas (Ed.). *Statistics teaching and resource library*. Minneapolis: University of Minnesota.
- Diaz, D. (2016). *Les facteurs influençant la réussite des activités collaboratives médiées par les TICE dans une situation de formation universitaire à la statistique*. Tese de doutorado, Université Lumière - Lyon II, Lyon, França.
- Freund, J. E. (2006). *Estatística aplicada: economia, administração e contabilidade* (11a ed.). Porto Alegre: Bookman.
- Gal, I. (2002). Adult's statistical literacy: meanings, components, responsibilities. *International Statistical Review*, 70(1), 1-25.

- Gal, I., & Garfield, J. (1997). *The assessment challenge in statistic education*. Amsterdã: IOS Press.
- Gal, I., & Ginsburg, L. (1994). role of beliefs and attitudes in learning statistic: towards an assessment framework. *Journal of Statistics Education*, 2, 64-89.
- Gonçalez, N. (2002). *Atitudes dos alunos do curso de pedagogia com relação à disciplina de estatística no laboratório de informática*. Tese de doutorado, Universidade Estadual de Campinas, Campinas, SP, Brasil.
- Gonsalves, E. P. (2003). *Iniciação à pesquisa científica*. Campinas: Alínea.
- Grácio, M. C. C., & Oliveira, E. F. T. (2005). O ensino de estatística na UNESP/Campus de Marília. *Educação Matemática em Revista*, 11(17), 9-15.
- Gras, R. (1979). *Contribution à l'étude expérimentale et à l'analyse de certaines acquisitions cognitives et de certains objectifs didactiques en mathématiques*, Thèse d'Etat, Université de Rennes 1.
- Gras, R. (2004). L'analyse implicative: ses bases, ses développements. *Educação Matemática Pesquisa*, 4(2), 11-48.
- Gras, R., & Almouloud, S. A. (2005). A implicação estatística usada como ferramenta em um exemplo de análise de dados multidimensionais. *Educação Matemática Pesquisa*, 4(2), 75-88.
- Gras, R., Régnier, J.-C., Guillet, F. (2009), *Analyse Statistique Implicative: Une méthode d'analyse de données pour la recherche de causalités*. Toulouse : Éditions Cépaduès.
- Gras, R., Régnier, J.-C., Lahanier-Reuter, D., Marinica, C., Guillet, F. (2017), *Analyse Statistique Implicative. Des Sciences dures aux sciences humaines et sociales*. Toulouse : Éditions Cépaduès.]
- Gras, R., Régnier, J.-C., Marinica, C., Guillet, F. (2013), *Analyse Statistique Implicative. Méthode exploratoire et confirmatoire à la recherche de causalités*. Toulouse : Éditions Cépaduès.
- Hand, D. J. (1998). Breaking misconceptions - statistics and its relationship to mathematics. *The Statistician*, 47(2), 245-250.
- Ignácio, S. A. (2010), Importância da Estatística para o processo de conhecimento e tomada de decisão, *Nota Técnica IPARDES*, n° 6, Curitiba.
- Lerman, I. C. (1981). *Classification et analyse ordinale des données*. Paris: Dunod.
- Mantovani, D. M. N., & Viana, A. B. N. (2004). Ensino de estatística para cursos de graduação em administração de empresas – novas perspectivas. *Anais do Seminário de Administração*, São Paulo, SP, Brasil, 7.
- Oriol, J.-C. (2007). *Formation à la statistique par la pratique d'enquêtes par questionnaires et la simulation: étude didactique d'une expérience d'enseignement dans un département d'IUT*. Tese de doutorado, Université Lumière-Lyon II, Lyon, França.

- Régnier, J.-C. (2003). À propos de la formation en statistique. Approches praxéologiques et épistémologiques de questions du champ de la didactique de la statistique. *Revue du Centre de Recherche en Education*, 22, 157-201.
- Régnier, J.-C. (2007). *Vocabulaire de statistique*. Lyon: ISPEF - Université Lumière Lyon 2.
- Régnier, J.-C., Almouloud, S. A., & Gras, R. (2013). Analyse statistique implicative: cadre théorique et applicatif pour l'exploration sémantique et non symétrique des données. *Anais do Colloque International D'analyse Statistique Implicative*, São Paulo, SP, Brasil, 7.
- Régnier, J.-C., Gras, R., & Bailleul, M. (2012). L'analyse statistique implicative: de l'exploratoire au confirmatoire. *Anais do Colloque International D'analyse Statistique Implicative*, Caen, França, 6.
- Rumsey, D. J. (2002). Statistical literacy as goal for introductory statistics courses. *Journal of Statistics Education*, 10(3).
- Sgoutas-Emch, S. A., & Johnson, C. J. (1998). Is journal writing an effective method of reducing anxiety towards statistics?. *Journal of Instructional Psychology*, 25(1), 49-57.
- Silva, C. B. (2007). *Pensamento estatístico e raciocínio sobre variação: um estudo com professores de matemática*. Tese de doutorado, Pontifícia Universidade Católica de São Paulo, São Paulo, SP, Brasil.
- Snee, R. D. (1988). Mathematics is only one tool that statistician use. *The College Mathematics Journal*, 19, 30-32.
- Stuart, M. (1995). Changing the teaching of statistics. *The Statistician*, 44(1), 45-54.
- Tufféry, S. (2010). *Data mining et statistique décisionnelle: l'intelligence des données*. Paris: Editions Technip.
- Vergnaud, G. (1990a). Epistemology and psychology of mathematics education. In P. Necher, & Kilpatrick, J. (Eds.). *Mathematics and cognition* (pp. 14-30). Cambridge: Cambridge University Press.
- Vergnaud, G. (1990b). La théorie des champs conceptuels. *Recherches em Didactique des Mathématiques*, 10(23), 133-170.
- Wada, R. S. (1996). *Estatística e ensino: um estudo sobre representações de professores de 3º grau*. Tese de doutorado, Universidade Estadual de Campinas, Campinas, SP, Brasil.

# Annexe 1 : Questionnaire n°1

1/3



Programa de Pós-Graduação em Ensino de Ciências e Matemática



Ano letivo 2016 – Centro Universitário Franciscano

Estudo das dificuldades e facilidades de aprendizagem em estatística

Questionário apresentado aos estudantes que estão cursando a disciplina de Estatística

Data : ...../...../2016

Você pode optar por permanecer anônimo.

Nome :	Sobrenome :	Matricula :
--------	-------------	-------------

## 1 Vous et votre parcours antérieur de formation

[V1] Homme (entourez la réponse)	Femme	[V2] Date de naissance :	Codage pour anonymat :
-------------------------------------	-------	--------------------------	------------------------

[V3] Vous êtes-vous déjà trouvé(e) confronté(e) à des situations qui demandent des connaissances en statistique ? : (Entourez votre réponse)

1_Jamais	2_Rarement	3_Quelquefois	4_Souvent	5_Très souvent	6_Presque chaque jour
----------	------------	---------------	-----------	----------------	-----------------------

[V4] Quand des connaissances en statistique sont requises, cela vous pose-t-il des problèmes ? : (Entourez votre réponse)

1_Pas du tout	2_Rarement	3_Quelquefois	4_Souvent	5_Très souvent	6_Toujours
---------------	------------	---------------	-----------	----------------	------------

[V5] Semestre

## 2 Vous et la statistique

[V6] Aujourd'hui qu'évoque pour vous la statistique ?

(Donner trois mots)

--	--	--

(Compléter par une phrase)

--

[V7] Quel niveau d'intérêt personnel attribuez-vous à une formation en statistique en Graduation ?

0	1	2	3	4	5	6	7	8	9	10
le plus faible										le plus fort

[V8] Quel niveau d'utilité attribuez-vous à une formation en statistique en Graduation?

0	1	2	3	4	5	6	7	8	9	10
le plus faible										le plus fort

[V9] A ce jour que signifient pour vous les termes suivants ?

Représentatif	
Significatif	
Estimation	
Test d'hypothèses	
Probabilité	
Sondage	

[V10] A ce jour que pensez-vous des sondages d'opinion ?

--

[V11] Dans la vie sociale actuelle, quelle utilité attribuez-vous à la statistique ?

--

[V12] Selon vous, quelles sont les trois **difficultés majeures** que vous estimez devoir affronter dans le cours de statistique

Diff_1 :	Diff_2 :	Diff_3 :
----------	----------	----------

[V13] Selon vous, quels devraient être les trois **apports majeurs** du cours de statistique ?

App_1 :	App_2 :	App_3 :
---------	---------	---------

[V14] Avez-vous déjà suivi un enseignement de statistique ? : (*Entourer votre réponse*)

1_Jamais	2_ à niveau très superficiel	3_ à un niveau de quelques rudiments	4_initiation qui m'a donné des bases	5_initiation à un niveau avancé	6_J'ai une formation de statisticien(ne) ou presque
----------	------------------------------	--------------------------------------	--------------------------------------	---------------------------------	---

2.1 [V15] A ce jour votre sentiment envers la statistique ?

**Consignes :** Chacune des phrases suivantes exprime un sentiment envers la statistique. Il vous est demandé ici de comparer votre sentiment personnel avec celui exprimé en cochant l'une des 4 modalités selon votre degré d'adhésion.

1	2	3	4
Je ne suis pas du tout d'accord	Je ne suis plutôt pas d'accord	Je suis plutôt d'accord	Je suis tout à fait d'accord

1. Je n'ai pas un bon niveau en statistique	1	2	3	4
2. La statistique est quelque chose que j'apprécie énormément	1	2	3	4
3. Je me sens insécurisé(e) quand je me force à faire de la statistique	1	2	3	4
4. J'ai une relation définitivement positive avec la statistique : j'aime et j'apprécie cette matière	1	2	3	4
5. Quand j'étudie la statistique, je n'arrive plus à penser clairement car tout ce bloque dans ma tête	1	2	3	4
6. Quand j'entends le mot statistique, je ressens un sentiment d'aversion	1	2	3	4
7. Je n'aime vraiment pas les mathématiques	1	2	3	4
8. Penser à l'obligation de résoudre un problème de statistique me rend nerveux	1	2	3	4
9. La statistique est fascinante et divertissante (amusante)	1	2	3	4
10. Je n'ai jamais aimé la statistique et c'est la matière qui me fait le plus peur	1	2	3	4
11. Le sentiment que j'éprouve envers la statistique est bon	1	2	3	4
12. J'ai peur de la statistique car j'ai peur des mathématiques	1	2	3	4
13. Je trouve la statistique très intéressante et j'aime aller en cours de statistique	1	2	3	4
14. La statistique me donne l'assurance et me stimule	1	2	3	4
15. La statistique me donne l'impression que je suis perdu(e) dans une pièce pleine de nombres dont je ne trouverais pas la sortie	1	2	3	4
16. Je me sens bien en statistique et j'aime beaucoup cette discipline	1	2	3	4
17. Je suis toujours stressé(e) pendant les cours de statistique	1	2	3	4
18. La statistique est une des matières que j'aime vraiment étudier à l'université ou à l'école	1	2	3	4
19. Je n'aime pas la statistique et ni aller en cours de statistique	1	2	3	4
20. La statistique me provoque un sentiment d'indécision qui résulte de ma peur de ne pas être capable de réussir	1	2	3	4
21. Je me sens plus heureux (se) en statistique qu'avec n'importe quelle autre discipline	1	2	3	4
22. La statistique me rend inquiet(e), mécontent(e), irrité(e) et impatient(e)	1	2	3	4
23. Je suis compétent(e) en mathématique	1	2	3	4
24. J'aime vraiment la statistique.	1	2	3	4

2.2 [V16] A ce jour quelle est votre définition de la statistique ?

--



