



HAL
open science

Études des phénomènes de transposition didactique de la statistique dans le champ universitaire et ses environnements : une contribution à la pédagogie universitaire

Marthe-Aline Jutand

► **To cite this version:**

Marthe-Aline Jutand. Études des phénomènes de transposition didactique de la statistique dans le champ universitaire et ses environnements : une contribution à la pédagogie universitaire. Education. Université de Bordeaux, 2015. Français. NNT : 2015BORD0366 . tel-01270556

HAL Id: tel-01270556

<https://theses.hal.science/tel-01270556>

Submitted on 8 Feb 2016

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

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

THÈSE PRÉSENTÉE
POUR OBTENIR LE GRADE DE

**DOCTEUR DE
L'UNIVERSITÉ DE BORDEAUX**

ÉCOLE DOCTORALE - SOCIÉTÉS, POLITIQUE, SANTÉ PUBLIQUE
SPÉCIALITÉ - SCIENCES DE L'ÉDUCATION

Par Marthe-Aline CHARTIER épouse JUTAND

**ÉTUDES DES PHÉNOMÈNES
DE TRANSPOSITION DIDACTIQUE DE LA STATISTIQUE
DANS LE CHAMP UNIVERSITAIRE ET SES ENVIRONNEMENTS**

Une contribution à la pédagogie universitaire

Sous la direction de Monsieur Bernard SARRAZY, Professeur

Soutenu le 2 décembre 2015

Membres du jury :

M. MARCHIVE, Alain,
Professeur d'Université Emérite, Université de Bordeaux, **président**

Mme GEGOUT, Anne,
Professeur d'Université, Université de Lorraine, **rapporteur**

M. REGNIER, Jean-Claude,
Professeur d'Université, Université Lyon 2, **rapporteur**

M PERIER, Pierre,
Professeur d'Université, Université Rennes 2, **examineur**

M. SARRAZY, Bernard,
Professeur d'Université, Université de Bordeaux, **directeur**

REMERCIEMENTS

Ce travail est un point d'étape de mon parcours à l'université aux multiples rencontres, je tiens à exprimer ma reconnaissance à toutes celles et tous ceux qui m'ont accompagnée et aidée à avancer sur ce chemin.

Tout d'abord je tiens à remercier tout particulièrement Bernard Sarrazy qui a su m'accorder sa confiance en acceptant de m'encadrer durant ces trois années. J'ai appris énormément auprès de toi, tu as su m'apprendre à regarder différemment les choses, tu m'as fait découvrir un mode de pensée bien différent de celui dont j'avais l'habitude. Sincèrement il y a eu un avant cette période de thèse et il y aura un après. Je ne souhaite qu'une chose c'est que nous puissions continuer à travailler ensemble, car tes qualités humaines, ta franchise et ta sensibilité sont un vrai moteur pour celles et ceux qui marchent dans tes traces. Merci.

Je remercie sincèrement Anne Gegout et Jean-Claude Régnier d'avoir accepté d'être rapporteurs de mon travail. J'ai eu l'immense bonheur de croiser vos chemins et d'apprécier ces temps d'échanges. Les regards croisés que vous portez l'un et l'autre sur la statistique et sa didactique trouveront, je l'espère, des éléments d'intérêt dans mon travail et je suis dans l'attente de vos retours pour continuer à avancer sur ce chemin.

Je tiens à remercier Alain Marchive et Pierre Périer d'avoir accepté d'être membres du jury, j'espère que cette thèse saura capter votre attention, et je sais que l'ensemble de vos commentaires me seront fort utiles pour poursuivre ma réflexion.

Un merci tout particulier à Marianne Savès pour ta relecture précise de ce travail mais aussi et surtout pour ton amitié sans faille. Je n'ai qu'un vœu : que nous puissions encore longtemps travailler ensemble.

Je désire remercier mes amis et collègues tout particulièrement de l'ISPED avec qui nous avons construit tant de choses depuis plus de 20 ans. Il est vraiment difficile de dresser une liste tant chacun de vous a participé de près ou de loin à la construction de ma réflexion et à mon épanouissement professionnel. Merci à Roger Salamon de m'avoir permis d'intégrer cette famille.

Mais il y a aussi, et surtout, les étudiants qui ont été mon moteur, et à qui je dis un immense merci. Ce sont vos difficultés, vos questionnements, vos attentes et vos envies qui m'ont toujours donné envie de comprendre ce qui se jouait pendant l'enseignement et comment nous pouvions construire de nouvelles réflexions pédagogiques pour vous permettre de devenir de bons professionnels responsables.

Merci à Valérie et Hélène pour la très belle aventure humaine que nous poursuivons ensemble depuis 4 ans pour faire reconnaître l'importance de la recherche en sciences infirmières et paramédicale. C'est toujours un plaisir de travailler avec vous et j'espère que cela continuera. Merci Christine d'avoir rejoint notre trio dans cette formation, car tu sais combien j'apprécie de travailler avec toi.

Un grand merci aussi à l'ensemble des membres du groupe enseignement de la statistique de la SFdS, car les échanges passionnants et passionnés que nous avons pu avoir ont nourri ma réflexion, j'ai appris beaucoup grâce à vous pendant mes 5 années de présidence du groupe. Et merci aussi à Jean-Michel, Anne, Nathalie, Servane, et tous les autres sans qui la SFdS ne serait pas une aussi belle société savante.

Un immense merci à ma famille, à mes parents, Nicole et Camille, à Etienne et Fabienne, pour tout l'amour dont vous m'avez toujours entourée et pour votre présence, votre soutien inconditionnel et votre regard attentif malgré les épreuves. Merci Janine et Claude pour votre affection et bienveillance.

Merci à mes deux princesses, Alix et Agathe, pour votre tendresse et tout simplement votre présence si sacrée pour moi, je vous adore.

Merci Christophe pour ton soutien constant, tes encouragements et ton amour. Sans toi tout cela n'aurait pas pu être réalisé.

RESUME - ABSTRACT

Titre : Études des phénomènes de transposition didactique de la statistique dans le champ universitaire et ses environnements : une contribution à la pédagogie universitaire

Résumé

La spécificité de la statistique, en tant que science d'analyse et de décision, suscite l'intérêt de l'ensemble des domaines scientifiques et sociétaux. Elle en devient ainsi une discipline protéiforme s'appuyant sur des dimensions partagées, mais aussi sur des formulations spécifiques. Dans les deux situations étudiées, doctorat et expertise, le chercheur doit développer des capacités à expliciter les étapes d'une démarche statistique et à communiquer de manière compréhensible à des fins de diffusion et d'appropriation. L'analyse des thèses met en lumière des régularités d'usage de concepts et de mots, mais aussi, de nombreuses spécificités disciplinaires. L'analyse de la situation d'expertise, quant à elle, souligne le travail didactique mis en œuvre par le chercheur afin de rendre possible le transfert de connaissance à destination des politiques décisionnaires.

Les différents éléments et analyses ayant constitué cette thèse ont pour objet de participer à la réflexion sur l'enseignement de la statistique à l'université et sur les stratégies pédagogiques à développer. Notre analyse met en évidence l'importance de former les étudiants à la diffusion et la traduction de résultats d'études statistiques. Tout ceci conduit à la question de la transformation des enseignements de la statistique à l'université afin de développer les temps et des méthodes d'apprentissage de diffusion, à visée d'explicitation, des savoirs statistiques et des résultats auprès de publics variés.

Mots clés : statistique, enseignement, transposition didactique, pédagogie universitaire

Title: The phenomena of didactic transposition of statistics in the academic fields: a contribution to the University Teaching

Abstract:

Statistics, with its particularity being a science based on analysis and decision, generates interests from all scientific and social domains. Thus, statistics is a many-sided discipline based on shared dimensions, but also on specific formulations.

Researchers must develop specific skills to explain each step of a statistical approach in both situations studied: doctoral theses and expertise. Additionally, results need to be communicated clearly and precisely by a researcher in order to assure their comprehension and the appropriation of knowledge. The analysis of doctoral theses highlights a regularity in the use of statistical concepts and statistical language, but also numerous specificities for academic disciplines. Moreover, the analysis of expertise highlights that researchers set up didactic approaches to improve the transfer of knowledge to decision-making politics.

This thesis aims to take part in the reflection about i) the teaching of statistics at the university and ii) the pedagogical strategies that needs to be developed by highlighting the teaching strategies proven to be useful for statistical education. Our analysis clearly shows the importance of qualifying students in the dissemination and communication of results from statistical analyses. Taken together, there is a clear need to develop new teaching strategies for statistics at the university aiming at developing teaching methods facilitating explicit communication of statistical expertise and results addressing a wide range of audiences.

Keywords: statistics, teaching, didactic transposition, pedagogy, higher education

Unité de recherche

[Laboratoire LACES – Laboratoire Cultures Education Sociétés
EA-4140, Université de Bordeaux
3 ter, place de la Victoire, 33076 Bordeaux cedex]

SOMMAIRE

REMERCIEMENTS	2
RESUME - ABSTRACT	4
SOMMAIRE	7
LISTE DES ACRONYMES	9
AVANT-PROPOS	11
INTRODUCTION	16
CHAPITRE 1 – SITUATIONS DIDACTIQUES DE LA STATISTIQUE POUR L’ENSEIGNANT- CHERCHEUR	20
A. MISSIONS DE L’ENSEIGNANT-CHERCHEUR : LA PLACE DE LA DIDACTIQUE.....	21
B. QUELLES SPECIFICITES DE LA TRANSPOSITION DIDACTIQUE POUR LES MISSIONS DE L’ENSEIGNANT- CHERCHEUR ?	30
C. POURQUOI PARLER D’ÉCOLOGIE DIDACTIQUE DES SAVOIRS POUR LA SCIENCE STATISTIQUE ? ...	37
D. SYNTHESE.....	55
CHAPITRE 2 – LA STATISTIQUE : UNE DISCIPLINE BIEN SPECIFIQUE	56
A. LA RECONNAISSANCE DISCIPLINAIRE DE LA STATISTIQUE	56
B. LE CADRE DE LA STATISTIQUE	63
C. LES NOTIONS FONDATRICES.....	72
D. LA PLACE DE LA STATISTIQUE DANS LA COMPREHENSION DU MONDE !	88
E. SYNTHESE.....	92
CHAPITRE 3 – DIFFERENTES SITUATIONS DE LA DIDACTIQUE DE LA STATISTIQUE	93
A. LA STATISTIQUE AU SERVICE DE LA RECONNAISSANCE D’UN JEUNE CHERCHEUR PAR UNE COMMUNAUTE DE CHERCHEURS	93
B. LE CHERCHEUR DANS UNE SITUATION D’EXPERT ET DE DIDACTICIEN DES POLITIQUES : CONSEQUENCES DE LA GUERRE DU GOLFE (1990 – 1991) SUR LA SANTE DES MILITAIRES FRANÇAIS.....	154
C. SYNTHESE.....	183
CHAPITRE 4 – ENJEUX DE L’ENSEIGNEMENT DE LA STATISTIQUE A L’UNIVERSITE	185
A. L’UNIVERSITE ET L’APPROCHE COMPETENCE	185
B. LES ACTEURS DE L’ENSEIGNEMENT DE LA STATISTIQUE A L’UNIVERSITE	198

C. QUELLES STRATEGIES PEDAGOGIQUES POUR ENSEIGNER LA STATISTIQUE ?.....	205
D. SYNTHESE.....	213
CONCLUSION	214
REFERENCES	217
LISTE DES THESES ETUDIEES	224
INDEX	227
TABLE DES MATIERES	228
LISTE DES TABLEAUX	232
LISTE DES FIGURES	233
LISTE DES SCHEMAS	234
LISTE DE PUBLICATIONS	235

LISTE DES ACRONYMES

ACM	Analyse des correspondances multiples
ACP	Analyse en composantes principales
AFC	Analyse factorielle des correspondances
APC	Approche par compétences
ASA	American Statistical Association
CNRS	Centre National de la Recherche Scientifique
CNU	Conseil National des Universités
DEUG	Diplôme d'études universitaires générales
INED	Institut National d'Etudes Démographiques
INRA	Institut National de Recherche Agronomique
INSEE	Institut National de la Statistique et des Etudes Economiques
INSERM	Institut National de la Santé et de la Recherche Médicale
InVS	Institut de Veille Sanitaire
ISPED	Institut de Santé publique, d'Epidémiologie et de Développement
LAP	Licence d'administration publique
LQAS	Lot Quality Assurance Sampling
MAPI	Mission d'appui à la pédagogie et à l'innovation
MASS	Mathématiques appliqués et sciences sociales
MIASH	Mathématiques et informatique appliquées aux sciences humaines et sociales
MST	Maîtrise de sciences et techniques
OMS	Organisation mondiale de la santé
PRAG	Professeur agrégé
PU-PH	Professeur des universités – Praticien hospitalier
QCM	Questionnaire à choix multiple
SFdS	Société Française de Statistique
STAPS	Sciences et techniques des activités physiques et sportives
UFR	Unité de formation et de recherche

AVANT-PROPOS

Après de nombreuses années d'enseignement et de cheminement dans le milieu de la statistique, j'ai eu le souhait de prendre le temps de regarder sous un autre angle les pratiques et les usages de cette discipline, de prendre de la distance par rapport à ma propre pratique d'enseignante et de statisticienne. Cet avant-propos est l'occasion de vous présenter les étapes qui m'ont amenée jusque-là.

Septembre 1993 fut pour moi une arrivée dans un monde que je ne connaissais pas et que j'allais explorer, et essayer de comprendre : l'université. Après avoir enseigné trois ans au niveau collège et lycée, j'avais eu l'immense bonheur d'être retenue sur un poste de PRAG mathématiques-statistique à l'UFR de santé publique de l'Université Bordeaux 2, pour enseigner principalement la statistique. Si initialement la raison de ce bonheur était surtout due à la possibilité de me rapprocher de mon conjoint, avec le recul, je me dis que beaucoup d'événements liés à l'obtention de ce poste me permettent maintenant d'avoir de nombreuses autres raisons de satisfaction.

Comment enseigner la statistique en santé publique, quand votre formation s'est déroulée uniquement au sein d'une UFR de mathématique, en maîtrise de mathématiques dites « pures », puis en année de préparation à l'agrégation option analyse numérique ? Le challenge était osé, surtout pour ceux qui avaient pris le risque de me recruter ! L'ignorance permet beaucoup de choses surtout si elle est associée à la curiosité ! Je comprends cependant le désarroi des enseignants de mathématiques du secondaire qui, sans formation, se sont vus dans l'obligation d'enseigner des notions complexes de statistique, tout particulièrement en terminale scientifique avec l'intégration du théorème de Moivre Laplace, de l'intervalle de fluctuation et de l'intervalle de confiance¹.

¹ Ayant participé à la rédaction du document ressources « Probabilités et statistique » pour la classe de terminale générale et technologique, j'évalue assez bien les difficultés rencontrées par les enseignants face à ces éléments.

http://cache.media.eduscol.education.fr/file/Mathematiques/11/5/LyceegT_ressources_Math_T_proba-stat_207115.pdf

Le travail auquel je me suis confrontée était donc multidimensionnel : acquérir les nouvelles connaissances à transmettre (savoir), dans le cadre de pratiques et dans un domaine de réalité (santé publique), comme l'indique Chevallard dans « Les processus de transposition didactique et leur théorisation » (1994).

Mes enseignements se déroulaient principalement auprès des étudiants de MST (maîtrise de sciences et techniques) de santé publique de l'UFR de santé publique. Il s'agissait de petites promotions de 25 à 30 étudiants (effectifs me permettant une acclimatation progressive au passage du lycée à l'université). La diversité des étudiants accueillis dans le cadre de cette formation est un élément qui a été, en analysant cela rétrospectivement, particulièrement favorisant pour mon intégration. En effet, nous accueillions dans cette formation des étudiants ayant obtenu des DEUG (Diplôme d'études universitaires générales) dans des disciplines aussi diverses que la biologie, l'économie, les mathématiques ... mais aussi des professionnels en reprise d'étude : des infirmiers, des kinésithérapeutes, voire des médecins français ou étrangers. Cette formation, qui avait une seule année d'existence lorsque je l'ai rejointe, était d'une grande originalité, et avait été construite de toute pièce en 1991 par les Professeurs Roger Salamon et Jacques Beylot, respectivement PU-PH de santé publique et PU-PH de médecine interne et directeur de l'UFR de Santé Publique lors de sa création en 1988. Il est important de replacer la création de cette formation dans son contexte. En effet, en France à cette époque, la santé publique était pensée uniquement comme une sous-discipline de la médecine, et il s'agissait de prouver que son développement devrait se nourrir d'une offre de formation originale basée sur la pluridisciplinarité (développement d'une équipe pédagogique pluridisciplinaire) et l'interdisciplinarité (originalité des méthodes pédagogiques favorisant des activités et projets interdisciplinaires) permettant ainsi de construire le cadre favorisant la mise en œuvre de la santé publique telle que définie par l'OMS (Organisation Mondiale de la Santé) en 1952.

La santé publique est la science et l'art de prévenir les maladies, de prolonger la vie et d'améliorer la santé et la vitalité mentale et physique des individus, par le moyen d'une action collective concertée visant à :

1. assainir le milieu,
2. lutter contre les maladies qui présentent une importance sociale,

3. enseigner les règles d'hygiène personnelle,
4. organiser des services médicaux et infirmiers en vue d'un diagnostic précoce et du traitement préventif des maladies,
5. mettre en œuvre des mesures sociales propres à assurer à chaque membre de la collectivité un niveau de vie compatible avec le maintien de la santé.

L'objet final étant de permettre à chaque individu de jouir de son droit inné à la santé et à la longévité.

Ces éléments étaient d'ailleurs spécifiés dans la plaquette de présentation de la formation de mars 1993².

Les spécificités de cette maîtrise tiennent :

- à l'originalité de l'UFR de Santé Publique de l'Université de Bordeaux 2 qui regroupe à côté de la Médecine les différents aspects de la Santé Publique, et qui associe à des universitaires des professionnels de la Santé
- aux orientations retenues. En effet, cette maîtrise est ouverte essentiellement aux non médecins et comporte une formation de base en Médecine et Biologie. L'enseignement proposé met l'accent sur des bases méthodologiques fortes, et fait largement appel aux professionnels. Elle a une orientation très interdisciplinaire
- à l'ouverture à des étudiants d'origine diverse dans des perspectives de réorientation, de formation permanente, ou de transfert de technologie.

Il n'existe aucune MST de Santé Publique.

Avoir participé pendant 10 ans à cette formation puis à la construction du master de santé publique, qui a vu le jour en 2000³, fut une expérience des plus enrichissantes, me permettant d'étudier et de comprendre l'importance de la communication et du sens des mots dans une activité nécessitant plus que la pluridisciplinarité, l'interdisciplinarité, mais aussi l'importance pour l'enseignant de prendre conscience de l'existence d'obstacles didactiques due souvent à la polysémie des mots mais pas uniquement. Ce point sera étudié tout particulièrement dans le cadre de cette thèse.

² Présentée dans actualité et dossier en santé publique du 2 mars 1993 - www.hcsp.fr/Explore.cgi/Telecharger?NomFichier=ad021313.pdf

³ Durant la construction de cette formation, j'étais directrice adjointe de l'ISPED, Institut de Santé Publique, d'Epidémiologie et de Développement

J'enseigne ou ai enseigné principalement la statistique inférentielle, les techniques d'enquête et théorie des sondages (de l'élaboration des outils à l'analyse prenant en compte les plans de sondage) ainsi que l'analyse factorielle, auprès d'étudiants de licence MASS (mathématiques appliqués et sciences sociales), de première et seconde années du master santé publique quelle que soit la spécialité, de seconde année du master de sciences de l'éducation, mais aussi d'étudiants en formation paramédicale. La diversité des étudiants rencontrés, tant dans leurs pré-requis en mathématiques que dans leurs projets professionnels, m'a convaincue de l'importance et de la faisabilité de former tout citoyen à la démarche statistique ; et que, pour pouvoir le réaliser, il était nécessaire de prendre en considération l'environnement de chacun et le cadre global de la formation dans lequel cette activité pédagogique se déroule, afin de ne pas apporter un temps de transmission de savoir statistique dénaturé. Contextualiser les enseignements de statistique est donc un élément central selon moi, pour en faciliter la compréhension, même si cela s'avère insuffisant. La seconde difficulté découle du souci de ne pas dénaturer le savoir en le transmettant : comment passer du « *savoir savant* » au « *savoir enseigné* », selon la formulation de Chevallard (1985).

Parallèlement à ces temps d'enseignement et bénéficiant d'un environnement scientifique très dynamique, j'ai eu l'opportunité de participer à différents projets de recherche en épidémiologie, dans des domaines aussi variés que la nutrition (Allès et al., 2011, 2012 ; Barberger-Gateau et al., 2005 ; Bourdel-Marchasson et al., 2009; Féart et al., 2007, 2011 ; Samieri et al., 2008) ou les problèmes de santé des militaires français ayant participé à la Guerre du Golfe 1990 - 1991 (Bensefa et al., 2006; Salamon et al, 2006; Verret et al., 2008). J'ai aussi été sollicitée plus ponctuellement dans le cadre de travaux de recherche pour une aide spécifique sur des analyses statistiques ou la construction de plans d'échantillonnage, comme par exemple pour l'étude du lien entre des expositions environnementales et le mésothéliome (Camiade et al., 2013) ou des associations entre les stratégies de coping et la survie après cancer du sein (Cousson-Gélie et al., 2007). Les échanges d'explicitations et de compréhension de la problématique et de la mise en œuvre des méthodes appropriées dans le cadre de ces travaux ont été particulièrement riches mais aussi parfois compliqués : comprendre, proposer l'analyse, la développer et

communiquer les résultats statistiques, voici les enjeux toujours remis en question lorsqu'on change d'habitat⁴. Trouver les bons mots, pour lesquels le sens est partagé n'est pas une étape facile. Le terme de *représentativité* est celui pour lequel les échanges ont toujours été les plus compliqués, car pouvant faire référence à des objets complètement différents ; cela sera discuté dans le cadre de cette thèse.

Les différentes situations évoquées ci-dessus, qu'il s'agisse d'enseignement ou de participation à une étude scientifique et à la rédaction d'un article ou d'une communication orale, peuvent être considérées comme des situations de transposition didactique se nourrissant du milieu dans lequel elles se construisent. Pourtant les outils et techniques statistiques sont parfois présentés dans les enseignements hors d'un cadre d'usage, par une approche trop souvent purement calculatoire vidant le sens de ces objets. Il est donc nécessaire de s'interroger quant à la spécificité de l'enseignement de la statistique et aux conditions de transposition dans des contextes divers, afin d'aider au développement de la démarche statistique de nos étudiants.

En tant que directrice scientifique de la Mission d'Appui à la Pédagogie et à l'Innovation (MAPI) de l'Université de Bordeaux depuis presque deux ans, je suis d'autant plus sensibilisée à la problématique de l'adaptation de stratégies pédagogiques accompagnant la réflexion didactique avec la prise en compte des spécificités disciplinaires, et donc à la contextualisation.

Pour toutes les raisons précédemment exposées, je conclurai cet avant-propos en me permettant de reprendre une formulation de Bernard Sarrazy qui résume parfaitement mon impression : Baignée depuis de nombreuses années dans l'environnement universitaire, ce travail de thèse m'a permis de clarifier « *ma propre équation idéologique* » (Sarrazy, 2002) en m'interrogeant du triple point de vue de mes rapports avec les institutions scientifiques, éducatives et politiques.

⁴ Cette notion sera précisée dans le cadre de l'écologie didactique des savoirs présentée par Chevallard.

INTRODUCTION

Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write.

Herbert George Wells (écrivain britannique, 1866–1946)

La statistique est un domaine scientifique ayant la particularité de se développer dans différents espaces et pour différentes raisons. C'est cette double diversité, spatiale et fonctionnelle, qui en fait tout son intérêt et sa richesse, sa force et parfois sa faiblesse. Cette discipline a été longtemps rattachée principalement aux mathématiques, mais a pris, semble-t-il, depuis quelques décennies son autonomie et a acquis une réelle reconnaissance au sein de nombreux domaines, jusqu'à devenir une science d'aide à la décision.

Comme le souligne Desrosières dans la « *Politique des grands nombres* », « *la statistique et les calculs des probabilités occupent une place essentielle parmi les outils d'invention, de construction et de preuves des faits scientifiques dans les sciences de la nature comme les sciences sociales* » (Desrosières, 2000, 19). La spécificité de la statistique, en tant que science d'analyse et de décision induite par l'existence de données et de démarche de compréhension de phénomènes observés, suscite l'intérêt de l'ensemble des domaines scientifiques et sociétaux. Elle en devient ainsi une discipline protéiforme s'appuyant sur des dimensions partagées, mais aussi parfois des formulations spécifiques.

L'usage et le développement de la science statistique au sein de différents domaines scientifiques construisent ainsi des objets répondant à des besoins propres à un domaine disciplinaire. Ils prennent une dimension transverse en étant décontextualisés, et peuvent revenir avec force dans certains giron disciplinaires. Rappelons par exemple que le test de Student a été développé par William Gosset, statisticien à la Brasserie Guinness à Dublin dans un but de contrôle de la qualité de la production de bière, puis qu'il a été diffusé *via* la publication d'un article, sous le nom d'auteur de Student, dans la revue *Biometrika*. Aujourd'hui, tout étudiant ayant

eu un minimum de 15 heures d'enseignement de statistique a connaissance de ce test, en connaît globalement le cadre d'usage, et les grandes étapes calculatoires.

Ainsi, certains objets et notions de statistique semblent prendre un statut amphibie pouvant évoluer dans des environnements extrêmement divers. Ils sont alors connus par des chercheurs d'univers différents, mais quelles fonctions y assurent-ils ? Comment ont-ils pu ou ont-ils dû s'adapter à ces environnements divers ? Est-ce, d'ailleurs, le même objet ?

La statistique est une science qui inquiète. D'autant plus quand des spécialistes des sondages tel que George Gallup⁵ se permettent d'indiquer « *Par la statistique, je pourrais prouver l'existence de Dieu* », et qu'un premier ministre anglais ajoute « *Je ne crois aux statistiques que lorsque je les ai moi-même falsifiées* » (Winston Churchill⁶).

Pour toutes ces raisons, et pour défendre le sérieux de cette discipline face à tant de méfiances, il est particulièrement important de pouvoir comprendre ce qui se joue dans la transmission d'informations concernant cette science, qu'il s'agisse des résultats obtenus ou des démarches mises en œuvre. Le scientifique doit se considérer comme membre d'une communauté, et pas uniquement comme sujet rationnel et lucide, et se doit d'œuvrer dans la capacité à communiquer son savoir et à faciliter les échanges au sein de cette communauté dont les frontières peuvent varier (Stengers, 1993).

Formuler avec clarté, précision et concision les démarches statistiques mises en œuvre et les résultats obtenus sont des éléments ayant un rôle social et stratégique pour le scientifique. Mais les spécificités de formulation dans les différents milieux concernant les objets eux-mêmes et leurs usages dans le cadre de la statistique sont à prendre en considération. Le scientifique a ainsi un rôle social dans l'accompagnement à la compréhension des résultats statistiques, la difficulté majeure d'appropriation étant due au halo d'incertitude les entourant toujours.

⁵ (1901 – 1984) statisticien et sociologue américain, il créa son entreprise de sondage d'opinion.

⁶ (1874 – 1965) homme d'état britannique, premier ministre de 1940 à 1945.

On comprend alors l'importance de la place de la didactique au sein de toute discipline, comme composante intégrée et non comme science extériorisée de cette discipline. Ceci est tout particulièrement vrai pour ce qui est de la statistique, tant son enseignement ne peut omettre le fait que cette discipline, en raison de la suspicion qu'elle suscite de manière presque implicite dans la pensée commune, se doit d'être accompagnée d'un perpétuel effort d'explicitation de ce qu'elle est et ce qu'elle n'est pas. Nous nous appuyerons sur la formulation de Chevallard et Wozniak:

Dans la recherche « non didactique » en une discipline donnée, il y a bien entendu du didactique, parce qu'il y a toujours une *intention didactique*. Mais celle-ci vise en principe un public de pairs ou de quasi-pairs, configuration qui pose au chercheur « non didacticien » des problèmes didactiques regardés en général – peut-être trop vite – comme fort éloignés de ceux que s'efforce de résoudre, ordinairement, le didacticien qui se réfère à l'institution scolaire. C'est dans cette perspective, soulignons-le en passant, qu'il faudrait situer l'affirmation – refusée aujourd'hui encore en nombre de disciplines, qui n'arrivent pas même à la *penser*, faute déjà de s'être dépris de l'*habitus* sécessionniste engendré par le grand partage historique des sciences – que la didactique d'une discipline donnée doit être regardée comme une composante de cette discipline même (Chevallard & Wozniak, 2003).

Nous aborderons donc, dans une première partie de notre thèse, le cadre conceptuel et théorique choisi et expliciterons les concepts inhérents à la problématique de la question de la didactique de la statistique et de sa spécificité, en lien principalement à la théorie de la transposition didactique. Un intérêt particulier sera accordé à la question de l'environnement didactique et notamment sur les différents espaces didactiques dans la vie d'un chercheur.

Dans le deuxième chapitre, nous montrerons la spécificité de la didactique de la statistique, les dimensions qui la définissent intrinsèquement et les spécificités de cette science notamment celle de pouvoir se mêler au plus près des autres disciplines et d'en absorber les spécificités en les déclinant dans sa propre démarche.

Dans une troisième partie, nous aborderons la question de la transposition didactique sous deux focales distinctes mais entretenant l'idée commune selon laquelle la statistique en tant que telle est "naturelle" mais ses usages dans diverses institutions peuvent révéler de spécificités de communautés. Nous considérerons dans un premier temps les régularités et différenciations des usages de la statistique

dans les thèses de doctorat de diverses disciplines et, par ailleurs, la démarche d'explicitation de démarche dans un cadre d'expertise en direction du politique.

La quatrième partie sera consacrée à l'enseignement de la statistique à l'université et aux défis que doit relever la pédagogie universitaire, ce qui existe, ce qui est fait, ce qu'on pourrait y faire. Il s'agira d'étudier comment la mise en place d'une approche programme a incité les enseignants à des recompositions de territoire qui pourraient ainsi ouvrir de nouveaux espaces pédagogiques. Nous nous interrogerons sur la possibilité d'ouvrir la discussion quant à la place de la science statistique dans un espace transversal aux disciplines, en évitant le terme de « socle commun ».

La déclaration de Bruno Latour dans « *La vie de laboratoire* » traduit assez bien la difficulté qu'il a fallu affronter tout au long de cette période de thèse et conclura cette introduction.

« Mais comment maintenir la distance et l'indépendance du jugement, lorsque l'on est soi-même un chercheur, un Occidental, un intellectuel ? Quelqu'un peut-il être assez ignorant des sciences pour jeter un regard vraiment neuf sur l'activité savante ? C'est sur ce point qu'il faut vraiment travailler la question, discipliner le regard, se forcer à la distance. » (Latour, 2008, 24).

CHAPITRE 1 – SITUATIONS DIDACTIQUES DE LA STATISTIQUE POUR L'ENSEIGNANT- CHERCHEUR

Etudier les phénomènes de transmission d'un savoir statistique, qu'il s'agisse d'une démarche, d'un test statistique, d'une notion telle que la représentativité, d'un résultat d'estimation ou de tout autre chose, nécessite de considérer et comprendre les environnements dans lesquels cela se déroule, mais aussi d'approcher et de cerner les différentes postures institutionnelles en jeu. La spécificité de la science statistique et de la diffusion de ses savoirs a induit le développement d'un cadre théorique spécifique que représente la didactique de la statistique, dans lequel nous situerons ce travail de thèse qui prend en compte les différentes dimensions sous-jacentes au phénomène.

« La didactique de la statistique est entendue comme cadre théorique d'étude des processus de communication, de diffusion et d'acquisition de la statistique, notamment en situation scolaire ou universitaire, qui ne peut être réduite à la statistique » (Régnier, 2005)

La didactique de la statistique ne se limite pas à l'espace de l'enseignement, mais ses domaines d'exercices doivent être vus de manière plus globale. Ainsi en dehors du contexte de l'enseignement, toute personne ayant conduit des études statistiques devra, par exemple, présenter et discuter les résultats obtenus à l'issue d'une étude statistique pour accompagner une prise de décision. Dans ce cas, elle se devra d'explicitier la réflexion et la démarche statistiques par lesquelles le travail a été réalisé et les résultats obtenus. Les résultats issus d'une démarche statistique ne doivent pas être isolés lors de leur diffusion d'une explicitation du développement y ayant conduit. Le résultat final n'a de sens que comme résultante d'une démarche contrôlée dans sa mise en œuvre. L'accès libre de plus en plus fréquent de fichiers de données issues d'enquêtes statistiques nécessite la présentation conjointe de

métadonnées précisant le cadre de l'étude et l'ensemble de la démarche (l'outil de recueil, l'effectif, la stratégie d'échantillonnage...).

Ainsi, la didactique de la statistique s'intéresse, à côté de la trinité classique « enseignant, étudiant et statistique », au pôle des pratiques sociales professionnelles de référence qui constitue une réelle spécificité du système didactique de la statistique appelée *la praxéologie* (Régnier 2003).

Comprendre les enjeux de la transmission du savoir, de la démarche statistique mise en œuvre mais aussi des résultats obtenus à l'issue d'une approche et démarche statistique dans des configurations diverses est la dimension que nous souhaitons appréhender. Il s'agira de ne pas omettre le fait que les connaissances à propos desquelles se nouent les relations didactiques doivent être traitées comme des objets vivants en tant qu'ils évoluent et changent en fonction de leur environnement d'application.

Le contexte dans lequel se situe ce travail de thèse est celui de l'activité des enseignants-chercheurs, non nécessairement statisticiens mais ayant une pratique et un usage de la statistique. L'activité statistique le conduit à la nécessité de diffuser ses connaissances, ses savoirs et ses pratiques de la statistique en tenant compte d'environnements potentiellement très différents.

A. Missions de l'enseignant-chercheur : la place de la didactique

Etudier la diffusion des connaissances et des savoirs statistiques dans l'environnement de la recherche nécessite de prendre en considération les acteurs, le savoir mais aussi le cadre de la diffusion c'est-à-dire l'environnement didactique. Nous proposons de nous appuyer sur la proposition d'Ernest Boyer qui caractérise en quatre dimensions la tâche des enseignants-chercheurs dans le but de faire bénéficier

à l'ensemble de la société des avancées de la recherche (E. L. Boyer, 1997). Il s'agit donc de :

- la découverte de la connaissance (*scholarship of discovery*), – recherche – innovation – création du savoir ;
- l'intégration des connaissances (*scholarship of integration*) - les publications ;
- l'application de la connaissance (*scholarship of application*) - transfert de la connaissance vers la société civile à travers l'expertise et la consultation ;
- l'enseignement (*scholarship of teaching*), centré sur la transformation et la transmission de la connaissance.

Dans le cadre présenté par la structuration de ces quatre situations, le chercheur se retrouvera fréquemment en situation de transmettre, de diffuser, d'enseigner et de former. La diffusion peut être en direction d'autres chercheurs de sa propre communauté, de chercheurs d'une autre communauté disciplinaire, d'étudiants ou de membres de la société civile mais aussi de politiques. Pour qu'il y ait transfert de connaissances de la recherche vers la société et que ce transfert puisse réellement avoir un impact sur les réalités sociales, le chercheur doit développer des compétences de communication, d'explicitation et d'accompagnement à l'apprentissage spécifique des connaissances variées.

En effet, les enjeux de chacun de ces temps didactiques différent, et appellent la mise en œuvre de cadres didactiques propres, et une réflexion différenciée quant aux transpositions didactiques spécifiques.

1. Différents temps didactiques dans la vie de l'enseignant-chercheur

La soutenance de thèse peut être décrite comme un rite de passage au plein sens de Bourdieu (1982). L'objet thèse correspond à l'œuvre permettant une reconnaissance d'affiliation à une communauté, il offre un premier accès à un sas

permettant de se diriger vers un devenir professionnel au sein duquel la dimension recherche sera présente.

La soutenance de thèse correspond à un acte de transmission singulier, au travers duquel les qualités didactiques du doctorant sont évaluées, bien qu'il ne le perçoive pas nécessairement ainsi, ni nécessairement les membres du jury car il est difficile en ce cas de faire la différence entre le contenu, sa maîtrise et sa transmission. Les membres du jury de thèse ne l'expriment que rarement de manière explicite, considérant avant tout leur rôle de jury comme garant de l'évaluation du doctorant quant à ses capacités et ses habiletés à développer des projets de recherche pour pouvoir être reconnu apte pour candidater au concours de maître de conférence, seconde étape de reconnaissance. La didactique n'étant pas perçue dans la plupart des situations comme faisant partie intégrante de la discipline, le jury ne positionnera pas son évaluation sur les habiletés didactiques du thésard et ne l'interrogera pas en ce sens directement. Cependant, durant la soutenance de thèse, le doctorant souhaite mettre à l'épreuve la reconnaissance de son travail par des chercheurs reconnus et ciblés de la communauté disciplinaire à laquelle il souhaiterait appartenir, mais aussi sa capacité à produire du savoir savant pouvant nourrir la réflexion de leur communauté scientifique. Il s'agit pour lui d'une phase de dépersonnalisation, en vue d'une reconnaissance institutionnelle de la qualité et de l'intérêt de son travail.

La diffusion de sa recherche est au centre de la carrière d'un chercheur. Il doit porter à la connaissance de ses pairs sa production, ses réflexions scientifiques ainsi que ses résultats selon des formes acceptées et comprises par sa communauté. Cependant le chercheur doit parfois diffuser des connaissances, non plus uniquement dans le cadre de sa propre communauté avec les codes et les règles qu'il connaît, mais en prenant le risque de partager au-delà de son habitat dans le cadre, par exemple, de travaux collaboratifs de recherche nécessitant dès lors un travail de dialectique partagée. De par ses compétences et connaissances potentiellement reconnues dans des sphères plus larges que sa propre communauté, l'enseignant-chercheur peut être sollicité par la société en tant qu'expert ; cette situation l'amènera à devoir partager les résultats de son expertise dans un environnement différent dans lequel les buts et les enjeux ne correspondent pas à ceux de la recherche. Ainsi l'enseignant-chercheur doit appréhender ces éléments spécifiques afin que la

transposition didactique soit facilitée. Il s'agit d'une situation didactique où le demandeur est en attente d'éléments de compréhension et d'apprentissage, mais aussi de résultats lui permettant de répondre à une interrogation bien souvent à des fins d'aide à la prise de décision. Il ne s'agit donc en rien d'un travail de vulgarisation scientifique, mais de la mise en œuvre d'une explicitation de la démarche scientifique pour répondre à une question précise. Il y a un niveau d'expertise sollicité par le politique, mais avec la nécessité de communiquer de l'information ainsi obtenue en explicitant la démarche.

Le cadre le plus classiquement reconnu pour étudier les situations didactiques est le temps d'enseignement dans une institution dont cela est la mission. Il s'agit ainsi pour l'enseignant-chercheur de transposer des savoirs, dont il peut choisir de sa propre initiative les contenus dans la plupart des cas, ce qui en fait une différence majeure avec la situation connue dans l'enseignement primaire ou secondaire. L'écart entre les savoirs savants ou experts et les savoirs à enseigner peut être très étroit, selon le niveau de l'enseignement et la proximité avec le propre cadre de recherche de l'enseignant-chercheur. Cette situation très particulière pour lui sera à prendre en considération, car la difficulté du nécessaire travail de transposition entre le savoir savant et le savoir enseigné peut représenter un obstacle didactique.

Les enseignants-chercheurs doivent développer une dialectique propre à chaque domaine d'intervention que sont l'enseignement, la recherche et l'expertise. La situation d'enseignement correspond au souhait d'accompagner des étudiants dans l'acquisition d'un certain nombre de savoirs, ne faisant pas encore partie de la communauté scientifique et n'ayant pas encore acquis l'ensemble des spécificités communautaires. Selon la spécialisation de l'année de formation, l'enseignant-chercheur devra plus ou moins distancier le savoir enseigné du savoir savant/expert considéré comme source. Dans la situation d'expertise, qui a pour raison d'être l'apport des connaissances de l'enseignant-chercheur au service de la société, il doit être en capacité de diffuser et donc de communiquer, mais surtout de faire comprendre les éléments attendus et le défi, dans le cadre de la statistique, est de

sensibiliser et faire comprendre la réalité de la démarche statistique avant de diffuser des résultats bruts.

Les trois situations didactiques représentant les phases d'enseignement, de recherche et d'expertise sont explicitées sur le schéma 1.

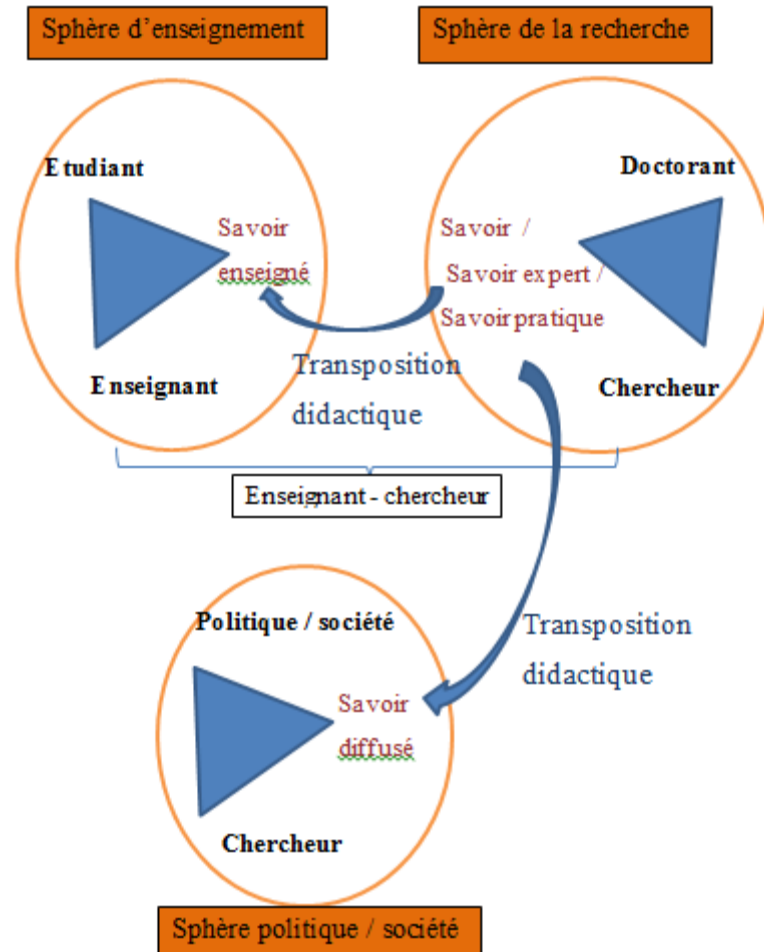


Schéma 1 : Différentes situations didactiques dans les missions d'un enseignant-chercheur

Ce schéma présente les circuits de transposition didactique vus du point de vue du chercheur et souligne les points essentiels discutés dans le cadre de ce travail. Les savoirs deviennent savants dans la sphère de la recherche mais peuvent être initiés dans le cadre de situations d'expertises offrant l'opportunité de développer de nouveaux savoirs. Nous n'avons pas intégré les phénomènes de transpositions didactiques dans la sphère recherche entre communautés scientifiques, ni ceux se

jouant au sein d'une communauté particulière. Nous nous sommes limitée aux situations investiguées par la suite et qui soulignent les situations didactiques de l'enseignant-chercheur lors de changement de types de sphères.

Dans chaque situation didactique du schéma 1, le statut de l'étudiant diffère mais il s'agit cependant, toujours d'un individu abordant, avec une structuration particulière, des connaissances compatibles (ou non) et en accord (ou non) avec ce qu'il connaît. Chaque situation est caractérisable par un contrat didactique le plus souvent implicite.

2. Différentes sphères didactiques pour l'enseignant-chercheur

Etudier ces « différents temps de diffusion des connaissances et des savoirs nécessite l'examen des critères de reconnaissance des phénomènes par lesquels se manifeste une action didactique » (Sarrazy⁷, 2005)

Il y a plusieurs façons de choisir une typologie des différents environnements ou sphères dans lesquels les phénomènes didactiques peuvent être étudiés. Nous nous limiterons dans cette présentation aux situations didactiques dans le domaine universitaire et de la recherche. Nous considérerons donc les trois types de sphères présentées dans le schéma 1, les sphères de la recherche et de l'enseignement supérieur qui correspondent aux deux ancrages de l'enseignant-chercheur, et la sphère sociale et politique qui peut le solliciter en tant qu'expert.

La sphère recherche peut être subdivisée pour l'enseignant-chercheur en deux espaces, d'une part, sa communauté ou espace de création, et, d'autre part les autres environnements appartenant toujours à la recherche mais correspondant à d'autres communautés disciplinaires en interaction plus ou moins directe avec sa propre communauté, et pouvant le solliciter en tant que partenaire ou collaborateur. Par la suite, nous nommerons ce second espace « communauté partenaire d'usage ».

⁷ « Questions à la théorie anthropologique du didactique du point de vue de la théorie des situations et de l'anthropologie wittgensteinienne » Bernard Sarrazy, lors du premier congrès international sur la théorie anthropologique du didactique. Société, école et Mathématiques : apport de la TAD. Baeza du 27 au 30 octobre 2005.

- La communauté du chercheur correspond à son environnement scientifique de proximité à l'intérieur duquel il développera une production de savoirs experts, au sens de Perrenoud (1998). Le choix de se positionner dans le cadre proposé par Perrenoud de production de savoirs experts, se justifie par la spécificité de la science statistique trouvant tout son sens dans la mise en œuvre de démarches au-delà de l'usage d'outils isolés. Au sein de cette sphère, des enjeux de partage de savoir avec ses pairs se jouent. Il s'agit de dépersonnaliser ce savoir qui deviendra dès lors un savoir reconnu par l'ensemble de sa communauté. Ce savoir pourra ou non être diffusé au-delà des frontières de son propre environnement, pour une transmission mais aussi avec la possibilité de vivre une décontextualisation suivi d'une recontextualisation.
- La « communauté partenaire d'usage » va ici correspondre aux communautés de chercheurs n'étant pas à l'origine de la création du savoir mais pouvant l'utiliser dans leur propre environnement de recherche. La connaissance que le chercheur-usager utilise peut être issue des enseignements reçus dans le cadre de sa formation initiale, ou ultérieurement dans le cadre de la formation continue, mais aussi d'échanges lors de projets interdisciplinaires par exemple. Cette transposition de savoir et son adaptation dans un nouveau domaine de recherche peut conduire dès lors à la création d'un nouveau savoir expert. Le développement de savoirs experts sera favorisé par ces interconnexions entre communautés à l'intérieur de la sphère de recherche. Ces moments d'échanges et de diffusion nécessitent la transposition des savoirs d'un domaine à l'autre en le nourrissant grâce aux cadres scientifiques et épistémologiques de chaque environnement. Dès lors ce n'est plus le même savoir que l'on retrouve dans chaque sphère mais bien des savoirs issus de racine commune ayant vécu des développements spécifiques.

La sphère de l'enseignement est quant à elle composée de deux profils d'acteurs, les enseignants et les étudiants liés par la transmission de savoirs.

Le savoir enseigné dépend de plusieurs facteurs liés entre eux : conception épistémologiques dominantes dans les communautés savantes, relations

culturelles entretenues avec ces domaines par le public, finalités sociales fixées à cet enseignement. (Johsua S., 2003)

La marque déposée par les enseignants-chercheurs sur les savoirs-enseignés est inhérente à leurs domaines de recherche et d'expertise. Le choix des définitions, des cadres conceptuels dans lesquels ils développent leurs enseignements, les stratégies pédagogiques qu'ils mettent en œuvre sont autant de signaux caractérisant l'état d'esprit et la personnalité de l'enseignant-chercheur. En tant que producteur de savoir, il a ainsi une filiation à un courant de pensée, à une spécialisation qui transparaîtra naturellement dans sa démarche pédagogique et ne pourra être atténuée qu'à la suite d'efforts de transposition didactique, si cela s'avère nécessaire pour assurer une transmission plus appropriée du savoir facilitant la capacité de compréhension de l'étudiant.

Le cadre de l'enseignement supérieur diffère de l'enseignement primaire et secondaire par de nombreux aspects. Tandis que l'enseignant du secondaire est contraint de suivre les programmes nationaux dans le respect des réglementations, l'enseignant-chercheur est libre dans sa démarche d'enseignement selon l'article L952-2 du décret n°84-431 du 6 juin 1984 fixant les dispositions statutaires communes applicables aux enseignants-chercheurs et portant statut particulier du corps des professeurs des universités et du corps des maîtres de conférences.

Dans l'accomplissement des missions relatives à l'enseignement et à la recherche, ils jouissent, conformément aux dispositions de l'article L. 952-2 du code de l'éducation, d'une pleine indépendance et d'une entière liberté d'expression, sous les réserves que leur imposent, conformément aux traditions universitaires et aux dispositions du code de l'éducation, les principes de tolérance et d'objectivité.

L'enseignant-chercheur est maître du contenu de son cours, même si la ligne directrice, les objectifs pédagogiques ainsi que les thèmes de son cours ont pu être discutés en équipe pédagogique pour une intégration cohérente d'ensemble, ce point fera l'objet de discussion dans le chapitre 4. Il pourra ainsi donner la direction qu'il souhaite avec une bien plus grande liberté que celle des enseignants du primaire ou du secondaire qui auront quant à eux su prendre une plus grande autonomie dans le choix de modalités pédagogiques. Le développement de nouvelles stratégies et

modalités pédagogiques variées au sein de l'enseignement supérieur est un enjeu du développement du nouveau courant de la « *pédagogie universitaire* ».

Les enseignants-chercheurs sont sollicités parfois par les membres de la sphère « politique – sociale » pour leur capacité d'expertise, afin d'accompagner le transfert de connaissance au plus proche de la société. L'enseignant-chercheur a un rôle important en tant qu'expert dans la vie de la société et de cette manière une influence sur la prise de certaines décisions, il se doit d'en évaluer avec justesse sa responsabilité. On lui demande alors de mesurer à sa juste valeur l'attente exprimée de la société quant à la réponse scientifique ainsi que les retombées possibles. Il s'agit donc de considérer l'environnement dans lequel son savoir prendra sens. Le statut du savoir est ici le savoir diffusé, et le passage de savoir expert au savoir diffusé est une situation de transposition didactique, ciblée sur la nécessité de compréhension de l'information, mais aussi, et surtout en statistique, de la compréhension de la démarche. Le savoir diffusé comme le savoir enseigné sont à considérer en regard à l'approche suspicieuse que porte parfois le public sur la statistique et les statistiques. Le savoir diffusé doit être déterminé en fonction de la finalité sociale qu'il faut cerner précisément pour permettre un cadrage précis, clair et partagé. Il est important de souligner qu'il est préférable que l'enseignant-chercheur préjuge de l'ignorance de son interlocuteur car, comme le soulignent Latour et Woolgar dans « *La vie de laboratoire* » la prise en compte de l'ignorance nécessite de spécifier l'implicite (Latour, 2008).

La question de la transmission d'un savoir interroge tout particulièrement le choix de ce savoir. Que doit-on ou que veut-on transmettre ? De quel savoir parle-t-on : savoir théorique, connaissance intégrée, pratique contextuée, mise en œuvre pratique ... ? Qui accompagne la transposition didactique et quels choix épistémologiques sont réalisés. Mais aussi pourquoi ce choix de savoir ? Qu'est ce qui influence ce choix et quel impact de l'environnement sur le choix ? Il s'agit d'un ensemble de questions accompagnant de manière logique les réflexions concernant la théorie de la transposition didactique.

B. Quelles spécificités de la transposition didactique pour les missions de l'enseignant-chercheur ?

La théorie des processus de transposition didactique initiée par Verret (1975) et repris par Chevallard (1985) est au centre des enjeux présentés précédemment. S'intéresser à la transposition didactique consiste à identifier les phénomènes et leurs effets lors des différentes étapes de nomadismes du savoir des sphères savantes au savoir des sphères d'enseignement (*lato sensu et stricto sensu*). Comme le souligne Chevallard il s'agit de repérer les différentes transformations du savoir et de ses statuts.

1. Cadre général

Chevallard distingue deux étapes dans la transposition didactique : la transposition didactique externe (passage du savoir savant au savoir à enseigner) et la transposition didactique interne (passage du savoir à enseigner au savoir enseigné) (1985). La transposition externe nécessite deux étapes importantes qui sont la phase de décontextualisation du savoir savant issu de la sphère de production et la phase de recontextualisation pour construire le savoir à enseigner ou à transmettre dans la sphère scolaire selon Reuter (2007). Le centre opérationnel de ce processus est ainsi appelé la noosphère⁸. Pour la transposition didactique de la statistique dans l'environnement universitaire et sociétal, la noosphère ne correspond pas précisément à ce que l'on peut concevoir comme telle dans l'enseignement secondaire, la raison principale étant celle de l'indépendance des enseignants-chercheurs. Nous avons déjà souligné cette situation et nous reviendrons sur ce point ultérieurement.

⁸ La noosphère – du grec *noos* (esprit, pensée) et *sphaira* (corps rond) – désigne littéralement la partie de l'espace où se manifeste, se crée une pensée. Chevallard introduit le terme dans le vocabulaire didactique pour désigner un espace, une sorte de *sas* dans lequel « s'opère l'interaction entre le système d'enseignement et l'environnement sociétal. » (1991, 24-5) ; elle correspond à une partie de l'espace social dans laquelle les représentants du système d'enseignement (les membres des commissions ministérielles, les représentants de la société – parents d'élèves, spécialistes de la discipline, auteurs de manuels, des revues scientifiques ou militantes...) « pensent », négocient, débattent de ce qu'il convient de faire ou de changer dans le système d'enseignement. En d'autres termes, la noosphère désigne « l'ensemble des personnes et groupes intéressés à la création et à la communication des savoirs d'un certain domaine » (Brousseau, 1998, 341).

L'approche proposée par Chevallard, en 1985, centrée sur le savoir savant comme élément de base pour le processus, peut sembler trop restreinte pour le champ de la statistique. En effet le processus de transposition didactique ne doit pas concerner uniquement les savoirs, *stricto sensu*, mais prendre aussi en considération toute l'importance de la démarche statistique. En 1994, Chevallard précise dans « *Les processus de transposition didactique et leur théorisation* » que :

« Ce qui doit être transposé est bien cette trinité – domaine de réalité, pratique, savoir - dont l'ensemble ne peut exister, à l'instar d'un nœud borroméen⁹, si l'un quelconque de ses éléments se perd ». (Chevallard, 1994)

Dans le cadre de cette définition, la proposition faite par quelques didacticiens d'étendre la théorie de la transposition didactique aux savoirs experts pour Johsua (1996) et aux savoirs professionnels pour Rogalski et Samurçay (1994), semble plus appropriée au champ de la statistique, de par la définition même de la statistique en tant que sciences des données. Perrenoud complète ainsi le cadre de la théorie de la transposition didactique en indiquant qu'il est important :

« de reconnaître l'existence et la diversité des savoirs liés à une forme d'expertise professionnelle ou plus globalement de maîtrise pratique. De tels savoirs experts existent et sont à l'œuvre dans la plupart des pratiques sociales, qu'ils aient ou non des fondements dans les savoirs savants, que ces fondements soient ou non connus des praticiens ». (Perrenoud, 1998)

Plusieurs éléments sont spécifiques à la statistique en tant que discipline : il s'agit de transmettre l'importance de la structuration de la démarche statistique et donc de chaque étape, d'être en mesure de percevoir et comprendre le régulier mouvement entre les données empiriques et l'usage de modèles, d'être en mesure d'apprécier et d'accepter la situation toute particulière de la présence constante de l'incertitude et de renoncer à un usage systématique de l'idée de la vérité. L'ensemble de ces éléments devient difficilement envisageable sans une prise en considération de la pratique et non uniquement du savoir hors de tout contexte.

Expliquer le test du chi-deux d'indépendance par une approche algorithmique déroulant les étapes de détermination des hypothèses (H_0 et H_1) à

⁹ En mathématiques, les nœuds borroméens constituent un entrelacs de trois cercles qui ne peuvent être détachés les uns des autres même en les déformant, mais tel que la suppression de n'importe quel cercle libère les deux cercles restants

tester, des conditions d'application du test, de la rédaction de la formule de la statistique de test en en précisant la loi qu'elle suit sous H_0 , la zone de rejet et les critères de conservation ou rejet de l'hypothèse de départ H_0 , ne pourra pas être un élément suffisant pour la transmission du savoir statistique que représente le test du chi-deux dans son usage. Car la présentation de ce processus seul isole l'outil des situations d'usage (le savoir, des connaissances), qui détermine tout son sens statistique et donc son intérêt. Pour réaliser un test statistique, il ne suffit pas de connaître les étapes de l'algorithme de calcul, on retrouve ici le célèbre paradoxe de la règle chez Wittgenstein (1961) et son corollaire chez Brousseau (1990) avec celui du contrat (car dans ce cas, tout logiciel de statistique serait apte à remplacer le statisticien). Ainsi, on sait que le test du chi deux nécessitera certaines conditions d'application, comme par exemple des effectifs théoriques suffisants. Un usage du test du chi-deux hors contexte peut signaler la possibilité de réunir les modalités ayant des petits effectifs. Bien que cette stratégie soit d'un point de vue calculatoire cohérente, elle peut conduire l'utilisateur à appliquer cette technique sans aucun recul et aboutir au final à des regroupements rendant l'analyse ininterprétable et ne répondant en rien à la question initiale.

Il est donc nécessaire que le test statistique enseigné soit une notion non pas hors de tout contexte mais comprise avec les sept étapes présentées dans le tableau 1 qui permettent ainsi de maintenir le sens et l'intérêt de l'objet.

Tableau 1 : Les sept étapes d'un test statistique

Etapes	Contenu
1	Définition de la question et modélisation impliquant la mise en forme des données, définition des modalités des variables considérées pour pouvoir répondre à la question
2	Dans l'environnement théorique, pris en situation de modèle, détermination des hypothèses nulle et alternative du test (H_0 et H_1)
3	Définition et choix d'une statistique de test T adéquate et de sa loi de probabilité sous la condition de l'hypothèse nulle comme réalité
4	Etude du comportement de T sous H_1 et déduction de la forme de la zone de rejet R
5	Calcul de cette zone R pour un niveau puis confrontation aux données et/ou calcul de la p-valeur du test sur les données.
6	Conclusion statistique : conservation ou rejet de l'hypothèse de départ H_0 et commentaire éventuel sur la p-valeur
7	Conclusion stratégique : décision que l'on va prendre une fois éclairé par le résultat statistique

Comme dans la plupart des situations statistiques, la première étape d'un test consiste en la modélisation du problème, en fonction de la situation et de la question posée, c'est-à-dire le passage des données au modèle. Cette étape correspond précisément à l'étape angulaire de l'usage de test ou modèle dans une démarche statistique : elle nécessite une bonne compréhension du problème posé. Ainsi, le cadre du test statistique du chi-deux implique de se positionner face à une question dont les éléments descriptifs ont été présentés avec pertinence à partir de tableaux de contingence mais aussi de tableaux de répartition conditionnés sur l'une des deux variables, selon la question formulée.

La première étape et la deuxième étape traduisent ainsi l'aspect pratique contextualisé, qui avait été oublié dans la présentation initiale.

La définition que donne Perrenoud de la transposition didactique souligne que les pratiques ne peuvent pas être exclues de la théorie de la transposition

didactique dans la mesure où « *la transposition des savoirs savants n'est qu'un cas particulier, certes pertinent et intéressant, mais qui n'épuise pas le réel* » (1998). Elle semble plus proche de la situation particulière de la statistique. Selon lui, les deux sources de la transposition didactique de la statistique seraient donc les savoirs et pratiques ayant cours dans le cadre d'intérêt. L'objet, savoir ou pratique, à enseigner se décline en curriculum formel, objectifs et programme, et l'objet enseigné en curriculum réel et contenu de l'enseignement.

Perrenoud, à la différence de Chevallard, souligne dans sa vision de la transposition didactique l'importance de s'intéresser aux apprentissages. Nous reviendrons sur cette approche de Perrenoud dans le chapitre 4, afin d'aborder la question de l'enseignement de la statistique à l'université en regard aux changements actuels des paradigmes générateurs de l'offre de formation notamment par la mise en place du système d'accréditation des formations universitaires.

La statistique est une discipline ayant comme matière première des données, qu'elle aura parfois à recueillir et à partir desquelles elle construira une stratégie afin de répondre à un ensemble de questionnements ou problématiques. Cette stratégie se construira à partir d'un ensemble d'outils adéquats qui seront mis en œuvre selon des pratiques appropriées et développées dans différents environnements. La praxéologie de la statistique, en tant qu'étude de l'ensemble des actions spécifiques de l'usage de la statistique et subissant l'influence de spécificités disciplinaires, est une des composantes importantes pour la didactique de la statistique.

2. Déroutement et spécificité

L'enseignant-chercheur de l'université est dans une situation très particulière dès lors qu'il s'agit pour lui d'enseigner et de diffuser son propre savoir ainsi que sa propre pratique dans un environnement spécifique qui peut être plus ou moins éloigné, physiquement, sociologiquement et disciplinairement du sien.

La décomposition en deux temps de la transposition didactique proposée par Chevallard (transposition interne et transposition externe) n'est pas clairement transposable dans ce cadre, l'enseignant-chercheur, en raison de son indépendance pouvant être vu comme seul maître à bord dans le cadre de son enseignement. Selon

les situations d'enseignement la distance entre le savoir expert et le savoir enseigné peut sembler courte : il peut s'agir par exemple d'une situation d'enseignement en direction d'étudiants de master deuxième année suivant un parcours propre à la spécialité de l'enseignant-chercheur, formation au sein de laquelle d'ailleurs se trouve potentiellement son futur doctorant. Cependant, quelle que soit cette distance les étapes de transformation seront toujours présentes et traduiront en quelques sortes la nécessité pour l'enseignant-chercheur d'accompagner à la compréhension des démarches, des résultats et des savoirs qui font partie intégrante de son activité de chercheur et qui peuvent lui être singuliers, afin d'accompagner l'étudiant dans l'apprentissage des savoirs et des connaissances.

La situation de l'enseignant-chercheur est, du point de vue de la transposition didactique, assez singulière. En effet, toute transposition nécessite tout à la fois une dépersonnalisation du savoir et, des conditions de sa production une décontextualisation, comme condition nécessaire à son usage dans des contextes nouveaux. Nous soutiendrons ici que la position institutionnelle l'emporte sur la dimension personnelle. L'enseignant-chercheur devra opérer ses propres transpositions (sphère de la production – sphère d'usage) selon des processus isomorphes à des situations dans lesquelles le producteur de savoir est distinct du diffuseur. L'enseignant-chercheur est par définition assujetti à plusieurs institutions. Dans la sphère recherche il est assujetti à son rattachement recherche qui sera un laboratoire avec ses codes, ses références. Cette situation est particulièrement bien décrite par Bruno Latour dans « *La vie de laboratoire* » suite à la période d'observation qu'il a consacré à l'étude anthropologique d'une communauté de chercheurs (2008). Sur son second mi-temps, l'enseignant-chercheur s'assujettit à une institution d'enseignement qui pour des raisons évidentes de diffusion possède des codes, des usages et des pratiques différents. Nous présentons ce cheminement dans le schéma 2.

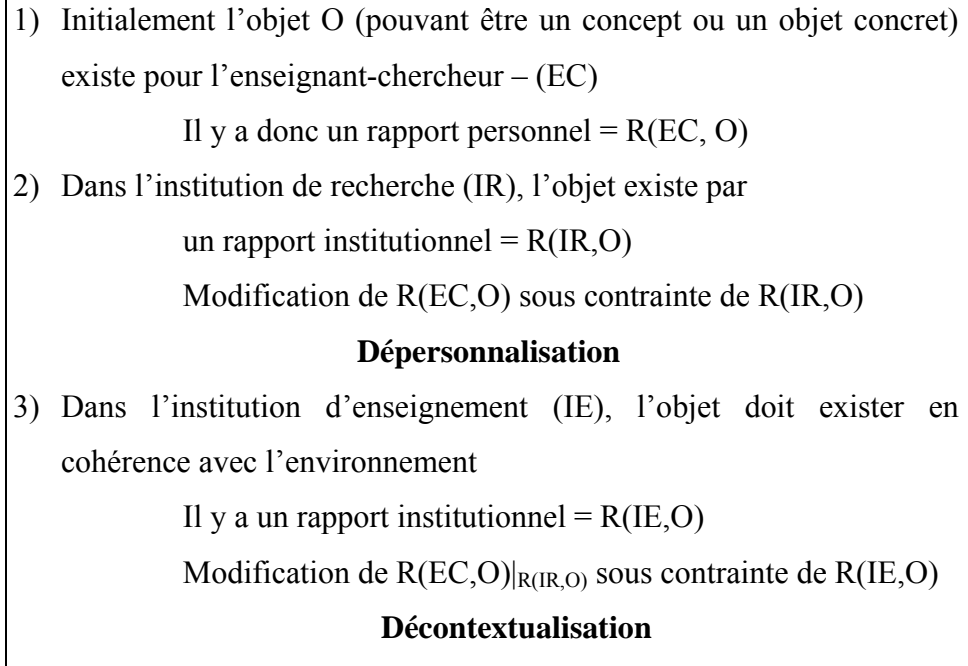


Schéma 2 : Cheminement de l'objet savoir d'un enseignant-chercheur

Cette situation est ainsi tout à fait particulière car, de par leur situation de bi-appartenance à deux environnements spécifiques, très spécialisés mais à mission très différentes, les enseignants-chercheurs se trouvent en situation d'un double changement du rapport à leur propre savoir.

Tout enseignant-chercheur est rattaché à une section CNU (Conseil National des Universités) qui est disciplinaire. Il sera sollicité pour intervenir dans des enseignements rattachés à la discipline correspondant à sa section. Concernant la statistique, une seule sous-section de CNU contient ce terme dans son intitulé, il s'agit de la sous-section 4604 - Biostatistiques, informatique médicale et technologies de communication de la section 46 - Santé publique, environnement et société du CNU Santé. Cependant, la plupart des enseignants-chercheurs en statistique sont issus de la sous-section 26 « Mathématiques appliquées et application des mathématiques » du groupe 5. Ainsi l'enseignement de statistique peut être assuré par des enseignants rattachés à des sections de CNU très différentes. La spécificité de la statistique en tant que discipline n'a pas d'écho sur une unicité institutionnelle puisqu'elle peut être "logée" au sein d'autres disciplines, tout en ayant des fondements communs la caractérisant comme une discipline à part entière.

L'enseignant-chercheur qui participe aux enseignements de statistique à l'université n'a pas toujours suivi un cursus spécifique de formation disciplinaire en statistique. Il a suivi, dans la plupart des cas, des enseignements de statistique dans le cadre de sa propre formation disciplinaire, puis a développé des compétences en statistique suite à des acquis professionnels, des formations continues et des apprentissages lors de travaux interdisciplinaires. Ses travaux de recherche ont pu ainsi faire reconnaître ses compétences en statistique et le conduire à se proposer ou à être sollicité pour intervenir dans le cadre d'enseignements s'appuyant sur la science statistique.

De même, l'enseignant-chercheur sollicité en tant qu'expert sur une thématique pourra expliciter des points spécifiques de statistique sans être pour autant "statisticien", mais en étant reconnu comme légitime dans un usage spécifique de la statistique au service de la question formulée pour l'expertise.

C. Pourquoi parler d'écologie didactique des savoirs pour la science statistique ?

L'usage de la statistique de plus en plus présent tant au sein des médias que de l'information scientifique lui confère une situation toute particulière et souligne l'importance et les enjeux de sa didactique. Ce qui est appelé ou présenté comme un élément, outils, savoir de statistique dans ces différents échanges d'informations peut couvrir des notions bien différentes et être présenté sous des formats très variés en raison de l'environnement même de sa création. Comment réussir à percevoir les différences sous-jacentes, les subtilités si l'on n'est pas préparé ? S'intéresser aux fonctions didactiques dans le champ de la statistique nécessite de regarder l'influence des diversités d'environnement dans le système didactique.

Comme le souligne Chevallard « *le concept et la résistance qu'il rencontre amènent à la lumière une autre vérité du fonctionnement didactique : on ne comprend pas ce qui se passe à l'intérieur du système didactique si l'on ne prend pas en compte son extérieur* » (1985).

1. Ecologie didactique des savoirs

Cet extérieur peut être défini comme l'environnement écologique du savoir qui va être impliqué dans un système didactique. L'étude écologique de cet environnement consiste à comprendre l'habitat lui-même en tant que localisation et composition. Il s'agit d'une compréhension principalement statique du système, mais la dimension considérée comme dynamique correspondra à la dénomination de *niche* par Chevallard, et couvrira la dimension des usages et de la mise en œuvre.

Ils (ndlr : les écologistes) « distinguent, s'agissant d'un organisme, son habitat et sa niche. Pour le dire en un langage volontairement anthropomorphe, l'habitat, c'est en quelque sorte l'adresse, le lieu de résidence de l'organisme. La niche, ce sont les fonctions que l'organisme y remplit ; c'est en quelque façon la profession qu'il y exerce. [...] Que les objets de savoir nouvellement introduits, plongés tout à coup dans un écosystème différent, allaient changer de « profession » pour s'adapter à un environnement neuf où ils devraient entrer en interrelation avec de nouveaux « partenaires » pour composer avec eux des « associations » jusque-là inédites. » (Chevallard, 1994)

Perez commente cette notion en proposant la formalisation suivante :

Selon cet auteur (ndlr Chevallard), deux notions guident l'analyse, l'habitat et la niche d'un objet :

Un savoir donné S se retrouve en divers types d'institutions I qui sont pour lui, en termes d'écologie des savoirs, autant d'habitats différents. Or, à considérer ces habitats, on aperçoit immédiatement que le savoir considéré vient régulièrement y occuper des niches bien distinctes [...] L'articulation de ces deux notions permet de définir une méthodologie pour décrire le rapport institutionnel à $S\bar{y}$ ¹⁰, soit le $R(I; S\bar{y})$. Il s'agit d'examiner, dans les lieux où $S\bar{y}$ est enjeu didactique, quels sont les autres objets présents, et quel est son rôle dans le système d'objets avec lequel il est en relation. Dès lors, nous rendrons compte du lieu où il est présent, du comment on entre en rapport avec cet objet, ce qui est fait, et ce qui attendu de cet objet (le rapport institutionnel). En étudiant ainsi les conditions d'existence, l'analyse écologique permet de mieux cerner les relations et modes de vie de $S\bar{y}$. (Perez, 2010)

L'étude du système didactique dans une situation particulière d'enseignement ou de partage de connaissances doit intégrer des éléments anthropologiques de l'environnement. Les objets de savoirs ne doivent pas subir la même transposition selon le cadre didactique dans lequel ils seront sollicités. Les enjeux de la transposition ne peuvent pas être étudiés de manière isolée. Ils se

¹⁰ L'original est noté Sy mais nous préférons barrer le y qui n'a semble-t-il pas de sens particulier.

doivent d'être compris en regard des autres éléments intervenant dans l'environnement.

Dans le cadre de l'enseignement de la statistique, la place de la statistique non paramétrique, par exemple, est bien différente selon les disciplines. Très présente et nécessaire en physiologie pour des raisons d'expérimentation sur faibles effectifs, elle est une stratégie de dernier recours en épidémiologie, discipline ayant pour habitude de s'intéresser aux études sur grands effectifs. Les justifications d'usage, la reconnaissance de sa validité seront donc à mettre en regard de l'ensemble des autres objets avec lesquels elle partage l'environnement.

2. La statistique dans des environnements différents

La place transversale de la statistique, tant au niveau disciplinaire que de la citoyenneté, implique une pluralité des environnements dans lesquels elle apparaît. Un habitat possible de la statistique se reconnaît par la présence de statisticiens ou par la production de supports mettant en évidence une démarche statistique ou le besoin de résultats statistiques en vue de prises de décision.

Les enjeux didactiques dans le cadre de chaque environnement seront fonction du système didactique mis en place et de la prise en considération des éléments constitutifs de l'environnement. Les transpositions didactiques devront donc être pensées en prenant en considération l'environnement dans son ensemble. Dans le cadre de la statistique, chaque environnement a un rapport différent : il peut s'agir d'un rapport utilitaire, d'un rapport scientifique ou d'un rapport décisionnaire.

Dans les instituts nationaux tels que l'INSEE (Institut national de la statistique et des études économiques), InVS (Institut de Veille sanitaire), l'INED (Institut national des études démographiques), ... la place centrale de la statistique est évidente comme science de la production et de l'analyse de données au service de l'Etat. L'enjeu principal est ainsi la veille de données, la problématique de la gestion des données manquantes et la capacité à correctement estimer les paramètres populationnels. Il s'agit de l'environnement historique du développement de la statistique, pour lequel l'approche est avant tout descriptive. Les questions centrales concernent les problématiques d'enquête et la mise en œuvre de plan de sondage.

Selon les communautés de chercheurs, les éléments phares de la statistique diffèrent, la statistique inférentielle pour certains domaines, la statistique bayésienne pour d'autres, induisant des stratégies d'analyse différentes. Longtemps les équipes d'épidémiologistes ne pensaient que par la mise en œuvre des modèles de régression, il n'était pas question d'introduire les approches par analyse de données à la française ni les stratégies de classification. Il a fallu attendre les années 2000 pour que les deux stratégies finissent par trouver des points d'accroche grâce à l'introduction de modélisation considérant des variables latentes. Les usages dans une communauté de telles ou telles stratégies de modèles ne peut être compris qu'en prenant en considération les phases historiques et les éléments épistémologiques les complétant.

Un savoir S peut être reconnu au sein de différentes institutions, mais, dans la plupart des situations, une institution sera reconnue comme le berceau de ce savoir noté P(S). Les savoirs statistiques ne naissent pas tous du même berceau comme certains pourraient le penser, mais, bien au contraire de célèbres histoires montrent la diversité des lieux de production. L'analyse factorielle a été ainsi présentée par le psychologue anglais Charles Spearman (1863 – 1945) lors de sa présentation du facteur g, pour facteur général. Une analyse factorielle construite à partir de données sur un échantillon d'étudiants a permis de mettre en évidence une association entre toutes les performances mentales indiquant ainsi la présence d'un seul facteur d'habileté général.

Cependant, le savoir une fois créé va germer au sein de différentes institutions et se développera en fonction du terreau rencontré, c'est-à-dire selon la proximité avec d'autres savoirs, mais aussi de la richesse des pratiques et potentiellement de l'émulation à l'usage de ce savoir. Les transformations d'usage pourront conduire ainsi jusqu'à la création de nouveaux savoirs, permettant la prise en compte de situations plus complexes. L'immigration d'un savoir de l'habitat A à un habitat B peut être due à différentes situations. Il peut s'agir tout simplement d'une curiosité d'un habitant de B, qui souhaite étudier la possibilité d'utiliser une stratégie analytique inhabituelle dans son habitat mais classique dans d'autres.

Ainsi en épidémiologie, les méthodes d'analyse des données les plus utilisées initialement étaient les méthodes de régression, pour lesquelles la variable à expliquer est définie par la question posée en amont de la mise en place de l'étude, et les variables explicatives sont *a priori* déjà ciblées. Il s'agit d'une démarche d'épidémiologie analytique ayant pour raison première la recherche de facteurs de risque ou de facteur associés à une pathologie. Les études épidémiologiques étaient ainsi construites pour répondre à des questions précises de caractérisations de facteurs de risque, le nombre de variables recueillies était défini et souvent raisonnable. Les situations devenant de plus en plus complexes et multidimensionnelles, la caractérisation *a priori* de facteurs associés devenait plus incertaine et les interactions entre facteurs difficiles à gérer. Une notion, connue depuis Spearman en psychologie, a ainsi fait une apparition remarquable en épidémiologie : il s'agit des facteurs latents (variables non observées directement mais pouvant être considérées comme un facteur composite de variables recueillies et observées). Les méthodes multidimensionnelles ont ainsi fait leur apparition au sein des stratégies d'analyse afin de pouvoir définir de nouveaux facteurs composites permettant d'expliciter une part importante de l'hétérogénéité de l'ensemble de variables.

L'épidémiologie nutritionnelle est d'ailleurs un domaine tout à fait remarquable quant à l'évolution massive de l'usage de ce type de méthodes à partir des années 2000 ; il s'agissait de pouvoir déterminer des profils de comportements alimentaires nécessitant la prise en compte d'un maximum d'informations recueillies par des techniques classiques de recueil par rappel à 24h¹¹, ou fréquentiel¹². Le développement des méthodes d'analyse factorielle ont nécessité la prise en compte des spécificités des mesures considérées dans ce cadre qui ne correspondaient pas aux situations rencontrées lors du développement de ces méthodes initialement plus fréquemment utilisées dans le champ de la psychologie (Samieri et al., 2008). Il s'agissait dès lors de définir une niche pour ce nouveau savoir qui venait d'emménager, et cela impliquait le développement de nouvelles compétences pour

¹¹ Bilan de la consommation de l'enquête sur la totalité de la journée précédant le jour de l'étude.

¹² Type de questionnaires permettant de recueillir la fréquence moyenne à laquelle les enquêtés ont consommé certains types d'aliments, de médicaments, de soins ou utilisé des transports au cours d'une certaine période.

les chercheurs afin de s'approprier les méthodes et de les décliner si nécessaire selon leur spécificité.

3. Des situations de changement d'habitats rencontrés

Dans le cadre de notre activité de statisticienne au sein de l'université, nous avons rencontré plusieurs situations de transfert de méthodes statistiques habituellement utilisées dans un environnement vers un autre environnement telles que celle présentée ci-dessus.

Cette situation concernait le développement de l'usage de la méthode LQAS (Lot quality Assurance Sampling), il s'agissait d'une méthode classiquement utilisée dans le monde industriel ayant pour objectif de permettre de détecter les éléments de production risquant de ne pas répondre aux critères d'acceptabilités, à un moindre coût et rapidement, cela concernant donc les contrôles de qualité (Dodge & Romig, 1929). L'OMS a publié dès 1991 des rapports recommandant l'usage de cette méthode dans les situations d'évaluation vaccinale, la méthode permettant dans un délai rapide de pouvoir conclure au risque d'une mauvaise couverture vaccinale et ainsi d'intervenir plus rapidement à trouvé assez rapidement un terrain de développement. Il est apparu assez rapidement que d'autres usages de cette méthode pouvaient être développés au sein de la santé, dans une période où la question de démarche qualité pointée son nez. Ainsi en 2000, suite à plusieurs expériences de mise en œuvre et afin de pouvoir permettre le transfert de cette méthode en France et dans le domaine de la santé, nous avons publié l'article « *La technique de sondage par lots appliquée à l'assurance qualité (LQAS) : méthodes et applications en santé publique* » dans la revue d'épidémiologie et de santé publique » en tant que note pédagogique, offrant ainsi à cette méthode une diffusion plus simple auprès des professionnels de santé publique français (Jutand, 2000). Cet article consistait à présenter la méthode le plus didactiquement possible en l'illustrant d'exemples du champ de la santé. Il a conduit à notre participation dans la rédaction d'un second article « *Lot quality assurance sampling for screening communities hyperendemic for Schistosoma mansoni* » mettant en œuvre la méthode pour repérer rapidement les

zones endémiques à Madagascar du ver plat **Schistosoma mansoni** parasite de l'Homme vecteur de la bilharziose intestinale. (Rabarijaona, 2003)

Note pédagogique

La technique de sondage par lots appliquée à l'assurance qualité (LQAS) : méthodes et applications en santé publique

Lot quality assurance sampling: methods and applications in public health

M.-A. JUTAND, R. SALAMON

Institut de Santé Publique, d'Épidémiologie et de Développement, Université Victor Segalen Bordeaux 2, 146, rue Léo Saignat, 33076 Bordeaux. (Tirés à part : M.-A. Jutand)

Lot Quality Assurance Sampling (LQAS), developed to meet industrial quality control needs, has been applied to health surveys. The WHO used this method to assess immunization coverage. The sampling strategy was developed to classify lots as acceptable or unacceptable. Lot sampling is an efficient, simple and time-efficient procedure for quality assurance. Under certain conditions, efficiency can be improved with double sampling. We describe the method and its theoretical basis and illustrate applications of LQAS in epidemiological surveillance and quality control of medical records. The advantages and disadvantages of this method are presented.

LQAS. Sampling studies.

La méthode « lot quality assurance sampling » (LQAS) est issue du monde industriel et trouve de nombreuses applications dans le secteur de la santé publique. L'exemple classique de l'utilisation de cette méthode est proposé par l'OMS pour l'évaluation vaccinale. Il s'agit de discriminer différentes zones géographiques en deux groupes : celles jugées suffisamment vaccinées et celles jugées insuffisamment vaccinées. Cet échantillonnage par lots offre un outil de jugement de qualité, simple et peu coûteux en temps. L'utilisation du LQAS à deux degrés peut, sous certaines conditions, améliorer l'efficacité. Cet article se propose de rappeler les fondements de cette méthode et de présenter des exemples d'applications dans le domaine de la surveillance épidémiologique, mais aussi dans le cadre du Programme de Médicalisation du Système d'Information. Les avantages et difficultés de la méthode sont présentés.

LQAS. Échantillonnage.

INTRODUCTION

Dans certaines situations d'étude, il est nécessaire de recueillir l'information d'intérêt à partir d'échantillons de petite taille. Dans ce cas, les investigateurs ne sont pas intéressés par l'estimation d'un paramètre avec une certaine précision mais par la position possible de ce paramètre par rapport à une valeur de référence. La technique

de sondage d'évaluation de la qualité des lots (Lot Quality Assurance Sampling ou LQAS) est une méthode d'échantillonnage utilisée pour répondre à cet objectif. Le développement du LQAS est dû au monde de l'industrie, pour lequel il constitue un moyen économique de contrôle de qualité dans l'identification des lots de produits inacceptables, en un minimum de temps [1]. Cette problématique existe aussi dans le domaine

biomédical. Par exemple, la méthode du LQAS peut permettre de sélectionner les régions où la qualité de la couverture vaccinale est jugée insuffisante et où une intervention est donc nécessaire [2-4]. Ce travail a pour objectif de rappeler les principes fondamentaux du LQAS et d'en illustrer l'intérêt dans le secteur de la santé à travers quelques exemples.

PRINCIPE DE LA MÉTHODE DU LQAS

La méthode LQAS consiste à juger de la qualité d'un lot (est-il « acceptable » ou non) à partir de l'étude d'un échantillon de faible taille extrait aléatoirement de ce lot. Un lot sera qualifié d'acceptable s'il contient une proportion d'éléments défectueux inférieure à une valeur P_0 ; sinon il sera qualifié d'inacceptable. Cette proportion seuil P_0 est une référence fixée par l'utilisateur ou imposée (de manière réglementaire par exemple).

Le sondage effectué sur le lot ne cherche pas à estimer la fréquence des éléments défectueux qu'il contient mais simplement à juger si la valeur de référence P_0 est ou non dépassée.

Nous appellerons N la taille du lot, n la taille de l'échantillon extrait aléatoirement du lot et $(d^* + 1)$ le nombre d'éléments défectueux dont l'observation sur l'échantillon permet de rejeter le lot.

EXPOSÉ MÉTHODOLOGIQUE

DÉFINITION DES HYPOTHÈSES DE TEST

La méthode du LQAS est basée sur la réalisation d'un test dont l'hypothèse nulle est :

H_0 : le lot n'est pas acceptable, la proportion d'éléments défectueux dans le lot est supérieure ou égale à P_0 ,

l'hypothèse alternative étant :

H_a : le lot est acceptable, la proportion d'éléments défectueux dans le lot (P_d) est inférieure à P_0 .

Les conséquences liées aux conclusions du test sont présentées dans le *tableau I*. D'un point de vue pratique, il s'agira donc pour l'enquêteur d'observer les éléments un à un lors du tirage au sort et de déclarer, à partir de critères clairement établis *a priori*, si chacun est défectueux ou non. Dès que l'on atteint $(d^* + 1)$ éléments défectueux, on juge le lot inacceptable, même si l'effectif n n'a pas été inclus. Si le nombre d'éléments défectueux est inférieur ou égal à d^* après avoir étudié les n sujets de l'échantillon, le lot est jugé acceptable avec une probabilité d'erreur de conclusion inférieure ou égale à α . Le choix d'une valeur faible pour α implique de retenir les seuls lots ayant peu d'éléments défectueux, d'où une forte sélection et un risque élevé de refus à tort des bons lots.

FONDEMENTS THÉORIQUES

La distribution du nombre d d'éléments défectueux dans un échantillon de taille n tiré aléatoirement d'un lot de taille N , contenant une proportion P d'éléments défectueux, suit une loi hypergéométrique. Cette loi prend en considération le fait que le tirage est sans remise et ainsi, que la probabilité qu'un élément défectueux appartienne à l'échantillon change à chaque sélection de sujet.

Cette loi hypergéométrique est définie par :

TABLEAU I. — Probabilités associées aux différentes décisions possibles lors de l'application de la méthode LQAS.

Décision	Réalité	
	$H_0 : P \geq P_0$ (Lot non acceptable)	$H_a : P < P_0$ (Lot acceptable)
Le lot est jugé non acceptable	$1 - \alpha$ test sensible aux mauvais lots	β mauvaise décision = « risque pour l'industriel ou le décideur »
Le lot est jugé acceptable	α mauvaise décision = « risque pour le consommateur ou la communauté »	$1 - \beta$ test reconnaissant les lots convenables

$$\text{Probabilité (d = k)} = \frac{C_{NP}^k \times C_{N(1-P)}^{n-k}}{C_N^n}$$

$$\text{Probabilité (d ≤ k)} = \sum_{i=0}^k \frac{C_{NP}^i \times C_{N(1-P)}^{n-i}}{C_N^n}$$

$$\text{Espérance (d)} = n \times P$$

$$\text{Variance (d)} = n \times P \times (1 - P) \times \frac{N - n}{N - 1}$$

On considère que pour N grand et pour un taux de sondage de moins de 10 %, c'est-à-dire n considéré négligeable devant N, on peut réaliser une bonne approximation de la loi hypergéométrique par la loi binomiale de moyenne (nP) et de variance [n × P × (1 - P)].

L'approximation par une loi normale d'espérance (nP) et de variance $\left[n \times P \times (1 - P) \times \left(\frac{N - n}{N - 1} \right) \right]$

peut être acceptée lorsque N est grand, que P n'est ni trop voisin de 1 ni trop voisin de 0 (ces conditions seront jugées valides si P est supérieur à 0,10 ou inférieur à 0,90) et que n satisfait aux conditions nP et [n × (1 - P)] supérieures à 10 [5, 6]. Si une de ces approximations de la loi hypergéométrique est acceptable, les calculs des valeurs n et d* deviennent alors plus simples.

DÉTERMINATION DE LA TAILLE D'ÉCHANTILLON (n) ET DE LA LIMITE DE REJET (d*)

Il s'agit donc de définir le nombre maximum d'éléments devant appartenir à l'échantillon et le nombre d'éléments défectueux (d* + 1) de l'échantillon nécessaire à la décision du rejet du lot.

La détermination de d* doit satisfaire à la condition suivante : « La probabilité d'obtenir au plus d* éléments défectueux dans l'échantillon de taille n sous l'hypothèse nulle doit être inférieure ou égale à α ».

En pratique deux cas sont possibles, selon que n est choisi *a priori* pour des raisons de faisabilité ou que n est calculé à partir des valeurs α et β fixées *a priori* [4]. Les calculs nécessaires aux déterminations des valeurs de n et d* ont été effectués à partir du logiciel Excel.

PREMIER CAS : n FIXÉ A PRIORI

Cette situation est fréquemment rencontrée lors d'études multicentriques où plusieurs lots doivent être analysés à partir de ressources limitées. La taille n de l'échantillon d'un lot étant fixée *a priori*, on détermine la valeur d* à partir de la fonction de répartition du nombre d'éléments défectueux parmi n éléments. Cette détermination s'effectue simplement par calcul de la fonction de répartition de la loi hypergéométrique de paramètres (N, n, P₀). Lorsque N est inconnu et supposé grand, on effectue ce calcul à partir de la loi binomiale de paramètres (n, P₀).

Supposons que la population d'intérêt soit composée de lots de taille 150 (N) et que l'on souhaite déterminer avec un seuil d'erreur maximum de 5 % (α) les lots ayant moins de 50 % (P₀) d'éléments défectueux, l'effectif maximum observable dans chaque lot étant fixé à 10 éléments, la valeur de d* calculée est alors égale à 2. C'est-à-dire qu'un lot est rejeté dès que l'échantillon de dix éléments extrait aléatoirement possède au moins trois éléments défectueux. La probabilité de ne pas rejeter un lot qui cependant possède 50 % d'éléments défectueux est égale à 0,049 (tableau II). La valeur de d* est donc déterminée pour répondre au mieux à la contrainte liée au risque α.

TABLEAU II. — Détermination de la valeur seuil d* pour l'application de la méthode LQAS avec N = 150, n = 10 et P₀ = 0,5.

Nombre k d'éléments défectueux parmi 10 éléments	Probabilité d'observer au plus k éléments défectueux
0	0,001
1	0,009
2	0,049
3	0,163
4	0,373
5	0,627
6	0,837
7	0,951
8	0,991
9	0,999
10	1,000

DEUXIÈME CAS : n NON FIXÉ A PRIORI

Cette situation correspond au cas classique de détermination de taille d'échantillon lors de la comparaison d'une proportion observée à une valeur théorique en fonction de α , β , P_0 et P_a définis a priori. P_a correspond à la proportion maximum d'éléments défectueux que l'enquêteur accepte pour juger un lot de bonne qualité, cette valeur est bien sûr inférieure à P_0 . β est donc le risque de refuser des lots qui satisfont pourtant au critère de qualité défini ci-dessus, c'est-à-dire des lots ayant une proportion d'éléments défectueux inférieure ou égale à P_a .

Sous les hypothèses de convergence de la loi hypergéométrique vers la loi normale, on détermine la taille minimale nécessaire de sujets pour répondre à ces conditions puis on détermine la valeur de d^* à partir des formules suivantes :

$$n = \left(\frac{U_{2\alpha}\sqrt{P_0 \cdot (1-P_0)} + U_{2\beta}\sqrt{P_a \cdot (1-P_a)}}{P_0 - P_a} \right)^2$$

$$\text{et } d^* = nP_0 - U_{2\alpha} \cdot \sqrt{n \cdot P_0 \cdot (1-P_0)}$$

avec U_k représentant la valeur pour laquelle la probabilité qu'une variable aléatoire suivant une loi normale centrée réduite lui soit supérieure soit égale à $k/2$.

Si les hypothèses de convergence ne sont pas vérifiées, il est nécessaire d'effectuer les calculs à partir de la loi hypergéométrique afin de choisir le meilleur couple $(n ; d^*)$ répondant aux contraintes fixées.

Par exemple, lors d'une étude de qualité concernant des lots de taille 1000, on souhaite rejeter les lots ayant plus de 30 % d'éléments défectueux dans 95 % des cas et accepter les lots ayant moins de 10 % d'éléments défectueux dans 90 % des cas. Ce problème détermine donc les valeurs suivantes nécessaires aux calculs de n et d^* : $P_0 = 0,3$, $\alpha = 0,05$, $P_a = 0,1$ et $\beta = 0,1$ et $N = 1000$. En utilisant l'hypothèse de normalité, on obtient le couple (33 ; 6) et le couple (32 ; 5) en utilisant la loi hypergéométrique ; N étant grand, les résultats sont proches. Par contre, si la taille des lots est égale à $N = 100$, alors le couple déterminé à partir de la loi hypergéométrique devient (26 ; 4), tandis que le couple obtenu à partir de la loi normale est bien sûr non

modifié. La détermination du couple $(n ; d^*)$ dépend, comme nous l'avons vu, des valeurs des paramètres α , β , P_0 et P_a . Pour illustrer les calculs ci-dessus, on peut, pour chaque couple de valeurs $(n ; d^*)$ tracer la « courbe de puissance » des points $(P ; \gamma)$, où P représente la proportion d'éléments défectueux dans le lot et γ correspond à la probabilité d'avoir au plus d^* éléments défectueux dans un échantillon de taille n tiré aléatoirement de ce lot. Lorsque P a pour valeur P_0 , cette probabilité correspond au risque de première espèce, et si P est inférieur à P_0 cette probabilité correspond à la puissance du test (fig. 1).

À partir du tracé des courbes correspondant à différents couples $(n ; d^*)$, on peut sélectionner celles répondant au mieux aux valeurs α et β fixées a priori, c'est-à-dire passant au plus près des points $(P_0 ; \alpha)$ et $(P_a ; \beta)$. Par exemple pour les valeurs $P_0 = 0,6$ et $\alpha = 5\%$ pour une population de taille 2000, on remarque que parmi les couples proposés celui offrant la meilleure puissance est (45 ; 21)

EXEMPLES

Les exemples suivants vont illustrer l'intérêt de cette méthode dans le secteur de la santé pour lequel les éléments étudiés sont généralement des individus. Un exemple classique d'utilisation de la méthode du LQAS est l'évaluation de la couverture vaccinale de certains sites géographiques. L'Organisation Mondiale de la Santé a comparé cette méthode à la méthode du tirage en grappes utilisée fréquemment par les programmes de vaccination, lors d'une enquête à Bobo Dioulasso au Burkina Faso [2]. Les lots considérés comprennent des enfants justiciables de la vaccination dans un lieu déterminé. Un seuil élevé et un seuil bas de non-couverture vaccinale sont déterminés correspondant respectivement aux valeurs de P_0 et P_a . Dans cette problématique, les enquêteurs souhaitent être en mesure de détecter avec une forte probabilité les sites ayant une bonne couverture vaccinale pour pouvoir intervenir médicalement sur les tous les autres sites. Les seuils, ainsi que les risques de première et seconde espèces, choisis pour calculer les valeurs du couple $(n ; d^*)$ satisfaisant à cette problématique, sont : $P_0 = 0,6$, $\alpha = 0,10$,

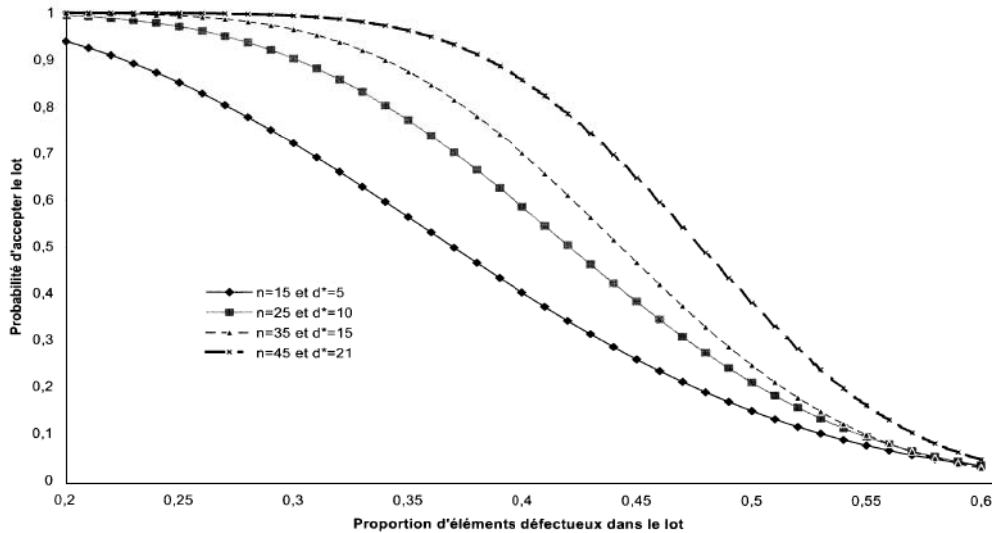


FIG. 1. — Courbes de puissance pour divers couples (n ; d^*) avec $P_0 = 0,6$ et $\alpha = 0,05$ pour des lots de taille 2000.

$P_a = 0,15$ et $\beta = 0,02$, les valeurs obtenues sont 11 pour n et 4 pour d^* . Il s'agit donc pour les enquêteurs de définir les zones géographiques qui détermineront les lots et de sélectionner aléatoirement sur chaque lot de manière successive au maximum 11 sujets. Dès que 5 sujets seront notés comme mal vaccinés, l'enquêteur classera la zone comme mal vaccinée et interviendra pour effectuer une vaccination. L'utilisation de la méthode LQAS a permis de gagner du temps sur le terrain par rapport à une enquête de programme élargi de vaccination (PEV) effectuée sur 30 grappes ; elle a de plus permis une couverture de l'ensemble de la ville. Le choix de cette méthode est pertinent dans la mesure où il y a une forte hétérogénéité des taux de vaccination entre les différentes zones géographiques du territoire considéré.

Une application similaire a été effectuée à Madagascar en 1998 dans le cadre du Programme National de Lutte contre la Bilharziose [7]. Il s'agissait d'identifier au moindre coût des communautés hyperendémiques d'enfants de 5 à 14 ans. Pour dépister ces localités il est nécessaire d'enquêter dans chaque communauté car la transmission du schistosome dépend des caractéristiques spécifiques de l'environnement, comme

les points d'eau utilisés par la population et la présence de mollusques qui constituent l'hôte intermédiaire du parasite. Il est matériellement impossible de déployer des méthodes classiques de sondage en raison de l'effectif nécessaire de l'échantillon et du coût que ceci entraînerait. L'objectif de ce travail était d'évaluer la capacité de la méthode LQAS à bien reconnaître les foyers et d'étudier la possibilité de déléguer l'échantillonnage au personnel de terrain, en comparant les informations exhaustives obtenues dans 38 écoles à des échantillons aléatoires obtenus sur ces données. Les zones peuvent être hyperendémiques (prévalence supérieure à 60 % = P_0), mésoendémiques (prévalence entre 30 % et 60 %) ou hypoendémiques (prévalence inférieure à 30 % = P_a). Des échantillons de taille 16 ont été constitués à partir des données des 38 écoles, la valeur de d^* satisfaisant aux conditions $P_0 = 0,6$, $\alpha = 0,10$, $P_a = 0,30$ et $\beta = 0,2$, vaut 6. En comparant les résultats de la méthode LQAS aux résultats obtenus de manière exhaustive, 8 % des zones hyperendémiques et 20 % des zones hypoendémiques ont été mal classées. Dans le même ordre d'idée, la répartition du virus HTLV 1 est extrêmement hétérogène à l'intérieur d'un même pays et il peut être intéressant de

rechercher des foyers de forte prévalence. C'est le cas dans une étude au Bénin [8] où nous appliquons actuellement la méthode LQAS à l'échelle de toute une région.

MÉTHODE LQAS À DEUX DEGRÉS

La méthode LQAS peut être perfectionnée dans les situations de forte variabilité de la qualité des lots en effectuant un échantillonnage en deux étapes. Dans un premier temps on détermine, par la méthode LQAS à un degré, les valeurs n et d^* en fonction des contraintes définies *a priori* P_0 , α , P_a et β .

Pour détecter rapidement les lots ayant une proportion très faible ou très forte d'éléments défectueux, on peut définir une étape intermédiaire permettant une première discrimination des lots à partir d'échantillons aléatoires de taille inférieure à n mais supérieure à d^* , soit n_1 . Le choix de n_1 doit permettre un gain de temps suffisamment important dans le déroulement de l'étude. La détermination de la valeur seuil d^*_1 est effectuée comme dans la méthode LQAS à un degré pour une valeur du risque de première espèce α_1 inférieure à celle de la première étape. α_1 doit être choisi très inférieur à α ; on peut suggérer que pour un α égal à 0,05 le choix de 0,01 pour α_1 est convenable.

L'enquêteur effectue dans un premier temps un échantillonnage de n_1 éléments; si le nombre d'éléments défectueux d est inférieur ou égal à d^*_1 le lot est jugé acceptable; si d est supérieur à d^*_1 , le lot est refusé, sinon il est nécessaire de poursuivre l'échantillonnage.

La deuxième étape consiste à poursuivre l'échantillonnage tant que le nombre d'éléments défectueux ne dépasse pas d^* , jusqu'à l'inclusion

de n éléments. On conclut que le lot est inacceptable lorsque $(d^* + 1)$ éléments défectueux sont observés, dans le cas contraire on considère que le lot est acceptable (*tableau III*).

Par exemple, on souhaite déterminer, dans une population structurée en lots d'environ 1 000 éléments chacun, les lots ayant plus de 30 % d'éléments défectueux avec un risque d'erreur α de l'ordre de 5 % et les lots ayant moins de 10 % de mauvais éléments (avec une puissance de test de l'ordre de 90 %). L'enquêteur décide d'utiliser la méthode LQAS à deux degrés car une forte hétérogénéité entre les lots est prévisible. Il détermine tout d'abord le couple $(n; d^*)$ correspondant à la méthode LQAS à un degré et obtient $n = 32$ et $d^* = 5$. Puis les valeurs du couple $(n_1; d^*_1)$, correspondant à la première étape d'échantillonnage, sont calculées pour que le risque de première espèce α_1 associé soit inférieur ou égal à 0,01 (valeur choisie inférieure à α afin de minimiser le risque d'erreur d'acceptation du lot à tort); on obtient alors $n_1 = 20$ et $d^*_1 = 1$. Si l'enquêteur, parmi les 20 premiers éléments sélectionnés, observe au plus 1 élément défectueux il acceptera le lot; s'il en observe plus de 5 il le rejettera. Sinon il continuera l'échantillonnage jusqu'à l'inclusion maximum de 32 éléments et le lot sera considéré acceptable si au plus 5 éléments défectueux ont été observés.

DISCUSSION

La méthode du LQAS est principalement utilisée afin de déterminer, rapidement et à moindre coût dans une population définie, des groupes d'éléments ou de sujets (lots, zones géographi-

TABLEAU III. — *Décision de rejet ou non du lot aux différentes étapes de la méthode du LQAS à deux degrés, en fonction du nombre d'éléments défectueux observés.*

Étape	Nombre d'éléments défectueux observés		
	Inférieur ou égal d^*_1	Entre d^*_1 et d^*	Supérieur ou égal à $d^* + 1$
1) on observe au plus n_1 éléments	Lot jugé acceptable	Pas de décision; il faut procéder à la seconde étape	Lot jugé inacceptable
2) on observe au plus n éléments ($n > n_1$)		Lot jugé acceptable	Lot jugé inacceptable

ques...) risquant de ne pas répondre à des critères d'acceptabilité.

Les applications en santé publique les plus classiques de la méthode du LQAS concernent la surveillance épidémiologique des maladies transmissibles dans les pays en développement.

Néanmoins, il existe d'autres possibilités d'utilisation. Par exemple, nous avons pu vérifier l'intérêt de cette méthode pour le contrôle de qualité de données dans le cadre du programme de médicalisation du système d'information (PMSI) au sein du service d'information médicale (SIM) du centre hospitalier universitaire de Bordeaux.

Actuellement, la seule obligation en matière d'analyse de la qualité du recueil est l'analyse de 100 séjours tirés au sort chaque semestre (un tirage au sort et un seul réalisé par une procédure informatique nationale) ; les dossiers sont codés en insu par le SIM pour son propre établissement. Le but de cette démarche est de tester le caractère de sur- ou de sous-cotation en points ISA (Indice Synthétique d'Activité), ce qui permet l'application d'un facteur de correction par la tutelle en plus ou moins, lors de l'arbitrage budgétaire définitif. Cette méthode ne permet pas l'analyse interne de l'établissement, en particulier des différents services. La qualité du recueil de données doit être régulièrement surveillée et le sondage LQAS peut être intéressant par sa rapidité et sa facilité d'application, ce qui permet l'application de cette méthode pour chaque service in situ et régulièrement.

Il s'agit donc de tirer au sort un échantillon de dossiers par service et de réviser, par exemple, pour un séjour donné le RUM (Résumé d'Unité Médicale), le Diagnostic Principal, la CMD (Catégorie Majeure de Diagnostic) et le GHM (Groupe Homogène de Malades), puis de confronter ces nouvelles mesures avec celles écrites sur le dossier. Un lot est ici défini comme un ensemble de dossiers traités par un service donné sur une période de 6 mois. On considère possible, dans un premier temps, d'analyser 9 dossiers-patient par lot. Dans cet exemple, P_0 représente la proportion de dossiers de bonne qualité et est fixée à 0,60 ; α est fixé à 10 %. Sous ces conditions, le seuil de bons dossiers en dessous duquel on considérera que le service

n'est pas satisfaisant est de 4. Par cette méthode, on discrimine principalement les très mauvais services jugés peu satisfaisants quant à la qualité de leurs dossiers. Ce type d'analyse peut être effectué en routine par le SIM pour évaluer régulièrement la qualité des données des services en complément de l'échantillon demandé au niveau national. L'exemple ci-dessus souligne la ressemblance évidente entre la méthode LQAS et le sondage stratifié ; il s'agit, dans les deux cas, de découper la population en groupes et d'effectuer un échantillonnage dans chaque groupe. L'efficacité de ces deux méthodes est de plus améliorée lorsque les lots ou strates sont des parties homogènes de la population mais hétérogènes entre elles. Cependant un sondage stratifié permet d'effectuer des estimations, ce qui n'est pas le cas généralement de la méthode LQAS puisqu'on peut rejeter un lot sans avoir inclus les n sujets prévus initialement par le calcul de taille nécessaire. Si l'on souhaite estimer un paramètre de la population il faut alors, pour chaque lot, examiner les n sujets et regrouper les informations en tenant compte de la répartition des lots dans la population comme dans le cas d'un échantillon stratifié.

L'attrait de cette méthode est bien évidemment l'utilisation d'échantillons de faible taille. Les échantillons ont donc finalement des tailles variables selon les lots mais toujours inférieure ou égale à la valeur n calculée au départ. La décision de rejeter le lot pouvant intervenir après chaque inclusion, il est indispensable que la sélection de chaque sujet se fasse de manière aléatoire, rendant ainsi l'échantillon représentatif à chaque inclusion. Cette condition est souvent difficile à respecter et est une limite à ne pas négliger lors de l'application de la méthode LQAS.

En conclusion, on peut retenir que la méthode LQAS est bénéfique en coût et en temps et facile à appliquer, mais nécessite quelques mesures de précautions. Son application est intéressante afin de discriminer rapidement des groupes en « acceptable ou non » à partir de seuils définis *a priori* ; le gain obtenu par cette méthode sera d'autant plus important qu'il y a une forte hétérogénéité entre les lots.

RÉFÉRENCES

1. Reinke WA. Applicability of industrial sampling techniques to epidemiologic investigations: examination of an underutilized resource. *Am J Epidemiol* 1991; 134: 1222-32.
2. World Health Organization. Expanded Programme on Immunization (EPI) Lot Quality Assurance Survey to assess immunization coverage. *Wkly Epidemiol Rec* 1995; 70: 261-4.
3. Lanata CF, Black RE. Lot quality assurance sampling techniques in health surveys in developing countries: advantages and current constraints. *World Health Stat Q* 1991; 44: 133-9.
4. Lemeshow S, Taber S. Lot quality assurance sampling: single and double sampling plans. *World Health Stat Q* 1991; 15: 71-81.
5. Abboud N, Audroing JF. Convergences stochastiques. In : *Probabilités et inférence statistique*. Paris : Nathan, 1989 : 105-23.
6. Saporta G. Convergences des suites de variables aléatoires. In : *Probabilités analyse des données et statistique*. Paris : Technip, 1990 : 57-64.
7. Rabarijaona LP. Validation de l'emploi des méthodes LQAS pour l'identification des communautés cible du Programme National de Lutte contre la Bilharziose à Madagascar. [Mémoire de D.E.A. d'Épidémiologie et Intervention en Santé Publique]. Bordeaux : Institut de Santé Publique, d'Épidémiologie et de Développement, 1998.
8. Houinato D, Verdier M, Preux N. Intra familial clustering and 4-years follow-up of asymptomatic human T-cell leukemia virus type I (HTLV 1) infection in Benin (West-Africa). *Int J Epidemiol* 1998; 27: 146-52.

Ainsi au-delà du savoir produit, les modifications d'usage induites par les changements d'environnement peuvent engendrer des modifications d'activités et pratiques, voire de métiers ou tout au moins impliquer des réajustements d'activités.

4. Catégorisation de statisticiens

L'évolution de la présence de la statistique dans de nombreux domaines a impliqué un développement important du nombre d'utilisateurs de cette science. Tous ne sont pas des statisticiens professionnels, et le degré d'usage d'éléments, d'objets de technique statistique ou plus exactement de la dynamique d'un esprit statistique est très variable. Cependant, le bon usage nécessite une connaissance du cadre d'usage de l'outil, et une habileté dans sa mise en œuvre selon l'habitat considéré.

L'académie des sciences repérait dans son rapport quatre grandes catégories de statisticiens (Malliavin, 2000) :

- les chercheurs en statistique, dont le rôle consiste à la création/production de résultats scientifiques nouveaux, sont principalement originaires des mathématiques et de l'informatique pour les développements méthodologiques,
- les statisticiens spécialistes – biostatisticien, statisticien actuariaire ..., qui doivent s'approprier la connaissance scientifique du domaine d'application, la statistique n'étant pas la finalité de leur travail,
- statisticiens officiels, ayant pour rôle le recueil, la gestion et l'interprétation et dont les missions sont essentielles pour l'exercice gouvernemental,
- les ingénieurs statisticiens et statisticiens d'entreprise qui utilisent des méthodes statistiques existantes dans un contexte professionnel.

Il est assez intéressant de souligner que dans cette proposition en quatre catégories, l'enseignant de statistique n'apparaît pas. Que peut suggérer ce manque ? L'idée que l'enseignement de la statistique n'est pas forcément assuré par

des statisticiens ? Cette situation est d'ailleurs celle actuelle de l'enseignement de la statistique dans le secondaire où l'enseignement de statistique est assuré par les professeurs de mathématiques. Nous ne reviendrons pas sur la controverse qui a animé les différentes communautés et nourris de nombreuses publications « qui doit enseigner la statistique ? »

La statistique comme cela a déjà été largement dit est une discipline qui trouve tout son sens dans l'application, que ce soit pour initier de nouvelles théories, et appliquer avec adaptations des théories déjà existantes dans d'autres situations. Des domaines tels que l'agronomie-biologie, la médecine – épidémiologie, l'économie ou les sciences de l'assurance de par l'existence soit de bases de données très denses, mais aussi de recherche d'associations et de causalité ont été particulièrement adaptées au développement de méthodes spécifiques. Tant est si bien que le métier de statisticien s'est décliné en spécialités, présentées sur le site de la société savante de statistique française SFdS¹³. L'approche écologique de la didactique de la statistique trouve dans ces éléments toute sa place. Il y a les bio-statisticiens, (chargés de collaborer avec les cliniciens pour tous les essais cliniques depuis la mise au point du protocole jusqu'à l'analyse des données), les actuaires (au cœur du métier d'assureur : ils évaluent les risques et examinent les conséquences financières des comptes futurs), les fiabilistes (ils doivent comprendre comment ça marche et comment ça tombe en panne, modéliser et calculer les probabilités d'occurrence des événements non désirés, identifier et hiérarchiser les scénarios et les actions les plus efficaces en terme de rentabilité c'est le statisticien industriel)... Communiquer sur cette richesse des métiers dans le domaine de la statistique a été un enjeu pour la SFdS afin d'attirer les jeunes générations, il s'agissait de leur présenter la science statistique non pas comme un domaine uniquement attaché aux mathématiques mais bien en interaction avec l'ensemble des domaines que ce soit de la météorologie aux finances. En 2011, l'ONISEP et la SFdS publient une brochure intitulée « *Les métiers de la statistique : un secteur à découvrir, des portraits de*

¹³Site web SFdS http://www.sfds.asso.fr/55-Quelques_metiers_pour_les_statisticiens consulté le 25 aout 2015

professionnels, les stratégies de formations. »¹⁴, et associent à cette diffusion des présentations vidéos accessibles sur le site. La diffusion de cette ressource est large, auprès des lycéens mais aussi des enseignants du secondaire. La sensibilisation de la place de la statistique dans le devenir citoyen et professionnel est ainsi activée et est accompagnée parallèlement à la modification des programmes de lycée qui développe une place non négligeable à l'enseignement de la statistique.¹⁵

Il est, cependant, aussi important de concevoir que la statistique n'est pas uniquement une science développée par les statisticiens mais elle est un art à l'usage d'un ensemble d'acteurs statisticiens ou non, ce qui en fait une réelle difficulté. Il s'agit donc comme nous le verrons dans le chapitre suivant de poser le cadre devant être partagé pour un usage correct de la statistique. L'usage de la statistique nécessite une adhésion aux règles communes ainsi définies pouvant dès lors être associées à des spécificités de niches.

¹⁴ Accessibilité à la plaquette - <http://www.sfds.asso.fr/images/zoom-statistique-2011.pdf> consulté le 25 aout 2015

¹⁵ Notre participation à la rédaction du document « Ressources pour la classe terminale générale et technologique « Probabilités et statistique », de février 2012, MENJVA/DGESCO nous a permis de concevoir la difficulté pour les enseignants de mathématiques du secondaire de s'approprier l'esprit statistique, et l'importance de leur proposer des situations contextualisées pour introduire l'ensemble des notions.

D. Synthèse

L'enseignant chercheur, de par les quatre dimensions de sa tâche, se trouve confronté à la nécessité de développer une plasticité lui permettant de travailler des capacités de communication, de diffusion, de transfert et d'accompagnement à l'apprentissage dans différents domaines pour lesquels son positionnement passera d'enseignant à expert. Les situations didactiques diffèrent et à chaque situation de nouveaux enjeux apparaissent car il est dans toutes les situations celui qui sait. Dans ce cadre général, il est ainsi important de repérer les spécificités disciplinaires qui peuvent accompagner la transposition de savoirs statistiques, ceux sont ces notions que nous allons explorer dans la suite. Le double rôle de l'enseignant-chercheur l'oblige à savoir diffuser son savoir au sens savoir savant mais aussi savoir expert dans des environnements très différents. Cela implique qu'il réussisse à dépersonnaliser son savoir et le contextualiser en prenant en considération les environnements et les situations didactiques qu'il rencontre dans toutes leurs différences disciplinaires et d'enjeux.

CHAPITRE 2 – LA STATISTIQUE : UNE DISCIPLINE BIEN SPECIFIQUE

A. La reconnaissance disciplinaire de la statistique

1. "Statistique" : de quoi parle-t-on ?

On sait depuis Wittgenstein (1961) que le sens des mots n'est pas dans leur signification mais dans l'usage qui en est fait selon les communautés. Que révèle le mot "statistique"? En lisant un panel de citations de personnalités sur la thématique de la statistique, nous pouvons avoir une idée de ce que chacun fait porter à ce terme et de la variabilité des jeux de langage auxquels il participe.

«La statistique a démontré que la mortalité dans l'armée augmente sensiblement en temps de guerre.» (Alphonse Allais (1854 – 1905) Journaliste, écrivain et humoriste français.)

« La raison d'être des statistiques, c'est de vous donner raison. » (Abe Burrow (1910 – 1983), humoriste, américain.)

« Il y a trois sortes de mensonges : les mensonges, les sacrés mensonges et les statistiques. » (Mark Twain (1835 – 1910), écrivain et humoriste, américain)

« Les statistiques sont vraies quant à la maladie et fausses quant au malade ; elles sont vraies quant aux populations et fausses quant à l'individu. » (Léon Schwartzberg (1923 – 2003), cancérologue, français, ministre délégué de la santé en 1988)

« Dans toute statistique, l'inexactitude du nombre est compensée par la précision des décimales. » (Alfred Sauvy, (1898 – 1990), économiste, démographe et sociologue français, il fut le premier directeur de l'INED jusqu'en 1962.)

« La statistique est la première des sciences inexactes. » (Edmond et Jules de Goncourt, (Edmond – 1822-1896 ; Jules – 1830 – 1870), écrivains français, co-fondateurs de l'Académie Goncourt.)

De son singulier à son pluriel, de sa validité à sa fiabilité et jusqu'à ses usages, la statistique interroge, inquiète mais semble cependant bien présente dans l'esprit de tous qu'ils soient scientifiques, politiques, écrivains ou humoristes. Comment aborder ce domaine, qui semble si éloigné du cadre classiquement attendu dans le champ des sciences, et, tout particulièrement, des sciences

mathématiques, c'est à dire un espace de vérité, d'exactitude et d'objectivité ? Pourquoi a-t-on un tel retard en France pour l'apprentissage de la statistique, alors qu'elle fait partie depuis bien longtemps de la culture anglo-saxonne ?

Interrogeons les jeux de langage associés au mot "statistique".

2. Des données à la démarche

L'Académie de sciences introduit son rapport de 2000 sur « la statistique » par l'encadré ci-dessous :

Les statistiques sont des dénombrements de sujets, d'objets, d'événements dans une population ou des sous-populations. La statistique est une démarche permettant de recueillir, de traiter et d'interpréter les données qu'on recueille dans divers domaines où celles-ci présentent une caractéristique essentielle : la variabilité. Ainsi en est-il par exemple dans le domaine de la santé, où les individus diffèrent les uns des autres. Aboutir à une conclusion dans le domaine de la science ou de la décision à partir de cas tous particuliers est une gageure qui nécessite bien une démarche spéciale. C'est le rôle de la statistique. Il en est de même notamment en agronomie dans l'industrie ou l'économie.

La statistique peut certes porter sur des données récoltées par des statistiques et celles-ci peuvent prêter à l'utilisation de la statistique, mais les deux démarches sont essentiellement différentes.

Les statistiques existent depuis des siècles, la statistique par contre repose sur un mode de pensée original qui ne s'est réellement développé qu'à partir du XIXe siècle. (Malliavin, 2000)

La différence de sens, proposée ci-dessus, entre le nom commun statistique au singulier et au pluriel correspond, comme cela est indiqué, au passage historique d'une statistique initialement basée sur le recueil et dénombrement de données, mais sans développement spécifique, à un mode de pensée qui n'a connu son essor qu'en se fondant principalement sur le développement de la théorie des probabilités¹⁶.

Cette distinction entre le pluriel et le singulier est acceptée par la communauté de statisticiens francophones, mais n'a pas de sens en anglais, le terme "statistics" étant utilisé pour les deux acceptions.

Desrosières, dans son ouvrage « *La politique des grand nombres* », précise les perceptions possibles concernant la diversité de ces mots.

¹⁶ Il y a un terme statistique au singulier qui n'est pas indiqué dans la formulation de l'Académie des Sciences, il s'agit du terme employé pour représenter une mesure obtenue à partir d'une fonction mathématique définie à partir des données d'un échantillon. On parlera ainsi par exemple de la statistique du test de Student.

«La diversité des usages des mots « statistique » et « statistiques » reflète la tension entre les points de vue de réalité et de méthode. Pour les uns, c'est une activité administrative d'enregistrement de données diverses, conduisant à des chiffres incontestables repris par le débat social et orientant l'action. Pour les autres, c'est une branche des mathématiques, enseignée à l'université et utilisée par d'autres scientifiques : biologistes, médecins, économistes, psychologues ». (Desrosières, 2000, 21)

Nous pouvons souligner que ces deux perceptions sont associées à des dualités : administrative *versus* académique, résultats incontestables *versus* sciences en association forte avec les probabilités, intégrée fortement à la société et à la notion de décision *versus* présente pour un usage disciplinaire.

Considérer la statistique comme activité administrative correspond au prolongement historique de l'usage des statistiques dans le cadre d'une science d'aide à la décision. Cette statistique-là niche principalement au sein des instituts nationaux, tels que l'INSEE, l'InVS, l'INED. La vision académique de la statistique, quant à elle, correspond plus à un développement d'une branche des mathématiques à destination de l'analyse des données, de types mathématiques appliquées, mise au service des autres disciplines. Cependant, cette définition a maintenant 15 ans et l'évolution croissante des usages de la statistique tend à l'éloigner des mathématiques et à la rapprocher de l'informatique. L'indépendance prise par la statistique en tant que discipline est de plus en plus marquante dans les 10 dernières années.

Considérons, dans un premier temps plus spécifiquement, le substantif « statistiques ». Il correspond aux données construites à partir des données recueillies dans un contexte bien précis, fondées principalement sur des étapes de dénombrement mais pas uniquement. Les statistiques peuvent ainsi correspondre aux premiers résultats descriptifs obtenus à partir d'une étude, d'un recueil : moyenne, écart type, coefficient de corrélation ... Ces résultats ne peuvent être commentés, discutés et donc présentés qu'avec la connaissance du cadre de recueil et ceci fait dire à Deheuvels que « *les statistiques englobent la collecte et la gestion des grands fichiers d'observations* » (Malliavin, 2000). Il s'agit d'un message important soulignant le non-sens de parler de statistiques uniquement pour des résultats qui seraient totalement externalisés de leur cadre de recueil. Pourtant, cette

situation est des plus fréquentes ; nous pourrions prendre comme exemple des statistiques épidémiologiques telles que celles vous indiquant que la prévalence du diabète en France est de 4,7% en 2013 sans vous préciser de quelle population il s'agit, ni si cette prévalence a été déterminée à partir de données déclaratives, de mesures biologiques ou de données issues des bases de données de tous les régimes d'assurance maladie. Cette définition des statistiques pourrait ainsi être comprise comme étant le couple comprenant l'information des métadonnées et le fichier de données ainsi que les premiers indicateurs descriptifs. Elle souligne un risque important associé parfois à la diffusion de fichiers de données sans éléments d'information précisant le contexte même du recueil mais aussi les étapes de gestion ainsi que l'ensemble des informations concernant les outils de recueil¹⁷.

Le terme statistique, au singulier, est plus spécifiquement associé au processus complet du recueil à l'analyse et l'interprétation des données. Il ne s'agit donc plus uniquement de résultats, mais bien de l'ensemble du raisonnement et des phases d'actions qui devront être mises en œuvre pour satisfaire l'obtention de résultats appropriés. Régnier insiste sur le positionnement bicéphale de la statistique entre science mathématique et statistique appliquée, et souligne que la statistique peut être vue :

« Comme un domaine scientifique qui se développe dans une tension dialectique entre la statistique mathématique et la statistique appliquée à [...] Cette relation dialectique est elle-même en tension dialectique avec les statistiques au sens des données construites. » (Régnier, 2003)

Il perçoit donc l'intérêt de considérer la statistique comme un tout composé de deux dimensions, l'une mathématique, basée dans sa plus grande partie sur l'approche mathématique-probabiliste, et l'autre appliquée pouvant être déclinée selon différents environnements disciplinaires. Sa conception d'une approche conjointe de la statistique mathématique et de la statistique appliquée est présentée dans le schéma ci-dessous. Ces deux approches sont complémentaires et peuvent surtout s'interpeller lors de différentes phases du raisonnement scientifique. Il ne

¹⁷ L'analyse de fichiers de données statistique sans aucune connaissance du contexte de recueil ne peut devenir qu'un pur exercice de manipulation mais ne permet pas une analyse statistique contextualisée. Le découpage ainsi proposé par Deheuvels scinderait ainsi la démarche statistique en deux éléments phares : le recueil et la conservation de l'information et la phase d'analyse dans laquelle l'étape d'interprétation serait intégrée.

s'agit pas de considérer ces deux dimensions comme s'excluant, mais bien d'en comprendre la complémentarité et la nécessité de leur existence mutuelle. La capacité de transmission entre les deux environnements nécessite un travail collaboratif avec une connaissance mutuelle des compétences et cadres d'activités de chacun. Il s'agit de l'enjeu d'une bonne mise en application des modèles et du respect des usages de méthodes en adéquation avec l'environnement. Il s'agit ainsi de faciliter le transfert de la problématique formulée dans un environnement précis à partir le plus souvent de données à un modèle, en facilitant le passage du concret à l'abstrait, puis de pouvoir transposer l'information dans la situation concrète sans en perdre la cohérence et la scientificité de la mise en œuvre.

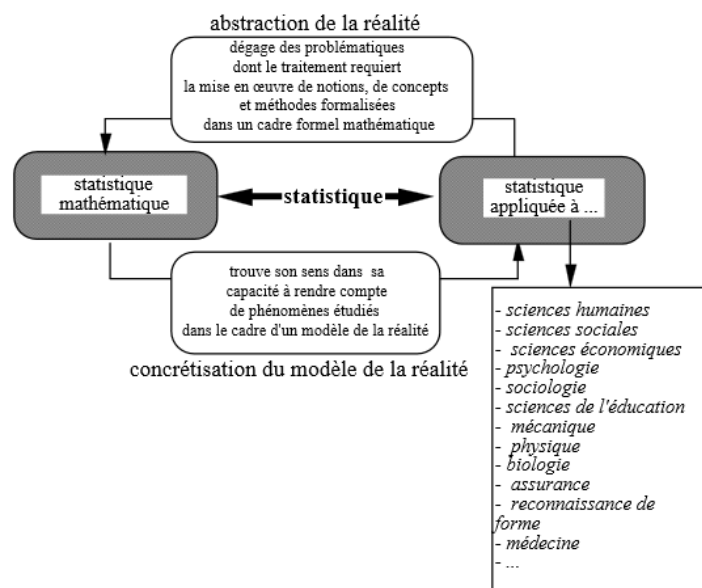


Schéma 3 : Conception de la statistique selon Régnier (Régnier, 2003)

Les tensions dialectiques, signalées par Régnier, ont cours depuis de nombreuses années et mettent ainsi en évidence qu'il ne s'agit pas uniquement d'une querelle de mots mais bien de conception même de la notion de la discipline et de ce qu'elle couvre, ainsi que de sa structuration. Comment, dans ce cas, imaginer que ces multiples conceptions peuvent ne pas avoir un impact sur la représentation, que chacun se fait, de ce qu'est ou n'est pas la statistique ? Peut-on alors envisager la statistique comme une discipline, et de quel type de discipline s'agit-il ?

3. La statistique : une discipline

La question de la place de cette discipline a été posée de nombreuses fois et depuis longtemps. Ainsi Penglaou répondait à la question « *la statistique existe-t-elle en tant que discipline autonome ?* » (Penglaou 1937), en proposant la définition suivante : « *une discipline, ou un faisceau de méthodes, pour la détermination sur le plan de la connaissance approchée de certains ensembles ou masses* ». La statistique n'est pas uniquement un ensemble de méthodes mais se base sur « *une logique cohérente, une base scientifique, mathématique et probabiliste solide* » (Dodge 1993). Selon le rapport de la Commission de Recherche sur l'Enseignement des Mathématiques :

« La statistique traite de données expérimentales ou d'observation, à étudier dans leur contexte (data with contexts) : sa spécificité est d'établir des liens entre ces données et la théorie mathématique des probabilités, d'expliquer ainsi le passé et de prévoir l'avenir. » (Kahane, 2000, 51)

C'est pour cette raison que la statistique nécessite une façon différente de raisonner, car les données chiffrées ne représentent pas uniquement un nombre mais portent avec elles les données du contexte auquel elles appartiennent ; elles sont nombre et sens comme le soulignent Cobb et Moore (Cobb & Moore, 1997).

Ils définissent la statistique comme « *a methodological discipline. It exists not for itself but rather to offer to other set of ideas and tools for dealing with data* ». Il s'agit donc d'une discipline de développement méthodologique ayant comme raison d'être de développer des objets et des démarches pour analyser des données.

Le lien, très fort, de la statistique avec les mathématiques est ainsi souligné, cependant la statistique a su se constituer comme discipline, à part entière, au cours du siècle dernier, ayant su gagner ses lettres de noblesse dans des domaines et disciplines aussi variés que la médecine (tout particulièrement avec le développement de l'épidémiologie), l'économie (développement de l'économétrie), les finances (développement de l'actuariat), ou encore la biologie (développement de la biostatistique). Les liens initialement très forts entre statistique et probabilités ont tendance à s'estomper, le rapprochement statistique – mathématiques est pour

cette raison de moins en moins central pour les domaines appliqués. Tout cela suggère des modifications considérables du métier de statisticien, avec une place de plus en plus importante de l'informatique. Ces transformations ont un impact non négligeable sur l'organisation même des formations de statisticien au sein desquelles la place de l'informatique est de plus en plus présente ; le développement exponentiel de méthodes permettant la gestion et l'analyse des Big Data en est l'exemple le plus frappant.

La pluralité des domaines d'usage de la statistique questionne la capacité de diffusion appropriée des méthodes et la qualité potentielle des usages. Ainsi Deheuvels va jusqu'à indiquer dans le rapport « La statistique » :

L'emploi systématique de telles méthodes statistiques¹⁸ par des spécialistes d'autres disciplines est propre à engendrer erreurs et confusions dans le cas où celles-ci sont appliquées en dehors d'une interface constante avec des statisticiens ». (Deheuvels, 2000)

Comme pour toute discipline, il ne peut y avoir de bons usages que si les connaissances et les habiletés concernant la science statistique sont suffisantes. L'article de Ioannidis « Why most published research findings are false » a pointé comme causes principales de fausses conclusions dans les articles publiés en médecine le manque de compétences en statistique des auteurs (Ioannidis, 2005). Brousseau souligne d'ailleurs en présentation de son dossier 13 « Les Cours de Statistique du COREM » sur son site internet¹⁹, la nécessité de connaissances suffisantes pour user de la statistique

« En effet, ces méthodes (ndlr statistiques) ne valent que dans la mesure où les connaissances de ceux qui les utilisent sont adéquates. De même qu'une boussole indique la direction du nord mais ne garantit pas d'arriver où on veut aller, les statistiques sont des clés inutiles à qui ne possède pas la serrure qu'elles ouvrent. »

Cette question concernant le manque de compétence et d'habileté de certains usagers de la statistique est ainsi régulièrement soulignée par les statisticiens reconnus (Fine, 2012). Deheuvels souligne, dans sa contribution au

¹⁸ Statistiques bayésiennes et fiduciaires.

¹⁹ <http://guy-brousseau.com/category/5dossiers-thematiques/>

rapport « La statistique », l'importance d'une proximité avec les statisticiens, et cette proximité doit aussi être intellectuelle, ce que nous appellerons ultérieurement « l'esprit statistique ». Tout ceci souligne la nécessité que les usagers de la statistique ne la perçoivent pas uniquement comme une boîte à outils. Les éléments caractéristiques et forts de la statistique doivent être connus par les usagers, afin de concevoir la mise en œuvre cohérente de démarches statistiques et de se porter garant d'interprétations correctes à l'issue des analyses. Nous repérons dans la suite les éléments connus et présentés comme les éléments fondamentaux pour la praxis de la statistique.

B. Le cadre de la statistique

Ce qui caractérise la spécificité de la statistique selon Deheuvels (Deheuvels, 2000) est le fait qu'elle combine les données et leur traitement, dans une approche et démarche statistiques et en nécessitant chez le professionnel un esprit statistique. Mais il est important de pouvoir définir la démarche statistique et les étapes la constituant, et de caractériser les qualités de l'esprit statistique.

La statistique peut, dans certaines situations, être appelée au service de l'élaboration d'une théorie scientifique, c'est d'ailleurs cet élément que nous avons souhaité aborder en nous intéressant à l'usage de la statistique dans les productions de thèses. Le rôle de la statistique dans la réflexion de construction de théorie scientifique va ainsi être abordé.

1. La place de la statistique dans certains développements de théorie scientifique

La théorie scientifique est définie par les éléments suivants : elle se présente sous la forme d'un énoncé universel, porte sur la totalité des événements d'intérêt en tous lieux et en tout temps. Elle explique le comportement des choses et peut prédire des événements à venir. Elle est basée sur l'expérience. Ainsi, la statistique est une discipline tout à fait adéquate pour être appelée dans la recherche de reconnaissance d'une théorie scientifique.

La conception de la théorie scientifique peut être déterminée par une approche dite d'inductivisme, approche ayant un certain nombre de limites comme nous le verrons, ou une approche hypothético-déductive. D'autres approches sont débattues au-delà de ces approches-là mais elles dépassent largement le cadre théorique de notre travail.

a) *Des données à la théorie scientifique*

Le développement de théorie scientifique est, selon certains auteurs, basé sur l'observation objective et sans préjugés des données. Il s'agit de pouvoir ainsi établir des faits qui pourront, dès lors, nourrir une réflexion et, par induction, permettre la construction d'une théorie scientifique. Certaines conditions autorisant l'induction sont attendues : la généralisation ne peut être acceptée que si les faits et constatations sont multiples dans des situations variables et que toutes les observations satisfont la théorie proposée. Il faut cependant souligner que l'induction est un raisonnement non fondé logiquement mais uniquement basé sur l'observation.

L'existence de cette théorie scientifique permettra ainsi aux chercheurs de développer les situations de déduction et ainsi de pouvoir tirer des conséquences, qu'il s'agisse d'explications ou de prédictions.

Les méta-analyses peuvent être considérées comme des méthodes développant ce type d'approche d'inductivisme scientifique. La méta-analyse est une démarche statistique combinant les résultats d'une série d'études indépendantes sur un problème donné. Elle permet à partir d'une analyse plus précise des données par l'augmentation du nombre de cas étudiés, de pouvoir induire une loi ou une théorie.

Le terme est utilisé pour la première fois par Glass en 1976 (Glass, 1976) mais le concept est plus ancien (Cochran 1954). Ce type de démarche est tout particulièrement utilisé dans le cadre médical, afin de faciliter le transfert de connaissances et ainsi la prise de décision. Une organisation œuvrant à titre non lucratif est fondée en 1993 sous la direction de Iain Chalmers, il s'agit de la collaboration Cochrane²⁰. Elle a vu le jour à la suite d'un besoin d'organiser de manière systématique les informations concernant la recherche médicale, une des questions posées dans certaines situations étant de savoir si l'ensemble des informations permettait de les monter au niveau de preuves scientifiques pour la prise de décision médicale, voire de théories scientifiques.

Bien que la proposition soit alléchante, un certain nombre de difficultés sont associées à cette approche inductive. Il s'agit tout d'abord de la nécessité d'avoir une multiplicité de faits variés. L'exemple de l'approche par méta-analyse permet de pointer précisément ce souci. Les méta-analyses sont construites dans la plupart des cas à partir des études ayant conduit à une publication scientifique, malheureusement, pour des raisons qui seront discutées ultérieurement, les études ne conduisant pas à des conclusions statistiquement significatives sont rarement publiées ce qui, de fait, met en défaut le premier critère pour la définition d'une théorie scientifique.

La seconde difficulté concernant la mise en œuvre d'une approche inductive pour une définition de théorie scientifique est de caractériser les faits qui doivent être observés et qui seront jugés significatifs, ainsi que l'angle d'observation si le cadre théorique n'est pas défini. Qu'il s'agisse d'expérimentation ou d'étude observationnelle, le chercheur ne peut pas recueillir l'ensemble des éléments pouvant décrire la situation ou intervenir sur ce qui se passe. Son regard et son recueil sont de fait guidés par la ou les théories sous-jacentes à son approche. Il est donc illusoire de penser que les faits ou observations réalisés seront en amont de la théorie pour en être producteurs

Les analyses exploratoires telles que l'ACP, l'AFC et l'ACM ainsi que la statistique descriptive sont des méthodes qui peuvent être mises au service d'une approche inductive puisque l'intérêt de ces méthodes est de ne pas nécessiter

²⁰ Site Fondation Cochrane - <http://www.cochrane.org>

l'existence d'hypothèse *a priori* des distributions de variables ainsi que des associations que nous chercherions à comprendre. Il faut cependant souligner que le choix même des variables considérées dépend de la question sous-jacente à la structuration du recueil de données, tout recueil d'informations étant marqué de l'empreinte d'une théorie.

L'approche par une modélisation ne peut en aucun cas être mise au service de cette démarche puisque, de fait, elle nécessite des choix *a priori* de modèles et de variables d'intérêt.

b) Modèle hypothético-déductif

Dans cette approche, le choix de la théorie scientifique questionnée précède l'observation ou l'expérimentation, elle guide donc le choix de la formulation des hypothèses à tester, qui, à l'issue de la démarche, seront réfutées ou acceptées.

C'est dans ce cadre de modèle hypothético-déductif que les tests statistiques ont été développés en tout premier par Fisher (les tests de significativité) puis par Neyman et Pearson (Schwartz, 2012).

Pour Fisher, il s'agissait de proposer une théorie des tests dont la seule issue était de rejeter l'hypothèse nulle ou de ne pas pouvoir conclure, aucune hypothèse concurrente n'étant proposée et seul le risque d'erreur de rejet étant défini. Cette approche correspond à la théorie de Popper, qui souligne que la spécificité des théories scientifiques est qu'elles sont falsifiables et non pas vérifiables. De plus Popper propose de pouvoir utiliser aussi le terme de *réfutable* à la place de *falsifiable*. Tout cela est donc basé sur le fait qu'à partir des observations ou des faits on ne peut pas déduire des théories mais par contre on peut en déduire si elles sont fausses.

Dans le cas des tests statistiques l'induction ne se fait pas en termes de vérité mais de vraisemblance, car les conclusions sont toujours entachées d'un risque d'erreur. Cette spécificité sera ainsi à prendre en considération dans le cadre de décision.

Pearson et Neyman (1928) ont quant à eux proposé de compléter les *tests de significativité* de Fisher en ajoutant un second risque d'erreur associé à l'acceptation

de l'hypothèse sous un autre modèle : il s'agissait dès lors des *tests d'hypothèses*. Cette modification faisait tout particulièrement rentrer la statistique dans le cadre des sciences de la décision.

Cependant, Fisher insiste sur le fait que les observations ne peuvent pas valider une hypothèse, et que dire le contraire consiste en un abus de raisonnement.

« For the logical fallacy of believing that a hypothesis has been proved to be true, merely because it is not contradicted by the available facts, has no more right to insinuate itself in statistical than in other kinds of scientific reasoning[...] It would therefore add greatly to the clarity with which the tests of significance are regarded if it were generally understood that tests of significance, when used accurately, are capable of rejecting or invalidating hypotheses, in so far as they are contradicted by the data: but that they are never capable of establishing them as certainly true. » R. A. Fisher

Car l'erreur de logique consistant à croire qu'on a démontré qu'une hypothèse était vraie, simplement parce que les données dont on dispose ne la contredisent pas, n'a pas plus de raison de s'insinuer dans le raisonnement statistique que dans tout autre raisonnement scientifique[...] Cela clarifierait beaucoup la vision que l'on a des tests d'inférence si on comprenait qu'en en faisant un bon usage on est capable de rejeter ou d'invalider une hypothèse, dans la mesure où elle est contredite par les faits : mais les tests ne peuvent jamais établir avec certitude qu'une hypothèse est vraie. (Schwartz, 2012)

Ces évolutions ont fait couler beaucoup d'encre, et, bien que certains auteurs soulignent l'intérêt de les réunir, cette réflexion est loin de faire l'unanimité. Michel Armatte souligne pour cette raison les dangers d'un usage hybride des tests statistiques (selon l'approche de Fisher ou de Pearson-Neyman), souvent dû à une non-connaissance de la différence épistémologique. Il est important d'ailleurs de souligner que cette présentation n'est presque jamais spécifiée dans les livres scolaires de statistique.

« Or ce contre quoi les différents auteurs contemporains mettent en garde, c'est la diffusion massive d'une version hybride de ces deux approches qui confond falsification et décision, qui suppose 2 hypothèses mais n'en spécifie qu'une, qui ne se soucie plus de randomisation, qui prend $\alpha = 5\%$ comme une norme, qui utilise indifféremment la p-value ou la valeur critique, et qui interprète $1 - p$ comme un degré de corroboration de H_1 . Or cette interprétation associée à la règle implicite de ne publier que les effets significatifs provoque de nombreux biais de publication. Dans une discussion de l'article de Denis, nous avons déjà argumenté sur la diversité disciplinaire des méthodologies hybrides construites et surtout pratiquées – elles ne sont pas les mêmes et n'ont pas les mêmes fonctions en psychologie et en économie – et nous avons conclu que l'hybridation, loin d'être une tare épistémologique, était sans doute une nécessité des pratiques de la recherche empirique. » (Armatte, 2006)

Actuellement, de nombreuses revues (par exemple la sixième édition du manuel de publication de l'*American Psychological Association* parue en 2009), pour les diverses raisons signalées ci-dessus souhaitent combattre la suprématie de la « p-value », et insiste pour que les auteurs n'indiquent pas uniquement les résultats de « p-value » significatives (au sens inférieur au seuil d'erreur choisi) mais aussi les estimations ainsi que les intervalles de confiance associés (Fine, 2013a).

2. Démarche statistique

La démarche statistique est ainsi définie comme une approche hypothético-déductive selon différents auteurs, reposant sur l'enchaînement de plusieurs étapes indiquées ci-dessous (Gattuso, 2011) :

- i. questionnement scientifique,
- ii. planification de l'expérience (étude du contexte),
- iii. collecte des données, conception des systèmes d'information,
- iv. analyse de ces données et interprétation,
- v. conclusion et restitution ou communication des résultats,
- vi. re-questionnement, esprit critique.

La phase 2 ainsi présentée ne doit pas être entendue au sens unique de l'expérience comme origine de toute connaissance valide selon la théorie de l'empirisme. Il s'agit d'une présentation plus large, pouvant inclure une approche par simulation, par exemple.

Nous retrouvons ainsi les éléments déjà cités précédemment ; cependant, les première et dernière étapes définissent la démarche de recherche scientifique et non spécifiquement une approche statistique.

Ainsi, comme Bachelard le souligne dans « *La formation de l'esprit scientifique* », la formulation du problème caractérise, avant tout, l'esprit scientifique dans son sens large.

« Avant tout, il faut savoir poser des problèmes. Et quoi qu'on dise, dans la vie scientifique, les problèmes ne se posent pas d'eux-mêmes. C'est précisément ce sens du problème qui donne la marque du véritable esprit scientifique. Pour un esprit scientifique, toute connaissance est une réponse à une question. S'il n'y a pas eu de question, il ne peut pas y avoir de connaissance scientifique. Rien ne

va de soi. Rien n'est donné. Tout est construit. » (p. 14) [...] il faut réfléchir pour mesurer et non mesurer pour réfléchir » (p255) (Bachelard, 2004).

Bien que les étapes 1 et 6 soient fondamentales, mais non spécifiques à la statistique, nous nous attacherons aux quatre étapes centrales caractérisant la démarche statistique, mais n'ayant bien évidemment d'intérêt que dans la contextualisation nécessaire des phases 1 et 6. Les quatre étapes intermédiaires sont présentées dans le rapport GAISE²¹ [rapport américain présentant des recommandations pour l'enseignement de la statistique] comme les quatre étapes de la démarche statistique devant être enseignées tout au long des cursus scolaires, c'est-à-dire : formuler une question, collecter des données, analyser les données et enfin interpréter les résultats (Fine, 2013).

Etre acteur d'une démarche statistique implique donc le fait de participer à tout ou partie des phases de la démarche, d'en accepter le déroulement et surtout d'accepter le cadre conceptuel sous-tendant l'usage de savoirs statistiques. La pratique et professionnalité du « statisticien », au sens non de métier mais d'activité, peuvent être définies comme fondées sur des verbes d'actions tels que recenser, sonder, repérer, analyser [...] Une démarche statistique ne peut se dérouler correctement dans sa globalité que si les participants ont pu conjointement travailler en partageant des valeurs communes, dénommé « esprit statistique » (Régnier, 2005).

²¹ Guidelines for Assessment and Instruction in Statistics Education édité par American Statistical Association

3. Culture statistique

La statistique se définit par un ensemble de fondamentaux, par des principes et une démarche, mais tout cela ne prend corps que dans un ensemble de comportements partagés, de manière d'agir, de manière de penser que nous appellerons culture statistique.

Un des éléments fondamentaux de la manière de penser en statistique est défini par Régnier comme le « *rapport que le sujet entretient à l'égard de l'incertitude et de l'erreur, facteurs inhérents à tout acte humain de prise de décision.* » (Régnier 2003), il parle de l'"esprit statistique". Cette distanciation, nécessaire, par rapport à un résultat observé due à l'acceptation de l'incertitude, est un comportement fondamental pour les usagers de la statistique, tout particulièrement s'ils sont en recherche de preuve. Nous pouvons préjuger de l'importance de l'environnement communautaire et sociétale dans la capacité à développer ou non ce rapport à la situation. Ainsi, pour reprendre Régnier, « *l'esprit statistique requiert un renoncement à l'usage systématique de l'idée de vérité pour chercher à maîtriser celle de vraisemblance, de plausibilité.* » (2005). La statistique a pour objet d'aider à appréhender l'incertitude et à revendiquer le droit à l'erreur, elle propose de quantifier le risque d'erreur subi ou considéré comme acceptable.

Une partie de la difficulté de l'enseignement de la statistique vient justement du fait que l'incertitude semble, au moins en France, anti-pédagogique, et que l'enseignant en tant que « celui qui sait », ne peut pas présenter du savoir entaché de risque d'erreur ou un savoir expert ne conduisant pas à l'exactitude. Cet état d'esprit est très difficile à faire accepter, d'autant plus, quand cet enseignement de la statistique est présent dans le cadre d'un enseignement de mathématiques, les mathématiques étant perçues et comprises comme sciences de l'exactitude obtenue par un raisonnement irréprochable et un résultat ne pouvant pas faire débat.

La mise en situation que Brousseau avait proposée à des élèves de l'école primaire de Michelet de Talence en 1974 correspondait précisément à une situation didactique permettant de développer la culture statistique des enfants. Il s'agissait de leur proposer une bouteille contenant 5 boules, et dont le bouchon transparent laissait apparaître une boule à la fois. Les enfants devaient « deviner » le nombre de boules

blanches et noires contenues dans la bouteille, sans jamais pouvoir ouvrir cette bouteille pour « vérifier ». Ils devaient donc se "convaincre" jusqu'à ce qu'ils jugent inutile une vérification empirique. Comme le précise Brousseau, découvrir le contenu réel de la bouteille n'aurait rien apporté pour valider la démarche mise en œuvre bien au contraire puisque le risque aurait été de remettre en question une stratégie qu'ils auraient développée et qui correspondait complètement à une stratégie de modélisation cohérente et pertinente dans ce type de situation²².

Les situations d'études statistiques pour lesquelles au final l'information exacte est diffusée, mettent en difficulté le développement de l'esprit statistique. Prenons pour exemple le cas des sondages pré-électorales qui sont des situations pouvant mettre en défaut l'esprit statistique puisqu'il s'agit de faire une prévision dont le statut (vrai ou faux) sera finalement obtenu, par un dénombrement des votants selon leurs votes. Il est dans ce cas difficile de renoncer à l'idée de vérité ou d'exactitude du résultat, puisque l'estimation obtenue est comparée au résultat obtenu et qu'une différence entre l'estimation et le résultat entraîne, à tort, une remise en question immédiate par les observateurs, de la qualité de la démarche. Cette situation met clairement en évidence la difficulté pour la plupart d'entre nous de différencier la démarche du résultat, et la vraisemblance de la vérité.

Wittgenstein souligne ce point dans le « *Cahier bleu* », en s'interrogeant sur la situation d'un sourcier déclarant « sentir l'eau à un mètre cinquante sous terre » :

« Maintenant, nous n'aurions pas ressenti ces difficultés si le sourcier avait dit qu'il avait appris à estimer la profondeur de l'eau, disons, en creusant pour trouver de l'eau à chaque fois qu'il avait une sensation particulière, et en établissant ainsi une corrélation entre de telles sensations et des mesures de profondeur. Nous devons donc examiner la relation entre le processus qui consiste à apprendre à estimer, et l'acte d'estimation. L'importance de cet examen tient à ce qu'il s'applique à la relation entre l'apprentissage du sens d'un mot et l'utilisation de ce mot. Ou, plus généralement, à ce qu'il montre les différentes relations possibles entre une règle donnée et son application ». (Wittgenstein, 1996)

Il est nécessaire d'explicitier ce que signifie pour l'utilisateur le mot *estimer*, à quoi il se réfère et quel est finalement le lien entre son résultat et sa règle. Cette

²² Site personnel Guy Brousseau - <http://guy-brousseau.com/1500/probabilites-et-statistiques-31-lecons-experimentees-a-l%E2%80%99ecole-j-michelet-de-talence-1974/>

situation est à mettre en regard avec l'exemple décrit ci-dessus concernant la situation des sondages politiques. Il y a des méthodes d'enquête permettant de recueillir correctement des informations, des fonctions spécifiques pour estimer appelées *estimateurs*, et il y a l'estimation, c'est-à-dire la valeur obtenue à partir des données recueillies sur un échantillon singulier. Dans cette situation de sondage politique, le couperet tombe lorsque ce qui est regardé est l'écart entre l'estimation et la valeur finale connue. Mais pourtant la question de la qualité de la méthode ne se situe pas dans le résultat. Ce qui doit être interrogé est la qualité de la praxéologie de « l'usager statisticien » et non le résultat. Malheureusement le but de ces travaux statistiques étant le résultat, et la démarche étant comprise comme un moyen pour y arriver, il n'est pas toujours fait grand cas de cela.

La difficulté majeure en statistique est ainsi d'accepter de reconnaître la vraisemblance d'une solution au détriment d'une vérité parfois autoproclamée. Il n'y a pas de certitude en statistique, même si les méthodes sont efficaces, il est important de savoir garder une attitude critique, c'est cette attitude qu'il faudra faire acquérir aux étudiants (Gattuso, 2011).

C. Les notions fondatrices

Il est important de repérer les éléments fondateurs de la discipline statistique qui interrogent ou rendent difficile l'accès à cette discipline et qui sont pourtant nécessaires pour les appliquer et les utiliser convenablement. Tous les statisticiens s'accordent à souligner qu'il s'agit en tout premier lieu de trois éléments qui sont : la notion de *variabilité*, la question de *représentativité* et enfin la *significativité*. Nous pouvons, d'ailleurs, souligner le fait que ces trois concepts correspondent aux éléments de la problématique de l'incertitude. Incertitude dans la construction de l'échantillon, et dans le résultat de cette démarche d'échantillonnage ; incertitude dans le recueil des données ; incertitude dans la détermination des résultats au regard des deux incertitudes précédentes. Le concept de variabilité est souvent vu comme central et constitue un élément pouvant susciter de grandes discussions entre statisticiens quant aux stratégies pédagogiques à développer pour faire comprendre

ce concept, qui de plus est multiforme comme nous le verrons ci-après. Mais, il n'est pas le seul concept fondamental de la discipline statistique.

« Deux autres concepts fondateurs de la statistique : représentativité et significativité. La variabilité est la raison épistémologique majeure de la statistique. La représentativité fonde la validité des énoncés produits par l'analyse statistique en tant qu'étude de la variabilité des univers statistiques au regard d'un (des) caractère(s) propre(s) au(x) phénomène(s) étudié(s) à partir de la partie (échantillon). La significativité fonde le degré de confiance du choix dans la décision. » (Régnier, 2005)

1. Pourquoi la variabilité est-elle l'élément incontournable de la statistique ?

« Le rapport (ndlr GAISE) insiste sur le fait qu'une formation à la statistique devrait présenter ces différents types de variabilité car, dans la pratique de la statistique, la résolution d'une question statistique et la prise de décision dépendent de la compréhension, de l'explicitation et de la quantification de la variabilité dans les données. » (Fine, 2013b)

La statistique, en tant que science de la description et de la compréhension de l'information issues de données doit, de fait, prendre en considération l'omniprésence de la variabilité dans les ensembles de données, qu'il s'agisse de données recueillies, mesurées, issues d'expériences ou obtenues de manière exhaustive. Cette caractérisation est discutée dans le rapport GAISE. Le rôle du statisticien est de pouvoir comprendre et dompter cette variabilité qui peut se décliner en différents types afin de ne pas décrire uniquement des tendances centrales.

a) *Cadre général*

Différents niveaux de variabilités expliquent l'hétérogénéité des données, et ce sont ces phénomènes que le chercheur souhaite comprendre et expliquer. Il y a tout d'abord la variabilité que nous pourrions décrire comme *naturelle*, qui correspond à la variabilité entre individus ou sujets d'études, appelée encore *variabilité inter-sujets*. Les sujets inclus dans une étude n'étant pas des clones les uns des autres, ils sont différents entre eux, et l'objet de la recherche peut consister à comprendre les facteurs pouvant expliquer ces différences. Il s'agit donc dans la plupart des cas de déterminer la variance puis de la décomposer.

Il y a un second niveau de la variabilité naturelle que nous nommerons *variabilité intra-sujet*, venant du fait que certaines données ne sont pas stables dans la mesure. Nous pouvons prendre comme exemple la mesure de la tension artérielle lors d'une visite médicale (fluctuation importante chez certains patients entre le début et la fin de la consultation, appelée *effet-blouse-blanche*), ou des informations concernant les opinions pouvant, de par la définition de l'opinion, être très influencées par des facteurs environnementaux. A ces deux variabilités intra et inter-sujets, que nous pouvons caractériser de naturelles et de préexistantes, viendront se greffer des variabilités induites ou subies qu'il s'agira de repérer, et selon les situations de pouvoir quantifier.

Un recueil de données est issu d'une étude d'observation ou expérimentale. Une situation d'expérimentation signifie que le responsable de l'étude construit, selon les objectifs même de l'étude, une situation favorisant la possibilité de répondre aux attentes scientifiques. La construction même de l'organisation de l'étude peut induire une variabilité additive sur les données, car on peut imaginer des situations d'étude nécessitant une surreprésentation de certains profils de sujets. Cette *variabilité* que nous pourrions qualifier d'*artificielle*, doit être quantifiable lors de l'analyse pour être en capacité d'en tenir compte et ne pas risquer des surestimations lors des calculs statistiques.

Dans la situation d'enquête en population, réalisée sur un échantillon aléatoire, une *variabilité due au hasard* va se greffer aux données car, bien que la méthode de sélection aléatoire puisse être précisément explicitée dans le respect des règles statistiques, il s'avère qu'un seul échantillon sera finalement obtenu et que cette réalisation sera entachée d'une variabilité due au hasard. Cette variabilité pourra dans la plupart des cas être estimée en modélisant les principes d'échantillonnage.

Un troisième niveau de variabilité est induit par les outils de mesure utilisés dans le cadre de l'étude : elle est nommée *variabilité de mesure*. Les outils de mesure peuvent être plus ou moins fiables ; un instrument est dit *fiable* si les mesures faites sur le même sujet diffèrent peu, voire pas du tout. Il est donc nécessaire, lors de la mise en place d'une étude, de standardiser les outils et stratégies de recueil afin de contrôler au mieux cette variabilité de mesure.

Décrire un ensemble de données consiste à résumer l'information présente dans une source de données, en en faisant ressortir les éléments perçus comme essentiels. En règle générale, le premier indicateur communiqué correspond à une information dite de *tendance centrale*, qui consiste à résumer en un indicateur quantitatif un ensemble d'informations. Cet indicateur peut être une fréquence, une moyenne, une médiane, un centre de gravité²³, etc. Sans variabilité des données, cet indicateur peut suffire. Cependant, ce type de synthèse de l'information est très appauvrissant au regard de la richesse présente dans un recueil d'informations, et rend presque impossible l'analyse poussée des associations et interactions entre les différentes informations. Il s'agira de comprendre les relations entre variables, et de les décrire graphiquement ou analytiquement. Les mesures de variabilité associées à ces informations sont multiples et peuvent différer selon les habitats d'usage. Cependant, la question sous-jacente est toujours de déterminer et de choisir la mesure ou la représentation graphique qui sera la plus informative.

Les éléments que nous venons d'explicitier réduisent la question de variabilité au niveau de chaque caractéristique, c'est-à-dire en considérant les différences entre sujets, variable par variable. Il est cependant souvent intéressant de se poser la question de la variabilité de manière globale, en considérant les données dans leur ensemble. Il est assez facilement compréhensible que la variabilité inter-sujets est d'autant plus élevée que le nombre de caractéristiques est plus important. Ainsi, la question même de la variabilité est induite par le choix du chercheur dans la définition des dimensions qu'il souhaitera explorer. Cette question est étudiée en situation ci-après.

²³ Nous reviendrons ultérieurement sur les différentes terminologies usitées.

b) Un exemple de problème de variabilité – les études de cas multiples en neuropsychologie

En 2005, nous avons eu la chance de pouvoir longuement échanger avec un chercheur psychologue canadien sur la question concernant les études de cas multiples en neuropsychologie et les questions méthodologiques en statistique y afférant. La question était posée ainsi : la plupart des études en neuropsychologie étaient des études de cas, donc réalisées sur très peu de personnes et sujettes à de nombreuses critiques scientifiques concernant leur validité intrinsèque. Il s'agissait alors d'aider à répondre à la question : comment peut-on développer l'usage d'études de cas multiples, comme solution possible à cette problématique ? Ce travail nous a conduit à la rédaction d'une publication dans la Revue de Neuropsychologie « *Plan expérimental des « études de cas multiples » : rationnel et méthode d'analyse des données* » (Rainville, 2006). Il s'agissait dans cet article de souligner les différents éléments statistiques soulevés par la mise en place d'études de cas, qu'il s'agisse des problèmes liés à la variabilité ou ceux liés à la représentativité, mais aussi de diffuser des éléments de compréhension concernant des méthodes statistiques appropriées. La deuxième partie de cet article, « méthodes d'analyse », explique les tenants et aboutissants des analyses factorielles et des classifications par une approche purement explicative, sans développement mathématique.

La question de l'importance de la prise en compte de la variabilité pour les études en neuropsychologie était au cœur de cette étude et mettait en évidence les différents niveaux de la variabilité vus dans le paragraphe précédent.

« Les problèmes méthodologiques de la neuropsychologie concernant la validité des observations s'inscrivent dans un débat plus vaste, propres aux recherches cliniques avec peu de sujets participant aux études. Pour l'essentiel, l'argumentation gravite autour des notions de variabilité des individus, de la formation des groupes expérimentaux et/ou d'études de moyennage des performances et la possibilité de généralisation des données. » (Rainville, 2006)

Nous ne souhaitons pas reproduire dans le corps du texte de cette thèse la totalité de l'article, qui est très long, mais nous en extrairons les parties particulièrement importantes pour mettre en lumière la question des différents niveaux de variabilité et de leur importance dans l'étude de situation pathologique.

Trois éléments concernant la variabilité étaient particulièrement repérés : il s'agissait tout d'abord d'une variabilité naturelle dans la situation pathologique, puis de l'impact d'un nombre important de mesures sur le regard porté à la variabilité inter-individus et enfin du lien entre la variabilité et l'analyse.

Le premier élément concernait donc la grande variabilité pouvant être expliquée par la situation pathologique. Il est souligné dans l'article que :

« En neuropsychologie, le problème de la variabilité s'accroît. Aux facteurs de variabilité propre à la population normale s'ajoutent ceux qui découlent du processus pathologique. La nature de la lésion ajoute une source de variabilité considérable, rendant la formation d'un groupe homogène plus aléatoire».

Ce point met ainsi en relief la relation inhérente entre la variabilité de la population d'étude et la structuration de l'étude que l'on souhaite construire. La connaissance *a priori* de la variabilité naturelle des sujets est donc un élément pouvant alimenter le choix stratégique de la méthode, et la construction même de l'échantillon en fonction de la question de recherche.

Le second point était caractérisé par le fait que, en neuropsychologie les chercheurs ont l'habitude de faire passer de nombreux tests à leurs patients, induisant ainsi une spécification de chaque sujet et rendant d'autant plus difficile la recherche de facteurs communs. Un élargissement de l'observation individualise de fait les observés, et ceci correspond ainsi à la situation des études de cas, c'est-à-dire une étude approfondie et fine d'un cas particulier induisant de fait la problématique de la généralisation.

« L'augmentation du nombre d'épreuves, en fonction des critères retenus et de leur sensibilité, accroît la capacité à discriminer les sujets. La variabilité inter-individuelle est donc augmentée, ce qui peut poser problème dans la mesure où le principe de généralisation se fonde sur une faible variabilité inter-individuelle. Il y a donc une « contradiction » apparente entre l'exigence d'augmenter le nombre de tests, ainsi que leur sensibilité, pour tracer un profil fonctionnel des sujets le plus précis possible, et les exigences de l'homogénéité. »

Cet élément est tout particulièrement intéressant, car la question posée ici concerne le fait de s'intéresser à la variabilité pour une caractéristique ou plus

largement la variabilité au sens global. Ce qui est indiqué implicitement ici est que la variabilité ne peut pas être considérée comme un concept général mais qu'il s'agit bien de la variabilité conditionnée à un espace d'informations d'intérêt choisies délibérément.

Pour la discussion concernant le lien entre la variabilité et l'analyse nous reprendrons ici la partie de l'article y référant dans sa totalité.

« La variabilité des individus et le “moyennage” des performances »

Le débat entourant le problème de la variabilité des individus débouche sur ceux de la méthode d'analyse des données et du moyennage des performances cognitives étudiées. Cet aspect est abordé différemment selon l'approche du chercheur. Dans les recherches (de groupe), la variabilité est vue comme un élément à contrôler (fiabilité et fidélité de la mesure) de façon à étudier les conduites potentiellement significatives. Dans les analyses, le rationnel sous-jacent veut que les facteurs non désirés jouent dans un sens pour un certain nombre de sujets, dans le sens opposé pour d'autres. Quand les données sont moyennées, ces influences vont s'annuler, en principe, pour ne laisser apparaître que le phénomène étudié. Le plan expérimental et le niveau de confiance statistique vont donc permettre de dégager les relations fonctionnelles entre les variables dépendantes et indépendantes. Pour d'autres chercheurs, en particulier dans la recherche clinique, la variabilité est considérée comme une dimension devant être intégrée dans l'étude des phénomènes. Elle ne découle pas nécessairement des conditions expérimentales et toute la question est de savoir comment on peut la contrôler et l'étudier (Chassan, 1967; Hersen et al., 1984; McReynolds & Kearns, 1983). Plutôt que de tenter d'en faire disparaître l'influence, elle est vue comme une variable qui intervient dans le phénomène étudié, voire un objet d'étude en soi. L'alternative au traitement statistique de la variabilité inter-sujets est de la mettre en évidence et tenter d'en déterminer la cause. L'identification des sources de variabilité dans un sous-groupe de sujets peut conduire à leur compréhension dans la population (Hersen et al., 1984; Sidman, 1960).

En neuropsychologie, pour les tenants des études de cas, l'étude de la variabilité est également importante en soi. De nouveau, l'argument central veut que le moyennage des grands groupes puisse faire perdre une quantité importante d'informations du fait de la complexité des individus. Au sein d'un groupe, cette stratégie peut cacher l'existence de profils de performances particulièrement intéressants (Zeki, 1993). Plus spécifiquement, l'analyse des performances moyennes risque de masquer les dissociations, témoins de l'existence d'un sous-composant. Par ailleurs, l'association des troubles peut tout simplement découler de la juxtaposition de sous-groupes de patients présentant des troubles différents (Seron, 1993, 1994). Selon McCloskey et Caramazza (1988a), les variations qualitatives rendent le moyennage sans intérêt. Il n'est pas approprié quand les sujets varient qualitativement dans les mécanismes cognitifs étudiés. Il n'est pas une solution avec les problèmes associés aux différences individuelles (hétérogénéité des sujets). L'observation des profils de performance est susceptible de conduire au développement de modèles et de générer de nouvelles hypothèses fonctionnelles. Les méthodes

d'analyse usuelles, comme le moyennage des performances, ne sont pas adaptées à cette situation puisqu'elles masquent les profils (Vallar, 1999). Enfin, les moyennes ne reflètent pas le fonctionnement du système cognitif d'un sujet. Elles ne sont pas le reflet du fonctionnement «moyen» d'un système cognitif (McCloskey & Caramazza, 1988a). Dans la perspective de l'étude de cas, le calcul d'une moyenne n'a de sens que s'il y a une seule façon d'obtenir ce score. »

La question de la prise en compte de la variabilité dans une étude induit naturellement la question de la représentativité de l'échantillon, le concept de représentativité faisant référence à la capacité d'estimer correctement les paramètres de la population et donc sa variabilité. La représentativité correspond au second concept central de la statistique.

« Une objection importante soulevée à l'endroit des études de cas concerne la généralisation des données. Dans les études de groupe, les résultats obtenus doivent permettre des généralisations sur l'ensemble de la population à laquelle ces individus appartiennent. Une des précautions méthodologiques consiste à former un groupe d'individus représentatif de la population dont il est tiré. La sélection aléatoire des sujets parmi la population d'intérêt permet d'éviter tout biais de sélection. Tout en tenant compte de la fluctuation d'échantillonnage, le sous-groupe ainsi formé est représentatif de la population. »

2. Vous avez dit représentatif !

« Hommage à JC Deville - introduction de son intervention « Vous avez dit représentatif ? » lors du colloque « Au royaume des sondages » (Bruxelles, 8-10 novembre 1988)

Un professionnel des sondages est contacté par une personne, et le dialogue débute ainsi :

- Je voudrais savoir comment faire pour constituer un échantillon représentatif.
- Qu'est-ce que vous appelez représentatif ? répond le professionnel
- Je pensais que vous me le diriez [...]
- Le problème, c'est que je ne sais pas trop !
- Dites-moi au moins la taille de mon échantillon pour qu'il soit représentatif ?
- C'est très simple, divisez le budget total dont vous disposez pour votre enquête par le coût de réalisation d'un interview. » (Selz, 2013)

Ce type d'échanges est légion dans la vie d'un statisticien, qui doit réussir à faire expliciter au demandeur le sens précis de sa demande mais aussi à lui faire comprendre que la taille de l'échantillon n'est pas une fin en soi et qu'il y a avant tout une démarche de sélection puis un effectif pour satisfaire à une attente de précision.

a) *Quel sens est donné à ce terme représentativité ?*

Le niveau de compréhension du terme *représentatif* est bien différent selon l'angle avec lequel on l'étudie : la méthode de construction de l'échantillon, l'échantillon lui-même, les estimateurs ou les estimations obtenues à partir de l'échantillon. Marion Selz a coordonné la rédaction collective d'un ouvrage « *La représentativité en statistique* » qui offre une présentation historique et transverse de la question de la représentativité (Selz, 2013) que nous ne reprendrons pas dans le cadre de ce travail. Nous ne soulignerons que certains points nous semblant majeurs.

Dans la plupart des situations, ce qui est jugé correspond uniquement à l'échantillon obtenu et non à la démarche. L'échantillon est alors jugé représentatif ou non selon des critères parfois tout à fait subjectifs du style « *il y a trop de personnes de petite taille dans cet échantillon pour que les résultats obtenus soient jugés fiables* ». Ce type de considérations remet en question l'approche statistique, puisque ce n'est pas la démarche statistique qui est construite dans le cadre d'une stratégie de représentativité mais uniquement des résultats. La démarche statistique n'a pas pour raison de s'intéresser à un échantillon réalisé, mais bien plus à la démarche mise en œuvre pour permettre de construire une stratégie qui assurera l'obtention d'échantillon²⁴ permettant de recueillir les informations qui seront utilisées pour déterminer, à partir d'une fonction estimateur adéquate, une estimation et son niveau d'incertitude pour une population définie.

Tout l'art du statisticien consiste à être en mesure de concevoir une méthode qui lui permette de contrôler le niveau d'incertitude de cette estimation en considérant la méthode de construction de l'échantillon, et ainsi de pouvoir construire une variable aléatoire, issue des variables d'échantillonnage, sans biais. Le point majeur dans la réalisation d'une étude statistique est de pouvoir obtenir des estimateurs sans biais et consistants. C'est ainsi que le statisticien entend la définition donnée par le Larousse 2015 de l'adjectif « représentatif » :

²⁴ La démarche statistique, développée en vue d'obtenir des estimations, doit expliciter la méthode théorique et non pas uniquement le résultat.

« qui possède les caractéristiques principales de l'ensemble dont il relève, et peut, à ce titre, le représenter ».

La formulation de cette définition est consensuelle dans sa syntaxe mais cependant bien différente dans le sens attribué par chacun. C'est pour cela que Tillé indique dès le début de son ouvrage « Théorie des sondages », les précautions qui doivent être prises dans l'usage de la notion de représentativité, en soulignant que :

« Voir invoquer la représentativité dans un rapport d'enquête pour justifier de la qualité d'un sondage peut presque à coup sûr laisser soupçonner que l'étude a été réalisée dans une méconnaissance totale de la théorie de l'échantillonnage. Le concept de représentativité est aujourd'hui à ce point galvaudé qu'il est désormais porteur de nombreuses ambivalences. Cette notion, d'ordre essentiellement intuitif, est non seulement sommaire mais encore fautive et, à bien des égards, invalidée par la théorie. Raison pour laquelle ce terme sera volontairement évité dans cet ouvrage. » (Tillé, 2001)

Le terme représentatif est le plus souvent montré haut et fort comme l'élément de légitimité de tout ce qui suivra lors de présentation de résultats d'étude, et tout particulièrement toutes les données chiffrées. C'est ce type de réaction qui fait dire à Tillé que ce terme est galvaudé.

Pour la plupart des personnes, non statisticiennes dans l'âme, un échantillon est représentatif si sa structure en termes de répartition est identique à celle de la population dont il est issu. Cependant, cette ressemblance ne pouvant pas être assurée sur l'ensemble des informations caractérisant une population (le nombre de caractéristiques serait d'ailleurs infini) il est, du choix de chacun, de définir les caractéristiques pour lesquelles il semble nécessaire d'assurer cette ressemblance. Cette situation est, d'ailleurs, celle utilisée par l'ensemble des sociétés de sondage pour justifier de la pertinence d'utiliser des sondages par quotas. Comme l'indique Desabie la méthode des quotas repose sur l'hypothèse suivante :

« les différents caractères que l'on peut observer dans une population n'étant pas indépendants entre eux un échantillon identique à la population dans laquelle il est prélevé en ce qui concerne la distribution statistique de certains caractères importants sera également peu différent de la population en ce qui concerne la distribution statistique des caractères qui ne sont pas contrôlés. » (Desabie, 1966)

Travailler à partir de sous-parties d'une population²⁵, c'est-à-dire un échantillon, induit le fait de vouloir généraliser des résultats sur la population alors qu'ils n'auront été obtenus que sur un sous-groupe particulier ; pour ce faire, il est donc nécessaire de prendre un certain nombre de précautions lors de la construction du plan d'échantillonnage, qui peuvent être déclinées différemment selon les auteurs.

Cette question de représentativité est un point revenant régulièrement et pour la plupart des personnes, elle est synonyme de « structure comparable » au sens utilisé pour la mise en œuvre de sondage par quota. Ce terme est réellement entendu de façon très différente selon les acteurs : c'est d'ailleurs ce qui fait dire à Tillé qu'il est préférable de ne pas l'utiliser, point auquel nous adhérons car l'utilisation de ce terme dans le cadre d'échange avec des chercheurs non statisticiens, des étudiants ou toute autre personne complique toujours les échanges.

b) Préciser plutôt les processus de sélection

Pour toutes les raisons invoquées ci-dessus, bien que le terme *représentatif* soit par de nombreux auteurs présenté comme une notion fondatrice de la statistique il semblerait que son utilisation soit difficile tant les représentations de chacun diffèrent. L'enjeu pour le statisticien est bien d'accompagner la compréhension de la stratégie d'échantillonnage, en insistant sur les choix stratégiques effectués et justifiables dans le cadre d'un objectif clair. Il s'agit bien de travailler la phase 2 de la démarche statistique.

Le choix fait par Tillé de ne pas utiliser le terme *représentativité* dans un des ouvrages incontournables pour toutes personnes souhaitant s'intéresser à l'élaboration d'échantillon, est un signal important pour l'ensemble de la communauté statistique. Le message doit donc passer par une présentation plus précise du processus, et ne doit pas se limiter à un mot « passe partout » tel que *représentatif*, qui veut dire tout et rien à la fois. Ainsi, on peut souligner que, dans cette évolution, l'année 2011 a été une date importante, mettant en lumière le débat sur la question des sondages électoraux. Faisant suite à de nombreuses consultations dont celles de statisticiens de l'INSEE et d'académiques, un rapport du Sénat,

²⁵ Nous employons ici le terme population au sens statistique c'est-à-dire l'ensemble d'unités statistique d'intérêt pour l'étude statistique dans laquelle l'échantillon sera sélectionné.

« *Sondages et démocratie : pour une législation plus respectueuse de la sincérité du débat politique* », (Portelli, 2010) a été publié, soulignant un certain nombre de recommandations. Il s'agissait de répondre à une demande de transparence méthodologique de plus en plus forte sur le processus de constitutions de ces estimations d'intentions de vote. Ce qui est remarquable dans ce rapport c'est qu'il comprend une annexe méthodologique, très pédagogique, à l'usage de non statisticiens rédigée par Pascal Ardilly (Groupe de travail sur les sondages en matière électorale, 2011).

Comme nous venons de l'indiquer, le développement de recueil de données nécessite dans de nombreuses situations l'élaboration de processus d'échantillonnage. La théorie des sondages est une théorie compliquée, associant une approche probabiliste donc très abstraite à des situations très concrètes et de terrain, donc associées à de nombreuses contraintes opérationnelles. Pour toutes ces raisons l'explication et la diffusion de ces procédés est vraiment un point important. Dans ce cadre, nous avons dans ce cadre participé à la rédaction de plusieurs documents et articles à destination principalement des acteurs du milieu sanitaire. Il ne s'agissait point de développer de nouvelles stratégies, mais de proposer des documents rédactionnels facilitant la compréhension et la mise en œuvre. Pour cela la présentation d'exemples appropriés et d'intérêt pour les acteurs étaient fondamentales : « *Initiation aux méthodes de sondage: application à la pharmaco-épidémiologie* » (Fourrier-Réglat, 2006), et le chapitre 22 « *Techniques de sondage et calcul de la taille d'échantillon* » (Jutand, 2012) de l'ouvrage collectif « *Epidémiologie de terrain : Méthodes et Applications* » (Dabis & Desenclos, 2012).

La structuration de l'échantillon construit étant connue et les données ayant été recueillies, la phase ultime de la démarche statistique est bien de pouvoir analyser les données et de pouvoir répondre aux questions ou hypothèses formulées dans le cadre de l'étude. L'attente de ce travail est bien le résultat des analyses statistiques, et tous les éléments de construction de l'étude sont développés au service des résultats. C'est à ce moment-là que le couperet fatidique arrive : significatif ou non significatif ?

3. La significativité : sésame de la vérité

La notion de significativité en statistique, ou de niveau de signification, est associée aux résultats de tests statistiques et plus généralement à la théorie des tests statistiques. Nous utiliserons ci-dessous la définition des tests d'hypothèses, au sens de Neyman et Pearson Il faut préciser qu'

« un test statistique est un mécanisme qui permet de trancher entre deux hypothèses au vu des résultats d'un échantillon. Soient H_0 et H_1 ces deux hypothèses, dont une et une seule est vraie. [...] Un test statistique se présente sous forme d'une règle de décision binaire formulée *a priori*, c'est-à-dire avant d'avoir fait les observations, et intangible dès que le risque α a été fixé. Les données recueillies ne peuvent modifier cette règle et ne sont utilisées *qu'a posteriori* : on est, ou on n'est pas, dans la région critique. Cette manière interdit toute manipulation des résultats et garantit l'objectivité de la décision : elle convient bien dans les problèmes de réglementation ou de normalisation. » (p336) (Saporta, 2011)

Saporta précise avec justesse l'importance de l'*a priori* dans la formalisation du test, la décision prise à partir des données devant uniquement découler d'une règle de décision prise et acceptée de tous en amont.

Malheureusement, un usage abusif des tests statistiques, facilité par le développement des logiciels plus simples à l'utilisation mais aussi par la nécessité de résultats statistiques pour faciliter les publications, a induit dans de nombreuses situations des mésusages. Il est aussi important de souligner que ce qui est attendu dans la plupart des cas n'est pas l'explication de la démarche mais bien le résultat de la prise de décision, la question étant « est-ce ou non significatif ? ». Nous reprenons ci-dessous le cadre général pour expliciter les éléments de l'usage abusif.

a) Probabilités d'erreur

Soient H_0 et H_1 les deux hypothèses entre lesquelles le test devra trancher au vu des données de l'échantillon. La situation de décision se traduit ainsi par le tableau classiquement connu par tous les étudiants ayant suivi un cours de statistique.

Tableau 2 : Probabilités d'erreur de décision

		Vérité	
		H_0 vraie	H_0 fausse
Prise de Décision	non rejet de H_0	<i>bonne conclusion de probabilité $(1-\alpha)$</i>	<i>erreur de seconde espèce de probabilité β</i>
	rejet de H_0	<i>erreur de 1^{ère} espèce de probabilité α</i>	<i>bonne conclusion de probabilité $(1-\beta)$ appelée puissance du test</i>

Les deux erreurs ne sont pas symétriques et il s'agit d'essayer de les contrôler lors de la mise en œuvre du test, ceci afin de pouvoir fixer le niveau d'incertitude accepté pour la décision. Il est de règle de fixer la probabilité α et ainsi de donner un rôle prééminent à l'hypothèse nulle. Les valeurs les plus classiques de probabilité pour α sont 0,05, 0,01 ou 0,1. Les habitudes de choix de seuil sont très dépendantes du contexte disciplinaire, ceci traduisant naturellement le niveau d'acceptation des risques encourus lors d'une mauvaise décision. La détermination de ce niveau de probabilité, ou erreur de première espèce, détermine la zone de rejet de l'hypothèse nulle, donc le critère de décision.

Cependant, classiquement les logiciels de statistique n'indiquent pas s'il y a rejet ou non rejet, mais indiquent plus exactement la probabilité de dépassement de la valeur de la statistique de test utilisée ou valeur de prise de décision. Cette probabilité est classiquement appelé *p-value* (parfois traduit *valeur de p*) et peut être nommée par certains *niveau de signification*.

b) Niveau de signification

Le niveau de signification peut être vu comme la probabilité que, sous H_0 , le hasard (ou fluctuations d'échantillonnage) donne une différence au moins aussi grande que celle qui a été observée. Contrairement à la région critique dont la détermination ne dépend pas de l'échantillon sélectionné, le degré de signification est propre à cet échantillon.

Le niveau de signification correspond au résultat que tout logiciel donnera lors de l'usage de l'algorithme du test statistique, et la démarche consistera ainsi à comparer la valeur du niveau de signification et celle de la probabilité d'erreur. Ainsi il sera indiqué que le résultat est significatif si le niveau de signification est inférieur à la probabilité d'erreur.

Dans les formulations utilisées dans les articles publiés, des auteurs peu scrupuleux de précision peuvent souligner la différence significative sans pour autant préciser les valeurs obtenues à partir des données et qui sont comparées dans le cadre du test, ni d'ailleurs préciser le test statistique employé. Dès lors, cela peut suggérer deux situations différentes : soit le terme *significatif* est employé dans un autre contexte que le contexte du test statistique soit il s'agit d'un oubli, volontaire ou non. Il est important de souligner que, dans le cadre d'une étude, les chercheurs statisticiens peuvent conclure à une différence significative d'un point de vue statistique sans que pour autant cette différence ait un sens clinique ou sociologique. Le résultat obtenu correspond uniquement à un des éléments de décision statistique, mais en rien à une décision définitive par rapport à la question contextualisée initiale. Cette différence de sens est parfois totalement omise dans certains articles qui fondent uniquement leur raisonnement et leur décision au vu des résultats statistiques significatifs. Dans une situation médicale, par exemple des patients peuvent voir une amélioration significative de leur état de santé suite à la prise d'un nouveau médicament alors que les épidémiologistes ne mettent pas en évidence une différence statistiquement significative entre le nouveau traitement et le traitement de référence pour des raisons d'effectif parfois trop faible. A l'inverse, suite à une enquête épidémiologique, un écart d'une semaine de durée de survie moyenne entre deux groupes de patients ayant eu des traitements différents peut être montré comme significativement différent de zéro suite à un test d'hypothèse sans que cet écart

puisse être considéré comme ayant un sens médical d'amélioration de la survie des patients.

Mais, que se cache-t-il derrière ce terme si ce n'est uniquement le fait de pouvoir enfin dire « j'ai vu quelque chose qui semble crédible ? ».

Un parallèle peut être fait avec ce que nous avons précédemment indiqué pour ce qui concerne la représentativité : il s'agit de la notion de référence car pour tout usage du terme *significativité*, il va s'agir de pouvoir définir cela en fonction d'une valeur de décision prise par la communauté dans un cadre précis et sans *a priori*.

Ainsi, dire qu'une relation est significative induit pour le statisticien qu'une analyse aura été effectuée sous l'hypothèse d'un modèle mathématique probabiliste afin de déterminer l'usage d'une statistique de test suivant une loi ainsi définie. Puis il s'agira de déterminer la probabilité d'observation du résultat sous cette hypothèse de modèle et de comparer cette probabilité avec le seuil que les chercheurs auront prédéfini. La significativité est donc dépendante du choix du seuil initial.

c) *Remarques*

Il est important de souligner la part arbitraire et subjective dans la décision même du seuil d'erreur permettant de conclure à la significativité ou non. Il faut insister sur le fait que certaines habitudes de niches communautaires existent ce qui permet d'une certaine façon de normaliser la pratique, et ainsi d'éviter des décisions ciblées. Le second point fondamental à souligner est le lien très fort entre l'effectif de l'échantillon étudié et la valeur du niveau de signification. Ainsi, à différence identique, la décision pourra être « significative » si l'étude est réalisée sur 5000 sujets et non significative pour une étude réalisée sur 100 sujets. Le réajustement du choix de la valeur du seuil d'erreur en fonction de la taille de l'échantillon est un élément rarement discuté dans les articles (Poitevineau, 2004). Bien d'autres éléments peuvent intervenir pour l'interprétation de ces résultats et il est donc vraiment important de bien sensibiliser les usagers et futurs usagers à ces difficultés possibles.

Deux autres éléments spécifiques à la diffusion sont importants à souligner, il s'agit tout d'abord de l'effet de résultats significatifs comme reconnaissance d'intérêt pour la publication. Comme nous l'avons évoqué, les articles pour lesquels le résultat principal de l'étude est non significatif sont des articles peu souvent soumis à des revues par leurs auteurs, cette auto-sélection induisant un biais de publication, appelé parfois *phénomène de l'iceberg*. Les chercheurs considèrent que montrer quelque chose d'intéressant pour leur communauté passe par le fait de montrer un résultat de test statistique significatif avec une vraisemblance acceptable : il est ainsi fait totalement abstraction de toutes les autres situations. Il est vécu comme échec ou inintérêt un résultat de test ne conduisant pas à une décision de rejet de l'hypothèse nulle selon le critère de seuil de rejet. Il est remarquable de constater en tant qu'enseignant la déception pour les étudiants de ne pas obtenir des p-values inférieures au seuil. Le second point concerne l'abus d'utilisation d'adjectif attribut pour caractériser la signification, qui peut être importante, légère, etc. ces critères étant souvent choisis par les auteurs selon la valeur de p. Il est réellement crucial comme le souligne Poitevineau de préciser que

« le test ne dit rien quant à l'intensité de l'effet parent. Un résultat significatif n'est qu'une indication de l'existence de l'effet supposé ; un résultat non significatif un constat d'ignorance ». (2004)

Ces différents éléments montrent combien il est important d'accompagner la compréhension de ces différents concepts que sont la variabilité, la représentativité et la significativité au sein de l'ensemble de la population. Et qu'une méconnaissance de ces éléments induit une difficulté à la compréhension des informations issues d'études statistiques. La quantification de l'information diffusée au public nécessite une prise de conscience du travail collectif à mener auprès de la population et la nécessité du développement de l'exemplarité.

D. La place de la statistique dans la compréhension du monde !

Le citoyen d'aujourd'hui se doit de comprendre et d'interpréter les données statistiques qui lui sont régulièrement communiquées par tout un ensemble de

média (Gattuso, 2011). En tant qu'acteur de sa société, il se doit d'être en capacité de faire la part des choses qui lui sont communiquées, de faire la différence entre les éléments de statut vérifiables et ceux pouvant être vraisemblables.

La place de la statistique comme instrument de décision questionne la place et le rôle de cette science dans la recherche scientifique et dans la formation universitaire au sens large. La statistique est une science à la frontière entre différents domaines et d'aide à la compréhension (Académie des sciences, 2000). Il est donc nécessaire de former le citoyen à la compréhension, et le communicant à la transposition de sa connaissance pour la diffusion du savoir.

1. Communiquer

Les informations sont de plus en plus souvent traduites en données numériques, représentant souvent un indicateur descriptif, mesurant une association, une hétérogénéité, obtenue à partir d'un ensemble de données. Ces données sont elles-mêmes recueillies selon des procédures plus ou moins complexes et plus ou moins explicitées ; faire appel à la connaissance de la démarche statistique ou des outils statistiques est donc nécessaire tant pour évaluer le contexte du recueil de l'information (sélection de l'échantillon, choix des outils et modes de recueil, ...), que pour juger de la pertinence dans les choix de modèles théoriques sous-jacents à l'analyse faite et de l'interprétation qui est faite des résultats.

La diffusion des informations issues d'études statistiques va nécessiter le développement de la capacité de communication de tous les acteurs à des fins de compréhension du ou des publics visés. Les éléments fondamentaux sur lesquels il est nécessaire de travailler la transmission sont la présentation dans un premier temps de la démarche puis celle des résultats. Il est impératif d'insister sur le fait que tout résultat statistique se définit au niveau d'une population et non à un niveau individuel, et sur le statut incertain des résultats obtenus à partir de telles enquêtes, qui ne remet néanmoins pas en cause la crédibilité de l'information. Il ne s'agit pas d'un résultat en tout ou rien mais bien d'un élément d'aide à la décision. Cet élément est clairement un obstacle à la compréhension et nécessite un travail

d'explicitation à ne pas omettre. Au final, il s'agit d'aider le citoyen à s'y retrouver dans cette forêt d'informations pouvant parfois sembler se contredire.

Comme l'indique Gauvrit, « la statistique a cette particularité extraordinaire d'être convaincante même pour des gens qui affirment ne pas la comprendre » (Gauvrit 2007). Vu ainsi, il s'agit donc d'une science extrêmement dangereuse qui jouerait sur un sentiment de croyance et non pas de réflexion. L'enseignant a donc pour mission de donner les éléments et des stratégies de compréhension.

2. Place de l'apprentissage pour former le citoyen

Il y a donc nécessité de pouvoir introduire ces notions tout au long de la formation. C'est d'ailleurs dans ce cadre que l'association des statisticiens américains (ASA) a rédigé le rapport GAISE (Franklin, 2007) qui propose le cadre d'un curriculum statistique de la maternelle à la terminale, soulignant par exemple l'importance de la présentation des différents types de variabilité tout au long de la formation afin de permettre la conception de l'incertitude.

Dans le cadre de la formation universitaire, il s'agit de former les étudiants à la compréhension des résultats scientifiques concernant leur domaine de connaissance, à la capacité de mettre en œuvre un certain nombre de méthodes statistiques employées dans le cadre de leur communauté scientifique et professionnelle mais aussi de les former à la capacité de réaliser une analyse critique des informations et articles qu'ils seront amenés à lire tout au long de leur vie professionnelle et scientifique. Le développement de ces capacités passera par une connaissance précise des contraintes associées aux différentes méthodes utilisables.

Il y a nécessité à tous niveaux de s'assurer que nos citoyens comprennent les messages diffusés, et tout effort de communication est un élément favorisant la compréhension. Expliciter, tout particulièrement, le cadre de la statistique auprès de l'ensemble des publics est un élément majeur.

La statistique est une science qui, à ce jour, semble s’immiscer dans tous les espaces de formation mais aussi de la société dans son terme le plus global. Il est presque devenu impossible à un étudiant universitaire futur statisticien ou non de faire l’impasse sur un nombre d’heures d’enseignement en statistique pouvant varier de quelques heures à des unités d’enseignement spécifiques. Cet enseignement peut être isolé et vu comme un enseignement disciplinaire, à part entière, ou être une sous-partie d’enseignement disciplinaire que ce soit avec une approche économique, biologique, sociologique ou médicale.

Les discussions qui peuvent naître entre deux chercheurs de disciplines différentes travaillant sur un même sujet traduisent rapidement le fait que chacun « s’accroche » à des notions ou des concepts différents dans le champ de la statistique.

E. Synthèse

La statistique, comme toute discipline, a des fondements nécessaires à son bon usage, et il ainsi central qu'ils soient compris et connus de tous les usagers, mais aussi expliqués à l'ensemble de la population confrontée plus que jamais à une diffusion d'informations quantitatives issues pour nombre d'entre elles d'une démarche statistique. Appliquer une approche statistique ne signifie pas utiliser des outils statistiques pour manipuler des données recueillies pour estimer des paramètres de tendance centrale ou de dispersion, ou faire « tourner » des modèles, mais il s'agit bien de respecter une démarche statistique dans un esprit statistique, c'est-à-dire en travaillant dans un respect de l'incertitude, dans le renoncement à la vérité au bénéfice de l'idée de vraisemblance. Plusieurs éléments sont fondamentaux dans le cadre de la mise en œuvre d'une approche statistique de données : concevoir les différentes dimensions de la variabilité omniprésente, comprendre les processus de recueil d'informations et la dangerosité de l'usage du terme représentatif de par la diversité de compréhension de ce terme, analyser avec précaution les données sans jamais oublier l'importance de la description et la connaissance du contexte de mise en œuvre de tests statistiques, afin d'employer et d'interpréter les résultats avec pertinence, en spécifiant les situations de description ou de prise de décision.

CHAPITRE 3 – DIFFÉRENTES SITUATIONS DE LA DIDACTIQUE DE LA STATISTIQUE

Dans les activités d'un enseignant-chercheur, il y a des temps particulièrement importants et nous avons choisi de nous arrêter sur deux d'entre eux : la thèse et l'expertise dans le cadre d'une sollicitation de politiques. Il s'agit de définir et discuter, pour chacune des situations, tout particulièrement l'environnement et le changement de contexte.

Dans le cadre des thèses, la communauté dans laquelle se déroule la transposition est constituée de scientifiques du même environnement, nous avons dans ce cas une communauté constituée autour de savoirs partagés. Le but du doctorant consiste à faire reconnaître son travail personnel, sa thèse, pour être lui-même reconnu dans cette communauté non plus en tant qu'étudiant mais en tant que chercheur. Il se doit dès lors d'utiliser le cadre de transmission de son savoir usité dans le domaine dans lequel il souhaite se faire reconnaître. Dans le second cadre, la situation est différente car il s'agit de la demande faite par des politiques d'expertise à un scientifique, à laquelle il répond en réalisant une explication de situation et de démarche pour accompagner sa réponse. Dans les deux situations nous nous intéressons particulièrement aux enjeux liés à la diffusion de démarches ou contenus statistiques.

A. La statistique au service de la reconnaissance d'un jeune chercheur par une communauté de chercheurs

La thèse de doctorat est le premier objet de reconnaissance des qualités scientifiques du doctorant par ses pairs scientifiques ; la validation de son doctorat passera donc par l'évaluation par le jury de son écrit principalement et de son oral. Il s'agit dès lors pour le doctorant de rédiger un écrit mettant en valeur ses acquis et ses compétences pour rentrer dans le milieu des chercheurs. La science statistique peut

ou non être présente dans le cadre de sa thèse, peut-être pour aider à développer des éléments d'argumentation de la question de recherche posée, ou dans le cadre de la mise en œuvre d'une démarche statistique lui permettant de répondre à sa question. Quels sont dès lors les usages de la statistique qu'il présentera dans le cadre de ce travail comme point d'étape et de passage entre le milieu du novice à celui du monde des chercheurs ? Peut-on ainsi reconnaître des régularités d'usage inter ou intra discipline ? La statistique est-elle à ce point protéiforme pour en reconnaître les besoins et les outils ?

1. Choix du champ de recherche

« il faut étudier (ndlr selon Latour en reprenant Collins) les sciences actuelles en train de se faire, et en pleine controverse, de façon à sortir définitivement du confort intellectuel des historiens qui arrivent comme les carabiniers de la chanson toujours un peu trop tard. Au lieu d'étudier les sciences « sanctionnées » il faut étudier des sciences ouvertes et incertaines. » (Latour, 2008)

Approcher la question des usages et de la formulation de la science statistique par l'étude des corpus de thèse était un choix effectué pour approcher la pluridisciplinarité sans avis *a priori* des usages mais dans un contexte de recherche. Comme l'indique Latour, il est intéressant d'étudier les sciences en construction et non sanctionnées car ainsi nous pouvons palper l'hétérogénéité de la production avant le formatage induit par les cultures des communautés de recherche et les contraintes éditoriales et de diffusion pouvant étouffer et normaliser les créations rédactionnelles, mais aussi sélectionner les éléments diffusés, comme nous avons pu le spécifier avec l'hégémonie de la p-value. Le niveau de précision est ainsi réduit, et le résultat de tout ceci est une perte de formulation et une standardisation de l'information diffusée.

Les thèses sont donc les écrits présentant la production scientifique de novices souhaitant intégrer la "cour des grands". Il ne s'agit pas uniquement d'un travail pour leurs propres satisfactions personnelles mais bien en quelques sortes d'une dépersonnalisation de leurs savoirs produits. Ils sont dans l'attente de la reconnaissance de leur travail par leurs pairs scientifiques respectifs, et donc de la reconnaissance de leurs productions scientifiques comme pouvant venir nourrir

l'ensemble des savoirs de leurs communautés de rattachement souhaité (Lonka et al. 2014). L'ensemble des thèses correspond en quelque sorte à un espace de sciences ouvertes et pour lesquelles l'avenir est complètement incertain. Ces docteurs deviendront-ils des grands chercheurs ? Peut-être mais nombreux sont aussi ceux qui ne fréquentent la recherche que le temps de leur thèse. Néanmoins, en France, cette reconnaissance doit aussi passer par une seconde étape : la reconnaissance au CNU. Il s'agit dans ce cas d'une reconnaissance d'appartenance à une « tribu » particulière et disciplinaire.

Etudier les usages dans les thèses, c'est aussi voir le résultat de tout un processus de formation mis en œuvre dans le cadre de la formation du supérieur, et cela peut soulever des pistes de recommandations pour la construction de l'intégration de l'enseignement de la statistique dans l'offre de formation.

Le temps de la rédaction de la thèse scientifique est un temps important pour le doctorant car ce document va constituer la trace de sa contribution à la science ou plus largement à l'ensemble de la société potentiellement. Il s'agit d'une production écrite constituant la passerelle entre la période d'apprentissage et le devenir du scientifique reconnu en tant que tel. Il s'agit de l'œuvre personnelle du doctorant, de l'élément lui donnant l'opportunité d'être reconnu. Cet objet nous intéresse en tant que traces laissées par le doctorant relativement à cet événement que représente la période de doctorat.

2. **Se faire reconnaître dans sa communauté**

a) *Premier pas : la thèse*

Cette production doit permettre au doctorant d'être reconnu par les pairs grâce à la diffusion de sa production originale et l'évaluation de sa pertinence en regard des connaissances déjà existantes et partagées.

Dans cette période initiatique, plusieurs temps de reconnaissance vont exister avec différents éléments à produire. Il s'agit de la soutenance de thèse, la présentation en colloque/congrès et la rédaction d'articles. Certaines productions sont

plus individuelles que d'autres. Ainsi la rédaction de la thèse est principalement un travail individuel même s'il s'appuie sur des temps de travaux collaboratifs, principalement en association avec le directeur ou les directeurs de thèse. Cette relation de travail s'exprime sur les trois années, voire plus, de doctorat, et elle évoluera dans le temps avec des relations didactiques pouvant passer du vertical à l'horizontal. Le doctorant poursuit tout au long de sa période de thèse une période de formation, qu'il s'agisse de formations spécifiques à son domaine de recherche ou de formations lui permettant d'acquérir de nombreuses connaissances et compétences transverses et variés afin de le préparer à l' « après thèse ».

Il s'agit donc d'une phase de constitution d'un savoir savant, ou plus exactement de la formulation discursive de ce savoir. Cela implique la contrainte de la dépersonnalisation, pour le partage nécessaire de cet apport scientifique avec la communauté permettant ainsi son élévation au statut de savoir savant, reconnu et validé. La constitution des jurys de thèse par le directeur de thèse et le doctorant correspond à la mise en scène du scénario de reconnaissance. Quels sont les courants de pensée les plus sensibles au travail réalisé par ce doctorant ? Quels sont les personnalités les plus sensibles aux concepts sur lesquels il s'est appuyé pour développer les éléments de savoir de sa thèse ? Qui sont les chercheurs qui pourront aider ce futur docteur dans le déroulement de sa carrière par leur jugement d'intérêt ? Et quelles sont les disciplines portées par les membres du jury au sens CNU ? Ces questions ne concernent pas directement notre travail mais l'enjeu pour le doctorant est de se faire apprécier par la qualité de son travail, l'outil utilisé dans ce cadre étant un document rédigé dans lequel sera présenté l'ensemble du cheminement de son travail de doctorat.

b) Les autres étapes de reconnaissance

Cette étape est fondamentale comme premier temps de reconnaissance, car, dans le système français elle n'est pas la seule. Si le doctorant veut devenir enseignant, il devra demander son rattachement à une ou plusieurs disciplines du CNU²⁶. Il y a 11 disciplines divisées en 52 sections. Il s'agit donc d'être reconnu

²⁶ CNU Le Conseil national des universités est une instance nationale régie par le [décret n° 92-70 du 16 janvier 1992](#). Il se prononce sur les mesures individuelles relatives à la qualification, au recrutement et à la carrière des professeurs des universités et des maîtres de conférences régis par le [décret n°84-431 du 6 juin 1984](#) fixant les

dans ce cadre plus spécifiquement par les communautés disciplinaires, qui ont leurs propres règles et qui reconnaissent la capacité à la personne de pouvoir enseigner et chercher. Ceci correspond à la situation de l'écologie du système didactique, c'est-à-dire l'habitat dans lequel le futur docteur souhaite pouvoir nicher.

Un autre élément de reconnaissance de l'intérêt du travail du doctorant est la capacité à être accepté pour présenter un poster ou une conférence lors d'un congrès ou colloque avec comité scientifique. Le comité scientifique jugera de la pertinence et de l'apport du travail du doctorant pour la ou les communautés représentées lors de cette manifestation. Enfin le dernier niveau de reconnaissance par les pairs est l'acceptation pour diffusion d'un article dans une revue à comité de lecture. Il s'agit de la reconnaissance suprême, tout particulièrement lorsque la revue a un bon "facteur d'impact". Ces facteurs d'impact mesurent la visibilité d'une revue, et peuvent être vus par certains comme des indicateurs de qualité de la revue. Un article publié dans une revue à fort facteur d'impact est perçu comme un article ayant été jugé comme pouvant attirer des lecteurs par le comité éditorial et ayant respecté les consignes pour les auteurs (*Author guidelines*). Bien que beaucoup décriée, la prise en compte des facteurs d'impact dans le cadre du développement de carrière d'un chercheur a une influence plus ou moins importante selon les disciplines. Ils sont cependant utilisés comme critère pertinent d'évaluation par les adeptes d'une évaluation *quantitative* de la recherche.

c) *Des reconnaissances interconnectées*

Durant la période de doctorat, la plupart du temps, le doctorant aura participé à des colloques et pourra avoir contribué à la rédaction d'article. Dans ce dernier cas, il peut alors avoir présenté sa thèse à partir de ces articles, et dès lors, ces articles en constituent le corpus principal ; l'architecture de ce type de thèse est alors très différente, s'appuyant en majeure partie sur des documents ayant déjà reçu un label de reconnaissance lors de la publication des articles. Il faut souligner que cette reconnaissance peut avoir été faite dans un environnement de chercheurs différent de celui, ou ceux, des membres du jury.

dispositions statutaires communes applicables aux enseignants-chercheurs et portant statut particulier du corps des professeurs des universités et du corps des maîtres de conférences.

Le directeur de thèse, dans cette relation duale avec son doctorant, est le représentant de la communauté scientifique à laquelle le doctorant souhaite vraisemblablement se rattacher. Il accompagne donc le thésard vers des horizons pertinents de recherche pouvant être d'intérêt pour la communauté qu'il représente. L'accompagnement de cette période se traduit de manière simpliste par le schéma 4 ci-dessous.

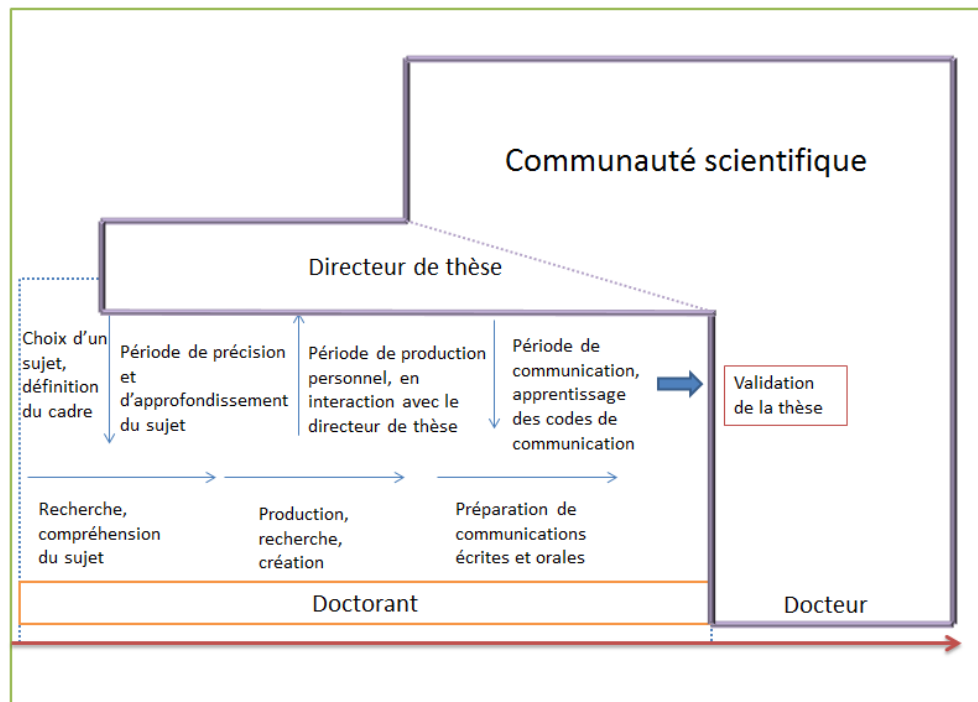


Schéma 4 - Cheminement dans le cadre de la thèse

Le sujet de recherche nécessite tout d'abord d'être défini avec précision, et impliquera un travail de l'état de l'art et la construction d'une réflexion induite. Ce travail sera mené par le thésard en interaction étroite avec le directeur. La phase de production, création, découverte mise en œuvre débutera et nécessitera dès lors des phases d'échanges de discussion. Quand des éléments intéressants semblent présents, il s'agit de pouvoir les confronter à des jugements extérieurs ; un travail de diffusion scientifique est alors en marche nécessitant une appropriation par le doctorat des codes singuliers de la communauté mais aussi de ceux des différents supports.

3. La place de la statistique dans les thèses

Les éléments de statistique présents dans les thèses, sont des indicateurs de la place que joue la statistique dans la discipline ou au sein de la communauté scientifique de rattachement du doctorant et de son directeur de thèse.

Les différentes influences pouvant jouer dans l'usage ou non d'objets de savoirs statistiques dans le cadre de la thèse peuvent être de plusieurs ordres. La thèse est la production d'une personne et est ainsi la mise en lumière de ses compétences²⁷ associant ses savoirs, ses connaissances, ses savoir-faire dans le cadre de la recherche. Les influences transparaissant dans la rédaction de la thèse sont donc issues de ce qui a été appris durant l'ensemble de ses années d'étude, il s'agit dès lors d'une influence académique, mais aussi de ce qui a été acquis par l'apprentissage qui s'est effectué pendant la période de doctorat, plutôt focalisé sur l'environnement de recherche et sur un écosystème bien spécifique incluant la proximité avec les porteurs scientifiques directs de la science.

Il s'agit dans ce travail de repérer, au sein de chaque thèse, les éléments se référant à la science statistique, ces éléments constituant dès lors le corpus d'étude qui a été étudié. Le choix ayant été fait, pour des raisons exposées ultérieurement, de diversifier le plus possible les disciplines, il n'est pas envisageable de se pencher sur les interprétations scientifiques.

Nous espérons avoir pu assurer la distanciation nécessaire à ce travail.

Il s'agissait dans le cadre de ce travail d'étudier les régularités et les dissemblances d'usage de démarche statistique au sein de l'objet thèse, mais aussi de repérer les spécificités disciplinaires. Cette étude est descriptive et nous y avons décrit les démarches et techniques statistiques utilisées, ainsi que les outils utilisés et les niveaux de preuve appréciés dans le cadre de mise en œuvre. Nous avons aussi relevé les usages de terminologies et les spécificités linguistiques en fonction des disciplines, ainsi que la formalisation des présentations.

²⁷ Compétence à considérer au sens de Tardif : « Un savoir-agir complexe prenant appui sur la mobilisation et la combinaison efficaces d'une variété de ressources internes et externes à l'intérieur d'une famille de situations. » (Tardif, 2006, p. 22)

a) Méthodes d'exploration

(1) Terrain d'étude

Le terrain d'étude sur lequel nous avons travaillé correspond à l'ensemble des thèses dans le cadre des écoles doctorales des établissements sur le site de Bordeaux. Les thèses sélectionnées ont été choisies dans la population de thèses soutenues entre janvier 2012 et décembre 2013, effectuées au sein d'une école doctorale rattachée à l'une des Universités de Bordeaux (Université de Bordeaux 1, Université Bordeaux 2 - Bordeaux Segalen, Université Bordeaux Montaigne – Université Bordeaux 3 et Université Montesquieu - Bordeaux 4)²⁸ et déposées sur "theses.fr" quel que soit le domaine disciplinaire.

Nous avons privilégié la diversité disciplinaire des thèses, car seuls des *a priori* d'usage auraient pu nous conduire à un choix et il s'agissait, bien au contraire, de développer une approche d'observation objective. Nous pourrions envisager dans le prolongement de ce travail de nous centrer sur certaines disciplines pour une approche plus spécifique à un environnement. Dans le cadre de ce travail, il n'était pas envisageable d'étudier l'ensemble des thèses correspondant aux critères retenus pour des raisons de faisabilité et nous avons dû sélectionner un échantillon de thèses. Les thèses ont été repérées à partir du moteur de recherche "theses.fr"²⁹ à partir de leur rattachement à une des écoles doctorales bordelaises. Il faut préciser que les disciplines sont différentes entre les écoles doctorales et souligner également qu'une école doctorale peut couvrir des champs disciplinaires très variés. Prenons l'exemple de l'école doctorale « Société, sciences politiques et santé publique », de ex-l'Université Bordeaux Segalen, qui englobe des disciplines aussi variées que la sociologie, l'épidémiologie, la psychologie, la biostatistique avec des sujets de thèses tels que « *Education et discipline au collège* » (Garcia, 2013), « *Comportements dangereux au volant chez les jeunes. Le rôle des mécanismes d'influence sociale dans le groupe* » (Gheorghiu, 2012), « *Capacités de*

²⁸ En janvier 2014, l'université de Bordeaux a vu le jour suite à la fusion de trois des universités bordelaises à l'exception de l'université de Bordeaux Montaigne. Cette thèse a été débutée durant le chantier de fusion.

²⁹ Moteur de recherche, vise à donner un point d'entrée commun pour toutes les thèses soutenues en France, et celles en cours de préparation : <http://www.theses.fr/>

réserve, vieillissement cérébral et maladie d'Alzheimer » (Foubert-Samier, 2013), etc. Pour ces différentes raisons, nous avons choisi de réaliser un échantillonnage permettant de prendre en compte les différentes écoles doctorales pour assurer une diversité des disciplines au sein de notre échantillon et donc la variabilité d'usage.

L'intérêt de l'ouverture à l'ensemble des disciplines étant bien évidemment de pouvoir satisfaire à l'objet même de l'étude et donc d'être en capacité d'étudier l'impact plus ou moins prégnant de la culture disciplinaire, mais aussi les usages partagés.

(2) Quelle sélection de thèses ?

Le nombre de thèses soutenues, entre le 1 janvier 2012 et le 30 décembre 2013 dans une des écoles doctorales des 4 universités de Bordeaux était de 612.

Une sélection a été réalisée avec un taux de sondage global de 12%³⁰ en stratifiant selon les écoles doctorales afin d'assurer l'hétérogénéité disciplinaire au sein de l'échantillon d'étude (cf. tableau 3).

Les thèses rédigées dans leur totalité dans une langue autre que le français n'ont finalement pas été étudiées dans le cadre de ce travail, l'analyse des usages, à partir d'un support rédigé, ne pouvait être réalisée qu'à partir de thèses rédigées dans la même langue. Cela représente trois thèses parmi les 76 sélectionnées.

Néanmoins, les thèses comprenant dans le corps du texte des publications scientifiques, en anglais ont été conservées pour l'analyse, afin d'étudier la place différenciée de l'analyse statistique entre les différentes parties de la thèse.

³⁰ Il avait été évalué la faisabilité de réaliser une recherche textuelle sur un peu plus de 60 thèses avec plus de 10 par université.

Tableau 3 : Répartition des thèses soutenues entre le 1^o janvier 2012 et le 31 décembre 2013 dans une des écoles doctorales des universités de Bordeaux et de l'échantillon d'étude

Ecoles doctorales	Effectif total	Echantillon sélectionné		Hors critères d'inclusion		Inclus
		N	n	%	n	
Total général	612	76	12%	3	4%	73
université Bordeaux 1	319	39	12%	2	5%	37
<i>Mathématiques et Informatique</i>	65	7	11%			7
<i>Sciences chimiques</i>	80	9	11%	1	11%	8
<i>Sciences physiques et de l'ingénieur</i>	127	19	15%	1	5%	18
<i>Sciences de la vie et de la santé</i>	9	1	11%			1
<i>Sciences et environnements</i>	38	3	8%			3
université Bordeaux II - Segalen	150	17	11%	1	6%	16
<i>Sciences, technologie, santé</i>	79	8	10%			8
<i>Sociétés, Politique, Santé publique</i>	71	9	13%	1	11%	8
université Bordeaux 3 - Montaigne	86	10	12%			10
université Bordeaux 4 - Montesquieu	57	10	18%			10
<i>Droit</i>	18	2	11%			2
<i>Science politique</i>	10	4	40%			4
<i>sciences économiques, gestion et démographie</i>	29	4	14%			4

Dans un premier temps, nous avons appliqué une méthode d'interrogation de ce corpus de thèses. Et nous avons localisé des extraits illustrant notre thème « science statistique » à partir de formes que nous avons sélectionnées en tant que marqueurs d'une pratique ou démarche statistique.

Pour chaque thèse sélectionnée, une étude de corpus a été réalisée afin de repérer la présence ou non d'une activité statistique ou tout simplement d'informations, de données statistiques au sein de la revue de la littérature par exemple.

Il s'agissait donc de travailler dans un premier temps par une approche de la terminologie employée et de ses spécificités.

Comme l'indiquent Silvia Pavel et Diane Nolet dans leur « Précis de terminologie », « *le mot terminologie signifie un ensemble de mots techniques appartenant à une science, un art, un auteur ou un groupe social* » (Pavel, 2002). Il s'agit dans ce travail de pouvoir repérer précisément ces usages et d'en étudier les stabilités et hétérogénéités selon l'appartenance à une communauté, donc le domaine d'utilisation.

Les disciplines connexes et les technologies convergentes peuvent se partager certains concepts et les termes qui les désignent. Parfois, le même concept peut recevoir des désignations différentes selon le domaine d'emploi où le même terme peut désigner des concepts différents dans d'autres spécialités. (Pavel, 2002)

Les termes sont-ils utilisés au sens de la langue commune, celle dont on se sert dans le quotidien, ou plutôt en tant qu'usage dans la langue spécialisée, communication sans ambiguïté dans un domaine particulier du savoir ou de la pratique ? Il s'agit ainsi de repérer un vocabulaire et des usages linguistiques qui lui sont propres. Ces usages peuvent-ils ainsi spécifier des épistémologies différentes ?

(3) Choix de l'algorithme de recherche de mots

La forme d'entrée qui se propose naturellement est le thème lui-même. Nous avons donc débuté la recherche par le suffixe "statis"³¹, afin de pouvoir couvrir les termes : "statistique", "statistiques", "statistiquement" et "statisticien" mais aussi les termes anglais "statistics", "statistical" ou "statistically", principalement pour les thèses construites à partir d'articles publiés dans des revues scientifiques internationales. Cette première recherche permet de repérer rapidement les thèses ayant ou non l'usage de notions statistiques dans leur textes. Cependant au cours de ce repérage, nous avons pu nous apercevoir que la non-présence de "statis" dans le corps du texte ne signifiait pas systématiquement le non-usage d'objet de la science statistique, car la démarche statistique peut être mise en œuvre sans que l'auteur le spécifie ; pour cette raison, il nous est apparu nécessaire de pouvoir compléter cette première recherche par la recherche systématique de quelques autres termes, caractérisant les éléments fondamentaux de la statistique.

Ainsi, une recherche sur la notion de variabilité a été effectuée en se référant à sa place cruciale en tant que fondatrice de la statistique. Afin d'assurer une recherche exhaustive de la notion de variabilité, nous avons choisi de rechercher les mots ayant la racine "varia", non pour le sens de ce mot, mais plus spécifiquement pour la possibilité de repérer ainsi l'ensemble des termes suivants : "variable", "variabilité", "variance", "variablement" et "variation".

³¹ Nous ne nous intéressons pas ici à l'étymologie du terme statistique.

Pour les thèses ne répondant pas aux critères de recherche ci-dessus, avant de les exclure définitivement de la recherche et pour s'assurer de l'absence d'une démarche statistique (au sens travail de compréhension de données statistiques ou développement d'un travail statistique), une lecture du résumé et de la table des matières a été réalisée.

Pour les thèses retenues, un relevé de corpus a été réalisé, en suivant un algorithme de recherche de termes. Pour chaque mot présenté ci-dessous, une recherche a été réalisée sur l'ensemble de la thèse, les occurrences ont été lues dans leur contexte afin d'évaluer s'il s'agissait d'une occurrence employée dans le cadre d'une démarche statistique, ou d'une interprétation de résultats statistiques. Sous ces conditions, la phrase ou le paragraphe incluant le terme était reproduit dans un fichier en repérant le numéro de page afin de pouvoir assurer le retour possible à la source.

Les termes utilisés dans l'algorithme couvrent les champs qui ont été présentés dans le chapitre précédant et qui sont caractéristiques d'une démarche statistique :

- "*statis*" : les raisons de ce choix ont été exposées précédemment,
- "*échantillon*" : afin de contextualiser les données recueillies, il est important d'avoir une bonne connaissance de la méthode de sélection des sujets étudiés et de la composition de l'échantillon. Les mots "échantillon", "échantillonné" et "échantillonnage" ont été ainsi repérés. La spécification de la méthode d'échantillonnage définit souvent le type d'étude réalisée. La caractérisation de la stratégie d'échantillonnage est un élément devant faire partie des meta-données d'un fichier comme cela a été justifié,
- "*moyenne*" : il s'agit de l'indicateur de tendance centrale le plus fréquemment employé en statistique. Ce terme est très fréquemment utilisé dans des formulations de type "en moyenne" qui n'ont pas été retenues,
- "*varia*" ; ce n'est pas tant le mot "varia" qui a été recherché mais tous les mots construits à partir de cette racine ("variance", "variabilité", "covariance", etc.), la notion de diversité étant à la base de la plupart des

développements statistiques,

- "*regress*" et "*correl*" : il s'agissait de repérer les termes construits sur ces deux racines ("régresser" et "régression", et "corrélér" et "corrélation") et traduisant l'intérêt pour des études d'associations et pour la possible mise en œuvre de modélisation. Les occurrences dans des formulations utilisant les termes "régresser" et "régression" au sens de diminuer n'ont pas été retenues.
- "*Gauss*" : il s'agissait de repérer la présence ou non de référence à des lois classiques, les hypothèses de variables suivant des lois de Gauss ou loi normale étant des conditions d'applications d'un certain nombre d'objets statistiques. Cependant, les usages linguistiques dans certaines disciplines conduisant à l'usage du groupe nominale « *loi normale* » à la place de "*loi de Gauss*", il a été décidé de mener une recherche complémentaire sur le terme "normal". Cela permettait de repérer, par exemple, des usages de normalisation (activité fréquente dans la manipulation de données permettant de standardiser les données),
- "*parametre*" et "*estim*" : ces deux termes étaient recherchés parallèlement, car l'étude du terme « paramètre » n'avait d'intérêt pour notre étude que s'il s'agissait d'une recherche d'estimation de paramètre. "*Estim*" a été utilisé pour être en mesure de repérer les emplois des termes "estimer", "estimateur" et "estimation",
- "*proba*" et "*aleatoire*" : il s'agissait de repérer dans ce cadre si des associations ou rapprochement avec les notions de probabilités avaient été réalisées. Le terme "hasard" a été ajouté dans un second temps comme réajustement d'analyse,
- "*représenta*" : ce terme permet de repérer les mots tels que "représentativité" et "représentatif", qui traduisent une notion importante pour la diffusion des résultats statistiques, en caractérisant les échantillons repérés précédemment. Cette racine nous permettait aussi de repérer le terme "représentation" et de sélectionner les occurrences de l'acception « *représentations graphiques* »,
- "*significat*" : il s'agit de repérer les termes "significativité",

"significativement", "significatif" et "significative" qui correspondent à l'idée d'obtention de résultats vraisemblables et plausibles, pouvant être des éléments d'aide à la décision.

Afin de compléter cette recherche, nous avons recueilli les informations concernant les usages de logiciels ayant permis la mise en œuvre de situations statistiques. Puis, afin de s'assurer de ne pas avoir omis d'éléments importants et de compléter les éléments de praxéologie mis en œuvre dans le cadre de ces travaux de recherche, une lecture transversale de la thèse a été réalisée tout particulièrement en s'appuyant sur les résumés et sommaires et en faisant certains zoom de lecture sur les parties dont le titre contenait un des termes repérés par l'algorithme.

Dans l'analyse du corpus, la source des *verbatim* reproduits indiquera le nom de l'auteur, l'université dans laquelle il a soutenu sa thèse (B1 pour université de Bordeaux 1, B2 pour université Bordeaux 2 Segalen, B3 pour université Bordeaux 3 Montaigne, et B4 pour université Bordeaux 4 Montesquieu) complétée de la discipline dans laquelle la thèse a été soutenue.

4. Analyse de corpus

a) Sélection des thèses pour l'analyse de corpus

Parmi les 73 thèses répondant aux critères de sélection, 18 ne comportaient aucune notion de démarche statistique, dont 11 aucune occurrence des termes de l'algorithme. Les 7 autres thèses intégraient de rares références à la statistique, quelques données de moyennes et de fréquences ainsi que, pour certaines, la présentation d'élaboration d'échantillonnage (dans une approche d'expérimentation) mais pas de suivi de démarche et d'analyse statistiques.

Cependant, la non présence de démarche statistique ne signifie pas pour autant un manque d'intérêt porté à la question de l'importance de données dans la démarche de recherche, mais peut parfois traduire, tout simplement, l'inexistence de données recueillies.

« Le manque de données, de recherches et de statistiques sur les spectateurs et les potentiels spectateurs du cinéma en général et des cinémas sud-africains en

particulier est une entrave infranchissable au développement des publics et à la connaissance du marché local. » (Landau, 2012)

Certains auteurs justifient l'absence d'une démarche statistique simplement en soulignant la méfiance nécessaire quant aux usages des techniques statistiques, de par leurs limites.

« Les recherches quantitatives et les techniques statistiques sont évitées et critiquées.(...) Le théoricien historico-structural Barrington Moore a attaqué l'influence de la révolution behaviouriste, mis l'accent sur les données statistiques dans un grand nombre de recherches et souligné les limites des méthodes statistiques. » (Ofrim, 2012)

Pour la suite de ce travail, nous considérerons plus spécifiquement les 55 autres thèses. L'analyse du corpus a permis de conclure à un usage de la statistique soit dans un objectif d'explicitation de l'état de la question et du contexte de la problématique ou pour l'argumentation même de la thèse soutenue pour trois thèses sur quatre. La place de la statistique au sein de ces 55 thèses étant très variables d'une thèse à l'autre et cela indépendamment de la discipline.

Parmi les 55 thèses, 29 étaient des thèses inscrites dans l'une ou l'autre des écoles doctorales de l'université Bordeaux 1 (dans la population initiale plus de 50% des thèses de Bordeaux étaient affiliées à l'Université de Bordeaux 1 : 319/612), 15 de l'Université Bordeaux 2, cinq de l'Université Bordeaux 3 et six de l'Université Bordeaux 4. (cf. tableau 4).

Tableau 4 : Répartition des thèses sélectionnées pour l'étude selon leur appartenance à des écoles doctorales des universités de Bordeaux et la place de la statistique dans le corpus

Ecoles doctorales	Effectifs	place de la statistique					
		sans référence à la statistique		avec peu de référence		des références à la statistique	
		n	%	n	%	n	%
Total général	73	18	25%	11	15%	44	60%
Bordeaux 1	37	8	22%	9	24%	20	54%
<i>Mathématiques et Informatique</i>	7	1	14%	2	29%	4	57%
<i>Sciences chimiques</i>	8	2	25%	2	25%	4	50%
<i>Sciences physiques et de l'ingénieur</i>	18	4	22%	4	22%	10	56%
<i>Sciences de la vie et de la santé</i>	1	0	0%	0	0%	1	100%
<i>Sciences et environnements</i>	3	1	33%	1	33%	1	33%
Bordeaux 2	16	1	6%	0	0%	15	94%
<i>Sciences, technologie, sante</i>	8	0	0%	0	0%	8	100%
<i>Societes, Politique, Sante publique</i>	8	1	13%	0	0%	7	88%
Bordeaux 3 - Montaigne humanités	10	5	50%			5	50%
Bordeaux 4	10	4	40%	2	20%	4	40%
<i>Droit</i>	2	1	50%	1	50%	0	0%
<i>Science politique</i>	4	3	75%	1	25%	0	0%
<i>sciences économiques, gestion et démographie</i>	4	0	0%	0	0%	4	100%

b) *Doctorants et statisticiens, quel partenariat ?*

Il est important de souligner que, parmi les 55 thèses étudiées, 14 d'entre elles contenaient dans le corps du texte l'intégralité des articles soumis ou publiés dans une revue scientifique. Cette particularité de rédaction se retrouvait principalement dans les thèses des écoles doctorales de l'université de Bordeaux 2 Segalen avec huit thèses sur 15.

L'étude des corpus de thèses rédigées à partir d'articles a été plus difficile car la rédaction d'articles, particulièrement dans les domaines scientifiques, fait l'objet d'un travail collaboratif pour lequel la part spécifique du doctorant est plus difficile à identifier. De plus les consignes aux auteurs des différentes revues conduisent à une régularité de présentation et une homogénéisation dans les formulations. Tout publiant a été amené à construire son article tant sur la forme mais aussi parfois sur le fond en fonction de la politique et des habitudes de la revue à laquelle il souhaitait soumettre son article. Dans la présentation des *verbatim* ci-après nous spécifierons dès lors qu'il s'agira d'extraits d'articles. Tous les articles présents dans le corps des thèses étaient rédigés en anglais, les doctorants

introduisaient les articles par une présentation plus ou moins succincte du contexte, et proposaient après l'article une présentation synthétique des principaux résultats.

Parmi les 55 thèses, neuf doctorants bien qu'usant de notions statistiques et même ayant une réelle démarche statistique n'utilisent à aucun moment de leur thèse le terme statistique. Est-ce que comme Monsieur Jourdain, le doctorant pourrait déclarer si nous l'en informions « *Par ma foi, il y a plus de trois ans que je fais de la statistique sans que j'en susse rien, et je vous suis le plus obligé du monde de m'avoir appris cela* ». Ou le sait-il déjà mais ne l'a-t-il pas indiqué considérant que l'explicitier n'apporterait rien à son travail, ou qu'au contraire cela nécessiterait d'être en capacité de prouver des compétences de statisticien qu'il ne juge pas posséder ?

Certains doctorants expriment clairement leur gratitude quant à l'aide en statistique apportée par des collègues, témoignant ainsi du fait qu'il s'agit d'éléments ou étapes dans leur travail qu'ils ne pensent pas avoir pu mener seul, ce qui leur permet aussi de ne pas assumer totalement cette partie de leur thèse.

« Les statistiques sont devenues moins pénibles à tes côtés. » (Al Kaddissi, 2012)(p6 – B1 – Ecotoxicologie)

« Pour son aide répétée dans les traductions, pour ses compétences en statistiques et pour ses connaissances outre-Atlantique, je tiens à remercier Claudia. » (Garcia, 2013)(p6 - B2-sociologie)

« A Mélanie, pour ton aide en statistiques et pleins d'autres choses » (Foubert-Samier, 2013) (p8 - B2 –Epidémiologie)

« Enfin, dans un troisième volet (réalisé par un économètre), nous avons mobilisé des méthodes économétriques d'estimation des effets propres pour estimer l'effet propre ou causal des PDM sur certaines variables objectives que nous avons pu mesurer. » (Sergent, 2013) (p165 - B4 - Science politique)

La collaboration est souvent une étape nécessaire dans un travail de recherche, et le doctorant apprend dans le cadre de sa thèse à reconnaître les compétences de chacun et à développer des collaborations. Cependant la rédaction de la thèse est un travail individuel et il s'agit, pour le doctorant, de mettre en évidence dans ce cadre sa capacité à développer un travail de recherche. Il n'est donc pas habituel que le doctorant précise dans le corpus de la thèse ses activités précises ou les partenariats dont il a bénéficié, les remerciements étant justement l'espace permettant de souligner ce point.

Pour les thèses rédigées sur articles, il est difficile d'évaluer l'intensité de collaboration entre le doctorant et les autres acteurs, et ainsi la part propre au doctorant pour le développement statistique. Un seul doctorant exprime précisément que la partie statistique a été traitée par des professionnels. Pour les autres doctorants nous n'avons pas pu déterminer leur part personnelle dans cette activité statistique mise en œuvre, cependant nous nous sommes intéressée spécifiquement ici à la formulation et à l'expression, et ne pouvons pas étudier directement la praxéologie mais simplement la façon d'exprimer ce qui a été fait, les résultats et les décisions prises à l'issue de l'analyse.

« J'ai personnellement participé à la conception de l'étude, l'encadrement du relevé des données, la collaboration avec l'équipe statistique, la rédaction de l'article dont j'étais le premier auteur et l'auteur correspondant. » (Boyer, 2012) (p53 - B2 – Pharmacologie)

c) *Place et rôle de la statistique dans les thèses*

Plusieurs doctorants soulignent l'importance de ne pas considérer les statistiques seules, mais de tenir compte d'un ensemble d'informations, tant pour leur approche spécifique que plus largement. Conscients de ces limites, ils traitent l'approche statistique en connaissance de causes.

« L'importance de la ville en tant que lieu de croissance et d'avenir pour le genre humain ne se mesure pas seulement au niveau de la production des statistiques (p25) [...] En Afrique la réalité urbaine a été interrogée sous divers aspects mais le poids de la statistique est difficile à évacuer, d'autant que l'organisation des Nations Unies considère d'abord la taille démographique.(p61) [...] D'un continent à l'autre, d'un pays à l'autre, la ville apparaît comme multiforme et plurielle, il ne fait donc aucun doute que le critère statistique, dans la multiplicité de ses chiffres, ne fait guère l'unanimité tant au niveau des spécialistes que des institutions. (p63) (Diouf, 2013) (B3 - Sciences de l'information et de la communication)

Dans certaines thèses, les bénéfices d'une approche hybride associant une approche qualitative (recueil d'information par le biais d'entretien ou d'observation) et une approche quantitative (recueil *via* un questionnaire, ou base de données déjà existante) sont sous-entendus mais peu explicités. La diversité des recueils assurerait une meilleure compréhension, et selon les auteurs de ces thèses, ce serait l'approche qualitative qui permettrait de donner du sens aux résultats quantitatifs.

« Sans écarter les données quantitatives, nous privilégions une analyse qualitative parce qu'elle nous permet d'interpréter au mieux les données sur la modernité et l'espace urbain, telles qu'elles sont livrées de manière brute dans la presse. (p31) [...] Pour explorer cette « vérité », un questionnaire recueille des données auprès de journalistes, habitués à traiter la matière urbaine. Ces données sont utilisées dans l'interprétation de certains éléments du corpus. » (Diouf, 2013) (p41 – B3 - Sciences de l'information et de la communication)

Les approches hybrides et croisées, permettant de recueillir des données de natures multiples et d'analyser les situations sous différents angles, sont souvent citées, mettant ainsi en évidence la place de l'interdisciplinarité dans ces travaux de recherche, et la nécessité du développement de compétences multiples pour les doctorants.

« Les approches comparatives employées, visuelle et statistique, sont complémentaires. [...] L'approche statistique (calcul de la variance relative et classification hiérarchique) fonctionne particulièrement bien avec le référentiel thermique du quartz. Elle discrimine les prélèvements suivant leur intensité TL et la position des pics de luminescence et les regroupe autour d'une ou plusieurs références thermiques dont les courbes TL ont l'allure la plus proche. » (Brodard, 2013) (p203 - B3 - Physique des archeomateriaux)

« Mais ce travail ne doit pas avoir pour seule vocation d'enrichir l'histoire locale, il est aussi et surtout une étude de cas. Pour les communautés précisément, donnant plusieurs exemples, Jean Delumeau rappelle l'intérêt des études de cas, qui contredisent les idées reçues et les statistiques d'ensemble. (p5) [...] A la différence d'autres études, nous ne disposons pas de nécrologes, ni de lettres circulaires permettant d'approcher de manière statistique les causes prétendues d'entrée en religion. Nous devons en revanche essayer de compter sur des sources qualitatives et externes. » (Tissot, 2012)(p320 – B3 - Histoire moderne et contemporaine)

d) "Statis" dans tous ses états

Le terme statistique est fréquemment employé en tant qu'adjectif afin de préciser des éléments développés pour répondre aux spécificités que peut susciter une démarche statistique :

- « couplée à des outils statistiques adaptés », « environnement statistique R » ou « logiciels statistiques bayésiens comme OxCal », exprimant ici le ou les logiciels développés spécifiquement pour le traitement des données et les analyses statistiques,

- « tests statistiques » indiquant une démarche permettant de rejeter ou de ne pas rejeter une hypothèse en fonction d'un échantillon de données. Il s'agit ainsi pour l'auteur de préciser qu'il travaille avec des outils caractéristiques d'activités statistiques.

« C'est pour cette raison qu'une théorie de la diffusion multiple a été développée par G. Molière [73]. C'est une théorie statistique qui permet de moyenniser l'effet de multiples collisions que rencontre un électron le long de sa trajectoire. » (Beaucourt-Jacquet, 2012) (p52 - B1 - Astrophysique, plasmas, nucléaire)

Parfois, cette association de l'adjectif statistique à un nom commun ne correspond pas précisément à une spécification du mot initial mais plus exactement à un mot composé construit pour représenter un nouvel objet dont l'existence est propre à un domaine et pour une « tribu ».

« Dans ce système sans interactions avec le solvant et sans orientations privilégiées des liaisons chimiques de la chaîne carbonée, la chaîne adopte une conformation favorable entropiquement, appelée pelote statistique. » (Cuenca, 2012) (p59 – B1 - Physique-chimie)

Ces terminologies sont souvent très spécifiques et déterminent des objets dans des environnements très contraints. Leur présentation en début de thèse correspond à la présentation des éléments et objets de contexte, qui peuvent dès lors préciser une appartenance à une « tribu » ainsi spécifiée « ceux qui savent ».

La mise en œuvre d'une démarche statistique par le doctorant se construit autour de différents éléments plus ou moins présents et détaillés dans la thèse. Il s'agit tout d'abord de repérer les données qui seront étudiées, et qui sont soit déjà existantes dans des bases de données ou qui ont dû être recueillies.

Puis il s'agit de présenter les « outils statistiques » auxquels le doctorant fera appel pour la description et l'analyse de ses données. Ces informations étaient signalées lorsque la statistique n'était pas une fonction classiquement connue.

« La courbe est construite en suivant la statistique bi variée développée par Le Goff [...] qui est une extension de la statistique de Fisher [...] statistique hiérarchique bayésienne ([...] statistique de Mc Fadden et Reid. » (Hervé, 2012) (p32 – B3 – Physique d'archéomatériaux)

Ces statistiques sont citées dans le corps du texte, mais la définition et la formule mathématique associée sont rarement présentes. Ainsi l'auteur suggère que le lecteur comprend précisément de quoi il parle. Le nom porte-il à lui seul l'information suffisante pour être explicite ? Cela suggère donc que le discours est en direction uniquement des sachants. Il faut souligner à cette étape que peu de thèses comportent des écritures mathématiques permettant de préciser les statistiques utilisées ou les modèles.

Le dernier élément d'une démarche statistique est la phase de synthèse, et de prise de décision. La précision et les éléments d'explication de chacune de ces étapes sont caractéristiques d'une niche, éléments que nous allons décrire ci-après.

e) Quel niveau de précision dans la présentation du recueil de données ?

Les caractéristiques des données étudiées diffèrent de manière importante entre les thèses et les disciplines. Il y a ainsi une grande variabilité des outils utilisés. Etudier des données caractérisant des phénomènes chimiques ou biologiques, ou celles caractérisant des phénomènes historiques ou des comportements humains nécessite des méthodes spécifiques. L'explication et la description des outils et modalités de recueil sont des éléments permettant de comprendre la stratégie, et d'évaluer sa capacité à mesurer correctement ce qui était prévu. Cependant, les habitudes de diffusion sur cette étape de la démarche statistique semblent bien différentes d'une discipline à l'autre.

Dans le cadre de disciplines telles que l'écotoxicologie, la biologie, l'astrophysique, le laser, les protocoles d'expérimentation et de recueil de données sont très précis. Ainsi les moyens mis en œuvre sont explicités, les caractéristiques des outils techniques utilisés (marque, caractéristiques techniques ...) sont spécifiés et toutes les phases précisées. Ces éléments d'expérimentations, de protocolisation sont considérés au centre de la pratique professionnelle et font partie des paramètres de validation de la méthode par les pairs, pouvant représenter près d'un quart à un tiers de la matière scientifique de la thèse.

« (extrait résumé) Durant cette étude, les gènes *mt* et *atp6* codant respectivement pour la metallothionéine (une protéine de stress) et l'ATP synthase ont été clonés, séquencés puis enregistrés dans GenBank sous les

numéros d'accessions respectifs GU220368.1 et GU220369.1. De même, nous avons étudié les impacts au niveau histologique. Nous avons tenté de lier ces effets entre eux et à la bioaccumulation du Cd dans les branchies et l'hépatopancréas, organes cibles de l'exposition. Dans un second temps, nous avons étudié les mêmes paramètres biologiques après exposition à différentes durées (4, 10, 30 et 60 jours) et concentrations (0,1- 40 μM) de l'uranium (U) et appliqué la même démarche. De plus, nous avons essayé de différencier la chimio- et la radiotoxicité de ce radionucléide en exposant des lots d'écrevisses soit à l'U appauvri soit au ^{233}U (présentant une activité spécifique plus élevée) au travers des mêmes critères d'effets. Enfin nous avons comparé les effets du Cd et de l'U après exposition aiguë (40 μM -10 jrs) et chronique (0,1 μM -60 jrs). » (Al Kaddissi, 2012) (B1 – Ecotoxicologie)

Les terminologies employées sont précises, les méthodes de codage et de repérage sont spécifiées, la répétition de mesures dans les mêmes conditions est un point souligné comme important pour la validité et la fiabilité de la donnée. Les éléments de précision dans ce type de recueil ne peuvent être appréciés que par des experts du sujet.

« L'expérience du guidage d'un faisceau d'électrons avec deux impulsions laser a été réalisée sur l'installation laser pet watt Vulcan au Rutherford Appleton Laboratory (RAL) à Didcot en Angleterre. Elle est basée sur la proposition d'un groupe international dans le cadre du projet FCI HiPER. Cette expérience nous a permis d'obtenir les conditions de guidage en fonction du rapport des intensités et du délai entre les deux impulsions. » (Beaucourt-Jacquet, 2012) (p6 - B1 - Astrophysique, plasmas, nucléaire)

De nombreuses manipulations précises, permettant le recueil d'informations sont nécessaires et doivent être décrites : il est d'ailleurs fréquent qu'un chapitre complet de la thèse soit réservé à cette description. L'expérimentation est ainsi très encadrée et ne correspond pas à une simple observation, mais à une préparation d'un environnement permettant de recueillir les données. Les thèses propres aux sciences dites expérimentales sont plus sensibles à ce cadre.

Les recueils concernant des comportements humains sont principalement de deux types : l'observation ou la déclaration. Les éléments de précision concernant l'observation sont très dépendants de la discipline et de l'objectif même sous-jacent au recueil. Ainsi une observation de comportement dans un objectif de modification technique d'un outil peut être réalisée avec un recueil de données précises et

techniques, comme dans la thèse de Cohé qui recueillera un nombre très important de données.

« (extrait résumé) Les premières études ont été réalisées afin de comprendre comment l'utilisateur réagit pour manipuler un objet virtuel 3D avec une surface tactile sans lui imposer de techniques d'interaction particulières. » (Cohé, 2012) (B1 – Informatique)

Dans le cas des études en paléontologie et préhistoire, le recueil de données d'un comportement ne se fait pas, bien évidemment, pendant le comportement, mais se nourrit de l'observation de l'environnement dans lequel il a eu lieu (il y a des milliers d'années). Des instruments de mesures spécifiques ont été dès lors développés et précisément décrits comme cela est indiqué dans les deux verbatims ci-dessous.

« L'étude de certains foyers de la grotte des Fraux a permis de tester le potentiel de paléothermomètres fondés sur ces deux propriétés indépendantes à savoir la TL des grains de quartz et le magnétisme des oxydes de fer contenus dans les sédiments. Le paléothermomètre TL a été élaboré en comparant les signaux TL d'échantillons provenant de foyers archéologiques à ceux de références thermiques chauffées en laboratoire. » (extrait résumé) (Brodard, 2013) (B3 - Physique des archeomateriaux)

« La question principale à laquelle nous avons tenté de répondre est la suivante: est-ce que la variabilité des matières premières (caractéristiques pétrographiques et morphologie des supports) a conditionné la production des hommes préhistoriques dans le temps et dans l'espace ? La mise en place de fiches descriptives technologiques spécifiques pour chaque catégorie d'objets nous a permis d'analyser les caractères techniques propres à chaque produit et de sélectionner les variables principales qui ont joué un rôle important dans l'évaluation des contraintes et ont guidé la production lithique. » (extrait résumé) (Santagata, 2012) (B1 – Préhistoire)

Dans le cas des disciplines de sciences humaines et sociales, les recueils étant plus qualitatifs, le niveau de précision de la mesure n'a pas les mêmes enjeux que dans les situations expérimentales. Ainsi, le recueil de données peut se faire par questionnaires ou entretiens mais il est intéressant de souligner la préférence dans la plupart des cas de la mise en place de recueil multi-sources.

« L'étude s'intéresse à la formation initiale et continue des futurs enseignants telle qu'elle est prescrite par la tutelle et telle qu'elle est mise en œuvre par l'institut de formation. Appuyée sur les données recueillies à partir des différents outils utilisés à cet effet, à savoir quatre questionnaires, deux interviews et l'analyse de la maquette du cursus de la formation initiale, la recherche met en lumière le décalage qui existe entre le prescriptif et le

descriptif de la formation. » (extrait du résumé) (Benabed, 2012) (B2 - Etudes anglaises : langue de spécialité, didactique de la langue)

Dans certaines situations, la question n'est pas de recueillir les données, mais plutôt de pouvoir sélectionner les données intéressantes, cette situation est particulièrement existante maintenant et cette direction pour la statistique est très porteuse : ceci peut être intégrée au développement des *big data*.

« Étant donné le grand volume de données disponibles autour de l'exploitation des aéronefs, ce travail vise à analyser les historiques des événements associés aux aéronefs afin d'en extraire des prévisions de maintenance [...] Ainsi, plus que dans les autres domaines tels que le marketing ou la médecine, les données se rapportant à la vie d'un système en aéronautique sont particulièrement hétérogènes et très volumineuses. Il est pratiquement impossible de par leur volume et l'importance de leur flux, à un expert ou à une équipe d'experts de pouvoir les analyser et les exploiter dans un délai raisonnable (extrait résumé) [...] Prenant en compte le grand volume d'informations disponibles concernant l'utilisation et l'exploitation des véhicules, il y a lieu de les exploiter par l'application d'outil statistique et d'analyse de tendance afin d'étudier le comportement des équipements, d'en améliorer la gestion et d'en optimiser la maintenance, le mode d'utilisation et même le design. » (Ben Zakour, 2012) (p36 – B1-Informatique)

La qualité des résultats peut aussi, dans un second temps nécessiter une validation par les experts, afin de ne pas laisser le résultat dépendre uniquement de l'association de données recueillies dans des situations plus ou moins vierges de tout « bruit » et d'une formule mathématique.

« Tous les résultats statistiques obtenus avec des enregistrements bruités ou non ont été validés manuellement avec la participation des biologistes. » (Quotb, 2012) (p83 - B1 – Electronique)

La recherche en sciences expérimentales au niveau de la thèse nécessite un haut niveau de précision qui doit être dit et explicité : le doctorant se doit de spécifier chaque étape permettant à l'expert de comprendre la gestion de la variabilité de mesures. La problématique est différente en sciences humaines, car, dans cet ensemble de disciplines il y a finalement peu d'usage d'instruments techniques, et le contrôle de la variabilité de mesures passe avant tout par une standardisation des modes de recueil et par la formation des enquêteurs.

f) *Vous avez dit échantillonnage ...*

Recueillir des données nécessite de définir les éléments sur lesquels les mesures sont effectuées. Les éléments constituent la population d'intérêt et selon les études, le chercheur peut recueillir les informations sur tout ou partie de la population. Que la population d'intérêt corresponde à une population d'individus, de médicaments ou d'objets, ou qu'elle corresponde à un élément naturel (eau, air, sable ...) ou à un matériau (aluminium, fer ...), le repérage d'éléments la composant pour effectuer des mesures est nécessaire mais diffère de manière radicale.

« Les benzodiazépines telles que le diazépam sont aussi régulièrement retrouvées dans les échantillons d'eau de stations d'épuration mais toujours à des concentrations limitées de l'ordre de quelques nanogrammes par litre. » (Belles, 2012) (p40 - B1 - Chimie analytique)

« Le procédé, une fois mis au point, a fait l'objet d'une évaluation de ses capacités et de ses limites sur différents types d'échantillons (plaques en aluminium et plaques composites en carboneépoxy). » (Leleux, 2012) (p12 - B1 - Mécanique et ingénierie)

« Cette biotechnologie récente permet d'analyser le niveau d'expression des gènes transcrits dans une cellule, un tissu, un organe, ou encore un organisme, à un moment donné et dans un état donné par rapport à un échantillon de référence. » (Al Kaddissi, 2012) (p77 - B1 – Ecotoxicologie)

Dans les différentes approches ci-dessus, le terme *échantillon* est défini comme une partie d'un matériau, ou d'un ensemble dense, voire infini. Il est dès lors considéré comme représentant en tant que représentation miniature de l'ensemble d'intérêt.

Dans le cadre d'étude en population humaine, l'échantillon correspondra à l'ensemble des individus sélectionnés par des processus de sélection diverses aléatoires ou non.

Les méthodes d'échantillonnage, qu'il s'agisse d'échantillonnage au sens sélection d'un morceau de matériau ou d'éléments ou d'échantillonnage parmi une population d'individus, sont des éléments fondamentaux dans la démarche statistique. Elles ont souvent des spécificités selon les disciplines et les caractéristiques de la population d'intérêt, ainsi que selon l'objectif de l'étude. Plusieurs *verbatim* soulignent l'importance de cette étape pour la qualité des données et donc de l'analyse.

« Parmi ces étapes, la préparation d'échantillon est nécessaire et cruciale mais souvent critique et une méthodologie spécifique à la préparation d'échantillon peut s'avérer d'une grande aide pour maîtriser l'analyse. (p14) [...] Pour avoir une bonne information statistique, ces tests sont réalisés sur un nombre important d'échantillons pouvant aller jusqu'à 80 pièces pour un seul test. Ces essais de fiabilité sont donc longs et coûteux. » (p102) (Aubert, 2012) (B1 – électronique)

« Pour avoir une idée de la progression rapide de l'arachide, il faut se reporter aux statistiques de la production, avec des années choisies par nous-même au hasard. » (Tandjigora, 2012) (p237 - B4 - Sciences économiques)

Un élément partagé dans l'ensemble des thèses ayant fondé une partie de leur travail sur un échantillon est le souci de la justification de la qualité de cet échantillonnage pour une assurance de la réussite de l'analyse.

« L'une des étapes les plus critiques dans l'assurance de la réussite de l'analyse est la préparation d'échantillon. » (Aubert, 2012) (p10 - B1 – Electronique)

« Même si l'échantillon de patients ayant une SEP-PP était plus faible que celui des patients avec une SEPRR, la force de notre étude a été d'utiliser un échantillon de sujets contrôlés strictement appariés et de taille importante pour réaliser la comparaison du fonctionnement cognitif entre les patients atteints de SEP-PP et ceux ayant une SEP-RR avec la même méthodologie. » (Ruet, 2012) (p127 - B2 - Sciences, technologie, sante. Neurosciences)

« Un problème demeure néanmoins avec ces données. Il s'agit d'un fichier au 1/20ème. En d'autres termes, il s'agit d'un échantillon et non de données exhaustives contrairement aux fichiers iris. Ceci pose un problème dans le sens où tout travail sur échantillon implique de subir l'impact de la variation aléatoire. Il faudra donc faire attention à ne pas prendre pour une évolution d'intensité un simple effet de variation aléatoire. » (Dasré, 2012) (p369) B4 – Demographie)

La sélection peut être choisie en fonction du coût direct, mais aussi du coût qui est associé au recueil. Elle peut être plus ou moins facile à obtenir, puis induire plus ou moins de difficultés et de coût pour le recueil de l'information. L'étape importante lorsque l'étude est réalisée sur une population dénombrable est la détermination de l'effectif de l'échantillon. Les critères de calcul de la taille de l'échantillon ne sont pratiquement jamais précisés, et la détermination de l'effectif n'est que rarement argumentée. Il semble que l'effectif dépende principalement du caractère de faisabilité.

« Seize sujets se sont portés volontaires, quatre hommes et douze femmes âgés de 19 à 60 ans (moyenne = 31.56, écart-type = 10.31). » (Cohé, 2012) (p66 - B1 – Informatique)

« Les participants à notre recherche étaient des enseignants stagiaires (n=200), des enseignants en service (n=100), des enseignants formateurs (n=20), des inspecteurs (n=12) et décideurs de la formation (n=2), qui avaient été sélectionnés de façon volontaire (les caractéristiques de cet échantillonnage seront précisées ci-après). » (Benabed, 2012) (p176 - B2 - Etudes anglaises : langue de spécialité, didactique de la langue)

« Les différentes caractéristiques morphologiques des cinq matériaux ont été étudiées : le nombre de couches dans l'épaisseur, les dimensions des torons (étude statistique effectuée sur une population d'un millier d'échantillons), » (Dupin, 2013) (p16 - B1 - Mécanique et ingénierie)

« Pour rappel : nous simulons 10000 industries contenant chacune 5 plateformes, l'échantillon global est donc de 50000 plateformes. » (Querbes-Revier, 2012) (p159 – B4 - Sciences économiques)

Les faibles effectifs sont régulièrement explicités avec des justifications parfois peu convaincantes ; les auteurs discutent souvent le fait qu'une augmentation d'effectif serait nécessaire pour confirmer les observations.

« Enfin, pour chaque condition, trois batteries identiques subissent le même test ce qui constitue la taille minimale pour un échantillon représentatif jusqu'à ce que des autopsies soient pratiquées sur certaines batteries au cours du vieillissement. » (Eddahech, 2013) (p64 - B1 – Electronique)

« Afin de pouvoir conclure, il serait nécessaire de discuter de ces éléments à partir d'un échantillon plus conséquent de faunes de ces niveaux. » (Campmas, 2012) (p326 - B1 - Préhistoire et géologie du Quaternaire)

Un effectif insuffisant peut aussi être un élément d'argumentation de non usage d'éléments ou d'outils d'analyse statistiques.

« Dans leur article, l'utilité des sommets, des bases et de la surface en général de ces outils est mise en évidence par rapport aux autres méthodes. Bien que non représentatives d'un point de vue statistique (total de moins de 100 pièces) ces études semblent indiquer qu'il s'agit d'un instrument polyvalent conçu pour des taches en dehors de l'habitat (p55) [...] Les produits d'aménagement de chaîne Levallois ont parfois un résidu cortical. Leur nombre limité ne permet pas de les analyser statistiquement : leurs caractéristiques seront analysées avec celles des produits sans cortex (p126) [...] Le petit nombre de produits ne permet pas d'analyser statistiquement toutes leurs caractéristiques morphométriques (p131) [...] Le petit nombre d'éclats de préparation ne permet pas une analyse statistique de leurs caractéristiques. » (Santagata, 2012) (p182) (B1 – Préhistoire)

La taille de l'échantillon est présentée comme pouvant être un facteur limitant, mais le problème est rapidement évacué, la thèse correspondant dans la plupart des cas, principalement en biologie et chimie, à une étape de la recherche

parmi d'autres et non à une fin ; ainsi la proposition d'agrandir l'échantillon pour confirmer les résultats est implicite pour certains. Ne serait-ce pas un moyen de dire « je dois encore continuer ce travail en étant reconnu dans votre communauté » ?

Le terme de *représentativité* revient fréquemment pour justifier de la qualité d'un échantillon, mais sans aucune précision de ce qui est entendu par ce terme. Dans une situation l'auteur déclare que l'échantillon n'a pas été réalisé dans cet objectif.

« Nous avons préféré choisir quelques établissements et inclure tous les individus volontaires de ces résidences plutôt que de choisir plus d'établissements avec un échantillon plus petit dans chacun d'eux. Cet échantillon n'ayant pas vocation à être représentatif de la population, il est plus intéressant de recueillir les discours de la plupart des individus fréquentant un même établissement. (133) (...) Nous avons ainsi un petit échantillon non significatif qui ne garantit pas d'approcher une quelconque représentativité de la population des enfants placés. » (Denecheau, 2013)(p456 – UB2- Sciences de l'éducation)

La qualité de l'échantillonnage peut aussi être exprimée par les auteurs par la capacité à communiquer les données nécessaires pour répondre à la question de recherche : il s'agit du critère de qualité d'estimation.

« Pour avoir une statistique satisfaisante sur les profils de vitesse, l'acquisition de 25 à 50 séries d'images issues d'un Scan 1 est requis. » (Cuenca, 2012)(p103 – B1 – Physique-chimie)

g) Les différents outils d'analyse statistiques

Selon le rattachement du doctorant à son habitat, il appartient à une niche spécifique, et donc il est confronté à des pratiques spécifiques pour lesquelles il va se référer à un ensemble d'outils, classiques ou spécialisés, partagés par tous ou propre à sa discipline.

Les outils statistiques peuvent se décliner selon la classification suivante :

- des outils permettant de synthétiser l'ensemble de données recueillies, analyse descriptive ;
- les tests statistiques,
- les outils de modélisation,
- et les logiciels.

Il s'agit ci-dessous de pouvoir étudier les usages de notions importantes et classiques en statistique selon les différentes thèses. Les utilisations des notions moyenne, variabilité, tests statistiques et régression ont été principalement observées dans les thèses.

(1) Moyenne

La moyenne est une statistique très usitée et très rarement définie. Il s'agit du seul mot présent dans toutes les thèses, soit en tant qu'adjectif soit en tant que nom. Nous allons étudier son usage en termes de mot, d'adjectif, mais aussi les différents formats des présentations chiffrées.

Ainsi, le terme « moyenne » a été parfois indiqué sans adjectif attribut.

« A partir de l'examen postmortem de 872 cerveaux, Brayne et al. (Brayne, Ince et al. 2010) ont montré par ailleurs que les sujets de haut niveaux d'étude avaient en moyenne un poids de cerveau plus important que les sujets de bas niveau. » (Foubert-Samier, 2013) (p41 - B2 – épidémiologie)

« Moyennes et écarts types sont calculés avec n=10. » (Monier, 2012) (p58 - B1 - Biologie végétale)

« Les valeurs de résistivité présentées par la suite sont une moyenne sur trois échantillons provenant d'une même fibre. [...] Les valeurs des propriétés mécaniques présentées par la suite sont une moyenne d'essais sur deux échantillons issus de la même fibre. » (Michardière, 2013) (p75 - B1 - Chimie-Physique)

« Pour chaque échantillon, la valeur prise en compte est la moyenne de dix mesures. » (Ouvrard, 2013) (p88 - B1 - Physico-Chimie de la Matière Condensée)

« Les courbes représentent la moyenne et l'écart-type obtenus avec 4 surnageants de production différents (107) Moyenne et écart-type de 3 expériences indépendantes. » (Morello, 2012) (p121 - B2 - Biologie cellulaire et physiopathologie)

Dans ces *verbatim*, le mot « moyenne » représente le ou les résultats obtenus par calcul à partir d'une fonction mathématique dont la formule ou les formules ne sont pas précisées. Seule une thèse spécifie les formules de moyenne utilisées (Valderrama Ulloa, 2013) ; pour les autres auteurs la formule est implicite. Nous pouvons supposer que comme dans le langage courant moyenne signifie moyenne arithmétique et non moyenne géométrique ou harmonique.

Le terme moyen/moyenne est aussi utilisé en tant qu'adjectif, et nous pouvons nous demander quelle différence faire alors entre la formulation « le poids moyen », ou « la moyenne des poids » ? Un extrait semble donner une réponse tout à fait pertinente, faisant la distinction entre la formule et le résultat.

« Pour chacune des données analysées, il est nécessaire de préciser que : - l'humidité relative est la moyenne mensuelle des moyennes quotidiennes (jour et nuit); - la température moyenne mensuelle (°C) est la moyenne entre la température minimale et de la température maximale. » (Valderrama Ulloa, 2013) (p64 - B1 - Mécanique et ingénierie)

Cette formulation propose dès lors de différencier le résultat « température moyenne », de la méthode employée pour l'atteindre « moyenne des températures ». La moyenne définit donc bien une statistique au sens de formule mathématique et la température moyenne correspond au résultat de cette fonction pour les données de l'échantillon. Cette spécification de langage n'est pas partagée par les thèses.

Le but du calcul de la moyenne peut être très différent selon les situations. Il peut s'agir soit d'un indicateur permettant de résumer un ensemble d'informations, pour décrire voire comparer, soit d'une information devenant dès lors un élément de référence, on peut obtenir ainsi des estimations qui deviendront des références nationales, soit enfin des mesures permettant de comparer différents sous-groupes.

« Egalement, la différence des courants de réduction obtenue entre ces deux fibres met en valeur des problèmes de reproductibilité pour un même type de fibre. C'est pourquoi la comparaison des courants obtenus doit se faire avec les moyennes de ces derniers. » (Michardière, 2013) (p98 - B1 - Chimie-Physique)

« Ils constatent que les enfants placés ont des résultats plus faibles que la moyenne nationale en lecture et en mathématiques. » (Denecheau, 2013) (p15 - B2 - Sciences de l'éducation)

La présentation des résultats de moyenne peut être ou non associée à une information complémentaire permettant de résumer la distribution avec un peu plus d'éléments de précision ; il peut s'agir de caractéristiques de tendance centrale ou de dispersion selon les cas. L'exemple ci-dessous présente le « SD » pour « standard déviation » ou « écart-type » (classiquement dans les documents en français on note ET).

« Ils conduisent environ 7212.44km/an (S.D. = 8785.00 km, min 30 km – max 100000), ont eu en moyenne 0.60 amendes (S.D. = 1.55, min 0 – max 10), et depuis qu'ils ont obtenu le permis de conduire ont subi 0.75 accidents (S.D. =

0.99, min 0 – max 5) dont ils étaient le plus souvent coupables (0.42, S.D.= 0.70 ; min 0 – max 4). » (Description de l'échantillon) (Gheorghiu, 2012)(p65 - B2 – psychologie)

On peut retrouver le même niveau d'information par des représentations graphiques du type ci-dessous.

Ces informations peuvent être représentées graphiquement en présentant la caractéristique et la barre d'erreur représentant souvent l'écart type.

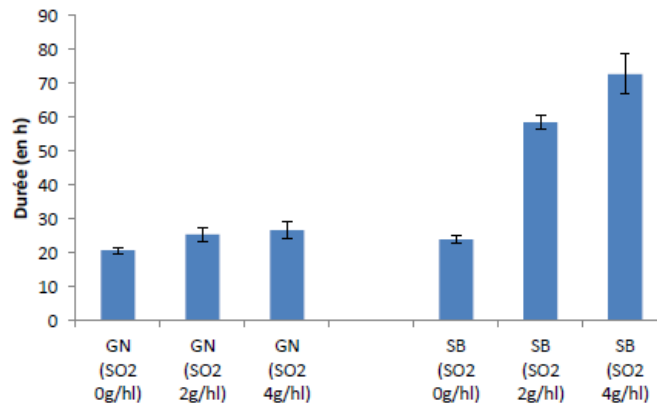


Figure 9. Phénotype des parents pour le paramètre étudié

Variation de la durée de la phase de latence en fonction de la concentration en SO₂ ajouté dans le milieu lors de la fermentation alcoolique pour les deux souches parentales utilisées dans cette étude. Les histogrammes et les écart-types sont calculés à partir de trois répétitions indépendantes.

Figure 1 : représentation barres d'erreurs et histogramme – p60 (Zimmer, 2013) (B2 – Enologie)

Cependant, des présentations de précisions complémentaires de la moyenne, par exemple, peuvent être moins explicites avec des écritures de type $\pm \Delta$, avec ou sans précision de ce que représente la valeur Δ .

« A titre d'exemple, les variations moyennes entre les modèles DMPC/CH en absence et en présence d'éthanol sont de 5 ± 1 kHz pour la région plateau. » (Furlan, 2013) (p133 - B1 - Chimie-Physique)

« Les données quantitatives continues seront décrites par des moyennes \pm ET si ils ont une distribution normale. » (Gheorghiu, 2012) (p224 - B2 – psychologie)

« Toutes les valeurs sont représentées comme la moyenne \pm l'erreur standard à la moyenne (SEM). » (Bonnan, 2012) (p111 - B2 - Sciences, technologie, sante. Neurosciences)

La notation « moyenne $\pm \Delta$ » représente des informations pouvant être très différentes mais non toujours explicites ; il s'agit de normes rédactionnelles propres à des disciplines, telles que la physique. Le sens commun de ce type d'écriture est de pouvoir informer d'un élément de variabilité qui peut correspondre, selon les cas à la mesure de variabilité entre les données initiales, appelée écart type ou *standard déviation* en anglais, ou à la mesure de précision quant à la mesure de l'indicateur, dans le cas présent appelé erreur standard à la moyenne ou *standard error* en anglais, et étant ainsi à comparer avec les présentations d'intervalles de confiance. La non précision des normes d'écriture sous-entend un niveau de communication implicite de par des usages partagés dans une communauté.

La perception de l'intérêt de l'écart type pour mesurer la dispersion d'une série de données est partagée, semble-t-il, entre les communautés, mais le format de la transmission de l'information n'est pas normalisé entre les différentes disciplines, et les lecteurs doivent se référer aux normes d'écriture de chaque discipline.

La mesure de la dispersion correspond globalement à la question de la variabilité.

(2) Mesure de la variabilité

La présentation de la variabilité diffère, comme nous venons de le voir ci-dessus, selon les cadres utilisés. Les indicateurs statistiques proposés dans les thèses pour traduire la variabilité des données autour d'une moyenne sont le plus fréquemment l'écart type, ou l'erreur standard.

L'écart-type est parfois dénommé différemment dans certaines thèses. Ce changement de nom semble associer à un usage spécifique : la formule mathématique correspond à l'écart type classique, mais la dénomination particulière traduit un effet de spécificité de l'usage et de la reconnaissance de ce que la formule représente dans le cadre de certaines disciplines. Les exemples de la rugosité et de la gigue en sont des cas particuliers bien spécifiques aux disciplines.

❖ La rugosité RMS « root-mean-square ».

La rugosité RMS σ_{rms} est définie comme suit :

$$\sigma_{rms} = \sqrt{\frac{1}{N} \sum_{i=1}^N [z(x_i) - \bar{z}]^2} \quad (II.4)$$

Où N est le nombre total de pixels, $z(x_i)$ la hauteur au point x_i , et la hauteur moyenne du profil définie par :

$$\bar{z} = \frac{1}{N} \sum_{i=1}^N z(x_i) \quad (II.5)$$

(Darwiche, 2012) (p57 - B1 - Laser, Matière et Nanosciences)

« De façon générale, la gigue³², ou jitter en anglais, est la variation indésirée d'une ou de plusieurs composantes d'un signal. Dans le domaine des lasers impulsions, la gigue temporelle est définie comme étant la fluctuation du taux de répétition de la cavité. Elle correspond donc à la fluctuation de l'intervalle de temps qui sépare deux impulsions d'un train qui sont consécutives ou pas. [...] Dans le cadre de ce travail, il a été choisi de définir cette valeur comme étant l'écart-type des fluctuations du retard pompe-sonde par rapport au retard attendu. » (Abbas, 2013) (p28 - B1 – Mécanique)

Les autres termes utilisés (« coefficient de variation », « variance relative ») sont quant à eux définis dans le texte, ce qui sous-entend que même dans leur communauté ces indicateurs et /ou terminologies ne sont peut-être pas fréquemment utilisés.

« Premièrement, nous étudions pour chaque geste l'erreur relative autour de la moyenne, connue sous le nom de coefficient de variation (CV). » (Cohé, 2012) (p96 - B1 – Informatique)

« La variance (V) correspond à la moyenne des carrés des écarts à la moyenne et la variance relative (VR) est le rapport de la variance au carré de la moyenne. » (Brodard, 2013) (p162 - B3 - Physique des archeomatériaux)

La prise en compte de la variabilité des données nécessite la compréhension des dimensions pouvant expliquer ces phénomènes de variabilité. Car comme nous l'avons spécifié dans le chapitre précédant plusieurs éléments peuvent intervenir et il s'agit donc de décomposer la variabilité afin de l'expliquer petit à petit, et dans certains cas de pouvoir y remédier. Qu'il s'agisse de la variabilité inter individuelle, de la variabilité de mesure, de la variabilité résiduelle, chercher à la mesurer et à

³² Mot utilisé en informatique, en mécanique représentant la variation aléatoire de caractéristique

l'expliquer si possible et voire à la contrôler est de la responsabilité du statisticien et son enjeu.

Repérage

« L'une des difficultés inhérentes aux études longitudinales réalisées chez des patients ayant une SEP réside dans la variabilité inter-patient au niveau de leurs performances cognitives. » (Dasré, 2012) (p85 - B4 – Démographie)

« La figure 79 représente le ratio entre l'écart-type à la moyenne et la moyenne afin d'évaluer les variations inter-individuelles de la bioaccumulation de la branchie et de l'hépatopancréas. » (Kaddissi, 2012) (p285 - B1 – Ecotoxicologie)

« La complexité s'est exprimée tant dans la variété des matières premières que dans la variabilité morphologique des outils façonnés ainsi que dans la variabilité des choix techniques mis en place pour façonner des outils sur galets qui paraissent très simples. » (Santagata, 2012) (p332 - B1 – Préhistoire)

Contrôle

« Lors de la résolution de l'algorithme RLS, on vise à éliminer l'erreur relative d'estimation du SOC et le bruit de mesure de la tension, assimilé à un bruit blanc de valeur moyenne nulle. » (Eddahech, 2013) (p56- B1 – Electronique).

« Pour les besoins de l'étude, il était nécessaire de sélectionner une cellule performante et robuste d'un point de vue élaboration, pour minimiser les risques de variabilité entre cellules susceptibles de masquer les effets de dégradation étudiés. //76 - On s'affranchit en particulier de la variabilité entre cellules et montage. » (Mansuy, 2012) (p37 - B1 - Physico-Chimie de la Matière Condensée)

« La barre d'erreur associée à la vitesse de chaque particule est déterminée à partir de l'écart-type du déplacement moyen de la particule entre les différentes images. Si l'écart type est trop important, la particule est écartée, i.e. le déplacement de cette particule n'est pas considéré, car il résulte d'une erreur de tracking. » (Cuenca, 2012) (p103 - B1 - Physique-chimie)

« Les individus sont facilement capturés en milieu naturel. Au stade adulte, ils fournissent assez de matériel biologique pour effectuer plusieurs analyses biochimiques sur un même spécimen. L'étude multi-critères d'effet gagne ainsi en cohérence en limitant la variabilité inter-individuelle. » (Al Kaddissi, 2012) (p97 – B1 – Ecotoxicologie)

Prendre en compte la variabilité pour améliorer l'analyse

« Les événements seront appliqués les uns après les autres, en fixant les autres phénomènes afin de ne pas accroître artificiellement la variance.(p60) [...] Comme tout indice statistique, l'indice de ségrégation est sensible à la variation aléatoire. (p82) [...] La question de la comparabilité des données d'un recensement à l'autre est un problème central des études statistiques portant sur les populations. Etudier la variation d'un phénomène dans le temps nécessite en effet une « parfaite » comparaison des données entre les différentes dates de collecte. » (p147) (Dasré, 2012) (B4 – Démographie)

La décomposition de la variance est donc un élément très important dans la démarche statistique, ayant pour raison de chercher à expliquer les raisons de la variabilité d'une variable. Cet élément permet ainsi de mettre en évidence les variables explicatives.

« Il est important de considérer que le handicap physique et les caractéristiques démographiques expliquent seulement 14% de la variance dans une étude incluant un large échantillon de 312 patients atteints de SEP. » (Ruet, 2012) (p147 - B2 - Sciences, technologie, sante. Neurosciences)

« Dans une étude de Wehlage et Rutter, les facteurs scolaires (aspirations, rendement scolaire et statut socioéconomique) correspondent à 89 % de la variance du modèle de décrochage étudié (1989). » (Denecheau, 2013) (p74 - B2 - Sciences de l'éducation)

(3) Mais qui est Gauss ?

Il est incroyable qu'un nom propre soit cité aussi souvent dans les thèses (cité dans 18 thèses). Le nom de Gauss³³ est lié à la démarche statistique, bien évidemment de par la place prise par loi de Gauss, dite aussi loi normale et représentée par la courbe en forme de cloche. Cette représentation est connue de tous les étudiants ayant réalisé au moins une vingtaine d'heures de statistique, car il s'agit d'une condition de validité pour l'application de nombreuses modélisations.

Cependant, il est remarquable de voir la diversité d'outils signalés dans l'une ou l'autre des thèses caractérisés par le fait d'être gaussien. Cette dénomination est cependant très rarement complétée de l'information définissant ce que cela signifiait précisément, et le lecteur non initié en tant que membre de la communauté peut être en difficulté pour rapprocher les notions de fenêtre gaussienne, de bruit blanc et la courbe en forme de cloche.

« On vient de démontrer que $AR(x; y)$ et $AI(x; y)$ suivent des statistiques gaussiennes. » (Le Cain, 2012) (p14 - B1 - Mathématiques appliquées)

« La figure 1.9 représente un exemple de grille de quantification optimale pour une variable aléatoire gaussienne centrée réduite en dimension 2. » (Brandejsky, 2012) (p36 - B1 - Mathématiques appliquées)

« Dans son équation Donoho utilise le fait qu'un bruit blanc gaussien peut être entièrement décrit par son écart type. » (Quotb, 2012) (p24 - B1 – Electronique)

³³ **Johann Carl Friedrich Gauß** né le 30 avril 1777 à Brunswick et mort le 23 février 1855 à Göttingen, est un mathématicien, astronome et physicien allemand. Il est surnommé « le prince des mathématiciens », il est considéré comme l'un des plus grands mathématiciens de tous les temps.

« Si l'astigmatisme d'un faisceau est trop important, un système optique ne pourra pas focaliser un faisceau gaussien selon les directions. » (Abbas, 2013) (p 15- B1 – Mécanique)

« L'analyse la plus simple a consisté à représenter l'ensemble des ISI au cours du temps ainsi que leur distribution pour chaque groupe neuronal (Figure 24 A-F). Cette première analyse nous a permis de mettre en évidence que la distribution des ISI des neurones PV semble gaussienne, » (Abdi, 2013) (p88 - B2 – Neurosciences)

« Ainsi pour l'émission, on définira tout d'abord la forme du signal temporel d'excitation, à savoir généralement un burst, avec une fréquence centrale f_0 , et fenêtré par une gaussienne, (---) où A est l'amplitude du signal et G est la fenêtre gaussienne. » (Leleux, 2012) (p78 - B1 - Mécanique et ingénierie)

« La tendance des enseignants à ajuster leurs exigences et leurs évaluations de façon à obtenir à peu près une distribution des résultats qui suit une courbe de Gauss autour d'une même moyenne (identifiée par Posthumus en 1947). » (Denecheau, 2013) (p158 – B2 - Sciences de l'éducation)

L'adjectif « gaussien » est un attribut très usité en statistique, permettant de traduire la notion de la répartition normale dans des espaces différents, ainsi que la modélisation de phénomènes aléatoires tout simplement. Selon les thèses, il y a présence ou non d'une explication mathématique de l'usage du terme.

(4) Méthodes descriptives et exploratoires

L'analyse des données débute classiquement par une partie descriptive, permettant de résumer et ainsi de percevoir ce qui est contenu dans l'ensemble de ces informations. La description des données implique la présentation de la répartition variable par variable faisant appel aux statistiques vues précédemment, mais il s'agit aussi de décrire les relations, associations entre les différentes variables. Cependant, cette étape d'analyse descriptive est souvent inexistante au détriment d'une démarche intégrant rapidement dans la présentation une approche hypothético-déductive ou parfois théorico-inductive. La partie descriptive est ainsi assez souvent pauvre en information, s'appuyant principalement sur des représentations graphiques.

Les représentations graphiques sont de types et de formats très différents selon les disciplines.

Quand l'objectif de la représentation graphique est de représenter la répartition selon les modalités d'une variable, certains auteurs utilisent les graphiques partageant des surfaces, comme par exemple les graphiques en secteurs, les

graphiques ternaires³⁴. L'organisation de la surface d'intérêt (disque, triangle) représente la répartition, et offre une lecture rapide et simple. Le choix de l'une ou l'autre des présentations correspond aux habitudes disciplinaires.

FIG. 5 : REPRÉSENTATIONS DES RISQUES INHÉRENTS AU TOURISME
(D'après les entretiens réalisés entre juin 2007 décembre 2010).

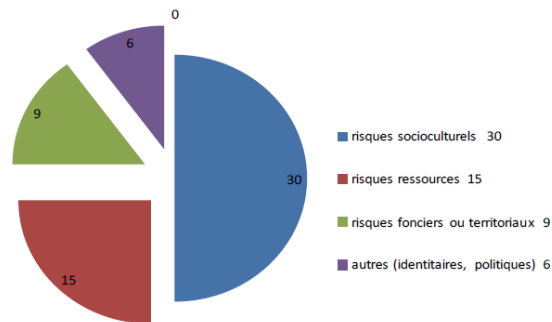


Figure 2 : diagramme en secteur – p214 -(Ndiaye, 2012) (B3 - Sciences de l'Information et de la Communication)

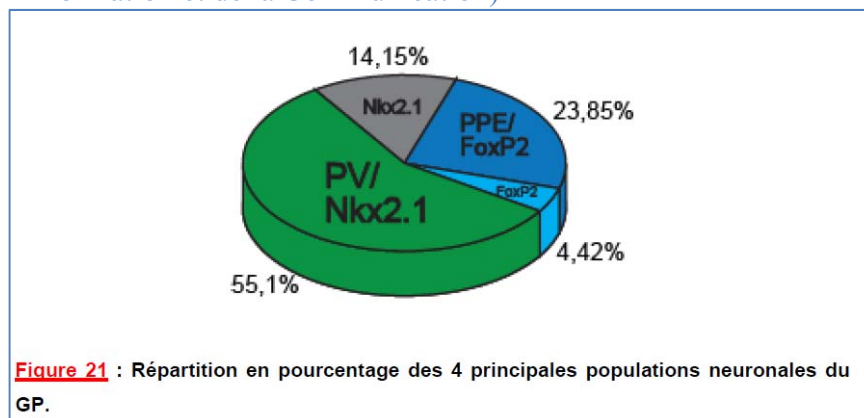


Figure 3 : exemple de "diagramme en secteur en 3D" – reproduction p 81(Abdi, 2013) (B2 – Neurosciences)

³⁴Il s'agit de la représentation de distribution selon trois modalités effectuée dans un triangle équilatéral. Chaque côté est gradué de 0 à 100% dans le sens des aiguilles d'une montre. Le point représentant la distribution est à l'intersection des trois droites parallèles aux côtés opposés.

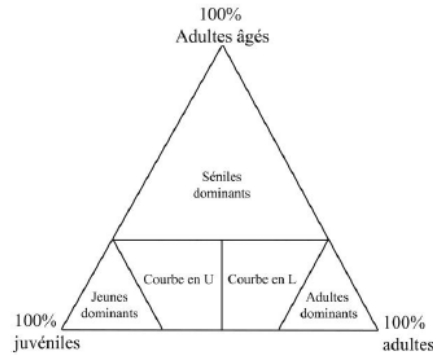


Figure 54. Diagramme ternaire proposé par M.C. Stiner (1990) avec les aires correspondant aux courbes d'abattage théoriques.

Figure 4 : exemple de représentation ternaire - reproduction p156 (Campmas, 2012) (B1 - Préhistoire et géologie du Quaternaire)

Pour l'ensemble des représentations graphiques, il est à souligner des manques d'informations (pas d'informations concernant les effectifs par exemple) ne permettant pas une compréhension du graphique isolé, dans la plupart des cas, et cela quelle que soit la discipline. La non-justification des choix de représentations graphiques sous-entend des habitudes communautaires de par la validation de la publication de la thèse.

TABLEAU 39 – Répartition des professeurs, par nombre de colles (clg Camus, hors SEGPA)

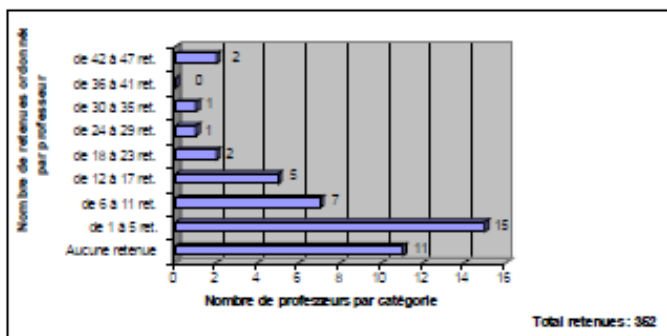


Figure 5 : quelle amplitude pour les intervalles? - reproduction p259 -(Garcia, 2013) (B2 – Sociologie)

Les représentations des répartitions de variables quantitatives sont réalisées par des « boîtes à moustaches », ces représentations permettent de représenter sur un même graphique plusieurs caractéristiques de tendance centrale et de tendance de

dispersion. Il est à souligner comme pour la représentation ci-dessous, que seule la connaissance préalable de ce type de représentation permet au lecteur d'en faire une lecture adéquate.

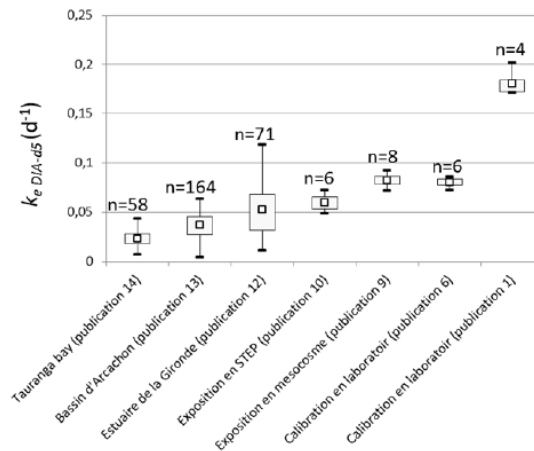


Figure 11 Comparaison des constantes d'élimination de la DIA- d_5 dans divers milieux.

Figure 6 : exemple de boîtes à moustache - reproduction p161 - (Belles, 2012) (B1 - Chimie analytique)

Les différentes représentations graphiques des thèses ne suivent pas de règles communes de qualité de présentation : les terminologies ne sont pas non plus partagées. Le terme histogramme n'est pas employé dans l'ensemble des thèses pour des représentations semblables. Classiquement un histogramme consiste en la représentation graphique de la distribution d'une variable quantitative continue suite à la répartition en classes des valeurs. Visuellement un histogramme correspond à une succession de rectangles accolés, traduisant la continuité de la variable, avec une information précise des valeurs minimales et maximales. La somme des surfaces des rectangles représentant l'effectif total de l'échantillon. Le terme histogramme est usuellement employé dans des situations moins spécifiques, par abus de langage. (cf. figure 7)

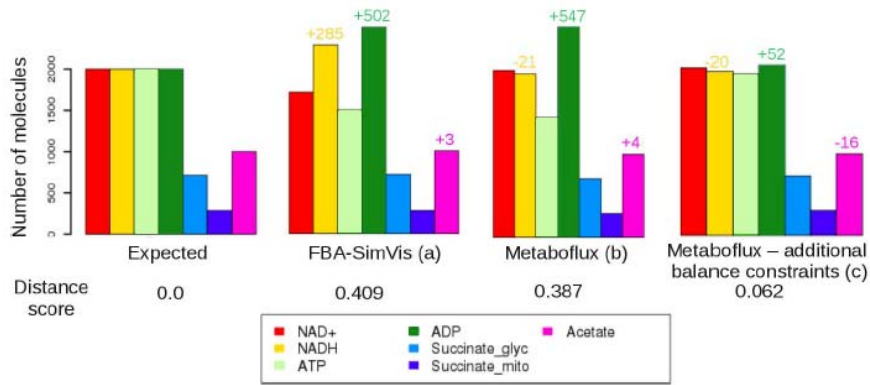


FIGURE 3.16 – Histogramme représentant le marquage final du FPN obtenu par FBA-SimVis et Metaboflux.

Nous considérerons ici le marquage final observé en moyenne pour 100 simulations du modèle PF pour les distributions de flux calculées par FBA-SimVis et Metaboflux. 1000 molécules (correspondant à des jetons au sein du FPN) de glucose et 2000 molécules d'ATP, d'ADP, de NAD^+ et de $NADH$ sont disposés en entrée du modèle. Le modèle de FBA-SimVis (a) et de Metaboflux (b) sont soumis à une contrainte sur la proportion des métabolites excrétés (succinate et acétate) correspondant respectivement à 36% pour succinate_glyc, 14% pour le succinate_mito et 50% l'acétate. Metaboflux (c) intègre une contrainte supplémentaire pour l'équilibre des rapports ATP/ADP et $NAD^+/NADH$. Le marquage final attendu correspond à 2000 molécules d'ATP, d'ADP, de NAD^+ et de $NADH$, 714 molécules de succinate_glyc, 286 molécules de succinate_mito et 1000 molécules d'acétate.

Figure 7 : exemple d'histogramme - reproduction p83 (Ghozlane, 2012) (B1 – informatique)

Il est à souligner une difficulté majeure et partagée sur l'ensemble des thèses quant à la capacité de représenter de manière synthétique et claire les informations sur une représentation graphique. Les graphiques diffèrent entre disciplines, ces dernières semblent être ainsi des niches spécifiques ayant des habitudes d'usages de graphiques différents.

(5) Intervalle de confiance

La notion d'intervalles de confiance permettant de définir un ensemble de valeurs vraisemblables du paramètre recherché sous certaines conditions est généralisé, dans certains habitats, par des terminologies spécifiques telles que « ellipses de confiance », « enveloppes d'erreur » pour des paramètres de dimension supérieure à 1. Il est à souligner que la notion d'intervalles de confiance est peu présente en regard de la présence des tests d'hypothèses

« La variation temporelle du champ magnétique est suivie par une série d'ellipses de confiance, calculées autour de la moyenne pondérée des directions

d'aimantation des structures pour une fenêtre de temps donnée. » (Hervé, 2012) (p32 - B3 - Physique des archeomateriaux)

« Par ailleurs, nous utilisons le programme statistique mis en place par T.E. Steele et T.D. Weaver (2002), T.D Weaver et al. , (2011b). Il permet, par des méthodes statistiques de ré échantillonnage, de placer sur les diagrammes ternaires des ellipses de confiance à 95 % afin de comparer statistiquement les échantillons. » (Campmas, 2012)(p159 – UB1 - Préhistoire et géologie du Quaternaire)

Par ailleurs, nous utilisons le programme statistique mis en place par T.E. Steele et T.D. Weaver (2002), T.D Weaver *et al.*, (2011b). Il permet, par des méthodes statistiques de rééchantillonnage, de placer sur les diagrammes ternaires des ellipses de confiance à 95 % afin de comparer statistiquement les échantillons. Ces ellipses reflètent la taille des échantillons utilisés (plus l'échantillon est conséquent plus l'ellipse est réduite). De plus, les échantillons sans chevauchement sont susceptibles d'avoir subi des histoires différentes (Figure 59).

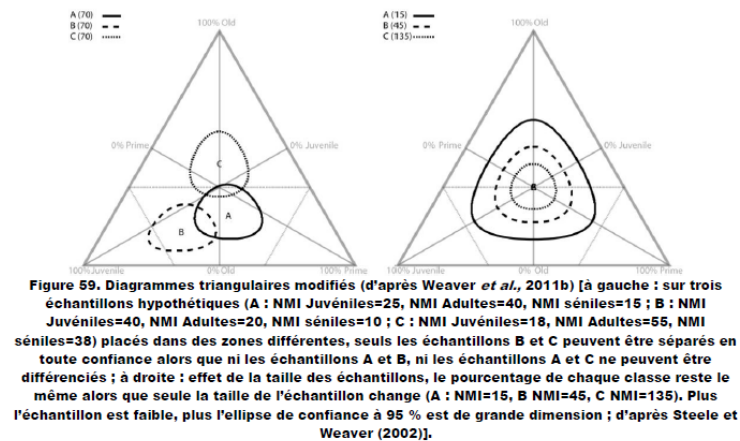


Figure 8 : exemples d'ellipses de confiance et interprétation - reproduction p159 (Campmas, 2012) (B1 - Préhistoire et géologie du Quaternaire)

L'explication de la représentation ci-dessus, figure 8, introduit la notion de test statistique au sens de Fisher, indiquant que des échantillons peuvent être considérés différents de par la non présence d'intersection entre leur ellipses de confiance.

La plupart des auteurs de thèses privilégient des approches tests d'hypothèses au sens de Neyman Pearson.

(6) Tests statistiques,

Les tests statistiques vont être présents tout particulièrement dans les thèses ayant une approche scientifique hypothético-déductive, pour lesquelles des hypothèses

auront été formulées, et le test statistique pourra alors être utilisé afin de permettre d'obtenir des éléments permettant de nourrir une prise de décision. Certains doctorants développent précisément les étapes de la démarche du test statistique, et spécifient surtout les hypothèses nulle et alternative ce qui dans l'ensemble des autres thèses utilisant des tests statistiques est implicite. Les formulations des hypothèses ne sont pas toujours rédigées en hypothèse de ce qui se passe théoriquement mais déjà en termes de formulation sur l'observation.

Troisième étape - Estimation de la significativité des enrichissements identifiés. La dernière étape consiste à déterminer la significativité des groupes par rapport aux données de l'organisme. Pour cela, deux hypothèses sont considérées : H_0 : L'ensemble identifié constitue une annotation non significative par rapport à l'organisme. H_1 : L'ensemble identifié constitue une annotation significative. Pour répondre à cette question, DAVID effectue un test de Fisher exact [Agresti 1992] modifié. Ce test consiste tout d'abord à calculer la table de contingence. (Ghozlane, 2012) (p126 - B1 – informatique)

Dans l'hypothèse où les paramètres sont pratiquement constants en fonctionnement normal, leur variabilité peut s'expliquer essentiellement par la présence des perturbations et de leur propagation. Ainsi, toute variation significative d'un ou plusieurs paramètres (par rapport à la valeur de référence) peut être interprétée comme l'apparition d'un défaut. Le problème que nous cherchons à résoudre consiste donc à choisir l'une des deux hypothèses suivantes :

- H_0 : le système est dans un état jugé normal, c'est-à-dire que les variations paramétriques observées ne sont pas provoquées suite à l'apparition d'un défaut, mais plutôt à cause du bruit et/ou des imprécisions de modèle ;
- H_1 : H_0 n'est pas vrai.

(Gheorghe, 2013) (p64 – B1 - Automatique et productique, signal et image)

Chaque test statistique a un cadre précis de contraintes d'utilisation particulières, tailles d'échantillon, type de distribution des variables car de par sa construction il nécessitera la détermination d'une statistique et de sa loi de probabilité. Ces contraintes sont parfois explicitement indiquées dans les corpus, mais dans la plupart des thèses ces informations sont passées sous silence. Par contre, ces explications sont presque toujours présentes dans les articles intégrés aux thèses sur articles, ce qui souligne le fait que dans le cadre d'une thèse l'explication de la partie méthode d'analyse de la démarche statistique ne composerait pas le point central attendu par le jury.

« Les tests statistiques classiquement utilisés pour l'analyse de résultat (analyse de variance à un ou plusieurs facteurs, Test de Student, Test de Duncan ou Test de Wilcoxon-Mann-Whitney (Au moins 5 répétitions indépendantes)) ont été

réalisés avec le tableur Excel (Microsoft, USA) ou avec le logiciel R-software (www.R.project.com) après avoir vérifié la normalité de distribution des résidus par le test de Shapiro-Wilk et l'homogénéité des variances avec le test de Levene (R's car package v 2.15.1). Le test de Duncan a lui été réalisé avec le package agricolae. » (Zimmer, 2013) (p79 – B2 – Œnologie)

« La taille de nos échantillons étant le plus souvent trop faible pour déterminer si la répartition des données suit une loi normale, nous avons utilisé des tests non-paramétriques. Pour les données appariées, nous avons utilisé le test de Wilcoxon et pour les données non appariées, celui de Mann-Whitney. Pour les comparaisons entre trois groupes, nous avons utilisé le test de Kruskal-Wallis avec une procédure post-hoc de Dunn. Les données ont été considérées comme significativement différentes pour des valeurs de p inférieures à 0,05. Tous les résultats dans les figures sont présentés sous la forme de la valeur moyenne \pm SEM (erreur standard à la moyenne). (p64) // Pour valider sur le plan statistique la régularité de la décharge d'un neurone, nous avons utilisé le test de normalité qui stipule que si la distribution des ISI a une valeur d'aplatissement « kurtosis » et d'asymétrie « skewness » entre -2 et 2, la distribution suit une loi normale (Figure 24 I). » (Abdi, 2013) (p88 – B2 – Neurosciences)

« La comparaison de variables quantitatives a été effectuée par l'utilisation du test T de Student lorsque la distribution des variables distribuées suivait une loi normale et du test de Kruskal-Wallis dans le cas contraire. » (Jaquet, 2012) (p82 - B2 - Epidémiologie)

Dans le cadre d'une recherche spécifique le chercheur est amené à élaborer les étapes de son analyse dans un but précis, et ces étapes doivent avoir été prévues de manière *a priori* au regard de son hypothèse et de son objectif d'étude et non pas au regard des données qui pourraient le détourner de sa question initiale et ce pour quoi il a recueilli telles données.

« Nous testons statistiquement la corrélation entre les pourcentages de survie et les densités osseuses pour tous les éléments squelettiques, ensuite nous excluons de ces tests de corrélation le squelette axial (sous-représenté dans les séries paléontologiques et archéologiques), puis nous ne considérons que les os longs (dont les extrémités sont préférentiellement détruites) et enfin nous nous concentrons sur les extrémités seules des os longs pour les taxons représentés par un nombre de restes autorisant une telle approche. Le test de corrélation utilisé est le coefficient de Spearman qui est un test non-paramétrique permettant de préciser l'existence et l'intensité de la corrélation entre deux variables quantitatives. Il porte sur les rangs du classement des valeurs et s'affranchit des conditions de normalité. Ces tests statistiques sont menés à l'aide du logiciel PAST. » (Campmas, 2012) (p172 – UB1 - Préhistoire et géologie du Quaternaire)

La présentation *a priori* de la démarche statistique d'analyse est une étape de validation de la démarche scientifique qui est présentée de manière systématique

dans les articles inclus dans certaines thèses et pas toujours dans les autres corpus de thèses.

(1) Outils de modélisation

Le terme modèle et modélisation couvrent des situations multiples avec des buts bien différents : au service de l'estimation d'association, ou de la recherche de facteurs explicatifs, mais aussi de la prédiction.

« La mesure d'association estimant le risque d'être infecté par le VIH chez les patientes atteintes de CIC par rapport à des patientes d'un groupe référent de comparaison a été estimée par l'intermédiaire d'un modèle de régression logistique non conditionnel. » (Jaquet, 2012) (p26 - B2 - Epidémiologie)

Il peut s'agir d'approche déterministe ou d'approche stochastique. Il s'agit de se référer à une base théorique et de pouvoir mettre en lien des éléments théoriques aux observations.

« Les modélisations probabilistes sont de plus en plus répandues dans des domaines aussi divers que la fiabilité industrielle, la finance ou encore la biologie. En effet, celles-ci permettent non seulement de rendre compte de phénomènes purement aléatoires mais également d'appréhender des systèmes déterministes pour lesquels on manque d'information ou de données observables ou qui sont soumis à des imprécisions de mesures ou encore à des aléas extérieurs. Enfin, certains systèmes déterministes du fait de leur grande complexité sont avantageusement assimilés à des modèles probabilistes. Les outils probabilistes adaptés à la description de l'évolution au cours du temps de tels modèles sont les processus stochastiques à temps continu. » (Brandejsky, 2012) (p9 – B1 – Mathématiques appliquées)

La réalité des situations spécifiques selon les domaines a induit le développement parallèle parfois d'outils de modélisation, ainsi sous des noms différents propres à chaque environnement disciplinaire, des méthodes similaires existent se basant sur les mêmes développements mathématiques. Le thésard peut ainsi parfois souligner sa connaissance épistémologique de l'outil ou méthode et des outils comparables existant dans d'autres espaces. Il justifie ainsi de la pertinence de son choix et sa connaissance de la méthode.

« Tout au long du XXème siècle, la SVD (ndlr décomposition en valeur singulières) a été réinventée à plusieurs reprises dans des domaines d'application différents et porte alors plusieurs autres noms. Elle réapparaît d'abord sous le nom d'Analyse en Composantes Principales (PCA - Principal

Component Analysis) en statistique. » (Godin, 2013) (p18 - B1 - Mécanique et énergétique)

Le développement de la modélisation se base sur le fait que la prise en compte de situations de plus en plus spécifiques complexifie et nécessite dès lors des développements qui peuvent être mathématique ou de programmation

« En connaissant l'impact que peut avoir l'effectif des subdivisions spatiales sur la variation aléatoire des indices statistiques, l'utilisation de ce modèle pour traiter des données moins agrégées imposera une reprogrammation du modèle. » (Dasré, 2012) (p59 - B4 – Démographie)

Cependant les références et formulations pouvant faire penser à une approche de type statistique mathématique se retrouvent quasiment uniquement dans le cadre de thèses en mathématiques appliquées.

« Les propriétés statistiques du speckle dépendent des propriétés du faisceau incident mais également des propriétés statistiques de la surface. Afin de simplifier l'analyse, certaines hypothèses doivent être faites concernant les propriétés statistiques de l'amplitude et la phase d'une contribution (p11) [...] On vient de démontrer que $AR(x; y)$ et $AI(x; y)$ suivent des statistiques gaussiennes et que $I(x; y)$ suit une loi exponentielle (p14). » (Le Cain, 2012) (B1 – Mathématiques appliquées)

La majorité des autres thèses ont une rédaction peu mathématisée des outils statistiques manipulés. Dans les autres thèses les modèles sont utilisés parfois comme des boîtes noires sans aucune explication des étapes. L'outil statistique est dès lors un moyen de répondre à une demande sans nécessité de comprendre l'ensemble du processus mathématique sous-jacent. Le thésard fait ainsi appel à la méthode citée mais sans en expliciter les éléments théoriques puisqu'il ne s'agit pas de sa discipline de rattachement.

Mettre en œuvre une modélisation nécessite une référence de modèle théorique associée à une réflexion et au choix d'un certain nombre d'informations : choix des variables prises en considération dans le modèle et critères de validation du modèle. Les critères de validation sont dépendants de la raison même du modèle, ils peuvent être purement probabilistes ou explicatifs.

« Le type de CHU n'a également pas été considéré dans le modèle multivarié car cette variable ne reflétait que l'orientation de chaque CHU en matière d'activité gynécologique. L'adéquation du modèle a été ensuite évaluée en utilisant le critère d'information d'Akaike (AIC), une valeur inférieure de l'AIC

suggérant une meilleure prédiction du modèle par rapport à un modèle précédent. Les facteurs qui n'étaient pas associés de façon significative avec la présence d'un CIC en retenant un seuil de significativité $<0,05$ et qui n'ajoutaient pas de valeur à la prédiction pour le modèle en se basant sur l'AIC étaient successivement supprimées. [...] En analyse multivariée, une procédure de type pas à pas descendante a été appliquée afin de déterminer le modèle prédisant le mieux la présence d'un CIC à partir des variables d'ajustement disponibles et susceptibles d'être associées au CIC. Les variables étaient retenues dans le modèle multivarié de départ si elles étaient associées à la présence d'un CIC en analyse univariée avec un seuil de significativité $\alpha < 0,20$. » (Jaquet, 2012) (p26 – B2 – Epidémiologie)

Les étapes ci-dessus consiste à déterminer le meilleur modèle dans un ensemble de modèles possibles tous construits sur le même modèle théorique mais pouvant être développés à partir de variables différentes choisies selon des critères statistiques, empiriques ou de bon sens selon la communauté.

Sept thèses ne font pas références à des modélisations. Pour les autres la situation la plus fréquente est une modélisation par régression linéaire ou logistique.

(2) Régression linéaire et corrélation

Les mots appartenant à la famille de corrélation sont fréquemment employés pour caractériser des situations d'association au sens de la définition première du Larousse : « *Relation existant entre deux notions dont l'une ne peut être pensée sans l'autre, entre deux faits liés par une dépendance nécessaire* ». En statistique, la corrélation a un sens plus restreint « *Liaison entre deux caractères (corrélation simple) ou plus (corrélation multiple) telle que les variations de leurs valeurs soient toujours de même sens (corrélation positive) ou de sens opposé (corrélation négative)* ». Cette définition sous-entend une association entre deux variables quantitatives. Cet usage est, quant à lui, moins fréquent. Les exemples ci-dessous soulignent la place important de l'usage du terme corrélation au sens d'association.

« L'analyse rapide des résultats de l'ACM représentés selon les axes factoriels 1 et 2, montre une forte corrélation entre les caractéristiques des séjours dans la région de l'estuaire entrepris par des touristes qui ont pour motivation principale la pratique des activités récréatives (ActRec1) et qui ont choisi le camping (Camping) comme principal mode d'hébergement. La motivation principale de la visite des parents et amis (Parent1) se trouve fortement corrélée avec le choix du mode d'hébergement non marchand. » (Ben Othmen, 2013) (p111 - B4 - Sciences économiques)

« Les résultats du test du khi2 montrent une très forte corrélation entre le sexe et la probabilité d'être collé. » (Garcia, 2013) (p249 – B2-sociologie)

« Il semble qu'il y ait une corrélation entre choix du mode d'aumône dotale et origine sociale. 53% de dots de la part de la noblesse d'épée au XVIIIe contre 100 % pour les officiers ; 50 % contre 75 % début XVIIIe siècle. » (Tissot, 2012) (p301 - B3 - Histoire moderne et contemporaine)

« Il y a une corrélation évidente et tangible de l'espace urbain, des TIC et de la modernité d'une part (p93) [...] La corrélation entre voies de communication, presse et espace est un élément déterminant dans le clivage existant entre les habitants d'un espace physique supposé être le même. (p142) » (Diouf, 2013) (B3 - Sciences de l'information et de la communication)

« Dans un premier temps, des paléothermomètres sont construits au moyen de deux approches : 1. la thermoluminescence (TL) fondée sur les modifications électroniques des grains de quartz ; 2. Le magnétisme basé sur les modifications minéralogiques des oxydes de fer contenus dans les sédiments. Ces deux approches sont indépendantes et la corrélation des résultats obtenus permet d'étudier leur potentiel comme paléo thermomètre. » (Brodard, 2013) (p53 - B3 - Physique des archeomatériaux)

Dans le cadre de régression et corrélation linéaires, les auteurs peuvent choisir de représenter les nuages de points pour présenter visuellement les effets d'association.

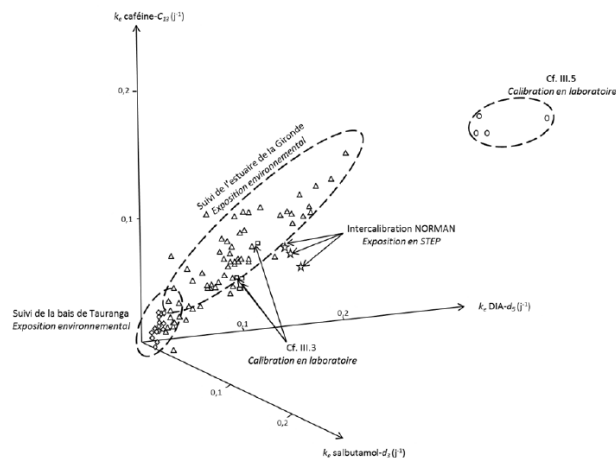


Figure 10 Corrélation entre les constantes d'élimination des trois PRC étudiés (n=145).

Figure 9 : exemple 1 de représentation d'un nuage de points en 3 dimensions (p160) (Belles, 2012) (B1 - Chimie analytique)

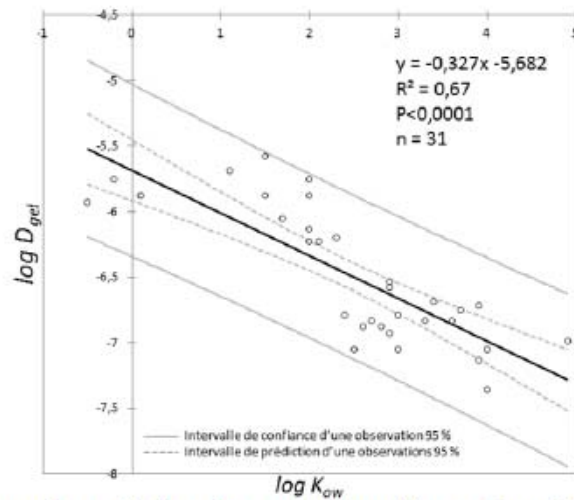


Figure 17 Corrélation entre le coefficient de partage octanol/eau et le coefficient de diffusion dans le gel polyacrylamide.

Figure 10 : exemple de représentation d'une régression linéaire (p263) (Belles, 2012) (B1 - Chimie analytique)

La mise en place d'une régression linéaire se base sur la méthode des moindres carrés, méthode étant un élément considéré de base en statistique. L'hypothèse de la normalité de la distribution de la variable à expliquer est nécessaire pour la mise en œuvre de cette méthode, et dans ce cas, soit le doctorant fonde son raisonnement en faisant l'hypothèse de cette normalité sans essai de validation, et sans réelle vérification même graphique, soit il met en œuvre des tests statistiques d'hypothèse pour tester si cette hypothèse de normalité est valide. Dans la plupart des situations, le doctorant omet de présenter les contraintes d'usage.

« L'ajustement de la courbe aux données est défini aux moindres carrés en supposant que les données suivent une loi d'échantillonnage gaussienne. » (Hervé, 2012) (p162 - B3 - Physique des archeo- matériaux)

« Avant de passer à l'interprétation des variables on a vérifié les conditions d'application de la régression. Presque aucune variable n'a une distribution normale en tenant compte du indicateur KS11. » (Gheorghiu, 2012) (p95 - B2 – psychologie)

(3) Avec quel moyen technique – logiciel ?

Pour la mise en œuvre de leurs analyses, un ou des outils techniques peuvent être nécessaires, et il s'agit tout particulièrement de logiciels permettant la mise en œuvre des méthodes que les doctorants souhaitent déployer. Dans 26 thèses,

le nom du ou des logiciels utilisés n'a pas été précisé, et, dans les autres situations il y a une réelle hétérogénéité de logiciels utilisés pouvant traduire des habitudes de communautés disciplinaires avec des logiciels statistiques, majoritairement libres, c'est-à-dire gratuits, très spécifiques qui ont d'ailleurs, pour la plupart, été développés par des chercheurs ou communautés de chercheurs (logiciel Kaleidagraph (logiciel payant pour l'analyse et la visualisation 2D de données expérimentales), logiciel libre PAST «PALaeontological STatistics», logiciel ImageJ (logiciel libre de traitement et d'analyse d'images particulièrement adapté aux applications en biologie et en médecine), ou des logiciels payants de statistique STATA, SAS, Statview et SPAD. Le logiciel le plus cité, dans notre échantillon par 4 doctorants, est le logiciel R, qui est un logiciel libre de traitement de données et d'analyses statistiques non spécifiques à un domaine d'application, disposant de nombreuses fonctions graphiques et ne cessant de se développer par la création d'extensions (appelés « packages ») par les membres de la communauté. Ces extensions permettent ainsi de couvrir un très large champ de méthodes. Ainsi, qu'il s'agisse de doctorants de Sciences économiques, d'écotoxicologue, d'œnologie ou d'informatique, ils trouvent les outils nécessaires à leur analyse.

Le passage entre les résultats communiqués par les sorties logiciels et la diffusion scientifique n'est pas une étape simple et nécessite une réelle transformation quant à la forme. Le travail de mise en forme est important et nécessite une bonne compréhension des étapes réalisées par le logiciel. La diffusion de sorties d'ordinateur brutes, c'est-à-dire sans mise en forme et sans épuration, existe dans certaines thèses. Plusieurs raisons peuvent être avancées : le thésard préfère présenter l'exhaustivité de l'information, ou le thésard a des craintes de ne pas sélectionner les informations importantes, ou il n'est pas complètement à l'aise avec la production de ces résultats. Ces trois pistes de réflexion ont un certain nombre de points communs basés tout particulièrement sur la relation que le doctorant entretient avec la statistique.

Tableau 3 – Attitude conditions d'application de l'analyse factorielle exploratoire – Roumanie

Matrice de corrélation	corrélations moyennes
Déterminant	1,15E-031 de 0.00001
KMO test	.811
Bartlett test	8584.89, p=0.000
Matrice anti-image	valeurs au-dessous de 0.50

Tableau 4 – Attitude conditions d'application de l'analyse factorielle exploratoire - France

Matrice de corrélation	corrélations moyennes
Déterminant	6.82E-015 de 0.00001
KMO test	.703
Bartlett test	2582.38, p=0.000
Matrice anti-image	valeurs au-dessous de 0.50

Figure 11 : exemple de sortie d'écran de logiciel non retravaillé (Gheorghiu, 2012) (p93- B2 – psychologie)

Nous pouvons remarquer que dans certaines situations les doctorants vont reproduire les résultats statistiques dans leur texte en reprenant à l'identique les informations chiffrées issues des logiciels, sans ajustement de l'information.

« Une ANOVA révèle une différence significative entre les moyennes de seuil de viabilité et les translocations ($P=0,02873$). » (Zimmer, 2013)³⁵ (p195 - B2 – Enologie)

h) Quelles décisions avec les données ?

Les résultats obtenus à l'issue de la phase d'analyse statistique sont l'objet de diffusion et de prise de décision tout particulièrement dans le cadre d'une démarche hypothético-déductive. Une analyse descriptive ponctuelle sur la quasi-totalité de la population d'intérêt semble souvent le moyen de délivrer un résultat le plus fiable possible avec les données accessibles.

« Le premier recensement en 1790 nous permet d'avoir une estimation fiable de la population des seigneuries de l'abbaye. » (Tissot, 2012) (p200 – B3 - Histoire moderne et contemporaine)

Mais la situation de recensement est peu fréquente, et l'auteur doit proposer des pistes de décision en fonction de ces observations et du cadre défini en amont. L'existence de la question de variabilité inhérente à la pratique statistique induit l'impossibilité de l'obtention d'un résultat en tout ou rien et donc tout résultat doit

³⁵ Comme le soulignait Alfred Sauvy « Dans toute statistique, l'inexactitude du nombre est compensée par la précision des décimales. »

être entendu avec son niveau d'incertitude, cette considération étant fondamentale dans l'analyse de résultats statistiques.

(1) L'argumentation du soi-disant non résultat

Il est difficile pour le doctorant d'accepter que parfois les résultats ne soient pas aussi clairs que souhaités pour diverses raisons, mais souvent pour des raisons d'effectifs.

« Il faudrait un plus grand nombre d'échantillons pour en déduire une tendance. » (Cuenca, 2012) (p34 - B1 - Physique-chimie)

Le doctorant semble parfois pousser par un *a priori* et sa déception peut se lire dans les formulations employées dans la conclusion des résultats. L'objectif du doctorant est de pouvoir découvrir, montrer, mettre en évidence, il peut, alors faire des choix de présentation, lui permettant de soutenir cette attente.

« Pour simplifier l'analyse des résultats, seules les valeurs les plus significatives sont présentées » (Aime, 2013) (p151 – B2 – Biochimie-Chimie)

« Pour accroître le sens de ces résultats, on peut analyser les bilans qui affichent les plus hautes plus-values. » (Garcia, 2013)(p194 – B2-sociologie)

« Cette estimation était à la limite de la significativité. » (Foubert-Samier, 2013) (p146 - B2 – épidémiologie)

Il est intéressant de remarquer que ce type de pratiques dépasse les frontières disciplinaires. Les contraintes du chercheurs seraient-elles partout les mêmes ?

(2) Prise de décision

Prendre une décision à l'issue d'un test d'hypothèse, c'est conclure au rejet ou non de l'hypothèse en fonction d'un seuil défini *a priori*. Le choix de ce seuil est, lorsqu'il est signalé, de 5% dans l'ensemble des thèses étudiées. Cependant, certains auteurs indiquent leur résultat non pas en se référant à un seuil défini *a priori*, mais en indiquant le seuil qu'il aurait fallu choisir pour pouvoir dire qu'il y a un résultat significatif, l'observation prédomine sur l'hypothèse. Dans cette situation la prédominance est donnée au calcul de la valeur de p. La valeur de p est utilisée dans une situation de vouloir quantifier l'"inquantifiable".

Dans les situations où les critères de décision sont précisés *a priori* les formulations peuvent prendre les formes ci-dessous.

« La régression quantile, elle, identifie la durée du séjour (moyen à long), le revenu et la motivation des touristes ainsi que le nombre d'enfants âgés de 10 à 14 ans dans le groupe de touristes comme significatifs à 1% au niveau du quantile de dépenses 25%. En outre, le choix de formule à plusieurs prestations est identifié comme statistiquement significatif au seuil de 5%. » (Ben Othmen, 2013) (p165 - B4 - Sciences économiques)

« Si la fréquence de décharge du neurone après l'hyperpolarisation est supérieure à la fréquence basale moyenne de ce neurone plus 3 fois l'erreur standard à la moyenne, nous avons considéré que le neurone possède un rebond-post-inhibiteur. » (Abdi, 2013) (p64 - B2 – Neurosciences)

Certaines décisions peuvent être issues d'analyse de représentations graphiques.

« En comparant une des relations [précédentes] avec des seuils fixes, il est possible de conclure sur la présence d'une anomalie dans le système. Une panne est détectée (hypothèse H_1 retenue) si le résultat d'un test est supérieur à un seuil, dont le choix est un compromis entre la sensibilité de détection et la faible probabilité de fausse alarme. Pour un des test, le seuil est choisi selon un intervalle de confiance donné qui dépend de la probabilité d'erreur (la probabilité du rejet de H_0 , quand H_0 est vrai). » (Gheorghe, 2013) (p66 - B1 - Automatique et productique, signal et image)

La prise de décisions est centrale dans le cadre de la mise en œuvre d'une démarche statistique au service de la justification et l'argumentation de la thèse développée par le doctorant. Les éléments de précision quant à la démarche de décision peuvent donc être plus ou moins spécifiés. Nous pouvons, cependant, souligner que les thèses issues des écoles doctorales de sciences et techniques présentent avec plus de précisions les démarches de décision.

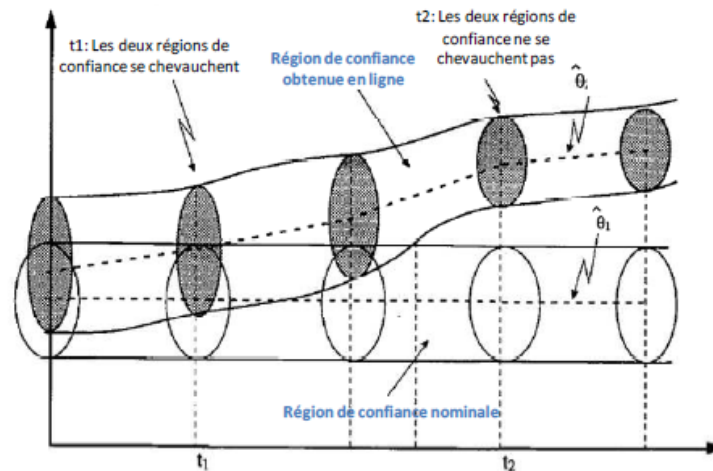


Figure III-1 : Disposition des ellipsoïdes de confiance pour la détection d'une panne

Test RC2
<p>(i) Vérifier que : $(\hat{\theta}_2(k) - \hat{\theta}_1)^T (C_2(k))^{-1} (\hat{\theta}_2(k) - \hat{\theta}_1) > \alpha$.</p> <p>Si l'inégalité n'est pas vérifiée, alors les deux régions se chevauchent et on a un fonctionnement normal. En conséquence, la procédure s'arrête.</p>
<p>(ii) Chercher avec la méthode de dichotomie, la racine négative unique λ_0 de $F(\lambda)$, où :</p> $F(\lambda) = (V(\lambda))^T (C_2(k))^{-1} V(\lambda) - \alpha$ $V(\lambda) = [\lambda (C_2(k))^{-1} - (C_1)^{-1}]^{-1} (C_1)^{-1} (\hat{\theta}_2(k) - \hat{\theta}_1)$ $F(-\infty) = -\alpha \text{ et } F(0) > 0$ <p>avec une valeur initiale $A < 0$ tel que $F(A) < 0$.</p>
<p>(iii) On considère la fonction: $W = [\lambda_0 (C_2(k))^{-1} - (C_1)^{-1}]^{-1} (C_2(k))^{-1} (\hat{\theta}_2(k) - \hat{\theta}_1)$</p> $\text{test} = \lambda_0^2 W^T (C_1)^{-1} W$ <p>Si $\text{test} > \alpha$, alors les deux régions ne se chevauchent plus. Une panne est ainsi détectée.</p>

Figure 12 : exemple de décision de test - (Gheorghe, 2013) (p80 - B1 - Automatique et productique, signal et image)

(3) Utilisations du terme significatif

Seulement 13 des 55 thèses n'utilisent pas le terme significatif dans un contexte statistique, c'est-à-dire lié à un contexte de résultats de tests d'hypothèse ou de modélisation. Pour les 42 autres, ce terme est employé à l'issue de la réalisation d'un test statistique ou de la mise en œuvre d'une modélisation. Cette information de résultats significatifs est communiquée soit dans le corps du texte ou directement sur des représentations graphiques.

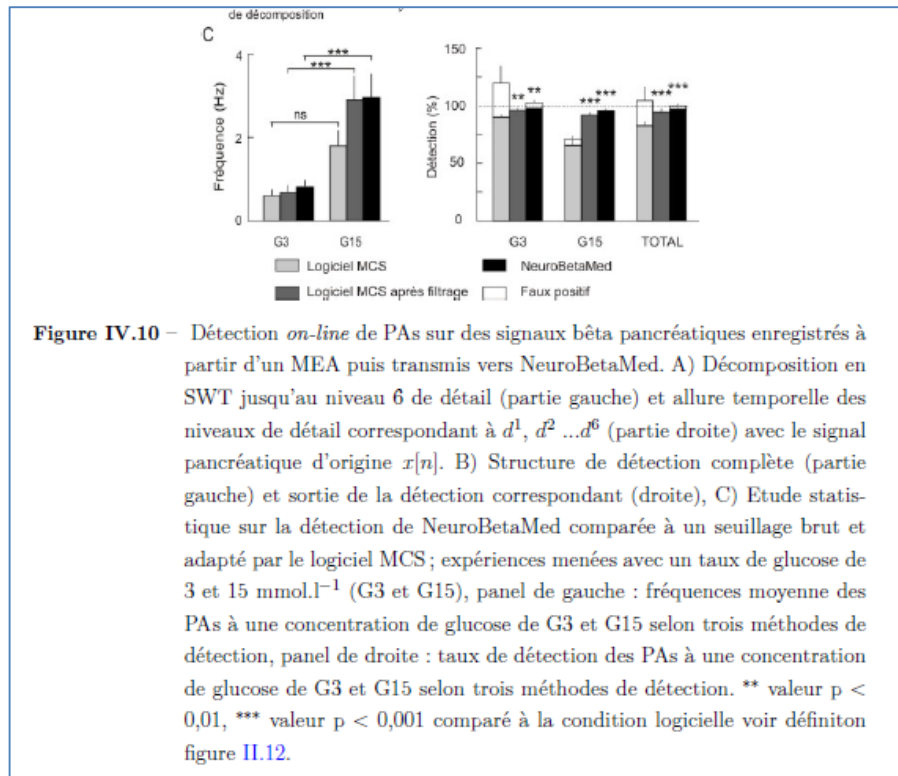


Figure 13 : reproduction p97 – fréquence – test (Quoth, 2012) (B1 – Electronique)

Certains doctorants justifient et explicitent l'usage qu'ils font du terme significatif. Ils précisent qu'un résultat est considéré comme significatif dès lors que la probabilité d'erreur de conclusion est inférieure à un seuil de décision défini *a priori* et communiqué.

“Results are given as mean \pm the standard error of the mean (SEM). All analyses were done using the R language and environment for statistical computing (R Development Core Team, 2009). For all statistical results, a probability of $P < 0.05$ was considered significant. Survival rate was presented using Kaplan Meier's estimator (Kaplan, 1983).” (Kaddissi, 2012) (p140 - B1 – Ecotoxicologie)

« Les données ont été considérées comme significativement différentes pour des valeurs de p inférieures à 0,05. Tous les résultats dans les figures sont présentés sous la forme de la valeur moyenne \pm SEM (erreur standard à la moyenne). » (Abdi, 2013) (p65 - B2 – Neurosciences)

« Les cercles de confiance se recoupent, ce qui indique qu'il n'existe pas de différence statistiquement significative entre les deux. » (Campmas, 2012) (p279 - B1 - Préhistoire et géologie du Quaternaire)

Le terme "*significatif*" est utilisé par les doctorants pour mettre l'accent sur la vraisemblance d'un résultat consistant souvent en la comparaison de différentes estimations ou d'une modélisation.

L'usage de la formulation « statistiquement significatif » précise qu'il y a eu mise en œuvre d'un test statistique avec la définition d'un seuil, tandis que signifier simplement qu'une différence est significative de manière subjective en considérant que deux choses semblent suffisamment différentes pour le doctorant ne porte pas le même sens de validation puisque cette seconde formulation ne dépend que de l'investigateur et l'algorithme de décision ne peut être reproduit à l'identique.

« Toutefois nous pouvons noter quelques écarts significatifs, les revenus de Marennes n'augmentent que de 43 % tandis que ceux de Saint-Sornin connaissent une forte augmentation de 150 %. » (Tissot, 2012) (p381 – B3 - Histoire moderne et contemporaine)

« Une exposition à 2.51 µg/L de Cd a entraîné une diminution de 50% des femelles gravides, une diminution significative du nombre de larves par portée (diminution de 27% à 1.51 µg/L ; et de 36% à 2.51 µg/L). » (Kaddissi, 2012) (p63 - B1 – Écotoxicologie)

Les résultats des tests statistiques dépendent des cadres contextuels dans lesquels ils sont utilisés, qu'il s'agisse du type de données, de leur effectif et du choix du test. Mais bien souvent la seule information qui est conservée dans la diffusion est le résultat lui-même, c'est-à-dire le fait que le test ait conduit ou non à une décision de différences statistiques. Le processus ayant permis d'obtenir cette décision disparaît tant et si bien que lorsque les doctorants font l'état de l'art et les revues de littérature les résultats des études antérieures sont transmis sans réelle précision quant à l'interprétation des résultats

« Entre 1990 et 2011, Sharp et Gatz ont dénombré 51 études montrant une liaison significative et 37 une absence de liaison. » (Foubert-Samier, 2013) (p9 - B2 – épidémiologie)

La décision prise par résultat probabiliste dans l'interprétation des résultats peut parfois sembler se jouer à peu de choses avec la prise en compte ou non de certaines variables sans toujours de justification explicite.

« Cependant, après ajustement supplémentaire sur le score au Mini-Mental State Examination (MMSE) et la présence de symptômes dépressifs à l'inclusion, la signification statistique disparaît. Cette diminution de l'association entre jeux de société et démence après ajustement supplémentaire, semble être en grande

partie due à l'effet du MMSE à l'inclusion, comme le montre des analyses complémentaires ajustant séparément sur le MMSE à l'inclusion et la dépression. » (Foubert-Samier, 2013) (p110 - B2 – épidémiologie)

i) Formulation du discours

La diffusion d'informations quant aux méthodes et démarches utilisées nécessite une formulation claire et précise, afin de développer une syntaxe au service de la sémantique et ceci afin de limiter les niveaux d'interprétation. Afin d'être reconnu au sein de sa communauté, le doctorant se doit d'explicitier avec clarté et qualité son savoir, pour que celui-ci soit correctement compris et apprécié. Dans la transmission du message, deux éléments se mesurent : ce que l'on dit et comment on le dit. Nous ne nous intéressons pas ici à ce qu'on veut dire qui n'a pas été étudié et qui nécessiterait un travail d'observation et d'échange avec les doctorants.

Certaines formulations utilisées par les doctorants pour décrire des résultats statistiques correspondent au "parlé" de statisticiens dans le « feu de l'action », qui par excès de simplification des formulations nomment les variables et informations uniquement par les libellés courts de leurs bases de données, induisant ainsi une perte de sens pour le lecteur extérieur.

« Leurs résultats pointent une relation statistique négative significative entre les dépenses touristiques et le nombre d'enfants rentrant dans la composition. » (Ben Othmen, 2013) (p176 - B4 - Sciences économiques)

« Le test du khi2 ne montre pas de liaison statistiquement significative entre la probabilité de coller et le sexe du professeur. » (Garcia, 2013)(p260 – B2- sociologie)

« L'effet des passagers est significatif jusqu'à un certain âge et puis il décroît en suggérant que la simple présence des pairs n'est pas suffisante pour entraîner des comportements à risque. [...] Cependant, les hommes de 20 à 25 ans et plus ont significativement plus d'accidents que les femmes de même âge. » (Gheorghiu, 2012) (p23 - B2 – psychologie)

5. Discussion

Nous pourrions reprendre à notre compte un certain nombre de remarques faites par les doctorants concernant cette analyse : peut-on considérer que notre échantillon de thèses est suffisant ? De quoi est-il réellement représentatif ? Et nous

pourrions nous justifier de ces choix par des éléments de réponse tels que : « *notre démarche ne nécessitait pas une représentativité de l'ensemble des thèses, mais devait nous permettre de repérer les régularités et les spécificités disciplinaires ou d'un autre niveau.* »

L'étude de ces 73 thèses initialement nous a permis ainsi de repérer un certain nombre d'éléments particulièrement intéressants.

La proportion de 55 thèses sur 73 se référant, selon les thèses, à des éléments statistiques ou à une totale démarche statistique est un marqueur de la place de cette discipline dans la construction de thèses développées par les chercheurs de nombreuses disciplines. Les thèses de sciences expérimentales, démographie, santé publique, et économie sont tout particulièrement impactées par ces usages mais il ne faut pas négliger la part occupée par la statistique dans les autres thèses.

Un certain nombre de régularités et de différences ont ainsi été repérées entre disciplines.

Une grande hétérogénéité a ainsi été soulignée quant au niveau d'explications et de précision lors de la présentation de l'échantillonnage et du recueil de données d'intérêt pour l'analyse statistique. Il est à souligner de réelles spécificités disciplinaires. Pour les disciplines de sciences et techniques, et de sciences de vie et de la terre, les processus de recueil, associés souvent à des situations expérimentales, sont présentés avec beaucoup de précision ; cette étape étant considérée pour ces disciplines comme centrales de la démarche statistique.

Quelles que soient les disciplines, les termes les plus classiquement connotés, tels que certains indicateurs de la statistique : moyenne, écart-type, corrélation, sont plutôt utilisés et formulés de la même façon mais rarement redéfinis, induisant dès lors que les lecteurs comprennent et partagent le même sens de manière implicite. Ces notions sont vues dans tous les enseignements de bases en statistique ou ont une existence en dehors des murs scolaires, ce qui peut ainsi en justifier cet usage. La notion de *moyenne*, par exemple, est ancrée dans l'esprit de tout un chacun par sa présence quotidienne dans tous les foyers grâce ou à cause de l'information de la météorologie. La notion de *corrélation* est, par contre, plus difficile à appréhender car, dans le cadre d'une approche statistique, elle sera associée à l'étude de relation

entre deux variables quantitatives (ou plus), alors que cette spécificité n'existe pas dans l'acception première de ce terme.

La moyenne, dans la majorité des thèses, caractérise de manière implicite la moyenne arithmétique, sauf lorsqu'il est précisé qu'il s'agit d'une autre formule de moyenne. Le terme "*moyenne*" est utilisé de manière indifférenciée pour représenter soit une statistique au sens fonction mathématique de données issues d'un échantillon, soit le résultat de cette fonction pour un échantillon spécifique, c'est-à-dire au sens d'indicateur. Les résultats de moyenne sont le plus souvent présentés accompagnés d'un indicateur représentant la notion de variabilité. Selon les habitudes disciplinaires cette présentation peut être une barre d'erreur représentant la notion de \pm un écart-type, représentant ainsi l'information du niveau d'hétérogénéité de la distribution de la variable étudiée. Ce type de présentation est particulièrement présent dans le cadre des thèses de l'école doctorale sciences physiques et de l'ingénieur, et principalement en physique et chimie. D'autres habitudes disciplinaires consistent à représenter l'information avec l'erreur standard de la moyenne. Dans cette situation, il ne s'agit pas de décrire la distribution de la variable initialement recueillie, mais de déterminer un niveau de précision de l'estimation de la moyenne ; il s'agit d'ailleurs d'une situation plus proche de la présentation que l'on trouverait en utilisant un intervalle de confiance. Le choix de telle ou telle forme de représentation n'est pas souvent justifié par les auteurs, mais est implicite pour les lecteurs de la même communauté. Cependant, il est à souligner que la rédaction des articles est beaucoup plus précise et suggère moins d'implicite. Cela peut suggérer que les codes disciplinaires ne passent peut être pas les frontières, ou que les revues scientifiques souhaitent offrir leurs éditions à un public plus large qu'une simple communauté.

Il est à souligner une plus forte hétérogénéité dans le cadre des représentations graphiques, tant pas le contenu que par la forme. Le but premier de l'utilisation de représentations graphiques est d'assurer une expression claire et compréhensible d'informations qui pourraient nécessiter un texte long et difficile. Cependant, la lecture des différentes thèses permet de mettre en évidence que l'usage de ce type d'outils de diffusion de résultats est difficile à appréhender, les graphiques proposés par les doctorants n'étant globalement pas jugés comme aidant pour la

compréhension. Il s'agit d'un point qui sera repris dans le chapitre 4 de cette thèse, en tant qu'élément important pour la construction de contenu de formation.

Les logiciels statistiques utilisés sont multiples et souvent différents d'une discipline à l'autre. Une culture disciplinaire d'usage des outils informatiques est réellement présente, et il est à noter que certains logiciels ont été directement développés par des chercheurs de la discipline, ceci marquant clairement l'existence d'outils statistiques propres à certaines disciplines nécessitant des développements informatiques spécifiques. Il est cependant important de souligner que le logiciel libre R, le plus cité dans des thèses de disciplines différentes permet de couvrir un champ très large d'usages de par le développement de paquets spécialisés par toute une communauté de chercheurs mettant en accès les programmes qu'ils développent dans le cadre de leur recherche. Des communautés nationales et internationales d'usage existent et sont particulièrement actives, de nombreux forums permettent à chacun de se former et de solliciter de l'aide si besoin.

La présentation des méthodes statistiques est plus ou moins explicitée selon les thèses et il en est de même de l'explication des résultats obtenus. L'analyse des contenus des différentes thèses suggèrent que le développement de la partie méthode statistique est proportionnel à l'habileté du doctorant dans ces usages et dans sa compréhension des objets. Ainsi, lorsque certains doctorants intègrent de façon brute les sorties de logiciel, on peut s'interroger sur le manque d'effort didactique. Est-ce une pratique due à l'inquiétude de ne pas tout transmettre au lecteur pour une analyse complète de la situation ? Cependant, ce type de comportement peut suggérer un manque de compréhension de la totalité de la sortie du logiciel et possiblement un défaut de compétence. Mais ceci ne peut en aucun cas être validé sans une approche plus complète dépassant la simple étude d'écrits comme cela est fait dans le cadre de cette thèse et ne peut se baser en l'état que sur le fait « *Ce que l'on conçoit bien s'énonce clairement, Et les mots pour le dire arrivent aisément.* »³⁶

La place importante de l'activité de modélisation est à souligner dans les thèses, avec une place assez importante des modèles de régression classiques dans les usages toutes disciplines confondues. Plusieurs doctorants signalent

³⁶ Nicolas Boileau, (1674) *l'Art Poétique*

s'être intéressés à des modèles utilisés dans d'autres communautés que leur communauté, et cela traduit réellement un effet de changement d'habitat des modèles qui, dès lors, évoluent souvent différemment dans chaque environnement mais cependant avec des bases et cadres communs. Il faut souligner cependant que les thèses utilisant des modèles plus sophistiqués ou moins classiques sont des thèses ayant dans leur ensemble une approche statistique assez riche, avec des développements d'écriture mathématique.

Il s'agit d'un élément important dans la formation de l'esprit statistique et du raisonnement statistique

« Le modèle devient un instrument cognitif du raisonnement statistique. Dans l'esprit statistique, ce raisonnement s'appuie sur une préférence donnée à une connaissance incertaine en tant que connaissance vraisemblable, plausible plutôt qu'à l'ignorance [...] Cette acceptation du risque dans un raisonnement à visée scientifique est nodale³⁷ dans l'esprit statistique et passe très probablement par une sorte de catharsis³⁸ analogue à celle dont parle Bachelard dans la formation de l'esprit scientifique. » (Régnier, 2005)

La place de la démarche statistique dans la démarche scientifique se traduit dans le cadre des thèses par un questionnement assez présent de la modélisation. Dans le cadre de sa thèse, le doctorant souhaite construire un nouveau savoir pouvant nourrir la recherche de sa communauté. La validation du statut savoir de sa production passe par des critères de construction pouvant varier selon les disciplines.

6. En conclusion

Au moins trois facteurs peuvent influencer l'importance de la place de la statistique dans la construction de l'argumentaire d'une thèse : la formation antérieure du doctorant et son propre jugement de ses compétences dans ce champ disciplinaire, la représentation par son directeur de thèse de la statistique comme

³⁷ Qui constitue le nœud, le point essentiel d'une question.

³⁸ Pour Aristote, effet de « purification » produit sur les spectateurs par une représentation dramatique. Toute méthode thérapeutique qui vise à obtenir une situation de crise émotionnelle telle que cette manifestation critique provoque une solution du problème que la crise met en scène.

discipline indispensable dans la mise en œuvre d'une recherche scientifique et, enfin, la reconnaissance par la communauté scientifique de la légitimité de la statistique.

Le travail que nous avons mené induit une difficulté que Isabelle Stengers a soulevé dans « L'invention des sciences nouvelles :

« En d'autres termes, le scientifique, de produit de son époque, est devenu acteur, et, s'il ne faut pas se fier, comme l'avait affirmé Einstein, à ce qu'il dit qu'il fait, mais regarder ce qu'il fait, ce n'est pas du tout parce que l'invention scientifique excéderait les mots, mais parce que les mots ont une fonction stratégique qu'il faut savoir déchiffrer. » (Stengers, 1993)

Il est à noter ainsi qu'un socle de connaissances, de savoirs et savoir-faire est partagé par tous les utilisateurs, mais se greffent dessus de réelles spécificités épistémologiques des champs disciplinaires, qui rendent nécessaires d'explicitier l'implicite dès lors qu'il est choisi de pouvoir faire un travail en interdisciplinarité.

« *Dans de nombreuses situation les chercheurs ne cherchent pas à travailler sur une dialectique commune* » disait G. Brousseau lors d'un séminaire organisé par l'équipe de recherche ADES du 14 octobre 2013. Ce constat peut expliquer la difficulté de la mise en place de projets de recherche interdisciplinaires. Mais au-delà des questions de communication, pour des projets interdisciplinaires il est aussi du rôle du chercheur de pouvoir devenir un interlocuteur de la société et des politiques en tant qu'expert pouvant permettre le transfert de connaissance et l'accompagnement de l'aide à la décision en s'appuyant sur ses connaissances.

C'est cette problématique que nous allons aborder ci-après.

B. Le chercheur dans une situation d'expert et de didacticien des politiques : Conséquences de la guerre du Golfe (1990 – 1991) sur la santé des militaires français

De nombreuses études statistiques sont réalisées, que ce soit pour mesurer la dangerosité de certains traitements ou de certaines expositions, ou pour prédire le futur président de la république, mais aussi pour estimer la taille de la population française dans 100 ans. Certains résultats surprennent, d'autres inquiètent, mais souvent il est bien difficile pour une personne non spécialiste de comprendre réellement la situation. Il est donc de la responsabilité des statisticiens, et de l'ensemble des scientifiques utilisant une démarche statistique, de décrire avec précision et d'explicitier leur démarche pour permettre une compréhension de l'information transmise. Ils peuvent donc être interpellés afin de préciser leur démarche, ou être sollicités en tant qu'experts pour aider à la compréhension d'une multitude d'informations qui peuvent sembler contradictoires.

Ainsi sollicité, le chercheur devra expliciter sa démarche ou celle des autres chercheurs en jeu, pour permettre de préciser le niveau d'incertitude dans les résultats diffusés. L'exemple classique de la météo est particulièrement marquant ; à la fin des informations données dans le cadre de la météo prévisionnelle, tout téléspectateur n'en retient qu'une pour le lendemain, et sera scandalisé si, le lendemain matin, au lieu des 18° attendu il ne lit que 10° sur son thermomètre. Les calculs prévisionnels réalisés sont d'une extrême complexité, mais comment réussir à faire évoluer les mentalités afin de faire accepter que la météorologie ne correspond pas à une science déterministe, que le battement d'aile de papillon peut faire changer le cours des choses et que dans le cadre de modèles aussi sophistiqués qu'ils soient le chercheur ne pourra pas contrôler l'ensemble ?

Il est difficile de comprendre sans aucun élément d'explication les raisons pour lesquelles il peut y avoir tant de différences de conclusion entre différents auteurs. Le chercheur se doit d'expliquer les notions de variabilité, d'incertitude contrôlée, afin que les résultats issus de recherche développée dans le cadre d'une démarche statistique puissent être écoutés et diffusés. Le manque de précision,

l'impression de résultats issus de boîtes noires, voire de boîtes truquées révélées induit automatiquement une perte de crédibilité dans les résultats diffusés dans ce cadre ; mais, plus largement les retombées pourront entacher la crédibilité de l'usage de la statistique. Le chercheur peut donc être invité à aider des politiques ou des personnes de la société à comprendre ce qui se passe. C'est dans un de ces cadres que s'est déroulée la situation présentée ci-après.

Elle concerne une demande faite à un professeur d'université – praticien hospitalier (PU-PH) de l'université de Bordeaux, pour une expertise de situation d'un problème très médiatisé, pouvant être considéré selon les angles comme un problème de société, un problème sanitaire, voire un problème politique : l'état de santé des militaires ayant été présents sur le théâtre des opérations en 1990 – 1991 lors de la guerre du Golfe.

Dans ce travail, nous nous intéressons tout particulièrement aux auditions du chercheur réalisées dans le cadre de la mission parlementaire, lieu et temps d'échanges intéressants entre les politiques et le chercheur. Quels sont les enjeux didactiques pour le chercheur s'exprimant en direction de politiques ? Dans un second temps nous présenterons l'article scientifique qui a été publié pour une diffusion scientifique et qui permettra de souligner les différentes approches de transposition dépendantes des sphères.

1. Contexte

En 2000, Avigolfe, association des victimes militaires de la guerre du Golfe, réussit à convaincre l'état français de s'intéresser à la question soulevée par les américains, depuis plusieurs années, concernant l'existence ou non d'un syndrome de la guerre du Golfe (1990 – 1991).

En octobre 2000, deux missions ont été constituées de manière parallèle, l'une diligentée par l'assemblée nationale constituée de politiques, et l'autre par l'exécutif, constituée de chercheurs. Le groupe d'expert était présidé par un chercheur de l'INSERM missionné par le Ministre de la Défense et le Secrétaire

d'Etat à la Santé et aux Handicapés pour présider « *un groupe de d'experts³⁹ chargé de faire des propositions au Gouvernement sur les modalités permettant d'améliorer les connaissances sur les conséquences de l'engagement dans le conflit du Golfe de 25 000 soldats français* ». Le rapport a ainsi été remis le 27 avril 2001, et dans un second temps et suite aux recommandations de ce rapport, le ministère a sollicité l'INSERM pour conduire une étude d'investigation de l'état de santé des militaires de la guerre du Golfe 1990 – 1991.

2. Auditions dans le cadre des missions concernant les conséquences de la guerre du Golfe 1990 - 1991

Dans ce cadre, le chercheur a été auditionné deux fois par la commission de la défense nationale et des forces armées dans le cadre de la mission d'information sur « les conditions d'engagement des militaires français ayant pu les exposer au cours de la guerre du Golfe et des opérations conduites ultérieurement dans les Balkans à des risques sanitaires spécifiques » :

- Audition le 8 novembre 2000, suite à la lettre de mission du 3 octobre 2000⁴⁰,
- Audition le 21 mars 2001, qui précède la remise du rapport « Rapport de mission du Groupe de travail chargé d'analyser les données sanitaires relatives aux anciens combattants français de la guerre du Golfe » du 27 avril 2001⁴¹.

L'ensemble des éléments de ces deux auditions ont été repris dans le Rapport d'information par la Commission de la défense nationale et des forces armées en conclusion des travaux d'une mission d'information sur les conditions d'engagement des militaires français ayant pu les exposer, au cours de la guerre du Golfe et des opérations conduites ultérieurement dans les Balkans, à des risques sanitaires spécifiques (Cazeneuve, Rivasi, & Lanfranca, 2001).

³⁹ Composition du Groupe d'experts indépendants : Roger Salamon, Annick Alperovitch, Françoise Conso, Marthe-Aline Jutand, Jean-Paul Boutin, Yves Coquin, Roland Laroche, Christophe Paquet, Frédéric Rouillon, Pierre Weinbreck

⁴⁰ <http://www.assemblee-nationale.fr/rap-info/i3055-04.asp>

⁴¹ Accessible url - <http://www.assemblee-nationale.fr/rap-info/i3055-06.asp>

L'objectif de la première audition⁴² était de définir les fonctions respectives des deux missions, et de préparer un communiqué commun précisant leur articulation. Le poids médiatique du syndrome de la guerre du Golfe était tel que l'inquiétude était forte pour les deux missions de ne pas être jugées objectives dans leur approche. Dès le début de l'audition, le président de la Mission d'Information Parlementaire indique que «*cette séance est une réunion de travail entre vous et nous, plutôt qu'une audition à proprement parler*». Cette audition fait partie des 10 auditions qui se sont déroulées d'octobre 2000 à décembre 2000, les autres auditions concernant plus spécifiquement des acteurs de la guerre du Golfe 1990 – 1991, et ayant un objectif de description concernant ce qui s'était réellement déroulé durant cette période. L'audition du président du groupe d'experts correspondait plus à un temps de discussion, d'échanges entre le groupe d'élus dans son ensemble et le responsable scientifique du groupe d'experts, seul, dans le but de communiquer aux journalistes et d'aider à clarifier les rôles de chacun (figure 14). Il s'agit d'un temps d'acculturation durant lequel chacune des parties apprend à comprendre l'autre à évaluer les différences et les complémentarités, se regarde et s'étudie.

« Nous pourrions sortir de cette réunion avec un communiqué commun de nature à lever les ambiguïtés qui pourraient persister dans l'esprit de certains concernant le champ d'investigation respectif de chacune des missions et la manière dont ces dernières pourraient travailler ensemble. » (audition du 8 novembre 2000)



Figure 14 : circuit des échanges et de la communication lors de la mise en place des missions

⁴² Présentée en début d'audition par le président de mission comme plus exactement une réunion de travail.

La seconde audition a été réalisée dans un contexte complètement différent puisqu'elle précédait d'un mois la remise du rapport du groupe d'experts aux ministres. Le chercheur était donc dans une situation plus délicate, ne pouvant pas révéler le contenu avant d'avoir remis le rapport aux ministres.

a) *La mission du groupe d'experts*

Comme l'indique la lettre de mission adressée au chercheur-président de la mission, il s'agit de faire un travail de synthèse des connaissances publiées concernant l'existence ou non d'un syndrome de la guerre du Golfe et de faire des propositions.

« Nous vous remercions d'avoir accepté de présider un groupe de travail chargé de faire des propositions au Gouvernement sur les modalités permettant d'améliorer les connaissances sur les conséquences de l'engagement dans le conflit du Golfe de 25 000 soldats français.

Nous vous demandons d'axer vos travaux de manière à pouvoir fournir une réponse aux deux questions suivantes :

- ▀ Que faut-il penser, à la lumière de la littérature internationale, de l'existence d'un syndrome spécifique lié à la guerre du Golfe et quelle en serait l'étiologie ?
- ▀ Quelles études convient-il de mener en France, compte tenu en particulier des données que vous jugeriez établies à partir de l'analyse de la littérature et des dossiers de malades anciens combattants de la guerre du Golfe ?

Le groupe que vous présiderez devra avoir une composition pluridisciplinaire et comporter des spécialistes à la fois civils et militaires.

La Direction centrale du Service de santé des Armées, la Direction générale de la Santé, l'Institut de veille sanitaire proposeront chacun un représentant pour participer aux travaux du groupe. Nous souhaitons disposer de vos conclusions dans un délai de six mois. »

(Lettre adressée au chercheur par Mme Gillot, Secrétaire d'état à la santé et aux handicapés et de M. Richard, Ministre de la Défense – 3 octobre 2000)

Le travail demandé consiste, en partie, en une synthèse de la littérature sur l'existence ou non du syndrome de la guerre du Golfe. Cette littérature est en grande partie composée d'articles de revues en épidémiologie, décrivant les résultats d'analyses statistiques permettant d'infirmer ou d'affirmer l'existence du syndrome de la guerre du Golfe. Mais comment définir la notion de syndrome ? Il s'agit d'une notion complexe qui peut couvrir de multiples situations, et le repérage d'un syndrome peut être issu de différentes approches : une approche médicale de par la

connaissance et l'expertise de médecins ayant reconnu des associations de symptômes pouvant être associés à un facteur causal, mais aussi une approche statistique par la recherche d'association entre symptômes. L'objet principal d'étude pour les deux missions se situe dans la validation ou non de l'existence du syndrome de la guerre du Golfe, les enjeux étant sanitaires mais aussi financiers de par la possible reconnaissance de maladie professionnelle. Il est donc important d'assurer la qualité nécessaire de ce travail et surtout de situer l'argumentation de l'existence ou non sur un niveau scientifique. Ce point sera un leitmotiv pour le président de la mission parlementaire. D'autres éléments importants de la démarche statistique seront étudiés dans ce cadre.

b) Le scientifique : caution de la qualité du discours et des décisions

Plusieurs éléments de cet échange soulignent combien les politiques estiment l'importance et le rôle de validation des chercheurs. Dès le début de l'audition, ils sont en demande d'une validation par l'expert de leur proposition de démarche scientifique. Ainsi, durant cet échange, il y a d'un côté le sachant/expert et, de l'autre, les politiques en situation de demande de la vérification de la pertinence de leur raisonnement. L'expert est attendu comme la personne caution d'un bon raisonnement.

Le politique a ainsi une inquiétude concernant la qualité scientifique du travail de sa mission, et comme l'enjeu sociétal est important il a le souci de placer leur réflexion dans un cadre de qualité scientifique irréprochable afin que les conclusions des missions ne puissent pas être remises en question. Il souligne dans le texte que « *le premier principe est donc celui de la rigueur scientifique* ».

Dans l'extrait ci-dessous, il est remarquable de voir le politique soulignant son inquiétude au début de l'extrait par « *je ne pense pas que, scientifiquement, la mission d'informations parlementaire soit en mesure de conclure à l'existence ou non d'un syndrome* ». Il souligne ainsi la nécessité d'un travail collaboratif en

déplaçant la partie de la validation scientifique dans l'espace des experts et de la recherche.

Président de la mission parlementaire : En outre, il est un point sur lequel je voudrais beaucoup insister [...] : je ne pense pas que, scientifiquement, la mission d'information parlementaire soit en mesure, notamment compte tenu des délais qui lui sont impartis et des matières qu'elle doit aborder, de conclure à l'existence, ou non, d'un syndrome.

La mission d'information parlementaire pourra définir très précisément les conditions dans lesquelles les militaires français se sont trouvés exposés à des risques, mais c'est au terme d'études épidémiologiques qu'il sera éventuellement possible de conclure sur la question d'un syndrome.

Je voudrais que cette position soit ratifiée par les membres de la mission et que l'expert scientifique que vous êtes, Professeur, donne son avis sur la validité d'un tel raisonnement.

[...]

De la préparation de cette audition avec le Professeur ____, et d'autres conversations que j'ai pu avoir avec des épidémiologistes, il ressort – [...] - que seules des études épidémiologiques approfondies permettraient de parvenir à une vérité scientifique incontestable.

Je demande que nous soyons rigoureux et que, par conséquent, nous aboutissions à des résultats scientifiquement validés. Je ne souhaite pas - [...] - que nous nous comportions comme des parlementaires qui, pour dire ce qu'une certaine presse aurait envie d'entendre, affirmeraient de façon définitive des choses qui, scientifiquement, n'auraient pas été prouvées. Je pense que ce ne serait, ni sérieux, ni crédible, ni convenable. (Audition du 8 novembre 2000)

Le cadre est ainsi clairement présenté par le président de la mission parlementaire, positionnant les deux missions dans leurs rôles respectifs, mais soulignant l'importance d'une compréhension partagée des concepts et son souhait d'un usage précis et partagé des mots. Il a ainsi le souci que le discours des deux missions soit basé sur des terminologies cohérentes afin de contrôler autant que faire se peut la perte de signifié par la médiatisation.

Président de la mission parlementaire - « Je souhaite que nous soyons attentifs aux concepts, que nous utilisons leur épaisseur sémantique et leur exacte signification. Je comprends bien que les mots ont une signification et une perception sociales. Par la médiatisation, ils perdent de leur puissance ou de leur acception scientifique. [...] Je voudrais donc, M le Professeur, que l'exigence scientifique qui est la vôtre dans l'approche du dossier, soit partagée par notre mission d'information et que nous prenions acte du principe qu'il nous revient de dire des choses précises qui soient concordantes avec celles que vous pourrez affirmer au terme de vos travaux. » (Audition du 8 novembre 2000)

Cette audition est donc majoritairement dirigée dans un but de structure de la démarche conjointe et non pour un échange d'informations de données ou de réponses. Cependant, certains élus ont une attente plus tournée sur le souhait d'obtenir une réponse en tout ou rien

Membre 1 de la mission parlementaire - « Les experts du groupe que vous présidez et qui, à l'inverse de nous, sont des spécialistes, doivent quand même être en mesure de se forger une opinion, sinon une intime conviction, sur l'existence ou non du syndrome dit « du Golfe » et éventuellement de faire des propositions que ce soit en matière d'indemnisation ou autre. » (audition du 8 novembre 2000)

Pour cet élu, la recherche bibliographique aussi intéressante qu'elle soit ne peut pas être le seul résultat de cette mission et, dans sa remarque, il est mis en évidence qu'il attend des réponses venant des scientifiques qui, selon lui, ont des opinions, voire des convictions. Il ne place donc pas le scientifique en tant que garant d'une démarche scientifique, mais plus comme le puits de savoir qui peut donner une réponse rapidement car, implicitement, il a déroulé le raisonnement qu'il faut. Il sous-entend que le politique et la société peuvent faire confiance les yeux fermés aux chercheurs experts, et que la question posée attend une réponse quelle qu'elle soit, pourvu qu'elle vienne d'un chercheur. Dans le cas présent la question formulée est « *peut-on parler d'un syndrome de la guerre du Golfe ?* ».

Lors de la seconde audition, les membres de la commission ont été assez clairs sur l'impact économique potentiel des décisions qui devront être prises à l'issue de ce travail. La question de l'existence ou non d'un syndrome de la guerre du Golfe étant associée à la question d'indemnisation des militaires si cette pathologie existe et est reconnue comme maladie professionnelle.

Membre 2 de la mission parlementaire - Pensez-vous que l'on pourra réaliser une typologie, avec un certain nombre de symptômes, qui pourrait être prise en compte pour indemniser les militaires, comme cela a été fait aux Etats-Unis ? En effet, nous avons un problème avec le code des pensions militaires d'invalidité qui ne prend pas en compte ce type de maladies à long terme : de nombreux soldats ont une grande difficulté à faire admettre qu'il existe un lien de causalité entre leur maladie et le fait d'avoir participé à la guerre du Golfe. (Audition du 21 mars 2001)

c) *Comment définir le syndrome de la guerre du Golfe ?*

Au cours de l'audition le président du groupe d'experts indique en répondant au membre 2 de la mission, sa préférence pour parler de *signes* et non de *syndrome* afin d'éviter les sur-représentations de la question des syndromes.

Président du groupe d'experts - « Vous avez raison de parler de symptôme car il est probablement préférable de parler de signes que d'un syndrome, ce dernier terme, malgré l'attachement que lui portent les médias, risquant de nous compliquer singulièrement la tâche » (audition du 8 novembre 2000)

Cependant, il souligne que la question qui lui a été posée concerne le syndrome de la guerre du Golfe, et que cette notion devra être définie afin de pouvoir répondre à son existence ou non. Mais un point important est souligné : il s'agit de la difficulté de renier l'existence de quelque chose qui a été nommé. La difficulté pour les scientifiques et les politiques dans le cadre de ces missions c'est de devoir répondre à la question de l'existence ou non d'un objet qui a une existence inhérente à l'existence d'un mot le représentant.

Président du groupe d'experts - Or, aujourd'hui, on est bien obligé d'imaginer que la réponse à cette question (*ndlr* : « y va-t-il un syndrome de la guerre du golfe ») est plutôt négative. Si c'est précisément le cas, il ne faut pas négliger que l'enjeu est une réalité sociale. Il faut bien définir la notion de syndrome pour dire que cette réponse négative ne signifie pas que les plaintes sont infondées. En d'autres termes, nous ne devons pas utiliser cette réalité sociale de manière quasiment malhonnête en disant : « puisque vous baptisez ainsi ce phénomène notre réponse est non ! ». En outre, cette représentation sociale a quand même, pour des gens comme nous, une connotation dangereuse parce qu'elle signifie que le diagnostic est porté. Les gens qui parlent du « syndrome de la guerre du Golfe » ont déjà, dans leur esprit, apporté la réponse, ce qui est humainement compréhensible mais scientifiquement inacceptable ! (audition du 8 novembre 2000)

Le membre 2 de la mission souligne le fait que le mot syndrome a perdu le sens scientifique initial en étant transporté dans l'environnement médiatique et social. Il a pris un autre sens qui n'est cependant pas défini. Il est ainsi signalé que derrière ce terme, chacun peut y mettre un sens différent et aussi une existence différente.

Membre 2 de la mission parlementaire - Concernant le problème du « syndrome », le terme nous a tous dépassés dans la mesure où il est devenu médiatique et a quelque peu perdu sa signification scientifique. En tant que

scientifique, vous savez comme moi que les termes évoluent en fonction de la société et que, même quand ils ont une signification précise, s'ils en acquièrent une autre dans les représentations sociales, nous sommes obligés de suivre ces dernières. C'est l'évolution de la langue qui le veut ! C'est vrai qu'à défaut d'être scientifiquement correct le terme devient socialement admis quand les gens se l'approprient. (Audition du 8 novembre 2000)

Ainsi, le président et le groupe d'experts ont défini la notion de syndrome dans le rapport remis à l'issue de cette mission de la manière suivante :

« A la maladie, état morbide souvent bien délimité et dont on a identifié le facteur causal et/ou le(s) facteur(s) de risque, les cliniciens opposent le syndrome, assemblage ou construction de symptômes et de signes le plus souvent hétérogènes, dont la constatation reconnue par le clinicien sur un certain nombre de cas conduit celui-ci à considérer que cet assemblage constitue une entité à laquelle doit être associée une cause. La notion de syndrome, pour habituelle qu'elle soit dans le langage médical recouvre finalement de multiples situations et n'est pas si simple à définir. » (Salamon, 2001) (p10).

A côté d'une approche d'expertise médicale pour la reconnaissance d'existence d'un syndrome, une méthode statistique peut être utilisée pour étudier l'existence ou non d'associations fortes entre symptômes non associés à une maladie connue. Les approches statistiques, qui ont été utilisées dès le début des études américaines et hollandaises pour rechercher un syndrome de la guerre du Golfe, ont été des analyses factorielles, principalement des analyses de correspondances multiples, ou des analyses en composantes principales. Une explication de ce type de méthode a été intégrée au rapport de mission, sans que les auteurs développent l'écriture scientifique, mais essaient d'en simplifier au maximum l'explication.

Peut-être existe-t-il une spécificité dans la « combinaison » des signes présentés par les vétérans qui serait la possible expression d'un syndrome. La recherche d'une telle spécificité combinatoire doit reposer sur des méthodes statistiques multivariées de type « analyse factorielle » qui sont des méthodes classiques bien qu'assez sophistiquées. L'analyse factorielle est une technique appropriée pour chercher un nombre minimum de « facteurs » pouvant représenter une part importante de l'information contenue dans un ensemble de données mesurées de manière quantitative ou binaire. Il s'agit de réduire un nombre important d'informations à quelques grandes dimensions. Les informations recueillies pour un individu sont résumées en un nombre limité de « facteurs » construits comme combinaisons linéaires des variables initiales et traduisant leurs associations. On utilise le terme de variables latentes pour parler de ces facteurs, qui existent au plan conceptuel seul et qui ne sont pas mesurées directement. (p16)

Lors de l'audition, le président du groupe d'experts n'avait cité qu'une seule fois cette méthode statistique, en donnant quelques éléments d'explication, qui dans ce contexte n'ont pas été repris par les élus.

Président du groupe d'experts - Les scientifiques britanniques, par exemple, qui ont fait un travail remarquable en comparant la situation de soldats ayant fait la guerre du Golfe et ceux ayant participé aux événements en Bosnie-Herzégovine [...] ont pu démontrer qu'ils présentaient à peu près les mêmes types de troubles mais que ces derniers étaient quantitativement plus importants chez les soldats du premier groupe. Pour y parvenir ils se sont livrés à des factor analysis, dites en français « analyses factorielles pour établir des regroupements. » (Audition du 8 novembre 2000)

Ces analyses statistiques sont sophistiquées, et l'explication de cette méthode d'analyse à l'oral dans le cadre d'une audition est tout à fait irréalisable. Il faut d'ailleurs souligner que ce point n'est pas non plus relevé dans le cadre de l'audition du 21 mars 2001.

Lors de la seconde audition du 21 mars 2001, le président du groupe d'experts signale que l'analyse de la littérature conduit à un rejet de la notion de syndrome.

Deuxièmement, de manière assez nette, sans toutefois recueillir l'unanimité, la notion de « syndrome » est rejetée. Ce n'est pas le fond du problème que vous vous posez. Cette question est intéressante sur un plan médico-légal, médiatique ou politique, mais elle ne touche pas à l'essentiel. Sur cinq auteurs très connus, les quatre plus sérieux rejettent cette appellation de « syndrome », et un, moins sérieux - il n'a pas de groupe témoin - lui accorde du crédit. Il n'y a donc pas unanimité, mais la grande majorité des auteurs de l'*establishment* scientifique américain dénie la notion de syndrome, avec une cause unique. J'ai donc écrit dans mon rapport que cette question n'était pas d'un intérêt majeur. (Audition du 21 mars 2001)

Il s'agit donc d'une décision prise suite à une synthèse de l'information existante dans un autre contexte et à partir de publications étrangères. Les défauts de publications sont connus, et le groupe d'experts ne pouvait déterminer un résultat à partir d'une synthèse de la littérature presque unanime.

d) Argumenter des propositions de stratégies d'enquête

Au-delà de la question même de l'existence ou non du syndrome de la guerre du Golfe, un point important a été soulevé : il s'agit du cadre de la réalisation d'enquêtes, et de la définition des personnes à interroger. Dans le cadre de cette audition des interrogations ont été soulevées concernant la mise en œuvre possible, suite à cette première étape, d'études épidémiologiques.

Membre 3 de la mission d'information parlementaire - « ceci étant, pour mener ces études, encore faut-il travailler sur l'ensemble de la population des soldats Français exposés soit environ 9000 personnes, ce qui constitue un chiffre extrêmement important et sous quelle forme pourrait-on obtenir ces réponses ? », (Audition du 8 novembre 2000)

Président du groupe d'experts - « l'épidémiologiste peut parfaitement vous répondre qu'il faut, au lieu de travailler sur 25000 personnes, travailler sur un échantillon représentatif de 2000 [...] Il vaut mieux avoir 2000 personnes tirées au hasard mais en représentant 20000 que 17000 parce que dans ce cas, il en manquera 3000 pour que l'étude soit complète [...] j'aurai tendance à dire en tant que chercheur, je me contenterai d'un échantillon plus petit qui me permettrait d'aller plus vite mais j'estime [...] que nous ne pouvons pas nous permettre de ne faire qu'un travail épidémiologique en laissant se débrouiller seuls ceux qui n'auront pas été tirés au sort ». (Audition du 8 novembre 2000)

Le président du groupe d'experts souligne ainsi qu'une approche décontextualisée de la question pourrait conduire à des éléments de réponse qui, même si leur recevabilité était entière de par la qualité de la démarche scientifique, ne seraient pas acceptables face à la question sociétale. Cette remarque est ainsi reprise par le membre 2 de la mission d'information.

Membre 2 de la mission d'information parlementaire - L'attente de l'opinion obéit à une approche psychologique très forte. Si les épidémiologistes nous disent qu'ils prennent un échantillon de 1000 ou de 2000 personnes à partir duquel ils vont tirer des conclusions scientifiques, beaucoup vont se sentir frustrés de n'avoir pas été sélectionnés pour faire partie de l'échantillon. » (Audition du 8 novembre 2000)

Lors de la seconde audition, le président du groupe d'experts a exposé une proposition quant aux études pouvant être demandées, en explicitant les raisons scientifiques et sociales.

Président du groupe d'experts - Nous allons proposer de réaliser une analyse exhaustive de la situation des 25 000 vétérans. On ne peut pas en effet se payer

le luxe, dans une situation d'inquiétude, de ne pas offrir à tous les anciens combattants de la guerre du Golfe la possibilité de se faire examiner - dans un hôpital civil ou militaire - avec un protocole qui sera standardisé par un comité scientifique. Une analyse statistique sera ensuite effectuée.

Contrairement aux Américains et aux Anglais, on ne procédera à aucune comparaison de militaires français déployés dans le Golfe, avec des militaires non déployés ou déployés en Bosnie-Herzégovine. En revanche, nous formulerons une autre proposition : parmi les 25 000 militaires, nous trouverons des « cas » - qu'il appartiendra de définir - et des « non-cas » que nous nous efforcerons de comparer par rapport aux expositions qu'ils ont subies. Or, je compte sur vous pour obtenir ce type d'information. De la sorte, la France pourrait réaliser une étude-cas témoin unique en son genre. Si parmi les « cas », on découvre davantage de militaires exposés à des risques que dans le groupe témoin, cela aura une valeur scientifique à même de valider ce que vous pouvez pressentir actuellement. (audition du 21 mars 2001)

Le terme « valeur scientifique » est souligné par le chercheur dans le cadre d'une approche de validation d'une hypothèse.

e) Argumenter les choix méthodologique pour une aide à la décision

Les éléments concernant plus spécifiquement les étapes d'analyse statistique et d'aide à la décision ont été discutés par le président du groupe d'experts plus spécifiquement lors de la seconde audition. Il s'agissait de proposer une démarche hypothético-déductive pour valider une hypothèse par une approche scientifique dans le cadre d'une démarche structurée. L'approche proposée correspond à la stratégie des tests d'hypothèse, comme le souligne le président du comité d'experts lors de la seconde audition en indiquant que le non rejet de l'hypothèse n'implique pas de fait qu'il n'y a pas de danger.

Président du groupe d'experts - Et je voudrais ajouter une précision importante : si l'on ne trouve aucun lien entre ce dont souffrent les militaires français et l'uranium appauvri, cela ne voudra pas dire que l'utilisation de ce dernier est sans danger. De la même façon, ce n'est pas parce que l'on aura prouvé que l'uranium appauvri peut être dangereux que les militaires qui ont participé à la guerre du Golfe et ont été en contact avec des armes en incorporant seront obligatoirement malades. (Audition du 21 mars 2001)

Le président du groupe d'experts souligne l'importance de la compréhension que les résultats de telles démarches ne doivent pas se lire en tout ou

rien, mais induisent de fait de l'incertitude, et qu'il est important d'en acquérir la subtilité. Les associations repérées ne traduisent pas un système déterministe indiquant que si il y a A alors il y aura B. L'association mesure une tendance sur la population ou le groupe, et le transfert sur l'individu ne peut être pensé qu'en termes de probabilité. La difficulté de la situation, fréquente pour les épidémiologistes, est de devoir étudier un phénomène sur une population quand la question de la société est la prise en charge individuelle.

Comme cela a été dit précédemment, la problématique sous-jacente à la mise en place de cette réflexion globale autour du syndrome de la guerre du Golfe était la question de l'indemnisation. Comme le souligne le président du groupe d'experts, les scientifiques peuvent apporter des éléments d'aide à la décision mais n'ont pas à prendre le rôle de décideur. Les éléments statistiques qui seront proposés comme éléments d'aide à la décision sont, comme tout résultat statistique, dans une situation de vraisemblance, associés de fait à des probabilités. Le pas nécessaire à réaliser par les politiques est de prendre une décision à partir de résultats statistiques ne pouvant pas être considérés comme éléments de vérité, mais des informations vraisemblables, l'incertitude du résultat étant l'élément indissociable de la statistique.

Président du groupe d'experts - Je vais vous décevoir, car je ne me sens pas capable, aujourd'hui, de répondre à votre première question concernant l'indemnisation des militaires à partir d'une typologie. Je ne sais d'ailleurs pas si c'est notre rôle, si un jour on en sera capable, et je ne sais pas comment on y répondra. En tout état de cause, le choix sera arbitraire : il vous appartiendra de le faire, à vous les politiques, les juristes, les militaires.

Il conviendra tout d'abord d'établir un lien de causalité : quelle est la probabilité de pouvoir imputer tel symptôme à tel événement ? Ensuite, en fonction d'une probabilité, il faudra décider si l'on est au-dessus ou non d'un seuil. Pour créer un modèle permettant de déterminer la probabilité que tel événement, dix ans après, ait un lien avec tel facteur que l'on connaît mal, il me faudra des années de travail et autant d'argent que les Américains ont donné à leurs chercheurs. A titre illustratif, ils ont établi un modèle extraordinaire ; un modèle de l'évolution atmosphérique des poussières de sable, avec je ne sais plus quoi dedans, peut-être de l'uranium appauvri ! Ils ont ainsi pu mesurer, avec la climatologie de l'époque, jour par jour, l'évolution du sable. Ce sont des modèles mathématiques incroyablement compliqués !

En ce qui nous concerne, il existe deux possibilités. Soit on essaie d'établir ce type de modèles, et je pense que l'on n'y arrivera pas ; soit on détermine les événements qui paraissent avoir joué, et ensuite ce sera un choix arbitraire qui ne pourra venir que de vous. Là, le scientifique ne peut rien faire. (Audition du 21 mars 2001)

Le rapport de la mission d'information qui a été diffusé le 15 mai 2001 reprenait en annexe la totalité des échanges des auditions, et soulignait dans le corps du texte l'intérêt méthodologique des études cas-témoin tout particulièrement.

A l'issue de la remise du rapport d'information de la mission parlementaire et du rapport du groupe d'experts, « l'enquête française sur la guerre du Golfe et ses conséquences sur la santé » a été lancée suite à la lettre de mission du 6 juin 2001 adressée par le cabinet du Ministre du Ministère de la Défense. L'objectif principal de cette étude était de décrire chez les militaires et civils français ayant participé aux opérations extérieures dans le Golfe Persique d'août 1990 à juillet 1991 : 1 - les symptômes et les pathologies présentés après le conflit, 2 - leur descendance après le conflit.

Il s'agissait d'une étude transversale à visée exhaustive nationale, réalisée sur la période d'octobre 2001 à juin 2004. Le comité de coordination était composé de cinq personnes : l'investigateur principal, le chef de projet médecin épidémiologiste, le statisticien (nous-même), le gestionnaire de projet et l'informaticien. Le rapport a été remis au ministre le 13 juillet 2004 (Salamon & al, 2004) contenant l'analyse descriptive ainsi que des études et analyses complémentaires. Les résultats concernant la question du syndrome de la guerre du Golfe étaient rédigés en fin de document et sont reportés ci-dessous.

Ces associations retrouvées pouvaient donc se résumer à trois groupes :

- un regroupement de signes respiratoires (toux, mal de gorge, respiration sifflante),
- un regroupement de signes neuropsychologiques (problèmes pour trouver les mots, troubles mnésiques, désorientation ...),
- un regroupement de signes psychologiques (dépression, panique ou angoisse).

Ces regroupements de symptômes observés à partir des données recueillies sont cohérents avec des associations de symptomatologie connues et ne semblent pas mettre en évidence de particularité, cependant des analyses complémentaires pourraient être effectuées afin de préciser cela.

Suivant différentes études américaines, les symptômes les plus fréquemment constatés chez les vétérans américains souffrant de pathologies et engagés durant le conflit du Golfe sont (par ordre décroissant) : la fatigue, les maux de tête, les éruptions cutanées, les douleurs musculo-articulaires, les pertes de mémoire, les troubles du sommeil, les douleurs thoraciques, ainsi que la toux, les diarrhées et autres symptômes gastro-intestinaux. Ces symptômes ont donné son nom de baptême à ce que les médecins américains ont appelé le « syndrome de la guerre du Golfe ». Cette notion a été largement reprise par les médias. Médicalement, un syndrome est une association de signes ou de symptômes

dont la constatation sur un certain nombre de cas suggère qu'elle n'est pas fortuite et qu'elle peut être liée à une cause spécifique. Les syndromes peuvent avoir un point d'ancrage anatomique ou psychologique. Les résultats précédents de l'enquête n'ont pas démontré l'existence d'un syndrome spécifique du Golfe, ce qui est concordant avec les résultats de la plupart des études étrangères. Il faut rapporter néanmoins que seul le quart des sujets concernés ont participé à l'enquête et que cette analyse ne porte évidemment que sur les sujets ayant participé aux opérations dans le Golfe Persique contactés et qui ont répondu au questionnaire.

Des publications ont été diffusées dans des revues internationales présentant ainsi la situation française. L'enquête menée devait ainsi satisfaire la demande institutionnelle qui avait été émise dans la lettre de mission, mais aussi présenter la situation française au sein des communautés de chercheurs travaillant sur cette thématique depuis déjà de nombreuses années pour certains d'entre eux. Il s'agissait, pour les enseignants-chercheurs et chercheurs ayant investi du temps dans cette étude, de pouvoir aussi faire reconnaître leur travail dans leur communauté de recherche, la production de savoir étant une de leur mission et la diffusion de ce savoir le moyen de mesurer leur activité.

Le premier article que nous avons publié à l'issue de la remise du rapport, en date de janvier 2006, était une description précise de l'organisation de l'étude, présentant le mode de recueil et le recrutement des 5 666 participants de l'étude, puis une description statistique selon leur armée de rattachement et les sites fréquentés pendant les événements (Irak, Koweït, Arabie Saoudite), ainsi que les expositions qu'ils ont pu subir (alerte chimiques et biologiques, feu de puits de pétroles, tempêtes de sable, usage d'insecticides). Leur état de santé et les symptômes ressentis avaient été recueillis et ont été décrits, puis une analyse factorielle a été réalisée pour étudier la question du syndrome de la guerre du Golfe. Les associations retrouvées étaient des associations classiques ne pouvant déterminer une nouvelle pathologie.

3. Article "Health consequences of the first Persian Gulf War on French troops"

Health consequences of the first Persian Gulf War on French troops

R Salamon,^{1*} C Verret,¹ M A Jutand,¹ M Bégassat,¹ F Laoudj,¹ F Conso² and P Brochard³

Accepted 20 December 2005

Summary Since 1993, many studies on the health of Persian Gulf War Veterans (PGWV) have been undertaken. These studies have concluded that there has been an increased mortality due to external causes, no excess of recognized diseases, and no effect on PGWV children. When compared with the non-deployed, PGWV have reported a higher frequency of infertility as well as different symptoms, but a specific Gulf War syndrome was not identified. In October 2000, the French government asked an independent working group to analyse the scientific literature on PGWV health. The group concluded that an exhaustive study of French PGWV was to be undertaken. The objectives of this study were to describe the exposures of PGWV in the operations theatre, to report on the symptoms and diseases that occurred in PGWV and their children during and after the military campaign, and to explore the possibility of a Gulf War syndrome. This exhaustive cross-sectional study, which included all civilians and troops who served in the Gulf from August 1990 to July 1991, began in January 2002. Data were collected by postal self-administered questionnaires. A standardized clinical evaluation was performed by 27 clinics of occupational diseases and nine military hospitals. Symptoms and diseases which appeared after the campaign are described.

To date, among 20 261 PGWV, 5666 participated in the study (28%). The most frequent symptoms described since the return from the Gulf were headaches (83%), neurological or psychological symptoms, and back pain. Apart from well-known symptoms associations (respiratory, neurocognitive, psychological and musculo-skeletal syndromes), no other cluster was highlighted by our analysis.

Keywords Gulf War, cross-sectional study, epidemiology, Gulf War Veterans, Gulf War syndrome

Background

From August 1990 to April 1991, the conflict with Iraq involved troops of a coalition constituted of almost 40 countries, mainly the United States (700 000 personnel), the United Kingdom (50 000) and France (20 000). The French troops were deployed nearby the US and UK troops and faced the same exposures.¹ However, French participants, as opposed to US and UK troops, did not receive vaccines against anthrax, botulism, or

plague, which have been associated with the self-reporting of post-deployment symptoms.^{1,2}

Since 1993, many studies on the health of Persian Gulf War Veterans (PGWV) have been undertaken, mainly in the US and the UK. US and UK mortality studies showed a lower mortality from all illnesses among Gulf War Veterans in comparison with non-deployed veterans, and an increased mortality due to external causes. This is consistent with the patterns of post-war mortality already observed in veterans of previous wars.^{3–6} Some authors found an association between Gulf War service and reported infertility,^{7,8} or miscarriage,⁹ but no effect was described on PGWV's children.^{10–13}

Compared with non-deployed veterans, PGWV have been reported to have no excess of recognized diseases,^{6,14–37} but were two to three times more likely to report fatigue, cognitive difficulties, headaches, myalgia and arthralgia, mood disturbance, and sleep problems.^{6,14–37}

Several studies, based on statistical factor analysis have been carried out to determine relationships among symptoms.^{38–50} In

¹ Department of Epidemiology, Public Health and Development (INSERM U593), Victor Segalen University Bordeaux 2, Bordeaux, France.

² Clinic of Occupational Diseases, Cochin Hospital, AP-HP, Paris 5 University, Paris, France.

³ Laboratory of Occupational and Environmental Health, Victor Segalen Bordeaux 2 University, Bordeaux, France.

* Corresponding author: INSERM U593, Université Victor Segalen Bordeaux 2, case no.11, 146, rue Léo-Saignat, 33076 Bordeaux Cedex, France. E-mail: Roger.Salamon@isped.u-bordeaux2.fr

these studies, factor structures were similar whether veterans were deployed or not deployed in the Gulf War.^{40–42,49} Despite different symptoms inventories and differences in statistical analyses, studies on Gulf War illness generally report from three to seven factors: (i) mood, cognition, fatigue, psychological symptoms; (ii) respiratory condition; (iii) neurological condition; (iv) musculoskeletal pain; (v) peripheral nervous system; (vi) gastrointestinal disorder; and (vii) mixed somatic complaints. However, no specific Gulf War syndrome has been retained.^{38–50}

Among those French PGWV who benefited from a close medical follow-up (general practitioners, psychiatrists) before 2000, no atypical acute pathological phenomenon was described.⁵¹ Since the military health network observed no new or unexplained pathology in the last decade, no specific monitoring or study of French PGWV was conducted. However, the first demand for compensation for a health problem related to the Gulf War was filed in 1995. In June 2000, a French Association of Gulf War Victims (Avigolfe) was founded. The Working Group in charge of the analysis of health data from French Gulf War Veterans, headed by one of us, was created in October 2000 by the French government. This group recommended an epidemiological study of all French PGWV in order to describe their complaints and to objectively measure their disorders through a standardized clinical examination.

The French Ministry of Defense, in collaboration with the Ministry of Health, solicited the INSERM in January 2002 to conduct an exhaustive investigation of all French PGWV (French Study on Persian Gulf War and its Health Consequences). The aim of this descriptive study was, mainly, to examine self-reported symptom data among Gulf War Veterans, to describe the main forms of exposure reported in the theatre, symptoms and diseases that appeared during and after the Persian Gulf mission, and to determine if unexpected statistical associations of such symptoms could suggest the presence of a new specific syndrome.

Methods

Population study

The French study on the Persian Gulf War and its health consequences was a cross-sectional exhaustive study which included all civilians and military personnel who served in the Persian Gulf from August 1990 to July 1991. The population list was established from several sources:

- a census conducted within the different military units between 2000 and 2001, provided by the army staff headquarters (nominative list of troops active at that time),
- the list of all decorated subjects for Persian Gulf War Operations, provided by the History Department of the Army (list created in 1991),
- a census of participating organizations for civilians.

In order to complete the nominative list, we also consulted operational records of regiments and their archives. The launch of the study was published in the newspapers. Finally, several subjects willing to participate to the study spontaneously contacted us and were added to the list, once we verified their participation in the Gulf War.

Current addresses were missing for many people. In order to retrieve missing addresses, the senior staffs of each army and the retirement pensions departments were contacted. Nominative lists were compared with telephone directories and all individuals were contacted by telephone in order to verify the eligibility criteria.

An explanatory letter stating the objectives of our study, of the study's objective, a consent form, a standardized postal questionnaire, a free-of-charge response envelope, and the list of the medical centres participating in the study were sent to all participants for whom a presumptive address was available.

The procedure for sending the questionnaire to every person eligible to participate was established independently from the army staff headquarters. It would have been possible to make the questionnaire available at military bases, but the nominal control and sending out of reminders would not have been possible, given the size of the military population and possible transfers between units.

People who did not answer the first letter or whose letter did not return stamped 'return to sender' within a 6-month period, received a reminder (an informative letter, a reply coupon listing reasons for not returning the questionnaire or refusal, and a prepaid envelope).

The National Commission of Data Processing and Civil Liberty approved this study, in agreement with article 15 paragraph 3 of the Law of January 6, 1978, concerning data processing, files, and civil liberty.

Questionnaire

Participants completed a 12-page postal self-administered questionnaire. The choice of a self-administered questionnaire was made due to the size of the population and their geographic distribution. Moreover, it gave participants a wider scope for answers than a controlled interview and they could take their time to answer the questions.

The questionnaire was developed on the basis of information already published in the authorities' reports (Working Group, Information Mission), and with reference to questionnaires used in PGWV morbidity studies. The questionnaire was reviewed by the Working Group's experts. Since the document needed to be as comprehensible as possible, it was pre-tested on five subjects of the target population and some questions were rephrased. It explored the following fields:

- socio-demographic characteristics (gender, date of birth, military, professional and civil status);
- military history (army, date of joining army, foreign operations before and after the Gulf War);
- living conditions during Persian Gulf mission, dates of departure and return, places of operation, and exposure (sandstorms, smokes from oil well fires, chemical or bacteriological alerts, vaccinations, medication);
- health: diseases and symptoms appearing before, during and after the war. A list of 49 symptoms appearing after the mission was rated according to the Hopkins Symptom Checklist.⁵² The frequencies of the symptoms were coded in five levels (never, less than once a month, one to three times a month, less than once a day, daily), except to six of them which were coded yes or no. Perceived stress was

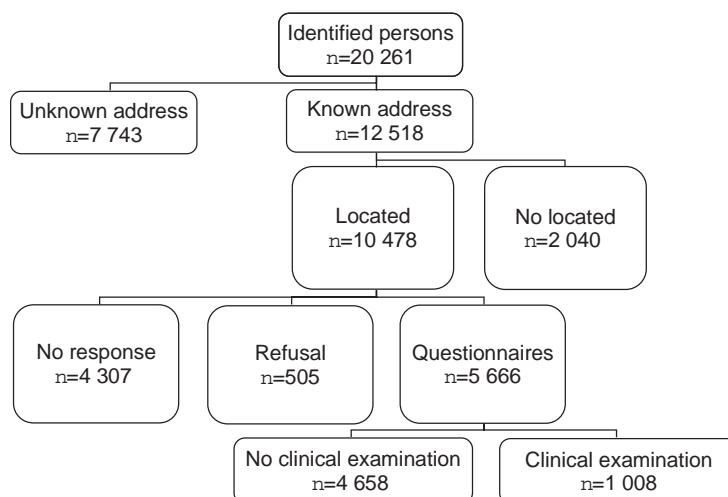


Figure 1 Population study

assessed with the four items of the Cohen and Williamson scale.⁵³ The ICD-10-CM was used to code the self-reported disease;

- children's health (miscarriages, number of ill children and diagnosis before and after the mission).

The questionnaire included 118 closed questions and 20 open questions concerning foreign operations and the illnesses present in participants or in their children before, during, or after the conflict. Furthermore, the inclusion of open questions obviously made data collection and analysis more difficult, but provided more substantial data (locations and exposures during the foreign operations, living conditions during the Gulf War mission...). Also, participants were allowed to provide further information if they wished to do so.

Medical examination

A free standardized clinical evaluation was performed by the 27 clinics of occupational disease in University Hospital centres and nine clinics of internal medicine in Military Hospitals. The participant of the study contacted one of the centres for an examination. When necessary, a consultation was conducted in French overseas departments and territories. These records were based on standardized clinical and para-clinical tests. Complementary tests were carried on when subjects presented specific signs.

A consultation report was regularly sent by all centres to the INSERM Unit 593 (laboratory exams results, exposures during the mission, medical diseases, and present health status), as well as data obtained by the Mini International Neuropsychiatric Interview.⁵⁴ The following syndromes were specified: fibromyalgia,⁵⁵ chronic fatigue syndrome,⁵⁶ post-traumatic stress disorders,⁵⁴ multiple chemical sensitivity,⁵⁷ and macrophagic myofasciitis.⁵⁸

On reception, the questionnaires and medical findings were made anonymous, coded, keyboarded and then analysed.

Statistical analyses

Statistical analyses were performed using SAS (SAS/STAT computers programs Cary, NC, SAS Institute, 2000) statistical software. The frequency distributions of the demographics and military service informations were examined. The response rates were very different among the branches of service. For this reason, living conditions and exposures during the mission were described stratified by branches of services.

In order to analyse the associations of symptom frequencies, we constructed the Pearson's correlation matrix among the 43 symptoms coded in five levels. We excluded respondents with missing values for any symptom and respondents who did not have any symptom at all. Principal components analysis was applied to the correlation matrix. Retained factors were defined by the Kaiser criterion (eigenvalues > 1). Moreover, a factor was retained if it had at least two symptom frequencies that loaded > 0.30 on this factor. The factor loading can be interpreted as a correlation coefficient measuring the association between the symptom frequencies and the factor.

Results

The survey began in February 2002. Data collection was completed in June 2004.

According to data obtained from the Ministry of Defense, a total of 20 261 French troops were deployed in the Gulf from August 1990 through July 1991. A current address was available for only 52% (10 478 subjects) and, among these, 5666 (54%) participated to our study (Figure 1). They belonged to different branches of service: 2695 served in the Army, 770 in the Navy, 1895 in the Air Force and 306 in other branches of service. The participation rate was 28% (5666/20 261) and varied according to the branch of service: 21% in the Army, 34% in the Air Force, and 41% in the Navy. Moreover, it was very difficult to contact people who had retired from the army between 1991 and 2002. The refusal rate was relatively low (5% of people contacted by

Table 1 Countries of deployment (a) and frequencies of self-reported exposures (b) during the Gulf War according to branches of service

	Army (n 5 2695)	Navy (n 5 770)	Air Force (n 5 1895)	Other ^a (n 5 306)	Total (n 5 5666)
(a) Countries deployed					
Saudi Arabia	80.0	21.2	89.8	90.2	75.8
Iraq	63.0	0.9	4.1	28.4	33.0
Kuwait	15.5	8.2	8.7	18.0	12.4
(b) Exposures					
Sandstorms	91.1	25.5	69.4	74.2	74.0
Chemical or biological alerts	80.7	9.7	56.7	73.9	62.7
Smoke from oil well fires	23.9	26.5	35.0	28.8	28.2
Insecticides	22.6	6.8	37.0	26.5	25.5

^a Other: state public force, military health service.

Table 2 Frequencies of self-reported medical conditions diagnosed after the Gulf War according to branches of service coded by ICD-10

Self-reported medical conditions (ICD-10)	Army (n 5 2695)	Navy (n 5 770)	Air Force (n 5 1895)	Other ^a (n 5 306)	Total (n 5 5666)
Infectious and parasitic diseases (A00–B99)	8.6	6.2	6.9	6.5	7.6
Neoplasms (C00–D48)	2.5	3.1	2.8	2.6	2.7
Diseases of the blood and blood-forming organs (D50–D89)	0.4	0.8	0.4	0.7	0.5
Endocrine, nutritional, and metabolic diseases, and immunity disorders (E00–E90)	6.3	7.0	9.3	7.8	7.5
Mental disorders (F00–F99)	7.2	4.3	5.4	11.1	6.4
Diseases of nervous system and sense organs (G00–G99)	3.7	2.1	2.7	3.3	3.1
Diseases of the eye and adnexa (H00–H59)	12.3	11.2	10.8	20.6	12.1
Diseases of the ear and mastoid process (H60–H95)	8.7	5.2	5.7	6.9	7.1
Diseases of the circulatory system (I00–I99)	6.1	4.9	7.8	13.4	6.9
Diseases of the respiratory system (J00–J99)	12.4	13.1	14.5	15.7	13.4
Diseases of the digestive system (K00–K93)	11.7	6.9	10.7	13.7	10.8
Diseases of the skin and subcutaneous tissue (L00–L99)	8.3	7.1	7.8	9.2	8.0
Diseases of the musculoskeletal system and connective tissue (M00–M99)	14.5	9.4	11.4	15.4	12.8
Diseases of the genito-urinary system (N00–N99)	6.6	3.8	5.6	6.9	5.9
Congenital anomalies; certain conditions originating in the perinatal period (Q00–Q99)	0.1	0.3	0.3	0.7	0.2
Symptoms, signs, ill-defined conditions (R00–R99)	13.0	6.2	9.3	13.7	10.9
Injury and poisoning (S00–T98)	11.1	5.8	9.4	11.1	9.8

^a Other: state public force, military health service.

June 2004), mainly because participants were only briefly present in the theatre or because they perceived themselves in good health. Only 1008 participants (18%) underwent a clinical examination.

Most of the 5666 subjects who completed the questionnaire were male (99.5%), with an average age of 41 years (standard deviation: 6 years) at the time of the completion of the survey, and 71% of respondents were still in service (73% in the Army, 87% in the Navy, 63% in the Air Force).

The average duration of deployment among respondents was 118 days, mainly in Saudi Arabia (76%) and in Iraq (33%). During their mission, subjects reported exposure mostly to sandstorms (74%) and chemical or bacteriological alerts (63%) (Table 1). The frequencies of self-reported exposures

were very different among branches of service, for example 91% of military serving in the Army were exposed to sandstorms whereas only 25.5% of military serving in the Navy were exposed.

While deployed in the Gulf, subjects suffered from the following symptoms and diseases, mainly diarrhoea (36%), headaches (28%), and difficulty to sleep (27%). Only 154 subjects were medically evacuated due to trauma or gastrointestinal infection.

Concerning self-reported medical diagnoses occurring after their mission (Table 2), 13.4% reported respiratory pathologies, 12.8% ophthalmic diseases, and 12.1% osteo-articular problems. Sixty-four cases of malignant disease were reported. Psychological problems, mainly anxiety or depression, were noted by

Table 3 Frequencies of self-reported symptoms appeared after the Gulf War by branches of service

Self-reported symptom	Army (n 5 2695)%	Navy (n 5 770)%	Air Force (n 5 1895)%	Other ^a (n 5 306)%	Total (n 5 5666)%
Headaches	85.9	80.6	79.4	84.0	82.9
Sleeping difficulties	72.5	60.0	72.2	75.8	70.9
Irritability	71.1	62.5	67.7	70.6	68.8
Backache	67.9	52.3	61.5	54.6	62.9
Memory difficulties	60.9	42.2	55.4	52.0	56.0
Fatigue	62.4	38.8	51.3	46.4	54.6
Difficulty to find words	56.5	41.8	53.1	46.1	52.8
Numbness or tingling	50.1	34.9	43.9	37.9	45.3
Gastralgia	46.6	36.0	39.6	44.8	42.7
Tinnitus	48.2	30.3	40.3	36.6	42.5
Sore throat	41.8	40.3	42.3	34.0	41.4
Joint pain	47.3	26.2	37.8	38.6	40.8
Diarrhoea	38.6	33.2	35.0	36.6	36.6
Fever	38.6	29.1	33.3	31.4	35.1
Muscle pain	39.8	20.5	32.2	29.4	34.1
Audition disorders	39.7	26.0	30.5	26.1	34.0
Hay fever or allergy	32.2	31.2	36.1	29.4	33.3
Sweating	36.3	20.9	30.8	25.5	31.8
Chest pain	34.1	23.9	27.7	23.5	30.0
Dizziness	33.5	22.6	26.2	22.5	29.0
Skin rashes	30.6	26.2	27.5	23.9	28.6
Problems with mouth, gums or teeth	31.8	22.7	27.9	19.9	28.6
Shortness of breath	31.0	20.1	25.4	21.6	27.1
Eyes trouble	29.2	19.2	25.1	21.6	26.0
Constipation	27.8	21.2	23.1	17.6	24.8
Rapid heart rate	27.5	17.1	22.9	27.1	24.5
Persistent cough	27.8	18.7	21.8	21.2	24.2
Depression	24.7	17.7	25.0	29.4	24.1
Difficulty with speech	27.1	14.9	20.3	20.6	22.8
Joint stiffness	26.9	12.2	19.8	18.6	22.1
Unintended weight gain . 10 lbs	25.8	14.0	18.4	24.2	21.6
Loss of strength	26.6	11.3	17.6	16.3	20.9
Nausea, vomiting	23.9	14.8	18.8	15.4	20.5
Hair loss	19.7	14.9	20.2	22.2	19.3
Panic or anxiety	21.2	12.7	17.9	17.3	18.7
Wheezing	22.2	11.8	15.2	11.4	17.8
Swollen glands	20.5	15.3	15.5	12.4	17.7
Confusion	21.9	10.5	15.9	8.5	17.6
Shaking	23.5	10.1	12.8	11.1	17.5
Chemical sensitivity	13.4	12.9	14.4	12.4	13.6
Frequent or painful urination	16.0	5.8	12.7	9.5	13.2
Joint swelling	15.1	6.1	11.0	11.8	12.3
Problems of swallowing	14.7	8.6	10.9	5.9	12.1
Impotence	12.2	5.3	12.2	10.8	11.2
Slowness of healing	10.2	3.8	6.6	5.9	7.9
Teeth loss	9.9	3.4	5.3	6.2	7.3
Unintended weight loss . 10 lbs	8.8	4.4	4.4	5.6	6.6
Problems during sexual intercourse	6.5	1.9	5.0	3.6	5.2
Tendency to bruise or bleed easily	4.9	1.4	2.6	2.3	3.5

^a Other: state public force, military health service.

Table 4 Principal component analysis—factor loading of the 43 symptoms scales of 5509 French PGWV

Symptom (percentage variance of each factor)	Factor 1 (29.97)	Factor 2 (4.21)	Factor 3 (3.56)	Factor 4 (3.07)	Factor 5 (2.91)	Factor 6 (2.80)	Factor 7 (2.65)
Difficulty to find words	0.56	-0.42	--	-0.33	--	--	--
Memory difficulties	0.63	-0.41	--	-1	--	--	--
Difficulty with speech	0.57	-0.36	--	--	-1	-1	--
Confusion	0.62	-0.30	--	--	--	-1	--
Wheezing	0.52	0.32	--	--	--	-1	-1
Persistent cough	0.51	0.36	--	--	--	--	-1
Sore throat	0.53	0.38	--	-1	--	--	--
Depression	0.55	-1	-0.38	-1	--	--	--
Panic or anxiety	0.56	-1	-0.35	-1	--	--	--
Muscle pain	0.68	--	0.37	--	--	--	--
Joint pain	0.66	--	0.47	--	--	--	--
Joint stiffness	0.59	--	0.47	--	--	--	--
Joint swelling	0.47	--	0.37	--	--	-1	--
Audition disorders	0.46	--	-1	-0.33	-0.34	0.31	-0.32
Tinnitus	0.51	--	--	-0.31	-1	0.34	-0.31
Rapid heart rate	0.56	--	--	--	--	--	-0.30
Shortness of breath	0.67	--	--	--	--	-1	-0.30
Headaches	0.49	--	--	--	0.31	-1	--
Impotence	0.43	--	--	-1	-0.35	--	--
Shaking	0.59	--	--	-1	--	--	--
Loss of strength	0.67	--	--	-1	--	--	--
Problems during sexual intercourse	0.40	--	--	0.32	-0.40	--	--
Diarrhoea	0.49	--	--	--	--	0.33	--
Swollen glands	0.43	--	--	--	--	--	--
Nausea, vomiting	0.56	--	--	--	--	--	--
Hay fever or allergy	-1	-1	--	-0.34	--	--	-1
Chemical sensitivity	0.38	-1	--	--	--	-1	-1
Problems of swallowing	0.50	-1	--	--	-1	--	--
Sleeping difficulties	0.56	-1	--	--	-1	-1	--
Fatigue	0.73	--	--	--	--	--	--
Dizziness	0.61	--	--	--	--	--	--
Problems with vision	0.58	--	--	--	--	--	--
Numbness or tingling	0.61	--	--	--	--	--	--
Backache	0.59	--	-1	--	-1	--	--
Irritability	0.61	-1	--	--	-1	--	--
Sweating	0.55	--	--	--	--	--	--
Frequent or painful urination	0.46	--	--	--	-1	-1	--
Gastralgia	0.54	--	--	--	--	-1	--
Chest pain	0.59	--	--	--	--	--	-1
Skin rashes	0.48	--	--	--	--	--	-1
Problems with mouth, gums or teeth	0.51	--	--	--	--	--	--
Constipation	0.41	--	--	--	--	--	-1
Fever	0.57	--	--	--	--	--	--

--: factor loading < 0.20; -1 : factor loading between 0.20 and 0.30.

7% of responders. Nonetheless, 86% of respondents considered themselves in good health.

The most frequent symptoms described since the return from the Gulf were headaches (82.9% of participants), neurological or psychological symptoms, and back pain (Table 3). The highest frequencies were reported in the Army and the lowest in the Navy.

A principal component analysis was performed for those subjects ($n = 5509$) who reported at least a symptom for several consecutive months. Seven factor axes were retained because their eigenvalue was > 1 (Kaiser criterion) and, at least for two symptoms, the factor loading was > 0.30 . These seven factors represented 49.2% of the explained variance (Table 4). Results from this analysis showed associations between six clusters of symptoms:

- i. Factors 1 and 2 showed two associations:
 - a. the association of neurocognitive symptoms (including difficulty to find words, memory difficulty, difficulty with speech, and confusion);
 - b. orthogonally, the association of respiratory symptoms (wheezing, persistent cough, and sore throat).
- ii. Factors 1 and 3 showed two associations:
 - a. the association of psychological symptoms (depression and panic or anxiety attacks);
 - b. orthogonally, the association of musculo-skeletal symptoms (muscle pain, joint pain, joint stiffness, and joint swelling).
- iii. For Factors 1 and 4, there was an association between auditive disorders and tinnitus;
- iv. For Factors 1 and 7, there was an association between rapid heart rate and shortness of breath.

The other symptoms had loaded only with the first factor, which constitutes a 'general' factor of symptoms without any specific interpretation. Apart from these nosologic associations, no other cluster, involving disparate organ symptoms, was highlighted by our analysis.

Discussion

To our knowledge, this study is the first study on French Gulf War Veterans, who experienced different risks during their mission. Its aim was, mainly, to describe self-reported symptom data among Gulf War Veterans and the association of symptoms. The data obtained by a self-administrated questionnaire and the clinical examination were very important to investigate the potential health impact of the Gulf War and to compare it with that reported by US and UK veterans.

Twelve years after the Gulf War, we were able to contact 52% of French PGWV, and 54% of them accepted to participate in this study. The frequencies of symptoms presented were slightly higher than those already published. Headaches after the mission have been described in 13–75% of veterans^{15,16,19,22,24–26,29,32,36} as opposed to 83% in our study. Back pain has been reported by 28–44% of veterans^{16,19,24,36} as opposed to 63% in our sample. Sleeping disorders have been reported by 6–56% of veterans,^{14,15,19,21–26,29,36} as opposed to 71% in our survey. Conversely, memory disorders, irritability, and fatigue were found with the frequencies

similar to those reported in other studies.^{15,19–26,31,36} Even if French troops did not face the same exposures as US or UK troops, a 'new' symptoms cluster, involving disparate organ symptoms, was not highlighted by our analysis. In our study, the associations of symptoms were quite similar to those derived among other PGWV and appropriate non-deployed control veterans.^{39–50}

If the only aim of the survey had been to investigate the existence of a specific Gulf War syndrome, then only a randomized sample could have been recruited as opposed to the entire population. However, this was not the case since the French Government wanted to offer the opportunity to all French PGWV to benefit from a free medical examination. Taking into account these elements as well as the fact that the study began 10 years after the mission, the epidemiological responses that can be drawn from this study are certainly limited and are only an appraisal of respondents' health.

The study limitations mainly concern selection and information bias. The study aimed to be exhaustive, and thus include all military and civil personnel who participated in the Persian Gulf War between 1990 and 1991. The nominal lists of people concerned were transmitted by the military services. A first selection bias could have been due to the fact that lists could have been incomplete, although it is not plausible that a subject could have been at the operations theatre without the military services being aware of his presence. Also, some subjects could not be contacted. A second selection bias could be due to a massive media coverage of complaints lodged due to health consequences of the Gulf War in foreign countries, so respondents, with health disorders or at least worried about them, were more willing to participate in our study. However this bias is compensated by the fact that 71% respondents were still in service, reported as the 'healthy-soldier effect' in other studies.^{3,59}

An information bias is inherent to observational studies measuring exposure which precede the study for 10 years and therefore health events which occurred after exposures. However, it is probable that symptoms or serious diseases, requiring specific care or of unusual frequency or intensity, were notified more often than events considered as slight or benign, or needing simple self-medication. Conversely, exposures may probably be more affected by information bias, particularly risks encountered during the Gulf War mission or during previous operations dating back more than 10 years. Finally, individuals with a general tendency towards negative perceptions of life may over-report exposures and symptoms, leading to a spurious association between exposure and symptoms.^{60,61} One solution to avoid this would be to consider collective and not individual exposure according to a location or an activity performed there (aerial offensive, forward push).

Conclusion

Until 2000, France did not receive any specific complaints and no study on French PGWV was undertaken. This first French survey on PGWV started in February 2002. Fourteen years after the end of the Gulf War (in 1990–91), it is very difficult to describe deployment circumstances, living conditions, and exposures and to link them to illnesses. Further studies, based on data collected, are in progress and results will be published in the near future.

Acknowledgements

The French Study on Persian Gulf War and its Health Consequences was funded by the French Ministry of Defense. We gratefully acknowledge the contribution of Véronique

Bigorie, Brigitte Manoux, Evelyne Mouillet, Eric Pin, Valérie Mazou, Sylvain Maurisset and Maria Winnock. Lastly, but most importantly, we sincerely thank the Gulf War Veterans for their participation.

KEY MESSAGES

- Many studies on the health of Persian Gulf War Veterans have been undertaken, mainly in the US and the UK. This study is the first on Persian Gulf War and its health consequences conducted in France.
- The most frequent symptoms described 10 years after the return from the Gulf were headaches, neurological or psychological symptoms, and back pain. These frequencies were similar or slightly higher than those of studies realized before our study.
- Apart from well-known symptoms associations (respiratory, neurocognitive, psychological and musculoskeletal syndromes), no other cluster was highlighted by our analysis.

References

- 1 Picard G. Au feu avec la Division Daguet. Glénat ed. Paris, 1991.
- 2 Hotopf M, David A, Hull L, Ismail K, Unwin C, Wessely S. Role of vaccinations as risk factors for ill health in veterans of the Gulf war: cross sectional study. *BMJ* 2000;**320**:1363–67.
- 3 Kang HK, Bullman TA. Mortality among U.S. veterans of the Persian Gulf War. *N Engl J Med* 1996;**335**:1498–504.
- 4 Macfarlane GJ, Thomas E, Cherry N. Mortality among UK Gulf War Veterans. *Lancet* 2000;**356**:17–21.
- 5 Kang HK, Bullman TA, Macfarlane GJ, Gray GC. Mortality among US and UK veterans of the Persian Gulf War: a review. *Occup Environ Med* 2002;**59**:794–49.
- 6 Gray GC, Gacksetter GD, Kang HK, Graham JT, Scott KC. After more than 10 years of Gulf War veteran medical evaluations. What have we learned? *Am J Prev Med* 2004;**26**:443–52.
- 7 Maconochie N, Doyle P, Carson C. Infertility among male UK veterans of the 1990–1 Gulf war: reproductive cohort study. *BMJ* 2004;**329**:196–201.
- 8 Sim M, Abramson M, Forbes A, Glass D, Ikin P, Ittak P *et al.* Australian Gulf veterans' health study 2003. Vol. 2. Commonwealth Department of Veterans' Affairs, 2003. Available at: www.dva.gov.au/media/publicat/2003/gulfwarhs (Accessed on July 25, 2005).
- 9 Doyle P, Maconochie N, Davies G, Maconochie I, Pelerin M, Prior S *et al.* Miscarriage, stillbirth and congenital malformation in the offspring of UK veterans of the first Gulf war. 2004;**33**:74–86.
- 10 Araneta MR, Destiche DA, Schlangen KM, Merz RD, Forrester MB, Gray GC. Birth defects prevalence among infants of Persian Gulf War veterans born in Hawaii, 1989–1993. *Teratology* 2000;**62**:195–204.
- 11 Cowan DN, DeFraités RF, Gray GC, Goldenbaum MB, Wishik SM. The risk of birth defects among children of Persian Gulf War veterans. *N Engl J Med* 1997;**336**:1650–56.
- 12 Kang H, Magee C, Mahan C, Lee K, Murphy F, Jackson L *et al.* Pregnancy outcomes among U.S. Gulf War veterans: a population-based survey of 30 000 veterans. *Ann Epidemiol* 2001;**11**:504–11.
- 13 Ryan MAK, PershynKisor MA, Honner WK, Smith TC, Reed RJ, Gray GC. The Department of Defense Birth Defects Registry: Overview of a new surveillance system. *Teratology* 2001;**64**:S26–29.
- 14 Joseph SC. A comprehensive clinical evaluation of 20 000 Persian Gulf War veterans. Comprehensive Clinical Evaluation Program Evaluation Team. *Mil Med* 1997;**162**:149–55.
- 15 Kizer KW, Joseph S, Moll M, Rankin JT. Unexplained illness among Persian Gulf War veterans in an Air National Guard unit: Preliminary report—August 1990–March 1995. *J Am Med Assoc* 1995;**274**:16–17.
- 16 Stretch RH, Bliese PD, Marlowe DH, Wright KM, Knudson KH, Hoover CH. Physical health symptomatology of Gulf War-era service personnel from the states of Pennsylvania and Hawaii. *Mil Med* 1995;**160**:131–36.
- 17 Schwartz DA, Doebbeling BN, Merchant JA, Barrett DH, Black DW, Burmeister LF *et al.* Self-reported illness and health status among Gulf War veterans: A population-based study. *J Am Med Assoc* 1997;**277**:238–45.
- 18 Kroenke K, Koslowe P, Roy M. Symptoms in 18 495 Persian Gulf War veterans. Latency of onset and lack of association with self-reported exposures. *J Occup Environ Med* 1998;**40**:520–28.
- 19 Proctor SP, Heeren T, White RF, Wolfe J, Borgos MS, Davis JD *et al.* Health status of Persian Gulf War veterans: self-reported symptoms. Environmental exposures and the effect of stress. *Int J Epidemiol* 1998;**27**:1000–10.
- 20 Roy MJ, Koslowe PA, Kroenke K, Magruder C. Signs, symptoms, and ill-defined conditions in Persian Gulf War veterans: findings from the Comprehensive Clinical Evaluation Program. *Psychosom Med* 1998;**60**:663–68.
- 21 Coker WJ, Bhatt BM, Blatchley NF, Graham JT. Clinical findings for the first 1000 Gulf war veterans in the Ministry of Defence's medical assessment programme. *BMJ* 1999;**318**:290–94.
- 22 Gray GC, Kaiser KS, Hawksworth AW, Hall FW, Barrett-Connor E. Increased postwar symptoms and psychological morbidity among U.S. Navy Gulf War veterans. *Am J Trop Med Hyg* 1999;**60**:758–66.
- 23 Murphy FM, Kang H, Dalager NA, Lee KY, Allen RE, Mather SH *et al.* The health status of Gulf War veterans: lessons learned from the Department of Veterans Affairs Health Registry. *Mil Med* 1999;**164**:327–31.
- 24 Kang HK, Mahan CM, Lee KY, Magee CA, Murphy FM. Illnesses among United States veterans of the Gulf War: a population-based survey of 30 000 veterans. *J Occup Environ Med* 2000;**42**:491–501.
- 25 Steele L. Prevalence and patterns of Gulf War illness in Kansas veterans: Association of symptoms with characteristics of person, place, and time of military service. *Am J Epidemiol* 2000;**152**:992–1002.
- 26 Gray GC, Reed RJ, Kaiser KS, Smith TC, Gastanaga VM. Self-reported symptoms and medical conditions among 11 868 Gulf War-era veterans. *Am J Epidemiol* 2002;**155**:1033–44.
- 27 Southwick SM, Morgan A, Nagy LM, Bremner D, Nicolaou AL, Johnson DR *et al.* Trauma-related symptoms in veterans of Operation Desert Storm: a preliminary report. *Am J Psychiatry* 1993;**150**:1524–28.
- 28 Perconte ST, Wilson AT, Pontius EB, Dietrick AL, Spiro KJ. Psychological and war stress symptoms among deployed and

- non-deployed reservists following the Persian Gulf War. *Mil Med* 1993;**158**:516–21.
- ²⁹ Newmark J, Clayton WL 3rd. Persian Gulf illnesses: preliminary neurological impressions. *Mil Med* 1995;**160**:505–07.
- ³⁰ Goldstein G, Beers SR, Morrow LA, Shemansky WJ, Steinhauer SR. A preliminary neuropsychological study of Persian Gulf veterans. *J Int Neuropsychol Soc* 1996;**2**:368–71.
- ³¹ Stretch RH, Bliese PD, Marlowe DH, Wright KM, Knudson KH, Hoover CH. Psychological health of Gulf War-era military personnel. *Mil Med* 1996;**161**:257–61.
- ³² Amato AA, McVey A, Cha C, Matthews EC, Jackson CE, Kleingunther R *et al*. Evaluation of neuromuscular symptoms in veterans of the Persian Gulf War. *Neurology* 1997;**48**:4–12.
- ³³ Engel CC, Ursano R, Magruder C, Tartaglione R, Jing Z, Labbate LA *et al*. Psychological conditions diagnosed among veterans seeking Department of Defense Care for Gulf War-related health concerns. *J Occup Environ Med* 1999;**41**:384–92.
- ³⁴ Wolfe J, Proctor SP, Erickson DJ, Heeren T, Friedman MJ, Huang MT *et al*. Relationship of psychiatric status to Gulf War veterans' health problems. *Psychosom Med* 1999;**61**:532–40.
- ³⁵ Dlugosz LJ, Hocter WJ, Kaiser KS, Knoke JD, Heller JM, Hamid NA *et al*. Risk factors for mental disorder hospitalization after the Persian Gulf War: U.S. Armed Forces, June 1, 1991–September 30, 1993. *J Clin Epidemiol* 1999;**52**:1267–78.
- ³⁶ Unwin C, Blatchley N, Coker W, Ferry S, Hotopf M, Hull L *et al*. Health of UK servicemen who served in Persian Gulf War. *Lancet* 1999;**353**:169–78.
- ³⁷ Barrett DH, Gray GC, Doebbeling N, Clauw DJ, Reeves WC. Prevalence of Symptoms and Symptom-based conditions among Gulf War veterans: current status of research findings. *Epidemiol Rev* 2002;**24**:218–27.
- ³⁸ Haley RW, Kurt TL, Hom J. Is there a Gulf War Syndrome? Searching for syndromes by factor analysis of symptoms. *J Am Med Assoc* 1997;**277**:215–22.
- ³⁹ Fukuda K, Nisenbaum R, Stewart G, Thompson WW, Robin L, Washko RM *et al*. Chronic multisymptom illness affecting Air Force veterans of the Gulf War. *J Am Med Assoc* 1998;**280**:981–88.
- ⁴⁰ Ismail K, Everitt B, Blatchley N, Hull L, Unwin C, David A *et al*. Is there a Gulf War syndrome? *Lancet* 1999;**353**:179–82.
- ⁴¹ Doebbeling BN, Clarke WR, Watson D, Torner JC, Woolson RF, Voelker MD *et al*. Is there a Persian Gulf War syndrome? Evidence from a large population-based survey of veterans and nondeployed controls. *Am J Med* 2000;**108**:695–704.
- ⁴² Knoke JD, Smith TC, Gray GC, Kaiser KS, Hawksworth AW. Factor analysis of self-reported symptoms: does it identify a Gulf War syndrome? *Am J Epidemiol* 2000;**152**:379–88.
- ⁴³ Cherry N, Creed F, Silman A, Dunn G, Baxter D, Semdley J *et al*. Health and exposures of United Kingdom Gulf war veterans. Part I: the pattern and extent of ill health. *Occup Environ Med* 2001;**58**:291–98.
- ⁴⁴ Bourdette DN, McCauley LA, Barkhuizen A, Johnston W, Wynn M, Joos SK *et al*. Symptom factor analysis, clinical findings, and functional status in a population-based case control study of Gulf War unexplained illness. *J Occup Environ Med* 2001;**43**:1026–40.
- ⁴⁵ Kang HK, Mahan CM, Lee KY, Murphy FM, Simmens SJ, Young HA *et al*. Evidence for deployment-related Gulf War syndrome by factor analysis. *Arch Environ Health* 2002;**57**:61–68.
- ⁴⁶ Everitt B, Ismail K, David AS, Wessely S. Searching for a Gulf War syndrome using cluster analysis. *Psychol Med* 2002;**32**:1371–78.
- ⁴⁷ Shapiro SE, Lasarev MR, McCauley L. Factor analysis of Gulf War illness: what does it add to our understanding of possible health effects of deployment? *Am J Epidemiol* 2002;**156**:578–85.
- ⁴⁸ Hallman WK, Kipen HM, Diefenbach M, Boyd K, Kang H, Leventhal H *et al*. Symptom patterns among Gulf War registry veterans. *Am J Public Health* 2003;**93**:624–30.
- ⁴⁹ Young HA, Simmens SJ, Kang HK, Mahan CM, Levine PH. Factor analysis of fatiguing syndrome in Gulf War era veterans: implications for etiology and pathogenesis. *J Occup Environ Med* 2003;**45**:1268–73.
- ⁵⁰ Nisenbaum R, Ismail K, Wessely S, Unwin C, Hull L, Reeves WC. Dichotomous factor analysis of symptoms reported by UK and US veterans of the 1991 Gulf War. *Popul Health Metr* 2004;**2**:8.
- ⁵¹ Malafosse A, Nugeyre M, Grosclaude B, Pierre A, Poyot G. Soutien sanitaire de l'opération Daguet. *Médecine & Armées* 1992;**20**:11–15.
- ⁵² Derogatis LR, Lipman RS, Rickels K, Uhlenhuth EH, Covi L. Hopkins symptom checklist. In: Guy W (ed). ECDEU assessment manual for psychopharmacology. Rockville, MD: US National Institute of Health, Psychopharmacology Research Branch, 1974, pp. 575–78.
- ⁵³ Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *J Health Soc Behav* 1983;**24**:385–96.
- ⁵⁴ Lecrubier Y, Weiller E, Hergueta T, Amorim P, Bonora LI, Lépine JP *et al*. Mini International Neuropsychiatric Interview French Version 5.0.0. Paris, France: INSERM, 1998.
- ⁵⁵ The American College of Rheumatology 1990. Criteria for the Classification of Fibromyalgia. Report of the Multicenter Criteria Committee. *Arthritis Rheum* 1990;**33**:160–72.
- ⁵⁶ Fukuda K, Straus S, Hickie I, Sharpe MC, Dobbins JG, Komaroff A *et al*. Chronic fatigue syndrome: a comprehensive approach to its definition and study. *Ann Intern Med* 1994;**121**:953–59.
- ⁵⁷ Multiple Chemical Sensitivity: 1999 Consensus. *Arch Environ Health* 1999;**54**:147–49.
- ⁵⁸ Chérin P, Laforêt P, Ghérardi RK, Authier FJ, Coquet M, Maisonobe T. La myofasciite à macrophages: description, hypothèses étiopathogéniques. *Rev Méd Interne* 1999;**20**:483–89.
- ⁵⁹ Haley R. Point: bias from the "healthy-warrior effect" and unequal follow-up in three government studies of health effects of the Gulf War. *Am J Epidemiol* 1998;**4**:315–23.
- ⁶⁰ Watson D, Pennebaker JW. Health complaints. Stress and distress: exploring the central role of negative affectivity. *Psychol Rev* 1989;**96**:234–54.
- ⁶¹ Macleod J, Davey Smith G, Heslop P, Metcalfe C, Carroll D, Hart C. Psychological stress and cardiovascular disease: empirical demonstration of bias in prospective observational study of Scottish men. *BMJ* 2002;**324**:1247–51.

4. Discussion

A partir d'une demande institutionnelle effectuée en octobre 2000 s'est constitué tout un cheminement d'étapes et de productions de ressources de connaissances à destination d'acteurs différents. Les chercheurs, présents initialement au sein du groupe d'experts puis dans la coordination de l'enquête, étaient confrontés à la situation de devoir répondre à la demande institutionnelle, qui était elle-même en situation de devoir répondre à une demande de la société, mais aussi de devoir nourrir d'informations la communauté de chercheurs internationaux ayant pour sujets de recherche l'état de santé des militaires présents sur le « théâtre des opérations »⁴³ de la guerre du Golfe 1990 - 1991. Les chercheurs étaient pressés par les politiques de les aider à trouver LA solution permettant de répondre à LA pression sociétale, venant tout particulièrement d'une association de défense des civils et des militaires actifs ou non, atteints de maladies de la guerre du Golfe ou des Balkans, AviGolfe⁴⁴. Présenter avec précaution les éléments qui pourraient être obtenus *via* les synthèses de la littérature puis les études mises en place étaient l'enjeu de la transmission du chercheur en direction des politiques, tout cela en vue de constituer une aide à la décision politique. Le chercheur tout au long des étapes (du rapport de mission au rapport d'enquête puis aux publications) a communiqué aux politiques un ensemble d'éléments issu de la synthèse de la littérature, puis de l'analyse statistique. La communication scientifique a été reportée ultérieurement, avec un décalage de publication de près de deux ans.

Le questionnement initial tournait principalement au début autour de la question du syndrome de la guerre du Golfe, et de son existence. De nombreuses publications internationales avaient été publiées suite aux études très volumineuses réalisées par les américains, les canadiens, ... Une première réponse issue de l'étude

⁴³ Terme employé par la communauté militaire.

⁴⁴ Association créée par Hervé Desplat et Christine Abdelkrim-Delanne (journaliste) – URL : <http://www.avigolfe.com/>

de la bibliographie donnait déjà les éléments de réponse soulignant le non fondement de l'appellation « syndrome de la guerre du Golfe ».

« Existe-t-il un syndrome de la guerre du Golfe ?

Il existe à l'évidence une augmentation nette et significative de « plaintes » chez les militaires ayant participé à la guerre du Golfe, mais il semble que ces plaintes soient très diverses, que leurs causes ne soient pas toujours précisées, et en tout cas certainement pas uniques. Ainsi, les auteurs s'accordent généralement pour reconnaître que n'apparaît pas clairement la notion d'un syndrome évident.

Peut-être existe-t-il une spécificité dans la « combinaison » des signes présentés par les vétérans qui serait la possible expression d'un syndrome. La recherche d'une telle spécificité combinatoire doit reposer sur des méthodes statistiques multivariées de type « analyse factorielle » qui sont des méthodes classiques bien qu'assez sophistiquées.

L'analyse factorielle est une technique appropriée pour chercher un nombre minimum de « facteurs » pouvant représenter une part importante de l'information contenue dans un ensemble de données mesurées de manière quantitative ou binaire. Il s'agit de réduire un nombre important d'informations à quelques grandes dimensions. Les informations recueillies pour un individu sont résumées en un nombre limité de « facteurs » construits comme combinaisons linéaires des variables initiales et traduisant leurs associations. On utilise le terme de variables latentes pour parler de ces facteurs, qui existent au plan conceptuel seul et qui ne sont pas mesurées directement.

Cinq équipes ont tenté de répondre à la question en utilisant de telles méthodes statistiques sur les symptômes recueillis par autoquestionnaires. Les cinq travaux cités ici sont considérés Outre-Atlantique comme des références importantes.

Haley a recherché les regroupements de symptômes uniquement chez un groupe de 249 vétérans américains (Haley, 1997). Les 4 autres articles comparent les regroupements obtenus par une analyse factorielle à partir d'un échantillon de militaires déployés sur le Golfe à ceux obtenus sur un groupe de militaires non déployés (Fukuda, 1998 ; Ismaïl, 1999 ; Doebbeling, 2000 ; Knoke, 2000). Ces différents auteurs utilisent le même outil statistique pour répondre à la question et prennent des critères identiques de construction du modèle. » (Salamon et al., 2001)

Une grande partie de ce qui s'est joué dans le cadre de cette expertise, puis de l'étude commanditée par le politique, avait été initiée par l'apparition d'un terme « syndrome de la guerre du Golfe » brandi par certains comme une nouvelle maladie mais sans fondement ni sanitaire (il existait selon les études étrangères des regroupement de symptômes mais sans réel spécificité), ni statistique (les études statistiques étrangères confirmaient les observations médiales de non spécificité). Il était ainsi de la mission du chercheur, dans la demande qui lui avait été formulée, de

déconstruire le construit s'il n'était qu'imaginaire. Cette dénomination n'était pas issue de nulle part et portait en elle le ressenti d'une situation de mauvais état de santé de certains militaires et civils, associé à un manque de reconnaissance du service rendu. Il s'agissait d'employer ainsi un terme pouvant permettre de dénommer en un groupe nominal la cause et la conséquence. Face à cette situation et à l'impact en terme financier que pouvait avoir l'apparition d'une nouvelle pathologie, les politiques en missionnant un groupe d'experts sur la question de l'existence du syndrome de la guerre du Golfe faisaient d'une certaine manière acte de dévolution de prise de décision aux chercheurs. Il s'agissait selon les politiques de savoir si l'énoncé de l'existence du syndrome de la guerre du Golfe était scientifique ou non. La grande difficulté pour tous les acteurs étant intervenus dans cette discussion était en réalité le point de départ de la question la non-définition même de ce que devait être le « syndrome de la guerre du Golfe », et donc une utilisation dans la pratique très différente d'un acteur à l'autre. Comme le souligne Wittgenstein dans le « Cahier Bleu »,

« l'utilisation dans la pratique est son sens [...] Le sens de l'expression dépend entièrement de la manière dont nous l'utilisons jour après jour. Ne nous imaginons pas que le sens est un lien occulte que l'esprit établit entre un mot et une chose, et que le lien contient l'usage tout entier du mot, comme on pourrait dire que la graine contient l'arbre ». (Wittgenstein, 1996)

La situation didactique pour les chercheurs n'était pas uniquement de concevoir la démarche statistique permettant de répondre à la question d'existence du syndrome de la guerre du Golfe, mais bien de formuler différemment la question pour étudier les signes et symptômes pouvant être associés à la présence sur les terrains d'opération.

Président du groupe d'experts – audition 1- Dire qu'il y a un syndrome équivaut à mon avis à rentrer dans un jeu qui d'emblée est faux et assez peu honnête, ce que ni les membres de l'Assemblée nationale, ni ceux d'une assemblée scientifique ne peuvent permettre. En revanche, on peut parfaitement refuser de parler de « syndrome » au profit des signes et des plaintes, [...]

Dans le cadre de cette étude, la démarche statistique consistait tout d'abord et avant tout à un travail descriptif. Suite à la lettre de mission d'octobre 2000, le premier travail statistique a consisté en une synthèse de la littérature, et une étude de la population des dossiers de plaintes. Puis, suite à la remise du rapport associé à ce

premier travail l'étude épidémiologique a été réalisée et l'analyse statistique a été descriptive et analytique, tout ceci afin de pouvoir décrire les associations de symptômes et signes, mais aussi les associations possibles entre les expositions environnementales et les conséquences.

Cette étude épidémiologique, réalisée suite à une demande de la société formulée auprès des politiques, souligne combien dans ce cadre la culture statistique participe à autre chose qu'à la seule réflexion scientifique. Le cadre de la demande sociale induit tout un environnement bien différent, et une nécessité de reconstruire la demande même de la question. Le chercheur doit, par sa rhétorique, expliquer aux politiques son cheminement scientifique pour répondre à la question qu'ils lui ont formulée et qui est elle-même issue de leur propre transposition de la demande sociétale. Nous avons signalé la différence d'écoute, de compréhension et d'attente entre les élus ce qui justifie l'importance de pouvoir expliciter sa démarche.

Sollicité en tant qu'expert, le chercheur a nécessité à comprendre les enjeux de la situation, non pas pour en jouer mais pour ajuster son discours et effectuer avec pertinence son travail de transposition didactique. Le chercheur est souvent appelé par les politiques quand, comme dans la situation présentée ici, le débat devient passionnel, politique, médical et parfois financier. Il se doit de conserver sa posture de chercheur et non de membre de la société, car c'est la demande qui lui est faite. Les éléments qu'il apporte sont de plusieurs niveaux, car il s'agit de réponses aux questions qui lui ont été posées mais aussi d'explications quant à la démarche. Il se doit de s'assurer de la bonne compréhension des informations transmises afin d'assurer un transfert de qualité.

C. Synthèse

Les deux situations, doctorat et expertise, correspondent pour le chercheur à des périodes lui permettant de tester ses capacités à expliciter les étapes et à en rendre le discours compréhensible à des fins de diffusion et d'appropriation. Il ne s'agit donc pas de situations de vulgarisation ou de simple communication, mais les enjeux sont bien des enjeux de transfert et de diffusion de culture, scientifique pourrions-nous dire.

Ces deux situations soulignent certaines compétences nécessaires aux chercheurs de capacités de transmettre des éléments propres à sa communauté en interne ou en externe dans une visée d'explication et de compréhension. La distance à parcourir pour son travail de transposition didactique semble pouvoir être plus ou moins grande. N'est-il pas plus simple pour un chercheur d'échanger avec ses pairs qu'avec des politiques potentiellement naïfs de sa culture disciplinaire ? Les transmissions intra-disciplinaires peuvent finalement se nourrir d'implicite, mais ce point devient impossible dans le cadre de situations telles que celle décrite précédemment lors de l'audition car les implicites pourraient conduire l'échange vers des incompréhensions tout simplement par l'absence de références partagées.

Ces observations confirment, selon nous, la remarque que nous avons soulignée en début de thèse et formulée par Chevillard et Wozniak (2003) qui consistait à souligner l'importance de considérer la didactique comme partie prenante de la formation à la statistique car la statistique est une discipline qui participe à de nombreuses réflexions scientifiques et sociétales comme vu précédemment, mais aussi qui envahit parfois l'environnement social et nécessite que le citoyen soit formé pour comprendre les notions sous-jacentes aux résultats qui lui sont exprimés.

De plus, le doctorat en tant que diplôme terminal d'une formation universitaire est une fenêtre intéressante pour évaluer les contenus pédagogiques réinvestis dans ce travail terminal, mais aussi pour évaluer les nouveaux éléments théoriques et techniques qui ont été nécessaires dans la réflexion et mise en œuvre et pour lesquels aucune formation n'avait été intégrée dans le cursus. Régnier indique « *C'est ce même sentiment de mauvaise préparation à la statistique auquel nous sommes confrontés dans l'accompagnement et la direction des travaux de recherche*

conduits par des étudiants accueillis en sciences humaines et sociales, qui alimente une part importante de nos questionnements didactiques et pédagogiques. » (Régnier 2012)

L'ensemble de ces éléments sont ainsi à prendre en considération dans la réflexion de la formation universitaire et en tenant compte des changements de paradigmes que semble vivre l'université. Les réflexions disciplinaires nourrissent ainsi la réflexion de la pédagogie universitaire, mettant ainsi en connexion les changements environnementaux de la formation universitaire, mais aussi des publics accueillis, les évolutions sociétales et professionnelles.

CHAPITRE 4 – ENJEUX DE L'ENSEIGNEMENT DE LA STATISTIQUE A L'UNIVERSITE

Enseigner la statistique à l'université est devenu presque incontournable afin de pouvoir former les étudiants aux différentes situations qui les attendent professionnellement, mais aussi et tout simplement en tant que formation citoyenne (Watson, 1997). Mais que doit-on enseigner et quelle place donner à l'enseignement de la didactique de la statistique ? Apprendre des savoir et des savoir-faire statistiques est un élément important, mais être en capacité d'explicitier les choix réalisés et les stratégies à des publics divers est fondamental pour cette discipline dont les résultats n'ont de sens que mis au regard de la méthode employée et des choix réalisés.

Se pencher sur la question de l'enseignement de la statistique à l'université nécessite de considérer les processus d'organisation et de construction de l'offre de formation ainsi que les réflexions de développement de la pédagogie universitaire.

Afin de pouvoir aborder ces points, il est important tout d'abord de comprendre les interactions entre le niveau national et le niveau local quant à la construction de l'offre de formation, mais aussi les étapes de construction des formations et les acteurs pouvant influencer le choix des contenus et des approches pédagogiques.

A. L'université et l'approche compétence

Chaque université est autonome dans la construction de son offre de formation et de ses contenus dans le respect du cadre national. Il n'y a donc pas d'injonction nationale pour la construction de l'offre de formation, mais une évaluation sur proposition faite par l'université au regard de l'offre de formation nationale et de la capacité de l'établissement à gérer et coordonner l'ensemble. Il s'agit du principe de l'accréditation mis en place avec la Loi du 22 juillet 2013 relative à l'enseignement supérieur et à la recherche (loi ESR). Ce processus renforce

ainsi l'autonomie pédagogique des établissements, qui se doivent cependant de respecter le cadre national des formations⁴⁵ qui précise les règles relatives à l'organisation des diplômes nationaux (licence, licence professionnelle et master) et fixe la liste des intitulés de ces diplômes.

Nous nous intéressons tout particulièrement à deux points inscrits dans le cadre national des formations prévu dans l'article L.613-1 du code de l'éducation et publié au journal officiel « Arrêté du 22 janvier 2014 fixant le cadre national des formations conduisant à la délivrance des diplômes nationaux de licence, de licence professionnelle et de master » : l'approche par compétence et la création de conseil de perfectionnement afin, d'étudier comment ces deux éléments sont importants dans la réflexion de l'enseignement de la statistique, mais aussi, de sa place dans la formation.

1. Approche par compétence

Dès l'article 4 du titre 1^{er} concernant les « dispositions communes aux diplômes nationaux de licence, de licence professionnelle et de master », il est indiqué que :

« L'organisation de la formation s'inscrit dans une logique d'apprentissage, permettant de prendre en compte les caractéristiques des étudiants en leur proposant des dispositifs pédagogiques qui favorisent la mise en activité, l'interaction avec les autres acteurs de la formation au-delà de l'équipe pédagogique, la mémorisation et la valorisation de leurs productions et réalisations, le développement de l'esprit critique, l'autonomie. Cette logique favorise la cohérence entre les unités d'enseignement, le décloisonnement des apprentissages afin de permettre à l'étudiant d'établir des liens au sein d'une même formation et entre ses expériences de formations. Elle incite les étudiants à mobiliser les savoirs et les compétences développés en formation dans de nouvelles situations. Les moyens pédagogiques mis en œuvre s'inscrivent dans cette logique d'apprentissage. »

La place de l'apprentissage dans le texte souligne l'importance à accorder à la mise en activité des étudiants de manière transversale à l'ensemble de la formation. Cette intégration de l'apprentissage dans les formations est en lien direct avec la définition de la transposition didactique proposée par Perrenoud (1998).

⁴⁵ Prévu dans l'article L.613-1 du code de l'éducation.

Mais, ce qui transparait dans cet article de l'arrêté est le changement de paradigmes que l'université vit. Comme le soulignent Chauvigné et Coulet :

« Le passage d'un paradigme fondé sur la transmission des savoirs académiques à un autre centré sur l'appropriation de ces savoirs et sur leur insertion dans des problématiques pratiques semble bien être, dans les universités, une des évolutions les plus remarquables des deux dernières décennies. » (2010)

Mouvement réformateur de l'éducation lancé en 1996 au Québec, le renouveau pédagogique se focalise sur la compétence comme levier de modernisation et de réussite des programmes scolaires. Ainsi, il s'est développé une approche par compétences (APC), opposée classiquement à la pédagogie par objectifs reposant sur des concepts clés, et dans laquelle s'inscrit la formalisation de référentiels de compétences.

D'ores et déjà, une nouvelle difficulté survient avec la définition de la compétence, notion centrale de l'APC. En effet, la littérature en présente une variété de définitions dont les perspectives théoriques diffèrent voire s'opposent parfois, mais notre thèse n'a pas pour objet de discuter précisément ce point de définition. L'absence de consensus sur le sens à donner au terme « *compétence* » est cependant source d'une confusion pouvant être minimisée par le choix d'une définition appropriée au cadre théorique. En l'espèce, l'APC est l'application pédagogique des théories constructivistes et de leurs prolongements socioconstructivistes, définissant l'apprentissage comme le processus actif de construction des connaissances mené par l'étudiant au fil d'activités, d'expériences et d'interactions avec son environnement. Nous retiendrons la définition de la compétence proposée par J. Tardif, en cohérence avec le choix qui a été fait au sein de l'Université de Bordeaux dans le cadre du processus d'accréditation. Cette définition nous paraît comme étant la plus pertinente au regard des fondements constructivistes de l'APC. Selon J. Tardif, la compétence

« est un savoir-agir complexe qui prend appui sur la mobilisation et la combinaison efficaces d'une variété de ressources internes et externes à l'intérieur d'une famille de situations. » (Tardif, 2006)

La mise en place de l'APC au sein des universités ne se fait pas sans mal ni réticences de la part des enseignants-chercheurs.

« Certains voient dans cette filiation un assujettissement de l'éducation aux intérêts économiques, au détriment d'une formation plus ouverte, critique, humaniste, citoyenne. » Chauvigné et Coulet (2010)

Cependant, les contenus des référentiels de compétences nous semblent être un bon marqueur de la place actuelle concédée à la statistique au sein des formations. De plus, ces éléments permettent de souligner le cadre de ce qui est attendu à l'issue de ces formations en tant que connaissances et savoir-faire, ainsi que l'état d'esprit dans lequel tout ceci est enseigné.

2. La place de la statistique dans les référentiels

Dans le cadre d'une approche par compétences, il est tout particulièrement important que les formations soient construites en tenant compte des connaissances antérieures des étudiants et des devenirs visés par la formation. Il est cependant utopique, et heureusement, d'imaginer que tous les étudiants arrivent avec les mêmes acquis et pour le même devenir. L'objectif est donc, tout principalement, de les aider à acquérir les notions statistiques ainsi que les démarches usitées dans le champ disciplinaire et les domaines visés par la formation. La notion de progression est néanmoins à prendre en considération. Nous nous intéressons ci-après tout particulièrement à la formation universitaire de niveau Licence.

a) *Liaison -3/+3, la connaissance des contenus de lycée*

L'article 15 du titre II « dispositions communes pour les diplômes nationaux de licence et de licence professionnelle »⁴⁶, souligne l'importance de la prise en compte de la continuité « *secondaire – supérieur* », nommé classiquement (-3/+3).

« La spécialisation progressive repose sur la mise en œuvre du principe de continuité, méthodologique et pédagogique, entre l'enseignement secondaire et l'enseignement supérieur. [...] L'architecture des parcours types en licence est élaborée par les établissements, au sein de leur offre de formation, conformément aux principes suivants :

- [...];
- Cette organisation repose sur des équipes pédagogiques pluridisciplinaires ;

⁴⁶ De l'« Arrêté du 22 janvier 2014 fixant le cadre national des formations conduisant à la délivrance des diplômes nationaux de licence, de licence professionnelle et de master »

— [...] Cette organisation se traduit donc par un socle commun défini en termes de compétences suffisantes pour permettre ces réorientations. »

Ainsi, l'enseignement de la statistique implique que les enseignants du supérieur aient connaissance des évolutions de contenu pédagogique des programmes de lycée afin d'assurer la continuité « -3/+3 » ; cette remarque est également vraie pour l'ensemble des disciplines enseignées durant le secondaire. L'enseignement de la statistique au lycée a été modifié en profondeur depuis les programmes de mathématiques de 2000. Ces programmes, initiés par le groupe technique disciplinaire en mathématiques sous la présidence de Claudine Schwartz-Robert, introduisaient une initiation à l'aléatoire et à la statistique inférentielle, en insistant, ainsi, sur l'importance de mettre en œuvre la pensée statistique.

La statistique a donc, petit à petit, pris sa place dans les cours de mathématiques, et dans les programmes. Ainsi, près de 20% du programme de mathématique de terminale S en 2012 concernent la partie « Probabilités et Statistique ». On peut souligner cependant que cette entrée en force ne s'est pas faite sans remous au sein du corps enseignant du secondaire de mathématique, dont la plupart des membres ne sont pas formés à cette discipline. Les contenus de programmes d'enseignement diffèrent selon la filière suivie durant le lycée, et il est donc difficile pour certaines formations de l'université, accueillant des bacheliers de tous types de bacs, de pouvoir définir les pré-requis communs de l'ensemble de leurs nouveaux étudiants. Les enseignants des premières années de l'université doivent, souvent, faire le choix de reprendre les notions de base en statistique en les contextualisant dans l'environnement disciplinaire de la formation. Les enseignants universitaires sont aussi parfois tout simplement ignorant des contenus actuels des programmes de lycée.

b) Le référentiel des mentions de licence

Le cadrage national des formations au niveau licence est accompagné, comme cela est indiqué dans l'article 3 de l'arrêté du 1er août 2011 relatif à la licence, de référentiels de compétences. Ces référentiels sont diffusés à titre indicatif, mais ils ont été élaborés en collaboration avec des représentants des différentes

communautés disciplinaires afin de pouvoir concevoir à Brest ou à Bordeaux des licences ayant des objectifs partagés et visant les mêmes compétences.

« Des référentiels de compétences sont définis pour une discipline ou un ensemble de disciplines à l'initiative du ministre chargé de l'enseignement supérieur, qui associe les conférences mentionnées à l'article L. 233-1 du code de l'éducation, les associations mentionnées à l'article L. 811-3 du code de l'éducation, les communautés scientifiques et les professionnels des secteurs concernés, et sollicite l'avis du comité mentionné à l'article 24.

La mise en œuvre des référentiels fait, au moins une fois tous les cinq ans, l'objet d'un examen par le comité mentionné à l'article 24 du présent arrêté. Cet examen contribue à l'évolution desdits référentiels.

La licence s'appuie sur des objectifs nationaux établis par les référentiels. Ceux-ci déclinent les compétences disciplinaires, linguistiques, transversales et préprofessionnelles que doivent acquérir les titulaires de la licence. » (article 3 de l'arrêté du 1er août 2011)

Sur le site du ministère de l'enseignement supérieur et de la recherche les référentiels de compétence sont, quant à eux, présentés de la manière suivante :

« Référentiels de compétences et non de formation, ils n'ont pas vocation à définir des contenus d'enseignement, des volumes horaires ni des méthodes pédagogiques, qui relèvent naturellement de l'autonomie des établissements et de la liberté pédagogique de leurs enseignants. En revanche, ils définissent des objectifs de formation, exprimés en termes de compétences à acquérir tout au long de la formation. »⁴⁷

L'article 6 précise l'organisation des référentiels de compétences en déclinaison en 4 types de compétences :

« La formation assure à l'étudiant l'acquisition d'un ensemble de connaissances et compétences diversifiées :

- des compétences disciplinaires, en premier lieu dans la ou les disciplines principales, mais aussi dans des disciplines connexes et, le cas échéant, dans des disciplines d'ouverture, afin de favoriser l'acquisition d'une culture générale ;
- des compétences linguistiques, se traduisant notamment par la capacité à lire, écrire et s'exprimer dans au moins une langue étrangère vivante ;
- des compétences transversales ou génériques, telles que l'aptitude à l'analyse et à la synthèse, à l'expression écrite et orale, au travail individuel et collectif, à la conduite de projets, au repérage et à l'exploitation des ressources documentaires, ainsi qu'au maniement des outils numériques ;

⁴⁷ Site internet du Ministère de l'éducation nationale, de l'enseignement supérieur et de la recherche
<http://www.enseignementsup-recherche.gouv.fr/cid61532/les-referentiels-de-competences-en-licence.html>

- des compétences préprofessionnelles, fondées sur la connaissance des champs de métiers associés à la formation, sur l'élaboration du projet personnel et professionnel de l'étudiant, ainsi que sur la capacité de ce dernier à réinvestir ses acquis dans un contexte professionnel. »

Un référentiel de compétences⁴⁸ pour le niveau licence a ainsi été publié par le ministère de l'éducation nationale, de l'enseignement supérieur et de la recherche en janvier 2015. Comme cela est indiqué dans le référentiel, le terme *compétence* étant polysémique, le choix d'une définition a été fait et devient ainsi la référence pour ce référentiel.

« Une compétence est l'aptitude à mettre en œuvre un ensemble organisé de savoirs, de savoir-faire et d'attitudes permettant d'accomplir un certain nombre de tâches. »

Afin d'étudier la place de la statistique au sein des compétences dans l'ensemble des mentions, nous avons réalisé une analyse des contenus de ce référentiel en étudiant les similarités et les dissemblances entre disciplines. Les six compétences suivantes, dites « transversales et linguistiques », sont communes aux 45 mentions.

- 1 utiliser les outils numériques de référence et les règles de sécurité informatique pour acquérir, traiter, produire et diffuser de l'information ainsi que pour collaborer en interne et en externe,
- 2 identifier et sélectionner diverses ressources spécialisées pour documenter un sujet,
- 3 analyser et synthétiser des données en vue de leur exploitation,
- 4 développer une argumentation avec esprit critique,
- 5 se servir aisément des différents registres d'expression écrite et orale de la langue française,
- 6 se servir aisément de la compréhension et de l'expression écrites et orales dans au moins une langue vivante étrangère.

La compétence « *analyser et synthétiser des données en vue de leur exploitation* » correspond à une compétence centrale de l'activité d'un statisticien. La place de cette compétence partagée par toutes les mentions est à voir comme un

⁴⁸ Ministère de l'Éducation nationale, de l'enseignement supérieur et de la Recherche, Référentiel de compétences des mentions de licences, janvier 2015

signal de l'importance de développer au sein des formations des enseignements et des apprentissages à l'analyse des données. Plus globalement, l'apprentissage de la démarche statistique permettrait de développer la capacité à construire un raisonnement en se fondant sur une question scientifique, la capacité à identifier et sélectionner diverses ressources spécialisées pour documenter un sujet et ainsi appréhender le contexte et planifier le recueil, la capacité d'analyse et de synthèse, puis la capacité à communiquer et restituer les résultats en langues française et étrangères, tout cela en développant une argumentation avec esprit critique. Il est dès lors assez clair que la statistique peut satisfaire la compétence 3 des compétences « transversales et linguistiques », mais aussi mettre en synergie les autres compétences précisées dans ce même socle⁴⁹.

Dans un second temps nous nous sommes penchée sur les compétences disciplinaires indiquées comme spécifiques pour chaque mention de licence et pouvant interpellier la statistique. La synthèse est présentée dans le tableau 5.

Pour 15 mentions parmi les 45, il n'y a pas de compétences disciplinaires s'appuyant précisément sur des notions, savoirs ou savoir-faire pouvant être développés dans le cadre d'un enseignement de statistique. Il s'agit des mentions : Administration publique - LAP / Droit / Gestion / Humanités / Histoire de l'art et archéologie / Philosophie / Théologie / Sciences du langage / Arts / Arts plastiques / Arts du spectacle / Lettres / Langues, littératures et civilisations étrangères et régionales / Langues étrangères appliquées / Lettres, Langues.

⁴⁹ Il est intéressant de souligner que ce terme n'a pas été employé dans ce référentiel suite aux réactions toujours assez véhémentes. Les compétences étant naturellement une construction individuelle, il est difficile de pouvoir réellement parler de bases communes : ce qui est commun à l'ensemble ici des licences, ce sont les éléments de savoir et de savoir-faire qui vont être intégrés au contenu pédagogique.

Tableau 5 : Place de la statistique dans les compétences disciplinaires indiquées dans le référentiel de compétences des mentions Licence de janvier 2015 – Ministère de l'enseignement supérieur et de la Recherche

Mention	Compétences disciplinaires liées à la statistique
Economie	Utiliser les <u>méthodes et outils d'observation</u> et <u>d'analyse</u> des phénomènes économiques (<u>qualitatives et quantitatives</u>)
Economie et Gestion	Utiliser les <u>méthodes et outils quantitatifs d'observation</u> et <u>d'analyse</u> de l'économie et des organisations
Administration économique et sociale	Utiliser les outils, méthodes et langages fondamentaux (<u>techniques quantitatives</u> et informatiques) appliquées aux sciences sociales
Science politique	<u>Rechercher, collecter et synthétiser</u> des informations sur les questions sociales et politiques
Sciences sanitaires et sociales	Mobiliser des données démographiques, sanitaires et <u>épidémiologiques</u> pour répondre à une demande de nature sociale et/ou sanitaire
Histoire	<u>Rassembler, mettre en forme et analyser</u> l'information historique au sein de documents de diverses natures (écrits, inventaires d'archives, iconographie, architecture, <u>statistiques...</u>) Utiliser les <u>outils spécifiques de l'étude des sources</u> d'information complexes (bibliothèques, ressources numériques, répertoires bibliographiques) ainsi que les <u>techniques d'enquête</u> dans le domaine
Géographie et aménagement	Mobiliser les modèles théoriques et méthodologiques propres à la discipline, <u>mobiliser des connaissances statistiques</u> , bibliographiques et cartographiques Réaliser la <u>collecte raisonnée de données</u> de terrain en utilisant les <u>outils de l'enquête</u> (<u>élaboration et dépouillement de schémas d'entretiens approfondis et de questionnaires</u>)

Mention	Compétences disciplinaires liées à la statistique
Sociologie	<p>Se servir aisément des <u>outils de l'enquête sociologique</u> (élaboration d'un questionnaire et d'un entretien approfondi passés en face à face ou en ligne ; <u>dépouillement des questionnaires, maîtrise de logiciels nécessaires à cet effet, usage des méthodes quantitatives et des approches démographiques</u>).</p> <p><u>Transcrire avec rigueur et régularité les enquêtes, établir un rapport complet et synthétique</u></p> <p><u>Lire des données statistiques</u> sous différentes formes, interpréter un tableau croisé, produire un document de synthèse de données statistiques</p>
Psychologie	<p><u>Mobiliser les bases méthodologiques et statistiques</u> pour l'analyse du comportement : objectif, problématique, hypothèse, méthode, analyse, implications</p> <p>Mettre en œuvre les principes fondamentaux des différentes méthodes en jeu dans la pratique de la discipline : méthodes d'observations, expérimentales et cliniques, méthodologies d'entretiens, <u>d'enquêtes, questionnaires, élaboration d'échelles</u>, pratique des tests, psychométrie et docimologie, dynamiques de groupe</p>
Sciences de l'éducation	<p>Mettre en œuvre les outils et méthodes d'analyse adéquats pour traiter une problématique d'éducation ou d'enseignement</p> <p><u>Mettre en œuvre un protocole de recueil de données</u></p> <p><u>Traiter les résultats par les méthodes quantitatives et/ou qualitatives</u></p> <p>Réaliser un rapport complet et synthétique</p>
Sciences sociales	<p>Définir une problématique d'enquête, <u>utiliser les outils nécessaires</u> (élaboration et dépouillement de questionnaire et <u>d'entretien</u>) et organiser une veille ou un observatoire de données</p>
Sciences de l'homme, anthropologie, ethnologie	<p>Construire un <u>protocole d'observation</u> : <u>élaboration d'un questionnaire et d'un entretien approfondi passés en face à face ou en ligne ; dépouillement, maîtrise de logiciels nécessaires à cet effet, usage des méthodes quantitatives et des approches démographiques</u></p>
Information – communication	<p>Choisir et appliquer une combinaison <u>d'outils, de méthodes et de techniques adaptés</u> (analyses sémiotiques, méthodes qualitatives et quantitatives en sciences humaines et sociales, techniques médiatiques, exploitation des ressources et architectures numériques, ingénierie documentaire), pour concevoir et réaliser des dispositifs d'information et de communication et en analyser la réception et les usages</p>

Mention	Compétences disciplinaires liées à la statistique
Informatique	<u>Caractériser les techniques de gestion de l'aléatoire</u> (probabilités et statistique) et leurs rôles dans le traitement de certaines données
Mathématiques et informatique appliquées aux sciences humaines et sociales (MIASH) Mathématiques	Se servir aisément des bases du raisonnement probabiliste ; mettre en œuvre une démarche statistique pour le traitement des données
Physique Chimie Physique, chimie Sciences de la vie Sciences de la Terre Sciences de la vie et de la Terre	Interpréter des données expérimentales pour envisager leur modélisation. <u>Valider un modèle par comparaison de ses prévisions aux résultats expérimentaux et apprécier leurs limites de validité.</u> Identifier les <u>sources d'erreur</u> pour <u>calculer l'incertitude</u> sur un résultat expérimental
Mécanique Génie civil Sciences pour l'Ingénieur	<u>Valider un modèle</u> par comparaison de ses prévisions aux résultats expérimentaux et <u>apprécier leurs limites de validité</u>
Sciences et techniques des activités physiques et sportives (STAPS)	Utiliser des outils et techniques d'analyse d'une activité physique (<u>collecte des données</u> par les outils adaptés, <u>traitement et interprétation des résultats</u>)
Electronique, énergie électrique, automatique	Utiliser des <u>outils</u> mathématiques (y compris le calcul numérique et matriciel) logiques et <u>statistiques</u> pour caractériser et piloter l'état et les tendances d'évolution d'un système <u>Valider un modèle</u> par comparaison de ses prévisions aux résultats expérimentaux, et <u>apprécier les limites de validité</u>
Sciences pour la santé	<u>Choisir et mettre en œuvre des outils théoriques permettant de s'appropriier les résultats des études expérimentales</u> (approches statistiques, par exemple)
Sciences et technologies	Se servir aisément des outils et méthodes de recueil, de traitement et d'analyse des données pour observer et analyser les phénomènes et/ou les comportements du sous domaine Valider un modèle par comparaison de ses prévisions aux résultats expérimentaux et apprécier les limites de validité

A la lecture des informations de ce tableau de synthèse, nous pouvons remarquer que la rédaction des compétences peut parfois se différencier uniquement par une information de contextualisation au sein d'une discipline : par exemple, nous

retrouvons « *Méthodes et outils d'observations* » déclinées pour des phénomènes économiques, de l'économie et des organisations, des sciences sociales, des questions sociales et politiques. Cette formulation suggère donc que les méthodes et outils d'observation peuvent être différents selon l'habitat dans lequel ils sont conviés. Cette information de contexte définit dès lors le type de méthodes sous-jacentes à l'usage de la statistique dans cette communauté. La liste de compétences ci-dessus ne définit pas directement le contenu de l'enseignement. C'est à chaque communauté ou équipe pédagogique de spécifier les éléments de savoir et savoir-faire à transmettre aux étudiants pour qu'ils soient capables de les mobiliser dans le cadre d'activités ou de missions professionnelles. Nous retrouvons donc au cœur de ces compétences des précisions concernant des étapes de la démarche statistique. Pour les disciplines de sciences humaines (sociologie, psychologie, sciences de l'éducation, sciences sociales et sciences de l'homme, anthropologie, ethnologie), les étapes les plus présentes sont les stratégies de recueil de données dans le cadre d'une approche de techniques d'enquête. Pour les disciplines de sciences et technologies, et sciences de la vie et de la terre, nous retrouvons plus spécifiquement l'approche modèle. Il faut aussi souligner que, pour ces seules disciplines, la notion d'erreur et d'incertitude est soulignée. Cela conforte ce que souligne Stengers : « *l'incertitude irréductible est la marque des sciences de terrains* » (Stengers, 1993) et particulièrement des sciences expérimentales.

Les spécificités disciplinaires que nous avons pu souligner lors de l'étude précédente des usages de la statistique dans les thèses se retrouvent à la lecture de ces compétences : une approche « techniques d'enquête » dans le cadre des disciplines se rapprochant du domaine des sciences humaines et sociales, et une « approche modèle » pour les disciplines du domaine sciences et techniques. Les usages lors de la thèse sont donc, comme cela avait été suggéré, influencés par l'ensemble de la communauté du doctorant, et donc par sa formation universitaire.

L'approche statistique est particulièrement vue dans le cadre de ce référentiel non pas en tant que démarche, mais comme une boîte à outils que l'on peut utiliser de manière différenciée, selon les besoins ou les attentes disciplinaires. Il est à souligner qu'une seule mention, la sociologie, souligne l'importance de savoir

lire des données et interpréter des tableaux. Ce point étant cependant nécessaire pour l'ensemble des autres activités statistiques, nous pouvons supposer que ce point est implicite, et que ces éléments seront réinterrogés dans les contenus, au bon vouloir des enseignants.

La capacité à synthétiser graphiquement les informations d'une étude est un autre élément non présent explicitement dans le référentiel, et peut être intégré de manière implicite dans l'usage du synthétiser. Cependant, suite à la synthèse de lecture des thèses, il est à souligner une certaine difficulté, que nous avons repérée en tant qu'enseignante de statistique, concernant la présentation des informations en représentation graphique. Ces dernières sont souvent surchargées, ne permettant pas une information synthétique, et manquant parfois des informations cruciales, telles que les effectifs ou parfois tout simplement les légendes. Le choix de l'esthétisme des couleurs ou des formes domine parfois par rapport à la qualité du contenu et du message devant être transmis. L'importance de développer des connaissances et des habiletés dans le choix et la mise en œuvre de représentations graphiques dans des buts d'explication et d'analyse des données n'est repérée dans le cadre du référentiel pour aucune des disciplines.

Les verbes d'action retrouvés dans les différentes compétences sont *rechercher, rassembler, recueillir, collecter, synthétiser, analyser, interpréter et rédiger*. Ces verbes d'action caractérisent certaines compétences et seront au centre d'objectifs pédagogiques de certaines formations disciplinaires. Ces verbes recouvrent ainsi les missions d'un statisticien, et il est intéressant de remarquer qu'au niveau licence nous pourrions signaler que seules les formations de « Mathématiques et informatique appliquées aux sciences humaines et sociales (MIASH) » et « Mathématiques » ont indiqué dans leur liste de compétences le fait d'accompagner leurs étudiants dans l'acquisition de savoir et savoir-faire pour le développement d'une démarche statistique. Ces deux mentions de licence proposent donc explicitement d'accompagner les étudiants vers un devenir possible de statisticien. Ainsi, la distanciation que nous avons signalée en début de thèse quant à la discipline statistique par rapport aux mathématiques ne semble finalement pas se

traduire clairement dans le cadre de l'offre de formation et des compétences associées.

B. Les acteurs de l'enseignement de la statistique à l'université

S'intéresser à l'enseignement de la statistique à l'université nécessite de se pencher sur le triangle didactique constitué par les étudiants, les enseignants et le savoir, sans oublier l'environnement de proximité de recherche et socio-économique.

1. La variété d'enseignants en statistique

Dans le cadre d'un enseignement de la statistique à l'université, trois profils d'enseignants peuvent intervenir, il s'agit d' :

- 1 enseignant ayant une réelle formation en statistique sur un parcours long, mais n'ayant pas nécessairement contextualisé leur pratique statistique dans l'environnement disciplinaire dans lequel il lui est demandé d'enseigner,
- 2 enseignant de formation mathématique ayant eu potentiellement très peu d'heures de formation en statistique mais une approche plus probabiliste, ayant ou non déjà une activité en interdisciplinarité avec le milieu disciplinaire de la formation,
- 3 enseignant de la discipline ayant une sérieuse pratique statistique dans le champ de la discipline, mais n'ayant pas eu un enseignement approfondi en statistique.

Ces enseignants ont dès lors des sensibilités à la statistique très différentes, plus ou moins théorisées et plus ou moins contextualisées. La diversité de référence des enseignants selon leur rattachement scientifique crée ainsi des spécificités épistémologiques qui ne sont pas clairement identifiées par les étudiants, car souvent non exprimées par l'enseignant. Prenons par exemple un enseignement intitulé « techniques de sondage ». Selon l'intervenant et le cadre de la formation, cet enseignement peut, entre autre, se fonder sur l'ouvrage de Tillé (2001), ouvrage

nécessitant des connaissances mathématiques ainsi qu'une aisance dans la manipulation de formules mathématiques, ou sur l'ouvrage « Initiation aux méthodes de sondage : application à la pharmaco-épidémiologie » que nous avons rédigé avec Annie Fourier Réglat, construit autour d'exemples contextualisés dans le champ de la pharmaco épidémiologie, avec uniquement les formules nécessaires pour les applications de ce domaine d'application (Fourier-Réglat and Jutand 2006).

Au-delà de l'intitulé même de l'enseignement il est donc important de pouvoir en préciser les contours, car les enseignements de statistique peuvent prendre trois directions : approche mathématiques, approche études de cas, approche logiciel plus applicative.

Les choix pédagogiques de l'enseignant sont nécessairement liés à la conception qu'il a aussi de l'identité de la discipline qu'il enseigne. Il peut être « simple » utilisateur de la statistique dans son contexte disciplinaire et ainsi avoir une relation utilitaire à la statistique, ou avoir une connaissance précise de l'objet manipulé dans une approche. L'enseignement de la statistique dans le cadre d'une formation disciplinaire, autre que statistique, nécessite cependant pour l'enseignant une double connaissance de la statistique et de l'environnement disciplinaire, afin de pouvoir réellement contextualiser l'enseignement de statistique et être en capacité d'aider l'étudiant à acquérir les éléments qui lui permettront de développer des compétences. Bien que, sur de nombreux points, la statistique puisse être comprise comme une discipline transverse, cela n'implique pas, comme nous avons pu le décrire précédemment, que l'ensemble des éléments soient aussi facilement transposables quel que soit l'environnement. L'analyse des thèses a ainsi pu mettre en évidence des spécificités disciplinaires.

Le fait que la formulation « *démarche statistique* » soit uniquement citée dans une des compétences de la mention licence de mathématiques souligne le fait, qu'en France, la statistique est encore considérée comme une branche des mathématiques. Cependant parallèlement à cela, comme le remarque Jeanne Fine (2012), les déclinaisons sectorielles de la statistique (statistique pour les biologistes, pour les financiers, ...) induisent un fort risque d'hyper-spécialisation et d'affaiblissement de l'identité de la discipline. La diversité des profils des

enseignants intervenant pour l'enseignement de la statistique témoigne ainsi ce point de potentielle crise identitaire de la statistique dans le cadre de l'enseignement.

Il est important de souligner que le parcours scolaire et professionnel de l'enseignant universitaire et ses activités de recherche sont des sources venant nécessairement nourrir son enseignement. Ainsi, chaque enseignant aura son propre style pédagogique, et les effets de transposition didactique interne seront de fait différents d'un intervenant à l'autre. Cette grande diversité peut être source de difficulté pour les étudiants qui n'identifient pas toujours « d'où parle » l'enseignant et qui, tout au long de son parcours, pourra se trouver confronter à des cadres d'enseignement de la statistique très différents abordant cependant des notions identiques. L'étudiant doit donc être en capacité de repérer les notions fondatrices de la statistique et les spécificités d'usages mais aussi de dénomination.

2. Favoriser les échanges entre enseignants

Il est ainsi important de pouvoir faire travailler en association les enseignants non statisticiens enseignant la statistique et des statisticiens, afin d'éviter une dilution de la statistique, non pour uniquement sauver cette discipline, mais bien pour assurer la qualité de sa diffusion et de son développement.

Le développement de l'usage du logiciel R au sein des formations aussi diverses que les formations de santé publique, de psychologie, de mécanique a permis depuis quelques années de faire se rencontrer au sein de colloques, de forum des doctorants, chercheurs et enseignants, spécialistes ou non de statistique. Ces rencontres ont pour objet un partage de pratique et de développement de « package », mais aussi une réflexion didactique de la diffusion de l'usage de ce logiciel. Créer ce type d'espaces d'échange autour d'un outil partagé peut devenir dès lors prétexte aux partages de connaissances entre les différents profils d'enseignants, et faciliter ainsi les interactions entre acteurs. Cette vision peut sembler naïve. Cependant, nous soutenons l'idée que la multiplication d'espaces physiques ou numériques de lieu d'échanges ouverts aux différents acteurs de l'enseignement de la statistique offrent l'opportunité de pouvoir faciliter la communication, et ainsi aider à la diffusion des savoirs et des savoirs faire disciplinaire et didactiques.

3. L'équipe pédagogique

Le travail d'enseignant dans le supérieur est principalement une activité individuelle (Berthiaume & Rege Collet, 2013, 17), cependant le développement de l'approche programme nécessite le développement du travail en équipe pédagogique, afin de construire la formation avec une cohérence interne, mais aussi une cohérence verticale pour le parcours de l'étudiant.

L'équipe pédagogique est alors responsable de l'organisation de la formation dans son ensemble ; les grandes lignes directrices sont partagées, mais le contenu d'un enseignement et son déroulement est toujours de la responsabilité de l'enseignant-chercheur, car *« ils jouissent, conformément aux dispositions de l'article L. 952-2 du code de l'éducation, d'une pleine indépendance et d'une entière liberté d'expression, sous les réserves que leur imposent, conformément aux traditions universitaires et aux dispositions du code de l'éducation, les principes de tolérance et d'objectivité. »*

L'évolution du travail en équipe pédagogique au sein des formations universitaires comme cela est fortement souhaité, peut permettre un travail de transposition partagée. Comme nous l'avons souligné précédemment lors de l'étude des thèses, les pratiques et usages diffèrent entre les thésards entre discipline, mais parfois aussi au sein d'une même discipline, chacun pouvant se référer plus ou moins à des spécificités épistémologiques. Afin de réduire les obstacles didactiques dans l'enseignement de la statistique, tout particulièrement dans le cadre de formations réalisées dans d'autres disciplines que la statistique, il est nécessaire que les différents enseignants intervenant puissent au minimum homogénéiser leurs écritures mathématiques s'il y en a.

Il ne s'agit en rien de vouloir uniformiser les pratiques, ce qui serait tout particulièrement contre-productif et inapproprié à la statistique qui doit être contextualisée pour l'apprentissage de sa mise en œuvre. Il ne s'agit donc pas de définir un socle commun aseptisé, mais bien d'ouvrir à la connaissance des autres pratiques, ainsi qu'à une approche épistémologique facilitant la connaissance des évolutions.

« La théorie des situations didactiques a permis de prévoir des obstacles didactiques en partant des obstacles épistémologiques historiques et de les expliquer. Nous en avons déduit que la conception d'une progression continue des enseignements par adjonction de savoirs avec le même vocabulaire et le même sens « *de la maternelle à l'université* » était contradictoire. Elle conduit à se limiter localement à des enseignements formels dont le sens incontrôlé crée des obstacles qui à terme conduisent à des échecs que nous avons identifiés et montrés. L'idée séduisante des « socles » qui pourraient être acquis définitivement, indépendamment de leur sens et des usages qui en seraient fait ultérieurement, relève de la même erreur. » (Brousseau, 2011)

4. Quid de la transposition didactique externe pour les formations de l'université ?

Comme nous l'avons indiqué précédemment, la question de la transposition didactique ne se pose pas précisément à l'université comme dans l'enseignement secondaire, et ceci pour plusieurs raisons déjà soulignées. Tout d'abord, les contenus pédagogiques des formations universitaires sont élaborés en toute autonomie d'une université à l'autre, et seuls des référentiels peuvent être des guides quant à l'élaboration des grandes lignes. La construction de l'offre de formation et les réajustements d'une année à l'autre méritent d'être discutés en équipe pédagogique et au sein du conseil de perfectionnement lorsqu'il existe.

L'article 5 du titre 1° « dispositions communes aux diplômes nationaux de licence, de licence professionnelle et de master » de l'arrêté du 22 janvier 2014 fixant le cadre national des formations conduisant à la délivrance des diplômes nationaux de licence, de licence professionnelle et de master souligne l'importance de mettre en place des espaces d'échange et de discussion permettant de faire évoluer la formation en considérant les contenus et les méthodes pédagogiques.

« Dans le cadre de la politique de l'établissement, des dispositifs d'évaluation sont mis en place pour chaque formation ou pour un groupe de formations, notamment à travers la constitution de conseils de perfectionnement réunissant des représentants des enseignants-chercheurs, des enseignants, des personnels bibliothécaires, ingénieurs, administratifs, techniciens, ouvriers, de service et de santé, des étudiants et du monde socioprofessionnel.

[...]

Ces dispositifs favorisent le dialogue entre les équipes pédagogiques, les étudiants et les représentants du monde socioprofessionnel. Ils éclairent les objectifs de chaque formation, contribuent à en faire évoluer les contenus ainsi que les méthodes d'enseignement afin de faciliter l'appropriation des savoirs,

des connaissances et des compétences et de permettre d'en améliorer la qualité. »

Dans la plupart des mentions de licences et de master des conseils de perfectionnement ont été institués. Ils sont le plus souvent composés d'enseignants-chercheurs, de professionnels du milieu socio-économique et d'étudiants et/ou diplômés. Le rôle de ce conseil est d'accompagner l'équipe pédagogique afin de faire évoluer les contenus et les méthodes, pour que l'enseignement soit en cohérence avec l'évolution de l'environnement professionnel ciblé par la formation, sans qu'il s'agisse de faire plier la formation aux exigences socio-professionnelles. Ce point est toujours très discuté au sein de l'institution, et il est important de rappeler que l'université n'est pas un institut de formation professionnelle.

De par la présentation des missions du comité de perfectionnement ainsi que sa composition, il peut être compris comme la noosphère d'une formation universitaire. Son positionnement est bien celui d'un sas dans lequel « *s'opère l'interaction entre le système d'enseignement et l'environnement sociétal.* » (Chevallard 1991, 24-5) : il est composé pour représenter les deux environnements académique et socio-professionnel. Il y est discuté des propositions de modifications et d'évolution de contenus ou de stratégies pédagogiques. Ce comité peut être considéré comme « *L'ensemble des personnes et groupes intéressés à la création et à la communication des savoirs d'un certain domaine* » (Brousseau, 1998, 341). Ce comité de perfectionnement cependant ne couvre pas la totalité des missions de la noosphère telle que présentée par Chevallard puisque l'enseignant-chercheur sera au final seul maître du contenu pédagogique.

« C'est elle (ndlr : la noosphère), dès lors, qui va procéder à la sélection des éléments du savoir savant qui, désignés par là comme « savoir à enseigner », seront alors soumis au travail de transposition ; c'est elle, encore, qui va assumer la partie visible de ce travail, ce qu'on peut appeler le travail externe de la transposition didactique, par opposition au travail interne, qui se poursuit, à l'intérieur même du système d'enseignement, bien après l'introduction officielle des éléments nouveaux dans le savoir enseigné. » (Chevallard, 1986)

Le conseil de perfectionnement est là pour accompagner principalement l'évaluation de la pertinence de l'organisation de la formation au regard de l'environnement ciblé comme habitat pour les futurs diplômés. Il s'agit d'analyser la

cohérence des savoirs à enseigner au regard des savoirs experts. Chevallard souligne que, dans le cadre de la noosphère, il est nécessaire que les compétences de chaque acteur soient « *finement délimitées, les registres assignés, les responsabilités départagées, les pouvoirs circonscrits* » (Chevallard 1982). Il est donc important que la composition du conseil de perfectionnement puisse être plurielle, avec des chercheurs, des représentants des habitats envisageables pour les devenirs des étudiants, ainsi que des enseignants et des étudiants et diplômés. Il s'agit dans cet espace de pouvoir confronter les différents points de vue, dans l'objectif de construire et développer la formation.

L'enjeu de ces conseils de perfectionnement pour l'enseignement de la statistique est de pouvoir déterminer les contenus devant impérativement être présents pour des raisons d'usages réguliers dans le cadre des habitats ciblés par l'offre de formation : il s'agit de pouvoir satisfaire à la demande ceci dans un but purement économique qui consiste à pouvoir permettre aux étudiants une intégration rapide sur le marché du travail. Plusieurs éléments sont donc plus spécifiquement discutés dans le cadre de ces conseils concernant l'enseignement de la statistique : les méthodes et modèles statistiques à présenter et les logiciels à utiliser. Cependant, un point important doit être systématiquement souligné et rappelé dans le cadre de ces discussions : il s'agit de la formation nécessaire non plus à des outils mais bien à la démarche statistique, seule façon de pouvoir armer l'étudiant face aux transformations incessantes des outils. Un étudiant ne sera bien formé en statistique que s'il a acquis le principe de la démarche statistique, les fondamentaux, et qu'il a développé son esprit statistique, les outils eux-mêmes n'étant que des moyens ponctuels à mettre en œuvre pour répondre à une question précise.

Ce point est selon nous essentiel, et ne peut transparaître dans la formation en statistique tout au long de la formation universitaire que s'il est repris par l'ensemble des enseignants intervenant en statistique.

C. Quelles stratégies pédagogiques pour enseigner la statistique ?

« Le rôle de l'enseignant du supérieur n'est plus de simplement exposer des notions reliées à son domaine d'expertise mais plutôt de concevoir des situations d'apprentissage lors desquelles les étudiants sont amenés à réfléchir aux notions présentées, à les utiliser, de façon à se les approprier » (Berthiaume, 2013)

Varié les stratégies pédagogiques est donc important dans l'enseignement supérieur afin de pouvoir associer des temps de diffusion de connaissance, de savoir-faire, mais aussi d'expérimentation contrôlée et non contrôlée. L'objectif étant avant toutes choses de pouvoir rendre l'étudiant autonome face à certaines pratiques, mais aussi face à certains problèmes clairement formulés ou non.

Selon Fine (Fine, 2012), « *il est clair que cette statistique (ndlr branche des mathématiques appliquées basées sur des principes probabilistes et ayant pour objet le traitement des données numériques) ne prépare pas à la pratique de la statistique de la même façon que les études théoriques de médecine ne préparent pas à la pratique de la médecine* » (ndlr) elles y contribuent. Il est nécessaire de mettre l'étudiant en activité dans des situations contextualisées afin de lui permettre de mobiliser les connaissances acquises et de développer ses propres compétences.

La formation à la statistique est un sujet qui a fait couler beaucoup d'encre quant aux curricula de ces formations mais de nombreux auteurs s'entendent sur la nécessité de construire une formation permettant d'aider les étudiants à acquérir une pensée statistique (Rumsey, 2002 ; Ben-Zvi, 2004).

1. Différentes approches pour enseigner la statistique ?

Après 22 ans d'enseignement de la statistique à l'université auprès de publics très variés, nous avons pu constater quatre obstacles didactiques majeurs en statistique qui sont :

- 1 les notations mathématiques et les formules – une symbolique et terminologie difficile,
- 2 la mise en relation entre un modèle et des données, entre le théorique et l'observé, entre l'abstrait et le concret,

- 3 la rédaction de la conclusion de tests statistiques, et la prise de décision, mais aussi la compréhension des résultats statistiques d'autres études, ce point induisant la capacité à développer un esprit statistique.
- 4 l'usage de logiciels ou de calculatrices.

Nous pourrions préciser chacun de ces éléments, qui ont été repris par différents auteurs, mais le point que nous soulignons ici concerne principalement la notion de diffusion, qu'il s'agisse de la capacité de compréhension des informations d'autrui ou de la diffusion de ses propres informations.

La statistique en tant que science du vraisemblable, incite l'usager à accepter de renoncer à l'usage de la vérité. Cette dimension dérange par son paradoxe qui est d'aider à décider en étant une science non basée sur la certitude. Il s'agit donc d'une science du risque, mais le risque est finalement pour qui ? Quel est le champ d'application de tel concept et de telle théorie, quel usage en fait le scientifique selon sa discipline ? Quel est le lien entre approche empirique et concept théorique : passage des statistiques à la probabilité ?

Tout cela mène à une réflexion ouverte sur l'ensemble des dispositifs pédagogiques concernant la statistique au niveau de l'université.

L'enseignement de la statistique au sein d'une formation peut ainsi être introduit selon différentes approches. Une de celles-ci peut être théorique plaçant la statistique comme une branche des mathématiques, accompagnée ou non de simulations (delMas, 1999 ; Yesilcay, 2000). Une autre approche peut consister en un enseignement appliqué à la discipline de référence de la formation suivie, et ceci selon différents ancrages ; il s'agit soit de partir de données, soit encore d'utiliser la résolution de problèmes, soit d'utiliser des analogies avec l'expérience de tous les jours. Armatte propose une typologie d'approches pédagogique de la statistique selon quatre approches (Armatte, 2010) :

- 1 une approche formelle hypothético-déductive, qui est d'ailleurs la plus fréquemment utilisée au sein de l'université française. Elle se décline toujours à peu près de la même façon, débutant par une approche de la statistique descriptive, puis un

enseignement de probabilité, pour atteindre les notions d'estimations et de tests statistiques d'hypothèse. Les exemples utilisés éludent la plupart du temps les questions associées au hasard, lissent tous les problèmes associés à l'étude de vraies situations. Cette méthode est peu encline à tester l'étudiant face à ses propres compétences ;

- 2 une approche expérimentale inductive, qui peut être initiée à partir de données réelles ou simulées, la simulation étant une démarche permettant de concevoir le concept de fluctuations d'échantillonnage. Cependant, cette approche expérimentale se doit d'être mise en relation avec une approche déductive afin de pouvoir repérer des régularités. La référence à des lois, des axiomes est bien nécessaire.
- 3 l'approche proposée par Fréchet (1955) qui consiste en une approche en deux phases permettant de corriger un biais de l'approche hypothético-déductive en débutant l'approche par des exemples, des données issues d'une activité sociale, et non plus des exemples pédagogiques construits, lissés de tout l'intérêt de travailler avec de vraies données, puis en complétant cela par une approche déductive.
- 4 enfin une approche historique, ayant pour intérêt de replacer les développements et les notions dans les cadres historiques, de pouvoir aussi en expliciter les régularités d'usage et les spécificités épistémologiques, et de pouvoir soulever les processus de raisonnement nécessaires à la création de notions.

La question concernant la place des approches 1 et 2, c'est-à-dire faut-il commencer par le théorique ou l'empirique, a conduit à de nombreuses controverses épistémologiques (Desrosières 2000). Cependant, il est dommage que la réflexion autour de ces différentes approches pédagogiques de la statistique est un élément trop rarement discuté entre enseignants de statistique à l'université, et pourtant comme Armatte le signale : « *puisque la didactique est plus un art et une philosophie qu'une science exacte, je milite clairement pour un pluralisme et un pragmatisme qui*

manifeste non seulement d'une épistémologie à la Feyerabend – « Tout est bon » – mais qui considère que nous avons tout à gagner à laisser l'apprenant choisir, en connaissance de cause, la pédagogie qui lui convient le mieux, et qu'aucune autorité ne peut nous imposer quelque chose en la matière. » (Armatte, 2010)

Ceci est d'ailleurs un leitmotiv dans tous les séminaires ou colloques de pédagogie universitaire « *il y a de nombreux profils d'apprenants, il faut donc diversifier nos stratégies pédagogiques.* » Permettre à l'étudiant d'être confronté à différents styles d'approche pédagogique lui offrirait l'opportunité durant son parcours de formation d'être éclairé sur les différentes stratégies et de prendre conscience de la proximité du développement de la statistique et de l'environnement social et politique. Ces relations ont été étudiées tout particulièrement par Desrosières et la lecture de son ouvrage « *La politique des grands nombres : histoire de la raison statistique* » mériterait d'être conseillée à tous les étudiants ayant pour projet d'avoir un usage de la statistique, et non uniquement en sociologie (2000).

2. Importance de former les étudiants aux notions fondamentales

Comme l'indique Gauvrit « la statistique a cette particularité extraordinaire d'être convaincante même pour des gens qui affirment ne pas la comprendre » (Gauvrit 2007). L'enseignant a donc pour mission de donner les éléments et des stratégies de compréhension. La capacité d'une lecture critique de l'information, quelle qu'elle soit grise, scientifique ou journalistique est une compétence nécessaire à l'issue d'une formation universitaire. Cette formation passe donc par l'apprentissage de notions fondamentales (caractéristiques de tendance centrale ou de dispersion ...), de savoirs statistiques classiquement connus (intervalle de confiance, tests d'hypothèse de comparaison de moyenne, chi deux ou régression), mais doit passer aussi par le développement du raisonnement statistique (Batanero, 2010 ; Gattuso, 2011; Scheaffer, 2006). Les représentations graphiques font aussi partie des éléments statistiques de description et d'analyse fondamentaux pour le développement d'un raisonnement statistique (Wild, 1999).

Nous avons pu souligner précédemment lors du focus sur les thèses qu'un certain nombre de maladresses existait, telles que des informations manquantes

(Hadjidemetriou, 2002 ; McClain, 1999) ou des représentations fausses (Pereira Mendoza, 1990; Lee, 2003; Bakker, 2004) ou des représentations surchargées ne permettant pas de dégager visuellement des éléments d'analyse marquant. Ces défauts sont des classiques selon les auteurs, mais cependant il est important de pouvoir mettre tout en œuvre pour y remédier.

Il s'agit donc pour l'enseignant d'être en capacité de faire comprendre aux étudiants de manière conjointe les bases du raisonnement déductif-probabiliste et du raisonnement inductif, mais aussi de les former à l'analyse contextualisée pour être en capacité de mettre en œuvre la démarche statistique adéquate (Fine, 2012).

La formation des éléments cités précédemment est indispensable pour être en capacité de réaliser une lecture critique d'article, car elle nécessite une bonne compréhension de la question posée et des moyens susceptibles d'être utilisés.

Les enjeux de la formation à la statistique à l'université sont donc importants car comme nous avons pu le souligner précédemment près de trois quart des thèses soutenues construisent une part de leur raisonnement sur des résultats statistiques et l'université se doit de développer la formation à la recherche et par la recherche.

3. **Former à la diffusion des démarches et résultats**

Développer les apprentissages de la démarche statistique à partir de données contextualisées et d'études de cas est une étape importante pour accompagner les étudiants dans l'acquisition des éléments qui les aideront à développer leur propre compétence. Un axe doit être tout particulièrement travaillé avec les étudiants : il s'agit de la communication et diffusion des savoirs et des résultats. Pour cette raison, il s'agit de travailler avec les étudiants non pas uniquement la mise en œuvre de procédures statistiques, mais aussi la formulation concernant la stratégie et les résultats statistiques (Batanero, 2007). Développer des habiletés didactiques chez les étudiants contribuent à l'amélioration de la communication en statistique à laquelle ils pourront être confrontés (Batanero, 2006).

Une des compétences soulignée dans le référentiel licence de sociologie, est la lecture des données et l'interprétation des tableaux croisés. Gal soulignait que cette compétence clé était finalement assez peu prise en considération (2002).

« The term 'statistical literacy' refers broadly to two interrelated components, primarily (a) people's ability to interpret and critically evaluate statistical information, data-related arguments, or stochastic phenomena, which they may encounter in diverse contexts, and when relevant (b) their ability to discuss or communicate their reactions to such statistical information, such as their understanding of the meaning of the information, or their concerns regarding the acceptability of given conclusions. These capabilities and behaviors do not stand on their own but are founded on several interrelated knowledge bases and dispositions. »

Evaluer la capacité des étudiants sur les deux dimensions que sont la production et la diffusion est un point assez facilement réalisable tout au long de l'enseignement de statistique dans le cadre d'études de cas, d'exercices pratiques ou de lectures critiques d'articles. Il est assez fréquent de rencontrer des étudiants n'ayant pas de difficulté majeure à manipuler des outils statistiques sophistiqués faisant appel à des programmations, et se trouvant pourtant en difficulté pour interpréter correctement même des éléments descriptifs simples.

Evaluer la compréhension réelle qu'un étudiant a d'une situation et d'analyses statistiques passe par l'évaluation de sa capacité à synthétiser et interpréter les résultats. Il s'agit de mettre en œuvre le travail de communication inhérent à la diffusion d'informations. Il est possible de leur présenter par exemple les vidéos réalisées par Rosling dans le cadre de Gapminder⁵⁰, pour leur présenter ce que peut permettre une diffusion pertinente de données statistiques.

Ce développement de compétence nécessite l'introduction au sein de l'enseignement d'éléments d'analyse réflexive (Régnier, 2002) faire travailler les étudiants sur l'explicitation de leurs étapes d'analyse afin de pouvoir les décrire et ainsi les expliquer à d'autres est réellement un travail trop peu souvent réalisé lors d'une formation.

⁵⁰ « organisation non gouvernementale à but non lucratif, dont le but est la promotion du développement durable et des objectifs du millénaire pour le développement de l'ONU. La démarche est originale : l'ONG veut y contribuer à sa manière, par la popularisation d'un outil de visualisation statistique » comme l'indique le site Internet, <http://www.gapminder.org/> .

Dans le cadre d'un enseignement en master 2 de sciences de l'éducation que nous assurons, nous avons proposé en 2014 de faire travailler les étudiants sur la rédaction d'un QCM. Il était demandé à chaque étudiant de rédiger un QCM de cinq questions liées à l'enseignement abordé tout au long des 24 heures de formation de statistique. Chaque question du QCM devait être accompagnée de quatre réponses possibles, dont une et une seule était valide. Le cadre ainsi posé, une séance de TD de 2 heures était consacrée aux échanges entre étudiants. Chaque étudiant proposait sa partie de QCM aux autres étudiants qui devaient dans un premier temps chercher à y répondre. Puis venait un temps de discussion autour de chaque question. Il s'agissait donc pour l'étudiant dans un premier temps d'une étape de décentration passant par la communication brute de son QCM. Puis les échanges argumentés des autres étudiants au sujet de chaque question et tout particulièrement quant à la formulation des réponses, lui permettait de recentrer son raisonnement. Ces échanges furent particulièrement riches et les étudiants ont pu appréhender au plus près leur capacité à retranscrire leur compréhension de notions statistiques, mais aussi leur capacité à rédiger des interprétations de résultats statistiques et à repérer les formulations correctes et comprises par les autres. Ce type d'activité a pu ainsi mettre à l'épreuve leur compréhension et leur capacité à rédiger et reconnaître les formulations correctes. Nous envisageons de reconduire cette expérience dans notre enseignement mais au-delà d'en étudier plus précisément les enjeux et les principes dans un objectif analytique pour notre propre recherche opérationnelle.

La statistique consiste principalement en la synthèse et l'analyse de données, dans de nombreux cas dans une visée de décision. *Recueillir, synthétiser, décider* sont des verbes d'action soulignant l'importance de l'interaction avec plusieurs acteurs, et, de fait, d'une compréhension partagée des processus et des enjeux. L'utilisateur de la statistique doit donc être en capacité évidemment de faire, mais aussi d'expliquer non pas uniquement les résultats obtenus mais surtout la démarche statistique dans laquelle ces résultats ont été obtenus, les résultats statistiques ayant la spécificité de devoir être compris comme résultats vraisemblables et non pas comme certitude.

Comme nous l'avons indiqué déjà de nombreuses fois la statistique est appelée au service d'autres disciplines pour parfois argumenter des choix, des décisions. En tant que science de la décision et cependant science devant accepter le statut d'incertitude, elle doit être réalisée en connaissance de cause et avec méthode et précision. Les éléments devant ainsi être particulièrement présents dans le cadre des enseignements de statistique ne sont pas uniquement des éléments de méthodes sophistiquées mais bien au contraire des travaux de rédaction, de compréhension d'article et de formulation pour des publics variés.

D. Synthèse

L'enseignement de la statistique à l'université niche dans des habitats bien différents et sous influence de communautés disciplinaires ayant leurs spécificités. Les contenus et stratégies pédagogiques dépendent de la discipline, mais aussi tout particulièrement des profils des enseignants et de l'environnement pédagogique.

Développer des lieux et temps d'échanges entre enseignants favorisant la confrontation et la diffusion des différentes approches permet d'accompagner l'étudiant dans sa compréhension des spécificités d'usage de la statistique qu'il sera amené à rencontrer tout au long de sa formation, et au-delà, parfois tout au long de sa vie professionnelle

Favoriser les temps de travaux rédactionnels et oraux, les temps d'interprétations de résultats statistiques d'articles scientifiques ou de productions personnels pouvant conduire à des confrontations entre étudiants sont autant de méthodes pédagogiques permettant de développer les aptitudes pour synthétiser, diffuser et communiquer, mais aussi pour développer une approche d'analyse réflexive de sa propre pratique. Le développement de la littérature statistique au sein des formations est un élément à promouvoir.

Le développement de l'enseignement de la statistique à l'université ne doit pas uniquement se fonder sur l'intégration de nouvelles méthodes sophistiquées mais doit avant tout favoriser la capacité des étudiants à expliquer leur activité statistique autant que leurs résultats.

CONCLUSION

Les différentes analyses de cette thèse ont pour objet de nourrir la réflexion sur l'enseignement de la statistique à l'université et éventuellement elles permettront d'envisager des stratégies pédagogiques à développer afin d'accompagner l'étudiant dans sa compréhension de l'usage de la statistique. Il s'agissait non seulement d'étudier la place de la didactique dans l'usage de la statistique mais de prendre en compte ses spécificités disciplinaires.

Reprenons les principaux éléments établis par la posture anthropo-didactique que nous avons adoptée dans ce travail.

Les doctorants ont un usage important d'outils statistiques qu'il s'agisse d'approche descriptive ou analytique pour développer les arguments de leur thèse. Selon les disciplines, des étapes de la démarche statistique seront plus ou moins développées, et nous avons montré, dans ce contexte, la prégnance des habitudes des diverses communautés disciplinaires : une approche modèle pour les disciplines en sciences et techniques, et une approche descriptive pour les sciences humaines. Le doctorant développe dans sa rédaction les étapes devant être considérées par la communauté comme essentielles ou plus spécifiques, les autres étapes restant parfois implicites car comprises tacitement par les membres de la communauté. Le doctorant choisit donc de réaliser un focus sur une partie particulièrement attendue, car "culturellement" sensible dans sa discipline. Ces spécificités disciplinaires apparaissent aussi dans les usages des termes que nous pourrions nommer "classiques" en statistique tels que : moyenne, corrélation, significativité, et dans l'usage de vocabulaire propre à une discipline pour identifier les notions partagées. Les difficultés lexicales sont donc doubles : les mots ont des sens différents et donc des usages différents, mais aussi, des mots différents sont utilisés pour désigner des notions similaires.

Si les implicites peuvent subsister dans le cadre des rédactions de thèse en direction d'une communauté disciplinaire, nous avons remarqué qu'il n'en est pas de même dans la rédaction des articles scientifiques (articles sur lesquels s'appuient

certaines thèses), ni sur les travaux d'expertise en direction du champ politique. L'explicitation, non seulement des résultats mais aussi de toute la démarche statistique pour les obtenir, est un élément fondamental pour conforter la vraisemblance et la validité des résultats. Le déroulement de l'audition analysée dans la thèse a mis en évidence la difficulté rencontrée par le scientifique pour expliciter les enjeux d'une approche statistique ne permettant pas de valider de façon binaire mais uniquement d'apporter des éléments que les responsables politiques devront mettre au regard de la demande sociale afin de prendre une décision. L'expert dans cette situation a fait un réel travail de transposition didactique afin de reformuler, mais surtout, de repositionner les objets et stratégies statistiques utilisés dans sa pratique dans la sphère politique pour que les décideurs puissent les appréhender dans l'usage qui est le leur et qui, de plus, doit contribuer à la prise de décision.

Les phénomènes qui apparaissent dans les situations que nous avons étudiées sont des projets didactiques au plein sens de Brousseau c'est-à-dire "un projet social de faire approprier par un sujet ou par une institution un savoir constitué ou en voie de constitution" (1998).

L'étude des enjeux liés à la diffusion des savoirs et usages de la statistique nécessite donc une compréhension des environnements et des acteurs destinataires. La transposition didactique ne doit pas être uniquement perçue comme le seul passage d'un "savoir savant" à un "savoir-enseigné" mais bien comme la nécessité d'intégrer les interfaces dans lesquelles doit se jouer la diffusion. La situation, au plein sens broussaldien, est l'instrument de cette transposition. Elle correspondrait alors dans cette perspective à ce que Bernard Sarrazy appelle une transposition didactique "horizontale", c'est-à-dire à une diffusion des savoirs entre diverses institutions, qui entretiennent entre elles des besoins de connaissance (cf. les situations d'expertises, d'auditions...), et non à une transposition "verticale" (du "savoir savant" au "savoir enseigné") au plein sens chevallardien du terme.

On comprendra alors, l'importance pour le citoyen de concevoir et d'appréhender les questions statistiques, il est, selon nous, de la mission de l'université de former à la capacité de concevoir le monde et d'y participer de ce point de vue. L'enseignant devrait donc équiper l'étudiant pour y participer et lui

fournir des clefs de sa compréhension. Développer une harmonie dans le cadre de son enseignement entre apport de connaissances et mise en œuvre contextualisée est nécessaire. La diversité de pratiques et de formulations nécessite que l'étudiant ait une bonne connaissance des notions fondamentales de la statistique pour pouvoir développer cette flexibilité dans l'usage (et donc le sens) en évitant un double écueil d'un positivisme étroit (focalisation sur le mot) ou à l'inverse d'un relativisme inconsistant (où tout se vaudrait).

La statistique ne peut s'exprimer que confrontée aux données qui elles-mêmes ne sont pas données mais produites selon une démarche qui anticipe un traitement et de leurs enjeux. Elle permet de construire des connaissances à partir de données en les synthétisant, en les analysant et en confrontant leur comportement à des modèles théoriques. Les difficultés de la démarche statistique sont donc multiples et des choix méthodologiques doivent nécessairement être faits par le statisticien, il est donc nécessaire qu'il soit en capacité de justifier, d'argumenter et d'explicitier ses choix, c'est selon nous cet aspect que doit viser l'enseignement de la statistique à l'université. Aujourd'hui cet aspect n'est pas vraiment présent dans les référentiels de formation ; on peut le regretter mais dans tous les cas cette absence souligne l'importance de positionner la didactique de la statistique comme partie prenante de la discipline statistique. Les étudiants doivent apprendre à traduire leur démarche et leurs résultats en prenant actes des institutions auxquelles ils s'adressent leur enjeu et les effets de cette diffusion.

RÉFÉRENCES

- Allès, B., Samieri, C., Féart, C., Jutand, M.-A., Laurin, D., & Barberger-Gateau, P. (2012). Dietary patterns: a novel approach to examine the link between nutrition and cognitive function in older individuals. *Nutrition Research Reviews* 25(2): 207-222.
- Allès, B., Samieri, C., Lorrain, S., Jutand, M.-A., Carmichael, P.-H., Shatenstein, B., & Barberger-Gateau, P. (2011). P026 Profils de comportement alimentaire des personnes âgées en Aquitaine et au Québec. *Cahiers de Nutrition et de Diététique* 46(Suppl 1): S64-S65.
- Armatte, M. (2006). Contribution à l'histoire des tests laplaciens. *Mathématiques et Sciences Humaines* (176): 117-133.
- Armatte, M. (2010). Le rôle de l'histoire dans l'enseignement de la statistique. *Statistique et Enseignement* 1(2): 23-47.
- Arteaga, P., Batanero, C., Contreras, J.M., & Canadas, G.R. (2015). Statistical graphs complexity and reading Levels: a study with prospective teachers. *Statistique et enseignement* 6(1): 3-23.
- Bachelard, G. (2004). *La formation de l'esprit scientifique: contribution à une psychanalyse de la connaissance*. Paris : Editions J. Vrin. 257 p.
- Bakker, A., Biehler, R., & Konold, C. (2004), Should young students learn about box plots? In: G. Burrill & M. Camden (Eds.), *Proceedings of the International Association for Statistical Education (IASE) Roundtable on Curricular Development in Statistics Education*. Auckland: IASE. pp. 163-173.
- Barberger-Gateau, P., Jutand, M.-A., Letenneur, L., Larrieu, S., Tavernier, B., Berr, C., & 3C Study Group. (2005). Correlates of regular fish consumption in French elderly community dwellers: data from the Three-City study. *European Journal of Clinical Nutrition* 59(7):817-825.
- Batanero, C. & Díaz, C. (2006). *Methodological and didactical controversies around statistical inference*. Actes du 36ièmes Journées de la Société Française de Statistique. [CD ROM]. Paris: Société Française de Statistique.
- Batanero, C., & Díaz, C. (2007). *Training future statisticians to teach statistics*. International Statistical Institute, 56th Session. 8 p.
- Ben-Zvi, D. & J. Garfield (Eds.) (2004), *The challenge of developing statistical literacy, reasoning and thinking*. Dordrecht :Kluwer Academic Publishers.
- Bensefa, L., Causse, E., Verret, C., Choudat, D., Brochard, P., Jutand, M. A., & Conso, F. (2006). Enquête française sur la Guerre du Golfe et ses conséquences sur la santé : resultants du bilan médical réalisé chez 1 008 vétérans. *Archives des Maladies Professionnelles et de l'Environnement* 67(2): 338-339.

- Berthiaume, D., & Rege Colet, N. (Eds.). (2013). *La pédagogie de l'enseignement supérieur: repères théoriques et applications pratiques*. Berne: Editions Peter Lang. 345 p.
- Bourdel-Marchasson, I., Rolland, C., Jutand, M.-A., Egea, C., Baratchart, B., & Barberger-Gateau, P. (2009). Undernutrition in geriatric institutions in South-West France: policies and risk factors. *Nutrition* 25(2): 155-164.
- Bourdieu, P. (1982). *Les rites comme actes d'institution*. Actes de la recherche en sciences sociales, 43(1), 58-63.
- Boyer, E. L. (1997). *Scholarship reconsidered: priorities of the professoriate*. New York: Wiley, Carnegie Foundation for the Advancement of Teaching. 160 p.
- Brousseau G. (1990). Le contrat didactique et le concept de milieu: Dévolution. *Recherches en Didactique des Mathématiques* 9(3) :309-336.
- Brousseau G. (2011). Théorie des situations didactiques en mathématiques. *Education & Didactique* 5(1):101-104
- Camiade, E., Gramond, C., Jutand, M.-A., Audignon, S., Rinaldo, M., Imbernon, E., & Lacourt, A. (2013). Characterization of a French series of female cases of mesothelioma. *American Journal of Industrial Medicine* 56(11): 1307-1316.
- Cazeneuve, B., Rivasi, M., & Lanfranca, C. (2001). *Rapport d'information par la Commission de la défense nationale et des forces armées en conclusion des travaux d'une mission d'information sur les conditions d'engagement des militaires français ayant pu les exposer, au cours de la guerre du Golfe et des opérations conduites ultérieurement dans les Balkans, à des risques sanitaires spécifiques*. Paris : Assemblée nationale. [Rapport d'information No. 3055].
- Chauvigné, C., & Coulet, J.-C. (2010). L'approche par compétences : un nouveau paradigme pour la pédagogie universitaire ? *Revue française de pédagogie. Recherches en éducation* (172): 15-28.
- Chevallard, Y. (1982). Pourquoi la transposition didactique ? Communication au *Séminaire de didactique et de pédagogie des mathématiques* de l'IMAG, Université scientifique et médicale de Grenoble. In : *Actes de l'année 1981-1982*. pp. 167-194.
- Chevallard, Y. (1985). *La transposition didactique - Du savoir savant au savoir enseigné* Grenoble: La pensée sauvage. 126 p.
- Chevallard, Y. (1986). La Transposition didactique: du savoir savant au savoir enseigné. *Revue Française de Pédagogie* 76(1): 89-91.
- Chevallard, Y. (1994). Les processus de transposition didactique et leurs théorisations. In *La transposition didactique à l'épreuve*, Grenoble : La pensée sauvage. pp. 135-180. Présenté à XIIe école d'été de didactique des mathématiques.
- Chevallard, Y., & Wosniak, F. (2003). Enseigner la statistique au secondaire. Entre genre prochain et différence spécifique.

- Cobb, G. W., & Moore, D. S. (1997). Mathematics, Statistics, and Teaching. *American Mathematical Monthly* 104(9): 801–823.
- Cousson-Gélie, F., Bruchon-Schweitzer, M., Dilhuydy, J. M., & Jutand, M.-A. (2007). Do anxiety, body image, social support and coping strategies predict survival in breast cancer? A ten-year follow-up study. *Psychosomatics* 48(3): 211-216.
- C. Verret, & Jutand, M.A. (2004). D1-6 L'Enquête française sur la guerre du Golfe et ses conséquences sur la santé. Contexte, mise en place et limites méthodologiques. *Rev Epidemiol Sante Publique* 52(2): 44-44.
- Dabis, F., & Desenclos, J.-C. (Eds) (2012). *Épidémiologie de terrain Méthodes et applications*. Paris: John Libbey Eurotext. 756 p.
- Deheuvels, P. (2000). La statistique : une science aux ramifications multiples. In : *La statistique*. Paris : Tec & Doc. pp 21-37.
- delMas, R., Garfield, J., and Chance, B. (1999), A model of classroom research in action: Developing simulation activities to improve students' statistical reasoning. *Journal of Statistics Education* [Online]
- Desabie, J. (1966). *Théorie et pratique des sondages*. Paris : Dunod. 181 p.
- Desrosières, A. (2000). *La politique des grands nombres : histoire de la raison statistique*. Paris: La Découverte. 456p.
- Dodge, H. F., & Romig, H. G. (1929). A method of sampling inspection. *The Bell System Technical Journal* 8(4): 613-631.
- Dutarte, P. (2011). Evolutions de la pratique statistique dans l'enseignement du second degré en France : du calcul statistique au développement de la pensée statistique. *Statistique et Enseignement* 2(1): 31-42.
- Féart, C., Jutand, M. A., Larrieu, S., Letenneur, L., Delcourt, C., Combe, N., & Barberger-Gateau, P. (2007). Energy, macronutrient and fatty acid intake of French elderly community dwellers and association with socio-demographic characteristics: data from the Bordeaux sample of the Three-City Study. *The British Journal of Nutrition* 98(5):1046-1057.
- Féart, C., Torrès, M. J. M., Samieri, C., Jutand, M.-A., Peuchant, E., Simopoulos, A. P., & Barberger-Gateau, P. (2011). Adherence to a Mediterranean diet and plasma fatty acids: data from the Bordeaux sample of the Three-City study. *The British Journal of Nutrition* 106(1): 149-158.
- Fine, J. (2012). Statistique, informatique, mathématiques et interdisciplinarité. *Statistique et Enseignement* 3(2): 33-59.
- Fine, J. (2013a). L'enseignement de la statistique en interdisciplinarité - Contribution aux débats. *Statistique et Enseignement* 3(2): 77-86.
- Fine, J. (2013b). Le rapport GAISE (U.S.) – Cadre d'un curriculum statistique de la maternelle à la Terminale. *Statistique et Enseignement* 4(1): 25-54.

- Fourrier-Réglat, A., & Jutand, M.-A. (2006). *Initiation aux méthodes de sondage: application à la pharmaco-épidémiologie*. Bordeaux: ARME-Pharmacovigilance.
- Franklin, C. Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for Assessment and Instruction in Statistics Education (GAISE)*. Report, a pre-K-12 curriculum framework. Alexandria, VA: American Statistical Association. 112 p.
- Frechet, M. (1955). *Les mathématiques et le concret*. Paris: PUF. 126 p.
- Gal, I. (2002). Adults' statistical literacy: Meanings, components, and responsibilities. *International Statistical Review* 70(1): 1-25.
- Gattuso, L. (2011). L'enseignement de la statistique: où, quand, comment, pourquoi pas? *Statistique et Enseignement* 2(1): 5-30.
- Gauvrit, N. (2007). *Statistiques, méfiez-vous!* Ellipses.
- Glass, G. V. (1976). Primary, Secondary, and Meta-Analysis of Research. *Educational Researcher* 5(10): 3-8.
- Groupe de travail sur les sondages en matière électorale. (2011). *Sondages et démocratie: pour une législation plus respectueuse de la sincérité du débat politique*. Paris: Sénat. [Rapport d'information No. 54].
- Hadjidemetriou C. & Williams J.S. (2001). Children's graphical conceptions: assessment of learning for teaching, *Proceedings of the 25th Conference of the International Group for the Psychology of Mathematics Education* 3: 89-96
- Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLoS Med* 2(8): e124.
- Johsua, S. (1996). Le concept de transposition didactique n'est-il propre qu'au mathématiques?, in Raisky, C. et Caillot, M. (dir) *Au-delà des didactiques, le didactique. Débats autour de concepts fédérateurs*. Bruxelles: De Boeck, pp. 61-73
- Johsua, S. (2003). *Introduction à la didactique des sciences et des mathématiques*. Paris: PUF.
- Jutand, M.-A., & Salamon, R. (2000). La technique de sondage par lots appliquée à l'assurance qualité (LQAS): méthodes et applications en sante publique. *Revue d'Épidémiologie et de Santé Publique* 48: 401-408.
- Jutand, M.-A., Le Strat, Y., & Ferley, J.-P. (2012). Techniques de sondage et calcul de la taille d'échantillon. In: *Épidémiologie de terrain: méthodes et applications*. Paris: J. Libbey Eurotext. pp 178-190.
- Kahane, J. P. (2000). *L'enseignement des sciences, rapport des sciences, rapport de la Commission de Réflexion sur l'enseignement des Mathématiques* (Editions Odile Jacob & CNDP). Paris: Ministère de l'Éducation nationale.
- Latour, B. (2008). *La vie de laboratoire: la production de faits scientifiques*. Paris: La Découverte. 300p.

- Lee, C., & Meletiou, M. (2003). Some difficulties of learning histograms in introductory statistics. In: *Proceeding of the Joint Statistical Meeting. Section on Statistical Education*. pp 2326-2333
- Lonka, K., Chow, A., Keskinen, J., Hakkarainen, K., Sandström, N., & Pyhältö, K. (2014). How to measure PhD. Students' conceptions of academic writing and are they related to well-being? *Journal of Writing Research*. pp 245—269.
- Malliavin, P. (Ed.). (2000). *La statistique*. Paris : Éditions Tec et Doc. [Rapport sur la science et la technologie n°8].
- McClain, K. (1999). Reflecting on students' understanding of data, *Mathematics Teaching in the Middle School* 4(6): 374-380.
- Murray, S., & Gal, I. (2002). Preparing for diversity in statistics literacy: Institutional and educational implications. In Phillips B. (Ed.). *Proceedings of ICOTS 6*. Cape Town: International Association for Statistical Education.
- Neyman J., Pearson E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference: part I. *Biometrika* 20A: 175–240
- Organisation Mondiale de la Santé. (1952). *Comité d'experts de l'administration de la Santé Publique - premier rapport*. Genève : OMS. [Rapports Techniques No. 55].
- Pavel, S., & Nolet, D. (2002). *Précis de terminologie*. Ottawa: Canadian Government Publishing.
- Penglaou, Ch. (1937). La statistique existe-t-elle en tant que discipline autonome ? *Journal de la société française de statistique* 78: 130-162
- Pereira-Mendoza, L., & Mellor, J. (1990). Student's concepts of bar graph: Some preliminary findings. In: D. Vere-Jones (Ed.). *Proceedings of the Third International Conference on Teaching Statistics* Voorburg: International Statistical Institute. pp 150-157.
- Perez, J.-M. (2010). La synthèse de textes dans l'enseignement supérieur, un objet à fort enjeu local : analyse comparative dans deux institutions, l'ENA et l'IUFM. *Éducation et Didactique* 4(2): 41–55.
- Perrenoud, P. (1998). La transposition didactique à partir de pratiques : des savoirs aux compétences. *Revue des Sciences de l'Éducation* 24(3): 487.
- Poitevineau, J. (2004). L'usage des tests statistiques par les chercheurs en psychologie: aspects normatif, descriptif et prescriptif. *Mathématiques et sciences humaines* 42 (167):5-25
- Portelli, H. et Sueur J.-P. (2010). Sondages et démocratie : pour une législation plus respectueuse de la sincérité du débat politique, rapport parlementaire : n°54
- Rabarijaona, L. P., Boisier, P., Ravaoalimalala, V. E., Jeanne, I., Roux, J. F., Jutand, M.-A., & Salamon, R. (2003). Lot quality assurance sampling for screening communities hyperendemic for *Schistosoma mansoni*. *Tropical Medicine & International Health* 8(4): 322–328.

- Rainville, C., Jutand, M.-A., & Bérard, J.-M. (2006). Plan expérimental des « études de cas multiples » : rationnel et méthode d'analyse des données. *Revue de Neuropsychologie* 16(2): 143-180.
- Régnier, J.-C. (2003). A 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* (22-23): 157-201.
- Régnier, J.-C. (2005). Formation de l'esprit statistique et raisonnement statistique. Que peut-on attendre de la didactique de la statistique ? In: Houdement, C., & Castela, C. (Eds). *Séminaire National de Didactique des Mathématiques*. Paris: Association pour la Recherche en Didactique des Mathématiques. pp 13-38.
- Reuter, Y., & Cohen-Azria, C. Eds. (2007). *Dictionnaire des concepts fondamentaux des didactiques*. Bruxelles: De Boeck. 272 p.
- Rogalski, J. et Samurçay, R. (1994). Modélisation d'un " savoir de référence " et transposition didactique dans la formation de professionnels de haut niveau, in Arzac, G., Chevillard, Y., Martinand, J.-L., Tiberghien, A. (dir.) *La transposition didactique à l'épreuve*, Grenoble, La Pensée sauvage Éditions, pp. 35-71.
- Rumsey, D. J. (2002). Statistical literacy as a goal for introductory statistics courses. *Journal of Statistics Education* 10(3): 1-12.
- Salamon, R., Verret, C., Jutand, M. A., Bégassat, M., Laoudj, F., Conso, F., & Brochard, P. (2006). Health consequences of the first Persian Gulf War on French troops. *International Journal of Epidemiology* 35(2): 479-87.
- Salamon, R. (2004). *Rapport de l'enquête française sur la guerre du Golfe et ses conséquences sur la santé*. Paris: La Documentation Française. 286 p.
- Salamon, R., Alperovitch, A., Conso, F., Jutand, M.-A., Boutin, J.-P., Coquin, Y., & Weinbreck, P. (2001). *Rapport de mission du groupe de travail chargé d'analyser les données sanitaires relatives aux anciens combattants français de la guerre du Golfe*. Paris: La Documentation Française. 140 p.
- Samieri, C., Jutand, M.-A., Féart, C., Capuron, L., Letenneur, L., & Barberger-Gateau, P. (2008). Dietary patterns derived by hybrid clustering method in older people: association with cognition, mood, and self-rated health. *Journal of the American Dietetic Association* 108(9): 1461-1471.
- Saporta, G. (2011). *Probabilités, analyse des données et statistique*. 3^{ième} édition. Paris: Ed. Technip. 656 p.
- Sarrazy, B. (2002). *Approche anthropo-didactique des phénomènes d'enseignement des mathématiques. Contribution à l'étude des inégalités scolaires à l'école élémentaire*. Note de synthèse pour l'habilitation à diriger des recherches, Université Victor Segalen Bordeaux 2, 162 p.
- Sarrazy B. (2005). – « Questions à la théorie anthropologique du didactique du point de vue de la théorie des situations et de l'anthropologie wittgensteinienne. » Actes du Premier Congrès International sur la *Théorie Anthropologique du Didactique*. Société, École et Mathématiques : Apports de la TAD. Baeza (Jaén – Espagne) du 27 au 30 octobre 2005 (A paraître).

- Scheaffer, R. L. (2006), Statistics and mathematics: On making a happy marriage. In: Burrill G. (Ed.), *NCTM 2006 Yearbook: Thinking and reasoning with data and chance*. Reston, VA: NCTM. pp 309-321.
- Schwartz, C. (2012). La preuve par les chiffres (evidence based) : de quoi s'agit-il ? *Statistique et Enseignement* 3(2): 3-21.
- Selz, M. (Ed.). (2013). *La représentativité en statistique*. Paris : INED. 136 p.
- Stengers, I. (1993). *L'invention des sciences modernes*. Paris: Flammarion. 211 p.
- Tardif, J. (2006). *L'évaluation des compétences : documenter le parcours de développement*. Montréal: Chenelière-éducation. 384 p.
- Tillé, Y. (2001). *Théorie des sondages : échantillonnage et estimation en population finie : cours et exercices avec solutions*. Paris: Dunod. 284 p.
- Verret, C., Jutand, M.-A., De Vigan, C., Bégassat, M., Bensefa-Colas, L., Brochard, P., & Salamon, R. (2008). Reproductive health and pregnancy outcomes among French gulf war veterans. *BMC Public Health* 8: 141.
- Verret, M. (1975). *Le temps des études* [Thèse de doctorat Lettres]. Lille: Atelier Reproduction des thèses. 2 vol. 937 p.
- Watson, J. M. (1997). Assessing Statistical Thinking Using the Media, In: Gal, I. & Garfield, J. B. (Eds.). *The Assessment Challenge in Statistics Education*. Amsterdam: International Statistical Institute. pp. 107-121
- Wild, C., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review* 67(3): 223-265.
- Wittgenstein, L. (1961). – *Tractatus logico-philosophicus* (Suivi des *Investigations philosophiques*). – [traduit de l'allemand par P. Klossowski]. Paris: Gallimard. 364 p.
- Wittgenstein, L. (1996). *Le Cahier bleu et le Cahier brun : Etudes préliminaires aux investigations philosophiques suivi de Ludwig Wittgenstein par N. Malcolm*, [traduit de l'anglais par G. Durand]. Paris: Gallimard, 1951. 423 p.
- Yesilcay, Y. (2000). *Research Project in Statistics: implications of a case study for the undergraduate statistics curriculum*. Journal of Statistics Education. 8 (2).

LISTE DES THESES ETUDIEES

Thèses bordelaises consultables sur theses.fr

- Abbas, A. (2013). Développement d'un dispositif pompe-sonde hétérodyne : application à l'imagerie en acoustique picoseconde.
- Abdi, A. (2013). Etude de la diversité neuronale au sein du Globus Pallidus : analyse neurochimique, électrophysiologique et manipulation optogénétique d'un sous-type neuronal chez le rongeur.
- Aime, A. (2013). Oligonucléotides amphiphiles et microARNs : mise en place de nanoplateformes à visée diagnostiques et thérapeutiques.
- Al Kaddissi, S. (2012). Comparaison de la réponse (en termes d'accumulation, d'impacts cellulaires et génétiques) de l'écrevisse *Procambarus clarkii* après exposition à un polluant métallique (cadmium) et un polluant radiologique (uranium 238 et 233).
- Aubert, A. (2012). *Analyse de défaillance de nouvelles technologies microélectroniques : nouvelles approches dans la méthodologie de préparation d'échantillon.*
- Beaucourt-Jacquet, C. (2012). *Étude expérimentale du guidage du faisceau d'électrons dans le cadre de l'allumage rapide de cibles de fusion.*
- Belles, A. (2012). *Développement et applications environnementales des échantillonneurs passifs pour la surveillance des écosystèmes aquatiques.*
- Benabed, A. (2012). *Les enjeux didactiques de la formation initiale et continue des enseignants d'anglais L2 en contexte algérien : entre objectifs institutionnels et attentes des enseignants.*
- Ben Othmen, A. (2013). *Tourisme de nature et financement de la préservation des sites naturels et aménités par leurs usagers : le cas de l'estuaire de la Gironde (France).*
- Ben Zakour, A. (2012). *Extraction des utilisations typiques à partir de données hétérogènes en vue d'optimiser la maintenance d'une flotte de véhicules.*
- Boyer, A. (2012). *Maîtrise de la résistance bactérienne : réflexions sur la phase empirique de l'antibiothérapie en réanimation.*
- Brandejsky, A. (2012). *Méthodes numériques pour les processus markoviens déterministes par morceaux.*
- Brodard, A. (2013). *Caractérisation thermique de structures de combustion par les effets de la chauffe sur les minéraux : thermoluminescence et propriétés magnétiques de foyers de la grotte des Fraux (Dordogne).*
- Campmas, E. (2012). *Caractérisation de l'occupation des sites de la région de Témara (Maroc) au Pléistocène supérieur et nouvelles données sur la subsistance des hommes du Paléolithique moyen d'Afrique du Nord :*

exemples des approches taphonomiques et archéozoologiques menées sur les faunes d'El Harhoura 2 et d'El Mnasra.

- Cohé, A. (2012). *Manipulation de contenu 3D sur des surfaces tactiles.*
- Cuenca, A. (2012). *Écoulements de fluides complexes dans des canaux sub-microniques .*
- Darwiche, A. (2012). *Microrhéomètre sur puce pour l'étude de l'écoulement d'un liquide proche d'une surface liquide.*
- Dasré, A. (2012). *Les mesures du regroupement spatial des populations : aspects méthodologiques et applications aux grandes aires urbaines françaises.*
- Denecheau, B. (2013). *Étude comparative de l'accrochage scolaire des enfants placés en France et en Angleterre : la suppléance familiale à l'épreuve de la question scolaire.*
- Diouf, A. M. (2013). *Medias et identité urbaine : la construction de l'idée de modernité dans les espaces urbains africains à travers la presse. le cas du Sénégal.*
- Dupin, C. (2013). *Etude du comportement mécanique des matériaux composites à matrice céramique de faible épaisseur.*
- Eddahech, A. (2013). *Modélisation du vieillissement et détermination de l'état de santé de batteries lithium-ion pour application véhicule électrique et hybride.*
- Foubert-Samier, A. (2013). *Capacités de réserve, vieillissement cérébral et maladie d'alzheimer.*
- Furlan, A. (2013). *Interactions entre les tannins et les lipides : impact possible sur le goût du vin.*
- Garcia, A. (2013). *Education et discipline au collège.*
- Gheorghe, A. (2013). *Détection robuste et précoce de l'embarquement et du grippage dans le système de commandes de vol.*
- Gheorghiu, A. (2012). *Comportements dangereux au volant chez les jeunes. Le rôle des mécanismes d'influence sociale dans le groupe.*
- Ghozlane, A. (2012). *Développement de méthodes bioinformatiques dédiées à la prédiction et l'analyse des réseaux métaboliques et des ARN non codants.*
- Godin, A. (2013). *Estimation sur des bases orthogonales des propriétés thermiques de matériaux hétérogènes à propriétés constantes par morceaux.*
- Hervé, G. (2012). *Datation par archéomagnétisme des terres cuites archéologiques en France au premier millénaire av. J.-C. : étalonnage des variations du champ géomagnétique en direction et intensité.*
- Jaquet, A. (2012). *Cancers du col de l'utérus et infection à VIH en Afrique de l'Ouest : Epidémiologie descriptive, déterminants et dépistage.*

- Le Cain, A. (2012). *Caractéristiques spatiales et temporelles d'une tache focale LMJ.*
- Landau, G.-J. (2012). *Les impacts de la démocratisation sur un secteur culturel : le cinéma sud-africain post-apartheid.*
- Leleux, A. (2012). *Contrôle non destructif de composites par ondes ultrasonores guidées, générées et détectées par multiéléments.*
- Mansuy, A. (2012). *Etude de la durabilité de cellules d'électrolyse de la vapeur d'eau à haute température : influence des paramètres de fonctionnement.*
- Michardière, A.-S. (2013). *Microélectrodes de nanotubes de carbone pour conversion d'énergie.*
- Monier, A. (2012). *Rôle des phosphatases acides dans le métabolisme et le stockage des sucres dans la vacuole chez Arabidopsis thaliana : approches in vitro et in silico.*
- Morello, A. (2012). *Génération et amélioration de protéines chimériques solubles associant un TCRg et la molécule pro-apoptotique FasL.*
- Ndiaye, A. (2012). *Communication, tourisme et développement durable au Sénégal : enjeux et risques.*
- Ofrim, D. M. (2012). *La Roumanie postcommuniste au prisme des théories de la transition démocratique.*
- Querbes-Revier, A. (2012). *Émergence et évolution de l'industrie des services numériques pour téléphones mobiles.*
- Quotb, A. (2012). *Méthodes et systèmes pour la détection adaptative et temps réel d'activité dans les signaux biologiques.*
- Ruet, A. (2012). *Prédiction du diagnostic et du pronostic aux stades précoces de la Sclérose en Plaques.*
- Santagata, C. (2012). *L'utilisation de roches autres que le silex au Paléolithique ancien et moyen : choix économiques, techniques et fonctionnels sur la base de l'étude des gisements de Sainte-Anne 1 (Haute-Loire, France) (MIS 5 et 6) et Notarchirico (Basilicata, Italie) (MIS 14 à 17).*
- Sergent, A. (2013). *La politique forestière en mutation : une sociologie politique du rapport secteur - territoire.*
- Tandjigora, A. K. (2012). *L'évolution économique et sociale comparée de deux régions sénégalaises dans le processus de colonisation, décolonisation et développement : le boundou et le gadiaga, 1885-1980.*
- Tissot, A. (2012). *Une abbaye de renom à l'époque moderne : l'Abbaye aux Dames de Saintes (fin du XVe siècle - début XIXe siècle).*
- Valderrama Ulloa, C. (2013). *Exploitation et réhabilitation de bâtiments tertiaires : une démarche énergétique d'amélioration environnementale.*
- Zimmer, A. (2013). *Etude du déterminisme génétique de la phase de latence chez Saccharomyces cerevisiae en conditions œnologiques. Impact des mécanismes de résistance au SO₂.*

INDEX

	Latour,19, 29, 35, 94
	Lonka, 95
A	
Ardilly,..... 83	
Armatte,67, 206, 207	
B	
Bachelard,68, 69, 152	
Batanero,208, 209, 217	
Ben-Zvi, 205, 217	
Berthiaume,201, 205, 218	
Bourdieu, 22	
Brousseau, ... 30, 32, 62, 70, 71, 153, 202, 203, 215	
C	
Chauvigné, 187, 218	
Chevallard, 12, 14, 15, 18, 30, 31, 34, 37, 38, 183, 203, 204	
Cochran, 65	
Coulet, 187, 218	
D	
Deheuvels,58, 59, 62, 63, 219	
Desabie, 81, 219	
Desrosières,16, 57, 58, 207, 208	
Dodge, 42, 61	
F	
Fine, 62, 68, 69, 73, 199, 205, 209	
Fisher,66, 67, 112, 133, 134	
Fourrier-Réglat,83, 199, 220	
Franklin, 90, 220	
Fréchet, 207	
G	
Gal,210, 220, 221, 223	
Garfield, 217, 223	
Gattuso,68, 72, 89, 208	
Gauvrit, 90, 208	
Glass, 65, 220	
J	
Johnsua, 28,31	
L	
Landau, 107, 226	
M	
Malliavin,52, 57, 58	
N	
Neyman,66, 67, 84, 133	
Nolet, 102, 103	
P	
Pavel, 102, 103	
Pearson,66, 67, 84, 133	
Penglaou, 61	
Perez, 38	
Perrenoud,27, 31, 33, 34, 186	
Poitevineau, 87, 88	
R	
Rainville, 76	
Rege Collet, 201	
Régnier, .2, 20, 21, 59, 60, 69, 70, 73, 152, 183, 210	
Rosling, 210	
Rumsey, 205	
S	
Salamon, 3, 12, 14, 156, 160, 163, 168, 180	
Saporta, 84, 222	
Sarraza,2, 15, 26, 215	
Schwartz,66, 67, 189, 223	
Spearman,40, 41, 135	
Stengers,17, 153, 196	
T	
Tardif, 99, 187	
Tillé,81, 82, 198	
V	
Verret, 30	
W	
Watson, 185	
Wittgenstein,32, 56, 71, 181	
Woolgar, 29	
Wozniak, 18, 183	

TABLE DES MATIERES

REMERCIEMENTS.....	2
RESUME - ABSTRACT.....	4
SOMMAIRE	7
LISTE DES ACRONYMES	9
AVANT-PROPOS.....	11
INTRODUCTION	16
CHAPITRE 1 – SITUATIONS DIDACTIQUES DE LA STATISTIQUE POUR L’ENSEIGNANT- CHERCHEUR	20
A. MISSIONS DE L’ENSEIGNANT-CHERCHEUR : LA PLACE DE LA DIDACTIQUE.....	21
1. <i>Différents temps didactiques dans la vie de l’enseignant-chercheur</i>	22
2. <i>Différentes sphères didactiques pour l’enseignant-chercheur</i>	26
B. QUELLES SPECIFICITES DE LA TRANSPOSITION DIDACTIQUE POUR LES MISSIONS DE L’ENSEIGNANT- CHERCHEUR ?	30
1. <i>Cadre général.....</i>	30
2. <i>Déroulement et spécificité</i>	34
C. POURQUOI PARLER D’ÉCOLOGIE DIDACTIQUE DES SAVOIRS POUR LA SCIENCE STATISTIQUE ? ...	37
1. <i>Écologie didactique des savoirs.....</i>	38
2. <i>La statistique dans des environnements différents</i>	39
3. <i>Des situations de changement d’habitats rencontrés.....</i>	42
4. <i>Catégorisation de statisticiens</i>	52
D. SYNTHESE.....	55
CHAPITRE 2 – LA STATISTIQUE : UNE DISCIPLINE BIEN SPECIFIQUE.....	56
A. LA RECONNAISSANCE DISCIPLINAIRE DE LA STATISTIQUE	56
1. <i>"Statistique" : de quoi parle-t-on ?.....</i>	56
2. <i>Des données à la démarche</i>	57
3. <i>La statistique : une discipline</i>	60
B. LE CADRE DE LA STATISTIQUE	63
1. <i>La place de la statistique dans certains développements de théorie scientifique.....</i>	64
a) Des données à la théorie scientifique	64
b) Modèle hypothético-déductif	66

2.	<i>Démarche statistique</i>	68
3.	<i>Culture statistique</i>	70
C.	LES NOTIONS FONDATRICES	72
1.	<i>Pourquoi la variabilité est-elle l'élément incontournable de la statistique ?</i>	73
a)	Cadre général	73
b)	Un exemple de problème de variabilité – les études de cas multiples en neuropsychologie	76
2.	<i>Vous avez dit représentatif !</i>	79
a)	Quel sens est donné à ce terme représentativité ?	80
b)	Préciser plutôt les processus de sélection	82
3.	<i>La significativité : sésame de la vérité</i>	84
a)	Probabilités d'erreur	85
b)	Niveau de signification	86
c)	Remarques	87
D.	LA PLACE DE LA STATISTIQUE DANS LA COMPREHENSION DU MONDE !	88
1.	<i>Communiquer</i>	89
2.	<i>Place de l'apprentissage pour former le citoyen</i>	90
E.	SYNTHESE	92
CHAPITRE 3 – DIFFERENTES SITUATIONS DE LA DIDACTIQUE DE LA STATISTIQUE		93
A.	LA STATISTIQUE AU SERVICE DE LA RECONNAISSANCE D'UN JEUNE CHERCHEUR PAR UNE COMMUNAUTE DE CHERCHEURS	93
1.	<i>Choix du champ de recherche</i>	94
2.	<i>Se faire reconnaître dans sa communauté</i>	95
a)	Premier pas : la thèse	95
b)	Les autres étapes de reconnaissance	96
c)	Des reconnaissances interconnectées	97
3.	<i>La place de la statistique dans les thèses</i>	99
a)	Méthodes d'exploration	100
(1)	Terrain d'étude	100
(2)	Quelle sélection de thèses ?	101
(3)	Choix de l'algorithme de recherche de mots	103
4.	<i>Analyse de corpus</i>	106
a)	Sélection des thèses pour l'analyse de corpus	106
b)	Doctorants et statisticiens, quel partenariat ?	108
c)	Place et rôle de la statistique	110
d)	"Statis" dans tous ses états	111
e)	Quel niveau de précision dans la présentation du recueil de données ?	113
f)	Vous avez dit échantillonnage	117

g)	Les différents outils d'analyse statistiques.....	120
(1)	Moyenne.....	121
(2)	Mesure de la variabilité	124
(3)	Mais qui est Gauss ?	127
(4)	Méthodes descriptives et exploratoires	128
(5)	Intervalle de confiance	132
(6)	Tests statistiques,	133
(1)	Outils de modélisation.....	136
(2)	Régression linéaire et corrélation.....	138
(3)	Avec quel moyen technique – logiciel ?.....	140
h)	Quelles décisions avec les données ?.....	142
(1)	L'argumentation du soi-disant non résultat	143
(2)	Prise de décision	143
(3)	Utilisations du terme significatif	145
i)	Formulation du discours.....	148
5.	<i>Discussion</i>	148
6.	<i>En conclusion</i>	152
B.	LE CHERCHEUR DANS UNE SITUATION D'EXPERT ET DE DIDACTICIEU DES POLITIQUES : CONSEQUENCES DE LA GUERRE DU GOLFE (1990 – 1991) SUR LA SANTE DES MILITAIRES FRANÇAIS.....	154
1.	<i>Contexte</i>	155
2.	<i>Auditions dans le cadre des missions concernant les conséquences de la guerre du Golfe 1990 - 1991</i>	156
a)	La mission du groupe d'experts	158
b)	Le scientifique : caution de la qualité du discours et des décisions	159
c)	Comment définir le syndrome de la guerre du Golfe ?	162
d)	Argumenter des propositions de stratégies d'enquête	165
e)	Argumenter les choix méthodologique pour une aide à la décision	166
3.	<i>Article "Health consequences of the first Persian Gulf War on French troops"</i>	169
4.	<i>Discussion</i>	179
C.	SYNTHESE.....	183
CHAPITRE 4 – ENJEUX DE L'ENSEIGNEMENT DE LA STATISTIQUE A L'UNIVERSITE		185
A.	L'UNIVERSITE ET L'APPROCHE COMPETENCE	185
1.	<i>Approche par compétence</i>	186
2.	<i>La place de la statistique dans les référentiels</i>	188
a)	Liaison -3/+3, la connaissance des contenus de lycée.....	188
b)	Le référentiel des mentions de licence	189

B.	LES ACTEURS DE L'ENSEIGNEMENT DE LA STATISTIQUE A L'UNIVERSITE	198
1.	<i>La variété d'enseignants en statistique</i>	198
2.	<i>Favoriser les échanges entre enseignants</i>	200
3.	<i>L'équipe pédagogique</i>	201
4.	<i>Quid de la transposition didactique externe pour les formations de l'université ?</i>	202
C.	QUELLES STRATEGIES PEDAGOGIQUES POUR ENSEIGNER LA STATISTIQUE ?.....	205
1.	<i>Différentes approches pour enseigner la statistique ?</i>	205
2.	<i>Importance de former les étudiants aux notions fondamentales</i>	208
3.	<i>Former à la diffusion des démarches et résultats</i>	209
D.	SYNTHESE.....	213
	CONCLUSION	214
	RÉFÉRENCES	217
	LISTE DES THESES ETUDIÉES	224
	INDEX	227
	TABLE DES MATIERES	228
	LISTE DES TABLEAUX	232
	LISTE DES FIGURES	233
	LISTE DES SCHEMAS	234
	LISTE DES PUBLICATIONS M.-A. JUTAND.....	235

LISTE DES TABLEAUX

Tableau 1 : Les sept étapes d'un test statistique	33
Tableau 2 : Probabilités d'erreur de décision.....	85
Tableau 3 : Répartition des thèses soutenues entre le 1 ^o janvier 2012 et le 31 décembre 2013 dans une des écoles doctorales des universités de Bordeaux et de l'échantillon d'étude	102
Tableau 4 : Répartition des thèses sélectionnées pour l'étude selon leur appartenance à des écoles doctorales des universités de Bordeaux et la place de la statistique dans le corpus.....	108
Tableau 5 : Place de la statistique dans les compétences disciplinaires indiquées dans le référentiel de compétences des mentions Licence de janvier 2015 – Ministère de l'enseignement supérieur et de la Recherche.....	193

LISTE DES FIGURES

Figure 1 : représentation barres d'erreurs et histogramme – p60 (Zimmer, 2013) (B2 – Œnologie).....	123
Figure 2 : diagramme en secteur – p214 -(Ndiaye, 2012) (B3 - Sciences de l'Information et de la Communication).....	129
Figure 3 : exemple de "diagramme en secteur en 3D" – reproduction p 81(Abdi, 2013) (B2 – Neurosciences).....	129
Figure 4 : exemple de représentation ternaire - reproduction p156 (Campmas, 2012) (B1 - Préhistoire et géologie du Quaternaire).....	130
Figure 5 : quelle amplitude pour les intervalles? - reproduction p259 -(Garcia, 2013) (B2 – Sociologie).....	130
Figure 6 : exemple de boites à moustache - reproduction p161 - (Belles, 2012) (B1 - Chimie analytique).....	131
Figure 7 : exemple d'histogramme - reproduction p83 (Ghozlane, 2012) (B1 – informatique).....	132
Figure 8 : exemples d'ellipses de confiance et interprétation - reproduction p159 (Campmas, 2012) (B1 - Préhistoire et géologie du Quaternaire).....	133
Figure 9 : exemple 1 de représentation d'un nuage de points en 3 dimensions (p160) (Belles, 2012) (B1 - Chimie analytique).....	139
Figure 10 : exemple de représentation d'une régression linéaire (p263) (Belles, 2012) (B1 - Chimie analytique).....	140
Figure 11 : exemple de sortie d'écran de logiciel non retravaillé (Gheorghiu, 2012) (p93- B2 – psychologie).....	142
Figure 12 : exemple de décision de test - (Gheorghe, 2013) (p80 - B1 - Automatique et productique, signal et image).....	145
Figure 13 : reproduction p97 – fréquence – test (Quotb, 2012) (B1 – Electronique).....	146
Figure 14 : circuit des échanges et de la communication lors de la mise en place des missions.....	157

LISTE DES SCHEMAS

Schéma 1 : Différentes situations didactiques dans les missions d'un enseignant-chercheur	25
Schéma 2 : cheminement de l'objet savoir d'un enseignant-chercheur	36
Schéma 3 : Conception de la statistique selon Régnier (Régnier, 2003)	60
Schéma 4 - Cheminement dans le cadre de la thèse.....	98

LISTE DES PUBLICATIONS DE M.-A. JUTAND

2013

1. Camiade, E., Gramond, C., **Jutand, M.-A.**, Audignon, S., Rinaldo, M., Imbernon, E., & Lacourt, A. (2013). Characterization of a French series of female cases of mesothelioma. *American Journal of Industrial Medicine* 56(11): 1307 - 1316.
2. Harjaningrum, A. T., Kartasasmita, C., Orne-Gliemann, J., **Jutand, M.-A.**, Goujon, N., & Koeck, J.-L. (2013). A qualitative study on knowledge, perceptions, and attitudes of mothers and health care providers toward pneumococcal conjugate vaccine in Bandung, West Java, Indonesia. *Vaccine* 31(11): 1516 - 1522.
3. Le Roux-Goglin, E., Dubus, P., Asencio, C., **Jutand, M.-A.**, Rosenbaum, J., & Mégraud, F. (2013). Hepatic lesions observed in hepatitis C virus transgenic mice infected by *Helicobacter hepaticus*. *Helicobacter* 18(1): 33 - 40.

2012

4. Allès, B., Samieri, C., Féart, C., **Jutand, M.-A.**, Laurin, D., & Barberger-Gateau, P. (2012). Dietary patterns: a novel approach to examine the link between nutrition and cognitive function in older individuals. *Nutrition Research Reviews* 25(2): 207 - 222.
5. Ekouevi, D. K., Kariyare, B. G., Coffie, P. A., **Jutand, M.-A.**, Akpadza, K., Lawson-Evi, A., & David, M. (2012). Feasibility and acceptability of rapid HIV screening in a labour ward in Togo. *Journal of the International AIDS Society* 15(2): 17380.
6. El Rhazi, K., Boly, A., Samieri, C., **Jutand, M.-A.**, Nejjari, C., & Barberger-Gateau, P. (2012). Comportements alimentaires. *Revue d'Epidémiologie et de Santé Publique* 60 : S129 - S130.
7. **Jutand, M.-A.** & Savès M. (2012). « Chapitre 3 - Quelles statistiques ? Rapports, proportions, taux et indices ». In: Dabis F. & Desenclos J.-C (Eds) « *Epidémiologie de terrain : Méthodes et Applications* ». Montrouge: John Libbey Eurotext. pp. 19 - 27.
8. **Jutand, M.-A.** & Savès M. (2012). « Chapitre 4 – Distributions statistiques et leurs paramètres ». In: Dabis F. & Desenclos J.-C (Eds) « *Epidémiologie de terrain : Méthodes et Applications* ». Montrouge: John Libbey Eurotext. pp. 28 - 38.
9. **Jutand, M.-A.**, Le Strat Y. & Ferley J.-P. (2012). « Chapitre 22 – Techniques de sondage et calcul de la taille d'échantillon ». In: Dabis F. & Desenclos J.-C (Eds) « *Epidémiologie de terrain : Méthodes et Applications* ». Montrouge: John Libbey Eurotext. pp. 178 - 90.
10. **Jutand, M.-A.**, Fagot Campagna A, & Ferley J.-P. (2012). « Chapitre 23 – Recueil des données épidémiologiques lors d'une enquête sur le terrain. ». In: Dabis F. & Desenclos J.-C (Eds) « *Epidémiologie de terrain : Méthodes et Applications* ». . Montrouge: John Libbey Eurotext. pp. 191 - 205.

11. **Jutand, M.-A.**, & Fagot Campagna A. (2012). « *Chapitre 24 – Organisation d'une enquête épidémiologique sur le terrain.* ». In: Dabis F. & Desenclos J.-C (Eds) « *Epidémiologie de terrain : Méthodes et Applications* ». Montrouge: John Libbey Eurotext. pp. 201 – 18.
12. **Jutand, M.-A.**, & Savès M. (2012). « *Chapitre 27 – L'inférence statistique* ». In: Dabis F. & Desenclos J.-C (Eds) « *Epidémiologie de terrain : Méthodes et Applications* ». . Montrouge: John Libbey Eurotext. pp. 248 – 84.
13. **Jutand, M.-A.** (2012). Sitographie – Tutoriels pour logiciels et cours interactifs de statistique. *Statistique et Enseignement* 2(2): 75 – 79.
14. Mamane, A., Bhatti, J. A., Savès, M., Alioum, A., **Jutand, M.-A.**, Hadiza-Jackou, D., & Sasco, A. J. (2012). La prise en charge du cancer du sein au Niger : connaissances, attitudes et pratiques des professionnels de santé non médecins de Niamey, Niger, 2010. *Journal Africain du Cancer* 4(3): 156 – 163.

2011

15. Allès, B., Samieri, C., Lorrain, S., **Jutand, M.-A.**, Carmichael, P.-H., Shatenstein, B., & Barberger-Gateau, P. (2011). P026 Profils de comportement alimentaire des personnes âgées en Aquitaine et au Québec. *Cahiers de Nutrition et de Diététique* 46(Suppl 1): S64 – S65.
16. Coureau, G., Perez, P., Benard, A., **Jutand, M.-A.**, Chêne, G., & Savès, M. (2011). Intérêt de deux formations en épidémiologie, à distance via Internet, pour la situation professionnelle des apprenants. *Revue d'Épidémiologie et de Santé Publique* 59: S29.
17. Féart, C., Torrès, M. J. M., Samieri, C., **Jutand, M.-A.**, Peuchant, E., Simopoulos, A. P., & Barberger-Gateau, P. (2011). Adherence to a Mediterranean diet and plasma fatty acids: data from the Bordeaux sample of the Three-City study. *The British Journal of Nutrition* 106(1): 149 – 158.
18. **Jutand, M.-A.**, & Ruiz-Gazen, A. (2012). L'enseignement à distance en statistique et l'usage de logiciels au travers de deux exemples. *Statistique et Enseignement* 2(2): 25 – 39.
19. **Jutand, M.-A.** (2011). Des logiciels au service de l'enseignement de la statistique. *Statistique et Enseignement*. *Statistique et Enseignement*, 2(2): 75 – 79.

2009

20. Bourdel-Marchasson, I., Rolland, C., **Jutand, M.-A.**, Egea, C., Baratchart, B., & Barberger-Gateau, P. (2009). Undernutrition in geriatric institutions in South-West France: policies and risk factors. *Nutrition* 25(2): 155 – 164.
21. Encrenaz, G., Kovess-Masféty, V., **Jutand, M.-A.**, Carmona, E., Sapinho, D., & Messiah, A. (2009). Use of psychoactive substances and health care in response to anxiety and depressive disorders. *Psychiatric Services* 60(3): 351 - 357.

2008

22. Dechamps, A., **Jutand, M.-A.**, Onifade, C., Richard-Harston, S., & Bourdel-Marchasson, I. (2008). Co-occurrence of neuropsychiatric syndromes in demented

and psychotic institutionalized elderly. *International Journal of Geriatric Psychiatry* 23(11): 1182 – 1190.

23. **Jutand, M.-A.**, Gallois, A., Léger, J., Pambrun, E., Rondi, F., Belot, M., & Marrimpoeuy, P. (2008). Échelle EDAAP 2 : validation statistique d'une grille d'évaluation de l'expression de la douleur chez les adultes ou adolescents polyhandicapés. *Motricité Cérébrale : Réadaptation, Neurologie du Développement* 29(3): 93 – 100.

24. Rondi, F., Marrimpoeuy, P., Belot, M., Gallois, A., Léger, J., Pambrun, E., & **Jutand, M.-A.** (2008). Échelle EDAAP 1. La douleur de la personne polyhandicapée : la comprendre et évaluer ses spécificités d'expression par une échelle. *Motricité Cérébrale : Réadaptation, Neurologie du Développement* 29(2): 45 – 52.

25. Samieri, C., **Jutand, M.-A.**, Féart, C., Capuron, L., Letenneur, L., & Barberger-Gateau, P. (2008). Dietary patterns derived by hybrid clustering method in older people: association with cognition, mood, and self-rated health. *Journal of the American Dietetic Association* 108(9): 1461 - 1471.

26. Verret C., **Jutand, M.-A.**, De Vigan, C., Bégassat, M., Bensefa-Colas, L., Brochard, P., & Salamon, R. (2008). Reproductive health and pregnancy outcomes among French gulf war veterans. *BMC Public Health* 8: 141.

2007

27. Cousson-Gélie, F., Bruchon-Schweitzer, M., Dilhuydy, J. M., & **Jutand, M.-A.** (2007). Do anxiety, body image, social support and coping strategies predict survival in breast cancer? A ten-year follow-up study. *Psychosomatics* 48(3): 211 – 216.

28. Féart, C., **Jutand, M.-A.**, Larrieu, S., Letenneur, L., Delcourt, C., Combe, N., & Barberger-Gateau, P. (2007). Energy, macronutrient and fatty acid intake of French elderly community dwellers and association with socio-demographic characteristics: data from the Bordeaux sample of the Three-City Study. *The British Journal of Nutrition* 98(5): 1046 – 1057.

2006

29. Bensefa, L., Causse, E., Verret, C., Choudat, D., Brochard, P., **Jutand, M.-A.**, & Conso, F. (2006). Enquête française sur la Guerre du Golfe et ses conséquences sur la santé : resultants du bilan médical réalisé chez 1 008 vétérans. *Archives des Maladies Professionnelles et de l'Environnement* 67(2): 338 – 339.

30. Fourier-Réglat, A., & **Jutand, M.-A.** (2006). *Initiation aux méthodes de sondage: application à la pharmaco-épidémiologie*. Bordeaux: ARME-Pharmacovigilance.

31. Rainville, C., **Jutand, M.-A.**, & Bérard, J.-M. (2006). Plan expérimental des « études de cas multiples » : rationnel et méthode d'analyse des données. *Revue de Neuropsychologie* 16(2): 143 – 180.

32. Salamon, R., Verret, C., **Jutand, M.-A.**, Bégassat, M., Laoudj, F., Conso, F., & Brochard, P. (2006). Health consequences of the first Persian Gulf War on French

troops. *International Journal of Epidemiology* 35(2): 479 - 487.

33. Winnock, M., Neau, D., Castera, L., Viot, J., Lacoste, D., Pellegrin, J. L., Dupon M., **Jutand, M.-A.**, Colombani F., Dabis F. & Groupe d'Epidémiologie Clinique du SIDA en Aquitaine. (2006). Hepatitis B vaccination in HIV-infected patients: a survey of physicians and patients participating in the Aquitaine cohort. *Gastroentérologie Clinique et Biologique* 30(2): 189 - 195.

2005

34. Barberger-Gateau, P., **Jutand, M.-A.**, Letenneur, L., Larrieu, S., Tavernier, B., Berr, C., & 3C Study Group. (2005). Correlates of regular fish consumption in French elderly community dwellers: data from the Three-City study. *European Journal of Clinical Nutrition* 59(7): 817 - 825.

35. Quaglia, A., **Jutand, M.-A.**, Dhillon, A., Godfrey, A., Togni, R., Bioulac-Sage, P., & Dhillon, A. P. (2005). Classification tool for the systematic histological assessment of hepatocellular carcinoma, macroregenerative nodules, and dysplastic nodules in cirrhotic liver. *World Journal of Gastroenterology* 11(40): 6262 - 6268.

2004

36. Thiessard, F., Saves, M., Alioum, A., Anglaret, X., Barberger-Gateau, P., **Jutand, M.-A.**, & Chêne, G. (2004). P17-13 Évaluation par les apprenants d'un dispositif de formation ouverte et à distance (FOAD) via Internet en épidémiologie. *Revue d'Épidémiologie et de Santé Publique* 52: 152.

37. Verret, C., **Jutand, M.-A.**, Begassat, M., Laoudj, F., & Salamon, R. (2004). D1-6 L'Enquête française sur la guerre du Golfe et ses conséquences sur la santé. Contexte, mise en place et limites méthodologiques. *Revue d'Épidémiologie et de Santé Publique* 52: 44.

2003

38. Daucourt, V., Saillour-Glénisson, F., Michel, P., **Jutand, M.-A.**, & Abouelfath, A. (2003). A multicenter cluster randomized controlled trial of strategies to improve thyroid function testing. *Medical Care* 41(3): 432 - 441.

39. Rabarijaona, L. P., Boisier, P., Ravaoalimalala, V. E., Jeanne, I., Roux, J. F., **Jutand, M.-A.**, & Salamon, R. (2003). Lot quality assurance sampling for screening communities hyperendemic for *Schistosoma mansoni*. *Tropical Medicine & International Health* 8(4): 322 - 328.

2001

40. Salamon, R., Alperovitch, A., Conso, F., Jutand, M.-A., Boutin, J.-P., Coquin, Y., & Weinbreck, P. (2001). Rapport de mission du groupe de travail chargé d'analyser les données sanitaires relatives aux anciens combattants français de la guerre du Golfe. Paris: La Documentation Française. 140 p.

2000

41. **Jutand, M.-A.**, & Salamon, R. (2000). La technique de sondage par lots appliquée à l'assurance qualité (LQAS): méthodes et applications en sante publique. *Revue d'Épidémiologie et de Santé Publique* 48: 401 - 408.