HAL
open science

# A propos de la formation en statistique. Approches praxéologiques et épistémologiques de questions du champ de la didactique de la statistique .

Jean-Claude Regnier

## To cite this version:

Jean-Claude Regnier. 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, 2002, 22/23, pp.157-201. halshs-00363427

**HAL Id: halshs-00363427**

**https://shs.hal.science/halshs-00363427v1**

Submitted on 19 Jul 2009

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

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

# A propos de la formation en statistique.

## *Approches praxéologiques et épistémologiques de questions du champ de la didactique de la statistique .*

Jean-Claude Régnier
L.I.R.C.A. - Département de Sciences de l'éducation -Université Lumière Lyon 2
jean-claude.regnier@univ-lyon2.fr

*Résumé :*

« *En France, l'absence de formation en statistique, dans les collèges, les lycées et de vastes secteurs de l'enseignement supérieur, conduit à des attitudes sociales aberrantes. (...)»* ce constat de Jean Dercourt situe bien le contexte dans lequel nous cherchons à construire et à résoudre, dans le champ de la didactique de la statistique, des problématiques relatives à la formation en statistique, et en particulier à des questions d'enseignement-apprentissage en milieu scolaire ou universitaire. Un regard historique nous montre qu'en France, la formation en statistique est une question déjà ancienne — remontant au XIX$^{ème}$ siècle — et récurrente, à laquelle peu de réponses ont été données, car elle ne mobilise réellement que peu d'intérêt des communautés scientifiques concernées. Nous portons ensuite notre attention sur le référent de la formation en statistique en tentant d'expliciter les objectifs d'intégration et de définir l'esprit statistique et son rapport à l'éducation statistique. Nous abordons dans cette perspective la question de la place de la statistique dans la formation en sciences humaines et sociales, envisageant tour à tour la statistique comme discipline de base, discipline de service, discipline d'ouverture, discipline-objet de la didactique, discipline-objet de la recherche en statistique dans son application à la recherche en sciences de l'éducation. Nous essayons de cerner le sens de l'interprétation statistique dans l'analyse statistique, et d'aborder le problème que soulève son enseignement et son apprentissage.Enfin nous tentons d'expliciter les questions que soulève la formation en statistique dans le champ de la didactique de la statistique. Un rapide état des lieux montre qu'en France, très peu de travaux de thèse ont été consacrés à de telles questions. Nous avons repris deux chantiers pionniers ouverts respectivement par Yves Chevallard (Chevallard 1978 ) et par Claudine Blanchard-Laville (Blanchard-Laville 1980) en relation à la rénovation de l'enseignement de la statistique en psychologie. À cette occasion, nous avons ré-abordé la question de la transposition didactique et pu retrouver le concept de *déterritorialisation* proposé par Yves Chevallard. Restant dans la perspective culturelle latine, le groupe espagnol de l'Université de Grenade dirigé par Carmen Batanero (Batanero 2001) constitue un pôle important pour le développement de la didactique de la statistique. L'analyse textuelle des titres des thèses et HDR du domaine de la didactique des mathématiques ou en périphérie, présentes dans les bases de données de l'I.N.R.P. et de l'ARDM montre que sur 269 seulement 9 relèvent de problématiques du champ de la didactique de la statistique, contre 24 en géométrie et 19 en algèbre. Si nous tenons compte en plus de l'évolution des programmes de mathématiques des lycées et des collèges, il nous semble nécessaire et pertinent que soient développées des recherches en didactique de la statistique.

*Mots Clés* : interprétation statistique, analyse statistique, raisonnement statistique, éducation statistique

## 1 Introduction.

Nous ne saurions mieux dire nos préoccupations à l'égard de la formation en statistique que ne le fait Jean Dercourt dans l'avant-propos du rapport de l'Académie des Sciences ( Acad. des Sc. 2000, p. ix ) « *En France, l'absence de formation en statistique, dans les collèges, les lycées et de vastes secteurs de l'enseignement supérieur, conduit à des attitudes sociales aberrantes. (...) alors que les résultats statistiques fournis par les médias s'accumulent tous les jours, les lecteurs et les auditeurs n'ont pas les moyens de les analyser comme ils le méritent.(...) Cette carence devient d'autant plus préoccupante que la statistique, comme toute science, évolue. Les utilisateurs, les clients et les citoyens doivent maîtriser cette information, et donc connaître les règles de la discipline et les possibles biais d'interprétation. Fort peu le font. La faiblesse de la statistique en France est, sans conteste, un verrou très solide entravant le développement économique et l'exercice des droits des citoyens*.»

En premier lieu, il nous semble judicieux de pouvoir resituer historiquement ce constat. Après quoi, nous tenterons d'expliciter le référent de la formation en statistique. Puis nous aborderons les questions soulevées par la formation en statistique dans le domaine de la didactique de la statistique.

## 2 La formation en statistique : une question déjà ancienne et récurrente.

La question de la formation en statistique et par conséquent celle de son enseignement en France ne sont pas nouvelles. La lenteur de l'implantation et du développement de cet enseignement peut nous sembler paradoxale dans un pays qui connut des savants comme Pierre Simon de Fermat (1601-1665), Blaise Pascal (1623-1662), Abraham de Moivre (1667-1754), Pierre Simon Marquis de Laplace (1749-1827), Siméon Denis Poisson (1781-1840) , Augustin Antoine Cournot (1801-1877) ou Jules Henri Poincaré (1854-1912), considérés parmi les contributeurs magistraux à la construction de la théorie des probabilités, une des sciences mères de la statistique mathématique. Dans un article sur l'*enseignement des statistiques en France du milieu du XIXème siècle à 1960*, (Affichard 1987 Tome 2 pp.811-823),Christian Morisson nous brosse un tableau fort instructif de l'évolution historique de cet enseignement dans le domaine de l'économie. Il mentionne, entre

autre, les sources de divers rapports et articles traitant de l'enseignement de la statistique, écrits par E. Cheysson (Cheysson, 1890), F. Faure, (Faure, 1893, 1894), R. Roy, (Roy, 1937), H. Bunle (Bunle, 1945), E. Morice, (Morice, 1961). Ainsi jusqu'en 1892, le seul enseignement officiel de statistique était le cours d'économie industrielle et statistique, institué en 1854 au Conservatoire national des arts et métiers. Cependant il est constaté que « *les responsables de ce cours, Burat puis de Foville, utilisaient des données de la statistique mais n'enseignaient pas, à proprement parler, les méthodes statistiques*. » ( Cheysson, 1890 cité par Morisson). Il conclut qu'« *il faut attendre les années 1960 pour que les enseignements de statistique fassent réellement partie de la formation des économistes universitaires*.» mais que les conditions requises pour que cet enseignement soit effectif et de qualité ne sont pas globalement réunies. « *Par suite, l'étudiant qui va utiliser les statistiques dans une activité professionnelle ou pour un travail de recherche se trouve mal préparé*.»

Dans les années 80, dans un article publié par l'UNESCO, Lennart Rade expose une série de constats émis depuis les années 50 à propos de la formation en statistique qui vont tous dans le sens de déplorer l'insuffisance de cette dernière au nom de sa nécessité grandissante dans notre société actuelle. En particulier, la conclusion d'un séminaire de l'Organisation Européenne de Coopération Économique organisé au cours de l'année 1959 en France, et consacré à l'enseignement des mathématiques soutient que « *Le calcul des probabilités élémentaires doit être considéré comme une branche des mathématiques susceptibles d'être enseignée dans les écoles secondaires. L'induction statistique doit être considérée comme une branche des mathématiques appliquées qui entre pour une part capitale dans les processus de décision conformes à l'esprit de la "méthode scientifique" et dont de très nombreux secteurs des sciences physiques et des sciences du comportement humain font un usage accru. Il faut admettre en outre que le raisonnement statistique acquiert une importance croissante dans le domaine des affaires publiques. Un enseignement élémentaire approprié du calcul des probabilités et de la statistique doit faire partie du nouveau programme des études secondaires, des cours préparatoires sur ces matières devront figurer aux programmes (...) des institutions formant les professeurs*.».

En 1963, W. Weaver[1] écrit « *La théorie des probabilités et la statistique sont deux domaines importants, intégrés à nos activités quotidiennes. Le monde de l'industrie, les compagnies d'assurance sont largement tributaires des lois probabilistes. La physique elle-même est de nature essentiellement probabiliste. Il en est de même des fondements de la biologie. Cependant, en dépit de cette importance, les responsables de l'enseignement n'ont pas encore admis le caractère universel de la théorie des probabilités et de la statistique. Il faut espérer que des éléments de la théorie des probabilités soient introduits dès que possible au niveau de l'enseignement secondaire...*» En introduisant *Les probabilités à l'école* (Glaymann, Varga 1973) avec cette citation, Maurice Glaymann et Tamas Varga exprimaient une certaine satisfaction en constatant qu'au cours de ces dix années écoulées, l'enseignement des mathématiques avait connu, tant en France que dans de nombreux pays, des réformes s'ouvrant à l'enseignement des probabilités. Ils ajoutaient alors *« et ce n'est peut-être plus qu'une question de temps pour qu'elles constituent une partie intégrante des programmes scolaires de tous les pays »* .

Force est de constater qu'en France, il aura fallu attendre les années 2000, c'est à dire encore plus de 25 ans, pour que la statistique s'inscrive explicitement et significativement dans l'ensemble des programmes de mathématiques des collèges et des lycées. Toutefois, si cette institutionnalisation d'un enseignement de statistique marque une réelle avancée, sa mise en œuvre demeure encore très problématique, compte tenu de la hiérarchisation des domaines à l'intérieur du champ des mathématiques. Par comparaison à ce que nous avions constaté (Régnier, 1983), nous dirions qu'elle est exposée au risque de subir un sort analogue à la trigonométrie : étudiée pour elle-même et reportée au dernier moment dans la progression annuelle. Trois mobiles majeurs semblent se dégager des enquêtes empiriques que nous avons menées :

- La formation des enseignants de mathématiques qui laisse peu ou pas de place à la statistique,
- Leur représentation relative à la statistique qui la situe plutôt comme une science mineure en marge du domaine des mathématiques, et pour laquelle ils ne voient pas d'usage réellement convaincant.
- La question de l'évaluation des connaissances acquises en statistique par les élèves.

---

[1] *Lady Luck* , N.Y. :Doubleday and Co, Garden City, pp 376-377, cité par (Glaymann, Varga, 1973 p.)

Dans ce sens, en analysant un échantillon de manuels de mathématiques de l'enseignement secondaire français, nous avons eu l'occasion de constater les dérives dans l'usage du concept d'*histogramme*. (Régnier 1998a).

Cette question de la formation en statistique se prolonge à l'Université. Pour nous, elle se pose d'une manière aiguë pour les étudiants du domaine des sciences humaines et sociales, en particulier pour ceux de licence et maîtrise de sciences de l'éducation. Ce qui nous frappe le plus, est la permanence de certains faits. En partant d'un constat d'enseignant relatif à la *considérable résistance des étudiants* de DEUG de psychologie *face à l'enseignement de la statistique*, et s'interrogeant sur *comment faire pour la vaincre*, Claudine Blanchard-Laville (Blanchard-Laville, 1980) a fondé sa thèse sur une recherche de la *compréhension de cette résistance*. Cette résistance qui se révèle d'abord au niveau cognitif : l'absence d'acquis élémentaires en mathématiques rendent problématiques les acquisitions nouvelles, se manifeste au-delà à travers le climat d'angoisse qui enveloppe cet enseignement. Ayant adopté une perspective clinique d'orientation psychanalytique, elle en conclut que « *cette attitude, fruit d'un passé de confrontation avec cette discipline tout au long du parcours scolaire, est profondément enracinée dans l'histoire personnelle de chaque sujet. Et l'on découvre progressivement, dans chaque cas, un roman singulier, fait de l'agencement subtil des différents facteurs qui sont intervenus dans le tissage de cette relation avec les mathématiques.* ».

Vingt ans avant, un article de L. Leboutet (Leboutet, 1958) rendait compte d'une étude sur les difficultés en statistique éprouvées par les étudiants de psychologie à partir de laquelle il rédigea un lexique de statistique. (Leboutet, Rouanet, 1959)

Vingt ans après, nous constatons le même phénomène de résistance au travers des manifestations très similaires. De tels constats ressortent de l'analyse d'un corpus de messages électroniques échangés au sein d'une liste de diffusion réunissant plus d'une centaine d'étudiants parmi ceux qui préparent une licence de sciences de l'éducation à distance avec le CNED et l'Université Lyon 2, et les 13 tuteurs impliqués dans ce dispositif pédagogique, au cours des années universitaires 2000-2001 et 2001-2002. (Régnier 2002a). Ils se confirment aussi au travers des résultats d'une enquête par questionnaire que nous avons répétée chaque année depuis 1992 auprès des étudiants de licence et de maîtrise de sciences de l'éducation, et dont le but est de faire émerger les représentations de différents concepts et techniques, tels que *représentativité*, *significativité*, *estimation*, *test d'hypothèse statistique*, *moyenne*, *variance*, etc., et d'expliciter le sens des attitudes à l'égard des méthodes statistiques ( Régnier 2002b *en préparation*).

## 3 La formation en statistique : son référent

En terme d'objectif d'intégration (De Ketele, 1983), une compétence doit s'exercer sur une situation complexe comprenant tant de l'information essentielle que secondaire et parasite et requérir des activités complexes intégratives des acquis antérieurs jugés fondamentaux : savoirs et savoir-faire. La situation d'évaluation du degré d'atteinte de l'objectif d'intégration doit être la plus proche possible des situations de vie sociale ou professionnelle auxquelles l'apprenant sera ultérieurement confronté. L'objectif d'intégration explicite des savoir-être et des savoir-devenir orientés vers la finalité du développement et de l'autonomie. En ce sens, la formation en statistique vise à rendre l'apprenant capable de :

- construire un modèle requis par une approche statistique congruent au modèle dans lequel une problématique est posée.
- décrire, traiter, analyser des données de manière pertinente dans le cadre du modèle construit pour conduire une étude.
- tenir un raisonnement intégrant l'idée de risque dans l'énoncé des conclusions.
- interpréter les phénomènes étudiés sur la base de données statistiques recueillies sur des faits et leurs relations.
- communiquer des résultats des analyses de données en faisant une distinction entre le modèle utilisé et la réalité qu'il est supposé représenter, entre les traitements conduits au sein du modèle et les interprétations reformulées dans le contexte au sein duquel est posée la problématique.
- poursuivre de façon autonome et personnalisée un apprentissage en statistique afin d'enrichir les acquis personnels actuels.

### 3.1 Formation de l'*esprit statistique* par la formation en statistique.

Nous aimerions pouvoir caractériser la formation de *l'esprit statistique* comme, en 1938, Gaston Bachelard (Bachelard 1989 p.22) a su le faire pour *l'esprit scientifique*, entendu dans son rattachement aux sciences physiques, à la chimie et à la biologie. Il parle aussi de *l'esprit mathématique* dont il fait le projet d'étudier la formation, du même point de vue critique. Selon son opinion, cette séparation est justifiée par le fait que : « *le développement de l'esprit mathématique est bien différent du développement de l'esprit scientifique dans l'effort de comprendre les phénomènes*

*physiques. En réalité, l'histoire de la mathématique est une merveille de régularité. Elle connaît des périodes d'arrêt. Elle ne connaît pas de périodes d'erreurs.* » il poursuit d'ailleurs en écrivant que (Bachelard 1989 p.22) : « *Aucune des thèses que nous soutenons dans ce livre ne vise donc la connaissance mathématique.* » En tant que branche des mathématiques, la statistique entrerait alors dans ce projet. Cependant, ce dont nous doutons, c'est de cette régularité d'un développement sans confrontation à l'erreur. Les débats et polémiques suscités par les notions d'*échantillon*[2] ou de *probabilité* et de leur valeur scientifique pourraient en constituer un contre-exemple. En 1836, A. A. Cournot situe le développement de la statistique dans une rupture. Ainsi écrit-il (Cournot, 1984) : « *La statistique est une science toute moderne : le génie des anciens ne se portait pas volontiers vers des travaux de précision ; les moyens de recherche et de communication leur manquaient »* De plus, il fonde épistémologiquement la statistique sur un principe de compensation dont les anciens ne paraissaient pas avoir soupçonné l'existence, qui : *« finit toujours par manifester l'influence des causes régulières et permanentes, en atténuant de plus en plus celle des causes irrégulières et fortuites.* » Nous pourrions chercher à savoir dans quelle mesure cette sorte de *loi des compensations*, souvent évoquée comme quasi naturelle, à laquelle, en 1836, A. A.Cournot fait référence, constitue un obstacle épistémologique à la compréhension et au développement de la statistique. Ainsi cette posture épistémologique permet de donner un certain sens à la notion de *moyenne* dont l'algorithme de calcul est conçu pour neutraliser les fluctuations dues aux causes fortuites lors de la mesure. Mais elle peut aussi en limiter la généralisation nécessaire à sa conceptualisation. Nous pensons alors à d'autres *moyennes* telles que moyenne harmonique, moyenne géométrique, moyenne quadratique... Un questionnement analogue concerne la notion de *probabilité*, avec la prégnance du modèle fréquentiste (dominant au XIX$^{ème}$ s., et même jusqu'en 1933, année de la publication de la théorie de Kolmogorov[3]) dans lequel la *probabilité* n'est autre que la *fréquence.*

---

[2] voir les débats au sein de l'I.I.S./I.S.I. à la fin du XIX$^{ème}$ siècle et au début du XX$^{ème}$ (Droesbeke et al. 1987 pp. 3-17)

[3] Dans son ouvrage (Popper 1988 Appendice II p.323-324) *La logique de la découverte scientifique*, K.R. Popper aborde, en 1938, cette problématique d'une théorie mathématique des probabilités. Dans l'appendice II, à propos de la note sur la probabilité publiée en 1938 dans *Mind*,

La formation de l'*esprit statistique* s'inscrit dans une dynamique qui permet aux êtres humains de se confronter aux obstacles — épistémologiques ou autres — au développement de la statistique en tant que science. Elle intègre une attitude de vigilance dans l'usage instrumental de la statistique dont Cournot disait déjà en 1836 que (Cournot 1984) : « *De nos jours (...) la statistique a pris un développement en quelque sorte exubérant ; et l'on n'a plus qu'à se mettre en garde contre des applications prématurées et abusives qui pourraient la décréditer pour un temps »,* ajoutant que la conséquence dommageable serait de : *« retarder l'époque si désirable où les données de l'expérience serviront de bases certaines à toutes les théories qui ont pour objet les diverses parties de l'organisation sociale.* » Dans cette opinion d'une étonnante actualité, nous mettons cependant en question aujourd'hui le caractère *certain* de telles *bases*, et nous dirions plutôt que les données de l'expérience traitées adéquatement et pertinemment par une approche statistique fournissent des bases *utiles*, *valides* ou *fiables* aux théories qui y recourent.

Nous postulons que ni l'*esprit statistique* n'est un don, ni sa formation n'est le résultat du développement naturel de l'être humain. (Régnier 1998b). La formation de l'*esprit statistique* se caractérise par la nature du rapport à l'incertitude et à l'erreur, considérées comme inhérentes à tout acte de prise de décision. Elle est instrumentée par une conceptualisation du risque encouru dans une prise de décision et par une modélisation de son contrôle, ne laissant plus l'exclusivité à une compréhension ou une explication fondées sur une conception spontanée du hasard et du déterminisme, empreinte parfois du fatalisme. Cette question est abordée par Jacques Bordier à propos des conceptions de la probabilité en étudiant *les conceptions de la probabilité que l'on se fabrique* (Bordier 1991 pp. 31-62) faisant suite aux travaux de Sylvette Maury (Maury 1986). Plus récents sont les travaux de Dominique Lahanier-Reuter qui explicite diverses conceptions du hasard issues de modes de connaissances différents circulant au cours de séquences d'enseignement et d'apprentissage concernant la modélisation probabiliste de situations statistiques. (Lahanier Reuter 1998)

---

*(suite de la note de la page précédente)*
sous l'intitulé : « Un ensemble d'axiomes indépendants pour la théorie de la probabilité. »il précise qu'il n'avait pas encore eu connaissance de la théorie de Kolmogorov. Il considère d'ailleurs que le système d'axiomes de Kolmogorov comme une des interprétations de son propre système.

La conclusion d'un rapport, réalisé en 1959 par la commission sur l'enseignement des mathématiques, instituée en 1955 aux U.S.A., mentionne que *« Si les mathématiques ont trait à des situations où les faits peuvent être déterminés, elles donnent aussi les moyens d'étudier, de comprendre et de maîtriser l'incertain. Parmi les plus récentes applications des mathématiques, beaucoup font appel à la théorie des probabilités et au raisonnement statistique.(...) La pensée statistique joue un rôle croissant dans la vie quotidienne des adultes instruits. L'initiation à la pensée statistique est un complément important à l'initiation à la pensée déductive.»* Ce point de vue renforce la perspective évoquée plus haut[4]. qui considère *l'induction statistique comme une branche des mathématiques appliquées, tenant une part capitale dans les processus de décision conformes à l'esprit de la méthode scientifique*. Par le schéma suivant, nous explicitons quelques éléments constitutifs de la formation de *l'esprit statistique*

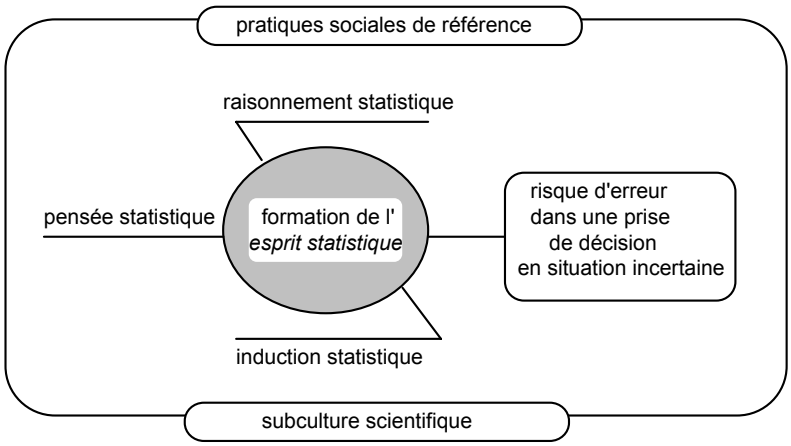


**Figure 1 : facteurs de la formation de *l'esprit statistique***

Dans la perspective actuelle des programmes de lycée, l'esprit statistique « *naît lorsqu'on prend conscience de l'existence de fluctuation d'échantillonnage ; en seconde l'élève constatera expérimentalement qu'entre deux échantillons, de même taille ou non, les distributions des fréquences fluctuent. (...) Le choix pédagogique est (...) d'aller de l'observation vers la conceptualisation et non d'introduire d'abord*

[4] Séminaire sur l'enseignement des mathématiques organisé en 1959 par l'Organisation Européenne de Coopération Economique.

*le langage probabiliste pour constater ensuite que tout se passe comme le prévoit cette théorie.* »[5]

La formation de *l'esprit statistique*, c'est aussi renoncer à l'usage systématique de l'idée de *vérité* pour chercher à maîtriser celle de *vraisemblance*. Dans *De la vérité autoproclamée à la vraisemblance reconnue* (Régnier 1998c), nous avons pris le parti de nous confronter à deux questions issues des échanges avec les étudiants :

- *Comment peut-on reconnaître et contrôler les risques encourus lors d'une prise de décision en situation incertaine ?*

- *Quelle place et quel rôle tiennent les notions de significativité d'un résultat et de représentativité d'un échantillon dans la prise de décision ?*

Les réponses que nous avons apportées, sont d'ordre instrumental. Les outils conceptuels, techniques de la statistique sont mis à la disposition de l'apprenant pour qu'il les conceptualise et les acquière afin de les transformer en *instruments psychologiques* (Vygostki 1985) (Rabardel 1995) actualisables dans le contrôle d'une classe de prises de décision en situation incertaine.

En filigrane de ces propos, figure la sollicitation de trois modes de raisonnement que sont la déduction, l'induction et l'éduction. Développer *l'esprit statistique* consiste en leur acquisition et en leur *manipulation consciente* pour conduire un *raisonnement statistique* et étayer l'activité d'*interprétation statistique* .

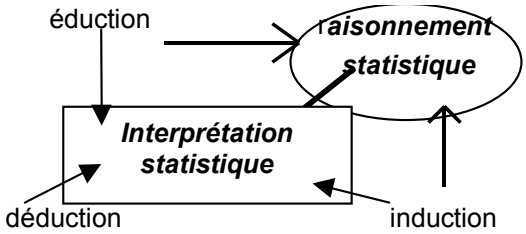


**Figure 2 : Interprétation et raisonnement statistiques sous contrainte de l'induction, l'éduction et la déduction.**

Quant à la *pensée statistique*, nous la replaçons dans la catégorisation des modes de pensée opérée par Philippe Meirieu (Meirieu 1987) et reprise par J-Pierre Astolfi (Astolfi 1993). Cinq modes y sont recensés :

[5] Document d'accompagnement du programme de Seconde / juin 2000 ; p. 8

| Pensée déductive | Pensée inductive | Pensée dialectique | Pensée divergente | Pensée analogique |
|---|---|---|---|---|

Au vu des caractéristiques fournies, la pensée statistique renvoie à un mode intégrant, de manière dominante, les deux modes de pensée inductive et analogique.

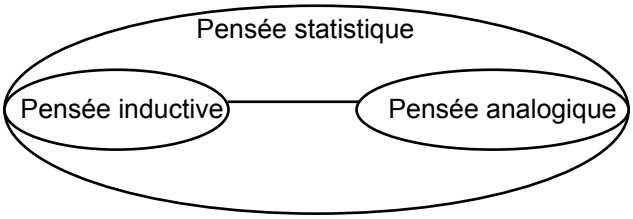


**Figure 3 : Mode de pensée statistique**

| *Caractéristiques principales de la pensée inductive* | *Caractéristiques principales de la pensée analogique* |
|---|---|
| Organiser les données pour chercher à les expliquer.<br>Déborder le niveau des faits pour accéder à celui des mécanismes explicatifs.<br>Rechercher des tendances, des régularités, des évolutions, des conservations. | Étendre à un domaine nouveau ce qui est établi dans un autre contexte.<br>Utiliser de manière systématique, puis critique, la comparaison et la métaphore. |

Ce mode de pensée statistique est particulièrement mis en œuvre dans l'usage des approches échantillonnales de la statistique inférentielle. Mais nous n'excluons pas pour autant l'intervention des autres modes évoqués, en périphérie du mode de pensée statistique, dans le cours d'une étude. Enfin, nous pensons que l'esprit statistique est à la fois un esprit de rigueur et un esprit de tolérance par le recours même à des propositions jugées vraisemblables et non plus vraies, au moyen de méthodes explicites.

### 3.2 Éducation statistique

Clarifions ce que nous entendons par *éducation statistique.* L*'éducation statistique* désigne à la fois un résultat actuel et une finalité du processus éduquer/éducation. Une requête au sein du corps du texte de l'Encyclopédie Universalis électronique (1998), s'est soldée par deux occurrences de *éducation mathématique* qui ne fournissent par ailleurs aucune caractérisation, mais aucune de *éducation statistique* !

Pour nous, *l'éducation statistique* représente d'abord l'intégration des éléments pointés précédemment à propos de la formation en statistique et de *l'esprit statistique.* Certes, cette éducation-là est visée par l'action de l'enseignement, mais

elle n'est pas le produit de cette seule action. De nombreuses autres situations dans divers contextes non-scolaires peuvent activer le processus d'apprentissage des sujets pour acquérir des connaissances, des compétences, des habiletés, des comportements propres à développer *l'éducation statistique*. Cependant il s'agit d'un produit en évolution permanente et par palier. Par exemple, la lecture d'un article de journal d'un quotidien régional peut constituer une telle situation (Régnier 1998d).

D'un autre côté, il serait sans doute intéressant de conduire une analogie avec les autres expressions : *éducation mathématique, éducation musicale, éducation artistique, éducation civique et éducation physique et sportive*. Nous ne le ferons pas ici compte tenu des limites que nous nous fixons. En revanche, l'analogie de *l'éducation musicale*, nous suggère de revenir à la question de l'*interprétation* en statistique.

Pour reprendre l'idée exprimée par Brigitte Escofier et Jérôme Pagès (Escofier & Pagès 1990 p.217-218), nous caractérisons *l'interprétation statistique* selon trois directions :

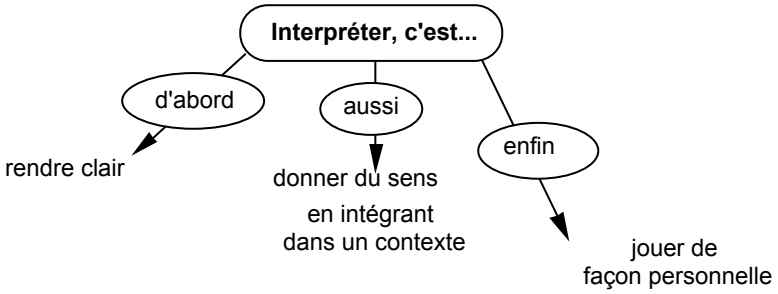


**Figure 4 : Interpréter, c'est...**

Dans le cadre de l'enseignement, la question de l'interprétation statistique émerge chez les étudiants. Dès que le niveau de formation en statistique permet de maîtriser quelques-uns des outils conceptuels et techniques de base de la statistique surgit une série de questions chez les étudiants dont le filigrane est la question de l'interprétation des données recueillies et traitées. À l'aide du schéma ci-dessous, nous tentons de situer ce questionnement dans l'articulation générale du parcours de formation vers un niveau d'éducation statistique qui s'élève à chaque étape.

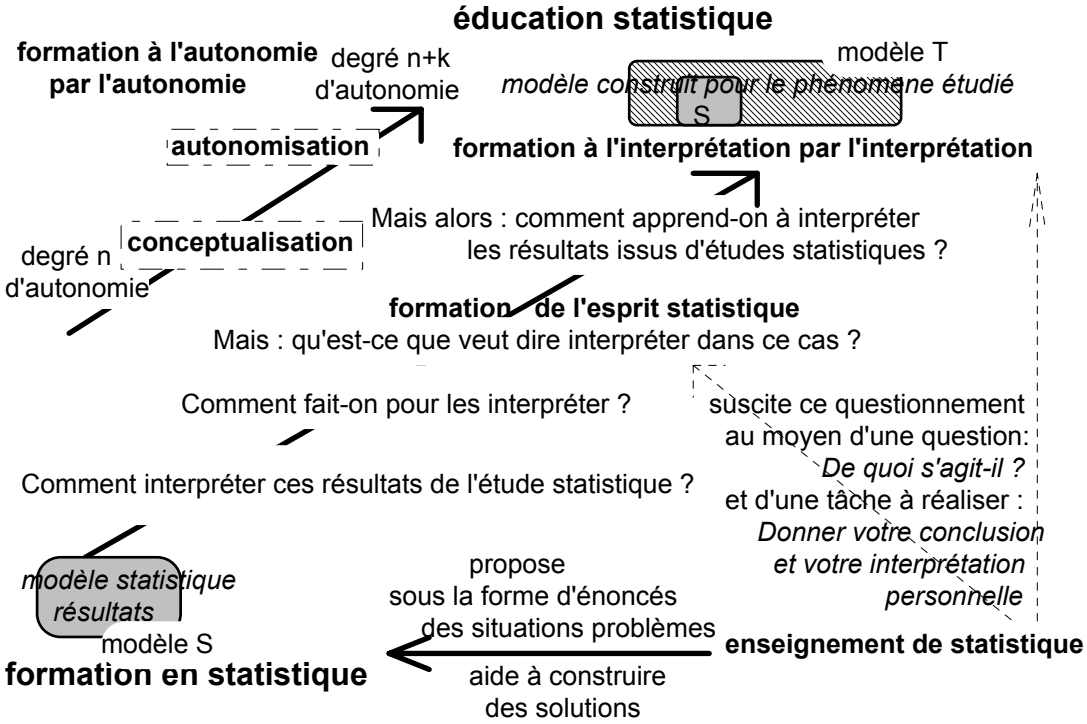


**Figure 5 : Schématisation de la place de la question de l'interprétation soulevée par les étudiants dans le parcours de formation et du rôle de l'enseignement.**

Les situations problèmes[6] proposées aux étudiants comportent un questionnement suggérant un traitement dont la réalisation conduit à la mise en œuvre d'outils statistiques, objets d'apprentissage. Au fil des années, il nous est apparu la nécessité pédagogique d'introduire des questions plus réflexives afin de provoquer chez l'étudiant un changement de comportement dans le traitement. En effet, malgré notre choix d'un habillage re-contextualisant les outils conceptuels et techniques de la statistique, nous avons observé que les étudiants s'en tenaient à la mise en œuvre de ces outils. Par exemple, la question portant sur la moyenne et l'écart-type d'une variable quantitative n'obtenait que les valeurs mathématiques comme réponse issue des algorithmes de calcul, sans aucune re-contextualisation. À la question : quel est l'âge moyen de cet échantillon ? Nous obtenions une réponse du type 15,5968777775555. Par ailleurs, les réactions spontanées manifestées par ces étudiants adultes face à ces énoncés de situations problèmes se traduisaient

[6] Le lecteur pourra prendre connaissance des énoncés de ces situations-problèmes sur notre site Web :

massivement par deux questions du type : *que faut-il calculer ou tracer ? Ça sert à quoi de calculer une caractéristique, d'estimer un paramètre ou de tracer un graphique...?* Nous avons alors introduit, de manière systématique, une question préalable sous la formulation : *de quoi s'agit-il ?* en présentant la lecture-compréhension des énoncés de situations problèmes par analogie à celle des articles scientifiques qu'ils ont à consulter. Puis pour chaque traitement, nous avons imposé qu'il se termine par une *conclusion* et une *interprétation personnelle*. Ces items sont aussi pris en compte dans l'évaluation finale comme le montrent les grilles d'évaluation qui accompagnent chaque énoncé de situation problème. (Régnier 1994a, 1994b, 1995a, 1995b). C'est alors que le questionnement relatif à l'interprétation est explicitement apparu dans la mesure où nous avons été directement interpellé par les étudiants dans les cours sur les critères de réussite de cette tâche.

Les exemples par lesquels cette question de l'interprétation statistique est soulevée, abondent (Régnier 2002b). Pour préciser notre pensée nous en rapportons quelques-uns uns :

•Supposons qu'une étude statistique descriptive ait été conduite sur un échantillon ou sur une population à partir d'une variable quantitative discrète. La moyenne et l'écart-type ont été calculées et le diagramme en bâtons de la distribution des fréquences a été tracé. Que signifie interpréter la moyenne et l'écart-type ou le diagramme en bâtons ?

•Si l'étude précédente a été conduite sur un échantillon aléatoire et que nous procédions à une estimation par intervalle de confiance à un niveau de 95%. Qu'interprète-t-on quand on lit la *fourchette* qui en résulte ?

•Si, sur ce même échantillon, nous étudions le croisement de deux variables quantitatives et que nous appliquions un test d'indépendance (du type test du Khi-deux), qu'est-ce qu'interpréter le rejet ou le non-rejet de l'hypothèse $H_0$ d'indépendance confrontée à l'hypothèse alternative composite $H_1$ de dépendance ?

•Si nous y étudions simultanément plusieurs variables qualitatives par une analyse factorielle des correspondances multiples suivie d'une classification — comme dans

---

*(suite de la note de la page précédente)*

http://perso.wanadoo.fr/jean-claude.regnier/ ou les télécharger au format PDF.

le cas de traitement des données textuelles assisté du logiciel SPAD_T -, qu'est-ce qu'interpréter les axes factoriels et les classes ?

Évidemment, en présentant ainsi ces questions, cela donne le sentiment d'une interprétation réalisée pour elle-même sans mise en correspondance. Mais celles-ci nous ont été renvoyées par les étudiants dans le contexte de l'enseignement de la statistique ou encore dans le travail d'étude en maîtrise de sciences de l'éducation.

À ce jour, à partir des réponses fournies lors des épreuves d'examen, nous avons constitué un corpus de conclusions et d'interprétations dont la caractéristique dominante est le recours à une formulation naïve ou tautologique, souvent peu pertinente, parfois contradictoire. C'est dans ce corpus que nous puisons des exemples à partir desquels nous faisons travailler les étudiants et que nous introduisons dans les livrets autocorrectifs que nous élaborons.(Régnier 1994a, 1994b, 1995a, 1995b)

Là encore nous percevons les limites de l'action d'enseignement. Comment enseigner l'*interprétation statistique* ? Il appert que cette question de l'interprétation est largement laissée à la charge de l'étudiant en comptant sur un travail autonome pour développer cette compétence. En restant dans une perspective d'enseignement, nous avons pris conscience de la difficulté à expliciter ce qu'était *l'interprétation en statistique* en dehors d'une approche empirique et praxéologique. Nous avons tenté une explicitation qui s'est limitée à une précision dans la grille d'évaluation.

**Tableau 3.2-1 : explicitation conclusion/interprétation personnelle**

| *Évaluation des items conclusion-interprétation* | |
|---|---|
| Conformément aux objectifs visés en cours, l'item «conclusion» correspond à deux niveaux | |
| *niveau 1*<br>(représenté par le point de vue «résultat»)<br>de la conclusion correspond à une description lisible et bien rédigée des résultats obtenus dans les traitements abordés. Ces résultats sont rapportés dans le contexte de la situation problème. (Par exemple : les nombres sont donnés dans un ordre de grandeur compatible avec le contexte, avec leurs unités ou encore les modalités sont explicitées en clair...) | *niveau 2*<br>(représenté par «raisonnement»)<br>correspond à un énoncé d'une interprétation personnelle fondée sur les résultats de l'étude et exprimée de façon pertinente. |

Si pédagogiquement nous avons pu fournir une réponse circonstanciellement adaptée, elle ne nous a pas vraiment satisfait. La formation à l'interprétation instrumentée doit être intégrée aux objectifs de l'enseignement de la statistique. Mais elle ne peut se réaliser que par l'activité même d'interprétation étayée par une formation en statistique au sein d'un groupe pour utiliser la confrontation des points de vue. En effet, la compétence d'interprétation paraît se développer, chez le novice, sous l'effet de confrontations de points de vue à la fois avec ses pairs et avec des experts. Nous poursuivons nos investigations pour construire des situations d'enseignement-apprentissage propices à l'activation du conflit socio-cognitif. Et dans une perspective vygotskienne, pour organiser des situations sociales favorables au développement de la compétence à l'interprétation statistique, la question demeure du repérage de la *zone du proche développement*. Toutefois, au-delà de ces questions pédagogique et didactique, demeure une autre question : à quelle théorie de l'interprétation peut-on rattacher *l'interprétation statistique* ?

Comme nous le pointons dans le schéma de la Figure 5, nous distinguons, d'une part, le modèle T construit pour étudier un certain phénomène et qui relève la plupart du temps d'un autre domaine que celui de la statistique ; d'autre part, le modèle statistique S adéquatement construit pour être intégré au modèle T. Le modèle S est inscrit dans une approche du type de la modélisation mathématique, à partir des outils conceptuels, techniques et méthodiques de la statistique. Sa fonction peut être *descriptive* ou *heuristique*, en référence à la statistique descriptive et exploratoire, *estimative*, ou alors de *tester des hypothèses*, en référence à la statistique inférentielle de l'estimation et des tests.

La première étape de *l'interprétation statistique* conduit à ce que nous nommons une *conclusion*, c'est à dire une ou plusieurs propositions énoncées en langage naturel en référence dominante au système formel représenté par le modèle S, et en référence secondaire au contexte déterminé au travers de l'habillage représenté par le modèle T.

La question se porte alors, d'une part, sur la valeur de vérité de la ou des propositions qui forment cette conclusion, en référence au modèle S. La valeur *vraie* est celle de la logique mise en œuvre en mathématiques. Elle se porte, d'autre part, sur les valeurs de pertinence et de cohérence de cette conclusion en référence au modèle T.

La seconde étape de *l'interprétation statistique* conduit à ce que nous nommons une *interprétation personnelle*, c'est à dire un texte — court dans le cadre de l'enseignement mais qui peut être long dans le cadre d'un mémoire de maîtrise — énoncé en langage naturel établissant une correspondance entre la conclusion — en tant que texte *vrai*, pertinent et cohérent — et le phénomène étudié représenté par le modèle T et son cadre théorique de référence. La question se porte alors sur les valeurs de validité, de pertinence et de cohérence de cette interprétation personnelle instrumentée.

Du point de vue de l'enseignement, dans l'aide à l'acquisition et au développement de la capacité d'interprétation statistique, si l'évaluation de la valeur de vérité en référence à la logique mathématique, nous semble assez aisée, il n'en est pas de même en ce qui concerne l'évaluation des valeurs de validité, de pertinence et de cohérence. Du point de vue praxéologique, cette évaluation nous paraît surtout être régulée par une *théorie-en-acte* (Vergnaud 1991) d*e* la pertinence et de la cohérence, qui s'apparenterait plutôt à une approche pragmatique. Dans l'énoncé même de nos objectifs de formation en statistique, nous justifions cette activité de l'interprétation statistique par sa double fonction de traitement de l'information et de communication des textes que les étudiants sont amenés à produire ainsi. Cette question de l'interprétation statistique a été abordée dans le cas des tests d'hypothèse statistiques par A. Vallecillos et C. Batanero (Vallecillos, Batanero, 1997). L'étude fait ressortir les difficultés de compréhension qui s'originent dans l'interprétation des probabilités conditionnelles et du niveau de signification. Pour notre part, nous aimerions mieux connaître l'articulation d'une conception de l'interprétation comme reconnaissance de forme sollicitant plutôt l'éduction, avec celle conçue comme élaboration de connaissances nouvelles mobilisant plutôt la déduction et l'induction. En effet, d'un point de vue praxéologique, l'action d'enseigner l'interprétation statistique ne peut se réaliser en s'inscrivant dans une ligne théorique fixe déterminée par une théorie de l'interprétation, comme cela peut être le cas dans une perspective de recherche.

### 3.3 Formation à l'autonomie du sujet

La perspective de nos propos précédents intègre le processus d'autonomisation, en concevant le développement de l'autonomie du sujet à partir d'activités sollicitant des conduites autonomes dans la résolution de problèmes méthodologiques faisant appel

à la statistique. Elle y intègre aussi un processus de conceptualisation sous-jacente à l'action. Nous donnons une place et un rôle importants à l'interprétation statistique dans le développement de l'éducation statistique. Nous considérons que le développement de la capacité à réaliser une interprétation statistique requerrait une élévation du degré d'autonomie du sujet tant vis à vis des connaissances statistiques que de l'enseignant. Dans le cadre qui nous intéresse, un seuil décisif du développement de l'autonomie est franchi quand l'étudiant est capable de transférer ses connaissances en statistique lors de l'activité de modélisation dans le cadre des travaux d'étude conduisant à l'élaboration de mémoire. C'est ce que nous avions commencé à explorer et présenté au sujet du *transfert de connaissance en statistique* (Régnier 1994c) Notre question centrale demeure de savoir comment une formation en statistique acquise dans les conditions d'enseignement décrites et attestée par la réussite à une épreuve écrite de validation universitaire peut-elle conduire l'étudiant à une utilisation différée de ses connaissances à son initiative. À l'heure actuelle, les données concernant ce phénomène, recueillies auprès d'étudiants au moyen d'entretien, nous laissent penser que, si effectivement un palier d'autonomie a été franchi, il n'en demeure pas moins qu'un nouvel étayage doit être fourni, en particulier, par l'enseignant chargé de diriger l'élaboration du mémoire qui, dans la plupart des cas, n'est pas un spécialiste de statistique, afin de faciliter le transfert de connaissances statistiques.

Nous poursuivons le projet d'un enseignement de la statistique qui, comme le dit Vygotski (Vygotski 1985 p.253) « *apporte plus que ce qu'il apporte immédiatement.* » c'est à dire un enseignement qui, non seulement, permettrait à l'apprenant d'acquérir des habiletés, mais qui lui donnerait « *la possibilité d'aller bien au-delà des résultats immédiats de l'apprentissage*. »

### 3.4 Place de la statistique dans la formation en sciences humaines et sociales.

En relation à notre pratique, nous nous intéressons davantage à la place de la statistique dans la formation en sciences de l'éducation. Actuellement nous entrevoyons pour la discipline *statistique*, au cinq positions :

- Discipline de base,
- Discipline de service, discipline-outil,
- Discipline d'ouverture,

- Discipline-objet de la didactique de la statistique,
- Discipline-objet de la recherche en statistique dans son application à la recherche en sciences de l'éducation,

Analysons maintenant ce qu'induit la prise en considération de chacune de ces positions.

### *3.4.1 La statistique comme discipline de base.*

Cette position s'appuie sur l'idée que la statistique est un domaine de connaissance indispensable de nos jours. Progressivement, les programmes de l'enseignement secondaire français intègrent à leurs objectifs, des savoirs relatifs à ce domaine. Cependant il n'a pas été institué un cours de statistique comme cela a été pour les mathématiques, les langues étrangères ou d'autres disciplines scolaires. Les connaissances visées sont abordées certes par le cours de mathématiques[7] mais également dans ceux de biologie, de géographie et d'économie. Un dossier publié par l'A.S.U.[8] fait un bilan fort instructif relatif à l'enseignement de la statistique à partir d'un inventaire de tous les lieux de formation. Il est clair que dans notre culture française, son enseignement ne s'impose pas de lui-même. Il résulte de l'effort déployé par des statisticiens professionnels ou amateurs, avant tout, convaincus de l'importance que revêt la statistique dans l'intelligibilité du monde au sein duquel nous vivons. Paradoxalement, bien que la statistique soit assez spontanément rattachée aux mathématiques, ce ne sont pas les mathématiciens ou les enseignants de mathématiques qui furent les plus actifs et militants. Dans cette communauté, la représentation dominante de la statistique s'apparente à un bricolage éloigné des mathématiques. Ceci explique en partie le fait que ce qui est prévu au sein des programmes d'enseignement des mathématiques des collèges et des lycées demeure négligé. Nous avons même constaté que lorsque la statistique est abordée, son approche est fort peu favorable à la construction d'une bonne représentation. Force est d'ailleurs de constater que la formation des enseignants de mathématiques a longtemps fait l'impasse sur ce domaine. La formation universitaire s'en tient à une

[7] Comme cela est prévu dans les nouveaux programmes actuels de mathématiques des lycées et des collèges, depuis la fin des années 90.

[8] Association pour la Statistique et ses Utilisations qui en fusionnant avec d'autres associations est devenue la Société Française de Statistique (SFdS).

théorie des probabilités conçue comme un cas particulier de la théorie de la mesure[9]. La gestion du transfert de connaissances acquises relativement à ces théories, vers les théories statistiques est laissée presque entièrement à la charge de l'étudiant puis de l'enseignant qu'il est devenu.

Or ce transfert est loin d'être immédiat. Selon nous, cela tient en particulier à cette tension dialectique évoquée dans notre conception et qui, pour désigner la statistique, nous conduit à recourir au signifiant complexe ci-contre :

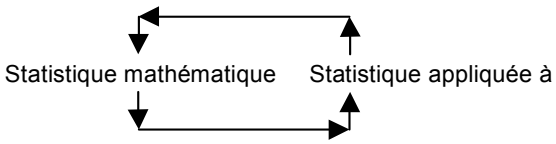


**Figure 6 : STATISTIQUE, un signifiant complexe**

De ce fait, le rattachement d'un fait observé à un concept statistique théorique, processus de modélisation, ou la réalisation d'un concept de la théorie par un phénomène de la pratique, processus d'interprétation, constituent des processus complexes qui requièrent des connaissances et des compétences caractérisant ce que nous appelons une *éducation statistique*. Celle-ci est le fruit d'une formation à la fois théorique et pratique instituée à l'école, mais aussi d'une pratique consciente menée hors de l'école et fondée sur les traitements de l'information massive quotidiennement déversée par les officines spécialisées au travers des médias à la demande des divers clients. Chercher à donner un sens aux résultats publiés d'une enquête d'opinion est tout à la fois un effet de l'éducation statistique et une cause pour son développement. Habitude et habileté sont indissociables d'une éducation statistique.

Une première visée est de générer chez l'individu des habitudes de vigilance telles que manifester une réaction critique, dresser une oreille attentive face à des propos se fondant sur des faits statistiques, s'interroger sur la façon dont les informations traitées ont été collectées, sur les traitements choisis, sur la façon dont ces traitements ont été conduits et les résultats obtenus, sur la conduite de l'interprétation et enfin sur la manière dont une décision a été prise en liaison avec l'interprétation.

[9] Une probabilité est une mesure de masse totale égale à 1.

Une seconde visée est de développer des habiletés qui lui permettent d'étayer, de justifier des propos issus du regard critique et de sa vigilance à l'égard de faits corroborés ou corrélés par des faits statistiques.

Ainsi la nécessaire vigilance à l'endroit des conclusions portant sur des imputations causales et l'habileté à réfuter ou conforter ces propositions, nous semblent fortement finaliser l'enseignement de la statistique en tant que discipline de base. Sous cet angle, sa place s'avère justifiée au sein du curriculum de sciences de l'éducation. Dans cette logique, les objectifs de référence que nous fixons actuellement, reprennent ceux visés par l'enseignement secondaire en ce qui concerne la statistique descriptive. Cela tient à la spécificité des étudiants de sciences de l'éducation dont la plupart ne maîtrise pas les notions et techniques de base même réduites à leur forme algorithmique. Par exemple, il en est ainsi des notions de moyenne, de variance, d'écart-type, de médiane ou d'histogramme. Puis nous les complétons par quelques-unes du domaine de la statistique inférentielle. À un stade plus utopique où les objectifs visés par l'enseignement secondaire seraient atteints par la grande majorité des étudiants, nous pourrions, sans négliger les nécessaires retours en arrière, envisager des objectifs de référence intégrant des connaissances et des outils plus récents. Nous pensons, par exemple, aux apports du domaine des analyses factorielles des données. C'est alors qu'un tel enseignement se doit d'intégrer un usage systématique et maîtrisé de logiciels informatiques. L'usage d'un tableur comme Excel constitue un premier niveau déjà fort riche dans les versions actuelles qui intègrent une large palette de fonctions statistiques et graphiques. Mais il convient de recourir à des logiciels de statistique même.

Toutefois depuis 1995, par le biais d'encadrement des recherches conduites par les étudiants de maîtrise, nous avons introduit l'usage du logiciel CHADOC[10], puis de celui SPAD_N[11] pour ce qui concerne les données numériques ou des données qualitatives fermées. Et même, dans le cours de méthodes qualitatives en maîtrise, nous abordons une introduction aux traitements de données textuelles assistés du

[10] Logiciel sous DOS produit par IUT Nice (Lemaire et al.)

[11] Logiciel sous Windows 32bits produit par le CISIA, Système Portable pour l 'Analyse des Données Numériques

logiciel SPAD_T[12]. Les résultats qui apparaissent dans quelques mémoires de maîtrise, sont particulièrement encourageants.

Paradoxalement, le recours à des traitements assistés par ordinateur, s'il décharge l'étudiant de tâches fastidieuses de calcul, impose un accroissement de sa vigilance pour produire des interprétations. L'espace dégagé par une moindre expertise dans le domaine calculatoire est à occuper au profit d'une plus grande expertise dans le choix des modèles de traitements et dans l'interprétation de leurs produits. Nous avons construit des exemples de calculs assistés par calculatrice donnant des résultats aberrants par effet d'approximation au voisinage des zones limites (Régnier 1998e)

Pour nous, la statistique en sciences humaines et sociales demeure une discipline de base dont l'enseignement a pour objectif la formation d'un utilisateur éclairé et autonome relativement aux outils fondamentaux apportés par cette discipline, sans toutefois avoir pour visée la formation d'un statisticien. À ce titre, nous considérons que son enseignement ne peut qu'être obligatoire.

### *3.4.2 La statistique comme discipline de service.*

Par discipline de service, nous entendons discipline-outil au service des études et des recherches en sciences de l'éducation. Certes en tant que discipline de base, la statistique a à remplir sa fonction de discipline-outil, mais l'inverse n'est plus vrai. En tant que discipline de service, nous pourrions éventuellement la considérer comme optionnelle, non sans risque. En tout état de cause elle resterait fortement conseillée et les travaux d'étude pourraient comporter systématiquement une approche statistique pour limiter l'évitement. À l'idée de *service* pourraient être rattachées celles d'*offre* et de *demande*. Dans le domaine des sciences de l'éducation, les enseignements sont *offerts*, dans le cadre d'un *service public*, supposant répondre à une *demande à la fois sociale et individuelle*. Depuis 1991, nous avons cherché à explorer et à comprendre cette *demande*. Par l'intermédiaire de questionnaires et d'entretiens, nous avons tenté d'identifier les représentations et les croyances de nos étudiants à l'égard de la statistique et de ses usages. L'étude : « *La statistique : ses représentations et ses usages didactiques et pédagogiques à l'école élémentaire*. »

[12] Logiciel sous DOS produit par le CISIA, Système Portable pour l 'Analyse des Données Textuelles

(Coutanson 1999) conduite par Bernard Coutanson sous notre direction confirme le déficit de formation en statistique chez les enseignants de l'école primaire et même parfois le désintérêt fondé sur des représentations des connaissances requises qui les placent à un niveau excessif de complexité. Notons que cette catégorie d'enseignants constitue une partie des étudiants traditionnellement accueillis en sciences de l'éducation.

Cependant offrir un enseignement de discipline-outil ne se réduit pas à enseigner des notions-recettes et des algorithmes-recettes à visée utilitariste. De nos investigations, il ressort que les questions articulées autour de l'idée centrale : de quoi avez-vous besoin qui se rattache à la statistique ? suscitent des réponses plutôt floues ou souvent pauvres pour pouvoir construire une offre de formation de statistique congruente à la somme des demandes individuelles, adressables en début de cycle universitaire. Quand elles sont sous-tendues par une logique de validation, c'est à dire celles des étudiants qui fixent le primat de l'obtention du diplôme sur la formation, ces réponses évoquent essentiellement l'évitement par la remise en cause de la fiabilité même des modèles statistiques ou par le coût excessif d'apprentissage. Sous-tendues par une logique de formation qui pose le primat de l'acquisition des connaissances et des compétences sur la certification, elles évoquent certes le bien-fondé mais en déplorant le coût d'apprentissage en relation à la durée et même la difficulté de concevoir clairement ce à quoi peut bien servir la statistique.

Durant les trois années universitaires de 1995 à 1998, nous avons assuré un *séminaire libre* d'une heure hebdomadaire, sorte de consultation statistique, visant à offrir à tout étudiant désirant recourir à des méthodes statistiques dans le cadre des travaux de recherche liés à la réalisation du mémoire[13], une assistance technique. Cette séance est aussi l'occasion de découvrir par l'usage, un logiciel de traitement statistique : CHADOC[14] Ces séances qui se déroulaient dans une salle équipée d'ordinateurs, constituèrent un cadre pour nos investigations sur les attentes des étudiants quant à la discipline-outil statistique. Elles nous ont apporté des

[13] Ici, l'élaboration du mémoire concerne la licence, la maîtrise et le DEA. À ces étudiants viennent s'ajouter des doctorants.

[14] extrait de la présentation du cours intitulé *Aide à la mise en pratique des approches statistiques* figurant dans le *Guide de l'étudiant de l'ISPEF*- Lyon 2

informations relatives aux effets de l'enseignement de statistique que nous assurions et à la question du transfert des connaissances. Ce que nous en avons tiré, c'est que la demande des étudiants ne s'exprime que rarement durant la période de construction de la problématique de leur mémoire. Certes, d'une manière confuse et vague, ils envisagent bien comment ils vont opérer la validation/réfutation de leurs hypothèses. Mais la réflexion méthodologique, c'est à dire celle portant sur les méthodes à mettre en œuvre, sur leur pertinence, sur leur faisabilité, ne parvient pas à anticiper suffisamment tôt les problèmes relatifs à la construction des données et à leurs traitements. Ce n'est qu'après avoir recueilli leurs données par questionnaire ou par entretien, que, démunis face à la masse d'informations, les étudiants lançaient leur appel à l'aide et venaient nous rencontrer. Notre intervention en tant qu'expert à ce niveau du *chemin*, peut s'avérer délicate. En effet des données de mauvaise qualité issues d'outils trop défectueux ou de protocoles trop peu fiables ou des données non pertinentes relativement à l'objet d'étude ne peuvent produire des résultats à la hauteur des attentes qu'exprime la demande de l'étudiant, même soumises à des traitements sophistiqués fondés sur des outils statistiques robustes[15]. L'offre ne peut alors satisfaire la demande. Nous avons pu observer chez certains étudiants un effet de retour qui les confortait dans leur désintérêt pour la statistique et dans l'inutilité de leur effort, dans la mesure où nous nous trouvions dans l'incapacité de mobiliser des traitements statistiques satisfaisants re-qualifiant les données. Les échanges que nous avions au cours des entretiens, finissaient par faire surgir qu'ils se réjouissaient de n'avoir pas trop investi pour apprendre la statistique. Toutefois nous avons pu aussi observer des conduites inverses par lesquelles les étudiants s'engageaient dans un processus d'apprentissage du type tâtonnement expérimental, par lequel ils ont accédé à des connaissances en statistique et à un niveau de compétence leur permettant de conduire des traitements statistiques en autonomie L'usage même du logiciel CHADOC a favorisé ce développement.

[15] La robustesse en statistique correspond au fait que des traitements peuvent être appliqués même si les conditions théoriques mathématiques ne sont pas absolument vérifiées. Les propriétés des outils robustes demeurent acceptables dans un voisinage des conditions théoriques. Exemple : un test qui requiert que la variable étudiée soit de type Laplace-Gauss, reste applicable pour une variable non gaussienne dont l'histogramme suggérerait un histogramme gaussien.

La pratique de l'enseignement se réalise principalement en situation frontale, autrement dit sous forme de cours magistraux face à un groupe très hétérogène[16] de plus d'une centaine d'étudiants. Les situations problèmes que nous avons progressivement construites pour accompagner cet enseignement, constituent chacune un cadre d'étude plausible transposable à une situation qu'un étudiant pourrait mobiliser dans son étude pour le mémoire. Chacune comporte une difficulté qui a émergé lors de son étude même. Citons, pour exemple, la question des données manquantes qui surgit d'une situation visant l'étude d'une variable « nombre de réussites à un QCM » dans laquelle nous avions introduit des absents dans le tableau de série statistique correspondante. Ou encore la question du lissage de l'histogramme qui, lors de la construction d'un polygone des fréquences, permit de révéler des obstacles de type bachelardien auxquels les étudiants se sont trouvés confrontés.

Quant à l'encadrement, il se déroule sur un mode plus individuel et personnalisé permettant un recours plus systématisé à un logiciel de statistique. Là, nous avons pu trouver matière à donner un sens à la nature des variables statistiques et l'importance du choix. En effet la constitution de la base de données, c'est à dire le tableau des séries statistiques, impose des déclarations préalables des variables statistiques selon leur nature qualitative ou quantitative. Cette compétence requise par l'usager pour qu'il soit autonome dans la première phase que constitue la saisie des données, correspond à la question posée systématiquement dans les documents « proposition de traitement » des situations problèmes : de quoi s'agit-il ? À l'heure actuelle, quand il s'agit des variables qualitatives, c'est le choix même du logiciel qui est en cause : le fait qu'elles soient qualitatives fermées va renvoyer à un logiciel (Spad_N, CHADOC) alors que des variables qualitatives ouvertes requièrent un autre logiciel (Spad_T).

[16] Cette hétérogénéité est à considérer selon plusieurs facteurs : âge, rapport à la statistique, parcours personnel et professionnel, niveau de connaissance et de compétence dans le domaine de la statistique, stratégies d'étude (logique de formation, logique de certification)

### *3.4.3 La statistique comme discipline d'ouverture.*

Nous entendons par discipline d'ouverture, le fait de considérer la statistique comme une discipline orientée vers le développement de la culture générale de l'individu sans intervenir explicitement dans la sélection ou dans les parcours de formation universitaire. Dans le cursus, elle peut alors être placée comme optionnelle et facultative. Ce cas de figure permettrait d'accueillir des étudiants volontaires et curieux. Pour attirant que cela soit pour l'enseignant, réduire l'enseignement de la statistique à cette position expose alors trop facilement les étudiants au choix de l'évitement des obstacles à franchir pour apprendre. Le contrat didactique qui se noue, est appelé à intégrer ouverture et exigence. Ceci constitue un défi plus délicat qu'il n'y paraît dans le contexte culturel d'une licence ou d'une maîtrise de science de l'éducation.

### *3.4.4 La statistique comme discipline-objet de la didactique*

La statistique conçue comme un des objets de la didactique de la statistique est liée aux sciences de l'éducation par deux aspects. Le premier tient au fait que nous rattachons la didactique disciplinaire à ce champ des sciences de l'éducation et qu'une recherche en didactique de la statistique oblige à explorer la discipline même. Le second est ordonné au fait qu'un enseignement de statistique est dispensé en licence et maîtrise. Ce dernier provoque des phénomènes dont nous avons déjà parlé dans les positions précédentes, et qui soulèvent des questions à l'égard de la statistique impliquant des prises de position dans le champ même des sciences de l'éducation. Une position est de remettre en cause le bien-fondé d'un enseignement obligatoire de statistique dans les cursus. Position que nous comprenons mais contre laquelle nous luttons avec des arguments méthodologiques et épistémologiques. (Régnier 1998c)

### *3.4.5 La statistique comme discipline-objet de la recherche en statistique dans son application à la recherche en sciences de l'éducation*

Nous pensons que la recherche en sciences de l'éducation peut rencontrer des problèmes méthodologiques qui appelleraient eux-mêmes des recherches dans le domaine de la statistique. L'histoire de la statistique illustre bien, par exemple, la part des recherches en psychologie expérimentale dans son développement. Certes la

panoplie des outils offerts actuellement paraît largement satisfaire les besoins des chercheurs en sciences de l'éducation. Mais rien ne dit que des points de vue nouveaux n'enrichiraient pas ces recherches. Par exemple, dans cette perspective, nous avons tenté d'importer une approche, par ailleurs bien connue à partir des travaux de Warner (Warner 1965), préservant l'anonymat des individus interrogés sur des questions sensibles (Régnier 1992).

Mais les questions qui nous intéressent le plus et devant lesquelles nous n'avons pas encore de réponse suffisamment générale sont celles attachées aux statistiques de rang. Il nous semble que, dans nombre de recherches en sciences de l'éducation, le rang constitue une variable particulièrement bien adaptée. Ainsi quand nous souhaitons savoir s'il existe une tendance à l'accord dans un groupe relativement à des choix préférentiels comme dans l'exemple ci-dessous, où il est demandé à cinq stagiaires de ranger par ordre décroissant d'importance les énoncés relatifs aux raisons qui les ont conduit à suivre un stage. Dans le cas présenté, les cinq individus ont fourni une réponse complète, c'est à dire une des 8 ! permutations. Le test de concordance de p rangements de n objets de Friedman ou celui de M. G. Kendall permettent de tester l'hypothèse $H_0$ d'indépendance des jugements contre l'hypothèse alternative $H_1$ de tendance à la concordance des jugements. La statistique utilisée est d'une forme mathématiquement connue et a été tabulée. Ceci permet de travailler à un niveau de risque de première espèce  connu. Des difficultés commencent à apparaître quand les individus posent quelques *ex æquo*.

**Tableau 2: Exemple de recueil des rangements préférentiels**

| | items | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ***individus*** | **n°1** | **n°2** | **n°3** | **n°4** | **n°5** | **n°6** | **n°7** | **n°8** |
| **Stagiaire01** | 1 | 3 | 5 | 8 | 6 | 7 | 2 | 4 |
| **Stagiaire02** | 4 | 1 | 7 | 5 | 6 | 2 | 8 | 3 |
| **Stagiaire03** | 4 | 1 | 6 | 3 | 2 | 8 | 5 | 7 |
| **Stagiaire04** | 1 | 6 | 4 | 2 | 3 | 8 | 7 | 5 |
| **Stagiaire05** | 5 | 2 | 3 | 4 | 7 | 8 | 6 | 1 |
| **Sommes des rangs** | 15 | 13 | 25 | 22 | 24 | 33 | 28 | 20 |

| **n°** | **Énoncés des items** |
|---|---|
| ***1*** | Actualiser ses connaissances théoriques professionnelles |
| ***2*** | Sortir de la routine quotidienne |
| ***3*** | Se confronter à des situations de formation |
| ***4*** | Compléter sa formation |
| ***5*** | Espérer une meilleure reconnaissance professionnelle |
| ***6*** | Se donner l'occasion de mieux comprendre les difficultés rencontrer dans l'exercice du métier |
| ***7*** | Avoir une occasion de rencontrer des collègues et d'échanger |
| ***8*** | Espérer améliorer les pratiques professionnelles quotidiennes |

Mais un problème est apparu lors d'une étude menée par un étudiant avec le protocole suivant que nous adaptons à notre exemple : *Choisir les quatre items qui correspondent le mieux à votre situation et les ranger dans l'ordre décroissant d'importance.* Dans cette circonstance, la statistique mise en œuvre requiert un traitement particulier pour obtenir un test d'hypothèse analogue au cas précédent.
Nous avons même vu, dans une autre étude, un cas beaucoup plus général dans la mesure où les individus n'avaient pas choisi le même nombre d'items à ranger et avaient mis des *ex æquo*. Nous n'avons pas trouvé, à ce jour, de traitement général de ce cas. Il reste une problématique statistique de construction d'un test de concordance des rangements de $k_p$ objets pris parmi n objets selon p critères synthétiques. Chacun des p individus détermine un critère synthétique et le nombre k varie d'un individu à l'autre.

Pour continuer, nous souhaiterions aborder la question des variables qualitatives et de leurs traitements. Ce type nous apparaît comme particulièrement adapté aux études des phénomènes éducationnels où la mesure quantitative s'avère souvent difficile et parfois peu pertinente. Mais alors plus de moyenne ni de variance ! L'inconfort semble ressurgir car nous avons pu constater maintes fois que des chercheurs n'hésitent pas à coder numériquement les modalités dans une visée quantitative. Du codage numérique à la mise en œuvre des traitements réservés aux variables quantitatives, il n'y a qu'un pas que d'aucun n'hésite nullement à franchir.

Nous ne savons pas encore en quoi la modalité « tout à fait satisfaisant » codée 4 pouvait être la somme de quatre modalités « tout à fait insatisfaisant » codée 1 ! Nous ne parvenons pas à comprendre que la satisfaction soit la somme d'insatisfactions élémentaires.

Plus nous aurons de connaissances relatives aux traitements des variables qualitatives, plus nous disposerons d'outils adaptés aux recherches en sciences de l'éducation.

Enfin en ce qui concerne les questions de la représentativité et de la significativité, nous pensons qu'ici aussi des recherches pourraient être conduites dans le domaine particulier des théories du sondage et de l'échantillonnage. En effet, les recherches en sciences de l'éducation font fréquemment référence à des échantillons de petite taille n obtenus, d'une part, difficilement par des méthodes aléatoires et, d'autre part, généralement par tirage sans remise au sein de population de taille finie N. Les conditions classiques d'application des tests statistiques ou des estimations statistiques ne sont guère remplies. Aussi l'exploitation d'outils plus adéquats serait à mieux considérer, et s'ils font défaut, ce serait l'occasion de développer de tels outils au travers d'une collaboration avec des statisticiens professionnels.

Pour achever cette discussion, nous reprenons cette phrase de Michel Develay (Develay 1992 p.90) « *Nous sommes convaincu que le didactique et le pédagogique constituent deux domaines en interactivité forte, au sein des sciences de la décision…* ». Or évoquer ces sciences de la décision renvoie aux sciences statistiques qui constituent un pourvoyeur de méthodes, de techniques et de notions particulièrement adéquates à la construction de modèles décisionnels. Tout un pan de la statistique est articulé à la thématique de la prise de décision en situation incertaine et sur informations partielles. Et là, il nous paraît utile de mieux cerner les outils pertinents à mobiliser dans la modélisation des situations didactiques et pédagogiques. Ces modèles seront d'autant plus performants qu'ils pourront intégrer des niveaux de complexité de plus en plus élevés. Dans la boîte à outils statistiques, cela correspond à la possibilité de prendre en compte simultanément plusieurs variables éventuellement de nature distincte. Ceci justifie à nos yeux l'intérêt d'une recherche en statistique appliquée à la recherche en sciences de l'éducation.

# 4 La formation en statistique : des questions pour la didactique de la statistique.

## 4.1 Promouvoir l'enseignement et la didactique de la statistique en sciences de l'éducation.

Notre parti pris est que la statistique en sciences humaines et sociales et, en particulier, en sciences de l'éducation, doit être un compromis entre la discipline de base et la discipline de service. Ainsi la formation en statistique devrait être obligatoire en raison de la thèse que nous défendons que l'élévation du niveau d'éducation statistique contribue à divers degrés au développement de l'autonomie de l'être humain, citoyen d'une communauté démocratique moderne qui fait grand usage des instruments statistiques. L'acquisition de cette formation reste principalement conçue par la médiation d'un enseignement dans le cadre de l'institution universitaire, mais adapté aux étudiants dont la grande majorité ne dispose que d'un faible bagage d'outils mathématiques et d'un niveau de compétences en mathématiques dépassant rarement celui de la classe de seconde. Qui plus est, comme nous l'avons évoqué plus haut, le rapport affectif qu'ils entretiennent avec les mathématiques et, par-delà, avec la statistique, est souvent négatif. Ici sont réunies les conditions les plus favorables à l'entretien d'une résistance face à un enseignement dont ils n'entrevoient pas d'autres finalités que celles issues de leurs représentations ; lesquelles se sont formées autour de concepts spontanés ou de concepts scientifiques des domaines des mathématiques et de la statistique restés à un niveau trop élémentaire de conceptualisation Le concept de moyenne est celui qui nous sert d'indicateur le plus pertinent pour illustrer le niveau élémentaire de la conceptualisation Notre expérience d'enseignement au Brésil dans un autre cadre culturel nous a aussi permis d'y réaliser les mêmes constats. Les enquêtes relatives au rapport à la statistique et aux représentations relatives à la statistique, que nous avons menées auprès d'étudiants de psychologie et de sciences de l'éducation, fournissent des résultats identiques à ceux que nous avons obtenus dans le cadre de l'enseignement en France. La question cruciale qui se pose à nous, est donc bien celle de savoir comment un enseignement de statistique, discipline de base/discipline-outil, peut-il être réalisé avec efficience pour des étudiants non-spécialistes des disciplines scientifiques dont les caractéristiques dominantes sont celles que nous avons constatées ci-dessus.

### 4.2 Un petit état des lieux de la didactique de la statistique.

À l'appui de notre projet de voir se développer le domaine de la didactique de la statistique, nous possédons deux références. Un épais rapport de Yves Chevallard (Chevallard 1978) intitulé *Notes pour la didactique de la statistique* aborde successivement le problème de la mesure, celui de la relation entre deux variables, la problématique générale de la statistique, la question de la transposition didactique dans l'enseignement de la statistique et la notion de moyenne en statistique comme problème de surdétermination en didactique. Ces études s'inscrivaient dans le cadre du projet ministériel de *rénovation de l'enseignement de la statistique aux étudiants en psychologie*. Yves Chevallard précise qu'il s'agit d'un travail en pluridisciplinarité et que « *les analyses présentées sont des analyses didactiques, ou très précisément, pour plusieurs d'entre elles, des analyses mathématiques ou épistémologiques à visée didactique.*» Au travers un propos dense conduit tout au long des 400 pages, nous y avons retrouvé nombre de nos réflexions et de résultats de nos observations dans notre posture d'enseignant-chercheur, praticien réflexif. En particulier, dans sa note sur la transposition didactique, Yves Chevallard caractérise l'analyse statistique par ses 5 étapes problématiques :

1. Les problèmes de recueil des données
2. Les problèmes de l'analyse qualitative des données recueillies et de leur réduction éventuelle.
3. Les problèmes de la quantification.
4. Les problèmes de la signification.
5. Les problèmes de l'interprétation.

Et c'est aussi à cette dernière étape qu'il identifie le plus de difficultés car elle concentre les effets pervers drainés par les précédentes.

Dans son propos, nous reconnaissons des éléments de notre propre conception de la statistique et en particulier la tension dialectique à laquelle est soumise la connaissance statistique. (Régnier 1998b p.9). Cette tension lie les deux dimensions de la statistique comme les deux pôles inséparables d'un aimant. Comme nous l'avons déjà dit plus haut, nous les nommons respectivement *statistique mathématique* et *statistique appliquée à….* S'il est impossible d'envisager la statistique mathématique hors du cadre théorique au sein duquel se développent

l'explicitation, la formalisation des notions, concepts, méthodes et des raisonnements de la statistique, nous ne pouvons pas non plus imaginer que cette théorie soit coupée de sa mise en œuvre dans d'autres cadres théoriques tels que la psychologie, l'économie, la mécanique, la médecine, etc. Le schéma (Figure 7) ci-dessous précise ce que nous avons désigné par le signifiant complexe : statistique. (voir Figure 6)

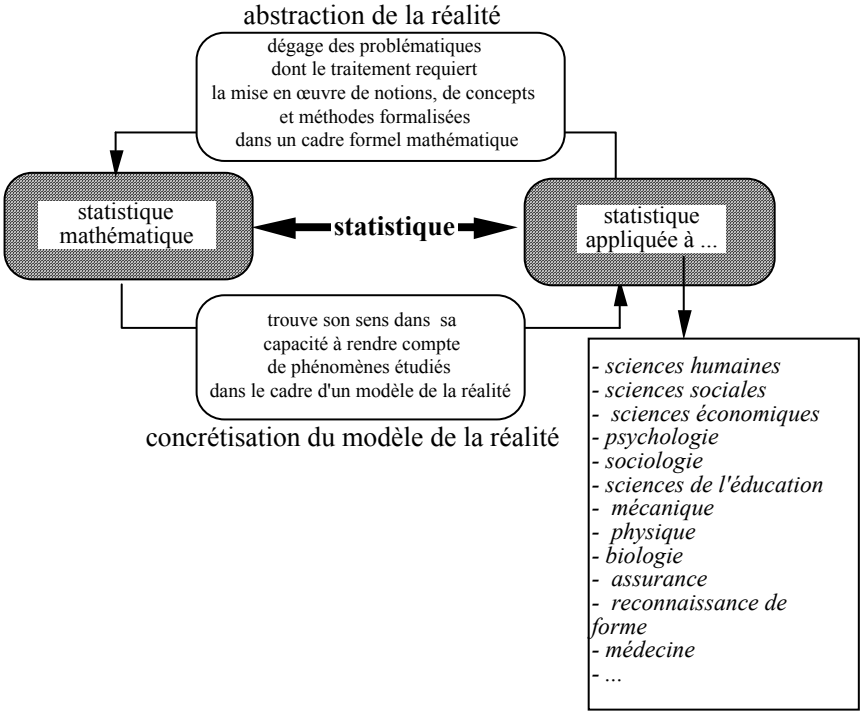


**Figure 7 : Schéma de notre conception de la statistique.**

Yves Chevallard insiste sur la nécessité de resituer l'analyse statistique dans le contexte où elle a été introduite. C'est ce que nous tentons de faire dans nos séquences didactiques avec les questions évoquées plus haut et auxquelles nous confrontons les étudiants à l'étape "traitement des données": De quoi s'agit-il ? puisque les données fournies ne sont pas le résultats de leur recueil ; Quelle sont votre conclusion et votre interprétation ? pour contrôler l'effet de l'oubli du problème que l'on pose et auquel on tente de répondre par la médiation de l'analyse statistique. Toutefois dans notre dispositif pédagogique, la conclusion intègre aussi la 5ème étape.

Dans le même temps, il met l'accent sur un effet de la transposition didactique dans l'enseignement de la statistique, qu'il nomme déterritorialisation, sur lequel nous insistons avec d'autres termes dans notre discours pédagogique. Depuis plusieurs années, nous cherchons, en fait, à résoudre le problème qu'il décrit ainsi : « *L'analyse statistique intervient toujours dans un champ de savoir structuré par certaines problématiques propres, — un territoire —, et la réduction didactique, en l'extirpant*

*de ce territoire la prive de son sens, et ne lui laisse guère qu'un intérêt le plus général, et donc le plus pauvre aussi. Tel est l'effet de cette déterritorialisation. Celle-ci a des effets d'autant plus pernicieux qu'elle est mieux cachée : un nombre imposant d'ouvrages se présentent aujourd'hui comme des introductions à la statistique pour le psychologue, pour le géographe, etc. . En fait, à peu de chose près, ces ouvrages disent tous la même chose. Le champ extra-statistique auquel ils se réfèrent chacun, n'est là que pour fournir des illustrations, et non des problèmes.* » . (Chevallard, 1978, *Transposition*… p.5)

La perspective épistémologique adoptée par Yves Chevallard le conduit à dégager la problématique propre à la statistique comme « *une visée s'efforçant de saisir dialectiquement, et de traduire en des énoncés validables d'un langage scientifique, les régularités et les perturbations dans les phénomènes marqués de variabilité.* » (Chevallard, 1978, *Transposition*… p.17).

Ainsi avec lui nous retrouvons un problème didactique fondamental : comment, dans le cadre d'un enseignement, peut-on amener les sujets-apprenants à entrer dans la problématique statistique ? Comment opérer une transposition didactique qui n'hypothèque ni le savoir statistique en sa complexité et sa richesse historiques, ni le sujet en sa *socialisation cognitive* ?

La seconde référence est constituée par *Didáctica de la Estadística* de Carmen Batanero (Batanero 2001). Au travers des 200 pages, elle brosse un panorama de la situation actuelle de la didactique de la statistique et de ses perspectives futures, esquisse les fondements épistémologiques, expose des recherches sur le raisonnement statistique et les difficultés d'apprentissage, aborde la question du curriculum de statistique et donne des exemples de projets pour la classe. Des nombreux points intéressants qu'elle développe, nous retenons les deux suivants, à propos de la spécificité de la statistique dans la didactique des mathématiques, « *La misma naturaleza de la estadística es muy diferente de la cultura determinista tradicional en clase de matemáticas. Un indicador de ello es que aun hoy día prosiguen las controversias filosóficas sobre la interpretación y aplicación de conceptos tan básicos como los de probabilidad, aleatoriedad, independencia o contraste de hipótesis, mientras que estas controversias no existen en álgebra o geometría. Las dimensiones políticas y éticas del uso y posible abuso de la estadística y la información estadística contribuyen, asimismo, a la especificidad de la campo* » et « *La naturaleza interdisciplinar del tema, que hace que los conceptos*

*estadísticos aparezcan en otras materias, como ciencias sociales, biología, geografía, etc., donde los profesores, a veces se ven obligados a enseñar estadística, lo que puede ocasionar conflictos cuando las definiciones o propiedades presentadas de los conceptos no coinciden con las impartidas en la clase de matemáticas.* » (Batanero 2001 p.7)

Ces dimensions de la statistique sont à prendre en considération dans l'approche de la problématique didactique énoncée précédemment.

À propos des manuels de statistique, elle rapporte que des recherches didactiques conduites montrent comment quelques erreurs conceptuelles et une approche pédagogique inadéquate apparaissent avec une fréquence plus forte qu'il n'est souhaitable. ( Sánchez-Cobo 1996; Ortiz, 1999). Ceci va dans le sens que nous avons exposé à propos de la notion d'histogramme et de son usage dans divers manuels scolaires (Régnier 1998a).

En ce qui concerne d'autres références, nous constatons avec Carmen Batanero que le nombre de recherches du domaine de la didactique de la statistique est encore très faible en comparaison de celles conduites dans les autres branches des mathématiques. En effet nous avons pris l'exemple de la base de données constituée par l'ARDM[17] et par l'I.N.R.P[18]. sous la rubriques : thèses et HDR en didactique des mathématiques. Nous obtenons 269 références entre les années 1968 et 2000.

[17] URL = http://www.ardm.asso.fr/equipe/france/these.html
[18] URL = http://www.inrp.fr/Didactique/These/Index.htm

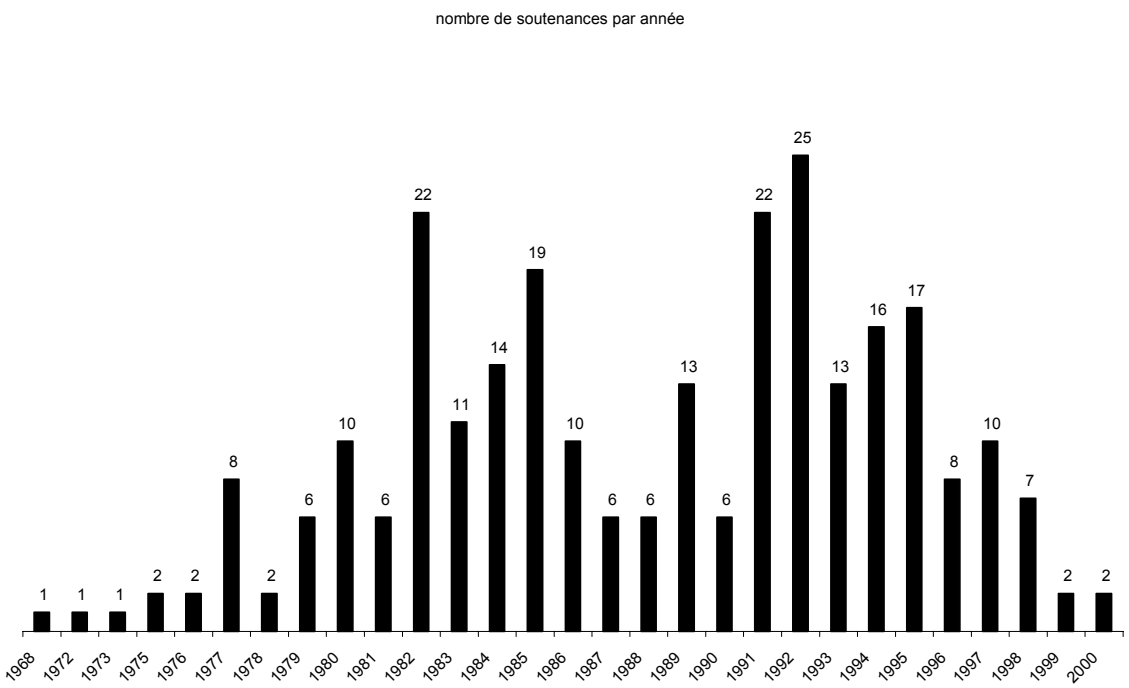


**Figure 8 : Nombre de thèses soutenues annuellement sur la période 1968-2000 inscrites sur le site de l'ARDM**

En analysant les titres et en mettant à part notre contribution (Régnier 2000), il appert que nous ne relevons que 8 thèses (voir Tableau 3) abordant des problématiques apparentées à des questions de didactique de la statistique, dans la mesure où la théorie des probabilités est un cadre dans lequel puise la statistique inférentielle. D'un point de vue plus restrictif, nous n'identifions que 2 thèses du champ de la didactique de la statistique :

- celle de Claudine Blanchard-Laville qui s'intéresse aux représentations des étudiants et au rapport affectif qu'ils entretiennent avec la statistique et les mathématiques
- celle Claudine Dhuin qui est centrée sur la notion d'estimation statistique.

La *simulation* est un processus mentionné 2 fois pour l'approche du concept de probabilité. Ce processus est un élément clé des programmes actuels de mathématiques du lycée. La notion *simuler une expérience* est prise dans le sens de « *choisir un modèle de cette expérience puis simuler ce modèle.*(…) *La simulation permettra de disposer d'échantillons de grande taille et d'observer des phénomènes appelant une explication dans le champ des mathématiques. Pour bien comprendre les mathématiques, il est utile d'apprendre quel type de questions sont à adresser à*

*cette discipline et aussi à reformuler ces questions dans le langage propre des mathématiques.»* [19]

**Tableau 3 : thèses pouvant être apparentées au champ de la didactqiue de la statistique**

| | | |
|---|---|---|
| 1980 | BLANCHARD-LAVILLE Claudine | Les étudiants de psychologie face à l'enseignement de statistiques ( analyse des réponses à un test de mathématiques et à des questionnaires d'opinion.) |
| 1982 | ALARCON Jesus | L'appréhension des situations probabilistes chez les élèves de 12-14 ans : résultats d'une enquête proposée à des élèves de 4e et de 5e. |
| 1985 | DHUIN Claudine | Compréhension de l'estimation statistique chez les étudiants de 1er cycle de gestion et d'économie appliquée. |
| 1986 | MAURY Sylvette | Contribution à l'étude didactique de quelques notions de probabilité et de combinatoire à travers la résolution de problèmes.. |
| 1989 | ZAKI Moncef | Traitements de problèmes de probabilités en situation de simulation. |
| 1991 | BORDIER Jacques | Définition, construction et expérimentation de logiciels de simulation sur ordinateur, pour l'enseignement de la probabilité. |
| 1992 | TOTOHASINA André | Méthode implicative en analyse de données et application à l'analyse de conceptions d'étudiants sur la notion de probabilité conditionnelle. |
| 1998 | LAHANIER REUTER Dominique | Étude de conceptions du hasard : approche épistémologique, didactique et expérimentale en milieu universitaire. |

Nous avons procédé à une analyse lexicométrique des 269 titres selon la méthodologie d'analyse des données textuelles ( Lebart, Salem, 1994) assistée par le logiciel SPAD_T. Il ressort que :

**Tableau 4 : Domaines étudiés en didactique**

| *Cadre théorique* | *Nombre de titres renvoyant au cadre* |
|---|---|
| Algèbre | 19 |
| Arithmétique | 2 |
| Passage Arithmétique à Algèbre | 2 |
| Analyse | 1 |
| Géométrie | 24 |
| Logique | 7 |
| Probabilités | 6 |
| Trigonométrie | 1 |

Les formes graphiques *statistique* et *statistiques* apparaissent respectivement 5 et 1 fois sur les 945 utilisées pour formuler les 269 titres. La forme *statistiques* figure dans le titre de la thèse de Claudine Blanchard-Laville. L'édition des contextes ( voir Tableau 5 ) met en évidence que le terme *statistique* est utilisé comme adjectif 4 fois

[19] *Op. Cit*. programme / juin 2000 ; p. 8-9

sur 5 et qu'à côté de l'emploi déjà repéré dans la thèse de Claudine Dhuin : *estimation statistique,* il renvoie par 3 fois l'*analyse statistique* en tant que méthode d'étude mise au service de la recherche en didactique des mathématiques et comme objet d'enseignement-apprentissage traité comme un objet de recherche de la didactique de la statistique.

**Tableau 5 : Contextes du mot: STATISTIQUE dans le titre de la thèse**

| | | | *Auteurs* |
|---|---|---|---|
| COMPREHENSION DE L ESTIMATION | **STATISTIQUE** | CHEZ LES ETUDIANTS DE 1ER CYCLE DE GESTION ET D ECONOMIE | DHUIN Claudine |
| IMPLICATION | **STATISTIQUE** | ET APPLICATIONS A L ANALYSE DE DEMARCHES DE PREUVE EN | LARHER Annie |
| ANALYSE | **STATISTIQUE** | IMPLICATIVE | BAILLEUL Marc |
| ETUDE DIDACTIQUE APPLICATION DE METHODES D ANALYSE | **STATISTIQUE** | MULTIDIMENSIONNELLE | POLO CAPRA Maria |
| DANS L ENSEIGNEMENT DES MATHEMATIQUES ET DE LA | **STATISTIQUE** | | REGNIER Jean-Claude |

Il nous semble donc nécessaire et pertinent de développer des travaux de didactique et de pédagogie de la statistique dans les perspectives théoriques que nous avons exposées à divers moments de notre propos. Un des buts de ces recherches est d'instrumenter l'enseignement de la statistique, tant celui qui s'adresse aux adolescents des collèges et des lycées que celui qui concerne les étudiants adultes, en développant une ingénierie didactique et pédagogique adaptée dont l'efficience serait expérimentable. La construction des séquences d'enseignement-apprentissage intégreraient, d'une part, la perspective développée au travers du triangle pédagogico-didactique complexifié (Régnier 2000, p.193), d'autre part, les processus d'autocorrection (Régnier 1983), d'auto-évaluation (Régnier 1999), d'apprentissage fondé sur le tâtonnement expérimental et le travail en autonomie ( Régnier 1988, 1991, 1994, 1998f ), d'instrumentation (Rabardel, 1995)

## 5 Conclusion

Nous étions parti du constat de Jean Dercourt relatif au déficit de formation en statistique en France. Ce contexte actuel s'inscrit dans une histoire. Un regard historique nous a montré qu'en France, la formation en statistique est une question

déjà ancienne — remontant au XIX$^{ème}$ siècle — et récurrente, à laquelle peu de réponses a été donnée, car elle ne mobilise réellement que peu d'intérêt des communautés scientifiques concernées. Et c'est bien dans ce contexte que nous cherchons à construire et à résoudre, dans le champ de la didactique de la statistique, des problématiques relatives à la formation en statistique, et en particulier à des questions d'enseignement-apprentissage en milieu scolaire ou universitaire.

En portant notre attention sur le référent de la formation en statistique, nous avons tenté d'expliciter les objectifs d'intégration et de définir l'esprit statistique et son rapport à l'éducation statistique. L'esprit statistique *naît lorsqu'on prend conscience de l'existence de fluctuation d'échantillonnage* et sa formation requiert de renoncer à l'usage systématique de l'idée de *vérité* pour chercher à maîtriser celle de *vraisemblance*. En abordant dans cette perspective la question de la place de la statistique dans la formation en sciences humaines et sociales, nous avons envisagé tour à tour la statistique comme discipline de base, discipline de service, discipline d'ouverture, discipline-objet de la didactique, discipline-objet de la recherche en statistique dans son application à la recherche en sciences de l'éducation. Nous avons essayé de mieux cerner le sens de l'interprétation statistique dans l'analyse statistique, et d'aborder le problème que soulève son enseignement et son apprentissage.

Enfin nous avons affronté l'explicitation de questions soulevée par la formation en statistique dans le champ de la didactique de la statistique. Un rapide état des lieux montre qu'en France, très peu de travaux de thèse ont été consacrés à de telles questions. Nous avons repris deux chantiers pionniers ouverts respectivement par Yves Chevallard ( Chevallard 1978 ) et par Claudine Blanchard-Laville (Blanchard-Laville 1980) en relation à la rénovation de l'enseignement de la statistique en psychologie. À cette occasion, nous avons ré-abordé la question de la transposition didactique et pu retrouver le concept de *déterritorialisation* proposé par Yves Chevallard. Restant dans la perspective culturelle latine, le groupe espagnol de l'Université de Grenade dirigé par Carmen Batanero (Batanero 2001) constitue un pôle important pour le développement de la didactique de la statistique.

L'analyse textuelle des titres des thèses et HDR du domaine de la didactique des mathématiques ou en périphérie, présentes dans les bases de données de l'I.N.R.P. et

de l'ARDM montre que sur 269 seulement 9 relèvent de problématiques du champ de la didactique de la statistique, contre 24 en géométrie et 19 en algèbre.

Si nous tenons compte en plus de l'évolution des programmes de mathématiques des lycées et des collèges, il nous semble nécessaire et pertinent que soient développées des recherches en didactique de la statistique.

**BIBLIOGRAPHIE**

ACADÉMIE DES SCIENCES (2000) *La statistique* Rapport sur la science et la technologie n°8, Paris : Éditions TEC&DOC

AFFICHARD, J. (Ed.) (1987) *Pour une histoire de la statistique :* Tome 1 *Contributions,* Tome 2 *matériaux*, Paris : Economica-INSEE.

ASTOLFI, J.P (1993) Styles d'apprentissage et mode de pensée*, in* J. Houssaye (dir.) *La pédagogie : une encyclopédie pour aujourd'hui*, Paris : ESF p.301-314

BACHELARD, G., (1989) *La formation de l'esprit scientifique,* Paris *:* Librairie philosophique J. Vrin, 14ème éd., 1989, 257 p, (1ère éd. 1938)

BATANERO, C. (2001) *Didáctica de la Estadística*, Granada : GEEUG

BLANCHARD-LAVILLE C. (1980) *Les étudiants de psychologie face à l'enseignement de statistiques (analyse des réponses à un test de mathématiques et à des questionnaires d'opinion.)* Thèse de Doctorat de 3ème cycle, Université Paris VII

BLANCHARD-LAVILLE, C. (1981) Les dimensions affectives de l'apprentissage des statistiques, *Éducation Permanente (61)* pp.41-62

BORDIER, J. (1991) *Un modèle didactique, utilisant la simulation sur l'ordinateur, pour l'enseignement de la probabilité*, Thèse de Doctorat de didactique des mathématiques, Université Paris 7

BUNLE, H., (1945) Interventions concernant l'enseignement de la statistique en France, *Journal de la Société de statistique*, (mars-avril)

CHEVALLARD, Y. (1978) *Notes pour la didactique de la statistique* I.R.E.M. d'Aix-Marseille. 395 p. + annexes

CHEYSSON, E., (1890) Rapport sur l'enseignement de la statistique, *Bulletin du Conseil supérieur de la statistique*, (4)

COURNOT, A. A., (1984) *A.A. Cournot, œuvres complètes.* Tome 1 : *exposition de la théorie des chances et des probabilités*, B. Bru (Ed.), Paris : Librairie J. Vrin, 385 p.,

COUTANSON, B., (1999) *La statistique : ses représentations et ses usages didactiques et pédagogiques à l'école élémentaire*. Mémoire de maîtrise de sciences de l'éducation de l'Université Lyon 2, dirigé par J-C Régnier

DE KETELE, J-M., (1983) Évaluation et objectifs d'intégration, *in* Contribution à la rénovation de l'enseignement agricole, *Actes de la session évaluation du 27 au 30 septembre 1983. Bulletin INRAP n°58, décembre 1983*, pp. 87-116

DEVELAY, M., (1990) *Didactique et Sciences de l'éducation: vers une épistémologie scolaire,* Note de synthèse HDR 172 p.

DROESBEKE, J-J, FICHET, B., TASSI, PH., (EDS) (1987) *Les sondages,* Paris : Ed. Economica 310 p

ESCOFIER, B., PAGÈS, J., (1990) *Analyses factorielles simples et multiples : objectifs, méthodes et interprétation*, Paris : Dunod 2ème éd. 267 p.

FAURE, F., (1893) Observations sur l'organisation de l'enseignement de la statistique, *Journal de la Société de statistique*, (janv.) pp. 25-29

FAURE, F., (1894) Rapport sur l'enseignement de la statistique et les programmes d'examen d'admission dans les administrations publiques. *Bulletin du Conseil supérieur de la statistique* (5)

GLAYMANN, M., VARGA, T., (1973) *Les probabilités à l'école*, Paris : CEDIC

LAHANIER-REUTER, D., (1998) *Étude de conceptions du hasard : approche épistémologique, didactique et expérimentale en milieu universitaire.* Thèse de Doctorat de l'Université Rennes 1

LEBART, L., SALEM, A. (1994) *Statistique textuelle.* Paris : Dunod

LEBOUTET, L., ROUANET, H., (1959) *Lexique de statistique,* Groupe d'études de psychologie de l'Université de Parsi Sorbonne.

LEBOUTET, L;, (1958) Étude sur les difficultés des étudiants en statistiques, *Bulletin de psychologie*, ( Tome XI, n°7-8, mars)

MAURY, S. (1986) *Contribution à l'étude didactique de quelques notions de probabilité et de combinatoire à travers la résolution de problèmes*. Thèse de Doctorat d'état, Université Montpellier 2

MEIRIEU, PH., (1987) *Apprendre...oui, mais comment ?,* Paris : Paris : ESF Ed.,

MORICE, E., (1961) L'enseignement de la statistique, *Paris 1900*, Paris :La Ville de Paris, pp.271-282

ORTIZ de HARO, J.J; (1999) *Significado de conceptos probabilísticos en los textos de Bachillerato.* Tesis Doctoral. Departamento de Didáctica de la Matemática. Universidad de Granada.

POPPER, K., R., (1988) *La logique de la découverte scientifique*, (1973) Paris : Éditions Payot, 480 p.

RABARDEL, P., (1995) *Les hommes et les technologies : approche cognitive des instruments contemporains.* Paris : Armand Colin

RÉGNIER, J-C (2002a) *L'enseignement de la statistique dans une formation à distance de licence de sciences de l'éducation en France*. XXXIV[e] Journées de la SFdS 2002, Bruxelles-Louvain

RÉGNIER, J-C (2002b) *Les difficultés des étudiants de Sciences de l'éducation face à l'enseignement de la statistique : questions pour la didactique de la statistique.* (en préparation)

RÉGNIER, J-C, (1983) *Étude didactique d'un test autocorrectif en trigonométrie,* Thèse de Doctorat de 3ème cycle, ULP Strasbourg

RÉGNIER, J-C, (1988) Étude didactique d'une méthode d'apprentissage fondé sur le tâtonnement expérimental de l'apprenant, *Annales de Didactique et de Sciences Cognitives,* séminaire de Didactique des Mathématiques de Strasbourg, pp 255-279

RÉGNIER, J-C, (1991) *Autonomie et travail personnel dans l'enseignement des disciplines scientifiques en Lycée : - Témoignage d'un travail conduit sur une année scolaire en mathématiques* Paris : MEN-DLC, Dijon: C.R.D.P., 167 p.

RÉGNIER, J-C, (1992) Respecter l'anonymat : suggestion pour une mise en œuvre de méthodes de collectes de données respectant le secret de la réponse., *la lettre du CLERSE*, (17), pp 10-19

RÉGNIER, J-C, (1994) Tâtonnement expérimental & Apprentissage en mathématiques*,* in P. Clanché, E., Debarbieux (Eds) *La pédagogie Freinet, mises à jour et perspectives*, P.U.Bordeaux. pp 135-153

RÉGNIER, J-C, (1994a) *Statistique : documents autocorrectifs et auto-évaluatifs d'aide à l'apprentissage*, livret n°1, Lyon : APPRENDRE-Sciences de l'éducation Collection "outils de base pour la recherche en éducation", Université Lumière, (2), 1994, 69p

RÉGNIER, J-C, (1994b) *Statistique : documents autocorrectifs et auto-évaluatifs d'aide à l'apprentissage*, livret n°2, Lyon : APPRENDRE-Sciences de l'éducation Collection "outils de base pour la recherche en éducation", Université Lumière,(3),1994, 73 p

RÉGNIER, J-C, (1994c) Le transfert de connaissances en statistique*, in* Les transferts de connaissances en formation initiale et continue *Actes préparatoires du colloque international :,* 29/9-2/10/94, Université Lumière Lyon 2, pp 260-263

RÉGNIER, J-C, (1995a) *Statistique : documents autocorrectifs et auto-évaluatifs d'aide à l'apprentissage,* livret n°3, Lyon : APPRENDRE-Sciences de l'éducation Collection "outils de base pour la recherche en éducation", Université Lumière, (4), 1995, 63p

RÉGNIER, J-C, (1995b) *Statistique : documents autocorrectifs et auto-évaluatifs d'aide à l'apprentissage,* livret n°4, Lyon : APPRENDRE-Sciences de l'éducation

Collection “outils de base pour la recherche en éducation”, Université Lumière, (5), 1995, 53 p

RÉGNIER, J-C, (1998a) Histogramme*, in* J-C Girard, D. Gros, P. Planchette, J-C Régnier, R. Thomas*, Enseigner la Statistique du CM à la Seconde, Pourquoi ? Comment ?* Villeurbanne: I.R.E.M. de Lyon pp. 21-41

RÉGNIER, J-C, (1998b) Finalités et enjeux de l’enseignement de la statistique, *in* J-C Girard, D. Gros, P. Planchette, J-C Régnier, R. Thomas*, Enseigner la Statistique du CM à la Seconde, Pourquoi ? Comment ?* Villeurbanne: I.R.E.M. de Lyon, pp 5-20

RÉGNIER, J-C, (1998c) De la vérité autoproclamée à la vraisemblance reconnue, *in* J-C Girard, D. Gros, P. Planchette, J-C Régnier, R. Thomas*, Enseigner la Statistique du CM à la Seconde, Pourquoi ? Comment ?* Villeurbanne: I.R.E.M. de Lyon, pp 107-118

RÉGNIER, J-C, (1998d) Lire un article de journal, *in* J-C Girard, D. Gros, P. Planchette, J-C Régnier, R. Thomas*, Enseigner la Statistique du CM à la Seconde, Pourquoi ? Comment ?* Villeurbanne: I.R.E.M. de Lyon, pp 127-133

RÉGNIER, J-C, (1998e) Danger ! Approximations…, *in* J-C Girard, D. Gros, P. Planchette, J-C Régnier, R. Thomas*, Enseigner la Statistique du CM à la Seconde, Pourquoi ? Comment ?* Villeurbanne: I.R.E.M. de Lyon, pp 99-106

RÉGNIER, J-C, (1998f) Méthode naturelle et tâtonnement expérimental, in N. Bizieau, J-F. Fouquer (Eds) *Célestin Freinet, l'ICEM, un choix pédagogique, un engagement social et politique*, Nantes : ICEM-Pédagogie Freinet, 1998, pp 312-325.

RÉGNIER, J-C, (1999) A auto-avaliação na prática pedagógica, *Avaliação*, Revista da rede de avaliação institucional da educação superior, ISSN 1414-4077, dez. 1999, pp 45-53

RÉGNIER, J-C, (2000) *Auto-évaluation et autocorrection dans l'enseignement des mathématiques et de la statistique. Entre praxéologie et épistémologie scolaire.* Note de synthèse pour HDR. Université Marc Bloch Strasbourg. 240 p.

ROY , R. (1937) L'enseignement de la statistique, Cinquantenaire de la *Revue d'économie politique*, 1887-1937

SANCHEZ-COBO, F.T., (1996) *Análisis de la exposición teórica y de los ejercicios de correlación y regresión en los textos de Bachillerato*. Memoria de Tercer Ciclo, Universidad de Granada

.

VALLECILLOS, A, BATANERO, C., (1997) Conceptos activados en le contraste de hipotesis estadísticas y su comprension por estudiantes universitarios , *Recherches en Didactique des Mathématiques* , vol.17/1 pp.29-48

VERGNAUD, G., (1991) Théorie des champs conceptuels, *Recherches en Didactique des Mathématiques* , vol.10/2.3 p.133-169

VYGOTSKI, L.S., (1985) *Langage et Pensée,* Paris : Messidor, Terrains/Éditions Sociales, 416 p. ( Traduction intégrale des textes russes de Vygotski par Françoise Sève, suivi d'un commentaire sur les remarques critiques de Vygotski par Jean Piaget )

WARNER, S.L., (1965) Randomized Response : a Survey Technic for Eliminating Evasive Answer Bias*, J.A.S.A.*