OBSTACLES A PRIORI À L'APPRENTISSAGE DE L'ANALYSE STATISTIQUE INFÉRENTIELLE

Philippe CALMANT[1], Marie DUCARME[2] et Maggy SCHNEIDER[3]

TITLE

Obstacles a priori to the learning of inferential statistical analysis

RÉSUMÉ

Cet article présente une amorce d'un travail de recherche plus global qui porte sur l'étude des conditions institutionnelles, cognitives et didactiques de l'introduction de la statistique inférentielle dans les programmes de l'enseignement secondaire belge. Nous proposons une analyse a priori des difficultés que pourraient rencontrer des élèves de ce niveau. Nous mettons au jour trois grands obstacles d'apprentissage : la non-prise en compte de la variabilité, le décodage d'ostensifs en termes de lecture « X-Y » plutôt qu'en termes de rapports d'aires et la difficulté à

Statistique et Enseignement, **2**(1), 43-59, http://www.statistique-et-enseignement.fr/ojs/

d'ingénierie désigne une forme de travail didactique, comparable à celle de l'ingénieur (cf. Brousseau, 2010 ; Chevallard, 1982). La théorie des situations didactiques (TSD) de Brousseau et la théorie anthropologique du didactique (TAD) de Chevallard, dont Schneider (2008) souligne et illustre la solidarité conceptuelle et méthodologique, fournissent des outils particuliers de construction d'ingénieries. Une telle construction se doit de respecter des contraintes méthodologiques. D'après Artigue (1988, p. 287-288), celle-ci suppose des analyses préalables :

- « l'analyse épistémologique des contenus visés par l'enseignement,
- l'analyse de l'enseignement usuel et de ses effets,
- l'analyse des conceptions des élèves, des difficultés et obstacles qui marquent leur évolution,
- l'analyse du champ des contraintes dans lequel va se situer la réalisation didactique effective,
- et bien sûr la prise en compte des objectifs spécifiques de la recherche. »

Nous nous intéressons ici exclusivement à « l'analyse des conceptions des élèves, des difficultés et obstacles qui marquent leur évolution ». Il nous a paru utile de nous baser sur une recherche antérieure (Calmant, 2004) relative à l'enseignement de la biostatistique en deuxième année d'université. Même si la thèse de Calmant relève de la biologie et concerne les étudiants universitaires, elle permet de mettre en évidence des obstacles à ce point robustes que, a priori, nous pressentons les rencontrer lors d'une introduction de la statistique inférentielle en secondaire. C'est en cela qu'elle présente un intérêt certain pour notre recherche. À travers cette étude, nous mettons au jour trois grands obstacles d'apprentissage : la non-prise en compte de la variabilité, le décodage d'ostensifs[4] en termes de lecture « X-Y » plutôt qu'en termes de rapports d'aires et la difficulté à concevoir des niveaux sémantiques plus abstraits. Pour les analyser, nous partirons de la typologie de Brousseau (1998), lequel distingue :

- les obstacles épistémologiques, qui sont des formes de connaissances devenues inopérationnelles dans un contexte nouveau et qui « sont ceux auxquels on ne peut, ni ne doit échapper, du fait même de leur rôle constitutif dans la connaissance visée ; on peut les retrouver dans l'histoire des concepts eux-mêmes » (Brousseau, 1998, p. 125) ;
- les obstacles didactiques, créés par les dispositifs d'enseignement eux-mêmes ;
- les obstacles ontogéniques, qui trouvent leur origine dans les limitations de l'apprenant à un moment de son développement cognitif.

Nous mettons à l'épreuve cette typologie en la complétant de ce que nous appellerons un obstacle cognitif.

[4] « Nous parlerons d'objet ostensif – du latin ostendere, 'montrer, présenter avec insistance' – pour nous référer à tout objet ayant une nature sensible, une certaine matérialité, et qui, de ce fait, acquiert pour le sujet humain une réalité perceptible. Ainsi en est-il d'un objet matériel quelconque et, notamment, de ces objets matériels particuliers que sont les sons (parmi lesquels les mots de la langue), les graphismes (parmi lesquels les graphèmes permettant l'écriture des langues naturelles ou constitutifs des langues formelles), et les gestes. » (Bosch et Chevallard, 1999 : 10)

2. Contexte de notre analyse a priori

Pour réaliser notre analyse a priori, nous prenons appui sur la thèse de Calmant (2004), entre autres sur ses interviews réalisées auprès d'étudiants universitaires en biostatistique, portant sur les tests d'hypothèses.

2.1. Recueil des données

Les interviews, réalisées par Calmant en 2003-2004, ont pour objectif premier de cerner la représentation que se font les étudiants de la matière dispensée, en cours d'enseignement ou après celui-ci. Le mot *représentation* fait écho, en un sens large, au concept de représentation mentale, introduit par Piaget en psychologie cognitive, qui recouvre un ensemble de connaissances ou intuitions exprimées, de confusions ou de réductions, d'incapacité à expliquer un paradoxe, de non-prise en compte de certains paramètres… Contrairement aux évaluations traditionnelles de type QCM qui jugent la *praxis*[5] (Chevallard, 1999), les interviews de Calmant permettent d'interroger les étudiants sur le *logos*, qui constitue un discours sur la pratique censé justifier et rendre intelligibles les techniques choisies pour réaliser les tâches données. Il semble qu'il s'agit là d'un registre qu'il faut explorer pour identifier des difficultés résistantes, susceptibles de se constituer en obstacles.

Pour permettre une méthode d'investigation souple, usage a été fait d'interviews semi-structurées comme définies par De Landsheere (1976, p. 83) : « Ici, l'enquêteur accorde moins d'importance à la standardisation qu'à l'information elle-même. Toutefois, il faut qu'en fin d'entretien, une série d'objectifs précis soient atteints. "Un schéma définit les principaux thèmes à explorer et prévoit éventuellement certaines questions ; mais la manière dont les thèmes seront amenés au cours de l'entretien, la façon dont les questions seront formulées et l'ordre dans lequel thèmes et questions apparaîtront ne sont pas fixés d'avance." (Maisonneuve et Duclot, 1962) ». Dans le cas présent, les étudiants ont été interrogés par deux. Cette méthode des « binômes », qui a déjà fait ses preuves dans plusieurs recherches, favorise la formulation des modèles implicites des interviewés.

Cependant, il ne faut pas perdre de vue que les propos recueillis sont ceux d'étudiants de deuxième année universitaire en sciences (sciences biomédicales, géologie, géographie, sciences vétérinaires). Il nous faut donc considérer en quoi diffèrent ces deux institutions[6] : celle du « cours de biostatistique à l'université » et celle d'un « cours d'introduction de la statistique inférentielle au niveau du secondaire ». Afin de situer les données recueillies et de les relativiser par rapport à l'institution « cours de biostatistique à l'université », il nous semble indispensable de positionner les tests d'hypothèses dans un contexte plus global dans lequel s'insère en général un cours de biostatistique.

---

[5] La *praxis* rend compte des pratiques mathématiques en termes de techniques permettant de réaliser des tâches données par la manipulation réglée de symboles mathématiques, de représentations graphiques ou autres…

[6] Ce terme est emprunté à Chevallard : « Une institution *I* est un dispositif social « total », qui peut certes n'avoir qu'une extension très réduite dans l'espace social (il existe des « micro-institutions »), mais qui permet – et impose – à *ses sujets*, c'est-à-dire aux personnes *x* qui viennent y occuper les différentes *positions p* offertes dans *I*, la mise en jeu de *manières de faire et de penser propres.* » (Chevallard, 2003)

2.2. Tests d'hypothèses en biostatistique

Selon Chevallard (1978a, p. 8), la statistique « met en contact, selon des modalités diverses, d'une part, des domaines extra-statistiques (et extra-mathématiques) qui proposent des problèmes à l'origine non statistiques (problèmes de biologie, de psychologie, de sociologie), vis-à-vis desquels la statistique intervient comme méthode de position et de résolution (...) ; d'autre part, une théorie mathématique (la théorie des probabilités, et sa spécialisation, la statistique mathématique) ». Dans ce cadre, la biostatistique apparaît comme une science hybride située à la croisée de la biologie (science du vivant) et de la statistique (science mathématique). Elle a pour objectif de fournir des outils mathématiques spécialisés capables d'apporter au biologiste une aide précieuse dans la perception et la quantification de son environnement matériel, organique ou inorganique.

La réalisation d'un test d'hypothèse en biostatistique peut être décomposée en différentes tâches pouvant être accomplies par une ou plusieurs personnes (cf. figure 1). Ainsi, le traitement des données statistiques va dépendre du commanditaire de l'analyse (entreprise pharmaceutique, laboratoire d'écologie…), véritable initiateur de la démarche scientifique et statistique. Celui-ci va transmettre ses idées au biostatisticien (1) chargé de les formaliser en fonction de règles statistiques strictes. Celles-ci vont le conduire à l'élaboration d'un protocole expérimental qu'il transmet ensuite à l'expérimentateur (2). Entre le recueil des données et le traitement de celles-ci, va s'instaurer une dialectique puisque « ce que l'on a recueilli va contraindre, de diverses manières, le traitement qu'on pourrait en faire ; inversement et conséquemment, avant de recueillir des données, il faudra savoir ce qu'on veut en faire, pour savoir quelles données au juste on cherchera, et comment on s'y prendra » (Chevallard, 1978a, p. 8). Les résultats chiffrés de l'expérience sont ensuite traités par le biostatisticien (3). Dans des cas simples auxquels nous nous limiterons ici, la méthode statistique appliquée a pour principale fonction de réduire les résultats expérimentaux à une valeur calculée à partir des observations, comparée à une frontière qui sépare celles de ces valeurs pour lesquelles on rejette l'hypothèse nulle H_0 de celles pour lesquelles on ne la rejette pas. Cette frontière se calcule à l'aide des lois de probabilités qui régissent, quand l'hypothèse H_0 est valide, la fonction de test (souvent des lois classiques telles que la loi normale, la loi du χ^2…) et en prenant en compte le seuil de signification adopté pour l'étude, c'est-à-dire un majorant imposé à la probabilité de conclure, à tort, au rejet de l'hypothèse nulle H_0 quand celle-ci est satisfaite. Après avoir fourni un effort de décontextualisation au sens de Crahay (2006) en projetant les résultats expérimentaux dans un système réduit propice à la prise de décision du test, le biostatisticien doit fournir un effort de recontextualisation afin de replacer la décision qui vient d'être prise dans son contexte. Le fruit de son analyse retourne ensuite au commanditaire (4). Formés à un champ spécifique d'application, les biostatisticiens peuvent être amenés à élaborer ou affiner leurs méthodes grâce au contact avec des chercheurs en Statistique plus généralistes ou plus fondamentalistes, en somme des « statisticiens professionnels » ; c'est à cette interface que Chevallard (1978a, p. 4) situe des intervenants qu'il qualifie de « méthodologues ». Bien sûr, cette séparation des différents « acteurs » est schématique et la réalité suppose des « allers-retours » liés au dialogue entre les parties prenantes : ainsi, dans un laboratoire, le commanditaire peut être aussi expérimentateur et même avoir des compétences en statistique qui le conduiront à intervenir auprès du biostatisticien dans la formalisation établie par celui-ci.

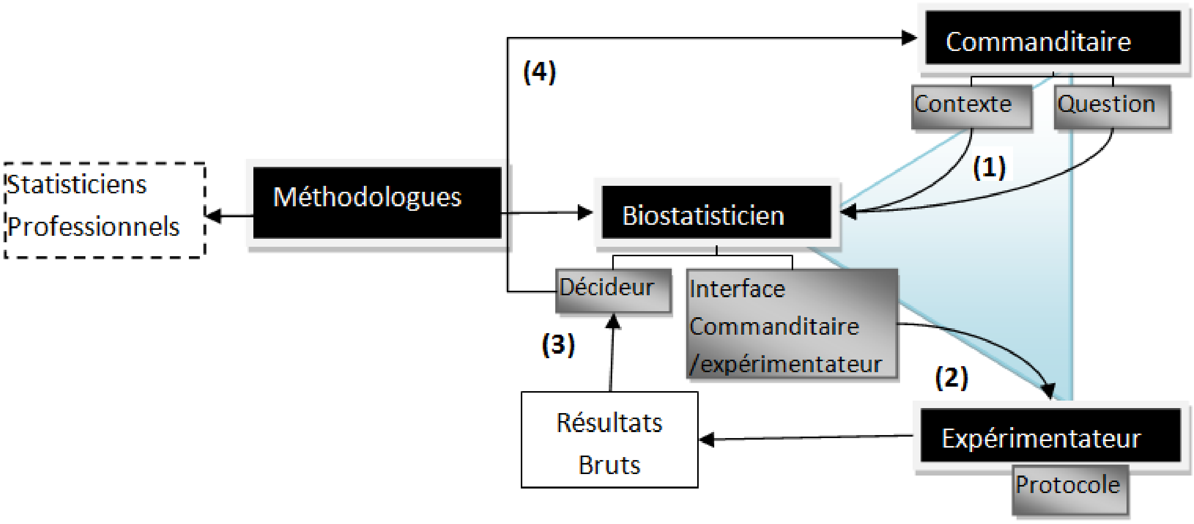


FIGURE 1 – *Répartition des tâches dans la réalisation d'un test d'hypothèse en biostatistique*

Lors d'un enseignement de biostatistique, il est important d'être au clair sur le rôle donné à l'étudiant. Où le positionne-t-on dans le triangle « commanditaire, biostatisticien, expérimentateur » ? Par exemple, dans le cours de biostatistique dispensé à l'université où Calmant a réalisé ses interviews, le rôle de l'étudiant s'apparente à celui du biostatisticien amateur, à cette différence qu'il n'est pas en mesure d'influer sur la conception du protocole expérimental. Comme un biostatisticien, il doit prendre en charge les données produites par l'expérience et le contexte qui les porte afin d'isoler la technique statistique adéquate. Il doit ensuite formuler ses résultats dans les termes du contexte comme s'il les transmettait au commanditaire de l'analyse statistique. L'étudiant pourra endosser le rôle d'expérimentateur lors de ses années de maîtrise et celui de commanditaire plus tardivement encore, par exemple s'il effectue un doctorat. Les élèves du secondaire, quant à eux, se situeront probablement dans une institution où la plupart du temps, en raison des contraintes rencontrées à ce niveau d'enseignement, ils ne joueront pas le rôle d'expérimentateur mais uniquement mimeront celui de décideur.

3. Description de trois obstacles majeurs

Notre analyse a priori nous permet de mettre en évidence trois difficultés majeures qui constituent de véritables obstacles dans l'apprentissage de la statistique : la non-prise en compte de la variabilité, le décodage d'ostensifs en termes « X-Y » plutôt qu'en termes de rapports d'aires et la difficulté à concevoir des niveaux sémantiques plus abstraits. Nous développerons chacun d'eux en mettant en lien des extraits d'interviews (cf. recueil des données) et nos analyses. Nous tenterons également d'en analyser la nature d'après la typologie de Brousseau.

3.1. Non-prise en compte de la variabilité

Le propre de la pensée statistique consiste en la gestion de la variabilité, omniprésente dans les phénomènes étudiés, comme le font remarquer Cobb et Moore (1997, p. 801) : « Statistics is a methodological discipline. It exists not for itself but rather to offer to other fields of study a coherent set of ideas and tools for dealing with data. The need for such a discipline arises from the omnipresence of variability ». De nombreuses études (Shaughnessy, 1997 ; Reading and Shaughnessy, 2005 ; Reading and Reid, 2010) ont révélé une difficulté conceptuelle chez les étudiants : celle de prendre en compte cette notion de variabilité. Nous l'analyserons ci-après en termes d'obstacle épistémologique.

Cette difficulté se manifeste dans les propos recueillis par Calmant. Nous reprenons ici deux morceaux d'interview significatifs. Dans le premier, un étudiant (E8) est interrogé sur la capacité d'un médicament à abaisser le taux de cholestérol.

> *E8 : « Supposons que l'on fait l'expérience et que l'on trouve 45. Cette valeur, je dois la comparer à un autre chiffre. »* Il ne semble pas être sûr de ce qu'il raconte. *« Il faut faire une mesure avant d'injecter le médicament et une mesure après le traitement. Si on a obtenu 45, on ne sait pas faire grand-chose avec cette seule valeur. Il faut voir si ce taux a augmenté ou diminué par rapport au taux qu'il avait avant l'administration du médicament. »*

Nous pouvons remarquer qu'à ce moment de la réflexion, E8 ne se prononce pas très explicitement sur le sens qu'il attribue à la valeur 45. S'agit-il d'un taux de cholestérol d'une population ou d'un échantillon ? L'interviewer intervient afin d'obtenir des précisions.

> *L'interviewer : « Supposons que l'on sait que le taux de cholestérol est de 50 en temps normal. En faisant ton expérience, tu as obtenu un taux de 45. Cela ne me dit toujours pas comment rédiger mes hypothèses… La personne qui va faire le test, de quels renseignements a-t-elle besoin pour le mener à bien ? En supposant que c'est une entreprise pharmaceutique qui a demandé à l'expérimentateur de faire cette étude. De quelles informations un biostatisticien a-t-il besoin pour faire un test d'hypothèses ? »*

> *E8 : « …des moyennes ? Comme on a obtenu une certaine moyenne sur un échantillon avant l'expérience, et qu'après on a obtenu une autre moyenne… le biostatisticien va demander à avoir ces deux mesures. En observant la moyenne avant le traitement (45), il va pouvoir savoir si cette différence est réellement due à l'injection du médicament ou bien si c'est dû au fait… qu'il a plus ou autre chose (les individus se sont enrhumés, etc.)… »*

De cet échange, il ressort clairement que E8 parle bien d'échantillon mais, et c'est là le point important pour notre propos, il ne prend pas en compte la variabilité, négligeant le concept d'écart-type qui en est pourtant une mesure fondamentale. Il se concentre exclusivement sur les moyennes. C'est la question suivante de l'interviewer qui incite E8 à émettre l'hypothèse que les moyennes seules ne suffisent pas dans un raisonnement statistique mais sans pour autant préciser par quoi les compléter.

> *L'interviewer : « Et ces seules moyennes suffisent-elles ? »*

> *E8 : « Probablement que non. Parce qu'il y a plein de 'trucs' à prendre en compte ! »*

Dans le second extrait, E11, contrairement à E8, prend en compte la variabilité mais indique vouloir éviter les valeurs extrêmes qu'il considère comme des erreurs susceptibles de fausser l'interprétation.

> *E12 : « Comme on ne peut jamais prendre tous les individus d'une population, forcément, il y a toujours une possibilité de se tromper. »*

L’interviewer : « Se tromper sur quoi ? »

[…]

E11 : « En disant que tous les oiseaux ont une taille de 60 cm et en même temps, on isole ceux qui sont hors normes puisqu’on ne peut pas vraiment les comparer véritablement à notre échantillon. On sait qu’il y en a qui seront anormaux et ceux-là ne nous intéressent pas trop parce qu’on sait que ça existe donc ceux-là on évite de les prendre dans l’échantillon… »

Ces valeurs extrêmes interpellent les statisticiens professionnels qui développent des stratégies pour les prendre en compte : soit pour détecter d’éventuelles erreurs d’encodage des données, soit pour ajuster le modèle initial au cas où les valeurs extrêmes l’influencent inconsidérément. C’est ce qui amène vraisemblablement l’interviewer à poser la question suivante.

L’interviewer : « Il n’existe pas un moyen de minimiser l’importance de ces valeurs [en parlant des valeurs extrêmes] ? »

E11 : « Ben en les enlevant, tout simplement ! »

Cet échange montre un rapport tranché de E11 vis-à-vis des valeurs extrêmes. Contrairement aux statisticiens professionnels, E11 ne mesure pas leur importance. Il les considère purement et simplement comme des erreurs par rapport à une norme.

Ces deux extraits d’interview, représentatifs de la pensée d’étudiants, nous interpellent tout particulièrement en ce qu’ils mettent en évidence le risque de ne pas prendre en compte la variabilité et celui de penser la variabilité en termes d’erreur, voire d’aberration. Ces dérives semblent avoir été présentes de tout temps. En effet, Chevallard (1978b) en retrouve des traces dans l’histoire de la statistique où il distingue deux moments historiques que nous résumons ci-après et qu’il convient de situer par rapport à l’usage de la statistique qui prévalait en sciences sociales ou humaines à l’époque où ce texte fut écrit.

Dans un premier moment historique, la statistique consistait à analyser des données recueillies sur des populations étendues afin de faire ressortir une tendance moyenne, des régularités. Süssmilch (1707-1767), démographe, dira : « […] l’ordre ne peut être révélé que par les grands nombres : les inexactitudes des petits chiffres disparaissent d’autant plus que les nombres que l’on compare sont plus grands » (cité par Chevallard, 1978c). Pour valider cette formulation, la statistique va faire intervenir une théorie extérieure, à savoir la théorie probabiliste des erreurs, qui va jouer un rôle de paradigme permettant, dans un premier temps, de penser commodément les permanences du réel. On doit l’impact de ce paradigme particulièrement aux travaux de Quetelet (1796-1874) qui s’est consacré à l’étude des êtres humains et aux paramètres qui les caractérisent (taille, poids…) sur base d’une méthode immuable : mesures d’individus, calcul de moyennes ou de taux. Ces travaux l’amènent à une conception précise d’une science de l’Homme qui influencera le développement ultérieur de la sociologie et qui met en avant-plan la notion « d’homme moyen ». La moyenne apparaît alors comme une norme et toute variation par rapport à cette dernière est considérée comme une erreur, une aberration. Chevallard (1978c, p. 65) explique que c’est le « mouvement d’annulation (ou l’oubli) des valeurs qui ne sont pas la moyenne, mais dont la moyenne est issue, qui fait ressembler la moyenne à une norme. Plus exactement : qui fait qu'une moyenne peut devenir, ou définir, une norme. » Ce concept d’homme moyen sera violemment critiqué, l’objection majeure tenant précisément au fait que les moyennes obtenues des divers paramètres (taille, poids…), selon lesquels Quetelet étudie l'homme, ne peuvent que donner un homme moyen aberrant.

C'est ainsi que, dans un second temps historique (vers le XX^e siècle), le modèle proposé par la théorie des erreurs deviendra un obstacle. Il va commencer à se désagréger lentement et à s'ouvrir aux fluctuations. Cette ouverture viendra de problématiques extérieures aux statistiques, telles que les théories de Galton et Binet qui ont donné naissance à la psychologie différentielle. A cette époque, « […] la moyenne a perdu sa prépondérance : elle n'est plus qu'un repère sur une échelle étalonnée (quartilage, centilage) ; cela est un acquis certain. Corrélativement, la conception de l'écart à la moyenne comme 'erreur' a subi l'effet de ce déplacement d'intérêt (de la moyenne vers l'ensemble de la distribution) : elle n'est plus recevable. » (Chevallard, 1978b, p. 51)

Ces deux moments mettent au jour le cœur de la problématique de la statistique que Chevallard présente comme une problématique « de la recherche et de la constitution d'une dialectique à caractère scientifique entre régularités et fluctuations, dans l'analyse de phénomènes marqués par un caractère de variabilité » (Chevallard, 1978b, p. 1). Il convient aujourd'hui d'ajouter un troisième moment historique, commençant environ au quatrième quart du vingtième siècle, quand l'usage intensif des simulations sur ordinateur a donné pour tous (même à des niveaux scolaires relativement bas) un nouvel accès à la notion de variabilité.

Parmi de multiples facteurs, psychologiques ou sociaux, induisant un « attrait pour les moyennes », qu'il n'y a pas lieu d'analyser ici, la difficulté ressentie par les étudiants de prendre en compte la variabilité apparaît comme le reflet du premier moment historique, où la moyenne occupait une place centrale et où la variabilité était totalement négligée ou considérée comme une erreur.

Cette analyse nous amène à qualifier la « non-prise en compte de la variabilité » d'obstacle épistémologique au sens de Brousseau[7]. Il constitue une connaissance qui possède son domaine de validité. Cependant, lors de l'apprentissage de la statistique, il va provoquer des erreurs persistantes, récurrentes. Cet obstacle est donc robuste, résiste et reparaît car il est inhérent au savoir et indépendant des individus. On en trouve, comme nous l'avons vu, des traces dans l'histoire. Cette résistance n'est pas sans conséquence didactique ; elle suppose un traitement spécifique d'explicitation et de rejet, qui s'apparente à la mise en place d'une nouvelle connaissance : « […] *un obstacle épistémologique est constitutif de la connaissance achevée* en ce sens que son rejet doit finalement être incontournablement explicité, et par conséquent il laisse des traces parfois profondes dans le système des connaissances » (Brousseau, 1998). Artigue (1988) et Schneider (2008) évoquent l'hypothèse qu'un obstacle épistémologique peut être renforcé par des dispositifs didactiques. Il serait intéressant de se demander dans quelle mesure l'enseignement consolide l'obstacle de la non-prise en compte de la variabilité et d'adopter un choix méthodologique qui permette de le surmonter plus facilement, mais nous ne pouvons développer ce point dans le cadre de cet article.

[7] Depuis les travaux originels de Brousseau (1983) et Duroux (1983), plusieurs auteurs ont travaillé sur les obstacles épistémologiques, surtout dans les domaines de l'analyse et des probabilités. Schneider (2008), sur qui nous prenons appui, a réalisé une analyse qui permet de résumer les caractéristiques qui se prêtent à ce genre d'obstacle.

3.2. Décodage d'ostensifs graphiques en termes « X-Y »

Dans l'enseignement de la statistique, tant descriptive qu'inférentielle, l'usage d'ostensifs graphiques est fréquent : diagramme en camembert, diagramme en bâtons, histogramme, nuage de points, polygone des effectifs, cloche de Gauss. Certains d'entre eux nécessitent une lecture « X-Y » : au départ d'une abscisse repérée sur un axe horizontal, on lit une ordonnée sur un axe vertical. D'autres, par contre, exigent une interprétation en termes de rapports d'aires : le renseignement utile nécessite le calcul d'aires sous une courbe. Cette interprétation semble poser problème aux étudiants, comme le montrent des coups de sonde réalisés auprès de futurs professeurs du secondaire ainsi que l'analyse des interviews de Calmant. Cette difficulté révélera un obstacle d'ordre plutôt didactique.

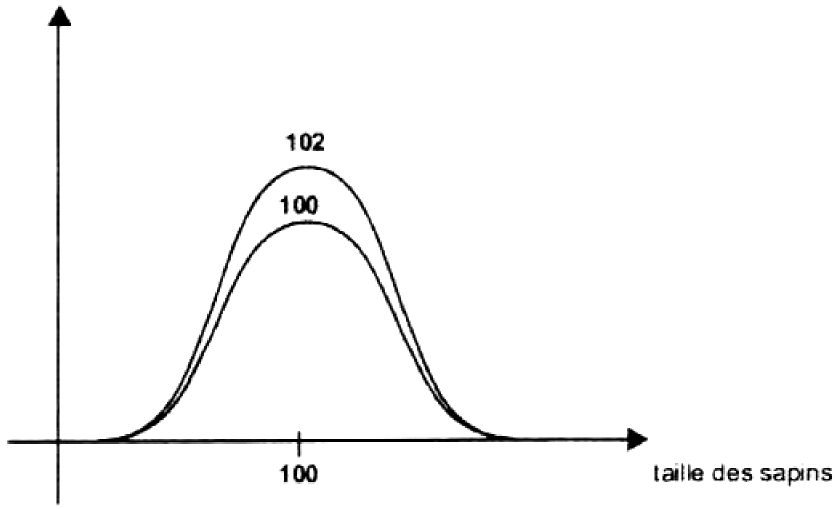
Calmant s'est intéressé plus particulièrement à l'ostensif qu'est la cloche de Gauss, fréquemment utilisée dans le cours de biostatistique. Celle-ci est le graphe d'une fonction représentant la densité de probabilité d'une variable aléatoire normale. L'axe des ordonnées ne fournit aucune information pertinente au sujet du renseignement utile recherché, à savoir une probabilité. Celle-ci sera en fait donnée par une aire sous la courbe entre deux bornes, étant donné que l'aire totale sous la cloche de Gauss vaut 1 comme probabilité de l'événement certain.

Dans les trois extraits suivants, il est demandé aux étudiants de préciser ce que représentent les axes « X-Y » dans le repère de la courbe de Gauss. L'axe des abscisses est généralement bien interprété en lui associant à juste titre le paramètre mesuré. Au contraire, l'axe des ordonnées, représentant une densité de probabilité qui est une notion abstraite, fait l'objet de mésinterprétations. Dans les extraits suivants, les étudiants cherchent en effet à lui donner un sens concret. La réaction la plus fréquente est de lui faire correspondre une quantité d'individus, comme le fait E18.

> *E18 dessine une cloche et place sur l'axe des abscisses les valeurs 94,100 et 108* [ces valeurs correspondent à des tailles de planches].
>
> *L'interviewer : « Alors, je te repose la question : qu'y a-t-il en X et en Y ? »*
>
> *E18 : « Il n'y a rien en Y, ben non !... C'est le nombre de planches ! Donc à 100 on a beaucoup de planches, et à 94 ben on en a moins. C'est ça ? »*

Malgré qu'E18 donne une interprétation à l'axe des ordonnées, il ne semble pas très sûr de sa réponse. Cette incertitude se manifeste également lorsqu'on l'interroge sur l'appartenance d'un échantillon de sapins, dont la moyenne de taille est de 102 cm, à une population de sapins dont la moyenne attendue est de (100 ± 10) cm.

> *L'interviewer : « Donc à quelque chose en X, je peux associer quelque chose en Y ?... Attends parce qu'on vient de dire que 100 était là. »* [En désignant sur l'axe des X le trait annoté « 100 »]



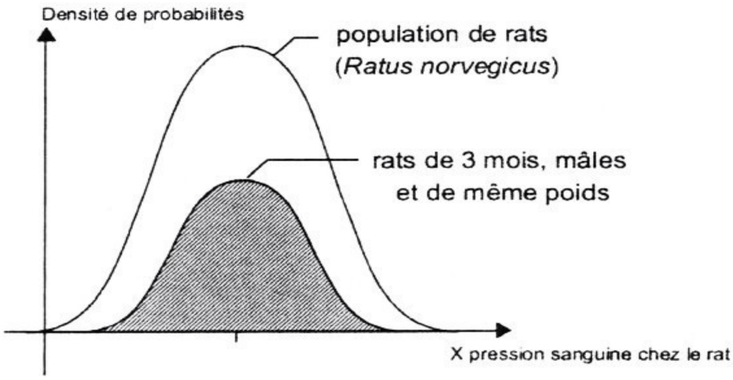
> *E18 : « Oui mais nous on a 102 donc ce n'est pas le bon ! C'est un autre graphique alors ! »*
>
> *L'interviewer : « […] Qu'est-ce que tu mets en X et en Y ? »*
>
> *E18 : « […] Ben la même chose si on veut. »*

La réponse évasive d'E18 illustre bien toute la difficulté ressentie par cet étudiant à donner une interprétation pertinente. Outre l'attribution d'une quantité d'individus à l'axe des ordonnées, d'autres étudiants lui font coïncider des probabilités. C'est ce que fait E13 en donnant la valeur « 1 » au sommet de la distribution gaussienne.

> *E13 : « Disons que cet axe-là* [il désigne du doigt un axe vertical], *on ne le dessine pas. C'est la probabilité de rencontrer la variable. Evidemment, le 1 sera peut-être ici, enfin ça dépend : si c'est une loi normale, ça variera en-dessous de 1et si c'est une variable centrée réduite, je crois que le 1 est situé ici, au niveau du maximum. »*

Le propos ci-dessus montre à nouveau clairement que la lecture spontanée de la courbe de Gauss se fait sous la forme d'une lecture indue « X-Y ». Même la tentative de certains étudiants d'interpréter l'aire sous la courbe comme une probabilité, peut se solder par un échec. C'est ce qui est illustré dans le quatrième extrait que nous proposons ci-après. L'étudiant E12 est invité à lire la courbe de Gauss afin de déterminer une probabilité.

> *L'interviewer : « Quelle est la probabilité pour un rat de trois mois mâle d'appartenir à cette sous-population de rats mâles de trois mois ? »*
>
> *E12 : « Disons que si on a toute la surface, ici* [E12 désigne la surface sous la courbe la plus grande du dessin] *de 100% pour la population de Ratus Norvegicus, et que celle-ci* [E12 désigne la courbe la plus petite] *est bonne pour ceux de trois mois, je hachurerais ceci et je ferais 100% moins cette surface-là* [en montrant la surface comprise entre les deux courbes] *pour avoir la probabilité que représentent les rats de trois mois mâles. »*



Bien qu'E12 précise sur son dessin que l'axe des ordonnées représente des densités de probabilité et qu'il parle de probabilité en termes d'aires sous la courbe, il semble attribuer des nombres d'individus à ces aires. En effet, selon lui, étant donné que la population des rats mâles de trois mois contient moins d'individus que la population entière de rats, l'aire sous la courbe doit s'en trouver réduite en taille. On peut donc supposer que l'étudiant manipule l'ostensif de la courbe de Gauss comme s'il s'agissait d'une variante d'un diagramme de Venn.

Comme le montrent ces quelques interviews, que l'on pourrait multiplier, la lecture en termes d'aires de la cloche de Gauss semble constituer une difficulté majeure pour les étudiants. Elle se retrouve également en statistique descriptive dans le cas de variables quantitatives continues. La distribution des fréquences (relatives) de telles variables peut être visualisée à l'aide d'un histogramme. Il est important, lors d'une lecture d'un histogramme, de se demander ce qui est représenté sur l'axe des ordonnées. Ceci n'est pas toujours très clair dans la littérature. Certains ouvrages présentent simplement une échelle de grandeur, soit chiffrée soit en pourcentage, sans indiquer ce qu'elle représente. D'autres mentionnent que sont représentés sur l'axe des ordonnées des effectifs et/ou des fréquences. Ce dernier cas, bien que largement répandu, peut susciter des ambiguïtés lors de la lecture par des novices de la statistique. En effet, une telle représentation peut conduire à une mauvaise compréhension de l'histogramme, entre autres en favorisant à nouveau une lecture « X-Y ». Celle-ci n'a pas de grandes conséquences quand toutes les classes sont de même longueur car, dans ce cas, l'aire et la hauteur vont exprimer la même information à une constante multiplicative près. Par contre, dans le cas de classes inégales, ce qui est rarement fait dans les manuels scolaires, elle peut engendrer des incidents à ne pas négliger.

Nous remarquons donc que la lecture indue d'ostensifs graphiques en « X-Y » est largement répandue en statistique. Cela nous conduit à nous interroger sur l'origine de cette fausse interprétation. Il nous semble que les erreurs décrites ci-dessus sont induites, jusqu'à un certain point du moins, par le système d'enseignement lui-même. Nous venons déjà d'en donner un exemple au sujet des histogrammes dans certains ouvrages. Nous pensons également que la lecture « X-Y », déjà bien ancrée dans l'esprit des étudiants dans le contexte des études de fonctions, peut faire écran à l'interprétation d'ostensifs graphiques en termes d'aires. Nous pourrions encore pointer l'absence d'activités qui inviteraient à faire apparaître un graphique comme constitué à partir de formes plus primitives de représentation. Ceci nous invite à qualifier le décodage d'ostensifs graphiques en « X-Y » d'obstacle didactique au sens de Brousseau.

Il est donc nécessaire de prêter attention à cet obstacle lors de la construction de notre ingénierie didactique, afin d'écarter l'usage d'ostensifs dont la valence instrumentale[8] serait occultée faute d'une valence sémiotique appropriée[9]. Par exemple, il serait éventuellement avantageux d'opter pour la boîte à moustaches, très utilisée dans les pays anglo-saxons, et d'envisager un usage plus élargi des intervalles de confiance, comme le préconisent de nombreux auteurs (Holender, 2008 ; Cumming, 2009).

3.3. Difficulté à concevoir des niveaux sémantiques plus abstraits

En statistique inférentielle, interviennent différents objets qui se distinguent fortement par une sorte d'ordre de grandeur : cela va d'un simple individu à la population d'individus en passant par l'échantillon et la population d'échantillons[10]. Ces objets peuvent être répartis en deux classes, la première regroupant ceux liés à l'expérience (individu, échantillon) et la seconde, ceux, plus abstraits, liés à la population associée (population d'individus, population d'échantillons). Ces quatre types d'objets doivent être bien différenciés dans une étude statistique, ce qui peut être subtil et présenter a priori quelques difficultés. Les relations entre ces quatre objets représentent quatre plans sémantiques (cf. figure 2 : 1, 2, 3, 4).

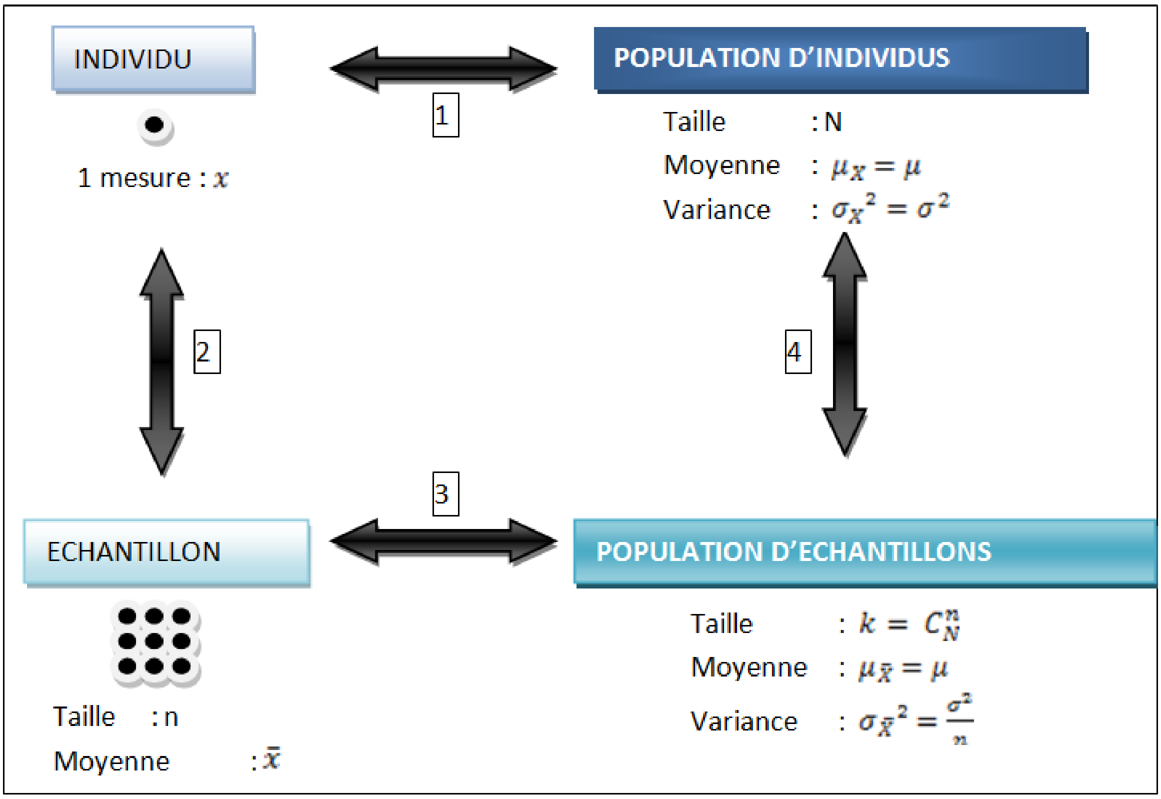


FIGURE 2 – *Quatre plans sémantiques dans les tests d'hypothèses*

[8] Une valence instrumentale relève de la praxis et fait de l'ostensif un véritable outil de travail se traduisant par une économie de pensée.

[9] Une valence sémiotique, grâce à laquelle l'ostensif apparaît comme signe, ou plutôt comme signifiant d'autres objets, est liée davantage au logos.

[10] Terminologie empruntée au cours dispensé aux étudiants concernés et pouvant entraîner des confusions. Nous y reviendrons.

Ces plans sémantiques interviennent dans le travail du biostatisticien quand il doit discerner ce qui est de l'ordre de la mesure expérimentale de ce qui appartient à celui de la population. Nous présentons ci-après des extraits d'interviews d'apprentis biostatisticiens qui sont confrontés à ces différents plans sémantiques. Les étudiants E13 et E14 sont amenés à distinguer la population d'individus et la population d'échantillons (cf. figure 2 : 4).

> *L'interviewer : « Si cette première courbe représente la population et que les deux suivantes représentent les distributions d'échantillonnage pour des échantillons de tailles 10 et 50, respectivement… Que se passe-t-il si je prends, en guise d'échantillon, la population ? »*
>
> *E13 : « Ben on obtient la courbe de la population ! »*

Nous constatons qu'E13 confond les courbes de la distribution de la population d'individus et de la distribution d'échantillonnage des moyennes. L'interviewer, qui avait déjà donné des indications graphiques afin d'amener la réponse, intervient à nouveau pour faire réfléchir E13.

> *L'interviewer : « Donc, à un moment donné, ça se resserre et puis, d'un coup, on va avoir la première courbe comme si la boucle était bouclée en quelque sorte ! »*
>
> *E13 : « Oui, je pense !...Euh non, en fait ce n'est pas logique ! En fait, si on prend la population comme échantillon, sa taille est infinie…[…] Ca va peut-être converger vers une valeur unique ! »*

E13 se rend compte que son raisonnement n'est pas logique. Il se sent alors incité à envisager une autre possibilité : la convergence vers une valeur unique. Sa réponse laisse croire qu'il se souvient d'une distinction entre ces deux ordres de grandeur. Cependant, l'explication donnée par son compagnon E14 montre qu'elle n'est pas pleinement comprise.

> *L'interviewer : « Comment l'expliques-tu ? »*
>
> *E14 : « Si on a utilisé tous les individus de la population pour l'échantillon, c'est que, …on aura toujours des exceptions… je ne sais pas comment le dire… On va avoir toutes les mesures possibles alors que sur des échantillons on va restreindre l'effet des valeurs extrêmes et donc on va tomber sur un pic… faible. »*

E14 et E13 ne se situent pas au bon plan sémantique. Ils confondent la population d'échantillons des moyennes et l'échantillon. De ce fait, ils mettent en relation la représentation de la population d'individus avec celle d'un échantillon qu'ils représentent également par une cloche de Gauss, ce que les propos suivants de E13 confirment.

> *E13 : « Je dirais qu'un échantillon, on le représente de la même manière [qu'une population n.d.l.r.] sauf que la courbe n'aura pas forcément la même allure… Ça dépend si on prend une loi centrée réduite ou une loi normale, mais la valeur centrale, si on prend une loi normale, sera peut-être différente. Elle peut être la même et ça veut dire que l'échantillon est très représentatif de la population, mais elle peut être différente et, là, ça peut être plus évasé sur les bords… Ça dépend… Moins on prend d'individus, moins on a de chance de se rapprocher de la population. »*
>
> *[…]*

L'interviewer : « Et donc l'échantillon va se représenter par une cloche ? »

E13 : « A priori oui. Si on choisit bien l'échantillon, il y aura toujours une valeur centrale qui aura une plus grande probabilité que les autres. Maintenant, c'est vrai que si on prend un échantillon de trois individus, la courbe va peut-être faire ça. »

Cet étudiant estime que l'allure de la cloche représentant l'échantillon varie en fonction de sa composition et plus particulièrement en fonction de sa taille, appliquant de manière inopportune la propriété qui lie la variance de la distribution d'échantillonnage à l'inverse de la taille de l'échantillon.

De la même manière que E13 et E14, E10, interrogé sur le même sujet, ne prend pas conscience de la différence qui existe entre un simple échantillon et la population d'échantillons.

L'interviewer : « Mais, on peut considérer que si la taille de l'échantillon augmente, la courbe se resserre tellement que finalement, je ne vais plus voir ma courbe. »

E10 : « Pas nécessairement. Plus « n » augmente et plus ça tend vers une distribution qui ressemble à celle de la population et… la courbe de la population, ça dépend à quoi elle ressemble. »

Malgré les propos de l'interviewer, qui donne quasiment la réponse, E10 reste sur ses positions et ses acquis. Ces propos d'étudiants ne sont pas des cas isolés et nous avons constaté qu'aucune des personnes interviewées n'est parvenue à expliquer clairement ce que devient la distribution d'échantillonnage des moyennes lorsque l'on prend en guise d'échantillon la population.

Les étudiants confondent les concepts de population d'individus et d'échantillon, et n'arrivent pas à concevoir la notion de population d'échantillons. Nous remarquons donc que ce qui ne relève pas directement de l'expérience mais demande une gymnastique intellectuelle abstractive pose problème. La non-prise en compte de certaines strates d'abstraction peut se révéler un véritable obstacle. Celui-ci n'est pas à proprement parler ontogénique, en ce sens que les étudiants universitaires, et même ceux de fin du secondaire, ont atteint le stade des opérations formelles, c'est-à-dire logiques et abstraites, telles que le raisonnement par hypothèses et déductions (Piaget, 1964). Il n'est donc pas lié aux limitations (neurophysiologiques entre autres) du sujet à un moment donné de son développement. L'obstacle que nous soulevons ici relève plutôt de la difficulté à manipuler des objets mathématiques faisant intervenir différents niveaux d'abstraction et à se situer au bon plan sémantique. Ce type d'obstacle n'est pas propre au présent contexte. Il a déjà été repéré par Rouche et Tossut (1985) à propos de la géométrie où les enfants sont amenés à travailler sur trois niveaux distincts : le plan, les droites, les directions. Ceux-ci sont définis successivement par des classes d'équivalence opérant sur des ensembles composés des objets précédents (ainsi une direction est la classe d'équivalence de droites parallèles). Les auteurs observent chez les élèves des difficultés à « grimper à l'échelle des ensembles ». Nous proposons donc d'ajouter à la typologie de Brousseau la notion d'obstacle cognitif pour désigner ces sauts d'abstraction que supposent les escalades successives d'objets mathématiques vers d'autres construits à partir de ceux-ci. Ce type d'obstacle peut être renforcé ou diminué par certaines pratiques didactiques. En particulier, le choix d'un certain vocabulaire peut prêter à confusion : ainsi, il vaudrait mieux parler d'ensembles d'échantillons plutôt que de populations d'échantillons ou encore parler d'effectif plutôt que de taille à propos des populations d'individus pour réserver le mot « taille » aux échantillons. Il faut donc en tenir compte dans toute stratégie d'enseignement, mais notre réflexion sur ce point précis reste à poursuivre.

4. Conclusion

L'analyse des interviews de Calmant (2004) réalisées auprès d'étudiants de deuxième année universitaire en biostatistique nous a permis de mettre au jour trois obstacles à l'apprentissage de l'analyse statistique inférentielle. Le premier, la non-prise en compte de la variabilité, nous est apparu comme un obstacle épistémologique, inhérent au savoir et dont le rejet se doit d'être pris en charge par un dispositif didactique adéquat. Le deuxième, le décodage d'ostensifs graphiques en termes « X-Y », s'est révélé étroitement lié aux dispositifs didactiques mis en place dans le cursus de l'enseignement secondaire actuel. Cet obstacle didactique devrait attirer l'attention de tout chercheur et de tout enseignant afin d'opérer des choix judicieux permettant de le surmonter. Le troisième, la difficulté à concevoir certains niveaux d'abstraction, nous a amenés à développer la notion d'obstacle cognitif et de compléter ainsi la typologie établie par Brousseau.

Ces trois obstacles majeurs constituent la base de notre analyse a priori de l'introduction de la statistique inférentielle en secondaire. Ils nous sont apparus comme fondamentaux et susceptibles d'être rencontrés dans l'institution « cours d'introduction de la statistique inférentielle au niveau du secondaire ». Forts d'avoir identifié et caractérisé ces obstacles, nous serons plus en mesure de poser des choix adéquats et éclairés pour la construction de notre ingénierie didactique. Néanmoins, d'autres analyses préalables restent encore à mener.

Références

[1] Antoine, Ph., J. Descy, M. Goffin et Ch. Van Hooste (2004), *Actimath 4-Manuel*, Van In, Wavre-Wommelgem.

[2] Artigue, M. (1988), Ingénierie didactique, *Recherches en didactique des mathématiques*, **9**(3), 281-307.

[3] Bertrand, J. (1889), *Calcul des probabilités,* Gauthier-Villars, Paris.

[4] Bosch, M. et Y. Chevallard (1999), *La sensibilité de l'activité mathématique aux ostensifs. Objet d'étude et problématique*, consulté le 16/09/2010 sur *http://yves.chevallard.free.fr/spip/spip/IMG/pdf/Sensibilite_aux_ostensifs.pdf* .

[5] Brousseau, G. (1983), Les obstacles épistémologiques et les problèmes en mathématique, *Recherches en didactique des mathématiques*, **4**(2), 165-198.

[6] Brousseau, G. (1998), *La théorie des situations didactiques*, La Pensée Sauvage, Grenoble.

[7] Brousseau, G. (2010), *Renaissance de la didactique des mathématiques*, Diaporama 2 du cours de didactique (2010), consulté le 25/02/2010 sur *http://guy-brousseau.com/111/introduction/* .

[8] Calmant, Ph. (2004), *Favoriser l'apprentissage des biostatistiques par le Web ? Essai de problématisation didactique d'une question issue du terrain*, Thèse de doctorat, Facultés Universitaires Notre Dame de la Paix, Namur.

[9] Chevalier, A. *et al.* (2002), *Référentiel de mathématiques*, De Boeck, Bruxellles.

[10] Chevallard, Y. (1978a), *Rénovation de l'enseignement de la statistique aux étudiants en psychologie, Premiers résultats, Sur la transposition didactique dans l'enseignement de la statistique*, exposé fait le 9 juin 1978 au Laboratoire de Psychologie du Travail du C.N.A.M., Département de Mathématiques, UER de Luminy et IREM d'Aix-Marseille, (non édité).

[11] Chevallard, Y. (1978b), *Rénovation de l'enseignement de la statistique aux étudiants en psychologie, note interne 5: La problématique générale de la statistique*, Laboratoire de Psychologie du Travail du C.N.A.M., Département de Mathématiques, UER de Luminy et IREM d'Aix-Marseille, (non édité).

[12] Chevallard, Y. (1978c), *Problèmes de surdétermination en didactique : la notion de moyenne en statistique*, Laboratoire de Psychologie du Travail du C.N.A.M., Département de Mathématiques, UER de Luminy et IREM d'Aix-Marseille, (non édité).

[13] Chevallard, Y. (1982), *Sur l'ingénierie didactique,* Texte préparé pour la deuxième Ecole d'Eté de Didactique des Mathématiques, Orléans.

[14] Chevallard, Y. (1999), L'analyse des pratiques enseignantes en théorie anthropologique du didactique, *Recherches en didactique des mathématiques*, **19**(2), 221-266.

[15] Chevallard, Y. (2003), Approche anthropologique du rapport au savoir et didactique des mathématiques, dans Caillot, M. et S. Maury (Eds), *Rapport au savoir et didactiques*, 81-104.

[16] Chevallard, Y. et F. Wozniak (2003), *Enseigner la statistique au secondaire – Entre genre prochain et différence spécifique,* Cours donné à la XII^{e} école d'été de didactique des mathématiques (Corps, 20-29 août 2003), consulté le 16/09/2010 sur *http://yves.chevallard.free.fr/spip/spip/article.php3?id_article=58* .

[17] Cobb, G. et D. Moore (1997), Mathematics, Statistics, and Teaching, *The American Mathematical Monthly*, **104**(9), 801-823.

[18] Crahay, M. (2006), Dangers, incertitudes et incomplétude de la logique de la compétence en éducation, *Revue Française de Pédagogie*, 154, 97-110.

[19] Cumming, G. (2009), Inference by eye: Reading the overlap of independent confidence intervals, *Statistics in Medicine*, 28, 205-220.

[20] Dagnelie, P. (2007), *Statistique théorique et appliquée-1. Statistique descriptive et bases de l'inférence statistique*, De Boeck, Bruxelles.

[21] De Landsheere, G. (1976), *Introduction à la recherche en éducation*, G. Thone, Liège.

[22] Depiereux, E. et Ph. Calmant (2004), *Biostastistiques (2^{e} candidature en sciences),* Notes de cours, Facultés Universitaires Notre Dame de la Paix, Namur.

[23] Dodge, Y. (2006), *Premiers pas en statistique*, Springer-Verlag France, Paris.

[24] Duroux, A. (1983), La valeur absolue : difficultés majeures pour une notion mineure, *Petit x*, 3, 43-67.

[25] Holender, D. (2008), Utilisation des intervalles de confiance au lieu des tests de signification : aspects épistémologiques et pratiques, *Psychologie du travail et des organisations*, **14**(1), 9-42.

[26] Krysinska, M. (2007), *Emergence de modèles fonctionnels comme outils de catégorisation de phénomènes divers : repères épistémologiques et didactiques,* Thèse de doctorat, Facultés Universitaires Notre Dame de la Paix, Namur.

[27] Maisonneuve, J. et M. Duclot (1962), Les techniques de la psychologie sociale, *Bulletin de Psychologie*, **15**(201).

[28] Piaget, J. (1964), *Six études de psychologie,* Denoël/Gonthier, Paris.

[29] Reading, C. and M. Shaughnessy (2005), Reasoning about variation, *The challenge of developing statistical literacy*, Part II, 201-226.

[30] Reading, C. and J. Reid (2010), Reasoning about variation: rethinking theoretical frameworks to inform practice, in Reading, C. (Ed.), *Data and context in statistics education: Towards an evidence-based society*, consulté le 1/02/2011 sur *http://www.stat.auckland.ac.nz/~iase/publications/icots8/ICOTS8_8E2_READING.pdf* .

[31] Rouche, N. et R. Tossut (1985), *Une géométrie pour tous les jours*, Groupe enseignement mathématique (GEM), Louvain-la-Neuve.

[32] Schneider, M. (1988), *Des objets mentaux aires et volumes au calcul des primitives,* Thèse de doctorat, Université catholique de Louvain.

[33] Schneider, M. (2008), *Traité de didactique des mathématiques – La didactique par des exemples et contre-exemples,* Les éditions de l'Université de Liège, Liège.

[34] Shaughnessy, M. (1997), Missed opportunities in research on the teaching and learning of data and chance, in Biddulph, F. and K. Carr (Eds), *Proceedings of the Twentieth Annual Conference of the Mathematics Education Research Group of Australia*, 6-22.

[35] Trousset, P. et J.-F. Morin (1994), *Mathématiques pour les sciences de la vie – 2. Probabilités et statistiques*, Ediscience international, Paris.