Aix Marseille Université
Master Mathématiques et Applications
Enseignement et Formation en Mathématiques
Parcours Didactique des Mathématiques

Mémoire de master de didactique des mathématiques
présenté et soutenu publiquement
par

Yannick Claudet

le 11 septembre 2013

La statistique descriptive en classe de troisième et de seconde. Un exemple de transposition didactique à l'œuvre

Directeur de mémoire :

Serge QUILIO

JURY :
Serge Quilio, maître de conférences, Université de Nice
Teresa Assude, professeur des Universités, Aix-Marseille Université
Lionel Vaux, maître de conférences, Aix-Marseille Université

Remerciements

Je tiens à remercier particulièrement Serge Quilio, directeur de ce mémoire pour son accompagnement toujours bienveillant et sans qui ce travail n'aurait pas pu voir le jour.

Je tiens à remercier Joël Briand d'avoir accepté avec gentillesse de m'avoir fait parvenir les documents que je n'aurai trouvé nulle part ailleurs.

Je remercie aussi Laurent, Karine et Priscilla qui ont accepté de participer à ce travail en prêtant, sans appréhension, les cahiers et les copies de leurs élèves et sans qui une partie de ce mémoire aurait été lettres mortes.

Un grand merci à l'équipe qui m'a permis de rentrer dans la didactique des mathématiques de par la qualité de leurs interventions et leurs exigences de travail : Yves Matheron, Teresa Assude, Alain Mercier, Yves Chevallard, Michèle Artaud, Serge Quilio, Tracy Bloor.

Un immense merci à mes collègues et camarades de promotion sans qui cette aventure aurait été moins belle.

Enfin un grand merci à mon épouse pour avoir assuré les mercredis après-midi quand « papa est à la fac ».

Table des matières

Introduction

Quand un professeur conçoit son enseignement, il s'interroge sur ce qui a été fait dans les classes précédentes (en tout cas il devrait le faire). Non seulement pour savoir ce que les élèves sont censés savoir et savoir-faire à l'issue des années précédentes mais aussi pour essayer de construire son enseignement dans le prolongement de ce que les élèves ont rencontré ou, tout au moins, qu'il espère qu'ils ont rencontré. L'enseignement de la statistique n'échappe pas à ce questionnement.

En regardant principalement l'enseignement de la statistique de la classe de troisième [1] et de la classe de seconde [2], nous nous sommes d'abord interrogés sur la façon dont le professeur pouvait réinvestir les notions présentes au programme de 3$^{\text{e}}$ et que l'on retrouve dans le programme de 2$^{\text{de}}$. Nous nous limiterons donc à la statistique descriptive, puisque la statistique inférentielle présente au programme de 2$^{\text{de}}$ ne l'est pas en classe de 3$^{\text{e}}$.

En regardant d'un peu plus près, nous nous sommes intéressés à ce qu'on peut trouver de commun ou non dans les enseignements de la statistique dans ces deux classes. Notre question de recherche s'est donc orientée vers l'observation de la transposition didactique au sein des classes de troisième et seconde de l'enseignement français.

Dans un premier temps nous avons analyser les praxéologies des instructions officielles mais aussi les commentaires apportés par les documents ressources.

Dans un deuxième temps, nous avons analysé les praxéologies de six manuels scolaires pour observer le savoir à enseigner tel qu'il est pensé et organisé par les concepteurs de manuels.

En dernier lieu, nous avons observé les traces écrites des élèves de quatre professeurs de mathématiques qui enseignent en collège et en lycée. Nous y avons donc observé la transposition didactique de la statistique telle qu'elle est produite par ces professeurs. Plus précisément nous avons mis au jour les praxéologies de ces professeurs pour observer le savoir enseigné et enfin le savoir appris par les élèves, tels que nous avons pu les voir au travers des cahiers de mathématiques, ainsi que les copies, de leurs élèves.

1. élèves de 14-15 ans
2. élèves de 15-16 ans

Chapitre 1

La statistique vue par la Didactique

La question qui nous intéresse n'est pas (totalement) nouvelle, plusieurs auteurs se sont intéressés à l'enseignement de la statistique. Que ce soit Guy Brousseau, Joël Briand, Yves Chevallard ou Floriane Wozniak, la statistique a déjà fait l'objet de publications. En partant des premières expérimentations des années 1970 jusqu'à la thèse de Floriane Mathieu-Wozniak de 2005, l'enseignement de la statistique, nous allons le voir, pose problème.

1.1 Les premières expérimentations

Dans le champ de la didactique, le premier à s'intéresser à l'enseignement de la statistique est Guy Brousseau. Au début des années 1970, il conçoit des séquences d'enseignement dont le compte rendu est donné dans *Description des 31 leçons expérimentées à l'école J. Michelet de Talence* (Brousseau, 1974). On y trouve essentiellement la description de l'organisation de l'expérimentation dont la question principale est : « Comment faire pour deviner la composition d'un sac contenant 5 billes ? »

Dans plusieurs articles, Guy Brousseau va reprendre ces expérimentations et les analyser de façon plus détaillées. Le premier qui nous intéresse *Notes sur l'enseignement des statistiques et/ou des probabilités dans la scolarité commune (6-14 ans)* (Brousseau, 2008) précise :

> La pratique des statistiques est très ancienne. Du point de vue anthropologique, elle fait partie de la culture des premières sociétés structurées et se développe avec le commerce. Par contre les connaissances théoriques qui en rendent compte sont très récentes et très sophistiquées. Et le rapport s'inverse, les probabilités sont primordiales pour l'étude de la plupart des chapitres importants de la statistique. Mais elles ne le sont pas pour la « pratique » de la statistique, qui se réduit pour beaucoup d'utilisateurs à l'usage d'algorithmes un peu complexes mais codifiés. (p. 2)

Nous le verrons, la pratique de la statistique est trop souvent réduite à une application d'algorithmes plus ou moins complexes. Mais c'est dans l'article *Alternatives en didactique de la statistique* (Brousseau, 2009) qu'il détaille un peu plus ses commentaires. En effet, la transposition qu'il a proposé lors des expérimentations à l'école primaire de la statistique part de l'expérience. En effet, les élèves doivent manipuler une bouteille opaque pour deviner la composition de la bouteille, qui ne sera jamais ouverte. C'est donc l'observation des fluctuations des séries statistiques qui a lieu en premier ; la loi des grands nombres suit puis la forme de convergence vers une fréquence théorique, l'intervalle de confiance, le test d'hypothèses et tout cela sans définitions formelles (rappelons que nous sommes à l'école primaire). Et ce sont bien des fréquences qui sont étudiées, jamais des probabilités ; les élèves ne réfléchissant jamais sur une série isolée. Nous verrons combien cette dernière remarque ne sera jamais mise en œuvre au collège ou au lycée.

Le dernier article à s'intéresser et commenter cette expérimentation est en fait une communication lors de l'école d'été de didactique des mathématiques de 2003. Nous le verrons donc plus loin.

Cette expérimentation sera reprise par Joël Briand en 1975 dont il fournira un compte rendu intitulé *Découverte des lois du hasard à l'école élémentaire* parue dans la revue PLOT de l'APMEP [1] (Briand, 1976). Il y décrit l'expérimentation qui a permis de prendre contact avec le hasard, percevoir intuitivement le rôle du nombre de tirages et l'importance de traiter la plus grande quantité d'informations possibles ainsi que d'aborder la stratégie générale du test d'hypothèse : affirmer une prévision et constater par l'expérience si celle-ce peut être infirmée ou non.

Par ailleurs Joël Briand reprendra cette expérimentation pour la transposer au lycée en classe de seconde. Le compte-rendu de cette expérimentation paraît dans la revue *Recherches en Didactique des Mathématiques (RDM)* (Briand, 2005). Une première transposition est effectuée par N. Pazat (Pazat, 2002) en 2002 dans le cadre d'un mémoire de PCL2. [2] Dans ce mémoire, le professeur stagiaire orientait son analyse plutôt sur la statistique inférentielle et les probabilités. Nous n'en parlerons pas plus ici puisque nous avons limité notre étude à la statistique descriptive.

C'est à la suite de l'observation des séances de ce professeur que Joël Briand poursuit l'idée de transposer ce qui avait été fait en 1974 pour l'introduire au lycée. L'année suivante (en 2003 donc) il demande à un professeur de seconde de reprendre à son compte cette expérimentation. Joël Briand, dans le résumé de l'article de RDM, précisait qu'il s'intéressait notamment « [...] à la construction de la notion d'expérience, à l'impact de la culture ambiante, de la culture personnelle des élèves et des nouveaux programmes de seconde [...] ». Il précise quelles étaient les principes de l'expérience de 1975 : « [...] aborder les statistiques et les probabilités dans leurs relations réciproques. » et « Pour cela, nous pensons que les données doivent être des informations recherchées intentionnellement par les élèves [...] » (p. 5). On le voit, la création d'une situation ne doit pas fabriquer des données pour être utilisées ; les données doivent être recherchées par les élèves parce qu'ils en auront besoin et non pas comme base pour effectuer des calculs.

Les conclusions de son article sont intéressantes à plus d'un titre au sujet de la transposition du travail de l'école primaire au lycée. Joël Briand dégage les obstacles que son stagiaire PCL2 a rencontré lors de son expérimentation : difficultés au niveau du vocabulaire, les élèves ayant déjà une connaissance sociale de certains termes statistiques ; difficultés de tenir une activité mathématique qui n'est pas reconnue comme tel par les élèves. Le fait d'expérimenter à la main ne conforte pas les élèves dans le fait qu'ils font des mathématiques. L'arrivée de l'outil informatique leur permettra d'en prendre conscience. Ceci doit nous faire réfléchir quant à l'introduction de la statistique bien avant et surtout sur la manière de l'introduire. Le paragraphe suivant est, à notre sens, encore plus représentatif des problèmes que nous allons soulever par la suite :

> Le professeur doit donc s'engager dans une quête, parfois personnelle, de l'intérêt à s'occuper de statistiques et faire en sorte qu'il permette à des élèves de seconde d'être des citoyens dotés de la culture de l'honnête homme en matière de statistiques. Il doit lutter contre l'idée que les statistiques, « tout le monde en fait un peu ». On a vu que les mots sont déjà utilisés dans les

1. Association des Professeurs de Mathématiques de l'Enseignement Public
2. Professeur de Collège et Lycée 2e année : ce sigle désigne les professeurs stagiaires des IUFM qui avaient réussi le concours et qui étaient donc élèves professeurs ; ils devaient rédiger un mémoire professionnel sur une situation d'enseignement

> pratiques sociales des adolescents, ce qui n'était pas le cas des élèves de l'école primaire. (p. 33)

On le voit, le professeur est amené à s'engager personnellement dans la quête de la statistique. Que ce soit du point de vue de sa formation mathématique, et nous verrons qu'il y a un vrai travail à mener, que du point de vue didactique. Le dernier paragraphe conclut sur la nécessité de faire rencontrer aux élèves les phénomènes de variabilité et de régularité, cela même que Yves Chevallard avait énoncé dans ses *Notes pour la didactique de la statistique* (Chevallard, 1978), ce que nous allons voir maintenant.

1.2 La transposition didactique

La transposition didactique apparaît sous l'égide d'Yves Chevallard, découlant du travail de Michel Verret, dans les années 1980 avec la sortie du livre *La transposition didactique. Du savoir savant au savoir enseigné* (Chevallard, 1985). L'auteur y pose les bases des différentes transformations qui s'opèrent sur le savoir entre le moment où il est produit par les mathématiciens (le savoir savant) et le moment où il est présenté aux élèves, le savoir enseigné.

Mais c'est dans ses *Notes pour la didactique de la statistique* (Chevallard, 1978) que nous allons regarder d'un peu plus près la transposition didactique de la statistique. Ces quelques 300 pages ont été rédigées dans le cadre d'une rénovation de l'enseignement de la statistique pour les étudiants en psychologie. Nous reproduisons ici les titres de ces notes :

- Note 0 : Le problème de la mesure des grandeurs ;
- Note interne 1 : Relation entre deux variables (échelles d'intervalles, échelles de rapport) ;
- Note interne 2 : Relation entre deux variables (échelles nominales, ordinales, d'intervalles) ;
- Note interne 3 : Relation entre deux variables - Situations de fonctionnement ;
- Note interne 5 : La problématique générale de la statistique ;
- Premiers résultats - Sur la transposition didactique dans l'enseignement de la statistique ;
- Problèmes de surdétermination en didactique - La notion de moyenne en statistique.

Nous nous intéresserons plus particulièrement à la première note ainsi qu'aux trois dernières. La première note s'intitule donc *Le problème de la mesure des grandeurs*. Chevallard montre que cette problématique est constituée de deux parties : la première concerne les aspects mathématiques tandis que la seconde concerne les aspects théoriques-expérimentaux. Il présente alors tout un travail sur « la constitution d'une *théorie des rapports de grandeurs*, qui fournisse un cadre mathématique adéquat au problème de la mesure des grandeurs » (p. 4). En partant d'Eudoxe, il en arrive à Stevens (nous faisons là un gros raccourci) ; nous reproduisons un peu longuement Chevallard :

> ... il faut selon Stevens, nettement distinguer les *opérations empiriques* que l'on peut effectuer sur les grandeurs à « mesurer », et les *propriétés mathématiques ou formelles des échelles de mesure*. On retrouve là deux des grands thèmes que nous avons examinés : d'une part le thème *opérationniste* ; d'autre part, le thème des mathématiques comme *langage des sciences*, avec sa neutralité axiologique et ses règles précises de syntaxe ; Stevens ajoute d'ailleurs à cela un troisième aspect : les *statistiques permises* selon les niveaux de mesure. Le *niveau de la mesure* est caractérisé par les *opérations empiriques* possibles ; celles-ci définissent le *groupe mathématique de transformations* admissibles sur la variable définie par la mesure ; à son tour ce groupe détermine les *statistiques invariantes* (pour ce groupe) qui peuvent dès lors être légitimement considérés (p. 92)

Sur la page suivante, il reproduit le tableau que contient un autre article de S. S. Stevens que nous reproduisons aussi tel qu'il est présenté par Yves Chevallard dans le tableau 1.1.

Échelle	Opérations empiriques de base	Structure mathématique de groupe	Statistiques permises (invariantes)	Exemples caractéristiques
Nominale	Détermination de l'égalité	Groupe de permutations $x' = f(x)$ où $f(x)$ désigne toute substitution bi-univoque	Nombre de cas Mode Mesures de l'« information » Corrélation de contingence	Numérotation des joueurs de football (américain) Attribution de numéros de modèles
Ordinale	Détermination de l'ordre : plus ou moins	Groupe isotonique $x' = f(x)$ où $f(x)$ désigne toute fonction monotone croissante	Médian Percentiles Corrélation d'ordre (type 0 : interprétée comme un test d'ordre)	Dureté des minéraux Qualités du cuir, du bois de charpente, de la laine, etc. Test d'intelligence (résultats bruts)
D'intervalle	Détermination de l'égalité des intervalles ou des différences	Groupe linéaire ou groupe affine $x' = ax + b$, $a > 0$	Moyenne Écart type Corrélation d'ordre (type 1 : interprétée comme r) Coefficient de corrélation de Pearson (r)	Température (degrés Fahrenheit et Celsius) Position sur une ligne Temps d'après le calendrier Énergie potentielle Test d'intelligence « standard scores » (?)
De rapport	Détermination de l'égalité des rapports	Groupes de similarité $x' = cx$, $c > 0$	Moyenne géométrique Moyenne harmonique Variation relative	Longueur, numérosité, densité, travail, intervalles de temps, etc. Température (degrés Kelvin) Sonorie (sones) Brillance (brils)

Tableau 1.1 – S. S. Stevens, Le quantitatif et la perception, *Bulletin de psychologie*, p.701

Dans sa note interne 5, Yves Chevallard définit la problématique générale de la statistique dans les termes qui suivent :

> . . . la problématique générale de la statistique est celle 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. Tel est le cadre général des problèmes de la statistique [...] et dans lequel viendront s'inscrire toutes les problématiques régionales ou locales ultérieures [...] (p. 1)

Ainsi nous voyons clairement apparaître l'importance de la variabilité dans l'étude statistique d'un phénomène. Plus encore c'est la dialectique entre régularités et fluctuations qui doit être recherchée. Cette dialectique se compose de plusieurs moments : le premier concerne la régularité derrière la variabilité et la mise en évidence de la tendance centrale ; le second moment met l'accent sur la prise en compte des fluctuations autour de cette tendance centrale. Mais ces problèmes vont ouvrir la porte à d'autres encore :

> L'analyse statistique, [...] comporte en effet les étapes suivantes :
>
> 1. les problèmes du *recueil des données* ;
> 2. les problèmes de *l'analyse qualitative*, et de la *réduction éventuelle*, des données recueillies ;
> 3. les problèmes de la *quantification* ;
> 4. les problèmes de la *signification* ;
> 5. les problèmes de *l'interprétation*. (p. 12)

Ces problèmes, nous allons les retrouver (ou non) dans le travail effectué par les professeurs s'agissant de leur enseignement de la statistique. Nous poserons alors à notre tour les questions : le professeur (ou les manuels) s'intéresse-t-il (et ses élèves avec lui) au problème du recueil des données ? S'intéresse-t-il à la qualité de ces données ? Pourquoi calculer la moyenne (ou la médiane) ? Est-elle significative dans l'étude proposée ? Comment l'interpréter ?

Dans la note suivante, intitulée *Sur la transposition didactique dans l'enseignement de la statistique*, Chevallard montre quelques difficultés de l'enseignement de la statistique. La première est liée au fait que la statistique naît de sa cohabitation avec des domaines extramathématiques. Cette première difficulté a trait donc au fait que les problèmes qui se posent sont à l'origine non statistiques. Et c'est là qu'intervient une seconde difficulté :

> Voilà donc un effet tout à fait clair de l'oubli de l'ensemble *des problèmes que l'on se pose véritablement*, et à quoi on tente de répondre *par le biais de l'analyse statistique*. Celle-ci intervient toujours dans un *champ de savoir structuré par certaines problématiques propres*, - ce que j'appellerai 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. (p. 15)

Chevallard la dénomme déterritorialisation et nous verrons qu'elle est systématiquement à l'œuvre dans l'enseignement secondaire français où l'on s'en tient à rester hors de la problématique de la statistique. Il conclut : « Ce qu'on met en évidence par ces observations, c'est la nécessité, pour un enseignement de statistique, d'amener les sujets à *entrer dans la problématique statistique* » (p. 19). Nous verrons que ce souci n'est toujours pas réglé de nos jours.

Une autre de ces difficultés est la dépersonnalisation du savoir. Citons Chevallard :

> La réduction textualiste du savoir, au-delà de la déterritorialisation (de la désyncrétisation, comme dit Michel Verret) des savoirs, a des conséquences beaucoup plus graves, du point de vue didactique : elle fait l'impasse sur ce qu'on peut désigner comme le *problème fondamental de la didactique*. C'est le problème, disons-le de manière un peu large, des rapports du *sujet connaissant* et du *savoir*. [...]

> Plus brièvement, on parlera de *la constitution du sens par le sujet*,à propos d'un concept ou d'un système de concepts (p. 16)

Quel sens les élèves se constituent-ils de la statistique (en général) et des questions qu'elle permet de poser ? La première question qu'il faudrait se poser est : est-ce que les élèves peuvent se poser ce genre de questions ?

Une troisième difficulté se présente : la programmabilité des savoirs. Or, nous le verrons, l'algorithmisation des savoirs est un trait important de l'enseignement de la statistique. Cependant Chevallard précise que : « Dans l'analyse statistique, en ses 5 étapes telles que nous les avons distinguées, les seules étapes vraiment algorithmisables sont l'étape 3, *le calcul d'indicateurs numériques*, et l'étape 4, l'examen de la *significativité* des indicateurs calculés » (p. 19). On le devine, cette troisième étape va devenir l'essentiel du travail statistique (si on peut le nommer ainsi) que nous rencontrerons dans les classes.

1.3 La XII^e^ école d'été de didactique des mathématiques

Guy Brousseau propose une communication (Brousseau, 2003) lors de cette école d'été. Elle reprend l'expérience de premier enseignement de 1973 en y apportant des commentaires supplémentaires liés à la Théorie des Situations Didactiques [3].

Guy Brousseau commence par faire la distinction qu'il convient d'opérer entre *statistique* et *statistiques*. Il donne d'ailleurs une définition de la statistique en ces termes : « La statistique est l'ensemble des méthodes *scientifiques* à partir desquelles on recueille, organise, résume, présente et analyse des données et qui permettent d'en tirer des conclusions et de prendre des décisions judicieuses. » (p. 8) tiré d'un ouvrage de Murray R. Speigel [4]. La dernière partie de cette citation est particulièrement importante : prendre des décisions. Nous verrons que cet aspect de la statistique n'est que très rarement rencontré, que ce soit dans le cours des professeurs ou dans les manuels. Pour la définition des statistiques, il reprend la définition de Littré : « ce qui a pour objet la statistique ».

Par la suite, Guy Brousseau expose les différents types de situations fondamentales (qui sont au cœur de la TSD) qui concerne la statistique et en dégage trois. Celui qui nous paraît le plus important est le troisième puisque la situation fondamentale « peut engendrer un *processus* qui aboutit à la connaissance de la notion par le jeu des questions qu'elle conduit à se poser et des réponses qu'elle appelle » (p. 19). Nous pouvons d'ailleurs rapprocher ce type de situation à ce que Yves Chevallard appellera dans le cadre de la Théorie Anthropologique du Didactique des Activités d'Étude et de Recherche (AER), voire même un Parcours d'Étude et de Recherche (PER).

Une remarque nous interpelle particulièrement : « [...] devoir calculer la moyenne et l'étendue d'une série de données numérique peut représenter la statistique, si nous pouvons justifier cette activité par une situation qui la rend nécessaire » (p.17). Autrement dit, si cela répond à une question que l'on se pose. Il est inutile de calculer moyenne ou étendue si ce n'est pour répondre à une question, pour prendre une décision. C'est cet aspect, nous le verrons, qui ne se retrouve que trop peu dans ce que nous avons pu observer dans les classes.

Une autre communication sera présentée par Yves Chevallard et Floriane Wozniak dans le cadre d'un cours (Chevallard et Wozniak, 2003a) et d'un atelier. C'est ainsi qu'ils font un constat

3. TSD ; on trouvera dans *Théorie des situations* (Brousseau, 1998) les éléments de base de cette théorie
4. Il s'agit de *Théorie et applications de la statistique*. Série Schaum 1972

qui nous intéresse au plus haut point : « Une distance quasi rédhibitoire se creuse entre savoir à enseigner et savoir enseigné - une distance qui apparaît comme une donnée lourde de l'enseignement actuel » (p. 13). C'est ainsi que la statistique enseignée va s'économiser de plusieurs ingrédients importants. Tout d'abord de la problématique de la statistique : les professeurs, nous le verrons dans la thèse de Floriane Mathieu-Wozniak, méconnaissent grandement la statistique dans ses aspects théoriques, on peut donc supposer qu'il en est de même de la problématique de la statistique.

En second lieu, même si les programmes insistent sur la nécessité d'utiliser des situations de la vie courante, c'est souvent à titre d'artefact et non pas pour que l'on se pose une véritable problématique à l'intérieur d'un domaine de réalité extramathématique [5]. Arrêtons-nous un instant sur cette notion. Dans le cadre de l'atelier de cette école d'été (Chevallard et Wozniak, 2003b), les auteurs souhaitaient d'une part étudier la place qui est allouée au DREM (objets, connaissances et savoirs) et les fonctions que ces objets, connaissances et savoirs assument ; d'autre part, étudier la place qu'occupe le concept de variabilité et les modalités de sa présence concrète. Toute question posée dans un DREM, par exemple apporter la réponse à une certaine tâche problématique t_p, comporte une tâche mathématique t qui doit rendre possible la construction d'une réponse à la tâche problématique t_p. Ils définissent alors l'*occasionnalisme* comme la mobilisation d'un DREM à titre de *decorum* ; l'enjeu de l'étude n'étant plus la réponse à apporter à t_p mais à son alter ego mathématique t.

Troisièmement, la réduction calculatoire de l'enseignement va finir d'étouffer cette statistique déjà amputée de sa raison d'être. Nous verrons aussi que l'interprétation des calculs effectués, lorsque ces interprétations sont présentes, ne relève en général que de l'illusion, les interprétations ayant tendance à devenir de véritables algorithmes prêts à l'emploi, ce que dénonçait d'ailleurs Yves Chevallard dans ses *Notes pour la didactique de la statistique* (*op. cit.*).

Un dernier article de Chevallard et Wozniak nous interpelle, même s'il n'est pas lié à l'école d'été, car il en découle directement. Dans *Enseigner la statistique : un problème de la profession* (Chevallard et Wozniak, 2007), les auteurs s'intéressent aux questions (génératrices) dont l'étude statistique :

> . . . 1) donne à voir la spécificité de l'apport de la science statistique à la compréhension du monde et 2) se réfère à des populations Ω et à des variables X pour lesquelles il est possible et réaliste d'obtenir des échantillons E se traduisant par des séries statistiques $X(E)$ appropriées à l'étude à conduire [...] » (p. 14)

On le voit, pour contrer la réduction calculatoire de la statistique, il faut poser des questions de haut niveau. Il faut donc des données disponibles et cela est parfois difficile à obtenir. Il est nécessaire alors de se tourner vers des données disponibles (par exemple l'INSEE [6]), la question étant choisie alors en fonction des données. Mais le travail sur de petits échantillons est possible aussi ; à condition de laisser de côté le passage à la limite et la stabilisation des fréquences ; la pensée probabiliste ayant tendance à déborder sur la statistique. En effet :

> Le danger que fait courir la proximité « culturelle » (dans l'enseignement des mathématiques) de la notion de probabilité à l'existence d'un enseignement de la statistique comme science de la variabilité est liée à deux contraintes massives, l'une propre à la statistique, l'autre beaucoup plus large. La première est le refoulement de la variation, qui pousse à mettre en avant le « passage à la limite ». Le second est au moins aussi profondément enraciné dans la culture de la profession : de

5. que nous noterons, à l'instar des auteurs DREM
6. Institut National de la Statistique et des Études Économiques

même que la « géométrie théorique » (dans le style euclidien), à travers ses rejetons contemporains, étouffe dans l'œuf la géométrie « expérimentale » au lieu de naître d'elle, de même la statistique théorique, fondée sur le concept de probabilité, asphyxie le travail statistique empirique et expérimental au lieu de le servir en proposant, à travers le fait central de la stabilisation des fréquences, une « algèbre des fréquences ». [...] le programme de 3e qui entrera en vigueur en septembre 2007 comporte une initiation à la notion de probabilité. Il y a là, pour la profession, un grand combat à livrer. (pp. 17-18)

Ce danger ne sera écarté qu'à condition de former les professeurs à cette variabilité et à la nécessaire distance qu'il faudra entretenir avec les probabilités pour qu'elles n'écrasent pas la statistique. Nous allons voir que la formation des professeurs est une des contraintes qui pèsent sur la profession et dont les conséquences sur l'enseignement de la statistique seront visibles, ce que nous observerons plus en détail.

1.4 Les conditions et les contraintes

En 2005, Floriane Wozniak publie sa thèse *Conditions et contraintes de l'enseignement de la statistique en classe de seconde générale. Un repérage didactique* (Wozniak, 2005). Dans son deuxième chapitre, elle étudie le changement apporté par le (nouveau à l'époque, en 2000) programme de mathématiques de la classe de seconde qui introduit notamment la fluctuation d'échantillonnage et la simulation. Ces deux thèmes ont été confortés, nous semble-t-il en tout cas, dans le nouveau programme de 2009.

Le chapitre 4 de la thèse s'intéresse aux conditions et contraintes qui pèsent sur l'enseignement de la statistique. L'auteure étudie les conditions et contraintes produites par les niveaux de détermination didactique que nous reproduisons ci-dessous :

$$\boxed{\text{Civilisation}} \\ \Updownarrow \\ \boxed{\text{Société}} \\ \Updownarrow \\ \boxed{\text{École}} \\ \Updownarrow \\ \boxed{\text{Pédagogie}} \\ \Updownarrow \\ \boxed{\text{Discipline}} \\ \Updownarrow \\ \boxed{\text{Domaine}} \\ \Updownarrow \\ \boxed{\text{Secteur}} \\ \Updownarrow \\ \boxed{\text{Thème}} \\ \Updownarrow \\ \boxed{\text{Sujet}}$$

Une des premières contraintes qu'elle identifie, et qui se situe au niveau de la civilisation est la rareté des statistiques en France comparée à d'autres pays comme, par exemple, les États-Unis. Or cette rareté va se retrouver au sein même de la société française (niveau de la société dans l'échelle ci-dessus), entre ceux qui ont droit de connaître la statistique et les autres. C'est ainsi qu'elle précise :

> La faible pénétration de la statistique dans la culture française « officielle », la non-familiarité avec le maniement statistique de l'information chiffrée doivent être regardés comme des données fondamentales qui imposent de sévères contraintes sur l'enseignement général de la statistique dans un cadre scolaire. (p. 204)

On le voit, la culture française en terme de statistique impose déjà une contrainte à l'enseignement, contrainte qui s'imposera ainsi aux professeurs. Par la suite, en étudiant les différents curriculums disponibles au moment de la réforme du programme de seconde de l'année 2000, elle met en exergue les différences qui existent entre les programmes des classes du lycée, qu'ils soient de la filière professionnelle, générale ou technologique. C'est alors qu'elle note : « On s'achemine ainsi vers un état de la transposition didactique des savoirs statistiques dont nous verrons qu'ils manifestent un *arrêt* historique du travail transpositif : présence côte à côte d'une théorie des probabilités et de notions de statistique dont les liens paraissent moins nécessaires que traditionnels » (p. 214). Par la suite, elle expose une autre contrainte : le savoir statistique pour les enseignants (en général) est fonction de la spécificité des types de tâches qu'ils auront à accomplir. Mais qu'en est-il exactement en mathématiques ? C'est l'objet de la partie suivante, où elle précise « Que la statistique soit regardée comme un savoir pour tous - ce qu'elle n'est pas aujourd'hui - est une condition pour que la culture *commune* inclue une sensibilité et une référence à la statistique » (p. 229). La statistique n'est pas un savoir pour *tous* mais un savoir pour tous ceux qui en auront l'utilité. Ce qui pose le problème (au niveau de la discipline) de la diffusion des savoirs au sein de la discipline mathématique. En effet, les mathématiques ne sont pas la seule discipline « consommatrice » de statistique. En effet, que ce soit les sciences économiques et sociales, la géographie, ou autres, la statistique est présente dans ces disciplines et chacune d'entre elles est amenée à s'intéresser à son enseignement. Quant à en faire une discipline à part entière, la société et l'École avec elle, n'est pas encore prête à franchir le pas. Une autre contrainte relevée est celle de la profession concernant les mathématiques et le commerce qu'elle devrait - mais qu'elle ne fait pas - entretenir avec l'extramathématique. Les autres disciplines ont un commerce avec la statistique au service de leur discipline, mais les enseignants mathématiques ne s'y fondent pas et même ont tendance à se replier sur leur discipline comme un tout duquel il ne faut pas s'éloigner. Et Floriane Wozniak de conclure « Ainsi serait levée une des contraintes qui affectent le plus durement, nous semble-t-il, l'enseignement de la statistique au second degré : l'interdit pesant sur un commerce non prédateur avec les autres savoirs » (p. 239)

Les contraintes sont relativement importantes pour le professeur qui devra enseigner la statistique. Elle s'interroge alors sur les conditions qui pourraient faire émerger une culture statistique. Or une des conditions pour que cette culture puisse être assumée par l'École est décrit par Floriane Wozniak dans le passage qui suit :

> La solution à cette difficulté essentielle semble devoir se trouver dans un travail de transposition didactique qui, à l'instar de ce qui a pu se passer en matière de statique ou de topographie, sélectionne un ensemble fini de types de situations du monde qui, à la fois, soient assez largement représentatifs de l'ensemble des situations du monde sur lesquelles la statistique intervient, et qui,

> en dépit de cela même, constituent un matériau adéquat pour motiver et alimenter l'étude d'une riche diversité de concepts et de « gestes » statistiques. Soulignons qu'il n'y a pas ici, à nos yeux, de *fatum* qui assombrirait le destin scolaire de la statistique, pas plus qu'il n'y en eût, autrefois, concernant la topographie ou la statique : tout tient dans un processus non immédiat *d'élaboration transpositive* qui ne saurait par principe (si l'on vise un enseignement *général* de la statistique) se contenter de reprendre *ne varietur* les élaborations *spéciales* réalisées dans des champs déterminés d'intervention de la statistique – biologie, économie, psychologie, sociologie, éducation, etc. Nous verrons plus loin que ce travail transpositif reste aujourd'hui largement à faire. (p. 235)

Ainsi c'est par un travail transpositif adéquat que la culture statistique pourra intégrer l'enseignement des mathématiques. C'est l'objet de notre travail et nous verrons, de par les observations que nous avons faites, que ce travail transpositif reste à faire, à l'instar de ce que notait Floriane Wozniak.

Le chapitre 5 s'intéresse plus particulièrement à la culture mathématique et à la formation des futurs professeurs. En prenant l'exemple de la formation donnée à l'IUFM d'Aix-Marseille durant la période 1999-2000 jusqu'à 2004-2005, elle relève les difficultés qu'ont les élèves professeurs avec la statistique. Elle identifie alors deux problèmes :

> Le premier, sur lequel on va revenir plus loin, est tout simplement celui de la mauvaise culture des futurs professeurs de mathématiques en matière de statistique, alors même que, dans le curriculum secondaire actuel, ils sont en charge d'une part importante de son enseignement. Le second est, si l'on peut dire, plus propre à la formation des professeurs de mathématiques : il s'agit de l'absence de connaissances – et a fortiori de maîtrise – sur les constituants essentiels de la théorie mathématique de la statistique

Nous l'avons vu plusieurs contraintes viennent peser sur la statistique et la première d'entre elles relevaient d'une mauvaise culture de la statistique en France, que l'on retrouve chez les futurs professeurs. De plus Wozniak soulève le problème de la formation des professeurs. Intéressons-nous à deux exemples de « cours » qui abordent l'aspect théorique de la statistique. Le site *les-mathématiques.net*[7] propose un cours pour l'agrégation externe de mathématiques où la partie statistique comporte 5 pages, quand la partie probabilités en compte 48. Guy Brousseau dans ses *Notes sur l'enseignement des statistiques et/ou des probabilités dans la scolarité commune (6-14 ans)* (Brousseau, 2008) relevait dans une note de bas de page :

> Ayant longtemps dirigé les séances d'exercices de probabilité et de statistiques pour les étudiants de Bordeaux 1, j'ai pu observer la cannibalisation des statistiques par les probabilités, et les deux ensemble par les mathématiques (par exemple, les statistiques consistaient, au langage près, en un cours d'analyse sur la théorie de la mesure, les probabilités, à l'étude des familles sommables etc.). La plupart de nos étudiants bornaient leur ambition à devenir professeurs. Résultat : au-delà de la combinatoire, ils n'avaient guère de familiarité avec les situations de ces deux domaines. Ils les tenaient d'ailleurs pour mineurs, même les plus utiles dans les sciences et dans la société. Au fond, comme leurs professeurs, ils ne se sentiraient redevables, vis-à-vis de leurs élèves, que des questions de mathématiques que l'on y rencontre. (p. 2)

Un autre exemple est la consultation de la plaquette d'Aix-Marseille Université au sujet de la licence mathématiques générales[8] qui montre que les six crédits alloués aux probabilités et statistiques ne seront abordés qu'au 4e semestre pour la première partie et que les six crédits de la deuxième partie ne peuvent être abordés qu'au dernier semestre de la licence et de manière

7. Que l'on peut consulter à l'adresse `http://www.les-mathematiques.net/pages/cours_zip.php`

8. Que l'on peut consulter et télécharger à l'adresse `http://maths-sciences.univ-amu.fr/formations/licences//licence-mathematiques?code=PRSMI301`

facultative. On notera, au passage, l'intitulé du cours *Probabilités et statistiques*. Encore faudrait-il que sur les 6 crédits, il y ait respect de l'égalité entre probabilités et statistiques. Nous ne pourrons pas le confirmer, nous en resterons donc là. Il reste donc un gros travail de formation à fournir.

La suite de ce chapitre s'intéresse particulièrement aux professeurs stagiaires de l'IUFM d'Aix-Marseille et aux questions qu'ils peuvent poser dans le cadre de leur seconde année de formation. Les exemples sont nombreux d'élèves professeurs qui ne sont pas encore familiers de la statistique :

> Le panorama un peu restreint dessiné par les questions posées est, par delà les tentations protestataires et une apparente distanciation, révélateur d'une absence réelle de connaissances chez les professeurs stagiaires au moment où ils doivent, bon gré, mal gré, s'engager dans l'enseignement de la statistique tel qu'il se propose en seconde. Dans certains cas, le questionnement est naïf, sincère et met le doigt sur des difficultés sur lesquelles on pourrait, en bonne méthode, attendre que des lauréats de concours de recrutement soient au clair. (p. 296)

Le chapitre suivant s'intéresse aux contraintes et conditions que produisent les manuels lorsque les professeurs s'appuient dessus. Nous irons, dans la partie suivante, examiner six manuels pour voir quelle transposition de la statistique ils proposent. Nous nous intéressons à un passage concernant la fluctuation d'échantillonnage que Floriane Wozniak expose en ces termes : « Les rédacteurs du programme le rappellent : " au collège, les élèves se sont familiarisés avec les phénomènes variables ". En principe, donc, cette première découverte a été faite, même si l'on n'est pas assuré qu'elle ait été bien faite » (p. 347). Nous irons voir plus particulièrement ce qu'il en est dans la troisième partie de cette étude.

Dans le dernier chapitre, on peut lire :

> Le type d'approche au sein duquel on peut réunir les propositions issues de l'enquête auprès des professeurs de mathématiques de l'académie du Rhône aussi bien que la proposition de module de formation d'Omar Rouan et Jean-Claude Régnier repose sur deux postulats relatifs aux moyens de modifier le système des conditions et contraintes sous lesquelles l'enseignement de la statistique pourrait fleurir au secondaire. Le premier postulat est que les changements recherchés passent par la *formation des professeurs* (et accessoirement par l'amélioration des programmes en tant qu'outil au service des professeurs) ; le second postulat énonce que le travail sur la statistique que réalisent les professeurs doit - notamment dans le cadre des actions de formation envisagées - porter sur un ensemble de questions *très large* par rapport au point de vue qu'ils semblent adopter spontanément devant le projet d'enseigner la statistique au collège et au lycée. (p. 403)

Ainsi, l'amélioration de l'enseignement de la statistique passe par un projet de coopération entre les divers professeurs qui enseignent séparément la statistique ; il faut aussi une formation des professeurs qui soit de nature a permettre l'accès à la statistique au travers de questions larges.

On le voit, les contraintes sont nombreuses à peser sur le professeur. Seule la *profession* peut se saisir du travail à fournir sur ces contraintes pour qu'un véritable enseignement de la statistique puisse voir le jour. Depuis cette étude, de nouveaux programmes pour la classe de troisième et de seconde ont été publié. Nous allons donc les analyser pour voir si la transposition de la statistique a évolué ou si elle en est restée à ce qu'elle était avec les précédents programmes.

Chapitre 2

L'observation du savoir à enseigner

C'est à partir de la lecture de ces textes que nous nous sommes attachés à analyser les programmes, les documents ressources, les manuels, les cours de professeurs et les copies des élèves.

Pour cela nous nous sommes aidés d'un concept important de la TAD, la notion de *praxéologie*. Toute activité mathématique (on pourrait dire toute activité tout court, d'ailleurs) peut se décomposer en des *tâches* qui peuvent relever d'un même *type de tâches* ; par exemple le type de tâches T « Calculer la moyenne d'une série » dont la tâche t « calculer la moyenne des notes de l'élève E » est un spécimen. Ce type de tâches (et donc la tâche avec lui) présente une technique τ qui permet d'accomplir ce type de tâches. On a ainsi la *praxis*, ou pour le dire autrement, le savoir-faire. Vient ensuite la technologie θ qui vient justifier la technique mais permet aussi de l'expliquer et enfin de la produire. Dernier maillon qui constituera le *logos*, le savoir, la théorie Θ qui vient à son tour justifier, expliquer et produire la technologie. La praxéologie (ici nommé ponctuelle car relevant du type de tâches T) est le quadruplet $[T, \tau, \theta, \Theta]$ [1].

Nous venons de voir ce qu'en TAD on appelle l'organisation mathématique ; intéressons-nous maintenant à l'organisation didactique. Elle est composée de six moments de l'étude que nous citons brièvement : le moment de la première rencontre, le moment de l'exploration du type de tâches et l'élaboration d'une technique, le moment de la constitution de l'environnement technologico-théorique, le moment du travail de la technique, le moment de l'institutionnalisation et le moment de l'évaluation. Nous nous sommes particulièrement référé à deux articles d'Yves Chevallard intitulés respectivement *Organiser l'étude : 1. Structures et fonctions* (Chevallard, 2002a) et *Organiser l'étude : 3. Écologie et régulations* (Chevallard, 2002b).

2.1 L'analyse des programmes et documents ressources

Pour trouver des informations sur le savoir à enseigner, tout professeur se tourne vers les instructions officielles : le programme et les documents ressources (que l'on nommait documents d'accompagnement auparavant).

2.1.1 Les programmes de troisième

En 2004 (Bulletin Officiel (B.O.) hors-série n^o 4 du 9 septembre 2004) paraît un nouveau programme pour la classe de 6^e qui va se poursuivre les années suivantes pour les classes de 5^e, 4^e et

1. Pour de plus amples informations, on pourra se référer à *Analyse des pratiques enseignantes et didactique des mathématiques : l'approche anthropologique* (Chevallard, 1998)

3^{e}. C'est donc le programme de 2007 qui nous intéresse dont la première édition paraît au B.O. hors série n^{o} 6 du 19 avril 2007 (MEN, 2007a)

L'*Introduction commune à l'ensemble des disciplines scientifiques* présente un passage qui concerne l'organisation et la gestion des données dont fait partie la statistique :

> *L'organisation et la gestion des données* sont indispensables pour comprendre un monde contemporain dans lequel l'information chiffrée est omniprésente, et pour y vivre. Il faut d'abord apprendre à lire et interpréter les tableaux, schémas, diagrammes, à réaliser ce qu'est un événement aléatoire. Puis apprendre à passer d'un mode de représentation à l'autre, à choisir le mode la plus adéquat pour organiser et gérer les données. [...] En demandant de s'interroger sur la signification des nombres utilisés, sur l'information apportés par un résumé statistique, sur les risques d'erreur d'interprétation et sur leurs conséquences possibles, y compris dans la vie courante, cette partie des mathématiques contribue à former de jeunes adultes capables de comprendre les enjeux et débats de la société où ils vivent. (p. 15)

Cette introduction porte de hautes ambitions pour l'enseignement de la statistique au collège. On retrouve d'ailleurs cette ambition dans l'*Introduction générale pour le collège - Mathématiques* qui complète : « [...] identifier et formuler un problème, conjecturer un résultat en expérimentant sur des exemples » (p. 19). Le programme de 3^{e} quant à lui insiste :

> La pratique des mathématiques en classe de troisième doit permettre aux élèves d'appréhender l'existence de lois logiques et développe notamment :
> - le sens de l'observation, l'imagination raisonnée, l'ouverture d'esprit ;
> - l'esprit critique : distinction entre le probable et l'incertain, situation d'un résultat ou d'une information dans son contexte, attitude critique et réfléchie vis-à-vis de l'information disponible ; (p. 28)

Entrons maintenant dans le corps du programme et plus particulièrement la partie *Organisation et gestion des données, fonctions.* Dans le chapeau introductif, on peut lire :

> Pour les séries statistiques, l'étude des paramètres de position est poursuivie : médiane et quartiles. Une première approche de la dispersion est envisagée. L'éducation mathématique rejoint ici l'éducation du citoyen : prendre l'habitude de s'interroger sur la signification des nombres utilisés, sur l'information apportée par un résumé statistique. De même, c'est pour permettre au citoyen d'aborder l'incertitude et le hasard dans une perspective rationnelle que sont introduits les premiers éléments relatifs à la notion de probabilité. (p. 29)

On le voit, de très hautes ambitions pour ces élèves. Malheureusement, il est à craindre que ces ambitions seront réduites à la portion congrue.

Allons voir le programme à proprement parlé. Il se présente sous la forme de 4 colonnes nommées : Connaissances, Capacités, Exemples d'activités, commentaires et enfin Commentaires spécifiques pour le socle. Nous reproduisons dans le tableau 2.1 la partie *Statistique.*

Nous remarquons alors la contradiction avec le chapeau introductif : la médiane et les quartiles seraient donc considérés par les concepteurs de programme comme relevant des caractères de dispersion ; nous pourrions croire à une erreur de mise en page, mais dans la colonne *Exemples d'activités, commentaires*, le programme insiste : « Le recours aux quartiles permet de préciser la dispersion d'une série par rapport à la seule notion d'étendue. La notion d'intervalle interquartile sera abordée en classe de première » (p. 34). L'ambiguïté sur la dispersion d'une série (quartiles ou écart interquartile) reste entière.

Connaissances	Capacités	Exemples d'activités, commentaires	Commentaires pour le socle
1.3. Statistique Caractéristiques de position	Une série statistique étant donnée (sous forme de liste ou de tableau ou par une représentation graphique :	*Il s'agit essentiellement de mettre en place des éléments de résumé des séries statistiques permettant de contrôler l'information apportée par la moyenne, abordée en quatrième.* Le travail est conduit aussi souvent que possible en liaison avec les autres disciplines dans les situations où les données sont exploitables par les élèves.	
Approche de caractéristiques de dispersion	– *déterminer une valeur médiane de cette série et en donner la signification ;* – *déterminer les valeurs pour les premier et troisième quartiles et donner la signification ;* – *déterminer son étendue.*	*Le fait que contrairement à la moyenne, la médiane ne dépend pas des valeurs extrêmes est dégagé.* *Le recours aux quartiles permet de préciser la dispersion d'une série par rapport à la seule notion d'étendue.* *La notion d'intervalle interquartile sera abordée en classe de première.*	Deux objectifs figurant dans la partie relative à la culture scientifique sont ici visés : – comprendre qu'à une mesure est associée une incertitude ; – comprendre la nature et la validité d'un résultat statistique.
Thèmes de convergence	Exprimer et exploiter les résultats de mesures d'une grandeur SVT, Histoire Géographie, Physique...	La notion de dispersion est à relier, sur des exemples, au problème posé par la disparité des mesures d'une grandeur, lors d'une activité expérimentale, en particulier en physique et chimie. L'utilisation d'un tableur permet d'avoir accès à des situations plus riches que celles qui peuvent être traitées « à la main »	

Tableau 2.1: Programme de 3^{e} de 2007 - p. 34

La dernière partie de ce programme s'intitule *Thèmes de convergence* qui se décompose en six thèmes que nous citons *in extenso* : « Énergie, environnement et développement durable, météorologie et climatologie, mode de pensée statistique dans le regard scientifique du monde, santé et sécurité » (p. 49). L'ensemble de ces thèmes fait appel aux mathématiques et plus particulièrement à la statistique (sauf le thème santé où il n'est pas fait allusion directement à la statistique). Or, nous le verrons, ces thèmes de convergence sont sous-exploités dans les cours des professeurs.

Intéressons-nous plus particulièrement au quatrième. Nous pouvons y lire une idée qui avancerait dans le bon sens puis son contraire dans la phrase qui suit : « Dans le cadre de l'enseignement des mathématiques, les élèves s'initient aux rudiments de la statistique descriptive [...] Ainsi sont mis en place les premiers éléments qui vont permettre aux élèves de réfléchir et de s'exprimer à propos de situations incertaines ou de phénomènes variables » (p. 57) ; la problématique de la statistique n'est-elle pas, ici, remise en question ? En effet, les élèves doivent d'abord apprendre les rudiments de la statistique descriptive pour pouvoir observer les phénomènes de variabilité.

Le 28 août 2008 paraît une seconde version des programmes du collège au B.O. spécial n^o 6 (MEN, 2008). Les modifications sont nombreuses et concerne une réorganisation de l'ancien programme en plus de la réécriture de certaines parties. Nous reproduisons ci-après (tableau 2.2), la partie statistique où les colonnes sont réduites au nombre de trois contre quatre auparavant, la partie Commentaires spécifiques pour le socle ayant disparu :

Connaissances	Capacités	Commentaires
1.3. Statistique Caractéristiques de position	– Une série statistique étant donnée (sous forme de liste ou de tableau ou par une représentation graphique : – *déterminer une valeur médiane de cette série et en donner la signification ;*	Le travail est conduit aussi souvent que possible en liaison avec les autres disciplines dans les situations où les données sont exploitables par les élèves. L'utilisation d'un tableur permet d'avoir accès à des situations plus riches que celles qui peuvent être traitées à la main
Approche de caractéristiques de dispersion	– *déterminer les valeurs pour les premier et troisième quartiles et donner la signification ;* – *déterminer son étendue.*	La notion de dispersion est à relier, sur des exemples, au problème posé par la disparité des mesures d'une grandeur, lors d'une activité expérimentale, en particulier en physique et chimie.
Thèmes de convergence	Exprimer et exploiter les résultats de mesures d'une grandeur	

Tableau 2.2: Programme de 3^e de 2008 - p. 57

On remarque donc que la médiane est reportée, sous la moyenne, au rang d'une caractéristique de position. En revanche, en ce qui concerne les quartiles, le programme ne revient pas sur la première version : ils sont toujours considérés comme des caractéristiques de dispersion. Cette bévue (nous pouvons l'appeler comme telle) n'est pas sans conséquence sur la transposition didactique, certains manuels reprenant à leurs comptes cette erreur ; et nous l'avons vu, les professeurs se réfèrent majoritairement aux manuels scolaires.

Intéressons-nous maintenant aux praxéologies présentes dans ce programme (nous prendrons la version de 2008) et plus particulièrement aux types de tâches que doit enseigner le professeur dans sa classe. Nous y trouvons essentiellement cinq types de tâches :

– T_1^P : déterminer la valeur médiane d'une série ;
– T_2^P : donner la signification de la médiane ;
– T_3^P : déterminer le premier et troisième quartile d'une série ;
– T_4^P : donner la signification des premier et troisième quartiles ;
– T_5^P : déterminer l'étendue d'une série.

Une première remarque s'impose de par son évidence : la signification de l'étendue est laissée de côté. Il faut s'attarder sur les *Commentaires* pour trouver une indication ; il y est précisé que la notion de dispersion est à relier au problème posé par la disparité des valeurs.

Le programme étant avare de commentaires, tournons-nous à présent vers le document ressources.

2.1.2 Le document ressources

Ce document paraît en janvier 2007 sur le site ÉdusSCOL de la DGESCO.[2] Il s'intitule *Ressources pour les classes de 6e, 5e, 4e et 3e du collège* (MEN, 2007b). Nous nous intéresserons plus particulièrement à celui qui concerne l'*Organisation et [la] gestion de données au collège*.

Le document ressources est prévu pour compléter le programme et apporter des indications supplémentaires sur les différentes parties du programme à enseigner. On y trouve aussi, parfois, des compléments mathématiques pour compléter la culture des professeurs. Les objectifs généraux proposent une organisation générale du cours de statistique :

> Il s'agit d'une part de continuer à initier les élèves de collège à la lecture, à l'utilisation et à la production de tableaux, de représentations graphiques, d'autre part de mettre en place les premiers outils de la statistique descriptive, en particulier la notion de résumé statistique à partir de l'étude de quelques caractéristiques de position et de dispersion. Il s'agit aussi, à travers ces premiers contacts, d'aider les élèves à percevoir que la mise en forme de l'information proposée résulte de choix qui en accentuent ou en atténuent certains aspects et donc de contribuer ainsi au développement de l'esprit critique indispensable dans la vie de tout citoyen. (p. 1)

L'insistance sur le calcul des paramètres statistiques est consommée. La notion de résumé statistique est dissociée de la dialectique variabilité - tendance centrale qui permettrait de justifier la recherche des caractéristiques de position et de dispersion.

Après avoir présenté l'intérêt de s'interroger sur la pertinence de telle ou telle représentation graphique, le paragraphe suivant nous intéresse particulièrement puisqu'il s'intitule *Le passage à la statistique* ; on y trouve notamment le passage suivant : « Ce sont les problèmes à résoudre et les calculs éventuels qui orientent le choix d'un type d'organisation et de représentation et introduisent alors la nécessité d'un traitement statistique » (p. 5). Où est la problématique de la statistique ici ? Pourquoi ignore-t-on volontairement le problème de la disparité des mesures, le problème du recueil des données. On voit apparaître que ce sont les calculs (mêmes s'ils sont éventuels) qui donnent la nécessité du traitement statistique : mais de quels calculs parle-t-on ? Le document ne le précise pas ; il y a fort à parier que les professeurs y liront le calcul de la moyenne, de la médiane ou encore des quartiles, le programme ne leur donnant pas plus de

2. Direction Générale de l'Enseignement Scolaire

liberté. La problématique statistique et les étapes d'analyse qui s'y réfèrent viennent d'être jetées au fond du placard sans espoir de retour.

L'exemple proposé dans le document est particulièrement représentatif de cette prise à rebours de la problématique de la statistique. Le document s'intéresse au recueil des données (p. 5 et 6), réalisé chaque jour, par une école de voile afin de mieux connaître son plan d'eau. On ne voit nulle trace de la question génératrice qui devrait pourtant être présente avant ce recueil des données. En fait, nous allons la découvrir, mais située trois pages plus loin (p. 8), après avoir déjà effectué un traitement statistique (calcul d'effectifs et de fréquences) et représenté les données (diagramme en bâtons). C'est là que les auteurs vont préciser :

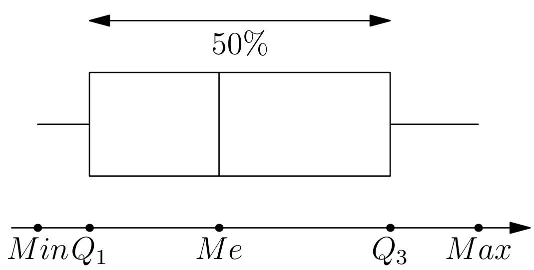
> Les responsables de l'école s'intéressent au nombre de jours où il est impossible de faire sortir les stagiaires pour des raisons de sécurité. Le travail porte alors, par exemple, sur le nombre de jours où le vent a dépassé un certain seuil. [...] Il est alors possible de rendre compte des observations sous différents aspects comme par exemple "la moitié du temps (les trois quarts du temps), le vent a été inférieur à 7 nœuds (10,7 nœuds)" (p. 8)

On le voit la question génératrice est là ! La nécessité du recueil des données s'impose alors (relevé tous les jours de la force du vent) puis de sa représentation (avec l'interrogation sur la pertinence si on ne classe pas les valeurs dans l'ordre croissant) puis de s'intéresser à la distribution des fréquences. L'essentiel y est, mais arrive trop tard ! De plus pas dans l'ordre qui conviendrait pour entrer dans la problématique de la statistique. Malheureusement, le document s'empresse de basculer vers les probabilités : « . . . et plus globalement à la distribution des fréquences pour construire un modèle de loi de probabilité (approche fréquentiste) » (pp. 8-9). La prégnance des probabilités que Guy Brousseau observait il y a quelques années est ainsi consommée.

La section suivante s'intéresse à la notion de résumé statistique. La sous-partie consacrée à la moyenne reprend, sans la mettre en contradiction, cette notion qui est bien connue des élèves de par le calcul de leurs moyennes trimestrielles ; à vrai dire, on pourrait dire que cette notion est même trop connue ; nous en profitons pour raconter une petite anecdote d'un professeur de seconde qui nous relatait, à propos de la statistique auprès de ses élèves, et qui leur posait la question suivante : « Pour vous, la statistique c'est quoi ? » et de lui répondre unanimement : « Calculer la moyenne ! ». On le voit, la chose est mal engagée.

La sous-partie qui concerne l'étendue attire notre attention ; nous reproduisons le premier paragraphe :

> Le seul paramètre relatif à la dispersion d'une série de données dans les programmes de collège est l'étendue. Ce premier élément concernant la notion de dispersion est rudimentaire. Il présente un inconvénient : sa très grande sensibilité aux extrêmes. La détermination des quartiles peut alors compléter la connaissance de la distribution, en considérant l'intervalle interquartile (voir schéma ci-dessous (p. 11)



Or le programme n'est pas très explicite sur la question de l'intervalle interquartile : cette notion est-elle abordée en troisième ? L'ancienne version du programme disait que non. Plus loin il est précisé : « Les diagrammes en boîte ne font l'objet d'aucune étude spécifique au collège » (p. 12) ;

quelle utilité présente alors ce diagramme s'il n'est pas exploitable (c'est-à-dire constructible et compréhensible par les élèves) ? On le verra, ce diagramme sera très présent dans les manuels scolaires que nous avons étudiés ; sans pour autant que l'intérêt d'une telle représentation soit vraiment dégagé : la seule justification qu'ils donneront sera la meilleure visibilité de l'écart interquartile.

Nous le voyons, le document ressources n'est pas plus bavard sur le savoir à enseigner que le programme. On aurait pu y voir des types de tâches qui permettrait de rentrer dans la problématique statistique. La rencontre n'aura pas lieu.

2.1.3 Le programme de 2^{de}

Nous allons maintenant nous intéresser au programme de 2^{de}, qui paraît l'année suivante au B.O. n^o 30 du 23 juillet 2009 (MEN, 2009a). Il succède à celui de 1999 pour tenir compte des modifications apportées par le nouveau programme de 3^{e} que nous venons d'évoquer. La troisième partie de ce programme s'intitule *Statistiques et probabilités.* Remarquons d'emblée que le titre comporte *Statistiques* au pluriel et non *Statistique.* Cela n'est pas anodin. Le chapeau introductif précise que « les cadres relatifs à l'enseignement des *statistiques* et des probabilités sont présentés séparément à la suite l'un de l'autre. Pour autant, ces enseignements sont en relation étroite l'un avec l'autre et doivent faire l'objet d'allers et retours » (p. 8) (c'est nous qui soulignons). On le voit, d'emblée les concepteurs du programme vont oublier les bonnes intentions qui étaient présentes dans le programme de 3^{e} ; le « s » à statistiques en est un exemple frappant ; mais aussi par la relation directe qu'ils imposent à se situer dans la perspective des probabilités. Au risque de faire de la statistique un simple pis-aller pour l'enseignement des probabilités.

Dans le paragraphe qui suit, nous retrouvons un résumé réducteur des objectifs définis dans le programme de 3^{e} :

> … **dans le cadre de l'analyse des données**, rendre les élèves capables
> - de déterminer et interpréter des résumés d'une série statistique ;
> - de réaliser la comparaison de deux séries statistiques à l'aide d'indicateurs de positions et de dispersion, ou de la courbe des fréquences cumulées. (p. 8)

Un objectif nouveau apparaît, celui de la comparaison des séries. C'est la seule mention qui en soit faite dans le programme. Nous reproduisons maintenant le corps du programme à proprement parlé dans le tableau 2.3.

On trouve donc les types de tâches suivants :
- T_1^P : Déterminer la moyenne d'une série ;
- T_2^P : Déterminer la médiane d'une série ;
- T_3^P : Déterminer les quartiles d'une série ;
- T_4^P : Calculer des fréquences à partir des effectifs ;
- T_5^P : Calculer des effectifs cumulés, des fréquences cumulées ;
- T_6^P : Représenter graphiquement une série ;
- T_7^P : Utiliser un logiciel (ou une calculatrice) pour étudier une série statistique ;
- T_8^P : Estimer une proportion inconnue à partir d'un échantillon ;
- T_9^P : Prendre une décision à partir d'un échantillon ;
- T_{10}^P : Simuler des expériences aléatoires ;
- T_{11}^P : Exploiter la simulation d'une situation concrète.

Nous laisserons de côté les types de tâches liés à la statistique inférentielle car nous n'avons pas pu avoir accès aux cahiers des élèves en fin d'année (la proximité des vacances ayant eu raison de la détermination des élèves à prêter leur cahier). Nous nous intéresserons aux sept premiers types de tâches qui concerne donc exclusivement la statistique descriptive.

CONTENUS	CAPACITES ATTENDUES	COMMENTAIRES
Statistique descriptive, analyse de données Caractéristiques de position et de dispersion • médiane, quartiles ; • moyenne.	• Utiliser un logiciel (par exemple, un tableur) ou une calculatrice pour étudier une série statistique ; • Passer des effectifs aux fréquences, calculer les caractéristiques d'une série définie par effectifs ou fréquences ; • Calculer des effectifs cumulés, des fréquences cumulées ; • Représenter une série statistique graphiquement (nuage de points, histogramme, courbe des fréquences cumulées).	L'objectif est de faire réfléchir les élèves sur des données réelles, riches et variées (issues, par exemple, d'un fichier mis à disposition par l'INSEE), synthétiser l'information et proposer des représentations pertinentes.
Échantillonnage Notion d'échantillon. Intervalle de fluctuation d'une fréquence au seuil de 95%* Réalisation d'une simulation	• Concevoir, mettre en œuvre et exploiter des simulations de situations concrètes à l'aide du tableur ou d'une calculatrice. • Exploiter et faire une analyse critique d'un résultat d'échantillonnage	Un échantillon de taille n est constitué des résultats de n répétitions indépendantes de la même expérience. A l'occasion de la mise en place d'une simulation, on peut : • utiliser les fonctions logiques d'un tableur ou d'une calculatrice ; ◊ mettre en place des instructions conditionnelles dans un algorithme. L'objectif est d'amener les élèves à un questionnement lors des activités suivantes : • l'estimation d'une proportion inconnue à partir d'un échantillon ; • la prise de décision à partir d'un échantillon.

* L'intervalle de fluctuation au seuil de 95%, relatif aux échantillons de taille n, est l'intervalle centré autour de p, proportion du caractère dans la population, où se situe, avec une probabilité égale à 0,95, la fréquence observée dans un échantillon de taille n. Cet intervalle peut être obtenu, de façon approchée, par simulation. Le professeur peut indiquer aux élèves le résultat suivant, utilisable dans la pratique pour des échantillons de taille $n \geqslant 25$ et des proportions p du caractère comprises entre 0,2 et 0,8 : si f désigne la fréquence du caractère dans l'échantillon, f appartient à l'intervalle $\left[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}}\right]$ avec une probabilité d'au moins 0,95. Le professeur peut faire percevoir expérimentalement la validité de cette propriété mais **elle n'est pas exigible**

Tableau 2.3: Programme de 2$^{\text{de}}$ de 2009 - p. 8

Attardons-nous quelques instants sur la partie *Échantillonnage* de ce programme, même si nous ne l'observerons pas dans les manuels, les cours et les copies. L'accent est mis sur la simulation mais peu sur la variabilité finalement. Comment les élèves vont-ils pouvoir rencontrer la variabilité et la dialectique variabilité - régularité qui permettrait de justifier les éventuels calculs qui permettraient de faire de la statistique ? Le programme ne répond pas à cette question et malheureusement, le document ressources n'est pas plus explicite à ce sujet, ce que nous allons voir maintenant.

2.1.4 Le document ressources de 2$^{\text{de}}$

Ce document ressources paraît en juin 2009 (MEN, 2009b) sur le site ÉdusSCOL de la DGESCO ; il a pour titre *Probabilités et Statistiques*. Encore une fois, le document ressources présente le même défaut que le programme en mettant un « s » à statistiques. L'introduction qui ouvre le document à le mérite de réparer cette erreur en désignant l'enseignement de la statistique (et des probabilités) comme essentiel pour la formation du citoyen. Plus important encore : « . . . lui donnant des outils pour comprendre l'information chiffrée, *décider et choisir*, de façon éclairée et participer au débat public » (p. 3) (c'est nous qui soulignons). En effet, l'enseignement de la statistique, comme nous l'avons vu aussi dans le programme de 3$^{\text{e}}$, doit permettre de faire des choix de prendre des décisions. Qu'en est-il vraiment quand on regarde les manuels ou les cours des professeurs ?

A propos des questions de fluctuation d'échantillonnage, un paragraphe de cette introduction mérite quelques réflexions. En effet les auteurs proposent d'utiliser un fichier de données réelles comme fil rouge pour mettre en œuvre les différentes notions du programme. Ils argumentent car, disent-ils : « En procédant ainsi, on limite le temps d'appropriation des données et les élèves peuvent plus rapidement se concentrer sur les outils mathématiques, la situation étudiée devenant plus familière » (p. 4) ; or ce qu'on voit là c'est précisément ce que nous voulons éviter : un passage trop rapide sur l'interrogation que les élèves doivent se poser sur le recueil des données et sur l'observation de la variabilité (notre propos n'est pas de dire qu'il faut changer de données systématiquement, au risque que les élèves soient perdus).

La deuxième partie de ce document s'intitule *Des statistiques aux probabilités*. Remarquons tout de suite le déséquilibre entre les sous-parties : deux pages (plus précisément les pages 5 et 6) sont consacrées à la statistique descriptive et l'analyse de données, quand huit pages sont consacrées aux probabilités (les pages 7 à 14), et cinq pages sont consacrées à la l'échantillonnage et la fluctuation d'échantillonnage ; enfin une dernière page est consacrée à l'évaluation. Deux pages sur quinze consacrées à la statistique descriptive (environ 13%) quand la statistique inférentielle en occupe un peu plus de 33% et les probabilités plus de 53%. Alors que faut-il en penser ? Certainement, les rédacteurs ont pensé que les professeurs iraient piocher dans le document ressources du collège, tel qu'ils l'ont préconisé dans l'introduction. Pourtant les professeurs sont, en grande majorité, plus familiers des probabilités que de la statistique ; une conséquence possible est une minoration de l'intérêt des professeurs pour la partie statistique descriptive où pourtant, nous l'avons remarqué, il reste un grand travail à reprendre.

Les deux pages que nous évoquons ci-dessus sont relativement avares en termes d'indications. Elles reprennent essentiellement ce qui se trouve dans le document ressources du collège en donnant quelques informations sur les fréquences cumulées croissantes. En effet la courbe des fréquence cumulées croissantes permet de retrouver une valeur approchée de la médiane ; méthode que les rédacteurs s'empressent de rejeter si « ... l'objectif est uniquement de déterminer une valeur médiane » (p. 6). Ils insistent davantage sur ce qu'elle permet de représenter la distribution des fréquences. Or, nous le verrons, la distribution des fréquences est rarement mise à contribution ; tout au moins l'analyse qu'il conviendrait de faire à son propos.

Nous ne relevons rien de plus sur la statistique descriptive dans ce document ressources qui s'avère donc bien pauvre pour le professeur soucieux de son enseignement.

Nous allons donc maintenant nous intéresser à six manuels scolaires, trois de la classe de 3^{e} et trois de la classe de 2de.

2.2 L'analyse praxéologique de six manuels

La question qui nous intéresse est l'observation de la transposition didactique à l'œuvre en classe de troisième et de seconde. Le savoir à enseigner, dont nous avons pu voir un premier aperçu dans les instructions officielles, est maintenant aux mains des auteurs de manuels scolaires. Nous nous sommes intéressés à trois manuels de troisième puis trois manuels de seconde et y avons mis en évidence les organisations mathématiques et didactiques concernant la statistique.

2.2.1 Les manuels de troisième

Les trois manuels sont récents, tous ayant été publié en 2012 : *Mathématiques collection zénius 3^{e}* (Aleixandre et al., 2012), *Triangle 3^{e}* (Chapiron et al., 2012) et *Nouveau prisme 3^{e}* (Du Roy et al., 2012) [3] Notre but n'est pas de faire une analyse exhaustive de tout ce qu'on peut trouver dans ces trois manuels, mais d'être suffisamment complet en dégageant les points communs ainsi que certaines spécificités qui montrent que les auteurs de manuel méconnaissent la problématique de la statistique, voire méconnaissent la statistique elle-même.

L'organisation mathématique

Les trois manuels commencent par étudier le type de tâches T_1^P : « Déterminer la médiane d'une série ». Intéressons-nous alors aux techniques qui sont présentées ; la notation utilisée est la suivante : $\tau_{i;j}^X$ est la technique numéro j utilisée dans le manuel X à propos du type de tâches T_i. Il en sera de même pour les technologies $\theta_{i;j}^X$ (ou plus généralement θ_i^X). Ainsi nous pouvons trouver :

– $\tau_{1;1}^M$:
  - écrire les valeurs dans l'ordre croissant ;

3. Dans ce qui suit, pour plus de facilité, nous repérerons chaque manuel par la première lettre de l'éditeur. Belin : B, Magnard : M et Hatier : H.

- déterminer le nombre total de valeurs, en les comptant (on le notera N pour plus de facilité) ;
- déterminer si N est pair ou impair ;
  - si N est pair, prendre pour médiane la moyenne des valeurs centrales ;
  - si N est impair, la médiane est la valeur telle qu'il y ait autant de valeurs avant qu'après.

– $\tau_{1;2}^{M}$:
- ranger la série dans l'ordre croissant ;
- déterminer le nombre total de valeurs ;
- déterminer si ce nombre est pair ou impair ;
- déterminer la valeur de la médiane ;
- interpréter le résultat à partir de la définition.

– $\tau_{1;3}^{M}$:
- calcul de l'effectif total ;
- déterminer si ce nombre est pair ou impair ;
- construire une ligne supplémentaire au tableau avec les effectifs cumulés croissants ;
- déterminer la médiane à partir des effectifs cumulés croissants ;
- interpréter le résultat à partir de la définition.

– $\tau_{1;1}^{H}$:
- saisir les données dans une feuille de calcul ;
- déterminer la médiane à l'aide de la formule du tableur.

– $\tau_{1;1}^{B}$:
- classer la série dans l'ordre croissant ;
- déterminer un nombre m tel que la moitié des valeurs lui soient inférieures et l'autre moitié lui soient supérieures ;
- prendre le plus grand nombre de la série inférieure à m et le plus petit de la série supérieure à m et en faire la moyenne ;
- ce nombre est la médiane (par convention).

– $\tau_{1;2}^{B}$:
- déterminer l'effectif total $N = 2n + 1$;
- déterminer si N est pair ou impair :
  - si N est impair, déterminer la $n + 1^{\text{e}}$ valeur ;
  - si les valeurs sont rangées dans l'ordre croissant, la valeur de rang $n+1$ est la médiane.
- on rédige une phrase de réponse.

Une première remarque s'impose, le manuel H ne présente pas de technique pour le calcul de la médiane « à la main » mais seulement à l'aide du tableur. Pour exemple, nous reproduisons l'activité d'*introduction* qui concerne T_1^P où l'on peut voir que les indications sont très vagues :

> Voici les effectifs de la section « jeunes » d'un club de rugby répartis suivant les âges. Quel est l'âge médian ?

Âge (en années)	12	13	14	15
Effectif	10	1	8	10

(p.149)

Comment un élève peut-il être amené à trouver l'âge médian ? Si l'activité a pour but d'explorer le type de tâches T_1, la rencontre n'aura pas lieu. L'élève est condamné à deviner ce qu'est la médiane, ou tout au moins d'aller voir les éléments technologiques dans les pages qui suivent ; ce qui est une manière de rencontrer le type de tâches mais l'exploration de la technique est réduite au strict minimum : son ostension.

Ainsi les techniques pour déterminer la médiane d'une série sont peu souvent correctes, parfois incomplètes (dans B, on ne traite que le cas où l'effectif est impair) ou alors plutôt vagues (par exemple voir dans le manuel M les techniques $\tau_{1;2}^{M}$ ou $\tau_{1;3}^{M}$). En ce qui concerne les éléments technologiques, nous trouvons :

– θ_1^M :

« Une série statistique étant **rangée dans l'ordre croissant**, on appelle **médiane**, la valeur qui partage cette série ordonnée en deux séries de même effectif. **Autrement dit**, il y a autant de valeurs inférieures ou égales à la médiane qu'il y a de valeurs supérieures ou égales à la médiane. » (p. 186)

– θ_1^H :

« On appelle médiane d'une série, un nombre tel qu'il y ait autant de valeurs inférieures à ce nombre que de valeurs supérieures » (p. 153)

– θ_1^B :

« La **médiane** d'une série est nombre, souvent noté **Med**, qui partage la série en deux groupes de même effectif tels que :

- les données du premier groupe sont inférieures ou égales à **Med** ;
- les données du deuxième groupe sont supérieures ou égales à **Med**. » (p. 162)

L'élément théorique Θ_1 est la médiane est la valeur qui minimise la fonction $f(X) = \frac{1}{n}\sum_{i=1}^{n}|x_i - X|$; la théorie étant inaccessible aux élèves.

Le second type de tâches T_2^P concerne la signification de la médiane. Intéressons-nous au manuel M. Les trois premières activités concernent les calculs de la médiane, de l'étendue et des quartiles. Il faut donc attendre la quatrième activité pour que les auteurs s'intéressent à la signification des paramètres calculés dans les activités précédentes :

> Dans une entreprise, le plus petit salaire est 1 200 € ; le plus haut salaire est 4 300 € ; la médiane est 2 400 € ; le premier quartile est 1 700 € et le troisième quartile est 3 000 €
>
> 1. Recopier et compléter chaque phrase à l'aide des informations précédentes :
>   (a) Au moins 25% des salaires sont compris entre ... et ... ;
>   (b) Au moins 50% des salaires sont compris entre ... et ... ;
>   (c) Au moins 75% des salaires sont compris entre ... et ... ;
>   (d) L'étendue de cette série est ... €.
> 2. Cette série peut être représentée à l'aide d'un schéma, appelé **diagramme en boîte**, avec :
>   – Min : la valeur minimale de la série ;
>   – Max : la valeur maximale de la série ;
>   – Q_1 : le premier quartile de la série ;
>   – Q_3 : le troisième quartile de la série ;
>   – Me : la médiane de la série.
>   Réaliser ce schéma pour la série des salaires, en commençant le tracé à 1 200 € et en utilisant comme échelle 1cm pour 400 € sur l'axe horizontal. (p. 185)

Cette interprétation des calculs va tourner à la simple recopie des éléments technologiques du cours avec des phrases « standardisées ». En quoi ces phrases sont-elles des interprétations pertinentes au regard de l'énoncé ? Posons-nous alors les questions suivantes : quelle est la question génératrice qui nécessite une étude statistique ? A priori, il n'y en a nulle trace ; on peut toutefois supposé que la question qui permettrait de s'intéresser aux salaires d'une entreprise est :

« Étant employé dans cette entreprise, est-ce que mon salaire est élevé par rapport à celui de mes collègues de travail ? » ; on le voit, cette question appelle de faire une enquête au sujet des salaires de l'ensemble du personnel de cette entreprise. Une fois les données recueillies, on peut faire un traitement statistique qui donne du sens. Ici ce n'est nullement le cas. Nous ne savons même pas combien de salariés cette entreprise comporte. Le seul but affiché par les auteurs est de faire construire le diagramme en boîte ; or nous l'avons vu, ce n'est pas un objectif prioritaire du programme. Que pouvons-nous dire du DREM ici ? Il sert uniquement de *decorum* à l'instar de ce que relevait Chevallard et Wozniak à l'école d'été de 2003. Avons-nous un semblant d'aperçu de la problématique de la statistique ? Nullement, puisque nous n'avons pas les données qui ont été recueillies.

Pour en revenir au type de tâches T_2^P, nous voyons que l'interprétation qui est faite de la médiane, de l'étendue et des quartiles n'apporte aucun intérêt scientifique. Dans le manuel H, aucune trace de ce type de tâches. Enfin dans le manuel B, nous retrouvons le même travers que dans le manuel M, à savoir que l'interprétation ne fait que reprendre l'élément technologique des paramètres calculés.

En ce qui concerne le type de tâches T_3^P, les choses ne sont pas différentes. Les techniques que nous trouvons sont les suivantes :

- $\tau^M_{3;Q_1;1}$:
  - déterminer la plus petite des valeurs pour laquelle 25% des valeurs lui sont inférieures ou égales.
- $\tau^M_{3;Q_3;1}$:
  - déterminer la plus petite des valeurs pour laquelle 75% des valeurs lui sont inférieures ou égales.
- $\tau^M_{3;Q_1;2}$:
  - ranger la série dans l'ordre croissant ;
  - calculer l'effectif total N ;
  - diviser N par 4 ;
  - déterminer l'entier supérieur à $\frac{N}{4}$;
  - déterminer le 1$^{\text{er}}$ quartile Q_1 correspondant à la $\lceil \frac{N}{4} \rceil^{\text{e}}$ valeur de la série.
- $\tau^M_{3;Q_3;2}$:
  - ranger la série dans l'ordre croissant ;
  - calculer l'effectif total N ;
  - calculer $\frac{3N}{4}$;
  - déterminer l'entier supérieur à $\frac{3N}{4}$;
  - déterminer le 3$^{\text{e}}$ quartile Q_3 correspondant à la $\lceil \frac{3N}{4} \rceil^{\text{e}}$ valeur de la série.
- $\tau^H_{3;Q_1;1}$:
  - calcul de la médiane ;
  - justifier un vrai-faux ;
  - définition des quartiles ;
  - construction d'un diagramme en boîte.
- $\tau^H_{3;Q_3;1}$: idem que $\tau^H_{3;Q_1;1}$.
- $\tau^H_{3;Q_1;2}$ et $\tau^H_{3;Q_3;2}$:
  - saisir les données dans une feuille de calcul ;
  - trouver les formules qui permettent de calculer les premier et troisième quartile ;
  - modifier une valeur de la série pour voir l'influence sur les quartiles.
- $\tau^B_{3;Q_1;1}$:

- trier les données dans l'ordre croissant ;
- déterminer la valeur de la série telle qu'au moins un quart des valeurs lui soient inférieures ou égales.

– $\tau^B_{3;Q_3;1}$:
  - trier les données dans l'ordre croissant ;
  - déterminer la valeur de la série telle qu'au moins trois quarts des valeurs lui soient inférieures ou égales.

– $\tau^B_{3;Q_1;2}$:
  - déterminer l'effectif total N ;
  - calculer $\frac{N}{4}$;
  - Q_1 est la valeur de rang $\lceil \frac{N}{4} \rceil$;
  - on rédige une phrase de réponse.

– $\tau^B_{3;Q_3;2}$:
  - déterminer l'effectif total N ;
  - Calculer $\frac{3N}{4}$;
  - Q_3 est la valeur de rang $\lceil \frac{3N}{4} \rceil$;
  - on rédige une phrase de réponse.

– $\tau^B_{3;Q_1-Q_3;3}$:
  - saisir les valeurs de la série dans une feuille de calcul ;
  - utiliser les fonctions du tableur pour calculer les premier et troisième quartiles ;
  - on note les résultats obtenus.

Les éléments technologiques que l'on peut trouver sont :

– $\theta^M_3 = \theta^B_3$:
  « Le **premier quartile** d'une série, noté Q_1, est la plus petite valeur de la série pour laquelle au moins 25% (ou un quart) des valeurs de la série sont inférieures ou égales à Q_1.
  Le **troisième quartile** d'une série, noté Q_3, est la plus petite valeur de la série pour laquelle au moins 75% (ou trois quarts) des valeurs de la série sont inférieures ou égales à Q_3. » (p. 186)

– θ^H_3 :
  « Le premier quartile(*), noté Q_1, est la plus petite valeur pour laquelle au moins 25% des valeurs sont inférieures ou égales à Q_1.
  Le troisième quartile, noté Q_3, est la plus petite valeur pour laquelle au moins 75% des valeurs sont inférieures ou égales à Q_3.
  (*) Dans cet ouvrage, la notion de quartile est approchée de façon empirique, conformément au document d'accompagnement des programmes (organisation et gestion des données) de 2007 » (p. 153)

En ce qui concerne la théorie, même si elle n'est pas accessible aux élèves, Θ_3 a à voir avec la fonction quantile. Nous n'en dirons pas plus ici. On remarquera aussi, la note relevée dans le manuel M sur le fait que la notion de quartile est approchée de façon empirique. Quel intérêt de le préciser alors qu'aucune remarque n'est faite sur le côté empirique des situations qui sont proposées ?

Quant au type de tâches T^P_4, la signification de ces deux paramètres est à l'image de ce qu'elle était dans le type de tâches T^P_2. Nous n'en parlerons pas plus ici.

Qu'en est-il du type de tâches T^P_5 ? L'étendue, est le seul paramètre de dispersion qui soit disponible au collège. Quel traitement lui réserve-t-on ? Et bien, pour le dire sans détour, c'est encore pire que pour les paramètres de position : il est réduit à sa plus stricte nécessité : le calcul.

Les techniques sont les suivantes :

- τ_5^M :
  - calculer la différence entre la plus grande et la plus petite valeur de la série.
- τ_5^H : la technique n'est pas précisée.
- τ_5^B :
  - déterminer la plus grande valeur de la série ;
  - déterminer la plus petite valeur de la série ;
  - calculer la différence entre la plus grande valeur et la plus petite valeur ;
  - on rédige une réponse de conclusion.

L'élément technologique θ_5 que l'on peu observer est : l'étendue d'une série statistique est la différence entre la plus grande et la plus petite valeur de la série.

Ainsi, aucun des manuels que nous avons observé ne montre l'intérêt de ce calcul. Toutes les activités se contentent de faire le calcul et s'en tiennent à cela. Le manuel B s'abstient même de faire une activité et se contente de donner dans la partie « Cours » l'élément technologique. Alors quel lien avec la dispersion des valeurs ? Aucune remarque ne fait allusion à la dispersion des valeurs en lien avec le calcul de l'étendue. Les interprétations sont souvent plates et n'apportent que peu d'intérêt, avec, ici encore, des phrases « standardisées ».

Nous avons parcouru l'ensemble des types de tâches qui étaient imposés par le programme. Intéressons-nous maintenant au type de tâches qui sont présents dans les manuels mais qui ne le sont pas dans le programme ou dans le document ressources (nous éliminons volontairement les type de tâches qui relèvent des programmes antérieurs ; par exemple « Calculer la moyenne d'une série à l'aide d'un tableur » qui relève de la classe de 4^e).

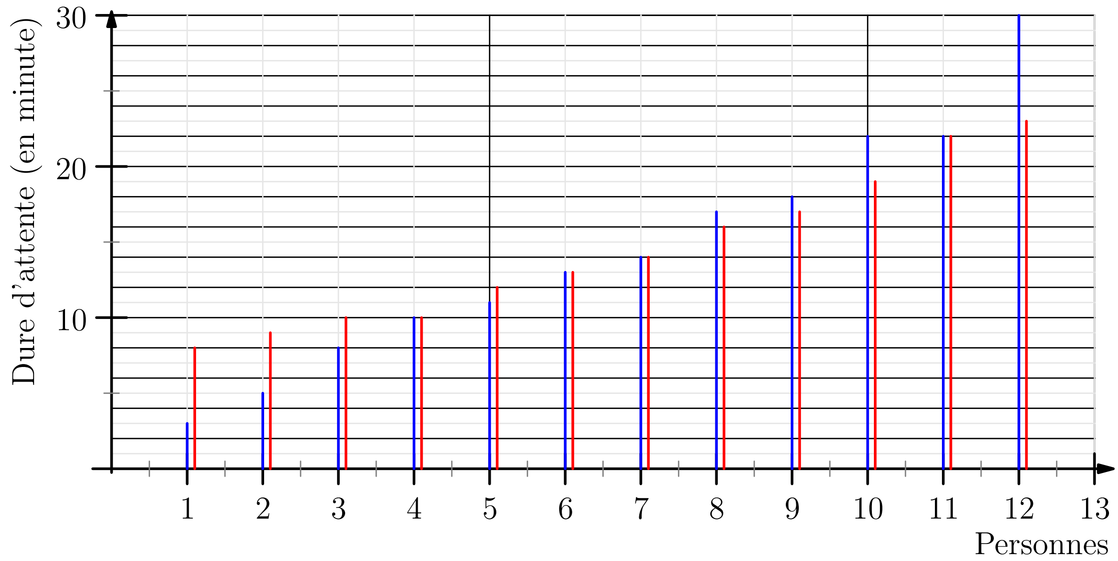
Dans le manuel M, nous ne trouvons pas de type de tâches complémentaires. En revanche, dans le manuel H, dans une section des activités qui s'intitule *Résoudre des problèmes*, nous trouvons les types de tâches suivants :

- T_1^H : interpréter la médiane, la moyenne et l'étendue d'une série ;
- T_2^H : comparer deux séries ;
- T_3^H : contredire une affirmation statistique ;
- T_4^H : analyser un diagramme en boîte.

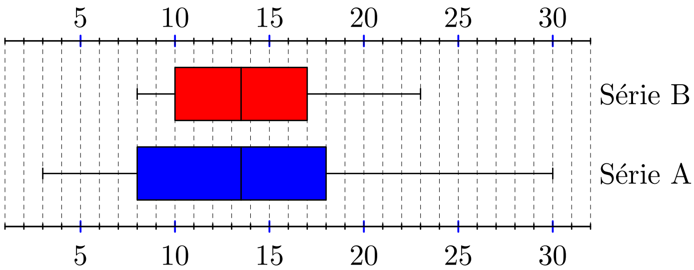
Or les techniques associées ne sont pas connues des élèves, voire même au programme d'une classe ultérieure.

Pour le premier type de tâches, la technique qui est associée est de répondre à un « vrai-faux » en regardant essentiellement si les valeurs numériques données dans les affirmations correspondent aux valeurs de la moyenne, de la médiane ou de l'étendue. Les affirmations sont les phrases standardisées que nous décrivions plus haut. Nous en resterons là pour ce type de tâches.

Le type de tâches T_2^H : « Comparer deux séries » dont le spécimen que nous avons pu observer est t_2^H : comparer « la durée d'attente au téléphone (en minute) relevée par 12 personnes auprès de deux entreprises A et B » (p. 152) n'est pas au programme de la classe de 3^e mais au programme de la classe de 2de. La tentation du diagramme en boîte a été la plus forte. On demande alors aux élèves de calculer les différents paramètres qu'ils connaissent : moyenne, médiane et quartiles et de construire le diagramme en boîte qui correspond. La dernière question est alors formulée en ces termes : « Quelles informations apportent les résultats des questions a) et e) pour comparer les deux séries ? » (*id.*). Mais quelle est la technique pour comparer deux séries ; qu'est-ce que les élèves doivent faire pour pouvoir comparer deux séries ? Est-ce qu'on attend d'eux qu'ils comparent les nombres deux à deux ? Ou doivent-ils prendre l'ensemble des valeurs ? Calculons les différents paramètres demandés ; on observe alors que les moyenne et médiane sont identiques pour les deux séries (plus exactement $\overline{x}_A = \overline{x}_B \approx 14,42$ et $Me_A = Me_B = 13,5$) seuls les 1er et 3e quartiles permettent de distinguer les deux séries du point de vue des résumés statistiques. Allons un peu plus loin et représentons les deux séries par deux diagrammes en bâtons (sur le même graphique) dont nous avons préalablement classé les valeurs dans l'ordre croissant :



Nous remarquons immédiatement que les valeurs des deux séries sont quasiment identiques, hormis pour les extrêmes, bien que la différence entre les valeurs semblent se « compenser » entre les petits temps d'attente et les grands temps d'attente (ce qui peut expliquer les résultats des médianes et des moyennes). On peut aussi observer que les personnes qui attendent peu attendent moins avec l'entreprise A qu'avec l'entreprise B (ce que vient confirmer le calcul de Q_1) tandis que les personnes qui attendent longtemps, attendent plus avec l'entreprise A qu'avec l'entreprise B (ce que confirme le calcul de Q_3). Quelle information supplémentaire peut alors apporter les diagrammes en boîte ? Nous les représentons :



Nous le voyons, nul besoin de recourir aux diagrammes en boîte qui viennent juste confirmer ce que nous savions déjà : la dispersion des valeurs est bien plus grande dans l'entreprise A que dans l'entreprise B ; ce qu'aurait d'ailleurs pu indiquer l'étendue, qui aurait alors pris un sens autre que le simple calcul qu'on lui prête habituellement. L'influence du document ressources est ici tout à fait palpable : ne nous privons pas d'un diagramme supplémentaire qui ravira les élèves (et peut être aussi leur professeur) : ils font de la statistique, après tout !

Le type de tâches T_3^H présent dans l'activité 9 se veut un contre-exemple en statistique. Elle se présente sous la forme suivante :

> Lu dans un journal :
>
> Une étude statistique montre que, dans un couple, une femme a une taille inférieure de 10% par rapport à celle de son conjoint.
>
> « C'est faux, répond Mathieu, ma mère est aussi grande que mon père. »
>
> Que pensez-vous de l'affirmation de Mathieu ? (p. 152)

Quelle technique est mise en avant ? Quelle technologie ? Nous ne pouvons que les imaginer, sans pour autant avoir la certitude que ce sont bien les intentions des auteurs. De plus quelle est la population étudiée ? Est-ce un échantillon représentatif ? Autant de questions qu'il conviendrait de se poser et qui sont absentes de cette activité.

L'organisation didactique

Sur le plan de l'organisation didactique, intéressons-nous à l'étude telle que nous avons pu l'observer dans ces manuels.

Du point de vue des DREM, nous avons vu que les auteurs de manuels ne sont pas des plus imaginatifs. Qu'est-ce qu'un élève connaît de mieux que l'école ? À les croire, on dirait pas grand chose. La réduction scolaire de la statistique que dénonçait Chevallard en 1978 est toujours d'actualité. C'est peut-être justement l'occasion de leur ouvrir les yeux sur le monde qui les entourent et essayer de les intéresser à autre chose que leurs notes et leurs moyennes. Malheureusement les activités où sont présents des DREM autres que l'univers scolaire sont peu nombreuses et souvent ne sont là qu'à titre d'exotisme ; sur 10 activités proposés par les auteurs du manuel H, cinq d'entre elles sont relatives à des notes et deux présentent des nombres sans DREM. Cette réalité est encore plus accentuée dans les parties exercices, où l'on peut relever nombre d'exercices avec pour consigne : calculer la médiane (ou la moyenne ou les quartiles) de telle ou telle série de nombres qui ne représentent rien du tout. Certes il faut des exercices de travail de la technique, mais cela n'oblige pas à ce qu'il y en ait trop. Pour s'en faire une idée, dans le manuel H, 22 exercices sur 58 sont des séries sans aucune indication sur la nature des nombres à exploiter ; parmi ceux qui restent (c'est-à dire 36), 13 concernent les notes obtenues à un devoir, au brevet, etc. Au total plus de la moitié des exercices que les élèves rencontreront seront, soit liés à l'univers scolaire, soit n'auront pas de signification particulière. [4]

Qu'en est-il de la question génératrice ? Dans aucun des trois manuels que nous avons observé, une telle question n'existe. Les activités et les exercices présentent des données sur lesquelles l'élève n'est pas amené à se questionner. Les données sont présentes comme des acquis : elles sont là, on les utilise. La problématique de la statistique est niée : pas d'observations de la disparité des mesures (même si dans quelques exercices, on peut trouver des situations où est évoquée le caractère expérimental des mesures) ni même de l'intérêt de chercher une tendance centrale. La dialectique disparité - tendance centrale n'existe pas.

La dernière observation que nous allons faire est celle qui concerne la partie *Exercices* de chacun de ces trois manuels. Dans la très grande majorité, la (ou les) première(s) question(s) qui suivent l'énoncé consiste à faire des *calculs*, que ce soit la moyenne, la médiane, l'étendue, les quartiles ou encore des fréquences. La statistique en est réduite à une surconsommation de calculs. On le voit, la tâche du professeur, qui consulte souvent les manuels quand il envisage son enseignement, est loin d'être aisée. Mais la Didactique lui permettrait d'y voir plus clair ; ce qui pose le problème de sa diffusion dans la profession.

Intéressons-nous aux moments de l'étude à proprement parlé. Les activités présentent les moments de première rencontre avec les types de tâches, mais la plupart du temps le moment exploratoire de la technique est réduit à sa plus simple expression : l'ostension. Souvent couplée, par ailleurs, avec le moment de constitution du bloc technologico-théorique, lui-même purement et simplement montré. Parfois, on rencontre certains types de tâches pour la première fois lors

4. Dans le manuel M, 31 exercices sur 87 sont sans DREM et 11 exercices sur 56 concernent l'univers scolaire. Dans le manuel B, 33 exercices sur 77 sont sans DREM et 8 sur 44 concernent l'univers scolaire.

de l'institutionnalisation du bloc technologico-théorique dans la partie *Cours* du manuel. Les techniques ne sont pas toujours institutionnalisées, c'est le cas par exemple dans les manuels H et B, et quand elles le sont, c'est souvent sous forme d'exercices corrigés. Les moments de travail de la technique sont présentés dans la partie *Exercices* de ces manuels. Ces exercices présentent aussi les moments d'évaluation de l'organisation mathématique présentée plus avant.

2.2.2 Les manuels de seconde

Les manuels de seconde sont moins récents que ceux de troisième, ils datent tous de 2010, c'est-à-dire l'année qui a suivi l'introduction du nouveau programme : *Maths 2^de^* (Arnoult et al., 2010), *Transmath 2^de^* (Barra et al., 2010) et *Maths repères 2^de^* (Choquer-Raoult et al., 2010).[5] Notons tout de suite que seul le manuel N sépare l'enseignement de la statistique en deux chapitres distincts mais qui se suivent (en l'occurrence les chapitres 5 et 6) sous les titres : *Statistiques* et *Simulation et échantillonnage*.

Le document ressources précise dans son introduction : « En conséquence, il s'agit de veiller à bien inscrire l'enseignement de la classe de seconde en continuité avec le collège » (MEN, 2007b, p. 3) et de préciser : « Il conviendra d'éviter des révisions systématiques et de proposer des situations permettant le réinvestissement des notions abordées dans les classes précédentes » (*op. cit.*). Allons voir ce que propose ces trois manuels.

L'organisation mathématique

Dans la partie *Activités* de ces manuels, nous rencontrons les types de tâches suivants :
- T_1 : déterminer la médiane et les quartiles d'une série ;
- T_2 : calculer la moyenne d'une série ;
- T_3 : comparer deux séries ;
- T_4 : déterminer l'intervalle de fluctuation au seuil de 95% ;
- T_5 : simuler une expérience aléatoire ;
- T_6 : estimer une proportion p inconnue.

La première remarque renvoie aux types de tâches présents dans le programme de 2^de^ en les comparant aux types de tâches des activités des manuels. En effet il y a peu de type de tâches du collège dans les activités ; les auteurs des manuels ont pensé que les types de tâches T_4^P calculer des fréquences à partir des effectifs, T_5^P déterminer des effectifs cumulés, des fréquences cumulées, T_6^P représenter graphiquement une série, et dans une certaine mesure T_7^P utiliser un logiciel (ou une calculatrice) pour étudier une série statistique renvoient à des savoirs-faire du collège et n'ont pas besoin d'être « redécouverts ».

Seuls les deux premiers types de tâches sont des reprises du programme de collège, les autres ont à voir avec l'échantillonnage et la simulation. Intéressons-nous au premier de ces types de tâches ; on ne le trouve que dans le manuel Be. La technique qu'on peut trouver à propos de la médiane est la suivante $\tau_{Me;1}^{Be}$:
- calculer les effectifs cumulés croissants ;
- calculer les fréquences cumulées croissantes ;
- construire la courbe des fréquences cumulées croissantes ;
- déterminer la médiane à partir de la courbe.

5. De la même manière, nous repérerons les manuels de la façon suivante : Nathan : N, Hachette : Ha et Belin : Be (pour les distinguer de ceux de 3^e^)

Or cette technique, cela est bien précisé dans le document ressources de 2^{de}, n'est pas du tout préconisé pour seulement calculer la médiane. Ici, les données sont regroupées en classe d'amplitudes inégales. C'est donc la seule façon de déterminer la médiane et les quartiles (en fait des valeurs approchées) puisque nous n'avons pas l'ensemble des valeurs. On le voit, la technique est encore incomplète : comment détermine-t-on graphiquement la médiane et les quartiles à partir de la courbe des fréquences cumulées croissantes ? Et question plus importante encore : pourquoi ? Quel est l'argument technologique qu'il y a derrière ? L'activité ne le dit pas et encore moins la partie *Cours*. On trouve une autre technique dans la partie *Les capacités* qui correspond à celle présentée au collège, $\tau_{Me;2}^{Be}$:

– déterminer le nombre total de valeur N ;
– 1^{er} cas : si N est impair ;
  • la médiane est la valeur du terme de rang $\frac{N+1}{2}$ dans la série ordonnée.
– 2^{e} cas : si N est pair ;
  • calculer la demi-somme des valeurs des termes de rang $\frac{N}{2}$ et $\frac{N}{2}+1$ dans la série ordonnée.
– $\tau_{Me;3}^{Be}$: (qui complète $\tau_{Me;1}^{Be}$)
  • tracer un repère gradué ;
  • l'axe des abscisses représente les modalités ;
  • l'axe des ordonnées représente les fréquences ;
  • placer les points de coordonnées (modalité ; fréquence) ;
  • relier les points par des segments ;
  • déterminer l'antécédent de 50% pour lire la médiane.

Dans le manuel N, dans la partie *Cours* (p. 113) on trouve la technique τ_{Me}^{N} :

– ranger la liste dans l'ordre croissant ;
– si N est l'effectif total :
  • si N est impair $N = 2k+1$, alors la médiane est la valeur de rang $k+1$;
  • si N est pair $N = 2k$, alors la médiane est la demi-somme des valeurs de rang k et $k+1$.

Enfin dans le manuel Ha, on trouve les techniques suivantes :

– $\tau_{Me;1}^{Ha}$:
  • déterminer l'effectif total de la série ;
  • déterminer la parité de cet effectif total ;
  • si l'effectif total est impair (de la forme $2n+1$), alors la médiane de la série statistique est le terme du milieu, c'est-à-dire le terme de rang $n+1$;
  • si l'effectif total est pair (de la forme $2n$), alors la médiane est la demi-somme des termes de rang n et $n+1$.
– $\tau_{Me;2}^{Ha}$:
  • calculer les fréquences cumulées croissantes ;
  • faire le graphique des fréquences cumulées croissantes ;
  • prendre l'ordonnée 0,5 ;
  • lire l'antécédent de la courbe des fréquences cumulées croissantes d'ordonnée 0,5.

Ainsi nous retrouvons les deux techniques utilisables pour calculer la médiane dans le cas où toutes les valeurs sont connues (caractère quantitatif discret) ou dans le cas où il y a un regroupement en classes (caractère quantitatif continu). En revanche, le fait que la technique du cas continu ne donne qu'une approximation de la médiane, n'est absolument pas relevé dans aucun manuel.

Pour les manuels N et Ha, la technologie $\theta_{Me}^{N,Ha}$ est :

– La médiane d'une série statistique est le nombre, noté Me, tel que :

- 50% au moins des valeurs des individus ont une valeur du caractère inférieure ou égale à Me ;
- 50% au moins des valeurs des individus ont une valeur du caractère supérieure ou égale à Me.

En revanche, pour le manuel Be, la technologie θ_{Me}^{Be} est :

– La médiane, noté M, est une valeur qui partage la série en deux groupes (ensembles, parties) de mêmes effectifs tels que :
  - l'un des groupes contient les données inférieures ou égales à cette valeur ;
  - l'autre groupe contient les données supérieures ou égales à cette valeur.

L'élément théorique Θ_{Me} est la médiane est la valeur qui minimise la fonction $f(X) = \frac{1}{n}\sum_{i=1}^{n}|x_i - X|$; la théorie étant bien sûr inaccessible aux élèves.

Intéressons-nous maintenant à la technique de calcul des quartiles :

– $\tau_{Q;1}^{Be}$:
  - calculer les effectifs cumulés croissants ;
  - calculer les fréquences cumulées croissantes ;
  - construire la courbe des fréquences cumulées croissantes ;
  - déterminer Q_1, Q_3 et la médiane à partir de la courbe.

– $\tau_{Q;2}^{Be}$:
  - calcul de Q_1 :
    - ∘ calculer $\frac{25}{100} \times N$;
    - ∘ calculer la valeur approchée par excès à un près du calcul précédent ; on le note N_1.
    - ∘ Q_1 est la $N_1{}^{\text{e}}$ valeur.
  - calcul de Q_3 :
    - ∘ calculer $\frac{75}{100} \times N$;
    - ∘ calculer la valeur approchée par excès à un près du calcul précédent ; on le note N_3.
    - ∘ Q_3 est la$N_3{}^{\text{e}}$ valeur.

– $\tau_{Q;1}^{Ha}$:
  - déterminer l'effectif total N ;
  - Calculer $\frac{N}{4}$ et $\frac{3N}{4}$;
  - lorsque le résultat n'est pas entier, arrondir à l'entier supérieur ;
  - on obtient la position des 1[er] et 3[e] quartiles de la série.

– $\tau_{Q;2}^{Ha}$:
  - calculer les fréquences cumulées croissantes ;
  - construire le graphique des fréquences cumulées croissantes ;
  - le premier quartile est l'abscisse du point d'ordonnée 0,25 ;
  - le troisième quartile est l'abscisse du point d'ordonnée 0,75.

On le voit, la prégnance des techniques par la courbe des fréquences cumulées est là. Qu'en est-il des technologies ? Nous trouvons dans les trois manuels l'élément $\theta_Q^{Be,N,Ha}$:

– Le **premier quartile**, noté Q_1, est la plus petite valeur de la série statistique telle qu'au moins 25% des données soient inférieures ou égales à Q_1 ;

– Le **troisième quartile**, noté Q_3, est la plus petite valeur de la série statistique telle qu'au moins 75% des données soient inférieures ou égales à Q_3.

Cet élément technologique étant imposé par le programme, nous n'y reviendrons pas. L'élément théorie Θ_Q a à voir avec la notion plus générale de *quantile*, qui est inaccessible aux élèves. On remarquera qu'aucune mention n'est faite du 2[e] quartile et de la différence qu'il peut entretenir avec la médiane.

Intéressons-nous au type de tâches T_2 dont on trouve seulement la trace dans le manuel Ha. On trouve trois techniques différentes :

– $\tau_{2;1}^{Ha}$:
  - Relever l'ensemble des notes du devoir ;
  - additionner toutes les notes ;
  - diviser par le nombre de notes ;
  - en déduire la moyenne du devoir.
– $\tau_{2;2}^{Ha}$:
  - Regrouper les notes et leurs effectifs correspondants ;
  - calculer la moyenne (pondérée) de l'ensemble des notes ;
  - en déduire la moyenne du devoir.
– $\tau_{2;3}^{Ha}$:
  - Regrouper les notes par classe d'amplitudes distinctes ;
  - calculer le centre des classes ;
  - calculer la moyenne (pondérée) de l'ensemble des classes ;
  - en déduire la moyenne du devoir.

On le voit, cette activité se propose de recenser toutes les techniques qui permettent de calculer la moyenne. Le but est de pouvoir aussi s'interroger sur la dernière technique et voir qu'elle ne donne qu'une valeur approchée (plus ou moins bonne) de la moyenne ; on lui préférera donc, lorsqu'on est en possession de l'ensemble des valeurs, la première ou la seconde de ces techniques. L'élément technologique θ_2 présent dans les manuels est le suivant :

– θ_2^{Be}

  « La moyenne d'une série statistique, calculée à partir des effectifs est :

$$\overline{x} = \frac{n_1x_1 + n_2x_2 + \ldots + n_px_p}{N}$$

  Elle se calcule également avec les fréquences : $\overline{x} = f_1x_1 + f_2x_2 + \ldots + f_px_p$. » (p. 136)

– θ_2^{Ha}

  « On considère une série statistique dont les valeurs du caractère sont $x_1, x_2, x_3, \ldots, x_p$ et les effectifs associés : $n_1, n_2nn_3, \ldots, n_p$.

  La **moyenne** de cette série statistique, notée $\overline{x}$, a pour valeur :

$$\overline{x} = \frac{n_1 \times x_1 + n_2 \times x_2 + \ldots + n_p \times x_p}{n_1 + n_2 + \ldots + n_p}$$

  » (p. 132)

– θ_2^{N}

  « La **moyenne** d'une série statistique est le nombre, noté $\overline{x}$, défini par :

$$\overline{x} = \frac{n_1x_1 + n_2x_2 + \ldots + n_px_p}{N}$$

  Effectif total $N = n_1 + n_2 + \ldots + n_p$ » (p. 113)

On le voit ces éléments technologiques n'en sont pas : ce sont les techniques qui permettent de calculer les moyennes. Un élément technologique qui pourrait convenir (à ce niveau) serait plutôt : la moyenne est la valeur qui serait telle que toutes les valeurs du caractère soient identiques. L'élément théorique Θ_2 est la moyenne est la valeur qui minimise la fonction $g(X) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - X)^2}$

Voila tout ce qui concerne la poursuite de l'enseignement de la statistique après le collège. Les types de tâches suivants concernent plus spécialement le programme de 2^{de}. Intéressons-nous alors au troisième type de tâches, qui est de comparer deux séries. Les techniques qu'on peut y trouver sont les suivantes :

- $\tau_{3;1}^{N}$:
  - représenter graphiquement la série ;
  - comparer les espérances de la vie des hommes et des femmes ;
  - examiner l'écart entre l'espérance de vie des hommes et des femmes sur les 20 dernières années.
- $\tau_{3;2}^{N}$:
  - saisir les données sur une feuille de calcul ;
  - les trier dans l'ordre croissant ;
  - utiliser les formules du tableur pour obtenir les paramètres des séries ;
  - saisir les données de la nouvelle série et calculer ses paramètres ;
  - comparer les paramètres des deux séries.
- $\tau_{3;1}^{Be}$ qui n'est pas explicitée. Quelques éléments laissent penser qu'il faut :
  - comparer la médiane et la moyenne :
    - si $Me = \overline{x}$, alors la distribution de la série est symétrique ;
    - sinon, la distribution de la série n'est pas symétrique.
  - déterminer l'importance de l'étendue et de l'intervalle interquartile ;
  - conclure.

On le voit, les techniques proposées par les manuels ne sont pas très explicites. En ce qui concerne la technologie qui justifierait cette technique, nulle trace et encore moins de la théorie.

Le quatrième type de tâches est de calculer des fréquences à partir des effectifs. Il est assez surprenant que ce type de tâches soit présent explicitement dans le programme de 2$^{\text{de}}$ puisqu'il est un type de tâches d'une classe antérieure, plus précisément de la classe de 5$^{\text{e}}$.[6] On peut supposer que ce type de tâches est présent à cause du type de tâches suivant : calculer des effectifs cumulés et des fréquences cumulées.

Alors que trouve-t-on de ce type de tâches dans les manuels ? Dans les activités des manuels, rien. il faut attendre le « Cours » pour trouver les éléments technologiques suivants :

- $\theta_{4;1}^{Be}$:
  - « **L'effectif** n_i associé à une modalité x_i est le nombre d'éléments dont le caractère est égal à cette modalité. On note N l'effectif de la population totale. » (p. 134)
- $\theta_{4;2}^{Be}$:
  - « La **fréquence** f_i de la modalité x_i est le quotient de son effectif n_i par le nombre total d'éléments, soit $f_i = \frac{n_i}{N}$. » (p. 134)
- $\theta_{4;1}^{Ha}$:
  - « L'effectif d'une valeur du caractère est le nombre de fois où cette valeur apparaît dans la série. » (p. 130)
- $\theta_{4;2}^{Ha}$:
  - « La fréquence d'une valeur du caractère est le quotient de l'effectif de cette valeur par l'effectif total. » (p. 130)
- Dans le manuel N, les éléments technologiques sont rappelés sans les mettre en exergue au travers d'un exemple.

6. On trouvera la mention du calcul des fréquences à la page 20 du programme paru au B.O. du 28 août 2008 (MEN, 2008)

C'est tout ce qu'on observe dans la partie « Cours » des manuels. La partie « Exercices » quant à elle ne donne pas d'exercices spécifiques de calculs de fréquences mais ce type de tâches est intégré dans des exercices ciblant un autre type de tâches.

Passons maintenant au cinquième type de tâches, calculer des effectifs cumulés et des fréquences cumulées. On notera, qu'il n'est pas précisé effectifs cumulés *croissants* ou *décroissants* et de même pour les fréquences cumulées. Quelles sont les techniques disponibles ?

- $\tau_{5;1}^{Be}$: cette technique n'est pas visible dans l'activité du manuel Be (p. 132) mais l'énoncé de la question demande explicitement de l'utiliser.
- $\tau_{5;2}^{Be}$: (effectifs cumulés croissants et fréquences cumulées croissantes)
  - calculer le nombre d'exploitations dont la surface est inférieure ou égale à 15 hectares ;
  - calculer le nombre d'exploitations dont la surface est inférieure ou égale à 30 hectares ;
  - procéder de même pour 45, 60 et 100 hectares ;
  - puis diviser les effectifs cumulés croissants par N, le nombre total d'exploitation.

  $\tau_{5;1}^{Ha}$: (fréquences cumulées croissantes)
  - ranger les valeurs dans l'ordre croissant ;
  - pour chaque valeur, ajouter sa fréquence à la fréquence cumulée croissante précédente.

  $\tau_{5;2}^{Ha}$: (fréquences cumulées décroissantes)
  - ranger les valeurs dans l'ordre croissant ;
  - pour chaque valeur, ajouter sa fréquence à la fréquence cumulée croissante précédente.

Le manuel N, ne propose pas de technique explicite, mais seulement de donner un exemple. Les éléments technologiques présents sont :

- $\theta_{5;1}^{Be}$:
  - « **L'effectif cumulé croissant** associé à une modalité x_i est égal au nombre d'éléments dont la modalité est inférieure ou égale à la modalité » (p. 134)
- $\theta_{5;2}^{Be}$:
  - « La **fréquence cumulée croissante** (resp. décroissante) associée à la modalité x_i est égale à la somme des fréquences dont la modalité est inférieure ou égale à x_i » (*id.*)
- θ_{5}^{Ha} :
  - « La **fréquence cumulée croissante** (respectivement **décroissante**) d'une valeur est la somme des fréquences des valeurs qui lui sont inférieures (respectivement supérieures) ou égales » (p. 130)

Pour le manuel N, la même remarque que pour les techniques s'impose, aucun élément technologique n'est présent. Quant aux autres, les éléments technologiques qu'ils donnent, ne sont en fait que les techniques. La théorie Θ_5 a à voir avec la fonction de répartition qui est inaccessible aux élèves.

Le sixième type de tâches est une reprise des programmes antérieurs, puisque les premières représentations graphiques sont étudiées dès la classe de 6e. Nous n'en trouvons qu'une seule trace dans le manuel Be à propos de la construction d'un histogramme. On trouve alors la technique suivante :

- $\tau_{6;1}^{Be}$:
  - toutes les classes ont la même amplitude $(b_i - a_i)$ et n_i son effectif associé ;
  - pour chaque classe, on représente un rectangle de largeur $b_i - a_i$ et de hauteur n_i.
- $\tau_{6;2}^{Be}$:
  - Les classes ont des amplitudes différentes ;
  - pour chaque classe, on représente un rectangle de largeur $b_i - a_i$ et de hauteur $h_i = \frac{n_i}{b_i - a_i}$.

On s'aperçoit rapidement que la technique est particulièrement édulcorée. La construction d'un histogramme est nettement plus complexe. Par exemple, une recherche sur Wikipedia donne l'extrait suivant à propos de la construction :[7]

> – Collecte des données
> La première phase est la collecte des données en cours de fabrication. Cette collecte peut être réalisée soit de façon exceptionnelle à l'occasion de l'étude de la variable soit en utilisant un relevé automatique ou manuel fait lors d'un contrôle réalisé dans le cadre de la surveillance du procédé de fabrication.
> Sans qu'il soit réellement possible de donner un nombre minimum, il faut que le nombre de valeurs relevées soit suffisant. Plus on dispose d'un nombre élevé de valeurs, plus l'interprétation sera aisée.
> – Nombre de classes
> La première opération est de déterminer le nombre de classes de l'histogramme. Généralement, dans le cadre d'une analyse de ce type, on utilise des classes de largeur identique.
> Le nombre de classes dépend du nombre de valeurs N dont on dispose.
> Le nombre de classes K peut être déterminé par une des formules suivantes : $K = 1 + \frac{10\log(N)}{3}$ ou plus simplement $K = \sqrt{N}$.
> Cependant, l'histogramme étant un outil visuel, il est possible de faire varier le nombre de classes. Ceci permet de voir l'histogramme avec un nombre différent de classes et ainsi de trouver le meilleur compromis qui facilitera l'interprétation. L'utilisation d'un logiciel dédié ou plus simplement d'un tableur facilite cette opération.
> – Intervalles de classe
> L'amplitude (minimale) w de l'histogramme est $w =$ valeur maximale $-$ valeur minimale.
> Il peut cependant être intéressant pour obtenir un histogramme plus parlant de choisir une amplitude plus large que l'amplitude minimale.
> L'amplitude h théorique de chaque classe est alors : $h = \frac{w}{K}$
> Il faut arrondir cette valeur à un multiple de résolution de l'instrument de mesure (arrondi à l'excès).

En revanche, on ne trouve nulle trace de l'élément technologique et encore moins de la théorie.

Le dernier type de tâches concerne la détermination des différents paramètres statistiques à l'aide de la calculatrice ou du tableur. C'est aussi un type de tâches qui a déjà été abordé au collège, notamment en 3e. On trouve les techniques suivantes :

- $\tau_{7;1}^{Be}$: (*Capacités*)
  - effacer les anciennes données sur la calculatrice ;
  - saisir les modalités ;
  - saisir les effectifs ;
  - afficher les caractéristiques à l'aide de la calculatrice.
- $\tau_{7;2}^{Be}$: (*Capacités*)
  - rentrer les données dans la feuille de calcul ;
  - calculer les produits $n_i \times x_i$;
  - faire la somme des produits précédents ;
  - diviser par l'effectif total ;
  - entrer la formule de la médiane ;
  - entrer la formule du premier quartile ;
  - entrer la formule du troisième quartile.

7. page consultée le 15 août 2013

– $\tau_{7;1}^{Ha}$: la technique consiste essentiellement à faire faire à la calculatrice, les techniques qui n'ont pas été vues par ailleurs au sujet du type de tâches T_4^P.

Seul le manuel Be propose ce type de tâches de façon claire. En revanche pas de trace des éléments technologico-théoriques.

L'organisation didactique

Intéressons-nous maintenant à l'organisation de l'étude. Le nombre d'activités concernant la statistique descriptive est fortement limité par les auteurs de manuels. Nous analyserons l'étude de la même façon que nous l'avons fait pour les manuels de 3^{e}.

En premier lieu, regardons d'un peu plus près les DREM. Le constat est un peu plus engageant que pour la classe de 3^{e} ; les situations rencontrées dans les activités sont les suivantes :

– espérance de vie des hommes et des femmes de 1817 à 2007 ;
– températures maximales du mois de juillet 2009 à La Rochelle ;
– Le parc des voiliers de moins de 24 mètres au 31 août 2008 ;
– les devoirs de maths que rend le professeur.

A première vue, on pourrait croire que l'univers scolaire est en nette régression ; allons voir la partie *Exercices* pour en savoir un peu plus. Dans le manuel Be, 15 exercices sur 49 sont sans DREM et 4 exercices sur ceux restants (donc 34) concernent l'univers scolaire. [8] La tendance semble se confirmer, les exercices sans DREM sont beaucoup moins fréquents que ce qu'on pouvait trouver dans les manuels de 3^{e} ; le plus inquiétant est que ces manuels sont plus *anciens* que ceux de 3^{e}. Faut-il y voir un retour en arrière ? Il faudra peut-être attendre de nouvelles éditions des manuels de 2de pour en être sûr. Cependant cette (petite) amélioration ne doit pas cacher que les DREM qui sont rencontrés ne servent que de décor à une véritable surconsommation de calculs ; comme nous l'avions déjà trouvé dans les manuels de troisième.

Comme nous avons pu voir plus de DREM dans ces activités et exercices, qu'en est-il des questions génératrices ? Malheureusement, cet aspect de la statistique est toujours autant négligé pour ne pas dire nié. Aucune question ne vient s'intéresser au recueil des données, à la nécessité d'écarter certaines valeurs qui pourraient être aberrantes (hormis dans le cas du calcul de la moyenne, pour montrer l'influence des valeurs extrêmes), de l'intérêt de résumer la série par la moyenne ou par la médiane, de s'intéresser à la distribution des fréquences, etc. En bref, la dialectique variabilité - régularités n'existe pas.

Nous le voyons, la problématique de la statistique est tout autant niée dans ces manuels que dans ceux que nous avons pu observer pour la classe de 3^{e}. Comment pourrait-on modifier ces manuels pour y voir aborder une véritable pensée statistique ? la question mériterait qu'on s'y arrête plus longtemps, ce que nous ne ferons pas ici.

Voyons maintenant les moments de l'étude à proprement parlé. Dans la plupart des manuels, les moments de première rencontre avec les types de tâches ont lieu dans les activités ; on note cependant que certains moments de nouvelle rencontre (par exemple le calcul des fréquences) ne se trouve que dans la partie *Cours* c'est-à-dire le moment de l'institutionnalisation. Les moments exploratoires des techniques sont relativement peu nombreux, voire dans certains cas réduits à la pure ostension de la technique agrémentée de la technologie. Les moments de constitution des blocs technologico-théoriques sont absents de la plupart des manuels, dans le sens où la constitution est réduite à de l'ostension. Les moments de travail des techniques restent nombreux ; mais

8. Dans le manuel N, 1 exercice sur 57 est sans DREM et 10 exercices sur 56 concernent l'univers scolaire. Dans le manuel Ha, 12 exercices sur 73 sont sans DREM et 4 exercices sur 61 concernent l'univers scolaire.

que dire du travail d'une technique que l'on a à peine explorer ? Le moment de l'institutionnalisation reste une (parfois) longue liste d'éléments technologiques se succédant les uns après les autres. Les moments d'évaluation sont présents au travers de quelques exercices (corrigés la plupart du temps) pour une auto-évaluation qu'il conviendrait d'interroger d'un point de vue didactique, mais ce n'est pas notre propos ici. L'organisation didactique ainsi présentée nous permet d'imaginer les effets d'un tel enseignement de la statistique. Nous en tirerons quelques conclusions un peu plus loin.

Voila un panorama assez peu engageant de l'enseignement de la statistique tel qu'il est pensé par les auteurs de manuels. Allons voir maintenant ce que les professeurs peuvent mettre en place dans leurs classes avec les conditions et les contraintes qui s'imposent à eux.

Chapitre 3

L'observation du savoir enseigné

Deux professeurs de troisième et deux professeurs de seconde ont accepté de nous transmettre les cahiers de leurs élèves quant à l'enseignement qu'ils ont donné sur la statistique. Pour plus de facilité, nous désignerons ces professeurs par leurs initiales.

3.1 Le cours de LC

Ce professeur enseigne dans une classe de 3e DP6 (qui vient d'être réformée et s'intitule maintenant Prépa-Pro) [1], qui est composée d'une vingtaine d'élèves. On trouvera le cours de ce professeur en annexe 5.

Le cours de ce professeur s'organise en deux parties. La première concerne la séquence 3 de l'année intitulée *Données et représentations statistiques*, tandis que la seconde, qui concerne la séquence 9, s'intitule *Calculs statistiques - Probabilités*. Cette distinction pourrait être intéressante car elle oblige à ne pas se focaliser sur l'aspect calculatoire de la statistique ; en revanche la proximité des probabilités dans la seconde partie est un signe de ce que Brousseau appelait la cannibalisation de la statistique par les probabilités.

3.1.1 L'organisation mathématique

La première séquence commence par le type de tâches T_1 : Préciser le vocabulaire lié à la statistique, qui présente les différents éléments technologiques dont auront besoin les élèves : notion de population, échantillon, individus, caractère, caractère qualitatif, caractère quantitatif.

Le second type de tâches abordé est T_2 : Représenter graphiquement une série statistique, dont trois spécimens sont étudiés :

- $t_{2;1}$: représenter une série par un diagramme en bâtons ;
- $t_{2;2}$: représenter une série par un diagramme circulaire ;
- $t_{2;3}$: représenter une série par un histogramme.

Les techniques associées sont les suivantes :

- $\tau_{2;1}$:
  - tracer l'axe des abscisses qui correspond au caractère étudié ;
  - tracer l'axe des ordonnées qui correspond aux effectifs ;
  - les hauteurs des bâtons sont proportionnelles aux effectifs.
- $\tau_{2;2}$:

1. Ces classes proposent un enseignement de Découverte Professionnelle durant 6 heures par semaine et sont principalement ouvertes dans les lycées professionnels

- la valeur de l'angle est proportionnelle à l'effectif ;
- pour trouver la valeur de l'angle, on peut faire un tableau de proportionnalité.

– $\tau_{2;3}$:
- un histogramme se compose de rectangles accolés ;
- si les classes ont la même amplitude, les hauteurs des rectangles sont proportionnelles aux effectifs.

Enfin les éléments technologiques convoqués sont :
– $\theta_{2;1}$ diagramme en bâtons ;
– $\theta_{2;2}$ diagramme circulaire ;
– $\theta_{2;3}$ histogramme.

Les éléments théoriques ne sont pas présents car inaccessibles aux élèves.

S'ensuit une liste de 8 exercices dont les élèves ne feront que les 7 premiers. On y retrouve les type de tâches T_1 et T_2 mais aussi des type de tâches qui n'apparaissaient pas dans le cours. Ces exercices contiennent essentiellement des type de tâches des programmes antérieurs ; nous les listons :

– T_3 : Identifier le caractère étudié ; la technique associée est τ_3 :
  - repérer dans l'énoncé quelle est l'information qui étudiée dans la population.

  L'élément technologique θ_3 est : le caractère est la modalité qui est étudiée sur la population.

– T_4 : Identifier la population statistique ; la technique associée est τ_4 :
  - repérer dans l'énoncé sur quels sont les individus qui sont étudiés.

  L'élément technologique θ_4 est : la population est l'ensemble des individus sur lesquels porte l'étude statistique.

– T_5 : Calculer l'effectif total d'une population ; la technique associée est τ_5 :
  - ajouter les effectifs de toutes les valeurs de la série.

  L'élément technologique θ_5 est : l'effectif total est la somme des effectifs des modalités du caractère.

– T_6 : Calculer la somme des produits $n_i \times x_i$; la technique associée τ_6 :
  - pour chaque valeur, multiplier la valeur du caractère par l'effectif ;
  - faire la somme des produits calculés.

  L'élément technologique θ_6 est : $N \times \overline{x} = \sum_{i=1}^{p} n_i \times x_i$.

– T_7 Déterminer la fréquence d'une modalité ; la technique associée τ_7 est :
  - prendre l'effectif de la modalité concernée ;
  - diviser par l'effectif total.

  L'élément technologique θ_7 est : $f_i = \frac{n_i}{N}$.

– T_8 : Déterminer la nature du caractère d'une série ; la technique associée est τ_8 :
  - déterminer le caractère étudié ;
  - si le caractère est un nombre alors il est quantitatif :
    - si les nombres sont isolés, alors il est quantitatif discret ;
    - sinon, il est quantitatif continu.
  - si le caractère n'est pas un nombre, alors il est qualitatif.

  L'élément technologique θ_8 : la nature du caractère.

– T_9 : Calculer un effectif dans un tableau connaissant l'effectif total et les autres effectifs ; la technique associée est τ_9 :
  - additionner tous les effectifs connus ;
  - soustraire à l'effectif total, la valeur calculée précédemment.

  L'élément technologique $\theta_9 = \theta_5$.

- T_{10} : Déterminer la valeur d'un angle au centre sur un diagramme circulaire ; la technique associée est τ_{10} :
  - prendre la fréquence correspondante au secteur considéré ;
  - multiplier cette fréquence par 360°.

  L'élément technologique θ_{10} est :
  - ○ le diagramme circulaire ;
  - ○ la fréquence : $f_i = \frac{n_i}{N}$.
- on retrouve le type de tâches T_7 : Déterminer les fréquences d'un caractère à partir des effectifs ; la technique associée est τ_7 et l'élément technologique θ_7.
- T_{11} : Déterminer les fréquences cumulées croissantes ; la technique associée est τ_{11} :
  - recopier la première fréquence ;
  - dans la case qui suit, ajouter au nombre précédent, la fréquence de la valeur du caractère qui suit.

  L'élément technologique θ_{11} : les fréquences cumulées croissantes.

On remarquera le dernier type de tâches T_{11} qui n'est pas au programme de 3e mais de la classe suivante.

La séquence suivante débute par l'étude du type de tâches T_{12} : « déterminer la médiane d'une série » avec l'étude de la tâche t_{12} : « déterminer la médiane d'une série de notes ». On note cependant que cette activité réintroduit des types de tâches des programmes antérieurs, que nous listons ici :

- T_{13} : déterminer l'effectif des valeurs inférieures à une valeur donnée de la série ; la technique associée τ_{13} :
  - déterminer les valeurs de la série qui sont inférieures à la valeur 10 ;
  - additionner les effectifs correspondants.

  L'élément technologique θ_{13} est l'effectif cumulé croissant.
- T_{14} : calculer la moyenne d'une série ; la technique associée est τ_{14} :
  - pour chaque valeur de la série, faire le produit de la valeur par son effectif ;
  - faire la somme des produits ;
  - diviser par l'effectif total.

  L'élément technologique θ_{14} est la définition de la moyenne.
- T_{12} : déterminer la médiane d'une série ; la technique associée est $\tau_{12;1}$:
  - ranger les valeurs de la série dans l'ordre croissant, les valeurs y étant notées autant de fois que leur effectif ;
  - déterminer la valeur de la série telle qu'il y ait autant de valeurs qui soient inférieures à cette valeur que de valeurs supérieures.

  L'élément technologique θ_{12} est la définition de la médiane.
- T_{15} : modifier une valeur de la série pour observer le comportement de la moyenne et de la médiane ; la technique associée τ_{15} :
  - modifier une valeur en l'exagérant ;
  - observer que la moyenne a été sensiblement modifiée ;
  - observer que la médiane n'a pas été modifiée.

  L'élément technologique θ_{15} est la propriété de la moyenne et de la médiane.

S'ensuit une liste de 5 exercices dont nous listons les types de tâches avec les techniques présentes :

- on retrouve le type de tâches T_{11} ; la technique présente est τ_{11} et les éléments technologiques θ_{11}.
- T_{16} : déterminer la fréquence inférieure à une valeur de la série ; la technique présente est

τ_{16} :

- repérer la valeur dans le tableau ;
- lire la fréquence cumulée correspondante.

L'élément technologique θ_{16} : fréquences cumulées croissantes.

– T_{17} : déterminer la fréquence supérieure ou égale à une valeur de la série ; la technique présente est $\tau_{17;1}$:
  - repérer la valeur dans le tableau ;
  - lire la fréquence cumulée correspondante ;
  - soustraire à 100%, la valeur trouvée précédemment.

  L'élément technologique θ_{17} : fréquence cumulée croissante.
– T_{18} : regrouper les valeurs d'une série en classe d'amplitude 5 ; la technique présente τ_{18} :
  - regrouper les notes par classe de 5 points d'amplitude ;
  - indiquer les effectifs correspondants.

  Élément technologique θ_{18} : caractère quantitatif continu.
– on retrouve le type de tâches T_7 avec la technique τ_7 et l'élément technologique θ_7.
– on retrouve le type de tâches T_2 dont la technique n'est pas explicitée mais l'élément technologique est θ_2.
– T_{12} : calculer la médiane d'une série ; la technique présente est $\tau_{12;2}$:
  - classer les données dans l'ordre croissant ;
  - déterminer l'effectif total $N = 24$;
  - la médiane correspond à la valeur de rang $\frac{N}{2} = 12$;
  - déterminer la classe qui correspond à la 12e valeur.

  L'élément technologique θ_{23} est la définition de la médiane.
– on retrouve les type de tâches T_{15} avec la technique $\tau_{15;2}$
  - Ajouter toutes les valeurs de la série ;
  - diviser par l'effectif total.

  L'élément technologique est θ_{15}.
– On retrouve le type de tâches T_{15} avec la technique $\tau_{15;1}$ et on a le même élément technologique θ_{15}.
– On retrouve le type de tâches T_2 ; la technique n'est pas précisée.
– On retrouve le type de tâches T_7 avec la même technique et les mêmes éléments technologiques.
– On retrouve le type de tâches T_{17} mais ici, deux techniques sont présentes :
  - $\tau_{17;2}$: faire la somme des fréquences des classes dont les valeurs sont supérieures ou égales à la valeur recherchée ;
  - $\tau_{17;3}$: faire la somme des effectifs des classes dont les valeurs sont supérieures ou égales à la valeur recherchée ; diviser par l'effectif total.

  L'élément technologique θ_{17} est la fréquence ainsi que l'effectif.
– On retrouve le type de tâches T_{15} avec la technique suivante $\tau_{15;3}$
  - pour chaque classe, déterminer le centre de la classe (demi-somme des extrêmes) ;
  - faire le produit du centre de classe par l'effectif ;
  - en faire la somme ;
  - diviser par l'effectif total.

  L'élément technologique est θ_{15}.
– T_{18} : déterminer les effectifs cumulés croissants ; la technique n'est pas présente mais l'élément technologique θ_{18} est la définition des effectifs cumulés croissants.

– On retrouve T_{12} avec la même technique et les mêmes éléments technologiques.

3.1.2 L'organisation didactique

Intéressons-nous maintenant à l'organisation de l'étude telle que ce professeur l'a proposé à ses élèves. La première séquence qui s'intitule *Données et représentations graphiques* est composée de trois parties qui sont respectivement : vocabulaire, représentations et données statistiques, et exercices. La forme générale du cours est constituée de « texte à trous » que les élèves doivent compléter sous la direction de leur professeur.

La première partie commence par présenter un moment d'institutionnalisation du vocabulaire, c'est-à-dire des éléments technologiques.

La seconde partie traite des différentes représentations graphiques ; on y trouve les moments d'institutionnalisation des éléments technologiques ainsi que quelques moments exploratoire des techniques.

La partie exercices est constituée de moments de travail de la technique mais aussi un moment d'évaluation de l'organisation mathématique mise en place par le professeur. Pour un certain nombre de types de tâches, le moment de (re)rencontre avec le type de tâches et exploratoire de la technique a lieu pendant les exercices.

La seconde partie, qui s'intitule *Calculs statistiques et probabilités* comporte quatre parties qui sont : calculs statistiques, exercices, notion de probabilités et exercices. Nous ne nous intéresserons qu'au deux premières parties.

Une activité d'introduction est présente avec un moment de première rencontre avec le type de tâches « déterminer la médiane » ; il y a une ébauche du travail de la technique mais le moment exploratoire, tout comme le moment d'institutionnalisation qui suit l'activité, est assuré par le professeur.

La seconde partie présente les exercices et les moments de travail de la technique et d'évaluation de l'organisation mathématique. Comme dans la séquence précédente, on trouve des type de tâches dont la première rencontre et l'exploration de la technique ont lieu dans cette partie.

3.2 Le cours de KB

Ce professeur enseigne dans une classe de 3^e^ d'une vingtaine d'élèves dans un collège marseillais. Nous n'avons pas eu accès à l'intégralité du cours, il nous manque l'activité d'introduction qui concernait la comparaison de séries ainsi que la partie des exercices. Cette première séquence de l'année s'intitule *Statistique*, on la trouvera en annexe 10.

3.2.1 L'organisation mathématique

Ce cours commence par établir la définition de la statistique et est suivie par la description de la démarche statistique :

– choisir une information à demander ;
– relever les réponses ;
– les écrire dans un tableau ;
– compter les effectifs ;
– calculer les fréquences de chaque valeur.

Ensuite un premier type de tâches est abordé T_1 : calculer une fréquence sous forme de pourcentage et présente deux techniques :

– $\tau_{1;1}$:
  - diviser l'effectif de la valeur par l'effectif total ;
  - écrire le résultat décimal sous forme de fraction avec au dénominateur 100 ;
  - écrire le résultat sous forme de pourcentage.
– $\tau_{1;2}$:
  - diviser l'effectif par l'effectif total ;
  - multiplier le numérateur par 100 ;
  - écrire le résultat sous forme de pourcentage.

L'élément technologique θ_1 est la définition de la fréquence : $f_i = \frac{n_i}{N}$.

La première partie de ce cours s'intitule *Moyenne* et présente le type de tâches T_2 : calculer la moyenne d'une série dont la technique τ_2 est :
- multiplier chaque valeur par son effectif ;
- additionner les produits ;
- diviser par l'effectif total.

L'élément technologique θ_2 est la définition de la moyenne.

La partie suivante s'intitule *Médiane* et présente le type de tâches T_3 : calculer la médiane d'une série ; la technique est seulement suggérée en indiquant qu'il faut classer les valeurs de la série dans l'ordre croissant. Seule la technologie θ_3 est donnée. S'ensuit un exemple de détermination de la médiane et de la moyenne. Nous avons donc la technique suivante τ_3 :
- ranger les valeurs de la série dans l'ordre croissant ;
- si l'effectif total N est impair : $N = 2k + 1$
  ◦ la médiane est la valeur de rang $k + 1$.
- si l'effectif total est pair : $N = 2k$
  ◦ la médiane est la demi-somme des valeurs de rang k et $k + 1$.

L'élément technologique θ_3 est la définition de la médiane donnée plus haut.

La troisième partie s'intitule *Étendue* et présente le type de tâches T_4 : calculer l'étendue d'une série dont la technique associée est τ_4 :
- déterminer le maximum ;
- soustraire le minimum au maximum.

L'élément technologique θ_4 est la définition de l'étendue.

La quatrième partie s'intitule *Quartiles* et présente le type de tâches T_5 : déterminer les quartiles d'une série ; on trouve la technique τ_5 :
- ranger la série dans l'ordre croissant ;
- calculer le quart de l'effectif total ;
- si le résultat n'est pas un entier, on prend la valeur suivante ;
- pour le 3^e^ quartile faire de même avec les trois quarts de l'effectif total.

L'élément technologique θ_5 est la définition des quartiles.

A la fin est présente une partie intitulée *Résumé statistique* qui présente les différents paramètres étudiés ci-dessus sous la forme d'un graphique qui ressemble au diagramme en boîte.

Un dernier type de tâches T_6 est abordé dans une feuille dactylographiée : déterminer les paramètres statistiques à l'aide du tableur ; la technique τ_6 est donnée sous forme d'un exercice corrigé :
- rentrer les valeurs dans la feuille de calcul ;
- rentrer les formules qui calculent les paramètres ;
- en déduire les résultats.

Les éléments technologiques θ_6 sont les différentes formules du tableur (MEDIANE() ; QUARTILE(), etc.) ainsi que les définitions de la médiane, de l'étendue, de la moyenne et des quartiles.

3.2.2 L'organisation didactique

L'organisation didactique de ce cours est forcément biaisé puisque nous n'avons pas l'intégralité du cours (il manque l'activité d'introduction ainsi que la partie exercices). La partie du cours à laquelle nous avons eu accès présente les moments d'institutionnalisation des éléments technologiques et des techniques liées aux types de tâches que nous venons de lister. Nous en resterons là en ce qui concerne ce cours.

3.3 Le cours de PN

Ce professeur enseigne dans une classe de 2^{de} d'un lycée polyvalent. Le cours de statistique est la deuxième séquence de ce professeur qui s'intitule *Statistique* ; on en trouvera une copie en annexe 13.

3.3.1 L'organisation mathématique

Le cours commence par une première activité au travers d'une question Q_1 : « Quelle est l'activité préférée des élèves de 2^{de} 10 le week-end ? ». On y trouve les types de tâches suivants :

- T_1 : définir la population d'une série statistique ; la technique associée τ_1 est :
  - relever dans l'énoncé sur quels individus porte l'étude.

  L'élément technologique θ_1 est la définition de la population.
- T_2 : déterminer le caractère étudié ; la technique associée τ_2 est :
  - relever dans l'énoncé sur quoi sont « interrogés » les individus de la population.

  L'élément technologique θ_2 est la définition du caractère.
- T_3 : déterminer la nature du caractère étudié ; la technique associée τ_3 est :
  - le caractère ne s'exprime pas à l'aide de nombres, donc la nature du caractère est qualitatif.

  L'élément technologique θ_3 est la définition du caractère qualitatif.
- T_4 : déterminer les valeurs prises par le caractère ; la technique associée τ_4 est :
  - repérer dans les données brutes, toutes les valeurs prises par le caractère étudié.

  L'élément technologique θ_4 est la définition de valeur.

S'ensuit une nouvelle question Q_2 : « Est-ce que les filles ont les mêmes goûts que les garçons ? » qui permet d'aborder le type de tâches T_5 : établir la distribution des fréquences dont la technique associée est τ_5 :

- relever les valeurs du caractère ;
- pour chaque valeur associer l'effectif correspondant ;
- déterminer l'effectif total de la série ;
- pour chaque valeur, déterminer la fréquence correspondante en divisant l'effectif considéré par l'effectif total.

L'élément technologique θ_5 est la définition de la fréquence $f_i = \frac{n_i}{N}$.

Le type de tâches suivant T_6 est : représenter une série à l'aide d'un diagramme circulaire ; la technique associée τ_6 est :

- pour chaque fréquence, multiplier cette fréquence par 360° ;
- diviser le résultat par 100% ;

- arrondir le résultat à l'entier le plus proche ;
- tracer un cercle ;
- pour chaque valeur construire le secteur d'angle calculé aux étapes précédentes.

Les éléments technologiques θ_6 sont les angles, les fréquences et le diagramme circulaire.

Une seconde activité suit avec comme question centrale Q_3 : « Combien de personnes vivent dans les foyers des élèves de 2^{de} ? » ; on y retrouve les types de tâches T_1 à T_3. Vient ensuite le type de tâches T_7 : déterminer les paramètres statistiques d'une série ; la technique associée τ_7 est :

- Étendue :
  - Prendre la valeur minimale de la série ;
  - la soustraire à la valeur maximale de la série.
- Mode :
  - déterminer la ou les valeurs de la série qui ont le plus grand effectif.
- Moyenne :
  - faire la somme des valeurs de la série ;
  - diviser par l'effectif total de la série.
- 1^{er} quartile :
  - prendre l'effectif total ;
  - le diviser par 4 ;
  - si le résultat n'est pas entier, prendre le nombre entier immédiatement supérieur ;
  - lire la valeur qui a pour rang le nombre entier précédent.
- 3^e quartile :
  - prendre l'effectif total ;
  - le multiplier par 3 et le diviser par 4 ;
  - si le résultat n'est pas entier, prendre le nombre entier immédiatement supérieur ;
  - lire la valeur qui a pour rang le nombre entier précédent.
- Médiane :
  - classe la série dans l'ordre croissant ;
  - diviser l'effectif total par 2 ;
  - si l'effectif total est pair : $N = 2k$
    - prendre les valeurs de rang k et $k + 1$ et en faire la demi-somme.
  - si l'effectif total est impair : $N = 2k + 1$
    - lire la valeur de rang $k + 1$.

Les éléments technologiques θ_7 convoqués sont la définition de l'étendue, du mode, de la moyenne, de la médiane et des quartiles.

A la suite, on trouve le type de tâches T_8 : calculer les effectifs cumulés croissants ; la technique associée τ_8 est :

- construire une ligne supplémentaire pour calculer les effectifs cumulés croissants ;
- recopier l'effectif de la première valeur ;
- dans la case suivante, ajouter au premier effectif, l'effectif de la 2^{de} valeur ;
- dans la dernière case, vérifier que l'on a bien l'effectif total.

L'élément technologique θ_8 est la définition des effectifs cumulés croissants.

On trouve ensuite trois types de tâches $T_{9;1}$ à $T_{9;3}$ qui ont tous trait aux effectifs cumulés :

- $T_{9;1}$: déterminer l'effectif cumulé qui vaut au plus une valeur définie ; la technique associée est $\tau_{9;1}$:
  - chercher la valeur définie dans le tableau ;

- lire l'effectif cumulé croissant correspondant.

– $T_{9;2}$: déterminer l'effectif cumulé qui vaut moins qu'une valeur définie ; la technique associée est $\tau_{9;2}$:
  - chercher la valeur définie dans le tableau et se placer sur la valeur immédiatement inférieure ;
  - lire l'effectif cumulé croissant correspondant.

– $T_{9;3}$: déterminer l'effectif cumulé qui vaut plus qu'une valeur définie ; la technique associée est $\tau_{9;3}$:
  - chercher la valeur définie dans le tableau ;
  - lire l'effectif cumulé décroissant correspondant.

  L'élément technologique θ_9 est θ_8.

On trouve ensuite le type de tâches T_{10} : déterminer les quartiles d'une série ; la technique présente est τ_{10} :
- prendre l'effectif total, le diviser par 4 ;
- si le résultat n'est pas entier, prendre le nombre entier immédiatement supérieur ;
- dans les effectifs cumulés croissants, lire la valeur dont le rang correspond au nombre calculé au-dessus, ce qui donnera le premier quartile ;
- faire de même pour l'effectif total multiplié par 3 et divisé par 4, pour obtenir le troisième quartile.

L'élément technologique θ_{10} est la définition des quartiles.

Dernière activité qui traite à son tour des type de tâches T_1 à T_3 ainsi que le type de tâches T_7. Un nouveau type de tâches T_{11} est abordé : représenter une série par un histogramme ; la technique associée τ_{11} est :
- choisir une unité d'aire ;
- tracer un axe gradué ;
- pour chaque classe représenter l'effectif à l'aide de l'unité d'aire choisie par un rectangle.

L'élément technologique θ_{11} est l'histogramme.

On retrouve alors le type de tâches T_{10} mais avec une technique différente $\tau_{10}^{'}$:
- calculer les fréquences cumulées croissantes ;
- représenter le polygone des fréquences cumulées croissantes :
  - ○ placer le point de coordonnées $(150; 0)$;
  - ○ pour chaque classe, prendre l'extrémité supérieure comme abscisse et l'effectif comme ordonnée ;
  - ○ relier les points à la règle ;
  - ○ se placer sur la fréquence cumulée croissante 50% et lire l'antécédent correspondant pour la médiane ;
  - ○ se placer sur la fréquence cumulée croissante 25% et lire l'antécédent correspondant pour le 1^er quartile ;
  - ○ se placer sur la fréquence cumulée croissante 75% et lire l'antécédent correspondant pour le 3^e quartile.

L'élément technologique est θ_{10} que nous avons vu plus haut.

On trouve alors un document qui donne les éléments technologiques accompagnés d'exemples et parfois de la technique qui lui est associée.

Trois fiches intitulées « Question du jour » viennent compléter ce cours. Ces fiches sont destinées aux élèves en début de cours pour se remémorer le travail fait dans les séances précédentes.Cela permet donc de reprendre les types de tâches vus dans les activités que nous venons de décrire. On trouve un type de tâches qui n'apparaît pas en tant que tel dans les activités : le type de tâches T_{12} : construire le tableau des fréquences cumulées croissantes et décroissantes ; on trouve deux techniques associées :

- $\tau_{12;1}$:
  - recopier la première fréquence de la première valeur ;
  - dans la case suivante prendre la fréquence précédente et lui ajouter la fréquence de la valeur considérée.
- $\tau_{12;2}$:
  - mettre 100% dans la première case ;
  - soustraire la fréquence de la case correspondante et copier la valeur trouvée dans la case suivante.

L'élément technologique θ_{12} est la définition des fréquences cumulées.

Un autre type de tâches $T_{13;1}$: déterminer le pourcentage des individus qui ont une valeur supérieure à une valeur donnée et $T_{13;2}$: déterminer le pourcentage des individus qui ont une valeur inférieure à une valeur donnée dont les techniques sont les suivantes :

- $\tau_{14;1}$:
  - chercher dans le tableau la valeur considérée ;
  - regarder la fréquence cumulée décroissante correspondante.
- $\tau_{14;2}$:
  - chercher dans le tableau la valeur considérée ;
  - regarder la fréquence cumulée croissante correspondante.

L'élément technologique θ_{14} est la définition des fréquences cumulées.

S'ensuit une partie exercices [2] qui reprend la plupart des types de tâches abordés dans les activités et les fiches « Question du jour » (QDJ). On trouve un type de tâches T_{15} : calculer la moyenne d'une série avec la technique τ_{15} :

- calculer le centre de chaque classe (demi-somme des extrémités de l'intervalle) ;
- multiplier chaque centre de classe par la fréquence correspondante ;
- en faire la somme.

L'élément technologique est θ_{15} la définition de la moyenne.

Un autre type de tâches qui n'est pas étudié auparavant T_{16} : construire le tableau de regroupement en classe avec la technique τ_{16} :

- écrire les classes avec pour amplitude 5 cm ;
- pour chaque classe dénombrer les tailles.

L'élément technologique θ_{16} est la définition des classes pour un caractère quantitatif continu.

Le dernier élément qui aborde deux types de tâches qui n'ont pas encore été vus plus haut : un module [3] qui s'intitule *Comparer deux séries statistiques.* On y trouve les types de tâches suivants :

- T_{17} : construire le diagramme en boîte d'une série statistique ; la technique associée τ_{17} est :
  - tracer un axe gradué horizontal ;
  - placer le minimum et le maximum ;
  - placer le 1[er] quartile, la médiane et le 3[e] quartile ; fermer la boîte ;
  - relier le 1[er] quartile au minimum et le 3[e] quartile au maximum.

2. Le manuel utilisé par la classe est *Travailler en confiance. Maths 2[de]* (Antibi, A. et al., 2010)
3. Ce terme désigne un dispositif qui avait été mis en place lors de la réforme de la classe de 2[de] de 2000 et qui a disparu lors de la réforme de 2010 au profit de l'accompagnement personnalisé

Les éléments technologiques θ_{17} sont la définition des médianes, quartiles et du diagramme en boîte.

– T_{18} : comparer deux séries ; la technique associée τ_{18} est :
  - comparer les médianes ;
  - comparer l'écart interquartile ou l'intervalle interquartile.

  Les éléments technologiques θ_{18} sont la définition de la médiane, des quartiles, de l'écart interquartile, de l'intervalle interquartile et du diagramme en boîte.

3.3.2 L'organisation didactique

La première activité par laquelle débute cette leçon présente un moment de première rencontre (nous devrions dire re-rencontre) avec des types de tâches vus dans les classes antérieures : T_1 à T_3. Une question est posée Q_1 : « Quelle est l'activité préférée des élèves d'une classe de 2^{de} ? » ; suite à un premier traitement statistique, une seconde question est posée Q_2 : « Est-ce que les filles ont les mêmes goûts que les garçons ? » ce qui permet de (re)rencontrer les types de tâches T_5 et T_6. Enfin une dernière question vient clore cette activité Q_3 : « Les jeux vidéos sont-ils davantage préférées par les filles ou par les garçons ? » ; le dernier traitement statistique permet alors de répondre complètement à la question en regardant les fréquences.

La seconde activité aborde les mêmes types de tâches au travers des questions Q'_1 : « Combien de personnes vivent dans les foyers des élèves de 2^{de} ? » ; on y (re)rencontre les types de tâches T_1 à T_3 ainsi que T_7. S'ensuit un exercice, moment de travail de la technique du type de tâches T_8 à propos de la médiane (avec la technique τ_{10}). On revient à l'activité avec la question Q'_2 : « Combien d'élèves ont au maximum 4 habitants dans leur foyer ? », la question Q'_3 : « Combien d'élèves moins de 4 habitants dans leur foyer ? », la question Q_4 : « Combien d'élèves ont plus de 4 habitants dans leur foyer ? », la question Q'_4 : « Combien d'élèves ont plus de 3 habitants ? » et Q'_5 : « Combien d'élèves ont au minimum 2 habitants ? » ; ces 5 questions permettent de (re)rencontrer les types de tâches T_{11} puis T_{10}. S'ensuit un exercice qui présente un moment de travail de la technique sur le type de tâches T_{10}.

La dernière activité présente les mêmes types de tâches mais au travers de la question Q''_1 : « Quelle est la taille des élèves d'une classe de 2^{de} ? » ; on y trouve les types de tâches T_7, T_{11} et T_{10} avec la technique τ'_{10} dans des moments de re-rencontre ainsi que des moments exploratoires des techniques. On trouve aussi des moments de constitution du bloc technologico-théorique.

Les « questions du jours » présentent des moments de travail de la technique mais en même temps des moments d'évaluation de l'organisation mathématique produite par le professeur. Les exercices qui suivent sont aussi des moments de travail de la technique.

La fiche « Vocabulaire de la leçon 2 » présente les éléments technologiques illustré d'exemples.

Enfin la partie exercices présente des moments de travail de la technique.

3.4 Le cours de YC

Ce professeur enseigne dans une classe de 2^{de} de lycée technique qui comprend 37 élèves. Le cours de statistique est la troisième séquence de l'année ; on en trouvera une copie en annexe 18.

3.4.1 L'organisation mathématique

Le premier type de tâches T_1 est calculer une fréquence dont la technique associée τ_1 est :

- diviser l'effectif n par l'effectif total N ;

- écrire la fréquence sous forme décimale, fractionnaire ou de pourcentage.

L'élément technologique θ_1 est la définition de la fréquence $f_i = \frac{n_i}{N}$.

On trouve alors deux éléments technologiques que sont la définition d'un échantillon (θ_2) et la définition d'une expérience aléatoire (θ_3). Dans la suite, est présent l'élément technologique θ_4, à savoir la définition de la fluctuation d'échantillonnage. Cet élément est suivi d'un exemple qui présente le type de tâches T_4 : déterminer la distribution des fréquences. La technique associée τ_4 est :

- pour chaque valeur diviser l'effectif par l'effectif total ;
- vérifier que la somme fait bien 1 (ou 100%).

L'élément technologique θ_4 est θ_1.

On trouve ensuite le type de tâches T_5 : déterminer l'intervalle de fluctuation avec la technique τ_5 :

- vérifier les conditions du théorème :
  - si $n > 25$;
  - si $p \in [0,2; 0,8]$.
- calculer l'intervalle $\left[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}}\right]$;
- conclure en disant que dans plus de 95% des cas, la fréquence d'apparition f appartient à l'intervalle calculé précédemment.

L'élément technologique θ_5 est la définition de l'intervalle de fluctuation.

On rencontre alors le type de tâches T_6 : construire le polygone des effectifs cumulés croissants avec la technique τ_6 :

- placer sur les abscisses les valeurs du caractère ;
- placer sur les ordonnées les effectifs cumulés croissants ;
- placer les points ayant pour abscisse l'extrémité b_i de chaque classe et pour ordonnée l'effectif cumulé correspondant ;
- relier les points par des segments.

L'élément technologique θ_6 est la définition des effectifs cumulés croissants et le polygone des effectifs cumulés.

Le type de tâches suivant est T_7 : construire un diagramme en barres dont la technique τ_7 n'est pas présente. L'élément technologique θ_7 est la définition du diagramme en barres.

Vient ensuite le type de tâches T_8 : construire un diagramme circulaire dont la technique τ_8 n'est pas non plus présente. L'élément technologique θ_8 est la définition du diagramme circulaire.

Le type de tâches T_9 est : construire un diagramme en bâtons, dont la technique τ_9 n'est pas présente ; l'élément technologique θ_9 est la définition du diagramme en bâtons.

Enfin, le type de tâches T_{10} concerne la construction d'un histogramme ; là encore, la technique τ_{10} n'est pas présente et l'élément technologique θ_{10} est la définition de l'histogramme.

La partie suivante présente le type de tâches T_{11} : calculer l'étendue d'une série ; la technique qui lui est associée τ_{11} est :

- déterminer la plus grande valeur de la série ;
- déterminer la plus petite valeur de la série ;
- soustraire la plus petite valeur à la plus grande.

L'élément technologique θ_{11} est la définition de l'étendue.

Le type de tâches suivant est T_{12} : calculer le mode d'une série dont la technique τ_{12} associée est :

- déterminer la (ou les) valeur(s) de la série qui a (ont) le plus grand effectif ;
- cette (ou ces) valeur(s) est (sont) le mode de la série.

L'élément technologique θ_{12} est la définition du mode.

Le type de tâches abordé est T_{13} : calculer la moyenne d'une série dont la technique associée τ_{13} est :

- multiplier les valeurs de la série par leur effectif ;
- faire la somme des produits ;
- diviser par l'effectif total.

L'élément technologique θ_{13} est la définition de la moyenne.

Le type de tâches T_{14} est : calculer la médiane d'une série ; la technique τ_{14} associée est :

- déterminer l'effectif total N ;
- si N est impair, la médiane est la valeur de la série qui a pour rang $\frac{N+1}{2}$;
- si N est pair, la médiane est la moyenne des valeurs de rang $\frac{N}{2}$ et $\frac{N}{2}+1$.

L'élément technologique θ_{14} est la définition de la médiane.

Le dernier type de tâches de la partie cours est T_{15} : calculer les quartiles d'une série ; la technique τ_{15} associée est :

- déterminer l'effectif total N ;
- pour le calcul de Q_1 :
  - ○ calculer $\frac{N}{4}$;
  - ○ si le résultat est entier, Q_1 est la valeur de la série de rang $\frac{N}{4}$;
  - ○ sinon, prendre la valeur entière immédiatement supérieure ;
- pour le calcul de Q_3 :
  - ○ calculer $\frac{3N}{4}$;
  - ○ si le résultat est entier, Q_3 est la valeur de la série de rang $\frac{3N}{4}$;
  - ○ sinon, prendre la valeur immédiatement supérieure.

L'élément technologique θ_{15} est la définition des quartiles.

Dans la partie exercices, 20 exercices sont proposés par le professeur. Seuls les exercices 5 à 17 sont traités. On retrouve le type de tâches T_4, avec la technique τ_4 explicitée :

- construire un tableau ayant pour première ligne les valeurs du caractère étudié ;
- pour seconde ligne les effectifs correspondants à chaque valeur ;
- pour troisième ligne la fréquence :
  - ○ calculer l'effectif total ;
  - ○ diviser chaque effectif de chaque valeur par l'effectif total.

L'élément technologique est θ_4.

On trouve un peu plus loin les techniques des type de tâches T_8 à T_{10} :

– τ_8 :
  - pour chaque valeur, multiplier la fréquence par 360° et diviser par 100% ;
  - tracer un cercle ;
  - tracer un rayon de ce cercle ;
  - placer le premier angle calculé puis les autres à la suite du nouveau rayon tracé.

– τ_9 :
  - placer en abscisses les valeurs du caractère étudié ;
  - placer en ordonnées les effectifs ;
  - pour chaque valeur, tracer un segment ayant pour hauteur l'effectif correspondant.

– $\tau_{10;1}$:
  - placer en abscisses les valeurs du caractère étudié ;
  - placer en ordonnées les effectifs ;

- pour chaque classe, tracer un rectangle de largeur l'amplitude de la classe (constante ici) et de hauteur l'effectif correspondant.

– $\tau_{10;2}$:
  - placer en abscisses les valeurs du caractère étudié ;
  - placer en ordonnées les effectifs ;
  - pour chaque classe, tracer un rectangle de largeur l'amplitude de la classe (constante ici) et de hauteur h_i telle que $h_i = \frac{n_i}{b_i - a_i}$. L'aire du rectangle étant proportionnelle à l'effectif correspondant.

On trouve cependant des types de tâches qui ne sont pas présents. En premier lieu, le type de tâches T_{16} est : déterminer l'effectif telle que les valeurs de la série soient inférieures ou égales à une valeur déterminée ; la technique τ_{16} associée est :

- repérer la valeur déterminée dans le tableau ;
- faire la somme des effectifs dont les valeurs sont inférieures ou égales à cette valeur.

L'élément technologique θ_{16} est la définition des effectifs cumulés croissants.

Vient ensuite le type de tâches T_{17} : déterminer la fréquence telle que les valeurs de la série soient inférieures ou égales à une valeur déterminée ; la technique τ_{17} associée est :

- repérer la valeur déterminée dans le tableau ;
- faire la somme des effectifs dont les valeurs sont inférieures ou égales à cette valeur ;
- diviser par l'effectif total.

L'élément technologique θ_{17} est θ_{16}.

On poursuit avec le type de tâches T_{18} : déterminer les effectifs cumulés croissants et les fréquences cumulées croissantes ; la technique τ_{18} associée est :

- recopier le premier effectif de la première valeur ;
- ajouter l'effectif suivant ;
- vérifier que dans la dernière case se trouve l'effectif total ;
- faire de même pour les fréquences.

L'élément technologique θ_{18} est la définition des effectifs cumulés croissants et des fréquences cumulées croissantes.

Le type de tâches suivant T_{19} est : déterminer l'effectif (ou la fréquence) telle que les valeurs de la série soient comprises entre deux valeurs déterminées ; la technique τ_{19} associée est :

- repérer les valeurs déterminées dans le tableau ;
- lire les effectifs cumulés croissants (ou les fréquences cumulées croissantes) correspondantes des valeurs déterminées ;
- faire la soustraction.

L'élément technologique θ_{19} est θ_{18}.

Enfin le dernier type de tâches T_{20} est : comparer deux séries dont la technique τ_{20} n'est pas présente ; seule une réponse est proposée. L'élément technologique θ_{20} est la comparaison de nombres.

3.4.2 L'organisation didactique

La séquence commence par une question Q posée par la classe dans le souhait d'une étude statistique « au long cours » : « Qu'est-ce qu'une grosse pomme de terre ? » : question que l'on retrouvera tout au long du cours.

On trouve un moment d'institutionnalisation du bloc technologico-théorique sur la fréquence ; suivi d'un exemple lié à notre question Q ; l'enquête n'a pas eu lieu ; le professeur a « inventé » les données. Ceci ouvre la voie vers une question Q' : « Sur les données précédentes peut-on dire que c'est encore vrai pour 100 autres pommes de terre ? » ; ce qui permet d'introduire la notion d'échantillon.

S'ensuit alors un moment de première rencontre avec la notion d'échantillon avec la question Q'. On y trouve un moment de constitution du bloc technologico-théorique concernant la notion d'échantillon et d'expérience aléatoire.

On trouve un moment d'institutionnalisation du bloc technologico-théorique avec un exercice ; chaque élève doit simuler le lancer d'une pièce 10 fois. Une question nouvelle apparaît en lien avec Q, la question Q'' : « Peut-on estimer la fréquence de pommes de terre qui mesurent entre 5 et 6 cm dans un échantillon de taille 100 ? ». La question Q'' permet d'énoncer le théorème de l'intervalle de fluctuation. Un exemple suit : moment exploratoire de la technique. Constitution du bloc technologico-théorique.

On trouve un moment d'institutionnalisation du bloc technologico-théorique illustré par un exemple en lien avec la question Q. On trouve alors un moment de (re)rencontre avec le type de tâches T_{17} ; moment de travail de la technique (ce type de tâches étant à la charge des programmes antérieurs ?).

Lors de l'étude des types de tâches concernant la représentation graphique des séries, on trouve un moment d'institutionnalisation du bloc technologico-théorique (programmes antérieurs) ; illustré par des exemples : moment du travail de la technique. S'ensuit un moment d'institutionnalisation du bloc technologico-théorique pour l'étendue et le mode d'une série ; suivi d'exemples : moments de travail de la technique. L'exemple est en rapport avec la question Q.

La section qui s'intéresse à la moyenne comporte une activité qui présente donc un moment de première (re)rencontre (programme de 4^{e}) en lien avec la question Q. C'est aussi un moment exploratoire de la technique, puis de constitution du bloc technologico-théorique.

Pour l'étude de la médiane, on trouve un moment d'institutionnalisation du bloc technologico-théorique ainsi que de la technique τ_{14}. Il est suivi d'un exemple qui présente un moment de travail de la technique.

Enfin pour l'étude des quartiles, on trouve l'institutionnalisation du bloc technologico-théorique et de la technique. L'exemple qui suit présente un moment de travail de la technique.

La partie des exercices présente des moments de travail de la technique.

Chapitre 4

L'observation du savoir appris

Suite à l'analyse des cours des professeurs, nous nous sommes intéressés aux copies des élèves de ces professeurs. Nous avons dans un premier temps analysé les praxéologies du sujet (essentiellement l'organisation mathématique) puis repérer les principales erreurs ou confusions dans les copies. Plus précisément, nous avons observé les techniques utilisées par les élèves dans leurs copies et les avons comparé à celles du cours.

4.1 Les copies des élèves de LC

4.1.1 Le sujet de la première évaluation

Les trois premiers exercices sont des questions de cours, nous n'en parlerons pas plus ici. Intéressons-nous au deux derniers exercices. L'exercice 4 s'intéresse à la taille des joueurs de basket. On trouve ainsi les types de tâches suivants :

- T_1 : déterminer un effectif connaissant les autres et l'effectif total dont le spécimen est la tâche t_1 : déterminer l'effectif de la classe $[200; 205[$; la technique τ_1 associée est :
  - faire la somme des effectifs connus ;
  - soustraire cette somme à l'effectif total ;
  - en déduire l'effectif manquant.

  L'élément technologique θ_1 est la définition de l'effectif total $N = \sum_{i=1}^{p} n_i$.
- T_2 : Tracer l'histogramme (à pas constant) d'une série dont le spécimen est la tâche t_2 : construire l'histogramme de la série répartition des tailles de joueurs de basket ; la technique τ_2 associée est :
  - construire le repère du plan en respectant les unités des axes imposées par l'énoncé ;
  - pour chaque classe, tracer un rectangle dont la hauteur est indiquée par l'effectif et la largeur par l'amplitude de la classe.

  L'élément technologique θ_2 est la définition de l'histogramme.

Le dernier exercice quant à lui s'intéresse au nombre de films regardés dans les salles de cinéma d'un gérant pendant le mois écoulé. On trouve les types de tâches suivants :

- $T_3 = T_1$; la technique et les éléments technologiques sont identiques ; la tâche proposée est t_3 : déterminer l'effectif de la valeur 6.
- T_4 : calculer les effectifs cumulés croissants d'une série dont la tâche t_4 : calculer les effectifs cumulés croissants du nombre de films regardés dans la salle du gérant durant le mois écoulé ; la technique τ_4 associée est :
  - recopier le premier effectif ;

- à cette valeur, ajouter l'effectif suivant ;
- recommencer jusqu'à arriver à la dernière case du tableau ;
- vérifier que l'on obtient l'effectif total.

L'élément technologique θ_4 est la définition des effectifs cumulés croissants.

– T_5 : déterminer l'effectif qui correspond à une valeur donnée dont la tâche t_5 : déterminer le nombre de personnes qui ont regardé un seul film le mois dernier ; la technique τ_5 associée est :
- dans le tableau, repérer la valeur cherchée ;
- en déduire l'effectif correspondant.

L'élément technologique θ_5 est la définition du tableau d'effectif ?

– T_6 : calculer une fréquence dont la tâche t_6 : calculer le pourcentage de personnes qui n'ont regardé qu'un seul film le mois dernier est un spécimen ; la technique associée τ_6 est :
- prendre l'effectif qui correspond à la valeur 1 ;
- diviser cet effectif par l'effectif total ;
- donner le résultat sous forme de pourcentage (en multipliant par 100% le cas échéant).

L'élément technologique θ_6 est la définition de la fréquence.

– T_7 : déterminer l'effectif cumulé qui correspond à une valeur inférieure ou égale fixée dont la tâche t_7 : déterminer le nombre de personnes qui ont regardé moins de 4 films le mois dernier est un spécimen ; la technique τ_7 associée est :
- repérer la valeur 4 dans le tableau ;
- regarder la valeur immédiatement inférieure ;
- en déduire l'effectif cumulé correspondant.

L'élément technologique θ_7 est la définition de l'effectif cumulé croissant.

4.1.2 L'observation des copies

Après avoir vu les praxéologies présentes dans ce sujet d'évaluation, observons les réponses des élèves, dont nous trouvons les copies en annexe 1. Les analyses de chacune des copies se trouvent en annexe 2. Nous en faisons une synthèse ci-après.

Dans la première question de l'exercice 4, les élèves ont majoritairement suivi la technique du cours τ_9^{LC} [1] ; seuls deux élèves n'ont pas suivi cette technique et ont répondu faux et un élève n'a pas répondu à la question. Quelles sont ces erreurs ? La première vient du fait que l'élève Anaïs a bien ajouté tous les effectifs du tableau mais en y incluant aussi l'effectif total ; nous reproduisons l'extrait [2] :

1. τ_i^{XY} désigne la technique du cours de LC qui correspond au type de tâches T_i
2. Nous reproduisons cette copie, mais en général, les erreurs présentées ont été trouvées dans plusieurs copies. Par ailleurs nous avons recopier les fautes d'orthographe sans les corriger

taille (cm)	effectif
[175 ; 180[	6
[180 ; 185[	6
[185 ; 190[	20
[190 ; 195[	26
[195 ; 200[	12
[200 ; 205[	X
	80

$$6 + 6 + 20 + 26 + 12 + 80 = 150$$

(Annexe 1, p.7)

Le professeur note alors que 80, correspond à l'effectif total. Peut-être l'erreur aurait-elle pu être évitée si dans le tableau était noté la mention du total à côté de 80 ?

La seconde erreur vient d'une mauvaise soustraction : inversion de l'effectif total et de la somme des effectifs connus : elle rétablira tout de même dans le résultat de la soustraction et dans le tableau.

La seconde question s'intéressait à la construction de l'histogramme associé au tableau de la question précédente. La plupart du temps l'histogramme est bien construit mais on ne peut pas voir la technique utilisée (même si on peut la « deviner ») ; seul un élève ne trace pas d'axe gradué et place les classes comme des abscisses à part entière. Le professeur fait souvent remarqué l'absence de titre pour le graphique ou d'étiquettes sur les axes.

Dans la première question de l'exercice 5, les élèves devaient calculer l'effectif manquant dans le tableau et compléter la colonne des effectifs cumulés croissants. Cette deuxième partie a été, dans la quasi totalité des cas, bien réussie beaucoup mieux que la première partie. La technique τ_9^{LC} utilisé dans l'exercice 4 s'appliquait encore ici et le pourcentage d'erreurs reste stable (aux alentours de 35%). Les élèves ont-ils contournés la difficulté liée à la première partie de la question en utilisant une autre technique (que nous ne pouvons voir, les calculs n'étaient pas demandés) qui leur permettaient de répondre correctement à la deuxième partie de la question ? Ils auraient pu, en effet, calculer les effectifs cumulés croissants jusqu'à la valeur 5 (incluse) puisqu'ils avaient toutes les données, puis partir de l'effectif total (donné dans l'énoncé) et « remonter » le tableau en soustrayant les effectifs aux effectifs cumulés et finir par compléter entièrement la colonne. Observons la copie de Tiphanie :

Nombre de films regardés	Effectifs	Effectifs cumulés croissants
0	50	50
1	60	110
2	120	230
3	40	270
4	50	320
5	30	350
6	36	386
7	20	406
8	10	416
Total	416	

(Annexe 1, p. 12)

On le voit l'erreur de la première partie de la question influe sur la réponse à la deuxième partie de la question. Cependant l'élève en calculant la ligne total, essaye de justifier son calcul, en montrant que la dernière case correspond à son total. Seulement, une meilleure lecture de l'énoncé lui aurait permis de voir que le total n'était pas 416 mais 400.

La seconde question demandait une lecture du tableau et plus précisément de la colonne des effectifs, ce qui a été bien réussi dans l'ensemble. Trois élèves confondent la colonne des effectifs avec la colonne des effectifs cumulés croissants et un élève confond avec l'effectif total.

Pour la troisième question, la technique τ_7^{LC} est mal maîtrisée par une moitié de la classe. Le plus difficile semble le passage de la fréquence sous forme décimale à la fréquence sous forme de pourcentage. Il est intéressant de noter que la technique qui permet ce passage n'est pas présente dans le cours du professeur LC. La copie d'Élodie, est à ce sujet particulièrement intéressante :

> Ce résultat en pourcentage est 15%. J'ai fait :
>
> $$\frac{60}{400} = 0,15 = 15$$
>
> (Annexe 1, p. 33)

Le professeur ne relève pas l'erreur et se contente de valider le calcul.

Pour la quatrième question, les deux techniques τ_{13}^{LC} et τ_7^{S} [3] sont présentes dans les copies des élèves mais beaucoup ont refait le calcul des effectifs cumulés croissants ; ils n'ont donc pas vu l'intérêt de la première question.

4.1.3 Le sujet de la seconde évaluation

Dans cette seconde évaluation, nous nous intéresserons aux deuxième et troisième exercices, le premier et le quatrième ne concernant pas directement notre étude.

Le premier type de tâches T_1 est : calculer un effectif à partir de l'histogramme dont la tâche t_1 : calculer l'effectif d'élèves dont la masse corporelle est comprise entre 40 et 50 à partir de l'histogramme est un spécimen ; la technique τ_1 associée est :

- lire sur l'histogramme le nombre de cases qui correspond à l'intervalle $[40; 50[$;
- lire l'unité dans la légende ;
- multiplier le nombre de cases par l'unité ;
- reporter la valeur dans le tableau d'effectifs.

L'élément technologique θ_1 est la définition de la fréquence et de l'histogramme

Le second type de tâches T_2 est : calculer une fréquence dont la tâche t_2 : calculer la fréquence des élèves qui ont une masse corporelle comprise entre 50 et 60 kg est un spécimen ; la technique $\tau_{2;1}$ associée est :

- prendre l'effectif concerné ;
- diviser par l'effectif total ;
- multiplier par 100%.

Une autre technique disponible $\tau_{2;2}$ est :

- faire la somme des fréquences connues ;
- soustraire à 100%, le calcul précédent.

L'élément technologique θ_2 est la définition de la fréquence.

Le troisième type de tâches T_3 est : calculer une fréquence qui vérifie certaines caractéristiques, dont la tâche t_3 calculer la fréquence des élèves dont la masse corporelle est comprise entre 60 et 80 kg est un spécimen ; la technique $\tau_{3;1}$ associée est :

- prendre la fréquence de la classe $[60; 70[$;
- l'ajouter à la fréquence de la classe $[70; 80[$.

Une autre technique $\tau_{3;2}$ possible est :

3. τ_i^S représente la technique que nous avons relevé dans le *sujet* à propos du type de tâches T_i

- prendre l'effectif de la classe $[60; 70[$;
- l'ajouter à l'effectif de la classe $[70; 80[$;
- diviser par l'effectif total ;
- multiplier par 100% le résultat.

L'élément technologique est la définition de la fréquence.

Le quatrième type de tâches T_4 est : calculer un effectif regroupant plusieurs classes dont la tâche t_4 : calculer le nombre d'élèves dont la masse corporelle est supérieure à 60 kg est un spécimen ; la technique $\tau_{4;1}$ associée est :

- prendre l'effectif de la classe $[60; 70[$;
- l'ajouter à l'effectif de la classe $[70; 80[$.

Une autre technique $\tau_{4;2}$ possible est :

- prendre la fréquence des élèves dont la masse corporelle est comprise entre 60 et 80 kg ;
- la multiplier par l'effectif total ;
- en déduire l'effectif recherché.

L'élément technologique est la définition de la fréquence.

Le dernier type de tâches de cet exercice, T_5 est : calculer la moyenne d'une série dont la tâche t_5 calculer la masse corporelle moyenne est un spécimen ; la technique $\tau_{5;1}$ associée est :

- calculer le centre des classes ;
- multiplier le centre de classe par l'effectif correspondant ;
- en faire la somme ;
- diviser par l'effectif total.

Une autre technique $\tau_{5;2}$ disponible est :

- calculer le centre des classes ;
- multiplier par la fréquence correspondante ;
- en faire la somme.

L'élément technologique θ_5 est la définition de la moyenne.

Le dernier exercice ne contient qu'un question. Le type de tâches T_6 est : calculer la médiane d'une série statistique dont la tâche t_6 : calculer la médiane des notes relevées dans l'ordre alphabétique est un spécimen ; la technique τ_6 associée est :

- classer la série dans l'ordre croissant ;
- déterminer l'effectif total N ;
- si $N = 2k$ est pair, alors la médiane est la demi-somme des valeurs de rang k et $k + 1$.

L'élément technologique θ_6 est la définition de la médiane.

4.1.4 L'observation des copies

Intéressons-nous aux réponses des élèves, dont nous trouvons les copies en annexe 3. Les analyses de chaque copie étant regroupées dans l'annexe 4. Nous en faisons la synthèse ci-après.

Le premier exercice, dans sa première question, présentait un type de tâches qui n'a pas été abordé dans le cours (au sens large : la partie *cours*, tout comme la partie *exercices*). On peut penser que comme la construction de l'histogramme est au programme de la classe de 5^{e}, les élèves ont déjà rencontré ce type de tâches. La question est posée de telle façon qu'il est assez difficile de savoir dans quelle mesure les élèves ont suivi telle ou telle technique, le professeur demandant juste de compléter le tableau.

Dans la seconde question, la technique utilisée par les élèves est encore transparente (lié à la formulation de la question par le professeur) ; on relève de nombreuses erreurs car les élèves ont confondu avec les effectifs cumulés croissants ou les fréquences cumulées croissantes. La copie de Rebecca est à ce sujet explicite :

Masse corporelle(en kg)	nombre d'élèves n_i	fréquence f_i (en %)	centre de classe x_i	*Produit*
[40 ;50[	40 ~~150~~	16		
[50 ;60[	110	~~126~~		
[60 ;70[	70 ~~100~~	28		
[70 ;80[	30	12		
Total	250			

(Annexe 3, p. 9)

La technique a été vue dans la partie exercices mais a été institutionnalisée sous une autre forme dans la partie cours.

La troisième question n'a pas été abordé par 5 élèves et 8 donnent une réponse sans calcul ; la technique qu'ils ont utilisé n'est donc pas apparente, les autres répondant correctement. Notons que ce type de tâches T_3^S n'est pas présent dans le cours du professeur.

La quatrième question est un échec pour une très grande majorité d'élèves (seuls 3 élèves ont répondu correctement) ; ce type de tâches n'a pas été vu en cours mais un type de tâches similaire est présent (type de tâches T_{13}^{LC}) mais cela ne suffit pas. La plupart du temps cela est dû à une erreur dans la première question comme on peut le voir, par exemple, dans la copie de Lucie :

Masse corporelle(en kg)	nombre d'élèves n_i	fréquence f_i (en %)	centre de classe x_i
[40 ;50[	~~16~~	16	
[50 ;60[	110		
[60 ;70[	~~94~~	28	
[70 ;80[	30	12	
Total	250		

$$\cancel{94 + 30 = 124}$$

Le nombre d'élève qui ont une masse supérieure à 60 kg est de 124. (Annexe 3, p. 17)

La technique est correcte mais les erreurs liées à la première question viennent rendre le résultat de la technique faux.

La dernière question s'intéresse au calcul de la moyenne. La technique utilisée reprend la technique $\tau_{15;3}^{LC}$ vue dans l'exercice 4 du cours. Pourtant cette question va être massivement un échec, seuls 4 élèves ont répondu correctement à la question. La technique demandée est un peu longue et beaucoup d'élèves ont fait faux dans un des passages de la technique, que ce soit le calcul du centre des classes, le produit $n_i \times x_i$, la somme des produits ou la division par l'effectif total. Un cas intéressant est celui de Teddy :

Masse corporelle(en kg)	nombre d'élèves n_i	fréquence f_i (en %)	centre de classe x_i	Produit $n \times x$
[40 ;50[	40	16	45	1 800
[50 ;60[	110	44	55	6 050
[60 ;70[	70	28	65	4 550
[70 ;80[	30	12	75	2 250
Total	250	100		14 650

$$\frac{100}{14\,650} = 68,2$$

(Annexe 3, p. 19)

On le voit, l'élève réussi parfaitement la complétion du tableau. Seule la dernière partie, celle de la division de la somme des produits par l'effectif total pose problème ; cet élève a divisé la fréquence totale (100%) par la somme des produits, ce qui ne correspond à aucune des techniques que nous avons pu voir.

En ce qui concerne l'exercice 3, qui n'est constitué que d'une question, il est massivement échoué : 3 élèves ne font pas l'exercice, 7 confondent la moyenne et la médiane et enfin 4 se trompent dans la technique. Pour les 6 élèves restant, 5 suivent la technique du cours ($\tau_{12;2}^{LC}$) qui est erronée dans le cas pair. Un seul élève réussira correctement l'exercice en utilisant la bonne technique ; ce qui pose la question : où a-t-il rencontré cette technique puisqu'elle n'est pas dans le cours ?

4.2 Les copies des élèves de KB

4.2.1 Le sujet du devoir maison

Ce devoir maison présente deux exercices. Dans le premier exercice, on trouve un premier type de tâches T_1 : déterminer la médiane d'une série » dont le spécimen t_1 : déterminer la médiane du nombre annuel de jours avec pics d'ozone ; la technique associée τ_1 est :

- ranger les valeurs de la série dans l'ordre croissant ;
- déterminer l'effectif total N ;
- calculer $\frac{N}{2}$:
  - si $\frac{N}{2}$ est pair, la médiane est la demi-somme des valeurs de la série de rang $\frac{N}{2}$ et $\frac{N}{2}+1$;
  - si $\frac{N}{2}$ est impair, la médiane est la valeur de la série de rang $\frac{N+1}{2}$.

L'élément technologique θ_1 est la définition de la médiane.

Le second type de tâches T_2 est : calculer la moyenne d'une série avec le spécimen t_2 : calculer la moyenne du nombre annuel de jours avec pics d'ozone ; la technique associée τ_2 est :

- faire la somme des valeurs de la série ;
- diviser par l'effectif total (nombre d'années de l'étude).

L'élément technologique θ_2 est la définition de la moyenne.

On trouve ensuite le type de tâches T_3 : comparer deux séries avec le spécimen t_3 : comparer les séries du nombre de jours avec pics d'ozone pour le département des Alpes-de-Haute-Provence et du Vaucluse ; la technique τ_3 est :

- comparer les paramètres statistiques calculés dans la question précédente.

L'élément technologique θ_3 est la définition de la moyenne, de la médiane et la comparaison de nombres.

Enfin le dernier type de tâches de cet exercice, T_4 est : justifier une valeur de la série dont le spécimen t_4 : justifier les valeurs élevées des deux séries ; la technique associée τ_4 est :

- déterminer l'année où les deux barres sont plus élevées que les autres ;
- réfléchir à un événement climatologique qui peut expliquer ces barres élevées.

Un second exercice présente en premier lieu le type de tâches T_5 : déterminer la série statistique dont un histogramme pourrait être la représentation ; la technique τ_5 est :

- déterminer une enquête statistique dont la série pourrait avoir comme représentation un histogramme ;
- préciser l'objet de cette enquête ;
- préciser l'échantillon de cette enquête.

L'élément technologique θ_5 est la définition de l'échantillon et de la population.

On trouve un deuxième type de tâches T_6 : compléter un tableau d'effectif à partir d'un histogramme avec le spécimen t_6 : compléter le tableau d'effectif à partir de l'histogramme ; la technique τ_6 est :

- construire un tableau à 5 colonnes ;
- le remplir en fonction de la question précédente.

L'élément technologique θ_6 est la définition d'effectif et caractère.

Le dernier type de tâches T_7 est reproduire et compléter un histogramme avec le spécimen t_7 : reproduire et compléter l'histogramme de l'étude choisie ; la technique associée τ_7 est :

- reproduire l'histogramme ;
- compléter à l'aide des informations imaginées dans les questions précédentes.

L'élément technologique θ_7 est la définition de l'histogramme.

4.2.2 L'observation des copies

Après avoir fait l'analyse praxéologique du sujet, intéressons-nous aux copies des élèves, que l'on trouvera en annexe 6. Les analyses détaillées se trouvent en annexe 7. Nous en faisons une synthèse ci-après.

Le premier exercice proposait dans sa première question de déterminer la médiane d'une série. La technique vue en cours (τ_3^{KB}) a été suivie par 2 élèves. Un élève applique la technique mais de façon incomplète (il ne classe pas les valeurs de la série dans l'ordre croissant). Tous les autres déterminent la médiane « à la main » en utilisant l'élément technologique θ_3^{KB} dont voici un exemple (tiré de la copie de Tessa :

Alpes de Haute Provence :

$$\overbrace{4\ 5\ 6\ 10}^{4}\boxed{13}\overbrace{14\ 16\ 19\ 37}^{4}$$

Le nombre médian est 13.
Vaucluse

$$\overbrace{5\ 7\ 9\ 10}^{4}\boxed{13}\overbrace{18\ 22\ 26\ 48}^{4}$$

Le nombre médian est également 13. (Annexe 6, p. 1)

La technique associée pourrait être :

- classer les valeurs dans l'ordre croissant ;
- compter le nombre de valeurs (N) ;
- diviser ce nombre par 2 ;
- compter $\lfloor \frac{N}{2} \rfloor$ valeurs ;
- positionner la médiane sur la valeur suivante ;
- compter qu'il y a bien $\lfloor \frac{N}{2} \rfloor$ valeurs après la médiane.

Pourquoi utiliser cette nouvelle technique qui nous semble plus complexe et surtout moins robuste dans l'absolue ?

Le calcul suivant est celui de la moyenne qui a été majoritairement réussi ; quelques élèves se trompent dans le calcul de l'effectif total. Deux élèves n'ont pas compris que l'étude portait sur des années et ont calculé le nombre de jours écoulés sur la période des 9 ans ($9 \times 365 = 3285$) comme effectif total ; on trouve un exemple dans la copie dont ni le nom, ni le prénom n'ont été noté :

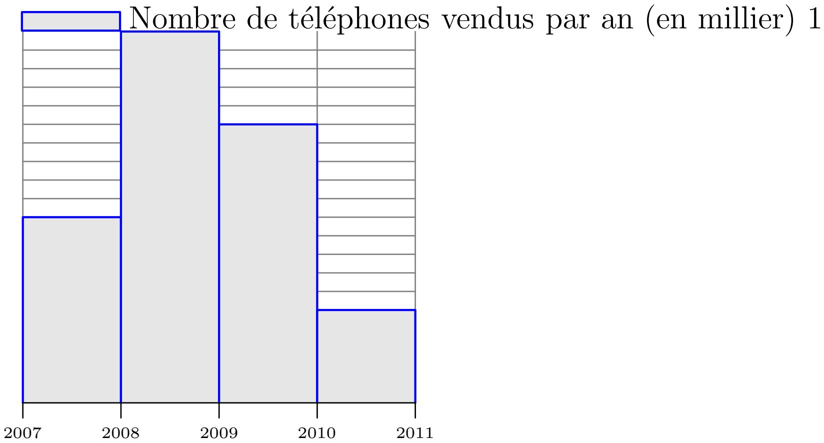
$5+7+9+10+12+22+26+48 : 3285 = 110,0246113$
En moyenne, il y a environ 110 de pics d'ozone par an. (Annexe 6, p. 4)

Dans la question suivante, le but était de comparer les deux séries de pics d'ozone. Beaucoup de copies montrent une comparaison des moyennes et des médianes entre elles, mais nous ne voyons que peu d'élèves conclurent de façon satisfaisante. Pour les autres, la conclusion est présente mais ils ne montrent pas la technique qu'ils ont utilisé pour arriver à ce résultat ; si comparaison il y a eu, elle n'apparaît pas dans les traces écrites. Or, rappelons-le, ce type de tâches n'est pas au programme de la classe considérée et, de plus, la technique qui permettrait d'aborder le type de tâches comparer deux séries, n'est pas présente dans le cours du professeur.

La dernière question demandait de donner une explication plausible aux deux valeurs qui étaient anormalement élevées en 2003. L'explication attendue était la canicule qui a été observé cette année là. Or seule la seule façon qu'ils avaient de trouver cette information était soit d'aller faire une recherche sur l'Internet, soit de fouiller dans leur mémoire, soit de demander de l'aide à un tiers.

Le second exercice a semblé déstabiliser les élèves. Dans la première question, il était demandé d'imaginer de quelle enquête statistique (plus précisément, quelle population était ciblée et quel était le caractère étudié) pouvait être issu d'un histogramme. Un seul élève a vu le caractère quantitatif continu associé à ce type de représentation ; pour tous les autres, ils ont imaginé des études portant sur des caractères quantitatifs discrets, voire même qualitatifs, dont l'histogramme n'est pas la représentation graphique usuelle. Beaucoup d'élèves ont utilisé la situation qui est présente dans la copie de Mathias :

Année	2008	2009	2010	2011
Nombre de téléphone vendu par an (en millier)	10	20	15	5



(Annexe 6, p.16)

La seconde question demandait de présenter les données de l'étude imaginée à la question précédente. La quasi totalité des élèves a présenté un tableau d'effectifs, certains calculant les fréquences associées.

La dernière question demandait de reproduire l'histogramme, en le complétant par les informations qui manquaient. Ce qui a été plutôt bien réussi par les élèves dans l'ensemble. Quelques élèves construisent un diagramme en barres (mais qui ne correspond pas non plus au caractère qu'ils ont choisi d'étudier), le professeur leur faisant la remarque que les « barres doivent être collées ».

4.2.3 Le sujet de l'évaluation

Cette évaluation comporte cinq exercices. Le premier exercice est de restituer la définition de la médiane, donc le type de tâches T_1 : écrire la définition de la valeur médiane d'une série statistique ; la technique τ_1 est : écrire l'élément technologique θ_1, la définition de la médiane.

Le second exercice présente le type de tâches T_2 : déterminer une valeur manquante d'une série avec le spécimen t_2 : déterminer le nombre d'adhérents d'un club de badminton ayant 16 ans ; la technique τ_2 est :

- déterminer l'effectif total ;
- faire la somme des effectifs connus ;
- soustraire la somme à l'effectif total.

L'élément technologique θ_2 est la définition de l'effectif et de l'effectif total.

Le type de tâches suivant est T_3 : déterminer la fréquence des valeurs de la série dont le spécimen t_3 : déterminer la fréquence des adhérents ayant 14 ans ; la technique τ_3 est :

- déterminer l'effectif total ;
- diviser l'effectif de la valeur 14 ans par l'effectif total ;
- multiplier par 100%.

L'élément technologique θ_3 est la définition de la fréquence.

Le type de tâches T_4 est : déterminer la fréquence des valeurs d'une série supérieures ou égaux à une valeur déterminée avec le spécimen t_4 : déterminer la fréquence du nombre d'adhérents ayant 15 ans ou plus ; la technique τ_4 est :

- faire la somme des effectifs des adhérents ayant 15 ans ou plus ;
- diviser par l'effectif total ;
- multiplier par 100%.

L'élément technologique θ_4 est la définition de la fréquence.

L'exercice 3 présente un premier type de tâches T_5 : calculer la moyenne d'une série avec le spécimen t_5 : calculer la taille moyenne des jeunes basketteurs ; la technique τ_5 est :

- faire la somme des valeurs ;
- diviser par l'effectif total.

L'élément technologique θ_5 est la définition de la moyenne.

Le second type de tâches T_6 est : calculer la médiane d'une série avec le spécimen t_6 : calculer la taille médiane des jeunes basketteurs ; la technique τ_6 est :

- classer les valeurs de la série dans l'ordre croissant ;
- déterminer l'effectif total N :
  - si N est impair, la médiane est la valeur de rang $\frac{N+1}{2}$;
  - si N est pair, la médiane est la demi-somme des valeurs de rang $\frac{N}{2}$ et $\frac{N}{2}+1$.

L'élément technologique θ_6 est la définition de la médiane.

Le troisième type de tâches T_7 est : déterminer les quartiles d'une série avec le spécimen t_7 : déterminer les quartiles de la taille des jeunes basketteurs ; la technique τ_7 est :

- classer les valeurs de la série dans l'ordre croissant ;
- déterminer l'effectif total N ;
- calculer $\frac{N}{4}$:
  - si $\frac{N}{4}$ est un entier alors le 1er quartile est la valeur de la série de rang $\frac{N}{4}$;
  - sinon prendre l'entier immédiatement supérieur comme rang de la valeur de la série qui sera le 1er quartile.
- calculer $\frac{3N}{4}$:

- ◦ si $\frac{3N}{4}$ est un entier alors le 3^{e} quartile est la valeur de la série de rang $\frac{3N}{4}$;
- ◦ sinon prendre l'entier immédiatement supérieur comme rang de la valeur de la série qui sera le 3^{e} quartile.

L'élément technologique θ_7 est la définition des quartiles.

Le quatrième type de tâches T_8 est : Interpréter les paramètres statistiques en langage usuel avec le spécimen t_8 : interpréter la moyenne, la médiane et les quartiles de la taille des jeunes basketteurs ; la technique τ_8 est :

- écrire : « la taille moyenne des jeunes basketteurs est : » ;
- recopier la valeur calculée à la question 1 ;
- écrire : « la taille médiane des jeunes basketteurs est : » ;
- recopier la valeur calculée à la question 2 ;
- écrire : « le premier quartile de la taille des jeunes basketteurs est : » ;
- recopier la valeur calculée à la question 3 ;
- écrire : « le troisième quartile de la taille des jeunes basketteurs est : » ;
- recopier la valeur calculée à la question 3.

L'élément technologique θ_8 est la définition de la médiane, de la moyenne et des quartiles.

Le type de tâches suivant est T_9 : calculer l'étendue d'une série avec le spécimen t_9 : calculer l'étendue des tailles des jeunes basketteurs ; la technique τ_9 est :

- déterminer le maximum de la série ;
- déterminer le minimum de la série ;
- soustraire le minimum au maximum.

L'élément technologique θ_9 est la définition de l'étendue.

Le quatrième exercice permet de retrouver le type de tâches T_5 avec le spécimen t'_5 : calculer la moyenne de la durée d'attente au téléphone ; la technique change τ'_5 :

- pour chaque valeur de la série, multiplier la valeur par son effectif ;
- faire la somme des produits ;
- diviser par l'effectif total.

L'élément technologique θ_5 est identique.

On retrouve ensuite le type de tâches T_6 avec le spécimen t'_6 : calculer la médiane de la durée d'attente au téléphone ; la technique change τ'_6 :

- déterminer l'effectif total N ;
- calculer $\frac{N}{2}$;
- déterminer la valeur de la série dont l'effectif cumulé croissant est inférieur ou égal à $\frac{N}{2}$.

L'élément technologique θ_6 reste inchangé.

On retrouve en dernier lieu le type de tâches T_7 avec le spécimen t'_7 : calculer les quartiles du temps d'attente au téléphone ; la technique change τ'_7 :

- déterminer l'effectif total N ;
- calculer $\frac{N}{4}$;
- déterminer la valeur dont l'effectif cumulé croissant est immédiatement supérieur ou égal à $\frac{N}{4}$;
- calculer $\frac{3N}{4}$;
- déterminer la valeur dont l'effectif cumulé croissant est immédiatement supérieur ou égal à $\frac{3N}{4}$.

L'élément technologique θ_7 est inchangé.

Le dernier exercice de cette évaluation permet de retrouver le type de tâches T_5 avec le spécimen t''_5 : calculer le nombre moyen de livres empruntés durant les 12 derniers mois ; la technique est τ_5.

On trouve ensuite le type de tâches T_{10} : déterminer la série dont l'effectif des valeurs est supérieur à une valeur fixée avec le spécimen t_{10} : déterminer la classe qui a le plus de grands lecteurs ; la technique τ_{10} est :

- pour la 1re série :
  - compter le nombre d'élèves qui ont lu plus de 5 livres.
- pour la 2e série :
  - la médiane représente 50% des valeurs ; ici cela représente la 13e valeur ;
- conclure.

L'élément technologique θ_{10} est la définition de la médiane et des effectifs.

On trouve le type de tâches T_{11} : déterminer le maximum d'une série avec le spécimen t_{11} : déterminer pour chaque série, l'élève qui a emprunté le plus de livres ; la technique τ_{11} est :

- déterminer le maximum de la 1re série ;
- la médiane de la 2e série est 8 donc le maximum est plus grand ;
- conclure que c'est la 2e série.

L'élément technologique θ_{11} est la définition de la médiane et du maximum.

4.2.4 L'observation des copies

Les copies observées se trouvent en annexe 8 et les analyses détaillées en annexe 9. Nous en faisons une synthèse ci-après.

Le premier exercice demandait la restitution de la définition de la médiane. Nous constatons qu'une grande majorité d'élèves n'a pas su restituer l'élément technologique θ_3^{KB} de façon satisfaisante. A contrario, il nous semble qu'a été retenu l'aspect « graphique » de la médiane, à savoir que c'est la valeur qui partage la série en deux.

Le second exercice, dans sa première question, demandait d'effectuer le type de tâches T_2^S. La totalité des élèves a réussi à répondre ; seuls deux élèves présentent la technique avant de l'appliquer et un seul justifie la technique qu'il utilise en donnant l'élément technologique associé.

La seconde question demandait dans un premier temps de calculer une fréquence. La technique qui prévaut dans la classe est la suivante :

- calculer le quotient $\frac{n_i}{N}$;
- multiplier par 100 ;
- dans la rédaction de la réponse, donner le résultat avec le symbole %.

La seconde partie de la question demandait de calculer une fréquence cumulée. Les élèves ont utilisé la même technique que le calcul précédent mais peu ont montré comment ils obtenaient l'effectif des adhérents ayant 15 ans ou plus. La copie de Killian montre le calcul suivant :

> 15 ans ou plus :
> $\frac{23}{30} \times 100 = 76,7$. Il y a 76,7% d'adhérents ayants 15 ans ou plus. (Annexe 8, p. 8)

On le voit, le calcul du « 23 » n'est absolument pas détaillé.

Dans l'exercice 3, la première question demande de calculer la moyenne. Hormis quelques erreurs dans la disposition des calculs et de frappes sur la calculatrice, une majorité d'élèves a réussi à répondre. On notera par contre que les fonctions statistiques de la calculatrice n'ont pas été utilisées par les élèves.

La question suivante demande de calculer la médiane. La technique vue en cours n'est que très rarement mise en œuvre. Les élèves préfèrent utiliser la technique suivante qui n'a pas été institutionnalisée dans le cours du professeur :

- classer la série dans l'ordre croissant ;
- donner N ;
- calculer $\frac{N}{2}$;
- sur la série classée, compter $\lfloor\frac{N}{2}\rfloor$ valeurs ;
- la valeur suivante est la médiane ;
- compter qu'il y a bien $\lfloor\frac{N}{2}\rfloor$ valeurs après la médiane.

On retrouve donc ce que nous avions déjà mis en avant dans le cas du devoir à la maison, comme nous le montre la copie de Valentin :

> $\overbrace{165, 165, 166, 170, 171, 174, 174,}^{7,5}\boxed{175}\overbrace{176, 176, 177, 178, 181, 184, 187}^{7,5}$
> (Annexe 8, p. 12)

Il est étonnant que cet élève ait choisi de diviser l'effectif total par deux ($15 \div 2 = 7,5$) et ait partagé la série en 7,5 et 7,5 en choisissant la valeur médiane comme étant simultanément dans les deux sous-séries.

La question qui suit concerne la détermination des quartiles. Les principales erreurs relevées sur la technique du cours sont un arrondi à l'entier inférieure à la place d'un arrondi supérieur lorsque le résultat du calcul n'est pas entier ou l'oubli de classer la série dans l'ordre croissant. Un élève, dans la détermination du 3e quartile, nous reproduisons sa copie (il s'agit de celle de Mélanie)

> $15 \div 4 = 3,75$
> Q_1 = 3e valeur : 166.
> $(15 \div 4) \times 3$ =11,25
> Q_3 = 5e valeur en partant de la fin = 177 (Annexe 8, p. 25)

La détermination du 3e quartile est surprenante : pourquoi calculer $\frac{N}{4} \times 3$ si ce n'est pour ne pas s'en servir. De plus, la 5e valeur en partant de la fin correspond à la 11e valeur c'est-à-dire l'arrondi inférieur du calcul précédent.

La question suivante s'intéresse à l'interprétation des résultats obtenus. En général, cette question a été très peu abordée par les élèves. Pour ceux qui ont répondu, les interprétations (qui en fait n'en sont pas) consistent à reprendre l'élément technologique et à donner la valeur calculée. Par exemple, dans la copie de Marie on trouve :

> La taille médiane signifie qu'il y a autant de basketteurs mesurant plus ou autant de 1,75m que de basketteurs mesurant autant ou moins de 1,75m. Le premier quartile est 170cm, cela signifie qu'un quart des valeurs sont inférieures ou égales à 170cm et Q_3 est égal à 178cm, cela signifie que les trois quarts des chiffres de la série sont égaux ou inférieurs à 178cm. (Annexe 8, p. 7)

La dernière question demande de calculer l'étendue. La quasi totalité des élèves a réussi a calculer l'étendue, un élève fait une erreur (inversion du maximum et du minimum) et quelques élèves oublient l'unité.

L'exercice 4 commence par demander un calcul de moyenne. Les élèves ont quasiment tous réussi en appliquant la technique du cours. On relèvera quelques oublis d'unités et des arrondis qui ne sont pas demandés. Un élève se trompe dans l'effectif total et un autre oublie que les valeurs ont des effectifs associés. En revanche la seconde question qui demande le calcul de la médiane est un échec. Beaucoup d'élèves ont confondu la technique du calcul dans le cas où l'effectif est pair et dans le cas où il est impair ; on en trouve trace dans la copie de Célia :

La durée médiane $16 \div 2$ donc la 8e valeurs. La durée médiane est 6 (Annexe 8, p. 22).

Les réponses données sont justes car les valeurs de rang $\frac{N}{2}$ et $\frac{N}{2}+1$ sont égales. L'autre erreur trop souvent rencontrée est le mauvais classement de la série (oubli de la multiplicité de certaines valeurs).

La question suivante s'intéresse à la détermination des quartiles et la grande majorité de la classe a utilisé la technique du cours. Les principales erreurs sont la confusion entre la valeur d'un quartile et la technique qui permet de l'obtenir (le rang qui permet de repérer sa position dans la série classée), un mauvais calcul de l'effectif total (ce qui conditionne le reste de la technique) et parfois certains élèves qui « inventent » une nouvelle technique (absolument fausse par ailleurs) pour déterminer les quartiles.

Le dernier exercice reprend le calcul de la moyenne dans la première question et la technique du cours est dans la grande majorité utilisée par les élèves. On note quelques erreurs d'arrondis ou des calculs mal présentés. Pour la seconde partie de la question, la comparaison des moyennes est assez souvent avancée mais les comparaisons ne sont pas explicitement notées sur la copie. Peu d'élèves tirent une conclusion de la comparaison. Dans la seconde question, la technique qui permet de calculer la médiane de la 1re série reste majoritairement non maîtrisée (technique incomplète ou confusion du cas pair et du cas impair). Ici encore, peu d'élèves font une comparaison explicite des valeurs pour argumenter la réponse qu'ils donnent. Enfin dans la dernière question, que peu d'élèves ont abordé, la comparaison de la médiane de la 2e série avec le maximum de la 1re série n'est pas un type de tâches courant, en tout cas, il n'a pas été abordé dans la classe. La technique n'ayant pas été vue, il est difficile aux élèves de « l'imaginer » ?

4.3 Les copies des élèves de PN

4.3.1 Le sujet de l'évaluation

Cette évaluation porte sur 2 chapitres du cours : celui qui concerne les fonctions (exercices 1 et 2) et celui qui concerne la statistique (exercices 3 et 4). Nous laisserons donc les exercices 1 et 2 de côté.

Le premier type de tâches T_1 est : calculer les fréquences dont le spécimen t_1 : calculer les fréquences des dépenses en euros des clients d'un magasin ; la technique associée est τ_1 :

- calculer la somme de tous les effectifs : effectif total ;
- pour chaque valeur de la série, prendre l'effectif et diviser par l'effectif total ;
- multiplier le résultat par 100%.

L'élément technologique θ_1 est la définition des fréquence, effectif et effectif total.

Le second type de tâches T_2 est : calculer les fréquence cumulées croissantes dont le spécimen t_2 : calculer les fréquences cumulées croissantes des dépenses des clients d'un magasin ; la technique associée τ_2 est :

- recopier la première fréquence de la première valeur ;
- puis ajouter la fréquence de la valeur suivante ;
- vérifier que la dernière valeur a pour fréquences cumulées croissantes 100%.

L'élément technologique θ_2 est la définition de la fréquence et des fréquences cumulées croissantes.

On trouve ensuite le type de tâches T_3 : déterminer le mode de la série avec le spécimen t_3 : déterminer le mode de la série des dépenses des clients d'un magasin ; une technique possible $\tau_{3;1}$ est :

- repérer l'effectif le plus grand ;
- en déduire le mode de la série.

Une autre technique possible $\tau_{3;2}$:

- repérer la fréquence la plus grande ;
- en déduire le mode de la série.

L'élément technologique θ_3 est la définition du mode et de la fréquence.

Le type de tâches T_4 est : calculer la moyenne d'une série dont le spécimen t_4 est : calculer la moyenne de la série des dépenses des clients d'un magasin ; une technique possible $\tau_{4;1}$ est :

- pour chaque classe, faire la demi-somme des extrémités de la classe : centre de la classe ;
- faire le produit du centre de la classe par l'effectif correspondant ;
- faire la somme de ces produits ;
- diviser par l'effectif total.

Une autre technique $\tau_{4;2}$ possible :

- pour chaque classe, faire la demi-somme des extrémités de la classe : centre de la classe ;
- faire le produit du centre de la classe par la fréquence correspondante ;
- faire la somme de ces produits.

L'élément technologique θ_4 est la définition des moyenne, fréquence, effectif et centre de classe.

Le type de tâches T_5 est : interpréter un paramètre statistique en langage usuel avec le spécimen $t_{5;1}$: interpréter le mode de la série en langage usuel et le spécimen $t_{5;2}$: interpréter la moyenne de la série en langage usuel ; la technique $\tau_{5;1}$ associée est :

- écrire : « le mode de la dépense des clients du magasin de vêtements est » ;
- copier le résultat de la détermination du mode de la série.

De même la technique $\tau_{5;2}$ est :

- écrire : « la dépense moyenne des clients dans ce magasin de vêtements est » ;
- recopier la moyenne calculée.

Les éléments technologiques $\theta_{5;1}$ sont la définition du mode et $\theta_{5;2}$ la définition de la moyenne.

Le type de tâches T_6 est : construire un histogramme dont le spécimen t_6 : construire l'histogramme de la série des dépenses des clients du magasin de vêtements ; la technique τ_6 est :

- tracer l'axe des abscisses en le graduant tous les cm ;
- représenter l'unité d'aire par 0,5 cm^2 représente un client ;
- pour chaque classe, représenter un rectangle tel que l'aire du rectangle soit proportionnel à l'effectif correspondant.

L'élément technologique θ_6 est la définition de l'histogramme.

Le type de tâches T_7 est : construire le polygone des fréquences cumulées croissantes dont le spécimen t_7 : construire le polygone des fréquences cumulées croissantes des dépenses des clients du magasin de vêtements ; la technique τ_7 associée est :

- tracer l'axe des abscisses en le graduant tous les cm de 10 euros ;
- tracer l'axe des ordonnées en le graduant tous les cm de 10% ;
- pour chaque classe $[a_i ; b_i[$, placer le point de coordonnées $(b_i ; n_i)$;
- rejoindre les points par des segments.

L'élément technologique θ_7 est la définition du polygone des fréquences cumulées croissantes.

Le type de tâches T_8 est : déterminer la médiane d'une série dont le spécimen t_8 : déterminer la médiane de la série des dépenses des clients du magasin ; la technique τ_8 associée :

- se placer sur l'axe des ordonnées au point $I(0; 50\%)$;
- tracer la parallèle à l'axe des abscisses passant par I ;
- le point d'intersection avec le polygone des fréquences cumulées croissantes est A ;

- tracer la parallèle à l'axe des ordonnées passant par A ;
- le point d'intersection avec l'axe des abscisses est B ;
- donner l'abscisse de B comme valeur approchée de la médiane.

L'élément technologique θ_8 est la définition de la médiane et du polygone des fréquences cumulées croissantes.

Le type de tâches T_9 est : déterminer les quartiles d'une série dont le spécimen t_9 : déterminer les quartiles de la série des dépenses des clients du magasin de vêtements ; la technique τ_9 est :

- se placer sur l'axe des ordonnées au point $I(0; 25\%)$;
- tracer la parallèle à l'axe des abscisses passant par J ;
- le point d'intersection avec le polygone des fréquences cumulées croissantes est C ;
- tracer la parallèle à l'axe des ordonnées passant par C ;
- le point d'intersection avec l'axe des abscisses est D ;
- donner l'abscisse de D comme valeur approchée du premier quartile ;
- se placer sur l'axe des ordonnées au point $I(0; 75\%)$;
- tracer la parallèle à l'axe des abscisses passant par K ;
- le point d'intersection avec le polygone des fréquences cumulées croissantes est E ;
- tracer la parallèle à l'axe des ordonnées passant par E ;
- le point d'intersection avec l'axe des abscisses est F ;
- donner l'abscisse de F comme valeur approchée du troisième quartile.

L'élément technologique θ_9 est la définition des quartiles et du polygone des fréquences cumulées croissantes.

Dans la question suivante, on retrouve le type de tâches T_5 avec les spécimens $t_{5;3}$: interpréter la médiane de la série en langage usuel ; la technique associée $\tau_{5;3}$ est :

- écrire : « la médiane de la série des dépenses des clients du magasin de vêtements est environ de » ;
- recopier la valeur de la médiane trouvée.

Le spécimen $t_{5;4}$: interpréter les 1er et 3^{e} quartile de la série en langage usuel ; la technique $\tau_{5;4}$:

- écrire : « 25% des valeurs de la série sont inférieures ou égales à » ;
- recopier la valeur du premier quartile ;
- écrire : « 75% des valeurs de la série sont inférieures ou égales à » ;
- recopier la valeur du troisième quartile.

Les éléments technologiques $\theta_{5;3}$ et $\theta_{5;4}$ sont les définitions de la médiane et des quartiles.

L'exercice 4 présente un premier type de tâches T_{10} : déterminer l'étendue d'une série dont le spécimen t_{10} : déterminer l'étendue de la série des températures moyennes mensuelles d'une ville ; la technique τ_{10} est :

- déterminer la température la plus grande ;
- déterminer la température la plus petite ;
- soustraire la valeur la plus petite à la plus grande.

L'élément technologique θ_{10} est la définition de l'étendue.

On retrouve le type de tâches T_4 avec le spécimen t'_4 : calculer la moyenne des températures moyennes mensuelles d'une ville ; la technique change τ'_4 :

- faire la somme des températures ;
- diviser par l'effectif total.

Ensuite, on retrouve le type de tâches T_8 avec le spécimen t'_8 : déterminer la médiane des températures moyennes mensuelles d'une ville ; la technique change τ'_8 :

- classer les valeurs de la série dans l'ordre croissant ;

- déterminer l'effectif total N ;
- si N est pair, la médiane est la demi-somme des valeurs de rang $\frac{N}{2}$ et $\frac{N}{2}+1$;
- si N est impair, la médiane est la valeur de la série de rang $\frac{N+1}{2}$.

A la question suivante, on retrouve le type de tâches T_9 avec le spécimen t'_9 : déterminer les 1er et 3^{e} quartiles de la série des températures moyennes mensuelles d'une ville ; la technique change τ'_9 :

- classer les valeurs de la série dans l'ordre croissant ;
- déterminer l'effectif total N ;
- calculer $\frac{N}{4}$:
  - si $\frac{N}{4}$ est un entier alors le premier quartile est la valeur de la série de rang $\frac{N}{4}$;
  - sinon, on prend l'entier immédiatement supérieur comme rang de la valeur qui correspond à Q_1.
- calculer $\frac{3N}{4}$:
  - si $\frac{3N}{4}$ est un entier alors le troisième quartile est la valeur de la série de rang $\frac{3N}{4}$;
  - sinon, on prend l'entier immédiatement supérieur comme rang de la valeur qui correspond à Q_3.

On trouve le type de tâches T_{11} : comparer deux séries avec le spécimen t_{11} : comparer la série des températures moyennes mensuelles de Barcelone et la série des températures moyennes mensuelles de Mexico ; la technique τ_{11} associée :

- pour chaque affirmation : trouver le paramètre statistique calculé dans la question 1 qui justifie l'affirmation.

L'élément technologique θ_{11} est la définition de la moyenne, de la médiane, de l'étendue et des quartiles.

4.3.2 L'observation des copies

L'analyse du sujet ayant été abordée, intéressons-nous aux copies des élèves (que l'on trouvera en annexe 11). Le détail des analyses se trouve en annexe 12. Nous en faisons une synthèse ci-après. Les deux premiers exercices évaluaient des types de tâches liés au domaine des fonctions, nous les laissons de côté.

Dans l'exercice 3, la première question demande de déterminer les fréquences de chaque valeur et de compléter le tableau. Comme le détail des calculs n'est pas demandé, la technique utilisée par les élèves n'est pas apparente. Pour le calcul des fréquences cumulées croissantes, la technique n'est pas plus apparente mais les élèves qui ont fait faux à cette partie de la question, laissent entendre que c'est la technique du cours ($\tau_{12;1}^{PN}$) qui a été utilisée. La seconde question demande de calculer le mode, la moyenne et d'en donner les interprétations. La technique utilisée par les élèves pour déterminer le mode de la série est celle du cours. En revanche, quelques élèves confondent la technique pour déterminer le mode et le mode lui-même : la confusion porte sur les effectifs des valeurs et les valeurs elles-mêmes, comme on peut le voir dans la copie de Garance :

> Le mode de cette série statistique est 48 (le plus grand effectif) et il a pour valeur 105 (le milieu de 90 à 120 environ)
> (Annexe 11, p. 19)

Non seulement le mode est le plus grand effectif, mais en plus il a deux valeurs (ici 48 et 105) ; la seconde étant le centre de la classe [90; 120[. Un élève pense à utiliser la technique alternative qui consiste à regarder les fréquences à la place des effectifs, mais ne le montre pas ; seule l'interprétation qu'il fait permet de dire qu'il a utilisé cette technique là. Le caractère étant quantitatif continu, certains élèves remarquent qu'on parle plutôt de classe modale dans ce cas précis. La seconde partie de la question concernait le calcul de la moyenne. Les élèves utilisent conformément à ce qui a été institutionnalisé, la technique qu'ils ont étudié avec le professeur. Les principales erreurs relevées concernent l'oubli du calcul du centre des classes ou le remplacement du centre d'une classe par une des extrémités. Les erreurs liées à une mauvaise utilisation de la calculatrice sont aussi assez fréquentes. Sur la même copie (Garance) :

$$\frac{30 \times 12 + 45,5 \times 18 + 75,5 \times 30 + 105 \times 48 + 135,5 \times 42}{150} = 135$$

(Annexe 11, p. 19)

Le calcul des centres n'est pas toujours identique : pour la première classe, l'extrémité inférieure est prise pour centre, puis pour les deux classes suivantes, il y a des décimales (45,5 et 75,5) alors que les valeurs auraient dû être entières ; la classe suivante, le calcul est juste et on retrouve une décimale pour la dernière classe. Les interprétations du mode et de la moyenne sont rarement faites car ce type de tâches n'a été que peu abordé dans la partie cours mais seulement vu dans la partie exercices. L'interprétation du mode est essentiellement une reprise de l'élément technologique. Quant à la moyenne, les élèves reprennent en général la question pour donner l'interprétation de ce paramètre.

La question suivante demande de tracer un histogramme, ce qui est dans la plupart des cas, plutôt réussi malgré l'absence de la légende (même si elle est imposée par l'énoncé) ou de l'étiquette sur l'axe. Quelques élèves construisent l'histogramme à partir des fréquences plutôt qu'à partir des effectifs. La question suivante est en revanche moins bien réussie. La construction du polygone des fréquences cumulées croissantes est entachée d'erreurs liées à l'utilisation des fréquences en lieu et place des fréquences cumulées croissantes ou à l'utilisation des centres de classe au lieu des extrémités b_i des classes $[a_i; b_i[$.

La dernière question demande de calculer la médiane et les quartiles puis d'en donner les interprétations. Hormis les élèves qui n'ont pas fait la question ou qui se sont trompés dans la construction du polygone des fréquences cumulées croissantes, les élèves ont suivi la technique du cours pour la détermination de la médiane et des quartiles. Les quelques erreurs rencontrées concernent la mauvaise lecture des valeurs sur l'axe. Mais est-ce vraiment une erreur ? En effet cette technique ne donne qu'une valeur approchée de la médiane ou des quartiles. Les interprétations des valeurs calculées sont souvent des phrases « standardisées » comme nous avons pu les voir dans le cours du professeur KB.

L'exercice 4 débute par le calcul de l'étendue. Les principales erreurs relevées sont soit l'inversion du maximum et du minimum dans le calcul, soit la confusion entre les valeurs de la série et les effectifs. La question suivante demande le calcul de la moyenne. Hormis quelques erreurs d'arrondis ou des problèmes de présentation des calculs, la technique du cours a été respectée. Ensuite, la question 1.c demandait de calculer la médiane, ce qui a été moins bien réussi. Souvent la technique est incomplète ou il y a confusion entre la technique du cas pair et du cas impair. Mais le plus significatif est qu'un nombre important d'élèves n'a pas abordé la question. En revanche, la détermination des quartiles est un peu mieux réussi même si le nombre d'élèves ayant répondu à la question est faible. On retrouve par ailleurs des techniques incomplètes ou des mixes de techniques ; par exemple dans la copie d'Alizé :

Q_1 Mexico : $12 \div 4 = 3 = \frac{13,9+14,1}{2} = 14$
Q_1 Barcelone : $12 \div 4 = 3 = \frac{10,3+12,4}{2} = 11,35$
(Annexe 11, p. 15)

On remarque de plus, que les égalités enchaînées sont nombreuses.

La question qui suit demande 4 comparaisons : les moyennes, les étendues, les quartiles et les médianes. On remarque cependant que les arguments apportés sont assez souvent les bons, mais que les comparaisons sont rarement explicites. Peut-être que la formulation de la question en est à l'origine.

4.4 Les copies des élèves de YC

4.4.1 Le sujet du devoir maison

Ce sujet de devoir à la maison comporte deux exercices.

Les deux premiers types de tâches T_1 et T_2 sont : déterminer la médiane d'une série et déterminer les quartiles d'une série dont les deux spécimens respectifs sont $t_{1;1}$: déterminer la médiane de la série M et $t_{2;1}$: déterminer les quartiles de la série M ; les techniques associées sont :

- τ_1 :
  - rentrer dans la calculatrice les données de la série M ;
  - afficher les paramètres de la série ;
  - donner celui qui correspond à la médiane.
- τ_2 :
  - rentrer dans la calculatrice les données de la série M ;
  - afficher les paramètres de la série ;
  - donner celui qui correspond aux premier et troisième quartile.

  Les éléments technologiques sont θ_1 :définition de la médiane et θ_2 : définition des quartiles d'une série ; ainsi que la calculatrice.

On retrouve dans cette question les deux types de tâches T_1 et T_2 avec les spécimens $t_{1;2}$: déterminer la médiane de la série P et $t_{2;2}$: déterminer les quartiles de la série P. Les techniques et technologies sont les mêmes que dans la question 1.

On trouve ensuite le type de tâches T_3 : comparer deux séries dont un spécimen est $t_{3;1}$: comparer les séries M et P ; la technique associée $\tau_{3;1}$ est :

- comparer les médianes ;
- comparer les quartiles.

L'élément technologique $\theta_{3;1}$ est la définition de la médiane, des quartiles et la comparaison de nombres.

Le dernier type de tâches de cet exercice est T_4 : interpréter le 3^{e} quartile en langage courant dont un spécimen est t_4 : interpréter le 3^{e} quartile de la série M ; la technique τ_4 associée est :

- lire le 3^{e} quartile dans la question 1 ;
- comparer avec le nombre annoncé dans la question :
  - si le 3^{e} quartile est inférieure alors la réponse à la question est non ;
  - si le 3^{e} quartile est supérieure alors la réponse à la question est non ;
  - si le 3^{e} quartile est égale alors la réponse à la question est oui.

L'élément technologique θ_4 est la définition du 3^{e} quartile et la comparaison de nombres.

En ce qui concerne l'exercice 2, on trouve un premier type de tâches T_5 : construire un diagramme en bâtons d'une série dont le premier spécimen est $t_{5;1}$: construire le diagramme en bâtons de la température moyenne du mois de janvier 2000 de la ville de Paris et le second spécimen $t_{5;2}$: construire le diagramme en bâtons de la température moyenne du mois de janvier

2007 de la ville de Paris ; la technique τ_5 associée est :

- construire un repère ;
- sur l'axe des abscisses placer les jours ;
- sur l'axe des ordonnées placer les températures ;
- la hauteur du bâton sera la température associée au jour du mois de janvier de l'année correspondante.

L'élément technologique θ_5 est la définition du diagramme en bâtons.

On retrouve par ailleurs le type de tâches T_3 à partir du spécimen $t_{3;2}$: comparer la série des températures de 2000 et la série des températures de 2007 ; mais cette fois-ci à l'aide de la technique $\tau_{3;2}$:

- pour chaque jour, comparer la hauteur des bâtons des deux séries ;
- conclure que l'une des séries a des températures plus importantes que l'autre sur l'ensemble du mois de janvier.

L'élément technologique $\theta_{3;2}$ est la comparaison de longueurs et la définition du diagramme en bâtons.

Dans la question qui suite, on retrouve le type de tâches T_5 et le spécimen $t_{5;2}$ avec la technique $t'_{5;2}$:

- classer les températures dans l'ordre croissant ;
- construire un repère ;
- sur l'axe des abscisses placer les jours ;
- sur l'axe des ordonnées placer les températures ;
- la hauteur du bâton sera la température associée au jour du mois de janvier de l'année 2007.

Ensuite, on trouve le type de tâches T_6 : calculer la moyenne d'une série avec le spécimen t_6 : calculer la moyenne de la série des températures moyennes de la ville de Paris au mois de janvier 2007 ; la technique τ_6 est :

- additionner les valeurs de la série ;
- diviser par l'effectif total.

L'élément technologique θ_6 est la définition de la moyenne.

Dans la question suivante, on retrouve les types de tâches T_1 et T_2 avec les techniques correspondantes ou à l'aide des techniques τ'_1 (les valeurs étant déjà rangées dans l'ordre croissant) :

- déterminer l'effectif total N ;
- si N est pair : la médiane est la demi-somme des valeurs de rang $\dfrac{N}{2}$ et $\dfrac{N}{2}+1$;
- si N est impair : la médiane est la valeur de la série de rang $\dfrac{N+1}{2}$.

et τ'_2 :

- déterminer l'effectif total N ;
- calculer $\dfrac{N}{4}$:
  - si c'est un entier, le 1^er^ quartile est la valeur de rang $\dfrac{N}{4}$;
  - sinon, prendre le rang immédiatement supérieur pour trouver Q_1.
- calculer $\dfrac{3N}{4}$:
  - si c'est un entier, le 3^e^ quartile est la valeur de rang $\dfrac{3N}{4}$;
  - sinon, prendre le rang immédiatement supérieur pour trouver Q_3.

Les éléments technologiques θ_1 et θ_2 restent inchangés.

La dernière question présente le type de tâches T_7 : déterminer si la valeur d'une série est une grande valeur avec le spécimen t_7 : déterminer la température pour dire s'il fait chaud à Paris au mois de janvier ; la technique associée τ_7 est :

- lire la valeur du 3e quartile.

L'élément technologique θ_7 est la définition du 3e quartile et la comparaison de nombres.

4.4.2 L'observation des copies

Après avoir analysé le sujet, intéressons-nous aux copies des élèves (que l'on trouvera en annexe 14) ainsi qu'à leur analyse (que l'on trouvera en annexe 15) ; nous en donnons une synthèse ci-après.

Le premier exercice et plus précisément les deux premières questions, demandait de calculer les médianes et quartiles de deux séries, pour tester l'efficacité d'un médicament. La première remarque que nous pouvons faire est que la consigne de l'énoncé demandait explicitement de le faire à la calculatrice et la quasi totalité des élèves l'ont fait à la main. Ceci nous permet de voir que la technique n'est pas maîtrisée par les élèves. Par exemple la copie de Coralie :

Médiane

12	12	13	13	13	13	13,5	13,5	14	14
14	14	14	14	14	14,5	14,5	14,5	14,5	14,5
14,5	15	15	15	15	16	16	17	18	18

$\frac{N}{2} = \frac{30}{2} = 15.$

Donc c'est la quinzième valeur.

La médiane est de 14.

(Annexe 14, p. 1)

La question suivante demande de comparer les séries. Nombre d'entre eux font la comparaison des médianes, mais peu en tirent une conclusion. Or le but de l'exercice était de tester l'efficacité du médicament, donc de prendre une décision ; peu d'élèves ont compris la question. Le type de tâches ayant été peu abordé avec les élèves (et la technique qui n'a pas été institutionnalisée) c'est peut-être une première réponse à cet échec. La dernière question pose question quant à son utilité, dans le contexte de l'exercice.

Le second exercice présente une étude des températures de la ville de Paris pour les mois de janvier en 2000 et 2007. La première question demandait de représenter graphiquement les deux séries sur le *même* graphique mais trop peu d'élèves ont suivi la consigne. La comparaison est, par contre, faite dans de nombreux cas et de manière satisfaisante. La question suivante demandait de construire le diagramme en bâtons pour la série de 2007 mais classée dans l'ordre croissant. Le graphique est bien construit mais l'interprétation qui en est faite n'est que rarement pertinente par rapport à la question posée. La question qui suit demande le calcul de la moyenne, ce qui est plutôt bien réussi dans l'ensemble. En revanche, l'interprétation qui suit reste très basique : si la température est supérieure à la moyenne, alors il fait chaud. Or la moyenne vaut 8,3° : peut-on vraiment dire qu'il fait chaud à Paris en janvier s'il fait 8 degrés ? Il faut aller chercher plus loin. La question suivante s'y intéresse en demandant de calculer la médiane et les quartiles où l'on retrouve les mêmes erreurs que dans le premier exercice. L'interprétation n'est absolument pas en adéquation avec la question de départ, les élèves se contentant de reprendre ce qu'ils avaient écrit plus avant dans le devoir.

4.4.3 Le sujet de l'évaluation

Cette évaluation présente trois exercices.

Dans le premier exercice, le type de tâches demandé dans cette question est T_1 : déterminer le caractère d'une série statistique dont la tâche t_1 : déterminer le caractère de la série statistique correspondant au résultat de l'enquête sur le nombre de pièces de l'habitation principale est un spécimen ; la technique τ_1 associée est :

- regarder dans l'énoncé quel est l'enjeu de l'étude.

L'élément technologique θ_1 est la définition du caractère statistique.

Dans cette question, le type de tâches est T_2 : déterminer la nature du caractère de la série statistique étudiée dont la tâche t_2 : déterminer la nature du caractère « nombres de pièces de l'habitation principale » est un spécimen ; la technique τ_2 associée est :

- dans l'énoncé regarder si les valeurs du caractère sont des nombres ou des qualités ;
- si les valeurs du caractère sont des nombres :
  - si les nombres sont isolés alors le caractère est quantitatif discret ;
  - sinon il est quantitatif continu.
- si les valeurs ne sont pas des nombres (donc des qualités), le caractère est qualitatif.

L'élément technologique θ_2 est la définition du caractère quantitatif.

Le troisième type de tâches T_3 est : déterminer le mode d'une série statistique dont la tâche t_3 : déterminer le mode de la série « nombre de pièces de l'habitation principale » est un spécimen ; la technique τ_3 associée est :

- repérer la valeur qui a le plus grand effectif ;
- le mode est la valeur du caractère correspondante.

Les éléments technologiques θ_3 sont la définition de l'effectif et la définition du mode.

Le quatrième type de tâches T_4 est : déterminer l'étendue d'une série statistique dont la tâche t_4 : déterminer l'étendue de la série « nombre de pièces de l'habitation principale » est un spécimen ; la technique τ_4 associée est :

- déterminer la plus grande valeur du caractère ;
- déterminer la plus petite valeur du caractère ;
- faire la différence entre la plus grande et la plus petite valeur du caractère.

L'élément technologique θ_4 est la définition de l'étendue.

Le dernier type de tâches est T_5 : construire la représentation graphique d'une série statistique dont la tâche t_5 : construire la représentation graphique de la série « nombre de pièces de l'habitation principale » est un spécimen ; la technique τ_5 associée est :

- déterminer la nature du caractère ;
- si le caractère est quantitatif :
  - si on veut illustrer la répartition des fréquences, on construit un diagramme circulaire ;
  - si on veut illustrer le mode de la série, on construit un diagramme en barres.
- si le caractère est quantitatif discret :
  - on construit un diagramme en bâtons.
- si le caractère est quantitatif continu :
  - on construit un histogramme.

Les éléments technologiques θ_5 sont les définitions des différents diagrammes.

Le second exercice permet de retrouver dans la première question les deux premiers types de tâches de l'exercice 1, T_1 et T_2 au travers des spécimens t'_1 : déterminer le caractère étudié sur les parties de jeu de fléchettes et t'_2 : déterminer la nature du caractère du jeu de fléchettes.

Dans la question suivante, on rencontre le type de tâches T_6 : calculer la moyenne d'une série statistique dont la tâche $t_{6;1}$: calculer la moyenne des points par partie pour Patrick et la tâche

$t_{6;2}$: calculer la moyenne des points par partie de Jonathan » sont des spécimens ; la technique associée τ_6 est :

- faire la somme des points ;
- diviser par le nombre de parties.

L'élément technologique θ_6 est la définition de la moyenne.

Le troisième type de tâches T_7 est : déterminer la médiane d'une série statistique dont la tâche $t_{7;1}$: déterminer la médiane des points pour Patrick et la tâche $t_{7;2}$ déterminer la médiane des points pour Jonathan, sont des spécimens ; la technique τ_7 est :

- ranger la liste dans l'ordre croissant ;
- déterminer l'effectif total N de la série :
  - si $N = 2k$ est pair, la médiane est la demi-somme des valeurs de rang k et $k+1$;
  - si $N = 2k+1$ est impair, la médiane est la valeur de rang $k+1$.

L'élément technologique θ_7 associé est la définition de la médiane.

Le quatrième type de tâches T_8 est : faire la somme des valeurs d'une série statistique dont la tâche $t_{8;1}$: faire la somme des points de Patrick et la tâche $t_{8;2}$: faire la somme des points de Jonathan sont de spécimens ; la technique τ_8 associée est :

- multiplier la moyenne par le nombre de parties.

L'élément technologique θ_8 est la définition de la moyenne.

Le cinquième type de tâches est T_9 : calculer le nombre de parties donc le score est supérieur à un nombre donné dont la tâche t_9 : calculer le nombre de parties dont le score est supérieur à 29 ; la technique τ_9 associée est :

- ranger la liste dans l'ordre croissant ;
- déterminer le nombre de valeurs supérieures à 29.

L'élément technologique θ_9 est la comparaison de nombres et la somme de nombres.

Dans le troisième et dernier exercice, le premier type de tâches rencontré est T_{10} : calculer la moyenne d'une série dont la tâche t_{10} : calculer la moyenne de la durée de vie de composants électroniques est un spécimen ; la technique τ_{10} associée est :

- multiplier chaque valeur par son effectif ;
- en faire la somme ;
- diviser cette somme par l'effectif total.

L'élément technologique est la définition de la moyenne.

Dans la question suivante, il y a deux types de tâches ; le premier T_{11} est : déterminer la médiane d'une série statistique dont un spécimen est la tâche t_{11} : déterminer la médiane de la durée de vie des composants électroniques ; la technique τ_{11} associée est :

- déterminer l'effectif total N ;
- si N est pair ($N = 2k$), la médiane est la demi-somme des valeurs de rang k et $k+1$
- si N est impair ($N = 2k+1$), la médiane est la valeur de rang $k+1$.

L'élément technologique θ_{11} est la définition de la médiane.

Pour le second type de tâches T_{12} : déterminer les quartiles d'une série statistique dont le spécimen est la tâche t_{12} déterminer les quartiles de la durée de vie des composants électroniques ; la technique τ_{12} associée est :

- déterminer l'effectif total N ;
- calculer $\frac{N}{4}$:
  - si $\frac{N}{4}$ est entier alors le premier quartile est la valeur de rang $\frac{N}{4}$; sinon le premier quartile est la valeur du rang entier qui suit $\frac{N}{4}$.

- calculer $\frac{3N}{4}$:
  - si $\frac{3N}{4}$ est un entier, alors le troisième quartile est la valeur de rang $\frac{3N}{4}$; sinon le troisième quartile est la valeur de rang entier qui suit $\frac{3N}{4}$.

L'élément technologiques est la définition des quartiles.

Le dernier type de tâches est T_{13} : répondre par vrai ou faux en justifiant à l'aide de paramètres statistiques ; la technique τ_{13} associée est :

- déterminer si le premier quartile est la valeur proposée :
  - si oui, alors la réponse est vraie ;
  - si non, la réponse est fausse.
- refaire de même pour la médiane puis le troisième quartile.

Les éléments technologiques θ_{13} sont la définition de la médiane et des quartiles.

4.4.4 L'observation des copies

L'analyse du sujet ayant été abordée, intéressons-nous aux copies des élèves (que l'on trouvera en annexe 16). Le détail des analyses se trouve en annexe 17. Nous en faisons une synthèse ci-après.

Dans la première question, la quasi totalité des élèves trouve le caractère étudié ; un seul élève donne l'effectif des valeurs comme caractère. En ce qui concerne la question suivante, elle est moins bien réussie par les élèves. Pour certains la réponse est incomplète (trouve que c'est un caractère quantitatif mais ne précise pas discret). D'autres ne répondent pas à cette question. D'autres encore trouvent un caractère qualitatif. Pour ceux qui n'ont pas répondu à cette question, on remarque un paradoxe quant à la réponse à la question 5 (partie justification) : ils n'ont pas répondu à cette question et pourtant ils arrivent à répondre correctement à la justification du diagramme en bâtons alors que c'est exactement la même chose qui est demandé. Ce qu'on peut observer dans le cours du professeur, c'est que ce type de tâches n'est que très peu travaillé avec les élèves : pas de technique institutionnalisée, notamment.

La troisième question est bien réussie dans l'ensemble ; on remarque que peu d'élèves justifient leurs résultats en présentant la technique : le professeur ne peut pas savoir comment ses élèves ont trouvé la réponse (technique « manuelle » ou à la calculatrice ?). Ce que nous venons de voir pour le mode, l'est moins pour l'étendue (question 4) où le calcul posé par les élèves révèle (le plus souvent) qu'ils ont effectué le calcul associé à la technique. Par contre on remarque que la technique est très mal maîtrisée par les élèves (type de tâches du programme de 3^{e}) : confusion entre les valeurs de la série et les effectifs correspondants. On peut alors remarquer que la technique n'est pas institutionnalisée : le professeur ne donnant que la technologie dans le cours.

Dans la cinquième question, la partie qui concerne la justification du choix du graphique a été traité en même temps que la question 2 ; on remarque de plus que certains élèves proposent un diagramme en barres mais construisent quand même un diagramme en bâtons. La construction du diagramme en bâtons est réussie dans la quasi totalité des copies. Par contre le choix des unités de l'axe des ordonnées représente un vrai problème pour la construction exacte du diagramme (même si ce n'est pas vraiment important : l'allure générale du diagramme donne une assez bonne idée de la distribution des valeurs). On peut aussi imaginer que la construction des axes avec une unité appropriée fasse partie d'un type de tâche qui est abordé dans un autre contexte, par exemple celui de la représentation des fonctions.

Le second exercice reprend la même question que dans le premier exercice. Ici encore on trouve quelques élèves qui confondent le caractère étudié avec l'effectif des valeurs. La partie de la question qui traite de la nature du caractère étudié appelle les mêmes commentaires que dans l'exercice 1. En revanche, on ne retrouve pas le paradoxe évoqué dans le premier exercice. La seconde question demande le calcul des moyennes, ce qui est en grande majorité fait. On retrouve le problème des techniques non apparentes : que ce soit la technique « manuelle » ou à la calculatrice (cette dernière est forcément peu apparente puisque tout se passe sur la calculatrice de l'élève ; mais l'évoquer sur la copie est une trace qui en principe devrait apparaître). En revanche la comparaison (explicite) des moyennes est très rarement faite par les élèves ; le professeur peut supposer qu'elle a été faite par rapport à la réponse apportée, lorsque celle-ci est juste. Ici encore la technique du type de tâches « Comparer deux séries » n'est pas institutionnalisée voire peu abordée dans le cours du professeur.

La troisième question demande le calcul des médianes et pose des problèmes aux élèves ; on trouve des techniques incomplètes chez bon nombre d'entre eux (même si le calcul aboutit à un résultat juste parfois) ou des techniques inversées (technique du cas pair dans le cas impair et inversement) : ici encore des élèves ne montrent pas la technique qu'ils utilisent. Quelques élèves déterminent la technique « à la main » (voir ci-dessous un exemple)

$$\overbrace{4\ 5\ 6\ 10}^{4}\ 13\ \overbrace{14\ 16\ 19\ 37}^{4}$$

La comparaison des médianes est aussi peu visible que celle des moyennes.

Dans la question suivante, beaucoup d'élèves refont la somme des scores (ne consomme pas trop de temps, il n'y a que 10 valeurs) sans voir le lien avec la moyenne :

$$\sum_{i=1}^{10} x_i = N \times \overline{x}$$

Peu d'élèves montrent la comparaison des sommes (en fait des moyennes) comme dans la deuxième question. La dernière question de l'exercice a été assez bien réussie par les élèves ; ils montrent assez bien l'égalité des scores supérieurs à 29 et en déduisent l'égalité entre Patrick et Jonathan. On pourrait s'interroger sur la manière de départager les deux joueurs : mais cette question n'est pas abordée dans l'exercice.

Le dernier exercice du sujet comporte l'utilisation de la calculatrice. Dans la première question, on remarque beaucoup d'erreurs sur la technique de calcul de la moyenne. Le problème de l'effectif total (noté 200 dans l'exercice alors que la somme fait 201 induit beaucoup d'erreurs chez les élèves ; ceci n'aura pas la même incidence sur le calcul des quartiles, par exemple. La deuxième question appelle les mêmes commentaires pour le calcul de la médiane. La dernière question a été peu souvent abordé par les élèves ; nombre d'entre eux donnent leurs réponses sans justification et sans faire le lien avec les paramètres calculés précédemment.

Chapitre 5

L'analyse des observations

5.1 La transposition didactique dans les instructions officielles et les manuels : le savoir à enseigner ?

Que dire du savoir à enseigner tel qu'il apparaît dans ces documents officiels ? Nous l'avons vu, les instructions officielles tentent de reprendre à leur compte différents ingrédients que la Didactique a mis en exergue : problématique de la statistique, les différentes étapes d'analyse d'un questionnement statistique en ne négligeant ni l'aspect mathématique, ni l'aspect extramathématique ; malheureusement, le tout est encore un peu flou et le professeur, malgré sa bonne volonté, souffre cruellement d'un manque de connaissances, qu'elles soient mathématiques ou didactiques.

Le programme, et les documents ressources qui s'y rapportent, ont essayé de prendre à leur compte la problématique statistique. Malheureusement, plusieurs facteurs font que cette tentative reste vaine même si des progrès sont à noter. Les questions génératrices ne sont pas une priorité des instructions officielles, elles arrivent même comme justification, après coup, des différents traitements statistiques que l'on aura pu effectuer avant. Or la question génératrice devrait être la base de toute étude statistique. La nécessité de s'intéresser à des domaines de réalité extramathématique n'est soulevé que de manière indirecte ; en effet le programme précise qu' « en demandant de s'interroger sur la signification des nombres utilisés, sur l'information apportés par un résumé statistique, sur les risques d'erreur d'interprétation et sur leurs conséquences possibles, y compris dans la vie courante, cette partie des mathématiques contribue à former de jeunes adultes capables de comprendre les enjeux et débats de la société où ils vivent. » (MEN, 2007a, p.15). Force est de constater que les enjeux et débats de la société où les élèves vivent sont relativement limités ; limités à leur univers scolaire ou leur environnement personnel direct (température, loisirs, etc.). Quel autre DREM avons-nous relevé, qui puisse les former ? A vrai dire pas grand chose, et seuls quelques exercices proposent de s'intéresser à des articles de journaux ou publications. L'échec est cuisant.

Un autre constat plutôt négatif est celui des thèmes de convergence. Les deux professeurs de 3^e^ que nous avons observé n'y ont pas fait de référence explicite. Seul un des deux professeurs fait allusion à la météorologie dans un de ces devoirs ; mais l'étude statistique qui est proposée se limite à des calculs de paramètres et une comparaison, tout comme nous pouvons le trouver dans le programme de 3^e^ : « Les mathématiques trouvent dans la météorologie des possibilités d'application tout à fait intéressantes. A partir de relevés de mesures, l'élève s'investit dans la construction de graphiques, l'utilisation des nombres relatifs, le calcul de moyennes... » (MEN, 2008, p. 7) ; nous sommes loin des ambitions du programme et retombons systématiquement dans la réduction calculatoire de la statistique.

Que peut-on dire de la première approche de la dispersion telle qu'elle est présentée dans le programme ? Malheureusement, le déni de variabilité est toujours aussi présent. Aucun des manuels que nous avons observé ne parle de variabilité, tout est fait comme si les grandeurs qui sont proposées à l'étude étaient fixes. Chevallard et Wozniak l'avait déjà relevé en 2003 à l'école d'été de didactique et nous pouvons constater que c'est toujours d'actualité. Les types de tâches relevés dans l'organisation praxéologique de ces manuels sont sans équivoque : les études proposées ne sont pas des études statistiques au sens où la didactique l'entend. Or le programme dans sa version de 2007 le notait clairement : « Exprimer et exploiter les résultats de mesures d'une grandeur » (MEN, 2007a, p. 35) dans la colonne *Capacités* et « La notion de dispersion est à relier, sur des exemples, au problème posé par la disparité des mesures d'une grandeur, lors d'une activité expérimentale, en particulier en physique et chimie » (*op. cit.*) dans la colonne *Exemples d'activités, commentaires*. Or ce qui précède est repris non pas dans le programme, mais dans la partie introduction du programme de 2008, que nous reproduisons :

> **Le problème de la variabilité de la mesure**
>
> De nombreuses activités dans les disciplines expérimentales (physique-chimie, sciences de la vie et de la Terre, technologie), basées sur des mesures, doivent intégrer la notion d'incertitude dans l'acte de mesurer et développer l'analyse des séries de mesures. Lors de manipulations, les élèves constatent que certaines grandeurs sont définies avec une certaine imprécision, que d'autres peuvent légèrement varier en fonction de paramètres physiques non maîtrisés. Plusieurs mesures indépendantes d'une même grandeur permettent ainsi la mise en évidence de la dispersion naturelle des mesures. Sans pour autant aborder les justifications théoriques réservées au niveau du lycée, il est indispensable de faire constater cette dispersion d'une série de mesures et d'estimer, en règle générale, la grandeur à mesurer par la moyenne de cette série. (p. 6)

Au passage, on notera que les mathématiques ne font pas partie des sciences expérimentales.[1] Le programme marque un appui fort sur la dispersion des mesures qu'il est *indispensable* que les élèves rencontrent. Vœu pieux s'il en est, car les manuels et les professeurs passent sans s'y arrêter. De ce point de vue, le programme a échoué. Les exemples donnés dans les documents ressources ne sont pas plus explicites sur ce sujet, la variabilité est évoqué mais les auteurs s'empressent de souligner qu' « il est possible aussi d'utiliser des données réelles directement fournies » (MEN, 2007b, p.1). Cela pourrait être intéressant si on se posait des questions adéquates à leur propos. Nous l'avons vu, aucun des professeurs n'est allé chercher ces données réelles directement disponible. Tout au plus, les professeurs s'intéressent-ils aux données qu'ils peuvent obtenir sans problème : les données qui concernent leurs élèves. Ainsi, on range au placard une phase importante du travail statistique : le recueil des données et les problèmes qu'il peut poser.

5.2 La transposition didactique dans le cours du professeur : le savoir enseigné

L'analyse des cours des professeurs que nous avons esquissé dans le chapitre précédent montre que la problématique statistique n'est pas connue des professeurs. Même si pour certains l'étude part d'une question, ce qui est plutôt positif, cette question ne mène pas à une interrogation sur la dialectique régularité - variabilité telle qu'elle devrait être présente. Les données utilisées sont trop souvent ramenés à l'univers des élèves, que ce soit l'univers purement scolaire ou leur univers personnel (taille, loisirs, etc.)

1. À ce sujet, on pourra voir *Le caractère expérimental de l'activité mathématique* (Chevallard, 1992)

Les types de tâches abordés sont conformes au programme des classes dont nous avons observé les traces écrites. Cependant, la tentation d'aller voir après est grande ; tout comme les manuels, le diagramme en boîte apparaît rapidement comme moyen de comparer deux séries par exemple. Mais nous l'avons vu, ces types de tâches sont insuffisamment travaillé (surtout du point de vue de la technique) et les évaluations à ce sujet sont en générales rarement abordées et même réussies par les élèves. C'est un fait qui nous interroge : comment le professeur peut-il travailler davantage ce type de tâches pour que la technique qui doit en émerger puisse être efficace ? D'ailleurs, est-ce qu'il n'y a qu'une technique possible et quels en sont les éléments technologiques ?

Mais revenons au savoir enseigné. Nous avons observé le cours des quatre professeurs et les premières remarques concernaient les questions génératrices qui sont encore trop peu présentes dans le savoir enseigné. Intéressons-nous plus particulièrement au domaine de réalité extramathématique. Les situations que nous avons étudié étaient soit rattaché à l'univers scolaire, soit rattaché à l'univers personnel des élèves. Pour d'autres, même si un effort a été fourni, le DREM n'est pas assez exploité, voire même ne sert que de décor à une étude statistique qui n'en aura que le nom. Chassez le naturel, il revient au galop. Alors que faire ? L'étude de ce problème par la Didactique devrait permettre de faire des progrès. Nous l'avons déjà soulevé, la diffusion de la Didactique est un problème.

5.3 La transposition didactique dans les copies : le savoir appris

Que nous apprend l'observation des copies de élèves ? Un premier élément que nous relevons est le calcul de la médiane. Que ce soit en 3^{e} où ce type de tâches est nouveau ou en 2de où son apprentissage est censé être consolidé, les élèves montrent assez rapidement qu'ils ne maîtrisent qu'assez mal cette technique. Pourtant la technique employée par les professeurs est correcte la plupart du temps. Est-ce que l'organisation de l'étude est mal pensée ? Nous ne le pensons pas ; cependant la confusion des techniques dans le cas où l'effectif total de la série est pair et où il est impair pose question. Est-ce que ce que retiennent les élèves de la technologie de la médiane est un obstacle à la technique qu'elle est censée justifiée ? En effet, nous l'avons vu, les élèves retiennent plus facilement que la médiane est la valeur qui sépare la série en deux que la véritable technologie qui va produire et justifier la technique. Pourtant, la détermination des quartiles posent sensiblement moins de problème. La technique de détermination est unique, ce qui explique davantage, à notre avis, que la détermination de la médiane soit si problématique.

Un second élément est lié aux techniques « transparentes » ; nous appelons ainsi des techniques que les élèves devraient produire sur les copies, que nous savons (ou parfois supposons avec suffisamment de certitude pour le considéré comme tel) qu'ils ont utilisé mais qui n'apparaissent pas de façon explicite. Or le problème qui apparaît est que la technique transparente devient une technique du cours. Le type de tâches comparer deux séries en est l'exemple le plus frappant. Un seul des professeurs a institutionnalisé une technique pour comparer deux séries à partir de la comparaison des diagrammes en boîte et de l'écart interquartile. Pour les autres, ce type de tâches est souvent évalué mais aucune technique n'a été institutionnalisé dans le cours. Le travail de la technique est réduit à un ou deux exemples dans la partie *Exercices* du cours. Comment les élèves vont-ils pouvoir appliquer une technique qu'ils n'ont pas vu ? La question qui se pose immédiatement alors : est-ce que ces techniques transparentes sont plus présentes dans ce domaine de la statistique que dans d'autres domaines des mathématiques ? Nous laisserons cette question en suspens.

Conclusion

Notre question de recherche s'orientait au départ sur une comparaison entre les différences que l'on peut observer entre les classes de troisième et de seconde concernant l'enseignement de la statistique. Notre étude a fait évoluer cette question vers l'étude de la transposition didactique de la statistique telle qu'on peut la trouver dans ces deux classes au travers des traces écrites des élèves de quatre professeurs. Nous nous sommes appuyés dans un premier temps, sur l'observation des praxéologies présentes dans les programmes et les documents ressources, et avons essayé de mettre en évidence la sous-estimation de la variabilité ainsi que le manque de prise en compte de la problématique de la statistique. Même si les documents officiels abordent ces thématiques, elles sont souvent peu mises en avant et un professeur pourrait avoir du mal à y voir clair. L'exemple des questions génératrices en est un exemple frappant : la question intervenant à un moment tardif dans l'étude statistique.

Dans un deuxième temps, nous avons observé les praxéologies de six manuels scolaires pour y relever la trop grande place prise par les calculs ; cela nous a conduit à considérer les domaines de réalités extramathématiques, que nous avons trouvé souvent rattachés à l'univers scolaire ou à l'univers des élèves (leurs tailles, leurs loisirs, etc.). La déterritorialisation que présentait Chevallard et plus que jamais à l'œuvre : les DREM sont trop souvent utilisé comme justification après-coup de l'étude statistique ou alors comme simple décor. Il reste un travail à faire pour que les manuels puissent s'ouvrir à une véritable prise en compte de la variabilité et de la problématique de la statistique.

Dans un troisième temps, nous avons considéré les cours de quatre professeurs de mathématiques qui enseignent en classe de troisième et de seconde et avons relevé l'absence, ou tout au moins le manque de visibilité, de la problématique de la mesure des grandeurs dans leurs cours. Certaines techniques prennent une place non négligeable dans l'enseignement au détriment d'autres qui sont peu abordées voire passées sous silence.

Enfin, nous avons étudié les copies des élèves de ces quatre professeurs et avons observé les effets de l'enseignement de la statistique tel qu'il a été proposé aux élèves et ce qu'ils en ont retenu. Certaines techniques qui sont sensées être maîtrisées à l'issue de la classe de 3^{e} ne le sont pas encore à l'issue de la classe de seconde. L'échec de la détermination de la médiane par ces élèves en est l'exemple le plus significatif. Nous avons observé aussi des techniques, que nous avons appelées *transparentes*, qui étaient demandées aux élèves mais qui n'apparaissent pas dans le cours du professeur.

Nous avons limité notre étude de l'enseignement de la statistique à l'enseignement français. Il nous faudrait aller voir dans d'autres pays si l'enseignement de la statistique est identique à celui que nous avons observé ou tout au moins si les conditions et les contraintes qu'a relevé Floriane Wozniak sur la statistique, telle qu'on peut la voir en France, sont encore présentes. Par ailleurs, les analyses de Wozniak se trouvent confortés dans nos propres analyses. En effet, les conditions et contraintes relevées en 2005, sont toujours d'actualité.

Guy Brousseau écrivait dans *Alternatives en didactique de la statistique* (Brousseau, 2009) :

> Le travail spontané de transposition didactique est lent, chaotique et il procure un accès assez sélectif aux connaissances. Les connaissances de statistiques doivent être vite et mieux diffusées. Les solutions à ce problème didactique sont actuellement recherchées dans des sciences qui apportent des renseignements intéressants mais très indirects comme la psychologie ou les neurosciences et qui ne prennent pas le processus spécifique de la création ou de la recréation d'une connaissance précise comme objet d'études théoriques et expérimentales. On peut espérer que la Didactique, en évitant les inférences douteuses auxquelles ces approches latérales nous condamnent, contribuera à améliorer la diffusion des la statistique. (p. 8)

Nous avons pu constater que cette diffusion ne s'est toujours pas opérée et qu'il reste du chemin à parcourir pour que l'enseignement de la statistique ressemble à une véritable étude statistique. Cela pose le problème de la diffusion de la didactique au sein de la profession.

Toutes ces observations ouvrent la voie à d'autres questions : est-ce que les techniques transparentes sont aussi des traits caractéristiques de l'enseignement de la géométrie ou des fonctions, par exemple, où est-ce que la statistique en est un révélateur ? Quelles sont les praxéologies utilisées par les autres professeurs (c'est-à-dire autres que les professeurs de mathématiques) ? Sont-elles aussi exsangues que celles que nous avons pu observer ? Constituent-elles un obstacle à l'enseignement de la statistique dans la classe de mathématiques ?

Bibliographie

ALEIXANDRE, D. et AL. (2012). *Mathématiques collection zénius 3^e^*. Magnard. Chap. 10, p. 183–204.

ANTIBI, A. et al. (2010). *Math 2^de^. Travailler en confiance*. Nathan.

ARNOULT, J. et AL. (2010). *Maths 2^de^*. Sous la dir. de Claude DESCHAMPS. Belin. Chap. 6, p. 130–157.

BARRA, R. et AL. (2010). *Transmath 2^de^*. Nathan. Chap. 5 - 6, p. 110–143.

BRIAND, J. (1976). « Découverte des lois du hasard à l'école élémentaire ». Dans : *PLOT* 3.

— (2005). « Une expérience statistique et une première approche des lois du hasard au lycée par une confrontation avec une machine simple ». Dans : *Recherche en Didactique des Mathématiques* 25.2.

BROUSSEAU, G. (1974). « Description des 31 leçons expérimentées à l'école J. Michelet à Talence ». Dans : *L'enseignement des Probabilités et des Statistiques. Compte-rendu de la 26e rencontre de la CIEAEM.*

— (1998). *Théorie des situations didactiques. (Didactique des mathématiques 1970-1990)*. Grenoble : La Pensée Sauvage.

— (2003). « Situations fondamentales et processus génétiques de la statistique ». Dans : *Balises en didactique des mathématiques*. Sous la dir. d'A. MERCIER et C. MARGOLINAS. Grenoble : La Pensée Sauvage, p. 165–194.

— (2008). *NOTES sur l'enseignement des statistiques et/ou des probabilités dans la scolarité commune (6-14 ans)*. URL : `http://guy-brousseau.com/1517/notes-sur-l%E2%80%99enseignement-des-statistiques-etou-des-probabilites-dans-la-scolarite-commune-6-14ans-2001-2008/`.

— (2009). « Alternatives en didactique de la statistique ». Dans : *(Inédit)*. URL : `http://guy-brousseau.com/1525/alternatives-en-didactique-de-la-statistique-2009/`.

CHAPIRON, G. et AL. (2012). *Triangle 3^e^*. Hatier. Chap. 8, p. 147–164.

CHEVALLARD, Y. (1978). *Notes pour la didactique de la statistique*. IREM Aix-Marseille.

— (1985). *La transposition didactique - Du savoir savant au savoir enseigné*. deuxième édition augmentée, 1991. Grenoble : La Pensée Sauvage.

— (1992). « Le caractère expérimental de l'activité mathématique ». Dans : *Petit x* 30.

— (1998). « Analyse des pratiques enseignantes et didactique des mathématiques : l'approche anthropologique ». Dans : *Analyse des pratiques enseignantes et didactique des mathématiques*. Sous la dir. de R. NOIRFALISE. IREM Clermont-Ferrand, p. 91–120.

— (2002a). « Organiser l'étude : 1. Structures et fonctions ». Dans : *Actes de la XIe école d'été de didactique des mathématiques*. Sous la dir. de J-L. DORIER et al. Grenoble : La Pensée Sauvage, p. 3–32.

Chevallard, Y. (2002b). « Organiser l'étude : 3. Ecologie et régulation ». Dans : *Actes de la XIe école d'été de didactique des mathématiques*. Sous la dir. de J-L. Dorier et al. Grenoble : La Pensée Sauvage, p. 41–56.

Chevallard, Y. et F. Wozniak (2003). *Enseigner la statistique dans le secondaire : des mathématiques mixtes pour penser la variabilité*. url : `http://yves.chevallard.free.fr/spip/spip/rubrique.php3?id_rubrique=4`.

— (2003a). « Enseigner la statistique au secondaire. Entre genre prochain et différence spécifique ». Dans : *Balises pour la didactique des mathématiques*. Sous la dir. d'A. Mercier et C. Margolinas. Grenoble : La Pensée Sauvage, p. 195–218.

— (2007). « Enseigner la statistique : un problème de la profession ». Dans : *Actes du XIVe colloque de la CORFEM*, p. 13–30.

Choquer-Raoult, A. et al. (2010). *Maths repères* 2^{de}. Hachette Education. Chap. 4, p. 128–165.

Du Roy, A. et al. (2012). *Nouveau prisme* 3^{e}. Sous la dir. de Nadine Jacob. Belin. Chap. 9, p. 158–175.

Mathieu-Wozniak, F. (2005). « Conditions et contraintes de l'enseignement de la statistique en classe de seconde générale. Un repérage didactique ». Thèse d'Etat. Université Claude Bernard - Lyon 1.

MEN (2007a). *Programme de mathématiques de la classe de* 3^{e}. Paris : CNDP.

— (2007b). *Ressources pour les classes de 6e, 5e, 4e, et 3e du collège. Organisation et gestion des données au Collège*. url : `http://eduscol.education.fr/cid45766/mathematiques-pour-le-college-et-le-lycee.html#lien1`.

— (2008). *Programme du collège. Programme de l'enseignement de mathématiques*. url : `http://www.education.gouv.fr/cid22120/mene0817023a.html`.

— (2009a). *Programme de mathématiques de la classe de seconde générale et technologique*. Paris : CNDP.

— (2009b). *Ressources pour la classe de seconde. Probabilités et Statistiques*. url : `http://eduscol.education.fr/cid45766/mathematiques-pour-le-college-et-le-lycee.html#lien0`.

Pazat, N. (2002). « Initiation aux statistiques inférentielles en classe de seconde ». Mém.de maîtr. IUFM d'Aquitaine.

Annexes

On trouvera ces annexes sur le CD associé à ce mémoire.

Annexe 1 : copies des élèves de LC (1^{re} partie) ;

Annexe 2 : analyses des copies des élèves de LC (1^{re} partie) ;

Annexe 3 : copies de élèves de LC (2^{e} partie) ;

Annexe 4 : analyses des copies des élèves de LC (2^{e} partie) ;

Annexe 5 : cours de LC (2 cahiers d'élèves) ;

Annexe 6 : copies des élèves de KB (DM) ;

Annexe 7 : analyses des copies des élèves de KB (DM) ;

Annexe 8 : copies des élèves de KB (DS) ;

Annexe 9 : analyses des copies des élèves de KB (DS) ;

Annexe 10 : cours de KB (2 cahiers d'élèves) ;

Annexe 11 : copies des élèves de PN ;

Annexe 12 : analyses des copies des élèves de PN ;

Annexe 13 : cours de PN (2 cahiers d'élèves) ;

Annexe 14 : copies des élèves de YC (DM) ;

Annexe 15 : analyses des copies des élèves de YC (DM) ;

Annexe 16 : copies des élèves de YC (DS) ;

Annexe 17 : analyses des copies des élèves de YC (DS) ;

Annexe 18 : cours de YC (1 cahier d'élèves).