

## LA COTE STANDARD COMME OBJET POUR RÉFLÉCHIR AUX NOTIONS STATISTIQUES AVEC LES FUTURS MAÎTRES

Stéphane CYR  
Université du Québec à Montréal  
Lucie DEBLOIS  
Université Laval

**Résumé :** Cette étude vise à identifier la compréhension des futurs maîtres de mathématiques à l'égard des formules impliquées dans l'apprentissage de certains objets statistiques. Pour ce faire, un premier groupe de 31 futurs maîtres ayant complété 3 des 4 années de leur formation initiale universitaire a répondu à un questionnaire sur le sujet. Un deuxième groupe de 63 futurs maîtres du même niveau a répondu à un autre questionnaire. Nos résultats laissent apparaître des difficultés importantes chez certains d'entre eux, difficultés pouvant avoir une incidence sur l'intervention en classe du secondaire, notamment lors de l'enseignement de la cote standard et éventuellement du coefficient de corrélation. Le but de cet article est de proposer une réflexion qui vise à analyser les difficultés observées.

**Mots-clés :** Statistique, Cote standard, Corrélation, Enseignement, Futurs maîtres.

### Introduction

Plusieurs travaux sur la didactique de la statistique font état, chez les élèves, d'un apprentissage en statistique orienté vers la maîtrise de techniques de calcul (Bakker, 2004). Les travaux de Gattuso (1997) sur la moyenne ont démontré l'effet pervers d'un enseignement dirigé vers l'application mécanique de formules toute faites. Elle a, entre autres, constaté que le fait de connaître et d'appliquer la formule de la moyenne n'est pas suffisant pour assurer une bonne compréhension du concept visé et ce, autant pour des élèves du secondaire que pour des futurs maîtres. Par exemple, à 15 ans, les élèves appliquent depuis longtemps une procédure numérique pour calculer des moyennes. Pourtant, très peu d'entre eux sont en mesure d'expliquer de façon précise ce qu'est la moyenne et en donner une définition exacte. En outre, des applications erronées de la moyenne apparaissent dans le cas de situations plus complexes.

Soulignant l'intérêt de se centrer d'abord sur le sens, Boyé et Comairas (2002) précisent que : " l'enseignement des statistiques ne se résumera pas à apprendre des formules et à les appliquer " (p.37). Ils reconnaissent que les élèves accumulent trop de formules sans signification pour eux, ce qui les conduit à considérer les statistiques comme une application d'algorithmes. Duperret (2001) mentionne que de telles applications mécaniques en statistique, non fondées sur une compréhension du sens, conduisent plus souvent à questionner " l'intérêt donné à l'enseignement des statistiques

si on ne mesure pas son rôle de formation scientifique et sociale, et s'il se réduit à quelques recettes " (2001, 9).

Nous présentons les résultats d'une étude exploratoire portant sur la compréhension de certains concepts de statistiques auprès de futurs maîtres de mathématiques du secondaire. En fait, nous cherchons à répondre à la question suivante : Comment les futurs maîtres parlent-ils de leurs connaissances à propos de certaines notions de statistiques et de leur enseignement durant les cours de didactique à l'université ? Plus particulièrement, nous avons cherché à cerner les significations (Brown *et al.*, 1989) que les futurs maîtres du secondaire accordent au sens, au rôle et à l'utilité des formules rencontrées dans l'apprentissage de ces objets statistiques tels que l'écart type et la cote standard.

## **1. Cadre théorique pour notre étude**

### **1.1 Les postures épistémologiques des futurs maîtres**

Pour appuyer nos observations et nos analyses, nous nous basons sur un cadre théorique qui prend en compte le jeu entre trois postures épistémologiques adoptées par les futurs maîtres durant une formation : la position de l'ancien élève, celle de l'étudiante ou de l'étudiant universitaire et celle de l'enseignante ou de l'enseignant (DeBlois et Squalli, 2002). Attardons-nous à chacune de ces postures.

L'ancien élève a l'habitude de travailler sur des exercices ou des problèmes pour lesquels l'enseignant ou l'enseignante détient les réponses, ce qui pourrait expliquer la recherche d'un 'bon' modèle d'enseignement. Cette posture conduirait à concevoir que l'apprentissage relève d'un processus 'transparent' : les élèves connaissent, apprennent, interprètent les connaissances à partir du moment où elles sont présentées clairement.. L'apprentissage des élèves n'ayant pas nécessairement lieu au même moment pour tous, cette posture risque d'entretenir des conceptions selon lesquelles, les élèves s'en tireront puisque étant passés par là, elle ou ils s'en sont finalement bien sortis. De plus, le risque est grand de réduire l'évaluation de l'apprentissage à la conformité à l'égard des connaissances attendues ou au résultat obtenu à une tâche. Le futur maître attend alors les signes d'un apprentissage sans considération du sens, du processus ou de la cohérence de ces signes chez l'élève. C'est "l'effet Jourdain" de Brousseau (1986).

Durant leur formation à l'enseignement, les futurs maîtres réalisent une variété d'activités dans le but de discuter, entre autres, la nature d'un concept mathématique, le développement de la pensée de l'élève, la relation qui se développe entre l'élève et le savoir mathématique, les dispositifs d'enseignement et, par conséquent, le rôle à jouer dans la classe. La posture de l'étudiant universitaire fait intervenir des connaissances spécifiques comme la manipulation de matériel didactique ou technologique, le travail coopératif, une réflexion sur les connaissances des élèves et sur leurs conceptions. La posture de l'étudiant universitaire pourrait conduire à considérer les connaissances didactiques de façon procédurale, croyant ainsi que ces dernières provoqueront nécessairement l'apprentissage des élèves. À nouveau, l'apprentissage des élèves d'une classe du secondaire n'ayant pas forcément lieu de la même manière et au même moment pour tous, un sentiment d'inhabileté pourrait conduire le futur maître à se

décharger de la responsabilité de faire un choix qui soit juste au moment d'une prise de décision en classe plutôt que de susciter un questionnement relatif aux raisons de cet échec.

Enfin, comme stagiaire dans une classe, le futur maître est invité à se préoccuper non seulement de stratégies d'enseignement, mais aussi de l'apprentissage de ses élèves et de leurs besoins socio-affectifs (Westerman, 1991). Il se situe alors dans une posture d'enseignant. Cette posture est caractérisée par la prise en charge de la responsabilité de la situation d'enseignement/apprentissage. Ainsi, lorsque le futur maître identifie ses intentions et qu'il analyse sa pratique, il endosse la responsabilité d'un enseignant pour se projeter dans sa pratique professionnelle. On pourrait parler en quelque sorte de la dévolution de la situation d'enseignement. Les difficultés observées chez les élèves d'une classe du secondaire pourraient conduire le stagiaire à utiliser les définitions et les axiomes comme des arguments d'autorité pour justifier les savoirs mathématiques impliqués dans la situation d'enseignement-apprentissage.

En conséquence, le futur maître est à la fois un ancien élève en mathématiques, une étudiante ou un étudiant universitaire en enseignement et une enseignante ou un enseignant stagiaire. Une étude a permis de reconnaître que, confronté au fait que les questions concernant des problèmes d'enseignement ou d'apprentissage n'ont pas nécessairement de réponses " toutes faites ", un écart entre la posture de l'ancien élève et de l'étudiant universitaire s'installe. Selon DeBlois et Squalli (2002), ce décalage crée une tension pouvant conduire à privilégier la posture de l'ancien élève, plus connue, contribuant ainsi à reproduire des stratégies d'interventions en classe du secondaire sans les discuter. C'est à ce décalage entre ces deux premières postures que notre étude s'intéresse puisqu'au moment où les futurs maîtres répondent à notre questionnaire, ils ne sont pas encore stagiaires en classe. Il nous semble important de mieux cerner les connaissances développées par les futurs maîtres à l'égard des concepts mathématiques pour prévoir des stratégies d'intervention qui favorisent non seulement l'apprentissage des mathématiques au primaire ou au secondaire, mais aussi l'appropriation des connaissances didactiques pertinentes.

## 1.2 La cote standard : outil ou objet ?

Dans la plupart des manuels de statistiques servant à la formation des maîtres en mathématiques au Québec<sup>1</sup> (Casella & Roger, 1990; Christensen, 1986) la cote standard est présentée comme un outil. Introduite parallèlement au concept de la courbe normale et aux calculs des probabilités de la distribution, cet outil sert à trouver la distance (en unité d'écart type) entre une donnée dans un échantillon et la moyenne de ce même échantillon. À l'aide de cette mesure, il est possible de calculer la probabilité qu'une donnée puisse se situer à une certaine distance de la moyenne. Toutefois, très vite, on

---

<sup>1</sup> En France, le terme *cote standard* n'apparaît pas explicitement dans les programmes. Par contre, cette notion est utilisée pour effectuer une transformation linéaire qui permet d'obtenir un résultat exprimant une *variable centrée réduite*. Plus formellement, elle est définie de la façon suivante : Considérons un échantillon de taille  $n$ , une variable aléatoire  $X$  et ses réalisations  $x_i$ . On appelle cote standard la quantité

$$z_{x_i} = \frac{x_i - \bar{x}}{s_x}$$

dans laquelle  $s_x$  est l'écart type de l'échantillon.

passer aux tests d'hypothèses et à la comparaison de moyennes tirées de distributions différentes. Dans ce dernier cas, la cote standard est généralement présentée de la façon suivante : on applique à chaque donnée  $x$  d'une distribution les opérations suivantes  $(x - \bar{x})/s_x$  (moyenne de 0 et écart type de 1).

L'idée de cote standard se trouve, en quelque sorte, noyée sous les concepts de courbe normale, de calculs de probabilités et de tests d'hypothèses. Or, à travers de telles explications, les futurs maîtres passent à côté d'un sens et d'un rôle de la cote  $Z$  que nous considérons essentiel : *comparer* deux données issues de distributions différentes. D'une part, appliquée à des distributions différentes, les données issues de ces distributions peuvent être comparées dans le but d'évaluer s'il y a présence ou non d'écart significatif. D'autre part, si les façons d'aborder la cote standard dans les manuels de statistiques comme un outil et non comme un objet sont parfois suffisantes pour des étudiantes et des étudiants universitaires, elles exigent des connaissances particulières qui ne sont pas abordées par nos élèves du secondaire (15-16 ans)<sup>2</sup>.

En effet, les opérations mathématiques réalisées sur des courbes normales sont absentes des programmes d'études secondaires. Il devient ainsi difficile pour nos étudiantes et nos étudiants universitaires de donner un sens à la cote standard à travers l'étude de distributions associées à des courbes normales. De plus, au Québec la cote standard apparaît explicitement comme un concept à étudier dans les programmes du secondaire. Elle n'est donc pas présentée comme un outil utile à des analyses plus complexes en statistique, mais comme un objet d'étude en soi.

Afin de clarifier le concept de cote standard, rappelons brièvement notre analyse conceptuelle (Cyr et DeBlois, 2002). La cote standard, comme *outil statistique*, permet entre autres, de standardiser des écarts pour comparer ces derniers lorsqu'ils sont issus de distributions différentes. Elle a ainsi comme fonction de rendre comparable des données prélevées dans des distributions différentes, transformant ainsi cet outil en objet conceptuel (Douady, 1986). Nous cherchons donc d'abord à répondre aux deux questions suivantes : Quels contextes pourraient être favorables à l'apprentissage de ce concept au secondaire ? Comment aborder ce concept et sa formule au niveau secondaire ?

Nous avons choisi d'étudier la consommation en essence de différents véhicules en ville et sur l'autoroute (litres consommés/100 km). Ayant en tête une préoccupation environnementale, nous cherchons à évaluer l'efficacité d'une voiture en comparant son rendement en ville et sur l'autoroute mais également, en le comparant aux autres véhicules.

---

<sup>2</sup> Dans le programme de mathématiques de cinquième secondaire de sciences naturelles (programme menant à des études supérieures en sciences), on écrit : « Calcul des mesures de tendance centrale, de position et de dispersion » (MEQ, 2004, p. 110). Les élèves ne font alors que calculer des cotes standards afin d'obtenir une mesure de position.

Types de voiture <sup>3</sup>	En ville	Écart à la moyenne	Sur l'autoroute	Écart à la moyenne
Intermédiaire hybride (manuelle)	4	6	4,2	2,9
Compacte (manuelle)	7,6	2,4	5,9	1,2
Familiale (automatique)	8,3	1,7	6,4	0,7
Intermédiaire (manuelle)	10,7	0,7	7	0,1
Grande berline (automatique)	11,8	1,1	7,0	0
Mini-fourgonnette (automatique)	12,2	2,2	8,2	1,1
Véhicule utilitaire 4x4 (automatique)	15,5	5,5	11	3,9

Dans ce tableau, nous constatons que l'intermédiaire hybride est la voiture qui possède la plus faible consommation en ville et sur l'autoroute. Toutefois, ce que nous cherchons à évaluer est l'efficacité d'un véhicule en ville et sur l'autoroute en comparaison avec les autres véhicules. Par exemple, nous savons que la voiture compacte (manuelle) possède une meilleure consommation sur l'autoroute (5,9l /100km) qu'en ville (7,6l /100km). Toutefois, par rapport aux autres véhicules, est-ce que son efficacité sur l'autoroute est meilleure qu'en ville ? Pour répondre à cette question, nous pouvons tout d'abord comparer la consommation d'essence de ce véhicule à la moyenne de consommation des autres véhicules en ville et sur l'autoroute.

La consommation moyenne des véhicules en ville est de 10,01l /100km alors qu'elle est de 7,1l /100km sur l'autoroute. La voiture compacte a donc un écart à la moyenne en ville de 2,41 (elle consomme 2,41 litres de moins par 100 km en ville que la moyenne des autres véhicules) et de 1,2 sur l'autoroute (elle consomme 1,2 litre de moins par 100 km sur l'autoroute que la moyenne des autres véhicules). Une première interprétation de ces écarts nous pousse à croire que la voiture compacte est plus efficace en ville que sur l'autoroute en comparaison avec les autres véhicules, du fait que son écart à la moyenne est plus grand en ville. Autrement dit, elle se démarque davantage des autres véhicules en ville que sur l'autoroute. Toutefois, nous ne pouvons arrêter notre analyse à ce seul calcul de l'écart à la moyenne compte tenu du fait que la moyenne de consommation en ville et la moyenne de consommation sur l'autoroute ne sont pas les mêmes. En effet, l'écart entre la consommation des différents véhicules est plus grand en ville. Nous devons en plus, tenir compte de l'écart des autres données par rapport à chacune des moyennes. C'est en quelque sorte ce que nous permet d'obtenir l'écart type. Nous pouvons d'ailleurs confirmer cet état de fait en calculant l'écart type de chacune des distributions.

Pour la consommation en ville, nous obtenons un écart type de 3,45 alors qu'il est de 2,3 sur l'autoroute. Il est donc moins surprenant d'avoir un écart à la moyenne élevé lorsque l'écart type est également élevé. Cet aspect vient, en quelque sorte, amoindrir

<sup>3</sup> Un air de changement. Comment acheter, conduire et entretenir son véhicule quand on se préoccupe de l'avenir de la planète. Association Québécoise de lutte contre la pollution atmosphérique. Ressources naturelles Canada (2004). *Énergide. Guide de consommateur de carburant*. Adresse Internet : <http://oee.nrcan.gc.ca/vehicules> . L'adresse a été modifiée. Le texte intégral n'est plus disponible sur ce site. Il est toutefois possible d'obtenir des données à l'adresse <http://oee.nrcan.gc.ca/transports/initiative-vehicules-personnels.cfm>

l'importance de l'écart à la moyenne dans le cas de la consommation de la voiture compacte en ville.

Réalisée avec des élèves du secondaire, une telle analyse a pour but de les amener à constater que l'écart de la donnée avec la moyenne ( $x - \bar{x}$ ) doit être relativisé en considérant les autres données de l'échantillon et plus particulièrement, leur écart avec la moyenne. Une façon de procéder est donc d'étudier un *rapport* particulier. Il s'agit d'établir un rapport entre l'écart à la moyenne, pour la donnée à considérer ( $x - \bar{x}$ ), et l'écart moyen des autres données. Donc, en divisant l'écart à la moyenne par l'écart type, nous relativisons cet écart à la moyenne en fonction des écarts des autres données de la distribution. Autrement dit, cette démarche permet de prendre en compte non seulement la position de la donnée par rapport à la moyenne, mais également, sa position *relativement* aux autres données de la distribution.

Ainsi, pour comparer les rendements des véhicules en ville et sur la route nous avons besoin d'un écart relatif à la moyenne ou du nombre de fois que l'écart type est contenu dans l'espace séparant la donnée de la moyenne, ce qui correspond à la cote standard. En reprenant notre exemple avec la voiture compacte, sa cote standard est de -0,70 en ville et de -0,52 sur l'autoroute. Nous pouvons maintenant, à l'aide de ce résultat, affirmer que la voiture compacte est plus efficace en ville que sur l'autoroute, en comparaison des autres véhicules. Autrement dit, même si la voiture compacte occupe, dans les deux situations, le deuxième rang en matière d'efficacité de consommation et qu'elle consomme moins sur l'autoroute qu'en ville, elle se démarque d'avantage des autres véhicules en ville que sur l'autoroute.

À partir de cet exemple, nous posons l'hypothèse selon laquelle le fait de reconnaître, d'une part que la soustraction ( $x - \bar{x}$ ) exprime un écart, et d'autre part, que ce dernier est mis en rapport avec un écart moyen ( $s$ ) pour le relativiser, pourrait favoriser une compréhension de la logique de la formule de la cote standard. La compréhension de la notion de cote standard comme objet implique donc une appropriation de ses éléments constitutifs : l'écart type et la moyenne.

### 1.3 L'écart type et l'écart moyen

Il est possible d'amener l'élève du secondaire à comprendre la signification de l'écart type, en lui permettant de reconnaître la nécessité de cette notion. Pour ce faire, il faut d'abord lui permettre de prendre conscience des limites de la moyenne, qui est une mesure de tendance centrale servant à décrire une population ou un échantillon. C'est donc à partir de la compréhension du concept de moyenne qu'apparaît celle d'écart type. Voyons comment.

Lorsque nous interprétons la moyenne d'un groupe de données, nous cherchons à décrire ce groupe à l'aide d'un minimum d'informations. Dans bien des cas, la valeur de la moyenne peut être suffisante pour se donner une représentation sommaire de la distribution. Cependant, cette mesure peut, à l'occasion, s'avérer insuffisante, ce qui permet de faire apparaître la nécessité de s'approprier de nouveaux outils pour répondre aux besoins ressentis. Par exemple, lorsque nous voulons effectuer une analyse plus précise du comportement d'un groupe de données et que ces données ne sont pas

regroupées autour de la moyenne ou, lorsque vient le temps de décrire et de comparer plusieurs distributions.

Prenons les trois groupes de données suivants qui représentent les buts marqués par chacun des joueurs de trois équipes différentes lors d'une saison de football :

Équipe A) 1 - 1 - 1 - 2 - 10 - 10 - 10    Étendue = 9  
 Équipe B) 4 - 4 - 5 - 5 - 5 - 6 - 6    Étendue = 2  
 Équipe C) 0 - 0 - 6 - 6 - 7 - 7 - 9    Étendue = 9

L'objectif ici est d'identifier l'équipe la mieux équilibrée à l'attaque (l'équipe dans laquelle il y a le moins d'écart entre le nombre de buts comptés par joueur). Dans les trois cas, la moyenne est de 5 et l'étendue des équipes A et C est la même alors que pourtant, les trois séries de données ont une distribution fort différente. Pour distinguer ces trois séries, il faudra une mesure supplémentaire. Dans le cas présent, la répartition des données autour de la moyenne est la caractéristique principale qui distingue ces trois groupes. Ainsi, l'écart de chacune des données par rapport à la moyenne devient l'élément central autour duquel nous pouvons introduire l'idée d'écart type. En effet, l'attention des élèves se porte sur la nécessité d'obtenir une valeur qui exprime la façon dont les nombres sont éloignés de la moyenne. À partir de la compréhension de la notion de moyenne, il est possible de réfléchir sur les écarts à la moyenne pour chacune des données. Les élèves reconnaissent alors que la somme de ces écarts est égale à zéro. En appliquant la valeur absolue sur ces écarts, nous controns ce problème et obtenons

un indice plus révélateur, en l'occurrence, l'écart moyen ( $e_m = \frac{|x - \bar{x}|}{n}$ ).

Pour le groupe A, nous trouvons 4,29  $((4+4+4+3+5+5+5) \div 7 = 4,29)$  et 2,86 pour le groupe C  $((5+5+1+1+2+2+4) \div 7 = 2,86)$ . Il est alors possible de dire, par exemple, que les joueurs de l'équipe A marquent en moyenne 4,29 plus de buts que le nombre de buts moyen ou 4,29 buts de moins que le nombre de buts moyen, alors que les joueurs de l'équipe C marquent en moyenne 2,86 buts de plus ou de moins que le nombre de buts moyen. Par conséquent, l'équipe C est légèrement plus équilibrée que l'équipe A parce que les marqueurs de l'équipe C sont plus groupés que ceux de l'équipe A. Avec un calcul semblable, nous trouvons pour l'équipe B un écart moyen de 0,57 démontrant ainsi que l'équipe B est de loin la plus équilibrée.

Cette situation permet de susciter, chez les élèves, une prise de conscience à l'égard de la pertinence de la notion d'écart moyen. Toutefois, il faut retenir que l'écart moyen est une notion qui ne permet pas de discriminer les données extrêmes. Par exemple dans le cas où :

A : 4,4,4,4,6,6,6,6  
 B : 1, 5,5,5,5,5,5, 9

La moyenne est de 5 et l'écart moyen est de 1 pour les deux distributions. C'est le calcul de l'écart type qui permettra de mettre en relief l'influence des données extrêmes. Il arrive que ces données extrêmes soient abandonnées dans les analyses car suspectes. Dans d'autres cas nous pouvons choisir de les conserver. À cette occasion, en élevant au carré l'écart à la moyenne pour chacune des données, il est possible d'attribuer une plus grande importance aux données extrêmes comme 1 et 9 dans la distribution B.

L'extraction de la racine carrée permet quant à elle de retrouver la dimension initiale de la grandeur en cause. L'utilisation de l'écart type plutôt que de l'écart moyen vient également du fait que le premier est plus facile à manipuler que le second. Comme le note Lafortune : " il est plus facile de calculer l'intégrale d'une expression contenant des racines carrées et des exposants au carré que de calculer l'intégrale d'une expression contenant une valeur absolue " (1998, p.332). Enfin, l'écart type offre un indicateur qui exprime la diversité de la distribution à partir de la moyenne plutôt qu'à partir d'une autre valeur comme la médiane. Pour développer de telles situations, nos étudiants universitaires doivent être en mesure d'identifier le sens, le rôle et l'utilité qu'ils attribuent aux formules rencontrées dans l'apprentissage de ces objets statistiques.

## 2. La méthode

Nous avons distribué un questionnaire écrit (Annexe 1) à un groupe de 31 futurs maîtres provenant d'une université québécoise. Ce questionnaire visait à connaître le sens qu'ils attribuent aux formules d'écart type et de cote standard. À noter que les étudiantes et les étudiants avaient tous auparavant suivi un premier cours de statistiques de niveau universitaire. Dans ce cours, ils se sont familiarisés avec les fondements et les concepts de l'inférence statistique générale. Ils abordent notamment les statistiques descriptives, la distribution des fonctions et transformations de variables aléatoires, la loi normale, de Student, khi-deux, de Fisher, les méthodes d'estimation classique, les estimateurs ainsi que les tests d'hypothèses<sup>4</sup>.

Nous avons analysé les réponses écrites d'abord en transcrivant chacune des réponses pour leur attribuer un numéro. Ensuite, nous avons regroupé les réponses dans lesquelles se dégageaient des idées communes. Ces catégories émergentes ont permis de cerner les différents sens que les futurs maîtres ont donnés aux réactions des élèves ou encore aux questions de contenu mathématique qui ont été proposées. L'analyse des réponses qu'ils nous ont offertes nous a permis d'une part, d'identifier la posture dans laquelle ils se situaient au moment de proposer une intervention et d'autre part, de cerner le sens accordé aux notions étudiées. Nous présenterons donc une interprétation des analyses effectuées.

## 3. Les résultats

### 3.1 La compréhension des notions de cote standard et d'écart type

Afin de dégager des éléments de réponse, nous nous sommes attardés sur la façon dont les futurs maîtres parlent de leurs connaissances mathématiques à propos des concepts de cote standard et d'écart type. À ce moment, les futurs maîtres sont placés dans la posture épistémologique de l'étudiant universitaire. En effet, ils doivent réfléchir

---

<sup>4</sup> Dans ces cours universitaires de premier cycle au Québec, il n'est pas de tradition d'aborder la moyenne, la médiane et l'écart type en terme de distances. Par exemple, la moyenne n'est pas présentée comme étant une valeur  $x$  minimisant la somme des carrés des différences entre une valeur  $x$  et chacun des  $x_i$ . De même, ils n'abordent pas l'écart type comme la distance minimum dans le cas précédent.

sur la nature du concept et situer la réflexion de l'élève dans un processus d'apprentissage.

### 3.1.1 Le rôle de la cote standard

Nous avons demandé à ce groupe de nous décrire le sens qu'ils attribuent à la cote standard. Cette question visait à faire définir cette notion. Parmi les réponses possibles des étudiantes et des étudiants, des indications à propos d'une mesure de position permettraient de repérer l'idée de comparer des données de différentes distributions ou encore de standardiser des écarts. Rappelons que cet objet statistique permet en effet de standardiser des écarts ou des données pour les comparer, même si elles proviennent de distributions différentes. Une définition plus fonctionnelle consisterait à décrire la cote standard comme étant une valeur représentant le nombre de fois que l'écart type est contenu dans l'écart séparant une donnée de sa moyenne. Or, dans les deux cas, l'interprétation nous renvoie à un résultat numérique relatif, qui n'est plus lié à une distribution en particulier. C'est ainsi que ce résultat numérique devient comparable à d'autres valeurs provenant de distributions quelconques. En effet, la cote standard est sans dimension et indépendante des unités utilisées.

Les résultats obtenus démontrent qu'aucun des répondants n'a référé complètement à l'une ou l'autre de ces descriptions. Près de la moitié d'entre eux n'est pas en mesure de donner une réponse. Parmi les autres réponses obtenues, une seule personne évoque l'idée d'une mesure permettant une comparaison avec d'autres groupes. Elle écrit : "Un outil servant à exprimer une comparaison entre les élèves d'une année à l'autre". Quatre autres formulent l'idée d'une comparaison ou d'un écart par rapport à la moyenne d'un groupe, sans toutefois faire référence à la standardisation de ces écarts, alors que sept personnes assimilent cette notion à une mesure de position dans un seul groupe. Ces derniers types de réponse sont incomplets puisqu'ils négligent l'un de ses rôles principaux ; celui de lui donner du sens aux yeux des élèves du secondaire par la standardisation des écarts. Enfin, cinq personnes expriment leur compréhension de cette notion en donnant une description des symboles mathématiques de la formule. Or, dans le cas où les futurs maîtres expriment leurs connaissances selon une définition, nous questionnons le type de réponses qu'ils pourraient offrir aux élèves durant le processus de construction des connaissances. Offriront-ils des arguments d'autorité comme nous l'avons évoqué dans le cadre théorique ou seront-ils en mesure de recourir à un questionnement pour guider les élèves ? Finalement, deux futurs maîtres assimilent la cote standard à la courbe normale.

Afin d'amener les étudiants à préciser leur pensée quant au sens de la cote standard, nous leur avons demandé de nous décrire quelle était, selon eux, l'utilité de la cote standard. Vingt-deux futurs maîtres notent que le rôle de la cote standard correspond à une comparaison. Toutefois, aucun d'entre eux n'évoque explicitement ou clairement une comparaison entre une donnée avec les données d'autres distributions. La moitié de ceux qui évoquent l'idée de comparaison rappelle que cette notion exprime une mesure de position. Par exemple, un étudiant inscrit : "C'est l'évaluation d'un résultat par rapport à l'ensemble des résultats d'un groupe". L'autre moitié commente la comparaison en utilisant les opérations de soustraction ou de division de l'une ou l'autre partie de la formule selon une différence ou un rapport. Ainsi, un futur maître écrit : "La différence d'un résultat par rapport à la moyenne de tous les résultats d'un groupe et ceci divisé par l'écart type". Ces propos se rapprochent davantage d'une description de

la formule que d'une interprétation du concept en contexte. Enfin, trois personnes ont assimilé la cote standard à la courbe normale ou aux résultats extrêmes d'une distribution. Par exemple, les étudiants expliquent : "D'être capable de situer un résultat par rapport à la courbe normale d'un échantillon." Ou "C'est pour ramener à une moyenne pour une courbe normale." ou encore "Éliminer les résultats extrêmes." Rappelons ici que même dans le cas d'une réponse qui aurait été jugée valable sur le plan mathématique (référence précise et complète avec la courbe normale), le sens de cette réponse est inadapté dans le contexte d'un cours de didactique où un des buts visés est une intervention dans une classe du secondaire. Ainsi, la relation avec la courbe normale est, dans ce contexte d'enseignement, inappropriée. Globalement, nous retrouvons donc le tiers des personnes qui ne s'est pas exprimé ou qui a exprimé une compréhension fragmentaire du rôle de la cote standard.

En utilisant le cadre théorique présenté au début de cet article, nous sommes en mesure d'ajouter des éléments à notre analyse. Si les réponses incomplètes ou encore valables sur le plan mathématique sont interprétées comme étant issues de la posture épistémologique de l'ancien élève, l'absence de réponse peut être interprétée du point de vue de la posture de l'étudiant universitaire. Dans ce cadre, cette absence de réponse pourrait être interprétée soit comme un manque de connaissances mathématiques, ce qui est possible mais peu probable, soit comme une difficulté à être le créateur d'une explication plausible pour soi-même. Pour réfléchir sur la nature du concept et situer la réflexion de l'élève dans un processus d'apprentissage, cette attitude est sollicitée.

### 3.1.2 La logique de la formule de la cote standard

Afin d'amener nos étudiantes et nos étudiants à préciser les raisons pour lesquelles la formule nous conduit à proposer une comparaison de données provenant de distributions différentes nous leur avons tout d'abord demandé de répondre à la question : "Pourquoi<sup>5</sup> nous permet-elle de faire ceci ?"<sup>6</sup>. Ainsi, d'un point de vue sémantique, ils pouvaient préciser que la comparaison était basée d'une part, sur l'écart des données à leur moyenne, mais plus important encore, que cet écart était standardisé en fonction de leurs écarts types respectifs. Or, 20 futurs maîtres ne donnent pas de réponse. Trois font référence à la courbe normale ou à l'utilisation de tests statistiques. La majorité de ceux qui donnent une réponse commentent la formule, ce qui correspond à une interprétation syntaxique de la question. Enfin, deux personnes sont en mesure d'exprimer leurs connaissances à ce sujet. Elles font alors référence à l'écart type selon une moyenne des écarts et perçoivent la cote standard comme un rapport entre deux écarts, ce qui pourrait leur donner des outils conceptuels pour discuter de cette notion en classe. Les résultats obtenus mettent ainsi en lumière la difficulté de 20 futurs maîtres à s'exprimer sur ce sujet, ce qui laisse penser que la cote standard n'est pas considérée pour comparer une donnée à d'autres données de distributions différentes.

---

<sup>5</sup> Vermeersch (2006) nous sensibilise à l'influence des questions commençant par « pourquoi ». En effet, elles ouvraient trop d'alternatives à la personne. Par conséquent, cette question ne conduit pas nécessairement à tenir compte du contexte, ce qui doit nous rendre prudents au plan des interprétations.

<sup>6</sup> Un argument d'autorité correspond à une affirmation sans justification. Nous avons choisi d'utiliser l'expression ceci afin de permettre aux étudiants de s'exprimer librement en fonction de la réponse donnée à la question précédente. Une trop grande précision de notre question aurait pu biaiser la réponse de l'étudiant.

Suite à cette question, nous nous sommes attardés plus spécifiquement aux différents termes de la formule de la cote standard. Nous avons cherché à vérifier la compréhension des étudiantes et des étudiants à propos du rôle de chacun de ces termes

(numérateur et dénominateur) dans la formule :  $z = \frac{x - \bar{x}}{s}$ . Nous leur avons tout d'abord

demandé de nous décrire le rôle de  $x - \bar{x}$  dans la formule (Que nous indique  $x - \bar{x}$ ? Quelle information nous donne ce numérateur?). Parmi les réponses possibles, les futurs maîtres pouvaient exprimer l'idée d'un écart entre une donnée et la moyenne d'une distribution.

Tout d'abord, huit personnes ne se sont pas prononcées sur la question. Près de trois quarts des répondants ont cherché à attribuer un sens à la soustraction : écart (10), distance (1), différence (8), position (3). Ces termes qui, au premier abord, peuvent apparaître comme synonymes, revêtent des sens différents qui pourraient influencer l'interprétation de la cote standard telle que développée dans notre analyse conceptuelle (section 1.2). Or, plusieurs de ces sens ne permettent pas d'interpréter la cote Z comme étant un rapport, puisqu'un rapport se définit comme étant une relation entre deux éléments de même nature, dans le cas présent : un rapport entre deux écarts. L'interprétation du résultat de  $x - \bar{x}$  pourrait donc avoir une influence sur la compréhension de la cote Z.

Nous avons donc cherché à cerner la relation qu'ils établissaient entre ces deux écarts en leur demandant d'expliquer la raison de la division de  $x - \bar{x}$  par l'écart type  $s_x$ . Une réponse qui manifeste une compréhension consisterait à mentionner que la présence du  $s$  au dénominateur permet de relativiser l'écart  $x - \bar{x}$  en fonction de la moyenne des écarts, rendant ainsi l'écart à la moyenne comparable à d'autres écarts de même distribution ou de distributions différentes.

À cette question, dix des futurs maîtres ont décrit la formule en la commentant. Par exemple, un étudiant écrit : "*La note moins la moyenne de la classe*". Trois autres ont justifié la présence de l'écart type en évoquant une position par rapport au groupe. Seulement quatre personnes ont formulé des connaissances se rapprochant de la notion de rapport, en considérant encore une fois, la contrainte imposée par l'enseignement de ce concept à des élèves du secondaire qui apprennent les statistiques. Par exemple, ils ont mentionné une comparaison avec l'écart moyen du groupe, amenant ici l'idée de comparaison entre des éléments de même nature. Par ailleurs, un nombre préoccupant de personnes (près de la moitié) n'ont pas été en mesure de donner une réponse à la question.

En conclusion, la cote standard semble être interprétée davantage comme un outil de classification à l'intérieur d'une seule distribution que comme un outil de comparaison dans le cas où il y a présence de plusieurs distributions. Toutefois, le questionnaire utilisé ne sollicitait pas explicitement les futurs maîtres à ce sujet. Cette difficulté pourrait aussi être issue de la logique que les étudiantes et les étudiants attribuent à la formule : les différents sens attribués à l'opération  $x - \bar{x}$  et l'absence de commentaires à propos du rôle joué par le  $s_x$  au dénominateur. Comment alors interpréter cette absence de commentaires sur la reconnaissance de la notion de rapport entre les deux écarts ? Nous avons cherché à connaître l'influence de ces connaissances

sur la création de contextes visant une compréhension de la logique de la formule chez les élèves du secondaire.

### 3.1.3 Comment prévoient-ils intervenir ?

Ainsi, nous avons cherché à connaître comment les futurs maîtres se proposent d'intervenir auprès de leurs élèves, sachant que prévoir une intervention n'est pas intervenir et qu'un écart peut surgir. Un exemple d'intervention consisterait à amener les élèves à comparer deux données issues de distributions différentes comme l'illustre l'exemple donné précédemment avec la consommation des automobiles. En ce faisant, les élèves pourraient constater que l'écart à la moyenne est insuffisant pour porter un jugement sur les positions de ces données dans les distributions. Le but étant d'en arriver à découvrir la nécessité de standardiser les écarts en fonction des écarts moyens.

Parmi les 31 étudiants questionnés, neuf d'entre eux ont mentionné qu'ils auraient recours à un exemple numérique. Toutefois, aucune précision n'est apportée sur le type d'exemple à proposer. D'ailleurs, aucun de ces répondants n'a mentionné qu'il utiliserait deux distributions dans son exemple numérique. Quatre autres étudiants ont fourni des réponses telles que "consulter les manuels scolaires", "en proposant une preuve ou en étudiant un point dans un graphique de nuage de points". Dans l'ensemble, aucune des réponses fournies par nos étudiantes et nos étudiants ne propose de situations permettant aux élèves de raisonner la formule. La plupart des étudiantes et des étudiants qui ont fourni une réponse ont donc adopté la posture de l'enseignant sans transiter par celle de l'étudiant universitaire (13 personnes), reproduisant ainsi des modèles connus. Un seul a adopté une posture épistémologique de l'ancien élève. Ce dernier a proposé une explication approximative de la cote standard. Finalement, un nombre très important d'étudiantes et d'étudiants (19) n'a pas été en mesure de fournir une réponse à la question, confirmant ainsi le peu d'outils dont ils disposent actuellement pour entrer dans la posture de l'étudiant universitaire qui se forme pour enseigner au secondaire.

## 3.2 L'écart type

### 3.2.1 L'utilité de l'écart type

Nous avons aussi cherché à connaître le sens que les futurs maîtres accordent à l'écart type en leur demandant de nous décrire l'utilité de ce concept (Selon vous, quelle est l'utilité de l'écart type ?). Les futurs maîtres pouvaient nous décrire la fonction de l'écart type comme : "une mesure exprimant la façon dont les nombres sont éloignés de la moyenne" ou "une mesure de dispersion des données d'une distribution" ou encore "un nombre exprimant un écart moyen des données par rapport à la moyenne". La notion d'écart moyen est fondamentale pour comprendre le sens de la formule de cote standard parce qu'elle permet de saisir l'idée de valeur relative de la cote standard.

Seulement cinq des futurs maîtres évoquent l'idée d'écart moyen. Près de la moitié des personnes qui ont pu donner une fonction comme "exprimer une mesure de dispersion" n'ont pas évoqué l'écart moyen. Ce type de réponse est adéquat. Toutefois, leur formulation semble assimiler la fonction de l'écart type davantage à une définition toute faite qu'à une réelle fonction de ce concept. Par exemple, un étudiant a répondu : "L'écart type est la moyenne des écarts de chaque donnée par rapport à la moyenne des valeurs de chaque donnée". Les 14 autres futurs maîtres assimilent le concept d'écart

type à celui de mesure de position, de courbe normale ou encore donnent un sens imprécis à leur énoncé. Ainsi, la moitié des futurs maîtres interrogés expriment difficilement leurs connaissances au sujet de cette notion.

En conclusion, près de la moitié des personnes interrogées ne s'expriment pas au sujet de l'utilité de l'écart type, comme c'était le cas pour la cote standard. À nouveau, nous choisissons d'analyser l'absence de réponse du point de vue de la posture de l'étudiant universitaire. Pour réfléchir sur la nature du concept et situer la réflexion de l'élève dans un processus d'apprentissage, une création personnelle est nécessaire. C'est ainsi que nous interprétons la difficulté qui semble émerger des réponses obtenues.

### 3.2.2 La logique de la formule de l'écart type

Nous avons posé la question suivante à ces mêmes personnes: " Dans la formule de l'écart type ( $s = \sqrt{\frac{(x - \bar{x})^2}{n}}$ ), pourquoi  $x - \bar{x}$  est-il mis au carré ? ". Dans ce cas, seulement deux personnes ont pu reconnaître qu'en ne mettant pas cet écart au carré, le nombre obtenu serait zéro. La majorité des futurs maîtres semble faire appel à un argument d'autorité: "La différence doit être positive." ou encore "Il faut éliminer le négatif.". Ce type d'explication, qu'il soit considéré comme syntaxique ou sémantique, permet de poser certaines hypothèses. Une d'entre elles serait que les futurs maîtres pourraient considérer cette formule comme un algorithme à appliquer pour trouver une réponse<sup>7</sup>. Une autre laisse penser que ces futurs maîtres entrevoient la possibilité d'un résultat négatif, ce qui est faux.

Dans la même voie que la question précédente, nous avons également demandé aux futurs maîtres de nous expliquer le rôle de la racine carrée dans la formule de l'écart type: "Dans cette même formule, comment justifiez-vous la présence de la racine carrée ?". Seulement douze des futurs maîtres interrogés reconnaissent l'importance de réaliser l'opération inverse "pour contrer l'effet des écarts au carré et se rapprocher d'une moyenne des écarts", une réponse à laquelle nous nous attendions. Toutefois, certaines formulations nous semblent problématiques entre autres, lorsque les futurs maîtres évoquent l'idée d'une somme d'écarts ("Élimine la mise au carré de la somme des  $(x - \bar{x})^2$ "). Cela pourrait laisser penser qu'ils considèrent que la racine carrée d'une sommation d'éléments au carré est équivalente à une sommation de ces éléments. Pourtant,  $\sqrt{x^2} \neq x$ . Dix personnes évoquent une raison qui ne fait pas référence à une logique interne de la formule ("Parce que c'est le radical de la variance"). À nouveau, la conception entretenue à l'égard de la notion de formule pourrait influencer leur difficulté à exprimer des connaissances mathématiques. En effet, si la formule est considérée comme un algorithme, ils peuvent ne pas sentir la nécessité de la raisonner. Enfin, onze personnes ne se sont pas prononcées, soit plus du tiers des répondants.

En conclusion, plus du quart des personnes n'a pu s'exprimer pour l'une ou l'autre des questions et parmi celles qui ont fourni une réponse, très peu ont pu s'exprimer à

<sup>7</sup> Vermeersch (2006) nous sensibilise à l'influence des questions commençant par « pourquoi ». En effet, elles ouvriraient trop d'alternatives à la personne. Par conséquent, cette question ne conduit pas nécessairement à tenir compte du contexte, ce qui doit nous rendre prudents au plan des interprétations.

l'égard du sens et de la fonction de l'écart type. De plus, un grand nombre de personnes ont proposé des définitions (dans le cas de la fonction de l'écart type) ou une description d'une formule sans référence au sens, ce que nous interprétons comme étant des manifestations d'arguments d'autorité. De tels arguments pourraient susciter chez les élèves du secondaire une conception selon laquelle les mathématiques viennent d'un ailleurs (Charlot, 1974). Il devient donc important de se questionner sur la nature des outils conceptuels à offrir pour supporter les futurs maîtres au moment d'entrer dans la posture épistémologique de l'étudiant universitaire. Sans ce passage, les étudiantes et les étudiants risquent de reproduire des modèles connus plutôt que de répondre aux besoins de leurs élèves.

## 4. Discussion

### 4.1 Retombées de la compréhension des notions statistiques au secondaire : l'exemple de la corrélation

De façon générale, un enseignement axé principalement sur les aspects plus théoriques des statistiques peut conduire à développer des connaissances procédurales. De telles connaissances pourraient contribuer à forger et à entretenir une posture épistémologique dans laquelle la manifestation d'un apprentissage se réduit à la conformité à l'égard des connaissances attendues. Nous avons posé l'hypothèse selon laquelle la posture épistémologique de l'ancien élève pourrait conduire à manifester une compréhension fragmentaire des notions sous-entendues dans les formules. Pour analyser la nature d'un concept, il devient nécessaire d'entrer dans la posture épistémologique de l'étudiant universitaire qui recrée pour lui les relations entre les fragments épars de ces connaissances ou entre les "îlots de connaissance" (Lemoine, 1989) élaborés durant son passage à l'école. Voyons d'abord quels sont ces îlots de connaissance.

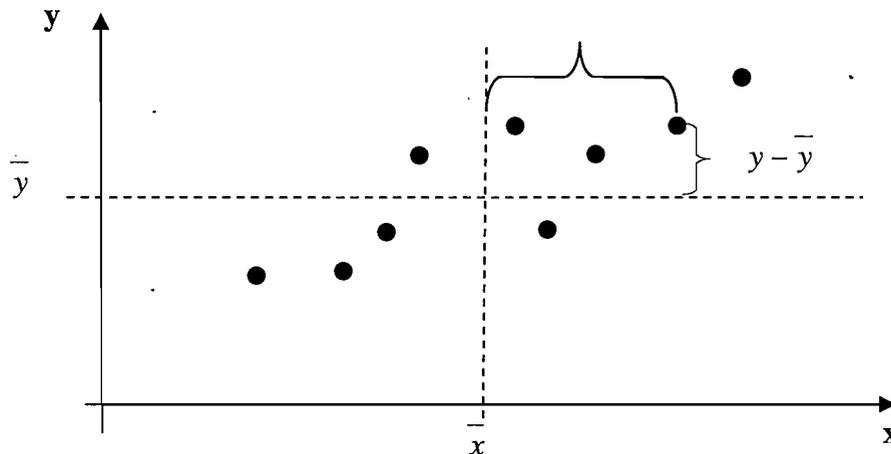
Dans un premier temps, nos résultats montrent que lorsque les futurs maîtres sont invités à s'exprimer sur le rôle de la cote standard, cette dernière est essentiellement considérée comme un outil de classification dans un seul groupe. Deuxièmement, la fonction de l'écart type est assimilée à une définition dans laquelle l'idée d'écart moyen n'est pas explicitée. Troisièmement, lorsque les futurs maîtres sont questionnés sur la logique sous-entendue par la formule de la cote standard, ils privilégient une opération mathématique au détriment du rapport. Quatrièmement, l'utilisation d'arguments d'autorité pour expliquer la formule de l'écart type apparaît. Dans ce cas, est-il possible que cette formule soit considérée comme un algorithme qui sert à trouver une réponse ? Enfin, l'inégalité entre la somme d'écart et la racine carrée d'une somme de ces écarts au carré pourrait poser problème.

Nous posons donc la question suivante : La reconnaissance des rôles particuliers de la cote standard et de l'écart type, comme nous l'avons abordé précédemment, pourrait-elle contribuer au développement d'une compréhension de notions statistiques comme la corrélation ? Cette question nous conduit à poser une réflexion préalable sur la compréhension de la notion de corrélation et sur son coefficient.

## 4.2 La compréhension du coefficient de corrélation

Deux compréhensions du sens nous paraissent essentielles : celle de la *notion* de corrélation (sens lié au contexte) et celle du *coefficient* de corrélation (sens statistique)<sup>8</sup>. Le premier type de compréhension exige la distinction entre une corrélation et une causalité (Duperret, 2001). En effet, deux phénomènes qui semblent intimement liés graphiquement en fonction de leur évolution ou de leur variation, n'indiquent en rien qu'un des phénomènes puisse influencer l'autre. La corrélation doit plutôt être comprise comme un indice de covariation entre deux variables ou entre deux phénomènes. Cette dernière définition fait donc intervenir l'idée de standardisation qui, dans la formule, est représentée par la présence de l'écart type au dénominateur.

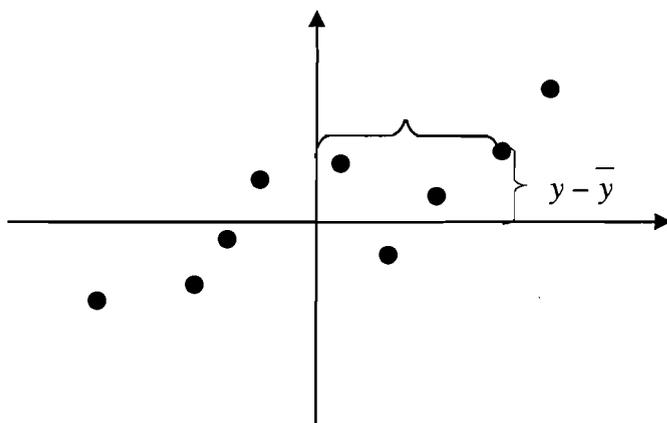
Cet aspect nous renvoie ainsi au second type de compréhension, qui lui, est associé au résultat numérique que nous fournit le coefficient de corrélation (sens statistique). Une interprétation du résultat numérique de la corrélation, ainsi qu'une compréhension profonde de cette lecture, nécessite de bien saisir la logique de la formule de la cote standard (écart standardisé). Weldon (2000) montre que le coefficient de corrélation  $r$  peut s'interpréter comme la moyenne d'un produit d'écarts standardisés (soit  $Z_x$  et  $Z_y$ )  $r = \frac{Z_x Z_y}{n}$ . Inspirés de Allard (1992) et de Lafortune (1998), nous avons cherché à interpréter la logique de cette formule à partir de la représentation graphique d'une corrélation entre deux phénomènes où  $Z_x = \frac{x - \bar{x}}{s_x}$  et  $Z_y = \frac{y - \bar{y}}{s_y}$



Dans ce graphique, nous avons représenté un nuage de points, avec les moyennes des deux variables en présence ( $\bar{x}$  et  $\bar{y}$ ) et l'écart entre les coordonnées du point et ces moyennes. Établir une corrélation entre deux variables revient à comparer le comportement d'une variable en fonction du comportement de l'autre. Pour comparer des données, elles doivent s'exprimer sous une même unité ou sur une même échelle. Le calcul de la cote standard de chacune des données permet de répondre à cette condition puisque cette notion "standardise des écarts". Sur le nouveau graphique obtenu à partir

<sup>8</sup> Dans ce contexte, le mot «statistique» réfère davantage à la science statistique qu'aux différents objets qui y sont étudiés.

des cotes standards de chacun des points, nous constatons que les couples ont subi une translation horizontale et verticale correspondant aux moyennes respectives en  $x$  et en  $y$ . Elles ont été 'centrées'. Ces translations s'expliquent par le fait que dans le calcul de la cote standard, on soustrait la moyenne à chaque coordonnée ( $x - \bar{x}$ ). Ceci a pour effet de ramener les moyennes de  $x$  et  $y$  à l'origine. Quant aux unités de l'ordonnée et de l'abscisse, elles sont maintenant exprimées en unités d'écart type. Cette transformation est due à la division de  $x - \bar{x}$  et de  $y - \bar{y}$  par leur écart type respectif ( $s_x$  et  $s_y$ ). Enfin, l'opération de multiplication entre  $Z_x$  et  $Z_y$  permet d'obtenir un seul indice indiquant une covariation entre deux variables standardisées.



Nous pouvons ainsi constater que les points qui sont dans les cadrans 1 et 3 ont des écarts à la moyenne ( $x - \bar{x}$  et  $y - \bar{y}$ ) qui sont dans les deux cas respectivement positifs et négatifs. De ce fait, la multiplication de  $x - \bar{x}$  par  $y - \bar{y}$  donne toujours un nombre positif<sup>9</sup>. De même, nous observons que les écarts à la moyenne en  $x$  et en  $y$  sont de signes contraires dans les cadrans 2 et 4 et que, par conséquent, la multiplication de  $x - \bar{x}$  par  $y - \bar{y}$  résulte en une valeur négative. Par ailleurs, comme l'écart type est toujours une valeur positive, la division de  $x - \bar{x}$  et  $y - \bar{y}$  par leur écart type respectif ne modifie pas le signe de ces valeurs. Il en résulte que  $Z_x Z_y > 0$  pour les points compris dans les cadrans 1 et 3, et de même,  $Z_x Z_y < 0$  pour les points appartenant aux cadrans 2 et 4.

Ainsi, lorsque la valeur du coefficient de corrélation est positive cela peut signifier que les points sont plus nombreux dans les cadrans 1 et 3 que dans les cadrans 2 et 4 ou encore que des valeurs extrêmes sont présentes dans les cadrans 1 et 3. De plus, à mesure que le nombre de points augmente dans les cadrans 1 et 3, cette valeur s'accroît en tendant vers 1. Le cas extrême correspond à celui où aucun point ne se trouve dans les cadrans 2 et 4, alors que les autres sont alignés sur une droite de pente positive passant par l'origine. Dans un tel cas, le coefficient de corrélation atteint sa valeur maximale de 1. Nous parlons alors d'une corrélation linéaire parfaite. À noter, comme le mentionne Saporta, que : "  $r$  ne mesure que le caractère *linéaire* d'une liaison et son

<sup>9</sup> La multiplication de ces deux valeurs est en fait un indice de covariation de chacun des écarts à la moyenne.

usage doit être réservé à des nuages où les points sont répartis de part et d'autre d'une tendance linéaire" (p. 133, 1990). Nous posons l'hypothèse selon laquelle cette interprétation du graphique en lien avec la formule ne peut être assimilée qu'à partir d'une compréhension des éléments qui composent la formule du coefficient de

corrélation soit, la cote standard et l'écart type, ( $Z_x = \frac{x - \bar{x}}{s_x}$  et  $s_x = \sqrt{\frac{(x - \bar{x})^2}{n}}$ ) dans le cas d'un échantillon de taille n.

### 4.3 Retombées des connaissances de cote standard et de l'écart type

Au Québec, depuis l'instauration du précédent programme d'étude au secondaire au milieu des années 90, la notion de coefficient de corrélation est enseignée en cinquième secondaire<sup>10</sup> (15-16 ans). En mathématiques et sciences naturelles, ces élèves sont initiés à la formule du coefficient de corrélation et à l'utilisation de celle-ci pour calculer des coefficients. Par contre, en mettant principalement l'accent sur le calcul des coefficients, réalisé à l'aide de logiciels, et sur l'interprétation de ces coefficients, les programmes et les manuels délaissent la compréhension du sens de la formule<sup>11</sup>.

Les étudiantes et les étudiants ayant suivi notre cours de didactique et participé à notre étude avaient déjà tous suivi un cours de statistique de niveau universitaire. Toutefois, les concepts statistiques enseignés au niveau secondaire ne sont que très peu abordés dans ce cours. De plus, les concepts statistiques qui sont traités le sont en tant qu'outil mathématique et non en tant qu'objet mathématique à enseigner. Dans un tel contexte, les étudiants développent souvent une compréhension essentiellement procédurale des statistiques, en ce sens qu'ils connaissent généralement les formules statistiques de base et la façon de les appliquer.

Nous avons proposé à un groupe de 63 futurs maîtres, d'une université différente de celle du premier questionnaire, un contexte susceptible d'être rencontré au moment d'enseigner dans une classe du secondaire. Nous leur avons demandé de décrire une intervention et ce, avant qu'une réflexion ne soit développée en classe de didactique. Nous avons proposé la question suivante : On a trouvé une corrélation forte et positive entre le nombre d'étudiants dans une école secondaire et le taux de réussite. Un élève interprète l'information en expliquant que s'il y avait moins d'élèves dans l'école, il réussirait mieux. Comment intervenez-vous ?

À cette question, parmi les réponses possibles, les étudiantes et les étudiants pourraient tenir compte à la fois du sens statistique (corrélation forte et positive) et du sens contextuel (malgré la corrélation forte et positive, absence probable de liens entre les deux phénomènes). Une corrélation forte et positive signifie qu'une variable évolue dans le même sens que l'autre en tendant vers une relation linéaire. Toutefois, le type de

<sup>10</sup> Dans le nouveau programme de formation, la cinquième secondaire correspond maintenant à la troisième année du second cycle du secondaire.

<sup>11</sup> Dans ce programme, aucune mention n'est faite quant à une compréhension des éléments constitutifs de la formule. On suggère plutôt d'avoir recours à des tableurs ou des calculatrices graphiques pour effectuer des modélisations ou pour obtenir le résultat numérique du coefficient de corrélation. À cet effet, le programme mentionne la chose suivante : «Il serait intéressant de calculer le coefficient de corrélation afin de valider le choix du modèle jumelé et de trouver la règle de la fonction correspondante» (MEQ, 2004, p.111).

relation impliquée par cette influence n'est pas nécessairement une relation de cause à effet. L'intervention des futurs maîtres pourrait porter sur la distinction entre une relation de cause à effet et une corrélation (sens contextuel) et en même temps sur l'interprétation erronée à la fois de la corrélation forte (sens statistique) et positive (sens statistique). Dans le cas présent, une intervention axée sur l'interprétation appropriée de la corrélation du point de vue d'un sens statistique impliquerait de conduire l'élève à illustrer, par un graphique, la corrélation positive de la situation proposée pour la comparer à l'interprétation de l'élève. En effet, l'élève a interprété la situation comme s'il s'agissait d'une corrélation négative.

Nos résultats montrent plutôt qu'une très grande majorité de personnes (plus de 90%) ne tient pas compte simultanément du sens statistique et du sens contextuel au moment d'intervenir. Une étude plus approfondie de leurs écrits fait ressortir deux postures épistémologiques : celle de l'ancien élève et celle de l'enseignant. L'ancien élève répond en réinterprétant la corrélation impliquée dans le problème alors que l'enseignant se préoccupe des stratégies d'intervention. Dans le premier cas, la posture adoptée semble conduire à privilégier une intervention où l'énoncé est paraphrasé ou répété. Les futurs maîtres qui se sont retrouvés dans cette posture répondent comme si la question leur était posée comme élève. Par exemple, ils écrivent : "L'élève parle d'une corrélation négative. Même s'il y a une corrélation forte, on ne peut pas nécessairement établir un lien de cause à effet". Les futurs maîtres qui se sont placés dans la posture épistémologique de l'ancien élève ont donc assimilé leur rôle d'enseignant à celui de l'ancien élève.

Quant à ceux qui ont adopté la posture de l'enseignant, ils manifestent une certaine distance entre leur rôle d'ancien élève. Ils écrivent : "Je lui expliquerais que s'il y avait moins d'élèves, la corrélation ne serait pas obligatoirement plus forte, car peut-être qu'il y aurait moins d'élèves forts par rapport au nombre d'élèves faibles". Cependant, si cette posture conduit à une plus grande variété de stratégies d'intervention, ces dernières restent vagues. Ainsi, ils proposent à l'élève de construire un graphique, lui définissent la notion, lui fournissent une explication de la situation. Il est aussi possible de constater qu'aucun futur maître n'a discuté des facteurs influençant les deux phénomènes : l'existence ou non d'une relation (sens contextuel), et la valeur qualitative de  $r$  (sens statistique). De plus, le questionnement suscité par ces étudiants cherche plutôt à faire définir les notions par l'élève qu'à guider le raisonnement. Le recours à des exemples est également évoqué, mais sans être présenté ni explicité. Finalement, on note dans leurs propos un grand nombre d'arguments d'autorité.

#### **4.4 Une hypothèse pour expliquer pourquoi les futurs maîtres ne tiennent pas compte des sens statistique et contextuel simultanément**

La posture de l'ancien élève invite le futur maître à susciter une explication par une réinterprétation ou une répétition de l'énoncé du problème. En adoptant cette posture, le futur maître développe une explication pour lui-même ou tente d'élaborer une intervention qui favoriser sa propre compréhension (auto-explication) plutôt que celle d'un élève. Une réflexion devient nécessaire pour permettre aux futurs maîtres de cerner comment ils ont pu surmonter leurs propres difficultés. Cette réflexion ne peut se réaliser que dans la posture de l'étudiante ou de l'étudiant universitaire. L'adoption de cette posture épistémologique pourrait ensuite les conduire à reconnaître l'apport d'une

analyse conceptuelle au moment de planifier une situation d'apprentissage pour élaborer des questions qui permettent de guider le raisonnement de l'élève du secondaire. Ainsi, l'identification des rôles des notions en jeu et du sens des opérations deviennent significatifs. Afin d'éviter que cette position ne conduise à considérer que les connaissances didactiques sont des dogmes il nous semble nécessaire d'amener les futurs maîtres à identifier des variables didactiques.

C'est à travers l'utilisation des outils statistiques permettant de résoudre des problèmes dans leur contexte ou de répondre à des questions que l'on se pose que ces derniers prennent tout leur sens. À titre d'exemple, certains sites Internet offrent, soit des bases de données, soit de construire des bases de données à partir desquels les élèves peuvent formuler leurs propres questions et discuter avec les enseignants des interprétations à accorder aux différentes représentations proposées (graphiques, tables de valeurs, coefficient de corrélation)<sup>12</sup>.

## 5. Ébauche d'une ingénierie didactique

Le travail de réflexion amorcé dans cette étude nous a conduits à mettre au point les premiers jalons d'une ingénierie didactique conçue non pas pour les élèves du secondaire mais pour les futurs maîtres de mathématiques. Pour l'élaboration de cette ingénierie, deux éléments ont guidé notre réflexion : 1- la formation mathématique préalable de nos étudiantes et de nos étudiants, 2- les situations favorisant chez les futurs maîtres une alternance entre les trois postures épistémologiques identifiées précédemment (DeBlois, sous presse).

Rappelons que dans le cas de la cote standard par exemple, cet outil statistique devient un objet d'étude au secondaire dont la portée demeure limitée. En effet, comme les statistiques inférentielles sont peu étudiées au secondaire, le point d'entrée pour la cote standard n'est pas le même que dans un cours de niveau universitaire. En plus de cet aspect, les futurs maîtres doivent garder à l'esprit la façon d'amener les élèves du secondaire à amorcer une réflexion qui les poussera à remettre en question leurs conceptions à l'égard de ces concepts.

Cette dernière considération nous conduit à souligner l'importance du changement de posture chez le futur maître. De la posture de l'ancien élève, il s'agit de le conduire à adopter une posture de l'enseignant, par un passage par celle de l'étudiant universitaire. Ainsi, nous proposerons d'abord un problème d'enseignement de manière à créer la nécessité de faire émerger les conceptions liées à l'enseignement et les connaissances statistiques de l'ancien élève pour les discuter. Nous présenterons une variété de représentations des élèves à l'égard de la cote standard, ce qui pourrait contribuer à accorder une plus grande importance à l'activité des élèves dans leur apprentissage et par conséquent, une modification de la conception de leur rôle d'enseignant.

---

<sup>12</sup> Au Canada, le site Internet de Statistique Canada ([www.statistiquecanada.ca](http://www.statistiquecanada.ca)) met à la disposition des usagers plusieurs bases de données à partir desquelles il est possible d'effectuer des analyses statistiques de base (graphique, calcul de moyennes, écart types et de coefficient de corrélation).

Ainsi, nous savons que près de la moitié des étudiantes et des étudiants questionnés ne donnent pas de définition à la cote standard alors que près de l'ensemble de l'autre moitié considère la cote standard comme une comparaison. En outre, la moitié de ceux qui évoquent l'idée de comparaison considèrent que son rôle consiste à exprimer une mesure de position alors que l'autre moitié commente la comparaison en utilisant les opérations impliquées dans la formule. Enfin, la majorité des futurs maîtres interrogés ne sont pas en mesure d'identifier les raisons de son rôle alors que quelques-uns font référence à la courbe normale ou à l'utilisation de tests statistiques. Il sera intéressant d'utiliser les conceptions exprimées en proposant le contexte de la consommation des voitures.

Par exemple, en donnant le tableau de la consommation des voitures en ville et sur l'autoroute une discussion démarre. Lors de cette discussion, la consommation moyenne des voitures sur l'autoroute et en ville peut être discutée. La reconnaissance de la limite de la moyenne pourra émerger à partir de l'étude d'une donnée issue de la consommation en ville et sur l'autoroute (moyennes et répartitions des données autour de la moyenne différentes). Questionner la répartition des données autour de la moyenne ainsi que l'écart entre chaque donnée et sa moyenne permettra d'introduire l'idée d'écart moyen, puis d'écart type, reléguant la notion de moyenne au statut d'objet. Ce faisant, il devient possible d'étudier la répartition des données autour de cet objet (moyenne). En donnant aux futurs maîtres le résultat du calcul de l'écart type (3,45 en ville et 2,3 sur l'autoroute), il devient possible de s'attarder sur l'interprétation à donner à ces résultats en leur faisant part de l'interprétation possible des élèves. Ainsi un élève du secondaire peut dire : "*Ce nombre exprime une mesure de dispersion autour de la moyenne*" alors qu'un autre affirme : "*Ces nombres expriment une mesure de position*" et qu'un troisième explique : "*Ces nombres expriment la moyenne des écarts de chaque donnée par rapport à la moyenne des valeurs de chaque donnée*". Afin de les sensibiliser à la diversité des représentations des élèves, nous pourrions leur demander ce que révèlent ces affirmations et chercher à en cerner les raisons. Les futurs maîtres pourront reconnaître que deux des affirmations sont vraies (1 et 3) pour ensuite discuter du caractère explicatif de chacune.

L'attention pourra enfin être portée sur l'effet de l'écart entre les moyennes de consommation d'essence en ville et sur l'autoroute. Il est alors possible d'amener les futurs maîtres à reconnaître qu'il devient difficile d'établir une comparaison entre les deux écarts en cause (3,45 en ville et 2,3 sur l'autoroute). C'est en effet à travers l'étude d'une situation où deux données sont issues de distributions différentes (moyennes et répartitions des données autour de la moyenne différentes) qu'il est possible de démontrer les limites de la moyenne et de l'écart type comme outils d'analyse. L'écart entre une donnée et la moyenne ( $x - \bar{x}$ ) doit être relativisé (standardisé). Il est possible de relativiser cet écart en établissant un rapport. L'écart type servant d'indice représentant cette répartition moyenne, les futurs maîtres pourraient reconnaître le rôle de l'écart type au dénominateur dans la formule de la cote standard. Ainsi, les consommations en ville et sur l'autoroute de la voiture compacte, par exemple, peuvent être comparées entre elles puisque le rapport est conservé relativement à la position qu'elles occupent dans leur distribution respective (l'écart de chaque donnée par rapport à la moyenne est transformé en unité d'écart type). Les futurs maîtres peuvent aussi reconnaître qu'établir un rapport entre deux écarts permet de considérer les autres données de l'échantillon et plus particulièrement, leur écart avec la moyenne. Nous

pourrions ajouter alors : "le rapport entre l'écart à la moyenne et l'écart type de la voiture compacte est de -0,70 en ville et de -0,52 sur l'autoroute. Que nous apprend cette information ?" Une simulation pourra permettre d'identifier certaines affirmations possibles d'élèves du secondaire. L'un dit: "Une mesure de position pour la consommation d'essence de véhicules en ville (classement)", un autre affirme : "Un écart ou une comparaison par rapport à la moyenne de consommation d'essence des véhicules", un dernier explique : "Un outil servant à exprimer une comparaison entre la consommation d'essence en ville et sur l'autoroute pour un véhicule donné". Afin de leur permettre d'identifier les caractéristiques pertinentes aux concepts, nous pourrions les questionner sur les limites de ces affirmations et les raisons de ces limites. Les futurs maîtres pourront reconnaître que la première affirmation, bien que vraie, ne permet de susciter la nécessité d'utiliser la cote standard dans le contexte de ce problème. La troisième affirmation, sans décrire la cote standard, précise l'utilité de cet outil statistique. En effet, comme outil, la cote standard permet de comparer des données issues de distributions différentes. Nous pouvons maintenant, à l'aide de ce résultat, affirmer que la voiture compacte est plus efficace en ville que sur l'autoroute, en comparaison des autres véhicules.

C'est ainsi que l'ancien élève pourra adopter la posture de l'étudiant universitaire pour reconnaître que :

- 1) l'écart à la moyenne est insuffisant pour porter un jugement sur les positions des données dans les distributions ;
- 2) l'écart type offre un indicateur qui exprime la diversité de la distribution à partir de la moyenne plutôt qu'à partir d'une autre valeur comme la médiane.
- 3) la cote standard n'est pas uniquement un outil de classification pour une donnée dans un seul groupe, mais permet aussi de standardiser les écarts en fonction des écarts moyens contribuant ainsi à réaliser des comparaisons ;
- 4) la formule de la cote standard correspond à un rapport plutôt qu'à une simple opération mathématique ;
- 5) La cote standard rend une distribution centrée réduite (moyenne de 0 et écart type de 1), ce qui fait de cet outil un objet : un écart relativisé.
- 6) on peut "raisonner" les formules d'écart type et de cote standard en discutant une variété de sens, de même que l'inégalité entre la somme d'écarts et la racine carrée d'une somme de ces écarts au carré.

Il devient donc possible, pour les futurs maîtres, de devenir *les créateurs d'une explication plausible pour eux-mêmes* à l'égard de la moyenne, de l'écart-moyen, de l'écart type et de la cote standard. L'idée fondamentale de cette ingénierie consiste à reconstruire le concept de cote standard en amenant les futurs maîtres à réfléchir sur *les façons de découvrir les limites des outils statistiques de base tout en introduisant la nécessité de nouveaux objets de pensée*. L'étude de la distinction entre enseignement et apprentissage permettra éventuellement à ces futurs maîtres de reconnaître des caractéristiques pertinentes à une situation d'enseignement apprentissage.

Cette compréhension de la cote standard comme écart relatif transforme l'outil en objet ce qui contribue à la compréhension, entre autres, de la formule du coefficient de corrélation. De plus, la compréhension du fonctionnement de la formule de la corrélation émergeant de l'interprétation des opérations à réaliser pourrait contribuer au développement du sens contextuel de la corrélation par la suite. Comme il a été

mentionné précédemment, il est possible de considérer la formule du coefficient de corrélation comme une moyenne de produits cartésiens d'écart relatif à la moyenne. L'étude d'un nuage de points appuyée par une compréhension de la cote standard comme objet (écart relatif) peut conduire à interpréter autrement la corrélation pour constater qu'il n'y a pas nécessairement de relation fonctionnelle ou de relation de cause à effet entre les deux variables en question, mais une covariation de deux variables ou deux phénomènes. C'est alors que se développe le passage d'une compréhension de la variance à la covariance. Reste ensuite à conduire les futurs maîtres à exprimer cette "création pour eux" en situation d'enseignement/apprentissage.

## 6. Conclusion

Certains principes nous semblent fondamentaux si nous désirons que les futurs maîtres, devenus enseignants, confrontent leurs élèves à des situations où ils auront à choisir des outils statistiques et à en évaluer leur pertinence et leurs limites. Il nous semble fondamental que les futurs maîtres puissent attribuer un sens à des formules pour sortir des sentiers battus. En les conviant à modifier le rapport qu'ils entretiennent avec ce savoir, nous dégagons des principes de même que leur influence sur les différentes postures épistémologiques.

Le cadre théorique retenu nous permet de considérer leurs difficultés à s'exprimer ou leur absence de réponse comme une difficulté à reconstruire pour eux des explications plausibles à partir des îlots de connaissance élaborés durant leur cheminement à l'école. Une compréhension approfondie des rôles des notions en jeu et du sens des opérations impliquées nécessite de modifier sa posture épistémologique. Ce mouvement crée une distance pouvant alors permettre la mise en place d'innovations comme la création de contextes numériques signifiants.

## Bibliographie

ALLARD, J. (1992). Une troisième voie dans l'enseignement de la statistique en sciences humaines. *Bulletin AMQ*, octobre, 19-26.

BAKKER, A. (2004). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal*, 3 (2), 64-83.

BOYE, A. ET COMAIRAS, M.-C. (2002). Moyenne, médiane, écart type : Quelques regards sur l'histoire pour éclairer l'enseignement des statistiques. *Reperes-IREM*, 48, 27-39.

BROWN, J. S., COLLINS, ET DUGUID. (1989). Situated cognition and the culture of learning. *Educational Researcher*. 28(1), 32-42.

CASELLA, G. & ROGER, L., B. (1990). *Statistical inference*. California: Wadsworth & Brooks/Cole Advanced Books & Software.

CHARLOT, B. (1976). Les contenus non mathématiques dans l'enseignement des mathématiques. *IREM de Nantes*.

- CHRISTENSEN, R. (1986). *La statistique : démarche pédagogique programmée*. Chicoutimi (Québec) : G. Morin Éditeur.
- CYR, S. et DEBLOIS, L. (2002). Donner du sens à la notion de corrélation à partir des connaissances antérieures. *Actes du 45<sup>e</sup> congrès du GRMS*, 38-42.
- DEBLOIS, L. et SQUALLI, H. (2002). Implication de l'analyse de productions d'élèves dans la formation des maîtres. *Educational Studies in Mathematics*, 50 (2), 212-237. Kluwer Academic Publishers. <http://www.kluweronline.com/issn/0013-1954>.
- DOUADY, R. (1986) Jeux de cadres et dialectique outil-outil. *Recherche en didactique des mathématiques* 7(2), Éditions la pensée sauvage.
- DUPERRET, J.-C. (2001). Des statistiques à la pensée statistique. *Publication IREM*, Université de Montpellier II.
- GATTUSO, L. (1997). La moyenne, un concept évident ? *Bulletin AMQ*, 37 (3), 10-19.
- LAFORTUNE, L. (1998). *Mathophilie : Mathématique 536* (tome 2). Montréal : Guérin.
- LEMOINE, G. (1989). La peur de ne pas savoir la réponse : sentiment partagé par les élèves en difficulté d'apprentissage en mathématiques et les enseignants. *Repères-IREM*, 12, 79-101.
- MINISTÈRE DE L'ÉDUCATION DU QUÉBEC (MEQ), (2004). *Programme de formation de l'école québécoise, 2<sup>e</sup> cycle* (version préliminaire). Québec : Gouvernement du Québec.
- SAPORTA G., 1990, *Probabilités, analyse de données et statistique*. Paris : Éditions Technip.
- WELDON, K. L. (2000). A Simplified Introduction to Correlation and Regression. *Journal of Statistics Education*, 8 (3), 1-7.
- WESTERMAN, D. (1991). Expert and Novice Teacher Decision-making. *Journal of Teacher Education*, 42 (4), 292-305.

## ANNEXE

## Questionnaire sur le sens des formules statistiques au secondaire

## Questionnaire A

## Question 1

a) Qu'est-ce que la cote standard ?  $Z = \frac{x - \bar{x}}{s}$

b) À quoi est-elle utile ?

c) Pourquoi nous permet-elle de faire ceci ?

## Question 2.

Dans la cote standard  $Z = \frac{x - \bar{x}}{s}$

a) Que nous indique  $x - \bar{x}$  ? (Quelle information nous donne-il ?)

b) Pourquoi  $x - \bar{x}$  est-il divisé par  $s$  ?

c) Comment expliquer ceci à un élève de cinquième secondaire ?

## Question 3

a) Selon vous, quelle est l'utilité de l'écart type ?

b) Dans la formule de l'écart type ( $s = \sqrt{\frac{(x - \bar{x})^2}{n}}$ ), pourquoi  $x - \bar{x}$  est-il mis au carré ?

c) Dans cette même formule, comment justifiez-vous la présence de la racine carrée ?

## Questionnaire B

On a trouvé une corrélation forte et positive entre le nombre d'étudiants dans une polyvalente et le taux de réussite. Une élève interprète l'information en expliquant que s'il y avait moins d'élèves dans l'école, ils réussiraient mieux.

Comment intervenez-vous ?