

Raisonnement statistique et formation citoyenne

VERMETTE, Sylvain

Université de Sherbrooke (étudiant au doctorat)

Canada

GATTUSO, Linda

Université du Québec à Montréal

Canada

INTRODUCTION

Notre époque est marquée par un grand développement de la communication. Les textes diffusés abondent de données statistiques, les journaux regorgent de graphiques illustrant des distributions. La statistique devient un outil de communication de plus en plus exploité au sein de notre société. (Konold & Higgins (2003), Cobb (1999)) Ainsi, chaque individu est confronté de plus en plus à ce phénomène grandissant, la statistique au quotidien. Donc, il est nécessaire que tout citoyen soit en mesure de comprendre et d'interpréter une communication statistique puisque nous sommes appelés à interpréter continuellement des données et des graphiques statistiques, à juger de la validité des résultats. En fait, sans ces savoirs, il s'avère difficile de nos jours de participer aux différents débats actuels concernant l'éducation, la santé, l'environnement, etc. (Konold & Higgins, 2003) Ainsi, les savoirs relatifs au champ des statistiques peuvent permettre aux élèves d'approcher la dimension culturelle du savoir mathématique et de l'activité mathématique afin de saisir leur importance dans la société d'aujourd'hui (sous-thème 2). Suite à ce qui précède, le questionnement suivant s'impose: l'enseignement actuel prépare-t-il nos élèves à ces activités? Les élèves sont-ils aptes à interpréter une distribution statistique?

La statistique est incluse dans le curriculum mathématique et enseignée par les enseignants de mathématiques dans plusieurs pays. Ainsi, les élèves apprennent les différents concepts mathématiques reliés à ce champ d'étude, mais trop souvent l'emphase est mise sur le calcul des différentes mesures (moyenne, médiane, écart type, etc.) plutôt que sur le développement des raisonnements statistiques nécessaires à la compréhension et à la communication des informations (perspective axée sur une préparation technique en lien avec le sous-thème 2).

Une analyse du curriculum du secondaire au Québec nous a révélé que les notions statistiques étaient présentées de façon morcelée d'année en année. (Gattuso, 2001) Dans la première année, on présente différentes représentations graphiques et tableaux aux élèves. En deuxième année, les statistiques sont mises de côté pour laisser place à l'étude des probabilités. En troisième année, les mesures de tendance centrale (moyenne, médiane et mode) ainsi que l'histogramme sont abordés. Finalement, les mesures de position (les quartiles, les centiles, etc.) sont vues dans la quatrième année et l'étude de la corrélation ainsi que l'écart type dans la cinquième année.

En aucun temps on intègre l'ensemble des notions apprises afin d'en arriver à une analyse complète. «Statistics requires a different *kind* of thinking». (Cobb & Moore, 1997). Il est possible que le fait de promouvoir une approche où l'on étudie de façon séparée les différents concepts en statistiques n'incite pas nécessairement à une analyse globale et au développement de la pensée statistique. Sans parler de l'inférence statistique, qui est d'une plus grande complexité, il est possible de lire et d'analyser adéquatement une distribution statistique en s'appuyant sur les concepts de base. Toutefois, selon notre expérience d'enseignant, les élèves ont tendance à percevoir les données comme étant uniquement des nombres, négligeant ainsi le contexte auquel ces nombres sont associés. Aussi, lors de la comparaison de deux distributions, les élèves se limitent régulièrement au calcul de la moyenne comme fin d'analyse.

Notre étude visait à découvrir si les élèves qui terminent leurs études secondaires étaient en mesure d'interpréter et d'analyser adéquatement une distribution statistique. Pour ce faire, nous avons tenté de vérifier si le mode de représentation de la distribution (1. graphique, 2. liste de données, 3. les mesures caractéristiques de la distribution) influençait les réponses des élèves, si les raisonnements des élèves tenaient compte du contexte de la situation dans leur analyse et si leur raisonnement révélait une pensée statistique telle que décrite dans Cobb & al. (1998). Par exemple, un raisonnement multiplicatif relié au contexte des données surpasse un raisonnement basé uniquement sur un seul point de la distribution ou sur une seule mesure (généralement la moyenne) ou encore sur un raisonnement additif.

MÉTHODOLOGIE

Un questionnaire comprenant 7 questions demandant aux élèves de comparer deux ou trois distributions et d'opter pour l'une d'elle fut présenté à 141 élèves de 5^e secondaire (11^e année au Québec) provenant de cinq écoles différentes. Les sept questions (voir en annexe) portent sur les contextes suivants présentés selon l'ordre d'apparition dans le questionnaire: durée de vie de deux modèles de piles, points obtenus par deux joueurs de basket-ball, température à Boston au mois d'avril et au mois de septembre afin d'organiser un voyage, volume de différents pains, vitesse des automobiles obtenue par radar, le taux de cholestérol avant et après une diète et traitement pour les personnes atteintes du VIH (« T-cells »).

Un groupe de chercheurs (McClain, 1999; McGatha, Cobb & McClain, 1999; Cobb, 1999) ont développé une expérimentation en classe basée sur certaines activités qui mettaient l'emphase sur la notion de distribution à partir de différentes représentations graphiques. Les résultats de l'expérimentation ont démontré notamment que les élèves avaient plusieurs types de raisonnements lors de l'analyse d'une distribution statistique. Ces élèves se trouvaient dans une classe de 7^e année. À partir des différentes descriptions rapportées suite à ces activités (id.), nous avons établi quatre niveaux de justification, de celle qui est déficiente à celle qui reflète une pensée statistique adéquate. Pour le niveau 0, nous avons placé toutes les justifications inadéquates ou hors champ comme : « nous ne pouvons comparer les deux distributions puisqu'elles ne comportent pas le même nombre de données ». Les justifications basées sur un seul point de la distribution ainsi que les raisonnements additifs furent regroupés dans le niveau 1. Par exemple, « la pile A est meilleure puisque la durée maximale pour une pile est plus grande que celle de la pile B » ou « le joueur B est meilleur puisqu'il a 5 parties au-dessus de 20 points ». Pour ce qui est du niveau 2, nous avons regroupé les justifications basées sur une mesure de tendance centrale ou une mesure de dispersion puisque celles-ci tiennent compte de l'ensemble de la distribution. Toutefois, l'utilisation d'une seule mesure n'est pas toujours considérée comme suffisante pour fin d'analyse. Le niveau 3 comporte les justifications qui utilisaient une combinaison de mesures de tendance centrale et de mesures de dispersion ou encore les raisonnements multiplicatifs qui pouvaient être exprimés par un pourcentage ou par des arguments qualitatifs comme : « la majorité de... ». De plus, chaque justification était codée de façon à savoir si le raisonnement de l'élève était relié ou non au contexte de la situation.

Par ailleurs, puisque nous voulions vérifier l'influence du mode de représentation de la distribution sur les raisonnements des élèves, les questions furent formulées sous trois formes différentes : une liste de données, une représentation graphique et les mesures caractéristiques la distribution (moyenne, médiane, les quartiles, le minimum, le maximum, l'écart type, l'étendue ainsi que la somme des données). Chacune des trois versions du questionnaire proposait une variété de modes de représentation. Cependant, puisque nous voulions savoir si tous nos groupes étaient homogènes, nous avons utilisé une représentation graphique identique pour les questions 5 et 7 et ce, pour les trois versions du questionnaire.

RÉSULTATS

Homogénéité

Tout d'abord, il fallait vérifier si nous pouvions considérer nos groupes comme étant homogènes. Pour ce faire, nous avons vérifié le niveau de justification ainsi que les liens avec le contexte de la situation pour les questions 5 et 7 qui étaient identiques pour l'ensemble des questionnaires. Nous avons pu constater qu'il n'y avait pas de différence significative pour ces variables entre les différents groupes. De ce fait, nous les avons donc considéré comme étant homogènes.

Contexte

La proportion des élèves qui ont relié leur justification au contexte de la situation est importante. En général, 72,93% des élèves l'ont fait. Toutefois, les résultats contredisent à première vue l'ordre de difficultés de présentation des questions, mais d'autres variables doivent également être considérées (taille des échantillons, etc.).

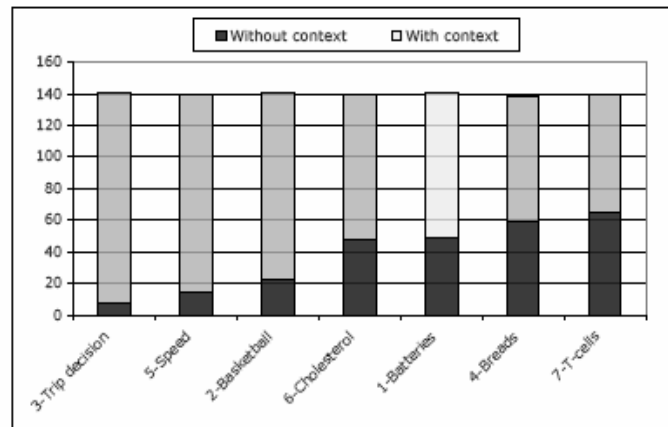


Figure 1. La référence au contexte selon chaque situation.

Niveau de raisonnement

Les résultats ont démontré globalement que 41% des élèves présentaient un raisonnement de niveau 2 qui tient compte de l'ensemble de la distribution. Ceci s'accorde avec l'enseignement de la statistique fait dans les écoles où l'emphase est souvent mise sur le calcul des mesures et où les notions sont présentées de façon morcelée rendant ainsi difficile la création de liens entre elles.

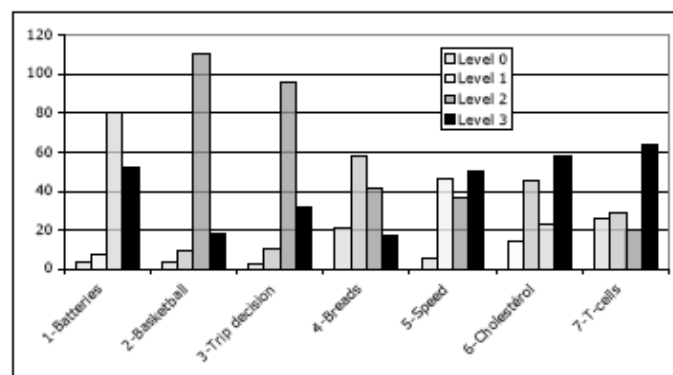


Figure 2. Nombre d'élèves selon le niveau de raisonnement atteint et ce, pour chaque question.

Aussi, si l'on considère que le niveau 3 reflète une pensée statistique adéquate, alors près du tiers des élèves ont justifié leur opinion à l'aide d'une combinaison de mesures ou par le biais d'un raisonnement multiplicatif. Il est important de souligner que certaines questions considérées a priori plus difficiles ont généré davantage de justifications de niveau 3.

CONCLUSION

En résumé, nous pouvons dire que certains élèves ont réussi à développer une pensée statistique adéquate. Aussi, l'utilisation du contexte est importante et pourrait être le résultat de l'emphasis mise sur les problèmes contextualisés dans les écoles. De façon générale, une réponse de haut niveau est reliée à l'utilisation du contexte. Les situations et le mode de présentation influence le niveau de justifications et le mode graphique semble donner plus de difficultés aux élèves.

Ces résultats suggèrent qu'il serait important d'utiliser des situations plus complexes en classe et surtout de donner plus d'emphasis aux représentations graphiques. De plus, nous nous rallions aux auteurs qui encourage les comparaisons entre plusieurs échantillons de données parce qu'elles incitent à focaliser sur les caractéristiques de la distribution (Konold, Higgins, 2003) et suscitent des raisonnements plus globaux (Russell, Schifter, & Bastable, 2002).

Nous aurons l'opportunité d'approfondir le tout lors de la présentation, expliquant du même coup plus en profondeur les résultats décrits au sein des deux graphiques présentés précédemment et ce, pour chacune des questions.

RÉFÉRENCES

Cobb, P. (1999). "Individual and collective mathematical learning: The case of statistical data analysis". *Mathematical Thinking and Learning*, 1, p. 5-44.

Cobb, W.G., Moore, D.S. (1997). "Mathematics, Statistics, and Teaching", *The American Mathematical Monthly*, Vol. 104, No.9, p. 801-823.

Gattuso, L., (2001). «L'insegnamento della statistica in Québec». *Induzioni*. (2) p. 43-56.

Konold, C. and Higgins, L.T. (2003). "Reasoning about data" In J. Kilpatrick, W.G. Martin and D. Schifter (eds). *A Research Companion to Principles and Standards for School Mathematics* (p. 192-215) Reston, VA: NCTM.

Konold, C., Pollastsek, A., & Gagnon, A. (1996). *Students' analyzing data: Research of critical barriers*. Paper presented at the Roundtable in mathematics education (p. 53-74). Dordrecht, The Netherlands: Kluwer.

McClain, K. (1999, April). *An analysis of the teacher's proactive role in supporting students' ability to reason statistically about data*. Paper presented at the Annual Meeting of American Education Research Association, Montreal, Canada.

McGatha, M., Cobb, P., & McClain, K. (1998, April). *An analysis of students' statistical understandings*. Paper presented at the Annual Meeting of the American Education Research Association. San Diego, CA.

McGatha, M., Cobb, P., & McClain, K. (1999, April). *Starting points of a learning trajectory*. Paper presented at the Annual Meeting of the American Education Research Association. Montreal, Canada.

Russell, S.J., Schifter, D. & Bastable, V. (2002). *Developing mathematical ideas: Working with data*. Parsippany, NJ: Dale Seymour Publications

Russell, S.J., Schifter, D. & Bastable, V. (2002) *Developing mathematical ideas: Working with data*. Parsippany, NJ: Dale Seymour Publications

Tukey, J W (1977). *Exploratory Data Analysis*. Reading, MA : Addison-Wesley

Vermette, Sylvain (2004). *Analyse et interprétation de distributions statistiques chez les élèves qui terminent leurs études secondaires*. Master dissertation, Department of mathematics, UQAM (Université du Québec à Montréal)

ANNEXE*

* Seulement une version de chaque question est présentée afin de ne pas alourdir inutilement le texte.

Question 1

Selon le tableau suivant, présentant la durée de vie (en heures) de 10 piles pour 2 modèles différents, quel modèle choisiriez-vous? Justifiez votre réponse.

Durée de vie (en heures) de 10 piles

Modèle 1	115	103	97	107	121	103	84	104	91	101
Modèle 2	115	74	115	113	111	45	115	106	115	96

Question 2

Un joueur doit être choisi pour la partie étoile de basket-ball. Le tableau suivant présente les points obtenus par les deux meilleurs candidats lors des huit dernières parties. En vous basant sur ces informations, quel joueur choisiriez-vous pour participer à cette partie ? Justifiez votre choix.

Points obtenus par les 2 meilleurs candidats

	Partie 1	Partie 2	Partie 3	Partie 4	Partie 5	Partie 6	Partie 7	Partie 8
Joueur A	11	31	16	28	27	14	26	15
Joueur B	21	17	22	19	18	21	22	20

Question 3

Certains étudiants ont planifié d'organiser un voyage à Boston pour la prochaine année scolaire. Deux (2) choix s'offrent à eux : une semaine en septembre ou une semaine au mois d'avril. Voici les températures (en Fahrenheit) de ces semaines au cours des trois dernières années.

SEPTEMBRE	<i>Année</i>	<i>Température</i>							<i>Moyenne</i>
	2001	77	71	80	75	73	79	77	Moyenne: 76
	2000	75	71	77	80	76	70	69	Moyenne: 74
	1999	79	71	74	79	77	73	72	Moyenne: 75

AVRIL	<i>Année</i>	<i>Température</i>							<i>Moyenne</i>
	2001	84	80	76	76	68	62	58	Moyenne: 72
	2000	79	83	87	90	89	86	81	Moyenne: 85
	1999	56	57	68	73	79	80	84	Moyenne: 71

Selon les étudiants, il serait préférable d'organiser le voyage au mois de septembre. Quels sont les critères qui peuvent justifier leur choix? Êtes-vous en accord avec leur choix? Justifiez votre réponse.

Question 4

Une entreprise produisant de la pâte à pain veut expérimenter trois méthodes de fabrication de la pâte, afin d'optimiser ses coûts de production. Elle s'intéresse, dans un premier temps, au volume de pâte obtenu après une heure d'action de la levure. Pour chacune des trois méthodes, un échantillon constitué de plusieurs morceaux de pâte à pain est fabriqué et le volume de chacun des pains est mesuré. Ainsi, on obtient plusieurs données mettant en relief le nombre de pains obtenus en fonction de leur volume respectif. En vous basant sur les informations ci-dessous, quelle méthode choisiriez-vous ? Justifiez votre réponse.

Mesures sur les trois distributions

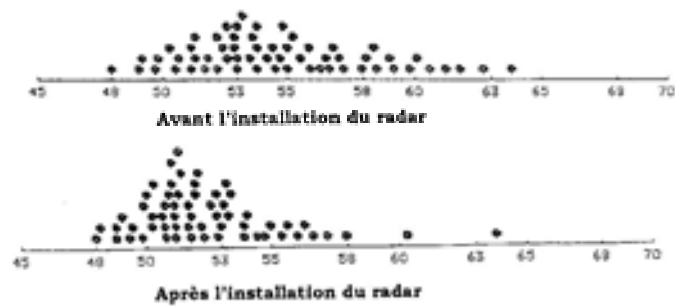
<i>Mesures</i>	<i>Méthode 1</i>	<i>Méthode 2</i>	<i>Méthode 3</i>
<i>Moyenne</i>	449,47	449,41	449,46
<i>Écart type</i>	102,40	102,43	102,47
<i>Médiane</i>	452	420	536
<i>Premier quartile</i>	375	382	361
<i>Troisième quartile</i>	524	492	536
<i>Borne inférieure</i>	125	325	225
<i>Borne supérieure</i>	725	1425	675

Question 5

L'introduction, à l'intérieur d'une zone de 50 km/h, d'un radar indiquant la vitesse des automobiles peut-il contribuer à ralentir la vitesse des automobilistes, contribuant par le fait même à réduire les risques d'accident? Les inspecteurs de police se sont interrogés sur cette question. Les graphiques ci-dessous illustrent la vitesse (en km/h) de 60 automobiles à l'intérieur d'une zone scolaire. Dans un premier temps, les inspecteurs se sont interrogés sur la vitesse des

automobilistes avant l'installation du radar et, dans un deuxième temps suite à son installation. En vous basant sur ces informations, croyez-vous que l'implantation d'un radar témoigne d'une grande utilité? Justifiez votre point de vue.

Vitesse (en km/h) de 60 automobile à l'intérieur d'une zone scolaire



(Cobb, 1999, p. 18)

Question 6

De nos jours, nous assistons à l'émergence d'une multitude de programmes de diètes. Un de ces programmes est désigné afin de réduire le taux de cholestérol pour les personnes susceptibles d'avoir des problèmes cardiaques. Le tableau suivant présente les différentes mesures relatives au taux de cholestérol pour 60 personnes ayant suivi le programme de diète. Les mesures sont réparties en deux catégories soit le taux de cholestérol avant l'application du programme et le taux de cholestérol un mois après avoir suivi le programme. En vous basant sur ces mesures, quelle est votre opinion sur l'efficacité de ce programme? Justifiez votre opinion.

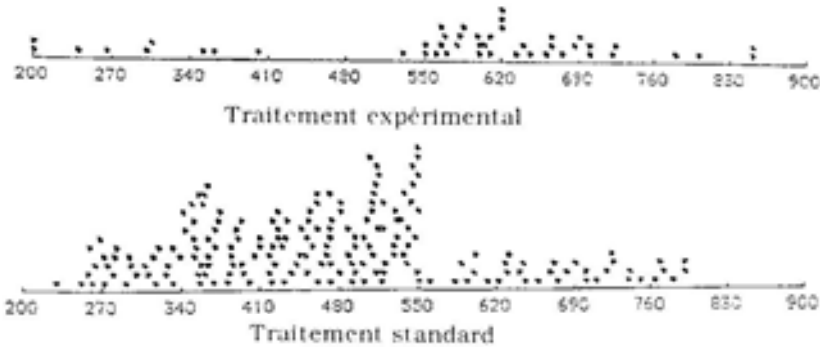
Mesures sur les deux distributions

Mesures	Avant le programme	Après le programme
Moyenne	231,05	209,75
Q1	209,5	185,5
Médiane	235,5	206,5
Q3	261,5	229
Minimum	121	124
Maximum	310	310
Étendue	189	186
Écart-type	43,51	39,14

Question 7

Plusieurs recherches expérimentales sont en vigueur afin d'en arriver à enrayer certaines maladies. Tel est le cas pour le sida. À l'intérieur de cette situation, vous aurez à comparer deux traitements pour des personnes atteintes du VIH. Dans le cadre d'une étude, 46 personnes ont eu l'occasion de suivre un programme expérimental et 186 personnes ont suivi le traitement standard. Dans chacun des cas, nous nous sommes interrogés sur le nombre de T-cell, une catégorie de globules blancs présents dans le sang qui améliore le système immunitaire. Donc, plus le nombre de T-cell est grand, meilleure est la condition de la personne. En vous basant sur les informations ci-dessous, que pensez-vous de l'efficacité du programme expérimental? Justifiez votre point de vue.

Nombre de T-cell chez certaines personnes atteintes du VIH



(Cobb, 1999, p. 21)