

Statistique et compréhension de la vie quotidienne: un objectif pour l'enseignement de la statistique

Alain Bihan-Poudec

► **To cite this version:**

Alain Bihan-Poudec. Statistique et compréhension de la vie quotidienne: un objectif pour l'enseignement de la statistique. 42èmes Journées de Statistique, 2010, Marseille, France, France. 2010. <inria-00494670>

HAL Id: inria-00494670

<https://hal.inria.fr/inria-00494670>

Submitted on 24 Jun 2010

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

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

STATISTIQUE ET COMPRÉHENSION DE LA VIE QUOTIDIENNE : UN OBJECTIF POUR L'ENSEIGNEMENT DE LA STATISTIQUE

Alain Bihan-Poudec

Institut des Sciences de l'Éducation d'Angers
Laboratoire de Recherche en Éducation et Formation
Université Catholique de l'Ouest
3, place André-Leroy – BP 10808
49008 ANGERS CEDEX 01
alain.bihan-poudec@uco.fr

Résumé : Un des objectifs assignés à l'enseignement de la statistique est la formation de futurs citoyens à même de décrypter les informations qu'ils reçoivent, notamment par les médias, et ce afin de favoriser leur esprit critique. Dans les cursus universitaires en sciences humaines et sociales, l'enseignement de la statistique remplit-il cet objectif ? Un exemple concret de validation dans un cursus de Licence en Sciences de l'Éducation nous permettra d'aborder cette question : partant de classements de films sur internet sur des sites cinématophiles, plusieurs questions étaient soumises aux étudiants : à quelle caractéristique de position correspond le nombre d'étoiles attribué aux films ? Quelles limites et quelles suggestions les étudiants pouvaient faire par rapport à cette procédure d'évaluation des films ? Au-delà des notes obtenues, les commentaires font apparaître des pré-conceptions surprenantes : déni de la variabilité des échantillons, nécessité de données quantitatives pour évaluer, croyance dans la nécessité d'une précision maximale, etc. ces *verbatim* confortent la nécessité pour l'enseignant d'identifier les conceptions préalables des étudiants.

Mots-clefs : enseignement de la statistique ; numérisme ; pédagogie universitaire ; pré-conceptions.

Abstract : One of the main aims of statistical education is the formation of future citizens able to decrypt the information they receive, particularly in the media, in order to promote critical thinking. In university courses in humanities and social, does statistics teaching fulfill this goal? An example of validation in a course of Bachelor in Educational Sciences will enable us to address this question. Hence ranking of a movie on Internet sites, several questions were presented to students: what is the meaning of the number of stars awarded to the film? What limits and what suggestions students could do about this evaluation process? Beyond assessment, comments reveal surprising pre-conceptions: Denial of the variability of samples, need quantitative data for evaluating, belief in the need for accuracy, etc. these *verbatim* reinforce the need for teachers to identify students' prior conceptions;

Keywords: misconceptions; statistical literacy; statistics education research; teaching statistics; teaching in university.

Un des objectifs assignés à l'enseignement de la statistique est la formation de personnes à même de décrypter les informations qu'ils reçoivent, notamment par les médias, et ce afin de favoriser leur esprit critique. Bref, il s'agit de former un *statistical citizen* (Rumsey, 2002).

Le numérisme¹ (*statistical literacy*) a été l'objet de communications (telle la journée d'études de la SFdS sur Angers, voir par exemple Padieu, 2006), d'articles, de livres. Ce terme peut être défini ainsi :

“...the term ‘statistical literacy’ refers broadly to two interrelated components, primarily (a) people’s ability to *interpret and critically evaluate* statistical information, data-related arguments, or stochastic phenomena, which they may encounter in diverse contexts, and when relevant (b) their ability to *discuss or communicate* their reactions to such statistical information, such as their understanding of the meaning of the information, or their concerns regarding the acceptability of given conclusions. These capabilities and behaviors do not stand on their own but are founded on several interrelated knowledge bases and dispositions...” (Gal, 2002, p.2-3).

Au niveau des modalités, Gattuso (2003) préconise une proximité des données utilisées lors des cours avec les informations présentes dans le quotidien. En effet, selon elle,

« L'enseignement de cette discipline ne peut se limiter à la transmission de techniques de calcul, mais doit mettre l'accent sur la compréhension de concepts et l'acquisition d'une pensée statistique à travers la pratique des méthodes utilisées dans cette discipline. [...] Trop souvent, les statistiques sont présentées hors de leur contexte "normal", c'est-à-dire les données réelles, et plutôt comme une procédure que comme une analyse globale d'une situation à l'aide d'outils statistiques. Hors [sic], il est indispensable que les élèves soient en mesure d'apprécier le rôle du raisonnement statistique qui permet de lire et interpréter des phénomènes réels sans vouloir apporter des réponses exactes mais en donnant toutefois la possibilité d'avoir une vision de la réalité qui ne soit pas que subjective » (Gattuso, 2003).

Qu'en est-il réellement ? Un enseignement permet-il l'acquisition de ce numérisme ? Quelles limites et difficultés sont-elles rencontrées ?

Incidentement, un contrôle en statistique auprès d'étudiants en Licence en sciences de l'éducation (troisième année) peut être l'opportunité d'aborder ces questions.

Un contrôle de statistique

Le sujet était organisé ainsi : il partait de la capture sur internet du classement d'un film (cf. annexe):



Madagascar 2

réalisé par Eric Darnell, Tom McGrath

avis des spectateurs ★★★★★

Animation (1h 35min)

¹ À l'instar de la commission générale de terminologie et de néologie de l'Académie française donnant comme équivalent français au terme « literacy » celui de « littérisme » (J.O. 30/08/2005). (Bihan-Poudec, 2008)

Rien ne permettant de dire si ces quatre étoiles correspondent au mode, à la médiane ou à la moyenne des avis des spectateurs, il était demandé aux étudiants de différencier ces dernières quant à leurs sens et propriétés.

Étaient soumis ensuite aux étudiants une distribution pour laquelle ils devaient déterminer les caractéristiques de position ; une distribution différente leur était alors soumise pour laquelle ils devaient opérer les mêmes calculs.

Les résultats ne différant pas, leur était demandé si les sites donnaient toutefois la même information, si les distributions étaient identiques.

Les étudiants étaient invités alors à déterminer les caractéristiques de dispersion (qui, elles, diffèrent) et à nouveau à comparer les distributions

Une cinquième question interrogerait sur le présupposé fait par les sites de considérer l'étoile comme étant une unité de mesure constante. Les étudiants pouvaient y répondre théoriquement et aussi pragmatiquement, en notant si les conclusions différaient selon que le classement du film relevait d'une échelle ordinale ou d'une échelle de rapport.

Ensuite, une troisième distribution était proposée : elle ne différait de la première que par les caractéristiques de position (translation d'une unité).

En fin, une dernière question les invitait à faire part de leurs remarques quant à la validité des évaluations et à suggérer des améliorations.

Quand les étudiants écrivent

Trente-deux copies ont été sélectionnées de par leurs développements aux questions ouvertes ; leurs réponses ont été classées en quatre catégories principales (même si parfois des liens apparaissaient entre ces dernières) :

Catégories :	de la subjectivité	la quantification	la variation comme erreur	l'échantillonnage
Fragments retenus	E1b, E3, E7a, E7b, E9, E11b, E13b, E14b, E16b, E19a, E21, E22B	E1d, E3a, E6a, E7b, E11a, E13b, E16b	E1a, E4a, E6a, E13c, E12a, E15a, E16a, E20a, E22a, E22c, E24a	E8c, E12a, E13a, E20a

Tableau 1 : répartition des fragments de discours des étudiants selon leur catégorisation

Place de la subjectivité

Une critique récurrente réside dans la dimension subjective de l'évaluation du film « *chaque spectateur a sa subjectivité, son avis qu'il est difficile de quantifier en nombre d'étoiles* » (E1b) ; elle se précise : « *on ne peut savoir ce que représente une étoile dans l'esprit des spectateurs* » (E3a), « *pourquoi obter [sic] pour une étoile plus que pour 2 étoiles ?* » (E6a) ; ou encore « *on ne connaît pas la valeur réelle d'une étoile. On ne sait pas à quoi elle correspond vraiment* » (E19a). Un étudiant (E9, fragment b) n'écarte pas le risque que pour le même avis deux spectateurs recourent à deux modalités distinctes.

Propriétés de mesure	À chaque individu correspond une modalité et une seule		Les modalités sont distinctes	Les modalités sont ordonnées	L'écart entre modalités a un sens	Le rapport entre modalités a un sens
	Niveau de mesure					
Échelle nominale	oui	oui	non	non	non	non
Échelle ordinale	oui	oui	oui	non	non	non
Échelle d'intervalles	oui	oui	oui	oui	oui	non
Échelle de rapports	oui	oui	oui	oui	oui	oui

Tableau 2 : propriétés des différents niveaux de mesure (cf. Stevens in Tiberghien, 1984, p. 57-59)

Ces remarques sont fondées et questionnent même l'idée de mesure (propriétés 1 et 2). Cependant, les suggestions d'amélioration du mode de recueil des avis montrent que les étudiants ne vont pas jusqu'à y renoncer, mais elles vont dans deux directions diamétralement opposées comme nous le verrons.

La quantification

Une des questions du contrôle interrogeait le statut de l'étoile comme mesure : correspond-elle à une unité de mesure constante ? Si oui, nous entrions dans le domaine du quantitatif, si non, nous serions en présence d'une échelle ordinale. La plupart des étudiants ont souligné le paradoxe selon lequel ils ne sont pas assurés d'avoir une unité de mesure constante mais que le quantitatif leur offrait la possibilité de calculer des indices de dispersion plus nombreux (écart moyen, écart-type) que s'ils n'avaient affaire qu'à un cas qualitatif. Toutefois, aucun n'est allé jusqu'à reconnaître que la conclusion de différence en hétérogénéité pouvait être obtenue avec l'étendue. Au travers des suggestions estudiantines d'amélioration, plusieurs positions se dégagent :

- remplacer les étoiles « par leurs significations » (E11a), tel par des mots comme « j'aime assez bien » ;
- remplacer les étoiles par des points (E3a), des notes (E13b, E25a) voire des classes de notes (E21a) ; un argument rencontré est « afin que tout le monde sache à quoi s'attendre » (E25a) ;
- établir une grille de critères (E1d) tels les effets spéciaux, l'intrigue, etc., quitte à « établir un coefficient d'importance à chaque critère pour enfin faire une moyenne et obtenir un chiffre décimal (E16b).

Plusieurs remarques peuvent être faites : l'orientation vers une mesure ordinale semble reposer sur un glissement sémantique « *la qualité du film me semble plus important que de quantifier à l'aide d'étoiles* » (E7b) ; la tendance à aller vers le connu (la notation) comme si le problème de la mesure se trouvait alors résolu (de manière illusoire, cf. les travaux en docimologie) ; quant au recours au quantitatif, il paraît être accompagné par la nécessité d'avoir une pensée analytique et une mesure précise. *Quid* alors d'un avis global du spectateur ?

La variation comme erreur

« *L'écart-type est en fait un indicateur de la « justesse » de ce que reflète en réalité ce nombre d'étoiles. Plus l'écart est grand, plus il y a d'écart entre ce qui est dit et la réalité* » (E1a).

« *Ce jugement [troisième distribution] est moins erroné que le premier. Nous avons le même nombre d'étoiles mais les écarts ont beaucoup moins importants (...). Il reflète plus la réalité* » (E13c)

« *Il [l'écart-type] ne doit pas être trop élevé pour signifier que la moyenne est correctement établie* » (E16a)

« *On peut dire que d'une distribution à l'autre l'estimation n'est pas faible. L'évaluation rend bien compte de l'ensemble de ces estimations mais la seconde distribution comprend des écarts trop importants entre les estimations... Les indices de dispersion éclairent donc les caractéristiques de position et confortent leur mauvaise fiabilité dans le cas présent* » (E22a et c)

« *Entre 2 et 4 étoiles existe un écart de 2 étoiles, l'unité de mesure étant une erreur [c'est nous qui soulignons], le rapport entre modalités a un sens* (E24a)

Tout se passe comme s'il existait une valeur vraie, réelle, telle que les écarts à cette valeur soient des erreurs. Comme s'il ne pouvait pas y avoir des appréciations différentes pour un même film ! Sans doute cette notion d'erreur peut-elle se comprendre, resituée dans la théorie

de l'échantillonnage, mais il s'agit ici de statistique descriptive. L'on pourrait se gausser de ces étudiants, y déceler un conformisme absolutiste, si ce n'est que des chercheurs, et non de moindres, ont postulé eux aussi l'existence d'une valeur vraie².

L'échantillonnage

Cette manière de concevoir la variation n'est pas sans conséquences sur la manière de comparer les sites. Ainsi, en observant que deux distributions ont les mêmes caractéristiques de dispersion mais diffèrent quant à leur distribution, est-il conclu que cela « *prouve bien que ces évaluations ne peuvent pas refléter [sic] la qualité des votes* » (E8c). Ou encore que le « *jugement [dans la troisième distribution] est moins erroné que le premier* » (E13c) au motif que les caractéristiques de dispersion y sont moins élevées. Un étudiant suggère -avec pertinence- de regrouper tous les avis mais c'est avec l'espoir que « *la note précise pourra être donnée* » (E20a). En d'autres termes, l'« idéal » des étudiants serait l'existence d'une seule valeur, vraie, réelle, absolue. Face à la variabilité des échantillons, soit ils « jettent le bébé avec l'eau du bain » en concluant à l'impossibilité de mesurer, soit en valorisant les échantillons ayant une moindre dispersion, les autres devenant « suspects »

Discussion

Pour aller plus avant dans cette recherche, il conviendrait de connaître le cursus antérieur de ces étudiants. En effet, il n'existe qu'une troisième année de licence en Sciences de l'Éducation, et les étudiants présents viennent de filières différentes (psychologie, communication, sociologie...). Notamment, il faudrait savoir s'ils ont eu des enseignements en statistique lors de leurs deux premières d'université ou si leur savoir se limite aux enseignements reçus en collège et lycée lors des cours de mathématique. Il serait paradoxal -et redoutable- de s'apercevoir que la conception de l'écart-type comme mesure de l'erreur soit le fait d'étudiants ayant suivi des cours sur l'échantillonnage. Nous aurions là une conception erronée mais acquise rétroactivement !

De plus, si notre analyse des *verbatim* présents dans des copies d'examen amène à identifier des éléments quant aux pré-conceptions des étudiants quant à la statistique, il ne faut pas écarter le risque que cette analyse amène une cohérence là où elle n'existe peut-être pas. Bref de construire artificiellement le discours d'un étudiant virtuel, gommant ainsi la possible divergence de conceptions.

Quoiqu'il en soit, les observations précédentes montrent bien que le challenge de l'enseignement de la statistique ne réside pas tant dans la maîtrise de ses procédures mais dans la capacité à savoir quand les utiliser (*cf.* Quilici et Mayer, 1996, p. 144) ; elles montrent aussi, à l'instar de Shaughnessy (2006) que les étudiants eux-mêmes sont une source inestimable pour enseigner.

Bibliographie

- [1] Rumsey, D. J. (2002). Statistical literacy as a goal for introductory statistics courses. *Journal of Statistics Education*, 10(3) Retrieved from <http://www.amstat.org/publications/jse/v10n3/rumsey2.html> database.
- [2] Bihan-Poudec, A (2008). L'apprentissage de la statistique à l'université : en recherche de sens ? In N. Bednarz et C. Mary (dir.), *L'enseignement des mathématiques face aux défis de l'école et des communautés. Actes du Colloque Espace mathématique francophone 2006*. Sherbrooke : Éditions du CRP.

² « Une fois la valeur vraie connue, on peut déterminer la précision d'un correcteur en confrontant sa notation avec la valeur vraie » (Piéron, 1963, pp. 22 sq.)

- [3] Padieu, R. (2006). In (Ed.), Pour une évaluation de la culture statistique. Communication à la Journée du Groupe "enseignement de la statistique" de la Société Française de Statistique, Université Catholique de l'Ouest, Angers, 22 mars.
- [4] Gal, I. (2002). Adults' statistical literacy: Meanings, components, and responsibilities. *International Statistical Review*, 70(1), 1-25.
- [5] Gattuso, L. (2003). *Les statistiques, un élément essentiel de la littéracie. une expérimentation d'enseignement des statistiques dans les écoles italiennes*. XXXV^{èmes} Journées de la statistique, Lyon, 2-6 juin. Retrieved from http://www.sfds.asso.fr/groupe/jslyon03/35JS_Lyon_Linda_Gattuso.pdf
- [6] Tiberghien, G. (1984). *Initiation à la psychophysique*. Paris: PUF.
- [7] Piéron (1963). *Examens et docimologie*. Paris, PUF
- [8] Quilici, J. L. et Mayer, R. R. (1996). Role of examples in how students learn to categorise statistics word problems. *Journal of Educational Psychology*, 88(1), 144-161.
- [9] Shaughnessy, J. M. (2006). *Student work and student thinking: An invaluable source for teaching and research*. 3rd International Conference on Teaching Statistics.

Annexe A : texte du contrôle de statistique

Traitez l'ensemble des questions

La statistique descriptive permet à son utilisateur d'organiser ses données, de les résumer et d'en tirer une vue d'ensemble. Elle devrait permettre aussi d'avoir un regard critique sur les informations que l'on reçoit.

Le sujet ci-dessous se situe dans cette perspective et prend comme point de départ une information banale telle que vous pouvez la trouver dans des quotidiens ou sur des sites Internet dédiés au cinéma. Ainsi avez-vous souvent, à côté des films présentés, une évaluation faite par les spectateurs comme celle-ci :



Madagascar 2

réalisé par Eric Darnell, Tom McGrath

avis des spectateurs 

Animation (1h 35min)

Il en ressort que le film a reçu quatre étoiles sur cinq possibles. Toutefois, pour un lecteur averti, cela n'est pas sans poser de problème car il ne sait pas si cet indice correspond à un mode, une médiane ou une moyenne.

Question 1 : quelle information chacune des caractéristiques de position mentionnées ci-dessus apporte-t-elle ? Montrez-en les différences.

En fait, les quatre étoiles ont été obtenues à partir de la distribution suivante :

Nombres d'étoiles	Nombre de spectateurs
2	6
3	8
5	20

Tableau 1 : distribution A

Question 2 : déterminez-en les caractéristiques de position (nous supposons qu'il existe une unité de mesure constante).

Un autre site fournit la distribution suivante (B) pour le même film :



Nombres d'étoiles	Nombre de spectateurs
0	6
1	1
5	27

Tableau 2 : distribution B

Question 3 : là aussi, déterminez les caractéristiques de position. En les confrontant avec celles que vous avez obtenues à la question 2, que constatez-vous ? Quelle évaluation (nombre d'étoiles) figurerait sur le site Internet ? Si quelqu'un vous disait que les sites donnent la même information quant à l'évaluation du film, que lui répondriez-vous en prenant en compte les distributions ?

Question 4 : déterminez les caractéristiques de dispersion pour les deux distributions (A et B). Quelles conclusions pouvez-vous en tirer de leur comparaison ?

Question 5 : la plupart des sites calcule un nombre moyen d'étoiles. Ce faisant, ils supposent qu'il existe une unité de mesure constante, l'étoile. D'après vous, est-ce justifié ? À partir des caractéristiques de position et de dispersion que vous avez déterminées pour les deux distributions (A et B), le fait de considérer l'évaluation comme quantitative apporte-t-elle un plus par rapport à une évaluation conçue comme ordinale ?

Un troisième site fournit les éléments suivants. Madagascar : avis des spectateurs   ; répartition des avis des spectateurs :

Nombres d'étoiles	Nombre de spectateurs
1	6
2	8
4	20

Tableau 3 : distribution C

En sont déduits les éléments suivants :

<i>Caractéristiques de position :</i>	mode : 4 étoiles	médiane : 4 étoiles	moyenne : 3 étoiles	
<i>Caractéristiques de dispersion :</i>	étendue : entre 1 et 4 étoiles, soit un écart de 3	écart interquartile : entre 2 et 4 étoiles, soit un écart de 2	écart moyen : 1,18 étoiles	écart type : 1,24 étoiles

Question 6 : en comparant ces indices avec ceux de la distribution A, quelles conclusions pouvez-vous tirer ?

Question 7 : à l'issue de ce contrôle, quelle remarque finale pouvez-vous faire sur la validité des évaluations des films par les spectateurs ? Quelles améliorations pourriez-vous suggérer ?