

D’OÙ VIENNENT LES DONNÉES ?

UNE LACUNE PARADOXALE EN ÉDUCATION STATISTIQUE

Noëlle Zendreras¹, Véronique Dubreil-Fréront¹, Jean-Marie Marion² et Alain Bihan-Poudec³

¹ *Faculté des Sciences humaines et sociales, PESSOA noelle.zendreras@uco.fr*

² *Faculté des Sciences, MAI, PESSOA*

³ *Faculté d’Éducation, PESSOA*

UCO, Université catholique de l’Ouest, campus d’Angers, 3, pl. André Leroy BP 10 808, 49008-Angers cx-1 France

MAI, équipe de recherche Mathématiques Appliquées et Informatique

PESSOA, équipe de recherche Pédagogies Socialisation et Apprentissages ;

- Projet StatEns : Pédagogie de la statistique à l’université -

Résumé. Cette communication rejoint l’idée d’une trop faible attention des curricula de Statistique au regard du recueil de données ainsi que des données elles-mêmes. La littérature du domaine de l’éducation statistique a déjà pointé l’inadéquation entre les contenus des cours et les besoins. Nos propres constats en sciences humaines et sociales et nos premiers résultats d’une analyse de contenu de manuels adaptés à ces filières nous amènent à étayer l’existence de lacunes et à relever notamment un glissement de la discipline Statistique vers le seul traitement des données. Alors que les questions des données (nature, propriétés), de leur recueil (échantillonnage, méthodes, outils) et de leurs qualités respectives constituent des piliers incontestables de cette science, ces pans se retrouvent relégués à des places minoritaires et excentrées. Cette situation constitue à nos yeux un véritable paradoxe, qui revêt des implications pour l’enseignement-apprentissage, mais aussi pour la recherche et la pratique. En effet, cette lacune suscite nombre de difficultés de conceptualisation et d’apprentissage et elle compromet aussi le développement du raisonnement, de l’esprit statistique. Des stratégies pédagogiques et didactiques palliatives sont discutées.

Mots-clés. Recherche en éducation statistique, données, recueil, qualité, traitement, apprentissage, supérieur, pédagogie, didactique, données réelles, projets, travaux collaboratifs.

Abstract. *Where do the data come from? A paradoxical gap in statistics education.* As statistics teachers of undergraduate students in social sciences, we analyse the contents and methodology of statistics books. We then observed that teaching statistics focuses on data analysis without any reference to applied context whereas statistics is not only a part of mathematical science, but a methodological discipline which gives tools for dealing with data. We further discuss the consequences on students’ statistical thinking.

Keywords. Statistics education research, data, collection, quality, analysis, learning, teaching statistics, undergraduate degree, pedagogy, didactics, real data, projects, collaborative works.

1 Constat introductif

Voici une question qui traverse rarement l’esprit des apprenants qui suivent un cours de Statistique : « d’où viennent les données » ? Justement et paradoxalement les données statistiques « ne sont pas données » : elles doivent être recueillies, elles constituent le fruit d’une collecte guidée par un protocole strict. Cependant dans les cursus de statistique, les données sont le plus souvent préalablement fournies, et leur existence même, voire leur qualité, ne sont pas questionnées.

Avec nombre d’auteurs du domaine de l’éducation statistique, nous constatons dans le supérieur une forte inadéquation entre les contenus des formations et les besoins réels des étudiants, futurs professionnels (Mosteller, *in* Moore, 1993 ; Lahanier-R., 2003). Parmi ces contenus défailants, et partant

de nos constats et de notre propre pratique en tant qu'enseignants en sciences humaines et sociales (SHS), nous pointons ici tout particulièrement la faiblesse de l'enseignement du « recueil des données » mais aussi le manque d'approfondissement du concept même de « donnée statistique ».

2 Une lacune paradoxale : les données et leur recueil

Ce constat prend pour nous la forme d'une grave lacune dans l'éducation statistique dans le supérieur, en France certes, pays où nous exerçons, mais aussi dans de nombreux autres pays.

Un glissement... Cette lacune éducative repose à notre sens sur le glissement de la discipline Statistique vers le seul traitement statistique, vers la seule analyse des données. Les premiers résultats de notre analyse de contenu d'une vingtaine de manuels de statistique destinés aux étudiants du supérieur non spécialistes nous confortent dans ce constat. Les apprenants sont largement formés à traiter les données mais non pas à s'interroger sur l'origine (quoi, où, qui ?), la nature (qualitative, quantitative), le niveau de mesure (nominal, ordinal, intervalles, rapport), l'outil de collecte (observation, entretien, questionnaire, appareillage), le type d'échantillonnage ou encore la qualité de ces données.

Des piliers fondamentaux ignorés. Ces deux pans de la Statistique, les données et leur recueil, pourtant piliers fondamentaux et fondateurs de cette science, demeurent ainsi minoritaires et fortement ignorés dans les cours pour les non spécialistes. Les concepts et procédures liés aux données et à leur génération, s'avèrent souvent abordés en tout début des *curricula* et sont considérés comme acquis tout au long des formations. Cet "oubli" de l'enseignement vis-à-vis des données et de leur recueil revêt des conséquences très préjudiciables : or, paradoxe, comment s'assurer de la qualité des résultats d'une recherche, d'une enquête, si les données, matériau même, substrat sur lequel se fondent les conclusions, ne sont pas contrôlées ? C'est dans cet oubli des piliers fondateurs, les données et leur recueil, que nous situons la lacune éducative paradoxale.

Un autre impératif négligé... La question de la qualité des conclusions statistiques d'une étude pâtit également d'un enseignement minoritaire. La considération de la fidélité (fiabilité, *reliability*) des données et de la validité (*validity*) de l'instrument de mesure bénéficie de trop peu de place et de temps dans les cours. Alors qu'impératif absolu, la qualité des données et de l'étude statistique se retrouve implicitement négligée aussi dans les enseignements. L'humour de Tesson en fig.1 illustre bien la déroute dans laquelle se trouveront les étudiants, futurs professionnels ...



Figure 1. Le département de contrôle de la qualité...
(avec l'aimable autorisation de l'auteur, Luc Tesson, dessinateur de presse et illustrateur).

3 Des particularités en sciences humaines et sociales

Les sciences humaines et sociales, qui englobent notamment la psychologie, les sciences de l'éducation, la sociologie, l'ethnologie, présentent des particularités notoires en ce qui concerne les unités statistiques interrogées et la nature des variables mesurées.

Des unités et des variables singulières. Encore plus que dans d'autres champs, en SHS les unités statistiques interrogées correspondent en large majorité à des êtres humains, des personnes observées ou questionnées par l'expérimentateur, le chercheur. Les variables mesurées s'avèrent plutôt de type qualitatif : il s'agit d'avis, opinions, idées, sensations, goûts, mais aussi compétences, performances, savoirs-faire, résolutions de tâches, raisonnements, actions intellectuelle ou motrices, ...

Nombre d'études en sciences économiques, gestion, marketing ou publicité, présentent ces mêmes particularités. Ces spécificités liées à la nature des unités statistiques et des variables interrogées impliquent des risques de subjectivité et de biais encore plus importants que dans d'autres champs et exigent donc des précautions méthodologiques encore plus rigoureuses.

Subjectivité et Biais. Les biais de subjectivité englobent ici celle du sujet et du chercheur, en passant par celle de l'expérimentateur ou encore du traiteur des données. Les biais inhérents au sujet interrogé s'étendent des biais de primauté aux biais de désirabilité sociale, en passant par ceux de centralité, d'acquiescement, d'ignorances ou par les sociaux-culturels. Calibrer l'avis ou l'idée d'un être humain s'avère finalement bien plus délicat que mesurer la longueur d'une feuille ou doser un taux sanguin.

Comme décrit dans la littérature, les étudiants des filières non spécialistes rencontrent en statistique de nombreuses difficultés conceptuelles, procédurales et de raisonnement ; les étudiants en SHS n'échappent évidemment pas à ce constat (Zendreras, 2010 ; Bihan-P. et Marion, 2014 ; Dubreil-F. et al., 2014 ; Zendreras et Marion, 2015).

Des difficultés spécifiques ? Dans notre pratique nous relevons de façon récurrente auprès de nos étudiants un certain nombre de difficultés liées aux particularités des données en SHS, dont nous en étayons ici deux. D'une part, les termes "quantitatif" et "qualitatif" suscitent nombre de confusions et de conceptions erronées. A titre d'exemple, les étudiants nous signalent souvent : « je ne peux pas réaliser de traitement statistique parce que mes données sont qualitatives ». Se faisant, ils semblent restreindre le traitement statistique au traitement des seules données dites quantitatives. D'autre part, l'incompréhension liée aux échelles de Likert (cf. fig.2), particulièrement utilisées en SHS, constitue à notre sens une autre difficulté inhérente. Les étudiants ne semblent pas avoir pris conscience qu'à la base de la construction d'une telle échelle se situe un abus méthodologique, souhaité et accepté, mais qui accentue la subjectivité et l'imprécision des mesures. Notamment, peu d'entre eux ont compris que ces échelles reposent souvent sur la transformation d'une variable qualitative mesurée au niveau ordinal en une "pseudo variable quantitative discrète".

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Totalement en désaccord (0)	Plutôt en désaccord (1)	Plutôt en accord (2)	Totalement en accord (3)

Figure 2 : Un exemple d'échelle de Likert ; le sujet est invité à émettre un jugement en cochant une case.

4 Discussions : esprit statistique et stratégies pédagogiques

De nombreuses évidences apparentes demandent à être discutées ici, à commencer par l'origine étymologique du vocable « donnée » (*data*) et par le sens accordé au terme en statistique. Il conviendrait aussi de baliser ce que renferme la Statistique et de préciser son statut et son positionnement, partant des points de vue premiers (Fisher, 1922) aux plus récents (Hahn, 2015). L'éducation statistique constitue également un champ en soi (Batanero, 2001 ; Franklin et al., 2007 ; Garfield et Ben-Zvi, 2008). Et ce champ s'organiserait sur deux dimensions majeures des connaissances (Shulman, 1988) : les disciplinaires (*subject-matter content knowledge*) et les didactico-pédagogiques (*pedagogical content knowledge*).

Il faut aussi s'interroger sur la place explicitement accordée en leur sein à la question des données et leur recueil. Car la conscientisation de l'importance des données et implicitement de leur nature, leur recueil, constitue un élément fondateur et indispensable pour l'installation du raisonnement et de l'esprit statistique, du *sentido estadístico*, des *statistical reasoning and thinking* (Wild et Pfannkuch,

1999 ; Régnier, 2006 ; Batanero et al., 2013). Comme le suggéraient Cobb et Moore en 1997 en distinguant la statistique des mathématiques : « *Statistics requires a different kind of thinking, because data are not just numbers, they are numbers with a context.* ».

Pallier la lacune éducative relative aux données et à leur recueil pourrait s'atteindre en partie en augmentant la masse horaire dévolue à la discipline ; mais le système éducatif, tendu et avare en temps, tardera à être convaincu. Il convient donc aussi de mettre en œuvre des stratégies d'enseignement différentes. Depuis les années 1990, nombre d'auteurs préconisent et pratiquent de nouvelles approches pédagogiques et/ou didactiques efficaces pour favoriser les apprentissages.

Données réelles, Projets et Collaboration. Certaines de ces approches se révèlent à notre sens capables de contribuer à pallier des lacunes éducatives au regard des données, leur nature et leur recueil. Il s'agit de stratégies telles que l'enseignement fondé sur des « données réelles » (*real data* ; Mosteller, in Moore, 1993), l'apprentissage « avec projets » (*con proyectos* ; Batanero et Díaz, 2011) et les travaux sur un mode collaboratif et interactif (Allard, 1992 ; Gallese et al., 2000).

5 Conclusion

Les sondages, les "statistiques", les travaux scientifiques, notamment en SHS, ont régulièrement à essuyer des critiques de la part des médias. A cet égard, Gauvrit (2007, p.198) souligne trois faiblesses, « trois lieux possibles de l'attaque statistique : le recueil des chiffres, le choix des méthodes, les interprétations » ; l'auteur met ainsi en première ligne l'objet même de cette communication, la collecte des données ("le recueil des chiffres"). Une éducation statistique s'attachant à une prise en compte plus vive de la place prépondérante des données et de leur recueil ne saurait que rendre la science Statistique et les sciences qu'elle sert encore plus « respectables », au sens de Moore (1992).

Références

- [1] Allard, J. (1992). Une troisième voie dans l'enseignement de la Statistique en Sciences Humaines. *Bull.AMQ*, 32(2), 19-26.
- [2] Batanero, C. (2001). *Didáctica de la Estadística*. Granada: Univ. de Granada.
- [3] Batanero, C. et Díaz, C. (dirs.). (2011). *Estadística con proyectos*. Granada: Univ. de Granada.
- [4] Batanero, C. et al. (2013). Sentido estadístico: componentes y desarrollo. *Números*, 83, 7-18.
- [5] Bihan-Poudec, A. et Marion, J.-M. (2014). De la représentation de la statistique chez des étudiants en Sciences humaines et sociales. *Actes des JdS-46*. SfdS, Rennes. <http://papersjds14.sfds.asso.fr/submission_240.pdf>
- [6] Cobb, G. et Moore, D. (1997). Mathematics, Statistics, and Teaching. *American Mathematical Monthly*, 104, 9, 801-823.
- [7] Dubreil-Frémont., V., Chevallier-Gaté, C. et Zendera, N. (2014). Students' conceptions of average and standard deviation. *Proc. of ICOTS-9*. IASE, Flagstaff. <http://icots.info/9/proceedings/pdfs/ICOTS9_C142_DUBREILFREMONT.pdf>
- [8] Fisher, R.A. (1922). On the mathematical foundations of theoretical statistics. *Phil. Trans. of the R. Society A*, 222, 309-368.
- [9] Franklin, C. et al. (2007). *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report, a pre-K-12 curriculum framework*. Alexandria (VA): American Statistical Association.
- [10] Gallese, E. et al. (2000). Problemática sobre la enseñanza y aprendizaje de la estadística en carreras no estadísticas. 5^{as} *Jornadas Investigaciones Fac. de Ciencias Económicas y Estadística*, p.310-320. Rosario (Arg): Univ. Nacional de Rosario.
- [11] Garfield, J. et Ben-Zvi, D. (2008). The Discipline of Statistics Education. In J. Garfield et D. Ben-Zvi, *Developing Students' Statistical Reasoning: connecting research and teaching practice*, p.3-19. NY: Springer.
- [12] Gauvrit, N. (2007). *Statistiques. Méfiez-vous!* Paris: Ellipses.
- [13] Hahn, C. (2015). La recherche internationale en éducation statistique: état des lieux et questions vives. *Statistique et Enseignement*, 6(2), 25-39. [En ligne] <<http://www.statistique-et-enseignement.fr/>>
- [14] Lahanier-Reuter, D. (1999). *Conceptions du hasard et enseignement des probabilités et statistiques*. Paris: PUF.
- [15] Moore, D. S. (1992). Teaching Statistics as a Respectable Subject. In F. Gordon et S. Gordon, *Statistics for the twenty-first century*, p.14-25. Washington D.C.: Mathematical Association of America.
- [16] Moore, D. S. (1993). A Generation of Statistics Education: An interview with Frederick Mosteller. *J. Statistics Educ.*, 1(1).
- [17] Régnier, J.-C. (2006). Formation de l'esprit statistique et raisonnement statistique. Que peut-on attendre de la didactique de la statistique? In C. Castela et C. Houdement, *Actes du Sémin. Nat. de Didactique des Math. 2005*, p.13-37. Paris: ARDM.
- [18] Shulman, L.S. (1988). Paradigms and Research Programs in the Study of Teaching: A contemporary perspective. In M.C.

Whittrock, *Handbook of research on teaching*, p.3- 35. NY: Macmillan.

- [19] Wild, C. et Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67 (3), 223-265.
- [20] Zendreras, N. (2010). Human sciences students' difficulties in parametric tests. A contribution to Statistics Education. Invit. Pap. *Proc. of ICOTS-8*. IASE, Ljubljana. <https://www.stat.auckland.ac.nz/~iase/publications/icots8/ICOTS8_7C2_ZENDRERA.pdf>
- [21] Zendreras, N. et Marion, J.-M. (2015). Évolution des conceptions de l'écart-type chez des étudiants en SHS. *Actes du CFIES-4*. SfdS, Bordeaux. <http://paperscfies2015.sfds.asso.fr/submission_17.pdf>

Nous réitérons nos remerciements à M. Luc Tesson, qui nous a aimablement autorisés à reproduire sa vignette « *Quality Carton* » ; en ligne : <[Tesson, dessinateur de presse et illustrateur](#)>.