
Bertrand, R. et Valiquette, C. (1986). *Pratique de l'analyse statistique des données*. Sillery, Québec: Presses de l'Université du Québec, 379 pages.

Évalué par: Michel Pépin, Ph.D., Ecole de Psychologie, Université Laval

Pratique de l'analyse statistique des données se veut essentiellement un manuel d'introduction à la statistique destiné aux chercheurs ou étudiants ayant peu ou pas de connaissances dans ce domaine. L'ouvrage aborde donc les mêmes thèmes ou sujets que l'on retrouve dans la plupart des manuels d'introduction à la statistique. Toutefois, contrairement aux autres manuels du même genre, les auteurs accordent une place prépondérante aux techniques d'analyse et de représentation graphique des données. Cet ouvrage s'inspire largement des travaux de Tukey (1977), lequel favorise une approche de l'analyse des données orientée davantage vers la description, la représentation graphique et l'examen approfondi des données. Dans cette approche, l'analyse statistique des données devient plus un art que la simple application de techniques.

Avant de présenter et discuter le contenu de chacun des sept chapitres constituant l'oeuvre, voici quelques remarques générales concernant l'ensemble de cet ouvrage.

Tel que mentionné en avant-propos, ce manuel en est un d'introduction à l'analyse statistique des données. Personnellement, je crois qu'il y a déjà trop de volumes d'introduction et on peut sérieusement questionner l'apport ou l'utilité de ce nouveau volume. Evidemment, cette critique doit être nuancée du fait que le volume soit rédigé en français. Les auteurs, en évitant volontairement les développements algébriques, les preuves et les démonstrations, ont opté pour une approche pédagogique qui a le mérite d'être simple et de favoriser une compréhension intuitive des principaux contenus. Ce choix m'apparaît cependant discutable si l'objectif du volume est de former des gens à l'analyse statistique des données. En effet, le lecteur intéressé à mieux comprendre ses intuitions demeure sur son appétit. C'est un peu comme si l'expert évitait d'aborder à fond certains sujets parce qu'il suppose qu'il ne sera pas compris de ses lecteurs; le ton plutôt paternaliste, adopté pour nous demander de nombreux actes de foi, finit par agacer à la longue. Dans l'ensemble, cependant, le langage simple et le style d'écriture facilitent la compréhension du lecteur.

Les auteurs utilisent à profusion les méthodes graphiques (diagramme en feuilles, en bâtonnets, nuage de points, etc.) permettant une représentation visuelle des données de même qu'une compréhension intuitive des analyses statistiques qui y sont reliées. En fait, en utilisant ces techniques graphiques de façon quasi-abusive, les auteurs visent à contrecarrer une tendance déplorable dans le domaine de l'analyse statistique des données, tendance qui consiste à appliquer aveuglément divers tests statistiques sans avoir pris la peine d'examiner attentivement les données recueillies. On nous rappelle, indirectement, que le bon sens a aussi sa place dans le domaine de l'analyse des données et ceci constitue un des points forts du volume. Ainsi, pourquoi utiliser un test statistique pour comparer deux moyennes (test t ou analyse de variance) lorsque le simple examen des données nous indique que les moyennes observées sont à

toutes fins pratiques identiques? Pourquoi appliquer des tests inférentiels quand les faits observés (analyse descriptive) ne les justifient pas? Un peu comme Howel (1982), entre autres, le professeur Bertrand tient un discours extrêmement sensé sur tous les sujets abordés dans ce livre. On devine une préoccupation constante, soit celle de présenter l'analyse de données comme une activité intelligente, non-automatique, où le choix des statistiques à utiliser est tributaire de l'examen approfondi des données recueillies.

Examinons maintenant brièvement le contenu de chacun des sept chapitres du volume.

Le chapitre premier présente quelques généralités reliées à l'analyse statistiques des données (les principales étapes d'une recherche, le rôle du consultant, le rôle de l'informatique et les types de méthodes d'analyse des données). Dans cette introduction, comme dans presque tous les chapitres d'ailleurs, on nous met fortement en garde contre une utilisation abusive des progiciels statistiques, tels SPSS, SAS, BMDP, etc., progiciels qui permettent d'obtenir rapidement toutes les analyses de l'arsenal de la statistique. Il est vrai, comme les auteurs le mentionnent, que l'analyse des données et le traitement informatique des données sont deux activités distinctes, la seconde étant un puissant outil au service de la première. Toutefois, ces progiciels demeurent à notre avis d'excellents outils. Il est étonnant de voir les auteurs se montrer si rébarbatifs à leur utilisation. Que des gens utilisent ces outils de façon abusive s'avère fort probable; cela n'est pas une raison valable pour les discréditer ou encore pour vouloir restreindre leur usage. Par rapport à l'utilisation de l'informatique, l'ouvrage se situe à contre-courant, se démarquant nettement de la tendance actuelle qui incite les auteurs à intégrer divers programmes informatiques dans les volumes de statistiques, comme le font, par exemple, Cohen et Holliday (1982).

Les auteurs distinguent deux types d'approche reliées à l'analyse des données: l'approche exploratoire et l'approche confirmatoire. Cette distinction risque de semer un peu de confusion car le terme exploratoire est généralement associé à un type d'études (étude exploratoire où il n'y a pas d'hypothèses de recherche formelles). En fait, on qualifie d'approche exploratoire l'ensemble des statistiques descriptives, cette dernière expression nous apparaissant nettement plus appropriée. D'ailleurs les auteurs semblent mal à l'aise avec cette terminologie car ils affirment par la suite que les deux approches apparaissent alors bien plus comme deux phases (ordonnées) d'une même analyse que deux types d'analyse distincts. On comprend toutefois facilement ce que les auteurs tentent d'exprimer: une analyse statistique de données devrait toujours débuter par un examen attentif des données à l'aide des statistiques descriptives (les représentations graphiques en faisant partie); par la suite, les techniques statistiques confirmatoires ou inférentielles peuvent être utilisées afin de confirmer ou infirmer, s'il y a lieu, les hypothèses établies ou les nouvelles hypothèses établies ou les nouvelles hypothèses qui pourraient découler de l'examen des statistiques descriptives.

Le deuxième chapitre, intitulé "Concepts de l'analyse exploratoire," est consacré essentiellement à la présentation de diverses techniques de représentation visuelle des données (diagramme en feuilles, diagramme en bâtonnets, histogramme, etc.). Les auteurs reprochent à la plupart des volumes de statistiques de ne pas accorder assez de place aux techniques de représentation

visuelle des données. On peut leur reprocher d'en abuser. A notre avis, certains types de diagramme (la méthode des carrés de Tukey, entre autres) nous semblent fastidieux et peu utiles, particulièrement lorsque le nombre de données est élevé. Il est certes important de bien examiner les données (la forme de la distribution, la tendance centrale, la dispersion, les valeurs extrêmes, etc.), mais cet examen peut être effectué plus simplement que nous le proposent les auteurs (examen de la distribution de fréquences, accompagné d'un histogramme et d'indices statistiques résumant l'information tels, moyenne, médiane, variance, étendue, scores extrêmes). De plus, lorsque le nombre de scores est élevé, il est sans doute préférable d'utiliser un progiciel ou un logiciel qui peut nous fournir rapidement tous ces éléments que nous pourrions examiner attentivement par la suite.

La chapitre trois, intitulé "Méthodes de l'analyse exploratoire," est axé sur les méthodes de transformation des données. Même si le contenu de ce chapitre présente un certain intérêt, nous pouvons questionner la pertinence d'un tel contenu dans ce manuel d'introduction, d'autant plus que les auteurs ne justifient pas vraiment l'emploi des procédures présentées.

Les chapitres 4, 5 et 6 sont à notre avis les chapitres les mieux réussis de ce volume. Le quatrième chapitre nous introduit de façon intelligente aux principaux concepts de la statistique inférentielle. On insiste sur les assises de la statistique inférentielle de même que sur la logique ou démarche générale commune à tous les tests statistiques.

Le cinquième chapitre traite de façon assez exhaustive d'une famille de techniques statistiques inférentielles très utilisées, soit l'analyse de variance. On y expose clairement les objectifs, les méthodes et les conditions d'application de l'ensemble des techniques associées à l'analyse de variance. Ce chapitre constitue une excellente introduction, et même plus, à l'analyse de la variance.

La sixième chapitre porte sur la corrélation et la régression; il s'agit là également d'une famille de techniques statistiques utilisées dans presque toutes les recherches et dans toutes les disciplines de recherche. Ici encore, on prend soin de bien exposer les objectifs, les méthodes, les conditions d'application, de même que l'interprétation des résultats découlant de ces techniques statistiques.

Enfin, le dernier chapitre du volume présente une étude de cas (une analyse de données) qui se veut représentative de l'approche générale suggérée par les auteurs en analyse de données (bien examiner les données à l'aide de graphes, porter une attention particulière aux données dites "aberrantes" (outliers) et procéder aux analyses statistiques appropriées). Malheureusement, l'objectif n'est pas atteint. En effet, l'auteur expose un problème de recherche et se propose d'analyser les données recueillies; après avoir présenté quelques statistiques descriptives (graphes, moyennes, écarts-types, corrélations), il oriente l'analyse uniquement vers l'examen des données dites aberrantes, sans apporter de réponses à la question de recherche à l'étude. En fait, on peut sérieusement questionner l'utilité de ce dernier chapitre inachevé.

De façon globale, Pratique de l'analyse statistique des données constitue une excellente initiation à l'analyse des données. Nous avons particulièrement apprécié les chapitres 4, 5, et 6 de ce manuel. Par contre, les trois premiers chapitres, portant sur l'analyse descriptive des données, insistent un peu trop sur certaines techniques de représentation visuelle des données. Ces techniques, nous en convenons avec les auteurs, sont certes très utiles et sous-utilisées dans

l'analyse des données. Toutefois, centrer l'analyse statistique principalement sur ces techniques (et l'étude des données extrêmes) nous paraît un peu limiter le champ de la statistique descriptive, d'autant plus que les techniques présentées deviennent vite fastidieuses lorsque nous devons analyser un grand ensemble de données.

Enfin, et c'est là la plus grande qualité de ce volume, les auteurs abordent le domaine de l'analyse statistique comme une activité rigoureuse, basée sur la compréhension de méthodes et techniques bien sûr, mais également guidée par les intuitions et le bon sens de celui ou celle qui analyse.

Références

- Cohen, L. et Holliday, M. (1982). *Statistics for the Social Scientists*. London: Harper & Row.
- Howell, D. C. (1982). *Statistical Methods for Psychology*. Boston: Duxbury Press.
- Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading, Mass.: Addison-Wesley.