



Cet ouvrage présente différentes techniques de collecte et de représentation des données : la statistique descriptive élémentaire, la statistique confirmatoire, les approches multivariées et la modélisation statistique. Il expose la possibilité de donner plus de robustesse aux méthodologies classiques des sciences de l'éducation en y adjoignant une démarche quantitative.

Les fondements de chacune des approches sont analysés le plus clairement possible, tout comme les raisons et la façon de les utiliser au travers d'exemples simples ou plus complexes. Ainsi, *Statistiques et traitement de données quantitatives en éducation* peut être utilisé comme un cours, pour les non-initiés, mais il peut aussi constituer un accompagnement pour des chercheurs déjà familiers avec ces concepts.

Dédié aux étudiants, doctorants, chercheurs ou professionnels concernés par les approches quantitatives dans le domaine de l'éducation, cet ouvrage expose des méthodes couramment utilisées dans les différentes sciences humaines et sociales et développe des approches spécifiques pour l'éducation.

**Introduction :** Le présent ouvrage jette les bases des principales approches qui permettent une analyse simple, puis plus élaborée, des données quantitatives obtenues à l'occasion d'une étude ou d'un travail de recherche. Il s'adresse, principalement, aux étudiants, enseignants et chercheurs qui travaillent dans le secteur de l'éducation mais il peut aussi être utile dans les différents domaines des sciences humaines et sociales.

L'ouvrage peut être utilisé comme un cours, il commence par la présentation des différentes méthodes utilisées pour récolter des données et, progressivement, il aborde l'essentiel des approches d'analyses quantitatives utilisées dans le domaine des recherches en éducation pour faire parler ces données.

A la base, le document est conçu pour des lecteurs non-initiés à ce type de démarches. Toutefois, il peut, également être très utile à des doctorants et même des chercheurs, déjà initiés, qui en ont uniquement une approche logicielle, et qui aspirent à mieux comprendre les fondements de ces méthodes afin de mieux les utiliser, d'aller plus loin et d'éviter certaines erreurs.

Contrairement à beaucoup d'ouvrages quelquefois difficiles à lire ou qui abordent une seule méthode, nous avons fait le choix de présenter un panel des principales approches qui peuvent être utilisées dans le domaine de l'éducation de façon simple. Pour autant, pour aller plus loin, il sera conseillé aux lecteurs motivés de consulter des ouvrages spécialisés.

Cet ouvrage n'est pas un ouvrage de mathématiques qui présente toutes les bases théoriques, souvent complexes, des méthodes utilisées, mais nous ne voulons pas, non plus, nous limiter à une présentation de recettes et de procédures d'utilisation. Nous avons recherché, à chaque étape, à avoir une présentation équilibrée avec le double objectif d'être abordable et de permettre l'exploitation des données en toute conscience. Lorsqu'il nous arrive, malgré tout, d'ouvrir des parenthèses mathématiques, il est alors indiqué que la lecture n'est pas indispensable.

Aujourd'hui, les étudiants et chercheurs utilisent des logiciels dans lesquels il suffit d'entrer les données puis comme on dit, d'appuyer sur le bouton. Cette démarche n'est pas sans danger si au préalable on ne connaît pas suffisamment les fondements des méthodes utilisées. Bien évidemment tout au long de nos présentations nous utilisons des logiciels mais nous avons fait le choix de ne pas leur consacrer une partie. En effet, leurs modes d'utilisation diffèrent, les uns des autres, et ils évoluent très vite au cours du temps. Quand vous lirez ce texte les logiciels que nous avons utilisés n'existeront peut-être déjà plus ou auront évolué et, certainement, il en existera d'autres plus performants et plus conviviaux. Dans tous les cas, avant de traiter des données, vous aurez à vous investir dans l'apprentissage de l'utilisation d'un logiciel et cet investissement préalable est indispensable. Avant d'utiliser une automobile il faut apprendre à la conduire...Toutefois nous présentons à plusieurs reprises les calculs à la main car ils permettent de suivre le cheminement théorique, étape par étape, en partant des données pour aboutir aux résultats souhaités et cette démarche est très formatrice.

Sans entrer dans les détails, on dira qu'il est indispensable de mettre en œuvre des analyses des données quantitatives dès lors que l'on dispose de données prélevées sur un nombre significatif d'individus pouvant aller de quelques dizaines à plusieurs milliers. Le chercheur ou l'étudiant récolte ses données, elles sont regroupées dans un tableau qui croise les individus prélevés avec les différents paramètres (variables) recueillis : le tableau Individus/Variables, [I/V]. C'est le point de départ du traitement des données car la plupart du temps il n'est pas directement interprétable et pourtant il va falloir en tirer un maximum d'informations. Pour cela le chercheur va procéder par étapes.

La première est la statistique descriptive élémentaire. Elle consiste à construire d'autres tableaux, plus explicites, extraits du tableau [I/V], puis à en faire des représentations graphiques y compris cartographique. De plus, pour le cas des variables numériques il est possible d'avoir une description plus précise en introduisant des indicateurs mathématiques : moyenne, variance, écarts types pour chacune des variables, et covariance, coefficient de corrélations pour les variables prises deux par deux. Après avoir mis en œuvre les outils de la statistique descriptive, le chercheur est en mesure de commencer à présenter les données, de les commenter, et de les confronter à ses hypothèses de départ...

La deuxième étape est la statistique confirmatoire, on dit aussi l'inférence statistique. A ce stade du travail, le chercheur est capable de présenter ses données de façon lisible, il a pu les commenter et en tirer des conclusions. Mais pour des raisons pratiques évidentes les données ont été prélevées sur des échantillons et non sur la population entière et rien ne dit que si l'on s'était adressé à d'autres

échantillons issus de la même population les conclusions auraient été les mêmes. Il doit donc se poser la question de savoir si les résultats obtenus, sur le ou les échantillons, sont généralisables à l'ensemble de la population. C'est la statistique confirmatoire qui traite de cette question. Elle repose sur des notions fondamentales de probabilité et sur les lois du hasard et des grands nombres. Il en découle des lois qui prédisent la probabilité pour qu'un évènement se produise dans une population. Il est alors possible d'établir des tables de probabilités qui sont à la base des tests statistiques (moyennes, Student, khi2, anova, test de corrélation,...) que le chercheur doit utiliser pour établir le caractère généralisable de ses résultats, à toute la population.

La troisième étape s'intéresse aux analyses multivariées qui permettent une observation globale des liens pouvant exister entre plus de deux variables (3, 4,..., n). Leur utilisation devient de plus en plus fréquente. Elles complètent la statistique descriptive élémentaire en permettant de mettre en évidence des liens multiples non soupçonnés et, en ce sens, elles vont plus loin dans l'analyse des données. C'est l'ACP (Analyse en Composante Principale), qui s'applique à des variables numériques, qui est à la base des méthodes multivariée. L'AFC (Analyse Factorielle des Correspondances) et l'AFCM (Analyse Factorielle des Correspondances Multiples) s'appliquent quant à elles à des variables qualitatives et elles s'appuient sur les bases théoriques de L'ACP.

La quatrième étape qui peut être utilisée est la modélisation statistique. Après avoir mis en évidence l'existence des liens entre les différentes variables on peut chercher à établir les relations mathématiques qui pourrait exister entre une des variables dite variable expliquée et une ou plusieurs autres variables dites variables explicatives. Cela revient à établir un modèle prédictif qui modélise l'évolution de la variable expliquée lorsque les variables explicatives évoluent, d'où le terme de modélisation statistique. Par exemple Pascal Bressoux essaie d'établir un modèle qui relie l'opinion qu'un professeur a sur ses élèves (variable expliquée) en fonctions de plusieurs variables explicatives comme les performances scolaires, le retard scolaire, la catégorie socio-professionnelle des parents, etc... La modélisation statistique peut aborder des situations très variées suivant la nature des variables explicatives et de la variable expliquée (régression normale ou logistique), suivant le type d'action directe ou indirecte des variables explicatives, ou encore la prise en compte des effets de contexte avec l'existence de plusieurs niveaux d'interactions (élèves, classe, école, ville...).

Le plan de l'ouvrage suit les différentes étapes que doit accomplir un chercheur à l'occasion d'une étude. Pour cela il se compose de six parties dont chacune comporte de deux à cinq chapitres. La première partie est consacrée aux recueils des données en éducation. La deuxième partie traite de la statistique descriptive élémentaire. La troisième partie est consacrée à la statistique confirmatoire dite aussi inférence statistique. On aborde ensuite dans la quatrième partie les approches multivariées puis dans la cinquième partie, la modélisation statistique. Enfin, la sixième partie, intitulée « vers une robustesse des études en éducation par l'approche quantitative », est consacrée à des présentations d'outils utilisés habituellement en éducation (et d'autres disciplines) mais qui gagnent en robustesse à prendre un caractère un peu plus quantitatif que d'ordinaire qu'il convient de formaliser. Les deux exemples cités ici concernent les représentations sociales en éducation puis les études des rapports aux savoirs. L'idée première est de montrer que beaucoup de méthodes peuvent être améliorées, transformées pour prendre un caractère quantitatif permettant alors aux études plus de reproductibilité....

## **Introduction**

### **Partie 1 : Le recueil des données en éducation**

Chapitre 1 : L'utilisation des bases de données existantes en éducation

Chapitre 2 : Le questionnaire d'enquête

Chapitre 3 : Les démarches expérimentales

## **Partie 2 : Statistique descriptive élémentaire et représentation des données**

Chapitre 1 : Tableaux et représentations graphiques

Chapitre 2 : Les indicateurs mathématiques

Chapitre 3 : Les méthodes de représentations spatiales des données

## **Partie 3 : La statistique confirmatoire**

Chapitre 1 : Les lois du hasard et des grands nombres - Les tests d'hypothèses

Chapitre 2 : Tests sur les moyennes-Tests de Student

Chapitre 3 : Analyse de la variance-ANOVA

Chapitre 4 : Analyse bi variée : Test de corrélation de Bravais-Pearson Erreur ! Signet non défini.

Chapitre 5 : Tests confirmatoires pour variables qualitatives : khi2 et comparaison de fréquences

## **Partie 4 : Analyses multivariées**

Chapitre 1 : L'analyse en composantes principales (ACP)

Chapitre 2 : Les analyses factorielles de correspondances

## **Partie 5 : La modélisation statistique**

Chapitre 1 : Modélisation linéaire bi variée simple

Chapitre 2 : Régressions linéaires multiples pour des variables explicatives quantitatives

Chapitre 3 : Modélisation avec des variables explicatives qualitatives

Chapitre 4 : Prise en compte des interactions entre variables

Chapitre 5 : Modélisations complexes

## **Partie 6 : Vers une robustesse des études en éducation par l'approche quantitative**

Chapitre 1 : Approche quantitative des représentations sociales en éducation

Chapitre 2 : Exemple d'approche quantitative des rapports aux savoirs

## **Bibliographie**

Partie 1 :

Boulan H., (2015). Le questionnaire d'enquête : Les clés d'une étude marketing ou d'opinion réussie. Dunod, fonction de l'entreprise, 240p.

Martin O., (2012). L'analyse quantitative des données : L'enquête et ses méthodes. Armand Colin, Edition : 3eme Ed, 128p

- Gérard F., (2015). Conduite d'enquête par questionnaire. Edition du robot furieux- Frédéric Gérard Edition, 122p
- Weil-Barais A. (1997) Les méthodes en psychologie : Observation, expérimentation, enquête, travaux d'étude et de recherche. Bréal, Grand Amphi Economie, 284p.
- Van der Maren J.M., (2014). La recherche appliquée pour les professionnels : Education, (para)médical, travail social. De Boeck, Méthodes en sciences humaines. 304p
- Partie 2 :
- Beguin M. et Pumain D., (2010). La représentation des données géographiques : Statistique et cartographie, 3e Ed. A. Colin, Paris, 256p.
- Pumain D. et Saint-julien Th. (2001). Les interactions spatiales., Paris, Armand Colin, 2001. 191p.
- Alinat S. et Barthes A. (2014). Apports de la cartographie aux recherches sur l'école rurale, in Barthes et Champollion, *L'école rurale et montagnarde en contexte méditerranéen : Approches socio-spatiales*. Volume 6. Besançon. Presses universitaires de Franche-Comté. pp.36-48
- Dagnélie P., (2007). Statistique théorique et appliquée : 1- Statistique descriptive et bases de l'inférence statistique. 2<sup>e</sup> ED., De Boeck. 460p.
- Py P. (2007)., Statistique descriptive, Economica. 353p.
- Py P., (2007). Exercices corrigés de statistique descriptive, Economica, 218p.
- Albarello L., Bourgeois E. et Guyot JL. (2007). Statistique descriptive- Un outil pour les praticiens et les chercheurs, 2<sup>e</sup> Ed., De Boeck, 174p.
- Maalouf JP., (2016). Introduction aux statistiques descriptives. , [www.xlstat.com](http://www.xlstat.com).
- Baggio S., Deline S. et Rothen S., (2017). Statistique descriptive-L1/L2 psycho. De Boeck, 220p.
- Baccini A., (2010). Statistique descriptive élémentaire, Publications de l'institut de mathématiques de Toulouse. 40p.
- Partie 3 :
- Dagnélie P. (2008). Statistique théorique et appliquée : 1. Statistique descriptive et bases de l'inférence statistique, 2<sup>e</sup> Ed. De Boeck, 488p
- Dagnélie P. (2011). Statistique théorique et appliquée : 2. Inférence statistique à une et à deux dimensions, 3<sup>e</sup> Ed. De Boeck, 715p.
- Guéguen N. (1997). Manuel de statistique pour psychologues. Dunod, 294p.
- Abdi H. (1987). Introduction au traitement statistique des données expérimentales. Presses universitaires de Grenoble, 419p.
- Ramousse R, Le Berre M. et Le Guelte L. (1996). Introduction aux statistiques- 3. Choisir le test statistique approprié. [www.cons-dev.org/elearning/stat/St2b.html](http://www.cons-dev.org/elearning/stat/St2b.html)
- Howell D.C. (2010). Fundamental statistics for the behavioral sciences, 7th. Ed. Wadsworth publishing, 672p
- Milot G . (2009). Comprendre et réaliser les tests statistiques avec R. De Boeck, 704p.
- Prum B (2010). La démarche statistique. Cépadués Editions, 348p.
- Partie 4 :
- Benzécri J.P., (1973). L'analyse des données Tome 1 : La taxinomie. Dunod, 610p
- Benzécri J.P., (1973). L'analyse des données Tome 2 : L'analyse des correspondances. Dunod, 619p.
- Baccini A., (2010). Statistique descriptive multidimensionnelle : Pour les nul. Publications de l'institut de mathématiques de Toulouse, 33p.
- Coupaud M., Larini.M, Castéra J. et Delsérieys-Pedregosa A., (2015). Méthode d'analyse exploratoire pour une étude comparative sur les représentations de la démarche d'investigation d'enseignants de collèges. Review of science Mathematics and ICT Education, p (99-114).
- Cornillon P.A. et all, (2012). Statistiques avec R : 3<sup>e</sup> édition revue et augmentée. Presses universitaires de Rennes : PUR, 296p.
- Dumolard P., (2011). Données géographiques : Analyse statistique multivariée. Hermès science publications, 208p.
- Stafford J. et Bodson P., (2005). Analyse multivariée avec SPSS. Presse de l'université de Québec, collection Mathématiques E, 245p.

Busca D., Toutain S., (2009). Analyse factorielle simple en sociologie : méthodes d'interprétation et études de cas. De Boeck, 259p.

Cibois P., (1994). Analyse en composante principale et analyse des correspondances, 4ième Ed. Que sais-je, Presse universitaire de France, PUF, 127p.

Escofier B. et Pages J., (2016). Analyse factorielle simple et multiple : cours et étude de cas. Dunod, 400p.

Munoz F., Borgner F., Clément P. et Graça S. Carvalho., (2009). Teachers' conceptions of nature and environment in 16 countries. *Journal of Environmental Psychology*, 29 p (407-413).

Costa R., 2013. Analyser les données en sciences sociales : de la préparation des données à l'analyse multivariée. Peter Lang, 301p

Partie 5 :

Bressoux P., (2010). Modélisation statistique appliquée aux sciences sociales. De Boeck, méthodes et sciences humaines, 464p.

Bressoux P. et Bianco M., (2004). Long term effect on pupils learning gains. *Oxford review of education*, 30 (3), 3, pp 27-345.

Aitkin M. et Longford N.T., (1986). Statistical modeling in school effectiveness studies. *Journal of the royal statistical society, series A*, pp 1-43.

Hoffmann J.P., (2003). Generalised linear model : an applied approach. Pearson, 1 édition, 216p

Raudenbush S.W., (1989). The analysis of longitudinal multilevel data. *International Journal of Educational research*, (13) 17, pp 721-740.

Blalock H.M., (1984). Contextual effect model : Théoretical and methodological issues. *Annual Review of Sociology*, 10, pp 353-372.

Bertrand F. et Maumy-Bertrand M., (2014). Initiation à la statistique avec R : cours et exercices corrigés. Dunod, 2<sup>e</sup> Ed. 421p.

Partie 6 a :

Abric J. C. (1994), *Pratiques sociales et représentations*. Paris : PUF.

Abric J. C. (1997), *Méthodologie de recueil des représentations sociales*, in Abric J.C. (Dir.) - *Pratiques sociales et représentations*, 2<sup>ème</sup> Ed., Paris, PUF, 59 - 82.

Barthes A. (2012). Comment réintroduire les savoirs face à l'éducation au développement durable ? Exemple des filières professionnelles d'aménagement des territoires. *Éducation au développement durable et à la biodiversité : concepts, questions vives, outils et pratiques*. UQUAM. Montreal: 206-229.

Barthes A., Alpe Y. (2016). *Utiliser les représentations sociales en éducation*. Paris : L'Harmattan, Collection Logiques Sociales, 220 p.

Bonardi C. et Roussiau N. (1999), *Les représentations sociales*. Paris. Dunod.

Bourdieu P. Chamboredon J-C et Passeron J-P. (1968). « Le métier du sociologue », Paris, Mouton, Ed. EHESS

Degenne A. et Verges P. (1973) Introduction à l'analyse de similitude, *Revue française de sociologie*, XIV, 471-512

Durkheim E. (1894). *Les règles de la méthode sociologique*. Paris : PUF

Flament C. (1987). *Pratiques et représentations sociales*. In Beauvois, J.-L. Joule, R.-V. et Monteil, J.-M. *Perspectives cognitives et conduites sociales. I. Théories implicites et conflits cognitifs*. Paris : PUF.

Flament C. (1994). Structure, dynamique et transformation des représentations sociales. In Abric, J. C. *Pratiques sociales et représentations*. Paris : PUF.

Flament C. (1995). Approche expérimentale de type psychologique dans l'étude d'une représentation sociale, *Les cahiers internationaux de psychologie sociale*, n°28, 4/1995, 67-76.

Jodelet D. (1994). *Les représentations sociales*, Paris, PUF, 1994

Jodelet D. (1989). Représentations sociales : un domaine en expansion. In JODELET, D., *Les représentations sociales*. Paris : PUF

Moliner P. (1996). *Images et représentations sociales. De la théorie des représentations à l'étude des images sociales*. Grenoble : PUG

- Verges P. (2001). L'analyse des représentations sociales par questionnaires. *Revue française de sociologie*, 42(3), 537-561.
- Verges P. Flament, C., (1997). Comparaisons de méthode : A propos de l'article « Apport des modèles graphique gaussiens en analyse de similitude », *Papers on social representations*, 1997, vol 6 - n°1, 73-87.
- Verges P. (1994). Approche du noyau central : propriétés quantitatives et structurales, *Structures et transformations des représentations sociales*, Guimelli, Ch. (ed), Delachaux et Niestlé, Lausanne, textes de base en sciences sociales, 1994, pp. 233-253.
- Partie 6 b :
- Bautier E., Rochex J-Y. (1998). L'expérience scolaire des nouveaux lycéens. Démocratisation ou massification ? Paris : Armand Colin.
- Beillerot J. (2000). Le rapport au savoir. Dans N. Mosconi, J. Beillerot & C. Blanchard-Laville (dir.), *Formes et formations du rapport au savoir* (pp. 39-57). Paris : L'Harmattan.
- Bourdieu P. (2000). L'inconscient d'école. *Actes de la recherche en sciences sociales*. 135(12), 3-5.
- Caillot M. (2001). Y a-t-il des élèves en didactique des sciences ? Ou quelles références pour l'élève ? Dans A. Terrisse (dir.), *Didactique des disciplines. Les références au savoir* (pp. 141-155). Bruxelles : De Boeck Université.
- Charlot B. (1997). *Du rapport au savoir. Éléments pour une théorie*. Paris : Anthropos.
- Charlot B. (1999). *Le rapport au savoir en milieu populaire. Une recherche dans les lycées professionnels de banlieue*. Paris : Anthropos.
- Charlot B., Bautier E. & Rochex J-Y. (1992). *École et savoir dans les banlieues et ailleurs*. Paris : Armand Colin.
- Chartrain J.-L. (2003). *Rôle du rapport au savoir dans l'évolution différenciée des conceptions scientifiques des élèves. Un exemple du volcanisme au cours moyen 2*. Thèse de doctorat non publiée, Université Paris 5, Paris.
- Chevallard Y. (1992). Concepts fondamentaux de la didactique : perspectives apportées par une approche anthropologique. *Recherches en didactique des mathématiques*, 12(1), 73-112.
- Kalali F. (2017). Rapports aux savoirs in Barthes, Lange, et Tutiaux-Guillon N. (dir.). *Dictionnaire critique des enjeux et concepts des éducations à*, L'Harmattan, *Paris*, 617 p.
- Venturini P. (2007). *L'envie d'apprendre les sciences. Motivation, attitudes, rapport aux savoirs scientifiques*. Paris : Édition Fabert.
- Vygotsky L. (1978) *Mind in Society : The Development of Higher Psychological Processes*.