

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.

À 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. Lorsque 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 cartographiques. 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. À 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, χ^2 , anova, test de corrélation, etc.) 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 multivariées. 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 pourraient exister entre une des variables dites 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 fonction 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, etc.).

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 chapitres dont chacun comporte de deux à cinq sections. Le premier chapitre est consacré aux recueils des données en éducation. Le deuxième chapitre traite de la statistique descriptive élémentaire. Le troisième chapitre est consacré à la statistique confirmatoire dite aussi inférence

statistique. On aborde ensuite dans le quatrième chapitre les approches multivariées puis dans le cinquième chapitre, la modélisation statistique. Enfin, le sixième chapitre, intitulé « Vers une robustesse des études en éducation par l'approche quantitative », est consacré 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...