



150 Photos et Schémas

Sites web

Cours Exercices Corrigés

STATISTIQUES POUR PSYCHOLOGUES

ANALYSES DESCRIPTIVES



Édition : Marie-Laure Davezac-Duhem
Fabrication : Anne Pachiaudi
Composition et mise en pages : Nord Compo
Conception couverture : Pierre-André Gualino
Relecture et correction : Isabelle Chave

Nos équipes ont vérifié le contenu des sites internet mentionnés dans cet ouvrage au moment de sa réalisation et ne pourront pas être tenues pour responsables des changements de contenu intervenant après la parution du livre.

© Dunod, Paris, 2010, 2015, 2020
11 rue Paul Bert, 92240 Malakoff
www.dunod.com
ISBN 978-2-10-082263-8

Le Code de la propriété intellectuelle n'autorisant, aux termes de l'article L. 122-5, 2° et 3° a), d'une part, que les « copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective » et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, « toute représentation ou reproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite » (art. L. 122-4).

Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles L. 335-2 et suivants du Code de la propriété intellectuelle.



150 Photos et Schémas

Sites web

Cours Exercices Corrigés

STATISTIQUES POUR PSYCHOLOGUES

ANALYSES DESCRIPTIVES



Jean-Marc Meunier

3^e édition



DUNOD

聖

觀自

菩薩

波羅

弟子

菩提

遙

僧

佛光山修持中

TABLE DES MATIÈRES

Préface

Introduction

1 – LES CONCEPTS STATISTIQUES

I. Individu, échantillon, population	6
II. Notions de variable et de facteur	6
III. La notion de mesure	10
IV. Notion de protocole	11
1. L'emboîtement	12
2. Le croisement	13
V. Les procédures	13
1. Les types d'objets	15
2. Les objectifs des méthodes statistiques	16

2 – LES PROCÉDURES APPLICABLES AUX VARIABLES

I. Pourquoi recoder une variable?	24
II. Le recodage par regroupement de modalités	26
1. Recoder une variable nominale	26
2. Recoder une variable ordinale	29
3. Recoder une variable numérique	31
III. Le recodage par transformation	35
1. Les transformations linéaires	36
2. La transformation Z	37
3. Les transformations non linéaires	38
4. L'inverse	40

3 – LES PROCÉDURES APPLICABLES AUX PROTOCOLES

I. Dériver un protocole d'analyse	46
II. Résumer numériquement un protocole	48
1. Indices de position.....	49
2. Indices de dispersion.....	53
III. Comparer des groupes d'observations	59
IV. Faire une distribution	61
1. Faire une distribution sur un protocole univarié.....	61
2. Faire une distribution sur un protocole bivarié.....	63
3. Un nouvel objet d'analyse.....	65

4 – LES PROCÉDURES APPLICABLES AUX DISTRIBUTIONS

I. Résumer une distribution	74
1. Résumer graphiquement une distribution.....	74
2. Cumuler une distribution.....	79
3. Résumer numériquement une distribution.....	81
4. Le choix des résumés numériques.....	89
II. Comparer les distributions de groupes d'observations	90
1. Cas d'une variable nominale.....	90
2. Cas d'une variable ordinale.....	92
3. Cas d'une variable numérique.....	95
III. Situer un individu dans une distribution	97
1. Situer un individu dans un décilage.....	98
2. Situer un individu avec un écart réduit.....	100
3. Situer dans une échelle normalisée.....	101
IV. Comparer une distribution à une distribution de référence	105
1. Comparaison à une distribution plate.....	106
2. Comparaison à une distribution normale.....	107

5 – L'ANALYSE DES PROTOCOLES BIVARIÉS

I. La notion de liaison	120
II. Étudier la liaison entre variables nominales	121
III. Étudier la liaison entre variables ordinales	128
IV. Étudier la liaison entre variables numériques	131

6 – L'UTILISATION DES LOGICIELS DE STATISTIQUES

I. Principes de fonctionnement d'un tableur	145
1. La feuille de calcul	145
2. Les formules	146
3. L'utilisation de références	146
4. La saisie d'un protocole	147
5. La structuration des données	148
II. Recoder une variable	149
1. Le regroupement de modalités	149
2. Le regroupement en classes	151
III. L'analyse des protocoles	152
1. Dériver un protocole d'analyse	152
2. Faire une distribution	155
3. Calculer les quartiles	155
4. Calculer la moyenne et l'écart-type	157
IV. L'analyse de distribution	158
1. Faire des graphiques	158
2. Cumuler une distribution	160
3. Résumer une distribution	161
4. Normaliser une distribution	163
V. L'analyse des protocoles bivariés	164
1. Réaliser une distribution bivariable	164
2. Calculer un Φ^2	166
3. Calculer un ρ de Spearman	168
4. Calculer un r de Bravais-Pearson	169

Annexes

Corrigés

Crédits photographiques



Préface

Les statistiques ont envahi notre univers quotidien, et comme elles sont très largement répandues, on peut avoir l'impression qu'elles reposent sur des notions qui vont de soi. C'est pourtant loin d'être le cas. Les statistiques sont une branche des mathématiques qui est apparue assez récemment et dont le développement a été grandement favorisé par l'informatique, qui fournit à la fois des facilités de recueil et des capacités d'enregistrement et de traitement de grandes masses de données. Mais la culture statistique suit-elle le développement technique ?

Une première difficulté est que les observations recueillies et analysées ne portent pas d'ordinaire sur l'ensemble des individus et plus généralement des objets concernés par l'étude, mais seulement sur une partie d'entre eux, ce que l'on appelle un échantillon. Il apparaît paradoxal que l'on puisse, à partir d'un échantillon d'un millier de personnes, évaluer l'opinion des millions de personnes que compte un pays, sur une question ou leur jugement sur un personnage politique. La justification est difficile à comprendre, car elle repose sur le calcul des probabilités et sur un certain nombre de conditions que doivent respecter les techniques de sondage et qui permettent d'appliquer les méthodes de calcul statistique à ce domaine.

Une seconde difficulté est que les techniques statistiques sont centrées sur l'analyse de la variabilité et que la variabilité est mal appréhendée par notre jugement naturel, qui retient ce qui est vu ou senti dans le présent et ce que nous livre notre mémoire, qui est par ailleurs très sélective. De la sorte, nous ne tirons guère d'informations de la variabilité et nous ne pensons guère qu'on peut le faire. La statistique oblige donc à construire de nouveaux concepts qui peuvent paraître abstraits, dans la mesure où ils ne sont pas dérivés de l'expérience quotidienne. Tel est le cas de la notion de distribution qui est à la base du raisonnement statistique et qui exprime la répartition des observations sur une échelle qui permet de les situer. Cette notion est complexe, car il n'en existe guère d'équivalent dans la vie quotidienne ; il est pourtant nécessaire que l'étudiant en acquière une bonne compréhension.

Une difficulté complémentaire vient de ce que les résultats de l'analyse statistique peuvent heurter les conceptions quotidiennes. Par exemple, si on fait une statistique sur le nombre d'enfants d'une famille, on peut trouver que la moyenne pour un pays est 2,2 ou 1,6. Cette valeur n'a pas de sens quand on parle d'une famille : comment se représenter la signification de cette valeur décimale ? Elle n'a de sens que relativement à l'ensemble des familles d'un pays puisqu'elle se réfère à la distribution du nombre

d'enfants par famille dans cette population et en cela n'a certainement rien d'intuitif. Faute d'avoir assimilé un certain nombre de concepts de base, des représentations erronées de la statistique peuvent se développer dans le public, comme l'idée répandue que l'on peut faire dire n'importe quoi aux statistiques. Les techniques statistiques permettent certes différentes analyses, mais la pertinence de ces analyses dépend du type de données auxquelles on a affaire. Le choix n'est pas aveugle et l'on peut en confrontant les conditions de validité de chaque analyse et les propriétés de la situation étudiée déterminer quel type d'analyse est optimale.

L'ouvrage de Jean-Marc Meunier se propose de donner en même temps que les techniques d'analyse une base conceptuelle qui permet des choix judicieux et une interprétation appropriée. Il est exigeant, mais il est à la portée d'un étudiant en sciences humaines, notamment d'un étudiant en psychologie, dans la mesure où l'introduction des notions est faite de façon très progressive et à partir de beaucoup d'exemples.

Une étape importante est le passage de l'observation comportementale à la mise en forme de ce qui a été observé en vue du traitement statistique. Cette étape est expliquée de façon très détaillée dans l'ouvrage à partir de l'analyse des différents types d'observations et l'étude des différentes façons de les coder en vue des analyses ultérieures.

Le concept de distribution, si important on l'a dit, est également présenté de façon très détaillée en lien avec les différents types de mesure dont on dispose. En effet, en psychologie, les mesures numériques, qui sont les plus faciles à traiter, sont en effet rares, sauf peut-être en psychologie expérimentale et en psychologie différentielle. On a affaire le plus souvent, que ce soit dans les échelles d'attitude ou d'opinion, ou dans les échelles de personnalité, élaborées à partir de questionnaires, qui sont utilisées en clinique, à des instruments de mesure dont les échelons sont simplement ordonnés et pour lesquels les notions de somme, et donc de moyenne, n'ont pas de sens. Il est d'autres cas où l'on ne dispose que de catégories, qui ne sont même pas ordonnées entre elles ou qui ne peuvent l'être que très partiellement, comme lorsqu'on fait juger des aliments selon un certain nombre de propriétés. Pourtant ces données sont traitables statistiquement ; on peut les résumer à l'aide de distributions, mais cela relève de la nature de l'échelle de mesure qui a servi au recueil de ces données. L'ouvrage de J.-M. Meunier met en relation de façon très précise les échelles de mesure, les procédures qui permettent de faire les distributions et les indices qui peuvent être calculés dans chaque cas pour pouvoir résumer au mieux la distribution. Les relations entre les échelles de mesure, les types de distribution et les types d'indices calculables sont présentées de façon suffisamment précise pour qu'un non-spécialiste des statistiques puisse faire des choix judicieux.

L'ouvrage est présenté de façon agréable avec une grande variété d'exemples et beaucoup d'illustrations qui retiennent l'attention et permettent d'éviter qu'un cours de statistiques soit rébarbatif. L'ouvrage sait aussi utiliser, quand c'est nécessaire, le formalisme des formules mathématiques, qui est indispensable pour décrire de façon non ambiguë les procédures de calcul, qui sont inévitablement complexes. Toutefois, la suite des calculs à effectuer est présentée de façon très progressive en liaison avec la formule, pour que l'étudiant puisse, à l'aide de sa calculatrice, reproduire la procédure et en comprendre la signification. Un étudiant doit savoir qu'un ouvrage de statistiques se lit calculatrice à la main, en reproduisant les procédures décrites et en faisant les exercices recommandés. Seul un apprentissage actif de traitement de données, comme en mathématiques la résolution de problèmes, permet une assimilation des procédures et des concepts sur lesquels elles sont basées.

La dernière partie introduit de façon approfondie à l'utilisation des tableurs. Les tableurs sont aujourd'hui l'outil indispensable à l'étudiant qui doit traiter des résultats de travaux pratiques, de travaux d'initiation à la recherche ou de travaux de recherche.

C'est un prolongement très heureux du cours, car la présentation des traitements sur des tableaux de données, qui est faite systématiquement dans le cours, se trouve exactement celle qui est adaptée pour traiter ces données à l'aide d'un tableur. Il devient donc possible de traiter des données conséquentes sans effort de calcul à la main. Les tableurs sont largement suffisants pour le calcul des statistiques descriptives qui font l'objet de cet ouvrage. Leur avantage est qu'ils permettent de suivre le déroulement des calculs et de découvrir éventuellement des incongruités, si la procédure a été mal choisie. L'utilisation de logiciels sophistiqués peut être dangereuse pour un étudiant peu averti, car elle exige une grande maîtrise conceptuelle du domaine pour une utilisation réfléchie et judicieuse. En période d'apprentissage, le tableur est plus transparent et c'est un des mérites de cet ouvrage de permettre de maîtriser cet outil et ainsi d'acquérir progressivement une autonome technique et conceptuelle dans la pratique statistique.

Jean-François Richard