

**Olivier Martin**

# **L'ANALYSE QUANTITATIVE DES DONNÉES**

**5<sup>e</sup> édition**

**ARMAND COLIN**

## Mise en pages : Nord Compo

Le pictogramme qui figure ci-contre mérite une explication. Son objet est d'alerter le lecteur sur la menace que représente pour l'avenir de l'écrit, particulièrement dans le domaine de l'édition technique et universitaire, le développement massif du photocopillage.

Le Code de la propriété intellectuelle du 1<sup>er</sup> juillet 1992 interdit en effet expressément la photocopie à usage collectif sans autorisation des ayants droit. Or, cette pratique s'est généralisée dans les établissements

d'enseignement supérieur, provoquant une baisse brutale des achats de livres et de revues, au point que la possibilité même pour

les auteurs de créer des œuvres nouvelles et de les faire éditer correctement est aujourd'hui menacée.

Nous rappelons donc que toute reproduction, partielle ou totale, de la présente publication est interdite sans autorisation de l'auteur, de son éditeur ou du Centre français d'exploitation du droit de copie (CFC, 20, rue des Grands-Augustins, 75006 Paris).



© Armand Colin, 2020 pour cette 5<sup>e</sup> édition

Armand Colin est une marque de  
Dunod Éditeur, 11 rue Paul Bert, 92240 Malakoff

ISBN : 978-2-200-62694-5

Le Code de la propriété intellectuelle n'autorisant, aux termes de l'article L. 122-5, 2<sup>o</sup> et 3<sup>o</sup> 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.

# Sommaire

<b>Introduction</b> .....	7
1. Pourquoi recourir aux outils statistiques ? .....	8
2. Le plan du manuel .....	8
<b>I. Produire des données « quantitatives »</b> .....	10
1. Les « données » du sociologue .....	11
2. Population, échantillon et individus .....	11
3. Les sources quantitatives en sociologie .....	12
3.1 Une source particulière mais courante : le questionnaire .....	13
3.2 Les autres sources quantitatives .....	13
3.3 L'analyse secondaire des données .....	15
4. La notion d'échantillon .....	16
4.1 Étudier toute une population ou un échantillon ? .....	17
4.2 Les échantillons au hasard .....	18
4.3 Les échantillons empiriques .....	22
4.4 Qu'est-ce qu'un échantillon représentatif ? .....	24
4.5 Remarques et conseils supplémentaires .....	26
4.6 L'échantillon comme prisme .....	28
4.7 À propos des « erreurs » et des « biais » .....	29
5. Que valent les informations issues d'un échantillon ? .....	31
5.1 Passer de l'échantillon à la population .....	32
5.2 Quelle est la « valeur » de l'estimation ? .....	32
5.3 L'intervalle et le niveau de confiance .....	38
5.4 L'usage pratique .....	39

5.5 Les outils de calcul de l'intervalle de confiance .....	42
5.6 Quelques généralités supplémentaires sur les tests statistiques.....	43
5.7 Usages et limites des tests statistiques en sociologie.....	45
<b>2. Concevoir et préparer les variables nécessaires à l'analyse.....</b>	<b>47</b>
1. Questions, variables et modalités.....	47
2. Variables qualitatives et variables quantitatives .....	47
2.1 Variables quantitatives.....	48
2.2 Variables qualitatives.....	49
3. De la nécessité de recoder les variables.....	50
3.1 Techniques de recodage 1 : regrouper des modalités.....	50
3.2 Techniques de recodage 2 : simplifier les variables quantitatives.....	51
3.3 Techniques de recodage 3 : simplifier les variables multiples .....	52
3.4 Techniques de recodage 4 : coder les matériaux qualitatifs .....	53
3.5 Techniques de recodage 5 : combinaison des variables.....	54
4. Passer des variables aux indicateurs théoriques : les variables synthétiques.....	55
4.1 Créer des variables synthétiques par combinaison.....	55
4.2 Créer des variables synthétiques par calcul de scores.....	57
4.3 Créer des variables synthétiques à partir de variables quantitatives .....	60
4.4 Créer des variables synthétiques par analyse factorielle .....	61

<b>3. Un intermède : saisir la diversité des situations</b> .....	63
<b>4. Analyser les relations entre deux variables</b> .....	66
1. Juger la différence entre deux pourcentages.....	67
1.1 L'intervalle de confiance d'une différence.....	68
1.2 Les formules de calcul et leur usage.....	69
2. Le test du khi-deux ( $\chi^2$ ).....	72
2.1 La notion d'indépendance.....	72
2.2 Évaluer l'effet du hasard.....	74
2.3 Une distance pour juger de la proximité entre tableaux.....	75
2.4 Le calcul de la distance et de l'effet du hasard.....	79
2.5 Une illustration pratique.....	83
2.6 Intérêts et limites.....	84
3. Le coefficient de corrélation linéaire.....	85
3.1 La notion de relation linéaire.....	86
3.2 La notion de covariation.....	87
3.3 La définition du coefficient de corrélation linéaire.....	88
3.4 Une illustration pratique.....	89
3.5 Intérêts, usages et limites.....	90
4. Une variable qualitative et une variable quantitative.....	91
4.1 Le principe de l'ANOVA.....	92
4.2 Mise en œuvre technique de l'ANOVA.....	93
4.3 Une illustration pratique.....	95
4.4 Compléments.....	96
<b>5. Analyser les relations entre plusieurs variables</b> .....	98
1. Automatiser le croisement de variables.....	99
2. Repérer et synthétiser les relations : l'analyse factorielle.....	103
2.1 Principes et intérêts généraux de l'analyse factorielle.....	103

2.2 La pratique de l'analyse factorielle .....	106
2.3 Rôles des variables.....	108
2.4 Usages .....	109
3. Classer les individus pour définir des types.....	110
3.1 Principes généraux.....	111
3.2 Usages .....	112
4. Décomposer les effets de chaque variable .....	113
4.1 L'effet brut, l'effet pur et le raisonnement « toutes choses étant égales par ailleurs » .....	113
4.2 Expliciter un modèle .....	115
4.3 Évaluer la pertinence d'un modèle.....	115
4.4 Remarques finales.....	118
<b>Quelques conseils pour conclure.....</b>	<b>120</b>
1. Comment utiliser intelligemment les outils statistiques ? .....	120
1.1 Ne pas oublier la réflexion sociologique.....	120
1.2 Ne pas croire en une recette unique.....	121
1.3 Accepter de tâtonner.....	121
1.4 Distinguer l'exploration et la « démonstration ».....	122
1.5 Ne pas chercher la preuve absolue : préférer l'accumulation d'indices .....	122
2. Comment rédiger un rapport ou un article quantitatif ? .....	123
<b>Pour en savoir plus.....</b>	<b>125</b>

# Introduction

Ce manuel présente les démarches et méthodes pour l'analyse quantitative des données en sociologie. Il répond notamment aux questions suivantes : qu'est-ce qu'une donnée ? Qu'est-ce qu'un échantillon ? Quelle confiance accorder à un échantillon ? Qu'est-ce qu'un échantillon représentatif ? Comment coder ou recoder des réponses à un questionnaire ? Comment analyser ces réponses ? Comment mesurer la relation entre deux variables ? À quoi servent les tests statistiques ? Existe-t-il des méthodes pour traiter des milliers d'informations en même temps ?

Ce manuel n'est pas un cours de statistique : nous avons pris le parti de ne présenter que les aspects utiles en sociologie. Ce parti pris a deux conséquences. Premièrement, nous privilégions les outils statistiques habituellement utilisés par les sociologues, au détriment des techniques enseignées dans les cours et manuels de statistique, mais peu ou pas utilisées en sociologie. Deuxièmement, nous favorisons l'exposé des principes, des logiques, des intérêts et des limites des outils statistiques, au détriment des aspects mathématiques et calculatoires. Par exemple, il est préférable de bien assimiler la notion d'intervalle de confiance, les questions qu'elle permet de traiter, son intérêt et ses limites, plutôt que de savoir justifier mathématiquement la formule qui permet de calculer cet intervalle (c'est le travail du mathématicien), plutôt que de savoir calculer sans comprendre le sens du calcul (c'est le travail des logiciels statistiques).

Pour nous, les aspects mathématiques et calculatoires sont secondaires, non pas au sens où ils ne servent à rien, mais au sens où ils ne doivent venir qu'après la présentation et la justification, en langage ordinaire, des fondements et de la portée de chaque outil statistique. Lorsque ce manuel aborde leurs aspects mathématiques et calculatoires, c'est uniquement à la suite de l'exposé de leurs principes et de l'examen de leur portée.

Dès lors, deux usages du manuel sont possibles : soit comme outil de formation à la compréhension des fondements logiques des outils statistiques du sociologue (sans compétence mathématique particulière) ; soit comme outil de référence pour la mise en œuvre pratique des outils statistiques (comment les utiliser, interpréter et calculer ?).

La « seule » chose que le manuel ne fait vraiment pas c'est fournir les justifications mathématiques exactes des outils statistiques. Ces justifications sont

d'ailleurs impossibles sans de solides connaissances en théorie des probabilités et en statistique mathématique – nous renvoyons le lecteur intéressé aux ouvrages cités en bibliographie finale. En général, le sociologue n'a pas besoin de connaître ces aspects pour bien utiliser les méthodes d'analyse quantitative des données.

## 1. Pourquoi recourir aux outils statistiques ?

La sociologie peut être vue comme une science étudiant les principes de variation sociale des caractéristiques individuelles, des comportements, des attitudes, des pratiques ou des opinions. Lorsqu'elle recourt aux données standardisées, par exemple issues d'un questionnaire, la sociologie fait appel à la statistique puisque celle-ci fournit des outils destinés à analyser de grands ensembles de données. Face à de tels ensembles, la science statistique dispose d'outils théoriques et pratiques permettant d'identifier ces variations (par exemple dans les pratiques culturelles des Français), de comparer ces variations entre divers groupes (les hommes et les femmes ont-ils les mêmes pratiques ?), de saisir les liens pouvant unir ces variations (relation entre les pratiques de lecture et les pratiques d'écoute de la musique), d'identifier les groupes « typiques » ayant des pratiques plutôt homogènes c'est-à-dire présentant peu de variations (les adolescents ont-ils des pratiques de lecture comparables ?) ou encore d'expliquer les principes de variations (la diversité des pratiques de lecture s'explique-t-elle par les différences d'âge ou de milieu social ?).

Ces notions de variation, de liens, de co-relation, d'explication, de typologie, de confiance, de significativité ou encore de comparaison, sont importantes en sociologie (comme dans toutes les sciences empiriques d'ailleurs) et trouvent des expressions et des formalisations pratiques grâce à la science statistique.

Ainsi, à titre d'illustration, l'idée de relation ou de lien s'exprime dans les notions statistiques de corrélation, de tableau croisé, de comparaison de pourcentages ou de test du khi-deux ; l'idée d'explication trouve une expression pratique dans la notion de régression ; l'idée de variation est bien incarnée dans la notion de variance ou dans le tri à plat... En somme, la statistique offre à la sociologie des instruments permettant d'opérationnaliser, c'est-à-dire de mettre en pratique, des questions que cette dernière se pose sur des faits sociaux.

## 2. Le plan du manuel

Ce manuel est découpé en deux grandes parties. La première d'entre elles, divisée en deux chapitres, expose les différents aspects de la production et de la



préparation des variables. Le chapitre 1 évoque les différentes sources de données statistiques en sociologie. Il aborde la question centrale de l'échantillonnage des individus ou situations analysées ainsi que la question de la validité des résultats établis sur un échantillon : dans quelle mesure une donnée issue d'une enquête auprès d'un échantillon d'individu renseigne-t-elle sur un phénomène dépassant le cadre du seul échantillon pour concerner l'ensemble de la population ? Le chapitre 2 est consacré aux aspects relevant de la préparation des données, du recodage, de la création de variables pertinentes pour le sociologue. Les contraintes pesant sur les protocoles de recueil et de constitution des données empiriques conduisent en effet à des variables primaires souvent imparfaitement adaptées au besoin de l'analyse : diverses opérations de préparation des variables sont nécessaires. Insistons ici sur un point essentiel : concevoir de « bonnes » variables c'est élaborer de « bonnes » catégorisations des faits sociaux. Il est donc indispensable de bien coder ou recoder les questions et variables pour réaliser des analyses sociologiques pertinentes. Les opérations de codage et recodage sont des aspects cruciaux à ne jamais négliger !

La seconde partie du manuel présente les outils d'analyse statistique proprement dits. Elle est également organisée en deux chapitres. Le premier d'entre eux (chapitre 4) expose les outils permettant d'analyser les relations en deux variables, quelle que soit leur nature. Le second (chapitre 5) aborde les outils destinés à l'analyse des relations entre plusieurs variables (trois, voire beaucoup plus).

Entre les deux parties, un court « intermède » permet de présenter un point central pour les raisonnements et outils statistiques : la notion de variabilité et les principaux outils de sa mesure (chapitre 3).

Ce manuel prend la suite de l'ouvrage *Le Questionnaire*, de François de Singly<sup>1</sup> consacré à la présentation de la démarche quantitative en sociologie, à sa méthode emblématique (le questionnaire) ainsi qu'à son premier outil statistique (le tableau croisé). Avec le présent manuel, vous allez acquérir des compétences plus générales sur l'élaboration des données, sur l'analyse des réponses, sur l'échantillonnage, sur la construction de variables pertinentes pour le sociologue, sur les raisonnements et outils statistiques indispensables en sociologie.

---

1. Armand Colin, 2016 (4<sup>e</sup> édition).

# 1

## Produire des données « quantitatives »

Pour étudier le social, le sociologue doit élaborer des outils (méthodes d'enquête, concepts, catégories, données) lui permettant de s'abstraire des cas particuliers, de se détacher des représentations individuelles (à commencer par la sienne). Ce travail d'objectivation est essentiel. Nous n'y revenons pas – le lecteur intéressé pourra consulter les ouvrages classiques<sup>1</sup> ou des manuels<sup>2</sup>. Nous nous attachons ici à préciser les aspects pratiques d'élaboration des données « quantitatives » en sociologie : les origines possibles de ces données, la construction des échantillons et enfin l'estimation de la fiabilité des résultats.

Avant d'entrer dans le vif du sujet, notons que l'adjectif « quantitatives » accolé à « données » ne signifie pas que ces données sont de nature quantitatives (elles n'expriment pas nécessairement des quantités) mais qu'elles sont très nombreuses (plusieurs dizaines de données sur plusieurs centaines d'individus) et qu'il faut donc les analyser avec des méthodes capables d'analyser des grandes quantités (les méthodes statistiques). En d'autres termes, l'approche quantitative en sociologie ne signifie pas une quantification (une mesure) de chaque individu !

---

1. Émile Durkheim, *Les Règles de la méthode sociologique*, Paris, PUF, 1992 (première édition 1895) ; Pierre Bourdieu, Jean-Claude Chamboredon et Jean-Claude Passeron, *Le Métier de sociologue*, Paris, Mouton, 1968.

2. François de Singly, Christophe Giraud, Olivier Martin, *Apprendre la sociologie par l'exemple*, Armand Colin, 2016 ; voir aussi François de Singly, *op. cit.*, 2016, notamment le chapitre 1, § 2.