



LE SOCIOLOGUE

Introduction
aux méthodes statistiques
en sociologie

THIERRY BLÖSS
MICHEL GROSSETTI

puf

024512781

30

LE SOCIOLOGUE

COLLECTION DIRIGÉE PAR SANDRINE BÉGIN

Introduction
aux méthodes statistiques
en sociologie

Introduction
aux méthodes statistiques
en sociologie

THÉOPHY BLOIS
MICHEL GROSSETTI



THÉOPHY BLOIS

16.
A2

1999-15534

ACCES

COLLECTION DIRIGÉE PAR GEORGES BALANDIER

Les religions et les sociétés
antiques

10/18
1971-1972

LE SOCIOLOGUE

Introduction aux méthodes statistiques en sociologie

THIERRY BLÖSS
MICHEL GROSSETTI



PRESSES UNIVERSITAIRES DE FRANCE

DL 26 AVR.99 17963

DES MÊMES AUTEURS

Thierry BLÖSS

- Les liens de famille. Sociologie des rapports entre générations*, Paris, PUF, coll. « Le Sociologue », 1997.
- La femme dans la société française*, Paris, PUF, coll. « Que sais-je ? », 1996 (1^{re} éd., 1994) (avec Alain Frickey).
- Éducation familiale et beau-parenté. L'empreinte des trajectoires biographiques*, L'Harmattan, coll. « Logiques sociales », 1996.
- Les rapports sociaux de sexe : permanences et changements*, Paris, La Documentation française, coll. « Problèmes politiques et sociaux », 1994.

Michel GROSSETTI

- La construction des politiques publiques locales. Le cas des reconversions industrielles*, L'Harmattan, coll. « Logiques politiques », 1998 (avec Christophe Beslay, Denis Salles, François Taulèlle, Régis Guillaume et Michel Daynac).
- Science, industrie et territoire*, Toulouse, Presses Universitaires du Mirail, coll. « Socio-logiques », 1995.
- (sous la direction de), *Université et territoire. Un système scientifique local, Toulouse et Midi-Pyrénées*, Presses Universitaires du Mirail, série « Villes et Territoires », 1994.



ISBN 2 13 049880 9

Dépôt légal — 1^{re} édition : 1999, avril

© Presses Universitaires de France, 1999
108, boulevard Saint-Germain, 75006 Paris

SOMMAIRE

INTRODUCTION	1
--------------------	---

PREMIÈRE PARTIE

Les différentes étapes dans la production des données d'enquête

CHAPITRE I. — <i>La construction des variables : introduction à la statistique descriptive</i>	11
I. Les variables qualitatives	15
II. Les variables quantitatives	19
1. Le regroupement en classes d'amplitudes égales	19
2. L'échelle ordinale	22
III. Types de variables, indices statistiques et représentations graphiques	23
1. Indicateurs de tendance centrale	24
2. Indicateurs de dispersion	25
3. Représentations graphiques	27
CHAPITRE II. — <i>La réalisation d'un sondage : introduction à la statistique inférentielle</i>	33
I. Distributions	34
II. Lois de probabilité	36
III. Du modèle à l'inférence : les sondages	39

DEUXIÈME PARTIE

*L'exploitation des données d'enquête :
les tableaux statistiques*

CHAPITRE I. — <i>La lecture des tableaux statistiques</i>	53
I. Lectures d'un tableau univarié	53
1. Rappels notionnels	53
2. Le recensement de données chronologiques	57
<i>a</i> - Variations absolues et variations relatives	58
<i>b</i> - La prévision des évolutions : la « règle des 70 »	65
II. Lectures d'un tableau de contingence	67
1. Les différents niveaux d'agrégation d'une variable	69
2. Les deux sens interprétatifs d'un tableau croisé	70
3. Tableaux d'indépendance : effectifs et pourcentages théoriques	77
4. Tableaux des écarts à l'indépendance	80
CHAPITRE II. — <i>Tests et mesure de la corrélation entre variables</i>	89
I. Corrélation de variables nominales : le test du χ^2 et ses dérivés	92
1. Formulation de l'hypothèse nulle et calcul des effectifs théoriques	93
2. Récapitulatif et compléments : la procédure du χ^2	96
3. Limites et précautions d'usage du χ^2	101
<i>a</i> - La « tolérance » de Fisher	102
<i>b</i> - La correction de Yates	103
4. Les spécificités de raisonnement d'un tableau à 4 cases	105
<i>a</i> - La formulation d'une hypothèse de travail unilatérale	105
<i>b</i> - Formule pratique de calcul du χ^2	109

5. Mesure de la corrélation : Coefficient Φ et coefficient de contingence.	110
6. Autres utilisations du χ^2	113
II. Corrélation de variables numériques : le coefficient de corrélation linéaire de Bravais-Pearson	113
1. Calcul du coefficient r	116
2. Équation et tracé de la droite de régression.	121
3. Objectifs de la régression linéaire.	124
4. Propriétés et limites du r de Bravais-Pearson.	130
5. Le coefficient de détermination r^2	130
6. Étude d'un diagramme de corrélation	134
7. Significativité du coefficient de corrélation	138
III. Corrélation de variables ordinales : le test de Spearman	139
1. Le règlement des <i>ex aequo</i>	142
2. Classements identiques et classements inverses	144
IV. Corrélation de variables mixtes.	146
1. Tests de comparaison de moyennes.	147
2. Le rapport de corrélation	151
3. L'analyse de variance	156
V. Recherche exploratoire de corrélations	158
CHAPITRE III. — <i>Analyse multivariée et analyse multidimensionnelle</i>	161
I. Introduction à l'analyse multivariée	161
1. L'épreuve d'une variable test	162
2. « Maîtrise » expérimentale et contextualisation historique	168
II. Modéliser	170
1. Principes de la modélisation.	170
2. Variables quantitatives : régression multiple	173
3. Un exemple de régression log-linéaire	176

III. Résumer l'information	183
1. Les techniques de classification	183
2. Les techniques de résumé graphique	195

ANNEXES :

Tableau simplifié des méthodes	215
Tables statistiques	217

INDICATIONS BIBLIOGRAPHIQUES	221
------------------------------------	-----

INDEX	223
-------------	-----

Introduction

Le mot « statistique »¹ possède en français deux sens distincts, l'un désignant les comptages et recensements (les statistiques du chômage par ex.), et l'autre référant aux méthodes d'analyse de situations comportant des aspects aléatoires. C'est dans le second sens que nous utilisons le terme dans cet ouvrage, dont l'objectif est de présenter les bases nécessaires pour utiliser ce type de méthodes.

On considère souvent que l'usage des statistiques en sociologie débute avec Émile Durkheim, qui fait dans *Les règles de la méthode sociologique* l'apologie de la méthode des « variations concomitantes » (que nous appellerions aujourd'hui analyse des corrélations) et fonde son étude du suicide sur l'analyse des variations de taux de suicide pour diverses populations. En fait, comme le montre Alain Desrosières², Émile Durkheim ne fait que prolonger et théoriser un mouvement existant. L'analyse statistique est en effet utilisée par l'astronome belge Adolphe Quetelet dans les années 1830 pour définir l'« homme moyen ». Quelques décennies plus tard, l'économiste et sociologue Vilfredo Pareto formalise la

1. Le terme « statistique » vient de l'allemand *Statistik*, lui-même formé sur une racine latine renvoyant à l'État.

2. A. Desrosières, *La politique des grands nombres. Histoire de la raison statistique*, La Découverte, coll. « Textes à l'appui / Anthropologie des sciences et des techniques », 1993. La lecture de ce très beau livre est indispensable à tous ceux qui souhaiteraient revenir sur la construction historique des méthodes d'analyse statistique, et notamment sur les débats auxquels elle a donné lieu, dont certains ne sont pas définitivement clos.

distribution des hauts revenus dans une loi statistique qui porte toujours son nom. Plus près de nous, c'est un sociologue, Jean Stoetzel, qui fonde après la seconde guerre mondiale le premier institut de sondage français (l'Institut français de l'opinion publique, IFOP).

Depuis Émile Durkheim, l'utilisation des méthodes statistiques en sociologie n'a cessé de se développer, allant de l'utilisation de données produites par les organismes gouvernementaux (Institut national des statistiques et études économiques (INSEE), Direction des études et de la prospective du ministère de l'Éducation nationale (DEP), etc.) à la réalisation d'enquêtes spécifiques par questionnaire, en passant par la quantification d'informations déjà existantes ou l'analyse secondaire de données traitées par ailleurs. Contrairement à ce que croient parfois les étudiants, l'enquête par questionnaire n'est donc que l'une des sources possibles pour l'utilisation de ces méthodes. L'analyse statistique n'est pas non plus réservée aux approches macrosociologiques. L'observation microsociologique peut donner lieu à la quantification de comportements¹ et l'analyse des entretiens peut être facilitée par des statistiques textuelles. Ainsi, si l'on a coutume d'opposer l'enquête par questionnaires aux techniques d'entretiens et d'observation, en argumentant que la première s'applique à de grands nombres, alors que les secondes ne portent que sur des nombres restreints d'unités, pour l'ensemble de ces instruments, la statistique n'est en réalité jamais bien loin, rendant ainsi leur opposition souvent superficielle. Comme le rappelle Jean-Claude Passeron², ces outils, quoique dif-

1. Pour décrire une scène de la vie hospitalière, on pourra compter le personnel soignant mobilisé pour une opération chirurgicale, le nombre de patients admis dans une journée au service des urgences, calculer la durée totale des interventions d'un chirurgien dans une journée, etc.

2. Cf. J.-C. Passeron, *Le raisonnement sociologique*, Paris, Nathan, coll. « Essais et recherches », 1991.

férents – c'est-à-dire quoique différemment situés sur l'échelle d'objectivation de l'information –, se distinguent moins souvent qu'on ne l'imagine par leur usage des descripteurs numériques (ou quantificateurs) et leurs résumés chiffrés de l'information.

L'utilisation des analyses statistiques a suscité (et suscite encore) parmi les sociologues de nombreux débats comme celui qui a opposé aux États-Unis la tradition de recherche souvent appelée « École de Chicago »¹ – utilisant de façon dominante l'observation, les entretiens ou des recensements officiels – aux tenants de la méthode du questionnaire et de l'analyse statistique, dont la figure la plus citée est Paul Lazarsfeld². L'enjeu de ces débats est généralement de discuter des limites d'une méthode dont le raffinement technique tend parfois à faire oublier les conditions de possibilité de son usage et les hypothèses implicites que cet usage implique.

Les statistiques ne constituent bien évidemment que l'un des outils dont dispose le sociologue pour étudier les phénomènes sociaux. L'observation, les entretiens, l'analyse historique et bien d'autres encore sont tout aussi légitimes. Un des grands intérêts des statistiques par rapport aux autres méthodes est de permettre d'évaluer l'importance d'un événement et d'apporter des preuves dans certains contextes. Les entretiens ou l'observation permettent de montrer l'existence d'un phénomène et de l'analyser en profondeur, mais ils se prêtent mal à une extrapolation vers d'autres parties de la population que celle qui a été étudiée : comment être sûr que ce que l'on observe ne se limite pas au petit groupe que l'on étudie ? En quantifiant

1. Yves Grafmeyer et Isaac Joseph, *L'École de Chicago, naissance de l'écologie urbaine*, Paris, Aubier, 1984.

2. P. Lazarsfeld, *Qu'est-ce que la sociologie ?*, Paris, Gallimard, 1970. Sur ces débats, voir par ex. C.-H. Cuin, F. Gresle, *Histoire de la sociologie*, La Découverte, coll. « Repères », n° 110, 1992 ; ou J.-M. Berthelot, *La construction de la sociologie*, PUF, coll. « Que sais-je ? », n° 2602, 1995.

ce qui dans les phénomènes observés est dû au hasard et ce qui est dû à d'autres causes, les statistiques permettent d'établir des régularités irréfutables par de simples contre-exemples. C'est le cas par exemple des inégalités de réussite scolaire selon l'origine sociale. Personne n'oserait prétendre à présent que toutes les catégories sociales sont à égalité devant le système éducatif. Raymond Boudon, Pierre Bourdieu et Jean-Claude Passeron, Christian Baudelot et Roger Establet et bien d'autres ont développé des théories expliquant ces inégalités, mais aucun ne les conteste.

Comme le montre ce dernier exemple, les statistiques se prêtent bien à l'étude de phénomènes de masse tels que la consommation, la réussite scolaire, les choix électoraux, toutes situations où l'on peut considérer, en première approximation, que les acteurs sociaux sont confrontés indépendamment les uns des autres à des situations semblables. Elle permettent dans ce cas de mettre en évidence des phénomènes dont elles ne fournissent en aucun cas l'explication. L'explication, c'est l'affaire des sociologues. Elles peuvent aussi être extrêmement utiles dans toutes les autres situations mais avec certaines limites, dont leur difficulté à prendre en compte les aspects longitudinaux ou les relations entre acteurs.

Il est donc nécessaire pour un sociologue de connaître suffisamment les statistiques pour s'en servir. L'existence de logiciels permet d'envisager cet apprentissage en y intégrant le moins possible de développements mathématiques. Un peu comme un conducteur doit savoir à quoi servent la pédale de frein et le levier de vitesses, mais n'est pas obligé de savoir démonter le moteur, un sociologue doit comprendre les principales notions de statistique pour mettre en œuvre des techniques dont il n'a pas besoin de connaître en détail les structures mathématiques.

Cet ouvrage est notamment destiné aux étudiants de sociologie confrontés à l'apprentissage parfois un peu

intimidant de savoirs qu'ils perçoivent comme relevant d'abord des mathématiques. En effet, un des contresens que commettent généralement les étudiants en sciences humaines (et plus généralement les utilisateurs des indices statistiques) est de penser que la principale difficulté des statistiques provient de l'existence de formules mathématiques jugées ardues. Or, dans la pratique, c'est plutôt l'incompréhension ou la méconnaissance du raisonnement sociologique qui s'avère constituer la réelle difficulté dans l'utilisation de ces indices. Quel intérêt y a-t-il à calculer tel indice ? Quel sens cette opération prend-elle dans le travail d'enquête ? Dans quelle mesure est-on fondé à utiliser tel coefficient plutôt que tel autre ? Ces questions ne trouvent pas toujours les réponses attendues. A croire que les efforts mis à réaliser les calculs mathématiques nécessaires devaient faire perdre aux utilisateurs leur lucidité analytique pourtant si indispensable dans ce genre d'exercice. « Je ne sais pas bien à quoi va servir mon calcul, mais je m'applique à le faire juste ! » Cette formule provocatrice (de notre invention) se veut ici un clin d'œil adressé aux différentes générations d'étudiants que nous avons côtoyés, et dont la principale erreur était d'oublier que les méthodes quantitatives qui leur étaient enseignées étaient avant tout appliquées, c'est-à-dire au service de raisonnements théoriques et de constats empiriques. Cet ouvrage s'adresse tout particulièrement à eux.

□ L'intention générale de cet ouvrage est de clarifier la place du raisonnement statistique en sociologie en montrant notamment l'importance des démarches en termes de comparaison et de corrélation, et leurs cortèges d'indices statistiques. Conçu tout particulièrement pour les étudiants de sociologie, de la première année d'étude au troisième cycle, ce manuel est donc avant tout un « instrument de travail » destiné à présenter un certain nombre de méthodes et de techniques statistiques nécessaires au traitement de leurs données. La présentation de ces méthodes

s'appuiera sur des exemples concrets extraits d'enquêtes dont les références seront mentionnées de sorte que le lecteur puisse se reporter à la problématique sociologique sous-jacente.

Cet ouvrage se veut également en filigrane un lieu d'explication du rôle de ces instruments, et de réflexion sur leurs conditions et leurs limites d'utilisation. Expliciter l'usage de la statistique en sociologie, en souligner la valeur, obligent en effet en même temps à en mentionner les limites. Tel est le dessein de cet ouvrage, même si nombre de critiques de type épistémologique que l'on pourrait adresser à la statistique dépassent le cadre de cet ouvrage, et donc ne seront pas abordées. Maîtrise mathématique des données sociologiques et lucidité d'utilisation sont par conséquent les deux aptitudes intimement liées que cet ouvrage se propose de transmettre aux utilisateurs. Le chapitre III de la seconde partie de ce livre en est une parfaite illustration, puisqu'il a pour objectif d'initier l'étudiant à la procédure « quasi expérimentale » du traitement multivarié, tout en signifiant qu'en sociologie on ne parvient jamais en réalité à « un contrôle suffisamment complet de toutes les variables pertinentes pour "raisonner toutes choses égales d'ailleurs" »¹, étant donné le nombre de variables interconnectées qui composent la réalité sociale, qui en fondent son caractère historique ou contingent, et qui rendent, au bout du compte, l'analyse sociologique irréductible à toute méthode expérimentale « en laboratoire ».

C'est pourquoi nous nous sommes centrés sur ce qui constitue le noyau des méthodes statistiques utilisées le plus habituellement par les sociologues. Cela définit un ensemble qui ne recouvre ni l'ensemble des méthodes sta-

1. M. Simon, Introduction à *Historiens et sociologues aujourd'hui*, Journées d'études annuelles de la Société française de sociologie, Éd. du CNRS, 1986, p. 15.

tistiques (nous ne détaillons pas par exemple la batterie considérable des tests très souvent utilisés par les psychologues ou certains modèles fréquents en économie), ni tous les usages des mathématiques en sociologie (comme la théorie des graphes dans l'analyse des réseaux sociaux)¹.

Éléments de *statistique descriptive* et méthodes de *statistique inférentielle* sont réunis dans cet ouvrage.

Par statistique descriptive, on entendra la présentation des méthodes de classement, de dénombrement, et de résumé de l'information des données numériques :

- *valeurs de tendance centrale et de dispersion* des variables selon leurs caractéristiques (Première partie, chap. I) ;
- *traitements élémentaires de tableaux croisés* (Deuxième partie, chap. I) ;
- *indices « analytiques » d'ajustement* (linéaire et non linéaire) d'une série croisée d'observations (Deuxième partie, chap. II, § 2 et 4) ;
- *analyses multidimensionnelles et multivariées* (Deuxième partie, chap. III) ;

Par statistique inductive, on entendra :

- la présentation des *méthodes de sondage* (constitution d'échantillons d'enquête : Première partie, chap. II) ;
- et l'utilisation des *tests d'hypothèses*, qu'il s'agisse des tests de corrélation de variables nominales (Deuxième partie, chap. II, § 1) ou ordinales (§ 3).

Comme manuel d'introduction, cet ouvrage est destiné à permettre au lecteur de combiner maîtrise du raisonnement expérimental et développement du raisonnement sociologique. Il s'agit pour nous, avant tout, de faire comprendre les idées et le raisonnement qui sont à la base des méthodes statistiques utilisées sur les logiciels les plus

1. A. Degenne, M. Forsé, *Les réseaux sociaux*, Paris, A. Colin, coll. « U », 1994.

répandus, la présentation détaillée des techniques mathématiques pouvant être trouvée dans des ouvrages plus techniques et plus volumineux¹.

En tant que sociologues, notre objectif serait sans doute atteint si nous permettions à quelques générations d'étudiants en sciences humaines d'accepter et de comprendre la contribution de la statistique élémentaire à l'analyse des faits sociaux.

1. A titre d'exemple et pour rester dans les ouvrages pour étudiants de sciences humaines, le livre assez complet (bien que restreint aux techniques enseignées habituellement dans les universités anglophones) de David C. Howell, *Méthodes statistiques en sciences humaines*, De Boeck Université, Bruxelles, 1998 (il s'agit de la traduction d'un ouvrage rédigé à l'origine pour les psychologues).

PREMIÈRE PARTIE

Les différentes étapes dans la production des données d'enquête

L'ambition, tout réelle de surcroît, doit être limitée, et limitée celle qui consiste à penser que l'objet de la recherche est de fournir des preuves définitives et absolues de validité des résultats d'enquête. Les instruments que l'on utilise ont leurs limites, d'abord, un champ bien défini de leur utilisation, la mesure des chiffres qui l'est et ensuite à ne donner que des renseignements pour servir de base à l'interprétation, et dans certains cas des conclusions descriptives ou indicatives.

Avant d'être réunies, l'information doit être contrôlée, et ce à partir de deux types d'ingrédients. Le premier est constitué par le connaissance que le sociologue peut avoir des phénomènes étudiés, à l'aide d'une démarche exploratoire, ou d'enquêtes préliminaires, par consultation de résultats produits par d'autres chercheurs, ou encore par supposition de ce que l'on connaît sur d'autres phénomènes. Le second est lié aux problèmes généraux ou spécifiques que le sociologue cherche à résoudre, ce qu'on appelle habituellement le problème-type. Il s'agit pas d'analyse purement exploratoire, sans questionnement préalable. Le regard sociologique est tou-

rigueur, le processus de détermination de la méthode statistique est un processus de recherche scientifique et non pas un processus de décision technique.

En fait, les méthodes statistiques sont des outils de recherche scientifique et non pas des outils de décision technique. Elles sont destinées à servir la recherche scientifique et non pas la décision technique.

Les différents types

Il existe trois types de méthodes statistiques : les méthodes descriptives, les méthodes d'inférence et les méthodes de régression. Les méthodes descriptives ont pour but de décrire les données, les méthodes d'inférence ont pour but de tirer des conclusions à partir de données échantillonnées et les méthodes de régression ont pour but de modéliser la relation entre deux variables.

1970

1970

Les méthodes statistiques sont classées en deux grandes catégories : les méthodes descriptives et les méthodes d'inférence. Les méthodes descriptives ont pour but de décrire les données, tandis que les méthodes d'inférence ont pour but de tirer des conclusions à partir de données échantillonnées. Les méthodes de régression sont une sous-catégorie des méthodes d'inférence et ont pour but de modéliser la relation entre deux variables.

CHAPITRE I

La construction des variables : introduction à la statistique descriptive

D'emblée, tout risque de malentendu doit être écarté, en priorité celui qui consiste à penser que l'objet de la statistique est de fournir des preuves irréfutables et absolues de validation des résultats d'enquête. Les instruments que l'on utilise ont leurs limites, c'est-à-dire un champ bien défini de mise en œuvre. Et les indices (chiffrés) que l'on est amené à produire ont fondamentalement pour objet de *résumer l'information*, et donc constituent des raccourcis descriptifs ou inductifs.

Avant d'être résumée, l'information doit être construite, et ce à partir de deux types d'ingrédients. Le premier est constitué par la connaissance que le sociologue peut avoir du phénomène étudié, à l'issue d'une démarche exploratoire, ou d'enquêtes antérieures, par réutilisation de résultats produits par d'autres chercheurs, ou encore par transposition de ce que l'on connaît sur d'autres phénomènes. Le second est lié aux problèmes généraux ou spécifiques que le sociologue cherche à résoudre, ce qu'on appelle habituellement la problématique. Il n'existe pas d'analyse purement exploratoire, sans questionnement préalable. Le regard sociologique est tou-

jours orienté et l'explicitation du questionnement est le préalable indispensable à toute démarche d'enquête. La construction de l'information en vue d'un traitement statistique consiste à mettre en équivalence des événements, objets ou informations par nature singuliers en décidant d'isoler une caractéristique stylisée que l'on va constituer en *variable*.

Pour mettre en œuvre les méthodes de l'analyse statistique, il faut en effet construire un *ensemble d'informations* qui se présente sous la forme d'un ensemble d'*unités statistiques* et d'un ensemble de *variables* décrivant ces unités statistiques. Les unités statistiques sont des entités comparables sur le plan des variables. Les variables sont des informations sur les unités statistiques, telles que les unités statistiques soient du point de vue de ces informations équivalentes ou ordonnables. On obtient alors un classique tableau de données :

Unités statistiques	Variables		
	Variable 1	Variable 2	Variable 3
Unité 1			
Unité 2			
Unité 3			
Unité 4			

En sociologie, les unités statistiques ne sont pas nécessairement des individus. Elles peuvent être des entités collectives (des entreprises par ex.), des événements (une interaction, un comportement), des objets (textes, images). L'important est que l'on puisse justifier la mise en équivalence de ces unités en ce qui concerne les variables que l'on va construire. Voici quelques exemples de tableaux très différents pouvant être analysés au moyen des mêmes méthodes :

<i>Individus interrogés</i>	<i>Variables</i>		
	<i>Age (en années)</i>	<i>Sexe</i>	<i>Profession</i>
Individu 1	32	Homme	Ouvrier
Individu 2	25	Femme	Cadre ou profession intellectuelle supérieure
Individu 3	56	Femme	Profession intermédiaire
Individu 4	43	Homme	Profession intermédiaire
etc.			

Dans ce tableau, l'information correspondant à la troisième variable est présentée sous la forme de catégories, ici les professions et catégories socioprofessionnelles de l'INSEE dans la version la moins détaillée. Supposons que le troisième individu soit une institutrice et le quatrième un infirmier. Puisque nous avons choisi d'utiliser une nomenclature simplifiée, les deux individus seront affectés à la même catégorie et donc considérés comme équivalents sur ce point. Si nous ne nous intéressons à la profession que comme indicateur général d'un niveau social, cette mise en forme est suffisante. Si au contraire nous nous intéressons à la profession en tant que telle, au secteur d'activité ou au temps de travail, alors il serait nécessaire d'utiliser une nomenclature plus détaillée.

<i>Entreprises</i>	<i>Variables</i>		
	<i>Effectif</i>	<i>Nombre d'ingénieurs</i>	<i>Présence de sections syndicales</i>
Entreprise 1	52	15	Oui
Entreprise 2	2 300	45	Oui
Entreprise 3	21	0	Non
Entreprise 4	143	87	Non
etc.			

Dans ce second exemple, l'unité statistique est l'entreprise et dans le troisième ce sont des textes, ce qui illustre le fait qu'en sociologie la base des traitements statistiques n'est pas forcément l'individu.

Dernier exemple :

<i>Tracts syndicaux</i>	<i>Nombre de mots</i>	<i>Variables</i>	
		<i>Présence de la thématique du temps de travail</i>	<i>Fédération</i>
Tract 1	2 500	Oui	CGT
Tract 2	1 658	Non	CGC
Tract 3	1 733	Oui	CFDT
Tract 4	3 142	Oui	Sud
etc.			

La pertinence de cette construction de variables ne constitue pas une qualité de statisticien, mais bien de sociologue qui doit pouvoir expliciter et soumettre à la critique ses choix en fonction du problème étudié.

Une fois collectées, les informations sont « saisies » sur un disque informatique, et ce en vue du traitement. Pour permettre cette saisie, on standardise préalablement cette information : c'est la phase du codage qui est une opération de standardisation numérique de l'information sociologique en vue de constituer une base de données : travail fastidieux et apparemment simple, la phase de codage des données consiste à transcrire les informations dans les catégories qui vont donner lieu au traitement informatique.

La phase de codage des données illustre parfaitement un des principes énoncés antérieurement, à savoir que les sciences sociales (dont fait partie la sociologie) travaillent non sur des données brutes, mais sur des données produites. Elle est partie intégrante de ce que l'on peut appeler

« le processus de production de l'information », dans la mesure où elle opère une formalisation des données. Le codage est donc une phase de traduction des énoncés en signes numériques : coder numériquement les modalités des variables afin de les dénombrer. Ces variables sont soit de type qualitatif, soit de type quantitatif. Dans les deux cas, on n'échappe pas au travail de classement, de regroupement de leurs modalités.

I. LES VARIABLES QUALITATIVES

Dans les tableaux que nous venons de voir, certaines variables prennent des valeurs qui sont des caractéristiques indiquées en toutes lettres (« ouvrier », « oui », « non », etc.). Il s'agit de variables qualitatives. On parle de variable qualitative lorsqu'il y a répartition des observations d'une population en un certain nombre de classes ; ce type de variable a de ce point de vue, comme toute échelle de mesure, deux propriétés :

- a) chaque observation doit appartenir à une classe et une seule ;
- b) deux observations appartenant à la même classe sont considérées comme équivalentes, ou identiques, pour ce qui concerne cette information.

Le sexe d'une personne, sa profession, sa situation de famille, la raison de son déménagement, la cause de son suicide... peuvent être considérés comme des caractères qualitatifs dans la mesure où leurs modalités (ou états) ne sont pas associés à un nombre, et donc mesurables. Les modalités « homme/femme », « ouvrier/employé » n'étant pas le résultat d'un critère statistique de comptage, on ne

Introduction aux méthodes statistiques en sociologie

THIERRY BLÖSS
MICHEL GROSSETTI

Depuis Emile Durkheim, l'utilisation des méthodes statistiques en sociologie n'a cessé de se développer. Cet ouvrage s'adresse tout particulièrement aux étudiants de cette discipline. Il a été conçu avant tout comme un « instrument de travail » destiné à présenter les différentes méthodes et techniques statistiques essentielles au traitement des données. La présentation de ces méthodes s'appuie sur des exemples concrets extraits d'enquêtes sociologiques. Quel intérêt y-a-t-il à calculer tel ou tel indice? Quel sens prend cette opération dans le travail d'enquête? Comment interpréter les résultats? Telles sont les questions auxquelles cet ouvrage apporte les réponses nécessaires à des étudiants de sociologie parfois déroutés par l'apprentissage des statistiques qu'ils identifient trop vite aux mathématiques.

Comme manuel d'introduction, cet ouvrage a pour objectif de faire comprendre la contribution de la statistique à l'analyse des faits sociaux. Il expose les principales techniques d'analyse utilisées dans les logiciels de traitement de données les plus répandus.

Thierry Blöss est professeur de sociologie à l'Université de Provence où il enseigne notamment les méthodes statistiques. Il est auteur de plusieurs ouvrages dont *Les liens de famille*, *Sociologie des rapports entre générations* (PUF, 1997), *Education familiale et Beau-parenté* (L'Harmattan, 1996) *La femme dans la société française* (avec A. Frickey, PUF, 1996).

Michel Grossetti est sociologue, chercheur au CNRS, et a une expérience de statisticien. Il a publié entre autres *Science, industrie et territoire* (Presses Universitaires du Mirail, 1995).

Maquette Laurence Verrier
En couverture: *Business City* (détail)
de J.P. Stora, 1995
© The Grand Design / Superstock



Participant d'une démarche de transmission de fictions ou de savoirs rendus difficiles d'accès par le temps, cette édition numérique redonne vie à une œuvre existant jusqu'alors uniquement sur un support imprimé, conformément à la loi n° 2012-287 du 1^{er} mars 2012 relative à l'exploitation des Livres Indisponibles du XX^e siècle.

Cette édition numérique a été réalisée à partir d'un support physique parfois ancien conservé au sein des collections de la Bibliothèque nationale de France, notamment au titre du dépôt légal. Elle peut donc reproduire, au-delà du texte lui-même, des éléments propres à l'exemplaire qui a servi à la numérisation.

Cette édition numérique a été fabriquée par la société FeniXX au format PDF.

La couverture reproduit celle du livre original conservé au sein des collections de la Bibliothèque nationale de France, notamment au titre du dépôt légal.

*

La société FeniXX diffuse cette édition numérique en accord avec l'éditeur du livre original, qui dispose d'une licence exclusive confiée par la Sofia – Société Française des Intérêts des Auteurs de l'Écrit – dans le cadre de la loi n° 2012-287 du 1^{er} mars 2012.

Avec le soutien du

