

LIBRES COURS  
ÉCONOMIE

# Statistiques descriptives

L'ÉCONOMIE ET LES CHIFFRES

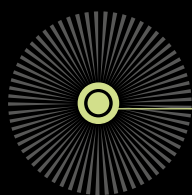
P. Bailly et C. Carrère

PUG

Les chiffres constituent une part majeure de l'information économique et sociale, en particulier les données statistiques, dont médias et gouvernants usent et abusent : elles sont un outil incontournable de connaissance de la réalité. Pour autant, les chiffres ne disent rien si on ne sait pas les faire parler.

Ainsi, la statistique suppose une bonne maîtrise de la théorie afin de produire des données cohérentes puis une analyse fine des résultats qui, seules, permettent de faire parler les chiffres.

En présentant l'essentiel des outils à connaître (les tendances centrales, de dispersion, de concentration, les ajustements, le traitement des chroniques, la présentation des indices), l'ouvrage constitue une excellente introduction aux traitements statistiques pour les étudiants de 1er cycle en économie-gestion (Licence, IUT, BTS). Il associe théorie et cas pratiques avec de nombreux exemples concrets et en contexte, afin d'éclairer la théorie par l'illustration.



**PIERRE BAILLY** enseigne les statistiques descriptives à la Faculté d'économie de Grenoble de l'université Grenoble Alpes.

**CHRISTINE CARRÈRE** est enseignante de statistiques.



Presses universitaires de Grenoble  
BP 1549 - 38025 Grenoble cedex 1  
ISBN 978-2-7061-2215-6 (e-book PDF)

# Statistiques descriptives



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.

Création de couverture : Corinne Tourrasse

© Presses universitaires de Grenoble, mars 2015  
5, place Robert-Schuman  
BP 1549 – 38025 Grenoble cedex 1  
pug@pug.fr / www.pug.fr

ISBN 978-2-7061-2215-6 (*e-book PDF*)

L'ouvrage papier est paru sous la référence ISBN 978-2-7061-2214-9.

Pierre Bailly  
Christine Carrère

# **Statistiques descriptives**

L'économie et les chiffres

Presses universitaires de Grenoble

## DANS LA MÊME COLLECTION

### **Droit**

- O. Soria, *Droit de l'environnement industriel*, 2013  
M. Pérès, *Droit et responsabilité en montagne. Jurisprudence relative aux activités sportives et touristiques en montagne*, 2006  
D. Mallet, P. Balme, P. Richard (dir.), *Réglementation et management des universités françaises*, 2005  
P. Pedrot (dir.), *Génétique, biomédecine et société*, 2005  
F. Servoin, *Droit administratif de l'économie*, 2001

### **Économie**

- R. Colliat et Y. Échinard (dir.), *Quelle fiscalité pour le XXI<sup>e</sup> siècle ? Contributions au débat*, 2014  
P. Bailly, C. Carrère, *Statistiques descriptives. Cours*, 2<sup>e</sup> édition, 2007  
P. Bailly, C. Carrère, *Statistiques descriptives. Exercices avec corrigés*, 2<sup>e</sup> édition, 2007  
H. Drouvot, *Le Made in Brésil. L'industrie brésilienne face à la mondialisation*, 2005  
C. Perret (dir.), *Perspectives de développement pour la Nouvelle-Calédonie*, 2002  
M. Lejeune, *Traitements des fichiers d'enquêtes. Redressements, injections de réponses, fusions*, 2001  
A. Samuelson, *Les grands courants de la pensée économique*, 5<sup>e</sup> édition, 1997

### **Gestion**

- P. Balme, J.-R. Cytermann, M. Dellacasagrande, J.-L. Reffet, P. Richard, D. Verhaeghe, *L'université française : une nouvelle autonomie, un nouveau management*, 2012  
B. Derrouch, *Gestion comptable et financière de l'entreprise*, 2005

### **Sciences politiques**

- Y. Deloye, O. Ihl, A. Joignant (dir.), *Gouverner par la science : perspectives comparées*, 2013  
G. Gourgues, *Les politiques de démocratie participative*, 2013  
M. Hollard, G. Saez (dir.), *Politique, science et action publique. La référence à Pierre Mendès France et les débats actuels*, 2010  
C. Bidégaray, S. Cadiou et C. Pina, *L'élu local aujourd'hui*, 2009  
M. Chauchat, *Vers un développement citoyen. Perspectives d'émancipation pour la Nouvelle-Calédonie*, 2006  
J.-L. Chabot, *Aux origines intellectuelles de l'Union européenne. L'idée d'Europe unie de 1919 à 1939*, 2005

### **Psychologie**

- E. Grebot, *Repères en psychopathologie*, 2002  
E. Grebot et I. Orgiazzi Billon-Galland, *Les bases de la psychopathologie – Éléments historiques, notionnels et théoriques*, 2001

### **Sciences**

- G. Dhont, B. Zhilinski, *Symétrie dans la nature*, 2011

### **Sociologie**

- R. Levy, C. Soldano et P. Cuntigh (dir.), *L'université et ses territoires. Dynamismes des villes moyennes et particularités de sites*, 2015  
A. Baron, *Innover dans les politiques sociales. Pratiques du changement*, 2013  
A. Baron, *Dynamiques territoriales de l'action sociale et médico-sociale*, 2010  
F. Moutet, *La Féminisation des effectifs chirurgicaux*, 2010

# Introduction

---

Les observations numériques, en particulier celles d'ordre statistique, participent massivement de l'information économique et sociale dont les médias et les gouvernants usent et abusent. Les données statistiques sont le complément indispensable pour les démonstrations économiques, elles justifient d'accepter ou de rejeter telle ou telle analyse. Elles constituent un outil incontournable de connaissance de la réalité économique<sup>1</sup>, en particulier parce qu'elles permettent de quantifier des phénomènes et donc d'estimer leur importance, ce que n'autorise pas une approche purement qualitative. Cependant, la statistique est tout autant science de la classification que de la quantification. En ce sens, elle ne se réduit pas à la production de nombres.

5

La maîtrise des méthodes du calcul statistique ne suffit pas pour produire des résultats intéressants ; la qualité des résultats est fonction de la qualité des données à l'entrée souvent exprimée par la formule anglaise *garbage in, garbage out*. L'évaluation de la qualité des données utilisées, associée à une bonne connaissance des concepts et des notions est primordiale pour apprécier la pertinence des résultats obtenus et en proposer un commentaire. Les chiffres fournissent la mesure de variables économiques, ils n'expliquent ni comment ni pourquoi elles prennent telle ou telle valeur, encore moins les conséquences économiques de la grandeur de celles-ci. La signification des chiffres est une des conditions premières de la compréhension de la situation économique.

Le champ de cet ouvrage se concentre sur la statistique dont l'objet est de fournir une description ; ni les lois de probabilités ni les statistiques inférentielles et de la décision ne seront traitées. L'inférence statistique, la prévision et l'estimation qui constituent la suite logique des domaines examinés ne seront pas traités dans cet ouvrage.

---

1. Cette réalité n'est pas indépendante des moyens utilisés pour l'appréhender.

## Qu'est-ce que la statistique ?

Le terme même de statistique est largement polysémique, quelques explications sont nécessaires pour préciser les différentes notions. Comme pour chaque champ de la connaissance, il existe un vocabulaire technique spécifique dont la compréhension est indispensable pour apprécier la pertinence des données utilisées. Les statistiques sont le produit d'une démarche, d'une construction théorique. Elles ne peuvent exprimer toute la complexité de la réalité économique qui elle-même ne peut être réduite aux aspects quantitatifs. Contrairement à la conception spontanée et illusoire qu'il suffit de faire une enquête pour obtenir les bonnes informations qui seraient déjà disponibles prêtes à se révéler à l'enquêteur impartial, le moindre questionnaire demande, pour apporter des données utilisables, une réflexion sur les dispositifs pertinents à mettre en œuvre en vue de cet objectif. Le modèle paradigmatique de cette conception reste le recensement, une opération difficile et coûteuse, qui fournit de précieuses et fiables observations avec, néanmoins, des ordres des marges d'erreur estimées.

Cet ouvrage portera sur la production d'informations statistiques dans le champ de l'économie et du social, il laisse donc de côté la statistique mathématique. La science statistique ne se réduit pas à une simple technique d'analyse et de présentation d'informations économiques et sociales quantifiées : un calcul, aussi simple soit-il, n'a de sens que par rapport à des données – validité, fiabilité, précision – et en vue d'un objectif.

Les données n'existent pas dans tous les domaines car la production de statistiques répond à une ou des demandes, plus ou moins vagues, qui doivent être formalisées par l'institution statistique et les statisticiens. Les choix des données à analyser et les orientations d'investigation sont arbitrés au niveau des décideurs, que ce soient des institutions publiques ou privées. S'ajoutent à ces contraintes « politiques » des contraintes budgétaires qui orientent la production de statistiques du système public comme des acteurs privés. Enfin les conditions de mobilisation, de recherche et de traitement des données dépendent de l'existence et de la qualité des organismes assurant ces missions. La neutralité et l'objectivité des statistiques sont néanmoins assurées par l'implication des professionnels de la production des données. La production statistique a pour premier objectif de fournir des observations puis dans un second temps d'assurer le traitement de celles-ci. Le second aspect est souvent le seul développé dans les manuels, ce qui tend à réduire les statistiques à un ensemble de techniques mathématiques plus ou moins complexes sans aucune interrogation préalable sur les observations faisant l'objet des calculs.



## Quelques définitions

La statistique est un ensemble de principes et de méthodes scientifiques pour recueillir, classer, synthétiser et communiquer des données numériques en vue de leur utilisation pour en tirer des conclusions et prendre des décisions. La diversité des usages du mot « statistique » reflète la double nature des pratiques sociales qui lui sont associées. À l'activité administrative d'élaboration des données se combine la réflexion scientifique mathématique. Autrefois imbriquées, les deux significations s'autonomisent au début du XIX<sup>e</sup> siècle. Le terme de statistique est riche de significations, au singulier c'est un ensemble de techniques mathématiques de traitement des données numériques. La statistique renvoie à une méthode scientifique, une branche des mathématiques<sup>2</sup> dont les principes découlent de la théorie des probabilités et qui a pour objet le groupement méthodique ainsi que l'étude des séries de faits ou de données numériques. Au pluriel, les statistiques sont synonymes de nombres, de données, d'informations numériques, elles indiquent une pluralité de phénomènes à travers les nombres attachés à l'appréhension de ceux-ci. Avec un article indéfini : une statistique est une série de nombres parmi d'autres séries possibles. De plus, le terme de statistique désigne fréquemment une série numérique, nous emploierons plus volontiers l'expression de distribution statistique.

La connaissance statistique est le rapport entre un besoin d'information et les moyens disponibles pour les produire. Elle se situe à la jonction d'une démarche théorique et d'une démarche empirique, un lieu assez peu confortable. La mesure dépend de conventions portant sur la définition de l'objet et les procédures de codages souvent d'origines administratives. C'est en particulier le cas des nomenclatures qui finissent parfois comme les catégories socioprofessionnelles par devenir hégémoniques dans l'appréhension de la réalité à décrire. Elle résulte également de l'application de concepts théoriques produits par des travaux scientifiques. Ces deux sources peuvent parfois fournir des résultats divergents. L'appréciation du chômage en est un exemple emblématique avec les deux estimations concurrentes, d'une part le nombre de demandeurs d'emploi en fin de mois, DEFM, mesurés par Pôle emploi dans une logique de gestion, et d'autre part le chômage au sens du Bureau international du travail (BIT), résultat d'enquêtes de l'INSEE pour des comparaisons internationales.

---

2. Ce n'est qu'avec l'autonomisation de ce champ de recherche, crée par Galton et Pearson et introduite en France par Lucien March, au début du XX<sup>e</sup> siècle que ce sens est admis. A. Girard, « La recherche en histoire de la statistique », *Courrier des Statistiques*, n° 32, octobre 1984.

Un calcul statistique n'a de sens que pour l'étude de populations importantes et que s'il permet d'obtenir un gain d'information ou de signification par rapport à une analyse immédiate des données. Les opérations de calcul réduisent notablement le volume des informations initiales. Les méthodes de la statistique descriptive ambitionnent de condenser un ensemble de données numériques en quelques indicateurs significatifs.

Le traitement statistique est aujourd'hui effectué automatiquement à l'aide d'ordinateurs. Le traitement de grandes quantités de données s'accomplit aisément et dans des délais très brefs, quasi instantanés. La compréhension, l'effort d'assimilation et d'appropriation des méthodes statistiques constituent, néanmoins, un préalable à une utilisation efficace et pertinente de ces outils. C'est pourquoi nous ne présentons aucun traitement sur ordinateur.

## Une courte histoire

Si le mot statistique est relativement récent puisqu'il semble avoir été introduit en Allemagne au xvii<sup>e</sup> siècle, la pratique des statistiques est par contre ancienne. Les grands empires antiques centralisateurs et unificateurs ont été, en raison de leur nature, confrontés à la nécessité du dénombrement des hommes et des biens pour les grands travaux et la levée des armées que ce soit en Mésopotamie, dans l'Égypte ancienne, en Chine antique comme dans l'Empire indien. Il s'agissait selon les époques de connaître la population pour la répartir sur le territoire, distribuer les terres, établir les rôles d'imposition, des corvées, des conscriptions militaires... Les dénombremments dans ces civilisations ont certes des significations fiscales ou militaires, mais aussi une portée magique ou religieuse, les unes et les autres fortement imbriquées.

En Grèce, les dénombremments distinguent les hommes libres, les métèques et les esclaves. La réflexion porte sur le nombre idéal de citoyens que doit comporter une cité et sur les moyens de le maintenir. À Rome, les recensements sont périodiques tous les cinq ans puis tous les dix ans. Ils permettaient de déterminer qui était citoyen romain, et de classer les citoyens d'après leurs revenus et de lever l'impôt. Ils étaient aussi le moyen de définir la place de chacun dans l'organisation politico-administrative militaire de la cité.

En France, le raffermissement du pouvoir royal avec les Carolingiens amène à un renouveau des inventaires (les capitulaires) de tous leurs biens (hommes, habitations, céréales, bétail). En Angleterre au xi<sup>e</sup> siècle, on procède à un ensemble de relevés afin de produire un cadastre. On recense les noms de lieux, de leurs tenanciers, le nombre d'occupants de chaque demeure, celui

des serfs, des hommes libres et de l'étendue des terres. Ce recensement ne concerne ni le clergé, ni les femmes, ni les enfants, ni les pauvres. Dans ces cas, l'unité sociale de base est le feu non la personne.

Le xvi<sup>e</sup> siècle est celui où les États se centralisent et s'unifient par la recherche d'une cohérence interne. C'est l'époque où règne le mercantilisme et où les auteurs cherchent à mesurer la richesse du Prince. J. Bodin, mercantiliste français, expose les avantages d'une meilleure connaissance de la population du royaume que ce soit pour la guerre ou pour la fiscalité. Le dénombrement des biens s'avère indispensable pour que la charge fiscale de chacun soit équitable, afin d'éviter des troubles et guerres civiles. La théorie mercantiliste développe la thèse selon laquelle l'État accroît sa force en favorisant l'enrichissement des citoyens. L'augmentation de la population, qui permet de maintenir de bas salaires, est, elle aussi, source de richesse. Il convient alors de procéder à des dénombrements ainsi que de contrôler les hommes. L'ordonnance de Villers-Cotterêts qui institue l'obligation de l'enregistrement des naissances, des morts et des baptêmes interdit aussi les coalitions ouvrières. En parallèle, l'enregistrement des baptêmes permettra à l'Église catholique de déceler les adeptes des autres religions (protestants, juifs...). Les recensements apparaissent donc comme un instrument privilégié du gouvernement suivant un des préceptes de Descartes « de faire partout des dénombrements si entiers et de revues si générales qu'il fut assuré de ne rien omettre ».

La pratique des dénombrements développe les réflexions théoriques sur les méthodes à mettre en œuvre. Le mouvement s'accroît au xvii<sup>e</sup> et xviii<sup>e</sup> siècle, parallèlement à l'achèvement de la concentration du pouvoir entre les mains du monarque. La naissance de la statistique administrative en Allemagne, en France et en Angleterre illustre la diversité des approches<sup>3</sup>. Les stratégies de recherche d'informations économiques et sociales françaises opposent l'arithmétique politique anglaise à la statistique descriptive allemande.

Les statisticiens allemands raisonnent du point de vue de la puissance et de l'activité de l'État. La statistique consiste à recenser tout ce qu'il y a d'intéressant pour l'État, elle est descriptive et non quantitative. L'objet principal de l'effort statistique est de classer, d'organiser des observations hétéroclites. Le résultat est la construction de nomenclatures, un des aspects de la statistique moderne aussi essentiels que la dimension quantifiée.

Le contexte anglais est tout différent, il existe une société civile distincte de l'État qui est une partie de la société, et non sa totalité comme en Allemagne. Dans les années 1660, un ensemble de techniques d'enregistrement et de calcul apparaissent sous le concept de « L'arithmétique politique » qui en

---

3. Alain Desrosières, *La politique des grands nombres*, Paris : La Découverte, 1993 (Textes à l'appui).

utilisant les inscriptions dans les registres (des baptêmes, des mariages...) fournit des données démographiques et économiques pour l'ensemble du pays. Ce souci quantitatif s'applique tout d'abord aux domaines de l'économie et de la démographie. La conception libérale anglaise de l'État interdisant les grandes enquêtes, les arithméticiens doivent recourir à des méthodes indirectes de calcul.

Les deux conceptions de la statistique se confrontent puis un processus d'homogénéisation et de codification, l'unification des systèmes de référence, se réalise. L'opposition entre les deux approches se traduira et par la construction de nomenclatures – pour décrire – et par le recours au calcul – pour mettre en relation – au sein d'une production statistique administrative.

L'effort statistique se développe au cours du XIX<sup>e</sup> siècle avec une logique scientifique et dans une démarche intimement lié au politique : l'expansion des statistiques administratives est le signe d'une politique interventionniste, tandis que les périodes de régression sont les signaux d'une ère de libéralisme. Les crises et les bouleversements économiques auront pour effet une extension de l'intervention centralisée, à la fois sous la pression du mouvement ouvrier – contre l'anarchie du marché – et du sentiment des pouvoirs publics que l'absence de régulation automatique par le marché pouvait conduire à sa disparition c'est-à-dire aux révolutions. Les statistiques guident l'action des pouvoirs publics d'où les critiques récurrentes des libéraux non pas contre les chiffres, mais contre les statistiques<sup>4</sup>.

Le besoin d'informations économiques s'explique du fait que, dans une économie décentralisée, les informations économiques ne sont pas immédiatement disponibles du fait de l'absence de liens institutionnels entre les producteurs et les consommateurs. L'explosion des statistiques au XIX<sup>e</sup> siècle, en Europe occidentale s'explique par l'expansion du capitalisme concurrentiel triomphant ; tout en se heurtant à l'opposition des libéraux pour qui les prix suffisent pour orienter les actions des agents. Dans le cours des transformations de l'économie, l'État prend de plus en plus de responsabilités dans l'économie, voire gère directement une partie des activités économiques, ce qui induit une forte croissance des besoins d'informations économiques et sociales et donc statistiques de la part des pouvoirs publics et des acteurs qui utilisent les mêmes indicateurs.

Le XIX<sup>e</sup> siècle s'est donné l'illusion de pouvoir tout connaître avec du temps et de l'énergie que ce soit la nature, l'homme ou la société. La connaissance statistique a participé de cette tentative et de cette illusion. Des auteurs

---

4. D'Adam Smith à Paul Fabra en passant par Jean-Baptiste Say et Friederich von Hayek, cf. « L'économie aveuglée » in *La cité des chiffres ou l'illusion statistique*, Paris : Autrement, 1992.

se préoccupent des relations entre la statistique et les sciences humaines, parmi ceux-ci le Belge Adolphe Quételet et le Français Augustin Cournot. Le premier se propose de connaître les phénomènes sociaux par leur unique aspect chiffré, la statistique appliquée aux actes humains devait constituer une science qu'il appelle physique sociale. Beaucoup de données statistiques publiées à l'époque nous paraissent actuellement sans intérêt. L'effort de mesure, de la recherche d'une évaluation de l'exactitude de la mesure conduit à développer la théorie des probabilités. Les méthodes de calcul s'affinent, la statistique s'affirme de plus en plus scientifique, le développement de la statistique continue en se complexifiant.

Le xx<sup>e</sup> siècle se caractérise par l'organisation de systèmes statistiques sous l'impulsion, voire sous la direction, des pouvoirs publics. La conception de l'information statistique se modifie. Sous la pression des groupes sociaux, l'information économique tend à être considérée comme un bien à la disposition du public. Le principe est donc la gratuité d'accès aux informations, sous réserve du coût du support. L'augmentation des demandes particulières en la matière conduit des organismes publics comme l'Institut national des études économiques (INSEE) à vendre l'information. Il s'agit là d'une inflexion importante qui modifie les conditions de diffusion de l'information statistique.

## Le système statistique

Un système statistique se définit par un ensemble de pratiques, de méthodes et d'institutions, de ce point de vue, il existe un système statistique français. Ce système s'appuie sur l'existence de grands fichiers et répertoires d'industries ou d'établissements, des nomenclatures d'activités de métiers, des règles juridiques contraignantes. La réalité du système statistique est relativement récente.

« Dans les faits, le système statistique français ne date que de la Libération, avec la création de l'INSEE, des services statistiques des ministères, des grands fichiers et répertoires d'individus et d'établissements, des grandes nomenclatures d'activités et de métiers et la systématisation des sondages. »<sup>5</sup>

Ce système est encadré par des règles juridiques contraignantes, en particulier le secret statistique.

Les systèmes statistiques contemporains naissent de la rencontre dans les années 1930-1940 de la statistique administrative – (la production des chiffres – et de la statistique scientifique – la production des méthodes.

5. Michel Lévy, *Comprendre les statistiques*, Paris : Seuil, 1979 p. 14.

Les prémisses apparaissent dans les années 1930 et les grands traits du système actuel se mettent en place à la Libération. Le système statistique français est le produit des besoins d'une gestion d'ensemble de l'économie nationale, spécialement de la planification, spécifique de la période de reconstruction de l'économie française. Il se constitue au cours de la Seconde Guerre mondiale puis s'organise et se développe durant la période de croissance des « trente glorieuses » (1945-1975), enfin il se transforme sous l'impact des grandes mutations contemporaines liées à la construction européenne, la mondialisation, etc.

## La production des statistiques

Des données n'existent qu'en rapport à des objectifs. C'est pourquoi la production de données répond donc à des demandes explicites ou implicites des groupes sociaux, des administrations... Si elles sont formalisées par les statisticiens, elles subissent des transformations et des modifications tenant autant aux choix des statisticiens eux-mêmes que des contraintes techniques ou sociales<sup>6</sup>. Les choix sont réalisés au niveau des décideurs, ensuite les professionnels mettent en œuvre les décisions prises. Toutes les informations ne seront pas produites, certains groupes sociaux influents peuvent bloquer la recherche ou la publication d'informations statistiques. Les statistiques disponibles dépendent du système statistique tel qu'il est organisé. Le producteur national dominant est l'INSEE dont l'essentiel des données se situe dans le cadre théorique de la comptabilité nationale. Avant tout traitement, il faut s'assurer de la fiabilité des informations disponibles, sinon un calcul aussi complexe soit il n'aurait aucun sens. Les statistiques disponibles résultent d'un travail pratique mais elles sont aussi le produit de l'histoire. En effet, en statistique, il faut toujours garder à l'esprit que le présent n'est que l'aboutissement provisoire de processus historique. Le recueil des informations permet d'obtenir des données en grand nombre qu'il faut organiser pour les rendre utilisables.

Les informations proviennent soit d'une procédure explicite de recherche par le biais d'enquêtes ou de sondages soit d'une collecte et d'une mise en forme d'informations préexistantes. Dans tous les cas, la valeur des observations dépend étroitement des conditions du rassemblement des données brutes.

---

6. Voir en ce domaine l'ouvrage de Michel Volle, *Le métier de statisticien*, Paris : Hachette, 1980.

## La recherche d'informations

---

L'enquête est une des méthodes courantes de recherche d'information, elle suppose d'avoir défini l'unité statistique enquêtée et la population de référence. Les recensements et les sondages sont les deux formes d'enquêtes réalisées directement auprès des détenteurs de l'information. Si les recensements sont exhaustifs, les sondages sont partiels. Certaines enquêtes combinent les deux techniques en fonction des sous-populations repérées. En statistique, un recensement est une étude exhaustive de toutes les unités statistiques de la population étudiée tandis que le sondage consiste à enquêter auprès d'une partie seulement de la population. Il faut faire attention car avec le temps, le terme de recensement, sous-entendu le recensement de la population d'un pays, a pris un autre sens courant. Il s'agit d'une enquête effectuée sur un échantillon de grande taille réalisée périodiquement et permettant une extrapolation des résultats qui constituent un « instantané » de la population d'un pays. Le recensement de la population en France est désormais organisé selon ce principe. L'ensemble des habitants des communes de moins de 10 000 habitants est recensé une fois tous les cinq ans par roulement. Dans les communes de plus de 10 000 habitants, chaque année une enquête est réalisée auprès de 8 % des habitants. Au bout de cinq ans tous les habitants des communes de moins de 10 000 habitants ont été recensés et 40 % des habitants des communes de plus de 10 000 habitants. Chaque année, les résultats de recensement sont produits à partir des cinq enquêtes annuelles les plus récentes.

Les recensements constituent l'opération statistique fondamentale. Le recensement n'est pas seulement un dénombrement, c'est aussi la mesure de certains caractères des individus de la population considérée. Le coût élevé des recensements en limite l'usage et la fréquence. Les recensements périodiques de la population constituent toujours une opération irremplaçable mais lourde et coûteuse.

Les sondages sont des enquêtes portant sur une fraction de la population. Cette technique est fondée sur le principe selon lequel les informations obtenues par l'interrogation d'un échantillon peuvent, sous certaines conditions, être généralisées à l'ensemble de la population. Cela peut consister à estimer certaines caractéristiques inconnues ou à faire des tests pour déterminer si des hypothèses ou des affirmations à propos de caractéristiques inconnues sont acceptables. La méthode des sondages présume l'existence de régularités au sein de la population concernée par les questionnaires. Un sondage tend à valoriser les modalités les plus courantes et à minimiser les signaux faibles significatifs. L'échantillon doit être construit de manière à rendre la

généralisation vraisemblable avec une bonne probabilité. Pour ce faire, deux méthodes sont possibles : les méthodes empiriques dont celle des quotas et les méthodes aléatoires ou probabilistes.

Il existe une autre méthode de l'échantillonnage dite de convenance qui consiste à interroger les individus sur le lieu d'achat ou d'activité ou dans la rue mais cette méthode, au petit bonheur la chance, n'offre aucune garantie scientifique. De plus, le sondage auprès de volontaires est un type de sondage qui demande à des individus de répondre : les téléspectateurs d'une chaîne particulière, ou les utilisateurs des réseaux sociaux, les lecteurs d'un journal. Là encore, aucune garantie scientifique ne peut être envisagée.

Dans la méthode des quotas, s'appuyant sur la connaissance de la répartition de la population selon des critères pertinents pour l'étude, il est possible de construire un échantillon représentatif possédant la même structure que la population mère. L'enquête doit respecter cette structure. Cette méthode est souple et rapide, mais le risque d'erreur est mal connu.

Les méthodes probabilistes désignent une technique par laquelle chaque unité de la population cible a une probabilité donnée, connue ou calculable préalablement (avant le tirage) d'appartenir à l'échantillon. Il devient alors possible de mettre en œuvre les techniques du calcul de probabilités pour réaliser des inférences sur l'ensemble de la population. Dans le sondage aléatoire simple, chaque individu de la population a une chance non nulle de faire partie de l'échantillon. Dans le sondage aléatoire stratifié, la population est subdivisée en strates, groupes homogènes selon un critère lié à la variable à estimer, par exemple les régions lors d'une enquête nationale. Puis au sein de chaque strate un échantillon est construit par sondage aléatoire simple. Dans le sondage en grappes, la population est divisée en plusieurs sous-ensembles, les grappes de l'échantillon sont choisies par sondage aléatoire simple et tous les individus appartenant aux grappes sélectionnées sont interrogés. Le sondage aréolaire sélectionne des aires au lieu de grappes. Dans un sondage à plusieurs degrés la population est divisée en grappes. L'échantillon est construit en tirant par sondage aléatoire simple au sein des grappes.

Les méthodes probabilistes exigent de disposer d'un recensement exhaustif de la population sous forme de répertoire comprenant la liste de toutes les unités. La constitution et la maintenance d'un répertoire sont des opérations complexes puisque toutes les unités sans exception doivent être répertoriées. La principale source d'erreur est constituée par l'absence de repérage d'unités. Pour les individus, seul l'INSEE dispose de cette base, mais en raison de la loi sur le secret statistique elle ne peut la communiquer à quiconque.



Le questionnaire est l'outil essentiel de toute enquête quantitative. Les questions retenues doivent permettre d'obtenir les informations recherchées. Pour cela, les personnes interrogées doivent être capables de répondre aux questions (elles doivent posséder les renseignements et estimer pouvoir les donner) et les comprendre sans ambiguïté. Les questions équivoques ou suggérant la réponse, comme c'est bien souvent le cas dans les sondages d'opinion, ne fournissent aucune information pertinente. Les résultats issus de ce genre d'enquêtes conduisent à des conclusions erronées. L'administration du questionnaire peut se faire par l'intermédiaire de différentes méthodes. Le mode d'administration le plus courant est le face-à-face entre l'enquêteur et le répondant. Actuellement, les tablettes ou les ordinateurs tendent à remplacer les questionnaires papier qui restent le mode le plus fiable d'administration. Les enquêtes via les réseaux sociaux de toute nature, outre les risques de discrimination qu'elles comportent, sont source de beaucoup de désillusions sur leur représentativité et leur fiabilité. Une fois les questionnaires recueillis, il faut en vérifier la cohérence et repérer les erreurs – il existe toujours des erreurs. Elles proviennent des enquêtés, dont les réponses peuvent être inappropriées (réponse imaginée par convenance) voire fausses (ressenti, mensonge...), des dispositifs d'obtention des informations (la représentativité des enquêtes via Internet est sujette à des interrogations, le recensement de 1968 fut particulièrement jugé peu fiable), du traitement de celles-ci (erreurs de codage), etc.

## La collecte des informations existantes

Au cours de leurs activités pour les besoins de celles-ci, les agents économiques produisent des informations numériques. Les comptabilités d'entreprises en sont un exemple. Les statistiques des administrations, des ministères représentent une seconde source abondante de données. Tous ces sous-produits des activités des agents sont irremplaçables. Cependant, les données obtenues répondent aux besoins des agents non à ceux des économistes. Les réalités mesurées dépendent de définitions légales ou réglementaires et non de concepts ou notions économiques. Elles ne sont pas toujours adaptées aux objectifs des analystes d'où la nécessité de traduire ces informations dans les cadres adéquats.

Cette introduction fournit un cadre général de la démarche statistique, la suite de l'ouvrage développe les méthodes et techniques de traitement des données.

## Organisation de l'ouvrage

Le premier chapitre présente les notions indispensables de la statistique descriptive et les représentations graphiques tandis que le deuxième donne les techniques utilisées pour l'analyse statistique des distributions à une dimension avec un accent mis sur les tendances centrales et les caractéristiques de dispersion. Le troisième chapitre fournit les outils classiques de l'étude des distributions à deux dimensions et de la mise en lumière des liens entre deux variables. Le chapitre suivant sera l'occasion d'une étude d'une forme particulière des séries à deux dimensions, dont le temps : ce sont les chroniques très présentes dans le champ de l'économie. Nous verrons enfin, dans un cinquième chapitre les indices, un des outils les plus indispensables et les plus controversés dans le champ de l'économie.

## Chapitre 1

# Les outils

Ce chapitre regroupe sous l'intitulé « Les outils » les concepts de base, les rudiments du vocabulaire des statistiques descriptives – les premières notions du calcul statistique – et enfin les principales représentations graphiques.

## Les concepts de base

Avant tout calcul statistique, il est nécessaire de disposer de données. Pour atteindre cet objectif, il s'agit de définir très précisément la population sur laquelle s'effectue l'enquête ainsi que les variables à analyser. Le type de ces variables conditionne les traitements statistiques qu'il sera possible de leur appliquer.

## La population et les unités statistiques

Dans le vocabulaire statistique, une population est un ensemble dont chaque élément est un individu ou une unité statistique. Les termes de population et d'individus sont employés aussi bien lorsqu'il s'agit d'un ensemble d'êtres humains : « la population résidente en France », « les salariés d'une entreprise », etc., que d'un ensemble d'objets inanimés : « la production automobile pour une année », « le stock des machines à une date donnée », et même d'ensembles abstraits ou des événements : « ensemble des jours d'une année », « la série du revenu national depuis vingt ans ». Chaque observation porte sur une unité statistique.

La population soumise à l'analyse statistique doit être définie avec précision afin que l'ensemble considéré soit déterminé sans ambiguïté de sorte qu'un individu quelconque puisse y être affecté sans incertitude. Quand on veut traiter de la population française au 1<sup>er</sup> janvier 2013 : il faut indiquer si

les étrangers résidant en France sont inclus et comment sont comptabilisés les Français résidants à l'étranger, il faudra alors préciser la signification du terme « résider ». Les chiffres de la population « française » correspondent aux nombres de personnes résidant en France métropolitaine et dans les départements d'outre-mer sous ces hypothèses, la population en France au 1er janvier 2013 est de 65,8 millions de personnes. Les habitants des collectivités territoriales d'outre-mer (Saint-Barthélemy Saint-Martin et Saint-Pierre-et-Miquelon Nouvelle-Calédonie, Polynésie Française, îles Wallis et Futuna) ne sont pas comptabilisés. De même les Français résidents à l'étranger, possédant parfois une double nationalité, ne sont pas décomptés. Par contre, les personnes de nationalité étrangère présentes sur le territoire appartiennent à la population de la France.

Comment définir les personnes employées dans une entreprise au 1er octobre 2013 ? Faut-il inclure les travailleurs à domicile, les travailleurs à temps partiel, les travailleurs intérimaires (salariés d'une autre entreprise), les stagiaires, les apprentis, les travailleurs « au noir » ? Doit-on comprendre les travailleurs absents pour maladie, en congé annuel ou en détachement ? L'effectif présent diffère en général de l'effectif théorique celui des personnes juridiquement salariées de l'entreprise.

Les règles qui définissent l'ensemble à étudier doivent permettre de dire sans ambiguïté si une unité appartient ou non au domaine. Pour chaque population statistique, il existe des conventions décrivant avec précision les limites de l'ensemble analysé. Il est exceptionnel que la population statistique et la population intuitive soient identiques.

## Les caractères et les modalités

Pour décrire une population, on classe les individus selon certains attributs que l'on appelle des caractères (sexe, genre) ou des variables (âge). Il est indispensable de ne retenir que les caractères les plus pertinents pour pouvoir décrire une population convenablement. Il convient de ne sélectionner qu'un nombre restreint de caractères pour obtenir une description synthétique. Le caractère est un critère de classement, il peut présenter plusieurs situations différentes, il prend plusieurs modalités. Les deux modalités du caractère « sexe » sont : masculin et féminin. Ce caractère qui peut prendre deux modalités est dit dichotomique, une illustration de ce type de caractère : on peut « être au chômage » ou « ne pas être au chômage ». Le nombre de modalités d'un caractère dépend de l'information disponible et du but de l'étude. Par exemple, l'état matrimonial peut comprendre diverses modalités : célibataire, marié, veuf, divorcé, union libre, pacsé ou deux modalités :

Lissage par les moyennes échelonnées.....	179
Lissage par les moyennes mobiles .....	180
La mesure de la saisonnalité.....	188
L'étude graphique.....	188
La dessaisonnalisation par la méthode des rapports à la moyenne mobile .....	190
Les rapports à la droite de tendance.....	192
La détermination de la tendance (T).....	202
Utilisation d'un ajustement .....	202
Ajustement par une droite.....	202
Principe de l'ajustement par une logistique .....	206
La composante cyclique.....	207
La méthode des résidus .....	208
La méthode du cycle moyen.....	208
<b>Chapitre 5</b>	
<b>Les indices</b> .....	211
Les indices simples .....	211
Définition .....	211
Propriétés des indices élémentaires .....	214
Les indices synthétiques.....	220
Les indices de Laspeyres .....	221
Les problèmes de comparaisons : les pondérations implicites .....	223
L'analyse d'un indice de valeur .....	226
Les indices de Paasche .....	228
L'indice de Fisher .....	230
Les indices chaînes .....	231
Les raccords d'indices .....	234
Les indices de prix .....	236
<b>Conclusion</b> .....	243
<b>Table des matières</b> .....	247

