Maurice LETHIELLEUX
Maître de conférences à l'université
Paris II-Panthéon-Assas

Statistique descriptive

8e édition

DUNOD

© Dunod, 2016
5 rue Laromiguière, 75005 Paris
www.dunod.com
ISBN 978-2-10-074564-7

Sommaire

Généralités

Les caractéristiques de tendance centrale ou de position

Les caractéristiques de dispersion

Les indices

Les séries statistiques à deux caractères

Les séries chronologiques

Enquête et simulation

FICHE 1

Utilisations de la statistique

I Objectifs

Dans un sens général, la « statistique » est l'ensemble des méthodes scientifiques à partir desquelles sont recueillies, présentées, résumées et analysées des données.
Dans un sens plus étroit, le terme de « statistique » est employé pour désigner des données ou des résultats obtenus à partir de ces données, on parle ainsi de statistiques démographiques, de statistiques sur les revenus, le chômage, etc.
Ceci correspond à la signification première du mot « state-istique », : ensemble des informations indispensables à l'État (dans la langue latine *Statisticum* signifie : qui a trait à l'État).
Les informations qui sont apparues très tôt indispensables à l'État étaient celles qui permettaient de recueillir des impôts et de recruter des conscrits pour les besoins des guerres. Ce n'est donc pas étonnant si les premiers chiffres recueillis concernaient les populations et l'économie.
Le ministre Colbert fit entreprendre la première grande enquête en 1664, dans la presque totalité des provinces françaises. Un certain nombre de renseignements concernaient des données économiques notamment dans le domaine de la finance : revue des principaux impôts, montant des impositions...
Le marquis de Vauban, s'intéresse de près aux connaissances chiffrées, il met en évidence avec des statistiques précises les problèmes économiques de son époque. Il publie en 1686 « Méthode généralle et facille pour faire le dénombrement des peuples »
Après l'économie et la démographie, la statistique s'est étendue à l'ensemble des sciences et elle est devenue une discipline scientifique faisant largement appel aux mathématiques (calcul algébrique, analyse, algèbre linéaire, calcul des probabilités) et à l'informatique pour les applications pratiques.
La statistique en tant que méthode d'analyse comporte deux niveaux :

- **La statistique descriptive**, objet de ce manuel, qui englobe un ensemble de méthodes, pour décrire avec des outils appropriés des ensembles nombreux et dégager l'essentiel de l'information qui en résulte. Elle utilise des modes de représentations graphiques comme des courbes de fréquences et des histogrammes. Elle utilise également des caractéristiques obtenues par un calcul algébrique :
  – indicateurs de valeur centrale : la moyenne, la médiane, le mode ;
  – indicateurs de dispersion autour d'une valeur centrale : variance, écart-type ;
  – indices qui résument l'évolution d'un ensemble de grandeurs : indices de prix, indices de quantités, indices en valeur, indices boursiers.

- **La statistique théorique ou mathématique** qui prend la suite de la statistique descriptive lorsque l'on peut énoncer ou élaborer des lois : loi binomiale, loi normale…

Le savant belge Adolphe Quételet (1796-1874) a défendu le principe d'une statistique scientifique s'appuyant sur le calcul des probabilités.

II L'essentiel à savoir

A. Prélèvement d'un échantillon et sondages

Un des objets de la statistique est d'étudier des caractères attachés à certains ensembles, qui constituent selon un terme emprunté à la démographie, une **population**. Une population est constituée d'un ensemble d'**individus**. Exemple de population : le parc automobile français. Un individu est une automobile, les **caractères ou variables** étudiées peuvent être la puissance, la consommation de carburant, l'âge, l'émission de CO_2...
Pour cela, on peut recueillir l'information sur chaque individu composant la population, c'est ce que l'on fait pour un **recensement**.
Cette façon exhaustive de procéder n'est pas toujours possible ou souhaitable à mettre en œuvre et on extrait l'information sur un certain nombre d'individus composant la population, c'est ce qu'on appelle prélever un échantillon ou encore faire un **sondage**.

B. Les raisons d'échantillonner

Elles sont très variées, en voici quelques-unes :

– le budget est limité et le coût de la collecte élevé ;
– il faut user ou détruire des éléments d'une fabrication pour en mesurer la qualité (exemple : résistance d'un moteur à l'usure) ;
– le manque de temps ne permet pas de recueillir l'information sur autant d'individus qu'on le souhaite ;
– les résultats sont recueillis avec plus de précision car ils sont plus facilement contrôlables du fait du nombre peu élevé d'observations ;
– le calcul d'une marge d'erreur est possible, voir l'application sur les intervalles de confiance.

C. L'inférence statistique

Les caractéristiques d'une population ne sont souvent connues qu'avec une certaine imprécision lorsque ces caractéristiques sont étudiées sur un échantillon. Les caractéristiques d'un échantillon reflètent en effet avec une certaine marge d'erreur les caractéristiques de la population.
La statistique **inductive ou inférentielle** consiste à induire des résultats sur une population à partir d'un échantillon en précisant si possible la marge d'erreur. Ceci fait appel au calcul des probabilités, donc à la statistique mathématique, qui n'est pas l'objet de ce manuel.
L'**intervalle de confiance** présenté en application dans cette fiche et sans justifications théoriques illustre par un exemple cette notion de **statistique inférentielle**.

D. Biais de mesure et biais de recrutement

Outre les imprécisions dues au fait que l'information est obtenue sur un échantillon, il existe deux autres sources de distorsions importantes sur les résultats obtenus.

• *Biais de mesure*

Les résultats sont mesurés avec des **erreurs**.

> **Exemple**
>
> L'appareil qui prend les mesures est défectueux, la personne qui prend les mesures n'est pas compétente. Lors d'une enquête sur les revenus, les personnes sondées ne déclarent pas la totalité de leur revenu. L'enquêteur n'est pas honnête et il invente les réponses. Tous ces exemples illustrent ce qu'on appelle un biais de mesure.

• *Biais de recrutement*

L'échantillon prélevé n'est **pas représentatif** de la population vis-à-vis du caractère étudié.

> **Exemple**
>
> Un échantillon prélevé pour connaître les intentions de vote contient 40 % de personnes de plus de 60 ans alors que cette proportion n'est que de 25 % dans la population. Si l'âge exerce une influence sur le choix des électeurs, on saisit facilement cette source d'erreur. En 1936, aux États-Unis trois sondages donnèrent Alf Landon vainqueur aux élections présidentielles alors que F. D. Roosevelt fut largement élu. Les échantillons constitués à partir d'annuaires étaient biaisés (les électeurs de Landon y étaient surepresentés).

III Compléments

A. Les méthodes de prélèvement d'un échantillon

Il existe de nombreuses méthodes pour prélever un échantillon, celles-ci sont guidées par des considérations pratiques, de facilité de traitement mathématique ou encore de coût.

• *Échantillon au hasard ou méthode probabiliste*

À chaque tirage d'un individu, chacun des N individus composant la population à la même chance d'être tiré. On dit que la probabilité est de $\frac{1}{N}$.

C'est cette méthode qui permet d'obtenir le plus facilement une mesure de la précision des résultats dans la population et également d'éviter des biais de recrutement (voir application sur les intervalles de confiance).

• Méthode des quotas

L'échantillon représente « en miniature » la population étudiée vis-à-vis des caractéristiques qui influent sur le phénomène analysé, par exemple même quotas selon l'âge, les revenus dans l'échantillon et dans la population. Ces échantillons sont souvent obtenus par téléphone ce qui risque d'introduire des biais.

• Méthode en cascade ou à plusieurs degrés

On tire au hasard un échantillon de quelques villes, puis dans chaque ville un échantillon de quelques quartiers, puis dans chaque quartier un échantillon de quelques individus.

• Méthode par grappes

Les grappes désignent des groupes d'individus qui habitent par exemple dans le même immeuble. Cette méthode de sondage consiste à tirer les grappes et ensuite les informations sont recueillies auprès de chaque individu de la grappe. Cette méthode simple et peu coûteuse donne les meilleurs résultats avec des grappes qui se ressemblent et avec des individus très différents à l'intérieur des grappes.

• Méthode par stratification

Les individus sont tirés au hasard dans des strates définies comme des groupes homogènes de la population. L'effectif de l'échantillon tiré dans une strate ne dépend pas spécifiquement de la taille de la strate comme on le ferait dans la méthode des quotas. L'effectif tiré dans une strate dépend de la variabilité connue ou estimée de la variable étudiée à l'intérieur de la strate. En effet, si tous les individus d'une strate se ressemblent beaucoup, il suffit de tirer un petit nombre d'individus dans la strate pour les connaître avec précision ; à contrario si les individus d'une strate sont très différents il faut en tirer un assez grand nombre pour les connaître avec précision. Cette méthode améliore la précision mais les calculs algébriques sont lourds, en particulier pour déterminer le nombre optimal d'individus à tirer dans chaque strate. L'échantillon final est la réunion des échantillons des différentes strates.

• Méthode à probabilités inégales

Lorsque des unités statistiques comme des villes sont de tailles très inégales, le nombre d'individus tirés dans chaque ville, est proportionnel à son nombre d'habitants.

B. La méthode des panels et des cohortes

• Méthode des panels

Cette méthode consiste à suivre le même échantillon avec une périodicité fixée ou non à l'avance. Par exemple tous les trois mois la consommation du même échantillon de 100 ménages est analysée. Les individus de l'échantillon sont choisis et volontaires, ils sont de ce fait assez disponibles pour répondre correctement aux questions de l'enquêteur. Cette méthode est très utilisée dans les techniques de gestion et de marketing.