OPENBOOK

Statistique et probabilités

en économie-gestion

Christophe **Hurlin**, Valérie **Mignon**

DUNOD

Les contenus complémentaires et les corrigés des exercices sont disponibles en ligne sur www.dunod.com/EAN/9782100780235 ou accesibles en flashant le QR code.

RESSOURCES NUMÉRIQUES

Conseiller éditorial : Lionel Ragot

Création graphique de la maquette intérieure : SG Créations
Création graphique de la couverture : Valérie Goussot et Delphine d'Inguimbert

Illustrations : Judith Chouraqui

Crédits iconographiques : p. 84 : Chee-Onn Leong – Fotolia.com ; p. 254 : Kashisu – Fotolia.com ; p. 290 : lenets_tan – Fotolia.com ; couverture : © August_0802 – www.shutterstock.com

11 rue Paul Bert, 92240 Malakoff
www.dunod.com

ISBN 978-2-10-078023-5

Sommaire

Avant-propos

Qu'est-ce que la statistique ? La statistique est une science recouvrant plusieurs dimensions. On emploie d'ailleurs très fréquemment le pluriel « statistiques » pour désigner cette discipline et témoigner ainsi de sa diversité. La statistique englobe la recherche et la collecte de données, leur traitement et leur analyse, leur interprétation, leur présentation sous la forme de tableaux et graphiques, le calcul d'indicateurs permettant de les caractériser et synthétiser... Ces différents éléments renvoient à ce que l'on a coutume de nommer la statistique descriptive, fondée sur l'observation de données relatives à toutes sortes de phénomènes (économiques, financiers, historiques, géographiques, biologiques, etc.).

Il arrive cependant fréquemment que les données représentatives du phénomène que l'on souhaite étudier ne soient pas parfaitement connues, c'est-à-dire pas toutes parfaitement observables, au sens où elles ne fournissent qu'une information partielle sur l'ensemble du phénomène que l'on analyse. Afin de pouvoir en réaliser une étude statistique, il est alors nécessaire d'inférer des informations à partir des quelques éléments dont on dispose. En d'autres termes, le statisticien devra effectuer des hypothèses concernant les lois de probabilité auxquelles obéit le phénomène à analyser. La statistique fait alors appel à la théorie des probabilités et est qualifiée de statistique mathématique ou encore de statistique inférentielle.

Un bref retour sur l'histoire. Même si le terme de « statistique » est généralement considéré comme datant du XVIII^{e} siècle[1], le recours à cette discipline remonte à un passé bien plus éloigné. On fait en effet souvent référence à la collecte de données en Chine en 2238 av. J.-C. concernant les productions agricoles, ou encore en Égypte en 1700 av. J.-C. en référence au cadastre et au cens. La collecte de données à des fins descriptives est ainsi bien ancienne, mais ce n'est qu'au XVIII^{e} siècle qu'est apparue l'idée d'utiliser les statistiques à des fins prévisionnelles. Ce fut le cas en démographie où les statistiques collectées lors des recensements de la population ont permis l'élaboration de tables de mortalité en Suède et en France.

Du côté des mathématiciens, les recherches sur le calcul des probabilités se sont développées dès le XVII^{e} siècle, au travers notamment des travaux de Fermat et Pascal. Même si Condorcet et Laplace ont proposé quelques exemples d'application de la théorie des probabilités, ce n'est qu'au cours de la deuxième moitié du XIX^{e} siècle, grâce aux travaux de Quételet, que l'apport du calcul des probabilités à la statistique fut réellement mis en évidence, conduisant ainsi aux prémisses de la statistique mathématique. Cette dernière s'est ensuite largement développée à la fin du XIX^{e} siècle et dans la première moitié du XX^{e} siècle.

Par la suite, grâce notamment aux progrès de l'informatique peu avant la deuxième moitié du XX^{e} siècle, de nouvelles méthodes d'analyse ont vu le jour, comme l'analyse multidimensionnelle permettant d'étudier de façon simultanée plusieurs types de données. La deuxième moitié du XX^{e} siècle est aussi la période durant laquelle plusieurs courants de pensée en statistique s'affrontent, notamment autour de la notion de probabilité.

1 On attribue en effet ce terme au professeur allemand Gottfried Achenwall (1719-1772) qui, en 1746, emploie le mot *Statistik* dérivé de *Staatskunde*.

Les domaines d'application de la statistique sont multiples. Initialement employée en démographie, elle est en effet utilisée dans toutes les sciences humaines et sociales comme l'économie, la finance, la gestion, le marketing, l'assurance, l'histoire, la sociologie, la psychologie, etc., mais aussi en médecine, en sciences de la terre et du vivant (biologie, géologie...), météorologie, etc. Cet éventail des domaines illustre ainsi toute la richesse de la statistique dont cet ouvrage vise à rendre compte.

En quoi ce manuel se distingue-t-il des autres ouvrages de statistique ? Tout en présentant de façon rigoureuse tous les développements théoriques nécessaires, cet ouvrage propose un exposé clair et pédagogique des différents concepts en les illustrant par de très nombreux exemples et cas concrets. Le lecteur sera ainsi à même de répondre à de multiples questions qui se posent au quotidien dans les domaines de l'économie, la finance et la gestion.

Chaque chapitre débute par des questions et exemples concrets, permettant de mettre en avant l'intérêt des concepts statistiques qui vont être étudiés. Afin de répondre à ces interrogations et traiter ces cas concrets, les différents outils et méthodes statistiques sont ensuite présentés. L'exposé est ainsi progressif, mêlant de façon harmonieuse définitions littéraire et mathématique. En fin de chapitre figurent des exercices qui permettent au lecteur d'évaluer et tester les connaissances acquises. Les exercices font l'objet de **corrigés très détaillés, disponibles en ligne sur www.dunod.com**, sur la page de l'ouvrage. Le lecteur trouvera également sur cette page Internet des annexes à télécharger reproduisant les principales **tables statistiques**, ainsi que de nombreux **compléments** relatifs à plusieurs chapitres de l'ouvrage.

Diverses rubriques spécifiques à la collection « Openbook » composent les chapitres. Outre les prérequis et les objectifs propres à chaque chapitre, une rubrique « Les grands auteurs » présente de façon synthétique un auteur clé dont les travaux ont profondément marqué le développement de la statistique. La rubrique « Focus » permet quant à elle de faire rapidement le point sur un concept fondamental, alors que la rubrique « Pour aller plus loin » offre la possibilité au lecteur d'approfondir un ou plusieurs points particuliers. La rubrique « En pratique » permet également au lecteur de se familiariser avec l'application concrète d'un concept ou d'une méthode. Enfin, la rubrique « Trois questions à... » illustre l'orientation résolument appliquée de l'ouvrage en donnant la parole à quelques grands acteurs du monde professionnel, nous expliquant la façon dont ils utilisent la statistique au quotidien.

Comment est organisé ce manuel ? Cet ouvrage a pour objectif de fournir au lecteur l'ensemble des connaissances que doit acquérir un étudiant au cours de son cursus de licence en économie-gestion ou de son cycle d'études Bac+3. Il couvre donc les trois années du cycle Bac+3 (licence ou bachelor). Il s'organise ainsi en trois parties, chacune étant relative à une année du cycle Bac+3. La première partie, correspondant au programme de la première année post-bac, traite de la statistique descriptive et comporte quatre chapitres. Le chapitre 1 étudie les distributions à un caractère et présente l'ensemble des concepts de base de la statistique descriptive : tableaux, graphiques et caractéristiques clés comme la moyenne, la variance, la médiane, etc. Le chapitre 2 étend l'analyse au cas de deux variables statistiques et porte ainsi sur les distributions à deux caractères. Le chapitre 3 offre une présentation des indices, très

utilisés en pratique. Le chapitre 4 propose quant à lui une introduction à l'analyse des séries temporelles en dotant le lecteur de l'ensemble des outils nécessaires pour l'analyse de l'évolution d'un phénomène au cours du temps.

La deuxième partie de l'ouvrage, correspondant au programme de la deuxième année du cycle Bac+3, relève du domaine de la statistique mathématique et se compose également de quatre chapitres. La notion fondamentale de probabilité fait l'objet du chapitre 5. Le chapitre 6 traite des variables aléatoires, c'est-à-dire des variables dont les valeurs sont soumises au hasard. L'étude de ces variables nécessite le recours à des lois de probabilité, dont les plus usuelles (lois normale, binomiale, de Student, de Poisson...) sont présentées au cours du chapitre 7. Le chapitre 8 clôt la deuxième partie par l'étude des propriétés de convergence.

La troisième partie de l'ouvrage, correspondant au programme de la dernière année du cycle Bac+3, traite de l'estimation et des tests. Le chapitre 9 est relatif à l'estimation, le chapitre 10 proposant quant à lui une description de l'une des méthodes les plus utilisées connue sous le nom de maximum de vraisemblance. La théorie des tests statistiques fait l'objet du chapitre 11, dernier chapitre du manuel.

Remerciements. Cet ouvrage est le fruit de divers enseignements de statistique dispensés par les auteurs en première, deuxième et troisième années de licence à l'Université d'Orléans et à l'Université Paris Ouest–Nanterre La Défense. Nous adressons nos remerciements à nos étudiants dont les questions et commentaires lors de nos cours ont naturellement contribué à la présentation pédagogique de ce manuel. Nous remercions Lionel Ragot pour la confiance qu'il nous a accordée en nous encourageant à rédiger ce manuel, ainsi que les éditions Dunod. Nous remercions très vivement nos collègues et amis Cécile Couharde, Olivier Darné, Emmanuel Dubois, Gilles Dufrénot, Elena Dumitrescu, Meglena Jeleva et Hélène Raymond pour leur relecture très attentive et pour leurs remarques et suggestions toujours très constructives. Emmanuel Dubois nous a également aidé pour la réalisation de certains graphiques dans la première partie de l'ouvrage, qu'il en soit chaleureusement remercié. Alina Catargiu, Axelle Chauvet-Peyrard, Andreea Danci, Damien Deballon, Laurent Ferrara, Yoann Grondin, Abdou Ndiaye, Ekaterina Sborets et Stéphanie Tring ont très gentiment accepté de répondre à nos questions, nous leur adressons nos plus vifs remerciements pour leurs contributions. Enfin, nous remercions très sincèrement nos familles pour leur soutien sans faille et leur patience lors de la rédaction de cet ouvrage.

À Séverine, Josiane, Emmanuel et Pierre.
À Tania et Emmanuel.

Table des matières

Partie 1

Statistique descriptive

Initialement employée en démographie dans le cadre des recensements de la population, la statistique descriptive est utilisée dans de nombreux domaines et disciplines, comme l'économie, la finance, l'assurance, le marketing, l'histoire, la géographie, la géologie, la biologie, la médecine, la météorologie, le sport, etc. Ce très large éventail de domaines d'application s'explique par le fait que dès lors que l'on dispose de données, c'est-à-dire d'observations, sur le phénomène que l'on souhaite étudier, il est nécessaire de les traiter afin de pouvoir les exploiter pour en extraire un certain nombre d'informations pertinentes. Tel est précisément l'objet de la statistique descriptive, qui permet de résumer et synthétiser l'ensemble des données étudiées au travers de graphiques, tableaux et divers indicateurs dont l'un des plus connus est la moyenne.

Au-delà de l'analyse d'un seul phénomène, la statistique descriptive permet aussi d'analyser et chiffrer la relation entre plusieurs phénomènes, c'est-à-dire plusieurs variables, et de mesurer l'intensité d'une telle liaison.

Chapitre 1

Quel est le salaire annuel moyen des hommes et des femmes en France ? Quelle est la proportion d'hommes et de femmes gagnant plus que ce salaire moyen ? À quel niveau de salaire se situe la plus grande partie de la population ? Les salaires ont-ils beaucoup fluctué ces cinquante dernières années ? Ont-ils suivi une évolution similaire pour les hommes et les femmes ? Les femmes sont-elles victimes d'inégalités salariales ?

La **statistique descriptive** permet de répondre à toutes ces questions. Elle permet en effet de résumer et synthétiser, par le biais de tableaux, graphiques et indicateurs statistiques, l'ensemble des données étudiées.

LES GRANDS AUTEURS

William Playfair (1759-1823)

Ingénieur et économiste écossais, **William Playfair** est considéré comme l'un des pionniers de la représentation graphique des données statistiques. Dans son ouvrage *Commercial and Political Atlas* paru en 1786, il introduit plusieurs représentations graphiques, comme celle retraçant l'évolution temporelle des intérêts de la dette publique britannique au cours du XVIII^{e} siècle ou encore le **diagramme en bâtons** lui permettant de comparer les importations et exportations de l'Écosse en 1781 à celles d'autres pays. Également crédité de l'invention du célèbre **histogramme**, les représentations graphiques proposées par Playfair figurent parmi celles les plus utilisées en statistique descriptive. Quelques années plus tard, son ouvrage *Statistical Breviary* paru en 1801 présente un schéma circulaire, connu aujourd'hui sous le nom de **représentation par secteurs** (ou « camembert »). ■

Distributions à un caractère

Plan

Pré-requis

→ **Connaître** les opérations mathématiques de base.

Objectifs

→ **Synthétiser, résumer et extraire** l'information pertinente contenue dans une série statistique.

→ **Représenter** graphiquement une distribution statistique.

→ **Construire** un tableau statistique.

→ **Définir** les indicateurs statistiques clés.

Le tableau 1.1 donne la valeur du salaire annuel net moyen en euros des hommes et des femmes en France de 1950 à 2010 (source des données : INSEE). La figure 1.1 représente graphiquement ces mêmes données : la courbe orange décrit l'évolution du salaire des hommes sur la période 1950-2010, la courbe grise étant relative à l'évolution du salaire des femmes sur la même période. Sans prendre en compte l'effet de l'inflation, on constate globalement une tendance haussière avec un niveau plus élevé du salaire pour les hommes que pour les femmes.

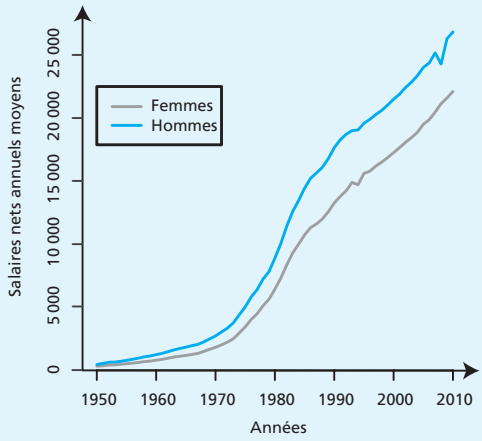
▼ **Tableau 1.1** **Salaire annuel net moyen en euros en France, 1950-2010**

Année	Femmes	Hommes	Année	Femmes	Hommes	Année	Femmes	Hommes
1950	310	444	1970	1 807	2 711	1990	13 258	17 643
1951	344	530	1971	2 002	3 020	1991	13 772	18 266
1952	402	622	1972	2 218	3 330	1992	14 225	18 708
1953	412	637	1973	2 487	3 746	1993	14 894	18 999
1954	462	694	1974	2 946	4 388	1994	14 703	19 054
1955	504	771	1975	3 424	5 009	1995	15 606	19 580
1956	550	854	1976	4 009	5 799	1996	15 782	19 896
1957	600	947	1977	4 465	6 380	1997	16 187	20 278
1958	669	1 051	1978	5 102	7 223	1998	16 506	20 607
1959	711	1 122	1979	5 616	7 804	1999	16 861	21 033
1960	789	1 227	1980	6 418	8 881	2000	17 259	21 498
1961	849	1 327	1981	7 298	10 041	2001	17 651	21 889
1962	941	1 460	1982	8 343	11 411	2002	18 072	22 422
1963	1 037	1 604	1983	9 287	12 587	2003	18 443	22 840
1964	1 099	1 714	1984	9 996	13 464	2004	18 858	23 360
1965	1 168	1 820	1985	10 718	14 430	2005	19 500	24 007
1966	1 240	1 935	1986	11 302	15 212	2006	19 866	24 370
1967	1 316	2 036	1987	11 590	15 639	2007	20 472	25 168
1968	1 479	2 231	1988	11 991	16 093	2008	21 135	24 287
1969	1 648	2 473	1989	12 561	16 776	2009	21 593	26 300
						2010	22 112	26 831

Source : INSEE.

De tels tableaux et graphiques visent ainsi à résumer et rendre lisible l'information contenue dans les données étudiées (ici le salaire). Ils doivent être complétés par le calcul de divers indicateurs statistiques qui nous permettront notamment de déterminer le niveau moyen du salaire sur la période considérée, le niveau du salaire tel que le nombre d'individus (hommes et femmes) percevant moins que ce niveau est identique au nombre d'individus gagnant plus, le niveau du salaire perçu par le plus grand nombre des individus étudiés, ou encore la dispersion, c'est-à-dire la variabilité, du salaire entre hommes et femmes et/ou au cours de la période d'étude. À cette fin, on

calcule des indicateurs dits de tendance centrale, de forme et de dispersion. Le recours aux indicateurs de concentration nous permet en outre de compléter l'analyse afin de quantifier précisément les inégalités de salaires entre hommes et femmes.



▲ **Figure 1.1 Évolution du salaire annuel net moyen en euros des hommes et des femmes en France, de 1950 à 2010**

1 Définitions et concepts fondamentaux de la statistique descriptive

L'objectif de la statistique descriptive est de résumer et synthétiser l'information contenue dans les données étudiées afin d'en déduire un certain nombre de propriétés. À cette fin, on utilise des tableaux et des graphiques (▶ section 1.2) et l'on calcule divers indicateurs ou caractéristiques (▶ section 2).

1.1 Définitions

1.1.1 Population, individus, échantillon

Une **population** est un ensemble, fini ou non, d'éléments que l'on souhaite étudier. Ces éléments portent le nom d'**individus** ou d'**unités statistiques**. Il peut s'agir par exemple d'êtres humains (adultes, enfants, chômeurs, salariés, etc.), d'animaux ou encore d'objets (entreprises, voitures, ordinateurs, incendies, accidents, etc.). Très souvent, la population que l'on souhaite analyser est très grande et il est usuel de se restreindre à l'étude d'un échantillon.

Un **échantillon** est ainsi un sous-ensemble de la population considérée qui doit posséder les mêmes caractéristiques statistiques que la population dont il est issu. À partir d'un échantillon dit **représentatif**, il est alors possible d'effectuer des analyses et d'en déduire des conclusions valables pour la population.

1.1.2 Caractères, modalités et variables statistiques

Caractères et modalités. Afin d'étudier les individus composant une population, on les classe en un certain nombre de sous-ensembles, appelés **caractères** ou **variables statistiques**. À titre d'exemple, si l'on étudie le personnel salarié d'une entreprise, on pourra retenir comme caractères le sexe, l'âge, la profession, le salaire, l'ancienneté dans l'entreprise, etc. Pour une voiture, on retiendra la puissance du moteur, le nombre de places assises, la couleur, le modèle... Les valeurs possibles prises par le caractère ou la variable sont appelées **modalités**. La variable « sexe » a ainsi deux modalités, masculin et féminin, mais les caractères peuvent avoir un très grand nombre de modalités. Notons que les modalités doivent être incompatibles – un individu ne peut pas appartenir simultanément à plusieurs modalités – et exhaustives – toutes les situations possibles doivent être recensées.

Une variable peut être **qualitative** ou **quantitative**. Dans le premier cas, les modalités ne sont pas des valeurs chiffrées, elles ne sont pas mesurables mais uniquement observables (sexe, nationalité, catégorie socio-professionnelle, etc.). Dans le cas d'une variable quantitative, les modalités sont mesurables : à chaque modalité est associé un nombre, c'est-à-dire une valeur chiffrée, représentant la mesure du caractère. Ainsi, la puissance d'un moteur, le nombre de places assises, l'âge, la taille, etc. sont des variables statistiques dont les modalités sont des nombres.

Variables statistiques qualitatives nominales et ordinales. Les variables qualitatives peuvent être **nominales** ou **ordinales**. Dans le premier cas, les modalités ne peuvent être ordonnées, contrairement au cas de variables ordinales. Des exemples usuels de variables nominales sont le sexe (modalités : masculin, féminin), l'état civil (modalités : célibataire, marié ou pacsé, veuf, divorcé), la couleur des yeux ou encore le groupe sanguin. Des variables comme le niveau d'études (avec, par exemple, comme modalités : sans diplôme, primaire, secondaire, universitaire) ou le niveau de satisfaction (peu satisfait, satisfait, très satisfait) sont des variables ordinales. Notons toutefois que le fait de pouvoir ordonner ou non les modalités d'une variable peut être sujet à débats. Prenons l'exemple de la variable « catégorie socio-professionnelle ». Si l'on a coutume d'ordonner comme suit les trois modalités « ouvriers », « employés », « cadres », il devient plus difficile d'ordonner les modalités « enseignant », « chercheur » et « responsable administratif » (en particulier si ces trois modalités correspondent au même niveau de diplôme et/ou de responsabilités).

Variables statistiques quantitatives discrètes et continues et regroupement en classes. Les variables quantitatives peuvent être discrètes ou continues. Une variable est dite **discrète** lorsque ses valeurs sont des nombres isolés dans son intervalle de variation. Il s'agit en règle générale de nombres entiers ; par exemple le nombre d'enfants par famille, le nombre de salariés d'une entreprise, le nombre d'automobiles vendues. Une variable est dite **continue** lorsqu'elle peut prendre toutes

les valeurs au sein de son intervalle de variation. On peut donner comme exemples la taille, le poids, la température, etc. Le nombre de valeurs possibles à l'intérieur de l'intervalle de variation étant infini, on les groupe par **classes**. Si l'on considère la variable de salaire annuel, on peut par exemple définir les classes suivantes : moins de 10 000 euros, de 10 000 à moins de 15 000 euros, de 15 000 à moins de 20 000 euros, de 20 000 à moins de 25 000 euros, de 25 000 à moins de 40 000 euros, plus de 40 000 euros. La longueur (ou l'étendue) de la classe, c'est-à-dire la différence entre l'extrémité supérieure et l'extrémité inférieure de la classe, est appelée **amplitude de la classe**. Elle peut être variable, comme dans l'exemple précédent, ou constante. Dans la mesure où il existe une infinité de valeurs au sein d'une classe, il est possible de calculer le **centre de classe** défini comme suit :

$$\text{Centre de classe} = \frac{\text{Extrémité inférieure} + \text{Extrémité supérieure}}{2} \tag{1.1}$$

EN PRATIQUE

La distinction variables discrètes/variables continues

Du fait de la précision limitée des mesures, il peut être difficile de distinguer entre variables discrètes et continues. On retient en conséquence fréquemment le groupement ou non en classes comme moyen de distinction : une variable continue est ainsi souvent telle que le nombre de ses valeurs est si important qu'il convient de les regrouper en classes afin de pouvoir l'étudier.

S'agissant des classes, mentionnons (i) que le nombre d'individus par classe doit être suffisamment important de sorte à limiter ou éliminer les variations accidentelles qui peuvent se produire si l'on retient un effectif trop faible et (ii) que les amplitudes ne doivent pas être trop importantes afin de conserver certaines particularités de la variable étudiée.

1.1.3 Fréquences et effectifs

Considérons une population comprenant N individus. Ce nombre est appelé **effectif total** de la population. On regroupe les N individus suivant les k modalités, notées $x_i, i = 1,\ldots,k$, de la variable x. À chaque modalité correspond un nombre d'individus $n_i, i = 1,\ldots,k$, appelé **effectif** (ou fréquence absolue)[1] de la modalité x_i. Dans le cas d'une variable quantitative ou qualitative ordinale, la somme des effectifs n_i pour $i = 1,\ldots,k$ est ainsi égale à l'effectif total de la population :

$$N = \sum_{i=1}^{k} n_i \tag{1.2}$$

La **fréquence** (ou fréquence relative) associée à une modalité x_i est définie comme le rapport :

$$f_i = \frac{n_i}{N} \tag{1.3}$$

1 Dans le cas d'une variable qualitative nominale, l'effectif n_i correspond au nombre de fois où la modalité x_i apparaît.

La fréquence donne la proportion d'individus de la population présentant la modalité x_i et est en général exprimée en pourcentage. En utilisant l'équation (1.2), on déduit immédiatement la propriété suivante :

$$\sum_{i=1}^{k} f_i = 1 = 100\ \% \tag{1.4}$$

La somme des fréquences f_i correspondant aux différentes modalités, notée F_i, est appelée **fréquence cumulée** :

$$F_1 = f_1 \tag{1.5}$$

$$F_2 = f_1 + f_2 \tag{1.6}$$

$$\ldots$$

$$F_i = f_1 + f_2 + \ldots + f_j + \ldots + f_i \tag{1.7}$$

soit :

$$F_i = \sum_{j=1}^{i} f_j \tag{1.8}$$

La fréquence cumulée F_i indique la proportion des individus pour lesquels la variable étudiée est strictement inférieure à x_{i+1}.

On définit de la même façon les effectifs cumulés :

$$N_i = \sum_{j=1}^{i} n_j \tag{1.9}$$

1.2 Tableaux statistiques et représentations graphiques

Les individus classés suivant les caractères et modalités forment une **distribution** (ou une **série**) statistique qui peut être synthétisée sous la forme de tableaux statistiques et de graphiques : une série représente ainsi la suite des valeurs prises par la variable étudiée. Ces tableaux sont à une dimension si l'on ne considère qu'un seul caractère et à deux dimensions si l'on retient deux caractères (▶ chapitre 2).

FOCUS

Variable statistique et variable aléatoire

Ainsi que nous l'avons vu, une variable est une entité pouvant prendre toutes les valeurs possibles au sein d'un ensemble de définition donné. Lorsque les valeurs prises par la variable sont soumises au hasard (par exemple, « pile » ou « face » dans le cas du lancer d'une pièce), on parle de **variable aléatoire** (▶ chapitre 6). Il convient de ne pas les confondre avec les **variables statistiques**, objet d'étude de ce premier chapitre. La distribution d'une variable statistique est une distribution *empirique*. Les différentes caractéristiques qui seront présentées dans ce chapitre se réfèrent à cette distribution empirique : fonction de répartition *empirique*, moyenne *empirique*, variance *empirique*, moments *empiriques*, etc. Dans la suite du chapitre, afin d'alléger la présentation nous omettrons généralement le terme « empirique », mais il convient de bien garder cette notion à l'esprit.