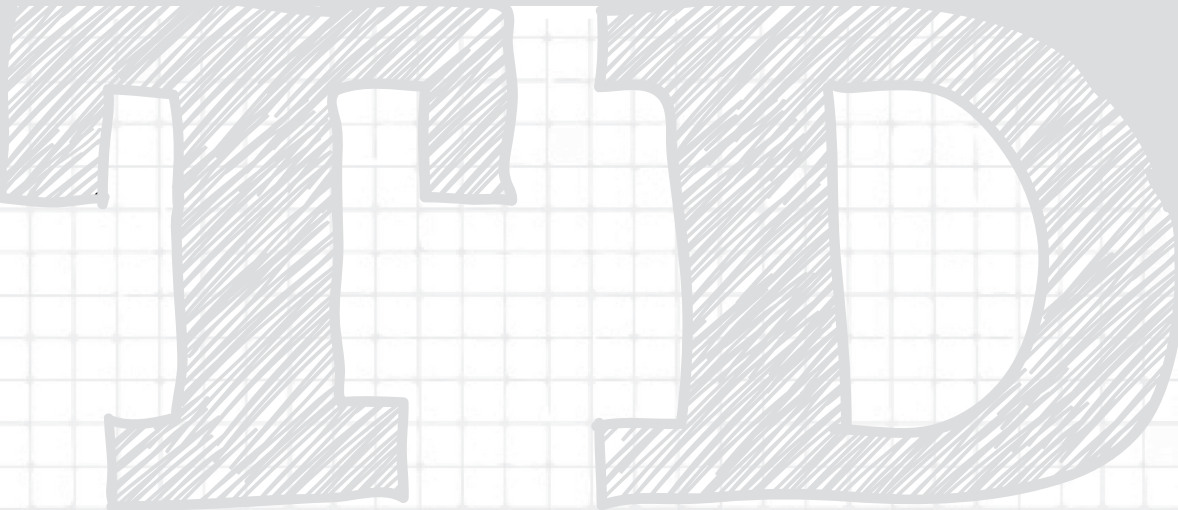




JEAN-LOUIS MONINO

STATISTIQUE DESCRIPTIVE

-  **QCM et exercices corrigés**
-  **4 sujets d'examen corrigés**
-  **Avec rappels de cours**

5^e ÉDITION

DUNOD

Ressources numériques

En complément de cet ouvrage, vous trouverez 2 sujets d'examen corrigés en ligne.

<http://www.dunod.com/contenus-complementaires/9782100758777>

Mise en page : Belle page

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

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

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

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

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



© Dunod, 2017

11, rue Paul Bert, 92240 Malakoff

www.dunod.com

ISBN 978-2-10-076387-0

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

TD Sommaire

Avant-propos	V
TD ① Vocabulaire de base	1
L'essentiel du cours	1
QCM	14
Réflexion	14
Entraînement	15
Solutions	19
TD ● Application Excel 2013. Les tableaux dynamiques	39
TD ② Résumés numériques des variables quantitatives	47
L'essentiel du cours	47
QCM	57
Réflexion	57
Entraînement	58
Solutions	61
TD ③ Indices. Concentration	85
L'essentiel du cours	85
QCM	93
Réflexion	94
Entraînement	94
Solutions	97
TD ④ Construction de tableaux	117
L'essentiel du cours	117
QCM	123
Réflexion	124
Entraînement	124
Solutions	127
TD ⑤ Résumés numériques des variables bidimensionnelles	139
L'essentiel du cours	139

QCM	145
Réflexion	146
Entraînement	146
Solutions	148
TD 6 Moindres carrés ordinaires	169
L'essentiel du cours	169
QCM	178
Réflexion	179
Entraînement	179
Solutions	181
TD 7 Qualité des ajustements	191
L'essentiel du cours	191
QCM	197
Réflexion	198
Entraînement	198
Solutions	201
TD Application Excel 2013. Les moindres carrés	217
TD 8 Séries chronologiques	229
L'essentiel du cours	229
QCM	236
Réflexion	237
Entraînement	237
Solutions	240
TD 9 Correction des variations saisonnières	255
L'essentiel du cours	255
QCM	260
Réflexion	261
Entraînement	262
Solutions	264
TD Application Excel 2013. Les séries chronologiques	281
TD Sujets d'examen	293
Énoncés	293
Corrigés	304
Annexe	335
Index	341

Avant-propos

Le monde est devenu numérique et les avancées technologiques ont démultiplié les circuits d'accès aux données, à leur traitement et à leur diffusion. Les données sont accessibles à tous et partout sur la planète. Le nombre d'internautes en 2014 était de 2,9 milliards soit 41 % de la population mondiale. Le besoin de connaissance se fait sentir afin d'appréhender cette multitude de données. Il faut renseigner, informer et former en masse. L'essor des technologies connexes, comme l'avènement de l'internet, des réseaux sociaux, des « cloud-computing » (usines numériques), a accru les volumes disponibles de données (Data). Actuellement, chaque individu crée, consomme, utilise de l'information numérique : plus de 3,4 millions d'e-mails envoyés dans le monde chaque seconde, soit 107 000 milliards annuels avec 14 600 mails par an et par personne, mais plus de 70 % sont des spams. Des milliards de contenus sont partagés sur les réseaux sociaux comme par exemple sur Facebook, plus de 2,46 millions chaque minute. La statistique est omniprésente, tous les jours, à travers les médias notamment, chacun d'entre nous se trouve exposé à une foule d'informations issues d'études quantitatives : sondages d'opinion, « baromètres » de popularité des hommes politiques, enquêtes de satisfaction, indicateurs de la comptabilité nationale,...

Les tableurs permettent d'effectuer des calculs parfois complexes en des temps très faibles ; aussi leur parfaite maîtrise nous paraît intimement liée aux objectifs de l'enseignement de la statistique descriptive. Les méthodes statistiques, notamment les plus sommaires telles que celles liées à la représentation graphique des données numériques, sont souvent choisies sans véritable connaissance de leur utilité et des limites de leur utilisation. Le développement de l'informatique a aussi pour conséquence de faire évoluer l'enseignement de la statistique. Il est à l'heure actuelle inconcevable de n'illustrer un cours de statistique qu'à l'aide de calculs effectués « à la main ». Pourtant, la plupart

des ouvrages dédiés aux étudiants sont conçus pour s'appliquer à un enseignement magistral de la statistique descriptive.

Tous ces constats nous ont amenés à proposer un ouvrage « charnière », s'inscrivant dans une approche nouvelle de l'enseignement de la statistique descriptive et destiné à un très large public ; un ouvrage permettant d'utiliser les possibilités offertes par les tableurs ou logiciels de statistique.

Compte tenu de l'expérience de l'auteur, enseignants dans les disciplines de sciences économiques et de gestion, *Statistique descriptive* – TD est particulièrement adapté aux étudiants des premiers cycles en licence d'économiques, gestion. Il veut cependant répondre aussi au besoin de tous les publics désireux d'assimiler et d'appliquer à des cas pratiques ou concrets les concepts principaux de la statistique descriptive : étudiants en médecine, pharmacie, psychologie, langues étrangères appliquées, mathématiques appliquées aux sciences sociales, mais aussi étudiants de cycles courts : BTS, DUT, etc. et de classes préparatoires. Les stagiaires de la formation continue y trouveront par ailleurs un outil adapté à leur souci de compréhension rapide et de pragmatisme.

Le lecteur intéressé par des prolongements de cet ouvrage est vivement invité à se connecter sur le site Internet http://www.portices.fr/Cours_Statistique où il trouvera d'autres éléments de cours, des exercices et des examens corrigés.

L'auteur remercie les étudiants qui ont testé les exercices de cet ouvrage durant l'année universitaire 2015-2017 : étudiants de l'UFR d'Économie, de l'École Sécurité Environnement Qualité de l'UFR de Droit et Sciences Politiques de l'Université de Montpellier. Je remercie plus particulièrement les étudiants des promotions de la « prépa » ENS Cachan du lycée Jean Mermoz de l'Académie de Montpellier.

Jean-Louis MONINO, Docteur en Sciences économiques I, Maître de Conférences à l'Université Montpellier I. Économètre, Responsable du C2i au sein de la formation continue de l'Université de Montpellier responsable du « Centre de Certification Numérique International ». Il est membre fondateur du réseau « Cercle K2 » et du Réseau de Recherche sur l'Innovation (RRI), Responsable du Diplôme Universitaire – Traitement de l'Information et Intelligence Économique –. Il enseigne la statistique, les méthodes probabilistes, l'analyse des données, les méthodes de sondage, le data Mining,... dans plusieurs universités françaises et à étrangères (Maroc « École des Sciences de l'Information », Burkina Faso « Fondation du « 2iE », Sénégal « Université Virtuelle du Sénégal »,...).

TD¹ Vocabulaire de base



Le terme statistique a un double sens en français :

- il désigne la totalité des données numériques (par exemple, la consommation des ménages français) ou non numériques (la situation matrimoniale) d'un ensemble ;
- il désigne également un ensemble cohérent de méthodes scientifiques qui permettent de résumer l'ensemble des données numériques ou non numériques (par exemple, le calcul de la consommation moyenne des ménages français).

1 ● Vocabulaire

1.1 ● La statistique

La statistique est un ensemble de méthodes scientifiques visant le recueil, l'organisation, la présentation de données, ainsi que la modélisation et la construction de résumés numériques. On parle de statistique descriptive lorsqu'on décrit et analyse un ensemble sans tirer de conclusion, de statistique inductive lorsqu'on tire des conclusions sur une partie d'un ensemble et que l'on tente d'étendre ces conclusions sur tout l'ensemble.

1.2 ● Population. Échantillon. Individu

Nous réalisons une enquête sur les consommateurs de la région de Nantes. Nous considérons que l'ensemble des consommateurs de la région nantaise est la population concernée par l'enquête.

Un ensemble d'éléments, d'objets ou d'individus, regroupés dans une catégorie bien définie est appelé population ou univers.

Mais prendre l'ensemble des consommateurs (population) pour étudier leur comportement vis-à-vis de leur consommation serait trop onéreux. Ainsi, lorsque la population devient trop grande pour être étudiée et analysée dans sa totalité, on considère une partie de cette population (échantillon). Bien sûr, cette partie doit contenir les propriétés fondamentales de la population (ou univers).

Un échantillon est un sous-ensemble de l'ensemble population. Il doit posséder les propriétés fondamentales de l'ensemble dont il est issu.

Le consommateur de la région de Nantes constitue un élément de la population et à ce titre, s'il est désigné pour répondre à un questionnaire, il peut être défini comme l'unité ou l'élément ou le consommateur de référence. Nous analyserons alors son comportement et sa façon d'exercer sa consommation dans la région.

Un individu, ou unité statistique, est un élément qui appartient à une population ou à un échantillon. Un individu est un élément constitutif d'une population ou d'un échantillon qui permet de donner des informations sur les données statistiques recherchées.

Pour recueillir les données statistiques (informations numériques ou non numériques) sur une population, on utilise soit les enquêtes exhaustives, soit les enquêtes partielles.

Une enquête est dite exhaustive lorsqu'on s'intéresse à l'ensemble des individus qui composent la population étudiée.

Une enquête est dite partielle lorsqu'on s'intéresse à une partie représentative des individus qui composent la population étudiée.

Une enquête exhaustive est appelée recensement. Toute enquête qui n'est pas exhaustive est partielle. Une enquête partielle est appelée enquête par sondage ou plus simplement sondage ; dans ce cas la représentativité joue un rôle déterminant dans la constitution de l'échantillon.

1.3 ● Caractères. Modalités

Nous avons réalisé une enquête sur 842 commerces de la région Languedoc-Roussillon en collaboration avec la CCI Languedoc-Roussillon¹. Les commerces ont répondu à la question « Dans quelles proportions a évolué votre chiffre d'affaires pendant les soldes d'été 2016 par rapport aux soldes d'été

1 <http://www.languedoc-roussillon.cci.fr/ecommunautes/a/e-develop/32105>

2015 ? ». Nous étudions la répartition de cette évolution selon les trois catégories d'achats¹. Nous obtenons le tableau ci-dessous :

Tableau 1.1 : Évolution de votre chiffre d'affaires pendant les soldes d'été 2016

Évolution du chiffre d'affaires	Divers	Équipement de la personne	Équipement du foyer	Total général
Évolution négative de > -20 %	11	41	5	57
Évolution négative de -10 à -20 %	20	61	11	92
Évolution négative de -5 à -10 %	15	74	6	95
Évolution négative de 0 à -5 %	34	123	14	171
Stabilité, pas d'évolution sensible	37	111	8	156
Évolution positive de 0 à 5 %	31	96	10	137
Évolution positive de 5 à 10 %	14	68	3	85
Évolution positive de 10 à 20 %	6	28	2	36
Évolution positive >20 %	0	9	4	13
Total général	168	611	63	842

Ainsi, « Évolution du chiffre d'affaires » est une des multiples questions que nous avons posées aux différents commerces retenus pour l'étude. L'« Évolution du chiffre d'affaires » correspond à un caractère.

Parmi les 842 commerces interrogés, 123 ont répondu avoir eu évolution négative comprise entre 0 à -5 % pour la catégorie « Équipement de la personne ».

Un caractère est une des multiples facettes selon laquelle on peut étudier un individu d'une population ou d'un échantillon. Un individu extrait d'une population ou d'un échantillon peut être étudié selon plusieurs caractères.

Les modalités sont les différentes situations possibles d'un caractère. Une modalité est donc une des réponses possibles à une question.

¹ L'équipement de la personne regroupe, dans la nomenclature commerciale, l'ensemble des produits permettant de fournir à une personne une parure. Il comprend les vêtements, chaussures, chapeaux, montres, etc. On l'oppose généralement à l'équipement de la maison, qui comporte des produits qui ne sont pas destinés à être sortis du foyer par l'utilisateur, notamment l'ameublement, l'électroménager, les arts de la table ou le linge de maison. (Source : Wikipédia)

1.4 ● Tableaux individus-caractères

Une base de données est un document sur lequel on inscrit l'ensemble des informations utiles à l'étude d'un phénomène. Dans notre exemple sur les consommateurs nantais, nous pouvons dresser une table ou base de données qui comporte 9 601 lignes et autant de colonnes qu'il existe de caractères utiles à l'enquête.

**Tableau 1.2 : Extrait d'une base de données
ou tableau individus-caractères**

Numéro	Sexe	Âge	Type d'occupation	Nombre d'actifs	Nombre de voitures
1	Homme	25	Locataire	3	0
2	Homme	35	Propriétaire	4	2
3	Femme	26	Locataire	3	1
4	Homme	38	Propriétaire	1	2
5	Femme	34	Locataire	3	2
6	Homme	33	Locataire	1	3
7	Homme	41	Propriétaire	3	1
8	Homme	64	Locataire	3	3
9	Homme	52	Propriétaire	1	2
10	Femme	39	Propriétaire	3	2
11	Homme	34	Propriétaire	3	2
12	Femme	27	Locataire	4	1
13	Femme	31	Propriétaire	4	1
14	Homme	38	Propriétaire	2	1
15	Femme	64	Propriétaire	2	1
16	Homme	66	Locataire	2	1
17	Femme	71	Propriétaire	2	1
18	Homme	70	Propriétaire	2	1
19	Femme	76	Locataire	1	0
20	Femme	25	Propriétaire	3	0

Un tableau individus-caractères est un recueil de données statistiques. C'est un document où l'on inscrit pour chaque individu d'une population ou d'un

échantillon les modalités de réponses à un ensemble de caractères retenus pour l'étude.

1.4 ● Caractères qualitatifs et quantitatifs

Dans le tableau 1.2, nous avons le caractère « nombre d'actifs » dans le ménage, qui correspond à un caractère quantitatif et le caractère « type d'occupation » qui correspond à un caractère qualitatif, être propriétaire ou locataire.

Lorsque les modalités d'un caractère ne sont pas mesurables, alors le caractère est dit qualitatif.

Lorsque les modalités d'un caractère sont mesurables, alors le caractère est dit quantitatif.

2 ● Opérateurs somme et produit

Afin de rendre plus compactes les écritures des résumés numériques, nous utilisons des opérateurs mathématiques. L'opérateur *sigma* sert à contracter l'écriture d'une somme de nombres. Il se note $\sum$ et prend la forme :

$$x_1 + \dots + x_i + \dots + x_n = \sum_{i=1}^{i=n} x_i$$

Il faut lire de bas en haut : « somme de i égal 1, à i égal n , de x_i ».

Afin de rendre plus compactes les écritures des produits, nous utilisons également un opérateur mathématique. L'opérateur *pi* sert à contracter l'écriture d'un produit de nombres. Il se note $\prod$ et prend la forme :

$$x_1 \cdot \dots \cdot x_i \cdot x_n = \prod_{i=1}^{i=n} x_i$$

Il faut lire de bas en haut : « produit de i égal 1, à i égal n , de x_i ».

Remarque 1 : se lit (de bas en haut) : « somme de i égale 1, à i égale n , de x indice i »

Remarque 2 : les variables et les indices sont dits muets :

$$\sum_{j=1}^{j=n} x_j = \sum_{i=1}^{i=n} x_i = \sum_{k=1}^{k=n} y_k$$

Propriété :

$$\sum_{k=1}^{k=n} (x_k + y_k)^2 = \sum_{k=1}^{k=n} (x_k^2 + y_k^2 + 2x_k y_k) = \sum_{k=1}^{k=n} x_k^2 + \sum_{k=1}^{k=n} y_k^2 + 2 \sum_{k=1}^{k=n} x_k y_k$$

3 ● Variables statistiques quantitatives et qualitatives

Dans le tableau 1.2, nous avons par exemple, le caractère « nombre d'actifs » qui définit la variable statistique X et correspond à un caractère quantitatif ; le caractère « type d'occupation » définit la variable statistique Y et correspond à un caractère qualitatif.

Les individus d'une population sont observés sous l'angle d'un caractère. Une variable statistique permet d'associer à chaque individu de la population une modalité de ce caractère. Une variable peut être de nature quantitative, on parle de variable quantitative ; ou de nature qualitative, on parle de variable qualitative.

Une variable statistique est dite quantitative si ses modalités sont quantifiables. Elle est dite qualitative si ses modalités sont non quantifiables. Les variables quantitatives peuvent être discrètes ou continues : lorsque les modalités numériques d'une variable prennent des valeurs dans l'ensemble $\mathbb{N}$ des entiers naturels (respectivement des valeurs dans l'ensemble $\mathbb{R}$ des réels) la variable est dite discrète (respectivement la variable est dite continue). Les variables qualitatives peuvent être ordinales ou nominales : lorsque les modalités d'une variable possèdent une relation d'ordre (respectivement ne possèdent pas une relation d'ordre) la variable est dite ordinale (respectivement la variable est dite nominale).

Le regroupement en classes est effectué sur des variables statistiques quantitatives chaque fois que le nombre de valeurs numériques prises par la variable est trop élevé ; cela permet un calcul aisé et rapide. On dit que la variable statistique est groupée en classes.

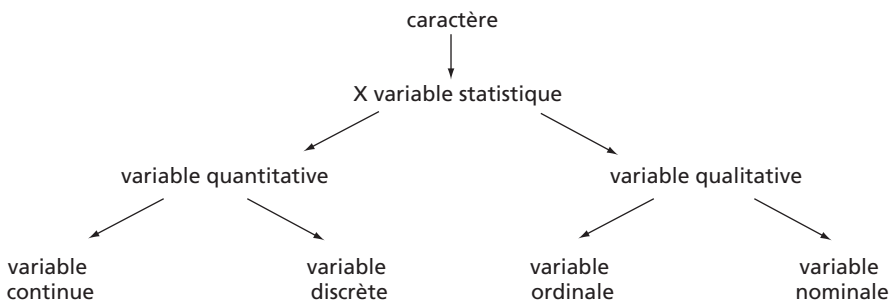


Figure 1.1 : **Caractère et variable statistique**

4 ● Séries statistiques. Distributions statistiques. Effectifs. Effectif total

Dans le tableau 1.2, le caractère « nombre d'actifs » définit la variable statistique X. La liste des 20 individus (taille de l'échantillon) qui ont répondu à ce caractère donne la série statistique du caractère. Dans le tableau 1.1, le caractère « âge » permet de définir une variable statistique Y ordonnée et groupée en classes, on parle alors de distribution statistique de la variable Y dont le caractère est « âge ».

L'effectif n_i d'une modalité i correspond au nombre d'individus qui possèdent cette modalité. L'effectif d'une modalité est également appelé fréquence absolue. L'effectif total N est égal à la somme des effectifs n_i .

$$N = n_1 + n_2 + \dots + n_i + \dots + n_r = \sum_{i=1}^{i=r} n_i$$

Ainsi, le couple (x_i, n_i) représente la distribution de la variable X où n_i est l'effectif et x_i la modalité du caractère (ou réponse) associé à la variable statistique X. L'indice r représente le nombre de réponses (modalités) que l'on peut faire sur le caractère étudié.

La liste des N observations faites pour un caractère d'une population P de taille N est appelée série statistique. Une série statistique ordonnée porte le nom de distribution statistique. Une distribution statistique est l'ensemble des couples (x_i, n_i) .

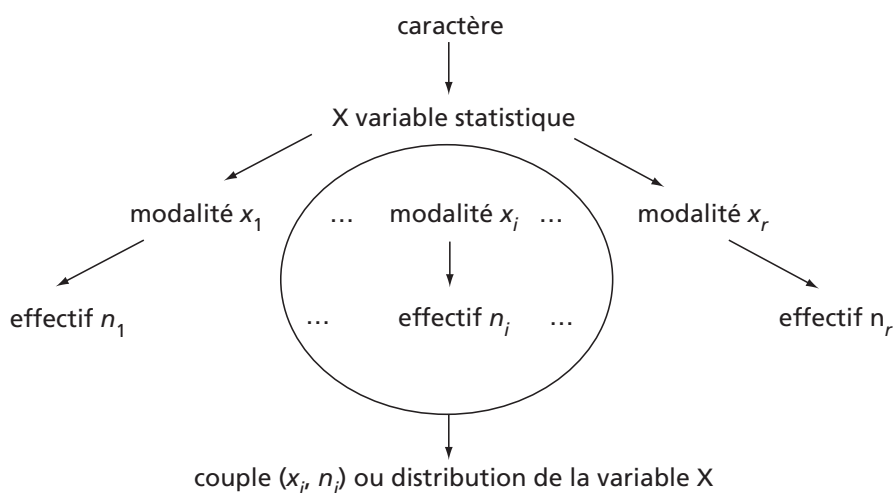


Figure 1.2 : Variable. Modalités. Effectifs. Distribution statistique

5 ● Séries chronologiques

Une variable statistique quantitative continue ou discrète X est dite temporelle lorsque les modalités x_i font référence au temps t_i . Les prises de mesure x_i sur la variable se font à intervalles réguliers et bien sûr une seule fois au cours du temps t_i .

6 ● Effectifs cumulés. Fréquences. Fréquences cumulées

6.1 ● Effectifs cumulés

La distribution en effectifs cumulés (x_i, N_i) , présente la somme des effectifs ayant au plus une modalité donnée x_i :

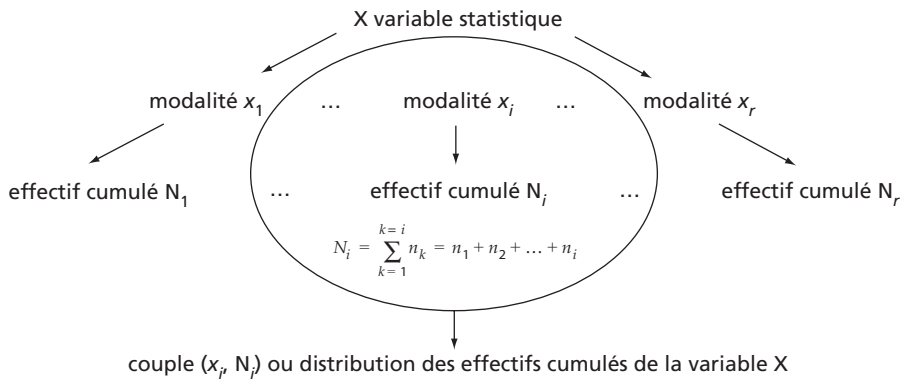
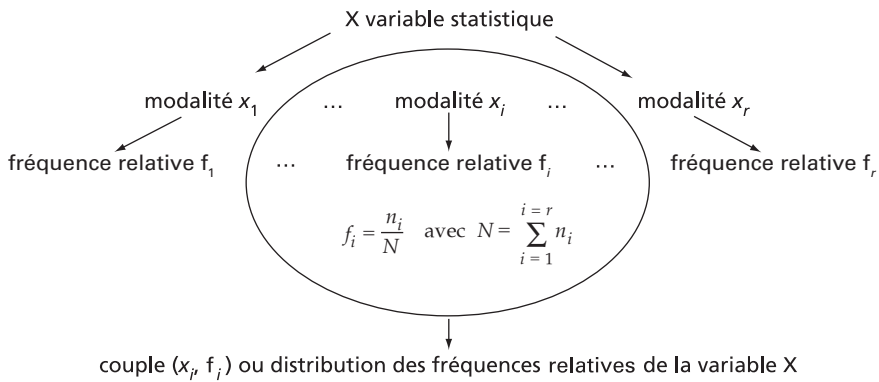


Figure 1.3 : **Distribution en effectifs cumulés**

6.2 ● Fréquences

La fréquence f_i associée à une modalité i est égale au rapport de l'effectif n_i par l'effectif total N : $f_i = \frac{n_i}{N}$

La fréquence est également appelée fréquence relative. Si l'on multiplie par 100 les fréquences, nous obtenons des fréquences en pourcentage notées $f_i \%$.

Figure 1.4 : **Distribution en fréquences**

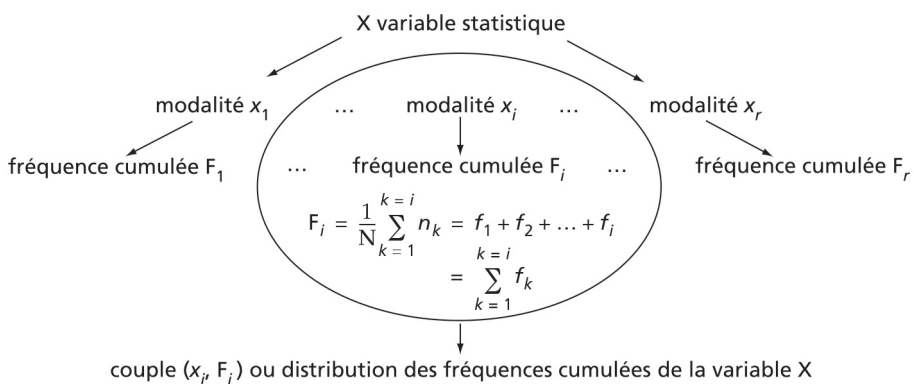
La somme des fréquences est égale à 1 :

$$\sum_{i=1}^{i=r} f_i = \sum_{i=1}^{i=r} \frac{n_i}{N} = \frac{1}{N} \sum_{i=1}^{i=r} n_i = \frac{1}{N} N = 1$$

6.3 ● Fréquences cumulées

Les distributions en fréquences cumulées (x_i, F_i) présentent la somme des fréquences ayant au plus une modalité donnée x_i .

Si l'on multiplie par 100 les fréquences cumulées, nous obtenons des fréquences cumulées en pourcentage notées $F_i \%$.

Figure 1.5 : **Distribution en fréquences cumulées**

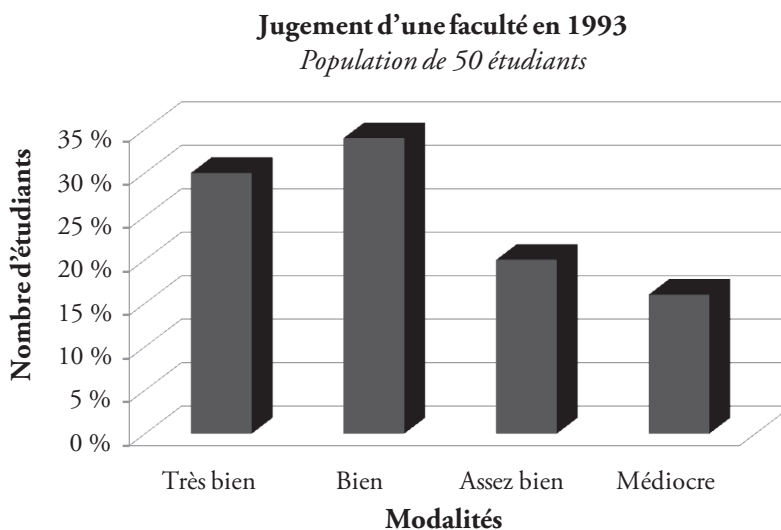
7 ● Principaux graphiques des variables statistiques

Pour résumer l'information contenue dans une distribution statistique (quantitative ou qualitative), on utilise en premier lieu des résumés graphiques. Chaque type de variable possède une ou plusieurs représentations graphiques.

7.1 ● Représentations graphiques des variables qualitatives

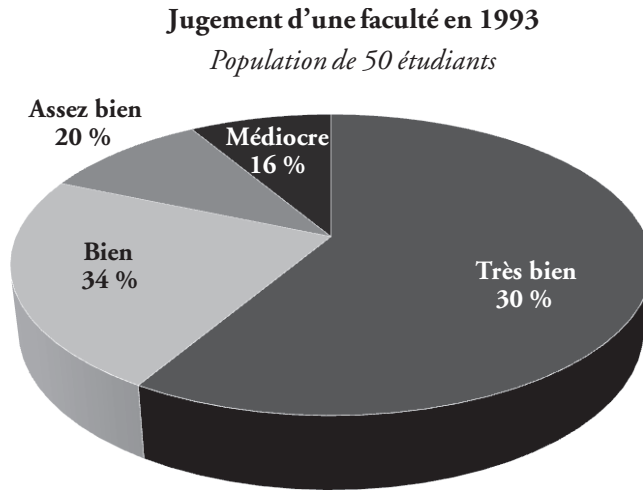
► Le diagramme en « tuyaux d'orgue »

Dans ce type de représentation, les rectangles (tuyaux) ont pour base une modalité et comme hauteur l'effectif (ou la fréquence). La base de chacun des rectangles (base sur l'axe des abscisses) ne possède aucune signification numérique puisque la variable est qualitative.



► Le diagramme en secteurs

Dans ce type de représentation, nous utilisons un disque (plus communément appelé camembert). Chacune des modalités est représentée par un secteur qui est proportionnel à l'effectif (ou à la fréquence).

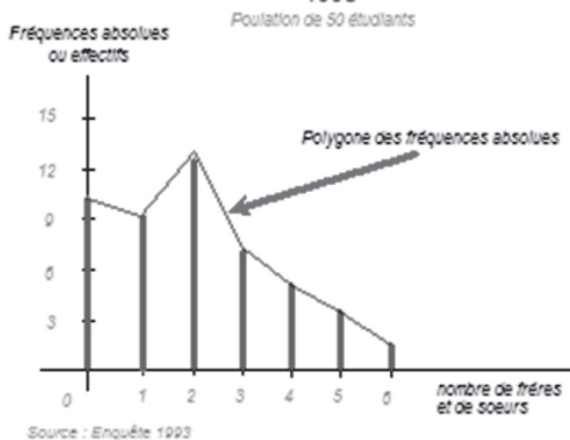


7.2 ● Représentations graphiques des variables quantitatives

► Le diagramme bâtons

Le diagramme bâtons est la représentation d'une distribution statistique discrète (x_i, n_i) ou (x_i, f_i) ou $(x_i, f_i\%)$, où chaque bâton a une hauteur proportionnelle à l'effectif n_i ou à la fréquence f_i (ou $f_i\%$) de la modalité x_i .

Nombre de frères et de soeurs dans l'enseignement supérieur en 1993
Population de 50 étudiants



► L'histogramme

L'histogramme est la représentation graphique d'une distribution statistique (x_i, n_i) ou (x_i, f_i) ou $(x_i, f_i\%)$ groupée en classes, où chaque classe est représentée