

Introduction à la statistique descriptive

Les méthodes de la statistique descriptive (statistique déductive) permettent de mener des études à partir de données exhaustives, c'est-à-dire concernant tous les individus de la population concernée par l'étude. Comme le rappelle André Vessereau (voir bibliographie), l'idée première et toujours fondamentale de la statistique descriptive est celle de dénombrement.

Quand les données ne concernent qu'un échantillon de la population, comme dans le cas des sondages, on a recours à la statistique inférentielle (statistique inductive), qui utilise la théorie des probabilités.

Globalement, la statistique reste très liée à la science du hasard, puisque les recensements nous fournissent des fréquences d'apparition auxquelles on fait jouer le même rôle qu'à la probabilité. Déjà, les manuscrits de Gottfried Leibniz, rédigés au début des années 1680, se situaient, à partir des travaux de John Graunt, dans la perspective d'une « synthèse entre science de la population et calcul des probabilités ».

Ce premier chapitre présente les principales clés de lecture de la statistique. La terminologie usuelle y est exposée, ainsi que la forme et le contenu des tableaux de données.

Deux annexes, proposées en fin de chapitre, sont consacrées à la prise en main d'Excel (annexe 1.1), ou de tout autre tableur équivalent, et de deux calculatrices graphiques, Texas Instrument et Casio (annexe 1.2) ou de toute autre calculatrice approchante. L'utilisation de ces outils facilitera la compréhension et la résolution de tous les exemples numériques des parties théoriques et des problèmes et exercices qui suivent.

1. Terminologie

Comme toute science, la statistique a son vocabulaire, qu'il est primordial de définir de façon rigoureuse afin d'indiquer le groupe sur lequel porte l'étude, les caractères ou variables relevés sur chacun des individus et les différents types de caractères.

1.1. La population

Le terme de population statistique est antérieur à la démographie et s'appliquait à l'origine à des catégories d'humains. Les populations n'étaient en effet pas pensées en bloc, leurs membres n'étant pas considérés comme égaux. Par exemple, on comptait les hommes en état de porter des armes, les individus soumis à l'impôt, etc. La démographie est venue plus tard, avec l'idée d'égalité des individus, qui a mené à la notion de recensement.

En statistique, le terme de population est plus général et peut désigner des humains, mais aussi des objets, des villes, des pays, des entreprises, des logements, etc., l'essentiel étant, comme pour la définition d'un ensemble en mathématiques, que l'on puisse dire clairement de tout élément qu'il appartient ou n'appartient pas à la population.

Les villes européennes de plus de 100 000 habitants, les voitures immatriculées en France, les départements français d'outre-mer sont autant d'exemples de population.

Définition

La **population statistique** est l'ensemble des éléments sur lesquels porte l'étude. Les éléments de la population sont appelés **individus statistiques** ou unités statistiques. La population constitue l'univers de référence de l'étude. Si la population comporte N individus, on notera $\Omega = \{\omega_1, \dots; \omega_N\}$, ω_i désignant pour i variant de 1 à N les individus qui la composent. Un **échantillon** de taille n est un sous-ensemble formé de n individus de la population ($n \leq N$).

La notion d'échantillon est fondamentale, car, en règle générale, la population entière n'est pas disponible ou observable. Dans ce cas, seul un échantillon est étudié et les résultats obtenus sont extrapolés à la population (voir P. Roger, chapitre 5). Par exemple, lorsqu'un magazine souhaite connaître la personnalité préférée des Français, il interroge seulement un échantillon de Français, généralement 1 000 individus, et non toute la population résidant en France métropolitaine, soit plus de 60 millions d'individus.

1.2. Notion de caractère ou variable statistique

Chaque individu d'une population peut être décrit relativement à un ou plusieurs caractères ou variables statistiques.

Définition

Une **variable statistique** (on parle aussi de **caractère statistique**), notée X , est une application définie sur une population statistique et à valeurs dans un ensemble M , appelé ensemble des modalités. Les **modalités** correspondent aux valeurs possibles de la variable statistique. Une variable statistique définit une partition sur une population, chaque individu appartenant à une et une seule modalité.

Si le nombre de modalités est noté r , l'ensemble des modalités de la variable X sera noté :

$$M = \{x_1; x_2; \dots; x_r\}.$$

Exemple 1.1 Une population statistique

Considérons les données suivantes concernant le nombre de femmes et d'hommes dans la population résidant en France métropolitaine en 2006 (en milliers) :

Femmes	Hommes
31 444	29 722

Source : Insee, recensement de la population, 2007 (champ : France métropolitaine)

La population étudiée est la population résidant en France métropolitaine recensée en 2006 et la variable étudiée est le sexe. Cette variable peut prendre deux valeurs possibles appelées modalités : féminin ou masculin. Ces modalités sont en général numérotées : si la variable étudiée, ici le sexe, est notée X , les deux modalités seront respectivement notées x_1 (pour féminin) et x_2 (pour masculin).

Une des premières opérations de la statistique consiste à recenser le nombre et/ou le pourcentage d'individus qui présentent une modalité déterminée d'une variable. C'est ainsi qu'à chaque modalité est associé un effectif et/ou une fréquence.

Définitions

L'**effectif** (aussi appelé fréquence absolue) de la modalité x_i est noté n_i et désigne le nombre d'individus de la population présentant la modalité x_i . L'effectif total de la population n est alors :

$n = n_1 + n_2 + \dots + n_r$, soit $n = \sum_{i=1}^r n_i$ (la somme des n_i pour i variant de 1 à r , et la lettre grecque sigma, \sum , désignant la somme).

La **fréquence** (par défaut fréquence relative) de la modalité x_i est notée f_i et est définie par : $f_i = n_i / N$; la fréquence exprime la proportion d'individus présentant une modalité donnée. Elle peut s'exprimer sous la forme d'un nombre décimal (en général avec une précision de quatre chiffres après la virgule) ou sous la forme d'un pourcentage.

Propriété

Soit X une variable à r modalités : $0 \leq f_i \leq 1$

$$\sum_{i=1}^r f_i = 1 \text{ (ou, en pourcentage : } \sum_{i=1}^r f_i = 100 \text{)}$$

Exemple 1.2 Effectifs et fréquences

Reprenons l'exemple précédent sur le sexe des individus de la population résidant en France métropolitaine. Les effectifs respectifs de ces modalités sont notés $n_1 = 31\,444$ et $n_2 = 29\,722$, avec $n = n_1 + n_2 = 61\,166$ milliers, effectif total de la population.

Les fréquences sont telles que $f_1 = n_1 / n = 31\,444 / 61\,166 = 0,5141$ et $f_2 = n_2 / N = 29\,722 / 61\,166 = 0,4859$, soit 51,41 % de femmes et 48,59 % d'hommes.

L'exemple 1.1 a mis en évidence une des deux natures des variables statistiques : la variable qualitative. Le sexe est une variable qualitative, car ses modalités ne sont pas des nombres. Une variable quantitative est une variable dont les modalités sont numériques.

Le poids d'un individu, l'âge, le nombre d'enfants par ménage, le salaire constituent des exemples de variables quantitatives.

1.3. Les variables qualitatives

Définition

Une variable statistique est dite de nature **qualitative** si ses modalités ne sont pas mesurables. Les modalités d'une variable qualitative sont les différentes catégories d'une nomenclature. Ces catégories doivent être exhaustives (chaque individu est affecté à une modalité) et incompatibles (un individu ne peut être affecté à plusieurs modalités) de façon à créer une partition.

Le sexe, la profession, l'état matrimonial sont quelques exemples de variables qualitatives. Pour ses enquêtes auprès des ménages, l'Insee utilise la nomenclature des Professions et catégories socioprofessionnelles (PCS-2003).

Les modalités d'une variable qualitative peuvent être classées sur deux types d'échelle : nominale ou ordinale. À ces deux types d'échelle correspondent deux types de variables qualitatives.

Variables qualitatives nominales

Les variables qualitatives nominales ne se mesurent pas. Cependant, leurs modalités peuvent être codées. L'ordre et l'origine de la codification sont arbitraires, cette codification pouvant être numérique, alphabétique ou alphanumérique. Les individus d'une même catégorie sont réputés « équivalents » pour la variable étudiée.

Définition

Une variable statistique qualitative est dite définie sur une échelle **nominale** si ses modalités ne sont pas naturellement ordonnées.

Exemple 1.3 Codage d'une variable qualitative nominale

Le tableau suivant indique les différentes catégories de la variable nominale Professions et catégories socioprofessionnelles (CSP) :

Code	Catégorie
1	Agriculteurs exploitants
2	Artisans, commerçants et chefs d'entreprise
3	Cadres et professions intellectuelles supérieures
4	Professions intermédiaires
5	Employés
6	Ouvriers
7	Retraités
8	Autres personnes sans activité professionnelle

Source : Insee, PCS-2003 (niveau 1 de la nomenclature)

Dans cet exemple, il n'y a pas d'ordre naturel entre les huit catégories, ou modalités, qui sont de simples étiquettes ; la variable qualitative « CSP » est définie sur une échelle nominale.

Variables qualitatives ordinales

Une échelle ordinale suppose l'existence d'une relation d'ordre total entre les catégories, c'est-à-dire que l'on peut opérer un classement de l'ensemble des catégories, de la plus petite à la plus grande (ou, inversement, de la plus grande à la plus petite).

Contrairement à ce qui se passe avec une échelle nominale, les expressions telles que « plus grand que », « précède », « se place après », etc. prennent un sens dans une échelle ordinale.

La codification peut être numérique, alphabétique ou alphanumérique, en association avec un sens de lecture. En cas de codage numérique, les opérations mathématiques sont dénuées de sens et l'écart entre les valeurs ne revêt aucune signification.

Définition

Une variable statistique qualitative est dite définie sur une échelle **ordinaire** si l'ensemble de ses modalités peut être doté d'une relation d'ordre.

1.4. Les variables quantitatives

Toute variable qui n'est pas qualitative ne peut être que quantitative. Les différentes modalités d'une variable quantitative constituent l'ensemble des valeurs numériques que peut prendre la variable.

Définition

Une variable statistique est dite de nature **quantitative** si ses modalités sont mesurables. Les modalités d'une variable quantitative sont des nombres liés à l'unité choisie, qui doit toujours être précisée.

Il existe deux types de variables quantitatives : les variables discrètes et les variables continues.

Ces variables ont en commun des modalités clairement ordonnées, pour lesquelles l'écart entre les valeurs possède une signification, et sur lesquelles il est possible de réaliser des opérations mathématiques telles que des calculs de moyennes, etc. Néanmoins, elles ont des propriétés et des traitements spécifiques qui nécessitent une étude séparée.

Variables quantitatives discrètes

Lorsque les modalités sont des valeurs numériques isolées, comme le nombre d'enfants par ménage, on parle de variable discrète¹.

Définition

Une variable statistique quantitative est dite **discrète** si l'ensemble de ses modalités est un ensemble fini ou dénombrable. Ainsi, l'ensemble des modalités peut être donné sous la forme d'une liste de nombres, $M = \{x_1 ; x_2 ; \dots ; x_i ; \dots\}$, finie ou infinie. Le plus souvent, les modalités appartiennent à l'ensemble N des entiers naturels ($N = \{0 ; 1 ; 2 ; \dots\}$). Cependant, une variable discrète peut prendre des valeurs non entières.

1. Du latin *discretus*, qui signifie « séparé » ; dans un ensemble discret, on peut séparer les éléments.

Variables quantitatives continues

Lorsque la variable, par exemple la taille d'un individu, peut prendre toutes les valeurs d'un intervalle, ces valeurs peuvent alors être regroupées en classes, et on parle dans ce cas de variable continue.

Définitions

Une variable statistique quantitative est dite **continue** si l'ensemble de ses modalités n'est pas dénombrable. Ainsi, une variable continue peut prendre toutes les valeurs d'un intervalle.

Pour étudier une variable statistique continue, on définit des classes ou intervalles de valeurs possibles. On peut ainsi **discrétiser** une variable continue (voir section 2.1). Les classes retenues constituent les modalités de la variable.

On appelle **amplitude de la classe** $[a_i ; b_i[$ le réel noté A_i , représentant la longueur de l'intervalle et défini par : $A_i = b_i - a_i$, a_i et b_i sont respectivement les bornes inférieure et supérieure de la classe n_i .

Le **centre de classe** de la classe $[a_i ; b_i[$ est le réel noté x_i représentant le milieu de l'intervalle et donné par : $x_i = (a_i + b_i) / 2$; c'est la moyenne arithmétique des bornes de la classe.

Le centre de classe est appelé à jouer un grand rôle dans les calculs, car le regroupement en classes constitue une perte d'information importante ; nous prendrons l'hypothèse de répartition uniforme à l'intérieur d'une classe, c'est-à-dire de concentration au centre des classes (voir chapitre 2).

Exemple 1.4 Calculs d'amplitudes et centres de classes

Le tableau suivant indique la structure par âges de la population féminine en France métropolitaine :

Âge	f_i (%)
Moins de 15 ans	17,5
15-24 ans	12,3
25-34 ans	12,7
35-44 ans	14,0
45-54 ans	13,6
55-64 ans	11,1
65-74 ans	8,6
75 ans ou +	9,1

Source : Insee, bilan démographique, 2006

Les modalités sont des intervalles qui, par convention, sont – à part pour la dernière classe – fermés à gauche et ouverts à droite. Ainsi, la première classe se note aussi : $[0 ; 15[$, la deuxième $[15 ; 25[$, etc.

Les classes ne sont pas de même amplitude, la première classe ayant une amplitude de 15 ans et les suivantes de 10 ans. Pour la dernière classe, dont l'amplitude n'est pas définie explicitement, la convention suivante est adoptée : en l'absence d'information, il lui est attribué l'amplitude de la classe précédente, $[65 ; 75[$, donc 10 ans, et elle est donc écrite : $[75 ; 85[$.

Le centre de la première classe est : $x_1 = (a_1 + b_1) / 2 = (0 + 15) / 2 = 7,5$ ans.

Cette distinction entre variable discrète et variable continue est parfois arbitraire, toute mesure étant discrète du fait de la précision limitée des instruments de mesure ou des arrondis. Cependant, la taille d'un individu, par exemple, est une variable continue du

fait que, indépendamment de la mesure, toute valeur de l'intervalle $[140 ; 150[$ peut représenter en centimètres la taille d'un individu. De même, il arrive qu'une variable discrète, comme le nombre d'habitants d'un pays, qui peut prendre un grand nombre de valeurs dans un intervalle soit considérée comme une variable continue.

En conclusion, toute étude de variable statistique devra être précédée d'une identification claire de la population, du caractère étudié et de sa nature, à savoir qualitatif ou quantitatif et, dans le cas quantitatif, discret ou continu.

2. Présentation des données

Les données statistiques sont issues de données brutes présentées sous forme de tableaux statistiques dans lesquels sont indiqués les effectifs et/ou les fréquences.

2.1. Distribution des effectifs ou des fréquences

Les tableaux statistiques contenant les effectifs et/ou les fréquences sont une première exploitation des données brutes.

Des données brutes au tableau statistique

Il est primordial de définir la population et de préciser avec rigueur la ou les variables relevées sur chacun des individus de la population ou de l'échantillon la représentant. Ensuite, quand les observations ont été recueillies, le premier travail consiste à les présenter, aussi clairement que possible, sous forme de tableau statistique. Ce tableau révèle la distribution statistique en présentant les couples de type $(x_i ; n_i)$, où les x_i sont les modalités et les n_i leurs effectifs respectifs, i entier variant de 1 à r , si r désigne le nombre de modalités du caractère. Il est également possible de présenter la distribution des fréquences, c'est-à-dire les couples de type $(x_i ; f_i)$.

Définitions

On appelle **données brutes** ou tableau élémentaire le tableau relevant pour chaque unité statistique la modalité de la variable étudiée.

Le **tri à plat** est la transformation qui permet de passer du tableau des données brutes au tableau de la distribution statistique présentant les modalités et les effectifs, les modalités étant classées par ordre croissant.

Discrétisation

Dans le cas d'une variable statistique quantitative continue, il est nécessaire de définir des classes pour pouvoir proposer un tri à plat.

Définition

On appelle **discrétisation** le découpage en classes d'une série statistique quantitative.

Ce découpage en classes pose de nombreuses questions : choix des amplitudes, amplitudes constantes ou variables, nombre de classes, etc. Nous ne rentrerons pas ici dans le détail de ces opérations (voir l'exercice 4 de ce chapitre).

2.2. Variables quantitatives : distribution des effectifs et des fréquences cumulés

Cette section concerne les variables quantitatives pour lesquelles le tableau statistique est réalisé, les modalités étant ordonnées dans l'ordre croissant. Les notions que nous allons définir sont liées à la notion de fonction de répartition, fondamentale en probabilité pour les variables aléatoires continues et sur laquelle nous reviendrons dans la section 3.3.

Reprenons l'exemple 1.4 et proposons de répondre à la question suivante : quelle proportion de la population féminine en France métropolitaine a moins de 35 ans ?

Nous pouvons affirmer que 42,5 % de la population féminine en France métropolitaine a moins de 35 ans, soit 17,5 % + 12,3 % + 12,7 %. Pour obtenir ce résultat, nous avons cumulé les fréquences des modalités inférieures ou égales à 34 ans.

Définitions

Effectifs cumulés croissants sur variable discrète : Si X désigne une variable quantitative discrète, on appelle effectif cumulé croissant, noté $n_{i,cc}$, le nombre d'individus statistiques pour lesquels X est inférieur ou égal à x_i .

$$\text{On a : } n_{i,cc} = n_1 \text{ et } n_{i,cc} = n_1 + n_2 + \dots + n_i = \sum_{k=1}^i n_k.$$

Si la série possède r modalités, x_r désignant alors la plus grande valeur de X , on a :

$$n_{r,cc} = n_1 + n_2 + \dots + n_r = \sum_{k=1}^r n_k = n, \text{ où } n \text{ désigne l'effectif total de la série.}$$

Fréquences cumulées croissantes sur variable discrète : Avec les mêmes hypothèses, on définit la fréquence cumulée croissante, notée $f_{i,cc}$, représentant la proportion d'individus statistiques pour lesquels X est inférieur ou égal à x_i .

$$\text{On a : } f_{i,cc} = f_1 \text{ et } f_{i,cc} = f_1 + f_2 + \dots + f_i = \sum_{k=1}^i f_k, \text{ ou encore } f_{i,cc} = \frac{n_{i,cc}}{n}.$$

Si la série possède r modalités, x_r désignant alors la plus grande valeur de X , on a :

$$f_{r,cc} = f_1 + f_2 + \dots + f_r = \sum_{k=1}^r f_k = 1 \text{ (ou 100 si les fréquences sont exprimées en pourcentage).}$$

Dans le cas d'une **variable quantitative continue**, les données sont groupées en classes $[a_i ; b_i[$, et on définit, de même que pour une variable discrète, $n_{i,cc}$ le nombre d'individus statistiques pour lesquels X est inférieur ou égal à b_i , et $f_{i,cc}$ la proportion d'individus statistiques pour lesquels X est inférieur ou égal à b_i .

Il est également possible de cumuler les effectifs et les fréquences dans le sens décroissant.

Définitions

Effectifs cumulés décroissants sur variable discrète : Si X désigne une variable quantitative discrète, on appelle effectif cumulé décroissant, noté $n_{i,cd}$, le nombre d'individus statistiques pour lesquels X est supérieur ou égal à x_i .

(Certains auteurs adoptent une convention différente : le nombre d'individus statistiques pour lesquels X est strictement supérieur à x_i).

$$\text{On a : } n_{i,cd} = n ; n_{i,cd} = n_i + n_{i+1} + \dots + n_r = \sum_{k=i}^r n_k, \text{ } r \text{ désignant le nombre de modalités, et } n_{r,cd} = n_r.$$

Fréquences cumulées décroissantes sur variable discrète : Avec les mêmes hypothèses, on définit la fréquence cumulée décroissante, notée $f_{i,cd}$, représentant la proportion d'individus statistiques pour lesquels X est supérieur ou égal à x_i .

On a : $f_{1,cd} = 1$; $f_{i,cd} = f_i + f_{i+1} + \dots + f_r = \sum_{k=i}^r f_k$, et $f_{r,cd} = f_r$, ou encore $f_{i,cd} = \frac{n_{i,cd}}{n}$.

Dans le cas d'une **variable quantitative continue**, les données sont groupées en classes $[a_i ; b_i[$, et on définit, de même que pour une variable discrète, $n_{i,cd}$ le nombre d'individus statistiques pour lesquels X est supérieur ou égal à a_i , et $f_{i,cd}$ la proportion d'individus statistiques pour lesquels X est supérieur ou égal à a_i .

Exemple 1.5 Calculs d'effectifs et fréquences cumulés croissants et décroissants

Le tableau suivant recense les enfants de moins de 6 ans en France métropolitaine :

Année	Moins de 2 ans	De 2 à 3 ans	De 4 à 5 ans
2006	1 577	1 550	1 511

Source : Insee, bilan démographique, 2006

Les effectifs cumulés croissants ($n_{i,cc}$), décroissants ($n_{i,cd}$), et les fréquences cumulées croissantes ($f_{i,cc}$), décroissantes ($f_{i,cd}$), correspondants sont les suivants :

Âge	n_i	$n_{i,cc}$	$n_{i,cd}$	$f_{i,cc}$	f_i	$f_{i,cd}$
[0 ; 2[1 577	1 577	4 638	0,3400	0,3400	1
[2 ; 4[1 550	3 127	3 061	0,3342	0,6742	0,6600
[4 ; 6[1 511	4 638	1 511	0,3258	1	0,3258
Total	4 638			1		

3. Représentations graphiques des séries à une variable

L'apparition des graphiques statistiques, liée à l'utilisation des coordonnées, doit essentiellement son origine au philosophe et mathématicien René Descartes (1596-1650). Ces graphiques constituent une synthèse visuelle indispensable de l'information contenue dans le tableau statistique.

Les graphiques utilisés dépendent de la nature de la variable. Nous utiliserons, pour représenter les distributions d'effectifs (ou de fréquences), les diagrammes circulaires (ou secteurs), les diagrammes en tuyaux d'orgue, les diagrammes en bâtons, les histogrammes et le polygone des effectifs. Pour les distributions cumulées, nous utiliserons les polygones des effectifs (ou des fréquences) cumulés croissants et décroissants.

3.1. Graphiques pour variables qualitatives

Les variables qualitatives – nominales ou ordinales – peuvent être représentées au choix à l'aide d'un diagramme circulaire ou à l'aide d'un diagramme en tuyaux d'orgue.

Diagramme circulaire

Le diagramme circulaire, également appelé « camembert », permet une représentation de la distribution d'une variable dans un cercle qui représente 100 % des modalités (voir figure 1.1).

Définition

Un **diagramme circulaire** est un graphique constitué d'un cercle divisé en secteurs dont les angles au centre sont proportionnels aux effectifs (ou aux fréquences). De fait, les aires des secteurs sont proportionnelles aux effectifs. L'angle α_i d'une modalité d'effectif n_i est donné en degrés

$$\text{par : } \alpha_i = \frac{n_i}{n} \times 360 = f_i \times 360 .$$

Il est également possible d'utiliser un graphique semi-circulaire formé d'un demi-cercle (180°).

Figure 1.1

Diagramme circulaire : proportion (en pourcentage) de bacheliers et non-bacheliers dans une génération en France métropolitaine et DOM, 2005.

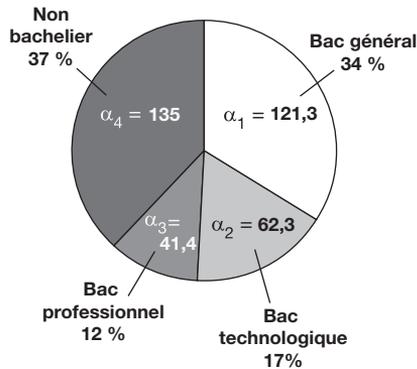


Diagramme en tuyaux d'orgue (en barres)

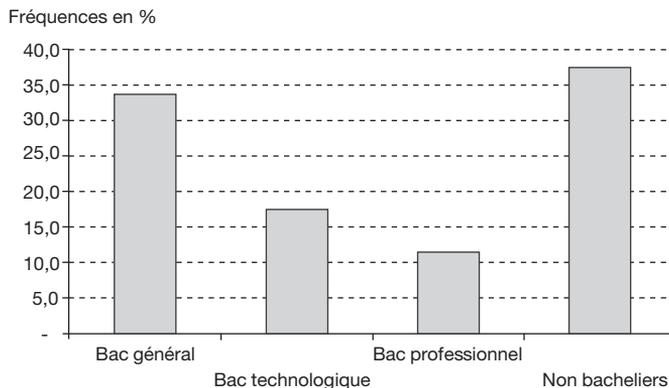
Le diagramme en tuyaux d'orgue est une représentation de la distribution d'une variable selon des rectangles horizontaux ou verticaux ayant tous une même base, de largeur arbitraire (voir figure 1.2).

Définition

Un **diagramme en tuyaux d'orgue** est un graphique qui à chaque modalité d'une variable qualitative associe un rectangle de base constante dont la hauteur est proportionnelle à l'effectif (ou à la fréquence). De fait, les aires des secteurs sont proportionnelles aux effectifs. Les rectangles sont en général disjoints, verticaux ou horizontaux.

Figure 1.2

Diagramme en tuyaux d'orgue : proportion (en pourcentage) de bacheliers et non-bacheliers dans une génération en France métropolitaine et DOM, 2005.



3.2. Graphiques pour variables quantitatives

La représentation graphique d'une variable quantitative dépend de sa nature : discrète ou continue.

Variables discrètes : diagramme en bâtons

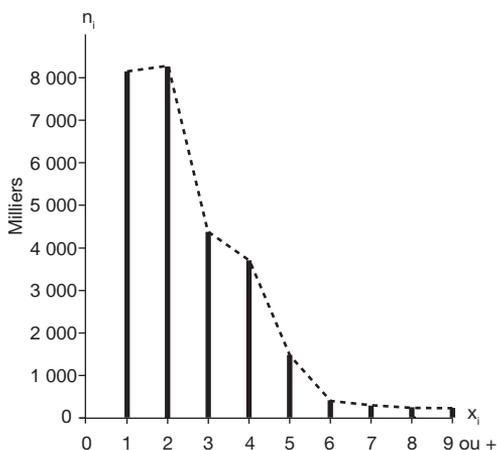
La distribution d'une variable quantitative discrète peut être représentée par un diagramme en bâtons (voir figure 1.3).

Définition

On appelle **diagramme en bâtons** un graphique qui à chaque modalité d'une variable quantitative discrète associe un segment (bâton) dont la hauteur est proportionnelle à l'effectif (ou à la **fréquence**).

Figure 1.3

Diagramme en bâtons et polygone des effectifs : nombre de personnes par ménage, France, 1999.



Variables continues : histogramme

En 2005, Monaco avait 32 543 habitants et le Japon, 127 417 244 (source : Institut national d'études démographiques). Bien sûr, les démographes diront que ces renseignements sont très largement insuffisants pour comparer la démographie des deux pays : il faut au

minimum s'intéresser aux superficies de ces deux pays et calculer pour chacun d'entre eux la densité de population, c'est-à-dire le nombre d'habitants au kilomètre carré. Avec une superficie de 2,02 km² pour Monaco et de 378 000 km² pour le Japon, les densités sont respectivement $d_1 = 32\,543 / 2,02 = 16\,110,40$ h/km² pour Monaco et $d_2 = 127\,417\,244 / 378\,000 = 337$ h/km² pour le Japon. Autrement dit, alors que la population de Monaco est la moins importante en taille, sa densité de population est plus importante que celle du Japon.

Cette notion de densité est essentielle pour les variables continues : il est absurde de comparer ou de représenter côte à côte des classes qui n'ont pas la même amplitude sans faire intervenir la densité. Ce principe est omniprésent lors de la réalisation d'un histogramme.

Définitions

Un **histogramme** est un diagramme composé de rectangles contigus dont les aires sont proportionnelles aux effectifs (ou aux fréquences) et dont les bases sont déterminées par les intervalles de classes.

Dans le cas d'une variable quantitative continue, on définit la **densité** d'effectif d_i d'une classe d'effectif n_i et d'amplitude A_i par : $d_i = n_i / A_i$ (ou, dans le cas des fréquences, f_i / A_i).

Lors de la réalisation d'un histogramme, il est indispensable de distinguer deux cas.

1. Si les amplitudes de classes sont égales, la hauteur des rectangles correspondra aux effectifs (ou aux fréquences) des classes.
2. Si les amplitudes sont différentes, afin de constituer l'histogramme, il est nécessaire de :
 - calculer, pour chaque classe, l'amplitude A_i ;
 - calculer la densité $d_i = n_i / A_i$ pour un histogramme des effectifs, et $d_i = f_i / A_i$ pour un histogramme des fréquences ;
 - affecter à chaque rectangle une hauteur proportionnelle à la densité d_i de la classe correspondante.

Soit $\min(A_i)$ l'amplitude minimale de classe, la hauteur est alors appelée « effectif corrigé » et notée $n_{ic} = d_i \times \min(A_i)$; cette convention revient à adopter $\min(A_i)$ comme unité d'amplitude de classe. Les classes ayant pour amplitudes $\min(A_i)$ sont alors représentées par des rectangles dont la hauteur est l'effectif. De même, il est possible de retenir comme hauteur la fréquence corrigée $f_{ic} = d_i \times \min(A_i)$, avec $d_i = f_i / A_i$ dans le cas d'un histogramme des fréquences. L'utilisation de $\min(A_i)$ est une convention facultative ; un histogramme est correct dès lors que les effectifs (ou les fréquences) corrigés sont proportionnels aux densités.

Remarques :

1. Dans un histogramme, les aires des rectangles permettent de comparer les effectifs et les hauteurs de comparer les densités.
2. La définition de la densité d'effectif peut s'écrire : $n_i = A_i \times d_i$; cette formule permet d'estimer l'effectif d'un intervalle, sous l'hypothèse de répartition uniforme à l'intérieur des classes (voir exemple 1.6).

Exemple 1.6 Réalisation d'un histogramme et d'un polygone des effectifs

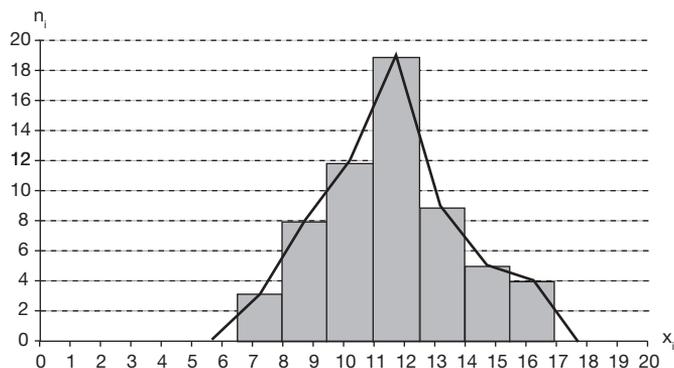
Le responsable des ressources humaines d'une entreprise a relevé la distribution statistique suivante correspondant à l'ancienneté du personnel cadre dans l'entreprise, exprimée en années :

Classes	Effectifs
[6,5 ; 8[3
[8 ; 9,5[8
[9,5 ; 11[12
[11 ; 12,5[19
[12,5 ; 14[9
[14 ; 15,5[5
[15,5 ; 17[4
Total	60

L'histogramme des effectifs est présenté avec, sur le même graphique, le polygone des effectifs tracé en courbe pleine (voir figure 1.4). Ce polygone permet de représenter la distribution sous la forme d'une courbe ; quand les amplitudes de classes sont égales, on l'obtient en joignant les milieux des bases supérieures de chaque rectangle de l'histogramme par des segments de droite. On adjoint généralement une classe d'effectif nul, de part et d'autre de l'histogramme, afin de respecter la règle de compensation des aires : l'aire totale du domaine situé entre l'axe des x et le polygone est égale à la somme des aires des rectangles de l'histogramme. Elle représente l'effectif total.

Figure 1.4

Histogramme et polygone des effectifs, classes de même amplitude : ancienneté du personnel cadre de l'entreprise.



Modifions légèrement cet exemple en regroupant les deux dernières classes en une seule. Ce regroupement permet de traiter le cas de classes d'amplitudes différentes, puisque ainsi la dernière classe est d'amplitude 3 contre 1,5 pour toutes les autres classes.

Classes	Effectifs
[6,5 ; 8[3
[8 ; 9,5[8
[9,5 ; 11[12
[11 ; 12,5[19
[12,5 ; 14[9
[14 ; 17[9
Total	60

Questions : tracez l'histogramme des effectifs et estimez la proportion de cadres ayant une ancienneté comprise entre 10 et 13,25 années.

Les classes étant d'amplitudes inégales, il est nécessaire de calculer les amplitudes (A_i), les densités (d_i), puis les effectifs corrigés ($n_i c$) pour chaque classe. Les résultats de ces calculs sont présentés à la figure 1.5.

Figure 1.5

Calcul des effectifs corrigés dans le cas de classes d'amplitudes inégales.

	A	B	C	D	E
1	Classes	n_i	a_i	d_i	$n_i c$
2	[6,5 ; 8[3	1,50	2,00	3,0
3	[8 ; 9,5[8	1,50	5,33	8,0
4	[9,5 ; 11[12	1,50	8,00	12,0
5	[11 ; 12,5[19	1,50	12,67	19,0
6	[12,5 ; 14[9	1,50	6,00	9,0
7	[14 ; 17[9	3,00	3,00	4,5
8	Total	60			

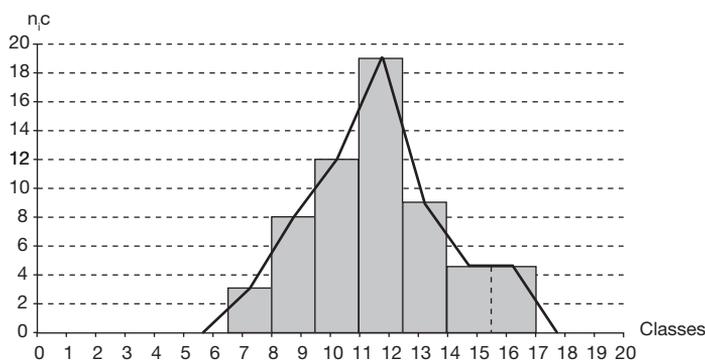
On peut alors tracer l'histogramme de la figure 1.6 à partir des effectifs corrigés, ainsi que le polygone des effectifs, en trait continu.

Pour tracer le polygone des effectifs, nous avons effectué un découpage artificiel en pseudo-classes d'amplitude 1,5, dont nous avons pris les milieux des bases supérieures de façon à respecter la règle de compensation des aires : les aires des triangles extérieurs au domaine délimité par le polygone sont égales à celles des triangles qui sont situés sous le polygone. Ainsi, l'aire totale du domaine situé sous le polygone des effectifs est égale à l'aire totale des rectangles de l'histogramme.

Ce qui est fait dans cet exemple à partir des effectifs peut également être réalisé à partir des fréquences, afin de tracer l'histogramme et le polygone des fréquences.

Figure 1.6

Histogramme et polygone des effectifs : classes d'amplitudes inégales.



Pour estimer l'effectif des cadres de l'intervalle [10 ; 13,25[, on décompose celui-ci en trois intervalles : [10 ; 11[, [11 ; 12,5[et [12,5 ; 13,25[. Il suffit alors de multiplier, pour chacun de ces intervalles, l'amplitude par la densité pour obtenir l'effectif total recherché (voir figure 1.6bis) :

Figure 1.6bis

Effectifs estimés et densité.

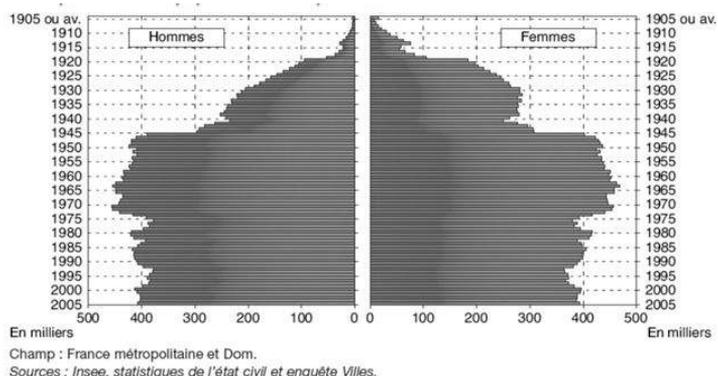
	A	B	C	D
1	Intervalles	a_i	d_i	n_i estimés
2	[10;11[1	8	8
3	[11;12,5[1,5	12,67	19,01
4	[12,5;13,25[0,75	6	4,5
5	Total			31,51

Ce qui donnera pour estimation de la proportion recherchée : $31,51 / 60 = 0,5251$, soit 52,51 %.

Enfin, il serait inconcevable de ne pas évoquer une variété d'histogramme, la pyramide, dont l'exemple le plus célèbre est la pyramide des âges (voir figure 1.7).

Cette variété d'histogramme, où les axes ont été modifiés (classes en ordonnées et effectifs en abscisses), est largement utilisée en démographie. Les classes sont annuelles. Les aires des rectangles représentent le nombre d'hommes ou de femmes vivants et nés l'année considérée, en lecture sur l'axe des abscisses.

Figure 1.7
Pyramide des âges.



3.3. Diagrammes cumulatifs

Les notions d'effectifs et de fréquences cumulés nous ont donné l'occasion d'introduire la notion de fonction de répartition, que nous définissons ci-après avant d'évoquer sa représentation graphique.

Définition

Si X est une variable quantitative, on introduit la **fonction de répartition**, qui à tout nombre réel x associe la proportion des individus de la population pour lesquels X est inférieur ou égal à x .

Pour tout x réel, $0 \leq F(x) \leq 1$ (les valeurs de F peuvent également être exprimées en pourcentage).

La première étape de la construction d'une fonction de répartition consiste donc à calculer les fréquences cumulées croissantes, en distinguant deux cas : le discret et le continu.

Fonction de répartition d'une variable discrète

La fonction de répartition d'une variable quantitative discrète est une fonction en escalier, c'est-à-dire constante par intervalle. De plus, elle est croissante de 0 à 1 et définie par :

- si $x < x_1$, $F(x) = 0$;
- si $x = x_1$, $F(x) = f_1cc$;
- si $x_1 \leq x < x_{i+1}$, $F(x) = f_1cc$;
- si $x \geq x_r$, $F(x) = 1$.

Fonction de répartition d'une variable continue

A priori, la fonction de répartition d'une variable continue n'est connue que pour les extrémités de classes. Cependant, si l'on admet l'hypothèse de répartition uniforme des observations au sein de chaque classe, on peut estimer les valeurs de $F(x)$ par interpolation linéaire. Cela revient à approximer la représentation graphique par une fonction affine par morceaux : concrètement, on trace la courbe en joignant deux points consécutifs connus par un segment de droite (cette courbe est aussi appelée ogive de Galton).

Avec cette hypothèse, $F(x)$ représente l'aire située sous l'histogramme des fréquences, à gauche de la valeur x .

Polygones des fréquences cumulées croissantes et décroissantes

Dans le cas d'une variable continue, on définit les polygones des fréquences (ou des effectifs) cumulées croissantes et décroissantes ; ils seront utilisés notamment pour déterminer la médiane de la série (voir chapitre 2).

Le polygone des fréquences cumulées croissantes commence au point de coordonnées $(a_1 ; 0)$, car la proportion de valeurs inférieures à a_1 est nulle. Il est obtenu en joignant les points de coordonnées $(b_i ; f_{i,cc})$ – il correspond à la restriction de la fonction de répartition aux valeurs de x inférieures ou égales à la borne supérieure de la dernière classe.

Le polygone des fréquences cumulées décroissantes s'obtient de la même façon, en adjoignant le point de coordonnées $(b_r ; 0)$, car, b_r désignant la borne supérieure de la dernière classe, la proportion de valeurs supérieures à b_r est nulle.

Exemple 1.7 Réalisation des polygones des fréquences cumulées croissantes et décroissantes

Le tableau suivant donne la structure de la population chinoise suivant l'âge :

0-14 ans	15-24 ans	25-59 ans	60 ans et plus
21,4 %	16,6 %	51,1 %	10,9 %

Source : ONU, 2005

Les fréquences cumulées croissantes et décroissantes sont calculées puis organisées pour correspondre aux bornes des classes (voir figure 1.8).

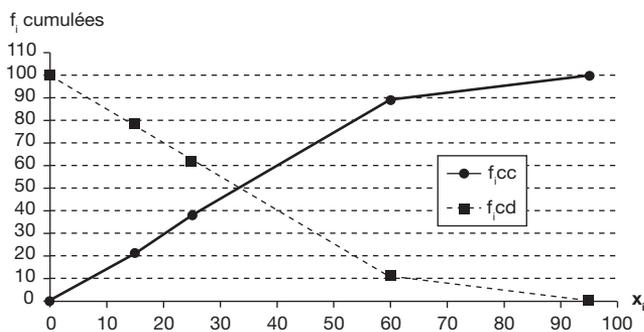
Figure 1.8

Plages de données des polygones des fréquences cumulées croissantes et décroissantes.

	A	B	C	D	E
1	Classes	f_i	Borne de classe	$f_{i,cc}$	$f_{i,cd}$
2			0	0,00	100,00
3	[0 ; 15[21,40	15	21,40	78,60
4	[15 ; 25[16,60	25	38,00	62,00
5	[25 ; 60[51,10	60	89,10	10,90
6	60 et +	10,90	95	100,00	0,00

À partir de ces données, il est possible de tracer les polygones des fréquences cumulées croissantes et décroissantes (voir figure 1.9).

Figure 1.9
Polygones des fréquences cumulées croissantes et décroissantes de l'âge de la population chinoise.



Conclusion

On retiendra de ce premier chapitre l'importance de la terminologie. On devra savoir identifier, dans un exercice, la population, les variables étudiées et leur nature : qualitative, quantitative discrète ou quantitative continue. On notera que le discret et le continu, en statistique comme en probabilité, nécessitent des traitements différents ; dans le cas continu, on retiendra l'importance de la notion de densité. Par ailleurs, on n'insistera jamais assez sur l'importance des représentations graphiques en statistique ; à l'issue de ce chapitre, on devra maîtriser notamment les histogrammes et les polygones des effectifs (ou des fréquences) cumulés croissants et décroissants.

Problèmes et exercices

Les problèmes et exercices suivants proposent la mise en application des notions exposées dans la première partie de ce chapitre.

- L'exercice 1 traite du passage d'une série brute à un tableau statistique.
- Les exercices 2, 3 et 5 s'attachent aux graphiques associés aux différentes natures de variables statistiques.
- L'exercice 4 s'intéresse à la discrétisation des données.



Exercice 1 : De la série brute à la présentation des statistiques

La liste suivante est composée de prénoms d'un groupe d'étudiants, suivis entre parenthèses du nombre de films que chacun d'entre eux a vus au cours du mois dernier :

Pierre (3), Paul (2), Jacques (2), Ralph (3), Abdel (1), Sidonie (2), Henri (0), Paulette (1), Farida (2), Laure (2), Kevin (0), Carole (3), Marie-Claire (0), Jeanine (3), Julie (2), Ernest (3), Cindy (3), Vanessa (2), José (1), Aurélien (1).

1. Déterminez :
 - a. la population étudiée ;
 - b. la variable étudiée.
2. Précisez :
 - a. la nature de la variable ;
 - b. les modalités de la variable.
3. Construisez le tableau statistique associé à la distribution des effectifs.
4. Représentez la distribution des effectifs par un diagramme en bâtons.
5. Calculez les effectifs :
 - a. cumulés croissants ;
 - b. cumulés décroissants.
6. Calculez les fréquences :
 - a. cumulées croissantes ;
 - b. cumulées décroissantes.

Solution

1. a. La population étudiée est le groupe d'étudiants.
b. La variable étudiée est $X =$ « nombre de films que chacun d'entre eux a vus au cours du mois dernier ».
2. a. La variable étudiée est quantitative discrète.

- b. L'ensemble M des modalités est $M = \{0 ; 1 ; 2 ; 3\}$.
3. Le tableau statistique associé est composé de deux colonnes :
- La première colonne comporte les modalités x_i de X.
 - La seconde colonne comporte les effectifs n_i associés à chacune de ces modalités.

Le tableau statistique associé à X est le suivant.

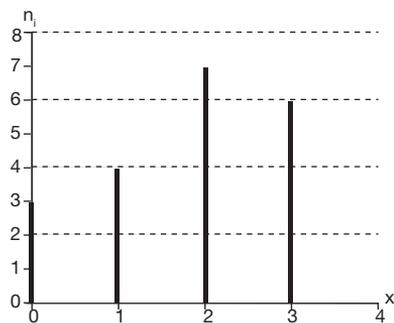
x_i	n_i
0	3
1	4
2	7
3	6

L'effectif total est $n = \sum_{i=1}^4 n_i$, soit $n = 20$.

4.

Figure 1.10

Diagramme en bâtons des effectifs.

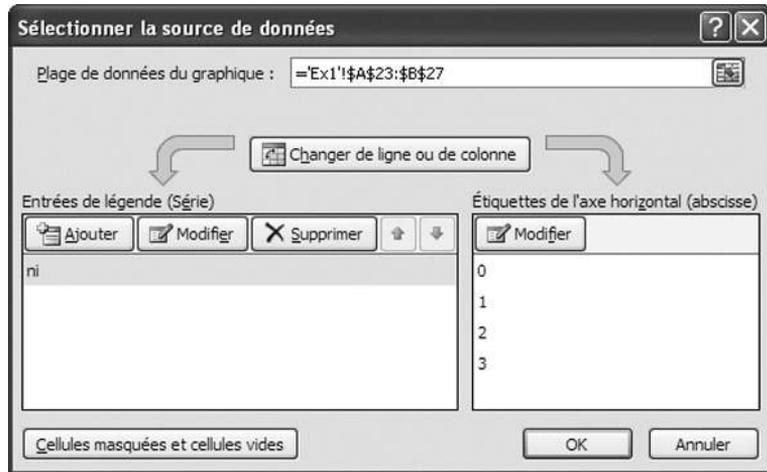


Le même diagramme en bâtons (voir figure 1.10) peut être réalisé sous Excel. Pour cela, placez le curseur sur une cellule vide. Dans la barre de menus, cliquez sur Insertion/Graphique. Sélectionnez le premier histogramme 2D (le plus à gauche). Notez que le mot « histogramme » est employé par Excel comme un terme générique désignant des barres verticales et non un histogramme au sens statistique.

Une zone de graphique blanche apparaît. Cliquez dessus, puis dans la barre de menus Excel, cliquez sur Création/Sélectionner des données. L'assistant graphique apparaît. Pour le renseigner correctement, comme cela est indiqué à la figure 1.11, suivez ces étapes :

- Saisissez la colonne des effectifs n_i dans le champ Plage de données du graphique. Pour cela, sélectionnez la plage B23:B27 de la feuille Ex1 à l'aide de la souris.
- Cliquez sur Modifier dans le champ Étiquettes de l'axe horizontal (abscisse). Dans la nouvelle boîte de dialogue qui s'ouvre, saisissez la colonne des modalités x_i dans le champ Plage d'étiquettes des axes. Pour cela, sélectionnez la plage A24:A27 de la feuille Ex1 à l'aide de la souris.
- Cliquez sur le bouton OK.

Figure 1.11
Sélection des données à représenter dans l'assistant graphique.



L'assistant graphique se ferme et le graphique apparaît (voir figure 1.12). Vous pouvez modifier les options d'affichage du graphique en appelant un menu par un clic droit sur la zone de graphique.

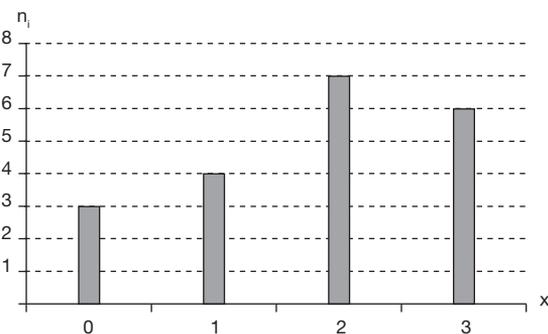
5. a. Soit $n_{i,cc}$ l'effectif cumulé croissant de la modalité i : $n_{1,cc} = n_1 = 3$, soit $n_{1,cc} = 3$; $n_{2,cc} = n_{1,cc} + n_2 = 3 + 4$, soit $n_{2,cc} = 7$; $n_{3,cc} = n_{2,cc} + n_3 = 7 + 7$, soit $n_{3,cc} = 14$; $n_{4,cc} = n_{3,cc} + n_4 = 14 + 6$, soit $n_{4,cc} = 20$.

b. Soit $n_{i,cd}$ l'effectif cumulé décroissant de la modalité i : $n_{1,cd} = n$, soit $n_{1,cd} = 20$; $n_{2,cd} = n_{1,cd} - n_1 = 20 - 3$, soit $n_{2,cd} = 17$; $n_{3,cd} = n_{2,cd} - n_2 = 17 - 4$, soit $n_{3,cd} = 13$; $n_{4,cd} = n_{3,cd} - n_3 = 13 - 7$, soit $n_{4,cd} = 6$.

Les résultats des effectifs cumulés croissants et décroissants se présentent dans un tableau obtenu en ajoutant deux colonnes au tableau statistique initial : les effectifs cumulés croissants $n_{i,cc}$ et les effectifs cumulés décroissants $n_{i,cd}$.

x_i	n_i	$n_{i,cc}$	$n_{i,cd}$
0	3	3	20
1	4	7	17
2	7	14	13
3	6	20	6

Figure 1.12
Diagramme en bâtons sous Excel.



6. Pour pouvoir calculer les fréquences cumulées croissantes $f_{i,cc}$ et décroissantes $f_{i,cd}$, il convient de calculer les fréquences f_i .

Soit f_i la fréquence de la classe i : $f_1 = \frac{n_1}{n} = \frac{3}{20}$, soit $f_1 = 0,15$; $f_2 = \frac{n_2}{n} = \frac{4}{20}$, soit $f_2 = 0,20$;

$f_3 = \frac{n_3}{n} = \frac{7}{20}$, soit $f_3 = 0,35$; $f_4 = \frac{n_4}{n} = \frac{6}{20}$, soit $f_4 = 0,30$.

- a. Soit $f_{i,cc}$ la fréquence cumulée croissante de la classe i : $f_{1,cc} = f_1 = 0,15$, soit $f_{1,cc} = 0,15$; $f_{2,cc} = f_{1,cc} + f_2 = 0,15 + 0,20$, soit $f_{2,cc} = 0,35$; $f_{3,cc} = f_{2,cc} + f_3 = 0,35 + 0,35$, soit $f_{3,cc} = 0,70$; $f_{4,cc} = f_{3,cc} + f_4 = 0,70 + 0,30$, soit $f_{4,cc} = 1$.
- b. Soit $f_{i,cd}$ la fréquence cumulée décroissante de la classe i : $f_{1,cd} = 1$, soit $f_{1,cd} = 1$; $f_{2,cd} = f_{1,cd} - f_1 = 1 - 0,15$, soit $f_{2,cd} = 0,85$; $f_{3,cd} = f_{2,cd} - f_2 = 0,85 - 0,20$, soit $f_{3,cd} = 0,65$; $f_{4,cd} = f_{3,cd} - f_3 = 0,65 - 0,35$, soit $f_{4,cd} = 0,30$.

Les résultats des fréquences cumulées croissantes et décroissantes se présentent dans un tableau obtenu en ajoutant deux colonnes au tableau statistique : les fréquences cumulées croissantes $f_{i,cc}$ et les fréquences cumulées décroissantes $f_{i,cd}$.

x_i	n_i	$n_{,cc}$	$n_{,cd}$	f_i	$f_{,cc}$	$f_{,cd}$
0	3	3	20	0,15	0,15	1,00
1	4	7	17	0,20	0,35	0,85
2	7	14	13	0,35	0,70	0,65
3	6	20	6	0,30	1,00	0,30

Exercice 2 : Représentations graphiques simples

Le tableau suivant indique la répartition des familles de l'île de La Réunion selon leur nombre d'enfants :

Nombre d'enfants	Nombre de familles
0	31 038
1	54 812
2	51 252
3	26 613
4 ou +	16 162

Source : Insee, recensement, 1999

- Déterminez :
 - la population étudiée ;
 - la variable étudiée.
- Précisez :
 - la nature de la variable ;
 - les modalités de la variable.
- Représentez la distribution par diagramme circulaire.

4. À la suite de la question précédente :
 - a. Calculez les effectifs cumulés croissants et décroissants.
 - b. Représentez la fonction de répartition.
5. Combien de familles sont composées de :
 - a. au moins 1 enfant ?
 - b. au plus 2 enfants ?

Solution

1. a. La population étudiée est composée des familles de La Réunion.
b. La variable étudiée est $X = \text{« nombre d'enfants »}$.
2. a. La variable étudiée est quantitative discrète.
b. L'ensemble des modalités de la variable étudiée est $M = \{0 ; 1 ; 2 ; 3 ; 4 \text{ ou } +\}$.
3. Pour réaliser un diagramme circulaire, il convient de tracer un cercle et de retenir pour chaque modalité i un secteur d'angle au centre : $\alpha_i = 360 \times f_i$ exprimé en degrés.

Pour la modalité 1, $f_1 = \frac{n_1}{n} = \frac{31038}{179877} = 0,1726$, donc $\alpha_1 = 360 \times 0,1726$, soit $\alpha_1 = 62,12^\circ$.

Pour la modalité 2, $f_2 = \frac{n_2}{n} = \frac{54812}{179877} = 0,3047$, donc $\alpha_2 = 360 \times 0,3047$, soit $\alpha_2 = 109,7^\circ$.

Pour la modalité 3, $f_3 = \frac{n_3}{n} = \frac{51252}{179877} = 0,2849$, donc $\alpha_3 = 360 \times 0,2849$, soit $\alpha_3 = 102,57^\circ$.

Pour la modalité 4, $f_4 = \frac{n_4}{n} = \frac{26613}{179877} = 0,1480$, donc $\alpha_4 = 360 \times 0,1480$, soit $\alpha_4 = 53,26^\circ$.

Pour la modalité 5, $f_5 = \frac{n_5}{n} = \frac{16162}{179877} = 0,0899$, donc $\alpha_5 = 360 \times 0,0899$, soit $\alpha_5 = 32,35^\circ$.

On vérifie que la somme des angles est bien de 360° .

Ces calculs sont effectués sous Excel, dans le tableau présenté à la figure 1.13, colonnes C et D.

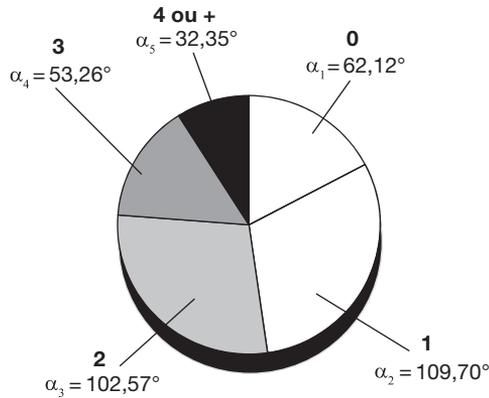
Figure 1.13

Résultats sous Excel.

	A	B	C	D	E	F	G
1	x_i	n_i	f_i	α_i	$n_{i\text{cc}}$	$n_{i\text{cd}}$	$f_{i\text{cc}}$
2	0	31 038	0,1726	62,12	31 038	179 877	0,1726
3	1	54 812	0,3047	109,70	85 850	148 839	0,4773
4	2	51 252	0,2849	102,57	137 102	94 027	0,7622
5	3	26 613	0,1480	53,26	163 715	42 775	0,9101
6	4 ou +	16 162	0,0899	32,35	179 877	16 162	1,0000

Le diagramme circulaire de la figure 1.14 est réalisé à partir de ces résultats.

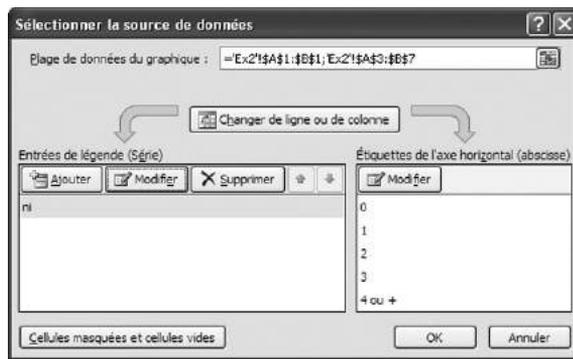
Figure 1.14
Réalisation d'un diagramme circulaire : répartition des familles de La Réunion selon leur nombre d'enfants.



Pour réaliser un diagramme circulaire sous Excel, sélectionnez, en mettant en surbrillance, les deux colonnes contenant les modalités n_i et les effectifs x_i , soit la plage A1:B6 de la figure 1.13. Puis cliquez sur Insertion/Secteurs dans la barre de menus. Sélectionnez le premier secteur 2D (le plus à gauche). Le graphique s'affiche alors immédiatement.

En sélectionnant la bonne plage de données dès le départ, Excel a automatiquement rempli l'assistant graphique adéquat, comme à la figure 1.15. Vous pouvez retrouver cette boîte de dialogue en sélectionnant le graphique et en cliquant sur Création/Sélectionner des données.

Figure 1.15
Données saisies dans l'assistant graphique.



Vous pouvez modifier les options d'affichage du diagramme en appelant un menu par un clic droit sur la zone de graphique.

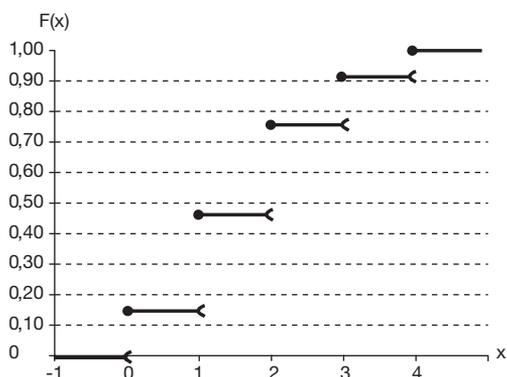
4. a. Soit $n_{i,cc}$ l'effectif cumulé croissant de la classe i : $n_{1,cc} = n_1 = 31\ 038$, soit $n_{2,cc} = 31\ 038$; $n_{2,cc} = n_{1,cc} + n_2 = 31\ 038 + 54\ 812$, soit $n_{2,cc} = 85\ 850$; $n_{3,cc} = n_{2,cc} + n_3 = 85\ 850 + 51\ 252$, soit $n_{3,cc} = 137\ 102$; $n_{4,cc} = n_{3,cc} + n_4 = 137\ 102 + 26\ 613$, soit $n_{4,cc} = 163\ 715$; $n_{5,cc} = n_{4,cc} + n_5 = 163\ 715 + 16\ 162$, soit $n_{5,cc} = 179\ 877$.

Soit $n_{i,cd}$ l'effectif cumulé décroissant de la classe i : $n_{1,cd} = n$, soit $n_{1,cd} = 179\ 877$; $n_{2,cd} = n_{1,cd} - n_1 = 179\ 877 - 31\ 038$, soit $n_{2,cd} = 148\ 839$; $n_{3,cd} = n_{2,cd} - n_2 = 148\ 839 - 54\ 812$, soit $n_{3,cd} = 94\ 027$; $n_{4,cd} = n_{3,cd} - n_3 = 94\ 027 - 51\ 252$, soit $n_{4,cd} = 42\ 775$; $n_{5,cd} = n_{4,cd} - n_4 = 42\ 775 - 26\ 613$, soit $n_{5,cd} = 16\ 162$.

Les résultats des effectifs cumulés croissants et décroissants se présentent dans un tableau obtenu en ajoutant deux colonnes au tableau statistique précédent : les effectifs cumulés croissants $n_{i,cc}$ en colonne E et les effectifs cumulés décroissants $n_{i,cd}$ en colonne F (voir figure 1.13).

- b. La fonction de répartition est réalisée à partir des fréquences cumulées croissantes ($f_{i,cc}$), calculées en colonne G du tableau statistique précédent (voir figure 1.13), sur du papier millimétré (voir figure 1.16).

Figure 1.16
Fonction de répartition
du nombre d'enfants des
familles de La Réunion.



5. a « Au moins 1 enfant » correspond aux familles qui ont 1, 2, 3 ou 4 et + enfants, ou encore toutes les familles sauf celles qui ont 0 enfant, c'est-à-dire toutes les familles sauf celles qui présentent la modalité x_1 de X. Le nombre de ces familles est l'effectif cumulé décroissant $n_{2,cd} = 148\ 839$, soit $179\ 877 - 31\ 038$. Ainsi, 148 839 familles sont composées d'au moins 1 enfant.
- b. « Au plus 2 enfants » correspond aux familles qui ont 0, 1 ou 2 enfants, c'est-à-dire les familles qui présentent les modalités x_1, x_2 ou x_3 de X. Le nombre de ces familles est l'effectif cumulé croissant $n_{3,cc} = 137\ 102$, soit $31\ 038 + 54\ 812 + 51\ 252$. Ainsi, 137 102 familles sont composées d'au plus 2 enfants.



Exercice 3 : L'histogramme

La Sécurité routière étudie l'accidentologie des passagers des véhicules de tourisme, âgés de 18 à 65 ans. Le tableau suivant indique le nombre de tués par tranches d'âge en 2005 :

Âge	Effectif
[18 ; 25[790
[25 ; 35[545
[35 ; 45[377
[45 ; 65[606

Source : ONISR, 2006

- Déterminez :
 - la population étudiée ;
 - la variable étudiée.
- Précisez :
 - la nature de la variable ;
 - les modalités de la variable.
- Dessinez l'histogramme de la distribution.

Solution

- La population étudiée est composée des passagers des véhicules de tourisme, âgés de 18 à 65 ans.
 - La variable étudiée est $X = \text{« âge des tués »}$.
- La variable étudiée est quantitative continue.
 - Les modalités de la variable étudiée sont les quatre classes suivantes : [18 ; 25[; [25 ; 35[; [35 ; 45[; [45 ; 65[.
- Nous calculons les amplitudes de classes (A_i), soit :
 $A_1 = 25 - 18 = 7$; $A_2 = 35 - 25 = 10$; $A_3 = 45 - 35 = 10$; $A_4 = 65 - 45 = 20$. Puisqu'elles sont différentes, il est nécessaire d'utiliser les densités pour réaliser l'histogramme.
Avec la calculatrice **TI 84**, conformément à la figure 1.17, saisissez les effectifs (n_i) dans la colonne L1 et les amplitudes (A_i) dans la colonne L2.
Pour calculer les densités (d_i) dans la colonne L3, placez le curseur sur l'en-tête de colonne L3. Indiquez $L3=L1\div L2$. Puis appuyez sur **ENTER**. La colonne L3 fait alors apparaître les densités (voir figure 1.17).
Les effectifs corrigés ($n_{i,c}$) sont obtenus en multipliant ces densités par l'effectif minimal, soit 7. Pour calculer les effectifs corrigés ($n_{i,c}$) dans la colonne L4, placez le cur-

seur sur l'en-tête de colonne L4. Indiquez $L4=L3 \times 7$. Puis appuyez sur **ENTER**. La colonne L4 fait alors apparaître les effectifs corrigés.

Avec la calculatrice **GRAPH 75**, saisissez les effectifs (n_i) dans List1 et les amplitudes (A_i) dans List2, placez le curseur sur l'en-tête de List3, appuyez sur **OPTN** pour accéder au menu List, et tapez la séquence : **F1****F1****1****÷****F1****2**, soit $List3=List1 \div List2$, puis validez avec la touche **EXE**. Placez le curseur sur l'en-tête de List4 et tapez la séquence : **F1****3****×****7****EXE**. List4 fait apparaître les effectifs corrigés (voir figure 1.18).

Figure 1.17
Calcul des densités avec la TI 84.

L1	L2	L3	3
790	7	112.85	
545	10	54.5	
377	10	37.7	
606	20	30.3	

L3(1)=112.8571428...			

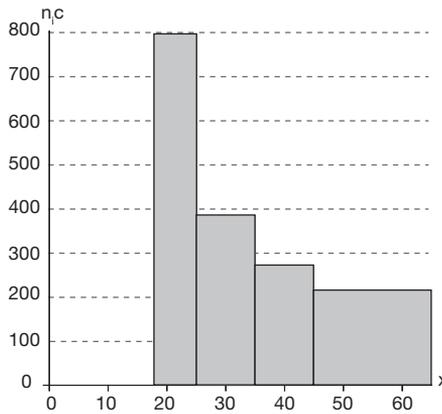
Figure 1.18
Calcul des effectifs corrigés avec la GRAPH 75.

	List 1	List 2	List 3	List 4
SUB				
1	790	7	112.85	790
2	545	10	54.5	381.5
3	377	10	37.7	263.9
4	606	20	30.3	212.1
				790

LIST EPLW CALC HYP PROB

L'histogramme des effectifs est ensuite tracé sur une feuille de papier millimétré (voir figure 1.19).

Figure 1.19
Histogramme des tués par tranches d'âge.



Exercice 4 : Discrétisation des données

L'Agence de l'environnement et de la maîtrise de l'énergie (ADEME) vous informe sur les émissions de CO₂ par habitant dans le monde en 2002 :

Pays	Émissions de CO ₂ (en tCO ₂ par habitant)
Asie du sud	0,82
Afrique	1,39
Amérique Latine	2,79
Chine	3,05
Europe centrale	5,68
CEI	5,97
Moyen Orient	6,04
Europe de l'Ouest	8,28
Japon	9,14
ASIE (NPI)	10,46
Australasie	12,2
Amérique du Nord	20,02

Source : ADEME, 2002

L'ADEME souhaite distinguer trois classes de pays, selon leur niveau d'émissions de CO₂ :

- ceux qui émettent moins de 6 tonnes par habitant ;
- ceux qui émettent de 6 à moins de 10 tonnes par habitant ;
- ceux qui émettent de 10 à moins de 22 tonnes par habitant.

1. Déterminez :
 - a. la population étudiée ;
 - b. la variable étudiée.
2. Précisez :
 - a. la nature de la variable ;
 - b. les modalités de la variable.
3. Construisez le tableau statistique associé. Pour cela, discrétisez le caractère étudié selon la classification souhaitée par l'ADEME.
4. Dessinez l'histogramme de la distribution.

Solution

1. a. La population étudiée est composée des régions du monde énumérées.
b. La variable étudiée est X = « émissions de CO₂ ».

2. a. La variable étudiée est quantitative continue.
- b. L'ensemble des modalités de la variable étudiée est $M = \{5 ; 1,39 ; 2,79 ; 3,05 ; 5,68 ; 5,97 ; 6,04 ; 8,28 ; 9,14 ; 10,46 ; 12,2 ; 20,02\}$.
3. Le tableau statistique associé est composé de deux colonnes :
 - La première colonne comporte les classes d'émission de CO_2 ,
 - La seconde colonne comporte les effectifs n_i affectés à chacune de ces classes.

Le tableau statistique associé à X est le suivant.

Émissions de CO_2	n_i
$[0 ; 6[$	6
$[6 ; 10[$	3
$[10 ; 22[$	3

4. Nous calculons ensuite les amplitudes de classes (A_i), soit :

$$A_1 = 6 - 0 = 6 ; A_2 = 10 - 6 = 4 ; A_3 = 22 - 10 = 12.$$

Avec la **TI 84**, saisissez les effectifs (n_i) dans la colonne L1 et les amplitudes (A_i) dans la colonne L2.

Pour calculer les densités (d_i) dans la colonne L3, placez le curseur sur l'en-tête de colonne L3. Indiquez $L3=L1\div L2$. Puis appuyez sur **ENTER**. La colonne L3 fait alors apparaître les densités.

Les effectifs corrigés ($n_{i,c}$) sont obtenus en multipliant ces densités par l'effectif minimal, soit 4. Pour calculer les effectifs corrigés ($n_{i,c}$) dans la colonne L4, placez le curseur sur l'en-tête de colonne L4. Indiquez $L4=L3 \times 4$. Puis appuyez sur **ENTER**. La colonne L4 fait alors apparaître les effectifs corrigés (voir figure 1.21).

Avec la **GRAPH 75**, saisissez les effectifs (n_i) dans List1 et les amplitudes (A_i) dans List2, placez le curseur sur l'en-tête de List3, appuyez sur **OPTN** pour accéder au menu List, et tapez la séquence : **F1****F1****1****÷****F1****2**, soit $List3=List1\div List2$, puis validez avec la touche **EXE** (voir figure 1.20). Placez le curseur sur l'en-tête de List4 et tapez la séquence : **F1****3****×****4****EXE**. List4 fait apparaître les effectifs corrigés.

Figure 1.20

Calcul des densités avec la GRAPH 75.

	List 1	List 2	List 3	List 4
SUB				
1	6	6		
2	3	4		
3	3	12		
4				

List 1 ÷ List 2
List L→M Dim Fill Seq ▾

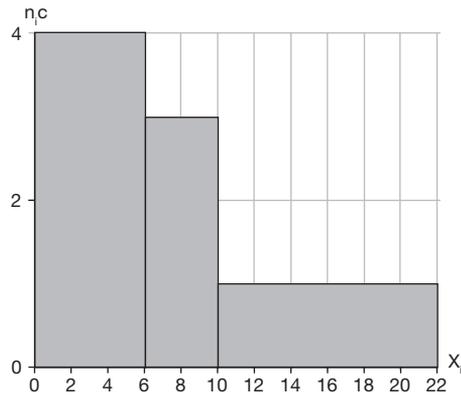
Figure 1.21

Calcul des effectifs corrigés avec la TI 84.

L1	L2	L3	3
6	6	.75	
3	4	.25	
3	12	.25	
-----	-----	-----	
L3(1)=1			

L'histogramme des effectifs est ensuite tracé sur une feuille de papier millimétré (voir figure 1.22).

Figure 1.22
Histogramme des pays selon leurs émissions de CO₂.



Exercice 5 : Les polygones

L'ADEME vous transmet le tableau suivant, qui recense les individus dans le monde selon le niveau de CO₂ qu'ils émettent :

Émission moyenne de CO ₂ (tonnes CO ₂ par habitant)	Population (millions)
[0 ; 2[2 205,79
[2 ; 4[1 809,21
[4 ; 6[401,26
[6 ; 8[172,46
[8 ; 10[590,05
[10 ; 16[112,48
[16 ; 22[319,84

Source : ADEME, 2002

1. Sur un même graphique :
 - a. Dessinez l'histogramme des fréquences de la distribution.
 - b. Dessinez le polygone des fréquences de la distribution.
2. À la suite de la question précédente :
 - a. Calculez les fréquences cumulées croissantes et décroissantes.
 - b. Représentez les polygones des fréquences cumulées croissantes et décroissantes sur un même graphique.

Solution

1. a. Une simple lecture du tableau permet de voir que les amplitudes de classes ne sont pas constantes, ce qui est confirmé par leur calcul en colonne C (voir figure 1.23). Les fréquences sont calculées en colonne D, puis les densités (d_i) en colonne E, en effectuant le rapport des fréquences sur les amplitudes. Enfin, les fréquences corrigées ($f_{i,c}$) sont obtenues en colonne F en multipliant ces densités par l'effectif minimal.

Figure 1.23

Résultats sous Excel.

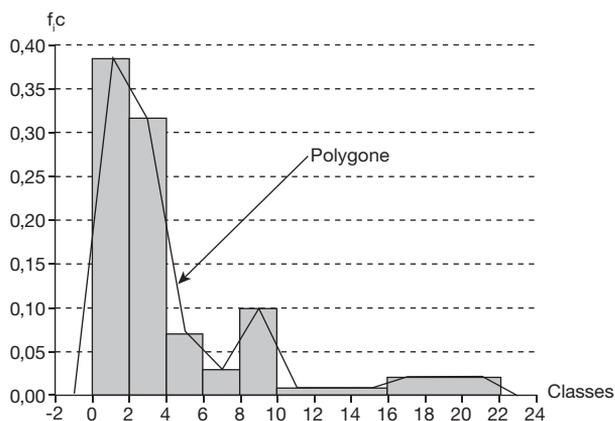
	A	B	C	D	E	F	G	H
1	Emissions de CO ₂	n_i	a_i	f_i	d_i	$f_{i,c}$	$f_{i,cc}$	$f_{i,cd}$
2	[0,2[2 205,79	2	0,3931	0,1966	0,3931	0,39	1,00
3	[2,4[1 809,21	2	0,3224	0,1612	0,3224	0,72	0,61
4	[4,6[401,26	2	0,0715	0,0368	0,0715	0,79	0,28
5	[6,8[172,46	2	0,0307	0,0154	0,0307	0,82	0,21
6	[8,10[590,05	2	0,1052	0,0526	0,1052	0,92	0,18
7	[10,16[112,48	6	0,0200	0,0033	0,0067	0,94	0,08
8	[16,22[319,84	6	0,0570	0,0095	0,0190	1,00	0,06

À partir de ces fréquences corrigées, il est possible de tracer l'histogramme des fréquences sur une feuille de papier millimétré (voir figure 1.24).

- b. Les classes sont d'amplitudes inégales. On procède à un découpage artificiel en prenant l'amplitude minimale, soit 2, pour unité d'amplitude. Le polygone des fréquences est alors obtenu en joignant à la règle les milieux des bases supérieures des rectangles du découpage précédent (voir figure 1.24).

Figure 1.24

Histogramme et polygone des fréquences des pays selon leurs émissions de CO₂.



2. a. À la suite du tableau Excel précédent, les fréquences cumulées croissantes ($f_{i,cc}$) sont calculées dans la colonne G et les fréquences cumulées décroissantes ($f_{i,cd}$) dans la colonne H (voir figure 1.23).

Ces calculs sont effectués selon le même principe que pour les effectifs cumulés croissants et décroissants, en remplaçant les effectifs par les fréquences.

- b. La présentation de ces résultats est légèrement modifiée pour faire apparaître dans un même tableau les fréquences cumulées croissantes et décroissantes de chacune des bornes des classes (voir figure 1.25).

Figure 1.25

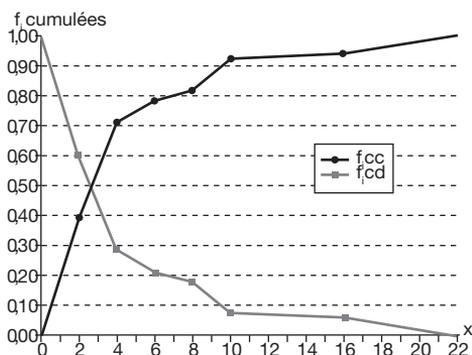
Données pour les polygones de fréquences cumulées.

34	Bornes de classes	f _{cc}	f _{cd}
35	0	0,00	1,00
36	2	0,39	0,61
37	4	0,72	0,28
38	6	0,79	0,21
39	8	0,82	0,18
40	10	0,92	0,08
41	16	0,94	0,06
42	22	1,00	0,00

Les courbes des fréquences cumulées croissantes et décroissantes de la figure 1.26 sont réalisées à partir de ce dernier tableau.

Figure 1.26

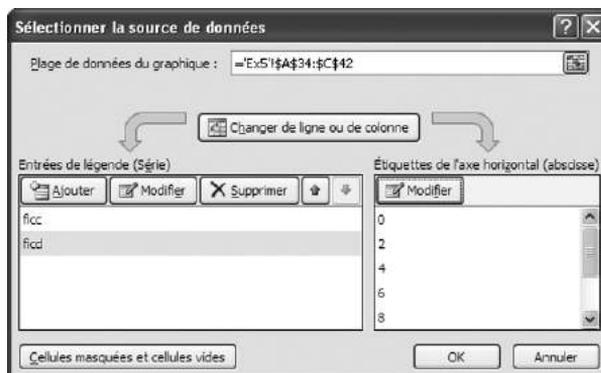
Polygones des fréquences cumulées croissantes et décroissantes des pays selon leurs émissions de CO₂.



Pour réaliser ces courbes des effectifs cumulés sous Excel, sélectionnez, en mettant en surbrillance, les deux colonnes contenant les modalités n_i et les effectifs x_i , soit la plage A34:C42 de la figure 1.25. Puis cliquez sur Insertion/Nuages de points dans la barre de menus. Sélectionnez le nuage de points avec courbes droites et marqueurs (en bas à gauche). Le graphique s'affiche alors immédiatement. En sélectionnant la bonne plage de données dès le départ, Excel a automatiquement rempli l'assistant graphique adéquat, comme à la figure 1.27. Vous pouvez retrouver cette boîte de dialogue en sélectionnant le graphique et en cliquant sur Création/Sélectionner des données.

Figure 1.27

Données saisies dans l'assistant graphique.



Annexe 1.1 Présentation du tableur (Excel 2003, 2007)

Quand vous ouvrez Excel, la zone de travail située au centre s'appelle le CLASSEUR.

La BARRE DE TITRE de la fenêtre affiche le nom du classeur – par exemple, « Classeur1 » – que vous devez renommer et enregistrer.

Un classeur comporte par défaut trois feuilles, dont le nom figure sur un ONGLET – par exemple, « Feuil2 ». Il est possible de renommer, d'insérer ou de supprimer une feuille en faisant un clic droit sur un des onglets et en choisissant Insérer, Supprimer ou Renommer dans le menu.

L'intersection d'une ligne et d'une colonne s'appelle une CELLULE. Une cellule est caractérisée par sa RÉFÉRENCE, colonne-ligne – par exemple, « B4 ».

La BARRE DE MENUS permet d'accéder aux différents menus déroulants : Accueil, Insertion, Mise en page, Formules, Données, Révision, Affichage, Développeur, Compléments.

Sous la barre de menus se trouvent les BARRES D'OUTILS, accessibles uniquement avec la souris. Lorsqu'on pointe sans cliquer sur les différents boutons, une info-bulle affiche le nom du bouton et sa fonction.

Sous les barres d'outils se trouve la BARRE DE FORMULE. Dans sa partie gauche apparaît la référence de la cellule active et dans la partie droite apparaissent les données, lors de leur saisie. Entre les deux, le symbole f_x (Insérer une fonction) désigne l'assistant fonction. Il comprend toutes sortes de fonctions, notamment statistiques, et sera extrêmement précieux pour les problèmes et exercices.

Pour saisir des données dans une cellule, placez la souris dessus, cliquez et entrez les chiffres ou les lettres voulus. Passez d'une cellule à une autre grâce à la souris ou aux touches , ,  et  du clavier.

Pour effectuer une opération mathématique, cliquez sur une cellule, tapez le signe  pour indiquer qu'il s'agit d'une formule de calcul, puis faites l'opération en utilisant les signes mathématiques du clavier : , ,  et . Par exemple, pour additionner une cellule à une autre, cliquez sur la cellule qui doit accueillir le résultat, tapez , cliquez sur la première cellule, tapez  puis cliquez sur la seconde cellule à additionner. Validez avec **ENTRÉE** pour faire apparaître le résultat. L'élévation à la puissance s'obtient en appuyant sur la touche accent grave, , suivie du nombre de la puissance désirée, ou en utilisant la fonction Puissance de l'assistant fonction.

Il existe trois types de références de cellules : pour passer d'un type à l'autre, utilisez la touche , qui procède par permutation circulaire, comme le montre cet exemple : saisissez  dans la cellule A2, placez le curseur de la souris à la suite de A1, contre le 1, et appuyez sur . Vous voyez alors apparaître : \$A\$1 (référence absolue). Si vous appuyez de nouveau sur , vous voyez apparaître successivement : A\$1, \$A1 (références mixtes) et enfin A1 (référence relative).

Références relatives : par défaut, sous Excel, les références des cellules sont « relatives ». Lorsqu'on recopie une formule d'une cellule à une autre, elle s'adapte automatiquement en fonction du déplacement en ligne ou en colonne.

- Si la formule $=B2 + B3$ est saisie en B4 puis recopiée en C4, elle devient $=C2 + C3$.
- Si la formule $=B2 + C2$ est saisie en D2 puis recopiée en D3, elle devient $=B3 + C3$.
- Si la formule $=B2 + C2$ est saisie en D2 puis recopiée en E3, elle devient $=C3 + D3$.

Références absolues : on peut figer la colonne et la ligne d'une cellule, en mettant le signe « \$ » devant la lettre de la colonne et devant le nombre de la ligne, afin que la cellule concernée reste identique en cas de recopie d'une formule. Cette cellule est alors définie par une référence absolue dans la formule.

- Si la formule $=B2 + \$C\2 est saisie en D2 puis recopiée en E3, elle devient $=C3 + \$C\2 .

Références mixtes : on peut aussi décider de ne figer que la colonne ou que la ligne d'une cellule, en positionnant le symbole « \$ » uniquement devant la lettre ou le nombre de la cellule. La cellule est alors définie par une référence mixte.

Si la formule $=B2 + \$C2$ est saisie en D2 puis recopiée en E3, elle devient $=C3 + \$C2$.

- Si la formule $=B2 + C\$2$ est saisie en D2 puis recopiée en E3, elle devient $=C3 + D\$2$.

La notion de fonction : Excel comporte des fonctions intégrées, identifiées par des noms de fonctions – par exemple, SOMME, PRODUIT, MOYENNE, RACINE... Les éléments sur lesquels porte la fonction sont appelés ARGUMENTS, se placent entre parenthèses et sont séparés par des points-virgules.

Pour utiliser une fonction : placez le curseur dans la cellule où vous souhaitez faire apparaître le résultat. Cliquez sur Insertion/Fonction (ou utilisez directement f_x), sélectionnez la catégorie de fonction souhaitée (dans cet ouvrage, Statistique ou Math & Trigo), puis la fonction désirée. Entrez les arguments en vous laissant guider par la boîte de dialogue Excel. Validez en cliquant sur **OK**.

Remarque : pour faire une somme, il est possible de se servir de l'icône Σ proposée par défaut dans la barre d'outils du menu Accueil. Cliquez sur la cellule où vous voulez faire apparaître la somme, cliquez sur l'icône Σ , puis sélectionnez les cellules dont vous souhaitez faire la somme, et validez avec **ENTRÉE**.

Les fonctions statistiques seront explorées lors de la correction des exercices.

Dans **Excel 2007**, la partie haute s'appelle le Ruban, et le Bouton Office situé en haut à gauche permet d'accéder aux fonctions de base (enregistrer, imprimer, etc.), mais aussi aux options et compléments. Pour accéder aux options, cliquez sur le Bouton Office, puis sur Options Excel.

Annexe 1.2 Présentation de deux modèles de calculatrice : GRAPH 75 (Casio) et TI 84 (Texas Instrument)

La calculatrice GRAPH 75 Casio

On accède au menu STAT par la touche **MENU** ; les colonnes sont notées List1, List2, etc.

Pour éditer un tableau : appuyez sur la touche **MENU**, et avec le pavé directionnel, sélectionnez STAT, puis validez avec la touche **EXE**. Saisissez alors les données et validez chacune d'elles par la touche **EXE**. Pour changer de liste, il suffit de se déplacer avec le pavé directionnel.

Pour effacer une colonne entière : placez le curseur sur l'en-tête de la liste (par exemple List1) que vous souhaitez effacer, appuyez sur la touche **F6**, puis sélectionnez DEL-A. Enfin, à la question « Delete List ? », répondez Yes en appuyant sur la touche **F1**. Pour accéder au menu LIST, tapez la séquence : **OPTN****F1**, puis faites défiler les fonctions avec la touche **F6**.

Pour effectuer la somme des termes d'une colonne, par exemple de List1, et l'afficher dans List2, placez le curseur dans la première cellule de List2, appuyez sur la touche **F6** pour voir apparaître la fonction Sum et tapez la séquence : **F1****OPTN****F1****F1****I**, soit SUM List1. Validez en appuyant sur la touche **EXE**, et la somme de List1 s'affichera dans List2.

Pour calculer les effectifs cumulés croissants d'une série, entrez les effectifs, par exemple dans List1, et placez le curseur sur l'en-tête de List2, pour voir afficher les effectifs cumulés croissants dans List2. Tapez alors la séquence : **OPTN****F1**, puis faites défiler les fonctions avec la touche **F6** pour faire apparaître la fonction Cuml. Tapez alors : **F3****OPTN****F1****F1****I**, soit Cuml List1. Validez en appuyant sur la touche **EXE** et la List2 affichera les effectifs cumulés croissants.

La calculatrice TI 84 Texas Instrument

Notations : les colonnes sont notées L1, L2, L3, L4, L5, L6. Les cellules sont identifiées par leur colonne, suivie de leur ligne entre parenthèses – par exemple, L1(2) indique la cellule figurant dans la première colonne, à la deuxième ligne.

Pour saisir un tableau : appuyez sur la touche **STAT**. Éditez le tableau en appuyant sur la touche **I**. Saisissez les données (validez chacune par la touche **ENTER**) en vous déplaçant avec le curseur.

Pour quitter l'éditeur de tableau : appelez la fonction **QUIT** par l'appui successif sur les touches **2ND** et **MODE**.

Pour effacer une colonne entière : placez le curseur sur l'en-tête de colonne Li que vous souhaitez effacer. Appuyez sur les touches **CLEAR** et **ENTER**.

Pour effectuer la somme des termes d'une colonne : placez le curseur dans la cellule (1) où vous souhaitez faire apparaître la somme. Appuyez sur les touches **2ND** et **LIST**, puis,

dans le menu MATH, appelez la fonction sum(. Indiquez la colonne Lj dont vous souhaitez faire la somme (par exemple, L1 est obtenu par $\boxed{2\text{ND}}$ et $\boxed{1}$) et validez avec $\boxed{\text{ENTER}}$

Pour effectuer la somme cumulée d'une colonne : placez le curseur sur l'en-tête de colonne Li dans laquelle vous souhaitez obtenir les effectifs cumulés. Appuyez sur les touches $\boxed{2\text{ND}}$ et $\boxed{\text{LIST}}$, puis, dans le menu OPS, appelez la fonction cumSum(. Indiquez la colonne Lj dont vous souhaitez faire la somme cumulée et validez avec $\boxed{\text{ENTER}}$

Bibliographie

BOLL M., *L'exploitation du hasard*, Que sais-je ?, PUF, 1947.

CALOT G., *Cours de statistique descriptive*, Dunod, Paris, 1969.

CHAREILLE P. et PINAULT Y., *Statistique descriptive*, Collection AES, Montchrestien, Paris, 1996.

DODGE Y., *Premiers pas en statistique*, Springer, 2005.

DROESBEKE J.-J., *Éléments de statistiques*, Éditions de l'université de Bruxelles, Ellipses, 2001.

LE BRAS H., *Naissance de la mortalité. L'origine politique de la statistique et de la démographie*, Gallimard/Le Seuil, Paris, 2000.

LIORZOU A., *Initiation à la pratique statistique*, Eyrolles, 1985.

PIATIER A., *Statistique descriptive et initiation à l'analyse*, Thémis, PUF, 1962.

ROGER P., *Probabilités, statistique et processus stochastiques*, Collection Synthex, Pearson Education, 2004.

ROHRBASSER J.-M. et VERON J., *Leibniz et les raisonnements sur la vie humaine*, INED, Paris, 2001.

VESSEREAU A., *La statistique*, Que sais-je ?, PUF, 1962.