

# DISTRIBUTIONS DE FRÉQUENCES ET TABLEAUX

- LE BUT DES DISTRIBUTIONS DE FRÉQUENCES
- ÉTAPES DANS L'ÉLABORATION D'UNE DISTRIBUTION DE FRÉQUENCES
  - LIMITES DE CLASSES RÉELLES • INTERVALLES DE CLASSES
  - CENTRE DE CLASSE • DISTRIBUTIONS DE FRÉQUENCES QUANTITATIVES
- GROUPEZ LES DONNÉES • CAS NON DÉCLARÉS OU NON MENTIONNÉS • PARTIES D'UN TABLEAU • ARRONDIS
- CALCUL DES POURCENTAGES



À LA FIN DE CE CHAPITRE VOUS DEVEZ ÊTRE CAPABLE DE :

- CONSTRUIRE UNE DISTRIBUTION DE FRÉQUENCES
- CONSTRUIRE UN TABLEAU STATISTIQUE
- ARRONDIR DES DONNÉES ET CALCULER DES POURCENTAGES

## CHAPITRE 3

# DISTRIBUTIONS DE FRÉQUENCES ET TABLEAUX

### Le but des distributions de fréquences

#### *le besoin de résumer les observations pour leur donner un sens*

Lorsqu'on entreprend une enquête statistique on se retrouve avec des séries d'observations de variables d'un certain nombre d'unités statistiques. En général, on observe à la fois les variables quantitatives et qualitatives de chaque unité. À partir d'une série d'observations on veut généralement établir un résumé de manière à leur donner un sens. On peut vouloir faire une estimation de la population entière des unités pour lesquelles notre groupe constitue un échantillon, ou bien on peut y voir simplement un moyen pratique de résumer les données de base pour réduire la quantité des informations de base à un niveau plus maniable.

#### *les distributions de fréquences résument les données*

On utilise fréquemment les distributions de fréquences comme moyen de résumer des données. On obtient une distribution de fréquences en divisant la série de valeurs observées en un certain nombre de classes (c'est-à-dire des groupes, ou des tranches de valeurs qui peuvent être attribuées aux observations), en répartissant les observations dans ces classes et en déterminant la fréquence de classe pour chaque classe (c'est-à-dire, le nombre d'observations dans chaque classe).

Une distribution de fréquences aide à comprendre la nature des données et nous indique la manière dont les données sont réparties dans la fourchette de valeurs possibles (c'est-à-dire la manière dont les données sont 'distribuées'), elle met facilement en évidence les classes qui ont des fréquences élevées ou des fréquences basses. La distribution de fréquences est complètement adaptée à certains cas et donne toutes les informations nécessaires à partir du jeu original des observations.

*Une distribution de fréquences est le groupement de données en catégories qui montrent le nombre d'observations dans chaque catégorie.*

#### *variables qualitatives*

Dans le cas des variables qualitatives, la distribution de fréquences est la fréquence des unités entrant dans les catégories de la variable quand les données auront été mesurées avec des échelles nominales ou ordinales. On trouve des exemples de ces fréquences chez les personnes classées par situation de famille, les personnes classées par activité économique, les ménages classés par type de logement et les entreprises classées par secteur industriel.

#### *données continues*

Les distributions de fréquences de données continues ont des intervalles de classe comme catégories. Ces intervalles de classe contiennent une fourchette de valeurs plutôt qu'une valeur simple, de manière à simplifier et résumer les données.

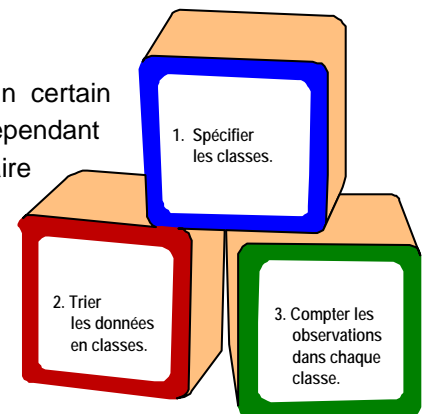
## Étapes de l'élaboration d'une distribution de fréquences

### 3 grandes étapes

Les trois grandes étapes qui forment l'élaboration d'une distribution de fréquences sont :

- 1 Spécifier les classes dans lesquelles les données seront groupées.
- 2 Trier les données en classes.
- 3 Compter le nombre d'observations dans chaque classe.

L'étape la plus compliquée est celle de la spécification des classes. Un certain nombre d'indications sont données pour aider à les spécifier. Il faut cependant souligner que ce sont seulement des conseils et que l'on devrait toujours faire preuve de bon sens.



### Spécifier les classes

#### données qualitatives

Si on travaille avec des données qualitatives, elles seront probablement déjà triées en classes, sur la base des catégories de réponses. Parfois, quand on a affaire à un grand nombre de catégories de réponses, on peut grouper ou combiner les observations. Par exemple, quand on fait des tableaux sur le nombre d'employés dans différents emplois, il est souvent utile de grouper les catégories d'emplois pour que le nombre de classes de professions soit maniable.

#### données quantitatives

Quand on travaille avec des données quantitatives, on peut passer par un grand nombre d'étapes avant de pouvoir spécifier les classes dans lesquelles les données seront groupées. On doit tenir compte de deux choses :

- 1 le nombre de classes nécessaires
- 2 la taille de chaque classe (la fourchette de valeurs dans chaque classe).

### NOMBRE DE CLASSES

#### *pas moins de 5 et pas plus de 20*

Quand on décide du nombre de classes le principe général est que *pas moins de 5 classes et pas plus de 20 classes* doivent être utilisées dans l'élaboration d'une distribution de fréquences. Trop peu de classes, ou trop de classes utilisées donnent peu d'informations sur la distribution des données.

#### *grand nombre d'observations*

On détermine également le nombre de classes par le nombre total d'observations, leur concentration et leur étendue. On peut préférer un grand nombre de classes quand le nombre d'observations est important. Si on a des valeurs extrêmes isolées, on essaye de ne pas leur laisser dicter le nombre de classes. On doit garder en mémoire que le but des distributions de fréquences est de résumer des données, pas de donner des détails comme les valeurs extrêmes. Si on a trop de classes, on peut aussi bien étudier les données d'origine, la distribution de fréquences ne donnera pas un résumé efficace.

**petit nombre de classes**

Si on a trop peu de classes, on cache de réelles différences entre les observations et on cause une perte substantielle d'information. Bien souvent on décide en même temps du nombre de classes et de leur taille.

**étendue des données**

La première chose à connaître est l'envergure des données — combien d'observations différentes notre distribution de fréquences doit-elle comporter ? On appelle l'envergure des données *l'étendue*, qui est la différence entre la plus petite et la plus grande des observations. L'étendue est alors utilisée pour spécifier les classes de manière à ce qu'elles divisent l'étendue (en général de manière homogène) des données.

**étendue**

**étendue** = la plus grande des observations - la plus petite des observations

**l'exemple des poissons**

Il est sûrement plus facile de comprendre comment définir le nombre de classes en prenant un exemple. L'exemple montrera aussi la relation entre le nombre de classes et la taille des classes — que nous couvrirons ensuite. On a choisi des données fictives de poissons pêchés dans une région donnée.

Tableau 3.1 Poids de 63 poissons (en kilos)

4,6	3,9	2,8	6,6	4,2	3,7	3,7	5,9
3,2	2,2	3,2	4,1	3,1	3,0	4,8	4,1
<u>2,1</u>	4,2	5,0	4,6	5,4	2,4	6,3	2,9
5,3	4,0	4,7	3,6	3,3	6,9	4,5	2,5
5,4	5,7	3,8	4,1	<u>7,9</u>	6,2	3,0	3,3
5,0	5,4	3,4	4,4	4,0	3,6	5,0	4,1
4,8	7,2	6,4	3,0	3,5	5,3	7,7	3,9
2,6	5,6	3,3	5,5	4,3	3,9	6,3	

Source : fictive

**taille de la classe**

La taille de la classe est la différence entre la valeur la plus basse de la classe et la valeur la plus basse de la classe suivante. En général on utilise les intervalles multiples de 5 ou 10, comme 5, 10, 20, 100 ou 1000.

On essaye aussi de garder des classes de taille identique, ce qui rend plus facile la compréhension et l'analyse des données. Quelquefois il vaut mieux ne pas avoir des classes de tailles identiques, au risque de se retrouver avec trop de classes, ou trop peu. En pratique, on détermine le nombre de classes et leur taille au même moment. D'abord on détermine le nombre de classes, et ceci sert de guide ou indique la taille des classes.

**nombre de classes du tableau 3.1**

Au tableau 3.1, la valeur la plus haute est 7,9 (souligné), la plus basse 2,1 (souligné). Donc l'étendue est :

$$\begin{aligned} \text{Étendue} &= 7,9 - 2,1 \\ &= 5,8 \text{ kg} \end{aligned}$$

L'étendue indique qu'on a un intervalle de 5,8 kilos en poids de la plus petite à la plus grande observation. Dans ce cas, l'étendue de 5,8 kg suggère 6 classes parce qu'on préfère avoir des intervalles de classes égaux.

#### **taille des intervalles de classes**

La taille des classes est estimée en divisant l'étendue par le nombre de classes :

$$\text{Taille approximative d'un intervalle de classe} = \frac{\text{étendue}}{\text{nombre de classes}}$$

#### **regardez le nombre de classes**

Ceci ne veut pas dire que la taille d'un intervalle de classe doit être ce chiffre, mais il donne une indication de la taille appropriée pour l'intervalle. Rappelez-vous que dans la pratique on détermine le nombre de classes et leur taille en même temps.

Pour les données du tableau 3,1, on a pensé que 6 classes seraient adéquates, ce qui donnerait un intervalle de classe approximatif de  $\frac{5,8}{6} = 0,97$  kg. Ce chiffre suggère un intervalle de 1 kg, et toutes les observations peuvent être couvertes par 6 classes.

## **LIMITES DES CLASSES**

#### **limites des classes**

On doit étudier ensuite les limites inférieure et supérieure des classes. On utilise la plus petite valeur et la plus grande pour spécifier les classes. Elles sont appelées les **limites de classe**. On a défini un intervalle de 1 kg pour les données du tableau 3,1. On a donc les classes :

2,0–3,0 kg  
3,0–4,0 kg  
4,0–5,0 kg  
etc.

Néanmoins, le problème avec de telles classes est qu'elles ne sont pas **mutuellement exclusives**. Dans quelle classe allez-vous placer un poisson de 3,0 kg ? Dans la classe 2,0–3,0 kg ou dans la classe 3,0–4,0 kg ? Pour résoudre ce problème on modifie les classes :

Tableau 3.2 Limites de classes des données sur les poissons (kg)

Classes	Limites de classes	Borne inférieure	Borne supérieure
2,0–2,9	2,0 & 2,9	2,0	2,9
3,0–3,9	3,0 & 3,9	3,0	3,9
4,0–4,9	4,0 & 4,9	4,0	4,9
5,0–5,9	5,0 & 5,9	5,0	5,9
6,0–6,9	6,0 & 6,9	6,0	6,9
7,0–7,9	7,0 & 7,9	7,0	7,9

Source : Tableau 3.1

Les limites de classes définissent l'étendue des valeurs incluses dans chaque classe après que les observations ont été enregistrées. Les bornes supérieure et inférieure définissent la plus petite et la plus grande des valeurs qui peuvent être incluses dans la classe. La classe 2,0–2,9 du tableau 3.2 comprend 10 valeurs : 2,0; 2,1; 2,2; 2,3; 2,4; 2,5; 2,6; 2,7; 2,8 et 2,9. Les 5 autres classes ont aussi 10 valeurs.

## DISTRIBUTION DE FRÉQUENCES

### *distribution de fréquences*

Maintenant on peut élaborer une distribution de fréquences à partir des données du tableau 3.1 avec 6 classes de 1 kg :

Tableau 3.3 Distribution de fréquences des données sur les poissons (kg)

Poids (kg)	Pointage	Fréquence
2,0–2,9	### //	7
3,0–3,9	### ### ### ////	19
4,0–4,9	### ### ### /	16
5,0–5,9	### ### //	12
6,0–6,9	### /	6
7,0–7,9	///	3
<b>Total</b>		<b>63</b>

Source : Tableau 3.1

### *autres critères des distributions de fréquences*

En plus de ces critères de base, on utilise quelques autres principes lors de l'élaboration des distributions de fréquences. Ceux-ci se rapportent au rythme auquel les observations se produisent dans les classes.

## RYTHME DES OBSERVATIONS

### *une fois, et une fois seulement*

On doit considérer deux autres critères quand on spécifie les classes :

- 1 Aucune observation ne peut être incluse dans plus d'une classe
- 2 Chaque observation possible doit être incluse dans une des classes.

### *exemple*

Il est particulièrement important que l'étendue de chaque classe soit définie pour que chaque observation puisse être classée dans un intervalle, et un seulement. On a vu à l'exemple des poissons que des classes qui semblent être a priori des classes raisonnables (2,0–3,0; 3,0–4,0; etc.) finissent par embrouiller. Voici un autre exemple, l'élaboration d'une distribution de fréquences pour les tailles d'un groupe d'étudiants. On peut avoir :

### *classes incorrectes*

1,00–1,10 mètres  
 1,10–1,20 mètres  
 1,20–1,30 mètres  
 1,30–1,40 mètres  
 et ainsi de suite.

### *compte double*

Un étudiant qui mesure exactement 1,70 mètres peut être classé dans deux classes différentes. Voici un bien meilleur jeu de classes :

### *classes correctes*

1,00 à moins de 1,10 mètres  
 1,10 à moins de 1,20 mètres  
 1,20 à moins de 1,30 mètres  
 1,30 à moins de 1,40 mètres  
 et ainsi de suite.

Dans ce cas il n'y a pas d'ambiguïté et une observation de 1,70 mètres ne peut entrer que dans une seule classe.

## Limites de classes réelles

### *données continues*

Les limites de classe définissent l'étendue des valeurs incluses dans chaque classe après que les observations ont été enregistrées. Pour les **données continues**, il peut y avoir des écarts entre la borne supérieure d'une classe et la borne inférieure de la classe suivante.

Rappelez-vous des classes établies au tableau 3.2. Rappelez-vous aussi qu'on essayait d'élaborer des classes d'une taille de 1 kg. Mais la différence entre les bornes inférieure et supérieure est de 0,9 kg.

Est-ce qu'on a fait une erreur ?

Prenez un poisson de 2,94 kg. Dans quelle classe est-il inclus ? Est-ce qu'il est hors critère selon lequel "chaque observation doit être incluse dans une des classes" ? Vous remarquerez qu'au tableau 3.1 toutes les informations ont été enregistrées en dixièmes de kilogrammes. On peut raisonnablement supposer que tous les poids des poisson ont été enregistrés au dixième de kilo près (après avoir été arrondis), alors un poisson qui a pesé réellement 2,94 kg sera enregistré sous le poids 2,9 kg, et donc inclus dans la première classe. En fait tous les poissons qui pèsent entre 2,90 kg et 2,95 sont arrondis et enregistrés sous 2,9 kg, et inclus dans la première classe. de même, tous les poissons pesant entre 2,95 kg et 3,00 kg sont arrondis et enregistrés sous 2,30 kg, et inclus dans la seconde classe.

#### **limites de classe réelles**

Ceci amène l'idée de **limites de classes réelles**. La limite de classe réelle spécifie l'étendue réelle des valeurs incluses dans la classe avant qu'on les ait arrondies. Pour les poissons du tableau 3.1, les limites de classes se définissent comme suit :

#### **limites de classes**

2,0–2,9 kg  
 3,0–3,9 kg  
 4,0–4,9 kg  
 5,0–5,9 kg  
 6,0–6,9 kg  
 7,0–7,9 kg

#### **limites réelles**

Les limites de classes réelles sont :

Tableau 3.4 Limites de classes et limites de classes réelles pour les données de poisson (kg)

Classe	Limites de classes	Borne inférieure réelle	Borne supérieure réelle
2,0–2,9	2,0–2,9	1,95	2,95
3,0–3,9	3,0–3,9	2,95	3,95
4,0–4,9	4,0–4,9	3,95	4,95
5,0–5,9	5,0–5,9	4,95	5,95
6,0–6,9	6,0–6,9	5,95	6,95
7,0–7,9	7,0–7,9	6,95	7,95

Source : Tableau 3.3

#### **données arrondies**

Remarquez que la distinction entre la limite de classe et la limite réelle de classe est importante seulement pour les données arrondies, qui apparaissent seulement dans les variables continues.

#### **données discontinues**

Pour les données discontinues les limites de classes et les limites réelles de classe seront les mêmes.



## Intervalles de classes

### *utilisez les limites réelles de classes pour trouver l'intervalle de classe*

On a défini les limites de classes et les limites de classes réelles de façon à éviter toute ambiguïté dans la manière d'attribuer des valeurs aux classes. L'**intervalle de classe** (c'est-à-dire l'étendue des valeurs dans chaque classe) est la différence entre les limites de classes réelles. Ça n'est pas nécessairement la différence entre les bornes supérieure et inférieure de la classe.

### *classes de tailles égales*

Autant que possible, les intervalles de classe devraient être les mêmes. Des intervalles de classes de même taille rendent plus facile la compréhension des distributions, et aident à tracer des diagrammes appropriés. Si on utilise des intervalles inégaux on rend difficile la comparaison entre deux fréquences de classes. Il est cependant quelquefois impossible d'éviter des intervalles de classes inégaux. Ils sont parfois nécessaires en cas de données très variables ou de confidentialité.

### *le bon sens prévaut*

N'oubliez-pas que le bon sens est de mise, bien que les grandes lignes de l'élaboration des distributions de fréquences présentées dans ce chapitre soient très utiles. Par exemple, que ferait-on si les observations du tableau 3.1 avaient une étendue de 0,1 kg à 7,9 kg ? Est-ce qu'on attribuerait des classes de même taille ?

Non — ça ne serait pas pratique. Prenez votre temps et réfléchissez aux classes que vous auriez avec cette étendue de donnée.

### *intervalles de classes ouverts*

Quand on prépare une distribution de fréquences il vaut mieux éviter les **intervalles de classes ouverts** (c'est-à-dire pas de borne inférieure ni supérieure). Néanmoins, dans de nombreux cas les très petites valeurs, ou les valeurs très élevées, qui rentrent dans les classes extrêmes, sont si différentes des autres observations de ces classes qu'on perd énormément d'observations lorsqu'on les rentre dans une classe extrême qui a une limite fixe. Dans ce cas de figure la meilleure façon de procéder sera de spécifier l'intervalle de classe ouvert, et de mettre toutes les valeurs des observations dans la classe, ou au moins d'indiquer leur valeur moyenne et l'étendue, ce qui évitera une perte d'informations pertinentes et permettra de plus amples calculs sur les données. Si, par exemple, aucune autre information que la fréquence n'est disponible pour la dernière classe ouverte, le point central et la limite supérieure devront être estimés en tenant compte des fréquences de classes précédentes.

## Centre de classe

### *définition*

Le centre de classe est le point central d'une classe, et on l'obtient en prenant la moyenne des bornes supérieure et inférieure de la classe.

### *formule*

$$\text{Centre de classe} = \frac{\text{Borne supérieure réelle} + \text{Borne inférieure réelle}}{2}$$

### *autres termes pour point central*

Dans l'exemple du tableau 3.5 ci-dessous, les centres de classe sont 2,45; 3,45; 4,45; 5,45; 6,45 et 7,45. On parle souvent des centres de classes en terme de valeur caractéristique de la classe, milieu de la classe ou valeur moyenne de la classe. Une fois que la distribution de fréquences est formée, on

suppose que toutes les données d'une classe ont la même valeur, qui est le centre de classe, à moins que la valeur moyenne de la classe soit disponible. Comme on va le voir dans les prochains chapitres, les centres de classe sont nécessaires pour résumer des données groupées.

Tableau 3.5 Distribution de fréquences des données de poissons pour un échantillon de 63 poissons

Classe	Limites de classes réelles	Centre de classe
2,0–2,9	1,95 & 2,95	2,45
3,0–3,9	2,95 & 3,95	3,45
4,0–4,9	3,95 & 4,95	4,45
5,0–5,9	4,95 & 5,95	5,45
6,0–6,9	5,95 & 6,95	6,45
7,0–7,9	6,95 & 7,95	7,45

Source : Tableau 3.2

## Distributions de fréquences quantitatives – grouper les données

### *perte d'informations pour les données quantitatives*

Remarquez qu'une fois qu'une distribution de fréquences est préparée, les identités des unités sont perdues et elles deviennent juste des points dans leurs classes respectives. De plus, pour une **variable quantitative**, la valeur réelle d'une unité est englobée avec celle des autres unités de la classe. Il y a alors une perte d'information en réduisant les données d'origine dans une distribution de fréquence, mais elle n'est pas cruciale vu la manière dont les données sont utilisées.

### *les distributions de fréquences résument les données*

La distribution de fréquences d'une variable quantitative indique immédiatement combien de valeurs, grandes, moyennes ou petites on trouve dans une population. Elle donne également au premier coup d'oeil une idée des valeurs les plus fréquentes. Elle éclaire considérablement sur la nature ou le profil de la population étudiée en faisant ressortir très clairement si les fréquences sont réparties uniformément ou sont concentrées sur des valeurs particulières.

Pour étudier des distributions de fréquences quantitatives, on a besoin de voir deux types de variables différents que nous avons couvert au chapitre 2, car les problèmes sont différents si on élabore une distribution de fréquences ou si on dessine des diagrammes.

### *variables continues et discontinues*

En général on distingue deux types de variables quantitatives, la première quand la variable peut prendre n'importe quelle valeur à l'intérieur d'une étendue donnée, et la deuxième quand la variable peut seulement prendre certaines valeurs. On appelle le premier type **variable continue** et le second **variable discontinue**. Dans la plupart des cas les variables discontinues sont des nombres entiers.

### *approximations*

En pratique, les données continues ne sont pas mesurées ou enregistrées de manière continue. On peut par exemple mesurer la taille d'une personne au millimètre près si on a un outil suffisamment précis. Avec un mètre ordinaire on peut prendre la mesure correcte au centimètre près. C'est le cas pour les données sur les poissons du tableau 3.1 où les poids des poissons sont mesurés au dixième de kilogramme près.

**données discontinues**

Avec les données discontinues on a deux possibilités. On peut définir une classe avec une valeur simple de la variable, ou bien on peut inclure une étendue de données. Les données détermineront quel type de classe sera utilisée. En voici deux exemples aux tableaux 3.6 et 3.7.

Tableau 3.6 Nombre de destinations de vacances par nombre de vacanciers

Destinations	Personnes
1	1.811
2	683
3	342
4	273
5	137
6	103
7	68
<b>Total</b>	<b>3.417</b>

Source : Cook Islands Visitor Survey 1991, Survey Report No. 13, TCSP, Table 23, p. 31.

Tableau 3.7 Revenus mensuels nets par nombre de ménages, Îles Salomon (1993 Income and Expenditure Survey)

Revenus mensuels nets	Ménages
0- 50	42.872
51-150	1.213
151-250	1.591
251-350	1.861
351-500	1.383
501-700	827
701-900	607
900 et plus	737
<b>Total</b>	<b>51.091</b>

Source : Solomon Islands Statistical Bulletin No. 18/95, Table 3.1.2, p8.

**comparer des populations différentes**

Le tableau 3.6 montre la distribution d'un certain nombre de destinations de vacances par le nombre de personnes qui visitent les Îles Cook. Dans cet exemple l'étendue des données est assez petite (1 à 7 destinations), donc on utilise le nombre réel de destinations pour définir chaque classe. Au tableau 3.7, on a la distribution des revenus mensuels nets des ménages des Îles Salomon. Dans cet exemple l'étendue des valeurs observées est large, de 0 à plus de 900, donc on utilise des classes qui contiennent une étendue de valeurs. Dans cet enquête, les revenus ont été collectés en tant que variable discontinue, c'est-à-dire des dollars entiers. Alors, pour les 1.213 ménages qui ont entre 51 dollars et 150 dollars de revenus nets mensuels, ils pourraient avoir des revenus de 51, 52, 53, 54 etc. jusqu'à 150 dollars. Des valeurs comme 51,29999 dollars, ou bien sûr toute autre valeur non entière entre 51 et 150 ne peuvent pas apparaître.

**intervalles de classes inégaux**

En général on préfère les classes à intervalles égaux, qui rendent l'interprétation et le dessin de diagrammes beaucoup plus facile. Néanmoins, quand on a un grand nombre de petites valeurs et un petit nombre de valeurs élevées, des intervalles de classes égaux ne sont pas vraiment appropriés. Au tableau 3.7 par exemple, on a beaucoup de petits revenus, et un petit nombre de revenus élevés. Si on utilise des intervalles égaux pour ces données, on va, soit perdre beaucoup d'informations, soit échouer dans notre tentative pour résumer les données. En d'autres termes, soit on aura seulement une ou deux classes avec toutes les observations, soit on aura tellement de classes que les données ne seront pas résumées. Dans ce cas on a besoin d'intervalles de classes inégaux pour résumer les données tout en gardant l'information essentielle.

**Cas non déclarés ou non mentionnés*****toujours inclure "non mentionné" dans la distribution de fréquences***

Lors de la collecte des données, il est toujours possible que pour certaines observations les informations nécessaires ne soient pas déclarées. Il est extrêmement souhaitable de voir apparaître ces cas dans une catégorie à part dans une distribution de fréquences, car cela peut donner une indication de l'ampleur du manque de données. Le traitement de ces cas dans des analyses ultérieures présente de nombreux problèmes car il y a une perte évidente d'informations. En voici un exemple :

Tableau 3.8 Répartition par âge et sexe des employés (en années complètes) à Fidji, 1993

Tranche d'âge	Nombre d'employés	
	Hommes	Femmes
15-19	2.170	1.209
20-29	17.075	9.659
30-39	17.257	9.454
40-49	11.669	4.819
50-59	4.774	1.249
Plus de 60 ans	498	104
Age non mentionné	17.472	4.632
<b>Total</b>	<b>70.915</b>	<b>31.126</b>

Source : Annual Employment Survey 1993, Bureau of Statistics, Fiji, Table 8, p. 19.

***'non mentionné' peut être très significatif***

Mis à part la simple perte d'information, les cas non mentionnés peuvent alerter les analystes sur les fausses conclusions possibles auxquelles mènerait l'étude des cas déclarés. Dans l'exemple ci-dessus, on peut soupçonner que la raison pour laquelle l'employeur ne mentionne pas l'âge des employés est de cacher les âges de plus de 60 ans ou de moins de 15 ans. Si ce soupçon se révélait vrai, la distribution de fréquence serait différente.

Pour d'autres aspects importants des distributions de fréquences, se référer à la fin de ce chapitre.

## Tableaux

### *tableaux statistiques*

En pratique, quand on prépare des rapports statistiques ou des publications, on applique un certain nombre de directives sur la mise en page et le format des distributions de fréquences. On les appelle tableaux statistiques de distributions de fréquences formatées. On utilise ces tableaux pour présenter des données unidimensionnelles (par exemple un tableau sur les répartitions par âge) ou bidimensionnelles (par exemple un tableau avec les tranches d'âges représentées dans les lignes, et les sexes dans les colonnes). On les utilise aussi pour montrer plus d'une variable — comme les tableaux qui comportent l'âge, le sexe et la région.

### *définition*

Un tableau est la disposition de données en lignes et colonnes. La forme la plus simple d'un tableau est une colonne ou une ligne de chiffres représentant le nombre d'unités entrant dans les catégories d'une variable unique. On l'appelle un *tableau à classification simple*.

### *conseils*

Voici les conseils que l'on peut donner pour la préparation ou la publication de tableaux statistiques :

- ☆ Donner une référence au tableau (comme un numéro d'ordre);
- ☆ Donner un titre clair;
- ☆ Mettre clairement les titres des lignes et des colonnes;
- ☆ Spécifier les unités des données du tableau (par exemple "kg");
- ☆ Inclure la source des données;
- ☆ Utiliser des lignes verticales et horizontales pour séparer les titres des données elles-mêmes;
- ☆ En général on ne sépare pas les colonnes par des lignes verticales, ni les lignes par des tracés horizontaux — ceci met trop de séparations dans le tableau;
- ☆ Séparez bien les données de manière à ce que le tableau soit très lisible;
- ☆ Utilisez des statistiques résumées (sous-totaux, totaux, moyennes) pour apporter des informations supplémentaires;
- ☆ Incluez des notes pour expliquer toute bizarrerie dans les données;
- ☆ Arrondissez les données de manière appropriée (en général à une ou deux décimales); et
- ☆ Assurez-vous que vous n'avez pas violé le secret professionnel en divulguant des informations personnelles ou délicates sur le plan commercial.

### *formater des tableaux*

Ces grands principes vous aideront à formater des tableaux :

- 1 Mettez en colonnes les chiffres les plus susceptibles d'être comparés.
- 2 Quand c'est faisable, mettez les colonnes avec les valeurs les plus élevées à gauche, et celles qui ont les chiffres les plus bas à la droite du tableau.

Un tableau qui aura été préparé en suivant ces conseils sera beaucoup plus facile à lire et à comprendre. Lorsque vous avez un doute, FAITES SIMPLE.

## Parties d'un tableau

**BOÎTE D'INFORMATION 3 : Parties d'un tableau**

(a) Numéro                      (b) Titre  
↓                                      ↓  
**Tableau 11 : Aide extérieure par principaux bailleurs, 1995**  
en milliers de AUD

(c) Note sous-titre                      (d) Titres                      (e) Titres de colonnes

Bailleur de fonds	Pays					Total
	Fidji	PNG	Samoa	Tonga	Vanuatu	
Australie	14.151	266.667	5.862	8.600	12.173	307.453
Nouvelle-Zélande	5.094	-	4.943	4.600	4.506	19.143
France	472	-	-	400	10.494	11.366
CE	19.245	18.841	1.667	-	4.593	44.345
Royaume Uni	377	-	-	-	3.333	3.711
USA	-	-	-	-	778	778
Canada	-	-	-	400	370	770
Japon	12.736	-	-	10.000	4.926	27.662
PNUD	660	-	977	700	2.519	4.856
BAD	-	-	-	-	4,099	4,099
Autres <sup>(1)</sup>	1.415	-	18.620	600	22.975	43.611
<b>Total</b>	<b>54.151</b>	<b>285.507</b>	<b>32.069</b>	<b>25.300</b>	<b>70.765</b>	<b>467.793</b>

(f) Cartouche

(g) Corps

(h) Note

(i) Source

<sup>(1)</sup> Tous les autres pays et autres organisations  
Source : SPESS, Communauté du Pacifique, 1998.

- (a) NUMÉRO DU TABLEAU Ceci sert à identifier le tableau et précède le titre. Si un rapport ou une publication contient plus d'un tableau, ils doivent tous être numérotés.
- (b) TITRE Le **titre** est placé au-dessus du corps du tableau. Il doit être bref et concis, mais complètement évident. Dans certains cas il est nécessaire de mettre plusieurs lignes de titre. Si un titre est trop long, alors on peut utiliser un titre en abrégé au-dessus du titre complet.
- (c) NOTE SOUS-TITRE Pour certains tableaux il est nécessaire d'inclure une **note sous-titre**. Elle est d'ordinaire imprimée en plus petits caractères que le titre et fournit des informations supplémentaires sur le tableau ou une de ses parties importantes. Les **notes sous-titres** sont souvent utilisées pour spécifier les unités des données dans le tableau, ou l'enquête dont les données proviennent.
- (d) TITRES Les variables dans les lignes et les colonnes du tableau doivent être définies par un titre.
- (e) TITRE DE COLONNE C'est la désignation en haut de chaque colonne qui explique ce que la colonne représente.
- (f) CARTOUCHE C'est la colonne de gauche du tableau avec les titres. Il indique la description de chaque ligne du tableau.
- (g) CORPS Le **corps** du tableau contient l'information qui est placée dans les cellules appropriées suivant les titres des lignes et des colonnes. Une **cellule** est l'intersection d'une ligne et d'une colonne.
- (h) NOTES DU BAS Elles fournissent des explications sur des chiffres isolés dans les colonnes ou les lignes. Elles sont placées en bas du tableau et sont d'habitude écrites en

caractères plus petits. Elles sont numérotées avec soit des lettres de l'alphabet ou des chiffres, de gauche à droite sur la page. Un **nouveau jeu de notes** doit être fourni à chaque tableau. Si une répétition très longue doit être évitée, alors seulement on peut écrire quelque chose comme "voir les notes en bas du tableau XX" au lieu de répéter une entière série de notes.

- (i) SOURCE Si les statistiques sont collectées à partir d'une source secondaire, c'est noté en-dessous du titre ou, le plus souvent, en-dessous des notes.

## ARRONDIR

### *pourquoi arrondir ?*

**Arrondir** est souvent la première étape pour simplifier et résumer des données statistiques. Arrondir correctement est essentiel si on veut qu'un tableau soit facile à comprendre.

On a souvent peur de perdre l'acuité des données quand on arrondit. Il existe deux arguments contre cette crainte.

### *acuité vs compréhension*

- 1 Si un ordinateur peut produire un chiffre avec plein de décimales, cela ne veut pas dire pour autant que ce chiffre est forcément exact. Un chiffre n'est pas plus exact que l'instrument qui l'a mesuré.
- 2 Même si le chiffre est super exact et nécessite plein de décimales, si le chiffre arrondi peut être compris au premier coup d'oeil, on opte naturellement pour la valeur arrondie.

### *règles*

Voici les grandes lignes directrices pour arrondir :

Les chiffres **plus petits que 5** sont arrondis à la **tranche inférieure**  
 Les chiffres **plus grands que 5** sont arrondis à la **tranche supérieure**  
 Les chiffres **égaux à 5** sont arrondis **au hasard** à la tranche inférieure ou à la tranche supérieure

### *exemple*

**93** est arrondi à 90  
**96** est arrondi à 100  
**95** est arrondi à 90 ou 100 au hasard

### *arrondir le chiffre 5*

Si la méthode utilisée pour arrondir la décimale 5 n'est pas la même que celle-ci, continuez à utiliser votre méthode. Le plus important c'est que la même méthode soit appliquée à tous les tableaux.

### *assurez-vous que les totaux sont corrects*

Au tableau 3.9 on voit que si on additionne les données arrondies des femmes on obtient 51.500, mais si on arrondit le chiffre du total brut (51.583) on obtient 51.600. En général le total arrondi doit être le même que le total non arrondi — si on présente seulement les chiffres arrondis, le total doit être 51.600.

Tableau 3.9 Population par état, États fédérés de Micronésie, 1994

Etat	Non arrondi		Arrondi à la centaine supérieure	
	Hommes	Femmes	Hommes	Femmes
Yap	5.565	5.613	5.600	5.600
Pohnpei	17.253	16.439	17.300	16.400
Kosrae	3.806	3.511	3.800	3.500
Chuuk	27.299	26.020	27.300	26.000
<b>Total</b>	<b>53.923</b>	<b>51.583</b>	<b>54.000</b>	<b>51.500</b>

Source : 1994 FSM Census of Population and Housing, Detailed Social and Economic Characteristics Report, 1996.

## Le calcul des pourcentages

### calculer des pourcentages

Il est souvent utile de présenter les données sous la forme de pourcentages. Pour changer un chiffre en pourcentage, on le divise par le total et on le multiplie par 100.

### pourcentage

$$\text{Pourcentage} = \frac{\text{chiffre}}{\text{total}} \times 100$$

### conseils

En général on n'a pas plus de deux décimales dans les pourcentages. Le total des pourcentages doit être 100. On doit aussi faire attention à indiquer, soit dans la colonne, soit dans une note en bas du tableau quel chiffre a été utilisé comme total — particulièrement si les totaux généraux n'ont pas été utilisés. Par exemple :

Tableau 3.10 Population par état, États fédérés de Micronésie, 1994

Etat	Nombre		Pourcentage de la population totale	
	Hommes	Femmes	Hommes	Femmes
Yap	5.565	5.613	5,3	5,3
Pohnpei	17.253	16.439	16,4	15,6
Kosrae	3.806	3.511	3,6	3,3
Chuuk	27.299	26.020	25,9	24,7
<b>Total</b>	<b>53.923</b>	<b>51.583</b>	<b>51,1</b>	<b>48,9</b>

Source : 1994 FSM Census of Population and Housing, Detailed Social and Economic Characteristics Report, 1996.

### pourcentage total

Au tableau 3.10 le titre de colonne indique clairement le pourcentage de la population totale — et le total pour les hommes (51,1 %) ajouté au total pour les femmes (48,9 %) est bien égal à 100 %.



## PLUS D'INFOS SUR LES DISTRIBUTIONS DE FRÉQUENCES

### *d'autres types de distributions de fréquences*

En plus des distributions de fréquences couvertes dans ce chapitre, il en existe trois autres types assez pratiques à utiliser dans l'analyse de données.

- ☆ les distributions de fréquences relatives,
- ☆ les densités de fréquences, et
- ☆ les distributions cumulatives.

### DISTRIBUTIONS DE FRÉQUENCES RELATIVES

#### *utiliser les quotients*

La distribution de fréquence relative est le quotient de la fréquence de classe à la fréquence totale. Les fréquences de classe relatives données au tableau 3.11 sont basés sur les données du tableau 3.3.

#### *fréquence de classe relative*

$$\text{Fréquence de classe relative} = \frac{\text{fréquence de classe}}{\sum \text{fréquences de classes}}$$

#### *notation*

Le signe  $\Sigma$  ou sigma signifie somme, ou total. C'est le symbole le plus souvent utilisé en statistique.

Tableau 3.11 Fréquence de classe relative pour les poids de poissons (kg)

Groupe de classe	Fréquence de classe	Fréquence de classe relative
2,0–2,9	7	0,11
3,0–3,9	19	0,30
4,0–4,9	16	0,25
5,0–5,9	12	0,19
6,0–6,9	6	0,10
7,0–7,9	3	0,05
<b>Total (<math>\Sigma</math>)</b>	<b>63</b>	<b>1,00</b>

Source : Tableau 3.3

## DENSITÉ DE FRÉQUENCE

### *on peut comparer les intervalles de classes inégaux avec les densités de fréquences*

Comme on l'a dit plus tôt dans ce chapitre, les intervalles de classe inégaux doivent être évités. Néanmoins, c'est quelquefois impossible, et quand on compare les fréquences de classes on doit être très vigilant à déterminer la nature de la distribution. Les fréquences de classes à intervalles inégaux ne peuvent pas être comparées de but en blanc. On peut cependant les comparer en divisant la fréquence d'une classe par la longueur intervalle de la classe. **Ce quotient est appelé densité de fréquence.** On peut aussi déterminer un intervalle de fréquence qui convienne en divisant les fréquences de classes par le quotient de l'intervalle de classe jugé approprié.

### *exemple*

Par exemple, si une distribution de fréquences a dix classes, les huit premières ayant un intervalle de 5 ans et les deux dernières de 10 ans, l'intervalle de classe le plus approprié pour déterminer la densité de fréquences sera 5 ans. Pour les huit premières classes la densité de fréquence sera égale à la fréquence de classe, car on divisera par un ( $5/5 = 1$ ). Pour les deux dernières classes la densité de fréquences sera une demie-classe, car on divisera par deux ( $10/5 = 2$ ). La densité de fréquence sera alors 'pour 5 ans'.

Prenons l'exemple du tableau 3.12. Il est utile de savoir calculer les densités de fréquences quand on construit des diagrammes avec des intervalles de classes inégaux.

Tableau 3.12 Âge des époux le jour du mariage, Guam, 1994

Âge au mariage	Intervalle de classe	Épouses		Époux	
		Fréquence de classe	Densité de fréquence (pour 5 ans)	Fréquence de classe	Densité de fréquence (pour 5 ans)
15 à 19 ans	5	124	124	37	37
20 à 24 ans	5	536	536	518	518
25 à 29 ans	5	391	391	359	359
30 à 34 ans	5	248	248	256	256
35 à 44 ans	10	218	109	272	136
45 ans et plus <sup>1</sup>	30	75	13	150	25
Total		1.592		1.592	

<sup>1</sup> La tranche d'âge maximum était de 70 à 74 ans, donc cet intervalle de classe égale :  $74 - 45 = 30$  (inclus).

Source : Office of Planning and Evaluation, 1995, Guam.

## DISTRIBUTIONS CUMULATIVES

### *plus ou moins qu'une valeur spécifique*

La distribution de fréquences donne des informations sur le nombre d'unités dans les différents intervalles de classes, en extrayant le nombre d'unités de petite taille, de taille moyenne et de grande taille. Cependant, dans de nombreux cas on s'attache seulement à trouver le nombre ou le pourcentage d'unités qui ont une valeur plus grande ou plus faible qu'une valeur spécifique.

### *applications*

Par exemple, quand on étudie la distribution de terres pour formuler une législation foncière, il est probablement important de connaître le nombre ou le pourcentage de propriétés qui ont une taille plus

grande ou moindre qu'une taille donnée. De même, quand on étudie la distribution des revenus de ménages ou d'individus, il est important de savoir quel nombre ou quel pourcentage d'individus ou de ménages a des revenus inférieurs ou supérieurs à une valeur donnée. Ce type d'informations peut être obtenu facilement à partir d'une distribution de fréquences en calculant les fréquences cumulatives ou les fréquences relatives cumulées.

#### **totaux progressifs de chaque intervalle de classe**

La **distribution cumulative de fréquences** est obtenue en calculant les totaux progressifs des fréquences à chaque intervalle de classe. On peut commencer soit au premier intervalle et aller vers le bas, soit au dernier intervalle et aller vers le haut, suivant le type de fréquences cumulées souhaité. Le premier type est appelé '**fréquence cumulative moins que**', et le deuxième est appelé '**fréquence cumulative plus que**'. Si les fréquences relatives sont cumulées à la place des fréquences au tableau 3.12, alors on obtient les **fréquences relatives cumulées moins que ou plus que**.

#### **exemples**

Voici deux exemples de fréquences cumulées aux tableaux 3.13 et 3.14.

Tableau 3.13 Distribution de l'âge des employés, Fidji, 1993

Tranche d'âge	Fréquence	Fréquence cumulative	
		(moins que)	(plus que)
15 à 19 ans	3.379	3.379	102.041
20 à 29 ans	26.734	30.113	98.662
30 à 39 ans	26.711	56.824	71.928
40 à 49 ans	16.488	73.312	45.217
50 à 59 ans	6.023	79.335	28.729
Plus de 60 ans	602	79.937	22.706
Non mentionné	22.104	102.041	22.104
<b>Total</b>	<b>102.041</b>		

Source : Enquête annuelle sur l'emploi 1983, Bureau of Statistics, Suva, Fidji

Tableau 3.14 Âges des époux mariés à Guam le jour du mariage, Guam, 1994

Tranche d'âge	Fréquence	Fréquence cumulative	
		(moins que)	(plus que)
15 à 19 ans	124	124	1.592
20 à 24 ans	536	660	1.468
25 à 29 ans	391	1.051	932
30 à 34 ans	248	1.299	541
35 à 44 ans	218	1.517	293
45 ans et plus	75	1.592	75
<b>Total</b>	<b>1.592</b>		

Source : Tableau 3.12

***fréquence cumulée – moins que***

La fréquence cumulée (moins que) nous permet de trouver facilement des informations comme le nombre de personnes de moins de 30 ans. Dans l'exemple du tableau 3.13, on peut voir que 30.113 personnes de moins de 30 ans ont été employées aux Îles Fidji en 1993, et que 56.284 personnes de moins de 40 ans ont été employées. Au tableau 3.14 on voit que dans l'année 1994 à Guam 124 femmes se sont mariées avant leur vingtième année, et 660 hommes se sont mariés avant l'âge de 25 ans. C'est une source d'information importante, et c'est en même temps très facile à lire.

***fréquence cumulée – plus que***

La fréquence cumulée (plus que) est tout aussi simple. Au tableau 3.13 on peut voir que 22.706 personnes ont été employées après leurs 60 ans et quand l'âge n'a pas été mentionné. Au tableau 3.14 on peut voir que 239 femmes d'au moins 35 ans et 541 hommes d'au moins 30 ans se sont mariés.

***les 'non mentionnés' sont généralement ignorés***

Ces sortes d'observations sont très utiles dans un grand nombre de cas, d'où le besoin de calculer les fréquences cumulées. On doit prendre soin de la manière de traiter les 'non mentionnés'. Il est généralement préférable de calculer les fréquences cumulées en excluant cette catégorie.

