

# REPRÉSENTATION GRAPHIQUE

- RÈGLES GÉNÉRALES POUR LA PRÉPARATION DE GRAPHIQUES
- DIAGRAMMES DE DISPERSION • GRAPHIQUES EN BARRES
- DIAGRAMMES À SECTEURS (CAMEMBERTS) • GRAPHIQUES LINÉAIRES
- HISTOGRAMMES • POLYGONES DE FRÉQUENCES



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

- CONSTRUIRE UN GRAPHIQUE
- COMPRENDRE LES DIFFÉRENCES DANS LA MANIÈRE DE CONSTRUIRE  
LES GRAPHIQUES POUR DES DONNÉES QUANTITATIVES ET LES  
GRAPHIQUES POUR DES DONNÉES QUALITATIVES
- MAÎTRISER UNE GAMME DE TECHNIQUES DE GRAPHIQUES

# CHAPITRE 4

## REPRÉSENTATION GRAPHIQUE

### Introduction

#### *Les diagrammes sont très faciles à comprendre*

Jusqu'à présent on s'est contenté de présenter les données sous forme de tableau. La présentation de données en diagrammes est souvent plus facile à comprendre. Mais on doit faire attention quand on dessine un graphique, car il est facile de faire une mauvaise représentation de données statistiques. Dans ce chapitre on présente les principaux types de présentations graphiques, ainsi que les types de données qui sont le plus facilement décrites par des graphiques.

### Règles générales pour la préparation de graphiques

#### *tout le monde préfère les graphiques*

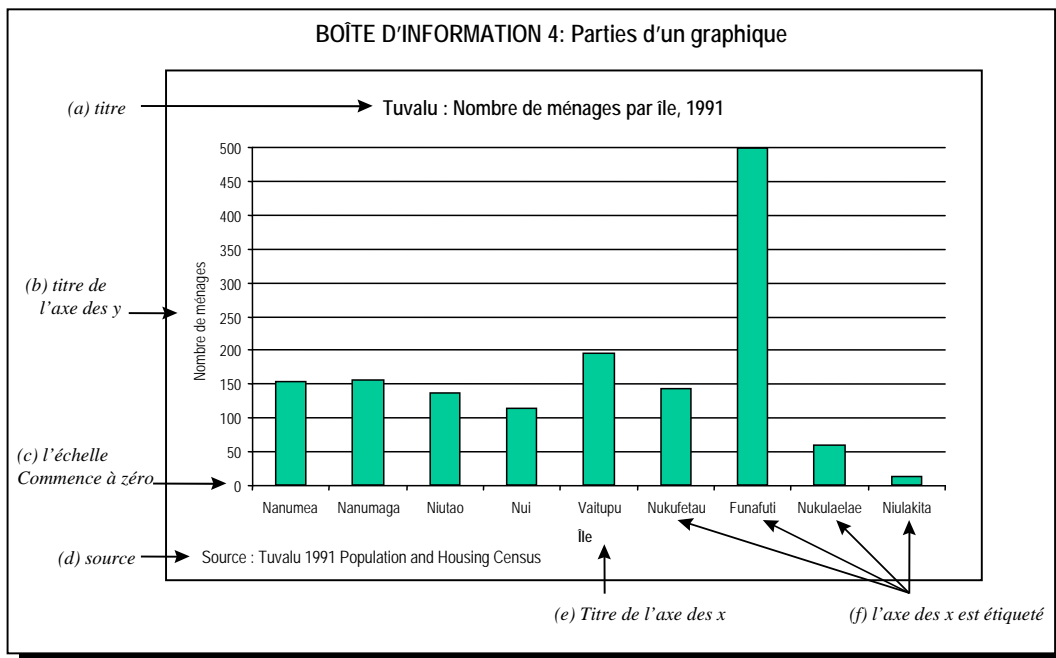
Beaucoup de gens préfèrent regarder un graphique plutôt qu'examiner un tableau de chiffres ou lire une page de texte. Une présentation graphique est souvent plus facile à comprendre qu'une présentation numérique — une image vaut 1000 mots, dans notre cas 1000 chiffres.

#### *règles générales*

On doit toujours se dire que des graphiques différents montrent des aspects différents des mêmes données. Quand on prépare des graphiques on doit suivre un certain nombre de règles pour s'assurer que le graphique peut être compris et interprété facilement :

- ☆ Les graphiques doivent avoir un titre clair et explicite.
- ☆ Les unités de mesure doivent être indiquées.
- ☆ Les graphiques doivent être simples et pas trop chargés.
- ☆ Tous les axes doivent être étiquetés clairement.
- ☆ Il faut inclure la source des données.
- ☆ L'échelle de chaque axe ne doit pas altérer ou cacher la moindre information. C'est-à-dire que le graphique doit montrer les données sans changer le message.
- ☆ Les graphiques doivent clairement montrer toute tendance ou différence entre les données.
- ☆ Il faut indiquer si l'échelle ne commence pas à zéro sur l'axe.
- ☆ Le type de présentation doit être choisi avec le plus grand soin, y compris la taille et la forme des symboles de points (astérisques, points ou carrés) et la méthode, s'il y en a une, pour relier les points (ligne droite, courbe, pointillés, etc.).

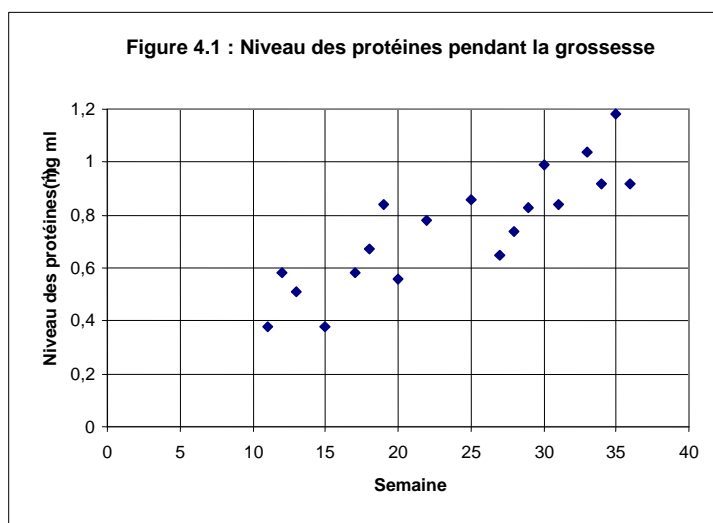
- ☆ Les graphiques doivent être précis à l'oeil nu — si une des valeurs sur le diagramme est 15 et l'autre 30, alors la seconde valeur doit paraître avoir deux fois la taille de la première.
- ☆ L'approche empirique est très utile pour améliorer les graphiques. Il est peu probable que vous trouverez le graphique approprié du premier coup.



## Diagrammes de dispersion

### il est facile de voir clairement les relations

Une des formes les plus simples de représentation graphique est le **diagramme de dispersion**. Les diagrammes de dispersion sont en général utilisés pour montrer une relation entre deux variables (données bidimensionnelles), pour déceler toute valeur aberrante ou pour déceler des grappes d'observations.



Source : fictive.

**valeurs aberrantes — l'inattendu**

La figure 4.1 illustre des données quantitatives bidimensionnelles où chaque observation consiste en deux mesures — le temps en semaines, et un niveau de protéines. Il montre le changement de niveau de protéines pendant la grossesse. On peut voir les **valeurs aberrantes** — c'est-à-dire les observations qui sont plus hautes ou plus basses que la structure générale des autres observations.

**Graphiques en barres****données qualitatives et quantitatives**

Les graphiques en barres sont utilisés pour décrire la **distribution de données qualitatives et quantitatives** quand les données sont groupées en intervalles de classes égaux. *La fréquence de la variable est représentée par la hauteur de la barre.*

**comparaisons**

Les graphiques en barres sont utiles quand on veut faire une comparaison entre deux variables — comme les différences entre les hommes et les femmes. Avec des données nominales l'ordre des catégories sur les axes est arbitraire, bien que les variables soient souvent classées de la plus grande à la plus petite fréquence.

**verticales ou horizontales**

Les graphiques en barres peuvent être verticaux ou horizontaux, bien qu'ils soient généralement verticaux avec les catégories de variables sur l'axe des x et leurs fréquences sur l'axe des y.

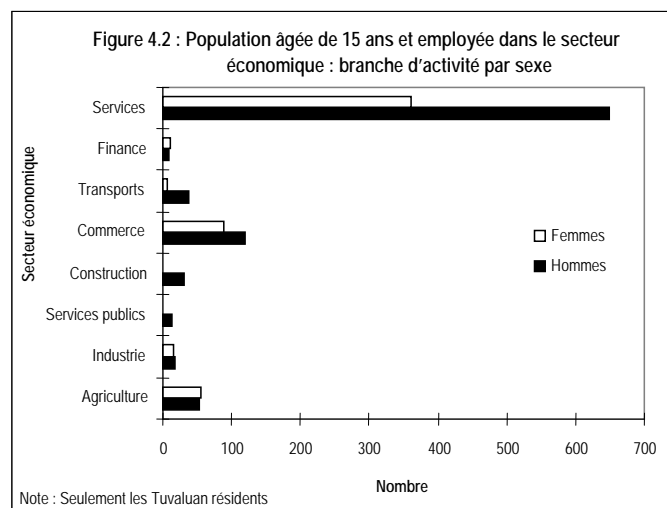
**ASTUCE**

**LES GRAPHIQUES EN BARRES VERTICALES SONT PLUS APPROPRIÉS POUR MONTRER LES DISTRIBUTIONS DE FRÉQUENCES ENTRE LES CATÉGORIES DE VARIABLES.**

**ON UTILISE LES GRAPHIQUES EN BARRES HORIZONTALES LORSQUE LES TITRES DES CATÉGORIES DE VARIABLES SONT TROP LONGUES POUR S'INSÉRER SUR L'AXE DES X ET QU'IL EST PLUS FACILE DE LIRE L'ÉCHELLE SUR L'AXE DES Y.**

**exemple**

Ce diagramme est un graphique en barres horizontales. Imaginez comme les titres seraient petits si on devait les aligner sur l'axe horizontal !



Source : Tuvalu 1991 Population and Housing Census.

## Diagrammes à secteurs (camemberts)

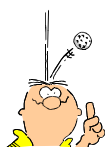
### tranche du camembert

Les diagrammes à secteurs sont ceux qu'on utilise le plus souvent pour représenter des données **qualitatives**. Chaque "tranche de camembert" représente la relative fréquence de classe de cette catégorie. Ils sont simples à construire, mais sont très efficaces quand vous avez **cinq classes ou moins** dans vos données.

### nombre de classes peu élevé

En général on utilise des camemberts quand on a un nombre de classes peu élevé et quand on s'intéresse à la fréquence relative de *chaque catégorie* (plutôt que les comparaisons avec les autres variables).

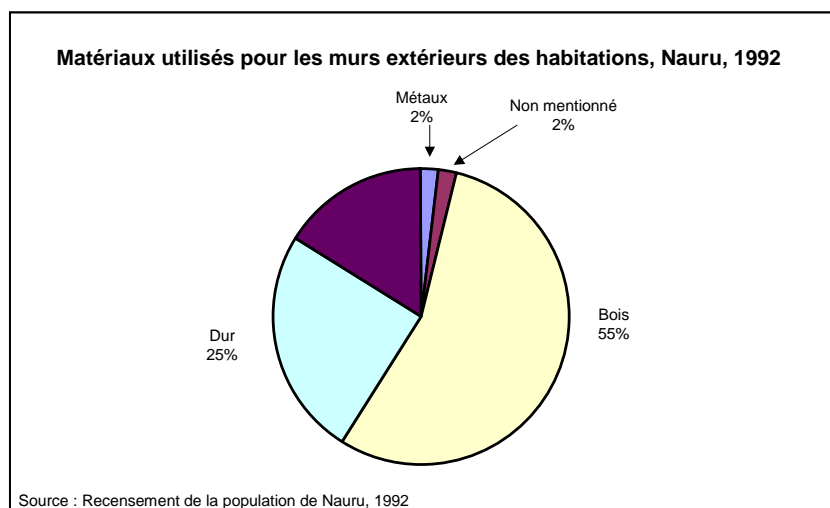
**ASTUCE**



**CONSEIL IMPORTANT** pour l'ÉLABORATION d'UN CAMEMBERT :  
LES SEGMENTS DOIVENT ÊTRE ORDONNÉS PAR TAILLE, pour FACILITER LA LECTURE ET LA COMPRÉHENSION DE L'INFORMATION.

### exemple

La figure 4.3 contient des données du recensement de la population de Nauru de 1992. Sur ce diagramme on montre le pourcentage de chaque variable. C'est une option utile quand on crée des diagrammes en secteurs puisqu'il n'y a aucune échelle qui permette la lecture, comme dans les graphiques en barres.



## Graphiques linéaires

### variables indépendantes et dépendantes

Les graphiques linéaires sont utilisés pour les données bidimensionnelles. Ils ont deux axes — l'axe horizontal ( $x$ ) sur lequel on met l'échelle des valeurs de la **variable indépendante**, et l'axe vertical ( $y$ ) sur lequel on met l'échelle de la **variable dépendante**. La variable indépendante est celle sur laquelle sont basées les prévisions, alors que la variable dépendante change par rapport à la variable indépendante. On peut prendre le temps comme exemple de variable indépendante — il ne dépend de rien. Par exemple, quand on regarde des décomptes de population de recensements successifs, on pourrait utiliser un graphique linéaire. Les années de recensement seraient la variable indépendante, et les décomptes de population seraient la variable dépendante.

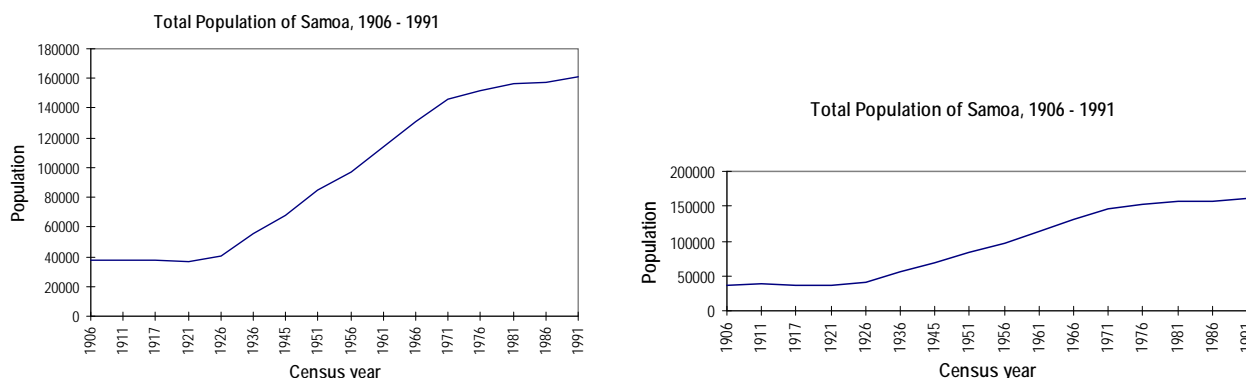
### usage courant

Les graphiques linéaires sont couramment utilisés pour répartir des données *sur une période donnée*, et quand on veut montrer des *tendances* dans les données. Quand on prépare un graphique en ligne on doit faire attention à ce que l'échelle de l'axe vertical ( $y$ ) ne fausse pas le sens des données.

### exemple

La figure 4.4 contient deux graphiques linéaires avec les mêmes données, mais avec deux échelles différentes sur l'axe des  $y$ . Le deuxième graphique donne une impression tout à fait différente du premier. En regardant le premier graphique on suppose que la population a augmenté de façon radicale entre les années 30 et les années 60, mais le deuxième graphique ne donne pas la même impression. Il est particulièrement important de garder la même échelle lorsqu'on compare des populations différentes.

Figure 4.4 : Population totale de Samoa, 1906–1991



Source : Census of Population and Housing, 1991, Western Samoa

## Histogrammes

### la surface de chaque barre est proportionnelle à une fréquence

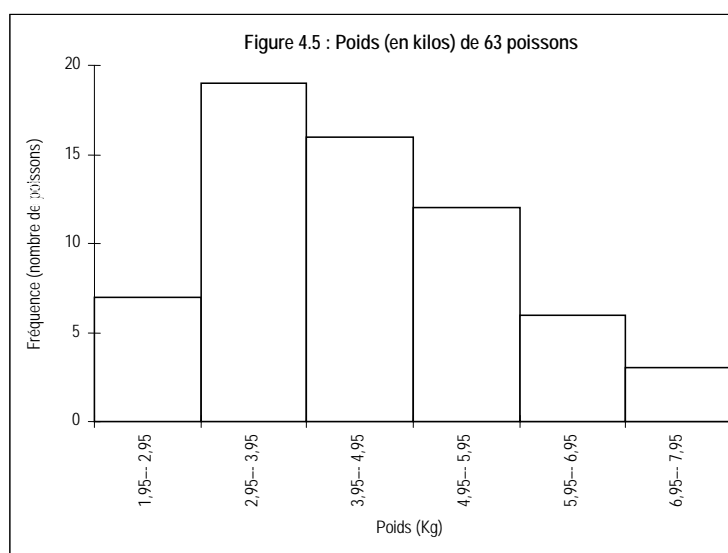
La méthode la plus commune pour représenter une distribution de fréquences de **données continues** (ou de données qu'on suppose continues) est d'utiliser un **histogramme**. Un histogramme est un graphique à barres verticales dont *la surface de chaque barre (pas nécessairement la hauteur des barres) est proportionnelle aux fréquences de classes* (alors que les graphiques en barres ont en même

temps la hauteur et la surface proportionnelles à la fréquence de classe).

### différent des graphiques en barres

Les histogrammes sont différents des graphiques en barres parce que les barres verticales sont toujours côte à côte, sans intervalles. Ceci pour illustrer la nature continue des données. En général un histogramme a des barres de largeurs égales, bien que la largeur des barres doive être ajustée quand les intervalles de classe ne sont pas égaux.

### exemple



Source : Tableau 3.1

### la surface est importante

Vous devez vous rappeler que quand vous tracez une distribution continue, c'est **la surface de chaque rectangle** (et pas la hauteur) qui est **proportionnelle à la fréquence**. C'est seulement dans le cas d'intervalles de classe égaux (comme à la figure 4.5) que la fréquence est proportionnelle à la hauteur du rectangle.

### classes inégales

Regardons un autre exemple d'histogramme. L'histogramme de la figure 4.6 (élaboré à partir du tableau 4.1) montre ce que nous devons faire avec une distribution à intervalles de classe inégaux et à intervalles de classes ouverts. Rappelez-vous que c'est **la surface des barres** qui est importante dans un histogramme. Ça veut dire que quand on a des classes inégales on doit s'assurer que la surface de chaque barre représente précisément sa fréquence. La classe 100–199 a deux fois la taille de la classe 50–99, aussi dans l'histogramme la largeur du rectangle 100–199 doit être double du rectangle 50–99.

### hauteur = densité de fréquence

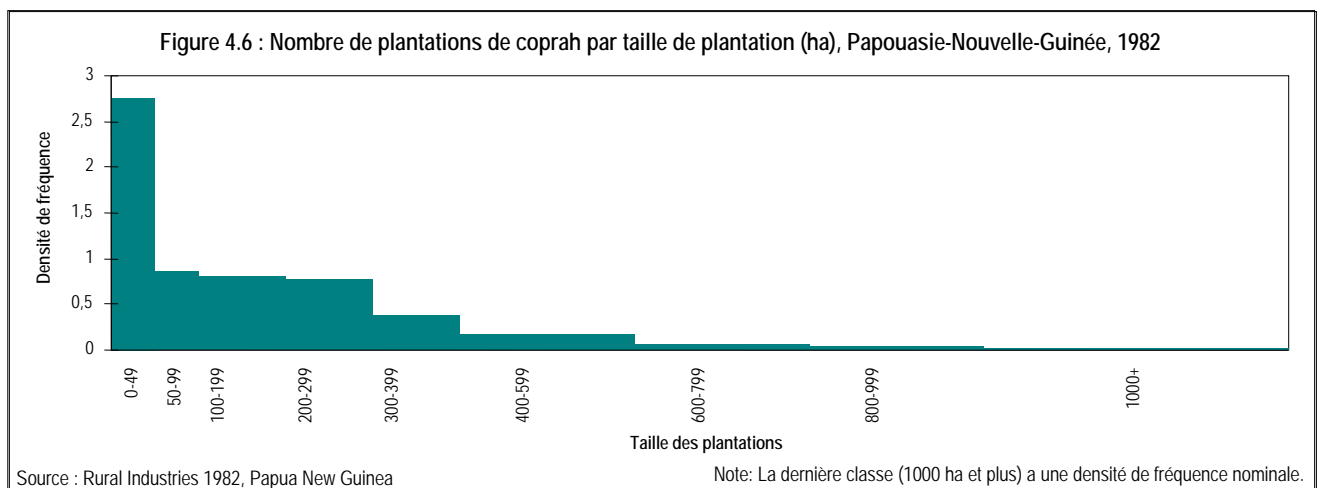
*Pour s'assurer que la surface d'un rectangle reflète la fréquence de classe, la hauteur du rectangle doit alors représenter la densité de fréquence.*

Notez que la méthode utilisée pour les intervalles de classe ouverts a fait l'objet d'une note en bas de page, pour être sûr que le lecteur soit informé.

Tableau 4.1 Nombre de plantations de coprah par surface plantée en Papouasie-Nouvelle-Guinée, 1982

Groupe de classe (ha)	Intervalle de classe	Fréquence de classe	Densité de fréquence
0-49	50	138	2,76
50-99	50	43	0,86
100-199	100	81	0,81
200-299	100	77	0,77
300-399	100	38	0,38
400-599	200	34	0,17
600-799	200	13	0,07
800-999	200	6	0,03
1.000 et +		8	–
<b>Total</b>		<b>438</b>	

Source : Rural Industries 1982, Papua New Guinea.



### classes ouvertes

À la figure 4.6, il y a aussi le problème du choix du traitement des classes ouvertes, ici la classe "1000 et +". On ne connaît pas la largeur ou l'intervalle de cette classe, et on n'a aucun moyen de le deviner. Dans certains cas, il est possible de supposer une limite supérieure — par exemple si on a des données sur l'âge au moment du mariage, on peut supposer une limite supérieure de 60, car il est improbable que des personnes de plus de 60 ans contractent le mariage.

À la figure 4.6 on n'a aucun moyen de connaître la limite supérieure de la classe 1000 et +, alors on ne peut pas calculer la hauteur du rectangle qui représente cette partie de la fréquence. Il serait faux de la laisser de côté, alors on doit décider quoi faire avec cette classe.

### deux façons de traiter les classes ouvertes

Le problème peut être traité de deux façons. D'abord, on peut supposer une limite supérieure pour la distribution et dessiner le rectangle en conséquence (comme à l'illustration de l'âge au moment du mariage). La seconde alternative est de dessiner le rectangle avec une hauteur nominale, mais de le laisser ouvert et de mettre en note la méthode utilisée, comme à la figure 4.6 ci-dessus. Ceci indique un intervalle ouvert. Cette seconde méthode, néanmoins, fonctionne uniquement quand la fréquence de



l'intervalle de classe ouvert est faible (comme c'est le cas à la figure 4.6), mais ce devrait toujours être le cas (de grandes classes ouvertes doivent toujours être évitées).

#### données discontinues

Quand on dessine un histogramme pour des **données discontinues** on a deux approches possibles. Regardons le tableau 4.2 ci-dessous :

Tableau 4.2 Taille des groupes d'emplois par nombre d'employés rémunérés, 1996

Taille des groupes d'emplois	Personnes touchant des revenus et salaires	Personnes par entreprise = fréquence ÷ centre de classe
1-4	5.053	= 5.053 ÷ 4 = 1.263,25
5-19	15.260	= 15.260 ÷ 15 = 1.017,33
20-49	15.312	= 15.312 ÷ 30 = 510,40
50-99	11.714	= 11.714 ÷ 50 = 234,28
100 et plus <sup>1</sup>	62.742	= 62.742 ÷ 400 = 156,85
<b>Total</b>	<b>110.081</b>	

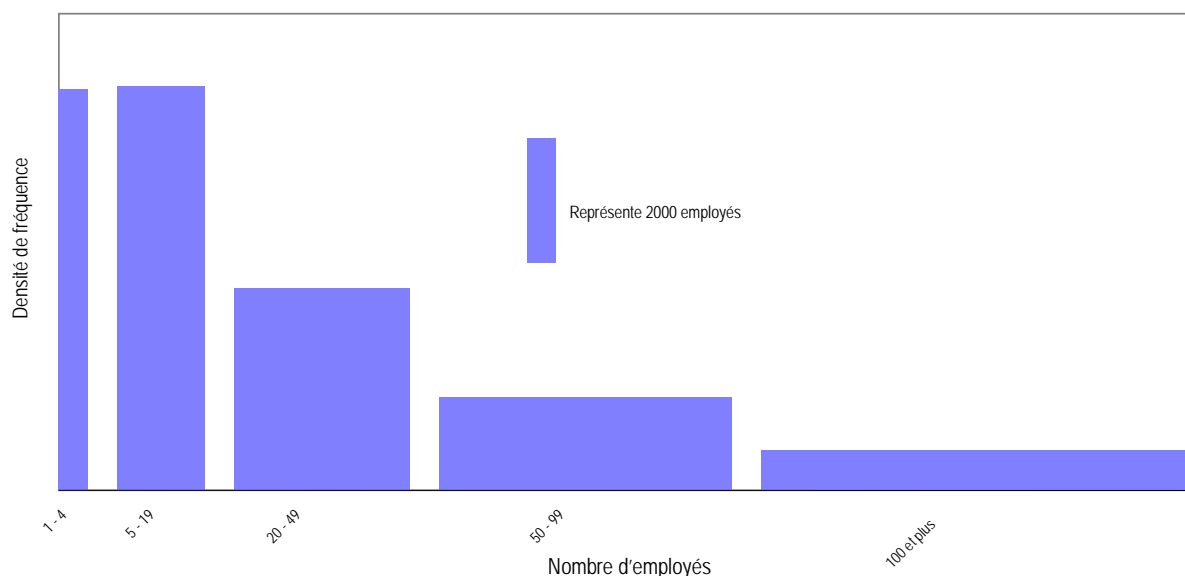
<sup>1</sup> Limite supérieure supposée de 500.

Source : 1996 Annual Employment Survey, Fiji Bureau of Statistics, Table 2, p. 15.

#### deux alternatives

La première approche est de dessiner un histogramme dont les rectangles sont séparés par des espaces. Cela montre au lecteur que les données sont discontinues. On voit cette approche à la figure 4.7 :

Figure 4.7 : Taille des groupes d'emplois par nombre d'employés rémunérés, Îles Fidji, 1996



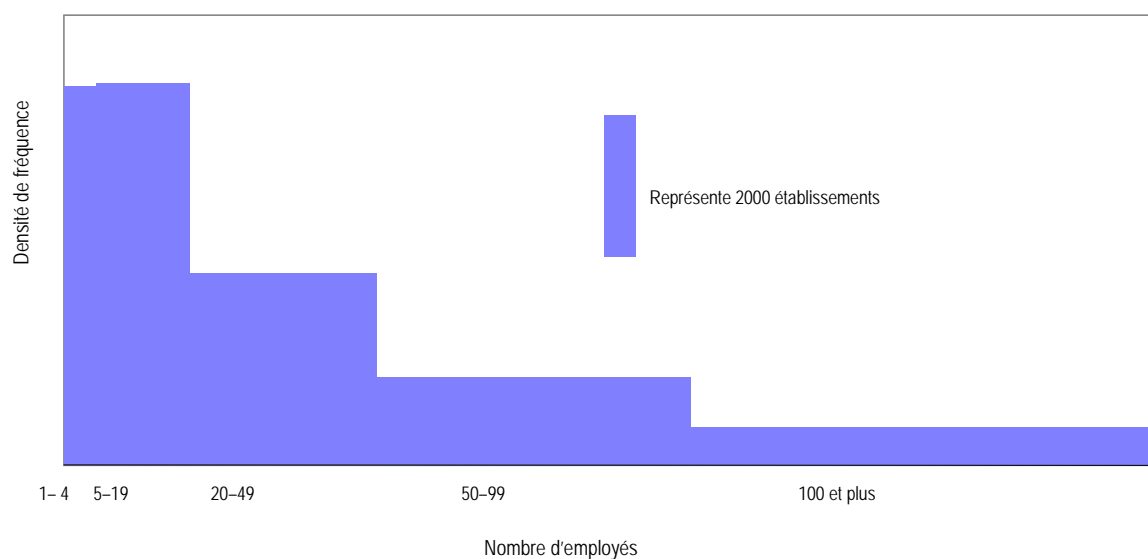
Source : Recensement annuel de l'emploi, 1996, Fiji Bureau of Statistics

Note : La classe 100 et plus a une densité de fréquence nominale

**seconde alternative**

La seconde approche consiste à procéder comme si les données étaient en fait continues. Si on a une distribution similaire à celle donnée au tableau 4.2, alors l'unité de mesure (dans notre cas des individus) est faible en comparaison de l'étendue des valeurs observées. Dans ce cas, on peut dessiner un histogramme approximatif en faisant comme si les données étaient continues. L'historgramme est dessiné à la figure 4.8 et ressemble un peu à celui de la figure 4.7 car l'échelle horizontale est divisée en intervalles inégaux. On doit être conscient du fait que *la figure 4.7 est seulement une approximation* parce que les données sont discontinues, donc on ne peut pas vraiment les représenter par des rectangles continus. Cependant, si on veut représenter la forme générale de la distribution, le diagramme est adéquat.

Figure 4.8 : Taille des groupes d'emplois par nombre d'employés rémunérés, Îles Fidji, 1996



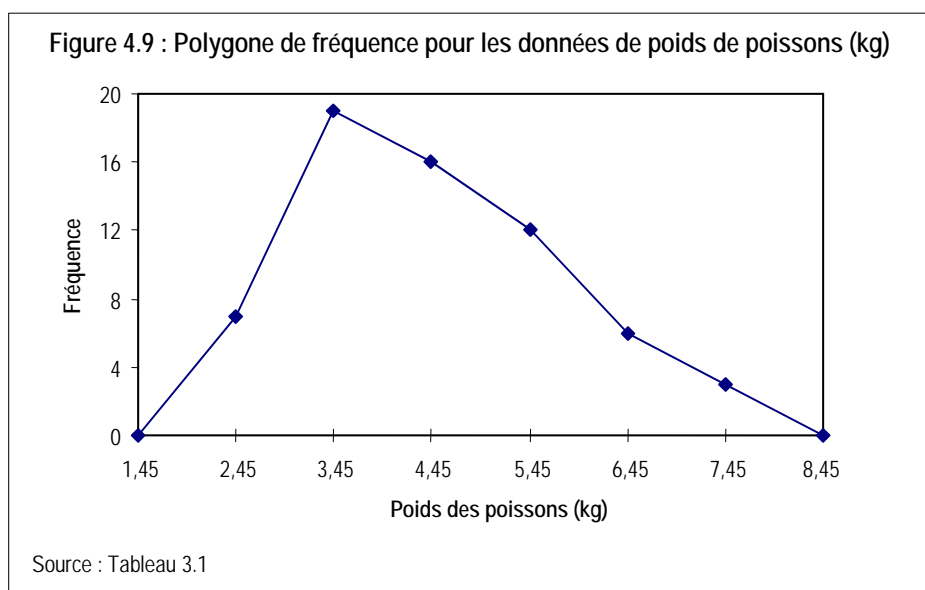
Source : Recensement annuel de l'emploi, 1996, Fiji Bureau of Statistics

Note : La classe 100 et plus a une densité nominale

**Polygones de fréquences****fréquences tracées au centre de classe**

Il existe un autre type de graphique qui convient aux **données continues**, ou aux *données discontinues qui peuvent être considérées comme approximativement continues*; c'est le **polygone de fréquence**. Dans ce graphique les fréquences de chaque classe sont tracées au centre de classe, et les points successifs joints par des lignes droites.

On donne un exemple d'un polygone de fréquence avec des intervalles de classes égaux à la figure 4.9 ci-dessous. Les données utilisées sont celles des poids de poissons du tableau 3.1.



### ***largeurs de classes inégales***

Quand les intervalles de classes sont inégaux, comme dans le cas d'un histogramme, on doit utiliser les densités de fréquences ou les densités de fréquences relatives.

### ***on doit commencer et finir à l'axe des x***

Le début et la fin des polygones doivent être rallongés sur l'axe horizontal, aux centres des classes inférieures et supérieures à celles couvertes par la distribution.

Pour une discussion sur la représentation graphique des distributions cumulatives de fréquences, voir "Plus d'info sur les distributions de fréquences".

## PLUS D'INFOS SUR LA REPRÉSENTATION GRAPHIQUE

### FRÉQUENCES CUMULÉES

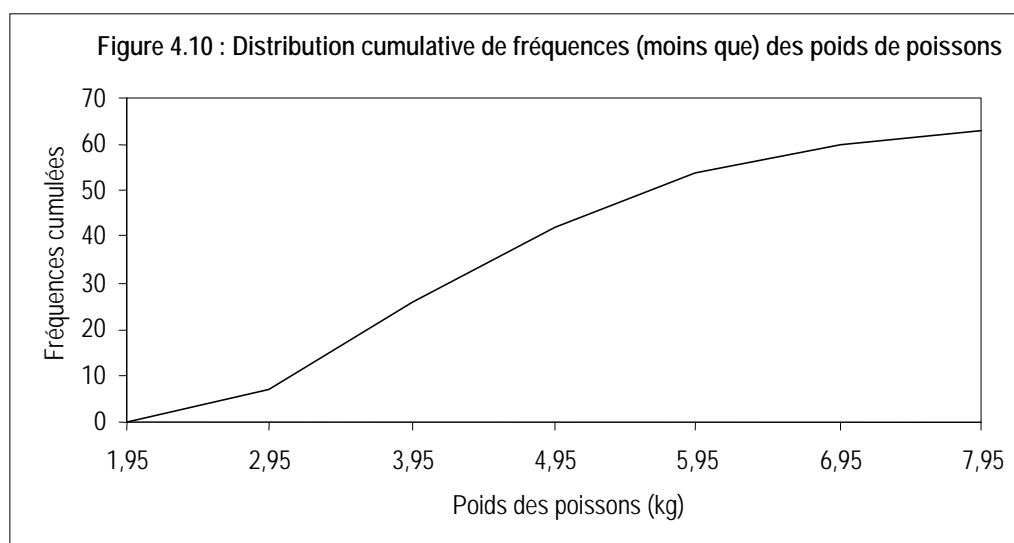
#### ogives

Dans certains cas il est préférable de construire une distribution cumulative de fréquences. De la même manière sa représentation graphique se révèle également très utile. Une distribution cumulative de fréquences des données sur les poissons est donnée au tableau 4.3. Une représentation graphique de distribution cumulative de fréquences est appelée une **ogive** (moins que ou plus que), et l'ogive moins que du tableau 4.3 est donnée à la figure 4.10.

Tableau 4.3 Distribution cumulative de fréquences des poids de poissons (kg)

Classe	Fréquence	Fréquence cumulée (moins que)	Fréquence cumulée (plus que)
2,0–2,9	7	7	63
3,0–3,9	19	26	56
4,0–4,9	16	42	37
5, 0–5,9	12	54	21
6,0–6,9	6	60	9
7,0–7,9	3	63	3
<b>Total</b>	<b>63</b>		

Source : Tableau 3.1



Source : Données du tableau 4.3

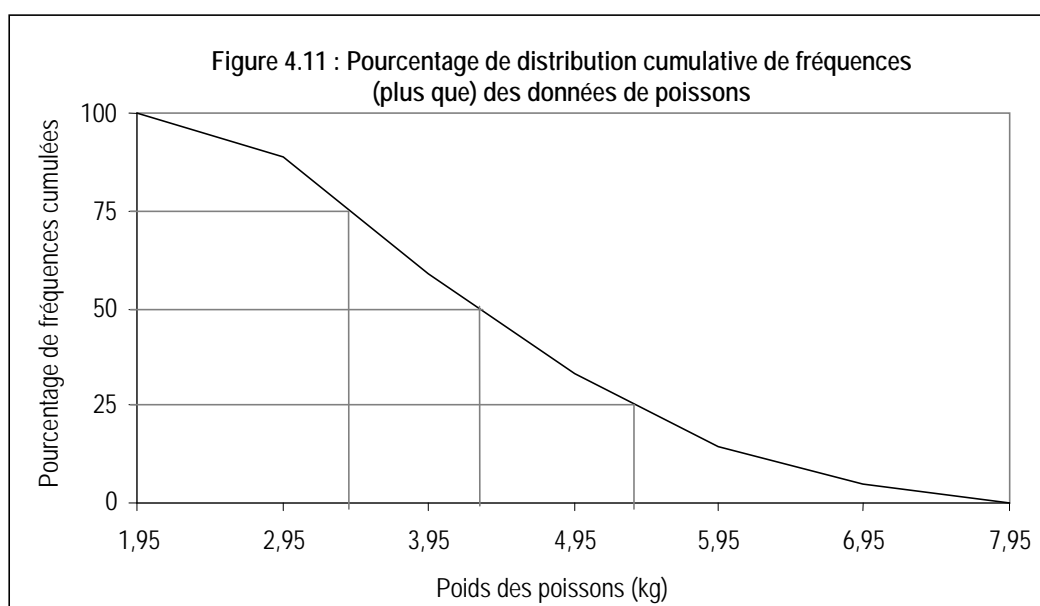
**ogive plus que**

La courbe de la figure 4.10 représente la distribution cumulative de fréquences moins que. On peut aussi dessiner un diagramme similaire pour la distribution cumulative de fréquences plus que. Celle-ci est donnée à la figure 4.11. Les données utilisées sont celles du tableau 4.4.

Tableau 4.4 Distribution cumulative de fréquences (plus que) des données de poissons (kg)

Classe	Fréquence	Pourcentage cumulatif de fréquence (plus que)
2,0–2,9	7	100,0
3,0–3,9	19	88,9
4,0–4,9	16	58,7
5,0–5,9	12	33,3
6,0–6,9	6	14,3
7,0–7,9	3	4,8
<b>Total</b>	<b>63</b>	

Source : Tableau 3.1



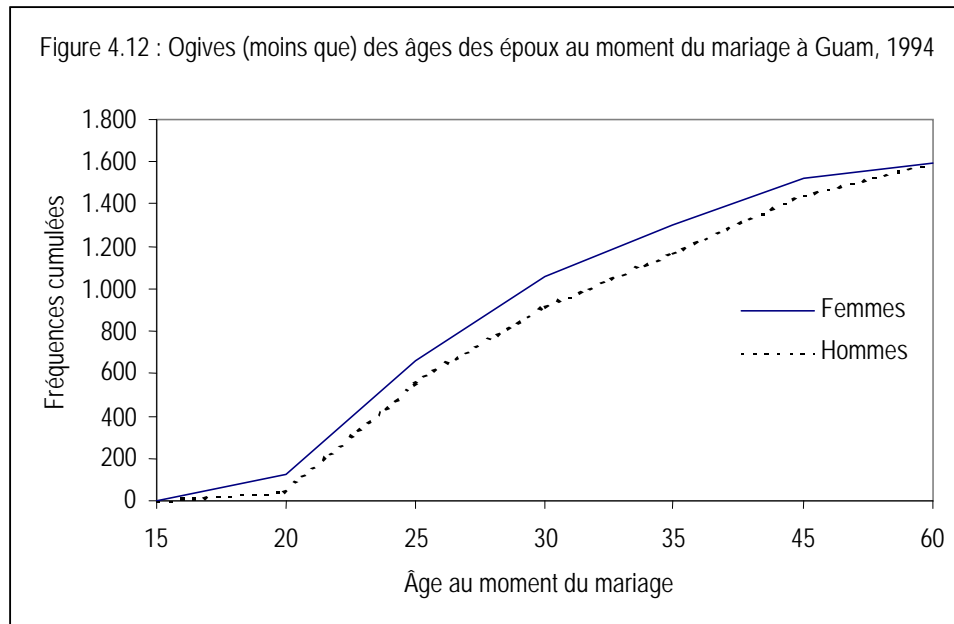
Source : Tableau 4.4

**les fréquences cumulées montrent la distribution des données**

À la figure 4.11, au lieu de tracer les fréquences réelles sur l'axe vertical, on a tracé les fréquences cumulées. On peut voir clairement sur ce diagramme que 100% des poissons pèsent plus d'1,95 kilos; 75% sont estimés à plus de 3,4 kilos; 50% sont estimés à plus de 4,4 kilos, et 25% sont estimés à plus de 5,4 kilos. Ceci est la représentation graphique de quartiles, et sera vue en détail au chapitre "Plus d'infos sur les mesures de position".

**deux distributions**

Finalement, regardons l'ogive des âges au moment du mariage à Guam du tableau 3.14.



Source : Tableau 3.14

**comparaisons entre les variables et les observations**

Comme les deux distributions ont la même base, il est utile de montrer leurs fréquences cumulées sur le même diagramme. La figure 4.12 montre clairement l'effet 'retard', c'est-à-dire la différence dans l'âge des hommes et des femmes au moment du mariage. Les hommes tendent à se marier plus tard que les femmes, comme on le voit à l'ogive ci-dessus.

