

Statistique descriptive

TALHI. R
Service de Biostatistique, Faculté de Médecine d'Oran

Plan du cours

- I- Introduction - Rappel
- II- Les tableaux statistiques
- III- Les représentations graphiques

Introduction

La distribution statistique représente des observations classées selon le ou les caractères étudiés.

Cette distribution est présentée sous forme de **tableaux statistiques** ou de **graphiques**.

Le type de représentation graphique dépend de la nature du caractère étudié.

Caractères et variables

Dans une population, par exemple celle des étudiants d'une faculté, les unités sont repérées par le nom et le prénom (liste) des étudiants. Si l'on souhaite étudier cette population, on va retenir certains critères d'étude comme la filière, l'âge, la taille, le poids, la nationalité, le groupe sanguin...

Parmi ces critères, certains sont **quantitatifs** et d'autres sont **qualitatifs**. Afin de différencier les deux types de critères, les critères qualitatifs sont appelés des **caractères** et les critères quantitatifs des **variables**.

Modalités d'un caractère

On désigne par **modalités** les différentes catégories d'un caractère qualitatif, et on qualifie de **valeurs** les différents chiffres d'une variable.

Exemple:

Caractère: sexe

Modalités: masculin, et féminin.

Un caractère est dit qualitatif

quand ses différentes modalités échappent à la mesure.

Exemple : couleur des yeux, profession, nationalité, groupe sanguin...

Le caractère qualitatif

Nominal:

Les classes sont prédéfinies, nommées mais pas ordonnées.
Exemple: le sexe, le groupe sanguin...

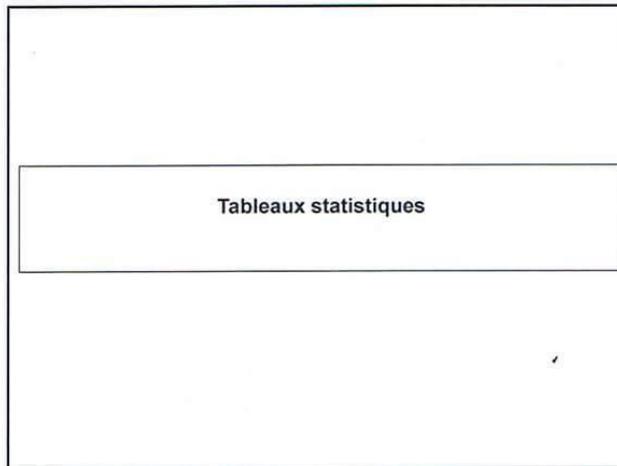
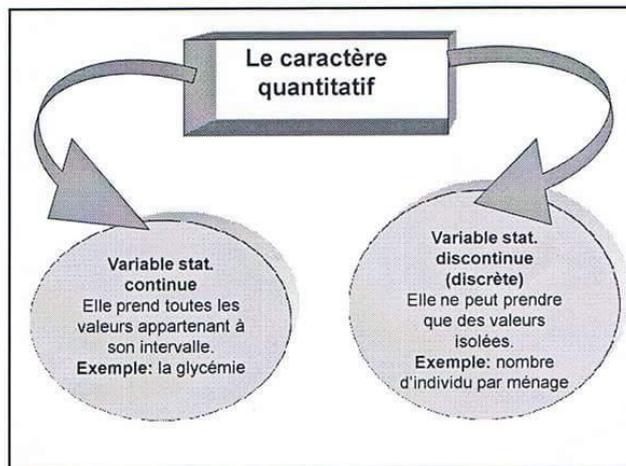
Ordinal :

Les classes peuvent être ordonnées selon une échelle de valeurs
Exemple: le niveau d'instruction, la forme clinique d'une maladie...

Un caractère est dit quantitatif

quand ses différentes modalités sont mesurables (exprimées par des chiffres).

Exemple : taille, poids, nombre d'enfants par famille...



Représentation tabulaire

Le caractère d'une population peut être présenté par un tableau, appelé tableau statistique des données.

Les observations sont résumées dans ce tableau où le nombre de dimensions est en liaison avec les modalités du caractère étudié.

Principe de base :

- Un tableau doit être simple, compréhensible
- Un tableau doit avoir :
 - un titre clair et explicite,
 - un numéro, et
 - une référence

Caractère Qualitatif Nominal

Tableau n°1 : Profil de la pathologie observée dans une consultation de médecine à Oran durant l'année 1997

Modalités du caractère (pathologie observée) x_i	Effectifs (fréquences absolues) n_i	Fréquences relatives f_i (%) ($f_i = n_i/n$)
Pathologie ORL	153	12
Intoxications	102	8
Pathologie traumatique	204	16
Pneumopathie	510	40
Dermatoses	255	20
Divers	51	4
Total	$n = \sum n_i = 1275$	$\sum f_i = 100$

Source : CHU d'Oran.

Caractère Quantitatif Discontinu

Nombre d'enfants x_i	Nombre de familles n_i	Fréquences relatives f_i (%)	Fréquences Cumulées f_i cum (%)
0	4	5	5
1	5	6	11
2	10	13	24
3	16	20	44
4	18	22	66
5	14	18	84
6	7	9	93
7	6	7	100
Total	80	100	

Tableau n°2 : Nombre d'enfants par famille

Caractère Quantitatif Continu

Tableau n° 3 : Les pesées de 50 nouveau-nés.

Poids (Kg) x_i	Effectifs n_i
2.0 - 2.5	2
2.5 - 3.0	4
3.0 - 3.5	6
3.5 - 4.0	30
4.0 - 4.5	8
TOTAL	50

Une des méthodes empiriques pour la détermination des classes d'amplitudes égales :

- **Nombre de classes** : $N = \sqrt{n}$
(n = effectif de l'échantillon)
- **Étendue** :
 e = limite supérieure de la série – limite inférieure de la série
- **Amplitude** : $k > \text{Étendue} / \text{Nombre de classes}$ ($k > e/\sqrt{n}$)

Exemple:

$n = 50$

limite supérieure de la série = 91 kg

limite inférieure de la série = 30 kg

Alors

$N = \sqrt{50} = 7$ classes

$e = 91 - 30 = 61$ kg

$k > 61/7 = 8,71 \rightarrow k = 9$ kg

Les classes de poids (kg)

30 – 39

40 – 49

50 – 59

60 – 69

70 – 79

80 – 89

90 – 99

Les représentations graphiques

Les Représentations Graphiques

Les représentations graphiques représentent l'image synthétique des données (information) présentées dans un tableaux statistique

Une présentation graphique est souvent plus facile à comprendre qu'une présentation numérique.

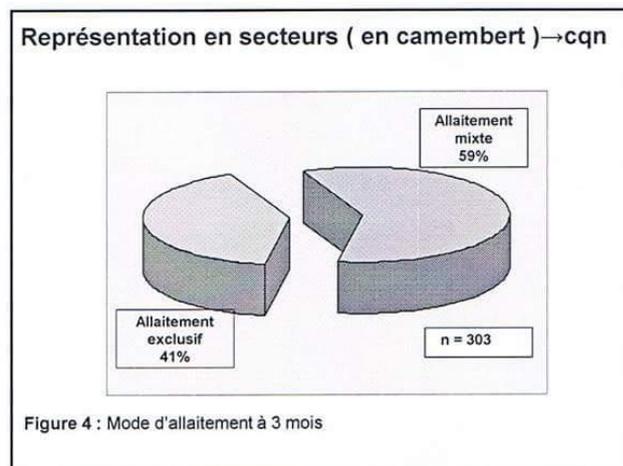
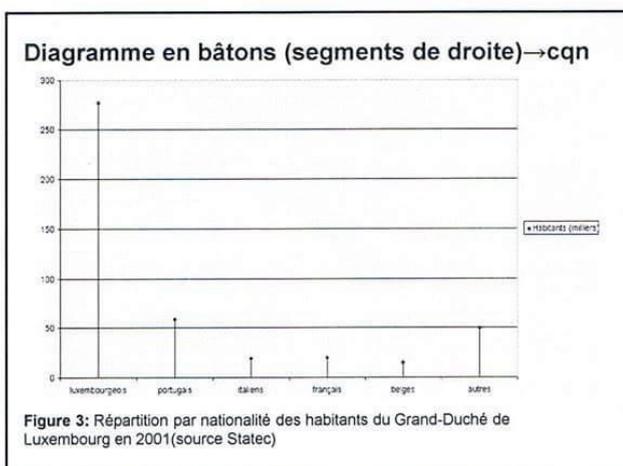
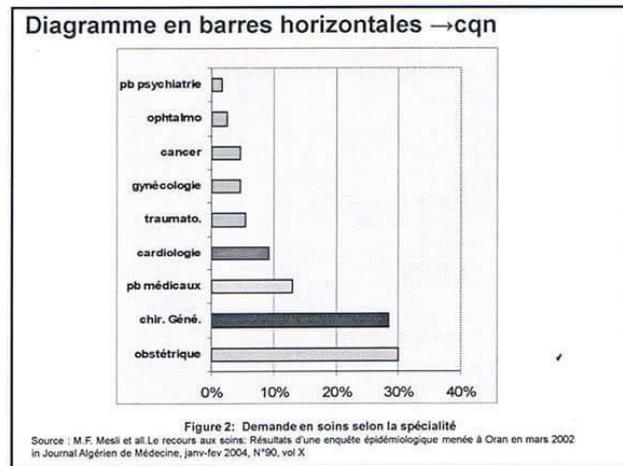
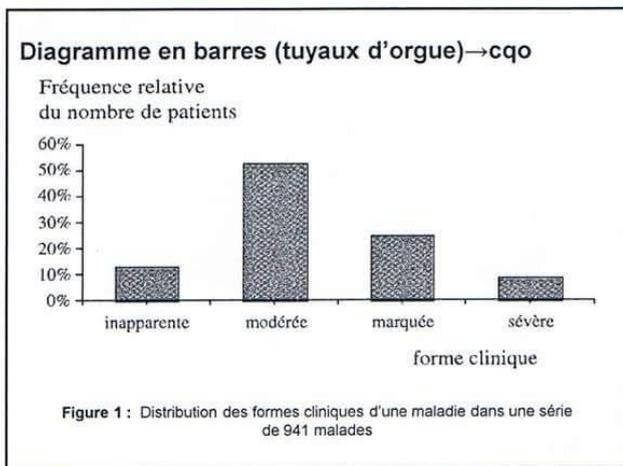
Le choix d'un type de graphe dépend de la nature des données du problème posé, et du caractère à présenter. .

Recommandations:

le graphique doit toujours être présenté clairement et respecter les normes suivantes :

- Un titre clair et explicite,
- Un numéro,
- Des axes
- Des unités de mesure
- Une échelle de chaque axe
- La source des données

Caractère Qualitatif



Caractère Quantitatif

Histogramme → vqc discrétisée

Effectif : nombre de sujets

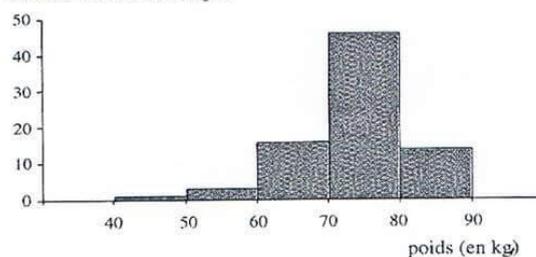


Figure 5: Distribution du poids chez 80 sujets adultes

Polygone des fréquences → vqc

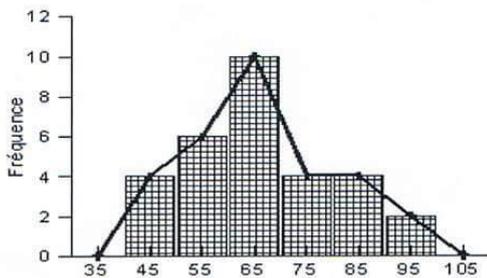


Figure 6: Répartition par tranche d'âge.....

Diagramme en bâtons (segments de droite) → vqd

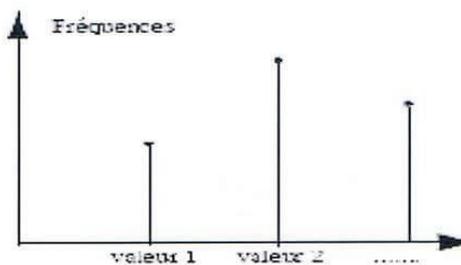


Figure 8:

Courbe des effectifs cumulés croissants (Sigmoide) → vq_c

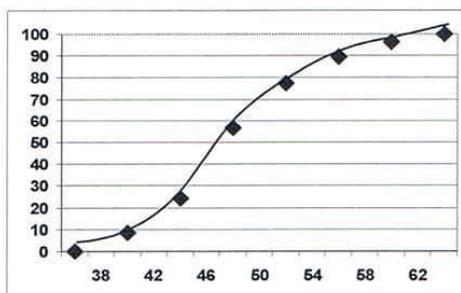


Figure 7: Courbe cumulative de la taille des membres inférieurs de 58 garçons.

Pyramide → distribution par âge et par sexe d'une population

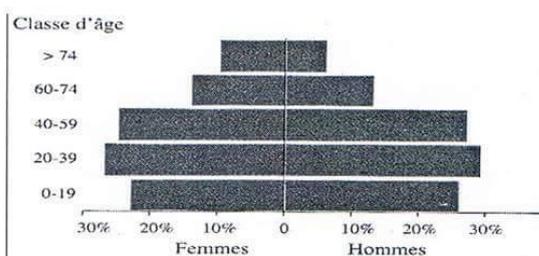


Figure 10 : Pyramide des âges - Population française en 1999

Les références:

- Mesli MF, Mokhtari A. Biostatistique. Édition mai 2007
- Cuggia M. Biostatistique descriptive. Université de Rennes 2006-2007
- Mazerolle F. Statistique descriptive. Gualino éditeur, Paris 2006.
- Ancelle T. Statistique Épidémiologie. Édition 2002
- Dabis F, Drucker J, Moren A. Épidémiologie d'intervention. Édition 1992
- Schwartz D. Méthodes statistiques. 1992