# SMA3
# M18 :Probabilités et Statistiques

Pr. Mostafa El Yassa

*Chap. 1 : Statistique descriptive*

*Chap. 2 : Eléments de Probabilités*

*Chap. 3 : Variables aléatoire et loi de Probabilité*

*Chap. 4 : Lois de probabilité classiques*

## Introduction

- Nous disposons des données suivantes :

10 11 10 11 9 9 9 12 10 13 11 14 15 10 11 12 13 11 9 8

- Pour exploiter ces valeurs nous devons les organiser :
  - Classement :

8 9 9 9 9 10 10 10 10 11 11 11 11 11 12 12 13 13 14 15

  - En tableau :

| | | |
|---|---|---|
| 8 | 1 | 5% |
| 9 | 4 | 20% |
| 10 | 4 | 20% |
| 11 | 5 | 25% |
| 12 | 2 | 10% |
| 13 | 2 | 10% |
| 14 | 1 | 5% |
| 15 | 1 | 5% |

La série ci-dessus concerne les notes de 20 étudiants.

On souhaite étudier ces donnée et en déduire les propriétés .

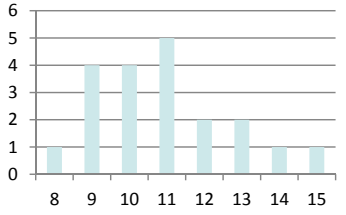
✓ La série est mise sous forme de tableau , ceci nous permet de constater que :

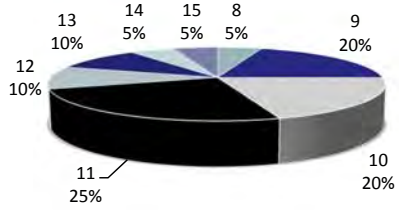
- 75% des notes sont comprise entre 9 et 12
- La note 11 est la plus fréquente

✓ À partir de cette série, on calcule quelques valeurs et indices :

- La moyenne des notes est 10,9
- L'étendue des valeurs est 7

✓ on peut aussi donnée des représentations graphiques :



## Introduction :
*définition*

- La statistique est un ensemble de méthodes et d'approches qui ont été élaborées afin de décrire au mieux et d'analyser le comportement de phénomènes aléatoires à partir d'une série d'observations (données).
- La statistique est utilisée dès qu'un phénomène est trop complexe ou encore trop bruité pour accéder à une description analytique et une modélisation déterministe

3

Statistique vient du mot latin status qui signifie état, situation. Les premières ébauches de la statistique remontent

aux recensements qui furent mis sur pied dès les premiers siècles de notre ère. Ce n'est pourtant qu'au 18e siècle

qu'elle se constitue comme une discipline scientifique autonome.

Aujourd'hui, la statistique est une branche des mathématiques appliquées en liaison avec le calcul des probabilités mais qui, à la différence de ce dernier, est basée sur des observations d'événements réels (statistique descriptive) à partir desquelles on cherche à établir des

hypothèses plausibles en vue de prévisions (statistique inférentielle) concernant des circonstances analogues.

## Introduction :
*méthodes et objectifs des statistiques*

- Les méthodes utilisées par les statistiques utilisent les mathématiques et font appel à l'outil informatique
- Les statistiques ont pour objectifs :
  - Intégrer l'aléatoire pour aider à prendre une décision
  - Prévoir en présence du hasard

4

- _____
- _____
- _____
- _____
- _____
- _____
- _____

## Introduction :
*approches*

- Statistique exploratoire, statistique descriptive ou analyse des données :
  - Ressortir des propriétés des données étudiées
  - Suggérer des hypothèses
- Statistique inférentielle ou statistique mathématique :
  - Étendre les propriétés constatées sur un échantillon à toute la population
  - Vérifier l'adéquation des hypothèses a priori ou issues d'une phase exploratoire

5

- ✓ Statistique exploratoire : Ensemble de méthodes dont l'objectif est la description des données à travers leur présentation synthétique, leurs représentation graphique et le calcul de résumés numériques
- ✓ Statistique inférentielle : des méthodes dont l'objectif est d'étudier un phénomène sur une population globale à partir de son observation sur un échantillon
- _____
- _____
- _____
- _____
- _____
- _____

## Introduction :
*Terminologie de base*

- Population ou champ d'étude Ω : c'est l'ensemble concerné par l'étude statistique
- Individu ω : c'est un élément ou membre de la population
- Échantillon : c'est un sous ensemble de la population sur lequel sont réalisées les observations
- Taille de l'échantillon : est le nombre d'individus de l'échantillon; c'est la cardinalité de l'échantillon

6

- _____
- _____
- _____
- _____
- _____
- _____
- _____

## Introduction :
*Terminologie de base*

- Enquête : opération qui consiste à observer tous les individus d'un échantillon
  - Recensement : est une enquête exhaustive; l'échantillon observé est la population entière
  - Sondage : est une enquête sur une partie de la population
- Variable ou Caractère : Ce qui est observé ou mesuré sur les individus d'une population. Une variable peut être quantitative ou qualitative

7

## Introduction :
*Terminologie de base*

- Variable ou Caractère : est ce qui est observé ou mesuré sur les individus d'une population.
  - Modalité : est une valeur possible prise par une variable.
  - Domaine d'une variable : ensemble des valeurs possibles ou des modalités est appelé le.
  - Une variable peut être quantitative ou qualitative
    - Une variable est dite quantitative lorsque ses modalités sont numériques sur lesquelles on peut effectuer des opérations algébriques
    - Une variable est dite qualitative lorsque ses modalités expriment une qualité et non pas une valeur numérique

Méthodes statistiques 8

- _____
- _____
- _____
- _____
- _____
- _____
- _____

Comme on utilise souvent le mot variable dans le cadre de l'étude de la statistique, il est important de comprendre sa signification. Une variable est une caractéristique observée ou mesurée sur les individus d'une population ou un échantillon.

La taille, l'âge, le revenu, la lieu de naissance, les années d'études, la couleur des yeux, sont tous des exemples de variables.

On distingue ainsi deux grandes catégories de variables: des variables qualitatives et des variables quantitatives. Par ailleurs, au sein de ces deux catégories, une distinction plus fine peut encore être envisagée, ce qui permet de considérer différents niveaux de mesure d'une variable et, par conséquent, différents types d'échelles.

## Introduction :

*Terminologie de base*

- Une variable peut être quantitative ou qualitative
  - Une variable est dite quantitative lorsque ses modalités sont numériques et sur lesquelles on peut effectuer des opérations algébriques. Une variable quantitative peut être discrète ou continue.
    Les variables statistiques discrètes sont des variables qui ne peuvent prendre que des valeurs isolées.
    Les variables continues peuvent prendre toutes les valeurs numériques possibles d'un ensemble inclus dans $\mathbb{R}$
  - Une variable est dite qualitative lorsque ses modalités expriment une qualité et non pas une valeur numérique. On peut coder les modalités une telle variable par des valeurs numériques.

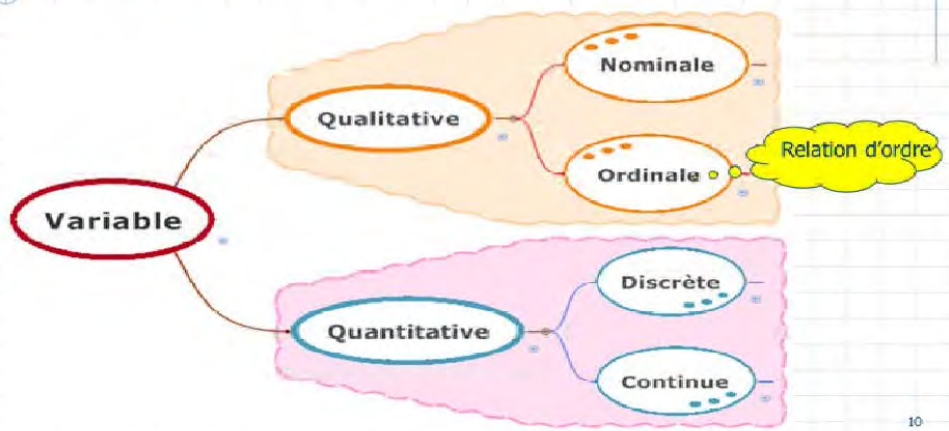
9

Comme on utilise souvent le mot variable dans le cadre de l'étude de la statistique, il est important de comprendre sa signification. Une variable est une caractéristique observée ou mesurée sur les individus d'une population ou un échantillon.

La taille, l'âge, le revenu, la lieu de naissance, les années d'études, la couleur des yeux, sont tous des exemples de variables.

On distingue ainsi deux grandes catégories de variables: des variables qualitatives et des variables quantitatives. Par ailleurs, au sein de ces deux catégories, une distinction plus fine peut encore être envisagée, ce qui permet de considérer différents niveaux de mesure d'une variable et, par conséquent, différents types d'échelles.

## Typologie des variables




10

Les données **qualitatives** définissent des échelles soit **nominales** soit **ordinales**.

- ✓ L'échelle **nominale** comporte un certain nombre de valeurs, dont la seule propriété est qu'elles sont toutes différentes les unes des autres. Par exemple : la variable **SEXE**, la variable **NATIONALITÉ**, la variable **COULEUR DES YEUX,** la variable **GROUPE SANGUIN**.
- ✓ Dans le cas d'une échelle **ordinale**, en revanche, les valeurs possibles peuvent être classées dans un ordre spécifique ou dans un ordre naturel quelconque, établie en fonction d'un critère donné. Par exemple : la variable **MENTION,** est ordinale parce que la valeur « Très bien » est meilleure que la valeur « Bien » et « Bien » est meilleur que « Assez bien », etc.

Les données **quantitative** ont des valeurs numériques, comme la variable **ÂGE** ou la variable **NOMBRE D'ENFANTS**. Toutefois, on ne considère pas que toutes les variables décrites par des nombres sont des variables quantitatives. Par exemple, lorsqu'on vous demande d'indiquer votre niveau de satisfaction par une valeur allant de 1 à 5, vous utilisez des nombres, alors que la variable « satisfaction » est en fait une variable ordinale. Les variables numériques peuvent être **continues** ou **discrètes**.

- ✓ Une variable est continue lorsque elle peut avoir un nombre infini de valeurs réelles. La distance, l'âge et la température en sont des exemples
- ✓ Une variable est discrète lorsque elle ne peut avoir qu'un nombre défini de valeurs réelles. Par exemple : la variable **NOMBRE D'ENFANTS**, la variable **NOTE.**

## Introduction :
## *plan d'échantillonnage*

- La façon de procéder pour sélection les éléments d'un échantillon est appelée plan d'échantillonnage; parmi les méthodes d'échantillonnage on a :
  - Échantillonnage aléatoire simple :
    Les individus sont choisis de telle sorte que chaque membre de la population a une chance égale de figurer dans l'échantillon.
  - Échantillonnage par quotas (L'échantillonnage stratifié):
    On subdivise la population en strates (groupes relativement homogènes) qui sont mutuellement exclusives. Dans chacune des strates, on choisit au hasard le nombre nécessaire d'individus de telle sorte que les proportions des strates dans la population soient respectées.

11

- ✓ Échantillonnage aléatoire : procédure d'échantillonnage telle que chaque individu de la population a la même chance (Probabilité) d'être choisi pour faire partie de l'échantillon.

- ✓ Échantillonnage par quotas: méthode d'échantillonnage raisonné fondé sur les caractéristiques de personnes interrogées; la répartition de ces caractéristiques descriptives de l'échantillon doivent être similaires à celles de la population totale. La structure de l'échantillon par quotas est définie a priori. Les quotas peuvent être fondés sur des proportions de la population. Si une population, par exemple, compte 80 hommes et 120 femmes et s'il faut en prélever un échantillon de 20 personnes, il se peut que vous vouliez respecter dans l'échantillon les proportions entre les sexes, ce qui donnerait 8 hommes et 12 femmes.

## Introduction :
## *Fluctuation d'échantillonnage*

- Considérons une population de 400 entiers compris entre 1 et 6 dont la distribution est :


| Valeurs | Effectifs |
|---|---|
| 1 | 37 |
| 2 | 95 |
| 3 | 134 |
| 4 | 79 |
| 5 | 28 |
| 6 | 27 |
| Moyenne | 3,117 |

- Nous allons extraire de cette population deux échantillons de 40 individus par un tirage aléatoire sans remise

12

- _____
- _____
- _____
- _____
- _____
- _____
- _____

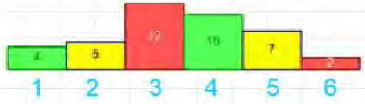
## Introduction :
### *Fluctuation d'échantillonnage*

- Échantillon 1 :

| Valeur | Fréquence |
|---|---|
| 1 | 3 |
| 2 | 11 |
| 3 | 13 |
| 4 | 10 |
| 5 | 1 |
| 6 | 2 |
| Moyenne | 3,025 |


- Échantillon 2 :

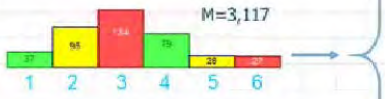
| Valeur | Fréquence |
|---|---|
| 1 | 4 |
| 2 | 5 |
| 3 | 12 |
| 4 | 10 |
| 5 | 7 |
| 6 | 2 |
| Moyenne | 3,425 |



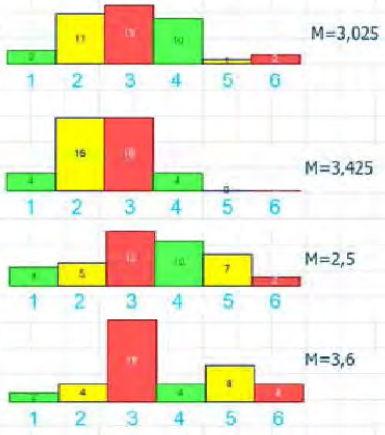

13

- _____
- _____
- _____
- _____
- _____
- _____
- _____

## Introduction :
### *Fluctuation d'échantillonnage*

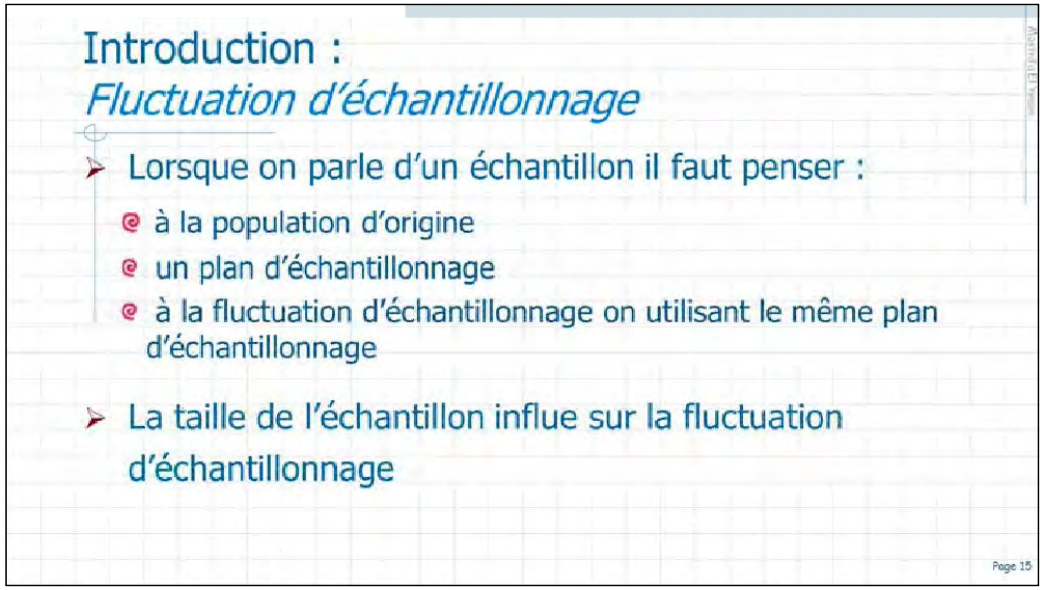


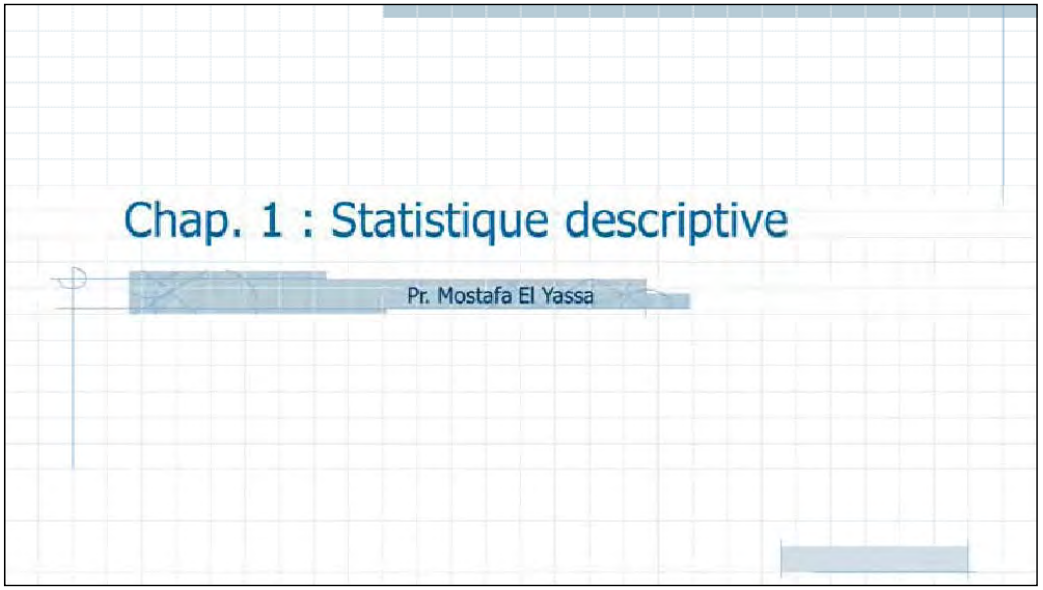

Population d'origine




14

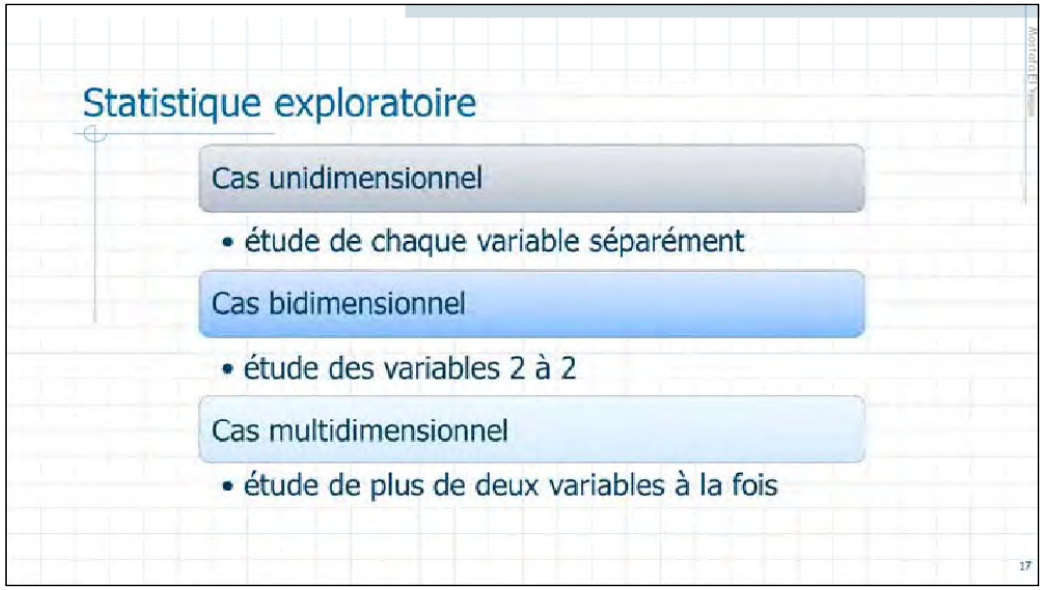
Quand on veut connaître la proportion p d'un caractère dans une grande population, il est long et coûteux de faire une enquête sur tous les individus. On prélève alors un **échantillon**.
Prélever un échantillon de taille n dans la population, c'est prendre simplement n individus (ou répéter n fois une expérience dans des conditions identiques) sur lesquels on mesure le caractère étudié.
La série statistique composée des n résultats obtenus constitue un échantillon de taille n.
Cette méthode ne peut pas fournir la valeur exacte de p, car des échantillons différents peuvent donner des proportions différentes.
Si on dispose de **plusieurs échantillons**, on peut observer ces **différences sur leurs distributions de fréquences**. C'est ce qu'on appelle la fluctuation d'échantillonnage et il suffit, pour l'observer, de prélever deux échantillons.

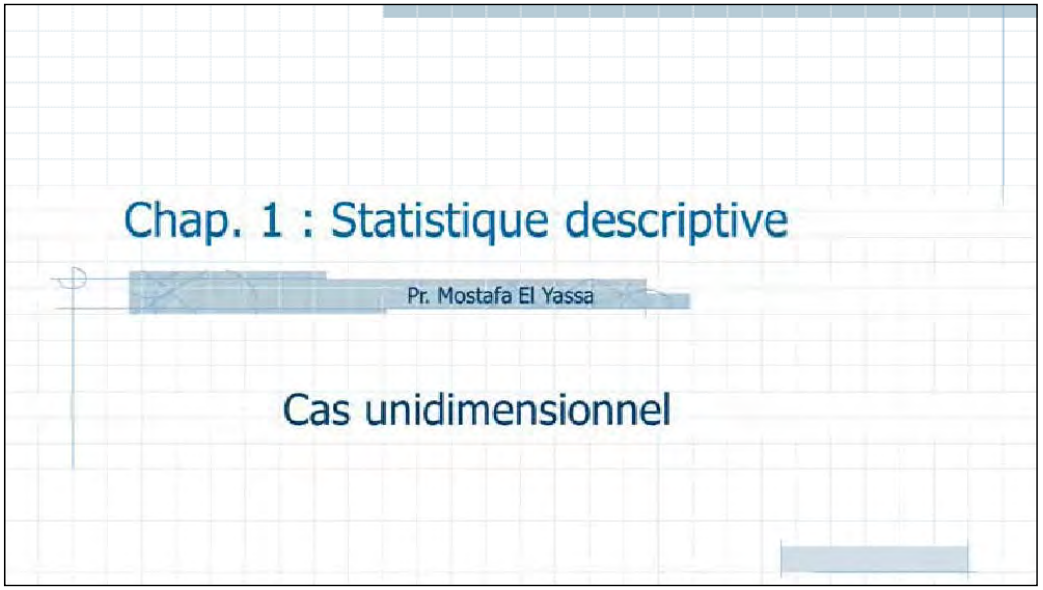



Le recours à un échantillon répond en général a la nécessité pratique (manque de temps, de place, évaluation destructive d'une production...) ou économique (coût trop élevé) d'éviter l'étude exhaustive de la population.

L'acte de sélection s'appelle l'échantillonnage. Comme il s'agit en d'être en mesure de généraliser des conclusions sur la population tout entière général à partir des résultats ou mesures obtenus sur l'échantillon, l'échantillonnage doit garantir la qualité des conclusions étendues.




En statistique exploratoire, on utilise trois techniques pour mettre en évidence les propriétés et les caractéristiques de l'échantillon étudié :

- ✓ présentation en tableau (tableau statistique)
- ✓ représentation graphique
- ✓ calcul de valeurs caractéristiques

En outre, les outils utilisés sont conditionnée par le type de la variable à étudier qui peut être :

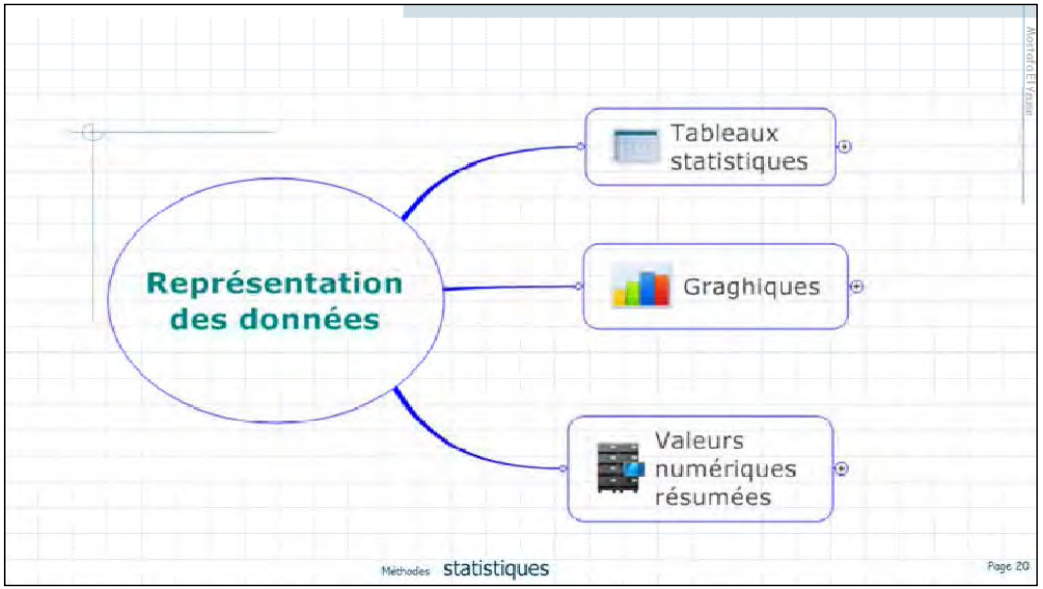
- ✓ qualitative nominale
- ✓ qualitative ordinale
- ✓ quantitative discrète
- ✓ quantitative continue

## Chap. 1 : Statistique descriptive

*Objectif*

- On dispose d'un échantillon de n individus notés $\omega_i$ ; i=1,...,n; X est une variable statistique et $X(\omega_i)$ est la valeur prise par cette variable sur l'individu $\omega_i$ ; $\{X(\omega_i)$ , i=1,...,n$\}$ est la série statistique brute
- La statistique descriptive est un ensemble de méthodes permettant de décrire, présenter, résumer des données souvent très nombreuses.
  Ces méthodes peuvent être numériques (tris, élaboration de tableaux, calcul de moyennes...) et mener à des représentations graphiques.

Page 19




Page 20

- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____

- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____

## Tableau statistique

*Variable qualitative nominale*

| Modalité | Effectif | Fréquence |
|---|---|---|
| $M_1$ | $n_1$ | $f_1$ |
| $M_2$ | $n_2$ | $f_2$ |
| $M_3$ | $n_3$ | $f_3$ |
| ⋮ | ⋮ | ⋮ |
| $M_{k-1}$ | $n_{k-1}$ | $f_{k-1}$ |
| $M_k$ | $n_k$ | $f_k$ |
| Total | $\sum_i n_i = n$ | $\sum_i f_i = 1$ |

Ligne des titres

Effectif de la modalité $M_3$

Fréquence de la modalité $M_3$ $f_2 = n_2 / n$

Ligne des totaux

21

- La première colonne comporte l'ensemble des **modalités** distinctes prises par la variable X .
- La deuxième colonne comporte les **effectifs** : le nombre d'occurrences ( apparitions) de chaque modalité, on les notes $\mathbf{n_i}$.
- Les effectifs peuvent être remplacés par les fréquences : $\mathbf{n_i / n}$ , $n = \Sigma n_i$
- On choisira les fréquences lorsque la taille de l'échantillon est grande
- _____
- _____

## Exemple de tableau statistique

*variable qualitative nominale*

On s'intéresse à la variable « état-civil » notée X et à la série statistique des valeurs prises par X sur 20 personnes. Considérons la série statistique suivante :

M; M; D; C; C; M; C; C; C; M; C; M; V; M; V; D; C; C; C; M

La codification est

C : célibataire, M : marie(e),
V : veuf (ve), D : divorcée.

**Tableau Statistique**

| $x_i$ | $n_i$ | $f_i$ |
|---|---|---|
| C | 9 | 0,45 |
| M | 7 | 0,35 |
| V | 2 | 0,10 |
| D | 2 | 0,10 |
| Total | 20 | 1 |

22

- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____

## Tableau statistique

*Variable qualitative ordinale ou quantitative discrète*

Ligne des titres

Modalités rangées par ordre croissant

| Modalité | Effectif | Effectif cumulé | Fréquence | Fréquence cumulée |
|---|---|---|---|---|
| $x_1$ | $n_1$ | $n_1$ | $f_1$ | $f_1$ |
| $x_2$ | $n_2$ | $n_1 + n_2$ | $f_2$ | $f_1 + f_2$ |
| $x_3$ | $n_3$ | $n_1 + n_2 + n_3$ | $f_2$ | $f_1 + f_2 + f_3$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| $x_{k-1}$ | $n_{k-1}$ | $n_1 + \cdots + n_{k-1}$ | $f_{k-1}$ | $f_1 + \cdots + f_{k-1}$ |
| $x_k$ | $n_k$ | $n$ | $f_k$ | 1 |
| Total | $\sum_{i=1} n_i = n$ | | $\sum_i f_i = 1$ | |

Ligne des totaux

Méthodes statistiques

Page 23

- ✓ La première colonne comporte l'ensemble des modalités ou des valeurs distinctes prises par X, rangées par ordre croissant
- ✓ La deuxième colonne comporte les effectifs. Les effectifs peuvent être remplacés par les fréquences $n_i/n$.
- ✓ La troisième colonne comporte les effectifs cumulés (ou les fréquences cumulées)

**L'avantage d'utiliser les fréquences et les fréquences cumulées est que la distribution ne dépend pas de la taille de l'échantillon.**

- _____
- _____
- _____
- _____
- _____

## Exemple de tableau statistique

*variable qualitative ordinale*

On interroge 50 personnes sur leur dernier diplôme obtenu. Considérons la série statistique suivante : Sd; Sd; P; P; P; P; P; P; Se; Se; P; P; Se; Su; Su; Su; Sd; Sd; Su; Su; Su; Su; P; P; Su; Su; Su; Su; Su; Su; Su; Su; Su; P; P; P; Su; Su; Su; Su; Se; Se; Se; Se; Se; Se; Se; Se; P;

La codification est :
P: Primaire, Se : Secondaire,
Sd : sans diplôme, Su : Supérieur

24

- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____

## Exemple de tableau statistique

*variable qualitative ordinale*

**Série statistique**

Sd; Sd; P; P; P; P; P;
P; Se; Se; P; P; P;
Su; Su; Su; Sd; Sd;
Su; Su; Su; Su; P; P;
Su; Su; Su; Su; Su;
Su; Su; Su; Su; Su; P;
P; P; Su; Su; Su; Su;
Se; Se; Se; Se; Se;
Se; Se; Se; Se;

**Tableau Statistique**

| $x_i$ | $n_i$ | $N_i$ | $f_i$ | $F_i$ |
|---|---|---|---|---|
| Sd | 4 | 4 | 0,08 | 0,08 |
| P | 14 | 18 | 0,28 | 0,36 |
| Se | 11 | 29 | 0,22 | 0,58 |
| Su | 21 | 50 | 0,42 | 1,00 |
| Total | **50** | | **1** | |

25

- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____

## Exemple de tableau statistique

*variable quantitative discrète*

Un quartier est composé de 50 ménages, et la variable $X$ représente le nombre de personnes par ménage.

Les valeurs observées de la variable sont :

1; 1; 4; 4; 4; 4; 2; 2; 2; 8; 5; 5; 5; 5; 2; 2; 3; 3; 3; 3; 3; 3; 3; 3; 2; 2; 2; 2; 3; 3; 3; 4; 4; 4; 4; 4; 4; 5; 5; 6; 6; 8; 1; 1; 1; 3; 3; 3; 6; 3.

26

- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____

## Exemple de tableau statistique

*variable quantitative discrète*

**Série statistique**

1; 1; 4; 4; 4; 4; 2;
2; 2; 4; 5; 5; 5; 5;
2; 2; 3; 3; 3; 3; 3;
3; 3; 3; 2; 2; 2; 2;
3; 3; 3; 4; 4; 4; 4;
4; 4; 5; 5; 6; 6; 5;
1; 1; 1; 3; 3; 3; 6;
3.

**Tableau Statistique**

| $x_i$ | $n_i$ | $N_i$ | $f_i$ | $F_i$ |
|---|---|---|---|---|
| 1 | 5 | 5 | 0,10 | 0,10 |
| 2 | 9 | 14 | 0,18 | 0,28 |
| 3 | 15 | 29 | 0,30 | 0,58 |
| 4 | 11 | 40 | 0,22 | 0,80 |
| 5 | 7 | 47 | 0,14 | 0,94 |
| 6 | 3 | 50 | 0,06 | 1 |
| **Total** | **50** | | **1** | |

27

- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____

## Tableau statistique

*Variable quantitative continue*

- On regroupe les valeurs observées en classes (intervalles)
- Il faut définir :
  - le nombre de classes k
  - les limites des classes (bornes des intervalles) $a_1, a_2, ..., a_{k+1}$
  - et les centres de classes $x_1, ..., x_k$

28

Une variable quantitative continue peut prendre une infinité de valeurs possibles.

Le domaine de la variable est alors IR ou un intervalle de IR. En pratique,

une mesure est limitée en précision. On peut alors traiter les variables continues comme des

variables discrètes. Cependant, pour faire des représentations graphiques et construire le tableau statistique, il faut procéder à des regroupements en classes. Le tableau regroupé en classe est souvent appelé *distribution groupée*.

La répartition en classes des données nécessite de définir *a priori* le nombre

de classes *k* et l'amplitude de chaque classe. En réglé générale, on choisit au moins cinq classes de même amplitude.

- _____
- _____
- _____
- _____

## Tableau statistique

*Variable quantitative continue*

| Classe | Centre | Effectif | Effectif cumulé | Fréquence | Fréquence cumulée |
|---|---|---|---|---|---|
| $[a_1, a_2[$ | $x_1$ | $n_1$ | $n_1$ | $f_1$ | $f_1$ |
| $[a_2, a_3[$ | $x_2$ | $n_2$ | $n_1 + n_2$ | $f_2$ | $f_1 + f_2$ |
| $[a_3, a_4[$ | $x_3$ | $n_3$ | $n_1 + n_2 + n_3$ | $f_2$ | $f_1 + f_2 + f_3$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| $[a_{k-1}, a_k[$ | $x_{k-1}$ | $n_{k-1}$ | $n_1 + \cdots + n_{k-1}$ | $f_{k-1}$ | $f_1 + \cdots + f_{k-1}$ |
| $[a_k, a_{k+1}]$ | $x_k$ | $n_k$ | $n$ | $f_k$ | $1$ |
| | Total | $\sum_{i=1} n_i = n$ | | $\sum_i f_i = 1$ | |

Les classes de regroupement

Dans ce tableau statistique :

- ✓ La première colonne comporte l'ensemble des classes (intervalles) de regroupement rangées par ordre croissant
- ✓ La deuxième colonne les centres des classes
- ✓ La troisième colonne comporte les effectifs. Les effectifs peuvent être remplacés par les fréquences $n_i/n$.
- ✓ La troisième colonne comporte les effectifs cumulés (ou les fréquences cumulées)

**L'avantage d'utiliser les fréquences et les fréquences cumulées est que la distribution ne dépend pas de la taille de l'échantillon.**

- _____
- _____
- _____
- _____

## Exemple de tableau statistique

*variable quantitative continue*

On mesure la taille en centimètres de 50 élèves d'une classe :

152 152 152 153 153 154 154 154 155 155 156 156 156 156 156
157 157 157 158 158 159 159 160 160 160 161 160 160 161 162
162 162 163 164 164 164 164 165 166 167 168 168 168 169 169
170 171 171 171 171

- Nous allons grouper les valeurs en cinq classes ayant la même amplitude 4 : [151,5 ; 155,5[ , [155,5 ; 159,5[ , [159,5 ; 163,5[ , [163,5 ; 167,5[ , [167,5 ; 171,5[

30

Nous allons grouper les valeurs en cinq classes ayant la même amplitude 4 :

[151,5 ; 155,5[ , [155,5 ; 159,5[ , [159,5 ; 163,5[ , [163,5 ; 167,5[ , [167,5 ; 171,5[

- _____
- _____
- _____
- _____
- _____
- _____

## Exemple de tableau statistique

*variable quantitative continue*

**Tableau Statistique**

| $[a_i ; a_{i+1}[$ | $x_i$ | $n_i$ | $N_i$ | $f_i$ | $F_i$ |
|---|---|---|---|---|---|
| [151,5 ; 155,5[ | 153,5 | 10 | 10 | 0,20 | 0,20 |
| [155,5 ; 159,5[ | 157,5 | 12 | 22 | 0,24 | 0,44 |
| [159,5 ; 163,5[ | 161,5 | 11 | 33 | 0,22 | 0,66 |
| [163,5 ; 167,5[ | 165,5 | 7 | 40 | 0,14 | 0,80 |
| [167,5 ; 171,5[ | 169,5 | 10 | 50 | 0,20 | 1,00 |
| **Total** | | **50** | | **1,00** | |

31

- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____

## Tableau statistique

*Variable quantitative continue*

Il existent des formules qui permettent d'établir le nombre de classes :

- La règle de Sturge : $k = 1 + 3{,}3 \log_{10}(n)$
- La règle de Yule : $k = 2{,}5\sqrt[4]{n}$

L'amplitude de l'intervalle de classe est obtenue de la manière suivante :

- $Ic = \frac{x_{max} - x_{min}}{k}$

32

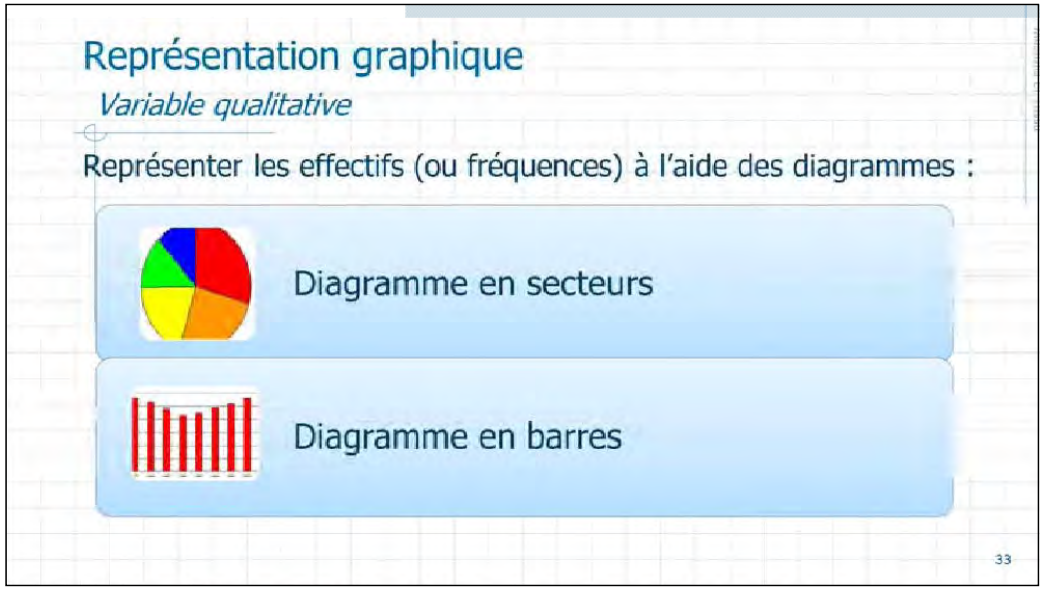
- Ici $x_{max}$ et $x_{min}$ désignent la plus grande et la plus petite valeur observées.
- Il faut arrondir le nombre de classe k à l'entier le plus proche.
- Par commodité, on peut aussi arrondir la valeur obtenue de l'amplitude de l'intervalle de classe.
- A partir de la plus petite valeur observée, on obtient les bornes de classes

en additionnant successivement l'amplitude de l'intervalle.

Exemple : si on applique le réglé de Yule à l'exemple précédent on trouve :

$k = 2{,}5\sqrt[4]{50}$ = *6,64 ≈7* et *Ic= 3*

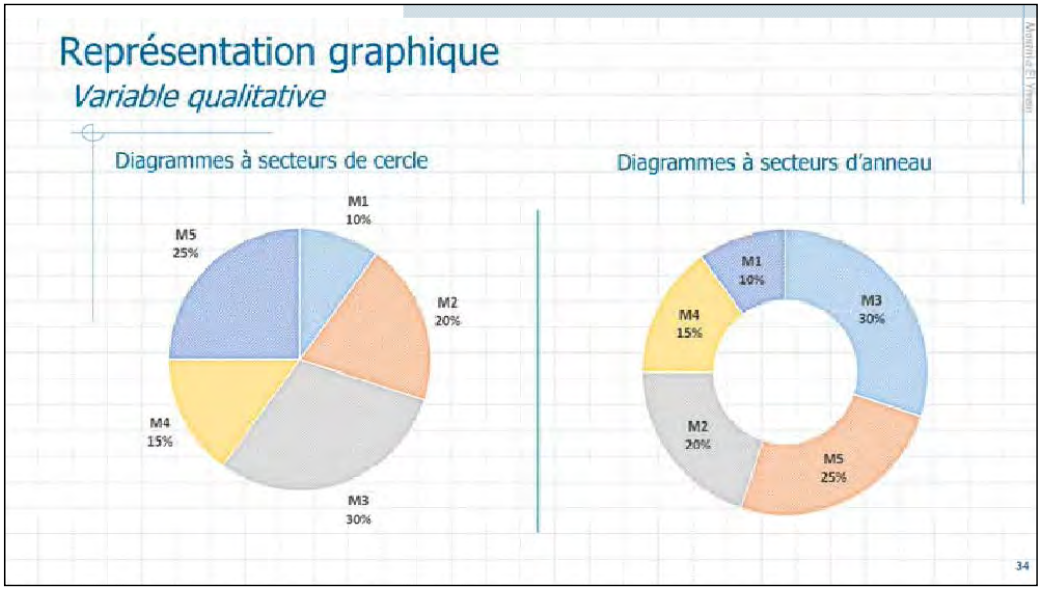
d'où les classes : [152 ; 155[ , [155 ; 158[ , [158 ; 161[ , [161 ; 164[ , [164 ; 167[ , [167 ; 170[ , [170 ; 173[




Un tableau statistique d'une variable qualitative nominale peut être représenté par deux types de graphique.

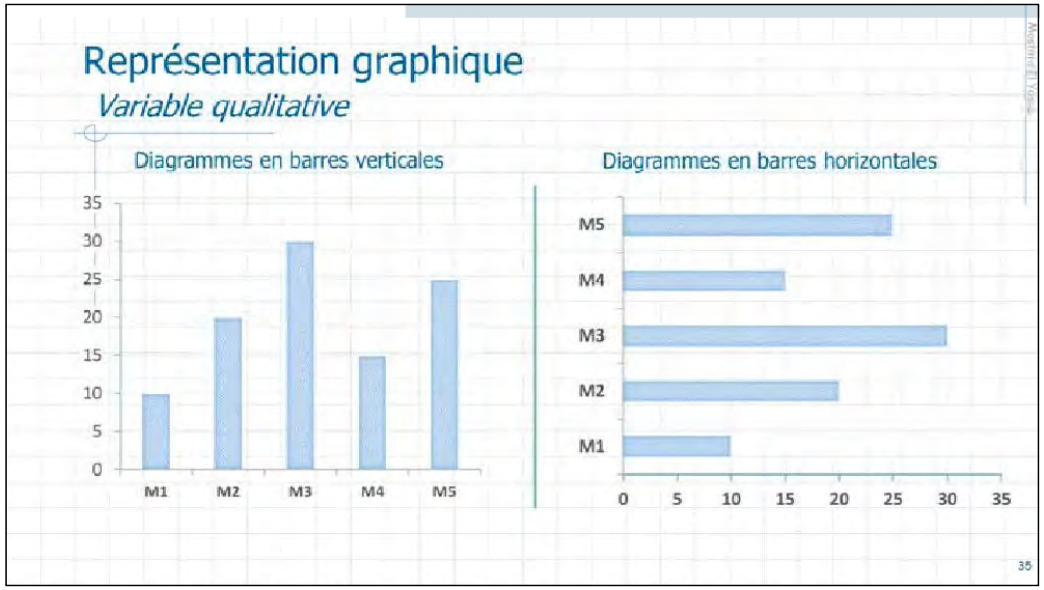
- Les effectifs sont représentés par un diagramme en barres et les fréquences par un diagramme en secteurs
- Le diagramme en barres peut être vertical ou horizontal

- _____
- _____
- _____
- _____
- _____



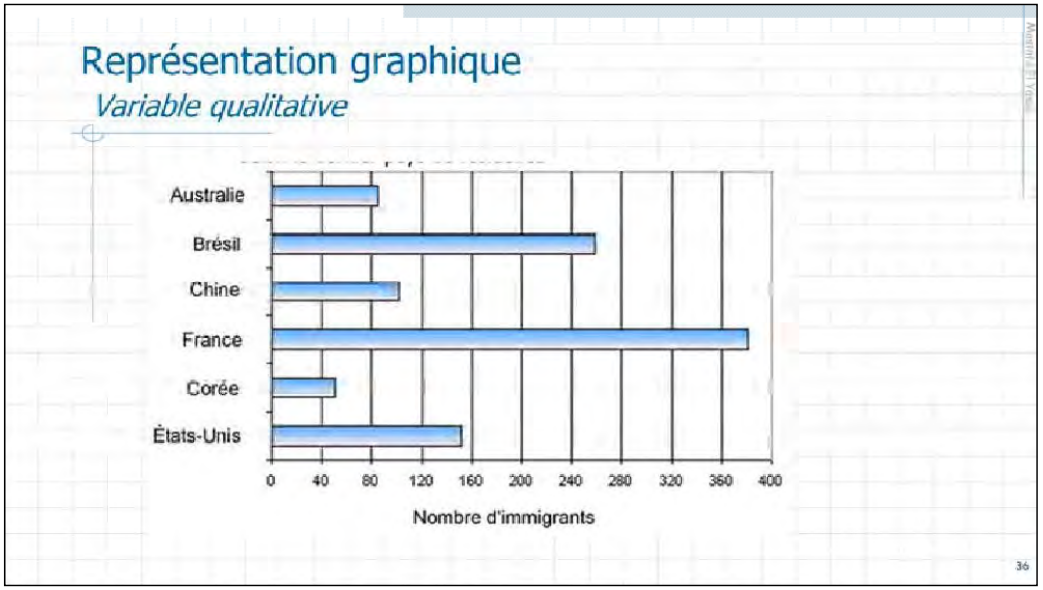

- Dans le diagramme en secteurs, chaque modalité est représenté par un secteur du cercle ou d'anneau dont la surface est proportionnelle à sa fréquence.
- Un diagramme circulaire est construit en convertissant la part de chacune des modalités en un pourcentage de 360 degrés; la formule de calcul est : ***angle en degré = fréquence x 360***
- Il est préférable de présenter un diagramme sectoriel en respectant l'ordre de grandeur des secteurs (du plus grand au plus petit) dans le sens horaire.

- _____
- _____
- _____
- _____
- _____
- _____




Le diagramme en barres peut être à barres verticales ou horizontales

- Chaque barre correspond à une modalité
  - Dans le diagramme à barres verticales la hauteur de chaque barre est égale à la valeur de l'effectif de la modalité correspondante
  - Dans le diagramme à barres horizontale la longueur de chaque barre est égale à la valeur de l'effectif de la modalité correspondante
- _____
- _____
- _____
- _____
- _____




Dans cet exemple nous avons un diagramme en barres horizontales représentant le nombre d'immigrants selon les pays;

- L'axe verticale correspond aux différentes modalités
- L'axe horizontale correspond aux effectifs (en milliers)
- _____
- _____
- _____
- _____
- _____

## Exemple de représentation graphique

*variable qualitative nominale*

Tableau Statistique

| $x_i$ | $n_i$ | $f_i$ |
|---|---|---|
| C | 9 | 0,45 |
| M | 7 | 0,35 |
| V | 2 | 0,10 |
| D | 2 | 0,10 |
| Total | 20 | 1 |

Diagramme en secteurs

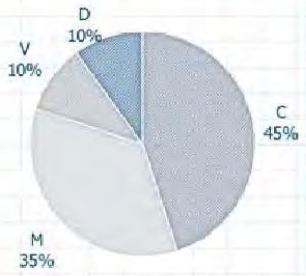
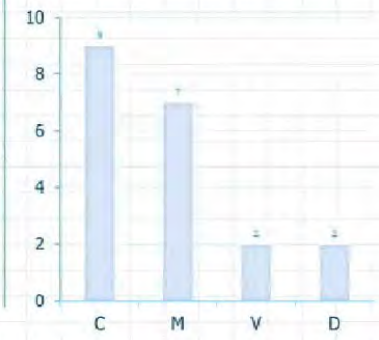



Diagramme en barres




37

## Représentation graphique

*Variable qualitative ordinale*

Représenter les effectifs (ou fréquences) et les effectifs cumulés (fréquences cumulées) :

Diagramme en secteurs des fréquences

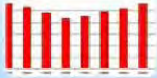


Diagramme en barres des effectifs

Diagramme en barres des effectifs cumulés

38

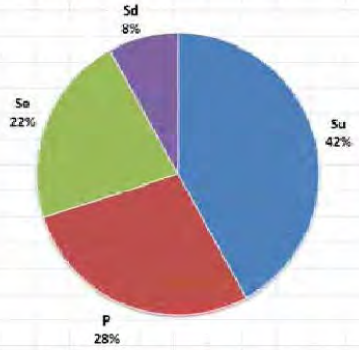
## Exemple de représentation graphique

*variable qualitative ordinale*

Tableau Statistique

| $x_i$ | $n_i$ | $N_i$ | $f_i$ | $F_i$ |
|---|---|---|---|---|
| Sd | 4 | 4 | 0,08 | 0,08 |
| P | 14 | 18 | 0,28 | 0,36 |
| Se | 11 | 29 | 0,22 | 0,58 |
| Su | 21 | 50 | 0,42 | 100 |
| **Total** | **50** | | **1** | |

Diagramme en secteurs

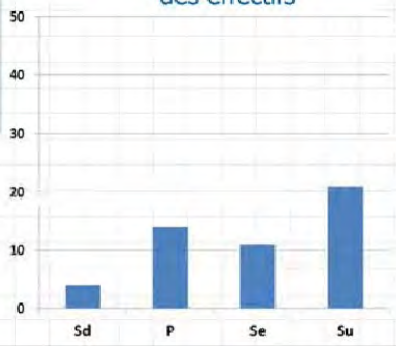



39

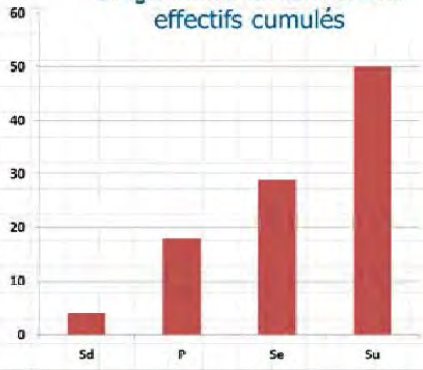
## Exemple de représentation graphique

*variable qualitative ordinale*

Diagrammes en barres des effectifs




Diagrammes en barres des effectifs cumulés




40

## Représentation graphique

*Variable quantitative discrète*

Représenter les effectifs (ou les fréquences) et les effectifs cumulés ( ou les fréquences cumulées) :

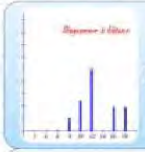


Diagramme en bâtons des effectifs

Diagramme en bâtons des effectifs cumulés

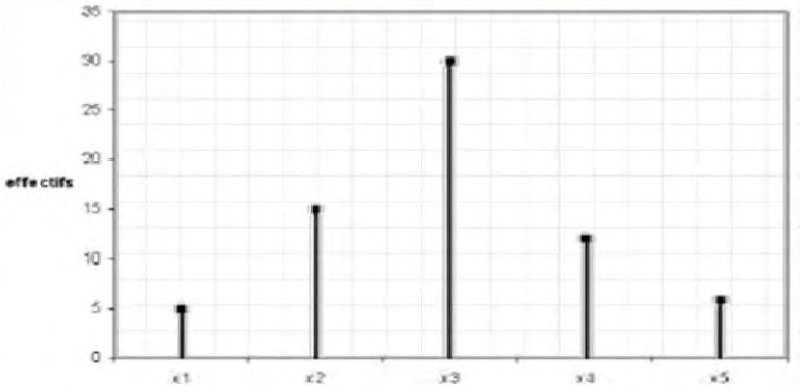
Méthodes statistiques 41

- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____

## Représentation graphique

*Variable quantitative discrète*

Diagramme en bâtons




42

Un diagramme en bâtons est une représentation graphique d'une série statistique de variable quantitative discrète.

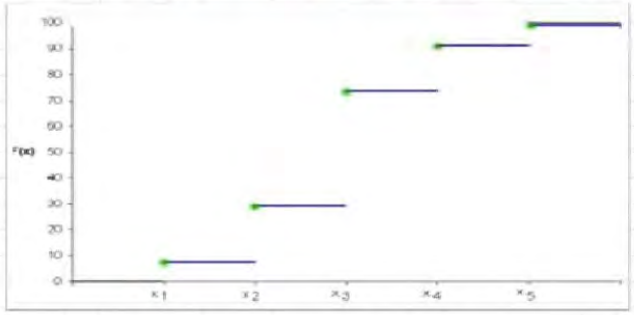
Il est constitué de segments de droite verticaux dont les hauteurs sont égales aux effectifs ou aux fréquences de chaque modalité.

- Sur l'axe des abscisses (horizontal) sont reportées les modalités de la série par ordre croisant.
- Sur l'axe des ordonnés sont reportées les effectifs ou les fréquence
- _____
- _____
- _____
- _____
- _____
- _____

## Représentation graphique

*Variable quantitative discrète*

Diagramme des fréquences cumulés



$$F(x) = \sum_{i=1}^{j} f_i \text{ , si } x_j \leq x < x_{j+1}$$

43

Un diagramme des fréquences cumulées est une représentation graphique de la fonction de répartition d'une série statistique de variable quantitative discrète :

$$F(x) = \begin{cases} 0\,, & si\ x < x_1 \\ F_j\ \text{, si } x_j \leq x < x_{j+1} \\ 1\,, & si\ x > x_n \end{cases}$$

Ce diagramme est constitué de segments de droite horizontale, paliers, dont les hauteurs sont égales aux fréquences cumulées des modalités.

Sur l'axe des abscisses sont reportées les modalités de la série.

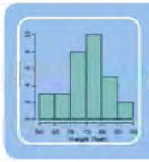
- _____
- _____
- _____
- _____
- _____



## Représentation graphique

*Variable quantitative continue*

Représenter les effectifs (ou les fréquences) et les effectifs cumulés ( ou les fréquences cumulées) :

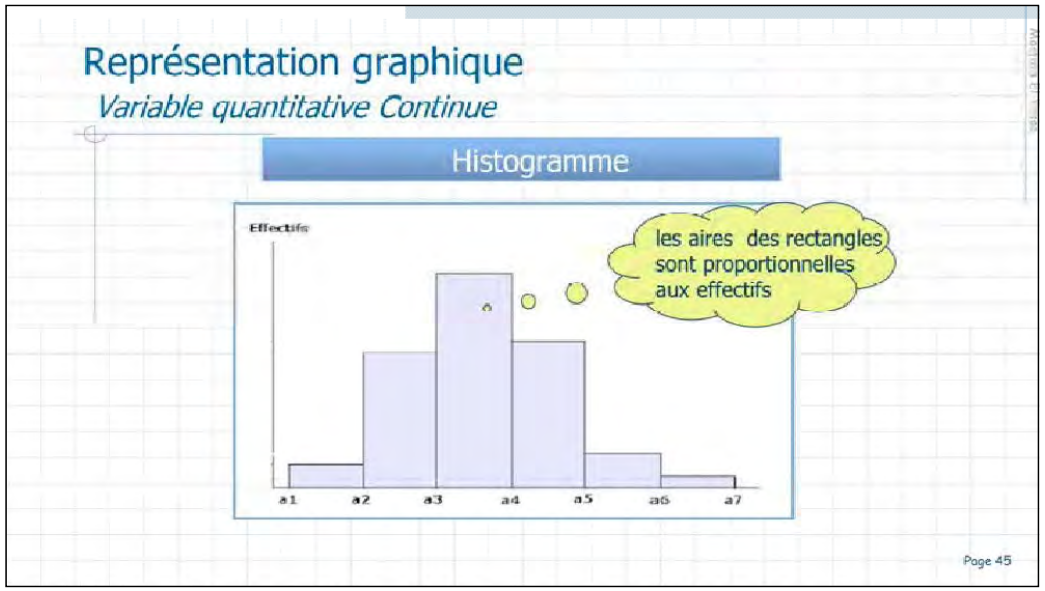


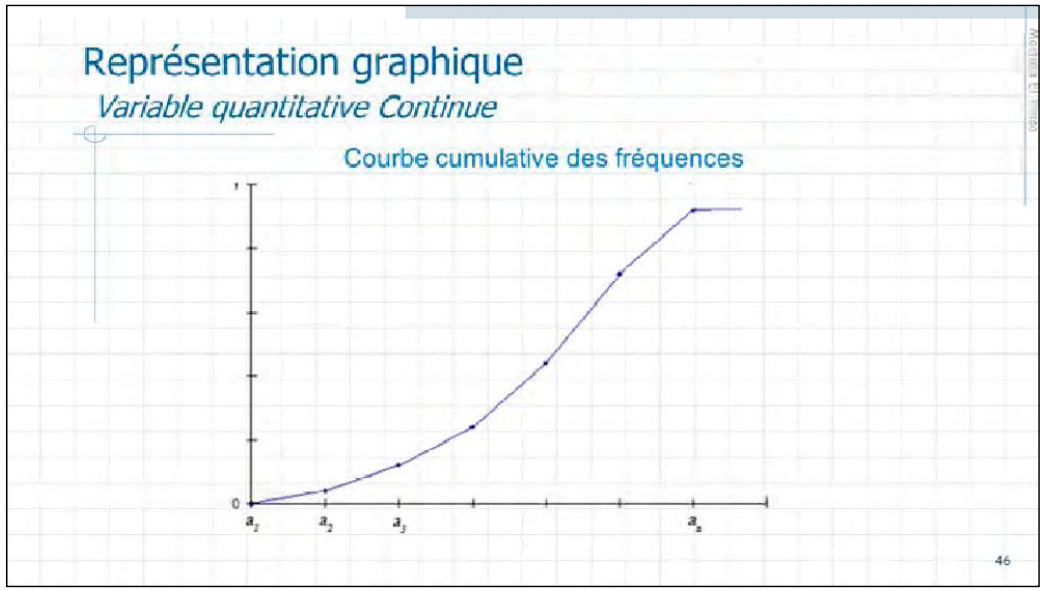
Histogramme

Histogramme des fréquences cumulées

44

- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____




- ✓ Un histogramme est une représentation graphique d'une série statistique de variable **quantitative continue**. Il est constitué de rectangles contigus dont les aires sont **proportionnelles** aux effectifs de chaque classe.
- ✓ Sur l'axe des abscisses sont reportées les bornes des classes de la série statistique.
- ✓ Notons $e_i$ l'étendue de la *i* ème classe, $f_i$ sa fréquence (ou son effectif) et $h_i$ la hauteur du rectangle de l'histogramme associé à cette classe. L'aire du rectangle de l'histogramme associé à cette classe est : $e_i \times h_i$
  Puisque les aires des rectangles sont proportionnelles aux effectifs des classes, on a $e_i \times h_i = k \times f_i$ où $k$ est le rapport de proportionnalité choisi.
  Ainsi, dans le cas où l'on choisit $k$=1, on aura $h_i = f_i / e_i$ .
- ✓ Lorsque les classes sont de même amplitude, la représentation d'un histogramme est analogue à celle d'un diagramme en bâtons : la hauteur d'un rectangle est proportionnelle à l'effectif de sa classe.
- ✓ Lorsque les classes ne sont pas de même amplitude :
  1. on calcule l'étendue de chaque classe, c'est-à-dire la longueur de chaque intervalle ;
  2. on choisit les dimensions du rectangle qui représente la classe de plus faible étendue ;
  3. on calcule le rapport de proportionnalité k
  4. Pour chaque classe, on calcule les hauteurs des autres rectangles en multipliant par k le rapport de la fréquence par l'étendue.

- Pour que chaque point, il faut prendre pour abscisses les limites supérieures des classes et, pour ordonnées, les fréquences cumulées correspondantes.
- Comme la variable statistique est continue, on tracera une courbe cumulative continue, et non une courbe en escalier, de façon qu'à une valeur de fréquence cumulée corresponde une et une seule valeur de variable.
- Entre deux points expérimentaux, on trace un segment de droite représentant l'interpolation linéaire
- _____
- _____
- _____
- _____
- _____
- _____
- _____

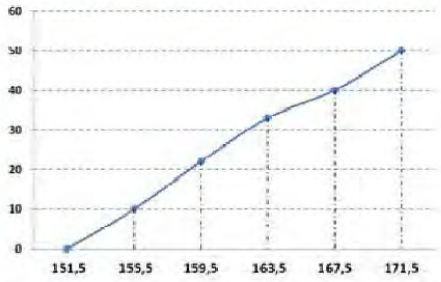
## Exemple de représentation graphique

*variable quantitative continue*

Tableau Statistique

| $[a_i ; a_{i+1}[$ | $n_i$ | $N_i$ |
|---|---|---|
| [151,5 ; 155,5[ | 10 | 10 |
| [155,5 ; 159,5[ | 12 | 22 |
| [159,5 ; 163,5[ | 11 | 33 |
| [163,5 ; 167,5[ | 7 | 40 |
| [167,5 ; 171,5[ | 10 | 50 |
| **Total** | **50** | |

Courbe cumulative des fréquences




47

Statistique exploratoire

# VALEURS CARACTÉRISTIQUES

48

- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____

## Valeurs caractéristiques
*(nombres résumés)*

Trois types de valeurs caractéristiques


Indicateurs de position


Indicateurs de dispersion


Indicateurs de forme

49

Les indicateurs sont utilisées pour résumer les caractéristiques de la série statistique étudiée :

- Les indicateurs de position permettent de savoir autour de quelles valeurs se situent les valeurs de la série statistique.
- Les indicateurs de dispersion permettent d'évaluer la répartition des valeurs de la série statistique autour des valeurs des indicateurs de position.
- Les indicateurs de forme caractérisent la forme de distribution des valeurs observées : symétrie, aplatissement, …
- _____
- _____
- _____
- _____
- _____
- _____

## Indicateurs de position
*Le mode*

- Est la modalité qui admet le plus grand effectif (la plus grande fréquence)
- Une distribution peut être unimodale ou plurimodale (bimodale, trimodale, …)
- Une distribution plurimodale peut avoir des modes locaux
- pour une distributions regroupées en classe, le calcul du mode se fait à partir de la classe modale

50

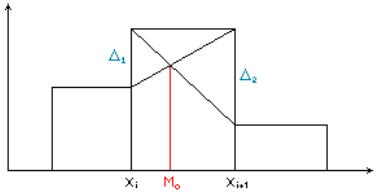
- Le mode est parfaitement défini pour une variable qualitative ou une variable quantitative discrète.
- Lorsque cette valeur est unique, on dit que la distribution est unimodale, dans le cas contraire on dit que la distribution est plurimodale.
- Une variable statistique peut présenter des modes locaux (modalités dont la fréquence est supérieure ou égale aux fréquences adjacentes). Cette situation est intéressante : elle met en évidence l'existence de plusieurs sous-populations, donc l'hétérogénéité de la population étudiée.
- Pour une variable quantitative continue ou pour une distributions regroupées en classe, le calcul du mode se fait à partir de la classe modale : c'est la classe dont la densité de fréquence est maximum :

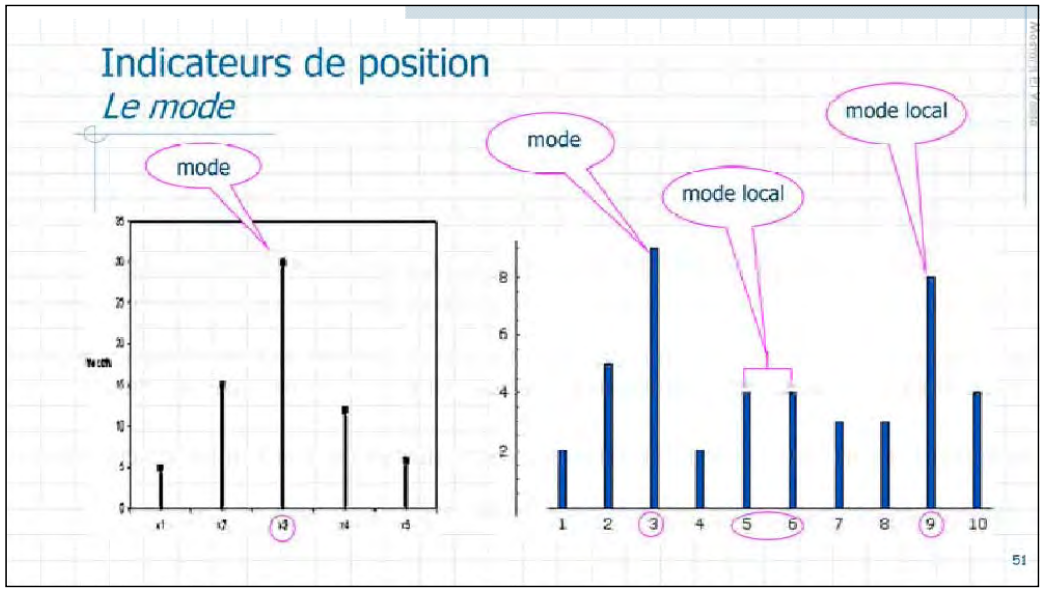
$$Densité\ de\ fréquence = \frac{Fréquence\ de\ la\ classe}{Amplitude\ de\ la\ classe}$$

Si les classes ont même amplitude la densité est remplacée par l'effectif ou la fréquence

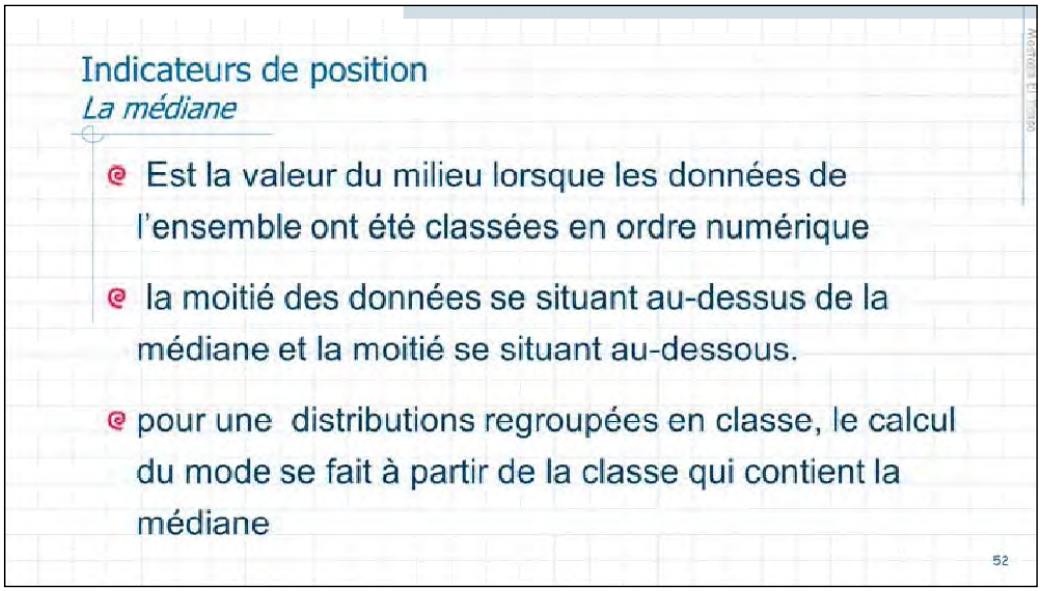
- La classe modale [ $a_i$, $a_{i+1}$ [ étant déterminée, le mode M vérifie :

$$M = a_i + \frac{\Delta_1}{\Delta_1 + \Delta_2}(a_{i+1} - a_i)$$






- ✓ La première distribution possède un seul mode égale X3 : c'est une distribution unimodale
- ✓ La seconde distribution possède plusieurs modes : un mode égale 3 et deux modes locaux (5 , 6) et 9. C'est une distribution plurimodale.
  On peut soupçonner l'existence de 3 sous-groupes dans l'échantillon observé.

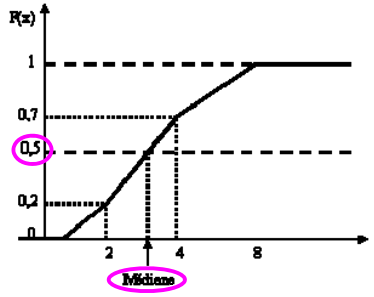



- ✓ Cette définition n'a de sens que si les modalités sont toutes ordonnées.
- ✓ Pour une variable quantitative continue ou une distributions regroupées en classe, le calcul de la médiane se fait à partir de la classe qui contient la médiane : la classe qui contient l'abscisse du point d'ordonnée n/2 ( ½ pour la fréquence ). Pour obtenir une valeur plus précise de la médiane, on procède à une interpolation linéaire à l'intérieur de la classe de la médiane.
- ✓ La classe de la médiane [ $x_i$, $x_{i+1}$ [ étant déterminée, La valeur de la médiane calculée analytiquement par la formule:

$$Médiane = x_i + (x_{i+1} - x_i)\frac{0{,}5 - F_i}{F_{i+1} - F_I} ;$$

$F_i$ *et* $F_{i+1}$ sont les valeurs des fréquences cumulées.

- ✓ Il est plus facile de lire sur les graphiques cumulatifs les abscisses des points d'ordonnée n/2 (effectif cumulé) ou ½ (fréquence cumulée).



## Indicateurs de position
*La Moyenne*

- La moyenne ne concerne que les variables quantitatives
- La moyenne est égale à la somme de toutes les valeurs observées divisée par le nombre d'observations
- Pour une distribution groupée en classes, le calcul de la moyenne utilise les centres des classes

53

Le calcul de la moyenne s'effectue directement sur les valeurs observées brutes ou bien en utilisant les effectifs ou les fréquences des observations. Ainsi, on trois formules.

✓ Utilisant les observations :

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i \; ;$$

✓ Utilisant les effectifs :

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i \; ;$$

✓ Utilisant les fréquences :

$$\bar{X} = \sum_{i=1}^{k} f_i x_i$$

Les $x_i$ désignent les valeurs prisent par la variable ou les centres des classes lorsque il s'agit d'une distribution groupée

## Indicateurs de position
*Exemple de calcul de la moyenne*

Tableau I

| $x_i$ | $n_i$ | $f_i$ | $n_i*x_i$ | $f_i*x_i$ |
|---|---|---|---|---|
| 1 | 5 | 0,10 | | |
| 2 | 9 | 0,18 | | |
| 3 | 15 | 0,30 | | |
| 4 | 11 | 0,22 | | |
| 5 | 3 | 0,06 | | |
| 6 | 7 | 0,14 | | |
| **Total** | **50** | **1** | | |

Tableau II

| $[a_i ; a_{i-1}[$ | $x_i$ | $n_i$ | $f_i$ | $n_i*x_i$ | $f_i*x_i$ |
|---|---|---|---|---|---|
| [151,5 ; 155,5[ | 153,5 | 10 | 0,20 | | |
| [155,5 ; 159,5[ | 157,5 | 12 | 0,24 | | |
| [159,5 ; 163,5[ | 161,5 | 11 | 0,22 | | |
| [163,5 ; 167,5[ | 165,5 | 7 | 0,14 | | |
| [167,5 ; 171,5[ | 169,5 | 10 | 0,20 | | |
| **Total** | | **50** | **1** | | |

Colonnes ajoutées pour calculer la moyenne

54

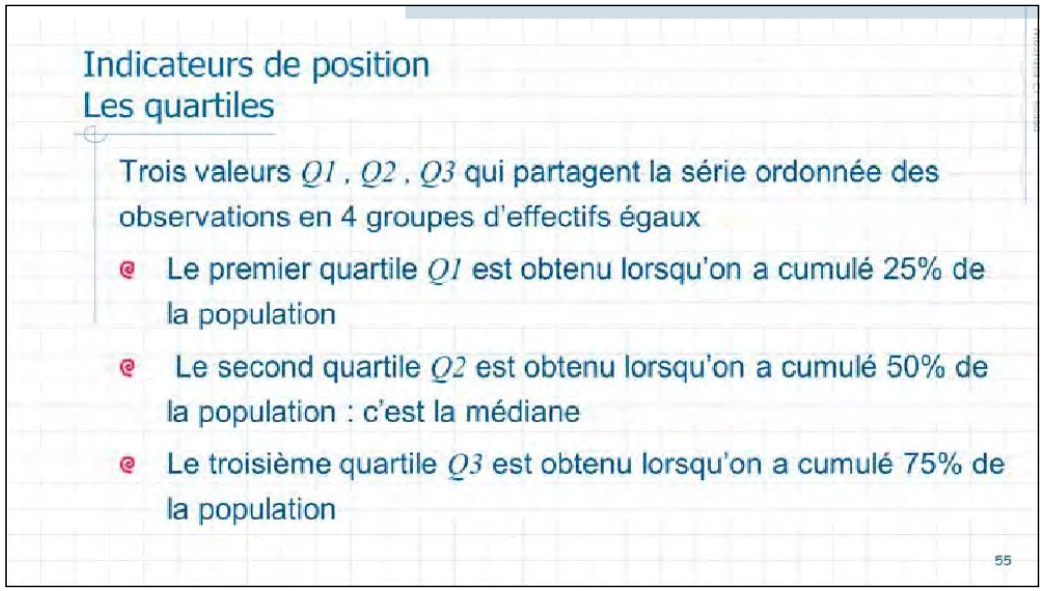
Calculons le moyenne de la série statistique du tableau *I* :

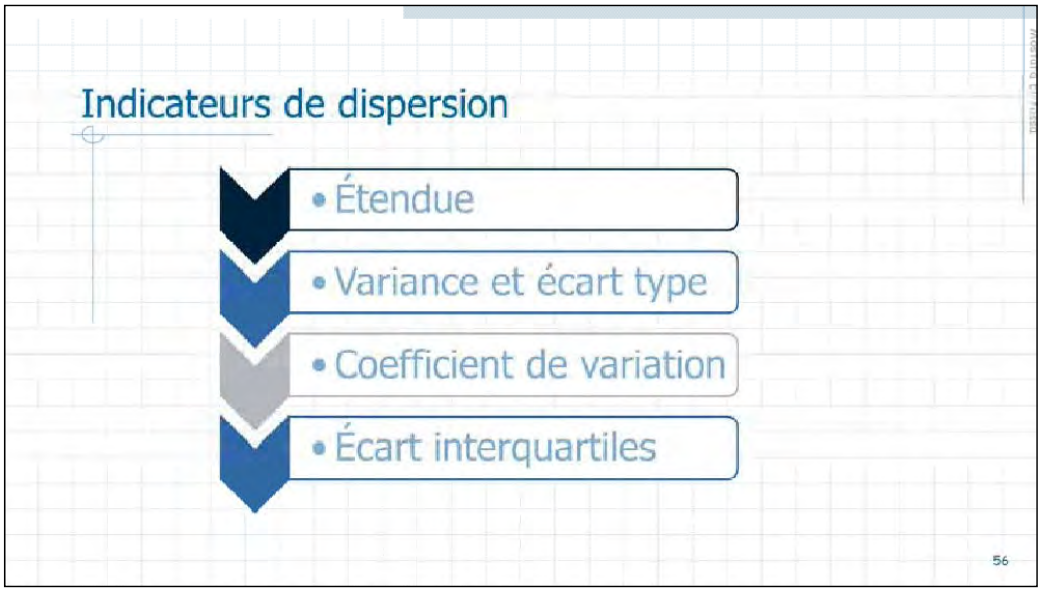
$$\bar{X} = \frac{1x5+2x9+3x15+4x11+5x3+6x7}{50} = 3{,}38$$

En général les calculs sont effectués dans le tableau statistique en ajoutant des colonnes :

| $x_i$ | $n_i$ | $f_i$ | $n_i*x_i$ | $f_i*x_i$ |
|---|---|---|---|---|
| 1 | 5 | 0,10 | **5** | **0.1** |
| 2 | 9 | 0,18 | **18** | **0.36** |
| 3 | 15 | 0,30 | **45** | **0.9** |
| 4 | 11 | 0,22 | **44** | **0.88** |
| 5 | 3 | 0,06 | **15** | **0.3** |
| 6 | 7 | 0,14 | **42** | **0.84** |
| **Total** | **50** | **1,00** | **169** | **3.38** |

La moyenne




Les quartiles permettent de séparer une série statistique en quatre groupes de même effectif (à une unité près).

Un quart des valeurs sont inférieures au premier quartile Q1.

Un quart des valeurs sont supérieures au troisième quartile Q3.

Comment déterminer les quartiles Q1 et Q3 d'une série de n valeurs ?

On calcule la quantité **q=n/4.**

Deux cas sont possibles: soit le résultat est entier (la division tombe juste), soit non.

**1ier cas :** q est un entier alors :

1. On classe les données par ordre croissant
2. Q1 est la valeur qui se trouve à la position q
3. Q3 est la valeur située à la position 3q

**2ème cas :** q n'est pas un entier alors :

1. On classe les données par ordre croissant
2. On arrondi q à l'entier supérieur, Q1 est la valeur qui se trouve à cette position
3. On arrondi 3q à l'entier supérieur pour trouver la position de la valeur de Q3.

**Exemple**

Prenons les valeurs rangées dans l'ordre croissant (n = 23 ) :

3-5-5-6-7-8-8-9-9-10-10-10-10-11-11-12-13-13-13-14-15-16-19

q=n:4 = 5,75 donc Q1 est la 6ème valeur de la série rangée dans l'ordre croissant donc Q1= 8,

3q = 17,25 donc Q3 est la 18ème valeur de la série rangée dans l'ordre croissant donc Q3= 13.

Pour une variable statistique quantitative réelle continue X, les quartiles Q1, Q2, Q3 sont les valeurs pour lesquels les fréquences cumulées de X sont respectivement 0,25, 0,50, 0,75. Ce sont les valeurs pour lesquelles l'ordonnée de la courbe cumulative des fréquences est respectivement égale à 0,25, 0,50, 0,75.

- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____

## Indicateurs de dispersion
### L'étendue

- L'étendue, notée $W$, est la différence entre la valeur maximale et la valeur minimale de la série statistique ;

  $W = x_{max} - x_{min}$
- L'étendue représente donc l'amplitude totale de la série.

57

## Indicateurs de dispersion
### *Ecart type*

- L'écart type σ mesure la dispersion autour de la moyenne.
- Pour calculer l'écart type, il faut calculer la variance :

$$\sigma^2 = var(X) = \frac{1}{n}\sum_i n_i(x_i - \bar{X})^2 = \sum_i f_i(x_i - \bar{X})^2$$

$$= \frac{1}{n}\sum n_i(x_i)^2 - (\bar{X})^2 = \sum f_i(x_i)^2 - (\bar{X})^2$$

58

- _____
- _____
- _____
- _____
- _____
- _____
- _____

L'écart-type (σ) s'exprime dans la même unité que les valeurs observées et mesure la dispersion autour de la moyenne.

Plus l'écart-type est grand, plus la dispersion de la distribution autour de la moyenne est importante. Plus l'écart-type est petit, plus la distribution est rassemblée autour de la moyenne. C'est une caractéristique de dispersion couramment utilisée car son traitement mathématique est facile.

En général les calculs sont effectués dans le tableau statistique en ajoutant des colonnes

| $x_i$ | $n_i$ | $f_i$ | $n_i*x_i$ | $f_i*x_i$ | $n_i*x_i*x_i$ | $f_i*x_i*x_i$ |
|---|---|---|---|---|---|---|
| 1 | 5 | 0,10 | **5** | **0,1** | **5** | **0,1** |
| 2 | 9 | 0,18 | **18** | **0,36** | **36** | **0,72** |
| 3 | 15 | 0,30 | **45** | **0,9** | **135** | **2,7** |
| 4 | 11 | 0,22 | **44** | **0,88** | **176** | **3,52** |
| 5 | 3 | 0,06 | **15** | **0,3** | **75** | **1,5** |
| 6 | 7 | 0,14 | **42** | **0,84** | **252** | **5,04** |
| **Total** | **50** | **1,00** | **169** | **3,38** | **679** | **13,58** |

$$var(X) = \frac{1}{n}\sum n_i\,(x_i)^2 - (\bar{X})^2 = \frac{1}{50}679 - 3{,}38^2 = 2{,}1556$$

$$\sigma = \sqrt{var(X)} = 1{,}468$$

## Indicateurs de dispersion
*Coefficient de variation*

- Le coefficient de variation (CV) est le rapport de l'écart-type à la moyenne;

$$CV = \frac{\sigma}{|\bar{X}|}$$

- Le coefficient de variation n'a pas d'unité
- Ce coefficient est souvent exprimé sous forme de pourcentage

59

- L'écart-type est un indicateur intéressant pour mesurer la dispersion d'une série . Mais il possède deux limites :
  1. il est exprimé dans l'unité de la variable dont il mesure la dispersion des valeurs. Ainsi, on ne peut pas comparer les dispersions de deux séries qui sont exprimées dans des unités différentes,.
  2. Il dépend de l'ordre de grandeur des valeurs mesurées (échelle de mesure)
- Le coefficient de variation n'a pas d'unité, il permet la comparaison de distributions de valeurs dont les échelles de mesure ne sont pas comparables.
- Il est généralement exprimé en pourcentage.
- Plus la valeur du coefficient de variation est élevée, plus la dispersion autour de la moyenne est grande.
- _____
- _____
- _____
- _____
- _____

## Indicateurs de dispersion
*Intervalle et écart interquartile*

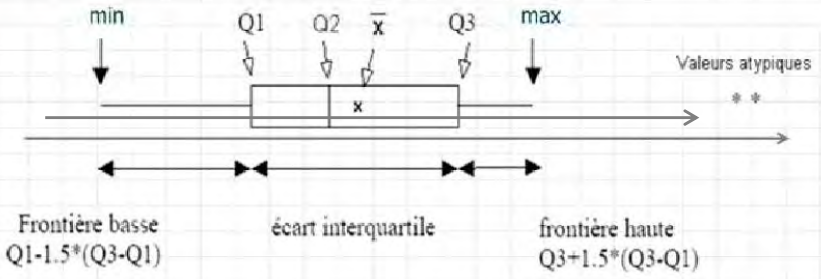
- On appelle intervalle interquartile l'intervalle *[Q1 ; Q3]*,
- L'écart interquartile est l'amplitude de l'intervalle *[Q1 ; Q3]*, c'est-à-dire le nombre *Q3-Q1*;
- L'écart interquartile est utilisé comme indicateur de dispersion. Il correspond à 50% des effectifs situés dans la partie centrale de la distribution

60

- L'écart interquartile correspond à l'étendue de la série statistique après élimination de 25% des valeurs les plus faibles et de 25% des valeurs les plus fortes. Cette mesure est plus robuste que l'étendue, qui est sensible aux valeurs extrêmes
- L'intervalle interquartile contient la moitié des valeurs centrales de la série statistique. Cet intervalle sert à mesurer l'étalement des données : il s'étend sur 50 % d'un ensemble de données et élimine l'influence des valeurs aberrantes (suppression des quarts supérieur et inférieur d'un ensemble de données)
- _____
- _____
- _____
- _____
- _____
- _____

## Représentation graphique : Boite à moustaches

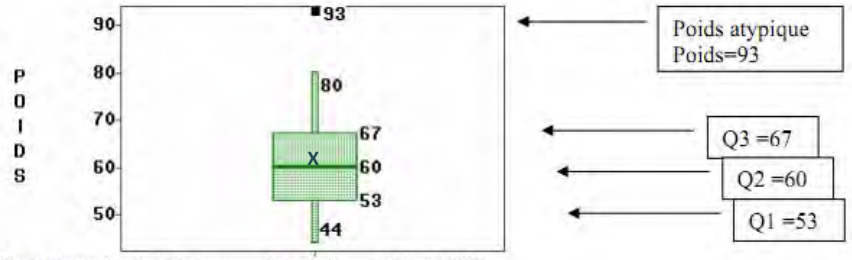
Graphique très pratique qui permet de résumer quelques caractéristiques




61

Ce diagramme est constitué de la façon suivante:

- On trace un axe des unités soit horizontalement ou verticalement
- On trace une "boîte" qui est un rectangle dont la longueur s'étend du premier quartile au troisième quartile, et qui est coupé par un trait vertical à hauteur de la médiane.
- On matérialise la position de la moyenne par une croix.
- De cette boîte partent deux traits dans le sens de l'axe : l'un va du premier quartile Q1 à la valeur minimale non aberrante de la série, l'autre du troisième quartile Q3 à la valeur maximale non atypique.
- Sur l'axe des moustaches, on représente également les valeurs aberrantes de la série. Toute valeur n'appartenant pas à l'intervalle [Q1-1,5(Q3-Q1) ; Q3+1,5(Q3-Q1)] est considérée comme aberrante




**Graphique 1** : *Boîte à moustaches de la variable POIDS*

## Paramètres de forme

*Coefficient d'asymétrie*

- Le coefficient d'asymétrie (Skewness) :

$$\gamma_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^3}{\sigma_X^{\,3}} = \frac{\sum_{i=1}^{n} f_i (X_i - \bar{X})^3}{\sigma_X^{\,3}}$$

- Il permet de déterminer la symétrie de la distribution des valeurs autour de la moyenne

62

- ✓ Si la répartition de l'échantillon ou de la distribution est symétrique autour de la moyenne, le coefficient d'asymétrie est nul.
- ✓ Dans le cas où il est positif, nous avons une asymétrie gauche, le graphique des fréquences est plus élevé à gauche ; l'asymétrie droite correspond au cas où il est négatif.

- _____
- _____
- _____
- _____
- _____
- _____

## Paramètres de forme

*Coefficient d'aplatissement*

- Le coefficient d'aplatissement (Kurtosis) est déterminé par la formule :

$$\gamma_2 = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^4}{\sigma_X^{\,4}} - 3 = \frac{\sum_{i=1}^{n} f_i (X_i - \bar{X})^4}{\sigma_X^{\,4}} - 3$$

63

## Les moments

- Les moments d'ordre $p$ centrés en zéro :

$$m'_p = \sum_{i=1}^{n} f_i X_i^p \; ; \; p = 1, \cdots$$

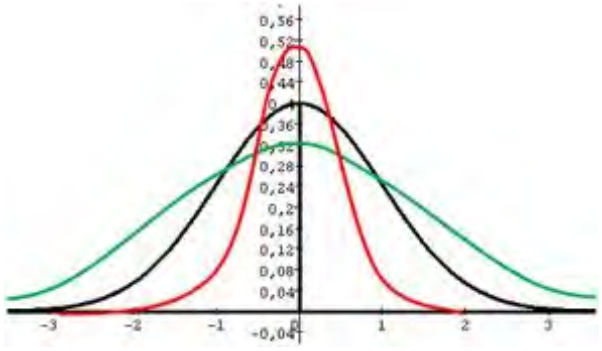
- Les moments, d'ordre $p$, centrés autour de la moyenne :

$$m_p = \sum_{i=1}^{n} f_i (X_i - m'_1)^p \; ; \; p = 1, \cdots$$

64

✓ Ce coefficient est très utile quand nous avons à faire à des échantillons symétriques. Dans ce cas :

- lorsque le coefficient d'aplatissement est nul nous disons que la répartition des observations est normale, c'est-à-dire que la courbe des fréquences a la forme d'une cloche comme la densité d'une loi Normale.
- Lorsqu'il est positif, nous avons une répartition sur-normale, c'est-à-dire moins aplatie qu'une densité normale.
- Lorsqu'il est négatif nous disons que la répartition est sous-normale, c'est-à-dire plus aplatie qu'une densité normale.




- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____