

## **Partie - II : Analyse des données**

### **Chapitre - 4 : Analyse préliminaire des données**

---

Prof. Abdelkamel ALJ

Updated: 2020/03/02

*FSJES - UMI*

Meknès

## 1. Introduction

## 1. Introduction

## 2. Paramètres et estimations

1. Introduction
2. Paramètres et estimations
3. Genres de données

1. Introduction
2. Paramètres et estimations
3. Genres de données
4. Mesures de récapitulation

## 1. Introduction

---

# 1. Introduction

L'analyse des données comprend le **résumé** des données et l'**interprétation** de leur **signification** pour donner des **réponses** claires aux **questions** qui ont **motivé** l'enquête.

## 2. Paramètres et estimations

---



## 2. Paramètres et estimations

### - Paramètres :

Un paramètre est une caractéristique de la population que l'utilisateur des données est intéressé à estimer, par exemple :

- ⊙ la moyenne de la population
- ⊙ l'écart-type de la population

## Estimations

Un estimateur est une formule de calcul de l'estimation du paramètre et l'estimation est la valeur de l'estimateur déterminée à l'aide des données de l'échantillon réalisé.

## Distribution d'échantillonnage

La distribution d'échantillonnage d'un estimateur est la distribution de toutes les valeurs différentes que l'estimateur peut avoir pour tous les échantillons possibles.

### 3. Genres de données

---

### 3. Genres de données

#### - Variables nominales :

Une variable nominale est une série de catégories qui sont simplement des étiquettes ou des noms sans lien mathématique entre eux.

## - Variables ordinales :

Une variable ordinale est une série de catégories ordonnées ou classées selon une échelle ou un continuum déterminé, et une catégorie en particulier peut précéder ou suivre une autre.

## - Variables discrètes :

Une variable discrète est une variable quantitative ayant des valeurs dénombrables.

## - Variables continues :

Une variable continue est une variable quantitative dont toute valeur dans une certaine étendue est possible.



## 4. Mesures de récapitulation

---

## 4. Mesures de récapitulation

Cette section présente les mesures récapitulatives suivantes :

- ⊙ distributions de fréquences (en tableau ou graphique)
- ⊙ mesures de tendance centrale (c.-à-d. moyenne, médiane ou mode)
- ⊙ mesures de l'étalement de la distribution (p. ex., variance, étendue)

## 4.1. Distributions de fréquences

- ⊙ La distribution de fréquences est la représentation la plus simple d'une variable elle donne la fréquence de chaque catégorie ou classe de valeurs.
- ⊙ Les distributions de fréquences peuvent aussi être représentées à l'aide de graphiques ou de diagrammes (une analyse visuelle des données).

Les graphiques et diagrammes suivants pourraient être ajoutés à un rapport sur les résultats de l'enquête :

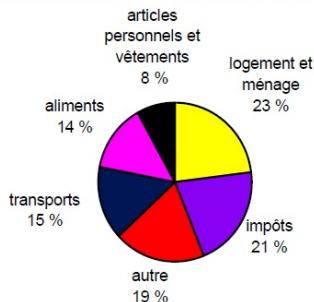
- ⊙ diagramme à secteurs
- ⊙ graphique à barres
- ⊙ diagramme en battons
- ⊙ histogramme et polygone des fréquences
- ⊙ diagramme à boîte et moustaches
- ⊙ graphique linéaire (nuage de points)

Le genre de diagramme à utiliser est déterminé par les données qu'il faut représenter et par le message qu'on veut souligner : ordre de grandeur, taille ou tendance.

## - Diagrammes à secteurs :

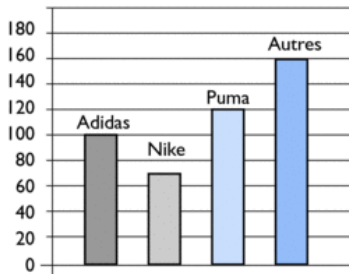
Un diagramme à secteurs est un cercle divisé en partitions comme une tarte pour afficher le pourcentage de la population dans différentes catégories d'une variable qualitative.

**Répartition des dépenses des ménages**



## - Diagramme en barre :

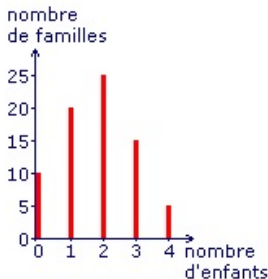
On appelle diagramme en barres un graphique qui, à chaque modalité de la variable qualitative associe un rectangle de base constante et dont la hauteur varie selon l'effectif.



La répartition de 450 collégiens selon la marque de leurs baskets

## - Diagramme en bâtons :

On appelle diagramme à barres un graphique qui, à chaque modalité de la variable quantitative discrète, associe un segment (bâton) dont la hauteur proportionnelle à l'effectif (ou à la fréquence) correspondant.

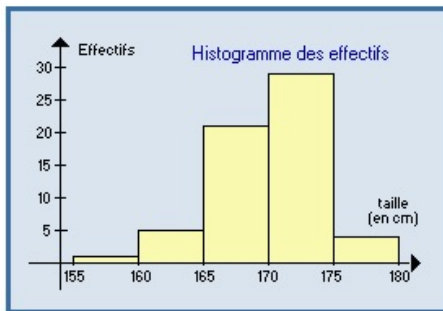


Le nombre d'enfants par famille dans une région maghrébine



## - Histogramme :

Graphique permettant de représenter une distribution continue regroupée en classes : rectangles juxtaposés dont les bases sont les classes, et les surfaces sont proportionnelles aux effectifs (ou fréquences) associés.



## 4.3. Paramètres de position

### - Le mode

Le mode est la valeur des données la plus fréquente. Il peut être appliqué à tous les genres de données.

## - La moyenne

La moyenne de la population pour un recensement est simplement la moyenne arithmétique pour les données quantitatives : la somme de toutes les valeurs d'une variable divisée par le nombre de valeurs.

## - La médiane

La médiane est la valeur du milieu d'une série de données disposées en ordre numérique (à partir de la plus petite jusqu'à la plus grande ou de la plus grande jusqu'à la plus petite).

## 4.4. Paramètres de dispersion

L'objectif d'une mesure de dispersion est de quantifier le taux de variabilité des données autour d'une valeur centrale.

### - Étendue

L'étendue d'une distribution est égale à la différence entre la plus grande et la plus petite valeur de cette distribution.

## - La variance

La variance, notée  $\sigma_x^2$  est la moyenne du carré des écarts à la moyenne.

## - L'écart type

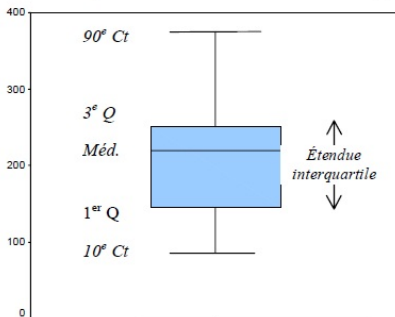
L'écart type, noté  $\sigma_x$  est la racine carré de la moyenne du carré des écarts à la moyenne, c'est à dire la racine carrée de la variance.

## - Diagrammes à boîte et moustaches :


Les statistiques sommaires peuvent aussi être présentées en un seul graphique récapitulatif : le diagramme à boîte et moustaches. Celui-ci est utilisé pour étudier la distribution et l'étalement des données.

## - Exemple :

**Distribution des prix de vente des maisons, juin 2002**  
**(Prix en milliers de dollars)**





 **"Méthodes et pratiques d'enquête".** Bibliothèque nationale du Canada (2010), No 12-587-X au catalogue. ISBN 978-1-100-95206-2.