



**Ecole Supérieure en Sciences Appliquées de Tlemcen**

# **Analyse de données**

## ***Chapitre 2: Statistique Descriptive***

**Présentée par:**

**Dr Imane NEDJAR**

L'objectif de la Statistique Descriptive est de décrire les données observées pour mieux les analyser

- **Description de données**
- **Valeurs centrales**
- **Indicateurs de dispersion**

# Description de données

The diagram illustrates the components of a data table. A blue arrow labeled 'Individu' points to the first column header 'Nom des Fournisseurs'. Another blue arrow labeled 'Variable' points to the 'Age' column header. A third blue arrow labeled 'Modalité' points to the 'Chiffre d'affaire' column header. The table contains 8 rows of data.

| Nom des Fournisseurs | Sexe | Age | Chiffre d'affaire |
|----------------------|------|-----|-------------------|
| Mohamed              | H    | 40  | Modéré            |
| Sarah                | F    | 50  | Important         |
| Ismail               | H    | 44  | Moyen             |
| Ilyes                | H    | 50  | Modéré            |
| Hanane               | F    | 35  | Important         |
| Ghouti               | H    | 60  | Moyen             |
| Yasmina              | F    | 55  | Modéré            |
| Fatima               | F    | 35  | Moyen             |

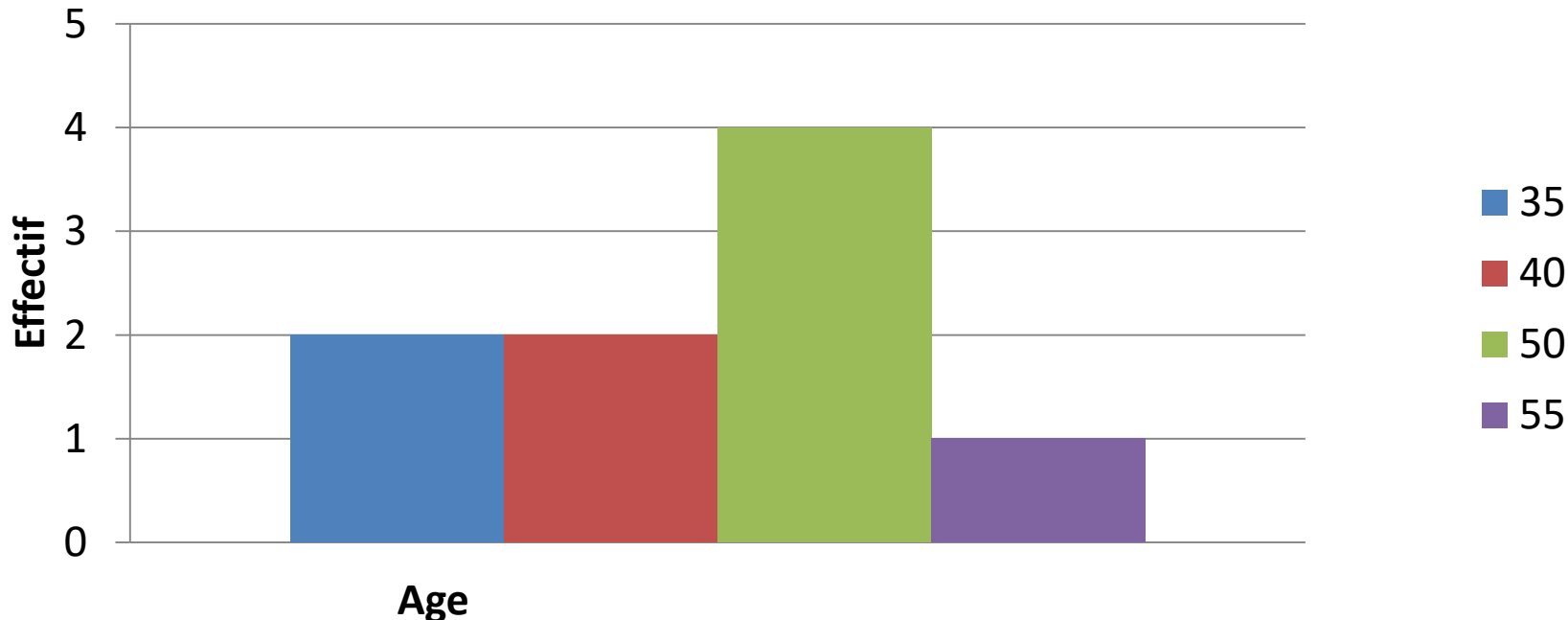
**Population** : Fournisseurs

# Description de données

## Effectifs

### La variable Age

| Age       | 35 | 40 | 50 | 55 |
|-----------|----|----|----|----|
| Effectifs | 2  | 2  | 4  | 1  |



| Nom des Fournisseurs | Sexe | Age | Chiffre d'affaire |
|----------------------|------|-----|-------------------|
| Mohamed              | H    | 40  | Modéré            |
| Sarah                | F    | 50  | Important         |
| Ismail               | H    | 40  | Moyen             |
| Ilyes                | H    | 50  | Modéré            |
| Hanane               | F    | 35  | Important         |
| Ghouti               | H    | 50  | Moyen             |
| Yasmina              | F    | 55  | Modéré            |
| Fatima               | F    | 35  | Moyen             |
| Karima               | F    | 50  | Important         |

# Description de données

## Fréquence

Fréquence de la modalité « M » d'une variable qualitative (FM)

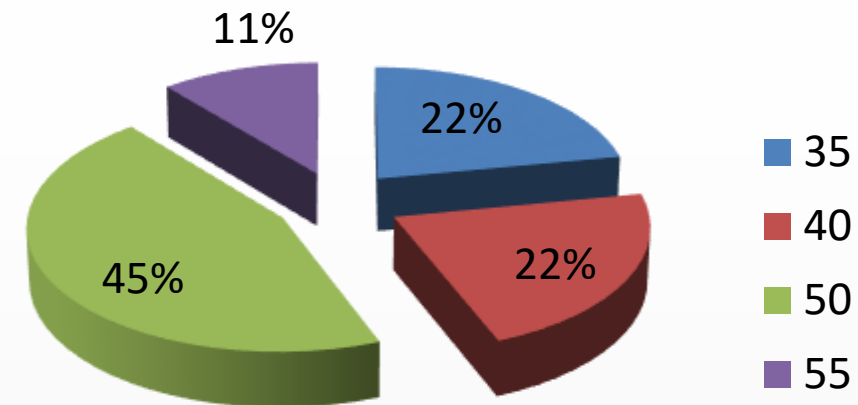
$$F_M = \frac{\text{Effe\_corresp\_M}}{\text{Effe\_Total}}$$

## Pourcentage

Pourcentage des individus correspondant à la modalité « M »

$$P_M = F_M \times 100$$

Diagramme en secteurs de l'Age



| Age | Effectifs                    | Fréquences                    | Pourcentage                  |
|-----|------------------------------|-------------------------------|------------------------------|
| 35  | 2                            | 2/9=0.22                      | 22%                          |
| 40  | 2                            | 2/9=0.22                      | 22%                          |
| 50  | 4                            | 4/9=0.45                      | 45%                          |
| 55  | 1                            | 1/9=0.11                      | 11%                          |
|     | <b>Total Effectifs « 9 »</b> | <b>Total Fréquences « 1 »</b> | <b>Total Pourcentage 100</b> |

# Valeurs centrales

## Le mode

Correspond à la valeur de la variable pour laquelle l'effectif (ou la fréquence) est le plus grand.

|           |    |    |    |    |
|-----------|----|----|----|----|
| Age       | 35 | 40 | 50 | 55 |
| Effectifs | 2  | 2  | 4  | 1  |

## La médiane

C'est le nombre qui sépare la série ordonnée en deux groupes de même effectif.

On écrit la liste de toutes les valeurs de la série par ordre croissant, chacune d'elle répétée autant de fois que son effectif

- Si le nombre de valeurs est impair, la médiane est la valeur du milieu
- S'il est pair, la médiane est la demi-somme des deux valeurs du milieu

Exemple 1 : 35 35 40 40 50 **50** 50 50 55 ->**50**

Exemple 2 : 10 20 40 50 ->**30**

# Valeurs centrales

## La moyenne arithmétique

On dispose d'une population de  $N$  individus et on observe  $X_1, X_2, \dots, X_n$  les valeurs d'une variable quantitative discrète  $X$  pour ces individus.

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

|           |    |    |    |    |
|-----------|----|----|----|----|
| Age       | 35 | 40 | 50 | 55 |
| Effectifs | 2  | 2  | 4  | 1  |

$$\bar{x} = 45$$

# Indicateurs de dispersion

---

## L'étendue

L'étendue  $E_x$  de la variable quantitative discrète  $X$  est la différence entre la plus grande et la plus petite des valeurs observées

$$E_x = \max(x_i) - \min(x_i)$$

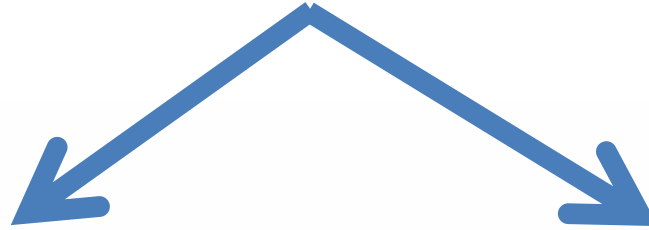
## La variance

- La variance est une mesure de la dispersion d'une série de données.
- Une variance faible indique que les nombres de la série de données sont proches l'un de l'autre.
- Une variance élevée indique que les nombres sont très distants.



# Indicateurs de dispersion

## La variance



La variance d'une population

La variance d'un échantillon

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

$$s^2(x) = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

→ On divise par  $n-1$  pour que l'écart-type de l'échantillon soit un bon estimateur de l'écart-type de la population.  
pour plus de détails voir → Loi de Bessel

# Indicateurs de dispersion

## La variance

|           |    |    |    |    |
|-----------|----|----|----|----|
| Age       | 35 | 40 | 50 | 55 |
| Effectifs | 2  | 2  | 4  | 1  |

$$\text{Variance} = \text{Moyenne}(n_i x_i^2) - \text{Moy}^2$$

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$



$$\sigma^2 = \frac{1}{N} \left( \sum n_i x_i^2 \right) - \mu^2$$

## L'écart-type

- L'écart-type est une mesure de la dispersion d'une série statistique autour de sa moyenne.
- Plus la distribution est dispersée c'est-à-dire les valeurs ne sont pas concentrées autour de la moyenne, plus l'écart-type sera élevé.

$$\textit{Ecart - type}(x) = \sqrt{\text{var}(x)}$$

# Indicateurs de dispersion

## L'écart-type

L'écart-type d'une population

L'écart-type d'un échantillon

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

$$s(x) = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$$

→ On divise par  $n-1$  pour que l'écart-type de l'échantillon soit un bon estimateur de l'écart-type de la population.  
pour plus de détails voir → Loi de Bessel

# Indicateurs de dispersion

