Faculté des sciences économique, commerciales, et sciences de gestion

Master 2 économie et gestion d'entreprise

# Cours 02: Analyse univarié

**DR BOUYACOUB BRAHIM**

**MAITRE DE CONFÉRENCES / DOCTEUR EN ÉCONOMIE MONÉTAIRE ET FINANCIÈRE**

## Les tests hypothèses statistiques (tests d'inférence)

Les tests d'hypothèses, ou tests d'inférence, ont pour objectif de mesurer l'effet d'une variable indépendante sur une variable dépendante, en fonction du nombre d'échantillons et en fonction de la nature des variables étudiées.

## Tests paramétriques

C’est l’ensemble des tests qui repose sur des familles de lois de probabilités pour lequel on fait deux hypothèses paramétrique H0 et H1.

- T test apparié
- T test à échantillons indépendants
- Corrélation/ analyse de régression
- Khi 2
- ANOVA à 1 facteur
- Analyse multivariée

## Tests non- paramétriques

Un test statistique utilisé dans le cas de variables indépendantes non métriques.

- TEST WILCOXON
- TEST MANN-WHITNEY
- LE TEST DE MCNEMAR
- LE KRUSKAL-WALLIS

## Description de la variable qualitative

La description d'une variable qualitative consiste à présenter les effectifs, c'est-à dire le nombre d'individus de l'échantillon pour chaque modalité de la variable, et les fréquences, c'est-à-dire le nombre de réponses associées aux modalités de la variable étudiée.

**Description de la variable quantitative**

Plusieurs critères permettent de décrire une variable quantitative :

- les mesures de la tendance centrale : moyenne, médiane, mode.
- les mesures de la dispersion : étendue, variance, écart type, coefficient de variation.
- les représentations graphiques : histogrammes, diagramme en bâton.

**Tableau statistique**

Pour crée un tableau statistique il faut déterminer :

1) Effectif totale n : le nombre de toutes les valeurs prises par la variable.
2) Effectif ni : nombre d'apparitions de la valeur xi dans la population ou dans l'échantillon.

$$\sum_{i=1}^{J} n_i = n_1 + n_2 + \ldots + n_J = n.$$

Fréquence fi associée à la valeur xi

$$\begin{cases} f_i = \frac{n_i}{n}, \\ \sum_{i=1}^{J} f_i = f_1 + f_2 + \ldots + f_J = 1. \end{cases}$$

Pourcentage pi associe à la valeur xi

$$\begin{cases} p_i = 100 \times f_i \ \%, \\ \sum_{i=1}^{J} p_i = p_1 + p_2 + \ldots + p_J = 100 \ \%. \end{cases}$$

Effectif cumulé Ni

$$\begin{cases} N_1 = n_1, \\ N_2 = n_1 + n_2, \\ N_3 = n_1 + n_2 + n_3, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ N_J = n_1 + n_2 + \dots + n_J = n. \end{cases}$$

Fréquence cumulé Fi

$$\begin{cases} F_1 = f_1, \\ F_2 = f_1 + f_2, \\ F_3 = f_1 + f_2 + f_3, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ F_J = f_1 + f_2 + \dots + f_J = 1. \end{cases}$$

**Exemple 1 : Variable qualitative nominale**

On note C : célibataire, M : marié, V : veuf, D : divorcé. On s'intéresse à la variable X= (état-civil) sur une population de n = 20 personnes.

Considérons la série statistique Suivante :

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| M | D | M | C | C | M | C | C | C | M | C | M | V | M | V | D | C | C | M | C |

## Tableau statistique :

| $x_i$ | $n_i$ | $f_i$ | $p_i\%$ | $N_i$ | $F_i$ |
|---|---|---|---|---|---|
| C | 9 | 0.45 | 45 | 9 | 0.45 |
| M | 7 | 0.35 | 35 | 16 | 0.75 |
| V | 2 | 0.10 | 10 | 18 | 0.85 |
| D | 2 | 0.10 | 10 | 20 | 1 |

Par exemple : le nombre d'apparition de la valeur x2 = M dans la série statistique est n2 = 7, sa fréquence est f2 = n2/n = 7/20 = 0,35, son pourcentage p2 = 100 * f2 =100 * 0,35 = 35 %, l'effectif cumulé N2 = n1 + n2 = 9 + 7 = 16 et la fréquence cumulée F2 = f1 + f2 = 0,45 + 0,35 = 0,75.

**Exemple 2 : Variable qualitative ordinale**

On interroge une population de n = 50 personnes sur leur dernier diplôme obtenu. On note : Sd : Sans diplôme, P : Primaire, Se : Secondaire, Su : Supérieur non-universitaire et U Universitaire.

Sd Sd Sd Sd P P P P P P P P P P P Se Se Su
Se Se Se Se Se Se Se Se Se Se Se Se Su Su Su
Su Su Su Su U U U U U U U U U U U U Su

## Tableau statistique

| $x_i$ | $n_i$ | $N_i$ | $f_i$ | $p_i$ | $F_i$ |
|---|---|---|---|---|---|
| Sd | 4 | 4 | 0.08 | 8 | 0.08 |
| P | 11 | 15 | 0.22 | 22 | 0.30 |
| Se | 14 | 29 | 0.28 | 28 | 0.58 |
| Su | 9 | 38 | 0.18 | 18 | 0.76 |
| U | 12 | 50 | 0.24 | 24 | 1 |

**Exemple 3 : Variable quantitative discrète**

Un quartier est composé d'une population de 50 ménages, et la variable X représente le nombre de personnes par ménage. Les valeurs de la variable sont :

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 |
| 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 3 |
| 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 4 |
| 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 5 |
| 5 | 5 | 5 | 5 | 5 | 6 | 6 | 6 | 8 | 8 |

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 | 10 | 39 | 0.20 | 0.78 |
| 5 | 6 | 45 | 0.12 | 0.90 |
| 6 | 3 | 48 | 0.06 | 0.96 |
| 8 | 2 | 50 | 0.04 | 1 |

**Les mesures de tendance :**

Les mesures de tendance centrale permettent de résumer un ensemble de données relatives à une variable quantitative. Elles permettent de déterminer une valeur «typique» ou centrale autour de laquelle des données ont tendance à se rassembler.

**Les indicateurs de tendance centrale sont :**

La moyenne : Somme des valeurs de toutes les observations divisée par l'effectif.

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_N}{N} = \frac{\sum_{n=1}^{N} X_n}{N}$$

La médiane : la valeur centrale d'une série statistique dont les valeurs observées ont été rangées dans l'ordre croissant, est la valeur qui partage la population étudiée en deux sous-ensembles de même effectif.

Pour calculer la médiane :

- On classe les valeurs de la série statistique dans l’ordre croissant :
- 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**

**Déterminer la médiane de la série :**

1, 4, 2, 5, 0

Premièrement, on classe les valeurs de la série statistique dans l'ordre croissant :

0, 1, 2, 4, 5

Il y a un nombre impair de valeurs, la médiane est donc la valeur du milieu.

0, 1, 2, 4, 5

**La médiane est 2.**

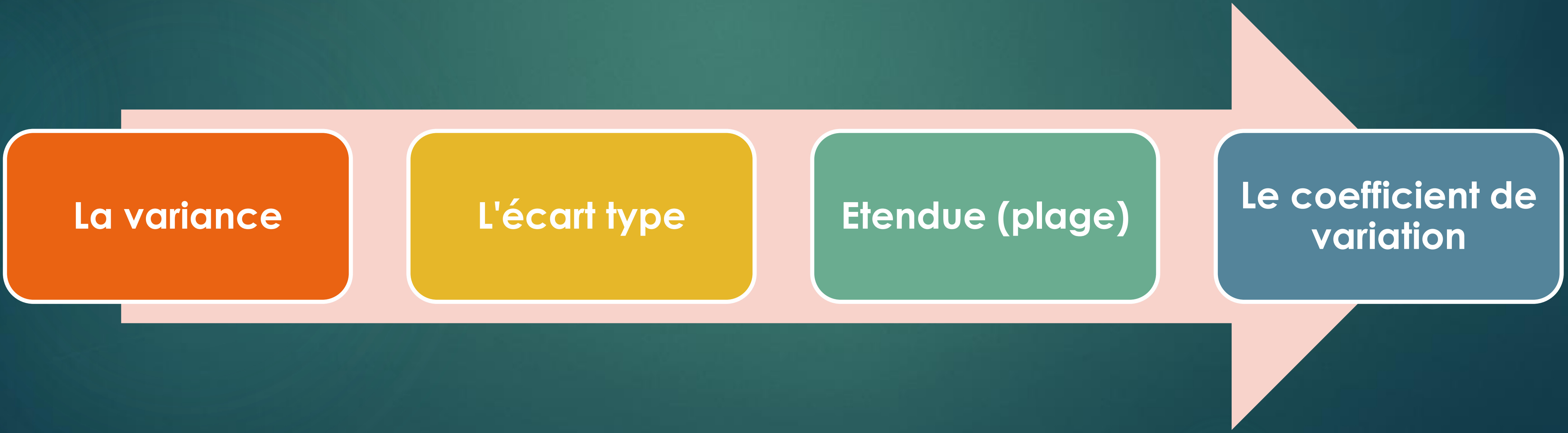
**Exemple 2**
**Déterminer la médiane de la série :**
10, 40, 20, 50
Premièrement, on classe les valeurs de la série statistique dans l'ordre croissant :
10, 20, 40, 50
Il y a un nombre pair de valeurs, on a donc 2 valeurs centrales. La médiane est alors la moyenne de ces deux valeurs.
10, 20, 40, 50
médiane= 20+40/2=60/2=30
Médiane=30
**La médiane est 30.**

Le mode : Le mode représente la valeur présentant la plus grande fréquence d’occurrence. Il est défini comme la valeur la plus fréquente dans la série d’observation. (Cette valeur n’est pas nécessairement unique).

**Les mesures de dispersion :**

Comme le nom l’indique, les indicateurs de dispersions permettent de mesurer comment les données se «répartissent».

Les indicateurs de dispersions sont :



**La variance :**

La variance est la mesure de la dispersion autour de la moyenne, égale à la somme des carrés des écarts par rapport à la moyenne, divisée par le nombre d’observations moins un.

$$\sigma^2 = \frac{1}{N}\sum_{n=1}^{N}(X_n - \bar{X})^2$$

**L'écart type :**

L'écart type est la mesure de la dispersion autour de la moyenne, exprimée dans la même unité que la variable. Il est défini comme la racine carrée de la variance.

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(X_n - \bar{X})^2}$$

**L’étendue :**

L’étendu d’une série statistique est l’écart entre sa plus grande valeur et sa plus petite. En d’autre terme, l’étendu c’est la différence entre la valeur la plus élevé et la valeur la plus bas.

$$e = \max X - \min X$$

**Le coefficient de variation :**

Le coefficient de variation est le rapport de l'écart type à la moyenne, exprimé en pourcentage. Son objet est de mesurer le degré de variation de la moyenne d'un échantillon à l'autre, lorsque ceux-ci sont issus de la même distribution.

CV=l'écart type/ la moyenne *100