



**RÉSUMÉ THÉORIQUE – FILIÈRE INTELLIGENCE ARTIFICIELLE**  
**OPTION ASSISTANT DATA ANALYST**  
**M201 - Maîtriser les Systèmes Décisionnels**



**105 heures**



# SOMMAIRE

## 01 – Approfondir les statistiques

Rappeler les notions essentielles  
Maîtriser les mesures de dispersion et de position  
Assimiler les probabilités et distributions avancées  
Vulgariser le langage R

## 02 – Visualiser les données décisionnelles

Créer divers types de graphiques  
Configurer les graphiques pour une analyse profonde

## 03 – Maîtriser les bases de données décisionnelles

Introduire les bases de données décisionnelles  
Explorer la modélisation dimensionnelle avancée  
Appliquer le processus ETL dans le contexte décisionnel

# MODALITÉS PÉDAGOGIQUES



1

## LE GUIDE DE SOUTIEN

Il contient le résumé théorique et le manuel des travaux pratiques



2

## LA VERSION PDF

Une version PDF est mise en ligne sur l'espace apprenant et formateur de la plateforme WebForce Life



3

## DES CONTENUS TÉLÉCHARGEABLES

Les fiches de résumés ou des exercices sont téléchargeables sur WebForce Life



4

## DU CONTENU INTERACTIF

Vous disposez de contenus interactifs sous forme d'exercices et de cours à utiliser sur WebForce Life



5

## DES RESSOURCES EN LIGNES

Les ressources sont consultables en synchrone et en asynchrone pour s'adapter au rythme de l'apprentissage

# PARTIE 1

## APPROFONDIR LES STATISTIQUES

**Dans ce module, vous allez :**

- Rappeler les notions essentielles
- Maîtriser les mesures de dispersion et de position
- Assimiler les probabilités et distributions avancées
- Vulgariser le langage R



**45 heures**

# CHAPITRE 1

## RAPPELER LES NOTIONS ESSENTIELLES

**Ce que vous allez apprendre dans ce chapitre :**

- Réviser rapidement les statistiques descriptives
- Identifier les types de variables statistiques
- Calculer les mesures centrales
- Appliquer pratiquement avec Python/Excel



**45 heures**



# CHAPITRE 1

## RAPPELER LES NOTIONS ESSENTIELLES

1. **Révision rapide des statistiques descriptives**
2. Types de variables statistiques
3. Mesures centrales
4. Application pratique avec Python/Excel



# 01 – Rappel les notions essentielles

## Révision rapide des statistiques descriptives



### Introduction aux statistiques

Les statistiques constituent un domaine essentiel des mathématiques appliquées qui se concentre sur la **collecte**, l'**analyse**, l'**interprétation**, la **présentation** et l'**organisation** des données. C'est un outil puissant utilisé dans une variété de disciplines pour prendre des décisions éclairées et tirer des **conclusions** basées sur des **informations numériques**.



# 01 – Rappeler les notions essentielles

## Révision rapide des statistiques descriptives



### Principales branches des statistiques

- **Statistique Descriptive** : Déterminer les **caractéristiques** d'une population. Elle utilise des mesures telles que la moyenne, la médiane, le mode, et d'autres outils pour synthétiser et visualiser les données de manière compréhensible.
- **Statistique Inférentielle** : Extrapoler et projeter les résultats numériques obtenus sur un **échantillon** à la population. Utilisant des échantillons représentatifs. Elle repose sur des concepts de **probabilité** pour extrapoler les résultats d'un échantillon à une population entière.



Décrire



Généraliser



## 01 – Rappeler les notions essentielles

### Révision rapide des statistiques descriptives



#### Objectif de la statistique descriptive

L'objectif de la statistique descriptive consiste à **exposer et décrire** chaque Individu , c'est-à-dire à **résumer numériquement** et/ou à **représenter graphiquement**, les données disponibles lorsqu'elles sont nombreuses ou proviennent d'un recensement.

#### Définition de la statistique descriptive

Statistiques descriptives: méthode d'analyse de données visant à **résumer** et **décrire** des caractéristiques d'un ensemble de données.  
Source: Dictionnaire Larousse.

A descriptive statistic is a **summary statistic** that quantitatively **describes** or summarizes features from a collection of information.



## 01 – Rappeler les notions essentielles

### Révision rapide des statistiques descriptives



Remarques :



Pour prédire : Utiliser les statistiques prédictives.

Pour généraliser : Utiliser les statistiques inférentielles.

➔ Pour **résumer** : Utiliser l'analyse descriptive.



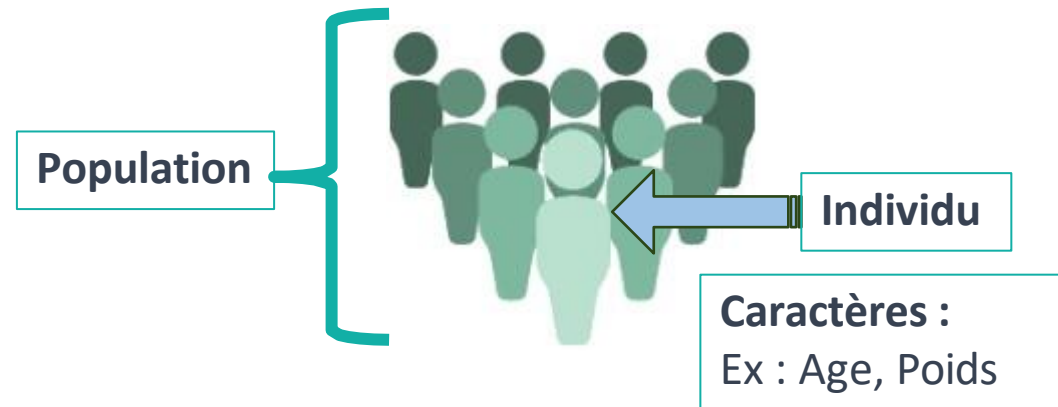
## 01 – Rappeler les notions essentielles

### Révision rapide des statistiques descriptives



#### Définitions Fondamentales

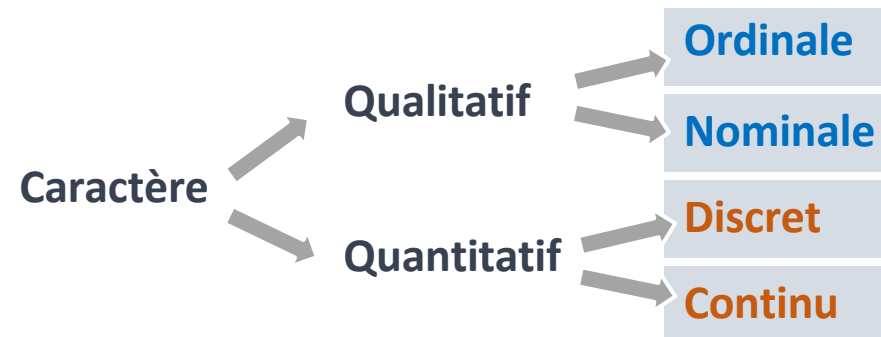
- L'ensemble étudié en statistique est appelé la **population**.
- Les éléments de cette population sont désignés comme **individus** ou **unités statistiques**.
- Les propriétés examinées sur les individus d'une population sont appelées les **caractères**.



### Types de variables statistiques

Dans le domaine des statistiques, les caractères étudiés peuvent être classés en deux catégories distinctes :

- **Caractères Qualitatifs** : Ces caractères englobent des modalités non mesurables numériquement, exprimées par des mots, ou des phrases. Ces caractères qualitatifs peuvent, à leur tour, être de nature **ordinaire** ou **nominale**
- **Caractères Quantitatifs** : Leur détermination conduit à un nombre ou une séquence de nombres. Ces caractères Quantitatifs peuvent, à leur tour, être de nature **discret** ou **continu**



### Types de variables statistiques

- **Caractères Qualitatifs Ordinaux** : Les modalités présentent un ordre spécifique. Un exemple concret serait le niveau de satisfaction des clients dans un sondage, où les réponses sont "Insatisfait", "Neutre", "Satisfait", "Très satisfait".
- **Caractères Qualitatifs Nominaux** : Les modalités n'ont pas d'ordre intrinsèque. Par exemple, cela inclut les couleurs
- **Caractères Quantitatifs Discrets** : Ils se limitent à prendre des valeurs spécifiques. Un exemple concret serait le nombre de personnes dans une famille. Vous ne pouvez pas avoir un nombre fractionnaire de personnes.
- **Caractères Quantitatifs Continus** : Ils peuvent prendre n'importe quelle valeur réelle. Un exemple concret serait le poids des personnes en kilogrammes. Le poids peut prendre n'importe quelle valeur réelle entre zéro et l'infini, par exemple, 60,54 kg, 72,30 kg, 85,01 kg, etc. Il existe une infinité de valeurs possibles dans cette plage.

## 01 – Rappel les notions essentielles

### Types de variables statistiques



### Types de variables statistiques



Parfois, pour classer les Caractères Quantitatifs Continus, nous pouvons créer des plages pour arrondir les données en Catégories Qualitatives Ordinales, par exemple, en regroupant les tailles des personnes en classes de "small", "medium", "large" et "Xlarge".



#### Introduction

Les mesures centrales sont utilisées pour déterminer **le point central** ou typique d'un ensemble de données. Trois mesures clés sont la **moyenne**, le **mode** et la **médiane**.



# 01 – Rappeler les notions essentielles

## Mesures Centrales



### Moyenne (mean / Average)

- **Définition :** La moyenne est la somme de toutes les valeurs, divisée par le nombre total de valeurs.
- **Formula :** 
$$\text{Moyenne} = \frac{\sum \text{Valeurs}}{\text{Nombre de Valeurs}}$$
- **Interprétation :** Représente la valeur moyenne de l'ensemble de données.
- **Exemple :**

Valeurs : 5,6,8,9,12,15

Moyenne = 5+6+8+9+12+15

$$\text{Moyenne} = \frac{55}{6}$$

Moyenne  $\approx$  9.16



#### Mode

- **Définition** : Le mode est la **valeur qui apparaît le plus fréquemment** dans un ensemble de données.
- **Relevance** : Indique la valeur la plus fréquente, **utile pour les données catégoriques**.
- **Exemple** : Dans une classe, les notes sont représentées par des mentions. La mention la plus fréquente est le mode.

Valeurs : Assez Bien, Assez Bien, Bien, Bien, Très Bien, Très Bien, Très Bien, Excellent.

Effectif : Assez Bien : 2, Bien : 2, Très Bien : 3, Excellent : 1.

Mode : Très Bien

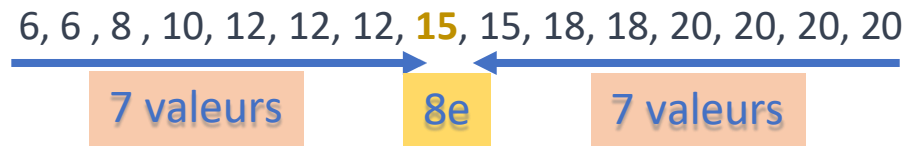
# 01 – Rappeler les notions essentielles

## Mesures Centrales



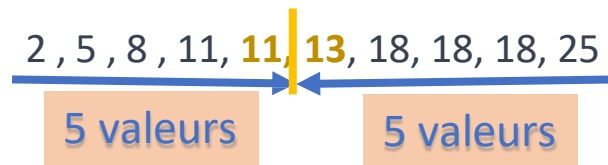
### Médiane

- **Définition** : La médiane représente la valeur qui partage l'effectif total en **deux effectifs égaux**, après les avoir classées par ordre croissant.
- **Exemple 1**: La longueur de la liste est impaire
- Imaginons un enseignant évaluant des devoirs surveillés avec les notes suivantes :



**Après le tri** et avec un effectif total de 15, la médiane correspond à la 8e valeur, soit 15.

- **Exemple 2**: La longueur de la liste est paire
- Un autre enseignant recueille la série de notes suivante :



**Après le tri** et avec un effectif total de 10, la médiane se situe entre la 5e et la 6e valeur, la médiane est la moyenne de ces deux valeurs, soit 12

## 01 – Rappeler les notions essentielles

### Application pratique avec Excel



### Exploration statistique des notes d'élèves en Excel : Mesures centrales

Cette application en Excel démontre l'utilisation des statistiques descriptives pour analyser les notes d'élèves.

	A	B
1	ID Stagiaire	Note
2	S1	10
3	S2	11
4	S3	12
5	S4	14
6	S5	15
7	S6	16
8	S7	17
9	S8	17
10	S9	18
11	S10	19
12	Moyenne	14,9
13	Mode	17
14	Mediane	15,5
15		

La moyenne, calculée à l'aide de la fonction `=MOYENNE(B2:B11)`, indique une répartition homogène des notes.

Utilisation de `=MODE.SIMPLE(B2:B11)` pour identifier le mode, mettant en lumière la note la plus fréquente.

La médiane, obtenue via `=MEDIANE(B2:B11)`, offre une vue centrale résiliente aux valeurs extrêmes.

## 01 – Rappeler les notions essentielles

### Application pratique avec Excel



#### Comparaison des Fonctions Excel : Français vs. Anglais

Les fonctions principales d'Excel restent cohérentes, mais les noms des formules peuvent varier entre les versions linguistiques.



Fonction	Français	Anglais
Moyenne	=MOYENNE()	=AVERAGE()
Mode	=MODE.SIMPLE()	=MODE.SNGL()
Médiane	=MEDIANE()	=MEDIAN()

## Exploration Statistique des Notes d'Élèves en Python

Cette application Python démontre l'utilisation des statistiques descriptives pour analyser les notes d'élèves.

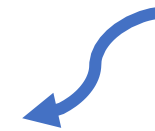
```
import statistics
# Données représentant les notes des élèves
notes = [17, 19, 15, 18, 14, 10, 12, 17, 11, 16]

# Calcul de la moyenne, du mode et de la médiane
moyenne = statistics.mean(notes)
mode_result = statistics.mode(notes)
median_result = statistics.median(notes)

# Affichage des résultats
print("Moyenne :", moyenne)
print("Mode :", mode_result)
print("Médiane :", median_result)

# Tri de la liste pour la comparaison
notes.sort()
print("Après le tri pour la comparaison :\n", notes)
```

Les méthodes incluent le calcul de la moyenne, du mode et de la médiane, ainsi que le tri de la liste pour une comparaison approfondie.



### Exploration Statistique des Notes d'Élèves en Python

#### Résultats et Interprétations :

La moyenne des notes s'établit à un niveau équilibré, indiquant une tendance générale au sein de la classe.

```
Moyenne : 14.9  
Mode : 17  
Médiane : 15.5  
Après le tri pour la comparaison :  
[10, 11, 12, 14, 15, 16, 17, 17, 18, 19]
```

Mode identifie la note la plus fréquente, soulignant une tendance dominante.

Médiane, résiliente aux extrêmes, offre une vue centrale de la distribution des notes.

5 valeurs

5 valeurs

Tri de la liste facilitant la comparaison visuelle des résultats.

### Consulter les Dernières Intégrations



Certaines fonctions de statistiques telles que la Moyenne, le Mode et la Médiane peuvent être utilisées à partir des packages scipy ou de Numpy, mais les packages Python sont soumis à des changements constants au fil du temps. Lorsque vous abordez ce cours, assurez-vous de vérifier les dernières intégrations disponibles.



## CHAPITRE 2

# MAITRISER LES MESURES DE DISPERSION ET DE POSITION

Ce que vous allez apprendre dans ce chapitre :

- Réviser les mesures de dispersion et de position
- Analyser les outliers
- Appliquer pratiquement avec Python/Excel



10 heures





# CHAPITRE 2

## MAITRISER LES MESURES DE DISPERSION ET DE POSITION

1. Rappeler les notions essentielles
- 2. Maitriser les mesures de dispersion et de position**
3. Assimiler les probabilités et distributions avancées
4. Vulgariser le langage R

