

4. Mesures de tendance centrale

Statistique Descriptive

Anicet Ebou

Institut National Polytechnique Félix Houphouët-Boigny

2026–2027

1. Pourquoi résumer par un centre ?
2. Moyenne arithmétique
3. Moyenne géométrique et harmonique
4. Médiane
5. Mode
6. Comparaison et choix de la mesure
7. Exercices

1. Pourquoi résumer par un centre ?
2. Moyenne arithmétique
3. Moyenne géométrique et harmonique
4. Médiane
5. Mode
6. Comparaison et choix de la mesure
7. Exercices

Un jeu de données brut est difficile à interpréter :

Rendements en t/ha (n = 20 parcelles)

3,2	4,1	2,8	5,0	3,7	4,4	3,1	4,9
3,5	2,6	4,2	3,8	5,1	3,3	4,6	3,0
4,0	3,6	2,9	4,3				

Une **mesure de tendance centrale** :

- ▶ résume la série en **une valeur typique**
- ▶ facilite la **comparaison** entre groupes
- ▶ constitue une **référence** pour détecter les valeurs atypiques

Trois familles de moyennes

Moyennes

Moyenne
arithmétique

Moyenne
géométrique

Moyenne
harmonique

Autres

Médiane

Mode

1. Pourquoi résumer par un centre ?
2. Moyenne arithmétique
3. Moyenne géométrique et harmonique
4. Médiane
5. Mode
6. Comparaison et choix de la mesure
7. Exercices

Moyenne arithmétique

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

Exemple : Rendements cacao (t/ha) sur 5 parcelles :

$$\{2,4 ; 3,1 ; 2,8 ; 3,5 ; 2,7\}$$

$$\bar{x} = \frac{2,4 + 3,1 + 2,8 + 3,5 + 2,7}{5} = \frac{14,5}{5} = 2,9 \text{ t/ha}$$

Quand les observations ont des **poids** différents w_i :

Moyenne pondérée

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

Rendement moyen d'une exploitation

3 variétés cultivées sur des surfaces inégales :

Variété	Surface (ha)	Rendement (t/ha)
Amelonado	12	1,8
Trinitario	8	2,4
Forastero	5	1,5

$$\bar{x}_w = \frac{12 \times 1,8 + 8 \times 2,4 + 5 \times 1,5}{25} = \frac{46,1}{25} = 1,84 \text{ t/ha}$$

Propriétés algébriques

- ▶ $\sum (x_i - \bar{x}) = 0$
(somme des écarts nulle)
- ▶ Si $y_i = a + b x_i$ alors $\bar{y} = a + b \bar{x}$
(invariance linéaire)
- ▶ Sensible aux **valeurs extrêmes**

Attention aux outliers

Rendements (t/ha) :

1,9 ; 2,1 ; 2,0 ; 2,3 ; **12,0**

$\bar{x} = 4,06$ t/ha, mais 4 parcelles sur 5 ont moins de 2,4 t/ha. La moyenne est ici **trompeuse**.

```
rendement ← c(2.4, 3.1, 2.8, 3.5, 2.7)

# Moyenne arithmetique
mean(rendement)

# Moyenne ponderee
surface ← c(12, 8, 5)
rdt_var ← c(1.8, 2.4, 1.5)
weighted.mean(rdt_var, w = surface)
```

1. Pourquoi résumer par un centre ?
2. Moyenne arithmétique
3. Moyenne géométrique et harmonique
4. Médiane
5. Mode
6. Comparaison et choix de la mesure
7. Exercices

Définition

$$G = \left(\prod_{i=1}^n x_i \right)^{1/n} = \exp \left(\frac{1}{n} \sum_{i=1}^n \ln x_i \right)$$

Quand l'utiliser ?

- ▶ Données issues d'un processus **multiplicatif** (taux de croissance, indices)
- ▶ Variables **log-normales** : concentrations de pesticides, tailles de colonies microbiennes

Taux de croissance d'une culture

Taux annuels : +20%, +5%, -10% soit facteurs 1,20 ; 1,05 ; 0,90

$$G = (1,20 \times 1,05 \times 0,90)^{1/3} = (1,134)^{1/3} \approx 1,043$$

Taux moyen : **+4,3% par an**

Définition

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

Quand l'utiliser ?

- ▶ Lorsque la variable est un **ratio** (vitesse, débit, rendement par unité de temps)
- ▶ Taux de travail, densités

Vitesse de traitement

Un pulvérisateur traite à 3 km/h l'aller et 5 km/h au retour.

$$H = \frac{2}{\frac{1}{3} + \frac{1}{5}} = \frac{2}{\frac{8}{15}} = \frac{30}{8} = 3,75 \text{ km/h} \quad , \text{ (et non } \frac{3+5}{2} = 4 \text{ km/h)}$$

Moyennes géométrique et harmonique en R

```
# Moyenne geometrique
facteurs ← c(1.20, 1.05, 0.90)
moy_geo ← exp(mean(log(facteurs)))
moy_geo

# Avec le package DescTools
# library(DescTools)
# Gmean(facteurs)

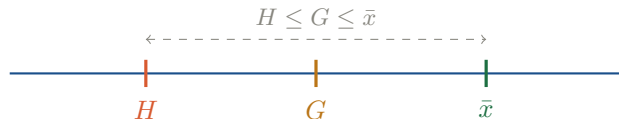
# Moyenne harmonique
vitesses ← c(3, 5)
moy_harm ← length(vitesses) / sum(1 / vitesses)
moy_harm

# Avec DescTools : Hmean(vitesses)
```

Relation entre les trois moyennes

Pour des valeurs positives non toutes égales :

$$H \leq G \leq \bar{x}$$



- ▶ Égalité si et seulement si **toutes les valeurs sont identiques**
- ▶ La moyenne arithmétique **surestime** pour les données multiplicatives

1. Pourquoi résumer par un centre ?
2. Moyenne arithmétique
3. Moyenne géométrique et harmonique
- 4. Médiane**
5. Mode
6. Comparaison et choix de la mesure
7. Exercices

Médiane

La médiane Me est la valeur qui **partage la série ordonnée en deux moitiés égales** : 50% des valeurs sont inférieures ou égales à Me , 50% sont supérieures ou égales.

Calcul pratique sur série triée :

- ▶ n impair : $Me = x_{(n+1)/2}$
- ▶ n pair : $Me = \frac{x_{n/2} + x_{n/2+1}}{2}$

Rendements arachide (t/ha) triés

$n = 7$: 1,2 ; 1,5 ; 1,8 ; **2,1** ; 2,3 ; 2,7 ; 3,0

$Me = x_{(7+1)/2} = x_4 = 2,1 \text{ t/ha}$

Quand les données sont présentées sous forme de **classes de fréquences** :

$$Me = L_{me} + \frac{\frac{n}{2} - F_{me-1}}{f_{me}} \times h_{me}$$

où :

- ▶ L_{me} : borne inférieure de la classe médiane
- ▶ F_{me-1} : effectif cumulé avant la classe médiane
- ▶ f_{me} : effectif de la classe médiane
- ▶ h_{me} : amplitude de la classe médiane

Avantages

- ▶ **Robuste** aux valeurs extrêmes
- ▶ Interprétation intuitive (50^e percentile)
- ▶ Valide pour variables **ordinales**

Revisiter l'exemple outlier

1,9 ; 2,0 ; 2,1 ; 2,3 ; 12,0

$\bar{x} = 4,06$ t/ha (biaisé)

$Me = 2,1$ t/ha (representatif)

Quand utiliser la médiane ?

Préférer la médiane lorsque la distribution est **asymétrique** ou contient des **valeurs aberrantes** : revenus, surfaces foncières, concentrations en polluants.

```
rendements ← c(1.9, 2.0, 2.1, 2.3, 12.0)

median(rendements) # 2.1
mean(rendements)   # 4.06

# Comparaison sur serie avec outlier
cat("Moyenne :", mean(rendements), "\n")
cat("Mediane :", median(rendements), "\n")
```

1. Pourquoi résumer par un centre ?
2. Moyenne arithmétique
3. Moyenne géométrique et harmonique
4. Médiane
5. Mode
6. Comparaison et choix de la mesure
7. Exercices

Mode

Le **mode** (ou valeur modale) est la valeur la plus fréquente de la série.

Types de distributions

- ▶ **Unimodale** : un seul mode
- ▶ **Bimodale** : deux modes
(deux groupes distincts)
- ▶ **Amodale** : toutes les valeurs sont
équifréquentes

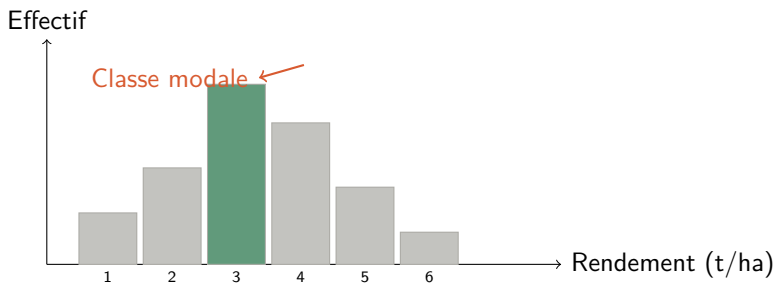
Type de sol observé

Type	Effectif
Ferralsol	3
Nitisol	8
Cambisol	5
Luvisol	2

Mode = **Nitisol**

Mode pour variables continues

Pour une variable continue, on parle de **classe modale** : la classe de plus grande fréquence.



R n'a pas de fonction `mode()` statistique native (c'est un mot réservé). Deux approches :

```
# Approche 1 : table de frequences
sol ← c("Ferralsol","Nitisol","Nitisol","Cambisol",
        "Nitisol","Luvisol","Cambisol","Nitisol")
tab ← table(sol)
names(which.max(tab)) # "Nitisol"

# Approche 2 : fonction personnalisée
mode_stat ← function(x) {
  tab ← table(x)
  names(tab)[which.max(tab)]
}
mode_stat(sol)
```

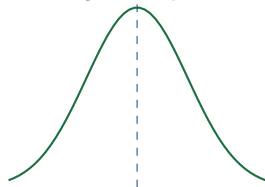
1. Pourquoi résumer par un centre ?
2. Moyenne arithmétique
3. Moyenne géométrique et harmonique
4. Médiane
5. Mode
6. Comparaison et choix de la mesure
7. Exercices

Moyenne, médiane, mode : lequel choisir ?

Critère	Moyenne	Médiane	Mode
Type de variable	Quantitative	Quant. ou Ord.	Tous types
Sensible aux outliers	Oui	Non	Non
Distribution symétrique	Idéale	Bonne	Variable
Distribution asymétrique	Biaisée	Recommandée	Indicative
Valeur unique	Toujours	Toujours	Pas toujours
Utilisation algébrique	Facile	Limitée	Difficile

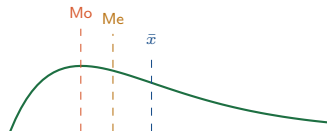
Position relative selon la forme de la distribution

Symétrique



$$\bar{x} = Me = Mo$$

Asymétrie droite



Asymétrie droite (positive) : $Mo < Me < \bar{x}$

Quelle mesure utiliser ?

Contexte	Description	Mesure
Rendement moyen	Distribution normale, pas d'outlier	$\bar{x}$
Surface des exploitations	Très asymétrique (quelques grandes fermes)	Me
Variété la plus cultivée	Variable qualitative	Mode
Croissance démographique	Taux de croissance annuels	G
Vitesse de pulvérisation	Ratio distance/temps	H

1. Pourquoi résumer par un centre ?
2. Moyenne arithmétique
3. Moyenne géométrique et harmonique
4. Médiane
5. Mode
6. Comparaison et choix de la mesure
7. Exercices

Données

Poids de gousses d'arachide (g) récoltées sur 10 plants :

25 ; 31 ; 18 ; 42 ; 27 ; 35 ; 22 ; 29 ; 38 ; 24

Questions :

1. Calculer la moyenne arithmétique.
2. Calculer la médiane.
3. La distribution semble-t-elle symétrique ? Justifier.
4. Quel indicateur recommandez-vous si l'un des plants avait produit 120 g ?

Exercice 4.2 - Données groupées

Tableau de distribution : teneur en azote (mg/kg)

Classes	Effectif f_i	Effectif cumulé F_i
[10 ; 15[	4	4
[15 ; 20[	11	15
[20 ; 25[	18	33
[25 ; 30[	12	45
[30 ; 35[	5	50

Questions :

1. Estimer la moyenne (utiliser le milieu de chaque classe).
2. Identifier la classe médiane et calculer la médiane interpolée.
3. Identifier la classe modale.

Ce qu'il faut retenir

- ▶ **Moyenne arithmétique** : sum/n , sensible aux extrêmes
- ▶ **Moyenne géométrique** : données multiplicatives
- ▶ **Moyenne harmonique** : données en ratio
- ▶ **Médiane** : robust, distribution asymétrique
- ▶ **Mode** : valeur la plus fréquente, toutes variables

Règle de décision rapide

- ▶ Distribution symétrique $\Rightarrow$ **moyenne**
- ▶ Outliers ou asymétrie $\Rightarrow$ **médiane**
- ▶ Variable qualitative $\Rightarrow$ **mode**
- ▶ Taux de croissance $\Rightarrow$ **géométrique**
- ▶ Vitesses, débits $\Rightarrow$ **harmonique**

Merci de votre attention

Statistiques Descriptives · Module 4

Equipe Bioinformatique et Biostatistique

Laboratoire de Microbiologie, Biotechnologie et Bioinformatique

Institut National Polytechnique Félix Houphouët-Boigny

