Méthodes quantitatives pour les sciences sociales

Séance 3 : Décrire une variable

Mattéo Lanoë

Printemps 2026

SciencesPo
CENTRE DE RECHERCHE SUR LES INÉGALITÉS SOCIALES

ined
INSTITUT NATIONAL D'ÉTUDES DÉMOGRAPHIQUES

Introduction et Objectifs

Le Plan de la séance :

1. Rappels de vocabulaire (Individu, Population, Variable).

Introduction et Objectifs

Le Plan de la séance :

1. Rappels de vocabulaire (Individu, Population, Variable).
2. Les mesures de **tendance centrale** (Mode, Moyenne, Médiane).

Introduction et Objectifs

Le Plan de la séance :

1. Rappels de vocabulaire (Individu, Population, Variable).
2. Les mesures de **tendance centrale** (Mode, Moyenne, Médiane).
3. Les mesures de **dispersion** (Étendue, Écart-type).

Introduction et Objectifs

Le Plan de la séance :

1. Rappels de vocabulaire (Individu, Population, Variable).
2. Les mesures de **tendance centrale** (Mode, Moyenne, Médiane).
3. Les mesures de **dispersion** (Étendue, Écart-type).
4. Visualisation et Pratique RStudio.

Introduction et Objectifs

Le Plan de la séance :

1. Rappels de vocabulaire (Individu, Population, Variable).
2. Les mesures de **tendance centrale** (Mode, Moyenne, Médiane).
3. Les mesures de **dispersion** (Étendue, Écart-type).
4. Visualisation et Pratique RStudio.

Objectif

Savoir résumer une longue liste de chiffres en quelques indicateurs clés pour décrire la réalité sociale.

1. Vocabulaire : Lire un jeu de données

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch
259	259	1	1	Ward, Miss. Anna	female	35.00	0	
680	680	1	1	Cardeza, Mr. Thomas Drake Martinez	male	36.00	0	
738	738	1	1	Lesurer, Mr. Gustave J	male	35.00	0	
28	28	0	1	Fortune, Mr. Charles Alexander	male	19.00	3	
89	89	1	1	Fortune, Miss. Mabel Helen	female	23.00	3	
342	342	1	1	Fortune, Miss. Alice Elizabeth	female	24.00	3	
439	439	0	1	Fortune, Mr. Mark	male	64.00	1	
312	312	1	1	Ryerson, Miss. Emily Borie	female	18.00	2	
743	743	1	1	Ryerson, Miss. Susan Parker "Suzette"	female	21.00	2	
119	119	0	1	Baxter, Mr. Quigg Edmond	male	24.00	0	
300	300	1	1	Baxter, Mrs. James (Helene DeLaudeniere Chaput)	female	50.00	0	
381	381	1	1	Bidois, Miss. Rosalie	female	42.00	0	
558	558	0	1	Robbins, Mr. Victor	male	*NA*	0	
701	701	1	1	Astor, Mrs. John Jacob (Madeleine Talmadge Force)	female	18.00	1	
717	717	1	1	Endres, Miss. Caroline Louise	female	38.00	0	
528	528	0	1	Farthing, Mr. John	male	*NA*	0	

R Global Environment

Data

titanic 891 obs. of 12 variables

Files Plots Packages Help Viewer Presentation

Zoom Export

1. Vocabulaire : Lire un jeu de données

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch
259	259	1	1	Ward, Miss. Anna	female	35.00	0	
680	680	1	1	Cardeza, Mr. Thomas Drake Martinez	male	36.00	0	
738	738	1	1	Lesurer, Mr. Gustave J	male	35.00	0	
28	28	0	1	Fortune, Mr. Charles Alexander	male	19.00	3	
89	89	1	1	Fortune, Miss. Mabel Helen	female	23.00	3	
342	342	1	1	Fortune, Miss. Alice Elizabeth	female	24.00	3	
439	439	0	1	Fortune, Mr. Mark	male	64.00	1	
312	312	1	1	Ryerson, Miss. Emily Borie	female	18.00	2	
743	743	1	1	Ryerson, Miss. Susan Parker "Suzette"	female	21.00	2	
119	119	0	1	Baxter, Mr. Quigg Edmond	male	24.00	0	
300	300	1	1	Baxter, Mrs. James (Helene DeLaudeniere Chaput)	female	50.00	0	
381	381	1	1	Bidois, Miss. Rosalie	female	42.00	0	
558	558	0	1	Robbins, Mr. Victor	male	NA	0	
701	701	1	1	Astor, Mrs. John Jacob (Madeleine Talmadge Force)	female	18.00	1	
717	717	1	1	Endres, Miss. Caroline Louise	female	38.00	0	
528	528	0	1	Farthing, Mr. John	male	NA	0	

R Global Environment

Data

titanic 891 obs. of 12 variables

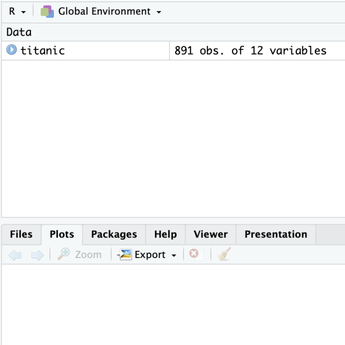
Files Plots Packages Help Viewer Presentation

Zoom Export

La Population : L'ensemble des individus concernés (Ex : Tous les passagers du Titanic).

1. Vocabulaire : Lire un jeu de données

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch
259	259	1	1	Ward, Miss. Anna	female	35.00	0	
680	680	1	1	Cardeza, Mr. Thomas Drake Martinez	male	36.00	0	
738	738	1	1	Lesurer, Mr. Gustave J	male	35.00	0	
28	28	0	1	Fortune, Mr. Charles Alexander	male	19.00	3	
89	89	1	1	Fortune, Miss. Mabel Helen	female	23.00	3	
342	342	1	1	Fortune, Miss. Alice Elizabeth	female	24.00	3	
439	439	0	1	Fortune, Mr. Mark	male	64.00	1	
312	312	1	1	Ryerson, Miss. Emily Borie	female	18.00	2	
743	743	1	1	Ryerson, Miss. Susan Parker "Suzette"	female	21.00	2	
119	119	0	1	Baxter, Mr. Quigg Edmond	male	24.00	0	
300	300	1	1	Baxter, Mrs. James (Helene DeLaudeniere Chaput)	female	50.00	0	
381	381	1	1	Bidois, Miss. Rosalie	female	42.00	0	
558	558	0	1	Robbins, Mr. Victor	male	NA	0	
701	701	1	1	Astor, Mrs. John Jacob (Madeleine Talmadge Force)	female	18.00	1	
717	717	1	1	Endres, Miss. Caroline Louise	female	38.00	0	
528	528	0	1	Farthing, Mr. John	male	NA	0	

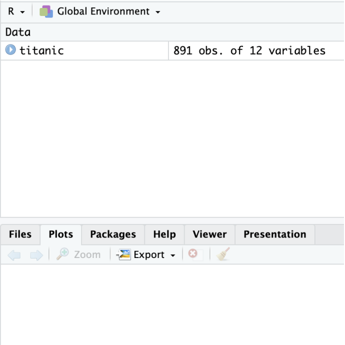


La Population : L'ensemble des individus concernés (Ex : Tous les passagers du Titanic).

L'Observation (Ligne) : L'unité statistique (Ex : Un passager).

1. Vocabulaire : Lire un jeu de données

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch
259	259	1	1	Ward, Miss. Anna	female	35.00	0	
680	680	1	1	Cardeza, Mr. Thomas Drake Martinez	male	36.00	0	
738	738	1	1	Lesurer, Mr. Gustave J	male	35.00	0	
28	28	0	1	Fortune, Mr. Charles Alexander	male	19.00	3	
89	89	1	1	Fortune, Miss. Mabel Helen	female	23.00	3	
342	342	1	1	Fortune, Miss. Alice Elizabeth	female	24.00	3	
439	439	0	1	Fortune, Mr. Mark	male	64.00	1	
312	312	1	1	Ryerson, Miss. Emily Borie	female	18.00	2	
743	743	1	1	Ryerson, Miss. Susan Parker "Suzette"	female	21.00	2	
119	119	0	1	Baxter, Mr. Quigg Edmond	male	24.00	0	
300	300	1	1	Baxter, Mrs. James (Helene DeLaudeniere Chaput)	female	50.00	0	
381	381	1	1	Bidois, Miss. Rosalie	female	42.00	0	
558	558	0	1	Robbins, Mr. Victor	male	*NA*	0	
701	701	1	1	Astor, Mrs. John Jacob (Madeleine Talmadge Force)	female	18.00	1	
717	717	1	1	Endres, Miss. Caroline Louise	female	38.00	0	
528	528	0	1	Farthing, Mr. John	male	*NA*	0	



La Population : L'ensemble des individus concernés (Ex : Tous les passagers du Titanic).

L'Observation (Ligne) : L'unité statistique (Ex : Un passager).

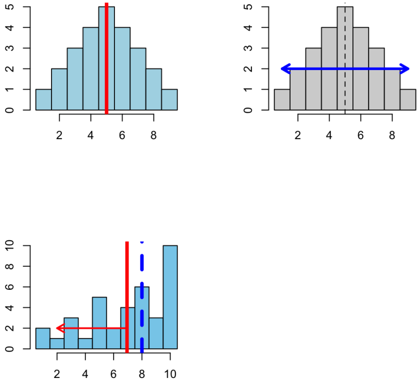
La Variable (Colonne) : Une caractéristique mesurée (Ex : Âge, Prix du billet).

2. Les mesures de tendance centrale

On cherche à résumer la distribution par une valeur "milieu".

Il existe trois logiques différentes :

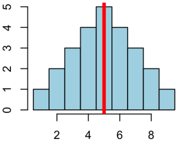
- **Le Mode** : La majorité (le plus fréquent).

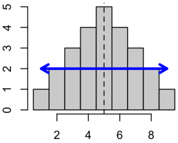


Quel indicateur résume le mieux ce graphique ?

2. Les mesures de tendance centrale

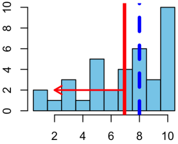
On cherche à résumer la distribution par une valeur "milieu".





Il existe trois logiques différentes :

- **Le Mode** : La majorité (le plus fréquent).
- **La Moyenne** : Le point d'équilibre (arithmétique).



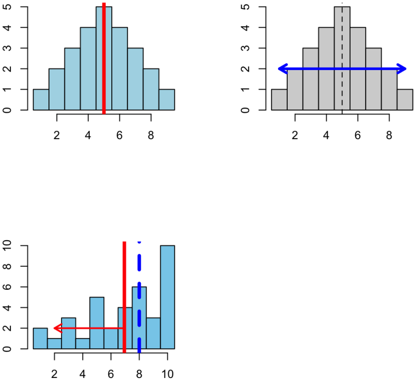
Quel indicateur résume le mieux ce graphique ?

2. Les mesures de tendance centrale

On cherche à résumer la distribution par une valeur "milieu".

Il existe trois logiques différentes :

- **Le Mode** : La majorité (le plus fréquent).
- **La Moyenne** : Le point d'équilibre (arithmétique).
- **La Médiane** : Le point de partage (50% / 50%).



Quel indicateur résume le mieux ce graphique ?

2.1 Le Mode : Définition

Définition

Le **mode** est la valeur la plus fréquente d'une variable sur une population.

2.1 Le Mode : Définition

Définition

Le **mode** est la valeur la plus fréquente d'une variable sur une population.

Exercice flash : Quel est le mode pour la série suivante : $(2; 3; \mathbf{2}; 4; \mathbf{2}; 5; 7; 8; \mathbf{2})$?

2.1 Le Mode : Définition

Définition

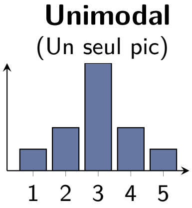
Le **mode** est la valeur la plus fréquente d'une variable sur une population.

Exercice flash : Quel est le mode pour la série suivante : $(2; 3; \mathbf{2}; 4; \mathbf{2}; 5; 7; 8; \mathbf{2})$?

$\rightarrow$ Le chiffre 2 apparaît 4 fois. C'est l'effectif le plus élevé.
Ici, le mode vaut 2.

2.1 Le Mode : Les cas de figure

Une distribution peut avoir un seul mode, plusieurs, ou aucun.

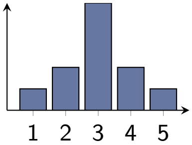


2.1 Le Mode : Les cas de figure

Une distribution peut avoir un seul mode, plusieurs, ou aucun.

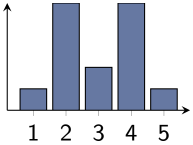
Unimodal

(Un seul pic)



Bimodal

(Plusieurs pics)

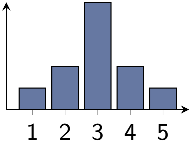


2.1 Le Mode : Les cas de figure

Une distribution peut avoir un seul mode, plusieurs, ou aucun.

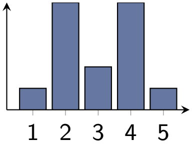
Unimodal

(Un seul pic)



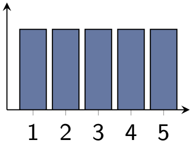
Bimodal

(Plusieurs pics)



Aucun mode

(Uniforme)



2.2 La Moyenne : Le centre de gravité

La moyenne est l'indicateur le plus connu et le plus utilisé.

Intuition

C'est une logique de **redistribution équitable**. Si on mettait toutes les valeurs dans un pot commun et qu'on redistribuait tout de manière égale à chaque individu, combien chacun aurait-il ?

2.2 La Moyenne : Le centre de gravité

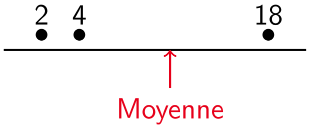
La moyenne est l'indicateur le plus connu et le plus utilisé.

Intuition

C'est une logique de **redistribution équitable**. Si on mettait toutes les valeurs dans un pot commun et qu'on redistribuait tout de manière égale à chaque individu, combien chacun aurait-il ?

Exemple : Si Paul a 0€ et Jacques a 100€, la moyenne est de 50€. Si on redistribue la masse totale (100€) équitablement sur 2 personnes, chacun a 50€.

Centre de gravité



La moyenne est le point où la planche tient en équilibre.

2.2 La Moyenne : Lire la formule mathématique

En statistiques, on utilise une notation précise. Pas de panique, c'est une simple phrase traduite en symboles.

Formule de la moyenne arithmétique ($\bar{x}$)

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

2.2 La Moyenne : Lire la formule mathématique

En statistiques, on utilise une notation précise. Pas de panique, c'est une simple phrase traduite en symboles.

Formule de la moyenne arithmétique ($\bar{x}$)

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

Décryptage :

- $\bar{x}$ (lire "x barre") : C'est le symbole de la moyenne.

2.2 La Moyenne : Lire la formule mathématique

En statistiques, on utilise une notation précise. Pas de panique, c'est une simple phrase traduite en symboles.

Formule de la moyenne arithmétique ($\bar{x}$)

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

Décryptage :

- $\bar{x}$ (lire "x barre") : C'est le symbole de la moyenne.
- N : Le nombre total d'individus (la taille de la population).

2.2 La Moyenne : Lire la formule mathématique

En statistiques, on utilise une notation précise. Pas de panique, c'est une simple phrase traduite en symboles.

Formule de la moyenne arithmétique ($\bar{x}$)

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

Décryptage :

- $\bar{x}$ (lire "x barre") : C'est le symbole de la moyenne.
- N : Le nombre total d'individus (la taille de la population).
- $\sum$ (Sigma majuscule) : C'est le symbole de la **SOMME**. Il dit : "Additionnez tout ce qui suit".

2.2 La Moyenne : Lire la formule mathématique

En statistiques, on utilise une notation précise. Pas de panique, c'est une simple phrase traduite en symboles.

Formule de la moyenne arithmétique ($\bar{x}$)

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

Décryptage :

- $\bar{x}$ (lire "x barre") : C'est le symbole de la moyenne.
- N : Le nombre total d'individus (la taille de la population).
- $\sum$ (Sigma majuscule) : C'est le symbole de la **SOMME**. Il dit : "Additionnez tout ce qui suit".
- x_i : La valeur de l'individu numéro i.

2.2 La Moyenne : Lire la formule mathématique

En statistiques, on utilise une notation précise. Pas de panique, c'est une simple phrase traduite en symboles.

Formule de la moyenne arithmétique ($\bar{x}$)

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

Décryptage :

- $\bar{x}$ (lire "x barre") : C'est le symbole de la moyenne.
- N : Le nombre total d'individus (la taille de la population).
- $\sum$ (Sigma majuscule) : C'est le symbole de la **SOMME**. Il dit : "Additionnez tout ce qui suit".
- x_i : La valeur de l'individu numéro i.

2.2 La Moyenne : Lire la formule mathématique

En statistiques, on utilise une notation précise. Pas de panique, c'est une simple phrase traduite en symboles.

Formule de la moyenne arithmétique ($\bar{x}$)

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

Décryptage :

- $\bar{x}$ (lire "x barre") : C'est le symbole de la moyenne.
- N : Le nombre total d'individus (la taille de la population).
- $\sum$ (Sigma majuscule) : C'est le symbole de la **SOMME**. Il dit : "Additionnez tout ce qui suit".
- x_i : La valeur de l'individu numéro i.

Traduction : "On additionne toutes les valeurs (x) et on divise par le nombre de personnes (N)"

2.2 La Moyenne : Exemple détaillé

Prenons les notes de l'élève **Charles** dans 7 matières.

Les données (x_i) : $8; 13; 14; 14; 14; 15; 20$.

Calcul étape par étape

1. On calcule la SOMME ($\sum x_i$) :

2.2 La Moyenne : Exemple détaillé

Prenons les notes de l'élève **Charles** dans 7 matières.

Les données (x_i) : 8; 13; 14; 14; 14; 15; 20.

Calcul étape par étape

1. On calcule la SOMME ($\sum x_i$) :

$$8 + 13 + 14 + 14 + 14 + 15 + 20 = \mathbf{98}$$

2.2 La Moyenne : Exemple détaillé

Prenons les notes de l'élève **Charles** dans 7 matières.

Les données (x_i) : $8; 13; 14; 14; 14; 15; 20$.

Calcul étape par étape

1. On calcule la SOMME ($\sum x_i$) :

$$8 + 13 + 14 + 14 + 14 + 15 + 20 = \mathbf{98}$$

2. On identifie l'EFFECTIF (N) : Il y a **7** notes.

2.2 La Moyenne : Exemple détaillé

Prenons les notes de l'élève **Charles** dans 7 matières.

Les données (x_i) : $8; 13; 14; 14; 14; 15; 20$.

Calcul étape par étape

1. On calcule la SOMME ($\sum x_i$) :

$$8 + 13 + 14 + 14 + 14 + 15 + 20 = \mathbf{98}$$

2. On identifie l'EFFECTIF (N) : Il y a **7** notes.

3. On DIVISE :

$$\bar{x} = \frac{98}{7} = \mathbf{14}$$

2.2 La Moyenne : Exemple détaillé

Prenons les notes de l'élève **Charles** dans 7 matières.

Les données (x_i) : $8; 13; 14; 14; 14; 15; 20$.

Calcul étape par étape

1. On calcule la SOMME ($\sum x_i$) :

$$8 + 13 + 14 + 14 + 14 + 15 + 20 = \mathbf{98}$$

2. On identifie l'EFFECTIF (N) : Il y a **7** notes.

3. On DIVISE :

$$\bar{x} = \frac{98}{7} = \mathbf{14}$$

La moyenne de Charles est de $14/20$.

2.2 Propriété critique : La sensibilité

La moyenne a un "défaut" sociologique majeur : elle est très influencée par les valeurs extrêmes (les *outliers*).

Exemple : Le Bar

Imaginez un bar avec 10 étudiants.

- Revenu mensuel moyen : **800 €**.

2.2 Propriété critique : La sensibilité

La moyenne a un "défaut" sociologique majeur : elle est très influencée par les valeurs extrêmes (les *outliers*).

Exemple : Le Bar

Imaginez un bar avec 10 étudiants.

- Revenu mensuel moyen : **800 €**.

Bill Gates entre dans le bar.

- Le revenu moyen passe soudainement à **150 millions d'euros**.
- Est-ce que les étudiants sont devenus riches ? Non.

2.2 Propriété critique : La sensibilité

La moyenne a un "défaut" sociologique majeur : elle est très influencée par les valeurs extrêmes (les *outliers*).

Exemple : Le Bar

Imaginez un bar avec 10 étudiants.

- Revenu mensuel moyen : **800 €**.

Bill Gates entre dans le bar.

- Le revenu moyen passe soudainement à **150 millions d'euros**.
- Est-ce que les étudiants sont devenus riches ? Non.

Conclusion :

- La moyenne "ment" quand les inégalités sont très fortes.
- Dans ce cas, on préfère la **Médiane** (qui resterait à 800€, car Bill Gates ne compte que pour "1" personne de plus).

2.3 La Moyenne Pondérée : Quand tout ne se vaut pas

Parfois, certaines valeurs ont plus d'importance (de "poids") que d'autres. On utilise des **coefficients**.

Formule

$$\bar{x}_p = \frac{\sum n_i x_i}{\sum n_i} = \frac{\text{Somme des (Note} \times \text{Coeff)}}{\text{Somme des Coeffs}}$$

2.3 La Moyenne Pondérée : Quand tout ne se vaut pas

Parfois, certaines valeurs ont plus d'importance (de "poids") que d'autres. On utilise des **coefficients**.

Formule

$$\bar{x}_p = \frac{\sum n_i x_i}{\sum n_i} = \frac{\text{Somme des (Note} \times \text{Coeff)}}{\text{Somme des Coeffs}}$$

Exemple : Semestre universitaire

- **Exposé** (Coeff 1) : L'étudiant a eu **18**/20.
- **Partiel final** (Coeff 4) : L'étudiant a eu **08**/20.

2.3 La Moyenne Pondérée : Quand tout ne se vaut pas

Parfois, certaines valeurs ont plus d'importance (de "poids") que d'autres. On utilise des **coefficients**.

Formule

$$\bar{x}_p = \frac{\sum n_i x_i}{\sum n_i} = \frac{\text{Somme des (Note} \times \text{Coeff)}}{\text{Somme des Coeffs}}$$

Exemple : Semestre universitaire

- **Exposé** (Coeff 1) : L'étudiant a eu **18**/20.
- **Partiel final** (Coeff 4) : L'étudiant a eu **08**/20.

Attention ! On ne fait pas $(18 + 8)/2 = 13$. Ce serait faux.

2.3 La Moyenne Pondérée : Quand tout ne se vaut pas

Parfois, certaines valeurs ont plus d'importance (de "poids") que d'autres. On utilise des **coefficients**.

Formule

$$\bar{x}_p = \frac{\sum n_i x_i}{\sum n_i} = \frac{\text{Somme des (Note} \times \text{Coeff)}}{\text{Somme des Coeffs}}$$

Exemple : Semestre universitaire

- **Exposé** (Coeff 1) : L'étudiant a eu **18**/20.
- **Partiel final** (Coeff 4) : L'étudiant a eu **08**/20.

Attention ! On ne fait pas $(18 + 8)/2 = 13$. Ce serait faux.

$$\bar{x}_p = \frac{(18 \times 1) + (8 \times 4)}{1 + 4} = \frac{18 + 32}{5} = \frac{50}{5} = \mathbf{10}$$

2.3 La Moyenne Pondérée : Quand tout ne se vaut pas

Parfois, certaines valeurs ont plus d'importance (de "poids") que d'autres. On utilise des **coefficients**.

Formule

$$\bar{x}_p = \frac{\sum n_i x_i}{\sum n_i} = \frac{\text{Somme des (Note} \times \text{Coeff)}}{\text{Somme des Coeffs}}$$

Exemple : Semestre universitaire

- **Exposé** (Coeff 1) : L'étudiant a eu **18**/20.
- **Partiel final** (Coeff 4) : L'étudiant a eu **08**/20.

Attention ! On ne fait pas $(18+8)/2 = 13$. Ce serait faux.

$$\bar{x}_p = \frac{(18 \times 1) + (8 \times 4)}{1+4} = \frac{18+32}{5} = \frac{50}{5} = \mathbf{10}$$

2.3 La Médiane : Définition

Définition

La **médiane** d'une variable est la valeur qui sépare la population en deux moitiés égales.

- 50% des observations sont inférieures à la médiane.
- 50% des observations sont supérieures à la médiane.

2.3 La Médiane : Définition

Définition

La **médiane** d'une variable est la valeur qui sépare la population en deux moitiés égales.

- 50% des observations sont inférieures à la médiane.
- 50% des observations sont supérieures à la médiane.

Méthode fondamentale

Étape 1 obligatoire : Ranger les valeurs dans l'**ordre croissant**.

2.3 La Médiane : Méthode de calcul (Cas Impair)

Méthode : Notons n la taille de la population.

- La médiane est la valeur de rang $(n+1)/2$.

2.3 La Médiane : Méthode de calcul (Cas Impair)

Méthode : Notons n la taille de la population.

- La médiane est la valeur de rang $(n+1)/2$.

Exemple 1 (n impair)

Série : $(8; 5; 10; -1; 19)$.

1. On trie : $-1; 5; \mathbf{8}; 10; 19$.

2.3 La Médiane : Méthode de calcul (Cas Impair)

Méthode : Notons n la taille de la population.

- La médiane est la valeur de rang $(n+1)/2$.

Exemple 1 (n impair)

Série : $(8; 5; 10; -1; 19)$.

1. On trie : $-1; 5; \mathbf{8}; 10; 19$.
2. Taille $n = 5$.

2.3 La Médiane : Méthode de calcul (Cas Impair)

Méthode : Notons n la taille de la population.

- La médiane est la valeur de rang $(n+1)/2$.

Exemple 1 (n impair)

Série : $(8; 5; 10; -1; 19)$.

1. On trie : $-1; 5; \mathbf{8}; 10; 19$.
2. Taille $n = 5$.
3. Rang : $(5+1)/2 = 3$. On cherche la 3ème valeur.

2.3 La Médiane : Méthode de calcul (Cas Impair)

Méthode : Notons n la taille de la population.

- La médiane est la valeur de rang $(n+1)/2$.

Exemple 1 (n impair)

Série : $(8; 5; 10; -1; 19)$.

1. On trie : $-1; 5; \mathbf{8}; 10; 19$.
2. Taille $n = 5$.
3. Rang : $(5+1)/2 = 3$. On cherche la 3ème valeur.

2.3 La Médiane : Méthode de calcul (Cas Impair)

Méthode : Notons n la taille de la population.

- La médiane est la valeur de rang $(n+1)/2$.

Exemple 1 (n impair)

Série : $(8; 5; 10; -1; 19)$.

1. On trie : $-1; 5; \mathbf{8}; 10; 19$.
2. Taille $n = 5$.
3. Rang : $(5+1)/2 = 3$. On cherche la 3ème valeur.

La médiane est 8.

2.3 La Médiane : Méthode de calcul (Cas Pair)

Méthode : Si n est pair.

- Par convention, la médiane est la **moyenne des deux valeurs centrales**.
- Celles de rangs $n/2$ et $(n/2)+1$.

2.3 La Médiane : Méthode de calcul (Cas Pair)

Méthode : Si n est pair.

- Par convention, la médiane est la **moyenne des deux valeurs centrales**.
- Celles de rangs $n/2$ et $(n/2) + 1$.

Exemple 2 (n pair)

Série : $(8; 5; 10; -1)$.

1. On trie : $-1; \mathbf{5}; \mathbf{8}; 10$.

2.3 La Médiane : Méthode de calcul (Cas Pair)

Méthode : Si n est pair.

- Par convention, la médiane est la **moyenne des deux valeurs centrales**.
- Celles de rangs $n/2$ et $(n/2) + 1$.

Exemple 2 (n pair)

Série : $(8; 5; 10; -1)$.

1. On trie : $-1; \mathbf{5}; \mathbf{8}; 10$.
2. Taille $n = 4$. Rangs : $4/2 = 2$ et 3.

2.3 La Médiane : Méthode de calcul (Cas Pair)

Méthode : Si n est pair.

- Par convention, la médiane est la **moyenne des deux valeurs centrales**.
- Celles de rangs $n/2$ et $(n/2) + 1$.

Exemple 2 (n pair)

Série : $(8; 5; 10; -1)$.

1. On trie : $-1; \mathbf{5}; \mathbf{8}; 10$.
2. Taille $n = 4$. Rangs : $4/2 = 2$ et 3.
3. On prend la moyenne entre la 2ème (5) et la 3ème (8) valeur.

2.3 La Médiane : Méthode de calcul (Cas Pair)

Méthode : Si n est pair.

- Par convention, la médiane est la **moyenne des deux valeurs centrales**.
- Celles de rangs $n/2$ et $(n/2)+1$.

Exemple 2 (n pair)

Série : $(8; 5; 10; -1)$.

1. On trie : $-1; \mathbf{5}; \mathbf{8}; 10$.
2. Taille $n = 4$. Rangs : $4/2 = 2$ et 3.
3. On prend la moyenne entre la 2ème (5) et la 3ème (8) valeur.

2.3 La Médiane : Méthode de calcul (Cas Pair)

Méthode : Si n est pair.

- Par convention, la médiane est la **moyenne des deux valeurs centrales**.
- Celles de rangs $n/2$ et $(n/2) + 1$.

Exemple 2 (n pair)

Série : $(8; 5; 10; -1)$.

1. On trie : $-1; \mathbf{5}; \mathbf{8}; 10$.
2. Taille $n = 4$. Rangs : $4/2 = 2$ et 3.
3. On prend la moyenne entre la 2ème (5) et la 3ème (8) valeur.

$$\text{Médiane} = \frac{5+8}{2} = \mathbf{6,5}$$

3.1 L'Étendue (Indicateur de dispersion)

Définition

L'**étendue** est la différence entre le maximum et le minimum d'une variable.

3.1 L'Étendue (Indicateur de dispersion)

Définition

L'**étendue** est la différence entre le maximum et le minimum d'une variable.

Élève	**Note**
Oriane	18
Albertine	15
Charles	11

Exemple de notes

Calcul :

3.1 L'Étendue (Indicateur de dispersion)

Définition

L'**étendue** est la différence entre le maximum et le minimum d'une variable.

Élève	**Note**
Oriane	18
Albertine	15
Charles	11

Exemple de notes

Calcul :

- Note maximum : 18
- Note minimum : 11

3.1 L'Étendue (Indicateur de dispersion)

Définition

L'**étendue** est la différence entre le maximum et le minimum d'une variable.

Élève	**Note**
Oriane	18
Albertine	15
Charles	11

Exemple de notes

Calcul :

- Note maximum : 18
- Note minimum : 11

$$\text{Étendue} = \text{Max} - \text{Min} = 18 - 11 = \mathbf{7}$$

3.2 L'écart à la moyenne

On aimerait quantifier : *"À quel point mes valeurs sont-elles éloignées de la moyenne ?"*
Prenons l'exemple de **Charles** (Moyenne $= 14$).

3.2 L'écart à la moyenne

On aimerait quantifier : *"À quel point mes valeurs sont-elles éloignées de la moyenne ?"*
Prenons l'exemple de **Charles** (Moyenne $= 14$).

Matières	Oriane	Albertine	**Charles**	**Écarts de Charles**
M1	13	8	**8**	$8 - 14 = -\mathbf{6}$
M2	13	8	**13**	$13 - 14 = -\mathbf{1}$
M3	14	14	**14**	$14 - 14 = \mathbf{0}$
M4	14	14	**14**	$\mathbf{0}$
M5	14	14	**14**	$\mathbf{0}$
M6	15	20	**15**	$15 - 14 = \mathbf{1}$
M7	15	20	**20**	$20 - 14 = \mathbf{6}$

3.2 Le problème de la somme nulle

Si on additionne les écarts de Charles :

3.2 Le problème de la somme nulle

Si on additionne les écarts de Charles :

$$(-6) + (-1) + 0 + 0 + 0 + 1 + 6 = \mathbf{0}$$

3.2 Le problème de la somme nulle

Si on additionne les écarts de Charles :

$$(-6) + (-1) + 0 + 0 + 0 + 1 + 6 = \mathbf{0}$$

Problème

Les écarts à la moyenne s'additionnent toujours à 0 ! Les "moins" compensent les "plus".

3.2 Le problème de la somme nulle

Si on additionne les écarts de Charles :

$$(-6) + (-1) + 0 + 0 + 0 + 1 + 6 = \mathbf{0}$$

Problème

Les écarts à la moyenne s'additionnent toujours à 0 ! Les "moins" compensent les "plus".

Comment éviter que les écarts ne se compensent ?

$\rightarrow$ On met tout **au carré** (car un carré est toujours positif).

3.3 La Variance

Définition

La **variance** (V) d'une variable est la moyenne des **écarts-à-la-moyenne au carré**.

3.3 La Variance

Définition

La **variance** (V) d'une variable est la moyenne des **écarts-à-la-moyenne au carré**.

Pour des valeurs x_i, on la calcule avec :

$$V = \frac{1}{n}\left((x_1 - \bar{x})^2 + ... + (x_n - \bar{x})^2\right)$$

3.3 La Variance

Définition

La **variance** (V) d'une variable est la moyenne des **écarts-à-la-moyenne au carré**.

Pour des valeurs x_i, on la calcule avec :

$$V = \frac{1}{n}\left((x_1 - \bar{x})^2 + ... + (x_n - \bar{x})^2\right)$$

Calcul pour Charles ($n = 7$, Moyenne=14)

$$V = \frac{1}{7}\left[(-6)^2 + (-1)^2 + 0^2 + 0^2 + 0^2 + 1^2 + 6^2\right]$$

3.3 La Variance

Définition

La **variance** (V) d'une variable est la moyenne des **écarts-à-la-moyenne au carré**.

Pour des valeurs x_i, on la calcule avec :

$$V = \frac{1}{n}\left((x_1 - \bar{x})^2 + ... + (x_n - \bar{x})^2\right)$$

Calcul pour Charles ($n = 7$, Moyenne=14)

$$V = \frac{1}{7}\left[(-6)^2 + (-1)^2 + 0^2 + 0^2 + 0^2 + 1^2 + 6^2\right]$$

$$V = \frac{1}{7}[36 + 1 + 0 + 0 + 0 + 1 + 36] = \frac{74}{7} \approx \mathbf{10,57}$$

3.4 L'Écart-type

Définition

L'**écart-type** (σ) est la racine carrée de la variance. Il permet de revenir à l'unité d'origine (ici, des points).

3.4 L'Écart-type

Définition

L'**écart-type** (σ) est la racine carrée de la variance. Il permet de revenir à l'unité d'origine (ici, des points).

$$\sigma = \sqrt{V}$$

3.4 L'Écart-type

Définition

L'**écart-type** (σ) est la racine carrée de la variance. Il permet de revenir à l'unité d'origine (ici, des points).

$$\sigma = \sqrt{V}$$

Calcul pour Charles

$$\sigma \approx \sqrt{10,57} \approx \mathbf{3,25}$$

3.4 L'Écart-type

Définition

L'**écart-type** (σ) est la racine carrée de la variance. Il permet de revenir à l'unité d'origine (ici, des points).

$$\sigma = \sqrt{V}$$

Calcul pour Charles

$$\sigma \approx \sqrt{10,57} \approx \mathbf{3,25}$$

Comparaison :

- Oriane : $\sigma \approx 0,76$ (Très faible $\rightarrow$ Homogène).

3.4 L'Écart-type

Définition

L'**écart-type** (σ) est la racine carrée de la variance. Il permet de revenir à l'unité d'origine (ici, des points).

$$\sigma = \sqrt{V}$$

Calcul pour Charles

$$\sigma \approx \sqrt{10,57} \approx \mathbf{3,25}$$

Comparaison :

- Oriane : $\sigma \approx 0,76$ (Très faible $\rightarrow$ Homogène).
- Albertine : $\sigma \approx 4,5$ (Très fort $\rightarrow$ Hétérogène).

3.5 Caractéristiques de l'écart-type (Résumé)

- **Toujours positif** ($\sigma \geq 0$).

3.5 Caractéristiques de l'écart-type (Résumé)

- **Toujours positif** ($\sigma \geq 0$).
- S'il est nul ($\sigma = 0$), alors toutes les valeurs sont identiques (constante).

3.5 Caractéristiques de l'écart-type (Résumé)

- **Toujours positif** ($\sigma \geq 0$).
- S'il est nul ($\sigma = 0$), alors toutes les valeurs sont identiques (constante).
- Il a la **même unité** que la variable (euros, mètres, points...). La variance n'a pas la même unité (*points*^{2}).

3.5 Caractéristiques de l'écart-type (Résumé)

- **Toujours positif** ($\sigma \geq 0$).
- S'il est nul ($\sigma = 0$), alors toutes les valeurs sont identiques (constante).
- Il a la **même unité** que la variable (euros, mètres, points...). La variance n'a pas la même unité (*points*^{2}).
- **Interprétation :**

3.5 Caractéristiques de l'écart-type (Résumé)

- **Toujours positif** ($\sigma \geq 0$).
- S'il est nul ($\sigma = 0$), alors toutes les valeurs sont identiques (constante).
- Il a la **même unité** que la variable (euros, mètres, points...). La variance n'a pas la même unité (*points*^{2}).
- **Interprétation :**
  - Écart-type **faible** $=$ Données serrées autour de la moyenne (*Homogénéité, exemple d'Oriane*).

3.5 Caractéristiques de l'écart-type (Résumé)

- **Toujours positif** ($\sigma \geq 0$).
- S'il est nul ($\sigma = 0$), alors toutes les valeurs sont identiques (constante).
- Il a la **même unité** que la variable (euros, mètres, points...). La variance n'a pas la même unité (*points*^{2}).
- **Interprétation :**
  - Écart-type **faible** $=$ Données serrées autour de la moyenne (*Homogénéité, exemple d'Oriane*).
  - Écart-type **fort** $=$ Données dispersées (*Hétérogénéité, exemple d'Albertine*).

3.5 Caractéristiques de l'écart-type (Résumé)

- **Toujours positif** ($\sigma \geq 0$).
- S'il est nul ($\sigma = 0$), alors toutes les valeurs sont identiques (constante).
- Il a la **même unité** que la variable (euros, mètres, points...). La variance n'a pas la même unité (*points*^{2}).
- **Interprétation :**
  - Écart-type **faible** = Données serrées autour de la moyenne (*Homogénéité, exemple d'Oriane*).
  - Écart-type **fort** = Données dispersées (*Hétérogénéité, exemple d'Albertine*).
- **Sensibilité :** Comme on utilise des carrés, il est très sensible aux valeurs extrêmes.

4.1 Les Quantiles : Définition

Définition

Les **quantiles** sont les bornes qui partitionnent les valeurs en **effectifs égaux**.

4.1 Les Quantiles : Définition

Définition

Les **quantiles** sont les bornes qui partitionnent les valeurs en **effectifs égaux**.

Les plus utilisés sont les **quartiles** (Q1, Q2, Q3) qui séparent les données en 4 quarts (25% chacun).

4.1 Les Quantiles : Définition

Définition

Les **quantiles** sont les bornes qui partitionnent les valeurs en **effectifs égaux**.

Les plus utilisés sont les **quartiles** (Q1, Q2, Q3) qui séparent les données en 4 quarts (25% chacun).

Attention !

Le but n'est **pas** de séparer la plage de valeurs (ex : 0 à 20) en quatre. Les quartiles ne sont pas 5, 10, 15. **Ils répartissent les effectifs (les gens), pas les valeurs.**

4.2 Les Quantiles : Méthode

Méthode

1. On range les valeurs dans l'**ordre croissant**.

4.2 Les Quantiles : Méthode

Méthode

1. On range les valeurs dans l'**ordre croissant**.
2. On trouve la médiane (C'est **Q2**).

4.2 Les Quantiles : Méthode

Méthode

1. On range les valeurs dans l'**ordre croissant**.
2. On trouve la médiane (C'est **Q2**).
3. La médiane de la première moitié est **Q1**.

4.2 Les Quantiles : Méthode

Méthode

1. On range les valeurs dans l'**ordre croissant**.
2. On trouve la médiane (C'est **Q2**).
3. La médiane de la première moitié est **Q1**.
4. La médiane de la seconde moitié est **Q3**.

4.2 Les Quantiles : Méthode

Méthode

1. On range les valeurs dans l'**ordre croissant**.
2. On trouve la médiane (C'est **Q2**).
3. La médiane de la première moitié est **Q1**.
4. La médiane de la seconde moitié est **Q3**.

4.2 Les Quantiles : Méthode

Méthode

1. On range les valeurs dans l'**ordre croissant**.
2. On trouve la médiane (C'est **Q2**).
3. La médiane de la première moitié est **Q1**.
4. La médiane de la seconde moitié est **Q3**.

Exemple IDH (Indicateur de Développement Humain) :

- $Q1 = 0,622$ (25% des pays sont en dessous).

4.2 Les Quantiles : Méthode

Méthode

1. On range les valeurs dans l'**ordre croissant**.
2. On trouve la médiane (C'est **Q2**).
3. La médiane de la première moitié est **Q1**.
4. La médiane de la seconde moitié est **Q3**.

Exemple IDH (Indicateur de Développement Humain) :

- $Q1 = 0,622$ (25% des pays sont en dessous).
- $Q2 = 0,762$ (La médiane : 50% des pays sont en dessous).

4.2 Les Quantiles : Méthode

Méthode

1. On range les valeurs dans l'**ordre croissant**.
2. On trouve la médiane (C'est **Q2**).
3. La médiane de la première moitié est **Q1**.
4. La médiane de la seconde moitié est **Q3**.

Exemple IDH (Indicateur de Développement Humain) :

- $Q1 = 0,622$ (25% des pays sont en dessous).
- $Q2 = 0,762$ (La médiane : 50% des pays sont en dessous).
- $Q3 = 0,862$ (75% des pays sont en dessous).

4.2 Les Quantiles : Méthode

Méthode

1. On range les valeurs dans l'**ordre croissant**.
2. On trouve la médiane (C'est **Q2**).
3. La médiane de la première moitié est **Q1**.
4. La médiane de la seconde moitié est **Q3**.

Exemple IDH (Indicateur de Développement Humain) :

- $Q1 = 0,622$ (25% des pays sont en dessous).
- $Q2 = 0,762$ (La médiane : 50% des pays sont en dessous).
- $Q3 = 0,862$ (75% des pays sont en dessous).

4.2 Les Quantiles : Méthode

Méthode

1. On range les valeurs dans l'**ordre croissant**.
2. On trouve la médiane (C'est **Q2**).
3. La médiane de la première moitié est **Q1**.
4. La médiane de la seconde moitié est **Q3**.

Exemple IDH (Indicateur de Développement Humain) :

- $Q1 = 0,622$ (25% des pays sont en dessous).
- $Q2 = 0,762$ (La médiane : 50% des pays sont en dessous).
- $Q3 = 0,862$ (75% des pays sont en dessous).

Note : Si on divise en 10 parts = Déciles. En 100 parts = Centiles.

4.3 Dispersion autour de la médiane

Comment mesurer la dispersion sans utiliser la moyenne (et donc sans variance) ?

4.3 Dispersion autour de la médiane

Comment mesurer la dispersion sans utiliser la moyenne (et donc sans variance) ?

L'intervalle interquartile (IQR)

C'est la différence entre le troisième et le premier quartile.

$$IQR = Q3 - Q1$$

4.3 Dispersion autour de la médiane

Comment mesurer la dispersion sans utiliser la moyenne (et donc sans variance) ?

L'intervalle interquartile (IQR)

C'est la différence entre le troisième et le premier quartile.

$$IQR = Q3 - Q1$$

- Il mesure la dispersion autour de la médiane.

4.3 Dispersion autour de la médiane

Comment mesurer la dispersion sans utiliser la moyenne (et donc sans variance) ?

L'intervalle interquartile (IQR)

C'est la différence entre le troisième et le premier quartile.

$$IQR = Q3 - Q1$$

- Il mesure la dispersion autour de la médiane.
- Il contient **50% des valeurs centrales** de la population.

4.3 Dispersion autour de la médiane

Comment mesurer la dispersion sans utiliser la moyenne (et donc sans variance) ?

L'intervalle interquartile (IQR)

C'est la différence entre le troisième et le premier quartile.

$$IQR = Q3 - Q1$$

- Il mesure la dispersion autour de la médiane.
- Il contient **50% des valeurs centrales** de la population.

4.3 Dispersion autour de la médiane

Comment mesurer la dispersion sans utiliser la moyenne (et donc sans variance) ?

L'intervalle interquartile (IQR)

C'est la différence entre le troisième et le premier quartile.

$$IQR = Q3 - Q1$$

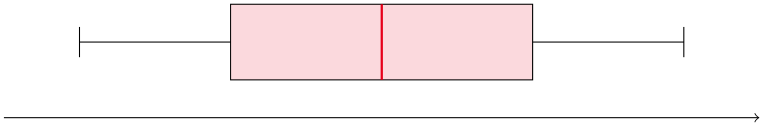
- Il mesure la dispersion autour de la médiane.
- Il contient **50% des valeurs centrales** de la population.

Pour l'exemple de l'IDH :

$$IQR = 0,862 - 0,622 = \mathbf{0,24}$$

4. Visualiser : La Boîte à Moustaches

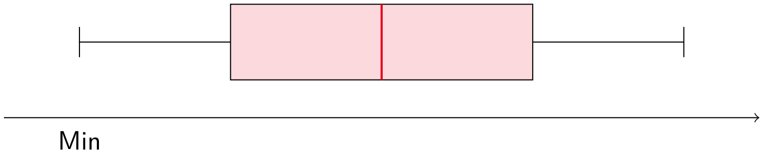
Pour résumer Centre + Dispersion + Outliers en un seul dessin.



- **La Boîte** contient 50% de la population centrale.

4. Visualiser : La Boîte à Moustaches

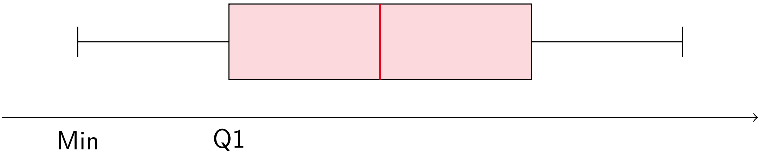
Pour résumer Centre + Dispersion + Outliers en un seul dessin.



- **La Boîte** contient 50% de la population centrale.
- **La Médiane** coupe la boîte.

4. Visualiser : La Boîte à Moustaches

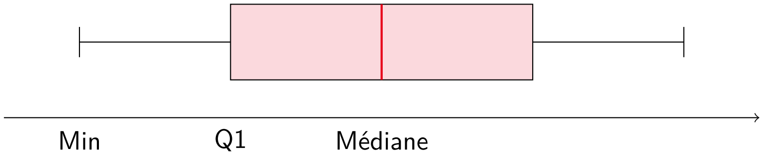
Pour résumer Centre + Dispersion + Outliers en un seul dessin.



- **La Boîte** contient 50% de la population centrale.
- **La Médiane** coupe la boîte.
- **L'écart inter-quartile** (largeur de la boîte) mesure la dispersion.

4. Visualiser : La Boîte à Moustaches

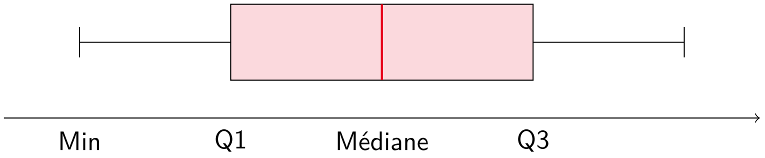
Pour résumer Centre + Dispersion + Outliers en un seul dessin.



- **La Boîte** contient 50% de la population centrale.
- **La Médiane** coupe la boîte.
- **L'écart inter-quartile** (largeur de la boîte) mesure la dispersion.

4. Visualiser : La Boîte à Moustaches

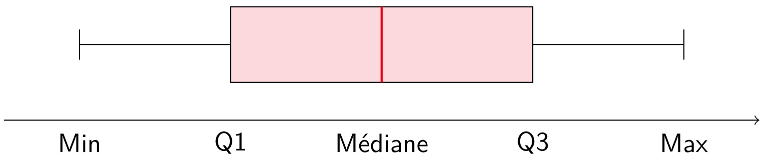
Pour résumer Centre + Dispersion + Outliers en un seul dessin.



- **La Boîte** contient 50% de la population centrale.
- **La Médiane** coupe la boîte.
- **L'écart inter-quartile** (largeur de la boîte) mesure la dispersion.

4. Visualiser : La Boîte à Moustaches

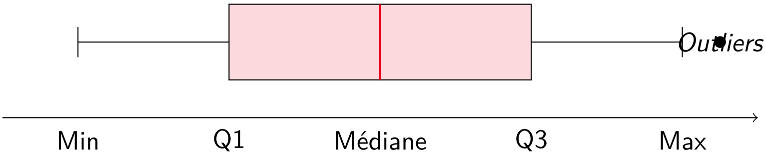
Pour résumer Centre + Dispersion + Outliers en un seul dessin.



- **La Boîte** contient 50% de la population centrale.
- **La Médiane** coupe la boîte.
- **L'écart inter-quartile** (largeur de la boîte) mesure la dispersion.

4. Visualiser : La Boîte à Moustaches

Pour résumer Centre + Dispersion + Outliers en un seul dessin.



- **La Boîte** contient 50% de la population centrale.
- **La Médiane** coupe la boîte.
- **L'écart inter-quartile** (largeur de la boîte) mesure la dispersion.