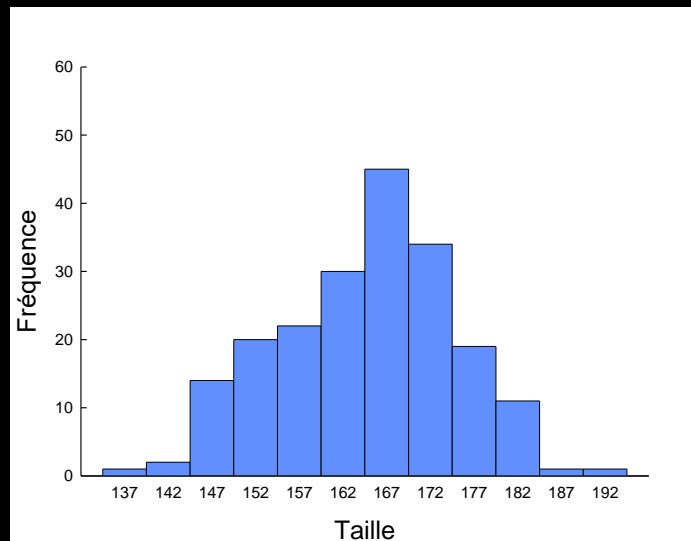


# Chapitre 2

## Statistiques descriptives univariées



# Statistique descriptive et inférentielle

---

## La statistique descriptive

Sert à synthétiser un ensemble de données  
(graphes, tableaux, moyenne...)

## La statistique inférentielle

Tirer des conclusions sur une population à partir d'un échantillon

**La statistique descriptive est  
l'ensemble des outils qui  
permettent de résumer des  
données obtenues sur une ou  
plusieurs variables**

# Techniques de statistique descriptive

---

1. Représentation graphique
  - Table de fréquence
  - Graphiques (circulaires, en barres, ...)
2. Mesures de tendance centrale et de dispersion
3. Mesures de relation (chapitre 3)

# Résultats de 50 sujets sur la variable nominale « Comportement fumeur »

|               |               |               |               |               |
|---------------|---------------|---------------|---------------|---------------|
| Fumeur        | Non-fumeur    | Ancien fumeur | Fumeur        | Non-fumeur    |
| Non-fumeur    | Non-fumeur    | Fumeur        | Non-fumeur    | Non-fumeur    |
| Ancien fumeur | Fumeur        | Non-fumeur    | Non-fumeur    | Non-fumeur    |
| Fumeur        | Non-fumeur    | Non-fumeur    | Ancien fumeur | Fumeur        |
| Non-fumeur    | Fumeur        | Fumeur        | Non-fumeur    | Non-fumeur    |
| Fumeur        | Non-fumeur    | Non-fumeur    | Non-fumeur    | Ancien fumeur |
| Non-fumeur    | Non-fumeur    | Fumeur        | Fumeur        | Non-fumeur    |
| Ancien fumeur | Fumeur        | Ancien fumeur | Non-fumeur    | Non-fumeur    |
| Non-fumeur    | Non-fumeur    | Non-fumeur    | Ancien fumeur | Non-fumeur    |
| Non-fumeur    | Ancien fumeur | Fumeur        | Non-fumeur    | Fumeur        |

# Résultats de 50 sujets sur la variable métrique « Temps à une tâche de résolution de problème »

|     |     |     |     |     |
|-----|-----|-----|-----|-----|
| 573 | 538 | 572 | 750 | 649 |
| 485 | 448 | 567 | 455 | 681 |
| 622 | 434 | 721 | 568 | 773 |
| 715 | 512 | 593 | 597 | 592 |
| 708 | 530 | 583 | 603 | 553 |
| 756 | 409 | 554 | 571 | 661 |
| 403 | 549 | 777 | 798 | 566 |
| 579 | 564 | 678 | 443 | 668 |
| 699 | 612 | 814 | 534 | 470 |
| 502 | 569 | 541 | 463 | 524 |

# Représentation graphique

---

## Représentation d'une variable nominale

- Tables de fréquence
- Graphiques circulaires
- Diagrammes en barre

## Représentation d'une variable ordinale ou métrique

- Tables de fréquence groupées
- Histogrammes

# Les fréquences (absolues)

---

Fréquences ou Occurrences

= Nombre de fois qu'une valeur apparaît dans un jeu de données

Ex: Fréquence de « fumeur » = 14



# La table de fréquence

---

**= Tableau qui indique les fréquences de chaque valeur d'un jeu de données**

**Quatre étapes de construction:**

- 1. Lister toutes les valeurs observées**
- 2. Compter les fréquences de chaque valeur**
- 3. Indiquer les fréquences à droite des valeurs**
- 4. Indiquer le total au bas du tableau**

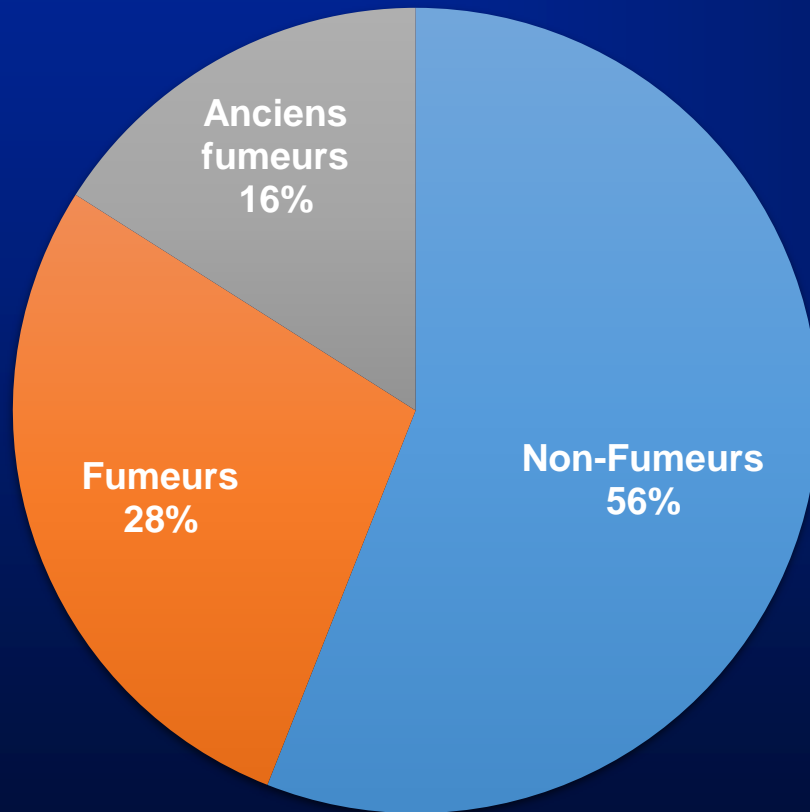
# Résultats de 50 sujets sur la variable nominale « Comportement fumeur »

|               |               |               |               |               |
|---------------|---------------|---------------|---------------|---------------|
| Fumeur        | Non-fumeur    | Ancien fumeur | Fumeur        | Non-fumeur    |
| Non-fumeur    | Non-fumeur    | Fumeur        | Non-fumeur    | Non-fumeur    |
| Ancien fumeur | Fumeur        | Non-fumeur    | Non-fumeur    | Non-fumeur    |
| Fumeur        | Non-fumeur    | Non-fumeur    | Ancien fumeur | Fumeur        |
| Non-fumeur    | Fumeur        | Fumeur        | Non-fumeur    | Non-fumeur    |
| Fumeur        | Non-fumeur    | Non-fumeur    | Non-fumeur    | Ancien fumeur |
| Non-fumeur    | Non-fumeur    | Fumeur        | Fumeur        | Non-fumeur    |
| Ancien fumeur | Fumeur        | Ancien fumeur | Non-fumeur    | Non-fumeur    |
| Non-fumeur    | Non-fumeur    | Non-fumeur    | Ancien fumeur | Non-fumeur    |
| Non-fumeur    | Ancien fumeur | Fumeur        | Non-fumeur    | Fumeur        |

# La table de fréquences

|                 | Fréquences | Pourcentages |
|-----------------|------------|--------------|
| Non-Fumeurs     | 28         | 56           |
| Fumeurs         | 14         | 28           |
| Anciens fumeurs | 8          | 16           |
| Total           | 50         | 100          |

# Le graphique circulaire

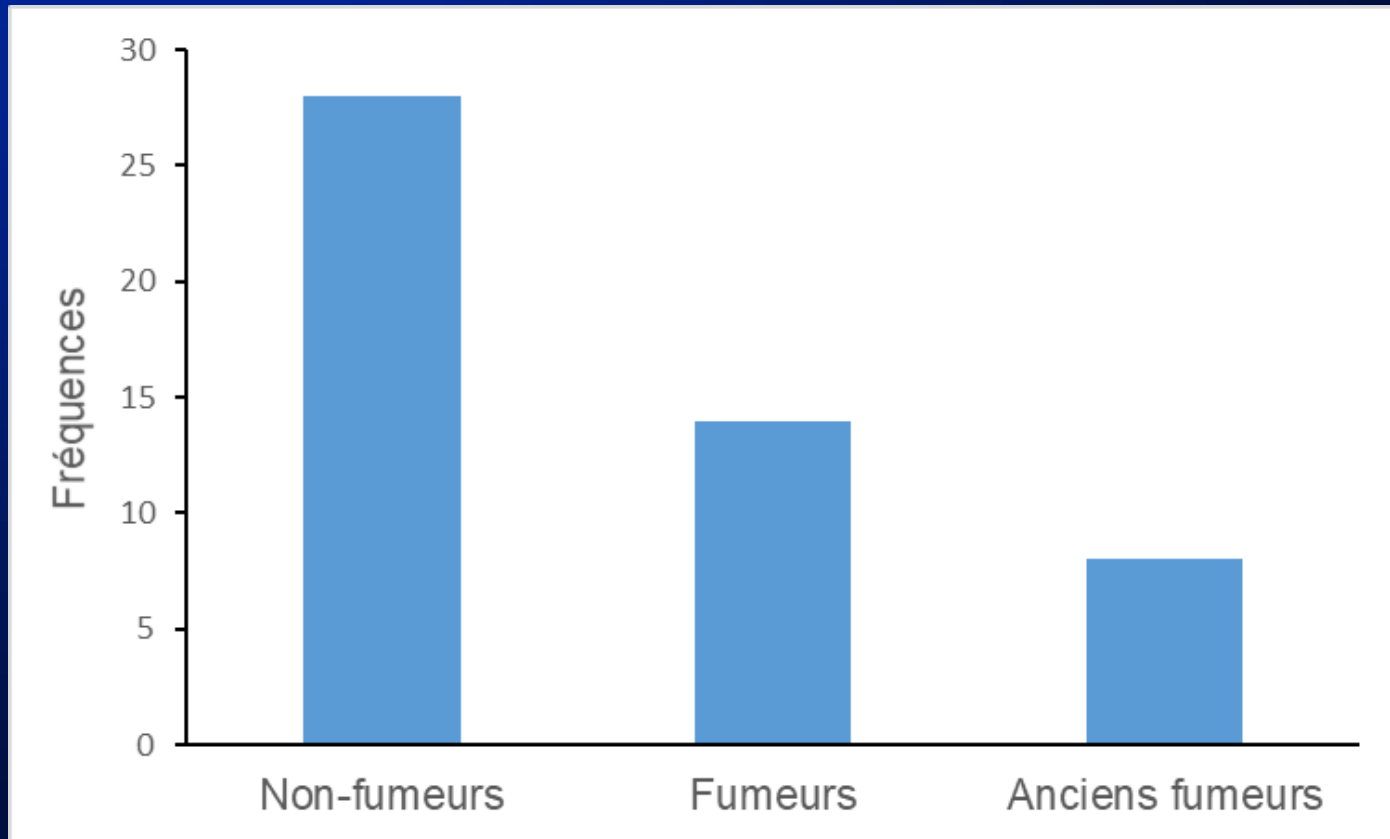


**Diagramme circulaire**

**Diagramme en secteurs**

**Camembert**

# Le diagramme en barres



# Remarques sur les pourcentages

## Fréquence relative ou proportion de fréquence

$$\text{Fréquence relative} = \frac{\text{Fréquence de la catégorie}}{\text{Somme des fréquences de toutes les catégories}}$$

**valeur entre 0 et 1**

## Pourcentage de fréquence

$$\text{Pourcentage de fréquence} = \frac{\text{Fréquence de la catégorie} \times 100}{\text{Somme des fréquences de toutes les catégories}}$$

**valeur entre 0 et 100**

# Remarques sur les pourcentages

Attention, les pourcentages peuvent être :

## 1. Les scores d'une variable métrique

Chaque sujet obtient un pourcentage

Ex: le sujet 8 a obtenu 50% à son examen

## 2. Une synthèse des valeurs d'une variable nominale

Le pourcentage porte sur plusieurs sujets

Ex: 40% des sujets de l'échantillon sont fumeurs

# Représentation graphique d'une variable ordinaire ou métrique

|     |     |     |     |     |
|-----|-----|-----|-----|-----|
| 573 | 538 | 572 | 750 | 649 |
| 485 | 448 | 567 | 455 | 681 |
| 622 | 434 | 721 | 568 | 773 |
| 715 | 512 | 593 | 597 | 592 |
| 708 | 530 | 583 | 603 | 553 |
| 756 | 409 | 554 | 571 | 661 |
| 403 | 549 | 777 | 798 | 566 |
| 579 | 564 | 678 | 443 | 668 |
| 699 | 612 | 814 | 534 | 470 |
| 502 | 569 | 541 | 463 | 524 |



# La table de fréquences groupées

**= Table de fréquences dans laquelle les valeurs ont été regroupées en intervalles**

**Quatre étapes de construction:**

- 1. Fixer le nombre et la taille des intervalles**
- 2. Indiquer tous les intervalles en ordre**
- 3. Indiquer les fréquences de chaque intervalle**
- 4. Indiquer les pourcentages ou fréquences cumulées**

# La table de fréquences groupées

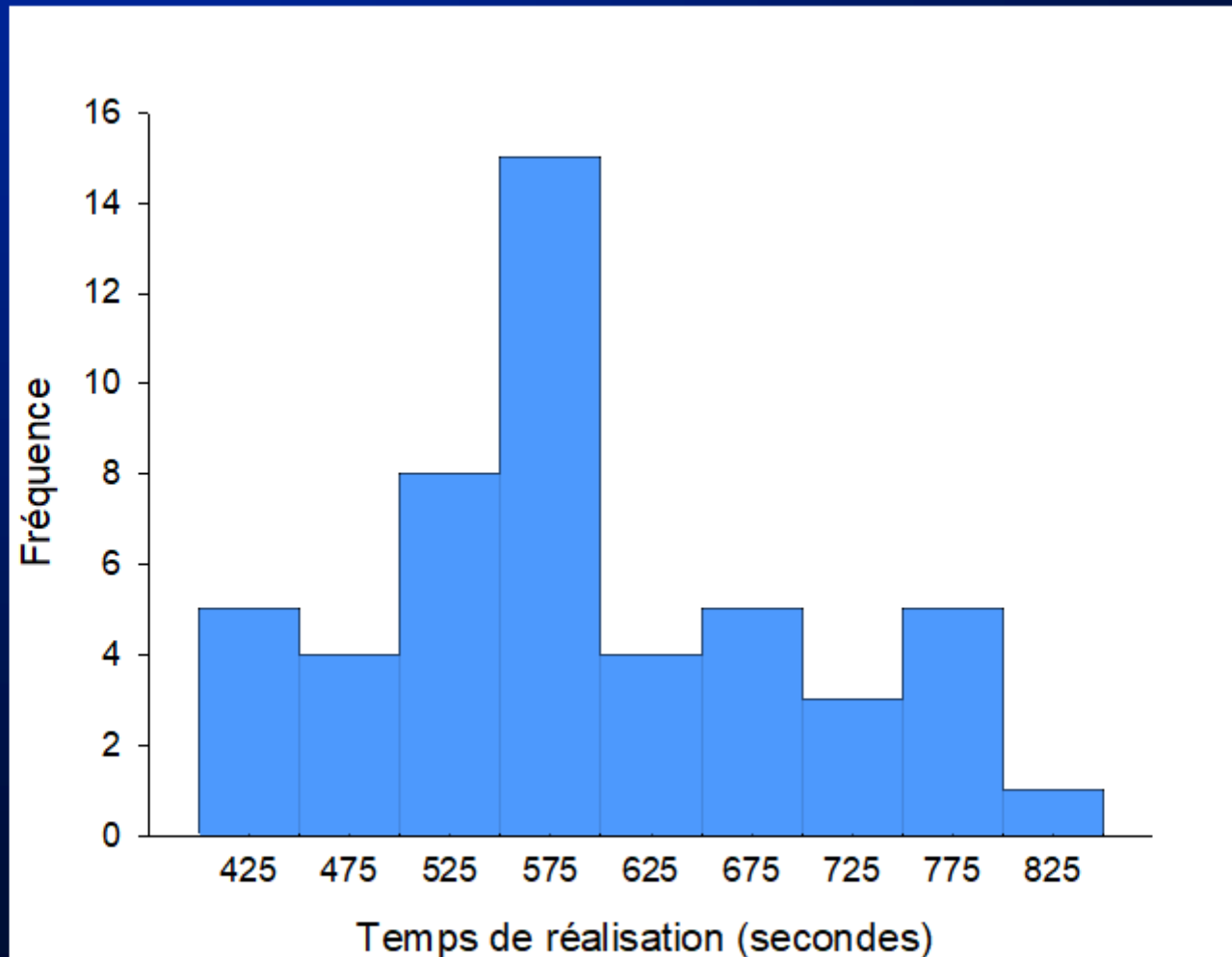
| Intervalle | Point central | Fréquence | Fréquence cumulée |
|------------|---------------|-----------|-------------------|
| 400-450    | 425           | 5         | 5                 |
| 450-500    | 475           | 4         | 9                 |
| 500-550    | 525           | 8         | 17                |
| 550-600    | 575           | 15        | 32                |
| 600-650    | 625           | 4         | 36                |
| 650-700    | 675           | 5         | 41                |
| 700-750    | 725           | 3         | 44                |
| 750-800    | 775           | 5         | 49                |
| 800-850    | 825           | 1         | 50                |

# La table de fréquences groupées

## Règles à respecter

- 5 à 20 intervalles de taille égale
- Taille des intervalles = étendue / nbr d'intervalles
- Ordre croissant ou décroissant
- Représenter tous les intervalles
- Convention pour les valeurs qui tombent sur les bornes
- Facultatif: ajouter les fréquences ou % cumulés

# Les histogrammes de fréquence



# Les mesures de tendance centrale

---

= le centre d'une distribution , la valeur typique

- Le mode
- La médiane
- La moyenne

# Le mode

---

**= la valeur la plus fréquente**

**Distribution unimodale**

**Distribution uniforme**

**Distribution bimodale**

**Distribution multimodale**

# Le mode (exemple)

Test de culture générale,  $X$  = résultat

| <u>X (résultat sur 20)</u> | <u>Fréquence</u> |
|----------------------------|------------------|
| 8                          | 5                |
| 9                          | 8                |
| 10                         | 15               |
| 11                         | 22               |
| 12                         | 25               |
| 13                         | 38               |
| 14                         | 54               |
| 15                         | 30               |
| 16                         | 32               |
| 17                         | 15               |
| 18                         | 8                |
| 19                         | 2                |
| 20                         | 1                |

# Le mode (exemple)

Test de culture générale, X = résultat

| <u>X (résultat sur 20)</u> | <u>Fréquence</u> |
|----------------------------|------------------|
| 8                          | 5                |
| 9                          | 8                |
| 10                         | 15               |
| 11                         | 22               |
| 12                         | 25               |
| 13                         | 38               |
| 14                         | 54               |
| 15                         | 30               |
| 16                         | 32               |
| 17                         | 15               |
| 18                         | 8                |
| 19                         | 2                |
| 20                         | 1                |

← Le Mode = 14



# Le mode

---

- **Mesure de tendance centrale pour les variables nominales**
- **Peu utile pour les variables métriques**
- **Utilité variable pour les variables ordinales**

# La médiane

---

**= la donnée centrale lorsque valeurs classées en ordre croissant**

**= le score par rapport auquel il y autant de scores qui sont plus grands que de scores plus petits**

# La médiane (exemple)

9 sujets ont été soumis à une épreuve de mémoire

$X$  = nombre d'erreurs commises

$X = 3, 5, 5, 7, 8, 9, 12, 13, 15$



La médiane

# Calcul de la médiane avec un nombre impair de scores

$X = 7, 12, 5, 9, 8, 5, 15, 13, 3$

Etape 1: Disposer les scores en ordre croissant

$3, 5, 5, 7, 8, 9, 12, 13, 15$

Etape 2: calculer la position médiane

$$\frac{N+1}{2} = \frac{9+1}{2} = 5$$

# Calcul de la médiane avec un nombre impair de scores

---

Etape 3: La médiane se trouve à la position médiane

= la cinquième valeur

3, 5, 5, 7, 8, 9, 12, 13, 15

# Calcul de la médiane avec un nombre pair de scores

**$X = 7, 12, 5, 9, 8, 5, 15, 13, 3, 15, 6, 11$  (N=12)**

**Etape 1: Disposer les scores en ordre croissant**

**3, 5, 5, 6, 7, 8, 9, 11, 12, 13, 15, 15**

**Etape 2: calculer la position médiane**

$$\frac{N+1}{2} = \frac{12+1}{2} = 6,5$$

# Calcul de la médiane avec un nombre pair de scores

Etape 3: La médiane se trouve à la position médiane

= la valeur entre la position 6 et 7

3, 5, 5, 6, 7, 8, ● 9, 11, 12, 13, 15, 15



**Médiane**

On prend la moyenne des 6ème et 7ème valeurs

**Médiane = 8,5**

# La moyenne arithmétique

= la somme des scores divisée par le nombre de scores

$$\overline{X} = \frac{\sum X}{N} = \frac{X_1 + X_2 + X_3 + \dots + X_N}{N}$$



# La moyenne (exemple)

Erreurs commises : 7, 12, 5, 9, 8, 5, 15, 13, 3, 15, 6, 11

N=12

$$\overline{X} = \frac{7+12+5+9+8+5+15+13+3+15+6+11}{12}$$

$$= \frac{109}{12} = 9,08$$

# Calculer la moyenne à partir d'un tableau de fréquence

| <u>X (résultat sur 20)</u> | <u>Fréquence</u> |
|----------------------------|------------------|
| 8                          | 5                |
| 9                          | 8                |
| 10                         | 15               |
| 11                         | 22               |
| 12                         | 25               |
| 13                         | 38               |
| 14                         | 54               |
| 15                         | 30               |
| 16                         | 32               |
| 17                         | 15               |
| 18                         | 8                |
| 19                         | 2                |
| 20                         | 1                |

$$\overline{X} = \frac{8+8+8+8+8+9+\dots+20}{255} = \frac{3473}{255} = 13,62$$

# Calculer la moyenne à partir d'un tableau de fréquence

| <u>X (résultat sur 20)</u> | <u>Fréquence</u> |
|----------------------------|------------------|
| 8                          | 5                |
| 9                          | 8                |
| 10                         | 15               |
| 11                         | 22               |
| 12                         | 25               |
| 13                         | 38               |
| 14                         | 54               |
| 15                         | 30               |
| 16                         | 32               |
| 17                         | 15               |
| 18                         | 8                |
| 19                         | 2                |
| 20                         | 1                |

$$\overline{X} = \frac{\sum f_i X_i}{N}$$

$$\overline{X} = \frac{(5 \times 8) + (8 \times 9) + \dots + (2 \times 19) + (1 \times 20)}{255} = \frac{3473}{255} = 13,62$$

# Choix de la mesure de tendance centrale

---

Dépend de quatre caractéristiques :

- Sensibilité aux scores extrêmes
- Stabilité
- Être un score réellement observé
- Utilisation dans les équations

# Choix de la mesure de tendance centrale

---

## Sensibilité aux scores extrêmes

- Mode et médiane sont insensibles aux scores extrêmes
- Moyenne est (très) sensible aux scores extrêmes
- Moyenne est problématique pour les variables métriques avec une distribution asymétrique

# Choix de la mesure de tendance centrale

---

## Sensibilité aux scores extrêmes – Exemple

**X = Temps de latence pour répondre à une cible**

**X = 360, 360, 390, 400, 420, 450, 470, 480, 520**

**Mode = 360**

**Médiane = 420**

**Moyenne = 427,78**

# Choix de la mesure de tendance centrale

---

## Sensibilité aux scores extrêmes – Exemple

**X = Temps de latence pour répondre à une cible**

**X = 360, 360, 390, 400, 420, 450, 470, 480, 2520**

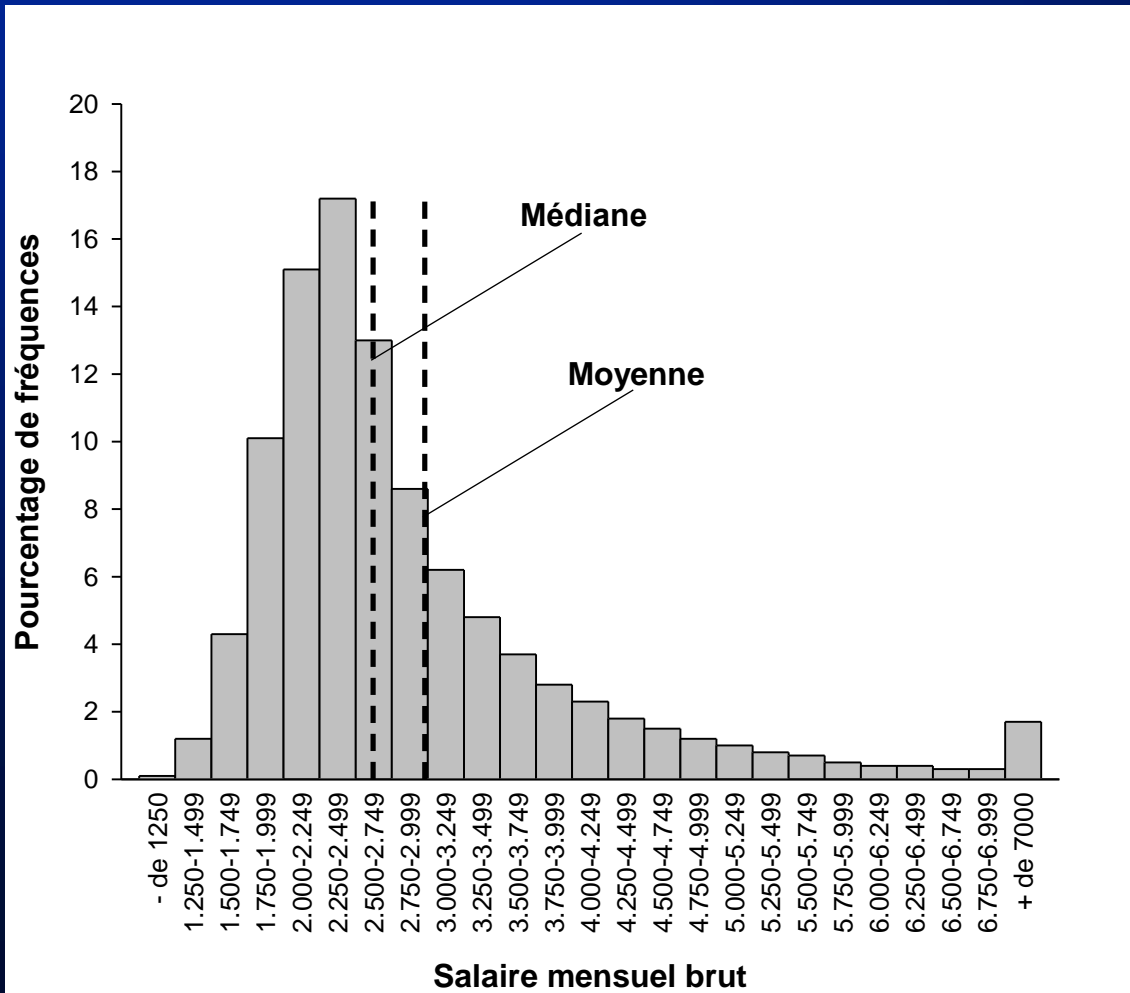
**Mode = 360**

**Médiane = 420**

**Moyenne = 650**

# Choix de la mesure de tendance centrale

## Sensibilité aux scores extrêmes



**Médiane est une meilleure mesure de tendance centrale pour les distributions asymétriques**



# Choix de la mesure de tendance centrale

---

## Stabilité

Si on tire successivement des échantillons dans une même population, les mesures sont-elles stables ?

**Moyenne > Médiane > Mode**

# Choix de la mesure de tendance centrale

---

## La mesure est une valeur réellement observée

- Mode est toujours une valeur observée
- Médiane est très souvent une valeur observée
- Moyenne peut être une valeur impossible

Ex: Dans un pays les femmes ont en moyenne 1,8 enfants

# Choix de la mesure de tendance centrale

---

## Utilisation dans les équations

- Moyenne est manipulable sous forme d'équation
- Mode et médiane ne s'écrivent pas sous forme d'équation

|                | Est utilisé principalement avec   | Avantages  | Inconvénients   |
|----------------|---|--|---|
| <b>Mode</b>    | <ul style="list-style-type: none"> <li>- Variables nominales</li> <li>- Variables ordinales non chiffrées</li> </ul>                  | <ul style="list-style-type: none"> <li>- Ne nécessite pas de valeurs chiffrées ordonnées</li> <li>- Insensible aux scores extrêmes</li> <li>- Est une valeur réelle</li> </ul> | <ul style="list-style-type: none"> <li>- Instable</li> <li>- Inutilisable en équation</li> <li>- Parfois peu représentatif</li> </ul> |
| <b>Médiane</b> | <ul style="list-style-type: none"> <li>- Certaines variables ordinales</li> <li>- Variables métriques avec forte asymétrie</li> </ul> | <ul style="list-style-type: none"> <li>- Insensible aux scores extrêmes</li> <li>- Est souvent une valeur réelle</li> </ul>  | <ul style="list-style-type: none"> <li>- Instable</li> <li>- Inutilisable en équation</li> </ul>                                      |
| <b>Moyenne</b> | <ul style="list-style-type: none"> <li>- La majorité des variables ordinales et métriques</li> </ul>                                  | <ul style="list-style-type: none"> <li>- Plus stable</li> <li>- Se prête bien aux équations</li> </ul>   | <ul style="list-style-type: none"> <li>- Sensible aux scores extrêmes</li> <li>- Est parfois une valeur impossible</li> </ul>         |

# Les mesures de variabilité

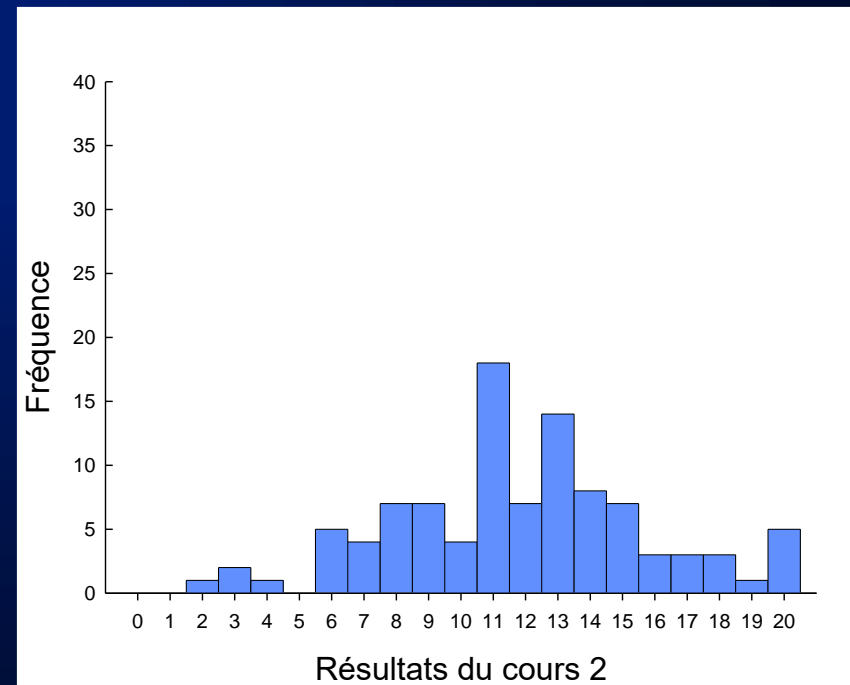
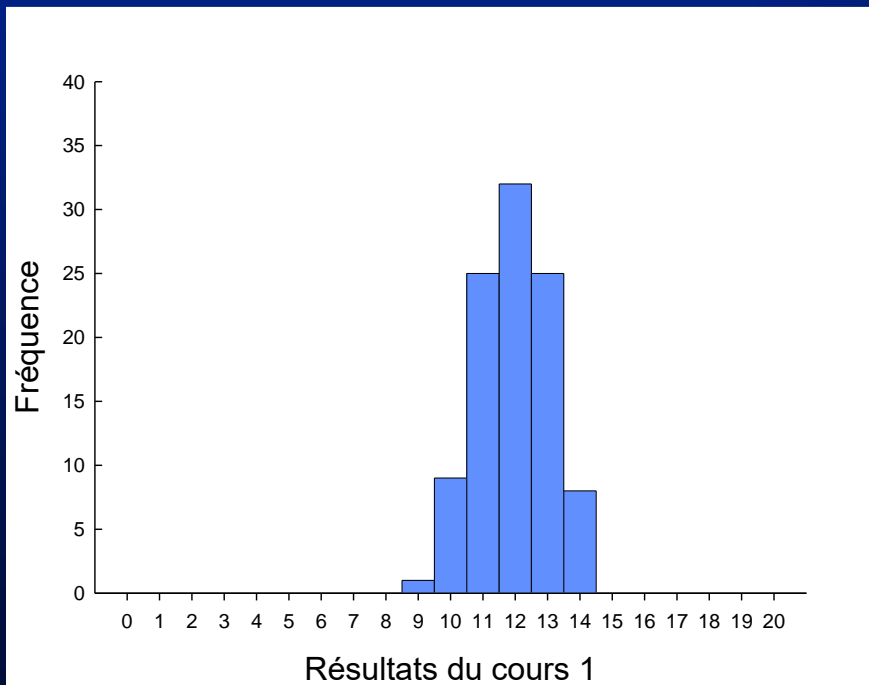
---

= la dispersion des données autour de la tendance centrale

1. L'étendue
2. L'écart moyen
3. L'écart absolu moyen
4. La variance
5. L'écart-type

# Illustration de la variabilité

## Résultats aux examens de deux cours



# L'étendue

---

**La distance entre le score le plus élevé et le score le moins élevé**

**Etendue = Plus grande obs. – Plus petite obs.**

**Mesure facile à calculer mais peu utilisée en raison de son inconvénient majeur**

# L'étendue (exemples)

## Exemple 1

$X$  = nombre d'erreurs commises lors d'un test de mémoire

$X = 3, 5, 5, 6, 7, 8, 9, 10, 12, 13, 15, 15$

Etendue =  $15 - 3 = 12$



# L'étendue (exemples)

## Exemple 2

$X$  = nombre d'erreurs commises lors d'un test de mémoire

$X = 3, 5, 5, 6, 7, 8, 9, 10, 12, 13, 15, 27$

Etendue =  $27 - 3 = 24$

# L'écart moyen

---

**Calculer les écarts par rapport à la moyenne**

**Faire la moyenne des écarts**

# L'écart moyen (exemple)

$X = 7, 12, 5, 9, 8, 5, 15, 13, 3, 15, 6, 10$

Moyenne = 9

Ecart = -2, 3, -4, 0, -1, -4, 6, 4, -6, 6, -3, 1

La moyenne de ces écarts est :

$$\frac{-2 + 3 - 4 + 0 - 1 - 4 + 6 + 4 - 6 + 6 - 3 + 1}{12} = \frac{0}{12} = 0$$

# L'écart moyen

---

**Calculer les écarts par rapport à la moyenne**

**Faire la moyenne des écarts**

**La moyenne des écarts est toujours zéro**

**L'écart moyen n'est pas  
un estimateur  
de la variabilité**

# L'écart absolu moyen (EAM)

---

**Calculer les écarts par rapport à la moyenne**

**Prendre les valeurs absolues de ces écarts**

**Faire la moyenne des écarts absolu**

# L'écart absolu moyen

---

$$EAM = \frac{\sum_{i=1}^N |\bar{X} - X_i|}{N}$$

# L'écart absolu moyen (exemple)

$X = 7, 12, 5, 9, 8, 5, 15, 13, 3, 15, 6, 10$

Moyenne = 9

Ecart = -2, 3, -4, 0, -1, -4, 6, 4, -6, 6, -3, 1

Ecart absolu = 2, 3, 4, 0, 1, 4, 6, 4, 6, 6, 3, 1

La moyenne de ces écarts est :

$$\frac{2 + 3 + 4 + 0 + 1 + 4 + 6 + 4 + 6 + 6 + 3 + 1}{12} = \frac{40}{12} = 3,33$$



# La variance

---

**Calculer les écarts par rapport à la moyenne**

**Mettre les écarts au carré**

**Faire la moyenne des écarts au carré**

**(= écart quadratique moyen)**

# La variance d'une population

---

$$\sigma^2 = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N}$$

# La variance d'une population

---

Formule rapide

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

# La variance d'un échantillon

---

$$S^2 = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N - 1}$$

**Estimateur de la variance**

# La variance d'un échantillon

Formule rapide

$$S^2 = \frac{\sum X_i^2 - \frac{(\sum X_i)^2}{N}}{N - 1}$$

# La variance d'un échantillon

Pour un échantillon, la formule dépend de l'objectif:

- Si estimation de la variance de population  
(statistiques inférentielles):  $n - 1$   
= estimateur de la variance
- Si simple mesure de variance de l'échantillon:  $n$

# Calcul de la variance de avec formule complexe

**$X = 7, 12, 5, 9, 8, 5, 15, 13, 3, 15, 6, 10$**

**Moyenne = 9**

**Ecart = -2, 3, -4, 0, -1, -4, 6, 4, -6, 6, -3, 1**

**Ecart au carré = 4, 9, 16, 0, 1, 16, 36, 16, 36, 36, 9, 1**

$$\frac{4 + 9 + 16 + 0 + 1 + 16 + 36 + 16 + 36 + 36 + 9 + 1}{12 - 1} = \frac{180}{11} = 16,36$$

# Calcul de la variance de avec formule rapide

$$X = 7, 12, 5, 9, 8, 5, 15, 13, 3, 15, 6, 10$$

$$\Sigma X = 7 + 12 + 5 + 9 + 8 + 5 + 15 + 13 + 3 + 15 + 6 + 10 = 108$$

$$\Sigma X^2 = 7^2 + 12^2 + 5^2 + 9^2 + 8^2 + 5^2 + 15^2 + 13^2 + 3^2 + 15^2 + 6^2 + 10^2 = 1152$$

$$s^2 = \frac{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}{n - 1} = \frac{1152 - \frac{108^2}{12}}{12 - 1} = \frac{180}{11} = 16,36$$



# Calculer la variance à partir d'un tableau de fréquences

$$S^2 = \frac{\sum_{i=1}^N f_i (X_i - \overline{X})^2}{N - 1}$$

# Calculer la variance à partir d'un tableau de fréquences

---

$$S^2 = \frac{\sum f_i X_i^2 - \frac{(\sum f_i X_i)^2}{N}}{N - 1}$$

# L'écart-type

## La racine carrée de la variance

Ecart-type de la population

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^N (X_i - \mu)^2}{N}}$$

Ecart-type d'un échantillon

$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N - 1}}$$

# La variabilité : résumé

---

**L'écart moyen: toujours égal à 0**

**L'écart absolu moyen: problème pour les équations**

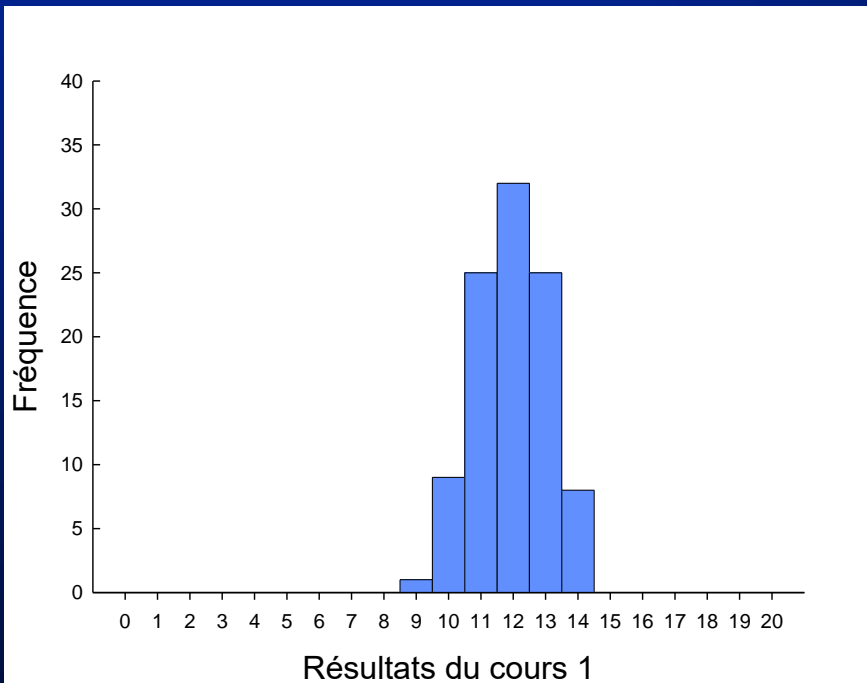
**Stratégie utilisée:**

**Ecart quadratique moyen (variance)**

**Puis racine carré = Ecart-type**

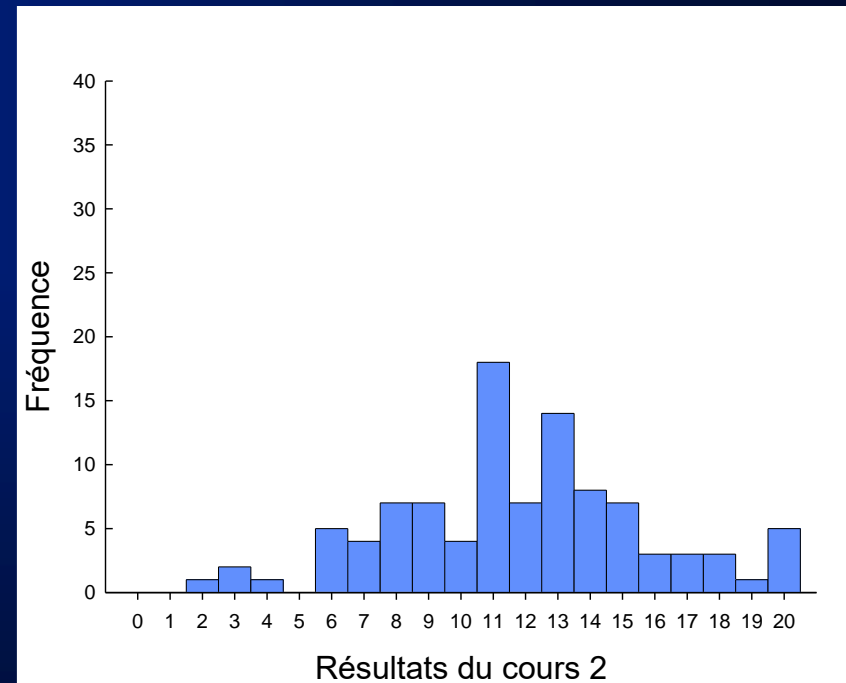
# Signification de l'écart-type

## Résultats aux examens de deux cours



$$\overline{X} = 12$$

$$S = 1,13$$



$$\overline{X} = 12$$

$$S = 3,92$$

# Rappel de notation

$\mu$  = la moyenne de la population

$\sigma^2$  = la variance de la population

$\sigma$  = l'écart-type de la population

$\overline{X}$  = la moyenne d'un échantillon

$S^2$  = la variance d'un échantillon

$S$  = l'écart-type d'un échantillon

$N$  = le nombre d'observations

# Rapporter les statistiques descriptives univariées dans un rapport écrit

---

Mesure de tendance centrale  $\pm$  variabilité

Exemple :

« Le temps moyen de résolution des problèmes pour l'ensemble des sujets ( $n = 50$ ) était de 590,52  $\pm$  104,77 (moyenne  $\pm$  écart-type) ».