DM / Activité de Mathématiques : Statistiques & Robustesse

Sur une copie double ou plusieurs rédigez en **bleu**, corrigez en **rouge**, et préparez vos questions pour le débrief en **vert**.

I. Vers la notation ∑ (Somme)

La lettre grecque Σ (Sigma majuscule) est utilisée pour condenser l'écriture d'une addition.

1. Série brute : Températures relevées à 8h (effectif total : N=12)

Voici les relevés : 8; 10; 9; 11; 13; 12; 10; 11; 12; 10; 14; 11.

- **Calculer** la moyenne $\bar{x}$. Remarque : pas besoin d'ordonner les valeurs
- **Notation :** On note x_i la $i^{ème}$valeur du relevé.
  Par exemple $x_1 = 8$, $x_4 = 11$ et $x_N = 11$
  On écrit $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{8+10+9+11+13+12+10+11+12+10+14+11}{N}$
- **Déterminer** Min, Q_1, Med, Q_3, Max.

2. Tableau d'effectifs : Nombre de livres lus par mois

Nombre de livres (x_i)	**0**	**1**	**2**	**3**	**4**	**5**	**6**
Effectif (n_i)	**2**	**5**	**8**	**12**	**7**	**4**	**2**
ECC							

- **Calculer** l'effectif total $N = \sum_{i=1}^{7} n_i$.
- **Notation :** La moyenne pondérée s'écrit :
  $\bar{x} = \frac{2\times0+5\times1+8\times2+12\times3+7\times4+4\times5+2\times6}{2+5+8+12+7+4+2} = \frac{\sum_{i=1}^{7} n_i x_i}{\sum_{i=1}^{7} n_i}$. **Effectuez** le calcul.
- **Déterminer** Q_1,Med,Q_3.
- Dessiner la boite à moustache correspondante.

3. Groupement en classes : Temps de trajet (min)

Temps $[x_i; x_{i+1}[$	**[0;10[**	**[10;20[**	**[20;30[**	**[30;40[**	**[40;50[**	**[50;60[**
Effectif (n_i)	4	14	22	18	9	3
Centre de classe (c_i)	5					
Fréquence (%)	5,7	20				
FCC (%)		25,7				

FCC correspond aux Fréquences Cumulées Croissantes. Par exemple dans la colonne [10 ;20[la FCC correspond à la réponse à la question quelle proportion de la série a un temps de trajet inférieur à 20. Pour obtenir cette quantité on va ajouter les Fréquences des colonnes [0 ;10[et [10 ;20[: $\approx 25{,}7\%$

- **Calculer** la moyenne en utilisant la notation : $\frac{1}{N}\sum_{i=1}^{N} n_i c_i$.
- **Polygone des FCC :** Construisez le polygone sur papier millimétré.
  - On place les points (BorneSup;FCC).
    Par exemple pour la première colonne on aura le point (10 ;5,7)
  - On rajoute le point (0;0). (en effet 0% de la population a un trajet inférieur à 0 minutes)
  - On relie les points obtenus pour obtenir une ligne brisée croissante.
- **Estimation graphique :** Déterminez Q_1,Med,Q_3 avec la méthode suivante:
  - Pour Q_1 on trace en pointillés une horizontale de hauteur 25%,
  - elle va couper le polygone en un point
  - une verticale en pointillés partant de ce point permet de lire une estimation de Q_1 sur l'axe des temps de trajet.
  - Même méthode pour Med (avec une verticale à 50%) et Q_3 (avec une verticale à 75%).

II. Analyse de la Dispersion & Robustesse

On considère une entreprise de 30 salariés. Leurs salaires mensuels (en €) sont les suivants :

- **29 salariés** gagnent tous entre 1 600 € et 2 400 €, avec une moyenne de **2 000 €**, une médiane de **1 950 €** et un écart interquartile de **300 €**.
- **Le 30ème salarié** est le PDG.

Cas A : Le PDG gagne 4 000 €. **Cas B : Le PDG gagne 40 000 €.**

1. **Calculs de comparaison :**
   - Calculez la nouvelle moyenne x¯ pour le Cas A, puis pour le Cas B.
   - Que deviennent la médiane et l'écart interquartile ($Q_3 - Q_1$) dans le Cas B par rapport au Cas A ? (Réfléchissez au rang des valeurs).
2. **L'Écart-type (σ) :** L'écart-type mesure la distance moyenne des salaires par rapport à la moyenne $\bar{x}$.
   - Dans quel cas (A ou B) l'écart-type sera-t-il le plus élevé ? Pourquoi ?
   - **Formule (pour info) :** $\sigma = \sqrt{\frac{1}{N}\sum ni(x_i - \bar{x})^2}$
3. **Synthèse sur la robustesse :**
   - Lequel de la **moyenne** ou de la **médiane** décrit le mieux le salaire "typique" d'un employé dans le Cas B ?
   - Pourquoi dit-on qu'un indicateur est "robuste" ? Citez les indicateurs robustes parmi ceux étudiés.

Corrigé de l'Activité : Statistiques

I. Vers la notation Σ

1. Série brute : Températures

- **Moyenne $\bar{x}$:** $\bar{x} = \frac{8+10+9+11+13+12+10+11+12+10+14+11}{12} = \frac{131}{12} \approx 10{,}92$ °C.
- **Notation ∑ :** Elle signifie que l'on additionne toutes les valeurs x de la première ($i = 1$) jusqu'à la douzième ($i = 12$).
- **Indicateurs (Série ordonnée : 8;9;10;10;10;11;11;11;12;12;13;14) :**
  - Min=8
  - $Q_1 : \frac{N}{4} = 3$. La 3ème valeur est **10**.
  - Med : $\frac{N+1}{2} = \frac{13}{2} = 6{,}5$. Entre la 6ème et 7ème valeur : $\frac{11+11}{2} = \mathbf{11}$.
  - $Q_3 : \frac{3N}{4} = 9$. La 9ème valeur est **12**.
  - Max=14

2. Tableau d'effectifs : Livres lus

Xi	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
ni	2	5	8	12	7	4	2	**40**
ECC	2	7	15	27	34	38	40	/

- **Moyenne :** $\bar{x} = \frac{0\times2+1\times5+2\times8+3\times12+4\times7+5\times4+6\times2}{40} = \frac{117}{40} = 2{,}925$ livres.
- **Quartiles et Médiane (N=40) :**
  - Q_1 : 40×0,25=10. La 10ème valeur est dans la colonne "2 livres" (car l'ECC passe de 7 à 15). Q1=2.
  - Med : $\frac{N+1}{2} = 20{,}5$. Les valeurs de rang 20 et 21 sont dans la colonne "3 livres" (ECC passe de 15 à 27). $Med = \frac{3+3}{2} = 3$.
  - Q3 : $\frac{3N}{4}$ =40×0,75=30. La 30ème valeur est dans la colonne "4 livres" (ECC passe de 27 à 34). Q_3=4.

3. Groupement en classes : Temps de trajet

Classes	**[0;10[**	**[10;20[**	**[20;30[**	**[30;40[**	**[40;50[**	**[50;60[**	**Total**
n_i	4	14	22	18	9	3	**70**
c_i	5	15	25	35	45	55	/
Freq (%)	5,7	20	31,4	25,7	12,9	4,3	**100%**
FCC (%)	5,7	25,7	57,1	82,8	95,7	100	/

- Pour les fréquences(%) : $\frac{effectif}{effectif\ total} 100 = \frac{effectif}{70} 100$
- Pour les FCC, fréquences cumulées on ajoute les fréquences comme on aurait pu le faire avec les effectifs cumulés.
- **Moyenne :** $x = \frac{5\times4+15\times14+25\times22+35\times18+45\times9+55\times3}{70} = \frac{1980}{70} \approx 28{,}3$min.
- **Polygone (Lecture graphique attendue) :**
  - Q_1 (à 25%) : environ **19 min**.
  - Med (à 50%) : environ **28 min**.
  - Q_3 (à 75%) : environ **37 min**.

II. Dispersion et Robustesse

1. Calculs des salaires :

- Somme des 29 salaires "de base" : 29×2000=58000 €.
- **Cas A (PDG à 4 000 €) :** $\bar{x}_A = \frac{58000+4000}{30} \approx 2066{,}67$ €.
- **Cas B (PDG à 40 000 €) :** $\bar{x}_B = \frac{58000+40000}{30} \approx 3266{,}67$ €.

2. Médiane et Écart Interquartile :

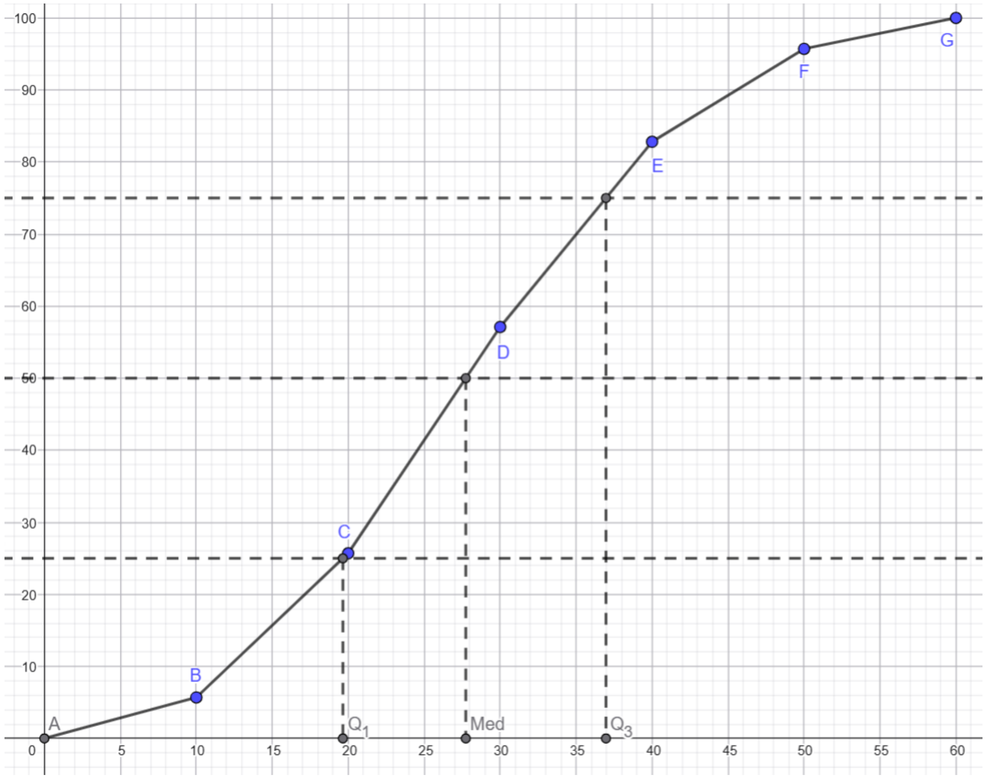
- L'effectif total est 30. La médiane se situe entre la 15ème et 16ème valeur. Comme les 29 premiers salaires sont compris entre 1 600 et 2 400 €, la 15ème et la 16ème valeur restent quasiment inchangées que le PDG gagne 4 000 ou 40 000 €.
- Q_1 (8ème valeur) et Q_3 (23ème valeur) restent également dans la tranche des 29 salariés.
- **Conclusion :** La médiane et l'écart interquartile sont identiques (ou presque) dans les deux cas.

3. Écart-type :

- L'écart-type sera bien plus élevé dans le **Cas B**. L'écart-type mesure la distance à la moyenne. Comme la moyenne a bondi à 3 266 € et qu'une valeur (40 000 €) est extrêmement loin de ce centre, l'écart-type augmente drastiquement.

4. Synthèse sur la robustesse :

- Le salaire "typique" est mieux représenté par la **médiane** (environ 1 950 €). La moyenne (3 266 €) est trompeuse car personne ne gagne ce montant (29 sont bien en dessous, 1 est très au-dessus).
- **Indicateurs robustes :** Médiane, Q_1, Q_3, Écart interquartile.
- **Indicateurs sensibles :** Moyenne, Étendue, Écart-type.



On peut lire $Q_1 \approx 19{,}7$, $Med \approx 27{,}7$ et $Q_3 \approx 37$