

DS d'entraînement :
Statistiques et Analyse de Données

Durée : 1h15 | **Calculatrice :** Autorisée (NumWorks conseillée)

Exercice 1 : Maîtrise de l'outil et des notations (5 points)

Un laboratoire analyse le taux de glucose (en g/L) de 15 patients à jeun. Série brute :
0,82;0,95;0,88;1,12;0,74;0,91;0,98;1,05;0,85;0,92;0,89;0,96;1,21;0,84;0,90.

- Calculs théoriques :** En utilisant la notation \sum , écrivez la formule de la moyenne \bar{x} puis calculez-la (arrondir à 10^{-2}).
- Paramètres de position :** Déterminez à la main Min, Q_1, Med, Q_3, Max .
- NumWorks :**
 - Saisissez la série dans l'application **Statistiques**.
 - Donnez la valeur de l'écart-type σ (arrondir à 10^{-3}).
 - Le laboratoire considère qu'un patient est "hors norme" si son taux est en dehors de l'intervalle $[\bar{x} - 2\sigma; \bar{x} + 2\sigma]$. Combien de patients sont hors norme ici ?

Exercice 2 : Comparaison de performances (6 points)

Deux entreprises, **TechLog** et **EcoFret**, comparent le temps de livraison (en jours) de leurs colis.

• **TechLog :**

Temps (j)	1	2	3	4	5	6
Effectif	12	25	48	32	18	5

• **EcoFret :**

$N=140$; $\bar{x}=3,2$ j ; $Q_1=2$ j ; $Med=3,5$ j ; $Q_3=4$ j ; $\sigma=0,8$.

- Calculez la moyenne et l'écart interquartile de la série de TechLog.
- Comparaison :**

- Quelle entreprise a le temps de livraison moyen le plus court?
- Laquelle semble la plus "régulière" (prévisible) dans ses livraisons ? Justifiez avec les indicateurs de dispersion.

- Interprétation :** Le directeur d'EcoFret affirme : "Au moins 75% de nos clients sont livrés en 4 jours ou moins". A-t-il raison ? Justifiez.

Exercice 3 : Étude de production (Classes et Polygone) (9 points)

Une usine de composants électroniques teste la durée de vie (en milliers d'heures) d'un échantillon de 200 processeurs.

Durée [h[[0;4[[4;8[[8;12[[12;16[[16;20[[20;24[
Effectif	10	30	75	55	22	8
FCC (%)						

- Tableau :** Complétez la ligne des Fréquences Cumulées Croissantes.
- Moyenne :** Calculez la durée de vie moyenne \bar{x} (montrez le calcul avec les centres de classes).
- Graphique :** Construisez le polygone des FCC sur votre copie.
 - Échelle :* Abscisse 1carreau = 2 000h / Ordonnée 1cm = 10%.
- Exploitation :**
 - Déterminez graphiquement Q_1, Med et Q_3 .
 - On estime qu'un processeur est de "Haute Qualité" s'il dure plus de 18 000h. Utilisez le polygone pour estimer le pourcentage de processeurs de Haute Qualité.
- Robustesse (Question ouverte) :** Si l'on découvrait que les 8 processeurs de la dernière classe [20;24[durent en réalité 100 000 heures (valeurs extrêmes), quels indicateurs parmi la moyenne et la médiane seraient les plus impactés ? Pourquoi ?

Corrigé détaillé du DS de Simulation

Exercice 1 : Glucose (5 points)

1. Moyenne théorique :

$$\bar{x} = \frac{1}{15} \sum_{i=1}^{15} x_i = \frac{0,82+0,95+\dots+0,90}{15} = \frac{13,97}{15} \approx 0,93 \text{ g/L.}$$

2. Paramètres de position (Série ordonnée) :

0,74;0,82;0,84;0,85;0,88;0,89;0,90;**0,91**;0,92;0,95;0,96;0,98;1,05;1,12;1,21

- **Min** : 0,74 | **Max** : 1,21.
- **Médiane** : $\frac{N+1}{2} = 8 \rightarrow$ 8ème valeur : **0,91**.
- Q_1 : $\frac{N}{4} = 15 \times 0,25 = 3,75 \rightarrow$ 4ème valeur : **0,85**.
- Q_3 : $\frac{N}{4} = 15 \times 0,75 = 11,25 \rightarrow$ 12ème valeur : **0,98**.

3. NumWorks et Écart-type :

- On trouve $\sigma \approx 0,113$.
- **Intervalle** $[\bar{x} - 2\sigma; \bar{x} + 2\sigma]$:
 $[0,93 - 0,226; 0,93 + 0,226] = [0,704; 1,156]$.
- **Patients "hors norme"** : Seul le patient à 1,21 g/L est en dehors de cet intervalle. Il y a donc **1 patient** hors norme.

Exercice 2 : TechLog vs EcoFret (6 points)

1. Analyse de TechLog (N=140) :

- $\bar{x} = \frac{1 \times 12 + 2 \times 25 + 3 \times 48 + 4 \times 32 + 5 \times 18 + 6 \times 5}{140} = \frac{454}{140} \approx 3,24 \text{ jours.}$
- **Écart interquartile** : Q_1 (35e val) = 2 ; Q_3 (105e val) = 4. $Q_3 - Q_1 = 2$ jours.

2. Comparaison :

- **Rapidité** : EcoFret ($\bar{x}=3,2$) est légèrement plus rapide que TechLog ($\bar{x}=3,24$).
- **Régularité** : On compare les écarts-types. Pour TechLog, la NumWorks donne $\sigma \approx 1,25$. EcoFret a un $\sigma = 0,8$. **EcoFret est plus régulière** car son écart-type est plus faible (les valeurs sont moins dispersées autour de la moyenne).

3. Interprétation :

- $Q_3 = 4$ jours signifie qu'au moins 75% des données sont inférieures ou égales à 4. Le directeur a donc **raison**.

Exercice 3 : Processeurs (9 points)

1. Tableau des FCC :

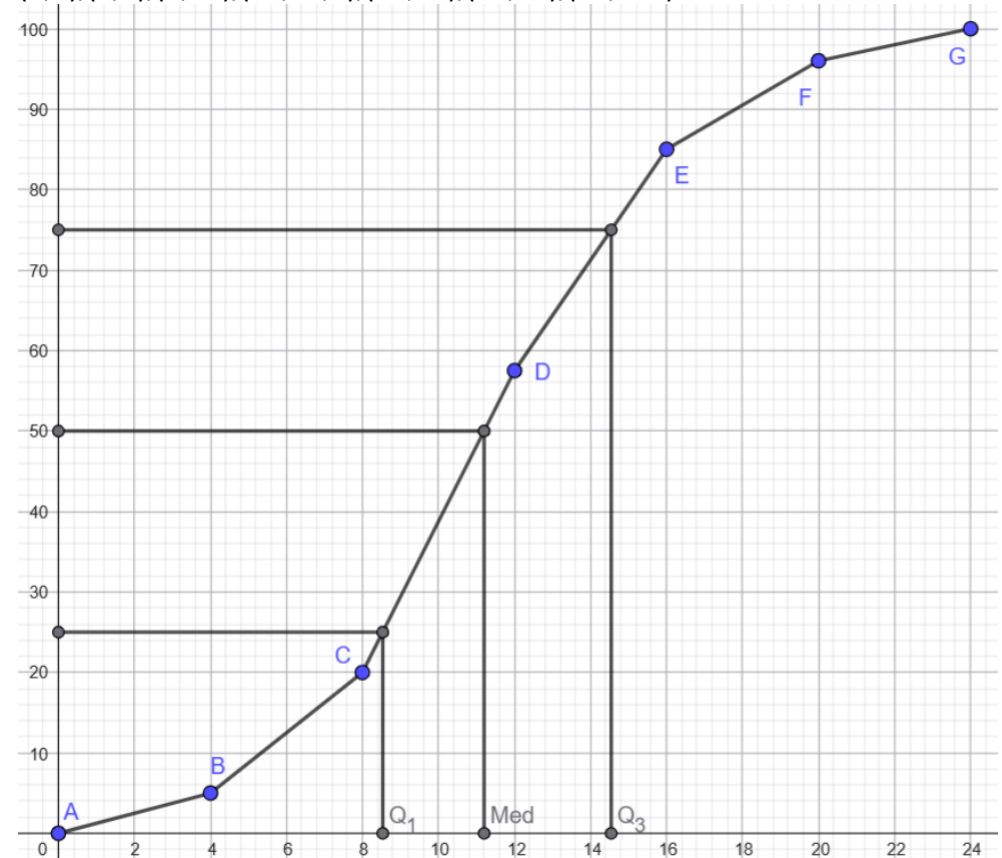
Durée	[0;4[[4;8[[8;12[[12;16[[16;20[[20;24[
Fréquence (%)	5	15	37,5	27,5	11	4
FCC (%)	5	20	57,5	85	96	100

Explication : $fréquence_{\%} = \frac{\text{effectif}}{\text{effectif total}} \times 100 = \frac{\text{effectif}}{200} \times 100 = \frac{\text{effectif}}{2}$

2. **Moyenne** : $\bar{x} = \frac{2 \times 10 + 6 \times 30 + 10 \times 75 + 14 \times 55 + 18 \times 22 + 22 \times 8}{200} = \frac{2312}{200} = 11,56$ (en milliers d'heures).

3. Polygone des FCC (Points à placer) :

(0;0),(4;5),(8;20),(12;57,5),(16;85),(20;96),(24;100).



4. Exploitation graphique :

- Q_1 (à 25%) : env. **8,5** (8 500 h).

- **Med (à 50%)** : env. **11,2** (11 200 h).
- Q_3 (**à 75%**) : env. **14,5** (14 500 h).
- **Haute Qualité (> 18 000 h)** : Sur le polygone, à $x=18$, on lit environ 90,5% de FCC. Cela signifie que 90,5% des processeurs durent *moins* de 18 000h. Le pourcentage de "Haute Qualité" est donc $100-90,5=9,5\%$.

5. Robustesse :

- La **moyenne** serait fortement impactée (elle augmenterait massivement car elle tient compte de la valeur numérique de chaque donnée).
- La **médiane** resterait identique (ou presque) car elle ne dépend que du rang des valeurs. Tant qu'il y a 8 valeurs dans la tranche supérieure, leur valeur exacte (24 000 ou 100 000) ne change pas la position de la valeur centrale.