

Statistiques descriptives (1,5h)

Éléments de correction du partiel (déc 2025)

Exercice 1 Dans une entreprise, les factures établies dans le courant d'un certain mois ont été classées d'après leur montant. Les résultats de cette opération ont été consignés dans le tableau suivant :

Montant (en €)	[90; 110[	[110; 130[	[130; 150[	[150; 170[	[170; 190[	[190; 210[
Nbr de factures	8	16	12	18	6	3

1. Que vaut l'étendue de la série ?

L'étendue vaut 120 €

2. Quel est le pourcentage de factures inférieures à 150€ ?

L'effectif total est 63 et l'effectif des factures inférieures à 150 € vaut 36.

Le pourcentage cherché vaut donc $\frac{36}{63} \times 100 \approx 57,14\%$

3. Calculer le montant de la facture moyenne dans cette entreprise.

Notons X la série, par la formule de la moyenne, on a :

$$\bar{X} = \frac{8 \times 100 + 16 \times 120 + \dots + 3 \times 200}{63} \approx 142,22 \text{ €}.$$

4. Calculer l'écart moyen de la série.

En notant E_m l'écart moyen, on a :

$$E_m = \frac{8 \times |100 - 142,22| + 16 \times |120 - 142,22| + \dots + 3 \times |200 - 142,22|}{63} \approx 22,86 \text{ €}.$$

5. Calculer la variance et préciser son interprétation ? Donner la valeur de l'écart type.

$$Var(X) = \frac{8 \times (100 - 142,22)^2 + 16 \times (120 - 142,22)^2 + \dots + 3 \times (200 - 142,22)^2}{63} \approx 737,92.$$

Puis, $\sigma(X) = \sqrt{Var(X)} \approx 27,16 \text{ €}$ qui est un indicateur de dispersion de la série autour de sa moyenne.

6. Tracer l'histogramme des fréquences cumulées croissantes.

7. Retrouver le résultat de la question 2 à l'aide de votre graphe.

Le résultat de la question 2 se lit comme l'image de 150 €.

8. Calculer la médiane. Retrouver graphiquement sa valeur, en expliquant votre démarche.

La médiane correspond à la 32^{ième} valeur, qui appartient donc à la classe [130; 150[. La valeur 130 € correspond à la 24^{ième}, donc $Me = 130 + (32 - 24) \times \frac{20}{12} \approx 143,33 \text{ €}$

9. Donner la définition des quartiles et les calculer.

Voir cours pour les définitions.

$$Q_1 = \text{la } 15,75^{\text{ième}} \text{ valeur} = 110 + (15,75 - 8) \frac{20}{16} \approx 119,69 \text{ €}$$

$$Q_3 = \text{la } 47,25^{\text{ième}} \text{ valeur} = 150 + (47,25 - 36) \frac{20}{18} \approx 162,5 \text{ €}$$

10. Le patron décide une remise de 5% sur toutes les factures. Calculez la nouvelle moyenne, ainsi que la nouvelle variance.

Soit X' la série diminuée de 5%. Par la linéarité de la moyenne et la pseudo linéarité de la variance, on a :

$$\bar{X}' = 0,95\bar{X} \approx 135,1 \text{ €} \text{ et } Var(X') = 0,95^2 Var(X) \approx 665,97 \text{ €}^2.$$

Exercice 2 Les questions sont indépendantes.

1. Donner un exemple de série X telle que $\bar{X} > Q_1$? Puis un exemple, où $\bar{X} < Q_1$. Détaillez.

La série $S_1 = \{1; 2; 3; 4\}$ vérifie $Q_1(S_1) = 1 < \bar{S}_1 = 2,5$.

La série $S_2 = \{-35; 2; 3; 4; 5; 6; 7; 8\}$ vérifie $\bar{S}_2 = 0 < Q_1(S_2) = 2$.

2. Le tableau suivant concerne la distance moyenne (en km) parcourue entre domicile et travail, par des employés d'entreprises situées dans 96 départements de France.

Min	Q_1	M_e	Q_3	Max
8,5	13,7	14,9	16,1	23,3

Dans combien de départements la distance moyenne parcourue est elle inférieure ou égale à 13,7 km ?

$Q_1 = 13,7 \text{ km}$ donc par déf du premier quartile, 25% des départements la distance parcourue est inférieure à 13,7 km.

ie : $0,25 \times 96 = 24$ départements

3. Proposez (en justifiant !) dans chaque cas une série (simple) ayant :

- (a) une moyenne de 10,

La série $X_1 = \{9; 11\}$ convient par exemple.

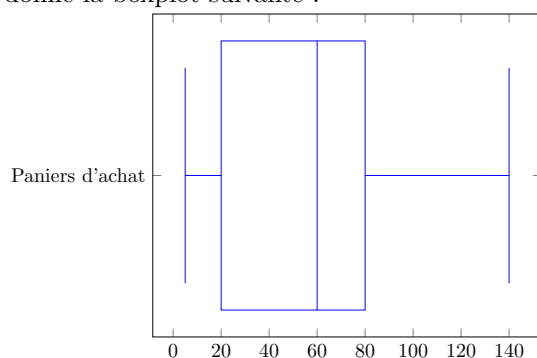
- (b) une variance de 9,

La série $X_2 = \{-3; 3\}$ a pour moyenne 0 et les écarts à la moyenne sont de ± 3 . Donc $Var(X_2) = 9$.

- (c) une moyenne de 10 et une variance de 9.

La série $X_3 = \{7; 13\}$ satisfait les deux conditions.

4. Une étude statistique des paniers d'achat (en €) d'un supermarché a donné la boxplot suivante :



Proposer 3 intervalles différents dans lesquels se trouve 50% des paniers d'achat.

Notons X la série. Les 3 intervalles suivants conviennent :

$$[\min(X); Me(X)] = [5; 60]$$

$$[Q_1(X); Q_3(X)] = [20; 80]$$

$$[Me(X); \max(X)] = [60; 140]$$

5. Les temps de retard (en *min*) d'un employé sont relevés sur 10 jours et consignés dans une série notée T . Malheureusement, les 2 données extrêmes ont été effacées et les 8 autres données sont 3, 16, 11, 7, 5, 10, 10, 9. Par ailleurs, on sait que $\bar{T} = 9 \text{ min}$ et que l'étendue vaut 17 *min*. Pouvez-vous retrouver les 2 données manquantes ?

Notons x et y respectivement le temps de retard le plus petit et le plus grand. L'énoncé se traduit par :

$$\begin{cases} y - x &= 17 \\ \frac{1}{10}(x + 3 + 16 + 11 + 7 + 5 + 10 + 10 + 9 + y) &= 9 \end{cases}$$

La résolution donne $x = 1$ et $y = 18$.

6. Un site internet permet à ses utilisateurs de donner leur avis sur des films en leur attribuant entre 0 et 5 étoiles selon leur niveau de satisfaction. Pour un certain film, les résultats sont les suivants :

Nombre d'étoiles	0	1	2	3	4	5
Nombre d'avis	59	178	505	1039	981	238

Calculer la variance et les quartiles.

Notons Z cette série, on a $\bar{Z} \approx 3,14$ puis $Var(Z) \approx 1,21$. Pour les quartiles, on peut s'aider des effectifs cumulés croissants :

Nombre d'étoiles	0	1	2	3	4	5
Nombre d'avis	59	178	505	1039	981	238
Nombre d'avis cumulés croissants	59	237	742	1781	2762	3000

L'effectif total est de 3000, donc le $Q_1 = 750^{\text{ième}} \text{ valeur} = 3$. De même, $Q_2 = 1500^{\text{ième}} \text{ valeur} = 3$ et $Q_3 = 2250^{\text{ième}} \text{ valeur} = 4$.

7. *Bonus* : Ecrire la commande qui permet de rentrer la série suivante $X = \{1; 2; 3\}$ en Rstudio, puis celle qui permet de calculer la médiane de cette série.

On tape : `>X=c(1,2,3)` pour définir la série X , puis `>median(X)` pour calculer la médiane.

