

STATISTIQUES

I Rappels : moyenne - médiane - quartiles

Exercice 01 (voir [réponses et correction](#))

Le service de contrôle d'une usine de meubles mesure la longueur en cm d'un élément.
On a obtenu les mesures suivantes :

| | | | | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 101 | 99 | 102 | 101 | 99 | 101 | 100 | 100 | 101 | 101 | 99 | 100 | 99 | 100 | 101 | 100 | 101 | 101 | 98 | 101 |
| 100 | 100 | 101 | 101 | 98 | 100 | 100 | 102 | 101 | 100 | 99 | 97 | 102 | 101 | 100 | 100 | 99 | 101 | 100 | 100 |
| 99 | 99 | 99 | 99 | 100 | 101 | 101 | 100 | 100 | 100 | 101 | 99 | 100 | 100 | 101 | 101 | 102 | 100 | 101 | 99 |
| 100 | 100 | 101 | 101 | 99 | 101 | 99 | 99 | 102 | 98 | 100 | 100 | 101 | 101 | 100 | 100 | 99 | 100 | 101 | 101 |
| 99 | 100 | 100 | 101 | 102 | 101 | 100 | 100 | 100 | 100 | 99 | 100 | 99 | 102 | 101 | 99 | 101 | 100 | 99 | 100 |

Ces mesures constituent une série statistique.
La longueur de l'élément est le caractère de la série. (Il s'agit d'un caractère quantitatif puisqu'il est mesuré par un nombre)
L'ensemble des mesures effectuées est appelé la "population".
Organiser ces données dans le tableau suivant :

| | | | | | | |
|------------------------------|----|----|----|-----|-----|-----|
| Valeurs | 97 | 98 | 99 | 100 | 101 | 102 |
| Effectifs | | | | | | |
| Effectifs cumulés croissants | | | | | | |

Calculer la moyenne m et la médiane M de cette série.
Déterminer le 1^{er} quartile Q_1 et le 3^{ème} quartile Q_3 et en déduire l'écart interquartile.
Que trouve-t-on pour ces valeurs en utilisant les fonctions statistiques d'une calculatrice ?

Définition

On considère la série statistique donnée par le tableau :

| | | | | | |
|----------|-------|-------|--|--|-------|
| Valeur | x_1 | x_2 | | | x_p |
| Effectif | n_1 | n_2 | | | n_p |

La **moyenne** de cette série, notée \bar{x} est le nombre réel :
$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_px_p}{n_1 + n_2 + \dots + n_p} = \frac{\sum_{i=1}^{i=p} n_i x_i}{\sum_{i=1}^{i=p} n_i}$$

Si on appelle f_i la fréquence de x_i , on a :
$$\bar{x} = f_1x_1 + f_2x_2 + \dots + f_px_p = \sum_{i=1}^{i=p} f_i x_i$$

Lorsque les données de la série sont rangées dans l'ordre croissant :

- La **médiane** de la série est la valeur partageant la population en deux groupes de même effectif.
- Le **premier quartile** de la série est la plus petite valeur Q_1 des termes de la série pour laquelle au moins un quart (25%) des données sont inférieures ou égales à Q_1 .
- Le **troisième quartile** de la série est la plus petite valeur Q_3 des termes de la série pour laquelle au moins trois quarts (75%) des données sont inférieures ou égales à Q_3 .
- L'**intervalle interquartile** est l'intervalle $[Q_1 ; Q_3]$.
- L'**écart interquartile** est l'amplitude de l'intervalle $[Q_1 ; Q_3]$, c'est-à-dire le nombre $Q_3 - Q_1$.

Remarque

Si la série comporte un nombre impair de termes la médiane correspond à une valeur de la série.
110 ; 110 ; 113 ; 113 ; 113 ; 115 ; **115** ; 116 ; 117 ; 120 ; 120 ; 121 ; 121. Cette série a pour médiane 115
Si la série comporte un nombre pair de termes on prend comme médiane la demi-somme de deux valeurs.
20 ; 20 ; 21 ; 21 ; 23 ; **23 ; 24** ; 24 ; 25 ; 26 ; 26 ; 30. Cette série a pour médiane 23,5

Exercice 02 (voir [réponses et correction](#))

Lors d'un test sur 215 élèves de première d'un lycée, on a obtenu la répartition suivante des notes (évaluation sur 10 points)

| | | | | | | | | | | | |
|-----------|---|---|---|---|----|----|----|----|----|----|----|
| Notes | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Effectifs | 3 | 5 | 8 | 7 | 33 | 42 | 37 | 51 | 12 | 10 | 7 |

1°) Compléter ci-dessous le tableau des fréquences et des fréquences cumulées croissantes (les résultats seront arrondis au millième près).

| | | | | | | | | | | | |
|---------------------------------|---|---|---|---|---|---|---|---|---|---|----|
| Notes | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Fréquences | | | | | | | | | | | |
| Fréquences cumulées croissantes | | | | | | | | | | | |

2°) Quel est le pourcentage des élèves ayant une note inférieure ou égale à 8 ?

3°) a) Calculer la moyenne m et la médiane M de cette série.

b) Déterminer le 1^{er} quartile Q_1 et le 3^{ème} quartile Q_3 . En déduire l'écart interquartile.

c) Que trouve-t-on pour ces valeurs en utilisant le mode statistique d'une calculatrice ?

4°) Représenter la série par un diagramme à barres.

Exercice 03 (voir [réponses et correction](#))

Une station-service a noté les achats de Gazole pour une journée donnée.

Les résultats, répartis en classes (intervalles), sont donnés dans le tableau ci-dessous.

| | | | | | | | | | | |
|---------------------------------|----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|------------|
| Volume en litres |]0 ; 10] |]10 ; 20] |]20 ; 30] |]30 ; 40] |]40 ; 50] |]50 ; 60] |]60 ; 70] |]70 ; 80] |]80 ; 90] |]90 ; 100] |
| Nombre de clients | 47 | 108 | 173 | 182 | 245 | 168 | 175 | 68 | 22 | 3 |
| Effectifs cumulés croissants | | | | | | | | | | |
| Fréquences cumulées croissantes | | | | | | | | | | |

1°) Calculer le volume moyen acheté par un client. (On utilisera pour calculer ce volume moyen le centre de chacune des classes et on arrondira le résultat au dixième)

2°) a) Compléter, dans le tableau, la ligne des effectifs cumulés croissants et la ligne des fréquences cumulées croissantes (à 10^{-3} près). Tracer la courbe des fréquences cumulées croissantes.

b) En utilisant le graphique de la question précédente, déterminer la valeur de la médiane, du 1^{er} quartile et du 3^{ème} quartile de cette série.

Exercice 04 (voir [réponses et correction](#))

On considère la série donnée par le nuage de points ci-contre représentant la durée de vie en années d'un modèle particulier de machine à laver noté modèle A.

1°) Calculer sa moyenne et déterminer sa médiane.

2°) En utilisant une calculatrice, donner son écart interquartile.

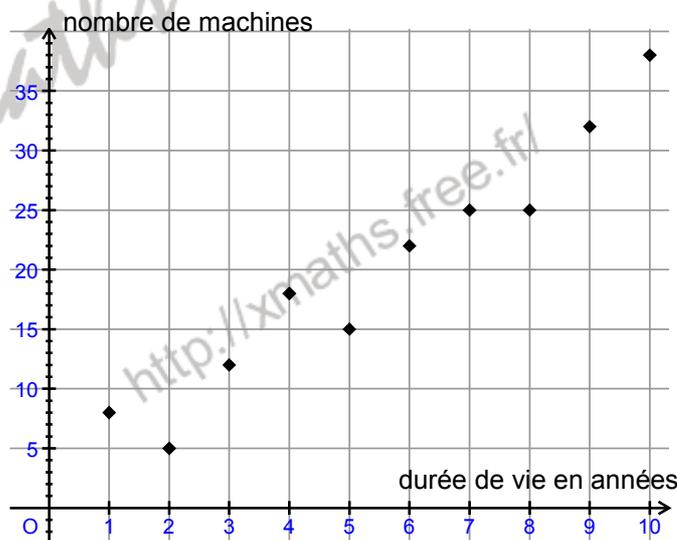
3°) Une série d'un autre modèle de machine à laver noté modèle B, a les paramètres suivants :

Moyenne : 6,84

Médiane : 7

Écart interquartile : 2

En comparant les paramètres des deux séries, quelles remarques pouvez-vous faire ?



II Variance - Écart-type

Définition

On considère une série statistique, de moyenne \bar{x} , donnée par le tableau suivant :

| | | | | | | | |
|------------------|-------|-------|-------|-----|-----|-----|-------|
| Valeur : x_i | x_1 | x_2 | x_3 | ... | ... | ... | x_p |
| Effectif : n_i | n_1 | n_2 | n_3 | ... | ... | ... | n_p |

La **variance** de cette série est :

$$V = \frac{n_1 \times (x_1 - \bar{x})^2 + n_2 \times (x_2 - \bar{x})^2 + n_3 \times (x_3 - \bar{x})^2 + \dots + n_p \times (x_p - \bar{x})^2}{n_1 + n_2 + n_3 + \dots + n_p} = \frac{\sum_{i=1}^{i=p} n_i \times (x_i - \bar{x})^2}{\sum_{i=1}^{i=p} n_i}$$

Si on appelle f_i la fréquence de x_i , on a :

$$V = f_1 \times (x_1 - \bar{x})^2 + f_2 \times (x_2 - \bar{x})^2 + f_3 \times (x_3 - \bar{x})^2 + \dots + f_p \times (x_p - \bar{x})^2 = \sum_{i=1}^{i=p} f_i \times (x_i - \bar{x})^2$$

L'**écart-type** de cette série est : $s = \sqrt{V}$

Remarques

- L'écart-type permet d'avoir une idée de la façon dont les valeurs de la série s'écartent par rapport à la moyenne. C'est une mesure de dispersion (comme l'écart interquartile).
Un écart-type faible correspond à une série concentrée autour de sa moyenne.
- Les calculs de moyenne, de variance et d'écart-type sont, pour des séries prenant un grand nombre de valeurs, des calculs compliqués. Mais ils sont facilement réalisés par les calculatrices utilisées en mode statistiques et les ordinateurs.

Exercice 09

 (voir [réponses et correction](#))

On considère la série

| | | | | | |
|------------------|-----|----|---|---|----|
| x_i | -10 | -5 | 0 | 5 | 10 |
| Effectif : n_i | 15 | 10 | 9 | 4 | 2 |

1°) Calculer sa moyenne \bar{x} .

2°) Compléter le tableau suivant et en déduire la variance V et l'écart-type σ de la série.

| | | | | | |
|-------------------------|-----|----|---|---|----|
| x_i | -10 | -5 | 0 | 5 | 10 |
| Effectif : n_i | 15 | 10 | 9 | 4 | 2 |
| $(x_i - \bar{x})$ | | | | | |
| $n_i (x_i - \bar{x})^2$ | | | | | |

3°) Retrouver les résultats des questions précédentes en utilisant le mode statistique d'une calculatrice.

Exercice 10

 (voir [réponses et correction](#))

On considère la série statistique suivante (masse en grammes des œufs de poule d'un élevage)

| | | | | | | | | | | | |
|------------------|----|----|----|-----|-----|-----|-----|----|----|----|----|
| Masse : x_i | 40 | 45 | 50 | 55 | 60 | 65 | 70 | 75 | 80 | 85 | 90 |
| Effectif : n_i | 16 | 20 | 75 | 141 | 270 | 210 | 165 | 63 | 21 | 12 | 7 |

1°) En utilisant une calculatrice en mode statistiques, déterminer pour cette série la moyenne \bar{x} et la médiane M, puis l'écart-type σ et l'écart interquartile e .

2°) On suppose qu'à la suite d'une erreur de transmission, l'une des données de la série n'a pas été correctement indiquée : un œuf d'une masse de 60g a été noté comme ayant une masse de 600g.

Quelle influence a cette erreur sur la moyenne, sur l'écart-type ?

Quelle influence a cette erreur sur la médiane, sur l'écart interquartile ?

III Diagramme en boîte

Définition

On considère une série, dont les valeurs sont **ordonnées** (rangées dans l'ordre croissant).

On appelle **premier décile** de la série la plus petite valeur D_1 des termes de la série pour laquelle au moins un dixième (10%) des données sont inférieures ou égales à D_1 .

On appelle **neuvième décile** de la série la plus petite valeur D_9 des termes de la série pour laquelle au moins neuf dixièmes (90%) des données sont inférieures ou égales à D_9 .

On appelle **intervalle interdécile** l'intervalle $[D_1 ; D_9]$.

On appelle **écart interdécile** l'amplitude de l'intervalle $[D_1 ; D_9]$, c'est-à-dire le nombre $D_9 - D_1$.

Construction d'un diagramme en boîte

Ce type de diagramme est aussi appelé diagramme de Tuckey, boîte à moustaches ou boîte à pattes. Il utilise la médiane, le 1^{er} et le 3^{ème} quartile, le 1^{er} et le 9^{ème} décile ou les valeurs extrêmes d'une série.

La construction ci-contre est faite pour une série caractérisée par :

médiane : 113

1^{er} quartile : 110

1^{er} décile : 108

3^{ème} quartile : 117

9^{ème} décile : 119

On choisit une graduation verticale permettant de représenter les différents paramètres de la série.

On pourra par exemple graduer entre 90 et 130.

Le "corps" du diagramme, c'est-à-dire la "boîte" est formée d'un rectangle ayant pour côté inférieur le 1^{er} quartile et pour côté supérieur le 3^{ème} quartile. À l'intérieur de ce rectangle on tracera un segment représentant la médiane.

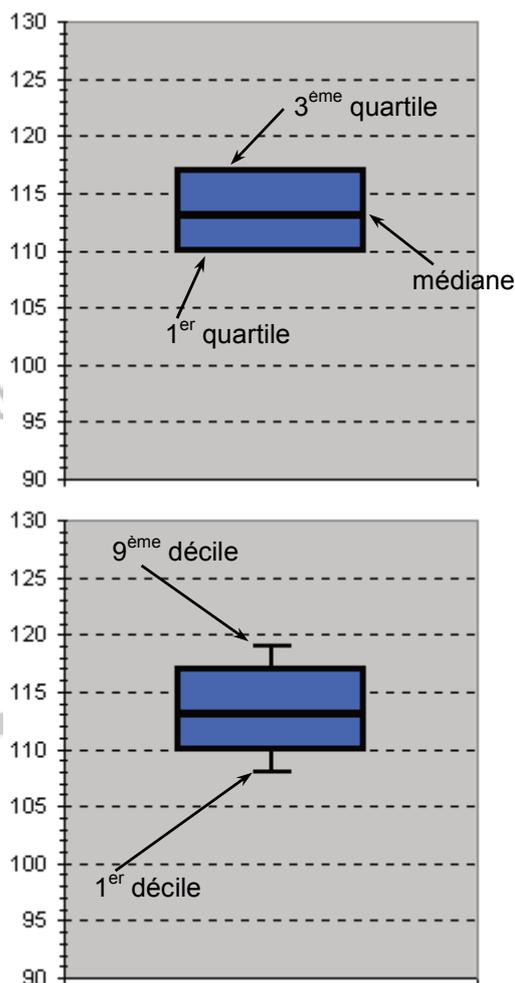
La largeur du rectangle n'est pas fixée, elle sera choisie de façon à obtenir un graphique "harmonieux".

Ce rectangle représente les données contenues dans l'intervalle interquartile.

On repère ensuite les hauteurs correspondant au 1^{er} et au 9^{ème} décile, et on trace deux "pattes" représentant les données contenues dans l'intervalle interdécile. (la largeur des pattes n'a pas d'importance).

Remarques

- On peut ensuite terminer le graphique, en faisant figurer par des points les données qui sont en dehors de l'intervalle interdécile.
- Si certaines données sont manifestement très éloignées, on ne les représentera pas, mais on pourra écrire leurs valeurs au-dessous du diagramme.
- Le graphique est parfois fait en dessinant des pattes correspondant non pas au 1^{er} et au 9^{ème} décile, mais aux valeurs extrêmes (ou au 1^{er} et au 99^{ème} centile).
- Une boîte et des "pattes" courtes indiquent que la série est assez concentrée autour de sa médiane. Au contraire une boîte et des "pattes" longues indiquent que la série est assez dispersée.
- Un des avantages de cette représentation, est qu'elle nécessite très peu de calculs.
- La représentation peut aussi se faire horizontalement, d'où l'appellation de "boîte à moustaches". La graduation se trouve alors sur l'axe horizontal.



Exercice 11 (voir [réponses et correction](#))

Construire le diagramme en boîte correspondant à la série initiale de l'exercice 10.

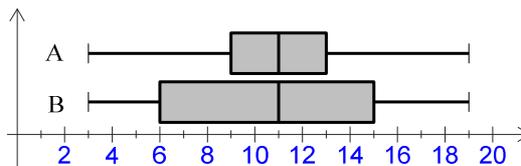
Exercice 12 (voir [réponses et correction](#))

On donne sur le dessin ci-contre les diagrammes en boîte de deux séries.

Ces séries sont les séries des notes de mathématiques de deux classes de Première S notées A et B.

Les "pattes" correspondent aux valeurs extrêmes.

Comparer ces deux séries. Que peut-on en conclure ?



Exercice 13 (voir [réponses et correction](#))

Une étude a été faite sur la longueur des épis d'une certaine variété de maïs.

Sur un échantillon de 210 épis, les résultats suivants ont été obtenus :

| | | | | | | | | | | | | | | |
|----------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| longueur en mm | 112 | 114 | 116 | 118 | 120 | 122 | 124 | 126 | 128 | 130 | 132 | 134 | 136 | 138 |
| nombre d'épis | 9 | 12 | 15 | 16 | 18 | 22 | 28 | 24 | 21 | 16 | 12 | 8 | 6 | 3 |

Les réponses aux questions doivent être soigneusement justifiées.

1°) Déterminer pour cette série, la médiane M, le 1^{er} quartile Q₁, le 3^{ème} quartile Q₃, le 1^{er} décile D₁ et le 9^{ème} décile D₉ de cette série.

2°) Représenter cette série par un diagramme en boîte.

Exercice 14 (voir [réponses et correction](#))

On considère la série ci-dessous correspondant aux résultats des participants lors d'une épreuve d'un concours.

| | | | | | | | | | | | | | | | | | | | | | |
|------------------|----|---|----|----|----|----|----|----|-----|-----|-----|-----|-----|----|----|----|----|----|----|----|----|
| Note sur 20 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| Nombre de copies | 12 | 7 | 10 | 22 | 43 | 35 | 48 | 84 | 141 | 174 | 123 | 146 | 121 | 55 | 76 | 28 | 12 | 5 | 1 | 2 | 2 |

1°) Faire le diagramme en boîte correspondant à cette série.

(Expliquer soigneusement l'obtention des différents éléments).

2°) En utilisant une calculatrice, donner la moyenne et l'écart-type de la série.

Exercice 15 (voir [réponses et correction](#))

Les diagrammes en boîte ci-contre représentent la répartition des salaires mensuels en euros dans 5 entreprises A, B, C, D et E.

Pour chaque série, les extrémités des pattes correspondent au premier et au neuvième décile, les valeurs maximales et minimales sont indiquées par des points.

Pour chacune des affirmations suivantes, indiquer **une** entreprise pour laquelle cette affirmation est vraie en justifiant votre réponse.

(1) La moitié au moins des employés ont un salaire qui ne dépasse pas 1500 €.

(2) L'échelle des salaires va de 1 à 10.

(3) 25% au moins des salariés gagnent moins de 1000 €

(4) Au moins trois salariés sur quatre ont un salaire supérieur à 2000 €

(5) Au moins 9 salariés sur 10 ont un salaire supérieur à 1400 €

(6) Le PDG (qui a le salaire le plus élevé) gagne quatre fois plus que la moitié au moins des salariés.

