

I. Vocabulaire des statistiques :

On appelle statistique, d'une part, un recueil de données concernant une population et d'autre part, les méthodes de traitement et d'interprétation de ces données.

1) Quelques définitions :

Une population est un ensemble sur lesquels vont porter les observations (ensembles d'objets, de personnes, d'entreprises, de machines, d'animaux...).

Chaque élément d'une population est appelé un **individu**.

Le nombre total d'individus d'une population s'appelle l'**effectif total de la population**.

Un échantillon est une partie de la population.

On appellera **caractère**, la propriété étudiée sur la population.

2) Présentation générale d'une série statistique :

Une série statistique se présentera généralement sous la forme d'un tableau, donnant pour chaque valeur du caractère, l'effectif correspondant.

| | | | | | | |
|------------------------------|-------|-------|-----|-----|-------|-------|
| Valeurs du caractère : x_i | x_1 | x_2 | ... | ... | x_p | TOTAL |
| effectifs : n_i | n_1 | n_2 | ... | ... | n_p | N |

L'effectif total de la série est la somme des effectifs de toutes les valeurs possibles du caractère étudié.

$$N = n_1 + n_2 + n_3 + \dots + n_p$$

La fréquence d'une valeur est égale au quotient de l'effectif de la valeur par l'effectif total de la population. La fréquence d'une valeur x_i est $f_i = \frac{n_i}{N}$

La fréquence s'exprime sous forme fractionnaire, décimale ou sous forme de pourcentage.

Exemple 1 Compléter le tableau ci-dessous, donnant pour chaque superficie, le nombre d'exploitations agricoles recensées.

| Superficie en ha | Nombre d'exploitations n_i | Fréquence f_i | Effectifs cumulés croissants |
|------------------|------------------------------|-----------------|------------------------------|
| 3 | 7 | 0,14 | 7 |
| 5 | 10 | 0,20 | 7 + 10 = 17 |
| 9 | 13 | 0,26 | 17 + 13 = 30 |
| 12 | 9 | 0,18 | 30 + 9 = 39 |
| 15 | 1 | 0,02 | 39 + 1 = 40 |
| 20 | 7 | 0,14 | 40 + 7 = 47 |
| 25 | 3 | 0,06 | 47 + 3 = 50 |
| TOTAL | N = 50 | 1 | |

Le caractère étudié est *la superficie en hectares d'une exploitation agricole*.

Quel est le nombre d'exploitations de moins de 20 hectares ? *Il y a 40 exploitations de moins de 20 hectares.*

Quel est le pourcentage d'exploitations ayant moins de 15 ha de superficie (strictement) ?

0,78 × 100 = 78. 78% des exploitations ont une superficie de strictement moins de 15ha.

Quel est le pourcentage d'exploitations ayant plus de 9 ha de superficie ?

0,4 × 100 = 40. 40% des exploitations ont une superficie de plus de 9ha.

II. Médiane et intervalle inter-quartile :

1) La médiane :

La médiane d'une série, notée Me , est la valeur du caractère qui partage la série en deux parties de même effectif.

Pour déterminer la médiane d'une série statistique, il faut **ranger les valeurs du caractère par ordre croissant**, chacune des valeurs figurant un nombre de fois égal à son effectif.

- Si le nombre de valeurs est **impair**, la médiane est la valeur centrale.
- Si le nombre de valeurs est **pair**, la médiane est la moyenne des deux valeurs centrales.

Exemples :

a) On considère la liste des prix en euros 4 ; 5 ; 8 ; 10 ; 12 ; 13 ; 14 ; 16 ; 17.
L'effectif est **9 (impair)** donc la médiane est **la 5^{ème} valeur**. La médiane est **12**.

b) On considère la liste des prix en euros 4 ; 5 ; 8 ; 10 ; 11 ; 11 ; 12 ; 13 ; 13 ; 14 ; 16 ; 17.
L'effectif est **12 (pair)** donc la médiane est **la moyenne de la 6^{ème} et la 7^{ème} valeur**.

La médiane est égale à $\frac{11+12}{2} = 11,5$.

c)

| | | | | | | | | |
|------------------------------|---|----|----|----|----|----|----|-------|
| Superficie en ha x_i | 3 | 5 | 9 | 12 | 15 | 20 | 25 | TOTAL |
| Nombre d'exploitations n_i | 7 | 10 | 13 | 9 | 1 | 7 | 3 | |

Déterminer la médiane de cette série.

Effectif total : 50. C'est un nombre pair donc la médiane est la moyenne entre la 25^{ème} et la 26^{ème} valeur qui sont toutes les deux égales à 9. Donc la médiane vaut 9.

Propriétés :

- **50% au moins des individus ont une valeur du caractère inférieure ou égale à la médiane.**
- **La médiane est une valeur insensible aux valeurs extrêmes. C'est un indicateur de tendance centrale. Elle va permettre de comparer deux séries.**

2) Les quartiles :

1er quartile: Q_1 est la plus petite valeur de la série telle qu'au moins 25% des valeurs soient inférieures ou égales à Q_1 .

3ème quartile: Q_3 est la plus petite valeur de la série tel qu'au moins 75% des valeurs soient inférieures ou égales à Q_3 .

Pour déterminer les quartiles d'une série statistique, il faut **ranger les valeurs du caractère par ordre croissant**, chacune des valeurs figurant un nombre de fois égal à son effectif.

On divise ensuite l'effectif total par 4 puis on arrondit à l'entier supérieur, ce qui nous donne le rang de la valeur égale à Q_1 .

Pour Q_3 , on divise l'effectif total par 4 et on multiplie le résultat par 3 puis on arrondit à l'entier supérieur, ce qui nous donne le rang de la valeur égale à Q_3 .

L'intervalle interquartile est $[Q_1, Q_3]$. Il contient au moins 50% des valeurs de la série.

L'écart interquartile est la différence $Q_3 - Q_1$, c'est la longueur de l'intervalle inter-quartile.

Les quartiles sont des indicateurs de position. L'écart inter-quartile est un indicateur de dispersion. Plus il est important, plus la série est dispersée. Il permet de comparer deux séries.

Exemples :

a) On considère la liste des prix en euros 4 ; 5 ; 8 ; 10 ; 12 ; 13 ; 14 ; 16 ; 17.

L'effectif est 9. $\frac{9}{4} = 2,25$ donc Q_1 est la 3^è valeur donc $Q_1 = 8$.

$2,25 \times 3 = 6,75$ donc Q_3 est la 7^è valeur donc $Q_3 = 14$.

b)

| | | | | | | | | |
|------------------------------|---|----|----|----|----|----|----|-------|
| Superficie en ha x_i | 3 | 5 | 9 | 12 | 15 | 20 | 25 | TOTAL |
| Nombre d'exploitations n_i | 7 | 10 | 13 | 9 | 1 | 7 | 3 | |

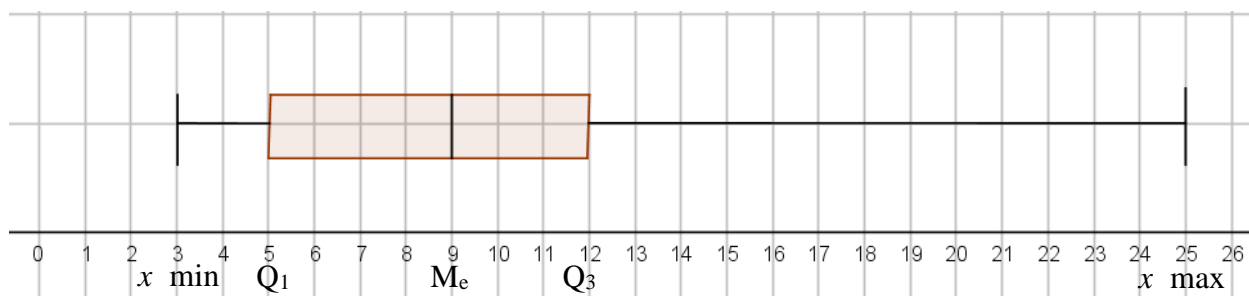
Déterminer les quartiles de cette série.

Effectif total : 50.

$\frac{50}{4} = 12,5$ donc Q_1 est la 13^è valeur de la série donc $Q_1 = 5$.

$12,5 \times 3 = 37,5$ donc Q_3 est la 38^è valeur de la série donc $Q_3 = 12$.

A partir de la médiane et des quartiles on peut construire le diagramme en boîte de la série.



3) La moyenne et l'écart-type

a) La moyenne :

| | | | | | | |
|----------------------------|----------------|----------------|-----|-----|----------------|-------|
| Valeurs du caractère x | x ₁ | x ₂ | ... | ... | x _p | TOTAL |
| effectifs : n _i | n ₁ | n ₂ | ... | ... | n _p | N |

La **moyenne pondérée** de la série, notée \bar{x} , est le nombre défini par :

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_p x_p}{n_1 + n_2 + \dots + n_p} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_p x_p}{N}$$

La **moyenne est un indicateur de position, elle est très sensible aux valeurs extrêmes.**

Exemple : 1) Calculer la superficie moyenne d'une exploitation.

| | | | | | | | | | |
|------------------------|----------------|---|----|----|----|----|----|----|-------|
| Superficie en ha | x _i | 3 | 5 | 9 | 12 | 15 | 20 | 25 | TOTAL |
| Nombre d'exploitations | n _i | 7 | 10 | 13 | 9 | 1 | 7 | 3 | |

$$\text{La superficie moyenne est : } \bar{x} = \frac{7 \times 3 + 10 \times 5 + 13 \times 9 + 9 \times 12 + 1 \times 15 + 7 \times 20 + 3 \times 25}{7 + 10 + 13 + 9 + 1 + 7 + 3} = \frac{526}{50} = 10,52.$$

2) On modifie la dernière colonne du tableau avec une superficie est de 100ha au lieu des 25. Calculer la nouvelle superficie moyenne.

$$\frac{7 \times 3 + 10 \times 5 + 13 \times 9 + 9 \times 12 + 1 \times 15 + 7 \times 20 + 3 \times 100}{7 + 10 + 13 + 9 + 1 + 7 + 3} = \frac{751}{50} = 15,02.$$

La nouvelle superficie moyenne est de 15,02 ha.

Linéarité de la moyenne :

- Lorsqu'on ajoute (ou en retranche) un même nombre k à chacune des valeurs du caractère, sans changer les effectifs, la moyenne augmente (ou diminue) de k.
- Lorsqu'on multiplie chacune des valeurs du caractère par un même nombre k, sans changer les effectifs, la moyenne est multipliée par k.

Exemples :

Dans une boutique la moyenne des prix est de 12 €.

1) Que devient le prix moyen dans cette boutique si le commerçant décide d'augmenter tous les prix de 1€ ?

Si le commerçant décide d'augmenter tous les prix de 1 € alors le prix moyen dans cette boutique sera de 13 €.

2) Que devient le prix moyen dans cette boutique si le commerçant décide de multiplier tous les prix par 1,1 et de leur rajouter 0,15€ ?

Si maintenant le commerçant décide de multiplier le prix de tous les articles de 1,1, et de leur rajouter 0,15€ alors le prix moyen sera $13 \times 1,1 + 0,15 = 14,3 + 0,15 = 14,45€$

b) L'écart-type :

| | | | | | | |
|----------------------------|----------------|----------------|-----|-----|----------------|-------|
| Valeurs du caractère x | x ₁ | x ₂ | ... | ... | x _p | TOTAL |
| effectifs : n _i | n ₁ | n ₂ | ... | ... | n _p | N |

L'écart-type d'une série est le nombre positif σ , défini par :

$$\sigma = \sqrt{\frac{n_1 (x_1 - \bar{x})^2 + n_2 (x_2 - \bar{x})^2 + \dots + n_p (x_p - \bar{x})^2}{n_1 + n_2 + \dots + n_p}} = \sqrt{\frac{n_1 (x_1 - \bar{x})^2 + n_2 (x_2 - \bar{x})^2 + \dots + n_p (x_p - \bar{x})^2}{N}}$$

Il se calcule à la calculatrice.

L'écart-type est un indicateur de dispersion. Plus il est important, plus la série est dispersée autour de sa moyenne. Il permet de comparer des séries.

Exemple : Calculer l'écart-type de cette série .

| | | | | | | | | | |
|------------------------|----------------|---|----|----|----|----|----|----|-------|
| Superficie en ha | x _i | 3 | 5 | 9 | 12 | 15 | 20 | 25 | TOTAL |
| Nombre d'exploitations | n _i | 7 | 10 | 13 | 9 | 1 | 7 | 3 | |

$$\sigma = \sqrt{\frac{7(3-10,52)^2 + 10(5-10,52)^2 + 13(9-10,52)^2 + 9(12-10,52)^2 + 1(15-10,52)^2 + 7(20-10,52)^2 + 3(25-10,52)^2}{7+10+13+9+1+7+3}} \approx 6,37$$

EXERCICES A FAIRE

Les parcours vert, orange et rouge sont bien .

Parcours vert :

- n° 34 page 283 Médiane et Quartile
- n° 35 page 283 Médiane et Quartile
- n° 37 page 281 Comparaison de deux séries connaissant la médiane et les quartiles
- n° 40 page 284 Médiane et Quartile
- n° 48 page 286 Moyenne et écart-type
- n° 51 page 286 Moyenne et écart-type
- n° 57 page 288 Moyenne pondérée
- n° 62 page 289 Comparaison de deux séries

Parcours orange :

- n° 33 page 282 Médiane et comparaison de séries
- n° 44 page 285 Comparaison de deux séries à l'aide des boîtes
- n° 45 page 285 Comparaison de deux séries à l'aide des boîtes
- n° 47 page 286 Moyenne et écart-type + comparaison de séries
- n° 49 page 286 Moyenne et écart-type + questions difficiles
- n° 54 page 287 Associer un résultat au couple (\bar{x} , σ)
- n° 61 page 289 Interpréter les indicateurs d'une série
- n° 68 page 291 Synthèse et recherche

Parcours rouge :

- n° 42 page 284 Médiane, inter-quartile et influence des changements de valeurs
- n° 43 page 285 Donner une série ayant des indicateurs fixés
- n° 50 page 286 Lecture de diagrammes en bâtons
- n° 58 page 288 Calcul littéral de \bar{x}
- n° 60 page 288 Donner une série ayant des indicateurs fixés, donnés par écran calculette
- n° 63 page 289 Commenter un article
- n° 65 page 290 Calcul littéral de \bar{x} + application