

Statistiques

Rappels de vocabulaire :

"Je suis pêcheur et je désire avoir des informations sur la taille des truites d'une rivière. Je décide de mesurer les 28 truites obtenues au cours des trois dernières semaines"



Voilà la liste des tailles obtenues arrondies au centimètre près :

22 - 24 - 25 - 23 - 25 - 25 - 24 - 24 - 24 - 17 - 20 - 23 - 23 - 25 - 24 - 17 - 24 - 25 - 23 - 18 - 20 - 22 - 24 - 25 - 20 - 23 - 24 - 25 -

Cette liste rassemble des **données**, il s'agit d' une **série statistique**.

La **population** choisie pour l'étude est l'ensemble des truites pêchées.

Elle comprend 28 **individus**.

Le **caractère** étudié est la "taille" des truites. Il est **quantitatif** car il s'exprime par un nombre. Dans ce chapitre, nous utiliserons toujours des caractères quantitatifs (et non pas **qualitatifs** comme "la couleur des écailles" par exemple)

La série comprend 7 **valeurs** différentes : 17; 18; 20; 22; 23; 24; 25.

L'**effectif** total de la série est 28. L'effectif correspondant à la valeur 17 est 2.

I) Présentation d'une série statistique:

Pour dépouiller les résultats d'une série, on regroupe les résultats dans un tableau précisant les effectifs de chaque valeur. Pour faciliter l'étude, on range les valeurs dans l'ordre croissant.

Ex : Tableau des effectifs

Taille (cm)	17	18	20	22	23	24	25
Effectifs	2	1	3	4	5	6	7

Une série peut aussi être définie par les fréquences de ses valeurs

Ex : Tableau des fréquences

Taille (cm)	17	18	20	22	23	24	25
Fréquences	$\frac{2}{28}$	$\frac{1}{28}$	$\frac{3}{28}$	$\frac{4}{28}$	$\frac{5}{28}$	$\frac{6}{28}$	$\frac{7}{28}$
	≈ 0,07	≈ 0,04	≈ 0,11	≈ 0,14	≈ 0,18	≈ 0,21	≈ 0,25

7 truites sur 28 ont une taille de 25 cm. La **proportion** de truites ayant cette taille est donc de $\frac{7}{28}$. $\frac{7}{28}$ est la **fréquence** correspondant à la valeur 25 cm. La fréquence est ici de 0,25; on peut l'exprimer aussi en pourcentage soit 25%.



La **somme** des fréquences est toujours **égale à 1**.

a) Effectifs cumulés croissants - Fréquences cumulées croissantes :

je souhaite connaître immédiatement le nombre de truites ou la fréquence correspondant à une taille inférieure ou égale à 22 cm. Il suffit d'ajouter tous les effectifs ou toutes les fréquences précédant cette valeur. C'est le principe des **effectifs cumulés croissants (eec)** ou des **fréquences cumulées croissantes (fcc)**.



Ex : Tableau des effectifs et effectifs cumulés croissants

Taille (cm)	17	18	20	22	23	24	25
Effectifs	2	1	3	4	5	6	7
Effectifs cumulés croissants (eec)	2	3	6	10	15	21	28

le nombre de truites de taille inférieure ou égale à 22 cm est 10



Ex : Tableau des fréquences et fréquences cumulées croissantes

Taille (cm)	17	18	20	22	23	24	25
Fréquences	≈ 0,07	≈ 0,04	≈ 0,11	≈ 0,14	≈ 0,18	≈ 0,21	≈ 0,25
Fréquences cumulées croissantes (fcc)	≈ 0,07	≈ 0,11	≈ 0,22	≈ 0,36	≈ 0,54	≈ 0,75	1

la fréquence correspondant à une taille inférieure ou égale à 22 cm est 0,36. 36% des truites ont une taille inférieure ou égale à 22 cm.



b) Regroupements par classes :

Pour faciliter l'exploitation et la présentation de certaines séries, on peut regrouper les valeurs en classes.

Reprenons notre exemple avec les données brutes de départ (avant d'arrondir au cm)
 22,2- 23,9 - 25,1 - 22,9 - 24,7 - 25,2 - 24,3 - 23,8 - 24,4 - 17,1 - 20,2 - 22,8 - 23,1 - 25,2 -
 22,4 - 17,2 - 23,3 - 24,8 - 23,2 - 18 - 20,2 - 22,4 - 23,9 - 25,1 - 20,2 - 22,7 - 22 - 24,8 -

Dans le tableau ci-dessous, la série est regroupée en en classes d'**amplitude** 2cm

Taille (cm)	[17;19[[19;21[[21;23[[23;25[
Effectifs	3	3	7	15
Fréquences	≈ 0,11	≈ 0,11	≈ 0,25	≈ 0,53

environ 53% des truites ont une taille comprise entre 20 et 25 cm !



II) Paramètres d'une série statistique :

a) Mesures de position (ou de tendance centrale) :

définition :

► La **moyenne** de la série statistique ci-contre est le nombre réel, noté \bar{x} , tel que :

valeur x_i	x_1	x_2	...	x_p
effectif n_i	n_1	n_2	...	n_p

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_px_p}{N}$$

N est l'effectif total de la série. $N = n_1 + n_2 + \dots + n_p$

► La moyenne peut se calculer à partir du tableau de fréquences :

valeur x_i	x_1	x_2	...	x_p
fréquence f_i	f_1	f_2	...	f_p

$$\bar{x} = f_1x_1 + f_2x_2 + \dots + f_px_p$$

Ex : Calculons la moyenne de notre exemple de départ.

Taille (cm)	17	18	20	22	23	24	25
Effectifs	2	1	3	4	5	6	7
Fréquences	≈ 0,07	≈ 0,04	≈ 0,11	≈ 0,14	≈ 0,18	≈ 0,21	≈ 0,25

avec les effectifs :

$$\bar{x} = \frac{2 \times 17 + 1 \times 18 + 3 \times 20 + 4 \times 22 + 5 \times 23 + 6 \times 24 + 7 \times 25}{28} \approx \mathbf{22,6 \text{ cm}}$$

avec les fréquences :

$$\bar{x} = 17 \times 0,07 + 18 \times 0,04 + 20 \times 0,11 + 22 \times 0,14 + 23 \times 0,18 + 24 \times 0,21 + 25 \times 0,25 \approx \mathbf{22,6 \text{ cm}}$$

définition :

La **médiane** M d'une série statistique est le nombre tel que :

- 50% au moins des individus ont une valeur du caractère inférieure ou égale à M
- 50% au moins des individus ont une valeur du caractère supérieure ou égale à M

ce nombre permet de couper la population étudiée en deux groupes comprenant le même nombre d'individus !



Ex : Reprenons notre exemple en modifiant la population.

► population de 5 truites rangées dans l'ordre croissant des valeurs

Taille (cm)	17	18	20
Effectifs	2	1	2

L'effectif total est 5. Il s'agit d'un nombre impair. La médiane correspond donc à la taille de la 3ème truite. **La médiane est $M = 18$ cm**

► population de 6 truites rangées dans l'ordre croissant des valeurs.

Taille (cm)	17	18	20	21
Effectifs	2	1	1	2

L'effectif total est 6. Il s'agit d'un nombre pair. La médiane correspond donc à une taille comprise entre la 3ème et la 4ème truite. On prend en général la demi-somme.

La médiane est $M = \frac{18 + 20}{2} = 19$ cm

Dans la pratique : 

Soit une série statistique d'effectif total N rangée dans l'ordre croissant des valeurs.

- Si N est impair, il s'écrit sous la forme $N = 2k + 1$ ($k \in \mathbb{N}$), alors M est la valeur située au rang $k + 1$
- Si N est pair, il s'écrit sous la forme $N = 2k$ ($k \in \mathbb{N}$), alors M est la demi-somme des valeurs situées au rang k et au rang $k + 1$

Ex : Reprenons la situation de départ

Taille (cm)	17	18	20	22	23	24	25
Effectifs	2	1	3	4	5	6	7

L'effectif total est 28. Il s'agit d'un nombre pair : $28 = 2 \times 14$
 M est la demi-somme des valeurs situées au rang 14 et 15.

La médiane est $M = \frac{23 + 23}{2} = 23$ cm

définition : Soit une série statistique d'effectif total N rangée dans l'ordre croissant des valeurs.

► **Le premier quartile Q_1** est la plus petite valeur Q_1 de la série telle **qu'au moins un quart** des valeurs de la liste **sont inférieures ou égales à Q_1**

► **Le troisième quartile Q_3** est la plus petite valeur Q_3 de la série telle **qu'au moins les trois quarts** des valeurs de la liste **sont inférieures ou égales à Q_3**

Ex : Calculons les quartiles de notre série statistique de départ.

Taille (cm)	17	18	20	22	23	24	25
Effectifs	2	1	3	4	5	6	7

L'effectif total est de 28.

$\frac{1}{4}$ de l'effectif correspond à 7 individus.

La plus petite valeur telle qu'au moins un quart des valeurs lui soient inférieures est 22.
Le premier quartile Q_1 est donc égal à 22.

je cherche la taille correspondant à la 7ème truite !!



$\frac{3}{4}$ de l'effectif correspond à 21 individus.

La plus petite valeur telle qu'au moins trois quarts des valeurs lui soient inférieures est 24.

Le troisième quartile Q_3 est donc égal à 24.

je cherche la taille correspondant à la 14 ème truite !



et si le quart ou les trois quarts des individus ne correspondent pas à des nombres entiers ?

Ex : Modifions les données pour étudier ce cas

Taille (cm)	17	18	20	22	23	24	25
Effectifs	1	4	3	4	7	2	4

L'effectif total est de 25. $\frac{1}{4}$ de 25 est égal à 6,25. $\frac{3}{4}$ de 25 est égal à 18,75

Le premier quartile est la valeur telle qu' **au moins** un quart des valeurs lui soient inférieures. Il s'agit donc de 20

je cherche la taille correspondant à la 7 ème truite !!

Le troisième quartile est la valeur telle qu' **au moins** trois quarts des valeurs lui soient inférieures. Il s'agit donc de 23

je cherche la taille correspondant à la 19 ème truite !

b) Mesures de dispersion :

définition : L'étendue d'une série est la différence entre la plus grande et la plus petite valeur de la série.

définition : L'écart interquartile est la différence entre le troisième et le premier quartile.

Ex : Reprenons la série statistique de départ.

Taille (cm)	17	18	20	22	23	24	25
Effectifs	2	1	3	2	5	8	7

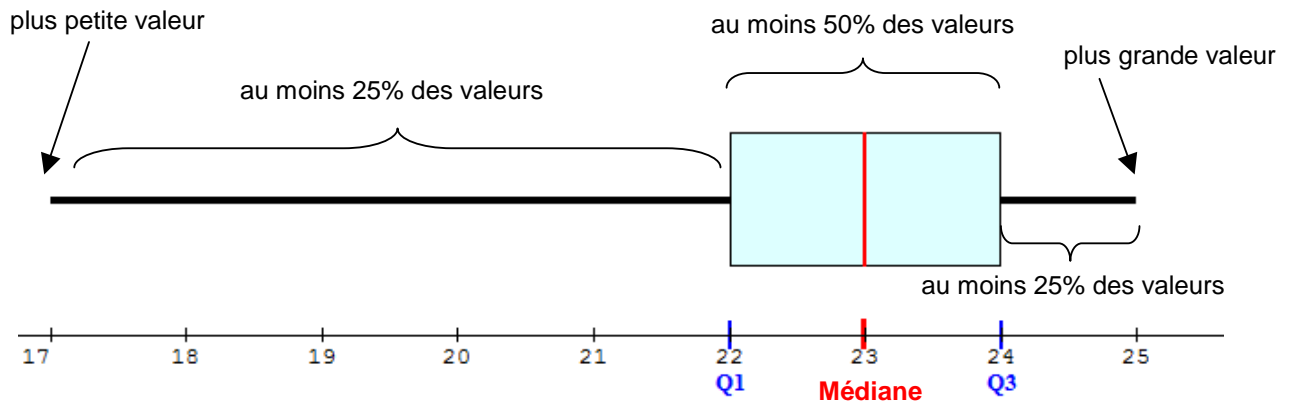
L'étendue de la série est : $25 - 17 = 8$ cm

L' écart interquartile est : $Q_3 - Q_1 = 24 - 22 = 2$

III) Résumé d'une série statistique :

On peut résumer tous les paramètres sur un diagramme "en boîte".

Ex : Résumons les indicateurs de notre première série statistique.



Près de la moitié des truites ont une taille entre 22 et 24 cm !
L'étendue de la série est de 8 cm.
La plus petite taille est 17 cm, la plus grande est 25 cm.
Près de la moitié des truites ont une taille inférieure ou égale à 23 cm !

