

Statistiques descriptives

Connaissances du collège nécessaires à ce chapitre

- ▶ Lire un diagramme en barres
- ▶ Lire un tableau à double entrée
- ▶ Calculer des fréquences
- ▶ Calculer des effectifs ou des fréquences cumulés



1. Définitions

■ DÉFINITION : Population, individu

Une **série statistique** est un ensemble de valeurs qui est recueilli pour étudier certains critères. La **population** d'une série statistique est l'ensemble des **individus** sur lesquels portent cette étude.

Exemple L'ensemble des mamans à qui le sondage était adressé peut être la population étudiée. Chaque maman ayant répondu à ce sondage est un individu de cette étude.

■ DÉFINITION : Caractère – Qualitatif/Quantitatif – Discret/Continu

Le **caractère** (ou variable) d'une série statistique est l'information étudiée.

- Lorsque le caractère ne prend que des valeurs numériques, il est **quantitatif** :
 - **discret** s'il ne peut prendre que des valeurs isolées (notes, âge, ...)
 - **continu** dans le cas contraire (poids, taille, ...). Dans ce cas, on effectue souvent un regroupement de valeurs par **classes**.
- Sinon, on dit qu'il est **qualitatif** (couleur des yeux, sport pratiqué, ...) : le caractère n'est pas mesuré par des nombres.

Exemple Dans la population des mamans de l'activité 1 :

- 1) le mode d'accouchement est un caractère qualitatif prenant trois valeurs : césarienne, sous péridurale ou naturel
- 2) la taille du bébé en centimètres arrondie au centimètre est un caractère quantitatif discret : il peut prendre les valeurs 46 ; 47 ; 48 ; 49 ; 50 ; 51 ; 52 ; ...
- 3) la masse du bébé en grammes est un caractère quantitatif continu. Il peut prendre toutes les valeurs comprises entre 2000 et 5000. On peut regrouper ces valeurs dans des classes d'étendue (ou amplitude) 500 g : [2000;2500[, [2500;3000[, [3000;3500[, [3500;4000[, [4000;4500[, [4000;4500[et [4500;5000[.

REMARQUE :

- La notation $[a; b[$ indique l'ensemble des nombres compris entre a et b , a inclus et b exclu.
- Les valeurs 2500, 3000, 3500, 4000 et 4500 ne doivent pas être comptabilisées deux fois. La valeur 3500 est comptabilisée dans la classe $[3500;4000[$ et non pas dans la classe $[3000;3500[$.

■ DÉFINITION : Effectif – Fréquence – Distribution des fréquences

À chaque valeur (ou classe) du caractère est associé un **effectif** n : c'est le nombre d'individus associés à cette valeur.

On peut aussi associer à chaque valeur de l'étude une **fréquence** : c'est la proportion (pourcentage) que représente cette valeur au sein de l'étude
(c'est-à-dire Fréquence = $\frac{\text{Effectif de la valeur (ou de la classe)}}{\text{Effectif total}}$)

Le tableau donnant les fréquences de toutes les valeurs du caractère d'une série statistique est appelé **distribution des fréquences**.



REMARQUE :

- La fréquence d'une valeur d'un caractère est un nombre rationnel compris entre 0 et 1.
- La somme des fréquences d'une série statistique doit toujours être égale à 1. On peut constater une légère imprécision si on utilise des valeurs approchées au lieu des valeurs exactes.
- Les fréquences peuvent être données sous forme fractionnelle, décimale ou en pourcentage.

Exemple

Note	7	8	9	10	10,5	11	11,5	12	14	16	18	19
Effectif	1	1	1	4	6	8	3	5	4	1	1	1

Le caractère étudié est dans ce cas la note à un devoir, c'est un caractère quantitatif.

L'effectif total est de 36. À titre d'exemple, l'effectif de la note 11,5 est 3.

Les fréquences des notes sont données ci-dessous :

Note	Fréquence	Note	Fréquence	Note	Fréquence
7	$\frac{1}{36} \approx$	10,5		14	
8		11	$\frac{8}{36} \approx$	16	
9		11,5		18	
10		12		19	

Correction

Note	Fréquence	Note	Fréquence	Note	Fréquence
7	$\frac{1}{36} \approx 0,0278$	10,5	$\frac{6}{36} \approx 0,1667$	14	$\frac{4}{36} \approx 0,1111$
8	$\frac{1}{36} \approx 0,0278$	11	$\frac{8}{36} \approx 0,2222$	16	$\frac{1}{36} \approx 0,0278$
9	$\frac{1}{36} \approx 0,0278$	11,5	$\frac{3}{36} \approx 0,0833$	18	$\frac{1}{36} \approx 0,0278$
10	$\frac{4}{36} \approx 0,1111$	12	$\frac{5}{36} \approx 0,1389$	19	$\frac{1}{36} \approx 0,0278$

Exemple Prenons comme caractère la masse du bébé (activité 1)

Masse (g)	Effectif	Fréquence
[2000; 2500[1	2,5%
[2500; 3000[5	12,5%
[3000; 3500[8	20%
[3500; 4000[15	37,5%
[4000; 4500[6	15%
[4500; 5000[5	12,5%
	40	100%



■ DÉFINITION : Effectif cumulé croissant

L'**Effectif cumulé croissant** (ECC) correspondant à une valeur x_i d'un caractère est la somme des effectifs des valeurs inférieures ou égales à x_i .

On définit de manière analogue les **fréquences cumulées croissantes** (FCC).

Exemple

Note	7	8	9	10	10,5	11	11,5	12	14	16	18	19
Effectif	1	1	1	4	6	8	3	5	4	1	1	1
ECC	1	2	3	7	13	21	24	29	33	34	35	36

Exemple À partir des éléments fournis ci-dessus, compléter le tableau des fréquences cumulées croissantes ci-dessous.

Note	7	8	9	10	10,5	11	11,5	12	14	16	18	19
Fréquence												
Fréq. CC												

Correction

Note	7	8	9	10	10,5	11	11,5	12	14	16	18	19
Fréquence	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{4}{36}$	$\frac{6}{36}$	$\frac{8}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
Fréq. CC	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{7}{36}$	$\frac{13}{36}$	$\frac{21}{36}$	$\frac{24}{36}$	$\frac{29}{36}$	$\frac{33}{36}$	$\frac{34}{36}$	$\frac{35}{36}$	1

Exemple Prenons comme caractère la masse du bébé (activité 1)

Masse (g)	Effectif	ECC
[2000; 2500[1	
[2500; 3000[5	
[3000; 3500[8	
[3500; 4000[15	
[4000; 4500[6	
[4500; 5000[5	

Masse (g)	Fréquence	FCC
[2000; 2500[2,5%	
[2500; 3000[12,5%	
[3000; 3500[20%	
[3500; 4000[37,5%	
[4000; 4500[15%	
[4500; 5000[12,5%	

Correction

Masse (g)	Effectif	ECC
[2000; 2500[1	1
[2500; 3000[5	6
[3000; 3500[8	14
[3500; 4000[15	29
[4000; 4500[6	35
[4500; 5000[5	40

Masse (g)	Fréquence	FCC
[2000; 2500[2,5%	2,5%
[2500; 3000[12,5%	15%
[3000; 3500[20%	35%
[3500; 4000[37,5%	72,5%
[4000; 4500[15%	87,5%
[4500; 5000[12,5%	100%





2. Paramètres d'une série statistique

Dans cette partie, on considère uniquement des séries statistiques à caractère quantitatif.

A. Mesures de position

■ DÉFINITION : Moyenne

La **moyenne** d'une série statistique à caractère quantitatif est la valeur que prendraient les valeurs de cette série si elles étaient toutes égales. Elle se note \bar{x} .

$$\bar{x} = \frac{\text{somme totale des valeurs prises par le caractère}}{\text{nombre de valeurs}}$$

■ PROPRIÉTÉ : Calcul de la moyenne avec les effectifs

Si x_1, x_2, \dots, x_p désignent les p valeurs possibles du caractère d'une série statistique et n_1, n_2, \dots, n_p désignent les effectifs correspondants, alors

$$\bar{x} = \frac{n_1 \times x_1 + n_2 \times x_2 + n_3 \times x_3 + \dots + n_p \times x_p}{n_1 + n_2 + n_3 + \dots + n_p}$$

Exemple À partir des notes vues précédemment

$$\bar{x} = \frac{7 \times 1 + 8 \times 1 + 9 \times 1 + 10 \times 4 + 10,5 \times 6 + 11 \times 8 + 11,5 \times 3 + 12 \times 5 + 14 \times 4 + 16 \times 1 + 18 \times 1 + 19 \times 1}{1 + 1 + 1 + 4 + 6 + 8 + 3 + 5 + 4 + 1 + 1 + 1} = \frac{418,5}{36} = 11,625$$

■ PROPRIÉTÉ : Calcul de la moyenne avec les fréquences

Si x_1, x_2, \dots, x_p désignent les p valeurs possibles du caractère d'une série statistique, et f_1, f_2, \dots, f_p désignent les fréquences correspondantes, alors

$$\bar{x} = f_1 \times x_1 + f_2 \times x_2 + f_3 \times x_3 + \dots + f_p \times x_p$$

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

$$\bar{x} = \frac{n_1 \times x_1 + n_2 \times x_2 + n_3 \times x_3 + \dots + n_p \times x_p}{n_1 + n_2 + n_3 + \dots + n_p}$$

$$\bar{x} = \frac{n_1 \times x_1}{N} + \frac{n_2 \times x_2}{N} + \frac{n_3 \times x_3}{N} + \dots + \frac{n_p \times x_p}{N}$$

$$\bar{x} = f_1 \times x_1 + f_2 \times x_2 + f_3 \times x_3 + \dots + f_p \times x_p$$

■ DÉFINITION : Médiane

La **médiane** d'une série statistique dont les valeurs sont rangées par ordre croissant est la valeur, notée Med , telle que 50 % au moins ont une valeur inférieure ou égale à Med et telle que 50 % des individus au moins, ont une valeur supérieure ou égale à Med .

■ DÉFINITION : Quartiles

Le **premier quartile** (noté **Q1**) d'une série statistique numérique est la plus petite valeur prise par le caractère telle qu'au moins 25 % des valeurs lui soient inférieures ou égales.
Le **troisième quartile** (noté **Q3**) d'une série statistique numérique est la plus petite valeur prise par le caractère telle qu'au moins 75 % des valeurs lui soient inférieures ou égales.