

Vocabulaire

La statistique étudie certaines caractéristiques : **caractères** ou **variables** d'un ensemble fini appelé **population**. Les éléments de cette population étudiée sont appelés **individus**. Une variable peut être :

- **Quantitative** : numérique et fait l'objet de calcul (âge, taille, poids, notes, nombres d'heures etc ...)
- **Qualitative** : c'est le contraire de quantitative.
- **Discrète** : si la variable ne prend qu'un nombre fini de valeurs (ces valeurs sont notées x_i).
- **Continue** : si la variable prend ses valeurs dans un intervalle (classe)

Exemples :

Supposons que l'on veuille faire une étude statistique sur les 50 notes attribuées par un jury à un examen. On dispose pour cette étude de la liste des notes obtenues :

2	14	10	16	20	19	7	5	13	14
6	9	16	13	12	3	7	8	18	12
4	8	15	10	8	11	13	9	9	13
5	8	14	5	11	12	2	1	7	1
6	12	3	11	19	17	18	3	0	4

On peut regrouper ces notes par ordre croissant : 0,1,1,2,2,3,3,3 ..., et construire le tableau suivant, dans ce cas la distribution est discrète :

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	2	2	3	2	3	2	3	4	3	2	3	4	4	3	1	2	1	2	2	1

Ou bien regrouper ces notes par **intervalle (classe)**, dans ce cas la distribution est continue

[0 ; 5[10
[5 ; 8[8
[8 ; 12[12
[12 ; 15[11
[15 ; 20[9
	50

Effectif

L'**effectif** d'une **classe** est le nombre d'individu de cette classe ou de cette modalité. On le note n_i
L'**effectif total** est la somme des effectifs de toutes les classes.

On le note souvent N , on a alors : $N = n_1 + n_2 + n_3 + n_4 + n_5 = 50$.

Autre notation :	$N = \sum_{i=1}^5 n_i$
------------------	------------------------

Fréquence

La **fréquence** f_i de la classes i ou de la variable x_i est le rapport n_i/N , la fréquence d'une classe est un nombre de l'intervalle $[0 ;1]$

Indicateurs de position

Mode d'une série statistique

Dans 'une série statistique continue, la **classe modale** est la classe du plus grand effectif :

Exemple:

Notes	Effectifs	Effectif
[0 ; 5[10	
[5 ; 8[8	
[8 ; 12[12	
[12 ; 15	11	
[15 ; 20	9	
	50	

Sur cette exemple, la classe modale est donc

Classe modale ? ▼

Dans le cas d'une série statistique discrète, le **mode** est la valeur de plus grand effectif :

Notes	Effectif	Effectif
0	1	
1	2	
2	2	
3	3	
4	2	
5	3	
6	2	
7	3	
8	4	
9	3	
10	2	
11	3	
12	4	
13	4	
14	3	
15	1	
16	2	
17	1	
18	2	
19	2	
20	1	

mode ? ▼

Moyenne

$$\bar{x} = \frac{\sum_{i=1}^N n_i x_i}{\sum_{i=1}^N n_i} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_N x_N}{n_1 + n_2 + \dots + n_N}$$

$n_1, n_2, n_3, \dots, n_N$ sont les effectifs correspondants aux variables $x_1, x_2, x_3, \dots, x_N$, si la série est discrète ou aux centres de classe $x_1, x_2, x_3, \dots, x_N$, si la série est continue.

Exemple:

Série discrète

Modalité	Effectif	Produit
10	2	
14	1	
15	4	
12	3	
11	1	
12	0	
Total		
	Moyenne	

Série continue

Notes	Effectifs	Centre	Produit
[0 ; 5[10		
[5 ; 8[8		
[8 ; 12[12		
[12 ; 15[11		
[15 ; 20[9		
	50	Total	
		Moyenne	

Médiane

La médiane est un paramètre de position, qui permet de couper la population étudiée en deux groupes contenant le même nombre d'individus.

Ce paramètre est utile pour donner la répartition du caractère étudié, car 50 % environ de la population étudiée a une modalité inférieure à la médiane et 50 % une modalité supérieure à la médiane.

Exemple

On fait une étude statistique sur les 50 notes attribuées par un jury à un examen, voici les résultats obtenus en classant ces notes par ordre croissant :

Variable discrète

Notes	Effectif	Effectif cumulé
0	1	1
1	2	3
2	2	5
3	3	8
4	2	10
5	3	13
6	2	15
7	3	18
8	4	22
9	3	25
10	2	27
11	3	30
12	4	34
13	4	38
14	3	41
15	1	42
16	2	44
17	1	45
18	2	47
19	2	49
20	1	50

Utilisons la colonne des **effectifs cumulés** pour déterminer la médiane : il y a 50 notes, la 25^{ème} note est 9 et la 26^{ème} : 10.

Voici la répartition des notes :

0; 1; 1; ...; 9; 9; **9**; **10**; 10; ... 19; 19; 20
└──────────┬──────────┘ └──────────┬──────────┘
25 NOTES 25 NOTES

Dans le tableau il n'y a pas de valeur partageant la série statistique en deux groupe de même effectif, (l'effectif total est pair) dans ce cas **l'intervalle médian** est **[9;10]** et on prend pour médiane le **centre** de cet intervalle : 9,5

Si la variable est continue (regroupement par intervalle des résultats) le calcul de la médiane se fait autrement :

Notes	Effectifs	Effectifs cumulés
[0 ; 5[10	10
[5 ; 8[8	18
[8 ; 12[12	30
[12 ; 15	11	41
[15 ; 20	9	50
	50	

Utilisons la colonne des effectifs cumulés pour déterminer la médiane : Il y a 50 notes, 50 % de l'effectif total c'est 25, la médiane est ici la note correspondant à l'effectif cumulé **25**.

D'après la colonne "effectif cumulé" :

- 18 personnes ont moins de 8
- 30 personnes ont moins de 12

La médiane se trouve donc dans l'intervalle [8;12[(appelé **classe médiane**).

Quartiles

Les 3 quartiles sont les 3 valeurs qui partagent la population totale en 3 parties d'effectifs égaux:

Le 1er quartile Q1 correspond à 25 % de l'effectif total

Le 2ème quartile Q2 correspond à la médiane, soit 50 % de l'effectif total

Le 3ème quartile Q3 correspond à 75 % de l'effectif total

Indicateurs de dispersion

Etendue d'une série statistique :

L'étendue d'une série statistique est la différence entre la plus grande et la plus petite valeur du caractère.

Exemple:

Notes	Effectif	Effectif cumulé
0	1	1
1	2	3
2	2	5
3	3	8
4	2	10
5	3	13
6	2	15
7	3	18
8	4	22
9	3	25
10	2	27
11	3	30
12	4	34
13	4	38
14	3	41
15	1	42
16	2	44
17	1	45
18	2	47
19	2	49
20	1	50

Notes	Effectifs	Effectifs cumulés
[0 ; 5[10	10
[5 ; 8[8	18
[8 ; 12[12	30
[12 ; 15	11	41
[15 ; 20	9	50
	50	

$20 - 0 = 20$,

20 est l'étendue de ces deux séries (continue et discrète)

Ecart inter-quartile: il est égal à $Q3 - Q1$

Variance :

Pour calculer la variance d'une série statistique on utilise la formule :

$$V = \frac{\sum_{i=1}^N n_i (x_i - \bar{x})^2}{\sum_{i=1}^N n_i} = \frac{n_1 (x_1 - \bar{x})^2 + n_2 (x_2 - \bar{x})^2 + \dots + n_N (x_N - \bar{x})^2}{n_1 + n_2 + \dots + n_N}$$

Il faut donc calculer d'abord la moyenne.

La variance peut être calculée aussi en utilisant la formule :

$$V = \frac{\sum_{i=1}^N n_i x_i^2}{\sum_{i=1}^N n_i} - \bar{x}^2$$

Ecart-type:

L'écart-type est le nombre noté σ tel que : $\sigma = \sqrt{V}$.