Caractéristiques d'une série statistique

Un tableau statistique ou un graphique sont parfois longs à consulter, sans permettre d'avoir une idée suffisamment concise de la distribution statistique observée.

Les caractéristiques d'une série permettent d'avoir une idée d'ensemble sur cette série. Elles servent aussi à comparer plusieurs séries statistiques.

Nous distinguerons les caractéristiques de **position** et les caractéristiques de **dispersion**.

1. Caractéristiques de position

1.1 Mode

On appelle **mode** ou **dominante** d'une série statistique la valeur qui a le plus grand effectif (ou la plus grande fréquence).

Un mode d'une série statistique discrète est une valeur du caractère qui correspond au plus fort effectif.

Une classe modale d'une série statistique continue est une classe qui correspond au plus fort effectif.

Remarque :

Une série statistique peut posséder plusieurs modes. Dans ce cas, on dit que la série est plurimodale.

Exemple : Considérons la série

Valeurs du caractère (notes) x_i	7	8	9	10	11	12
Effectifs (nombre d'élèves ayant la note) n_i	5	6	2	**7**	4	6

7 est le plus grand effectif donc le mode est 10.

Moyenne $\overline{x}$

La **moyenne** se calcule additionnant les différentes valeurs de la série statistique et en divisant ce nombre par l'effectif total de la série. L'expression mathématique de la moyenne s'écrit de la façon suivante :

Dans le cas où la série est donnée par

x_i	x_1	x_2	….	x_p
n_i	n_1	n_2	…	n_p
f_i	f_1	f_2		f_p

L'effectif total est $N = n_1 + n_2 + \ldots + n_p$.

La moyenne de la série est le nombre $\overline{x} = \dfrac{n_1 x_1 + n_2 x_2 + \ldots + n_p x_p}{N}$.

Remarques :

- On obtient aussi la moyenne à partir des fréquences $\bar{x} = f_1 x_1 + f_2 x_2 + \ldots + f_p x_p$, où

$$f_1 = \frac{n_1}{N} \quad , \quad f_2 = \frac{n_2}{N} \quad \ldots \text{ et } \quad f_p = \frac{n_p}{N}$$ sont les fréquences ;

- Pour une série statistique non regroupée suivant les effectifs, les effectifs de chaque variables sont tous égaux à 1, la moyenne est égale à :

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_p}{N} \quad ;$$

- Pour une série regroupée en classes, on utilise le centre de chaque classe comme valeur de x_i dans le calcul de la moyenne.

1.2 Médiane

La **médiane** d'une série statistique est le nombre qui partage les valeurs du caractère en deux parties de même effectif :

- les valeurs inférieures à la médiane ;
- les valeurs supérieures à la médiane.

Recherche pratique de la médiane :

On range les valeurs du caractère une par une dans l'ordre croissant (chaque valeur du caractère doit apparaître un nombre de fois égal à l'effectif correspondant).

- Si l'effectif total est impair, la médiane *M* est la valeur du caractère située au milieu.
- Si l'effectif total est pair, la médiane *M* est la demi-somme des 2 valeurs situées au milieu.

Exemple :

Dans une classe de 15 élèves, le professeur rend les copies par ordre croissant des notes : la note médiane est la note de la $8^{\text{ème}}$ copie.

Pour 14 élèves, la note médiane sera la moyenne des notes des $7^{\text{ème}}$ et $8^{\text{ème}}$ copies.

Remarque :

Dans le cas d'une série à caractère continu, la médiane est la valeur qui correspond à une fréquence cumulée de 0,5.

2. Caractéristiques de dispersion

2.1 Étendue

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

Ce paramètre est d'un intérêt limité car les valeurs extrêmes sont souvent accidentelles.

2.2 Écart moyen absolu e_m

L'écart à la moyenne de la valeur x_i est $x_i - \bar{x}$.

L'**écart moyen absolu** e_m d'une série statistique est la moyenne des valeurs absolues des écarts à la moyenne des valeurs :

Valeurs	x_1	x_2	…	x_p
Effectifs	n_1	n_2	…	n_p

Alors $e_m = \dfrac{n_1|x_1-\bar{x}|+n_2|x_2-\bar{x}|+\ldots+n_p|x_p-\bar{x}|}{N}$

Avantages et inconvénients :

L'écart moyen a l'avantage de faire intervenir tous les éléments de la série statistique, mais sa détermination nécessite beaucoup de calculs quand N est grand.

C'est un paramètre peu utilisé.

2.3 Variance V

La **variance** d'une série statistique est la moyenne des carrés des écarts à la moyenne :

$V = \dfrac{n_1(x_1-\bar{x})^2+n_2(x_2-\bar{x})^2+\ldots+n_p(x_p-\bar{x})^2}{N}$ où $\bar{x}$ est la moyenne et $N = n_1 + n_2 + \ldots + n_p$.

Méthode de calcul :

Même avec des valeurs observées x_i très simples, il se peut que le calcul de la variance *V* nécessite des calculs fastidieux.

On démontre que : $V = \dfrac{n_1 x_1^2+n_2 x_2^2+\ldots+n_p x_p^2}{N} - (\bar{x})^2$.

Cette expression de la variance permet de simplifier les calculs.

Remarques :

- Plus la variance est grande, plus la série est dispersée.

Plus la variance est petite (voisin de 0), plus la série est resserrée autour de la moyenne.

- La variance est une quantité positive ou nulle.

2.4 Écart-type σ

L'**écart-type** d'une série statistique est la racine carrée de la variance.

On a donc $\sigma = \sqrt{(V)} = \sqrt{\left(\dfrac{n_1(x_1-\bar{x})^2+n_2(x_2-\bar{x})^2+\ldots+n_p(x_p-\bar{x})^2}{N}\right)}$

De même que pour la variance :

- plus l'écart type est grand, plus la série est dispersée ;
- plus l'écart est voisin de 0, plus la série est resserrée autour de la moyenne.

Exemple :

											Totaux
Valeurs x_i	2	3	4	6	8	9	10	11	12	14	
Effectifs n_i	1	3	2	5	5	10	5	7	7	2	47
$n_i.x_i$	2	9	8	30	40	90	50	77	84	28	418
x_i^2	4	9	16	36	64	81	100	121	144	196	
$n_i\ x_i^2$	4	27	32	180	320	810	500	847	1008	392	4120

- La moyenne vaut $\bar{x} = 8{,}9$;
- La variance vaut $V = \frac{4120}{47} - 8{,}9^2 = 8{,}56$;
- L'écart-type vaut $\sigma = \sqrt{(V)} = \sqrt{(8{,}56)}$ = 2,93.