CARACTERISTIQUES D'UNE SERIE STATISTIQUE

1. Caractéristiques de position

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.
Considérons les notes des deux épreuves des tableaux 1 et 7.

Tableau 7. Notes sur 20 obtenues par des élèves de la Terminale au cours d'une épreuve 2.
13-17-08-07-08-05-09-09-07-11-07-08-10-12-07-05-15-08-07-
08-11-08-11-12-09-14-13-09-13-06-16-12-12-08-19-14-09
On souhaite en déduire une comparaison du niveau des élèves. Le plus simple est de caractériser chaque série par un seul nombre et de comparer. Plusieurs choix sont possibles pour déterminer ce nombre unique.

1.1 **Mode** (ou dominante)

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

Exemple. épreuve 1 : le mode est 14
épreuve 2 : le mode est 08

Dans le cas des séries classées, on considère des classes d'égale étendue et on appelle *classe modale* la classe qui a le plus grand effectif. Par convention, le mode est le centre de la classe modale.

Remarques :
- Il peut arriver qu'une série statistique présente plusieurs modes. On parle de série plurimodale.
- Le mode existe même lorsque le caractère est qualitatif.

1.2 Médiane

- Reprenons l'exemple des notes obtenues par les élèves de la Terminale au cours de l'épreuve 1, on constate que sur 40 élèves 20 on une note < 13 et 20 ont une note > 13. Par conséquent le "milieu" de la classe est 13 On dit que la série a une médiane 13.

- Définition :

La médiane est la valeur du caractère qui partage l'effectif total en deux parties d'effectifs égaux.

- Détermination de la médiane

i-. Si la population est peu nombreuse, la médiane est obtenue de façon immédiate :
- Cas d'une série constituée d'un nombre impair de valeurs rangées par ordre croissant. La médiane est le nombre qui est au milieu de la série. (Exemple : notes de l'épreuve 2)
05-05-06-07-07-07-07-07-08-08-08-08-08-08-08-09-09-09-09
09-10-11-11-11-12-12-12-12-13-13-13-14-14-15-16-17-19
La médiane est 09
- S'il y a un nombre pair de valeurs, la médiane est un nombre arbitrairement choisi entre les deux nombres consécutifs au milieu de la série. C'est le cas des notes de l'épreuve 1.

ii-. Si la population est nombreuse, on se sert du tableau des effectifs cumulés.
- Cas d'un caractère discret : La médiane est la valeur du caractère à partir de laquelle l'effectif cumulé atteint ou dépasse la moitié de l'effectif total (ou la fréquence cumulée atteint ou dépasse 0,5)

- Cas d'une répartition en classes : On détermine d'abord la classe à laquelle appartient la médiane et l'on convient de choisir la médiane en admettant que, dans cette classe, les valeurs du caractère sont uniformément distribuées.

- Détermination graphique de la médiane

On utilise la courbe des effectifs cumulés. La médiane est l'abscisse du point de la courbe ayant pour ordonnée la moitié N / 2 de l'effectif total. (ou l'ordonnée 0,5 d'une courbe de fréquence)

1.3 Moyenne arithmétique

- Définition :

La *moyenne arithmétique* d'une série statistique est égale à la somme des valeurs du caractère divisées par leur nombre.

i- Cas des données énumérées : $\bar{x} = \dfrac{x_1 + x_2 + ... + x_N}{N}$

ii- Cas d'une variable discrète : on utilise la moyenne pondérée

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + ... + n_c x_c}{N} = \frac{\sum_{i=1}^{c} n_i x_i}{\sum_{i=1}^{c} n_i}$$

exemple : notes de l'épreuve 1

iii- Cas où les valeurs sont regroupées en classes : les n_i valeurs de la i-ème classe sont supposées regroupées au centre de la classe.

- Présentation des calculs

Exemple : reprenons les notes de l'épreuve 1

classes	Centre x_i des classes	Effectif n_i	$n_i x_i$
] 0 , 2]			
] 2 , 4]			
] 4 , 6]			
] 6 , 8]			
]8, 10]			

classes	Centre x_i des classes	Effectif n_i	$n_i x_i$
]10,12]			
]12,14]			
]14,16]			
]16,18]			
]18,20]			
Total			

Tableau 8.

- Propriétés de la moyenne

Changement d'origine - changement d'échelle

Soit une série statistique (x_i, n_i). On démontre que, si $\bar{x}$ est la moyenne de cette série, alors pour tout réel **a** la moyenne de la série (x_i -a, n_i) est $\bar{x}$ - a.

Pour tout réel h, la moyenne de la série (hx_i, n_i) est $h\bar{x}$.

1.4 Comparaison du mode, de la médiane et de la moyenne

Chacun des trois paramètres de position présente des avantages et des inconvénients et le choix dépend de l'usage que l'on veut en faire.

- Le **mode** est facile à déterminer. Il est défini même lorsque le caractère est qualitatif mais il apporte une information réduite sauf lorsqu'une valeur du caractère l'emporte nettement sur les autres.

- La **moyenne** donne une information plus riche qui permet certaines décisions. D'autre part, si on connaît la moyenne et l'effectif de deux populations, on peut en déduire la moyenne de la population totale. Mais la moyenne est influencée par les valeurs extrêmes.
- La **médiane** est facile à déterminer. Et, elle est très peu influencée par les valeurs extrêmes.

2. Caractéristiques de dispersion

Considérons les notes obtenues en mathématiques par deux élèves :

Élève A	10	11	05	12	17
Élève B	11	10	11	10	13

Elles ont même moyenne (11) et même médiane (11). Cependant elles diffèrent profondément, la première série est beaucoup plus dispersée que la deuxième. Ceci montre que la moyenne (ou tout autre paramètre de position) ne suffit pas pour caractériser une série statistique. Il est donc important de résumer une série statistique non seulement par des caractéristiques de tendance centrale, mais par des caractéristiques de dispersion. Un paramètre de dispersion est d'autant plus élevé que les valeurs de la série seront dispersées.
On en définira de deux sortes : celles liées à la moyenne : écart absolu moyen et écart type ; celles liées à la médiane : écart interquartile, écart interdécile, etc.

2.1 Étendue

Nous avons déjà défini l'étendue d'une série statistique. Ce paramètre est d'un intérêt limité car les valeurs extrêmes sont souvent accidentelles.

2.2 Quantiles

Quartiles : Pour une série statistique dont les valeurs sont classées par ordre croissant, la médiane partage la série des résultats en deux parties de même effectif. On peut aussi partager en 4 parties de même effectif. On obtient ainsi 3 valeurs Q_1, Q_2 et Q_3 (avec Q_1 [Q_2 [Q_3.) appelées quartiles. Q_2 est la médiane.
L'intérêt des quartiles consiste à diminuer l'importance des extrémités d'une série statistique.

On appelle *écart interquartile* le réel $Q_3 - Q_1$. Il représente l'étendue d'une partie de la série initiale centrée sur la médiane, et contenant 50% de l'effectif total.
Remarque : c'est un paramètre facile à déterminer. Il a l'avantage par rapport à l'étendue, d'écarter les valeurs accidentelles. Mais il ne renseigne pas sur la répartition des valeurs à l'intérieur de son domaine.

Déciles, centiles : Les déciles et les centiles partagent en dix ou cent parties d'effectifs égaux. On les utilise lorsque l'effectif total est élevé.
Les quartiles, déciles et centiles sont appelées *quantiles*.
La détermination graphique des quartiles est analogue à celle de la médiane.