CARACTERISTIQUES D'UNE SERIE STATISTIQUE

Dans ce chapitre, on considère des séries à caractères quantitatifs discrètes ou continues (avec, dans le cas d'une série continue, l'hypothèse d'une répartition uniforme à l'intérieur de chaque classe)

Notations :

- $x_1 , x_2 , \dots , x_p$ sont les valeurs ou les centres des classes si ces valeurs sont regroupées en classe.
- $n_1 , n_2 , \dots , n_p$ sont les effectifs respectifs des valeurs $x_1 , x_2 , \dots , x_p$
- $f_1 , f_2 , \dots , f_p$ sont les fréquences respectives des valeurs $x_1 , x_2 , \dots , x_p$
- N est l'effectif total : $N = n_1 + n_2 + \dots + n_p$

1) EXEMPLES

Suite à un contrôle de mathématiques , dans une classe de 28 élèves , le professeur s'est amusé à classer certains résultats dans deux tableaux :

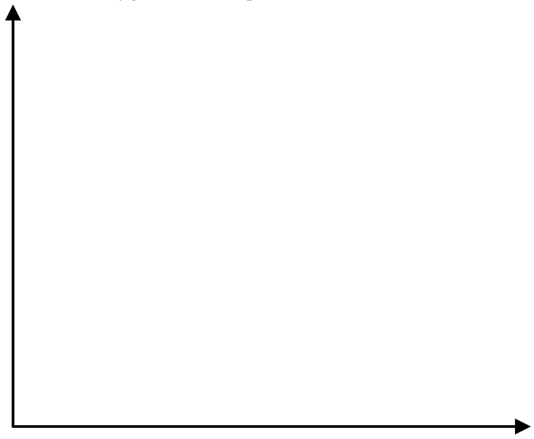
TABLEAU 1

notes	4	5	6	7	8	9	10	11	12	13	14	15	16	17
nb d'élèves (effectifs)	2	0	2	1	2	4	3	2	4	3	2	0	2	1
effectifs cumulés (croissants)														
effectifs cumulés (décroissants)														
fréquences														
fréquences cumulées (croissantes)														
fréquences cumulées (décroissantes)														
$n_i x_i$														
$n_i x^2_i$														

TABLEAU 2

temps de révision en h	[0;1[	[1;2[	[2;3[	[3;4[	[4;5[
Centres des classes					
effectifs	7	5	5	6	5
effectifs cumulés (croissants)					
fréquences en %					
fréquences cumulées (croissantes) en %					
$n_i x_i$					
$n_i x^2_i$					

Polygônes des fréquences cumulées croissantes



2) CARACTERISTIQUES DE POSITION D'UNE SERIE STATISTIQUE

A) CARACTERISTIQUES DE POSITION DE TENDANCE CENTRALE

- **Mode (ou classe modale)** : notation Mo
  Le mode, pour un caractère discret, est la valeur du caractère qui correspond au plus grand effectif.
  Pour un caractère continu, on parle de classe modale . Si les classes ont la même amplitude la classe modale est la classe qui correspond au plus fort effectif.
  Il peut y avoir plusieurs modes (ou plusieurs classes modales)
- **Médiane** : notation Me *(pour une série ordonnée par ordre croissant)*
  La médiane est une valeur Me du caractère qui partage la population en deux sous-ensembles de même effectif . Les éléments du premier sous-ensemble correspondent à des valeurs du caractère inférieures ou égales à Me, ceux du second correspondent à des valeurs du caractère supérieures ou égales à Me .
- **Moyenne** : notation $\overline{x}$

$$\overline{x} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_p x_p}{N} = \frac{1}{N}\sum_{i=1}^{p} n_i x_i \quad \text{ou} \quad \overline{x} = f_1 x_1 + f_2 x_2 + \dots + f_p x_p = \sum_{i=1}^{p} f_i x_i$$

Ex :

(tab1) :

Modes : Me = $\overline{x}$ =

(tab2) :

Classe modale : Me = $\overline{x}$ =

B) CARACTERISTIQUES DE POSITION NON CENTRALE : LES QUARTILES

On considère une série ordonnée par ordre croissant.

- **Le premier Quartile Q_1** d'une série statistique est la plus petite valeur de la série telle qu'au moins 25% des valeurs de celle-ci lui sont inférieures ou égales .
- **Le troisième Quartile Q_3** d'une série statistique est la plus petite valeur de la série telle qu'au moins 75% des valeurs de celle-ci lui sont inférieures ou égales .

Rem :
- Une série admet trois quartiles ; le deuxième, dont on ne fait pas usage en première, est associé à la valeur 50% .
- De nombreuses calculatrices considèrent les quartiles comme les médianes des deux séries obtenues après avoir partagé la série initiale par sa médiane … ce qui explique les différences constatées. Dans la pratique, ces différences ont peu d'importance vu la taille des séries.
- De la même façon, on peut définir les déciles d'une série statistique.

Les différentes situations sont regroupées dans le tableau ci-dessous :

Série discrète :
- Si $\frac{N}{4}$ est un entier, le premier quartile Q_1 est la valeur qui dans cette liste occupe le rang $\frac{N}{4}$ et le troisième quartile Q_3 est la valeur qui dans cette liste occupe le rang $\frac{3N}{4}$.
- Si $\frac{N}{4}$ n'est pas un entier, le premier quartile Q_1 est la valeur qui dans cette liste occupe le rang immédiatement supérieur à $\frac{N}{4}$ et le troisième quartile Q_3 est le valeur qui dans cette liste occupe le rang immédiatement supérieur à $\frac{3N}{4}$.

- **Ex** (tab 1) :

Série continue : (les valeurs sont regroupées par classe)
- Q_1 est la valeur correspondant à la fréquence cumulée croissante égale à 0,25 .
- Q_3 est la valeur correspondant à la fréquence cumulée croissante égale à 0,75 .

- **Ex** (tab 2) :

3) CARACTERISTIQUES DE DISPERSION D'UNE SERIE STATISTIQUE

A) ECART INTERQUARTILE

- **L'intervalle interquatile** d'une série statistique est l'intervalle [Q_1 ; Q_3]
- **L'écart interquatile** d'une série statistique est le nombre $Q_3 - Q_1$

- **Ex** (tab 1) : $Q_3 - Q_1 =$, e =
- **Ex** (tab 2) : $Q_3 - Q_1 =$, e =

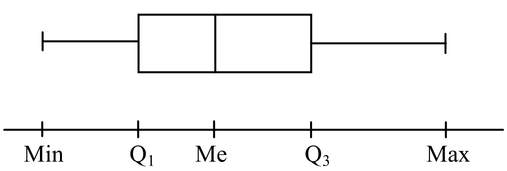
Rem :
- L'écart interquartile mesure la dispersion des valeurs autour de la médiane ; plus l'écart est petit, plus les valeurs de la série appartenant à l'intervalle interquartile sont concentrées autour de la médiane.
- Contrairement à **l'étendue** (notée e) qui mesure l'écart entre la plus grande et la plus petite valeur, l'écart interquartile élimine les valeurs extrêmes qui peuvent être douteuses, cependant il ne tient compte que de 50% de l'effectif …

On peut correctement résumer une série statistique par le couple : **(médiane ; intervalle interquartile)**

- **DIAGRAMME EN BOITES (boîtes à moustaches ou à pattes)**

Un diagramme en boite est un rectangle délimité par le premier quartile et le troisième quartile .

Pour l'obtenir, on trace un axe horizontal (ou vertical) sur lequel on place les valeurs de Q_1 , Q_3 et Me . L'un des côtés du rectangle a pour longueur l'écart interquartile, l'autre est quelconque. On complète ce diagramme en traçant deux traits horizontaux : l'un joignant Q_1 au minimun de la série et l'autre joigant Q_3 au maximun de la série.



B) VARIANCE ET ECART TYPE

- **La variance** V est la moyenne des carrées des écarts des valeurs x_i à la moyenne $\overline{x}$, c'est à dire :

$$V = \frac{1}{N}\left(n_1 (x_1 - \overline{x})^2 + n_2 (x_2 - \overline{x})^2 + \dots + n_p (x_p - \overline{x})^2 \right)$$
$$= f_1 (x_1 - \overline{x})^2 + f_2 (x_2 - \overline{x})^2 + \dots + f_p (x_p - \overline{x})^2$$

Ce nombre s'écrit aussi $V = \frac{1}{N}\sum_{i=1}^{p} n_i (x_i - \overline{x})^2 = \sum_{i=1}^{p} f_i (x_i - \overline{x})^2$

- **L'écart type** s est la racine carrée de la variance : $s = \sqrt{V}$

La fonction

$g : x \longmapsto \frac{1}{N}\sum_{i=1}^{p} n_i (x_i - x)^2$

posséde un minimum lorsque $x = \overline{x}$

Ce minimum est V .

Ex :

- (tab 1) : s = (tab 2) : s =

Rem :

- L'écart type est un paramètre plus fin que l'étendue, car il tient compte de la répartition des valeurs.
- L'écart type à la même unité que les valeurs de la série étudiée.
- L'écart type mesure la dispersion des valeurs de la série autour de la moyenne . Plus l'écart type est petit, plus les valeurs de la série sont concentrées autour de la moyenne.

- On peut correctement résumer une série statistique par le couple : **(moyenne ; écart type)**

Autre expression de la variance : (bien plus pratique)

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

La variance est égale à la moyenne des carrés des x_i diminuée du carré de la moyenne

Preuve :

On a $V = \frac{1}{N}\sum_{i=1}^{p} n_i (x_i - \overline{x})^2 = \frac{1}{N}\sum_{i=1}^{p} n_i (x_i^2 - 2x_i\overline{x} + \overline{x}^2) = \frac{1}{N}\sum_{i=1}^{p} n_i x_i^2 - \frac{1}{N}\sum_{i=1}^{p} 2n_i x_i \overline{x} + \frac{1}{N}\sum_{i=1}^{p} n_i \overline{x}^2$

Or $\frac{1}{N}\sum_{i=1}^{p} 2n_i x_i \overline{x} = \frac{2}{N}\overline{x}\sum_{i=1}^{p} n_i x_i = \frac{2}{N}\overline{x} \times N\overline{x} = 2\overline{x}^2$ et $\frac{1}{N}\sum_{i=1}^{p} n_i \overline{x}^2 = \frac{1}{N}\overline{x}^2 \sum_{i=1}^{p} n_i = \frac{1}{N}\overline{x}^2 \times N = \overline{x}^2$

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

4) INFLUENCE D'UNE TRANSFORMATION AFFINE DE DONNEES

On appelle image d'une série statistique S de valeurs $(x_1, x_2, \dots, x_p)$ ayant pour effectifs respectifs $(n_1, n_2, \dots, n_p)$ par une fonction f, la série statistique S' de valeurs $(f(x_1), f(x_2), \dots, f(x_p))$ ayant pour effectifs respectifs $(n_1, n_2, \dots, n_p)$

Soit S une série statistique, $f : x \longmapsto a x + b$ (avec $a \neq 0$) une application affine et S' l'image de S par f .
On a (avec les notations usuelles) :

$$\overline{x}' = a\overline{x} + b$$
$$\text{Si } a > 0 ,\ Q'_3 - Q'_1 = a\,(Q_3 - Q_1)$$
$$V' = a^2 V$$
$$s' = |a|\,s$$

- Si $f : x \longmapsto x + b$, alors la variance, l'écart type et l'écart interquartile sont inchangés .

- Ces propriétés permettent un changement d'origine (et/ou d'échelle) pour le calcul de la moyenne de l'écart interquartile et de l'écart type.

Preuve :

- On sait que la moyenne de la série S' est $\overline{x}' = a\overline{x} + b$

- Si $a > 0$, $Q'_1 = a Q_1 + b$ et $Q'_3 = a Q_3 + b$, on a alors :

$$a Q_3 + b - (a Q_1 + b) = a\,(Q_3 - Q_1)$$

- $V' = \frac{1}{N}\sum_{i=1}^{p} n_i (x'_i - \overline{x}')^2 = \frac{1}{N}\sum_{i=1}^{p} n_i (a x_i + b - a\overline{x} - b)^2 = a^2 \frac{1}{N}\sum_{i=1}^{p} n_i (x_i - \overline{x})^2 = a^2 V$
- Ainsi $s' = \sqrt{a^2 V} = |a|\,s$