

CARACTÉRISTIQUES D'UNE SÉRIE 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 : $N = n_1 + n_2 + \dots + n_p = \sum_{i=1}^p n_i$

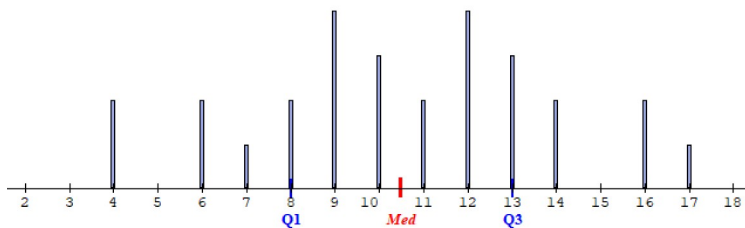
1) EXEMPLES

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 (série discrète)

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)	2	2	4	5	7	11	14	16	20	23	25	25	27	28
effectifs cumulés (décroissants)	28	26	26	24	23	21	17	14	12	8	5	3	3	1
fréquences	0,07	0	0,07	0,04	0,07	0,14	0,11	0,07	0,14	0,11	0,07	0	0,07	0,04
fréquences cumulées (croissantes)	0,07	0,07	0,14	0,18	0,25	0,39	0,5	0,57	0,71	0,82	0,89	0,89	0,96	1
fréquences cumulées (décroissantes)	1	0,93	0,93	0,86	0,82	0,75	0,61	0,5	0,43	0,29	0,18	0,11	0,11	0,04

Diagramme en bâtons



Boîte à moustaches

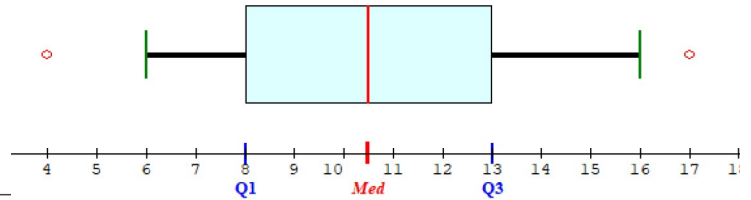
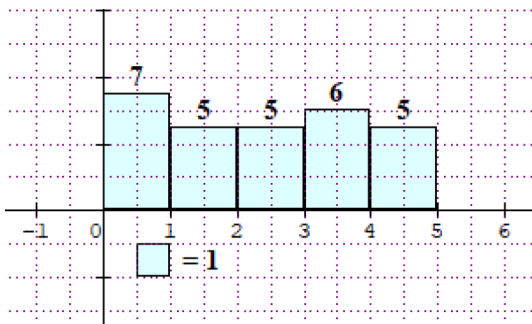


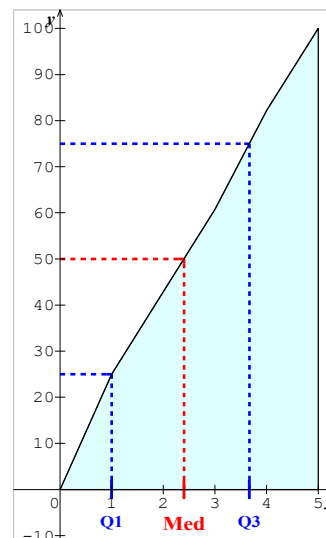
TABLEAU 2 (série continue)

temps de révision en h	[0;1[	[1;2[	[2;3[	[3;4[	[4;5[
centre des classes	0,5	1,5	2,5	3,5	4,5
effectifs	7	5	5	6	5
effectifs cumulés(croissants)	7	12	17	23	28
fréquences en %	25	17,9	17,9	21,4	17,9
fréquences cumulées(croissantes) en %	25	42,9	60,7	82,1	100

Histogramme



Polygone des fréquences cumulées croissantes



2) CARACTÉRISTIQUES DE POSITION D'UNE SÉRIE STATISTIQUE

A) CARACTÉRISTIQUES 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 $\bar{x}$

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

Exemples :

(tab1) :

(tab2) :

Modes : 9 et 12 $Me = 10,5$ $\bar{x} \approx 10,54$ (à 10^{-2} près)

Classe modale : $[0;1[$ $Me = 5$ $\bar{x} \approx 2,39$ (à 10^{-2} près)

Propriété : Linéarité de la moyenne

Soit a et b deux nombres réels.

Si la série (x_i) a pour moyenne $\bar{x}$, alors la série $(ax_i + b)$ a pour moyenne $\bar{x}' = a\bar{x} + b$

Preuve :

$$\bar{x}' = \frac{(ax_1 + b)n_1 + (ax_2 + b)n_2 + \dots + (ax_k + b)n_k}{N} = \frac{a(x_1 n_1 + x_2 n_2 + \dots + x_k n_k) + b(n_1 + n_2 + \dots + n_k)}{N}$$

$$\text{Ce qui donne } \bar{x}' = a \frac{(x_1 n_1 + x_2 n_2 + \dots + x_k n_k)}{N} + b \frac{n_1 + n_2 + \dots + n_k}{N} = a\bar{x} + b$$

Exemples : (tab1)

Si toutes les notes augmentent de 6, la nouvelle moyenne est alors 16,54

Si toutes les notes sont multipliées par 0,5, la nouvelle moyenne est alors 5,27

Si on multiplie toutes les notes par 0,5, puis on ajoute 6, la nouvelle moyenne est alors 11,27

B) CARACTÉRISTIQUES DE POSITION NON CENTRALE : LES QUARTILES

On considère une série ordonnée par ordre croissant.

Définitions :

- **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.

Remarques :

- 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 la valeur qui dans cette liste occupe le rang immédiatement supérieur à $\frac{3N}{4}$.

- **Exemple (tab 1) :**

$$Q_1 = 8$$

$$Q_3 = 13$$

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.

- **Exemple (tab 2) :**

$$Q_1 \approx 1,1$$

$$Q_3 \approx 3,7$$

3) CARACTÉRISTIQUES DE DISPERSION D'UNE SÉRIE STATISTIQUE

A) ÉCART INTERQUARTILE

Définitions :

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

- **Exemple (tab 1) :**

$$Q_3 - Q_1 = 5, e = 13$$

- **Exemple (tab 2) :**

$$Q_3 - Q_1 = 2,6, e = 5$$

Remarques :

- 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)

B) VARIANCE ET ÉCART TYPE (en anglais: standard deviation ... d'où la lettre s !)

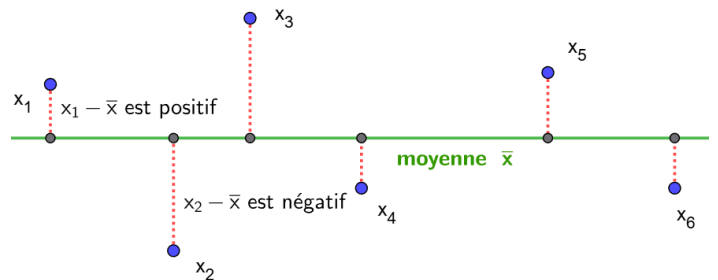
Pour chaque valeur x_i de la série, son « éloignement » de la moyenne peut se mesurer par la distance $|x_i - \bar{x}|$. La moyenne pondérée de toutes ces distances fournit un très bon paramètre de dispersion. On l'appelle écart absolu moyen. Mais puisque la valeur absolue ne se prête pas trop aux calculs, il n'y a pas d'application dans les résultats obtenus en statistique.

Définitions :

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

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

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



Exemples :

(tab 1) : $s \approx 3,35$ (à 10^{-2} près)

(tab 2) : $s \approx 1,45$ (à 10^{-2} près)

Remarques :

- L'écart type est un paramètre plus fin que l'étendue, car il tient compte de la répartition des valeurs.
- L'écart type a 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)

Propriété : Autre expression de la variance : (bien plus pratique)

La variance peut aussi être calculée par l'expression:

$$V = \frac{\sum_{i=1}^p n_i x_i^2}{N} - \bar{x}^2 \text{ (bien plus pratique)}$$

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

4) CROISEMENT DE DEUX VARIABLES QUALITATIVES

Définition :

- Un tableau croisé d'effectifs (ou tableau à double entrée) est un tableau donnant les effectifs portant sur deux caractères d'une même population.
Les valeurs de l'un des caractères sont représentées en ligne et celles de l'autre sont représentées en colonnes.
- L'**effectif marginal** pour une valeur d'un caractère est le nombre total d'individus présentant cette valeur du caractère.

Exemple :

Le tableau ci-dessous donne la répartition des élèves d'un lycée suivant leur sexe et leur classe.

	Seconde	Premières	Terminale	Total
Filles	321	325	336	982
garçons	332	327	329	988
Total	653	652	665	1970

Le cases grises
représentent les marges
du tableau

Il a 1970 élèves au total dans le lycée . Parmi eux, il y a 336 filles en classe de Terminale.
988 est l'effectif marginale des garçons du lycée.

Définitions :

- **La fréquence marginale** d'une valeur A d'un caractère est le quotient de l'effectif marginal de cette valeur par l'effectif total de la population. On a $f(A) = \frac{\text{Effectif marginal de } A}{\text{Effectif total}}$
En additionnant toutes les fréquences marginales d'un caractère on obtient 1.
- Soit A une valeur d'un des deux caractères et B une valeur de l'autre caractère.
La fréquence conditionnelle de la valeur A parmi la valeur B , notée $f_B(A)$, est le quotient du nombre d'individus présentant à la fois les valeurs A et B par l'effectif marginal de B .
On a : $f_B(A) = \frac{\text{effectif vérifiant à la fois } A \text{ et } B}{\text{effectif marginal de } B}$
En additionnant les fréquences conditionnelles parmi la valeur B de toutes les valeurs du premier caractère on obtient 1.

Exemple : On reprend l'exemple précédent.

La fréquence marginale des filles est $\frac{982}{1970} \approx 0,498$

La fréquence conditionnelle des premières parmi les garçons est $\frac{327}{988} \approx 0,33$