

Seconde	<u>Statistiques descriptives</u>	Année scolaire 2013/2014
---------	---	-----------------------------

Introduction :

L'un des principaux objectifs des statistiques descriptives consiste à rendre « lisibles » et exploitables des données souvent brutes en les résumant à l'aide de quelques caractéristiques.

I) Vocabulaire :

1) Population :

Ensemble de personnes ou d'objets ,appelés individus, définis par une propriété commune.

Exemples :

Les habitants d'une ville, les pièces mécaniques produites par une usine, les élèves d'une classe, etc...

2) Caractère :

Une population étant choisie, on peut étudier un élément qui les concerne : le caractère.

Les caractères peuvent être de deux types:quantitatifs ou qualitatifs.

a) Quantitatif :

Ce sont des caractères que l'on va pouvoir compter ou du moins quantifier.

Exemples :

- On s'intéresse à la population d'un village et on souhaite étudier le nombre d'enfants par foyer. (Population : habitants du village, Caractère : nombre d'enfants par foyer)
- Dans une usine de fabrication de pièces automobiles, on souhaite étudier le nombre de pièces défectueuses par jour de production. (Population : pièces automobiles produites, Caractère : Nombre de pièces défectueuses)

Un caractère quantitatif peut être continu (exemples : taille des élèves d'une classe, poids des enfants dans une crèche ,etc...) ou discret (exemples : nombre de frères et sœurs des élèves dans un collège, nombre d'écrans présents dans chaque maison d'un village, etc...)

b) Qualitatif :

Un caractère qu'on ne peut pas quantifier est dit qualitatif. Exemple : la couleur des cheveux.

Dans ce chapitre on ne travaillera qu'avec des caractères quantitatifs

II) Caractéristiques d'une série statistique :

1) Effectif et ECC :

a) Effectif :

On considère un caractère dont toutes les valeurs sont données sous la forme d'une série statistique. (on la note souvent (x_i))

Les différentes valeurs obtenues dans la série s'appellent les modalités de la série.

Le nombre de fois que chaque modalité apparaît dans la série s'appelle son effectif.

Exemple :

Voici les notes obtenues par les 36 élèves d'une classe de seconde :

12-9-7-15-2-16-20-11-11-9-5-16-5-2-3-13-15-8-9-18-16-15-15-1-2-18-16-15-14-11-10-7-14-13-10-16

On dit que ces données sont brutes.

On va « dépouiller » cette série en rangeant les notes et les effectifs correspondants dans un tableau :

Notes	1	2	3	5	7	8	9	10	11	12	13	14	15	16	18	20
Effectifs	1	3	1	2	2	1	3	2	3	1	2	2	5	5	2	1

Ensuite, on peut alors dresser le tableau des effectifs cumulés croissants (les ECC).

b) Les effectifs cumulés croissants : (ECC)

En partant de l'exemple précédent, pour obtenir les ECC, on ajoute les effectifs au fur et à mesure dans chaque colonne. Le dernier ECC obtenu doit être égal à l'effectif total.

Notes	1	2	3	5	7	8	9	10	11	12	13	14	15	16	18	20
Effectifs	1	3	1	2	2	1	3	2	3	1	2	2	5	5	2	1
ECC	1	4	5	7	9	10	13	15	18	19	21	23	28	33	35	36

Ce tableau permet de voir directement , par exemple, combien d'élèves ont au plus 10 : c'est-à-dire 15 élèves.

2) Fréquence et FCC :

a) Fréquence :

On considère un caractère dont les modalités sont : x_1, x_2, x_3 , etc... et les effectifs correspondants sont n_1, n_2, n_3 , etc...

On pose N = effectif total = $n_1 + n_2 + n_3 + \text{etc...}$

La fréquence d'apparition de la modalité x_1 notée f_1 est donnée par : $f_1 = \frac{n_1}{N}$

La fréquence d'apparition de la modalité x_2 notée f_2 est donnée par : $f_2 = \frac{n_2}{N}$

La fréquence d'apparition de la modalité x_3 notée f_3 est donnée par : $f_3 = \frac{n_3}{N}$, etc...

Remarques : - La fréquence ainsi définie est un nombre compris entre 0 et 1

- On donne parfois cette fréquence en pourcentages : il suffit de multiplier par 100 les fréquences obtenues précédemment.

Reprenons l'exemple précédent :

Notes	1	2	3	5	7	8	9	10	11	12	13	14	15	16	18	20
Fréquences	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{5}{36}$	$\frac{5}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

b) Fréquences cumulées croissantes (FCC) :

On procède de la même façon que pour les ECC.

Notes	1	2	3	5	7	8	9	10	11	12	13	14	15	16	18	20
Fréquences	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{5}{36}$	$\frac{5}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
FCC	$\frac{1}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{10}{36}$	$\frac{13}{36}$	$\frac{15}{36}$	$\frac{18}{36}$	$\frac{19}{36}$	$\frac{21}{36}$	$\frac{23}{36}$	$\frac{28}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	1

50 % des élèves ont une note inférieure ou égale à 11 (se voit directement)

3) Moyenne :

On considère un caractère x dont les modalités sont $x_1, x_2, x_3, \dots$ et les effectifs correspondants $n_1, n_2, n_3, \dots$, alors la moyenne des x_i notée $\bar{x}$ est donnée par :

$$\bar{x} = \frac{n_1 \times x_1 + n_2 \times x_2 + n_3 \times x_3 + \dots}{n_1 + n_2 + n_3 + \dots}$$

Exemple : En reprenant l'exemple des notes vu précédemment, on a :

$$\bar{x} = \frac{1 \times 1 + 2 \times 3 + 3 \times 1 + 5 \times 2 + \dots + 16 \times 5 + 18 \times 2 + 20 \times 1}{36} \approx \underline{\underline{11.08}}$$

Propriété :

On considère un caractère x dont les modalités sont $x_1, x_2, x_3, \dots$, les effectifs correspondants $n_1, n_2, n_3, \dots$, et les fréquences d'apparition correspondantes $f_1, f_2, f_3, \dots$

Alors, on a : $\bar{x} = f_1 x_1 + f_2 x_2 + f_3 x_3 + \dots$

Démonstration :

Si on pose $N = n_1 + n_2 + n_3 + \dots$ (effectif total)

$$\begin{aligned} \text{On a : } \bar{x} &= \frac{n_1 \times x_1 + n_2 \times x_2 + n_3 \times x_3 + \dots}{N} = \frac{n_1}{N} \times x_1 + \frac{n_2}{N} \times x_2 + \frac{n_3}{N} \times x_3 + \dots \\ &= f_1 \times x_1 + f_2 \times x_2 + f_3 \times x_3 + \dots \end{aligned}$$

4) Médiane :

Définition :

Si les modalités d'un caractère sont rangées dans l'ordre croissant, **la médiane** de ce caractère est une valeur qui partage l'ensemble des valeurs prises en deux groupes de même effectif.

Méthode :

On suppose que les modalités sont rangées dans l'ordre croissant.

Si on note N : l'effectif total.

Deux cas sont à envisager :

- Si N est impair :

La médiane de la série est la $\frac{N+1}{2}$ ième valeur (c'est donc une valeur de la série)

- Si N est pair :

On prend pour médiane la moyenne entre la $\frac{N}{2}$ ième valeur et la $(\frac{N}{2} + 1)$ ième valeur

Exemple :

Dans notre exemple, il y a 36 notes en tout :

1.2.2.3.5.5.7.7.8.9.9.9.10.10.11.11.11.12.13.13.14.14.15.15.15.15.15.16.16.16.16.16.18.18.20

18 notes

18 notes

On prend pour médiane la moyenne entre la 18^{ième} et la 19^{ième} note.

Donc : La médiane est égale à : $\frac{11+12}{2} = \underline{11,5}$

Remarque :

Pour déterminer la médiane, on peut également utiliser les ECC (voir les exercices)

5) Quartiles :

a) Premier quartile :

C'est la plus petite valeur du caractère telle qu'au moins 25% des valeurs lui sont inférieures.

Méthode :

- On calcule le quart de l'effectif total
- Ce dernier correspond au rang du premier quartile si les données sont rangées dans l'ordre croissant.

Remarque : Si l'effectif total n'est pas divisible par 4, on prend pour rang du premier quartile, l'entier qui suit le résultat trouvé.

Exemple : Supposons qu'il y ait 25 valeurs dans la série.

$\frac{25}{4} = 6,25$. Le premier quartile sera donc la 7^{ième} note de la série.

Calculons le premier quartile de la série des 36 notes :

- Calcul de la position : $\frac{36}{4} = 9$
- Le premier quartile est donc la 9^{ième} valeur : Donc $Q_1 = 7$

b) Troisième quartile :

C'est la plus petite valeur du caractère telle qu'au moins 75% des valeurs lui sont inférieures.

Méthode :

- On calcule les trois quarts de l'effectif total
- Ce dernier correspond au rang du troisième quartile si les données sont rangées dans l'ordre croissant.

Remarque : Si l'effectif total n'est pas divisible par 4, on prend pour rang du troisième quartile, l'entier qui suit le résultat trouvé.

Exemple : Supposons qu'il y ait 25 valeurs dans la série.

$25 \times \frac{3}{4} = 18,75$. Le troisième quartile sera donc la 19^{ième} note de la série.

Calculons le premier quartile de la série des 36 notes :

- Calcul de la position : $\frac{36}{4} \times 3 = 27$
- Le troisième quartile est donc la 27^{ième} valeur : Donc $Q_3 = 15$

6) Ecart et intervalle interquartile :

a) Ecart interquartile :

C'est l'écart entre le troisième et le premier quartile :

$$\text{Ecart interquartile} = Q_3 - Q_1$$

Dans l'exemple : Ecart interquartile = $15 - 7 = \underline{8}$

b) Intervalle interquartile :

C'est l'intervalle $[Q_1; Q_3]$.

Dans notre exemple : Intervalle interquartile : [8;15]

7) Etendue :

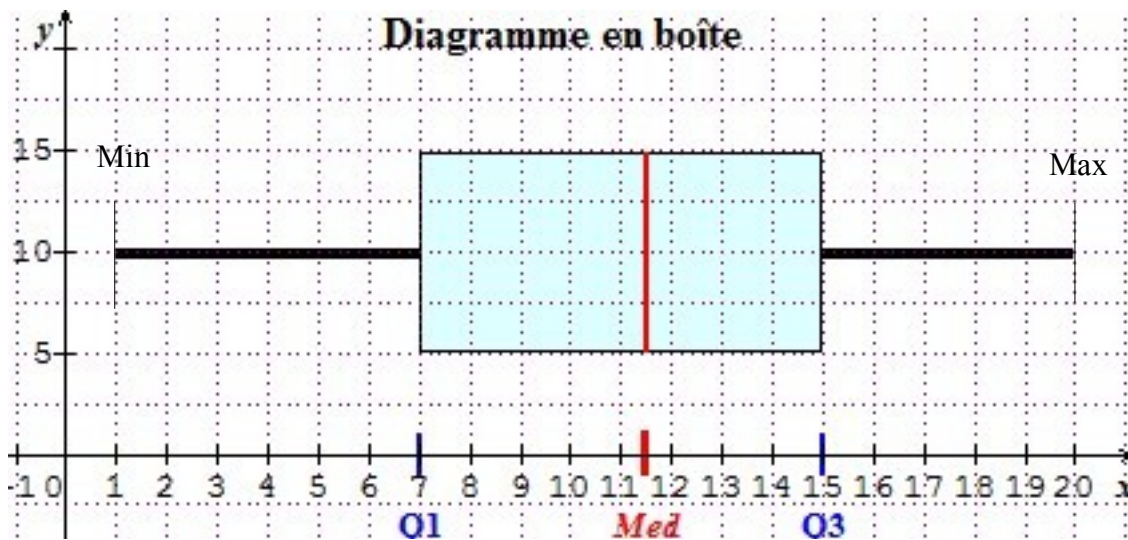
C'est l'écart entre la plus petite valeur du caractère et la plus grande.

Dans notre exemple :

$$\text{Etendue} = 20 - 1 = \underline{19}$$

Remarque :

On peut regrouper l'essentiel de ces paramètres dans un schéma appelé boîte à moustache ou diagramme en boîte ou diagramme de Tuckey :



III) Représentations graphiques :

1) Diagramme en bâtons :

2) Histogramme :

3) Polygone des ECC (ou des FCC) :