Chapitre XII

Statistique descriptive (1s)

Table des matières

I. Indicateurs de position (moyenne, médiane et quartiles) *2*

a. Moyenne 2

b. Médiane 3

c. Quartiles 3

II. Indicateurs de dispersion (étendue, écart interquartile et écart type) *4*

a. Étendue 4

b. Écart interquartile 4

c. Écart type 4

III. Avec la calculatrice *5*

L'objectif de ce cours est de vous donner les premiers outils permettant de synthétiser une série de valeurs (qui peut être très grande).
On utilisera deux types d'indicateurs : les indicateurs de position (moyenne, médiane et quartiles) et les indicateurs de dispersion (étendue, écart interquartile et écart type).

Le chapitre s'appuie sur la série du tableau ci-dessous qui présente le nombre de buts par match durant la Coupe du monde de football de 2018 :

Nombre de buts x_i	0	1	2	3	4	5	6	7
Nombre de matchs n_i	1	15	17	19	5	2	2	3
Effectifs cumulés	1	16	33	52	57	59	61	64

Les valeurs x_i du caractère étudié sont les "nombres de buts".
Les effectifs n_i correspondants sont les "nombres de matchs".

I. Indicateurs de position (moyenne, médiane et quartiles)

Lien vidéo : lien

a. Moyenne

Définition :
La moyenne $\bar{x}$ d'une série statistique dont les valeurs du caractère sont x_1, x_2, x_3, ..., x_k et les effectifs correspondants sont n_1, n_2, n_3, ..., n_k est égale à :

$$\bar{x} = \frac{x_1 \times n_1 + x_2 \times n_2 + \cdots + x_k \times n_k}{n_1 + n_2 + \cdots + n_k}$$

Exemple :
La moyenne de buts par match est égale à :

$$\bar{x} = \frac{1 \times 0 + 15 \times 1 + 17 \times 2 + 19 \times 3 + 5 \times 4 + 2 \times 5 + 2 \times 6 + 3 \times 7}{1 + 15 + 17 + 19 + 5 + 2 + 2 + 3} = \frac{169}{64} \approx 2{,}64$$

Propriétés :

- Moyenne pondérée de 2 groupes de valeurs :

On étudie un caractère dans une population partagée en deux sous-groupes d'effectifs respectifs n_1 et n_2.
Si la moyenne du caractère est $\overline{x_1}$ dans le premier sous-groupe et $\overline{x_2}$ dans le second, alors la moyenne du caractère dans la population entière est : $\bar{x} = \frac{n_1 \times \overline{x_1} + n_2 \times \overline{x_2}}{n_1 + n_2}$

- Linéarité de la moyenne :

On considère une série statistique de moyenne $\bar{x}$.
Si toutes les valeurs de la série sont multipliées par un même nombre a, alors la moyenne de la série est aussi multipliée par ce nombre a.
Si, à toutes les valeurs de la série, on ajoute un même nombre b, alors la moyenne est augmentée aussi de b.

Cela signifie que si à un DS, toutes les notes sont multipliées par 2, alors la moyenne de la classe à ce DS sera aussi multipliée par 2 et si le professeur décide de rajouter 3 points à tout le monde, alors la moyenne de la classe sera augmentée de 3 points aussi.

b. Médiane

Définition :
La médiane Me d'une série statistique de n valeurs <u>ordonnées</u> est :

- La valeur centrale si n est impair ;
- La demi-somme (moyenne) des deux valeurs situées « au milieu » si n est pair.

Remarques :

- Au moins la moitié des valeurs de la série est inférieur (ou supérieur) ou égal à la médiane.
- La médiane est peu sensible aux valeurs extrêmes.

En pratique, pour obtenir la médiane d'une série, on range les valeurs de la série dans l'ordre croissant. La médiane est la valeur qui partage la série en deux populations d'effectif égal.

<u>Exemple :</u>
Pour la série étudiée dans le chapitre, l'effectif total est égal à 64.
$\frac{64}{2} = 32$, donc la médiane se trouve entre la 32^{e} et 33^{e} valeur de la série.

On écrit les valeurs de la série dans l'ordre croissant :
0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 **2 2** 3 3 3 3 3 ....

La 32^{e} et la 33^{e} valeur sont égales à 2. La médiane est donc également égale à 2.
On en déduit que durant la Coupe du monde 2018, il y a eu autant de matchs dont le nombre de buts était supérieur ou égal à 2 que de matchs dont le nombre de buts était inférieur ou égal à 2.

Remarque :
Dans le cas d'un grand nombre de valeurs, il peut être plus judicieux d'utiliser la ligne des effectifs cumulées afin de retrouver la valeur cherchée.

Nombre de buts x_i	0	1	2	3	4	5	6	7
Nombre de matchs n_i	1	15	17	19	5	2	2	3
Effectifs cumulés	1	16	33	52	57	59	61	64

Nous avons ici la 1ère valeur seule

Nous avons ici de la 2ème à la 16ème valeur.

Nous avons ici de la 17ème à la 33ème valeur

c. Quartiles

Le <u>premier quartile</u> Q_1 est la plus petite valeur de la série telle qu'au moins 25% des valeurs sont inférieures ou égales à Q_1.
Le <u>troisième quartile</u> Q_3 est la plus petite valeur de la série telle qu'au moins 75% des valeurs sont inférieures ou égales à Q_3.

Exemple :
Pour la série étudiée dans le chapitre, l'effectif total est égal à 64.

Le premier quartile Q_1 est la 16[e] valeur. En effet, $\frac{1}{4} \times 64 = 16$.
Donc $Q_1 = 1$.
Le troisième quartile Q_3 est la 48[e] valeur. En effet, $\frac{3}{4} \times 64 = 48$.
Donc $Q_3 = 3$.

II. Indicateurs de dispersion (étendue, écart interquartile et écart type)

Lien vidéo : lien

a. Étendue

Définition :
L'étendue d'une série statistique est la différence entre la plus grande et la plus petite valeur de la série.

Ex. :
Ici, l'étendue de notre série est : 7-0=7.

b. Écart interquartile

Définition :
L'écart interquartile d'une série statistique de premier quartile Q1 et de troisième quartile Q3 est égal à la différence Q3 - Q1.

Ex. :
Pour la série étudiée dans le chapitre, l'écart interquartile est égal :
$Q_3 - Q_1 = 3 - 1 = 2$.

Remarques :
- L'écart interquartile d'une série mesure la dispersion autour de la médiane. Il contient **au moins 50% des valeurs** de la série.
- L'écart interquartile n'est pas influencé par les valeurs extrêmes de la série.

c. Écart type

Définition :
L'écart-type σ (sigma) est égal à :

$$\sigma = \sqrt{\frac{n_1 \times (x_1 - \bar{x})^2 + n_2 \times (x_2 - \bar{x})^2 + \cdots + n_k \times (x_k - \bar{x})^2}{n_1 + n_2 + \cdots + n_k}}$$

Ex. :
Pour la série étudiée dans le chapitre, l'écart type est égal à :

$$\sigma = \sqrt{\frac{1\times\left(0-\frac{169}{64}\right)^2+15\times\left(1-\frac{169}{64}\right)^2+17\times\left(2-\frac{169}{64}\right)^2+19\times\left(3-\frac{169}{64}\right)^2+5\times\left(4-\frac{169}{64}\right)^2+2\times\left(5-\frac{169}{64}\right)^2+2\times\left(6-\frac{169}{64}\right)^2+3\times\left(7-\frac{169}{64}\right)^2}{64}} \approx 1{,}56$$

L'écart-type possède la même unité que les valeurs de la série.
Ainsi pour la série étudiée, l'écart-type est environ égal à 1,56 buts.

Remarque :
L'écart-type exprime la dispersion des valeurs d'une série statistique autour de sa moyenne. Les valeurs extrêmes influencent l'écart-type.

III. Avec la calculatrice

Méthode : Déterminer les caractéristiques statistiques à l'aide d'une calculatrice

Déterminer la moyenne et l'écart-type de la série statistique étudiée dans ce chapitre.

Correction :
On saisit les données du tableau dans deux listes de la calculatrice :

TI-83 : Touche « stats » puis « 1:Edit ...»

Casio 35+ : Menu « STAT

On obtient :

L1	L2	L3	L4
0	1		
1	15		
2	17		
3	19		
4	5		
5	2		
6	2		
7	3		

On indique que les valeurs du caractère sont stockées dans la liste 1 et les effectifs correspondants dans la liste 2 (L1, L2,... s'obtiennent avec les touches numériques et non en tapant L puis 1 !):
TI-83 : Touche « stats » puis « CALC » et « Stats 1-Var ».
Stats 1-Var L1,L2

Casio 35+ : « CALC » (F2) puis « SET » (F6) :
1Var XList :List1
1Var Freq :List2
Puis touches « EXIT » et « 1VAR » (F1).

On obtient :

Stats 1-Var
$\overline{x}$ =2.640625
Σx=169
Σx²=603
Sx=1.577291103
σx=1.56492001
n=64