

## Chapitre 5 : Statistiques

### I. Introduction

Dans ce chapitre, on se propose de résumer une série statistique en **indiquant la répartition des données**. Deux questions peuvent être alors posées :

- Autour de quelle valeur centrale les données sont-elles réparties ?
- Quelle est l'importance de la dispersion des données autour de cette valeur centrale ?

Pour se faire, on utilisera un **paramètre de position** indiquant la tendance centrale de la série (médiane ou moyenne) et un **paramètre de dispersion** (écart interquartile ou écart type).

### II. Moyenne et écart-type

On considère une série statistique composée de  $p$  valeurs  $x_i$  d'effectifs  $n_i$ .

Valeurs	$x_1$	$x_2$	...	$x_p$	TOTAL
Effectifs	$n_1$	$n_2$	...	$n_p$	N

L'effectif total des valeurs de la série statistique est noté  $N = n_1 + n_2 + \dots + n_k = \sum_{i=1}^k n_i$ .

#### 1. Moyenne pondérée

**Définition** : La **moyenne pondérée** d'une série statistique est le nombre réel, noté  $\bar{x}$ , défini par :

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + n_3 x_3 + \dots + n_p x_p}{N} = f_1 x_1 + f_2 x_2 + \dots + f_p x_p$$

Exercice 1 : Le tableau ci-contre présente la répartition d'une chaîne de restaurants en fonction du nombre de ses employés :

Nombre d'employés	1	2	3	4	5	6	7
Effectifs	2	10	48	90	54	14	4

En détaillant vos calculs, déterminer la moyenne du nombre d'employés par magasin arrondi à 0,01 près.

Remarques :

- la moyenne permet de résumer une série statistique à l'aide d'un seul nombre
- la moyenne ne donne aucune information sur la dispersion des valeurs.

**Propriété:** Si on note  $f_i = \frac{n_i}{N}$  la fréquence de la valeur  $x_i$  alors  $\bar{x} = f_1 x_1 + f_2 x_2 + \dots + f_p x_p$  .

Démonstration :

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + n_3 x_3 + \dots + n_p x_p}{N} = \frac{n_1}{N} x_1 + \frac{n_2}{N} x_2 + \dots + \frac{n_p}{N} x_p = f_1 x_1 + f_2 x_2 + \dots + f_p x_p \quad \#$$

**Propriété: linéarité de la moyenne**

- Si toutes les valeurs d'une série statistique sont multipliées par un même réel  $a$  , sans changer les effectifs, alors la moyenne est également multipliée par  $a$  c'est à dire :

$$\text{Si } y_i = a x_i \text{ alors } \bar{y} = a \bar{x}$$

- Si on ajoute un même réel  $b$  à chacune des valeurs d'une série statistique, sans changer les effectifs alors la moyenne est également augmentée de la valeur  $b$  c'est à dire :

$$\text{Si } y_i = x_i + b \text{ alors } \bar{y} = \bar{x} + b$$

Remarque : on peut résumer les deux propriétés en une seule : si  $y_i = a x_i + b$  alors  $\bar{y} = a \bar{x} + b$  .

Démonstration :

- Notons  $y_i = a x_i$  d'effectif  $n_i$  . On a alors :

$$\bar{y} = \frac{n_1 a x_1 + n_2 a x_2 + n_3 a x_3 + \dots + n_p a x_p}{N} = \frac{a(n_1 x_1 + n_2 x_2 + \dots + n_p x_p)}{N} = a \frac{n_1 x_1 + n_2 x_2 + \dots + n_p x_p}{N} = a \bar{x}$$

- Notons  $y_i = x_i + b$  d'effectif  $n_i$  . On a alors :

$$\bar{y} = \frac{n_1(x_1 + b) + n_2(x_2 + b) + n_3(x_3 + b) + \dots + n_p(x_p + b)}{N} = \frac{(n_1 x_1 + n_2 x_2 + \dots + n_p x_p) + N b}{N}$$

$$\bar{y} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_p x_p}{N} + b = \bar{x} + b \quad \#$$

## 2. Variance et Écart-type

### Définitions :

- La **variance**, notée  $V$ , est égale à 
$$V = \frac{n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \dots + n_p(x_p - \bar{x})^2}{N}$$
- L'**écart-type** d'une série statistique est égal à  $\sigma = \sqrt{V}$

### Remarques :

- La variance et l'écart-type sont des nombres réels positifs.
- L'écart-type a la même unité que les  $x_i$ .
- L'écart-type mesure la dispersion des valeurs autour de la moyenne : plus l'écart-type est grand, plus les valeurs sont dispersées autour de  $\bar{x}$  et par conséquent moins la moyenne représente de manière significative la série statistique.
- Le couple (moyenne ; écart-type) permet de résumer les valeurs d'une série statistique. La moyenne est un caractère de position et l'écart-type est un caractère de dispersion.

Exercice 2 : Calculer l'écart-type de la série statistique définie à l'exercice 1, à 0,01 près.

Exercice 3 : Lors d'une rencontre sportive interacadémique d'athlétisme, on a relevé les âges des participants et obtenu le tableau suivant :

Age	15	16	17	18
Nombre d'athlètes	24	29	35	22

1. Déterminer l'écart-type de cette série statistique à l'aide de calculs détaillés.
2. Vérifier vos calculs à l'aide du mode STATS de votre calculatrice.

Exercice 4 : ci-dessous est écrit un programme en Python.

Compléter les lignes floutées 4, 8 et 9 puis expliquer la finalité de ces deux fonctions.

Remarque : la ligne 1 sert uniquement à importer l'instruction racine carrée « sqrt » de la bibliothèque « math » car elle n'est pas présente dans la bibliothèque de base de Python.

```

1 from math import sqrt
2
3 def moyenne(x1,x2,x3,x4,x5):
4     m = ((x1+x2+x3+x4+x5))/5
5     return(m)
6
7 def ecart(x1,x2,x3,x4,x5):
8     m = moyenne(x1,x2,x3,x4,x5)
9     v = (((x1-m)**2)+(x2-m)**2)+(x3-m)**2+(x4-m)**2+(x5-m)**2)/5
10    e=sqrt(v)
11    return(e)

```

## II. Médiane, quartiles et écart interquartile

### 1. Médiane

**Définition :** La médiane, notée  $Me$ , d'une série statistique est telle que :

- 50% au moins des individus ont une valeur du caractère inférieure ou égale à  $Me$
- 50% au moins des individus ont une valeur supérieure ou égale à  $Me$ .

**Définition :** Pour un caractère quantitatif discret, on peut calculer la médiane de 2 façons :

- On ordonne les valeurs de la série par ordre croissant
  - si la série est de taille  $2p + 1$ ,  $Me$  est la valeur du terme de rang  $p + 1$
  - si la série est de taille  $2p$ ,  $Me$  est la moyenne des valeurs des termes de rang  $p$  et  $p + 1$
- On construit le tableau des fréquences cumulées croissantes.
  - la médiane est la plus petite valeur dont la fréquence cumulée croissante dépasse 0,5.

Remarques :

- La médiane est une mesure de tendance centrale (comme la moyenne)
- Contrairement à la moyenne, la médiane n'est pas influencée par les valeurs extrêmes
- si l'effectif total est impair alors la médiane partage la série statistique en deux groupes de même effectif.

Exercice 5 : Un devoir commun de mathématiques pour 125 élèves de 1ère d'un lycée a donné les résultats suivants :

Note	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Effectifs cumulés	4	10	19	31	48	70	90	97	103	111	117	122	123	125

Déterminer rigoureusement la médiane  $Me$  de cette série statistique.

### 2. Quartiles

**Définition :**

- Le premier quartile d'une série statistique, noté  $Q_1$ , correspond au plus petit nombre de la série tel qu'au moins 25% des données soient inférieures ou égales à ce nombre.
- Le troisième quartile d'une série statistique, noté  $Q_3$ , correspond au plus petit nombre de la série tel qu'au moins 75% des données soient inférieures ou égales à ce nombre.

Remarques

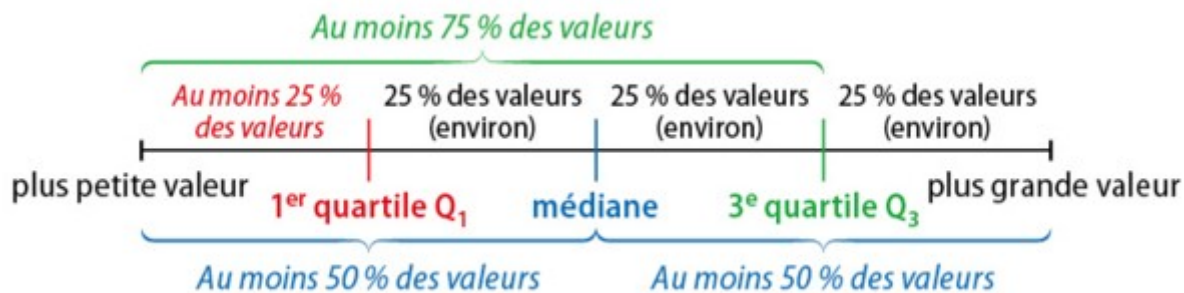
- Les quartiles et la médiane sont des **paramètres de position**
- Les quartiles et la médiane partagent la série en **4 groupes de (presque) même taille**.

**Définition :** On considère une série statistique dont les valeurs sont rangées par ordre croissant.

(1) Le rang de  $Q_1$  est le plus petit entier supérieur ou égal à  $\frac{N}{4}$ .

(2) Le rang de  $Q_3$  est le plus petit entier supérieur ou égal à  $\frac{3N}{4}$ .

En résumé :



Source : Indice Maths Seconde - Édition Bordas - Programme 2019

### 3. Écart interquartile

**Définition :**  $Q_3 - Q_1$  est l'écart interquartile

Remarques :

- L'écart interquartile est un paramètre de dispersion.
- $[Q_1 ; Q_3]$  est l'intervalle interquartile. Il contient environ 50 % des valeurs de la série statistique.

Exercice 6 :

1. Calculer rigoureusement  $Q_1$  et  $Q_3$  pour la série statistique de l'exercice 5
2. En déduire le pourcentage d'élèves ayant obtenu une note comprise entre 9 et 12.
3. Calculer l'écart interquartile

III. Étude d'une série statistique avec une calculatrice

Avec une calculatrice Casio	Avec une calculatrice Texas
<b>Saisie des données</b>	
<p>Accès au mode <b>statistique</b> : touche <b>MENU</b> puis sélectionner l'icône <b>STAT</b> puis <b>EXE</b>.</p> <p>Saisir les valeurs dans une liste, par exemple List 1, et les effectifs dans une autre liste, par exemple List 2.</p> <p>Utiliser la touche <b>EXE</b> pour passer d'une valeur à l'autre et les flèches pour se déplacer dans les listes.</p>	<p>Accès au mode <b>statistique</b> : touche <b>stats</b>.</p> <p>Choisir le menu <b>EDIT</b> puis sélectionner <b>1: Modifier</b> suivi de <b>entrer</b>.</p> <p>Saisir les valeurs dans une liste, par exemple L1, et les effectifs dans une autre liste, par exemple L2.</p> <p>Pour passer d'une valeur à l'autre ou pour changer de liste, se servir des flèches du curseur.</p>
<b>Calcul de paramètres statistiques – Séries avec effectifs</b>	
<p>Activer le menu <b>CALC</b> (touche <b>F2</b>), puis le menu <b>SET</b> (touche <b>F6</b>).</p> <p>Choisir List1 pour 1VarX List et List2 pour 1Var Freq : LIST s'obtient à l'aide de la touche <b>F2</b>, qu'on complète ensuite avec le numéro de la liste. Valider en appuyant deux fois sur la touche <b>EXE</b>.</p> <p>Sélectionner le menu <b>1-Var</b> (touche <b>F1</b>).</p> <p>Les paramètres s'affichent.</p> <p><math>\bar{x}</math> est la moyenne, <math>\Sigma x</math> est la somme des valeurs, <math>\sigma x</math> est l'écart-type et n est l'effectif total.</p> <p>On utilise les flèches du curseur pour avoir d'autres résultats.</p> <p>On peut alors lire la médiane (Med) les valeurs minimum (minX) et maximum (maxX) ainsi que les 1<sup>er</sup> et 3<sup>e</sup> quartiles (Q<sub>1</sub> et Q<sub>3</sub>).</p>	<p>À l'aide de la touche <b>stats</b>, choisir le menu <b>CALC</b>, puis sélectionner <b>1: Stats 1-Var</b>, suivi de <b>entrer</b>.</p> <p>Stats 1-Vars apparaît à l'écran.</p> <p>Compléter L1 pour Xliste et L2 pour ListeFréq (L1 et L2 s'obtiennent à l'aide des touches <b>2nde</b> <b>1</b> et <b>2nde</b> <b>2</b>). Puis sélectionner Calculer, puis <b>entrer</b>.</p> <p>Les paramètres s'affichent.</p> <p><math>\bar{x}</math> est la moyenne, <math>\Sigma x</math> est la somme des valeurs, <math>\sigma x</math> est l'écart-type et n est l'effectif total.</p> <p>On utilise les flèches du curseur pour avoir d'autres résultats.</p> <p>On peut alors lire la médiane (Med) les valeurs minimum (minX) et maximum (maxX) ainsi que les 1<sup>er</sup> et 3<sup>e</sup> quartiles (Q<sub>1</sub> et Q<sub>3</sub>).</p>
<b>Calcul de paramètres statistiques – Série brute</b>	
<p>Quand la série est une liste dont les effectifs sont tous 1, il suffit de saisir les valeurs dans une liste, par exemple List1 et de choisir « 1 » pour 1Var Freq.</p>	<p>Quand la série est une liste dont les effectifs sont tous 1, il suffit de saisir les valeurs dans une liste, par exemple L1, et de laisser vide la rubrique ListFreq.</p>
<b>Effacement des données</b>	
<p>Instruction QUIT (touches <b>SHIFT</b> <b>EXIT</b>).</p> <p>Placer le curseur sur une valeur de la liste à effacer, taper <b>F6</b> puis sélectionner <b>DEL-ALL</b> (touche <b>F5</b>).</p>	<p>Placer le curseur sur le nom de la liste à effacer, puis utiliser la touche <b>annul</b> suivi de <b>entrer</b>.</p>

 Les valeurs des quartiles ne sont pas nécessairement les mêmes que celles obtenues avec la méthode utilisée dans le cours.

Source : Indice Maths Seconde - Édition Bordas - Programme 2019

**IV. Étude d'une série statistique à l'aide d'un tableur**

**TP1 - Séries sans effectifs à l'aide d'un tableur**

Déterminer, à l'aide d'un tableur, les différents paramètres (moyenne, écart-type, médiane, 1<sup>er</sup> quartile, 3<sup>ème</sup> quartile, écart interquartile, minimum, maximum et étendue) de la série statistique suivante :

8 - 8 - 3 - 9 - 8 - 7 - 5 - 4 - 8 - 5 - 8 - 5 - 9 - 7 - 9 - 10

Solution

- On entre les données dans la colonne A du tableur (cellules A1:A16)
- Dans la colonne C, on écrit le caractère que le tableur va calculer dans la colonne D
- Dans la colonne D, on entre la formule correspondant au caractère souhaité (voir ci-dessous).

	A	B	C	D
1	8		Moyenne	=MOYENNE(A1:A16)
2	8		Écart-type	=ECARTYPEP(A1:A16)
3	3		Médiane	=MEDIANE(A1:A16)
4	9		1 <sup>er</sup> quartile	=QUARTILE(A1:A16;1)
5	8		3 <sup>e</sup> quartile	=QUARTILE(A1:A16;3)
6	7		Écart interquartile	=D5-D4
7	5		Minimum	=MIN(A1:A16)
8	4		Maximum	=MAX(A1:A16)
9	8		Étendue	=D8-D7
10	5			
11	8			
12	5			
13	9			

Source : Indice Maths Seconde - Édition Bordas - Programme 2019

On obtient les résultats suivants :

Moyenne	7,0625
Écart-type	1,98332896
Médiane	8
1 <sup>er</sup> quartile	5
3 <sup>e</sup> quartile	8,25
Écart interquartile	3,25
Minimum	3
Maximum	10
Étendue	7

Source : Indice Maths Seconde - Édition Bordas - Programme 2019

**TP2 - Simulation et calcul de la moyenne  $m$ , de l'écart-type  $\sigma$  et de la proportion en pourcentage des éléments appartenant à l'intervalle  $[m - 2\sigma; m + 2\sigma]$  à l'aide d'un tableur.**

Un jeu consiste à lancer six dés identiques équilibrés. On s'intéresse à la série statistique obtenue en calculant la somme des numéros inscrits sur la face supérieure de ces six dés.

1. Réaliser une simulation de 1000 lancers des six dés.
2. Calculer la moyenne  $m$  et l'écart-type  $\sigma$  de la série obtenue.
3. Déterminer le pourcentage des valeurs appartenant à l'intervalle  $[m - 2\sigma; m + 2\sigma]$

**Solution**

- Les colonnes A contient le n° du lancer. Il suffit d'étirer le 1 entré en A2 jusqu'à 1000.
- Les colonnes B à G du tableur contiennent les numéros des dés 1 à 6.
- On commence par simuler un lancer de dé en **B2=ALEA.ENTRE.BORNES(1;6)**
- On étire cette formule jusqu'en G2 pour obtenir un lancer des six dés.
- En H2, on calcule la somme des six faces supérieures des six dés. On entre la formule : **H2= SOMME(B2:G2)**
- On étire les formules B2 à G2 afin d'obtenir 1000 lancers.
- En K2 et en K3, on écrit : « Moyenne (m) » et « Ecart-type ( $\sigma$ ) »  
En L2 et en L3, on entre les formules **L2=MOYENNE(H2:H1001)** et **L3=ECARTYPEP(H2:H1001)**
- En K5, K6 et K7, on écrit : « Bornes de l'intervalle », «  $m - 2\sigma$  » et «  $m + 2\sigma$  »  
En L6 et L7, on entre les formules **L6=L2-2\*L3** et **L7=L2+2\*L3**
- Dans la colonne I, on va compter le nombre d'éléments appartenant  $[m - 2\sigma; m + 2\sigma]$  afin de pouvoir calculer la fréquence ds valeurs appartenant à cet intervalle. Pour cela, en I2, on entre la formule **I2=SI(ET(H2>=\$L6;H2<=\$L7);1,0)** afin de renvoyer la valeur 1 si le contenu de I2 est compris entre  $m - 2\sigma$  et  $m + 2\sigma$  et 0 sinon. On étire la formule entrée en I2 jusqu'à I1001
- Pour calculer la fréquence des valeurs appartenant à l'intervalle  $[m - 2\sigma; m + 2\sigma]$ , il suffit d'entrer **L9=SOMME(I2:I1001)/1000**.
- Pour effectuer une nouvelle simulation de 1000 lancers, il suffit d'appuyer sur la touche F9 du clavier ou bien d'utiliser la combinaison CTRL+Maj+F9 (sous Open Office).

	A	B	C	D	E	F	G	H	I	J	K	L
1	Lancer n°	Dé1	Dé2	Dé3	Dé4	Dé5	Dé6	Somme	Test			
2	1	1	4	4	3	3	1	16	1		Moyenne (m)	20,92
3	2	5	5	6	2	1	2	21	1		Ecart-type (s)	4,33
4	3	4	6	3	3	6	3	25	1			
5	4	3	2	2	3	6	1	17	1		Bornes de l'intervalle	
6	5	1	5	2	6	6	1	21	1		m - 2s	12,27
7	6	2	6	4	2	2	5	21	1		m + 2s	29,57
8	7	2	3	2	4	3	2	16	1			
9	8	6	2	2	3	3	1	17	1		Fréquence	95,5%
10	9	5	2	2	1	4	6	20	1		de valeurs dans l'intervalle	
11	10	4	4	4	4	3	2	21	1			
12	11	4	2	1	4	1	1	13	1			

Source : Indice Maths Seconde - Édition Bordas - Programme 2019