Seconde

Cours : Statistiques descriptives

I Série statistiques

Définition : ***Série statistiques***

Établir une **statistique**, c'est relever pour tous les « **individus** » d'une « **populations** » la valeur d'un « **caractère** ».

Définition : ***Série statistique***

Le caractère peut être • **qualitatif**, quand elle prend des valeurs non numériques ;
• **quantitatif**, quand elle prend des valeurs numériques.

Exemples :

Population	Individu	Caractère	Type du caractère
Lycée du BdA	Elève	Couleur des yeux	qualitatif
Classe	Elève	Note	quantitatif
Union européenne	pays	Nombre de commune	quantitatif

Pour la suite, on considère uniquement des séries statistiques à caractère quantitatif.

II Indicateur de position

II.1 La moyenne

Pour résumer une série statistique très importante (ou comparer deux séries statistiques), on peut calculer la moyenne.

Définition : ***Moyenne***

La **moyenne** d'une série statistique est le nombre réel qui pourrait remplacer toutes les valeurs de la série sans changer leur somme.

Exemples :

• Avec la série de données et les **effectifs correspondants** :

Valeurs	x_1	x_3	x_3	...	x_p
Effectifs	n_1	n_2	n_3	...	n_p

$$\text{Moyenne} = \overline{x} = \frac{n_1 \times x_1 + n_2 \times x_2 + \ldots + n_p \times x_p}{n_1 + n_2 + \ldots + n_p}$$

• Avec la série de données et les **fréquences correspondantes** :

Valeurs	x_1	x_3	x_3	...	x_p
Fréquences	f_1	f_2	f_3	...	f_p

$$\text{Moyenne} = \overline{x} = f_1 \times x_1 + f_2 \times x_2 + \ldots + f_p \times x_p$$

Remarque :
Si les données sont regroupées par classes, le calcul est le même que précédemment en prenant le centre de la classe comme valeur pour le calcul.

II.2 Les quartiles et la médiane

Définition : ***Médiane***

La **médiane** est un nombre réel qui sépare la **moitié** des valeurs les plus petites de la **moitié** des valeurs les plus grandes. La détermination de la médiane nécessite le classement des données :

- Si la série est de taille impaire (2n + 1), la médiane est la valeur du terme de rang n +1.
- Si la série est de taille paire (2n), la médiane est la moyenne des valeurs des termes de rang n et n +1.

Définition : ***Quartiles***

Les **quartiles** d'une série statistique sont les trois valeurs q_1, q_2 et q_3 du caractère qui partagent la population en quatre parties de même effectif.

- **Premier quartile** Q_1 **:** C'est la plus petite valeur de la série, telle qu'au moins 25% des données lui sont inférieures ou égales.
- **Deuxième quartile** Q_2 **:** C'est la médiane.
- **Troisième quartile** Q_3 **:** C'est la plus petite valeur de la série, telle qu'au moins 75% des données lui sont inférieures ou égales.

Pour calculer la **médiane** et les **quartiles**, il faut donc que les valeurs de la série soient **rangées dans l'ordre croissant** !

III Indicateurs de dispersion

Définition : ***Étendue***

L'étendue est la différence entre la valeur la plus grande et la valeur la plus petite du caractère étudiée.

L'étendue ne renseigne pas énormément sur la dispersion des valeurs puisque seules les deux valeurs extrêmes interviennent dans son calcul.

Définition : ***Variance et écart type***

Pour mesurer la dispersion des valeurs, on calcule la moyenne des carrés des écarts à la moyenne :

1) La **variance** est la moyenne des carrés des écarts de chaque valeur avec à la moyenne :

$$V = \frac{n_1(x_1-\bar{x})^2 + n_2(x_2-\bar{x})^2 + \ldots + n_p(x_p-\bar{x})^2}{N}$$

2) **L'écart type** est la racine carrée de la variance : $\sigma = \sqrt{V}$.

On utilise plus souvent l'écart type car il s'exprime dans la même unité que celle des valeurs de la série.

Plus l'écart type est proche de 0, plus les valeurs de la série sont **homogènes**, et à l'inverse, plus l'écart type est grand, plus les valeurs sont **hétérogènes**.

Définition : ***L'écart interquartile***

L'écart interquartile est (comme son nom l'indique) l'écart entre Q_1 et Q_3 : $Q_3 - Q_1$

L'écart interquartile permet donc de mesurer la distance qui sépare environ 50% des valeurs situées au centre de la série.

IV Étudier des séries statitiques

Pour étudier une série statistiques on utilisera, en fonction du regard que l'on souhaite porter sur la série, les indicateurs précédents :

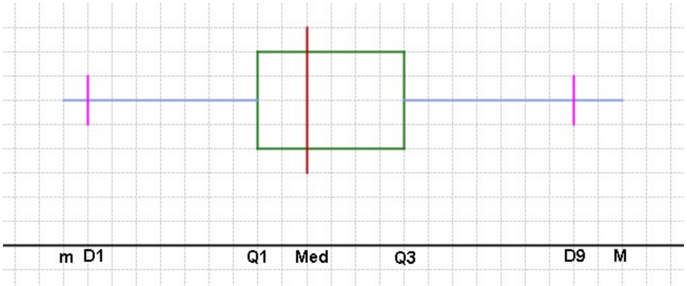
• Si l'on souhaite décrire un ensemble de valeurs centrales ou les valeurs les plus extrêmes, on utilisera les **quartiles** et le couple **médiane – écart interquartile** ;

• Si l'on souhaite décrire l'homogénéité ou l'hétérogénéité des valeurs, on utilisera le couple **moyenne – écart-type**.

Exemple 1 : *Exemple de représentation d'une série statistiques*

Pour décrire des ensembles de valeurs, il est d'usage de présenter cette observation sous une forme graphique après avoir calculer les **quartiles** et la **médiane**. On parle de :

- **diagramme en boîte** (de TUCKEY)
- ou **boite à moustaches**



Ici, on a aussi indiqué les **déciles** $D1$ (10% des valeurs) et $D9$ (90% des valeurs).

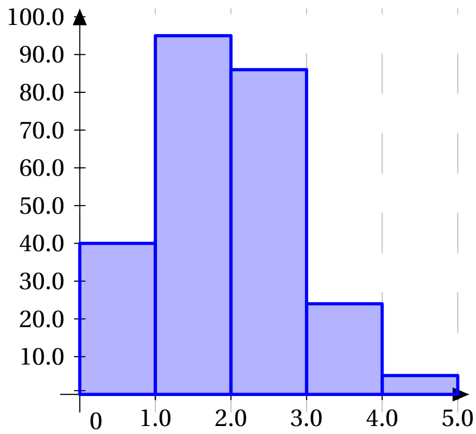
Exemple 2 : *Exemple d'étude statistique rangée par classes*

Une enquête sur le temps (en heures) de travail personnel quotidien des élèves de Seconde d'un lycée a donné les résultats suivants :

Les réponses ont été rangées dans 5 classes d'amplitude une heure.

Temps	$[0;1[$	$[1;2[$	$[2;3[$	$[3;4[$	$[4;5[$
Effectif	40	95	86	24	5

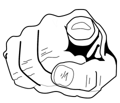
Il est alors possible de représenter ces données sous forme d'un **histogramme** :



Moyenne :

Même s'il n'est pas possible ici de connaître exactement la réponse de chaque élève, il est possible d'estimer la moyenne par calcul. L'effectif total est de 250 et on suppose que les réponses des élèves sont uniformément réparties dans chaque intervalle d'heure. On prend donc le **centre de chaque classe** :

Estimation de la moyenne : $\dfrac{40 \times 0.5 + 95 \times 1.5 + 86 \times 2.5 + 24 \times 3.5 + 5 \times 4.5}{250} = 1{,}936$ soit presque 2 h.



Dans l'exemple précédent, où se situent le premier et le troisième quartile ?

Exemple 3 : *Étude complète de deux séries statistiques*

Dans une même classe, les professeurs de français et de mathématiques comparent leur notes :

Notes	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Effectif en français	0	0	0	0	3	5	3	7	6	3	2	1	0	0	0
Effectif en maths	1	1	2	1	3	4	3	2	1	2	3	4	2	0	1

En français :

Moyenne : $= \dfrac{\dots\dots\dots\dots}{30} =$

Notes	Effectifs	$x_i - \overline{x}$	$(x_i - \overline{x})^2$	$n_i(x_i - \overline{x})^2$
3	0			
4	0			
5	0			
6	0			
7	3			
8	5			
9	3			
10	7			
11	6			
12	3			
13	2			
14	1			
15	0			
16	0			
17	0			
X	X	X	X	

Variance :

Écart-type :

Quartile 1 : $Q_1 =$

Médiane : $Q_2 =$

Quartile 3 : $Q_3 =$

Ecart interquartile : $Q_3 - Q_1 =$

En maths :

Moyenne : $= \dfrac{\dots\dots\dots\dots}{30} =$

Notes	Effectifs	$x_i - \overline{x}$	$(x_i - \overline{x})^2$	$n_i(x_i - \overline{x})^2$
3	1			
4	1			
5	2			
6	1			
7	3			
8	4			
9	3			
10	2			
11	1			
12	2			
13	3			
14	4			
15	2			
16	0			
17	1			
X	X	X	X	

Variance :

Écart-type :

Quartile 1 : $Q_1 =$

Médiane : $Q_2 =$

Quartile 3 : $Q_3 =$

Ecart interquartile : $Q_3 - Q_1 =$