CHAPITRE 10 : Statistiques descriptives

CHAPITRE 10 : Statistiques descriptives

1 Moyenne

1.1 Moyenne et moyenne pondérée

- Soit une série statistique de p valeurs distinctes $x_1, x_2, \dots . x_P$ et d'effectifs $n_1,\ n_2, \dots . n_P$

	x_1	x_2					x_p	
Valeurs x_i								TOTAL
Effectifs n_i								$n =$

L'effectif total est $n = n_1 + n_2 + \cdots + n_p$. La moyenne est

$$\boxed{\bar{x} = \frac{n_1x_1 + n_2x_2 + \cdots + n_px_p}{n}}$$

1.2 Linéarité de la moyenne

Si on multiplie chaque valeur par un réel a et qu'on additionne au résultat un réel b alors on obtient une deuxième série $ax_1 + b,\ ax_2 + b, \dots . ax_P + b$. Sa moyenne est $\boldsymbol{a\bar{x} + b}$.

2 Variance et écart-type

2.1 Variance

Deux professeurs M. X et M.Y ont donné les notes suivantes à un examen :

X: 7, 12, 11, 13, 8, 14, 15, 15, 14, 7

Y: 11, 12, 12, 11, 10, 13, 12, 12, 11, 12

On souhaite comparer les deux façons de noter. Remplir les tableaux d'effectifs suivants :

	x_1	x_2					x_p	
Valeurs x_i								TOTAL
Effectifs n_i								$n =$
n_ix_i								

Valeurs y_i					TOTAL
Effectifs n_i					$n =$
n_iy_i					

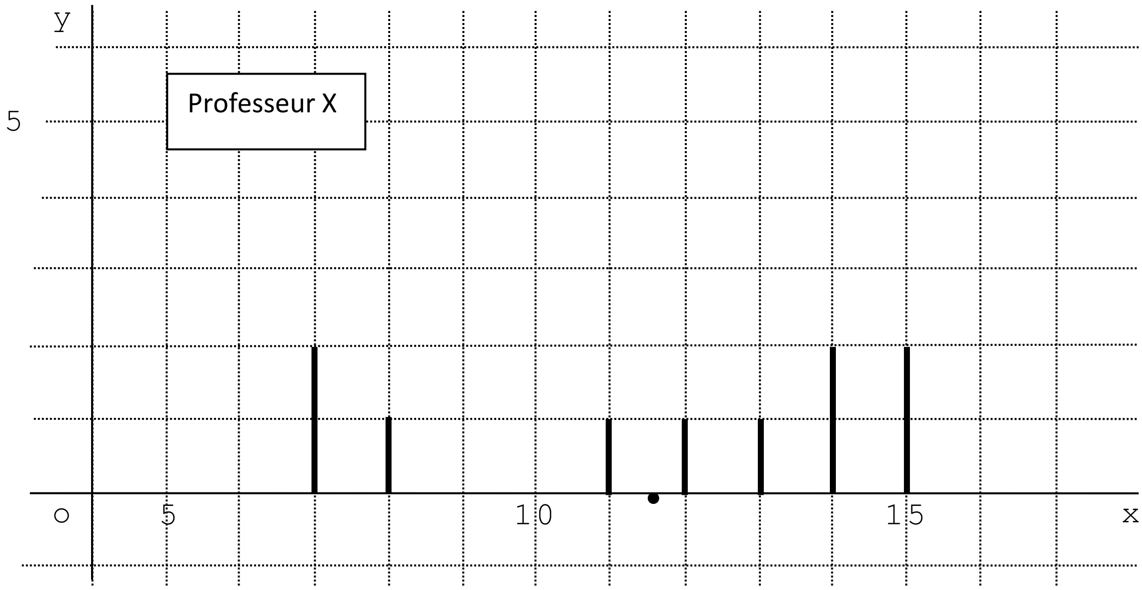
En utilisant la formule du calcul de la moyenne

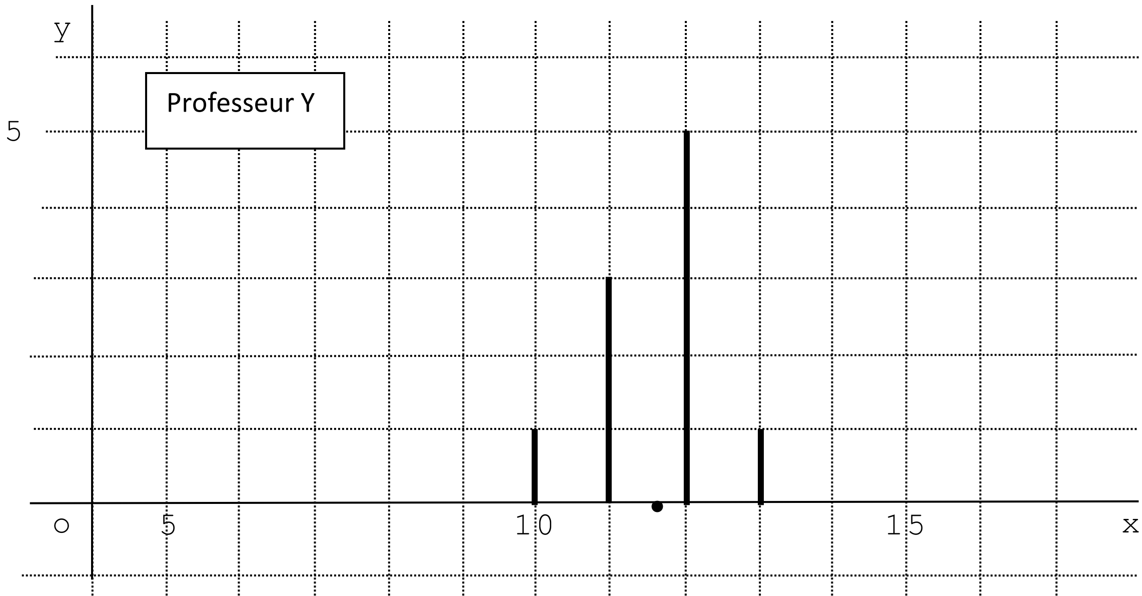
$$\bar{x} = \frac{n_1 \times x_1 + n_2 \times x_2 + \cdots + n_p \times x_p}{n}$$

calculer la note moyenne $\bar{x}$ pour le professeur X et la note moyenne $\bar{y}$ pour le professeur Y.

$$\bar{x} = \frac{\ldots \times \ldots + \ \ldots \times \ldots + \cdots + \ \ldots \times \ldots}{\ldots} = \cdots \qquad \bar{y} = \frac{\ldots \times \ldots + \ \ldots \times \ldots + \cdots + \ \ldots \times \ldots}{\ldots} = \cdots$$

Les deux moyennes sont les mêmes, mais on se rend compte au moyen d'un diagramme en bâtons que les façons de noter des deux professeurs ne sont pas les mêmes.





Bien que les deux séries aient la même position, visiblement le professeur Y note de façon moins dispersée. Comment calculer la dispersion ?

On peut calculer l'**étendue**

$$x_{max} - x_{min} =$$

$$y_{max} - y_{min} =$$

Mais il existe des indicateurs de dispersion qui tiennent compte de toutes les valeurs.

On peut penser à faire la moyenne des écarts de chaque valeur à la moyenne.

	x_1	x_2					x_p	
Valeurs x_i								TOTAL
Effectifs n_i								
$n_i(x_i - \bar{x})$								

$$\frac{n_1(x_1 - \bar{x}) + n_2(x_2 - \bar{x}) + \cdots + n_p(x_p - \bar{x})}{n} = \cdots$$

Les écarts négatifs compensent les écarts positifs.

- **La variance :** c'est la moyenne des carrés des écarts à la moyenne.

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

Exemple pour le professeur X :

$V_X = \ \ldots$

Inconvénient de la variance : son unité est celle du caractère étudié au carré.

2.2 Ecart type

C'est la racine carrée de la variance. Son unité est donc la même que celle du caractère étudié. On la note avec la lettre grecque σ (sigma) ou avec la lettre s

$$\sigma = \sqrt{V}$$

Pour le professeur X,

$\sigma_X = \cdots$

Exercice : calculer la variance V, l'écart type σ pour le professeur Y

3 Médiane, quartiles et écart interquartile

3.1 Quartiles

Il faut faire le tableau des effectifs cumulés croissants

Reprenons les notes du professeur X

X : 7, 12, 11, 13, 8, 14, 15, 15 14, 7

Valeurs x_i	7	8	11	12	13	14	15	TOTAL
Effectifs n_i	2	1	1	1	1	2	2	n = 10
Effectifs cumulés croissants	2	3	4	5	6	8	10	

la 3eme note est dedans

la 8eme note est dedans

- **La médiane** *M* est la valeur du caractère qui correspond à 50 % des observations. **Si l'effectif total *n* est impair**, la médiane *M* est la valeur centrale. **Si l'effectif total *n* est pair**, on convient que la médiane *M* est la moyenne des deux valeurs centrales.

- **Les quartiles** Q_1 et Q_3 sont les plus petites valeurs du caractère qui correspondent à **au moins** 25% et à au moins 75% des observations

3.2 Ecart interquartile

- **L'écart interquartile** est la différence $Q_3 - Q_1$

Exemple avec le professeur X :

3.3 Médiane

Comme n = 10 est pair :

$$M = \frac{5^e note + 6^e note}{2} \qquad M = \frac{12 + 13}{2} = 12{,}5$$

- **Premier quartile**

Il correspond à 25% des observations

$10 \times \frac{25}{100} = 2{,}5$

Donc la 3ème note est le premier quartile. $Q_1 = 8$

- **Troisième quartile**

Il correspond à 75% des observations

$10 \times \frac{75}{100} = 7{,}5$

Donc la 8ème note est le troisième quartile. $Q_3 = 14$

- L'écart interquartile est $Q_3 - Q_1 = 14 - 8 = \mathbf{6}$

L'écart interquartile est **un paramètre de dispersion**.

3.4 Résumé : comment trouver la médiane M, le 1er quartile Q_1 et le 3ème quartile Q_3

	Médiane M	1er quartile Q_1	3ème quartile Q_3
Série à 9 valeurs $\boxed{\frac{9+1}{2}}$	On fait le tableau des effectifs cumulés croissants. La médiane est la 5ème valeur	$\frac{25}{100} \times 9 = 2{,}25$ $\rightarrow Q_1$ *est la* 3*ème valeur*	$\frac{75}{100} \times 9 = 6.75$ $\rightarrow Q_3$ *est la* 7*ème valeur*
Série à 14 valeurs $\boxed{\frac{14}{2}}$	On fait le tableau des effectifs cumulés croissants. La médiane est la moyenne de la 7ème valeur et de la 8ème valeur	$\frac{25}{100} \times 14 = 3.5$ $\rightarrow Q_1$ *est la* 4*ème valeur*	$\frac{75}{100} \times 14 = 10.5$ $\rightarrow Q_3$ *est la* 11*ème valeur*

Série à 507 valeurs $\boxed{\frac{507+1}{2}}$	On fait le tableau des effectifs cumulés croissants. La médiane est la 254ème valeur	$\frac{25}{100} \times 507 = 126.75$ $\rightarrow Q_1\ est\ la\ 127ème\ valeur$	$\frac{75}{100} \times 507 = 380.25$ $\rightarrow Q_3\ est\ la\ 381ème\ valeur$
Série à 628 valeurs $\boxed{\frac{628}{2}}$	On fait le tableau des effectifs cumulés croissants. La médiane est la moyenne de la 314ème valeur et de la 315ème valeur	$\frac{25}{100} \times 628 = 157$ $\rightarrow Q_1\ est\ la\ 157ème\ valeur$	$\frac{75}{100} \times 628 = 471$ $\rightarrow Q_3\ est\ la\ 471ème\ valeur$