Séquence 8 : Statistiques à une variable

I) Généralités et représentations

Définition :

- L'ensemble sur lequel porte l'étude d'une série statistique s'appelle la population.
- Un élément de la population est un individu.
- L'objet étudié s'appelle le caractère de la série. Si le caractère prend des valeurs numériques, on dit qu'il est quantitatif. Sinon, il est qualitatif.
- Un caractère quantitatif peut être discret ou continu :
  → discret s'il prend des valeurs isolées ;
  → continu s'il peut prendre toute valeur dans un intervalle appelé aussi classe.
- La fréquence f d'une valeur est le quotient de l'effectif de cette valeur par l'effectif total.

$$f = \frac{\text{Effectif de la valeur}}{\text{Effectif total}}$$

Exemple :

Voici un tableau qui présente les salaires de l'entreprise A.

Salaire (en euros)	1500	1700	1900	2700	Total
Nombres d'employés	5	4	4	2	15

La fréquence des employés qui ont un salaire de 1700 € est $\frac{4}{15} \approx 0.27$ soit 27 % .

Diagramme bâton :

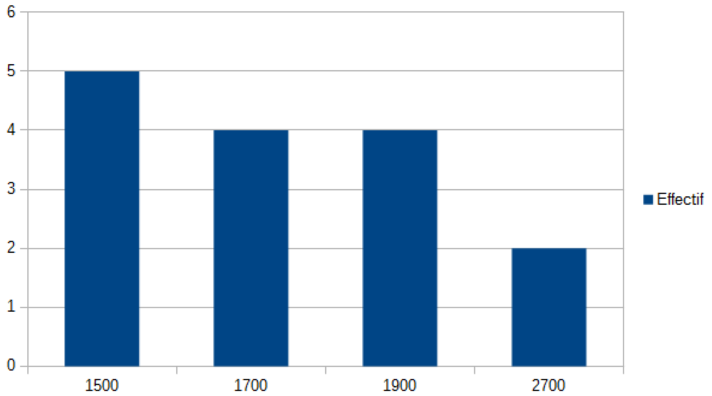
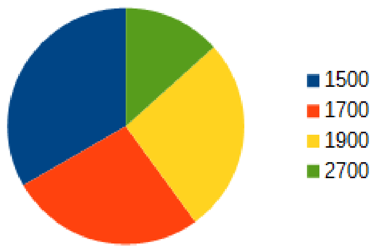


Diagramme circulaire :

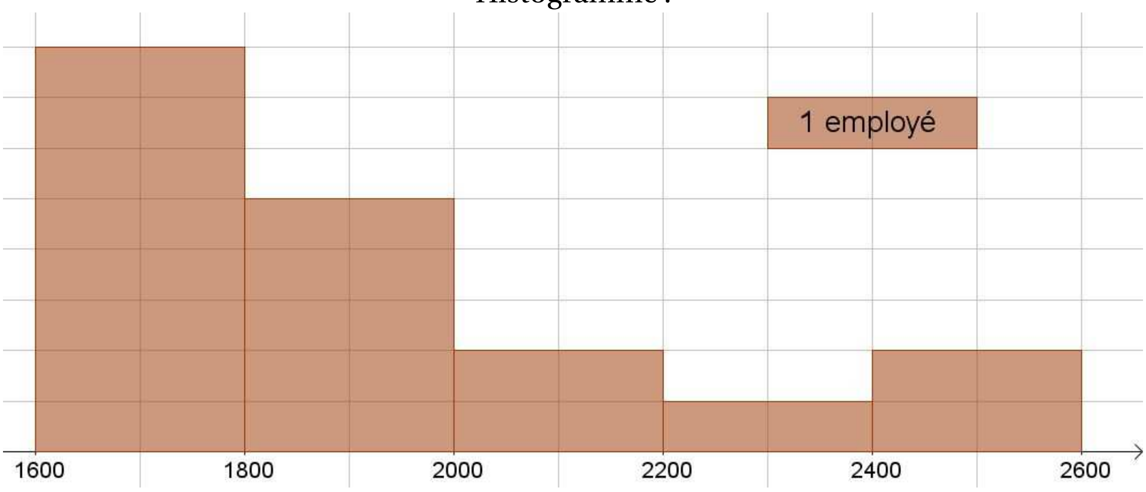


Voici un tableau qui présente les salaires de l'entreprise B.

Salaire (en €)	[1600 ; 1800[	[1800 ; 2000[	[2000 ; 2200[	[2200 ; 2400[	[2400 ; 2600[	Total
Effectif	5	8	2	1	2	18

Nous sommes ici dans le cas d'un caractère quantitatif continu (avec des intervalles appelés classes).

Histogramme :



II) Caractéristiques de position

1) Moyenne pondérée

On considère la série statistique donnée par le tableau ci-dessous. n_1, n_2, ... , n_p sont les effectifs.
On note $N = n_1 + n_2 + ... + n_p$ l'effectif total.

Valeur	x_1	x_2	...	x_p
Effectif	n_1	n_2	...	n_p
Fréquence	f_1	f_2	...	f_p

Définition :
La moyenne pondérée de la série est le nombre réel, noté $\overline{x}$, tel que :

$$\overline{x} = \frac{n_1 x_1 + n_2 x_2 + ... + n_p x_p}{N}$$

On peut calculer cette moyenne avec les fréquences et on a : $\overline{x} = n_1 f_1 + n_2 f_2 + ... + n_p f_p$

Exemple :
Reprenons l' exemple de l'entreprise A et déterminons la moyenne.

Salaire (en euros)	1500	1700	1900	2700	Total
Nombres d'employés	5	4	4	2	15

$$\overline{x} = \frac{1500 \times 5 + 1700 \times 4 + 1900 \times 4 + 2700 \times 2}{15} = 1820$$

Remarque 1 :
Dans le cas d'un caractère quantitatif continu (avec des intervalles appelés classes), pour le calcul de la moyenne, on doit utiliser comme valeur du caractère le centre de chaque intervalle.

Remarque 2 :
La moyenne est sensible aux valeurs extrêmes de la série statistiques.

Propriété :
On considère une série statistique dont la moyenne est $\overline{x}$.
Lorsque toutes les valeurs de la série sont transformées par une fonction affine $x \mapsto mx + p$, la moyenne de la nouvelle série est $m\overline{x} + p$

Exemple :
Reprenons l' exemple de l'entreprise A et déterminons la moyenne.
Le chef d'entreprise décide de multiplier par 1.20 les salaires des employés et d'ajouter une prime de 30 euros par mois.

Salaire (en euros)	1830	2070	2310	3270	Total
Nombres d'employés	5	4	4	2	15

Avant l'augmentation, la moyenne des salaires était de 1820 euros.

$\overline{x'} = 1.2 \times 1820 + 30 = 2214$

Après l'augmentation, la moyenne des salaires est de 2214 euros.

Remarque :
Si votre professeur multiplie par 2 toutes les notes d'un contrôle alors la moyenne sera multipliée par 2.
Si votre professeur ajoute 1 point à toutes les notes d'un contrôle alors la moyenne sera augmentée de 1 point.

2) La médiane

Définition :
La médiane d'une série de N valeurs rangées par ordre croissant est le nombre Me, tel que :
Au moins 50 % des valeurs de la série lui sont inférieures ou égales.
Au moins 50 % lui sont supérieures ou égales.

- Si N est impair, Me est la valeur centrale.
- Si N est pair, Me est la demi-somme des deux valeurs centrales.

Exemple 1 : (Effectif total impair)

Salaire (en €)	1500	1700	1900	2700	Total
Effectif	5	4	4	2	15
Effectifs cumulés croissants					

Illustration :

La 8^e valeur est 1700. La médiane est 1700.

Au moins 50 % des salaires sont inférieurs ou égaux à 1700 €.
Au moins 50 % des salaires sont supérieurs ou égaux à 1700 €.

Exemple 2 : (Effectif total pair)

Salaire (en €)	1600	1800	2000	2200	2280	Total
Effectif	7	2	2	2	1	14
Effectifs cumulés croissants						

Illustration :

La 7^e valeur est 1600. La 8^e valeur est 1800.

$$Me = \frac{1600 + 1800}{2} = 1700$$

La médiane est 1700.

Au moins 50 % des salaires sont inférieurs ou égaux à 1700 €.
Au moins 50 % des salaires sont supérieurs ou égaux à 1700 €.

Remarque :
Contrairement à la moyenne, la médiane ne tient pas compte des valeurs extrêmes de la série statistique.

3) Les quartiles

Définition :
Les valeurs d'une série statistique étant rangées par ordre croissant :

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

Exemple :
Déterminer le premier et le troisième quartile des salaires d'une entreprise donnés dans le tableau.

Salaire (en euros)	1500	1700	1900	2700	Total
Nombres d'employés	5	4	4	2	15

Premier quartile :
$\dfrac{N}{4} = \dfrac{15}{4} = 3.75 \approx 4$

Le premier quartile est la 4^e valeur. La 4^e valeur est 1500.
Au moins 25 % des valeurs sont inférieures ou égales à 1500 €.

Troisième quartile :
$\dfrac{3N}{4} = \dfrac{3 \times 15}{4} = 11.25 \approx 12$

Le troisième quartile est la 12^e valeur. La 12^e valeur est 1900.
Au moins 75 % des valeurs sont inférieures ou égales à 1900 €.

III) Caractéristiques de dispersion

1) Étendue et écart interquartile

Définition :
L'étendue est la différence entre la plus grande et la plus petite valeur d'une série.
L'écart interquartile est la différence $Q_3 - Q_1$ entre les troisième et premier quartiles.

Exemple :
Le tableau ci-dessous donne le salaire des employés d'une entreprise.
Déterminer l'étendue et l'écart interquartile.

Salaire (en euros)	1500	1700	1900	2700	Total
Nombres d'employés	5	4	4	2	15

Déterminons l'étendue :

Le salaire le plus élevé de l'entreprise A est 2700 et le salaire le plus bas est 1500
L'étendue des salaires de l'entreprise A est : 2700 - 1500 = 1200

Déterminons l'écart interquartile :

Premier quartile : $Q_1 = 1500$
Troisième quartile : $Q_3 = 1900$
Ecart interquartile : $Q_3 - Q_1 = 1900 - 1500 = 400$

L'écart interquartile est 400.

Remarque :
L'écart interquartile ne tient pas compte des valeurs extrêmes de la série, il mesure la dispersion autour de la médiane. Plus il est petit, plus les valeurs se concentrent autour de la médiane.

2) Variance et Écart-type

Définition :

- La variance de la série statistique est le nombre réel positif, noté V, tel que :

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

- L'écart-type est le nombre réel positif, noté σ tel que $\sigma = \sqrt{V}$.

Exemple :
On considère l'exemple précédent. Déterminer la variance et l'écart-type.

Salaire (en euros)	1500	1700	1900	2700	Total
Nombres d'employés	5	4	4	2	15

Rappel : La moyenne de cette série est 1820.

Déterminons la variance.

$$V = \frac{5 \times (1500 - 1820)^2 + 4 \times (1700 - 1820)^2 + 4 \times (1900 - 1820)^2 + 2 \times (2700 - 1820)^2}{15} = \frac{428800}{3} \approx 142933$$

La variance est de 142933.

Nous pouvons en déduire l'écart-type.

$\sigma = \sqrt{V} = \sqrt{142933} \approx 378$

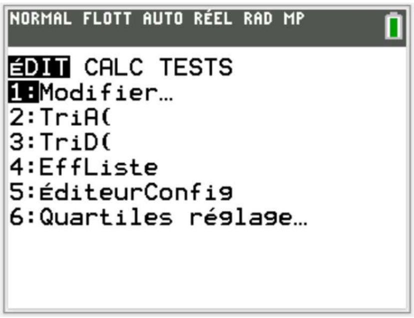
L'écart-type est d'environ 378.

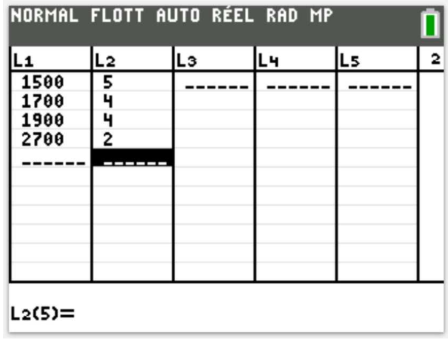
Remarque :
L'écart - type mesure la dispersion des valeurs autour de la moyenne : plus l'écart-type est petit, plus les valeurs se concentrent autour de la moyenne.

Fiche Calculatrice et Statistiques

Étape 1 : Cliquer sur Stats

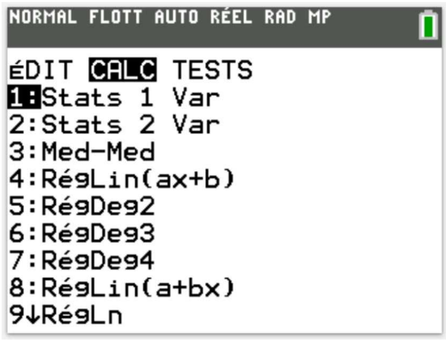
Étape 2 : Dans le sous menu EDIT , cliquer sur modifier puis saisir la série.





Étape 3 : Cliquer de nouveau sur Stats

Étape 4 : Aller dans le sous menu CALC puis Cliquer sur Stats 1 Var



Étape 5 : Compléter comme ci-dessous puis cliquer sur Calculer

Étape 6 : Vous obtenez les indicateurs de position et de dispersion souhaités.