Statistiques descriptives

Statistique descriptive, analyse de données Caractéristiques de dispersion : variance, écart-type. Diagramme en boîte.	• Utiliser de façon appropriée les deux couples usuels qui permettent de résumer une série statistique : (moyenne, écart-type) et (médiane, écart interquartile). • Étudier une série statistique ou mener une comparaison pertinente de deux séries statistiques à l'aide d'un logiciel ou d'une calculatrice.	On utilise la calculatrice ou un logiciel pour déterminer la variance et l'écart-type d'une série statistique. Des travaux réalisés à l'aide d'un logiciel permettent de faire observer des exemples d'effets de structure lors du calcul de moyennes.

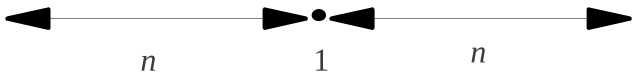
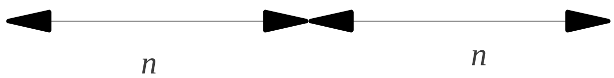
On étudie l'exemple suivant :
Les notes obtenues par une classe de 25 élèves lors d'un contrôle sont :
6-6-8-9-9-9-10-10-10-10-10-11-11-11-12-12-12-12-14-14-15-15-16-16-17
Ceci se traduit par le tableau suivant :

Notes	6	7	8	9	10	11	12	13	14	15	16	17
Effectifs	2	0	1	3	5	3	4	0	2	2	2	1
Fréquences	0,08	0	0,04	0,12	0,2	0,12	0,16	0	0,08	0,08	0,08	0,04

On se propose de résumer ces séries de nombres par des paramètres statistiques.

A. Médiane et quartiles

1- Médiane

> La médiane Me d'une série donne une valeur centrale de la série lorsque les données ont été classées dans l'ordre croissant.
>
> On distingue deux cas :
>
> - si l'effectif total est impair, égal à $2n+1$, alors la médiane est valeur de rang $n + 1$
>
> 
>
>
> - si l'effectif total est pair, égal à $2n$, alors la médiane est la moyenne entre les valeurs de rang n et $n+1$.
>
> 
>
>
> La médiane partage la population en deux parties de sorte que :
>
> - au moins 50% des individus correspondent à une valeur inférieure ou égale à la médiane.
>
> - au moins 50% des individus correspondent à une valeur supérieure ou égale à la médiane.

Exemple
Dans l'exemple donné, l'effectif total est de 25 élèves. La médiane est donc la note qui se trouve au 13ème rang, c'est à dire 11. Lorsque l'effectif total est important, il est intéressant d'utiliser les effectifs cumulés.

2- Quartiles

> Lorsqu'une série de données est rangée dans l'ordre croissant :
>
> - le 1er quartile est la plus petite valeur Q_1 telle qu'au moins 1/4 (ou 25%) des données de la liste sont inférieures à Q_1.
>
> - le 3ème quartile est la plus petite valeur Q_3 telle qu'au moins 3/4 (ou 75%) des données de la liste sont inférieures à Q_3.

Exemple

L'effectif total est 25.

$\frac{25}{4}=6,25$. C'est donc la 7ème note qui fournit Q1, soit Q1 = 10.

$\frac{3\times 25}{4}=18,75$. C'est donc la 19ème note qui fournit Q3, soit Q3 = 14

Utiliser les effectifs cumulés

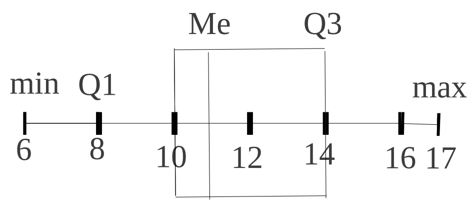
L'effectif cumulé d'une note N est le nombre de notes inférieures ou égales à N.

Notes	6	7	8	9	10	11	12	13	14	15	16	17
Effectifs	2	0	1	3	5	3	4	0	2	2	2	1
Effectifs cumulés	2	2	3	6	11	14	18	18	20	22	24	25

Comme Q1 est la 7ème note, elle correspond à un effectif cumulé de 11, donc Q1 = 10.
Comme Q3 est la 19ème note, elle correspond à un effectif cumulé de 20, donc Q3 = 14.

3- Écart inter-quartile et diagramme en boite

Pour évaluer la dispersion des données on peut utiliser l'écart inter-quartile qui est égal à Q3-Q1.
Le couple (médiane, écart inter-quartile fournit un bon résumé statistique.
On peut aussi résumer les données à l'aide d'un diagramme en boite (à moustache) qui montre le minimum, le 1er quartile, la médiane, le 3 ème quartile et le maximum (la population est partagée en 4 groupes d'effectifs équivalents)



B. Moyenne et écart type.

On considère une série statistique x_i, et on appelle n_i l'effectif associé au caractère x_i.

a) Effectif total

> L'effectif total de la série est $N = \sum n_i$

b) Moyenne

> La moyenne de la série est $\overline{x}=\frac{\sum n_i x_i}{N}$.

Remarque
La somme des écarts à la moyenne est égale à 0 : $\sum n_i(x_i-\overline{x})=0$, le total des valeurs supérieures à la moyenne est égal au total des valeurs inférieures à la moyenne.

c) Variance

> La variance sert à évaluer la dispersion des x_i autour de la moyenne. Elle donne la moyenne des carrés des écarts à la moyenne, elle est égale à $V=\frac{\sum n_i(x_i-\overline{x})^2}{n}$.

d) Écart-type

L'écart-type est la racine carrée de la variance; $\sigma=\sqrt{V}$. C'est une bonne unité pour mesurer l'écart à la moyenne.

Remarque
Le couple (moyenne, écart type) fournit un bon résumé statistique. Pour une distribution « normale » des données, on a environ 95% des données dans l'intervalle [$\overline{x}$ - 2 σ ; $\overline{x}$ + 2 σ] et 99% des données dans [$\overline{x}$ - 3σ ; $\overline{x}$ + 3σ]