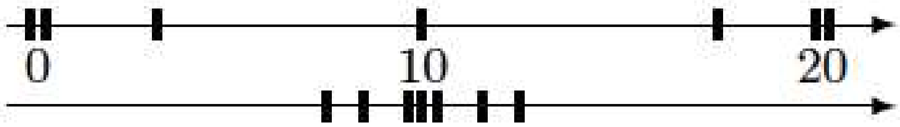
6. Paramètre de dispersion

Le principal inconvénient des mesures de position est qu'elles ne reflètent pas le niveau de dispersion des données, comme le montre l'exemple suivant.

Exemple : On considère la série des notes obtenues par deux élèves
Élève A : 20 ; 0 ; 17 ; 0 ; 20 ; 3 et 10.
Élève B : 10 ; 8 ; 10 ; 12 ; 9 ; 11 et 10.



Ces deux séries de notes présentent des moyennes et des médianes identiques, mais leurs profils diffèrent radicalement : les notes de A sont très dispersées, tandis que celles de B sont beaucoup plus regroupées.

Les notions présentées dans cette section ont pour objectif de quantifier la dispersion des données d'une série statistique

6. Paramètre de dispersion

6.3). L'étendue

L'étendue (ou amplitude) d'une série statistique correspond à la différence entre sa plus grande modalité et sa plus petite modalité. Elle se calcule ainsi : $e = x_{max} - x_{min}$

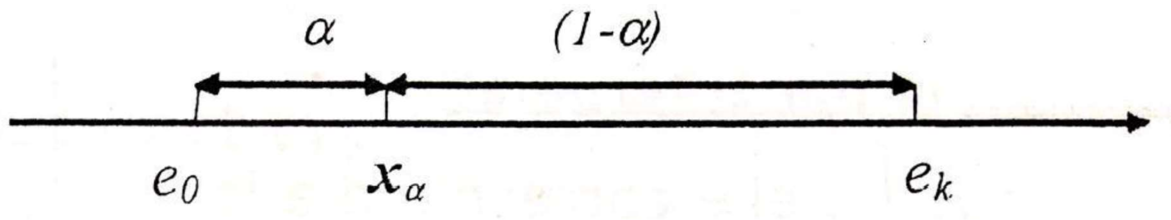
Exemple:

- Pour l'élève A, dont les notes sont {20, 0, 17, 0, 20, 3, 10}, l'étendue est calculée comme suit :$e = 20 - 0 = 20.$
- Pour l'élève B, dont les notes sont {10, 8, 10, 12, 9, 11, 10}, l'étendue est calculée comme suit :$e = 12 - 8 = 4.$

6. Paramètre de dispersion

6.3). Les quantiles

Le quantile d'ordre α (avec $0 \leq \alpha \leq 1$), noté X_α, d'une série statistique, est la valeur pour laquelle α représente la proportion des individus ayant une modalité inférieure ou égale à X_α.



On définie quatre types de quantiles :

- Quartiles
- Quintiles
- Déciles
- Centiles

6. Paramètre de dispersion

6.3). Les quantiles

a) Les quartiles : le premier quartile Q_1

Le premier quartile Q_1 est la valeur du caractère pour laquelle 25 % des observations sont inférieures et 75 % des observations sont supérieures ; $F(Q_1) = 0{,}25$.

- **Calcul du 1er quartile : variable quantitative discrète**

Si, pour tout i, $F_i \neq 0{,}25$, le premier quartile est la modalité x_i correspondant à la plus petite fréquence cumulée dépassant 0,25.

$\forall i, F_i(Q_1) \neq 0,25$
La plus petite valeur dépassant 0,25

x_i	n_i	ECC_i	f_i	F_i
0	2	2	0,18	0,18
1	3	5	0,27	0,45
2	4	9	0,36	0,81
3	2	11	0,18	1
Total	N=11	Au plus	≈ 1	Au plus

$\boxed{Q_1 = 1}$

6. Paramètre de dispersion

6.3). Les quantiles

a) Les quartiles : le premier quartile Q_1

- **Calcul du 1er quartile : variable quantitative continue**

Dans ce cas, le calcul de Q_1 nécessite d'abord de déterminer la classe contenant Q_1, notée $[e_{i-1};\ e_i]$, telle que : $F_{i-1} < 0{,}25 \leq F_i$, puis d'appliquer la formule suivante.

$$Q_1 = e_{i-1} + \frac{a_i}{f_i} \times [0{,}25 - F_{i-1}]$$

e_{i-1} : Borne inférieure de la classe Q_1. Et a_i : Amplitude de la classe.

Exemple : le tableau ci-dessous présente la répartition du personnel d'une entreprise en fonction de leur salaire mensuel en DH.

Classe contenant Q_1

Salaire mensuel en dhs	Effectifs n_i	ECC_i	Fréquences f_i	F_i	a_i
[1500 – 2000[	26	26	0,43	0,43	500
[2000 – 2500[	12	38	0,2	0,63	500
[2500 – 3000[	11	49	0,18	0,81	500
[3000 – 3500[	6	55	0,1	0,91	500
[3500 et plus[	5	60	0,08	1	500
Total	60	--	1	--	--

$$Me = 1500 + \frac{500}{0{,}43} \times (0{,}25 - 0) = 1790{,}7$$

$F_{i-1} = 0$ si c'est la première classe

01

Statistique descriptive a une dimension

6. Paramètre de dispersion

6.3). Les quantiles

b) Les quartiles : le troisième quartile Q_3

Le troisième quartile Q_3 est la valeur du caractère pour laquelle 75 % des observations sont inférieures et 25 % des observations sont supérieures ; $F(Q_3) = 0{,}75$.

Calcul du 1er quartile : variable quantitative discrète

Si, pour tout i, $F_i \neq 0{,}25$, le troisième quartile est la modalité x_i correspondant à la plus petite fréquence cumulée dépassant 0,75.

$\forall i, F_i(Q_3) \neq 0,75$
La plus petite valeur dépassant 0,75

x_i	n_i	ECC_i	f_i	F_i
0	2	2	0,18	0,18
1	3	5	0,27	0,45
2	4	9	0,36	0,81
3	2	11	0,18	1
Total	N=11	Au plus	≈ 1	Au plus

$\boxed{Q_3 = 2}$

6. Paramètre de dispersion

6.3). Les quantiles

b) Les quartiles : le troisième quartile Q_3

- **Calcul du troisième quartile : variable quantitative continue**

Dans ce cas, le calcul de Q_1 nécessite d'abord de déterminer la classe contenant Q_3, notée $[e_{i-1};\ e_i]$, telle que : $F_{i-1} < 0{,}75 \leq F_i$, puis d'appliquer la formule suivante.

$$Q_3 = e_{i-1} + \frac{a_i}{f_i} \times [0{,}75 - F_{i-1}]$$

e_{i-1}: Borne inférieure de la classe Q_1. Et a_i : Amplitude de la classe.

Exemple: Le tableau ci-dessous présente la répartition du personnel d'une entreprise en fonction de leur salaire mensuel en DH.

Salaire mensuel en dhs	Effectifs n_i	ECC_i	Fréquences f_i	F_i	a_i
[1500 – 2000[	26	26	0,43	0,43	500
[2000 – 2500[	12	38	0,2	0,63	500
[2500 – 3000[	11	49	0,18	0,81	500
[3000 – 3500[	6	55	0,1	0,91	500
[3500 et plus[	5	60	0,08	1	500
Total	60	--	1	--	--

Classe contenant Q_3

$$Me = 2500 + \frac{500}{0{,}18} \times (0{,}75 - 0{,}63)$$
$$= 2833{,}33$$

6. Paramètre de dispersion

6.3). Variance et écart type

La variance d'une série statistique $X = \{x_1, x_2, \dots, x_N\}$ est la moyenne arithmétique des carrés des écarts entre chaque valeur et la moyenne, définie par:

$$V(x) = \frac{1}{N}\sum_{i=1}^{k} n_i(x_i - \overline{x})^2$$

Ou encore :

$$V(X) = \frac{1}{N}\sum_{i=1}^{k} n_i x_i^2 - \overline{x}^2 = \sum_{i=1}^{k} f_i x_i^2 - \overline{x}^2$$

Où : N c'est le nombre d'observations de la variable X prennent k valeurs $x_1, x_2, \dots, x_k$, avec des effectifs $n_1, n_2, \dots, n_k$ où $n_1 + n_2 + \dots + n_k = N$

Dans le cas d'une variable continue, les x_i représentent les centres des classes $[x_i, x_{i+1}[$: $c_i = \frac{x_i + x_{i+1}}{2}$

6. Paramètre de dispersion

6.3). Variance et écart type

Remarque : Soient *a* et *b* deux nombres réels, on a:

$$V(X+b) = V(X)$$

$$V(aX+b) = a^2\, V(X)$$

Si une population P de taille n est constituée de m sous-populations $P_1, P_2, \dots, P_m$, de tailles respectives $n_1, n_2, \dots, n_m$, et de moyennes respectives $\overline{x}_1, \overline{x}_2, \dots, \overline{x}_m$, ainsi que de variances respectives $V_1, V_2, \dots, V_m$.

$$V(X) = \frac{1}{n}\sum_{i=1}^{m} n_i\, V_i + \frac{1}{n}\sum_{i=1}^{m} n_i\,(\overline{x}_i - \overline{x})^2$$

Avec $\overline{x} = \frac{1}{N}\sum_{i=1}^{1} n_i \overline{x}_i$ est la moyenne de la population P.

6. Paramètre de dispersion

6.3). Variance et écart type

L'écart type d'une série statistique $X = \{x_1, x_2, \ldots, x_N\}$, est la moyenne quadratique des écarts à la moyenne arithmétique, définie par:

$$\sigma_X = \sqrt{V(X)}$$

Remarque : L'écart type est exprimé dans la même unité que les observations, tandis que la variance est exprimée dans le carré de cette unité.

Exemple:

Considérons une série statistique ayant pour variance $V(X)$=22,67.

L'écart-type est :

$$\sigma_X = \sqrt{V(X)} = \sqrt{22{,}67} = 4{,}761$$

6. Paramètre de dispersion

6.3). Variance et écart type

Conclusion

- L'écart type (ou la variance) mesure la dispersion d'une série de données par rapport à sa moyenne.
- La variance et l'écart type prennent en compte toutes les valeurs d'une série statistique.
- Une faible variance (ou écart type) indique que les valeurs sont proches de la moyenne.
- Une variance (ou écart type) élevée signifie que les valeurs sont plus dispersées autour de la moyenne,
- La variance (ou l'écart type) est nulle uniquement lorsque toutes les valeurs sont identiques et égales à la moyenne.

6. Paramètre de dispersion

6.4). Coefficient de variation

Le coefficient de variation (noté $\boldsymbol{C_V}$) d'une série statistique $X = \{x_1, x_2, \dots, x_N\}$ est défini par la relation suivante :

$$C_V(X) = \frac{\sigma(X)}{\bar{X}}$$

Le coefficient de variation est une mesure de dispersion relative, car l'écart type est rapporté à la moyenne. Il est généralement exprimé en %.

Remarque

L'écart type seul ne permet généralement pas d'évaluer efficacement la dispersion des valeurs autour de la moyenne.

Par exemple, une distribution avec **une moyenne de 10 et un écart type de 1** (coefficient de variation de 10 %) sera beaucoup plus dispersée qu'une distribution avec **une moyenne de 1000 et un écart type de 10** (coefficient de variation de 1 %).

6. Paramètre de dispersion

Exemple d'application 1: Considérons la série statistique suivante (notes de Statistiques) :

x_i	n_i
9	1
10	3
11	2
12	4
13	1
14	1
15	2
16	1
Total	

Calculer l'étendu, la variance, l'écart type et le coefficient de variation.

6. Paramètre de dispersion

Exemple d'application 2: les classes de salaire (en dollar $) des employés d'une banque sont comme suit :

x_i	n_i
[30-40[	11
[40-50[	26
[50-60[	63
[60-70[	81
[70-80[	35
[80-90[	21
[90-100[	13
Total	

Calculer l'étendu, la variance, l'écart type et le coefficient de variation.

6. Paramètre de dispersion

Conclusion

Bien que les mesures de position, telles que la moyenne ou la médiane, soient utiles pour résumer un ensemble de données, elles ne suffisent pas à donner une image complète de leur structure. L'étude de la dispersion est donc indispensable pour mieux comprendre la variabilité des données. Les indicateurs comme la variance, l'écart type, et le coefficient de variation permettent d'évaluer cette dispersion, en offrant une vision plus fine de la répartition des valeurs autour de la moyenne. Cependant, ces mesures doivent être utilisées avec discernement, notamment lorsqu'elles sont influencées par des valeurs extrêmes. En combinant ces différents outils, on obtient une évaluation plus précise et équilibrée de la distribution des données.