SÉRIES STATISTIQUES

On considère une série statistique, par exemple une liste de notes d'élèves ou une liste de salaires d'employés d'une entreprise.

I – Quatre indicateurs de positions d'une série

On peut se demander s'il y a une « position centrale » autour de laquelle les valeurs de la série ont tendance à se concentrer. On peut prendre comme position centrale la moyenne de la série ou la médiane de la série. La médiane est une valeur qui partage la liste ordonnée des valeurs de la série en deux listes ayant à peu près le même effectif ; ce n'est pas forcément une valeur de la série et il y a parfois plusieurs choix possibles (dire *la* médiane est donc un abus de langage).

Les quartiles Q_1, Q_2 et Q_3 de la série sont des valeurs qui partagent la liste ordonnée des valeurs de la série en quatre sous-listes ayant à peu près le même nombre de valeurs (environ 25 % des valeurs). Q_1 est le 1er quartile de la série, Q_3 est le 3e quartile de la série, Q_2 est la médiane.

≈25 % ≈25 % ≈25 % ≈25 %

----------|----------|----------|----------> *liste ordonnée des valeurs de la série*

Q_1 Q_2 Q_3

II – Moyenne pondérée d'une série

Valeurs données avec leurs effectifs

Valeur	x_1	x_2	…	x_p
Effectif	n_1	n_2	…	n_p

$$\text{Moyenne} = \frac{n_1x_1 + n_2x_2 + \cdots + n_px_p}{n_1 + n_2 + \cdots + n_p}$$

Valeurs affectées de coefficients

Valeur	x_1	x_2	…	x_p
Coefficient	c_1	c_2	…	c_p

$$\text{Moyenne} = \frac{c_1x_1 + c_2x_2 + \cdots + x_px_p}{c_1 + c_2 + \cdots + c_p}$$

Valeurs données avec leurs fréquences

Valeur	x_1	x_2	…	x_p
Fréquence	f_1	f_2	…	f_p

$$\text{Moyenne} = f_1x_1 + f_2x_2 + \ldots + f_px_p$$

III – Linéarité de la moyenne

Propriété Quand on ajoute une même quantité à toutes les valeurs d'une série, la nouvelle moyenne s'obtient en ajoutant cette quantité à l'ancienne moyenne. Idem quand on multiplie par un même nombre.

IV – Trois indicateurs de dispersion d'une série

Plus un indicateur de dispersion d'une série est grand, plus les valeurs de la série sont dispersées, moins elles sont homogènes. Voici trois indicateurs de dispersion :

- l'étendue de la série (la différence entre la plus grande et la plus petite valeur) ;
- l'écart interquartile $Q_3 - Q_1$ de la série (il rend compte de la dispersion des 50 % des valeurs centrales autour de la médiane) ;
- l'écart type de la série que l'on calcule à l'aide d'une calculatrice.

V – Exemple

Retard (min)	0	2	4	6	8	10	15	20	25	120
Effectif	26	48	10	5	4	4	2	3	1	1

Calcul de la moyenne des retards : $\frac{26\times0+48\times2+10\times4+\cdots+1\times25+1\times120}{26+48+10+\cdots+1+1} \approx 4{,}6$

La moyenne des retards vaut donc environ 4 min et 36 s.

L'effectif total étant de $26 + 48 + 10 + \cdots + 1 + 1 = 111$, les quartiles s'obtiennent en partageant la liste des valeurs de la série en groupes de 28 valeurs ($4 \times 28 = 112$).

La dernière valeur du premier groupe (la 28e de la série) est 2, et la première du deuxième groupe (la 29e de la série) est aussi 2. Donc Q_1 vaut 2 min.

La dernière valeur du deuxième groupe (la 56e de la série) est 2, et la première du deuxième groupe (la 57e de la série) est aussi 2. Donc Q_2 vaut 2 min. Cette valeur est aussi la médiane car elle partage la série en deux sous-listes de taille à peu près égale.

La dernière valeur du troisième groupe (la 84e de la série) est 4, et la première du quatrième et dernier groupe (la 85e de la série) est 6. Donc on peut donner à Q_3 n'importe quelle valeur entre 4 et 6 min.

$120 - 0 = 120$, l'étendue vaut 120 min.

$Q_3 - Q_1 = 4 - 2 = 2$, l'écart interquartile vaut 2 min.

L'écart type vaut environ 11,9 minutes. Cette valeur élevée de l'écart type est surtout due au train en retard de 120 minutes. Sans ce train, l'écart type tomberait à 4,6 minutes.