

Statistiques à une variable — dispersion

2^{de}

Programme de seconde générale et technologique — BO spécial n°1 du 22 janvier 2019

Deux classes peuvent avoir exactement la même moyenne et n'avoir rien en commun : dans l'une, tout le monde est à 10 ; dans l'autre, la moitié est à 4 et l'autre à 16. Résumer une série par une seule valeur, c'est donc toujours mentir un peu. Le programme de seconde ajoute pour cette raison une seconde dimension — la **dispersion** — et deux outils pour la mesurer.

1 Résumer par une position

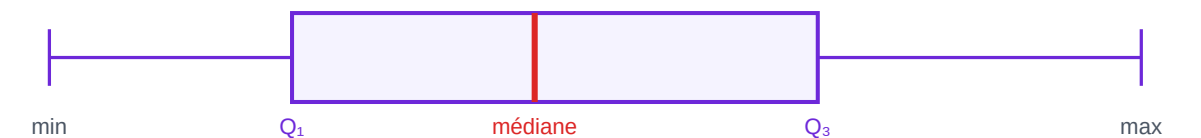
DÉFINITION

La **moyenne** $\bar{x}$ est la somme des valeurs divisée par leur nombre : elle répartit équitablement. La **médiane** partage la série ordonnée en deux effectifs égaux : autant de valeurs au-dessous qu'au-dessus. La moyenne se laisse tirer par une valeur extrême ; la médiane, non. C'est pourquoi on parle du **salaire médian** et non du salaire moyen.

DÉFINITION

Les **quartiles** partagent la série ordonnée en quatre parts d'effectifs égaux. Q_1 est la plus petite valeur telle qu'au moins 25 % des données lui soient inférieures ou égales ; Q_3 joue le même rôle pour 75 %. La différence $Q_3 - Q_1$ est l'**écart interquartile** : c'est l'étendue de la moitié centrale, débarrassée des valeurs extrêmes.

la boîte contient la moitié centrale des valeurs ; les moustaches vont aux extrêmes



la largeur de la boîte est l'**écart interquartile** : il mesure la dispersion sans les valeurs extrêmes.

Cinq nombres, un diagramme, et toute la répartition d'un coup d'œil.

2 Mesurer la dispersion

DÉFINITION

La **variance** V est la moyenne des carrés des écarts à la moyenne. On élève au carré pour deux raisons : empêcher les écarts positifs et négatifs de s'annuler, et donner plus de poids aux valeurs éloignées. L'**écart-type** σ est la racine carrée de la variance ; il a l'avantage décisif de s'exprimer dans la **même unité** que les données.

$$V = \frac{1}{n} \sum (x_i - \bar{x})^2 \quad \sigma = \sqrt{V}$$

EXEMPLE

Calculer la moyenne, la variance et l'écart-type de la série : 2, 4, 4, 4, 5, 5, 7, 9.

- ① La somme des huit valeurs.
 $2 + 4 + 4 + 4 + 5 + 5 + 7 + 9 = 40$
- ② La moyenne.
 $\frac{40}{8} = 5$
- ③ Les écarts à la moyenne sont $-3, -1, -1, -1, 0, 0, 2, 4$. On en prend les carrés et on les additionne.
 $9 + 1 + 1 + 1 + 0 + 0 + 4 + 16 = 32$
- ④ La variance est la moyenne de ces carrés.
 $\frac{32}{8} = 4$
- ⑤ L'écart-type en est la racine.
 $\sqrt{4} = 2$
les valeurs s'écartent en moyenne de 2 unités autour de 5.

$\bar{x} = 5, V = 4, \sigma = 2$

LE PIÈGE QUI COÛTE LE PLUS DE POINTS

Comparer une variance à une moyenne, ou l'annoncer avec l'unité des données. — la variance est une moyenne de **carrés** : si les données sont des centimètres, elle s'exprime en centimètres carrés. La comparer à la moyenne revient à comparer une aire à une longueur.

Le réflexe : toujours conclure sur l'**écart-type**, qui partage l'unité des données et se compare donc à la moyenne.

Écart-type et écart interquartile mesurent tous deux la dispersion, mais pas de la même façon. L'écart-type prend en compte **toutes** les valeurs, y compris les plus extrêmes, qu'il amplifie par le carré. L'écart interquartile les **ignore** délibérément. Sur une série contenant une valeur aberrante — une erreur de saisie, un cas exceptionnel — le second est bien plus fiable.

3 Comparer deux séries

UNE COMPARAISON EN DEUX TEMPS, JAMAIS UN SEUL

- ① **La position d'abord** : comparer deux moyennes, ou deux médianes — mais jamais une moyenne avec une médiane, qui ne mesurent pas la même chose.
- ② **La dispersion ensuite** : comparer deux écarts-types, ou deux écarts interquartiles.
- ③ **Conclure en une phrase qui dit les deux** : « la série B est en moyenne plus élevée, mais nettement plus dispersée ». une seule des deux informations ne répond jamais à la question posée.
- ④ **Interpréter dans le contexte** : une dispersion faible est un avantage pour une machine de production, un défaut pour un groupe d'élèves qu'on voudrait voir progresser.

À VÉRIFIER

- Ai-je **ordonné** la série avant de chercher médiane et quartiles ?
- La variance se calcule sur les **carrés** des écarts : ai-je élevé au carré avant d'additionner ?
- Ai-je conclu sur l'**écart-type**, dans l'unité des données ?
- Comparaison : ai-je donné la **position** et la **dispersion** ?
- Ai-je comparé des indicateurs de même nature entre eux ?
- Une valeur aberrante est-elle présente ? Alors l'écart interquartile est plus sûr.

À RETENIR

- **Moyenne** : répartis. **Médiane** : partage l'effectif en deux. La seconde résiste aux extrêmes.
- Q_1 et Q_3 découpent la série en quatre parts d'effectifs égaux.
- **Écart interquartile** $Q_3 - Q_1$: la dispersion de la moitié centrale.
- **Variance** = moyenne des carrés des écarts · **écart-type**
 $\sigma = \sqrt{V}$.
- L'écart-type est dans l'**unité des données** ; la variance ne l'est pas.
- Comparer deux séries = comparer une **position** ET une **dispersion**.
- Valeur aberrante : l'écart interquartile est plus fiable que l'écart-type.