

Statistiques — Moyenne, médiane, quartiles et dispersion

4^e

Programme du cycle 4 — BO spécial n°11 du 26 novembre 2015

Deux classes peuvent avoir **exactement la même moyenne** et n'avoir rien à voir : l'une groupée autour de 12, l'autre partagée entre 5 et 19. Un seul nombre ne décrit jamais une série. Ce chapitre ajoute ce qui manque : où se situe le milieu, et à quel point les valeurs s'écartent.

1 Moyenne et médiane ne disent pas la même chose

DÉFINITION

On multiplie chaque valeur par son **effectif**, on additionne le tout, et on divise par l'effectif total. C'est le point d'équilibre de la série : si toutes les valeurs étaient identiques, ce serait celle-là.

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_px_p}{n_1 + n_2 + \dots + n_p}$$

DÉFINITION

On range les valeurs **dans l'ordre croissant**, et on prend celle du milieu. Avec un effectif pair, la médiane est la demi-somme des deux valeurs centrales. La moitié de la série est en dessous, la moitié au-dessus.

neuf valeurs proches, une très grande : les deux indicateurs se séparent



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 quand on décrit une population : quelques très hauts revenus suffisent à déplacer la moyenne vers le haut sans que la situation du plus grand nombre ait changé. Choisir l'indicateur, c'est déjà interpréter.

2 Les quartiles : découper la série en quatre

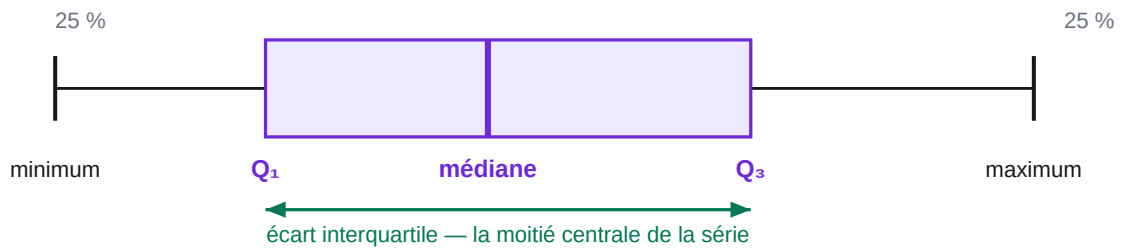
DÉFINITION

Une fois la série rangée, Q_1 est la plus petite valeur telle qu'**au moins un quart** des données lui soient inférieures ou égales ; Q_3 correspond aux **trois quarts**. Q_2 est la médiane. L'**écart interquartile** $Q_3 - Q_1$ contient la moitié centrale de la série : c'est la mesure de dispersion la plus robuste.

TROUVER Q_1 ET Q_3

- 1 Ranger les N valeurs dans l'ordre croissant. Sans cela, tout le reste est faux.
- 2 Calculer $\frac{N}{4}$. Si le résultat est entier, Q_1 est la valeur de ce rang ; sinon, on prend le rang **immédiatement supérieur**.
on arrondit toujours vers le haut : il faut atteindre le quart, pas s'en approcher.
- 3 Même chose avec $\frac{3N}{4}$ pour Q_3 .

le diagramme en boîte résume cinq nombres, et rien d'autre



La boîte contient la moitié des valeurs ; les moustaches, les extrêmes.

LE PIÈGE QUI COÛTE LE PLUS DE POINTS

Calculer la médiane sans avoir rangé la série. — la médiane est une position, pas un calcul. Prendre la valeur du milieu d'une liste en désordre donne un nombre qui ne signifie rien — et l'erreur ne se voit pas, puisque le résultat a l'air d'une réponse.

Le réflexe : réécrire la série rangée sur la copie, en entier. C'est deux minutes qui sécurisent toutes les questions suivantes.

Deux séries se comparent en deux temps : d'abord leur **position** — moyennes et médianes — puis leur **dispersion** — étendue et écart interquartile. Une réponse qui ne compare que les moyennes est incomplète : c'est précisément ce que ce chapitre existe pour dépasser.

3 Étendue et interprétation

DÉFINITION

La différence entre la plus grande et la plus petite valeur. C'est la mesure de dispersion la plus simple — et la plus fragile : **une seule valeur aberrante suffit à la tripler**, alors que l'écart interquartile ne bouge pas. Les deux se donnent ensemble, jamais l'une sans l'autre.

EXEMPLE

Deux séries de sept notes : A = 9 ; 10 ; 11 ; 12 ; 13 ; 14 ; 15 et B = 2 ; 11 ; 11 ; 12 ; 13 ; 13 ; 22.

- 1 Les deux ont la même médiane, 12, et des moyennes proches.
 $\frac{9+10+11+12+13+14+15}{7} = 12$
- 2 Étendue de A :
 $15 - 9 = 6$
- 3 Étendue de B :
 $22 - 2 = 20$
- 4 Pourtant l'écart interquartile de B vaut $13 - 11 = 2$, contre $14 - 10 = 4$ pour A : la moitié centrale de B est **plus groupée** que celle de A.

l'étendue et l'écart interquartile classent ces deux séries en sens inverse

DÉFINITION

Quand les valeurs sont trop nombreuses pour être listées — des tailles, des salaires, des durées — on les regroupe par **classes** : [150 ; 160[, [160 ; 170[... L'**histogramme** représente alors chaque classe par un rectangle dont l'**aire** est proportionnelle à l'effectif. C'est le diagramme des données continues.

REGROUPER UNE SÉRIE EN CLASSES

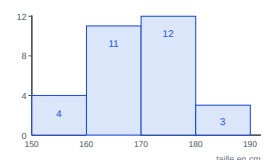
- 1 Repérer les **extrêmes** de la série, puis choisir des classes de même largeur qui les couvrent toutes.
- 2 Fixer la **convention de bord** et la garder : [150 ; 160[contient 150 mais pas 160, qui appartient à la classe suivante.
sans cette convention, un élève de 160 cm serait compté deux fois — ou zéro.
- 3 Compter les valeurs de chaque classe, et vérifier que la somme des effectifs redonne l'effectif total.

EXEMPLE

Les tailles de 30 élèves, regroupées en classes de 10 cm : [150 ; 160[: 4 élèves ; [160 ; 170[: 11 ; [170 ; 180[: 12 ; [180 ; 190[: 3.

- 1 Chaque rectangle a la largeur de sa classe — ici 10 cm pour toutes — et une hauteur proportionnelle à l'effectif. Les rectangles se touchent.
- 2 Contrôle : la somme des effectifs doit redonner le total.
 $4 + 11 + 12 + 3 = 30$
- 3 On perd l'information exacte — on ne sait plus qu'un élève mesurait 163 cm — et l'on gagne la forme de la répartition, invisible sur une liste de trente nombres.
c'est tout l'intérêt du regroupement, et sa seule perte.

la moitié des élèves se concentre entre 160 et 180 cm



Des rectangles jointifs : les classes se touchent parce que les valeurs sont continues.

EXEMPLE

Sur la série regroupée précédente, quelle est la taille moyenne des 30 élèves ?

- ① Les valeurs exactes sont perdues : on remplace chaque classe par son **centre** — 155, 165, 175, 185 — et l'on calcule une moyenne pondérée.
- ② Le numérateur.
 $4 \times 155 + 11 \times 165 + 12 \times 175 + 3 \times 185 = 5090$
- ③ On divise par l'effectif total.
 $\frac{5090}{30} \approx 169,7$
- ④ Ce résultat est une **valeur approchée** : la vraie moyenne des trente tailles exactes en diffère un peu. Le regroupement l'a rendue calculable, pas exacte.
plus les classes sont étroites, plus l'approximation est bonne.

la taille moyenne est d'environ 169,7 cm

LE PIÈGE QUI COÛTE LE PLUS DE POINTS

Dessiner un histogramme comme un diagramme en barres, avec des colonnes séparées et de même largeur. — les barres d'un diagramme en barres comptent des **catégories** distinctes — des couleurs, des matières — et se séparent. Les rectangles d'un histogramme couvrent des **intervalles** qui se touchent : un blanc entre deux classes signifierait qu'aucune valeur ne s'y trouve.

Le réflexe : des rectangles jointifs, et une largeur qui suit celle de la classe : une classe deux fois plus large se dessine deux fois plus large.

À VÉRIFIER

- Ai-je rangé la série avant de chercher médiane ou quartiles ?
- Ai-je divisé par la somme des effectifs, et non par le nombre de valeurs distinctes ?
- Le rang non entier a-t-il été arrondi vers le haut ?
- Ai-je comparé la dispersion, et pas seulement les moyennes ?

À RETENIR

- La **moyenne** est le point d'équilibre ; la **médiane**, la valeur du milieu.
- Ranger la série **avant** toute médiane ou tout quartile — sinon tout est faux.
- Une valeur extrême déplace la moyenne et laisse la médiane intacte.
- $Q_3 - Q_1$ est l'**écart interquartile** : la moitié centrale de la série.
- Rang non entier ? On arrondit **vers le haut**.
- Comparer deux séries, c'est comparer position **et** dispersion.
- Données groupées : on remplace chaque classe par son centre, et on parle d'estimation.