# Statistique à deux variables quantitatives

Fiche méthode

(Mathématiques complémentaires Terminale · Probabilités et statistique)

## La démarche pour étudier une série double

**1. Je regarde le nuage** : forme (alignée ou courbée), sens, points isolés, plage des abscisses.

**2. Je résume** : point moyen $G(\overline{x}\,;\overline{y})$, puis coefficient de corrélation $r$ pour juger l'alignement.

**3. J'ajuste** : droite des moindres carrés $y = ax + b$ si $|r|$ est proche de 1, sinon changement de variable puis retour à $y$.

**4. Je prévois et je critique** : interpolation ou extrapolation, réserve « si la tendance se poursuit », jamais de cause tirée de $r$.

**Le réflexe** — chaque résultat se donne avec ce qu'il signifie : « $r \approx 0{,}98$, donc les points sont presque alignés », « $x = 15$ est hors de la plage, c'est une extrapolation ». Un nombre seul ne rapporte pas tous les points.

## Cas 1 · Moyenne, variance, écart type

**Réflexe** — moyenne, écarts (somme nulle), carrés, variance = moyenne des carrés, écart type = racine ; changement d'unité y = ax + b : σ multiplié par |a|.

**Exemple.** Questions types : « Calculer la moyenne et l'écart type de la série ». Exemple : durées de quatre appels, en minutes : 3 ; 7 ; 8 ; 10.

$\overline{x} = \dfrac{28}{4} = 7$ min.

Écarts : $-4$ ; 0 ; 1 ; 3 ; carrés : 16 ; 0 ; 1 ; 9, de somme 26.

$V(x) = \dfrac{26}{4} = 6{,}5$ et $\sigma_x = \sqrt{6{,}5} \approx 2{,}55$ min.

**Attention** — la calculatrice donne σx (division par n) et sx (division par n − 1) : le cours utilise σx.

## Cas 2 · Nuage et point moyen

**Réflexe** — je décris le nuage (forme, sens, point isolé), puis G(x̄ ; ȳ) avec deux moyennes séparées.

**Exemple.** Questions types : « Représenter le nuage », « Calculer les coordonnées du point moyen ». Exemple : $x_i = 10$ ; 20 ; 30 ; 40 et $y_i = 3$ ; 5 ; 6 ; 10.

Le nuage est croissant, presque aligné, sans point isolé.

$\overline{x} = \dfrac{100}{4} = 25$ et $\overline{y} = \dfrac{24}{4} = 6$.

Point moyen : $G(25\,;6)$, qui n'est pas un point du nuage.

**Attention** — G n'est ni le milieu des points extrêmes, ni le point des médianes.

## Cas 3 · Droite des moindres carrés

**Réflexe** — a = Cov(x, y)/V(x), b = ȳ − a x̄ (ou calculatrice) ; je contrôle a x̄ + b = ȳ.

**Exemple.** Questions types : « Déterminer une équation de la droite de régression de $y$ en $x$ ». Exemple : $\overline{x} = 6$, $\overline{y} = 30$, $\mathrm{Cov}(x, y) = 15$, $V(x) = 4$.

$a = \dfrac{15}{4} = 3{,}75$ et $b = 30 - 3{,}75 \times 6 = 30 - 22{,}5 = 7{,}5$.

Droite : $y = 3{,}75x + 7{,}5$. Contrôle : $3{,}75 \times 6 + 7{,}5 = 30 = \overline{y}$.

À la calculatrice, on rédige : « une équation de la droite de régression de $y$ en $x$ est $y = 1{,}843x + 24{,}06$ » en recopiant les valeurs affichées avec les arrondis demandés.

**Attention** — on garde a et b non arrondis pour la suite ; la droite de x en y est une autre droite.

## Cas 4 · Coefficient de corrélation

**Réflexe** — r = Cov(x, y)/(σx σy) ; signe = sens du nuage ; |r| proche de 1 : ajustement affine justifié.

**Exemple.** Questions types : « Calculer $r$ et interpréter », « Peut-on parler de cause ? ». Exemple : $\mathrm{Cov}(x, y) = -12$, $\sigma_x = 4$, $\sigma_y = 3{,}5$.

$r = \dfrac{-12}{4 \times 3{,}5} = \dfrac{-12}{14} \approx -0{,}857.$

Nuage décroissant, tendance affine nette mais dispersion visible : l'ajustement est acceptable, avec prudence.

Cause : $r$ ne la prouve jamais ; on cherche une variable cachée ou un mécanisme.

**Attention** — r proche de 0 : pas de lien affine, mais un lien courbé reste possible.

## Cas 5 · Interpoler, extrapoler

**Réflexe** — plage [x min ; x max] ; dedans : interpolation ; dehors : extrapolation « si la tendance se poursuit ».

**Exemple.** Questions types : « Estimer $y$ pour… », « À partir de quand… ? ». Exemple : données pour $x \in [5\,;15]$, ajustement $y = -0{,}75x + 58$.

Pour $x = 9$ : $y = -0{,}75 \times 9 + 58 = 51{,}25$ ; $9 \in [5\,;15]$, c'est une interpolation.

Pour $x = 20$ : $y = -0{,}75 \times 20 + 58 = 43$ ; 20 est hors de la plage, c'est une extrapolation, valable seulement si la tendance se maintient.

**Attention** — diviser une inéquation par a < 0 change son sens ; pour des années, on donne le premier entier qui convient.

## Cas 6 · Changement de variable

**Réflexe** — nuage courbé : z = ln y, z = 1/x ou ln y en ln x ; |r| proche de 1 pour la nouvelle série ; retour à y.

**Exemple.** Questions types : « On pose $z = \ln y$… En déduire $y$ en fonction de $x$ ». Exemple : la régression donne $z = -0{,}12x + 4{,}5$.

$\ln y = -0{,}12x + 4{,}5$, donc $y = \mathrm{e}^{-0{,}12x+4{,}5} = \mathrm{e}^{4{,}5} \times \mathrm{e}^{-0{,}12x}$.

$\mathrm{e}^{4{,}5} \approx 90{,}0$ : $y \approx 90{,}0\,\mathrm{e}^{-0{,}12x}$.

À chaque pas, $y$ est multiplié par $\mathrm{e}^{-0{,}12} \approx 0{,}887$.

**Attention** — e^(αx + β) = e^β × e^(αx), pas une somme ; la réponse porte sur y, pas sur z.

## Les pièges

- Corrélation n'est pas causalité : r proche de 1 ne dit pas quelle variable agit sur l'autre.
- r proche de 0 veut dire « pas de lien affine », pas « aucun lien ».
- Une extrapolation sans la réserve « si la tendance se poursuit » est une réponse incomplète.
- Arrondir a et b trop tôt déplace les prévisions : arrondir à la fin.
- Un seul point aberrant peut faire chuter r : regarder le nuage, et ne l'écarter qu'avec une raison.
- Après z = ln y, répondre en y : e^(αx + β) = e^β × e^(αx).

*Ce document est la propriété intellectuelle de Ben, professeur particulier (benprofparticulier.fr), protégée par le Code de la propriété intellectuelle (art. L. 111-1 et L. 122-4). Il est mis à disposition pour un usage personnel de révision. Toute reproduction, diffusion, mise en ligne ou revente, totale ou partielle, sans autorisation écrite est interdite et constitue une contrefaçon (art. L. 335-2).*