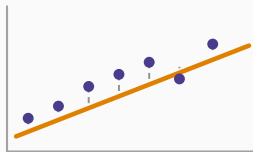


Introduction à l'économétrie

Chapitre 4 : Les moindres carrés ordinaires



Jaouad Madkour

Faculté d'Économie et de Gestion de Tanger · 27 août 2026

Le critère

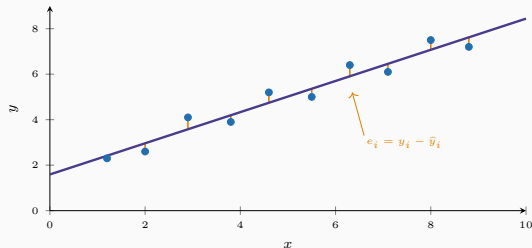
Ce que les formules disent

Ce qu'il faut retenir

Le critère

Trois questions sur une même figure

les moindres carrés : on minimise la somme des carrés des segments verticaux



Trois questions, trois réponses

Pourquoi vertical ?

Parce qu'on explique y par x . L'erreur porte sur y , pas sur x : la distance pertinente est celle-là.

Pourquoi au carré ?

Pour que les écarts positifs et négatifs ne se compensent pas, et pour pénaliser davantage les gros écarts. La valeur absolue conviendrait aussi, mais elle n'est pas dérivable.

Pourquoi cette droite-là ?

Elle est *unique* : annuler les deux dérivées donne un système de deux équations à deux inconnues, résolu au chapitre 11 du cours de mathématiques.

Ce que l'on minimise

$$\min_{\beta_0, \beta_1} S(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$$

On cherche la droite qui rend la somme des carrés des écarts verticaux la plus petite possible.

Un problème du cours de mathématiques

Une fonction de deux variables, dont on cherche le minimum : on annule les deux dérivées partielles.

C'est le chapitre 7 du cours de mathématiques pour la gestion, appliqué tel quel. Le système obtenu — deux équations, deux inconnues — se résout par la méthode du chapitre 11.

Annuler les deux dérivées

$$\frac{\partial S}{\partial \beta_0} = -2 \sum_i (y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\frac{\partial S}{\partial \beta_1} = -2 \sum_i x_i (y_i - \beta_0 - \beta_1 x_i) = 0$$

La solution

$$\hat{\beta}_1 = \frac{\text{cov}(x, y)}{V(x)} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Les salaires

$$\text{cov} = 9\,859,9 \quad V(x) = 35,574$$

$$\hat{\beta}_1 = \frac{9\,859,9}{35,574} = 277,17 \quad \hat{\beta}_0 = 7\,158,26 - 277,17 \times 10,142 = 4\,347,23$$

La consommation

$$\widehat{C} = 101,23 + 0,779 R$$

La propension marginale à consommer vaut **0,779** : comprise entre zéro et un, comme Keynes l'annonçait.

Ce que les formules disent

La covariance, normalisée

$$\hat{\beta}_1 = \frac{\text{cov}(x, y)}{V(x)} \quad \text{et} \quad r = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y}$$

La pente et la corrélation ont le **même numérateur**. Ils portent donc toujours le même signe, et l'un est nul si et seulement si l'autre l'est.

Le lien exact

$$\hat{\beta}_1 = r \frac{\sigma_y}{\sigma_x}$$

La corrélation est sans unité; la pente ne l'est pas. **La pente dit combien, la corrélation dit à quel point.** Ce sont deux questions distinctes, et le chapitre 5 y reviendra.

Vraies par construction, quelles que soient les données

1. $\sum_i e_i = 0$: les résidus somment à zéro.
2. $\sum_i x_i e_i = 0$: ils sont orthogonaux à la variable explicative.
3. La droite passe par le **point moyen** $(\bar{x}, \bar{y})$.

Elles ne prouvent rien sur le modèle

Ces trois propriétés découlent des équations normales : elles sont vraies même si le modèle est absurde.

Voir des résidus de somme nulle n'est donc pas un contrôle de qualité — c'est une identité. Le vrai diagnostic viendra du chapitre 13.

Pourquoi ce critère plutôt qu'un autre

Trois raisons, dans l'ordre d'importance

1. **La solution est unique et explicite** : deux formules, aucun algorithme.
2. **Elle est optimale** : le théorème de Gauss-Markov, chapitre 12.
3. **Elle est dérivable** : la valeur absolue ne l'est pas en zéro.

Sa faiblesse, et il faut la connaître

Le carré pénalise lourdement les grands écarts : une seule observation aberrante peut déplacer sensiblement la droite.

C'est le même compromis que celui de la moyenne contre la médiane, chapitre 7 de l'inférence. **Les moindres carrés sont efficaces et fragiles** — le chapitre 18 montrera comment repérer les points qui en abusent.

Ce qu'il faut retenir

Six points

1. On minimise la somme des **carrés des écarts verticaux** : $\sum_i (y_i - \beta_0 - \beta_1 x_i)^2$.
2. Verticaux parce qu'on explique y par x ; au carré pour ne pas compenser et rester dérivable.
3. $\hat{\beta}_1 = \text{cov}(x, y) / V(x)$ et $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$.
4. $\hat{\beta}_1 = r \sigma_y / \sigma_x$: la pente dit **combien**, la corrélation dit **à quel point**.
5. Trois propriétés algébriques, vraies par construction — donc sans valeur de diagnostic.
6. Les MCO sont **efficaces et fragiles** : une observation extrême pèse lourd.

La suite

La droite est calculée. Reste à savoir si elle vaut quelque chose : de quelle part de la dispersion rend-elle compte ?

Le chapitre 5 y répond par une décomposition que le cours d'inférence a déjà rencontrée sous un autre nom.