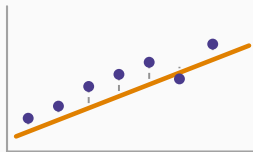


Introduction à l'économétrie

Chapitre 5 : La qualité de l'ajustement



Jaouad Madkour

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

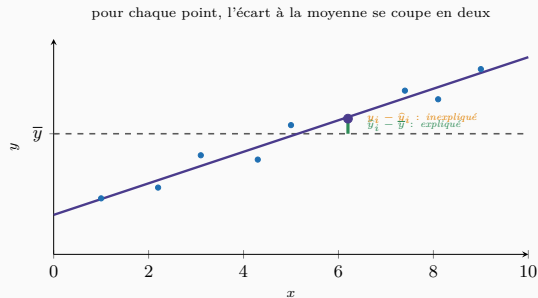
La décomposition

Le coefficient de détermination

Ce qu'il faut retenir

La décomposition

Couper l'écart en deux



La décomposition, terme à terme

$$y_i - \bar{y} = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i)$$

En élevant au carré et en sommant, le double produit s'annule — c'est une propriété des moindres carrés, non une approximation :

$$\underbrace{SCT}_{\text{totale}} = \underbrace{SCE}_{\text{expliquée}} + \underbrace{SCR}_{\text{résiduelle}}$$

C'est exactement la décomposition de l'analyse de la variance, chapitre 18 du cours d'inférence. L'ANOVA découpe selon des groupes ; la régression découpe selon une droite.

D'où $R^2 = SCE/SCT$, qui n'est autre que le η^2 de l'ANOVA.

L'identité fondamentale

$$\underbrace{\sum_i (y_i - \bar{y})^2}_{\text{SCT}} = \underbrace{\sum_i (\hat{y}_i - \bar{y})^2}_{\text{SCE}} + \underbrace{\sum_i (y_i - \hat{y}_i)^2}_{\text{SCR}}$$

Totale, expliquée, résiduelle.

Le double produit s'annule exactement

En développant le carré, il apparaît un terme croisé $2 \sum_i (\hat{y}_i - \bar{y}) e_i$.

Il est **nul** — conséquence directe des deux propriétés algébriques du chapitre 4. Ce n'est pas une approximation : c'est une identité, et elle ne vaut que pour les moindres carrés.

Le coefficient de détermination

Le R^2

$$R^2 = \frac{\text{SCE}}{\text{SCT}} = 1 - \frac{\text{SCR}}{\text{SCT}} \quad 0 \leq R^2 \leq 1$$

C'est la **part de la variance de y** dont le modèle rend compte.

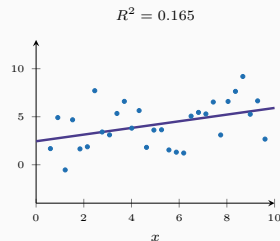
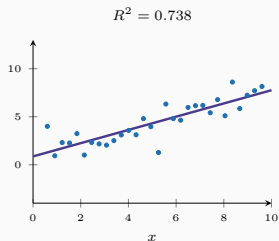
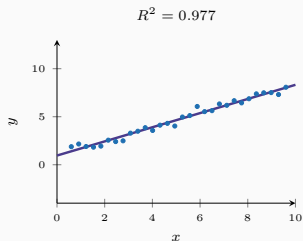
Trois identités à connaître

En régression simple, $R^2 = r^2$: le carré du coefficient de corrélation linéaire.

C'est aussi le η^2 de l'analyse de la variance, chapitre 18 de l'inférence — même quantité, autre découpage.

Et $R^2 = 0$ signifie que la droite est **horizontale** : la meilleure prévision est alors la moyenne.

Ce qu'il ne dit pas



Les trois droites ont la même pente. Seule change la dispersion autour d'elles.

Le R^2 mesure la *part de la variance expliquée*, non l'intensité de l'effet. Un R^2 faible avec une pente forte et significative est parfaitement possible — et c'est le cas ordinaire en économie des ménages, où l'on explique rarement plus de la moitié de la variance.

Ne jamais juger un modèle sur le seul R^2 . Un R^2 de 0,30 sur des données individuelles est un bon résultat ; un R^2 de 0,99 sur une série temporelle est souvent le signe que les deux variables croissent simplement avec le temps.

Un R^2 faible n'est pas un mauvais résultat

Sur des données **individuelles**, on explique rarement plus de la moitié de la variance : les comportements humains sont irréductiblement dispersés.

Nos salaires donnent $R^2 = 0,40$ avec la seule ancienneté. C'est un bon résultat, et la pente est significative à $t = 18,4$.

Un R^2 élevé n'est pas non plus un bon résultat

Notre enquête de ménages donne $R^2 = 0,849$: bien plus élevé que sur les salaires, et c'est normal — le revenu détermine la consommation de façon beaucoup plus directe que l'ancienneté ne détermine le salaire.

Un R^2 ne se compare qu'entre modèles portant sur la même variable expliquée. Comparer 0,40 sur les salaires et 0,85 sur la consommation ne dit rien : ce ne sont pas les mêmes phénomènes.

Quatre questions, dans cet ordre

1. Les **signes** des coefficients sont-ils conformes à la théorie?
2. Leurs **ordres de grandeur** sont-ils plausibles?
3. Sont-ils **significatifs**?
4. Enfin, quel est le R^2 ?

Le R^2 arrive en dernier, et ce n'est pas un hasard

Un modèle aux coefficients absurdes et au R^2 de 0,95 ne vaut rien. Un modèle aux coefficients justes et au R^2 de 0,30 enseigne quelque chose.

On ne publie pas un R^2 , on publie un coefficient et son intervalle.

La mesure oubliée

$$\hat{\sigma}_\varepsilon = \sqrt{\frac{\text{SCR}}{n-2}}$$

C'est l'écart-type des résidus, **dans l'unité de la variable expliquée**.

Elle est souvent plus parlante que le R^2

Sur nos salaires, $\hat{\sigma}_\varepsilon = 2\,008$ dirhams : le modèle se trompe typiquement de deux mille dirhams sur un salaire individuel.

Voilà une phrase qu'un directeur des ressources humaines comprend immédiatement. « $R^2 = 0,40$ » ne lui dit rien. **Le $n - 2$ au dénominateur est celui du chapitre 21 de l'inférence : deux paramètres estimés, deux degrés de liberté perdus.**

Ce qu'il faut retenir

Six points

1. $SCT = SCE + SCR$: le double produit s'annule, propriété exacte des MCO.
2. $R^2 = SCE/SCT$: la part de variance expliquée, entre 0 et 1.
3. En régression simple, $R^2 = r^2$ — et c'est le η^2 de l'ANOVA.
4. Un R^2 **faible** est normal sur données individuelles; 0,40 ici.
5. Un R^2 ne se compare qu'à **variable expliquée identique** : 0,40 et 0,85 ne se confrontent pas.
6. $\hat{\sigma}_\epsilon = \sqrt{SCR/(n-2)}$ est souvent plus parlant : 2 008 dirhams.

La suite

Le R^2 décrit l'échantillon. Il ne dit rien de la population — et c'est pourtant sur elle que porte la question économique.

Le chapitre 6 fait le pas, et il n'introduira rien : tout a été démontré au chapitre 21 du cours d'inférence.