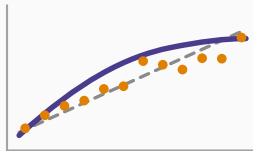


Économétrie avancée

Chapitre 6 : Régression multiple — approfondissements



Jaouad Madkour

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

Avant de commencer

Mise à l'échelle des données

Formes fonctionnelles avancées

Qualité d'ajustement et choix de modèle

Prédiction et analyse des résidus

Synthèse

Avant de commencer

Pourquoi ce chapitre ?

Les chapitres 2 à 5 ont posé les fondations : estimation, inférence, propriétés asymptotiques. Ce chapitre réunit des outils pratiques indispensables à toute application sérieuse — changer les unités de mesure, choisir la bonne forme fonctionnelle, comparer des modèles, et construire des prédictions fiables avec leur incertitude.

Mise à l'échelle des données

Effet du rééchelonnage

Si l'on multiplie une variable explicative x_j par une constante $c \neq 0$, son coefficient est divisé par c (l'effet économique total est inchangé). Si l'on multiplie y par c , **tous** les coefficients sont multipliés par c . Dans les deux cas, les statistiques t , les statistiques F , les intervalles de confiance et le R^2 sont **invariants** — seule l'unité de lecture des coefficients change.

Poids de naissance et tabagisme

Avec $\widehat{bwght} = 116.97 - 0.463 \text{ cigs} + 0.093 \text{ faminc}$ (poids en onces), passer le poids en livres ($bwght/16$) divise l'intercept et les pentes par 16, sans changer les statistiques t . Remplacer cigs par $\text{packs} = \text{cigs}/20$ multiplie son coefficient par 20 et son erreur-type par 20 — la statistique t , elle, ne bouge pas : le test de significativité ne dépend jamais de l'unité choisie.

Coefficients bêta

En centrant-réduisant y et chaque x_j (score $z = (x - \bar{x})/\hat{\sigma}_x$), on obtient les **coefficients bêta** $b_j = (\hat{\sigma}_j/\hat{\sigma}_y)\hat{\beta}_j$. Ils mesurent l'effet, en écarts-types de y , d'une hausse d'un écart-type de x_j – ce qui rend les magnitudes comparables entre régresseurs d'unités différentes.

Comparer des pommes et des oranges

Comparer directement $\hat{\beta}_{nox}$ (mesuré en parties par million) à $\hat{\beta}_{rooms}$ (mesuré en nombre de pièces) n'a aucun sens. Les coefficients bêta répondent à la question « quel régresseur bouge le plus y , en termes relatifs? » – par exemple $\widehat{z_{price}} = -.340 z_{nox} - .143 z_{crime} + .514 z_{rooms} - \dots$: le nombre de pièces domine largement l'effet de la pollution ou de la criminalité sur le prix standardisé.

Formes fonctionnelles avancées

Variation exacte en pourcentage

Dans $\log(y) = \beta_0 + \beta_1 \log(x_1) + \beta_2 x_2 + u$, l'approximation usuelle $\% \Delta y \approx 100 \beta_2 \Delta x_2$ se dégrade quand $\beta_2 \Delta x_2$ est grand. La variation **exacte** est

$$\% \Delta y = 100 [\exp(\beta_2 \Delta x_2) - 1].$$

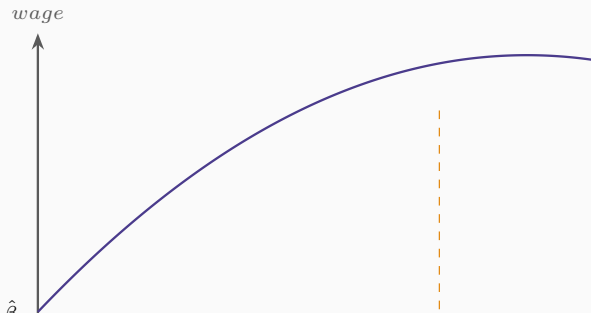
Une pièce de plus

Avec $\widehat{\log(\text{price})} = 9.23 - .718 \log(\text{nox}) + .306 \text{rooms}$, l'approximation donne +30.6% pour une pièce supplémentaire; le calcul exact donne $100[\exp(.306) - 1] \approx 35.8\%$ – un écart non négligeable dès que le coefficient est élevé.

Effet marginal et point de retournement

Dans $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2$, la pente varie avec x : $\Delta\hat{y}/\Delta x \approx \hat{\beta}_1 + 2\hat{\beta}_2 x$. Si $\hat{\beta}_1 > 0$ et $\hat{\beta}_2 < 0$ (forme parabolique), le maximum est atteint en

$$x^* = \left| \frac{\hat{\beta}_1}{2\hat{\beta}_2} \right|.$$



Effet croisé entre deux régresseurs

Dans $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + u$, l'effet marginal de x_1 dépend du niveau de x_2 : $\Delta y / \Delta x_1 = \beta_1 + \beta_3 x_2$. On évalue généralement cet effet à des valeurs intéressantes de x_2 (moyenne, quartiles), et β_3 se teste comme n'importe quel coefficient.

Recentrer pour interpréter l'intercept

Remplacer x_2 par $(x_2 - \bar{x}_2)$ dans l'interaction transforme le coefficient sur x_1 en effet marginal **à la moyenne** de x_2 , avec son erreur-type directement disponible – une astuce systématique pour obtenir un test de significativité à un point précis plutôt qu'à $x_2 = 0$, souvent hors du domaine des données.

Qualité d'ajustement et choix de modèle

R^2 ajusté (ou $\bar{R}^2$)

$$\bar{R}^2 = 1 - \frac{SSR/(n - k - 1)}{SST/(n - 1)} = 1 - \hat{\sigma}^2/[SST/(n - 1)].$$

Contrairement au R^2 usuel, qui ne peut jamais baisser quand on ajoute un régresseur, $\bar{R}^2$ pénalise l'ajout de variables inutiles : il augmente si et seulement si la statistique t de la nouvelle variable dépasse 1 en valeur absolue.

$\bar{R}^2$ pour comparer des modèles non emboîtés

Deux modèles sont **non emboîtés** si aucun n'est un cas particulier de l'autre (par exemple $\log(\text{sales})$ contre $\text{sales} + \text{sales}^2$ pour expliquer l'intensité de R&D). Le test F , réservé aux modèles emboîtés, ne s'applique pas ; $\bar{R}^2$ offre un critère simple — mais il ne permet **jamais** de comparer des modèles où la variable dépendante diffère (y contre $\log(y)$), car ils n'expliquent pas la même variation totale.

Le danger du sur-ajustement

Maximiser le R^2 peut conduire à contrôler pour des variables qu'il ne faut **pas** contrôler : dans un modèle expliquant les décès routiers par la taxe sur la bière, ajouter la consommation de bière comme régresseur détruit l'interprétation *ceteris paribus* de l'effet de la taxe, qui agit précisément *via* la consommation. La qualité d'ajustement ne remplace jamais le raisonnement économique sur ce qu'il faut tenir fixe.

Prédiction et analyse des résidus

Prédiction ponctuelle et son incertitude

Pour des valeurs $c_1, \dots, c_k$ des régresseurs, $\hat{\theta}_0 = \hat{\beta}_0 + \hat{\beta}_1 c_1 + \dots + \hat{\beta}_k c_k$ estime $E(y \mid x = c)$. En recentrant chaque régresseur $(x_j - c_j)$ et en ré-estimant, l'intercept de la nouvelle régression donne directement $\hat{\theta}_0$ et son erreur-type, d'où l'IC $\hat{\theta}_0 \pm t_{\alpha/2, n-k-1} \text{se}(\hat{\theta}_0)$.

Intervalle de prédiction pour une observation future

Intervalle de prédiction

Pour une observation individuelle y^0 (non l'espérance), l'erreur de prédiction $\hat{e}^0 = y^0 - \hat{y}^0$ a pour variance $\text{Var}(\hat{y}^0) + \sigma^2$: à la variance de l'estimation s'ajoute la variance de l'erreur elle-même, qui ne diminue pas avec n . L'intervalle de prédiction à 95 % est donc **toujours plus large** que l'IC pour l'espérance conditionnelle :

$$\hat{y}^0 \pm t_{.025, n-k-1} \text{se}(\hat{e}^0), \quad \text{se}(\hat{e}^0) = \sqrt{\text{se}(\hat{y}^0)^2 + \hat{\sigma}^2}.$$

GPA universitaire prédit

Pour un profil donné, $\hat{y}^0 = 2.70$ avec $\text{se}(\hat{y}^0) = .020$ et $\hat{\sigma} = .560$. L'IC pour l'espérance est étroit (2.66 à 2.74), mais l'intervalle de prédiction pour un étudiant individuel, $2.70 \pm 1.96(.560)$, s'étend de 1.60 à 3.80 : la quasi-totalité de l'incertitude vient de σ^2 , pas de l'estimation des paramètres.

Ce que révèlent les résidus

Le résidu $\hat{u}_i = y_i - \hat{y}_i$ mesure l'écart entre observé et prédit **compte tenu des facteurs contrôlés**. Un résidu très négatif sur un prix de vente immobilier signale un bien sous-évalué relativement à ses caractéristiques; en finance, l'analyse des résidus d'un modèle factoriel (chapitre 1, économétrie financière) identifie l'alpha spécifique à un actif après avoir neutralisé son exposition aux facteurs de risque communs.

Prédire y quand $\log(y)$ est la variable dépendante

Retransformation du log

Si $\widehat{\log y}$ est le prédicteur du modèle log-linéaire, $\exp(\widehat{\log y})$ **sous-estime** systématiquement $E(y \mid x)$. Sous normalité de u , la correction est $\hat{y} = \exp(\hat{\sigma}^2/2) \exp(\widehat{\log y})$; sans hypothèse de normalité, on estime plutôt un facteur $\hat{\alpha}_0$ par régression sans constante de y sur $m_i = \exp(\widehat{\log y_i})$, puis $\hat{y} = \hat{\alpha}_0 \exp(\widehat{\log y})$.

Application : prévoir une rémunération, prévoir un rendement

La retransformation du log est essentielle chaque fois qu'un modèle log-linéaire sert à prévoir un niveau plutôt qu'une variation relative — prévoir le salaire d'un dirigeant à partir de $\log(\textit{salary})$, ou reconstituer un prix ou une valeur de portefeuille à partir d'un rendement logarithmique prévu par un modèle GARCH (chapitre 4, économétrie financière). Oublier le facteur correctif biaise systématiquement la prévision vers le bas.

Synthèse

L'essentiel du chapitre 6

- Le rééchantillonnage des variables change les coefficients de façon prévisible mais laisse t , F et R^2 inchangés; les coefficients bêta rendent les effets comparables entre régresseurs.
- Logarithmes, quadratiques et interactions permettent des formes fonctionnelles riches; chacun a sa règle d'interprétation propre (effet exact, point de retournement, effet croisé).
- $\bar{R}^2$ pénalise la complexité et permet de comparer des modèles non emboîtés — mais jamais des modèles à variable dépendante différente.
- Prédiction : distinguer l'IC pour $E(y \mid x)$ (rétréci par n) de l'intervalle de prédiction pour une observation future (dominé par σ^2); la retransformation du log requiert un facteur correctif.

Vers le chapitre 7

Jusqu'ici, tous les régresseurs étaient quantitatifs. Le chapitre suivant introduit les **variables qualitatives** (indicatrices) — genre, catégorie, appartenance à un groupe — et montre comment leur donner une interprétation aussi rigoureuse que celle des variables continues.