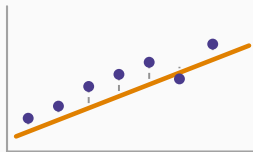


Introduction à l'économétrie

Chapitre 11 : Les variables qualitatives



Jaouad Madkour

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

Faire entrer une catégorie

Les interactions

Un dernier usage

Ce qu'il faut retenir

Faire entrer une catégorie

On ne régresse pas sur des mots

Le diplôme prend trois valeurs : Bac, Bac+2, Bac+5. La région en prend quatre, le sexe deux.

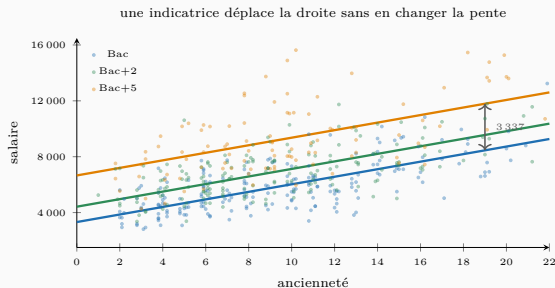
Les coder 1, 2, 3 serait une faute : cela imposerait que l'écart entre Bac et Bac+2 soit exactement le même qu'entre Bac+2 et Bac+5, et que l'ordre ait un sens arithmétique. Pour la région, ce serait absurde.

La variable indicatrice

Une **indicatrice** vaut 1 si la modalité est présente, 0 sinon.

$$D_2 = \begin{cases} 1 & \text{si Bac+2} \\ 0 & \text{sinon} \end{cases} \quad D_5 = \begin{cases} 1 & \text{si Bac+5} \\ 0 & \text{sinon} \end{cases}$$

Ce qu'elle fait au modèle



Trois modalités, deux indicatrices

$$\hat{S} = 3\,320 + 270 A + 1\,100 D_2 + 3\,337 D_5$$

Le Bac n'a pas d'indicatrice : c'est la **modalité de référence**, absorbée dans la constante. Les deux coefficients se lisent *par rapport à elle*.

Le piège des indicatrices

Mettre les trois rendrait leur somme égale à 1, donc colinéaire à la constante : le modèle serait insoluble. **Toujours $k - 1$ indicatrices pour k modalités.**

Ici les droites sont *parallèles* : le modèle suppose que le rendement d'une année d'ancienneté est le même pour tous. C'est une hypothèse — que le chapitre suivant lève.

Le piège de la colinéarité parfaite

Pour k modalités, on met $k - 1$ **indicatrices**, jamais k .

Les trois réunies auraient une somme constamment égale à 1, donc identique à la constante : la matrice $X'X$ serait singulière et le modèle n'aurait aucune solution.

La modalité écartée n'est pas perdue

Elle devient la **référence**, absorbée dans la constante. Les coefficients des autres se lisent **par rapport à elle**.

Ici : 1 100 dirhams de plus pour un Bac+2 que pour un Bac, 3 337 de plus pour un Bac+5. Le Bac ne disparaît pas — il est le zéro de l'échelle.

Il ne change rien, et il change tout

Prendre le Bac+5 en référence donnerait —3 337 pour le Bac : mêmes écarts, signes inversés. Le modèle est **strictement identique** — mêmes prévisions, même R^2 , même F .

Mais la lecture change. **On choisit comme référence la modalité la plus naturelle** — la plus nombreuse, ou celle qui sert de point de comparaison dans le débat.

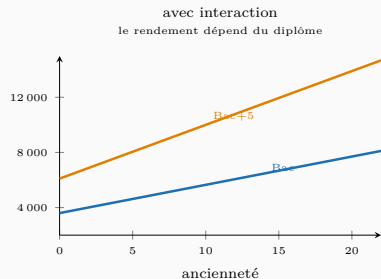
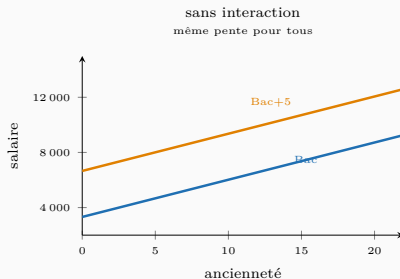
Les logiciels choisissent pour vous

R prend la première modalité dans l'ordre alphabétique : ici « Bac », par chance le bon choix.

Sur d'autres données, l'ordre alphabétique donne une référence absurde. **Vérifiez toujours laquelle a été retenue avant de lire les coefficients** : c'est une source d'erreur d'interprétation très fréquente.

Les interactions

Quand l'effet dépend du groupe



$$\hat{S} = \beta_0 + \beta_1 A + \beta_2 D_5 + \underbrace{\beta_3 (A \times D_5)}_{\text{le terme d'interaction}}$$

Une indicatrice seule déplace la droite ; multipliée par une variable, elle en change la pente. L'interaction répond à une question que le modèle additif ne pose même pas : *le rendement de l'ancienneté est-il le même selon le diplôme ?*

Le coefficient β_3 se teste comme les autres. S'il n'est pas significatif, on revient au modèle à droites parallèles — plus simple, et donc préférable.

Multiplier une indicatrice par une variable

$$S = \beta_0 + \beta_1 A + \beta_2 D_5 + \beta_3 (A \times D_5) + \varepsilon$$

- β_1 : le rendement de l'ancienneté **pour la référence**.
- $\beta_1 + \beta_3$: le rendement **pour les Bac+5**.
- β_3 : la **différence** de rendement entre les deux.

La question que l'interaction pose

Sans elle, le modèle **impose** que le rendement de l'ancienneté soit le même pour tous. C'est une hypothèse, jamais un fait.

L'interaction la lève et la teste : β_3 se teste comme n'importe quel coefficient. **S'il n'est pas significatif, on revient au modèle à droites parallèles** — plus simple, donc préférable.

À ne pas oublier

1. **Toujours conserver les termes simples** : si $A \times D_5$ est dans le modèle, A et D_5 doivent y être aussi.
2. **Ne jamais lire β_1 seul** dans un modèle avec interaction : il ne vaut que pour la référence.

L'usage politique de l'interaction

« L'effet d'une formation est-il le même pour les jeunes et les seniors ? » « Le rendement du diplôme diffère-t-il selon la région ? »

Ce sont des questions d'**hétérogénéité des effets**, et l'interaction est le seul moyen simple d'y répondre. Elles sont souvent plus intéressantes que l'effet moyen lui-même.

Un dernier usage

Une indicatrice pour une poignée d'observations

Une indicatrice qui vaut 1 pour un petit groupe atypique — un service, un site, une catégorie de contrats — permet de **lui laisser sa propre constante** sans qu'il déforme le reste.

Une alternative à la suppression

Face à quelques observations manifestement particulières, deux attitudes : les retirer du fichier, ou leur donner une indicatrice.

La seconde est presque toujours préférable. Elle conserve l'information, elle est visible dans la sortie, et elle se teste. Retirer des lignes sans le dire ne se voit nulle part — c'est la différence entre une décision de modélisation et une manipulation.

Ce qu'il faut retenir

Six points

1. Une **indicatrice** vaut 1 ou 0; coder 1, 2, 3 serait une faute.
2. Pour k modalités, $k - 1$ **indicatrices** : sinon colinéarité parfaite et modèle insoluble.
3. La modalité écartée devient la **référence**, absorbée dans la constante.
4. Changer de référence ne change **rien au modèle**, seulement la lecture.
5. Une **interaction** fait varier la pente selon le groupe; garder toujours les termes simples.
6. Une indicatrice pour un sous-groupe atypique vaut mieux que sa **suppression** : elle se voit et se teste.

La suite

Toute la mécanique est en place : estimer, mesurer l'ajustement, tester, prévoir, avec des variables de toute nature.

Reste à savoir **à quelles conditions tout cela est vrai**. La quatrième partie énonce les hypothèses, et la cinquième examine ce qui se passe quand elles tombent.