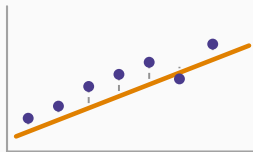


Introduction à l'économétrie

Chapitre 14 : La multicollinéarité



Jaouad Madkour

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

Le problème

Que faire

Ce qu'il faut retenir

Le problème

Multicolinéarité

Deux ou plusieurs variables explicatives sont **fortement liées entre elles**, sans l'être parfaitement.

Le calcul aboutit — H5 n'est pas violée — mais il devient instable.

La question à laquelle on ne peut plus répondre

Le *ceteris paribus* demande l'effet d'une variable **les autres étant constantes**.

Si deux variables bougent toujours ensemble, les données ne contiennent aucune observation où l'une varie et pas l'autre. **On demande au modèle de séparer ce que les données n'ont jamais séparé.**

on ajoute la *même* variable, exprimée en mois : corrélation 0,99994, VIF = 8 084

	M2 — ancienneté en années		M6 — années <i>et</i> mois	
	coefficient	<i>t</i>	coefficient	<i>t</i>
Ancienneté (années)	270,27	23,06	-1 404,96	-1,34
Ancienneté (mois)	—	—	139,58	1,59
Bac+2	1 099,57	6,93	1 095,86	6,92
Bac+5	3 336,99	18,11	3 336,00	18,13
R^2	0,6418		0,6436	

Les quatre symptômes de la multicolinéarité, tous visibles ici.

1. Les coefficients deviennent *absurdes* : une année d'ancienneté ferait perdre 1 405 dirhams. 2. Ils deviennent *non significatifs* alors que la variable compte manifestement. 3. Les écarts-types *explosent* : 1 051 contre 11,72. 4. Le R^2 , lui, ne bouge pas — il monte même très légèrement.

Le diagnostic tient dans cette dernière ligne. Le modèle prédit toujours aussi bien ; il est simplement devenu incapable de *séparer* les rôles. Demander l'effet des années à mois constants n'a aucun sens : c'est la même information deux fois.

Remèdes : retirer l'une des deux, les combiner, ou ne rien faire si le coefficient litigieux n'est pas celui qui intéresse. La multicolinéarité ne biaise rien — elle rend seulement imprécis.

Ils vont ensemble

1. Les **écarts-types explosent** : 1 051 contre 11,72.
2. Les coefficients deviennent **non significatifs** alors que les variables comptent.
3. Ils prennent des valeurs **absurdes** — ici, un signe négatif pour l'ancienneté.
4. Le R^2 **ne bouge pas** : 0,6418 contre 0,6436.

Le quatrième est le plus révélateur

Le modèle prédit toujours aussi bien. Il a seulement perdu la capacité d'**attribuer**.

C'est exactement le cas d'école du chapitre 10 : un F global significatif avec des t individuels qui ne le sont pas.

Le facteur d'inflation de la variance

$$\text{VIF}_j = \frac{1}{1 - R_j^2}$$

où R_j^2 est le R^2 de la régression de x_j sur **toutes les autres** variables explicatives.

Sur nos données

Modèle	Variable	R_j^2	VIF
M4	ancienneté	0,012	1,01
M4	Casablanca	0,492	1,97
M4	Tanger	0,519	2,08
M6	ancienneté	0,999876	8 084

Les repères usuels

VIF	Lecture
< 5	rien à signaler
5 à 10	à surveiller
> 10	multicolinéarité sérieuse

Ce que le VIF signifie littéralement

Un VIF de 4 signifie que la variance du coefficient est **quatre fois** plus grande qu'elle ne le serait si la variable était indépendante des autres. L'écart-type est donc doublé.

Notre modèle M4 est parfaitement sain : le plus grand VIF vaut 2,08. La **multicolinéarité n'est pas un problème dans nos données réelles** — il a fallu la fabriquer pour la montrer.

Que faire

Quatre options

1. **Retirer** l'une des variables redondantes.
2. Les **combiner** en un indice unique.
3. **Collecter plus de données** — cela réduit toutes les variances.
4. **Ne rien faire.**

La quatrième option est légitime, et souvent la bonne

La multicolinéarité **ne biaise rien**. Les prévisions restent excellentes, le R^2 est intact, et les coefficients non touchés par le problème restent parfaitement estimés.

Si le coefficient qui vous intéresse n'est pas parmi ceux qui sont brouillés, **il n'y a rien à corriger**. Sur notre modèle M6, les coefficients du diplôme sont inchangés au dirham près.

Retirer une variable est plus dangereux que la garder

Supprimer une variable colinéaire pour « nettoyer » le modèle revient à créer une **variable omise**.

On échange alors un problème de précision — sans biais — contre un problème de biais. **C'est très exactement le mauvais échange**, et le chapitre 17 dira pourquoi.

La règle de décision

Retirer une variable uniquement si l'on peut justifier **économiquement** qu'elle n'a pas sa place dans le modèle.

Jamais parce que son t est faible, jamais parce que son VIF est élevé. **La spécification se décide sur la théorie, pas sur les diagnostics.**

Ce qu'il faut retenir

Six points

1. La multicolinéarité, c'est demander de **séparer ce que les données n'ont jamais séparé**.
2. Quatre symptômes : écarts-types gonflés, t effondrés, coefficients absurdes, R^2 intact.
3. Le quatrième est le diagnostic : le modèle prédit bien mais n'attribue plus.
4. $VIF_j = 1/(1 - R_j^2)$; au-delà de 10, c'est sérieux. Nos données réelles : 2,08 au plus.
5. **Elle ne biaise rien** : ne rien faire est souvent la bonne décision.
6. Retirer une variable colinéaire crée une **variable omise** — un problème bien pire.

La suite

Passons à une violation qui, elle, est présente dans presque toutes les données individuelles — et que nos salaires exhibent de façon spectaculaire.