

ECONOMETRÍA

TEMA 4: MULTICOLINEALIDAD

2024-2025

1	Concepto, causas y consecuencias	2
1.1	Multicolinealidad exacta	3
1.2	Multicolinealidad aproximada	4
1.3	Consecuencias	8
1.3.1	Consecuencias sobre el análisis estadístico del modelo	8
1.3.2	Efectos nocivos sobre el análisis numérico del modelo	10
2	Detección de multicolinealidad aproximada	11
2.1	Factor de Inflación de la Varianza	11
2.2	Número de Condición	12
2.3	Otras medidas	13
3	Soluciones	14

CONCEPTO

- ▶ El problema de multicolinealidad consiste en la existencia de relaciones lineales entre dos o más variables independientes del modelo lineal uniecuacional múltiple.
- ▶ Dependiendo de cómo sea dicha relación lineal hablaremos de multicolinealidad perfecta (o exacta) o aproximada.
- ▶ Dentro de la multicolinealidad aproximada, distinguiremos entre multicolinealidad aproximada errática o sistemática y multicolinealidad aproximada esencial o no esencial.

MULTICOLINEALIDAD EXACTA

- ▶ **Concepto:** La multicolinealidad exacta o perfecta hace referencia a la existencia de una relación lineal exacta entre dos o más variables independientes.
- ▶ **Causas:** Dicho tipo de multicolinealidad se traduce en el incumplimiento de una de las hipótesis básicas del modelo uniecuacional múltiple: la matriz X no es de rango completo por columnas, esto es, $rg(X) < k$.
- ▶ **Consecuencias:** El incumplimiento de dicha hipótesis no permite invertir la matriz $X^t X$, por lo que el sistema normal

$$X^t X \cdot \beta = X^t y,$$

es compatible indeterminado, es decir, es imposible obtener una solución única para $\hat{\beta}$ (hay infinitas).

- ▶ **Ejemplo:** Un ejemplo sería la trampa de las variables ficticias.

MULTICOLINEALIDAD APROXIMADA

- ▶ **Concepto:** La multicolinealidad aproximada hace referencia a la existencia de una relación lineal aproximada entre dos o más variables independientes.
- ▶ **Causas:** Las principales causas que producen multicolinealidad aproximada en un modelo son:
 - relación causal entre variables explicativas del modelo.
 - escasa variabilidad en las observaciones de las variables independientes.
 - reducido tamaño de la muestra.

TIPOS DE MULTICOLINEALIDAD APROXIMADA

En función de las causas se hace la siguiente diferenciación:

- ▶ Según Spanos y McGuirk:

Multicolinealidad sistemática: debida a un problema estructural, es decir, a la alta correlación lineal de las variables exógenas consideradas.

Multicolinelidad errática: debido a un problema puramente numérico, es decir, a un mal condicionamiento de los datos considerados.

- ▶ Mientras que Marquandt y Snee:

Multicolinealidad no esencial: relación lineal de las variables exógenas con la constante (es sabido que se solventa centrando las variables).

Multicolinelidad esencial: relación lineal entre las variables exógenas (excluida la constante).

TIPOS DE MULTICOLINEALIDAD APROXIMADA

Luego, se podrían distinguir los siguientes cuatro casos:

Multicolinealidad	Sistemática	Errática
No esencial	1	2
Esencial	3	4

¿Puede existir multicolinealidad aproximada preocupante en el modelo de regresión lineal simple $\text{Salario} = \beta_1 + \beta_2 \cdot \text{Experiencia} + u$?

TIPOS DE MULTICOLINEALIDAD APROXIMADA

Luego, se podrían distinguir los siguientes cuatro casos:

Multicolinealidad	Sistemática	Errática
No esencial	1	2
Esencial	3	4

¿Puede existir multicolinealidad aproximada preocupante en el modelo de regresión lineal simple $\text{Salario} = \beta_1 + \beta_2 \cdot \text{Experiencia} + u$?

¿Y en el modelo $\text{Salario} = \beta_1 + \beta_2 \cdot \text{Experiencia} + \beta_3 \cdot \text{Edad} + u$?

TIPOS DE MULTICOLINEALIDAD APROXIMADA

Luego, se podrían distinguir los siguientes cuatro casos:

Multicolinealidad	Sistemática	Errática
No esencial	1	2
Esencial	3	4

¿Puede existir multicolinealidad aproximada preocupante en el modelo de regresión lineal simple $\text{Salario} = \beta_1 + \beta_2 \cdot \text{Experiencia} + u$?

¿Y en el modelo $\text{Salario} = \beta_1 + \beta_2 \cdot \text{Experiencia} + \beta_3 \cdot \text{Edad} + u$?

CONCEPTO, CAUSAS Y CONSECUENCIAS

MULTICOLINEALIDAD APROXIMADA

- ▶ **Consecuencias:** En este caso, no se incumplirá la hipótesis básica de que la matriz X sea completa por columnas ($rg(X) = k$), por lo que se podrá invertir X^tX y obtener los estimadores por mínimos cuadrados ordinarios. Sin embargo, el determinante de X^tX será muy próximo a cero, por lo que $(X^tX)^{-1}$ tenderá a tener valores altos y por tanto:
 - las varianzas de los estimadores son muy grandes.
 - al efectuar contrastes de significación individual no se rechazará la hipótesis nula, mientras que al realizar contrastes conjuntos si.
 - los coeficientes estimados serán muy sensibles ante pequeños cambios en los datos.
 - un coeficiente de determinación elevado.
- ▶ **Ejemplo:** La edad y la experiencia suelen presentar una alta relación ya que ambas evolucionan conjuntamente: a mayor edad se presupone mayor experiencia. Por tal motivo será difícil separar el efecto de cada una sobre la variable dependiente y que se produzca multicolinealidad debido a la relación causal existente entre dichas variables

EFECTOS NOCIVOS SOBRE EL ANÁLISIS ESTADÍSTICO DEL MODELO

Wissel, J. (2009). *A new biased estimator for multivariate regression models with highly collinear variables*. Ph.D. thesis.

Variable	Estimación	Desviación típica
Constante	5.469264	13.016791
Consumo	-4.252429	5.135058
Ingresos	3.120395	2.035671
Crédito Pendiente	0.002879	0.005764
R^2		0.9235
$\hat{\sigma}^2$		0.8695563
$F_{3,13}$		52.3
Estimación por MCO del crédito en Estados Unidos		

No se rechaza que $\beta_i = 0, i = 1, 2, 3$, y se rechaza que $\beta_2 = \beta_3 = \beta_4 = 0$.

EFECTOS NOCIVOS SOBRE EL ANÁLISIS ESTADÍSTICO DEL MODELO

Wooldridge, J.M. (2000). *Introductory Econometrics: A modern approach*. South-Western, Cengage Learning.

Variable	Estimación	Desviación típica
Constante	0.4404	0.09556
Tipos de interés a 3 meses	1.00569	0.01343
$R^2 = 0.9787, \quad \hat{\sigma}^2 = 0.2025, \quad F_{2,122} = 5611$		

Estimación por MCO de los tipos de interés a 12 meses en función de los tipos de interés a 3 meses

Variable	Estimación	Desviación típica
Constante	0.22471	0.0397
Tipos de interés a 3 meses	-0.62891	0.06582
Tipos de interés a 6 meses	1.59334	0.06394
$R^2 = 0.9965, \quad \hat{\sigma}^2 = 0.03330625, \quad F_{2,122} = 17371.66$		

Estimación por MCO de los tipos de interés a 12 meses en función de los tipos de interés a 3 y 6 meses

Correlación entre los tipos de interés a 3 y 6 meses: 0.9893021.

EFECTOS NOCIVOS SOBRE EL ANÁLISIS NUMÉRICO DEL MODELO

Wissel

C	C _p	I	I _p	CP	CP _p	β	β_p
4.7703	4.838464	4.8786	4.992323	808.23	839.7796	5.46926428	-12.844487481
4.7784	4.910650	5.0510	5.136480	798.03	789.1984	-4.252429358	3.324325322
4.9348	5.038435	5.3620	5.385545	806.12	836.3163	3.120395253	0.316260872
5.0998	5.044642	5.5585	5.635094	865.65	871.7818	0.002879118	-0.002397538
5.2907	5.350734	5.8425	5.838306	997.30	1006.9633		
5.4335	5.501856	6.1523	6.256959	1140.70	1148.8869		
⋮	⋮	⋮	⋮	⋮	⋮		
⋮	⋮	⋮	⋮	⋮	⋮		

Wooldridge

r3	r3 _p	r6	r6 _p	β	β_p
2.77	2.784788	3.02	3.027994	0.2247061	0.1845118
2.97	3.000442	3.43	3.463405	-0.6289143	-0.5209103
4.00	4.168316	4.32	4.423468	1.5933374	1.4875326
4.60	4.628050	4.68	4.721110		
4.16	4.148983	4.33	4.338835		
3.07	3.166463	3.50	3.588747		
⋮	⋮	⋮	⋮		
⋮	⋮	⋮	⋮		

2.1. FACTOR DE INFLACIÓN DE LA VARIANZA

Una de las medidas más usadas para detectar el grado de multicolinealidad existente es el Factor de Inflación de la Varianza (FIV) dado por:

$$FIV(i) = \frac{1}{1 - R_i^2}, \quad i = 2, \dots, p, \quad (1)$$

donde R_i^2 es el coeficiente de determinación de la regresión de \mathbf{X}_i sobre el resto de variables independientes, \mathbf{X}_{-i} .

Si esta medida es superior a 10 se supone que el grado de multicolinealidad presente en el modelo es preocupante.

El VIF no tiene en cuenta la relación de las variables exógenas del modelo, $\mathbf{X}_2 \dots \mathbf{X}_p$, con la constante, **1** Por tanto, no detecta la multicolinealidad no esencial.

2.2. NÚMERO DE CONDICIÓN

Otra medida muy extendida es el Número de Condición (NC), el cual viene dado por:

$$NC = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}}, \quad (2)$$

donde λ_{max} y λ_{min} son, respectivamente, los autovalores máximo y mínimo de $\tilde{\mathbf{X}}^t \tilde{\mathbf{X}}$ donde:

$$\tilde{\mathbf{X}} = [\tilde{\mathbf{1}} \ \tilde{\mathbf{X}}_2 \ \dots \ \tilde{\mathbf{X}}_p],$$
$$\tilde{\mathbf{1}} = \frac{\mathbf{1}}{\sqrt{n}}, \quad \tilde{\mathbf{X}}_i = \frac{\mathbf{X}_i}{\sqrt{\sum_{j=1}^n X_{ji}^2}}, \quad i = 2, \dots, p.$$

Si esta medida es superior a 20 se supone que el grado de multicolinealidad presente en el modelo es moderado y si es superior a 30 preocupante.

El NC tiene en cuenta la relación de las variables exógenas del modelo, $\mathbf{X}_2 \dots \mathbf{X}_p$, con la constante, $\mathbf{1}$.

2.3. OTRAS MEDIDAS

- ▶ **R**: Matriz de correlaciones simples de las variables independientes del modelo: Ignora por completo a la constante y proporciona información de las relaciones dos a dos, por lo tanto, sólo es capaz de detectar multicolinealidad aproximada del tipo esencial dos a dos.
- ▶ $\det(\mathbf{R})$: Determinante de la matriz de correlaciones: Recoge estructuras más complejas aunque sigue ignorando la relación con la constante, por lo que detecta multicolinealidad aproxima del tipo esencial. En este caso, valores próximos a cero indica que este problema es grave.
- ▶ $CV(X_i)$: Coeficiente de variación de las variables explicativas: Puede detectar una escasa variabilidad de las variables lo que podría provocar multicolinealidad aproximada de tipo no esencial.

3. SOLUCIONES

Algunas de las posibles soluciones al problema de multicolinealidad son las siguientes:

- ▶ mejora del diseño muestral extrayendo la información máxima de la variables observadas.
- ▶ eliminación de las variables que se sospechan son causantes de la multicolinealidad.
- ▶ en caso de disponer de pocas observaciones, aumentar el tamaño de la muestra.
- ▶ utilizar la relación extramuestral que permita realizar relaciones entre los parámetros (información a priori) que permita estimar el modelo por mínimos cuadrados restringidos.

Por otro lado, algunos autores sugieren tratar el problema de la multicolinealidad de forma mecánica y puramente numérica proponiendo métodos de estimación alternativos: estimación cresta, estimación alzada o residualización (entre otras).