

Teoría de la Multicolinealidad

Análisis Matricial, Diagnóstico y Remedios



Universidad Pedagógica y Tecnológica de Colombia

Facultad de Ciencias Económicas y Administrativas

Econometría Avanzada

Estudiante de Econometría

`estudiante@uptc.edu.co`

Viernes, 13 de febrero de 2026

Tunja, Boyacá - Colombia

Resumen

La multicolinealidad constituye uno de los problemas fundamentales en el análisis de regresión múltiple, afectando la precisión y estabilidad de los estimadores de mínimos cuadrados ordinarios (MCO). Este documento presenta un análisis riguroso y exhaustivo de la teoría de la multicolinealidad, abarcando desde sus fundamentos teóricos hasta las técnicas avanzadas de detección y remedios disponibles.

Se desarrolla el marco matricial completo del problema, incluyendo la formulación de la matriz de varianza-covarianza, el fenómeno de “explosión de varianza”, y los diagnósticos formales como el Factor de Inflación de Varianza (FIV), el Test de Farrar-Glauber, y el análisis de valores propios. Adicionalmente, se presentan soluciones prácticas que incluyen transformaciones de variables, regresión Ridge, LASSO, y técnicas de componentes principales.

El tratamiento combina rigor matemático con aplicabilidad práctica, proporcionando ejemplos detallados en modelos económicos relevantes como la función de producción Cobb-Douglas, modelos de inflación monetarista, y especificaciones de demanda de dinero.

Palabras clave: Multicolinealidad, Matriz de Varianza-Covarianza, Factor de Inflación de Varianza, Test de Farrar-Glauber, Regresión Ridge, Econometría.

Clasificación JEL: C10, C13, C20, C51, C52

Índice

Índice de figuras

Índice de tablas

1 Fundamentos Teóricos de la Multicolinealidad

1.1 Introducción y Motivación

La regresión lineal múltiple constituye la herramienta fundamental del análisis econométrico, permitiendo modelar relaciones complejas entre variables económicas. Sin embargo, cuando las variables explicativas están altamente correlacionadas entre sí, surge el problema de la **multicolinealidad**, que compromete la capacidad del investigador para aislar el efecto individual de cada regresor.

Multicolinealidad

En el modelo lineal $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, existe multicolinealidad cuando las columnas de la matriz \mathbf{X} son linealmente dependientes o casi dependientes, es decir:

$$\exists \lambda_1, \lambda_2, \dots, \lambda_k \in \mathbb{R} \quad (\text{no todos cero}) \quad \text{tal que} \quad \sum_{j=1}^k \lambda_j \mathbf{X}_j \approx \mathbf{0} \quad (1)$$

donde \mathbf{X}_j denota la j -ésima columna de \mathbf{X} .

1.2 Clasificación de la Multicolinealidad

1.2.1 Multicolinealidad Perfecta

La multicolinealidad perfecta representa el caso límite donde existe una relación lineal *exacta* entre las variables explicativas.

Consecuencias de la Colinealidad Perfecta

Si existe una combinación lineal exacta entre las columnas de \mathbf{X} , entonces:

- (I) El determinante $|\mathbf{X}'\mathbf{X}| = 0$
- (II) La matriz $(\mathbf{X}'\mathbf{X})$ es singular (no inversible)
- (III) Los estimadores MCO no existen o son indeterminados
- (IV) Los errores estándar tienden a infinito: $\text{se}(\hat{\beta}_j) \rightarrow \infty$

Demostración. Si $\sum_{j=1}^k \lambda_j \mathbf{X}_j = \mathbf{0}$ con $\boldsymbol{\lambda} \neq \mathbf{0}$, entonces $\mathbf{X}\boldsymbol{\lambda} = \mathbf{0}$, lo que implica que $\text{rango}(\mathbf{X}) < k$. Por consiguiente:

$$\text{rango}(\mathbf{X}'\mathbf{X}) = \text{rango}(\mathbf{X}) < k \quad (2)$$

Esto significa que $\mathbf{X}'\mathbf{X}$ no tiene rango completo, por lo tanto $|\mathbf{X}'\mathbf{X}| = 0$ y la inversa $(\mathbf{X}'\mathbf{X})^{-1}$ no existe. \square

La Trampa de la Variable Dummy

Consideremos un modelo con intercepto y variables dummy para género:

$$Y_i = \beta_1 + \beta_2 D_{1i} + \beta_3 D_{2i} + u_i \quad (3)$$

donde:

- $D_1 = 1$ si es hombre, 0 en caso contrario
- $D_2 = 1$ si es mujer, 0 en caso contrario

Observamos que $D_1 + D_2 = 1$ (la columna del intercepto), generando **colinealidad perfecta**.

Solución: Omitir una de las dummies (categoría de referencia).

1.2.2 Multicolinealidad Aproximada o Imperfecta

En la práctica econométrica, el caso más frecuente es la multicolinealidad *imperfecta*, donde:

$$|\mathbf{X}'\mathbf{X}| \approx 0 \quad (\text{cercano a cero, pero no exactamente cero}) \quad (4)$$

Nota

La multicolinealidad imperfecta es un problema de **grado**, no de existencia. Los estimadores son calculables, pero presentan:

- Varianzas muy grandes
- Alta sensibilidad a cambios en los datos
- Intervalos de confianza excesivamente amplios
- Inestabilidad numérica en la computación de $(\mathbf{X}'\mathbf{X})^{-1}$

1.3 Causas de la Multicolinealidad

La ?? resume las principales fuentes de multicolinealidad en datos económicos.

Tabla 1: Causas Principales de la Multicolinealidad

Causa	Explicación	Ejemplo
Series de tiempo	Variables económicas tienden a moverse juntas temporalmente	PIB, Consumo, Inversión
Especificación del modelo	Inclusión de variables que miden fenómenos similares	Ingreso y Riqueza
Micronumerosidad	Tamaño de muestra pequeño relativo al número de parámetros	$n = 30, k = 25$
Método de recolección	Muestreo en rango limitado de los regresores	Solo familias de alto ingreso
Términos polinomiales	X, X^2, X^3 altamente correlacionados	Especialmente si X tiene rango pequeño
Tendencias comunes	Variables crecen simultáneamente en el tiempo	Variables macroeconómicas

2 Notación Matricial del Modelo de Regresión

2.1 Especificación General del Modelo

Para un modelo de regresión lineal múltiple con k variables explicativas y n observaciones:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \quad (5)$$

donde:

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{n \times 1} \quad \text{Vector de observaciones de la variable dependiente}$$

$$\mathbf{X} = \begin{bmatrix} 1 & X_{12} & X_{13} & \cdots & X_{1k} \\ 1 & X_{22} & X_{23} & \cdots & X_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n2} & X_{n3} & \cdots & X_{nk} \end{bmatrix}_{n \times k} \quad \text{Matriz de diseño}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}_{k \times 1} \quad \text{Vector de parámetros poblacionales}$$

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}_{n \times 1} \quad \text{Vector de perturbaciones estocásticas}$$

2.2 La Matriz de Información

La matriz de información o matriz de momentos cruzados juega un papel central en el análisis de multicolinealidad.

Matriz de Información

La matriz de información se define como:

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} n & \sum_{i=1}^n X_{i2} & \sum_{i=1}^n X_{i3} & \cdots & \sum_{i=1}^n X_{ik} \\ \sum_{i=1}^n X_{i2} & \sum_{i=1}^n X_{i2}^2 & \sum_{i=1}^n X_{i2}X_{i3} & \cdots & \sum_{i=1}^n X_{i2}X_{ik} \\ \sum_{i=1}^n X_{i3} & \sum_{i=1}^n X_{i2}X_{i3} & \sum_{i=1}^n X_{i3}^2 & \cdots & \sum_{i=1}^n X_{i3}X_{ik} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n X_{ik} & \sum_{i=1}^n X_{i2}X_{ik} & \sum_{i=1}^n X_{i3}X_{ik} & \cdots & \sum_{i=1}^n X_{ik}^2 \end{bmatrix}_{k \times k} \quad (6)$$

Propiedades de $\mathbf{X}'\mathbf{X}$

La matriz $\mathbf{X}'\mathbf{X}$ posee las siguientes propiedades:

- (a) **Simetría:** $(\mathbf{X}'\mathbf{X})' = \mathbf{X}'\mathbf{X}$
- (b) **Definida positiva** (bajo ausencia de multicolinealidad perfecta)
- (c) **Dimensión:** $(k \times k)$
- (d) Su determinante mide el grado de colinealidad:

$$|\mathbf{X}'\mathbf{X}| = 0 \iff \text{Multicolinealidad perfecta} \quad (7)$$

2.3 El Estimador de Mínimos Cuadrados Ordinarios**Estimador MCO**

El estimador de mínimos cuadrados ordinarios se obtiene minimizando la suma de residuales al cuadrado:

$$\min_{\beta} \text{SRC} = \min_{\beta} (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta) \quad (8)$$

La solución viene dada por:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} \quad (9)$$

Condición de existencia: $(\mathbf{X}'\mathbf{X})^{-1}$ existe $\iff |\mathbf{X}'\mathbf{X}| \neq 0 \iff \text{rango}(\mathbf{X}) = k$

Nota**Efecto de la multicolinealidad en el estimador MCO:**

Cuando $|\mathbf{X}'\mathbf{X}| \approx 0$:

- Los elementos de $(\mathbf{X}'\mathbf{X})^{-1}$ son muy grandes
- $\hat{\beta}$ se vuelve altamente volátil
- Pequeños cambios en \mathbf{Y} provocan grandes cambios en $\hat{\beta}$
- Problemas de precisión numérica en el cálculo computacional

2.4 Valores Predichos y Residuales

Matrices de Proyección

Los valores predichos y residuales se expresan mediante matrices de proyección:

Valores predichos:

$$\hat{Y} = X\hat{\beta} = X(X'X)^{-1}X'Y = PY \quad (10)$$

donde $P = X(X'X)^{-1}X'$ es la **matriz sombrero** (*hat matrix*).

Residuales:

$$\hat{u} = Y - \hat{Y} = Y - PY = (I - P)Y = MY \quad (11)$$

donde $M = I - P$ es la **matriz aniquiladora** (*annihilator matrix*).

Propiedades de las Matrices de Proyección

Las matrices P y M satisfacen:

- (I) **Simetría:** $P' = P$, $M' = M$
- (II) **Idempotencia:** $PP = P$, $MM = M$
- (III) **Ortogonalidad:** $PM = MP = 0$
- (IV) **Complementariedad:** $P + M = I$
- (V) **Traza:** $\text{tr}(P) = k$, $\text{tr}(M) = n - k$

3 Matriz de Varianza-Covarianza y el Fenómeno de Explosión de Varianza

3.1 Definición y Estructura

La matriz de varianza-covarianza de los estimadores MCO es fundamental para la inferencia estadística.

Matriz de Varianza-Covarianza de $\hat{\beta}$

Bajo los supuestos del modelo lineal clásico:

$$\text{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1} \quad (12)$$

donde $\sigma^2 = \mathbb{E}(u_i^2)$ es la varianza homoscedástica del término de error.

3.2 Expansión Explícita para el Modelo Trivariado

Para el modelo $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$:

$$\sigma^2(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} \text{Var}(\hat{\beta}_1) & \text{Cov}(\hat{\beta}_1, \hat{\beta}_2) & \text{Cov}(\hat{\beta}_1, \hat{\beta}_3) \\ \text{Cov}(\hat{\beta}_2, \hat{\beta}_1) & \text{Var}(\hat{\beta}_2) & \text{Cov}(\hat{\beta}_2, \hat{\beta}_3) \\ \text{Cov}(\hat{\beta}_3, \hat{\beta}_1) & \text{Cov}(\hat{\beta}_3, \hat{\beta}_2) & \text{Var}(\hat{\beta}_3) \end{bmatrix} \quad (13)$$

Nota

Propiedades de la matriz de varianza-covarianza:

- **Simetría:** $\text{Cov}(\hat{\beta}_i, \hat{\beta}_j) = \text{Cov}(\hat{\beta}_j, \hat{\beta}_i)$
- **Diagonal:** Varianzas individuales
- **Fuera de la diagonal:** Covarianzas entre pares de estimadores
- **Definida positiva semidefinida** (bajo supuestos clásicos)

3.3 El Fenómeno de “Explosión de Varianza”

Fórmula de Varianza con Multicolinealidad

La varianza de un coeficiente individual puede descomponerse como:

$$\text{Var}(\hat{\beta}_j) = \frac{\sigma^2}{\sum_{i=1}^n (X_{ij} - \bar{X}_j)^2} \cdot \frac{1}{1 - R_j^2} = \frac{\sigma^2}{\text{STC}_j} \times \text{FIV}_j \quad (14)$$

donde:

- $\text{STC}_j = \sum_{i=1}^n (X_{ij} - \bar{X}_j)^2$ es la suma total de cuadrados de X_j
- R_j^2 es el coeficiente de determinación de la **regresión auxiliar** de X_j sobre todas las demás variables explicativas
- $\text{FIV}_j = (1 - R_j^2)^{-1}$ es el Factor de Inflación de Varianza

Demostración. Partiendo de la expresión matricial $\text{Var}(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$, y utilizando el teorema de Frisch-Waugh-Lovell, la varianza del j -ésimo coeficiente puede expresarse como:

$$\text{Var}(\hat{\beta}_j) = \frac{\sigma^2}{\hat{\mathbf{u}}_j' \hat{\mathbf{u}}_j} \quad (15)$$

donde $\hat{\mathbf{u}}_j$ son los residuales de la regresión de X_j sobre las demás variables. El coeficiente de determinación de esta regresión auxiliar es:

$$R_j^2 = 1 - \frac{\hat{\mathbf{u}}_j' \hat{\mathbf{u}}_j}{(\mathbf{X}_j - \bar{X}_j \mathbf{1})' (\mathbf{X}_j - \bar{X}_j \mathbf{1})} \quad (16)$$

Reorganizando:

$$\hat{\mathbf{u}}_j' \hat{\mathbf{u}}_j = (1 - R_j^2) \sum_{i=1}^n (X_{ij} - \bar{X}_j)^2 \quad (17)$$

Sustituyendo en la expresión de la varianza:

$$\text{Var}(\hat{\beta}_j) = \frac{\sigma^2}{(1 - R_j^2) \sum_{i=1}^n (X_{ij} - \bar{X}_j)^2} \quad (18)$$

□

3.4 Análisis del Factor de Inflación

La ?? ilustra el efecto de diferentes niveles de colinealidad sobre la varianza.

Tabla 2: Relación entre R_j^2 , FIV y Inflación de Varianza

R_j^2	FIV _j	Factor de Inflación	Interpretación
0.00	1.00	×1	Sin correlación (ideal)
0.25	1.33	×1,33	Duplica la varianza en 33 %
0.50	2.00	×2	Duplica la varianza
0.75	4.00	×4	Cuadruplica la varianza
0.80	5.00	×5	Quintuplica la varianza
0.90	10.00	×10	Umbral crítico
0.95	20.00	×20	Problema severo
0.99	100.00	×100	Catastrófico
1.00	∞	∞	Colinealidad perfecta

Advertencia

Consecuencia directa del FIV alto:

Si $R_j^2 \rightarrow 1$, entonces:

$$\text{Var}(\hat{\beta}_j) \rightarrow \infty \implies \text{se}(\hat{\beta}_j) \rightarrow \infty \quad (19)$$

Esto invalida las pruebas de hipótesis t :

$$t = \frac{\hat{\beta}_j}{\text{se}(\hat{\beta}_j)} \rightarrow 0 \quad (\text{incluso si } \hat{\beta}_j \neq 0) \quad (20)$$

Resultado: Se **acepta incorrectamente** $H_0 : \beta_j = 0$ cuando el parámetro es significativo.

3.5 Estimación Práctica

En la práctica, σ^2 es desconocido y se estima mediante:

$$\hat{\sigma}^2 = s^2 = \frac{\sum_{i=1}^n \hat{u}_i^2}{n - k} = \frac{\text{SRC}}{n - k} \quad (21)$$

Por lo tanto, la matriz de varianza-covarianza estimada es:

$$\widehat{\text{Var}}(\hat{\beta}) = s^2 (\mathbf{X}'\mathbf{X})^{-1} \quad (22)$$

Los errores estándar se obtienen de:

$$\text{se}(\hat{\beta}_j) = \sqrt{s^2 \cdot [(\mathbf{X}'\mathbf{X})^{-1}]_{jj}} \quad (23)$$

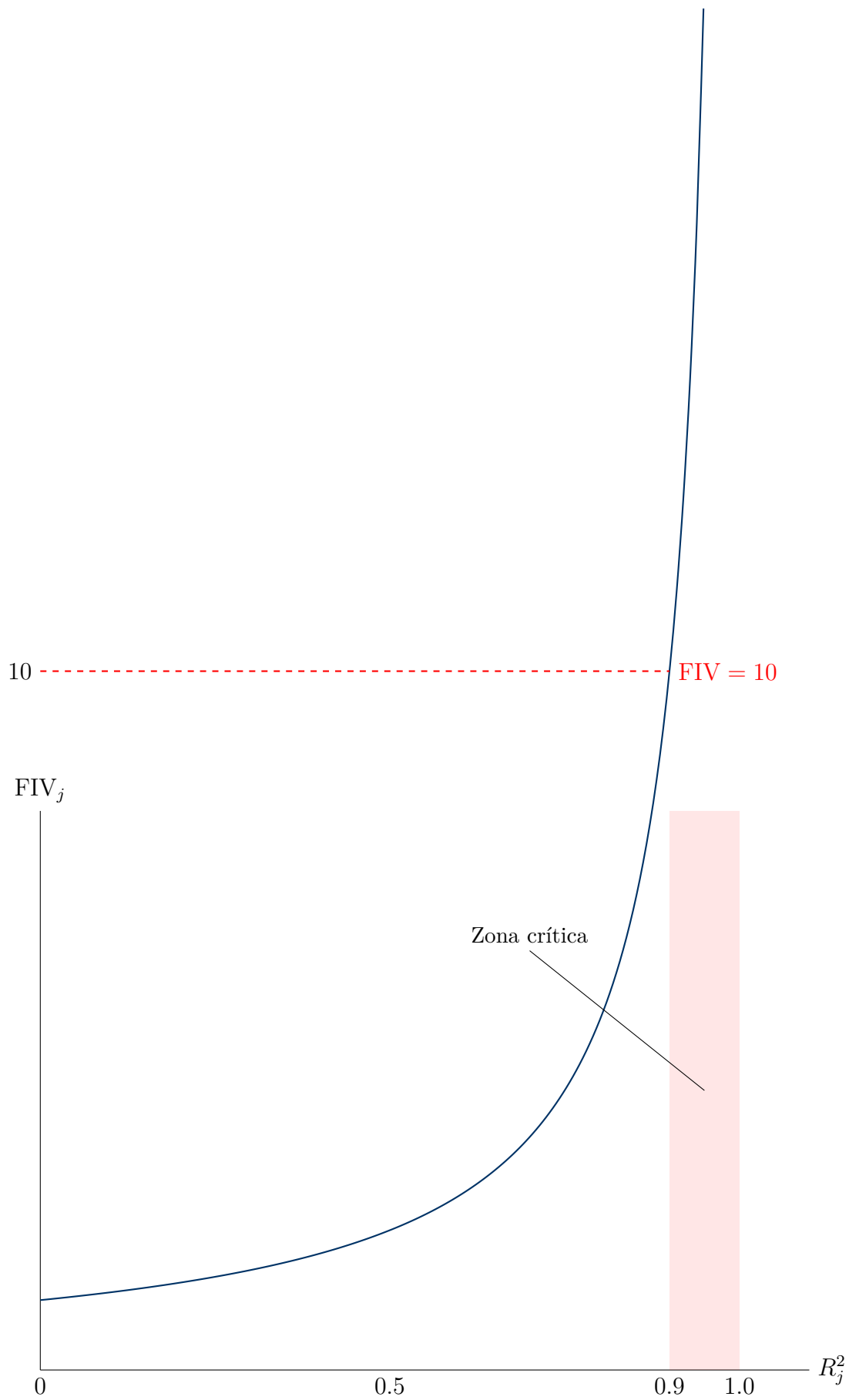


Figura 1: Relación entre R_j^2 y el Factor de Inflación de Varianza

4 Diagnósticos de Multicolinealidad

4.1 Síntomas Observables

4.1.1 La Paradoja del R^2 Alto con Coeficientes No Significativos

Este es el síntoma más característico de multicolinealidad severa.

Configuración Típica de Multicolinealidad

Consideremos el modelo:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + u \quad (24)$$

Resultados contradictorios:

- Coeficiente de determinación: $R^2 = 0,95$ (muy alto)
- Prueba F global: $F = 150,3$ con $p < 0,001 \implies$ El modelo es significativo
- Pruebas t individuales:

$$t(\beta_2) = 1,45 \quad (p = 0,152) \implies \text{No significativo}$$

$$t(\beta_3) = 0,87 \quad (p = 0,389) \implies \text{No significativo}$$

$$t(\beta_4) = -1,12 \quad (p = 0,268) \implies \text{No significativo}$$

Diagnóstico: Multicolinealidad severa entre X_2 , X_3 y X_4 .

4.1.2 Explicación Matemática

El estadístico F prueba la hipótesis conjunta:

$$H_0 : \beta_2 = \beta_3 = \beta_4 = 0 \quad (25)$$

mediante:

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \quad (26)$$

Este puede ser grande incluso cuando los estadísticos t individuales son pequeños debido a las **altas covarianzas** entre los estimadores, consecuencia de la multicolinealidad.

4.2 Factor de Inflación de Varianza (FIV)

4.2.1 Definición y Cálculo

Factor de Inflación de Varianza

Para cada variable X_j ($j = 2, 3, \dots, k$), el FIV se define como:

$$\text{FIV}_j = \frac{1}{1 - R_j^2} \quad (27)$$

donde R_j^2 se obtiene de la **regresión auxiliar**:

$$X_j = \delta_0 + \sum_{\substack{i=1 \\ i \neq j}}^k \delta_i X_i + v_j \quad (28)$$

Algorithm 1 Procedimiento para Calcular el FIV

Require: Matriz de datos \mathbf{X} con k variables explicativas

Ensure: Vector de FIV para cada variable

```

1: for  $j = 2$  to  $k$  do
2:   Regresar  $X_j$  sobre todas las demás variables  $X_i$  ( $i \neq j$ )
3:   Calcular  $R_j^2$  de esta regresión auxiliar
4:    $\text{FIV}_j \leftarrow \frac{1}{1 - R_j^2}$ 
5:   if  $\text{FIV}_j > 10$  then
6:     Señalar variable  $X_j$  como problemática
7:   end if
8: end for
9: return Vector  $[\text{FIV}_2, \text{FIV}_3, \dots, \text{FIV}_k]$ 

```

4.2.2 Criterios de Decisión

Tabla 3: Criterios de Interpretación del FIV

Valor de FIV_j	Diagnóstico	Acción Recomendada
< 5	Colinealidad leve	Ninguna acción
$5 - 10$	Colinealidad moderada	Monitorear
> 10	Colinealidad grave	Acción correctiva urgente
> 100	Colinealidad catastrófica	Rediseñar modelo

Nota

Nota sobre umbrales:

- Algunos autores usan el umbral de $\text{FIV} > 5$
- En series de tiempo, umbrales de 20-30 pueden ser tolerables
- El contexto del problema debe considerarse

4.3 Factor de Tolerancia (TOL)

Factor de Tolerancia

El factor de tolerancia es el inverso del FIV:

$$\text{TOL}_j = \frac{1}{\text{FIV}_j} = 1 - R_j^2 \quad (29)$$

Tabla 4: Interpretación del Factor de Tolerancia

Valor de TOL_j	Interpretación
$\rightarrow 1$	Independencia total (ortogonalidad)
$> 0,20$	Aceptable
$0,10 - 0,20$	Zona de precaución
$< 0,10$	Problema serio
$\rightarrow 0$	Colinealidad perfecta

4.4 Test de Farrar-Glauber

El test de Farrar-Glauber es un procedimiento de tres etapas para detectar, localizar y cuantificar la multicolinealidad.

4.4.1 Etapa 1: Test Chi-Cuadrado (Detección Global)

Estadístico de Bartlett

El estadístico global para detectar la presencia de multicolinealidad es:

$$\chi_{\text{calc}}^2 = - \left[n - 1 - \frac{1}{6}(2k + 5) \right] \ln |R_{XX}| \quad (30)$$

donde:

- n = número de observaciones
- k = número de variables explicativas (excluyendo intercepto)
- $|R_{XX}|$ = determinante de la matriz de correlación entre regresores

Distribución bajo H_0 :

$$\chi_{\text{calc}}^2 \sim \chi_{\left[\frac{k(k-1)}{2}\right]}^2 \quad (31)$$

Hipótesis:

H_0 : Las variables explicativas son ortogonales (no hay multicolinealidad)

H_1 : Existe multicolinealidad

Regla de decisión:

$$\text{Rechazar } H_0 \text{ si } \chi_{\text{calc}}^2 > \chi_{\alpha, gl}^2 \quad (32)$$

Ejemplo Numérico Completo del Test de Farrar-Glauber**Datos:**

- $n = 50$ observaciones
- $k = 3$ variables explicativas (excluyendo intercepto)
- $|R_{XX}| = 0,10$

Paso 1: Calcular el factor de corrección

$$\begin{aligned}
 \text{Factor} &= n - 1 - \frac{1}{6}(2k + 5) \\
 &= 50 - 1 - \frac{1}{6}(2 \times 3 + 5) \\
 &= 49 - \frac{11}{6} = 47,167
 \end{aligned} \tag{33}$$

Paso 2: Calcular el estadístico

$$\begin{aligned}
 \chi_{\text{calc}}^2 &= -47,167 \times \ln(0,10) \\
 &= -47,167 \times (-2,303) \\
 &= 108,65
 \end{aligned} \tag{34}$$

Paso 3: Grados de libertad

$$gl = \frac{k(k-1)}{2} = \frac{3(3-1)}{2} = 3 \tag{35}$$

Paso 4: Valor crítico Para $\alpha = 0,05$ y $gl = 3$:

$$\chi_{0,05,3}^2 = 7,815 \tag{36}$$

Paso 5: Decisión

$$\chi_{\text{calc}}^2 = 108,65 > 7,815 = \chi_{\text{crítico}}^2 \tag{37}$$

Conclusión: Se rechaza $H_0 \implies$ **Existe multicolinealidad significativa**

4.4.2 Etapa 2: Test F (Localización de Variables Problemáticas)

Una vez detectada la presencia de multicolinealidad, el siguiente paso es identificar cuáles variables específicas están involucradas.

Test F para Regresiones Auxiliares

Para cada variable X_j , ejecutamos la regresión auxiliar:

$$X_j = \delta_0 + \sum_{\substack{i=1 \\ i \neq j}}^k \delta_i X_i + v_j \quad (38)$$

El estadístico F para probar si X_j está linealmente relacionada con las demás es:

$$F_j = \frac{R_j^2 / (k - 2)}{(1 - R_j^2) / (n - k + 1)} \quad (39)$$

Distribución bajo H_0 (variable no colineal):

$$F_j \sim F_{(k-2, n-k+1)} \quad (40)$$

Interpretación:

- Si F_j es **significativo** $\implies X_j$ está altamente correlacionada con otras variables
- Si F_j es **no significativo** $\implies X_j$ no presenta colinealidad problemática

4.4.3 Etapa 3: Test t (Identificación de Pares Colineales)

La tercera etapa identifica qué pares específicos de variables están causando el problema.

Correlación Parcial

La correlación parcial entre X_i y X_j controlando por las demás variables se calcula como:

$$r_{ij \cdot \text{resto}} = \frac{-c_{ij}}{\sqrt{c_{ii}c_{jj}}} \quad (41)$$

donde c_{ij} son los elementos de la matriz $(R_{XX})^{-1}$.

Estadístico t para Correlaciones Parciales

El estadístico para probar la significancia de la correlación parcial es:

$$t_{ij} = r_{ij \cdot \text{resto}} \sqrt{\frac{n - k}{1 - r_{ij \cdot \text{resto}}^2}} \quad (42)$$

Distribución: $t_{ij} \sim t_{(n-k)}$

Interpretación: Si t_{ij} es significativo \implies El par (X_i, X_j) está altamente correlacionado

4.5 Análisis de Valores Propios

El análisis espectral de la matriz $\mathbf{X}'\mathbf{X}$ proporciona información adicional sobre la estructura de la multicolinealidad.

4.5.1 Descomposición Espectral

Descomposición en Valores Propios

La matriz $\mathbf{X}'\mathbf{X}$ (o su versión estandarizada) puede descomponerse como:

$$\mathbf{X}'\mathbf{X} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}' \quad (43)$$

donde:

- \mathbf{Q} es una matriz ortogonal ($k \times k$) de vectores propios
- $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_k)$ con $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k \geq 0$

4.5.2 Número de Condición e Índice de Condición

Número e Índice de Condición

El **número de condición** se define como:

$$\kappa = \frac{\lambda_{\text{máx}}}{\lambda_{\text{mín}}} \quad (44)$$

El **índice de condición** es la raíz cuadrada del número de condición:

$$\text{CI} = \sqrt{\kappa} = \sqrt{\frac{\lambda_{\text{máx}}}{\lambda_{\text{mín}}}} \quad (45)$$

Tabla 5: Criterios de Belsley, Kuh y Welsch para el Índice de Condición

Valor de CI	Diagnóstico
< 10	Sin multicolinealidad
10 – 30	Multicolinealidad moderada a fuerte
≥ 30	Multicolinealidad severa
> 100	Multicolinealidad extrema

5 Remedios y Soluciones para la Multicolinealidad

5.1 Filosofía General: ¿Actuar o No Actuar?

Nota

La decisión de aplicar remedios depende fundamentalmente del **objetivo del análisis**:

Objetivo	¿Importa?	Acción
Predicción	NO	No hacer nada
Inferencia	SÍ	Aplicar remedios
Interpretación económica	SÍ	Aplicar remedios

Advertencia

Recordatorio fundamental:

La multicolinealidad **NO** sesga los estimadores MCO. Los estimadores permanecen:

- **Insesgados:** $\mathbb{E}(\hat{\beta}) = \beta$
- **BLUE:** Siguen siendo los mejores estimadores lineales insesgados
- **Pero ineficientes:** Alta varianza reduce precisión

A Tablas Estadísticas

A.1 Distribución Chi-Cuadrado

Tabla 6: Valores Críticos de χ^2 para el Test de Farrar-Glauber

gl	$\alpha = 0,10$	$\alpha = 0,05$	$\alpha = 0,01$
1	2.706	3.841	6.635
2	4.605	5.991	9.210
3	6.251	7.815	11.345
4	7.779	9.488	13.277
5	9.236	11.070	15.086
6	10.645	12.592	16.812
10	15.987	18.307	23.209
15	22.307	24.996	30.578
20	28.412	31.410	37.566

B Código R para Diagnósticos

Listing 1: Cálculo de FIV en R

```
1 # Cargar librería
2 library(car)
3
4 # Supongamos que tenemos un modelo
5 modelo <- lm(Y ~ X2 + X3 + X4, data = datos)
6
7 # Calcular FIV
8 vif(modelo)
9
10 # Calcular Tolerancia
11 1 / vif(modelo)
12
13 # Matriz de correlación
14 cor(datos[, c("X2", "X3", "X4")])
15
16 # Determinante de la matriz de correlación
17 det(cor(datos[, c("X2", "X3", "X4")]))
```