REVISTA COLOMBIANA DE ESTADISTICA Nº 2 1980

MULTICOLINEALIDAD

POR : ALBERTO VARGAS N.
Instructor Asistente U.N.

ISMAEL RODRIGUEZ G.
Instructor Asociado U.N.

A. *Presentación del Problema.*

Multicolinealidad es el nombre dado al problema que aparece cuando alguna o todas las variables independientes en una relación están altamente correlacionadas una con otra. Aquí llega a ser muy difícil sino imposible, detectar sus influencias por separado y obtener estimadores razonablemente precisos de sus efectos relativos, esto es, los coeficientes de regresión parcial pueden no ser significantes aunque exista una relación estadística entre la variable dependiente y el conjunto de variables independientes.

Este fenómeno es típico en problemas económicos y de mercadeo, por ejemplo las variables independientes ingreso familiar y activos tenderán a ser altamente correlacionadas.

B. *Casos que se presentan.*

Se pueden considerar varios casos distintos :

Caso I : Existe una relación lineal exacta entre 2 variables (Perfecta multicolinealidad).

Supóngase que en el siguiente modelo :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$

Se cumple la siguiente relación exacta entre X_1 y X_2

$$X_1 = k_1 + k_2 X_2$$

Tomando desviaciones con respecto a la media aritmética tenemos :

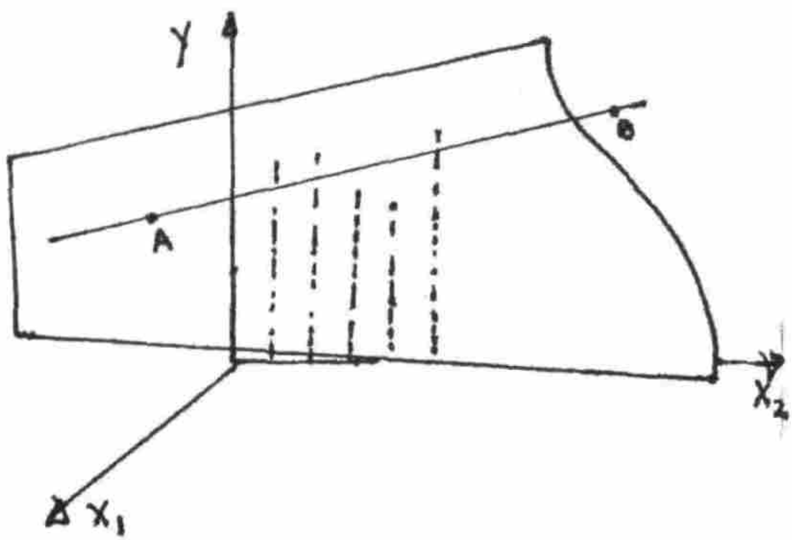
$$\mathbf{X}_1 = k_2\mathbf{X}_2 \quad y \quad \mathbf{X}'\mathbf{X} = \begin{pmatrix} \Sigma X_1^{2} & \Sigma X_1 X_2 \\ \Sigma X_2 X_1 & \Sigma X^2{}_2 \end{pmatrix} = \begin{pmatrix} k_2^{\,2}\, \Sigma X^2{}_2 & k_2 \Sigma X^2{}_2 \\ k\, \Sigma X_2^2 & \Sigma X_2{}^2 \end{pmatrix}$$

Entonces el determinante de $\mathbf{X}'\mathbf{X}$ es

$$|\mathbf{X}'\mathbf{X}| = k_2{}^2\, (\Sigma X_2{}^2)^2 - k_2{}^2\, (\Sigma X^2{}_2)^2 = 0$$

Esto implique que $(\mathbf{X}'\mathbf{X})^{-1}$ no existe y así no podemos estimar β .

La interpretación geométrica de este caso es interesante



La dispersión de puntos en el plano X_1X_2 ocurre exclusivamente sobre la línea $X_1 = k_1 + k_2X_2$. Los valores de Y simplemente elevan estos puntos dejándolos en un plano vertical; arriba y abajo de una línea recta en el espacio tridimensional. Al tratar de explicar Y, el hecho de que las variables X_1 y X_2 estén relacionadas literalmente hace que se pierda una dimensión, y el mejor ajuste mínimo cuadrático para Y no es un plano sino una línea, la línea AB para este caso.

Si se mantiene la misma estructura de correlación, es decir si se conserva la relación de X_1 y X_2, y nos limitamos a predecir sobre la línea AB de la figura no se presentan problemas especiales pero no podemos examinar como X_1 y X_2 afectan a Y individualmente; cualquier intento de definir β_1, β_2, los efectos marginales de X_1, X_2, implicaría tratar de explicar Y con un plano en lugar de con una línea y se encontraría que hay infinitos planos que pasan a través de la línea AB; cada uno de los cuales produce uha suma de cuadrados de los errores igual y pór lo tanto obtenemos el mismo ajuste para todos.

Caso II : Un caso menos extremo pero mucho más probable es el caso en que los valores de las variables independientes que aparecen en nuestra muestra están altamente correlacionadas, pero no perfectamente correlacionadas. Lo que sucede, en otros términos, es que una o más variables tienden a ser función lineal de una u otras variables con una pequeña disturbancia.

Supongamos el mismo modelo $Y_i = \beta_o + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$ con $X_{1i} = k_1 + k_2 X_{2i} + U_i$, $i = 1, 2, \ldots, n$ donde U_i es una pequeña disturbancia tal que $E(U_i) = 0$ y ademas $\Sigma X_{2i}U_i = 0$ $X_{1i} = k_2X_{2i} + U_i$ entonces :

$$\mathcal{X}^1\mathcal{X} = \begin{pmatrix} \Sigma k_2^2 X_{2i}^2 + \Sigma U_i^2 & \Sigma X_{2i}(k_2 X_{2i} + U_i) \\ \\ \Sigma X_{2i}(k_2 X_{2i} + U_i) & \Sigma X_{2i}^2 \end{pmatrix}$$

$$\mathcal{X}^1\mathcal{X} = \begin{pmatrix} k_2^2 \Sigma X_{2i}^2 + \Sigma U_i^2 & k_2 \Sigma X_{2i}^2 \\ \\ k_2 \Sigma X_{2i}^2 & \Sigma X_{2i}^2 \end{pmatrix}$$

$$|\mathcal{X}^1\mathcal{X}| = k_2^2(\Sigma X_{2i}^2)^2 + (\Sigma X_{2i}^2)(\Sigma U_i^2) - k_2^2(\Sigma X_{2i}^2)^2 = (\Sigma X_{2i}^2)(\Sigma U_i^2)$$

$|\mathcal{X}^1\mathcal{X}|$ es por tanto muy pequeño.

Entonces :

$$(\mathcal{X}'\mathcal{X})^{-1} = \frac{1}{(\Sigma X_{2i}^2)(\Sigma U_i^2)} \begin{pmatrix} \Sigma X_{2i}^2 & -k_2 \; \Sigma X_{2i}^2 \\ -k_2 \; \Sigma X_{2i}^2 & k_2 \; \Sigma \; X_{2i}^2 + \Sigma U_i^2 \end{pmatrix}$$

Luego los elementos de $(\mathcal{X}'\mathcal{X})^{-1}$ son muy grandes y por tanto los elementos de la *matriz* de covarianzas de $\hat{\beta}$ serán también muy grandes, así como también los *errores standar* de los $\hat{\beta}_i$ *lo cual nos* da valores poblacionales muy inciertos.

Debemos observar que es posible tener una relación que se ajuste muy bien, es decir, tener un R^2 alto, y aún así, al aplicar pruebas t para los β_i se puede obtener que ninguna de éstas son significativas por separado.

C. *Un método para detectar multicolinealidad.*

Como ya se ha dicho, se puede sospechar que existe una severa multicolinealidad cuando las *variables en conjunto son significativas* pero *individualmente* no lo son.

Prueba empírica de Klein : Para examinar la severidad de la colinealidad por pares de las variables independientes, Klein sugiere construir la matriz de correlaciones de las variables :

$$\overline{R} = (\rho_{ij}) \; ; \; \rho i_j = \frac{\sum_{k=1}^{n} (X_{ki} - \overline{X}_i)(X_{kj} - \overline{X}_j)}{\sqrt{\sum_{k=1}^{n} (X_{ki} - \overline{X}_i)^2 \; \sum_{k=1}^{n} (X_{kj} - \overline{X}_j)^2}}$$

$\overline{R}$ es la matriz de correlaciones.

Se sugiere como regla que si el $\max_j |\rho_{ij}| < R \;\; \forall_i$ donde R es el coeficiente de correlación múltiple, la multicolinealidad es tolerable.

D. *Soluciones al Problema de Multicolinealidad.*

1. *Eliminación de una o varias variables independientes del modelo.*

Se propone esta solución para disminuir la multicolinealidad y así reducir los errores standar de los coeficientes de regresión estimados de las variables que restan en el modelo. Esta acción sin embargo no ayuda a evaluar los efectos de las variables independientes por 2 razones. Primero : No se obtiene información acerca de las variables dejadas. Segundo: Las magnitudes de los coeficientes de Regresión para las variables independientes que quedan en el modelo son afectadas por las variables independientes no incluidas en el modelo; la influencia de las variables eliminadas pasaal término de error, de esta forma, el término de error quedará correlacionado con las otras variables y el estimador mínimo cuadrado deja de ser insesgado.

2. *Uso de información extraña.*

En algunos casos es posible obtener información adicional acerca de los parámetros del modelo; esta información puede ser conocimiento de la razón de algunos coeficientes; de los valores de algunos coeficientes, de los valores de alguna combinación de los coeficientes; o simplemente del signo de algunos coeficientes.

Este conocimiento puede provenir de investigaciones empíricas anteriores o de otras muestras.

Restricciones Lineales exactas : Mínimos cuadrados restringidos. Supóngamos que la información extraña es una restricción lineal exacta sobre los coeficientes.

$$\gamma = R\,\beta \tag{1}$$

Donde γ es $J \times 1$ conocido y R es una matriz $J \times k$ conocida, de rango $J < k$ y así tenemos J restricciones independientes sobre los elementos de β.

Casos especiales son :

i) Conocimiento de los valores de algunos elementos de β por ejemplo

$$\beta_1 = \beta^{*}_1 \quad \text{con} \quad \gamma = (\beta^{*}_1) \qquad R = (1, 0, \ldots, 0)$$

ii) Conocimiento de la razón de algunos elementos de β por ejemplo

$$\beta_1 / \beta_2 = c_1 \quad \text{y} \quad \beta_3 / \beta_2 = c_3 \quad \text{con}$$

$$\gamma = \binom{0}{0} \qquad R = \begin{pmatrix} 1 - c_1, 0 \ldots, 0 \\ 0, - c_3, 1\, 0 \ldots, 0 \end{pmatrix}$$

iii) Conocimiento del valor de una combinación lineal de los valores de los elementos de β por ejemplo

$$\beta_1 + \ldots + \beta_k = 1 \quad \text{con} \quad \gamma = (1) \qquad R = (1, 1, \ldots, 1)$$

Para incorporar esta infomación se propone el método de los mínimos cuadra dos restringidos.

Debe obtenerse el β que minimice la suma de cuadrados del error, es decir, $(\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)$ sujeto a la restricción $k_{\beta} - \gamma = 0$

Por consiguiente minimizamos

$$S = (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta) - 2\lambda'(R\beta - \gamma) \tag{2}$$

Donde λ es un vector $J \times 1$ de multiplicadores de Langrange. Derivando S con respecto a β e igualando a cero tenemos

$$\frac{1}{2}\frac{\partial S}{\partial \beta} = - \mathbf{X}'\mathbf{Y} + (\mathbf{X}'\mathbf{X})\beta^{*} - R'\lambda^{*} = 0$$

donde β^* y λ^* son los valores que minimizan a S.

Entonces $\beta^* = \widehat{\beta} + (\mathbf{X}'\mathbf{X})^{-1} R' \lambda^*$ (3)

Donde $\widehat{\beta}$ es $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$

Premultiplicando por R se obtiene

$$R\beta^* = R\widehat{\beta} + R(\mathbf{X}'\mathbf{X})^{-1} R' \lambda^* \qquad (4)$$

Imponiendo la restricción $R\beta^* = \gamma$ da :

$$\gamma = R\widehat{\beta} + R(\mathbf{X}'\mathbf{X})^{-1} R' \lambda^* \qquad (5)$$

$$\lambda^* = [R(\mathbf{X}'\mathbf{X})^{-1} R']^{-1} [\gamma - R\widehat{\beta}] \qquad (6)$$

Reemplazando (6) en (3) tenemos

$$\beta^* = \widehat{\beta} + (\mathbf{X}'\mathbf{X})^{-1} R' [R(\mathbf{X}'\mathbf{X})^{-1} R']^{-1} [\gamma - R\widehat{\beta}] \qquad (7)$$

Se ve que el estimador restringido β^* difiere del no restringido $\widehat{\beta}$ por una función lineal de la cantidad $\gamma - R\widehat{\beta}$ por lo cual $\widehat{\beta}$ decae para satisfacer la restricción.

Propiedades de β^*

$$\widehat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' [\mathbf{X}\beta + \epsilon] = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon$$

Reemplazando en (7) tenemos :

$$\beta^* = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon + (\mathbf{X}'\mathbf{X})^{-1} R' [R(\mathbf{X}'\mathbf{X})^{-1} R']^{-1} [\gamma - R\beta - R(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon]$$

$$\beta^{*} = \beta + \{ I - (\mathbf{X}'\mathbf{X})^{-1} R' [R(\mathbf{X}'\mathbf{X})^{-1} R']^{-1} R \} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \cdot \tag{8}$$

Ya que $\gamma - R\beta = 0$

Entonces

$E[\beta^{*}] = \beta$ Luego β^{*} es un estimador insesgado de β

$$\overset{\bullet}{V} = E[(\beta^{*} - \beta)(\beta^{*} - \beta)'] =$$

$$E[\{I - (\mathbf{X}'\mathbf{X})^{-1} R' [R(\mathbf{X}'\mathbf{X})^{-1} R']^{-1} R\} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \epsilon \epsilon' \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1}$$

$$\{I - R[R(\mathbf{X}'\mathbf{X})^{-1} R']^{-1} R(\mathbf{X}'\mathbf{X})^{-1}\}] =$$

$$= \sigma^{2} I \{(\mathbf{X}'\mathbf{X})^{-1} - (\mathbf{X}'\mathbf{X})^{-1} R' [R(\mathbf{X}'\mathbf{X})^{-1} R']^{-1} R (\mathbf{X}'\mathbf{X})^{-1}\}$$

$$= \sigma^{2} (\mathbf{X}'\mathbf{X})^{-1} \{I - R' [R(\mathbf{X}'\mathbf{X})^{-1} R']^{-1} R (\mathbf{X}'\mathbf{X})^{-1}\}$$

Luego $V^{*} = V - VR'(RVR')^{-1} RV$ donde $V = \sigma^{2}(\mathbf{X}'\mathbf{X})^{-1}$

Es fácil ver que hay una ganancia en eficiencia. Se sabe que V es definida positiva; puesto que R' es una matriz $k \times J$ de rango J, se sigue que RVR' es definida positiva y por tanto $(RVR')^{-1}$ también es definida positiva. Además $(RV)'(RVR')^{-1}RV$ es definida no negativa, así cada elemento de la diagonal de V^{*} es menor o igual que el correspondiente elemento de V.

Concluimos que la varianza de cada elemento β^{*} es menor o igual que la varianza del correspondiente de $\hat{\beta}$

Además β^{*} es el BLUE de β

En la práctica puede ser más conveniente computar β^{*}usando las restricciones para eliminar algunos elementos de β entonces se aplican mínimos cuadrados ordinarios y finalmente imponiendo la restricción se estiman los elementos de β que restan

Ejemplo : Supongamos el modelo $Y = \beta_1 X_1 + \beta_2 X_2 + \epsilon$ y sea la restricción $1 = -2\beta_1 + \beta_2$ sustituyendo la restricción en el modelo se tiene :

$$Y = \beta_1 X_1 + (2\beta_1 + 1) X_2 + \epsilon \qquad \text{o,}$$

$$Y - X_2 = \beta_1 (X_1 + 2X_2) + \epsilon$$

Lo cual puede escribirse como

$$Y^{*} = \beta_1 X^{*} + \epsilon$$

Regresando Y^{*} sobre X^{*}obtenemos β^{*}_1 e imponiendo la restricción da

$$\beta^{*}_2 = 2\beta^{*}_1 + 1$$

BIBLIOGRAFIA

1. Johnston, D. S. ***Econometric Methods.*** Mc. Graw Hill Book Co., Inc., New York, 1963.

2. Neter, J., and W. Wasserman. ***Applied linear models - Regression, Analysis of variance, and Experimental Designs.*** Richard D. Irwin, Inc. Homewood, Illinois, 1974.

3. Goldberger, A.S. ***Econometric Theory.*** John Wiley Sons. Inc. New York, 1964

4. Aldana V. Eduardo. ***Modelos Estadísticos.***