

3. MODELO DE REGRESIÓN LINEAL MULTIPLE

3.1 Notación matricial

Sean

\mathbf{Y} = Vector de observaciones de la variable dependiente Y .

\mathbf{X} = Matriz de observaciones de las variables independientes X_i , $i = 1, \dots, n$.

$\boldsymbol{\beta}$ = Vector de parámetros.

$\boldsymbol{\varepsilon}$ = Vector de errores.

La notación matricial aplicada al caso de regresión lineal simple es la siguiente

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ \dots \\ Y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & X_1 \\ \dots & \dots \\ 1 & X_n \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \dots \\ \varepsilon_n \end{pmatrix}.$$

En este caso, \mathbf{Y} es de dimensión $n \times 1$, \mathbf{X} es de dimensión $n \times 2$, $\boldsymbol{\beta}$ es de dimensión 2×1 y $\boldsymbol{\varepsilon}$ es de dimensión $n \times 1$.

La ecuación en la cual se basa el modelo de Regresión Lineal Múltiple es

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (3.1)$$

y los supuestos que dan soporte al modelo son los siguientes

$$E(\boldsymbol{\varepsilon}) = \mathbf{0} \quad (3.2)$$

y

$$\text{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}, \quad (3.3)$$

donde \mathbf{I} es la matriz identidad de dimensión $n \times n$.

Con respecto a los supuestos mencionados anteriormente, explícitamente se tiene que

$$E(\boldsymbol{\varepsilon}) = E \begin{pmatrix} \varepsilon_1 \\ \dots \\ \varepsilon_n \end{pmatrix} = \begin{pmatrix} E(\varepsilon_1) \\ \dots \\ E(\varepsilon_n) \end{pmatrix} = \begin{pmatrix} 0 \\ \dots \\ 0 \end{pmatrix} = \mathbf{0}$$

y

$$\text{Var}(\boldsymbol{\varepsilon}) = E[(\boldsymbol{\varepsilon} - \mathbf{0})(\boldsymbol{\varepsilon} - \mathbf{0})']$$

$$= E(\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}')$$

$$= \begin{pmatrix} E(\varepsilon_1^2) & E(\varepsilon_1 \varepsilon_2) & \dots & E(\varepsilon_1 \varepsilon_n) \\ E(\varepsilon_2 \varepsilon_1) & E(\varepsilon_2^2) & \dots & E(\varepsilon_2 \varepsilon_n) \\ \dots & \dots & \dots & \dots \\ E(\varepsilon_n \varepsilon_1) & E(\varepsilon_n \varepsilon_2) & \dots & E(\varepsilon_n^2) \end{pmatrix}$$

$$= \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix}$$

$$= \sigma^2 \mathbf{I}.$$

En el caso de regresión lineal simple se tienen las siguientes ecuaciones

$$Y_1 = \beta_0 + \beta_1 X_1 + \varepsilon_1$$

...

$$Y_n = \beta_0 + \beta_1 X_n + \varepsilon_n$$

de esta manera, con la notación matricial se obtienen los resultados que siguen

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{pmatrix}, \quad \mathbf{X}'\mathbf{Y} = \begin{pmatrix} \sum Y_i \\ \sum X_i Y_i \end{pmatrix}$$

y

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{Y}$$

que equivalen a

$$\begin{aligned} nb_0 + \sum X_i b_1 &= \sum Y_i \\ \sum X_i b_0 + \sum X_i^2 b_1 &= \sum X_i Y_i \end{aligned}$$

lo que conduce a las Ecuaciones Normales.

Además, al resolver la igualdad matricial para \mathbf{b} se obtiene que

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad (3.4)$$

con

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{n\sum (X_i - \bar{X})^2} \begin{pmatrix} \sum X_i^2 & -\sum X_i \\ -\sum X_i & n \end{pmatrix}$$

siempre que $\mathbf{X}'\mathbf{X}$ sea una matriz no-singular.

Si se hace el desarrollo de la ecuación (3.4) se obtiene el siguiente resultado

$$\begin{aligned} \mathbf{b} &= \frac{1}{\sum (X_i - \bar{X})^2} \begin{pmatrix} \sum X_i^2/n & -\bar{X} \\ -\bar{X} & 1 \end{pmatrix} \begin{pmatrix} \sum Y_i \\ \sum X_i Y_i \end{pmatrix} \\ &= \begin{pmatrix} \frac{\bar{Y} \sum (X_i - \bar{X})^2 + \bar{Y}n\bar{X}^2 - \bar{X} \sum X_i Y_i}{\sum (X_i - \bar{X})^2} \\ \frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\sum (X_i - \bar{X})^2} \end{pmatrix} \\ &= \begin{pmatrix} \bar{Y} - b_1 \bar{X} \\ \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \end{pmatrix} \quad (3.5) \end{aligned}$$

que son los estimadores de MCO obtenidos previamente.

En el caso general se obtiene

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

donde \mathbf{Y} es de dimensión $n \times 1$, \mathbf{X} es de dimensión $n \times p$, $\boldsymbol{\beta}$ es de dimensión $p \times 1$ y $\boldsymbol{\varepsilon}$ es de dimensión $n \times 1$, con p el número de parámetros del modelo ($1 \leq p \leq n$).

Cabe mencionar la existencia de dos versiones del modelo de regresión lineal múltiple. En la primera versión se tiene el modelo en el cual no se incluye el término constante, es decir, el valor de la ordenada al origen, pues éste se considera cero de antemano (por razones de carácter teórico del fenómeno en estudio). La expresión para este modelo es

$$Y = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon \quad (3.6)$$

En la segunda versión se tiene el modelo que incluye a la ordenada al origen. En tal caso la expresión es

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_{p-1} X_{p-1} + \varepsilon. \quad (3.7)$$

Nótese que en ambos casos aparecen p parámetros. Para el primer caso la matriz de observaciones \mathbf{X} es

$$\mathbf{X} = \begin{pmatrix} X_{1,1} & X_{2,1} & \dots & X_{p,1} \\ X_{1,2} & X_{2,2} & \dots & X_{p,2} \\ \dots & \dots & \dots & \dots \\ X_{1,n} & X_{2,n} & \dots & X_{p,n} \end{pmatrix} \quad (3.8)$$

y en el segundo caso

$$\mathbf{X} = \begin{pmatrix} 1 & X_{1,1} & \dots & X_{p-1,1} \\ 1 & X_{1,2} & \dots & X_{p-1,2} \\ \dots & \dots & \dots & \dots \\ 1 & X_{1,n} & \dots & X_{p-1,n} \end{pmatrix}. \quad (3.9)$$

A continuación se mostrarán algunos resultados de estadística multivariada, en los cuales se basan algunas de las demostraciones que se verán a lo largo del capítulo.

- Si Z es un vector o una matriz de variables aleatorias, denotado por $Z = [Z_{ij}]$ entonces $E(Z) = [E(Z_{ij})]$.

Es decir, el operador esperanza matemática se aplica elemento por elemento.

- $E(AZB + C) = AE(Z)B + C$

donde A , B y C son matrices de constantes y Z es una matriz aleatoria.

- Si V y W son vectores aleatorios de la misma dimensión, entonces

$$\text{Cov}(V, W) = [\text{Cov}(V_i, W_j)] = E\{[V - E(V)][W - E(W)]'\},$$

$$\text{Var}(W) = \text{Cov}(W, W),$$

$$\text{Cov}(AV, BW) = A\text{Cov}(V, W)B' \text{ con } A \text{ y } B \text{ dos matrices de constantes.}$$

- Si $E(W) = \theta$, $\text{Var}(W) = \Sigma$ y si A es una matriz simétrica de dimensión $n \times n$, entonces

$$E(W'AW) = \text{tr}(A\Sigma) + \theta'A\theta$$

(ver Seber, G.A.F. 1977, p. 13, Linear Regression Analysis, John Wiley and Sons).

- Como $\text{Cov}(W, V) = [\text{Cov}(V, W)]'$, entonces

$$\text{Cov}(BW, AV) = B\text{Cov}(W, V)A'$$

$$= B[\text{Cov}(V, W)]'A'$$

$$= A[\text{Cov}(V, W)B']'$$

3.2 Estimación por MCO

El método de MCO busca minimizar la SCR, que se expresa como

$$SCR = \sum e_i^2 = \mathbf{e}'\mathbf{e} \quad (3.10)$$

sujeta a la condición de que

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} \text{ para algún vector de estimadores } \hat{\boldsymbol{\beta}}.$$

Para solucionar este problema nótese que

$$\begin{aligned} \mathbf{e}'\mathbf{e} &= (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} \\ &= \mathbf{Y}'\mathbf{Y} - 2\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} \end{aligned}$$

Por lo tanto

$$\left. \frac{\partial \mathbf{e}'\mathbf{e}}{\partial \hat{\boldsymbol{\beta}}} \right|_{\hat{\boldsymbol{\beta}}=\mathbf{b}} = \mathbf{0} - 2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} \Big|_{\hat{\boldsymbol{\beta}}=\mathbf{b}} = \mathbf{0}$$

y, como se había visto antes,

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{Y} \quad (3.11)$$

lo cual implica que

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad (3.12)$$

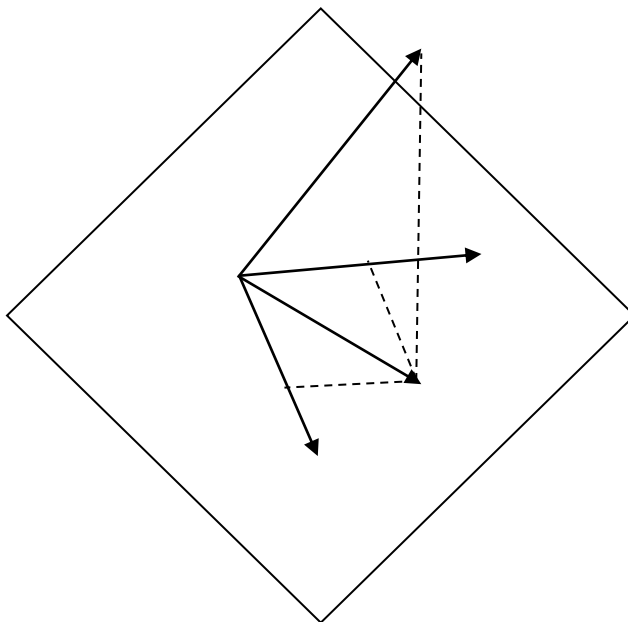
cuando $\mathbf{X}'\mathbf{X}$ es no-singular.

Nota: En el caso de regresión lineal simple, el modelo se puede escribir como

$$\mathbf{Y} = \beta_0\mathbf{1} + \beta_1\mathbf{X} + \boldsymbol{\varepsilon}$$

con $\mathbf{1}$ un vector de unos de dimensión $n \times 1$, $\mathbf{X} = (X_1, X_2, \dots, X_n)'$ el vector de valores de la variable X y $\boldsymbol{\varepsilon}$ el vector de errores también de dimensión $n \times 1$.

El plano donde está \mathbf{Y} queda definido por los vectores $\mathbf{1}$ y \mathbf{X} , o sea,



Los estimadores b_0 y b_1 hacen que el módulo del vector \mathbf{e} sea mínimo respecto a cualquiera otra proyección de \mathbf{Y} sobre el plano.

El vector \mathbf{e} será ortogonal a todos los vectores del plano y, en particular, a los vectores $\mathbf{1}$ y \mathbf{X} , de manera que se cumple

$$\mathbf{e}'\mathbf{1} = 0 \quad \text{y} \quad \mathbf{e}'\mathbf{X} = 0.$$

Además, se cumple que

$$|\mathbf{Y}|^2 = |\mathbf{e}|^2 + |\hat{\mathbf{Y}}|^2,$$

es decir,

$$\sum Y_i^2 = \sum e_i^2 + \sum \hat{Y}_i^2.$$

Ahora bien, los residuos de MCO están dados por

$$\begin{aligned}
\mathbf{e} &= \mathbf{Y} - \hat{\mathbf{Y}} \\
&= \mathbf{Y} - \mathbf{X}\mathbf{b} \\
&= \mathbf{Y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\
&= [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{Y} \\
&= (\mathbf{I} - \mathbf{M})\mathbf{Y}
\end{aligned}$$

con $\mathbf{M} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

Las matrices \mathbf{M} e $\mathbf{I}-\mathbf{M}$ tienen las siguientes propiedades :

i) \mathbf{M} e $(\mathbf{I}-\mathbf{M})$ son simétricas e idempotentes, ya que

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

y

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

ii) $(\mathbf{I}-\mathbf{M})\mathbf{X} = \mathbf{0}$, pues

$$\begin{aligned}
(\mathbf{I}-\mathbf{M})\mathbf{X} &= \mathbf{X} - \mathbf{MX} \\
&= \mathbf{X} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} \\
&= \mathbf{X} - \mathbf{X}
\end{aligned}$$

iii) $\text{tr}(\mathbf{I}-\mathbf{M}) = \text{tr}(\mathbf{I}) - \text{tr}(\mathbf{M})$

$$\begin{aligned}
&= n - \text{tr}[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] \\
&= n - \text{tr}[\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] \\
&= n - \text{tr}(\mathbf{I}_p) \\
&= n - p.
\end{aligned}$$

Por otro lado, algunas propiedades de los residuos de MCO son las siguientes

$$E(\mathbf{e}) = (\mathbf{I} - \mathbf{M})E(\mathbf{Y}) = (\mathbf{I} - \mathbf{M})\mathbf{X}\boldsymbol{\beta} = \mathbf{0} \quad (3.13)$$

y

$$\text{Var}(\mathbf{e}) = (\mathbf{I} - \mathbf{M}) \sigma^2 \mathbf{I} (\mathbf{I} - \mathbf{M})' = \sigma^2 (\mathbf{I} - \mathbf{M}) \quad (3.14)$$

así que los residuos tienen media cero, pero están correlacionados.

Además,

$$\begin{aligned} \mathbf{e}'\mathbf{e} &= (\mathbf{Y} - \mathbf{X}\mathbf{b})'(\mathbf{Y} - \mathbf{X}\mathbf{b}) \\ &= \mathbf{Y}'\mathbf{Y} - 2\mathbf{b}'\mathbf{X}'\mathbf{Y} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} \\ &= \mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y} + \mathbf{b}'(\mathbf{X}'\mathbf{X}\mathbf{b} - \mathbf{X}'\mathbf{Y}) \\ &= \mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y} \end{aligned}$$

por lo tanto

$$\mathbf{e}'\mathbf{e} = \sum Y_i^2 - \mathbf{b}'\mathbf{X}'\mathbf{Y}. \quad (3.15)$$

Asimismo, algunas de las propiedades de \mathbf{b} son las siguientes

$$\begin{aligned} E(\mathbf{b}) &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{Y}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\ &= \boldsymbol{\beta} \end{aligned} \quad (3.16)$$

y

$$\begin{aligned}
 \text{Var}(\mathbf{b}) &= \text{Var}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}] \\
 &= \text{Cov}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}] \\
 &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{Cov}(\mathbf{Y}, \mathbf{Y})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
 &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} [\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] \\
 &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}.
 \end{aligned} \tag{3.17}$$

Además, al igual que en regresión lineal simple, el estimador de MCO, \mathbf{b} , resulta ser MELI, de acuerdo con el Teorema de Gauss-Markov.

Ahora es necesario realizar una estimación de σ^2 , la idea es usar un “promedio” basado en la SCR. Dicha SCR se define como

$$\text{SCR} = (\mathbf{Y} - \mathbf{X}\mathbf{b})'(\mathbf{Y} - \mathbf{X}\mathbf{b}) \tag{3.18}$$

luego, como

$$\mathbf{Y} - \mathbf{X}\mathbf{b} = (\mathbf{I} - \mathbf{M})\mathbf{Y}$$

entonces

$$\text{SCR} = \mathbf{Y}'(\mathbf{I} - \mathbf{M})\mathbf{Y}$$

y, por lo tanto

$$\begin{aligned}
 E(\text{SCR}) &= E[\mathbf{Y}'(\mathbf{I} - \mathbf{M})\mathbf{Y}] \\
 &= \text{tr}[(\mathbf{I} - \mathbf{M})\sigma^2\mathbf{I}] + \boldsymbol{\beta}'\mathbf{X}'(\mathbf{I} - \mathbf{M})\mathbf{X}\boldsymbol{\beta} \\
 &= \sigma^2\text{tr}(\mathbf{I} - \mathbf{M}) \\
 &= \sigma^2(n-p).
 \end{aligned}$$

De esta manera, el estimador que se debe usar para que sea insesgado, es

$$S^2 = \frac{SCR}{n - p} \quad (3.19)$$

de tal forma que

$$E(S^2) = \sigma^2.$$

3.3 Pronóstico

Para un vector \mathbf{X}_0 , que contiene los valores de las variables explicativas de interés, se tiene que el pronóstico de la Y correspondiente se obtiene como estimación de $E(Y_0) = \mathbf{X}_0' \boldsymbol{\beta}$, o sea

$$\bar{Y}_0 = \mathbf{X}_0' \mathbf{b} = \mathbf{b}' \mathbf{X}_0. \quad (3.20)$$

La varianza de este pronóstico \bar{Y}_0 se expresa como

$$\text{Var}(\bar{Y}_0) = \text{Var}(\mathbf{X}_0' \mathbf{b}) = \mathbf{X}_0' \text{Var}(\mathbf{b}) \mathbf{X}_0$$

y, en consecuencia

$$\text{Var}(\bar{Y}_0) = \sigma^2 \mathbf{X}_0' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_0. \quad (3.21)$$

Así que el estimador de la varianza del pronóstico se define como

$$\bar{\text{Var}}(\bar{Y}_0) = S^2 \mathbf{X}_0' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_0 \quad (3.22)$$

cuando el pronóstico se refiere a la estimación del valor medio de Y , para \mathbf{X}_0 dado.

Por otro lado, la varianza del error del pronóstico de un valor individual de Y , o sea de $e_0 = Y_0 - \bar{Y}_0$, se obtiene como

$$\text{Var}(e_0) = \text{Var}(Y_0) + \text{Var}(\bar{Y}_0) - 2\text{Cov}(Y_0, \bar{Y}_0)$$

donde $\text{Cov}(Y_0, \bar{Y}_0) = 0$ porque \bar{Y}_0 es función lineal de Y_1, \dots, Y_n , que no están correlacionados entre sí, ni con Y_0 .

Por lo tanto,

$$\text{Var}(e_0) = \sigma^2 [1 + \mathbf{X}_0' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_0] \quad (3.23)$$

y su estimador es

$$\hat{\text{Var}}(e_0) = S^2 [1 + \mathbf{X}_0' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_0]. \quad (3.24)$$

3.4 Estimación por MV e inferencia estadística

Si se hace el supuesto de normalidad en los errores, éstos tendrían la siguiente distribución

$$\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}) \quad (3.25)$$

con N_n la distribución Normal n -variada. Entonces se sigue que

$$\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}) \quad (3.26)$$

y su función de densidad se expresa como

$$f_Y(\mathbf{Y} | \boldsymbol{\beta}, \sigma^2) = (2\pi)^{-n/2} \sigma^{-n} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \right\}$$

la cual resulta ser una función de verosimilitud cuando $\boldsymbol{\beta}$ y σ^2 son desconocidos y se desea estimarlos.

Por MV se tiene que

$$\max_{\boldsymbol{\beta}, \sigma^2} f_Y(\boldsymbol{\beta}, \sigma^2 | \mathbf{Y}) = f_Y(\tilde{\boldsymbol{\beta}}, \tilde{\sigma}^2 | \mathbf{Y})$$

y su solución es

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{b} \quad (3.27)$$

junto con

$$\tilde{\sigma}^2 = \frac{\text{SCR}}{n} . \quad (3.28)$$

En la ecuación anterior se puede apreciar que $S^2 \neq \tilde{\sigma}^2$, por lo que $\tilde{\sigma}^2$ tiene sesgo. Por ello es conveniente realizar un ajuste, a fin de corregir el sesgo. De esta manera se concluye que el estimador insesgado a utilizar debe ser

$$S^2 = \frac{n}{n-p} \tilde{\sigma}^2 . \quad (3.29)$$

Ahora bien, como \mathbf{b} es una combinación lineal de las Y 's se sigue que

$$\mathbf{b} \sim N_p(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}) \quad (3.30)$$

y también que

$$\frac{(n-p)S^2}{\sigma^2} \sim \chi^2_{(n-2)} \quad (3.31)$$

con S^2 y \mathbf{b} independientes.

A partir de las distribuciones para \mathbf{b} y S^2 es posible realizar inferencia estadística del tipo de intervalos de confianza para los elementos de $\boldsymbol{\beta}$, o del Análisis de Varianza que se muestra en la siguiente sección.

Los intervalos de $100(1-\alpha)\%$ de confianza para los parámetros individuales se expresan como

$$\beta_i \in \{b_i \pm t_{(n-p), \alpha/2} e\hat{e}(b_i)\} \quad (3.32)$$

expresión que es válida para $i = 0, \dots, p-1$ si hay constante en el modelo o para $i = 1, \dots, p$ si no la hay, donde

$$e\hat{e}(b_i) = \sqrt{(i+1) - \text{ésimo elemento en la diagonal de } S^2(\mathbf{X}'\mathbf{X})^{-1}} \quad \text{si hay constante}$$

$$= \sqrt{i - \text{ésimo elemento en la diagonal de } S^2(\mathbf{X}'\mathbf{X})^{-1}} \quad \text{si no la hay.}$$

Si se construyen intervalos individuales del $100(1-\alpha)\%$ de confianza para cada una de las β 's, la inferencia simultánea es comparable “aproximadamente” con una región del $100(1-\alpha)^p\%$ de confianza (y no con una del $100(1-\alpha)\%$).

3.5 Análisis de varianza

Cuando existe la constante en el modelo, se tienen $p-1$ variables independientes y la tabla de ANOVA resulta ser.

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F
Regresión sobre X_1, \dots, X_{p-1}	$SCE = \mathbf{b}' \mathbf{X}' \mathbf{Y} - n \bar{Y}^2$	$p-1$	$CME = \mathbf{b}' \mathbf{X}' \mathbf{Y} / (p-1)$	CME/CMR
Residual	$SCR = \mathbf{Y}' \mathbf{Y} - \mathbf{b}' \mathbf{X}' \mathbf{Y}$	$n-p$	$CMR = SCR / (n-p)$	
TOTAL	$SCT = \mathbf{Y}' \mathbf{Y} - n \bar{Y}^2$	$n-1$		

Cuadro 5. Tabla de ANOVA. Con término constante y $p-1$ variables independientes.

Si no existe término constante en el modelo, existen p variables independientes y se tiene la siguiente tabla de ANOVA.

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F
Regresión sobre X_1, \dots, X_p	$SCE = \mathbf{b}' \mathbf{X}' \mathbf{Y}$	p	$CME = \mathbf{b}' \mathbf{X}' \mathbf{Y} / p$	CME/CMR
Residual	$SCR = \mathbf{Y}' \mathbf{Y} - \mathbf{b}' \mathbf{X}' \mathbf{Y}$	$n-p$	$CMR = SCR / (n-p)$	
Total	$\mathbf{Y}' \mathbf{Y}$	n		

Cuadro 5. Tabla de ANOVA. Sin término constante y p variables independientes.

Debido al supuesto de normalidad, en el primer caso se tiene que el estadístico $F \sim F_{(p-1, n-p)}$ y sirve para realizar la siguiente prueba de hipótesis

$$H_0: \beta_1 = \beta_2 = \dots = \beta_{p-1} \quad \text{vs.} \quad H_A: \beta_i \neq 0 \quad \text{para alguna } i = 1, \dots, p-1$$

En el segundo caso, el estadístico F tiene la distribución $F_{(p, n-p)}$ y permite realizar la prueba de hipótesis

$$H_0: \boldsymbol{\beta} = \mathbf{0} \quad \text{vs.} \quad H_A: \boldsymbol{\beta} \neq \mathbf{0}.$$

3.6 Prueba de una hipótesis lineal general

Una hipótesis lineal general es de la forma

$$H_0: \mathbf{A}\boldsymbol{\beta} = \mathbf{C} \tag{3.33}$$

donde \mathbf{A} es una matriz de constantes de orden $q \times p$, con $1 \leq q \leq p$ y rango q , y \mathbf{C} es un vector de constantes de dimensión $q \times 1$.

Este tipo de hipótesis surgen, por ejemplo, cuando se tiene el siguiente modelo lineal.

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \varepsilon$$

y se requiere probar la hipótesis:

i) $H_0: \beta_1 = C_1, \beta_2 = C_2$ mientras que β_3, \dots, β_p son arbitrarios. En este caso la matriz \mathbf{A} tendría la forma

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \end{bmatrix}$$

y el vector $\mathbf{C} = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix}$, de tal manera que

$$\mathbf{A}\boldsymbol{\beta} = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \dots \\ \beta_p \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} = \mathbf{C}$$

También se podría pensar en probar la hipótesis

$$\text{ii) } H_0: \beta_1 = \beta_2 = \beta_3 = \dots = \beta_q .$$

que es equivalente a

$$H_0: \beta_1 - \beta_2 = 0, \beta_2 - \beta_3 = 0, \dots, \beta_{q-1} - \beta_q = 0$$

En este caso la matriz A sería

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & -1 & 0 & \dots & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots & 1 & -1 & \dots & 0 \end{bmatrix} \text{ de dimensión } (q-1) \times p$$

$$\text{y el vector } \mathbf{C} = \begin{bmatrix} 0 \\ \dots \\ 0 \end{bmatrix} .$$

Otro ejemplo podría ser

iii) $H_0: 3\beta_1 - 2\beta_2 = 1 + 4\beta_3 - \beta_1$

lo cual podría escribirse de forma alternativa como

$$H_0: 4\beta_1 - 2\beta_2 - 4\beta_3 = 1$$

que, para expresarse en la forma de hipótesis lineal general, daría origen a la matriz $A = \begin{bmatrix} 0 & 4 & -2 & -4 & 0 & \dots & 0 \end{bmatrix}$ de dimensión $1 \times p$ y al vector $C = 1$ (escalar).

En general, la suma de cuadrados residual mínima es

$$SCR = (\mathbf{Y} - \mathbf{X}\mathbf{b})' (\mathbf{Y} - \mathbf{X}\mathbf{b}) = (n-p) S^2 \quad (3.34)$$

y, en la hipótesis nula (la que se desea probar) se obtiene la suma de cuadrados residual mínima como

$$SCR_H = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_H)' (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_H) \quad (3.35)$$

en donde

$$\hat{\boldsymbol{\beta}}_H = \mathbf{b} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{A}' [\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{A}']^{-1} (\mathbf{C} - \mathbf{A}\mathbf{b}) \quad (3.36)$$

es el estimador de $\boldsymbol{\beta}$ que satisface la restricción $\mathbf{A}\boldsymbol{\beta} = \mathbf{C}$ (lo cual puede verificarse al premultiplicar por \mathbf{A}), entonces la estadística de prueba apropiada viene dada por

$$F = \frac{(SCR_H - SCR)/q}{SCR/(n-p)}$$

$$= \frac{(\mathbf{Ab} - \mathbf{C})' [\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{A}']^{-1} (\mathbf{Ab} - \mathbf{C})}{qS^2}$$

$$\sim F_{(q, n-p)} \quad (3.37)$$

que es exacta cuando se cumplen los supuestos del modelo, en particular si los errores aleatorios son independientes y se encuentran normalmente distribuidos.

Como ejemplo de la aplicación de esta prueba considérese la hipótesis

$$H_0: \beta_1 = \beta_2 = \dots = \beta_{p-1} = 0$$

en el modelo

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_{p-1} X_{p-1} + \varepsilon$$

es decir, se tendría $H: \mathbf{A}\boldsymbol{\beta} = \mathbf{0}$ con $\mathbf{A} = [\mathbf{0}, \mathbf{I}_{p-1}]$ de dimensión $(p-1) \times p$, o sea que $q = p-1$, entonces

$$SCR = \mathbf{Y}'\mathbf{Y} - \mathbf{bX}'\mathbf{Y}$$

y

$$SCR_H = \min_{\beta_0} \sum_{i=1}^n (Y_i - \beta_0)^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2 = \mathbf{Y}'\mathbf{Y} - n\bar{Y}^2$$

de tal forma que

$$F = \frac{(SCR_H - SCR)/(p-1)}{SCR/(n-p)} = \frac{(\mathbf{b}'\mathbf{X}'\mathbf{Y} - n\bar{Y}^2)/(p-1)}{(\mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y})/(n-p)} \sim F_{(p-1, n-p)}$$

que es la prueba usual del análisis de varianza.

Si la hipótesis lineal general es rechazada, se debe buscar entonces a qué se debió el rechazo. Una manera de hacer esto es mediante pruebas individuales del tipo

$$H_i: \mathbf{a}_i'\boldsymbol{\beta} = C_i \quad \text{para } i = 1, 2, \dots, q \quad (3.38)$$

usando pruebas t. Estas pruebas se basan en que

$$\mathbf{a}_i'\mathbf{b} \sim N(\mathbf{a}_i'\boldsymbol{\beta}, \sigma^2 \mathbf{a}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}_i) \quad (3.39)$$

así que el cociente

$$\frac{\mathbf{a}_i'\mathbf{b} - \mathbf{a}_i'\boldsymbol{\beta}}{S\sqrt{\mathbf{a}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}_i}} \sim t_{(n-p)} \quad (3.40)$$

de donde se sigue que

$$\mathbf{a}_i'\mathbf{b} \pm t_{(n-p), \alpha/2} S\sqrt{\mathbf{a}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}_i} \quad (3.41)$$

es un intervalo del 100(1- α)% de confianza para $\mathbf{a}_i'\boldsymbol{\beta}$.

Se debe pues observar si el valor C_i pertenece o no al intervalo que se acaba de construir, para rechazar o no-rechazar H_i . A este método se le denomina el de *Menor Diferencia Significativa* y hace uso de la siguiente partición

$$A\beta = \begin{bmatrix} \mathbf{a}_1'\beta \\ \mathbf{a}_2'\beta \\ \dots \\ \mathbf{a}_q'\beta \end{bmatrix} = \begin{bmatrix} C_1 \\ C_2 \\ \dots \\ C_q \end{bmatrix} = \mathbf{C} .$$

Además, puede suceder que la hipótesis general H_0 se rechace, pero que individualmente, ninguna H_i sea rechazada.

3.7 Mínimos Cuadrados Generalizados (MCG)

El método de MCG surge con la idea de extender la aplicabilidad de mínimos cuadrados al caso de matrices de varianza-covarianza de los errores que no sean necesariamente diagonales o con varianza constante.

Los supuestos en los que se basa el método de MCG son

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon \quad (3.42)$$

con

$$E(\epsilon) = \mathbf{0} \text{ y } \text{Var}(\epsilon) = \sigma^2\Omega \quad (3.43)$$

donde Ω es una matriz de varianza-covarianza ‘conocida’ positiva definida. Además se supone que X es de rango completo, es decir, que las columnas de la matriz X son linealmente independientes y que sus valores están fijos. Nótese que en el caso de tener $\Omega = I$ se obtendría el caso de MCO.

Supóngase que se usa MCO para estimar β , de tal forma que se obtiene

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \text{ y } \text{Cov}(\mathbf{b}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

Sin embargo, debido a los nuevos supuestos, se tiene que

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

así que

$$\mathbf{b} - \boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}$$

y por lo tanto

$$E(\mathbf{b}) = \boldsymbol{\beta}$$

o sea que \mathbf{b} sigue siendo insesgado, aun en el modelo general.

Para la varianza, se tiene que

$$\begin{aligned} \text{Var}(\mathbf{b}) &= E[(\mathbf{b}-\boldsymbol{\beta})(\mathbf{b}-\boldsymbol{\beta})'] \\ &= E[(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'(\mathbf{X}'\mathbf{X})^{-1}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

en esta última expresión se puede observar que

$$\text{Var}(\mathbf{b}) \neq \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

a menos que $\boldsymbol{\Omega} = \mathbf{I}$.

El método de MCG consiste en elegir una matriz \mathbf{V} no singular tal que

$$\mathbf{V}'\mathbf{V} = \boldsymbol{\Omega}^{-1} \quad \text{y} \quad \mathbf{V}\boldsymbol{\Omega}\mathbf{V}' = \mathbf{I} \quad (3.44)$$

la cual se sabe que existe porque $\boldsymbol{\Omega}$ es positiva definida.

Nota: Sea \mathbf{C} una matriz ortogonal y $\lambda_1, \dots, \lambda_n$ los valores propios de $\boldsymbol{\Omega}$. Entonces, se tiene que

$$\mathbf{C}'\boldsymbol{\Omega}\mathbf{C} = \boldsymbol{\Lambda} = \text{Diag}(\lambda_1, \dots, \lambda_n) \quad \text{y} \quad \mathbf{D} = \text{Diag}(1/\sqrt{\lambda_1}, \dots, 1/\sqrt{\lambda_n})$$

donde $V = D'C'$ es la matriz buscada, ya que es no-singular (por ser producto de matrices no-singulares y cumple con

$$V\Omega V' = D'C'\Omega CD' = D'\Lambda D = I$$

Por lo tanto,

$$V'(V\Omega V')V = V'V$$

de donde $V'V\Omega = I$ y $V'V = \Omega^{-1}$.

Así pues, si se realiza una transformación de los datos originales, de tal forma que se considere $VY = Z$, $VX = W$ y $V\epsilon = \delta$, se llega a

$$VY = VX\beta + V\epsilon \quad (3.45)$$

lo que implica que

$$Z = W\beta + \delta \quad (3.46)$$

el cual es un modelo lineal con variables transformadas Z y W . Nótese que

$$E(\delta) = VE(\epsilon) = 0$$

y

$$\text{Cov}(\delta) = E(\delta\delta') = E(V\epsilon\epsilon'V') = VE(\epsilon\epsilon')V' = \sigma^2 V\Omega V' = \sigma^2 I.$$

Por lo cual los supuestos de MCO se cumplen para Z , W y δ . En consecuencia, para estimar β se debe usar

$$\mathbf{b}^* = (W'W)^{-1}X'Z = (X'V'VX)^{-1}X'V'VY = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y \quad (3.47)$$

que resulta ser el estimador mínimo cuadrático generalizado. Con

$$\text{Var}(\mathbf{b}^*) = \sigma^2(W'W)^{-1} = \sigma^2(X'\Omega^{-1}X)^{-1}. \quad (3.48)$$

Para estimar σ^2 se usan los residuales del modelo donde se cumplen los supuestos de MCO, o sea

$$\mathbf{d} = \mathbf{Z} - \mathbf{W} \mathbf{b}^* = \mathbf{V}(\mathbf{Y} - \mathbf{X} \mathbf{b}^*) = \mathbf{V} \mathbf{e}^* \quad (3.49)$$

con

$$\begin{aligned} \mathbf{e}^* &= \mathbf{Y} - \mathbf{X} \mathbf{b}^* \\ &= (\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) - \mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{Y} \\ &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}\boldsymbol{\beta} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon} \\ &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} - \mathbf{X}\boldsymbol{\beta} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon} \\ &= [\mathbf{I} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}] \boldsymbol{\varepsilon} \\ &= (\mathbf{I} - \mathbf{M}^*) \boldsymbol{\varepsilon} \end{aligned}$$

donde \mathbf{M}^* es una matriz simétrica e idempotente tal que $\text{tr}(\mathbf{M}^*) = p$ y

$$\begin{aligned} (\mathbf{I} - \mathbf{M}^*)' \boldsymbol{\Omega}^{-1}(\mathbf{I} - \mathbf{M}^*) &= [\mathbf{I} - \boldsymbol{\Omega}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}']\boldsymbol{\Omega}^{-1}(\mathbf{I} - \mathbf{M}^*) \\ &= [\boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}] (\mathbf{I} - \mathbf{M}^*) \\ &= \boldsymbol{\Omega}^{-1}(\mathbf{I} - \mathbf{M}^*)(\mathbf{I} - \mathbf{M}^*) \\ &= \boldsymbol{\Omega}^{-1}(\mathbf{I} - \mathbf{M}^*). \end{aligned}$$

Además la SCR obtenida en el modelo de MCG se define como

$$\mathbf{d}'\mathbf{d} = \mathbf{e}^{*'} \mathbf{V}'\mathbf{V} \mathbf{e}^* = \mathbf{e}^{*'} \boldsymbol{\Omega}^{-1} \mathbf{e}^* = \boldsymbol{\varepsilon}' (\mathbf{I} - \mathbf{M}^*)' \boldsymbol{\Omega}^{-1} (\mathbf{I} - \mathbf{M}^*) \boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}' \boldsymbol{\Omega}^{-1} (\mathbf{I} - \mathbf{M}^*) \boldsymbol{\varepsilon}$$

por lo cual su esperanza resulta ser

$$\begin{aligned} E(\mathbf{e}^{*'} \boldsymbol{\Omega}^{-1} \mathbf{e}^*) &= E[\boldsymbol{\varepsilon}' \boldsymbol{\Omega}^{-1} (\mathbf{I} - \mathbf{M}^*) \boldsymbol{\varepsilon}] \\ &= E\{\text{tr}[\boldsymbol{\varepsilon}' \boldsymbol{\Omega}^{-1} (\mathbf{I} - \mathbf{M}^*) \boldsymbol{\varepsilon}]\} \end{aligned}$$

$$\begin{aligned}
 &= E\{\text{tr}[(I - M^*) \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}' \Omega^{-1}]\} \\
 &= \text{tr}\{E[(I - M^*) \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}' \Omega^{-1}]\} \\
 &= \text{tr}[(I - M^*) E(\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}') \Omega^{-1}] \\
 &= \text{tr}[(I - M^*) \sigma^2 \Omega \Omega^{-1}] \\
 &= \sigma^2 \text{tr}(I - M^*)
 \end{aligned}$$

o sea que

$$E(\mathbf{d}'\mathbf{d}) = E(\mathbf{e}^{*'} \Omega^{-1} \mathbf{e}^*) = \sigma^2(n-p). \quad (3.50)$$

Por lo tanto, un estimador insesgado de σ^2 en MCG viene a ser

$$S^{*2} = \mathbf{e}^{*'} \Omega^{-1} \mathbf{e}^* / (n-p) \quad (3.51)$$

y

$$\hat{\text{Var}}(\mathbf{b}^*) = S^{*2} (\mathbf{X}' \Omega^{-1} \mathbf{X})^{-1}. \quad (3.52)$$

Un problema fundamental de MCG es que Ω se supone conocida, si esto no es así no es factible estimar los $n(n+1)/2$ elementos distintos de Ω a partir de los n datos muestrales. Por ello lo que se hace en la práctica es postular una cierta estructura para Ω . Esto es útil, por ejemplo, para corregir el problema de varianza no constante y de correlación en los errores, como se verá más adelante.

3.8 Autocorrelación

El supuesto de MCO que $\text{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$ implica que no hay correlación entre los errores. Este supuesto se viola en muchas ocasiones debido a que las observaciones se realizan de manera ordenada respecto a alguna dimensión, por ejemplo el tiempo. En consecuencia se dice que existe autocorrelación en los errores.

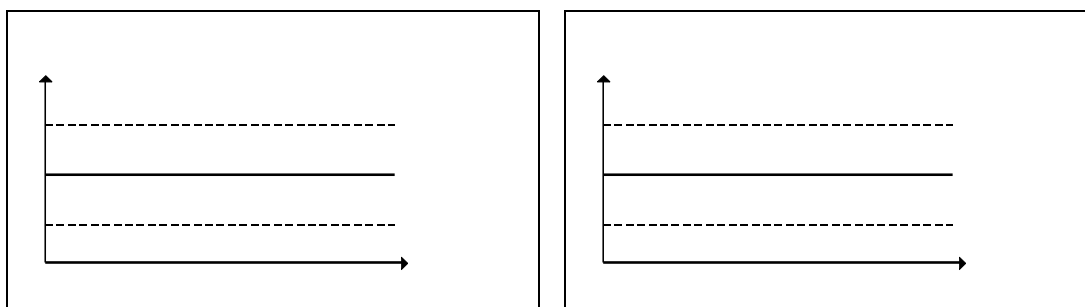
Consecuencias

Las consecuencias que tiene la violación del supuesto de no-correlación, pueden ser:

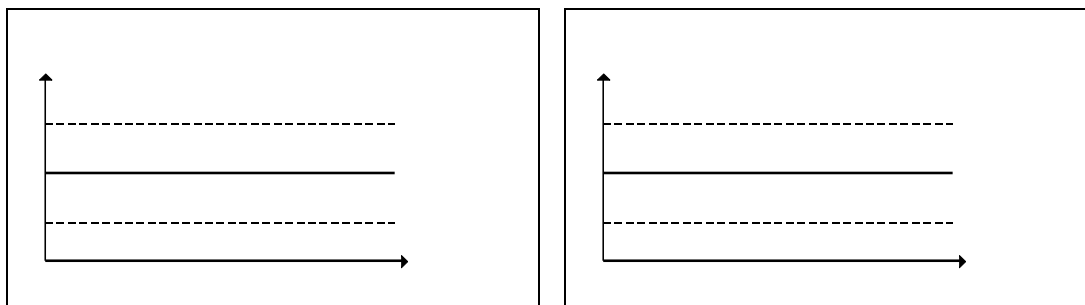
- 1.- Los estimadores de MCO seguirán siendo insesgados, pero su desviación estándar será menor que la real, lo cual hará que las pruebas t tiendan a rechazar la hipótesis nula con mucha frecuencia y se pensará que el modelo es bueno, sin que lo sea.
- 2.- σ^2 será subestimada al usar S^2 y el coeficiente de determinación será mayor que lo debido, por lo cual habrá una tendencia falsa hacia el optimismo.
- 3.- En general, las inferencias basadas en los estadísticos t y F serán inválidas.

Detección

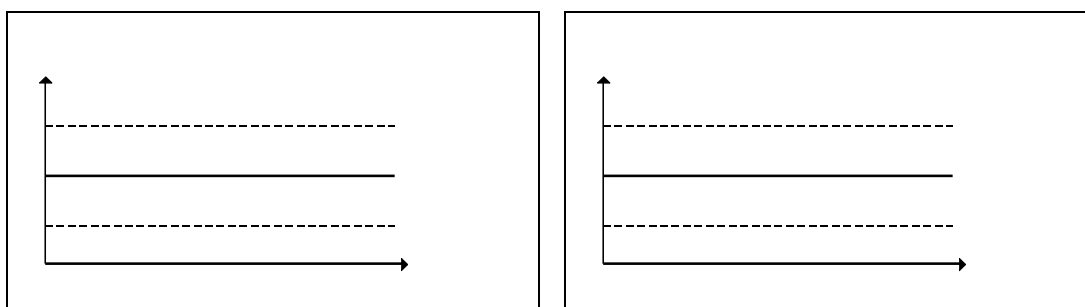
La detección de autocorrelación en los errores se puede efectuar visualmente, en la gráfica de residuos contra el tiempo (el orden) o en un diagrama de dispersión de los residuos contra ellos mismos, pero retrasados uno o más periodos. Esto se debe a que los residuos se comportan aproximadamente como los errores. De esta forma, en las citadas gráficas no debería presentarse ningún patrón sistemático, ya que ello puede ser indicio de que la autocorrelación está presente



No hay indicios de autocorrelación



Indicios de autocorrelación positiva



Indicios de autocorrelación negativa

Debido a que la presencia de autocorrelación invalida fuertemente los resultados que surjan del modelo (de hecho esta violación es de las más graves que pudieran ocurrir), es importante contar con un criterio numérico que permita detectar la violación de una manera más clara y objetiva que simplemente con una inspección visual de una gráfica. Por tal motivo, a continuación se presenta un estadístico de prueba desarrollado para tal fin.

Prueba de Durbin-Watson

Esta prueba es la más utilizada en la práctica hoy en día y se basa en el cálculo de

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} \text{ con } \bar{e} = 0. \quad (3.53)$$

- i) Para residuos correlacionados positivamente, las diferencias $(e_t - e_{t-1})^2$ tenderán a ser pequeñas en relación a su varianza, por lo cual d tomará un valor pequeño.
- ii) Para residuos que estén correlacionados negativamente, $(e_t - e_{t-1})^2$ será comúnmente mayor que la varianza de e_t , de forma tal que d será grande.
- iii) De no haber correlación entre e_t y e_{t-1} , d tomará un valor intermedio (alrededor de 2).
Esto se sigue de que $\bar{e} = 0$, ya que entonces

$$\begin{aligned} d &= \frac{\left(\sum_{t=2}^n e_t^2 + \sum_{t=2}^n e_{t-1}^2 \right) / (n-1) - 2 \sum_{t=2}^n e_t e_{t-1} / (n-1)}{\sum_{t=1}^n e_t^2 / (n-1)} \\ &\cong \frac{2 \bar{\text{Var}}(\varepsilon_t) - 2 \bar{\text{Cov}}(\varepsilon_t, \varepsilon_{t-1})}{\bar{\text{Var}}(\varepsilon_t)} \\ &= 2 (1 - \bar{\rho}) \end{aligned}$$

con $\rho = \text{Corr}(\varepsilon_t, \varepsilon_{t-1})$ el coeficiente de autocorrelación de primer orden.

Por consiguiente,

$$\bar{\rho} \rightarrow 0 \text{ implica que } d \rightarrow 2$$

$$\bar{\rho} \rightarrow -1 \text{ implica que } d \rightarrow 4$$

$\bar{\rho} \rightarrow 1$ implica que $d \rightarrow 0$.

La distribución de d la obtuvieron Durbin y Watson en función de la matriz de observaciones X , pero resultó ser tan complicada que tan solo se usan los límites inferior y superior, para declarar niveles de significación de la d calculada.

Sea entonces d_C la estadística calculada, mientras que d_U y d_L son los límites superior e inferior correspondientes al número de regresores involucrado y al tamaño de muestra, n .

Para probar si existe autocorrelación positiva al nivel de significancia del $100\alpha\%$, se procede como sigue:

si $d_C < d_{L, \alpha}$, se rechaza $H_0: \rho = 0$ en favor de $H_A: \rho > 0$

si $d_C > d_{U, \alpha}$, no se rechaza H_0

si $d_{L, \alpha} < d_C < d_{U, \alpha}$, no hay conclusión.

En el caso en el que no se llegue a una conclusión, posiblemente se requieran más observaciones para llegar a ella.

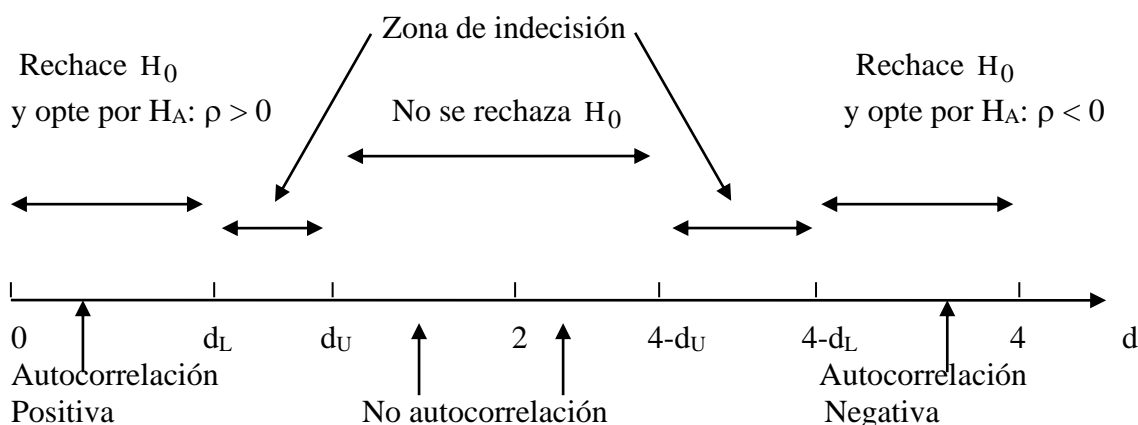
Para probar si hay autocorrelación negativa, se utilizan los valores $4 - d_U$ y $4 - d_L$ en lugar de d_L y d_U , de tal forma que:

si $d_C > 4 - d_{L, \alpha}$, se rechaza $H_0: \rho = 0$ en favor de $H_A: \rho < 0$

si $d_C < 4 - d_{U, \alpha}$, no se rechaza H_0

si $4 - d_{U, \alpha} < d_C < 4 - d_{L, \alpha}$, no hay conclusión.

Conviene ver la siguiente gráfica, donde se resumen las posibles conclusiones del procedimiento de prueba



Los límites d_U y d_L fueron proporcionados por Durbin y Watson mediante tablas de valores correspondientes a tamaños de muestra $15 \leq n \leq 100$ y a lo más cinco variables explicativas. Posteriormente, Savin y White ampliaron dichas tablas para $6 \leq n \leq 200$ y hasta 20 variables explicativas (véase la tabla B-5 de Johnston, 1984).

Debe recordarse que d no se debe utilizar si no hay término constante en el modelo. Tampoco debe usarse si existen observaciones aberrantes. Si se da el caso de observaciones faltantes, el numerador de d se debe de ajustar para no calcular las(s) diferencia(s) correspondientes. Además, si la variable dependiente aparece en forma retrasada como regresor, deberá usarse una modificación de la prueba, que fue sugerida por Durbin en 1970 y que dio origen a la prueba h de Durbin (véase Pindyck y Rubinfeld, 1981 p. 194).

Corrección

Si se detecta autocorrelación en el modelo, debe tenerse en cuenta que su causa puede ser una forma funcional incorrecta o bien el que se haya omitido alguna variable explicativa. Por ejemplo, si el modelo debiera ser

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t$$

pero se supuso incorrectamente que

$$Y_t = \beta_0 + \beta_1 X_{1t} + \varepsilon_t$$

resulta que, implícitamente, se tiene $\varepsilon_t = \beta_2 X_{2t} + u_t$ lo cual se podría ver reflejado en que los residuos muestren un patrón sistemático, posiblemente del tipo de autocorrelación.

Si existe convencimiento de que se ha detectado autocorrelación de primer orden, ésta puede corregirse con alguno de los diversos procedimientos que existen para corregirla. Algunos de ellos son, el de Hildreth-Lu, el de Cochrane-Orcutt, el de MCG y el de corrección AR(1).

1) Procedimiento de Hildreth-Lu

Si el error sigue un esquema autorregresivo de primer orden, se tiene que

$$\varepsilon_t = \rho \varepsilon_{t-1} + \delta_t \quad (3.54)$$

con δ_t un error que satisface los supuestos clásicos de MCO, Hildreth-Lu proponen que ρ sea estimado de manera simultánea con β . Para ello se tiene (por simplicidad de notación se presenta el caso de regresión simple, pero esencialmente lo mismo se hace en regresión múltiple)

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t \quad \text{y} \quad \rho Y_{t-1} = \rho \beta_0 + \rho \beta_1 X_{t-1} + \rho \varepsilon_{t-1}$$

así que

$$Y_t - \rho Y_{t-1} = (1-\rho)\beta_0 + \beta_1(X_t - \rho X_{t-1}) + \delta_t$$

o sea,

$$Y_t' = \beta_0' + \beta_1 X_t' + \delta_t \quad (3.55)$$

con $Y_t' = Y_t - \rho Y_{t-1}$, $\beta_0' = (1-\rho)\beta_0$ y $X_t' = X_t - \rho X_{t-1}$.

De esta manera se obtienen los estimadores de MCO

$$\bar{\beta}_0' = \frac{(\bar{Y}' - \bar{\beta}_1' \bar{X}')}{(1-\rho)} \quad \text{y} \quad \bar{\beta}_1' = \frac{\sum_{t=2}^n (Y_t' - \bar{Y}') (X_t' - \bar{X}')}{\sum_{t=2}^n (X_t' - \bar{X}')^2} \quad (3.56)$$

junto con

$$S^2 = \frac{1}{n-3} \sum_{t=2}^n \left[(Y_t - \rho Y_{t-1}) - \bar{\beta}_0' (1 - \rho) - \bar{\beta}_1' (X_t - \rho X_{t-1}) \right]^2. \quad (3.57)$$

Nótese que $\bar{\beta}_0'$, $\bar{\beta}_1'$ y S^2 resultan ser funciones de ρ . Así pues, para diferentes valores de ρ se tienen distintas S^2 , por lo cual se sugiere minimizar esta variable en función de ρ . Esto se puede lograr moviendo el valor de ρ dentro del intervalo $(-1, 1)$. De hecho se acostumbra elegir un valor mínimo de ρ , a partir del cual ir creciendo escalonadamente hasta alcanzar un valor máximo de ρ . Para cada uno de estos valores se calcula S^2 y se elige $\bar{\rho}$ como aquél que minimiza la suma de cuadrados.

2) Procedimiento de Cochrane-Orcutt

Este consiste en los siguientes pasos.

1.- Calcular tentativamente b_0 y b_1 en el modelo

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t \quad \text{con} \quad \varepsilon_t = \rho \varepsilon_{t-1} + \delta_t$$

2.- Obtener los residuos e_1, \dots, e_n y, a partir de ellos, estimar ρ mediante

$$\tilde{\rho} = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2} \quad (3.58)$$

3.- Construir las variables

$$Y_t' = Y_t - \tilde{\rho} Y_{t-1} \quad \text{y} \quad X_t' = X_t - \tilde{\rho} X_{t-1}$$

para obtener los nuevos estimadores de MCO, $\hat{\beta}_0$ y $\hat{\beta}_1$ en el modelo correspondiente.

4.- Obtener los nuevos residuos $\tilde{e}_1, \dots, \tilde{e}_n$ para estimar ρ mediante

$$\tilde{\tilde{\rho}} = \frac{\sum_{t=2}^n \tilde{e}_t \tilde{e}_{t-1}}{\sum_{t=1}^n \tilde{e}_t^2} \quad (3.59)$$

5.- Regresar al paso 3 y continuar así hasta lograr convergencia. De acuerdo con algún criterio de convergencia, por ejemplo, hasta que las estimaciones no cambien de manera notoria de una iteración a otra.

Cabe mencionar que el procedimiento de Cochrane-Orcutt es equivalente asintóticamente al de Hildreth-Lu. Además ambos procedimientos son aplicables al modelo de regresión lineal

múltiple, sin embargo, el procedimiento de Cochrane-Orcutt no realiza la estimación simultánea de ρ con los demás parámetros del modelo, mientras que el de Hildreth-Lu sí, por lo tanto, con muestras pequeñas es preferible el método de Hildreth-Lu.

3) Solución por MCG

Para este caso, supóngase que

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \text{ con } \varepsilon_t = \rho\varepsilon_{t-1} + \delta_t \text{ para } t = \dots, -2, -1, 0, 1, 2, \dots$$

además

$$E(\boldsymbol{\delta}) = \mathbf{0} \quad \text{y} \quad E(\boldsymbol{\delta}\boldsymbol{\delta}') = \sigma_{\delta}^2 \mathbf{I}. \quad (3.60)$$

Se supondrá también que $|\rho| < 1$. Así pues se tiene que

$$\begin{aligned} \varepsilon_t &= \rho\varepsilon_{t-1} + \delta_t \\ &= \rho(\rho\varepsilon_{t-2} + \delta_{t-1}) + \delta_t \\ &= \rho(\rho(\rho\varepsilon_{t-3} + \delta_{t-2}) + \delta_{t-1}) + \delta_t \\ &\dots \\ &= \delta_t + \rho\delta_{t-1} + \rho^2\delta_{t-2} + \dots \end{aligned}$$

Por lo tanto

$$E(\varepsilon_t) = E(\delta_t) + \rho E(\delta_{t-1}) + \rho^2 E(\delta_{t-2}) + \dots = 0. \quad (3.61)$$

Asimismo se llega a que

$$\begin{aligned}
 \text{Var}(\varepsilon_t) &= \text{Var}(\delta_t) + \rho^2 \text{Var}(\delta_{t-1}) + \rho^4 \text{Var}(\delta_{t-2}) + \dots \\
 &= \sigma_\delta^2 (1 + \rho^2 + \rho^4 + \dots) \\
 &= \sigma_\delta^2 \left(\frac{1}{1 - \rho^2} \right) \tag{3.62}
 \end{aligned}$$

y, de manera similar,

$$\begin{aligned}
 \text{Cov}(\varepsilon_t, \varepsilon_{t+1}) &= E(\varepsilon_t \varepsilon_{t+1}) \\
 &= E\left[\varepsilon_t (\rho \varepsilon_t + \delta_{t+1}) \right] \\
 &= E\left[\rho \varepsilon_t^2 + \varepsilon_t \delta_{t+1} \right] \\
 &= \rho \text{Var}(\varepsilon_t) \\
 &= \sigma_\delta^2 \left(\frac{\rho}{1 - \rho^2} \right)
 \end{aligned}$$

$$\text{Cov}(\varepsilon_t, \varepsilon_{t+2}) = \rho^2 \text{Var}(\varepsilon_t) = \sigma_\delta^2 \left(\frac{\rho^2}{1 - \rho^2} \right)$$

así que, de manera general

$$\text{Cov}(\varepsilon_t, \varepsilon_{t+h}) = \sigma_\delta^2 \left(\frac{\rho^h}{1 - \rho^2} \right) \text{ para } h = 0, 1, 2, \dots \tag{3.63}$$

Por consiguiente, en términos matriciales se llega a

$$\text{Var}(\varepsilon) = \frac{\sigma_\varepsilon^2}{1-\rho^2} \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ \dots & \dots & \dots & \dots & \dots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{pmatrix} = \sigma_\varepsilon^2 \Omega.$$

Además, se puede demostrar que

$$\Omega^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho & 0 & \dots & 0 \\ -\rho & 1+\rho^2 & -\rho & \dots & 0 \\ 0 & -\rho & 1+\rho^2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix} \quad (3.64)$$

y así se puede aplicar MCG con la matriz Ω citada. Se requiere, sin embargo, un valor preliminar para ρ , el cual puede provenir de la estadística de Durbin-Watson, ya que $d \approx 2(1-\hat{\rho})$.

4) Método AR(1)

El método AR(1) en los errores hace uso de la ecuación

$$Y_t = \beta_0(1-\rho) + \beta_1 X_t - \beta_1 \rho X_{t-1} + \rho Y_{t-1} + u_t \quad (3.65)$$

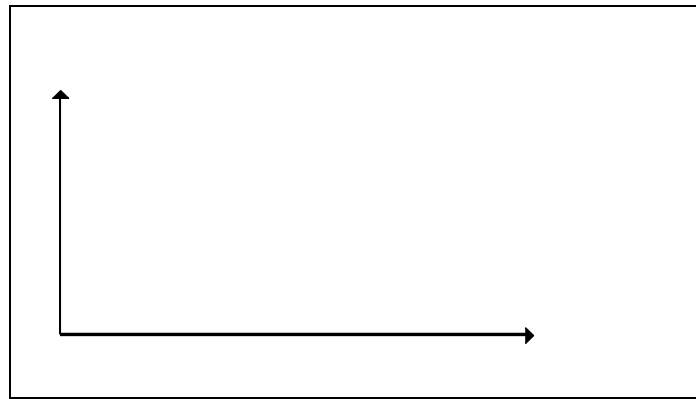
y estima simultáneamente los parámetros β_0, β_1 y ρ mediante un procedimiento no-lineal.

3.9 Heteroscedasticidad

Este problema se presenta cuando se viola el supuesto de varianza constante, es decir, cuando existe inestabilidad en la varianza; lo opuesto (varianza constante) se llama homoscedasticidad.

Este tipo de problema surge con frecuencia, cuando la respuesta que se estudia tiene un recorrido de observación muy extenso (digamos de 0 a 100,000 unidades). Entonces la variabilidad tiende a ser proporcional a la magnitud de las observaciones.

Por ejemplo, en un diagrama de dispersión de ingreso contra edad, se observará lo siguiente.



En este caso, las observaciones del ingreso para edades cercanas a veinte, muestran menor variabilidad que para las edades cercanas a 60.

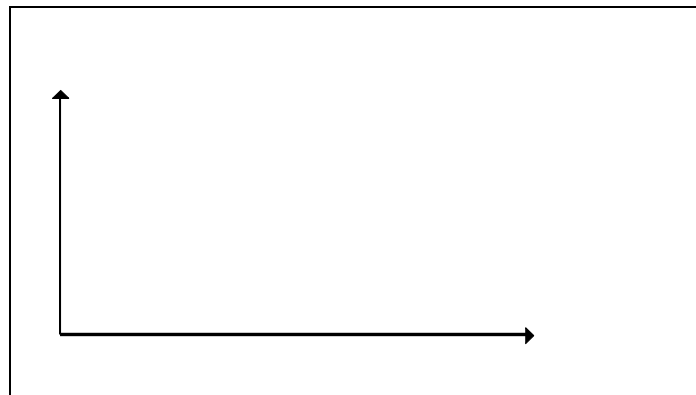
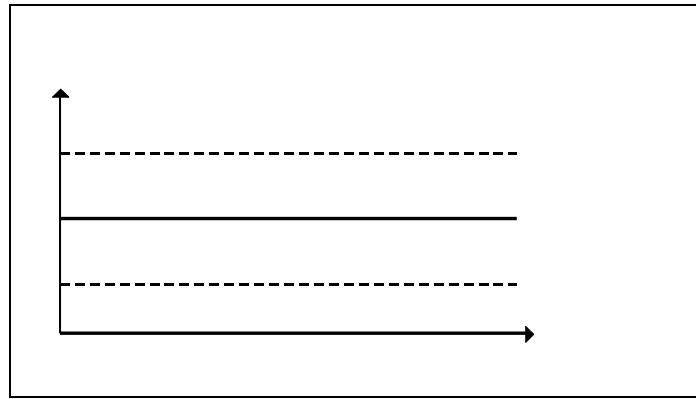
Consecuencias

El efecto que la violación al supuesto de varianza constante puede tener, se refleja en que las pruebas F, presentarán más resultados significativos de los que realmente deberían serlo. Por otro lado, el estimador MCO de β no será eficiente, puesto que existe en tal caso otro estimador lineal e insesgado cuya varianza es menor que la de \mathbf{b} (dicho estimador es el de MCG).

Detección

Se puede llevar a cabo mediante gráficas de residuos o bien a través de pruebas estadísticas desarrolladas para este fin. Tales pruebas se encuentran en el libro de Pyndick y Rubinfeld (1981).

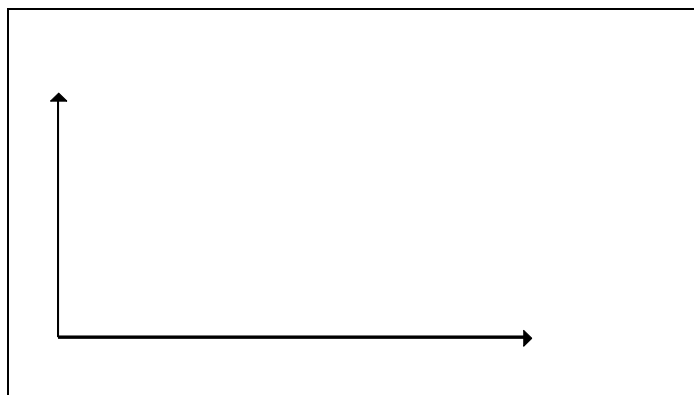
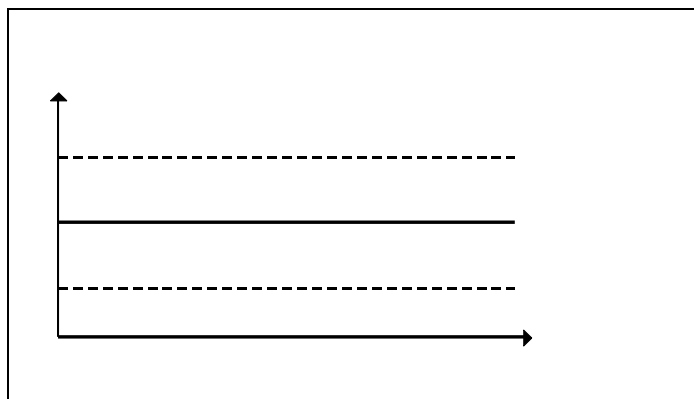
En la práctica una buena idea es graficar los residuos. En particular las gráficas contra \hat{Y} y las variables explicativas son bastante elocuentes al respecto.



Corrección

Cuando se presenta el caso de la gráfica (b), la manera de corregir la violación del supuesto es mediante una transformación potencia a los datos Y , ya que aparentemente la heteroscedasticidad está en función de los valores estimados de la variable dependiente.

Por otro lado, si lo que ocurre es que la varianza depende de alguna variable explicativa, lo que se debe hacer en tal situación es usar MCG. Supóngase que al trabajar con el modelo $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ se obtienen las siguientes gráficas.



Entonces, la variable involucrada en la falta de constancia de la varianza residual es X_2 . Si ahora se supone que la varianza de ε_i es directamente proporcional al cuadrado de X_{2i} , es decir, si se postula que

$$\text{Var}(\varepsilon_i) = kX_{2i}^2 \quad \text{para } i = 1, \dots, n \quad (3.66)$$

entonces la estructura de varianza-covarianza estará dada por la matriz

$$\Omega = \begin{pmatrix} X_{21}^2 & 0 & \dots & 0 \\ 0 & X_{22}^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & X_{2n}^2 \end{pmatrix} .$$

En tal caso puede utilizarse el método de MCG, que produce el estimador

$$\hat{\beta} = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} Y.$$

De hecho, cuando se conoce cuál de las variables explicativas es la relacionada con la varianza residual, el método de MCG se conoce como Mínimos Cuadrados Ponderados. Por ejemplo, en el caso de regresión lineal simple

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad \text{con} \quad \text{Var}(\varepsilon_i) = k X_i^2 \quad (3.67)$$

se procede a transformarlo de la siguiente forma

$$\frac{Y_i}{X_i} = \frac{\beta_0}{X_i} + \beta_1 + \frac{\varepsilon_i}{X_i} .$$

Así se obtiene entonces el modelo

$$Y_i' = \beta_0' + \beta_1' X_i' + \delta_i \quad (3.68)$$

en donde $Y_i' = Y_i/X_i$, $X_i' = 1/X_i$, $\beta_0' = \beta_1$, $\beta_1' = \beta_0$ y $\delta_i = \varepsilon_i/X_i$.

Además, nótese que

$$E(\delta_i) = E\left(\frac{\varepsilon_i}{X_i}\right) = 0, \quad \text{Var}(\delta_i) = E\left(\frac{\varepsilon_i}{X_i}\right)^2 = \frac{\sigma_\varepsilon^2}{X_i^2} = k$$

y

$$\text{Cov}(\delta_i, \delta_j) = \text{Cov}\left(\frac{\varepsilon_i}{X_i}, \frac{\varepsilon_j}{X_j}\right) = \frac{1}{X_i^2 X_j^2} \text{Cov}(\varepsilon_i, \varepsilon_j) = 0.$$

De esta forma se puede aplicar el método de MCO en el modelo para las variables transformadas.

3.10 No-linealidad

El hecho de que el modelo sea lineal en los parámetros implica que

$$Y = f(X_1, X_2, \dots, X_k; \beta_1, \beta_2, \dots, \beta_p; \varepsilon)$$

de forma que $\frac{\delta Y}{\delta \beta_j}$ no depende de β_i para toda i y $j = 1, \dots, p$. Si la teoría que sustenta el fenómeno en estudio indica un modelo no-lineal, el método de estimación a utilizar debe ser no-lineal (véase por ejemplo Pyndick y Rubinfeld, 1981).

Por otro lado, el supuesto de linealidad en la forma funcional indica que las variables independientes entran de manera lineal en el modelo.

Consecuencias

La violación del supuesto implicará esencialmente, sesgos en los valores estimados y en los pronósticos.

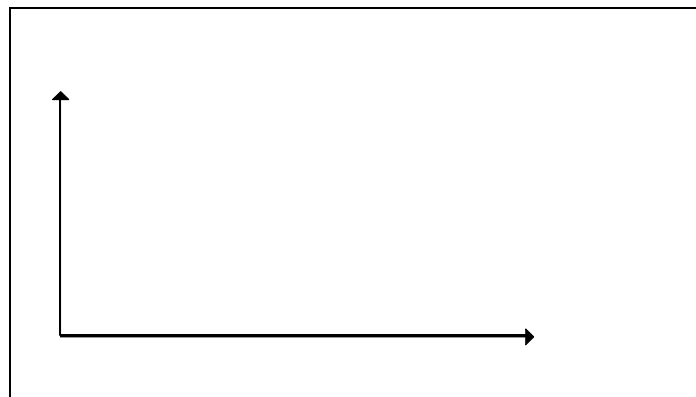
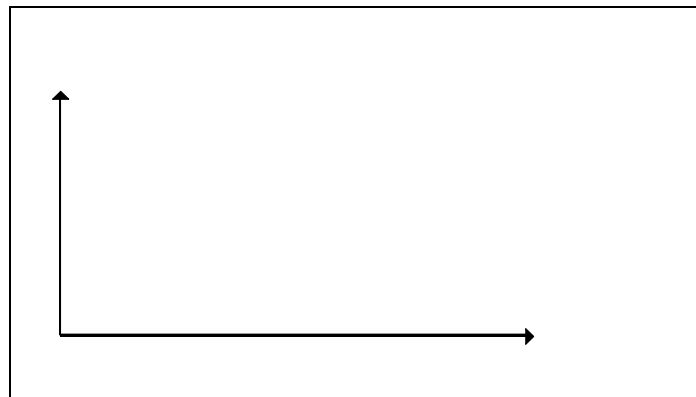
Detección

En la práctica, la violación al supuesto de linealidad en las variables, se detecta mediante gráficas, ya sea de los datos originales Y's o de los residuos.

Corrección

El remedio a la violación al supuesto de linealidad en las variables se basa en el uso de transformaciones y/o en la inclusión de otras variables explicativas.

Existen modelos no-lineales en las variables, que son intrínsecamente linealizables, o sea que mediante una transformación se vuelven lineales



Los modelos linealizables más comunes son los siguientes.

Modelo no-lineal

$$Y = f(\beta_0, \beta_1; X; \varepsilon)$$

Modelo lineal transformado

$$Y' = \beta_0' + \beta_1' X' + \varepsilon$$

$$Y = \beta_0 e^{\beta_1 X} \varepsilon$$

$$Y' = \log(Y) ; X' = X ; \beta_0' = \log(\beta_0) ; \beta_1' = \beta_1$$

$$Y = \beta_0 X^{\beta_1} \varepsilon$$

$$Y' = \log(Y) ; X' = \log(X) ; \beta_0' = \log(\beta_0) ; \beta_1' = \beta_1$$

$$Y = \frac{1}{(\beta_0 + \beta_1 X + \varepsilon)}$$

$$Y' = \frac{1}{Y} ; X' = X ; \beta_0' = \beta_0 ; \beta_1' = \beta_1$$

$$Y = \frac{X}{(\beta_0 + \beta_1 X + \varepsilon/X)}$$

$$Y' = \frac{1}{Y} ; X' = \frac{1}{X} ; \beta_0' = \beta_1 ; \beta_1' = \beta_0$$

No todos los modelos no-lineales pueden linealizarse. Por ejemplo el modelo

$$Y = \beta_0 e^{\beta_1 X_1} + e^{\beta_2 X_2} + \varepsilon$$

no es linealizable.

En otros casos, no conviene linealizar el modelo. Por ejemplo, considérese el modelo

$$Y = \frac{e^{-(\beta_0 + \beta_1 X + \varepsilon)}}{[1 + e^{-(\beta_0 + \beta_1 X + \varepsilon)}]} \quad \text{que es válido para } 0 < Y < 1$$

y que puede ser linealizado mediante $\log\left(\frac{Y}{1-Y}\right)$, pero al aplicar los métodos lineales al modelo transformado, se pierde la eficiencia respecto a la estimación del modelo original.

Para apreciar esto, sea Y_i una proporción observada de h_i experimentos independientes, de tal forma que

$$Y_i = \sum_{j=1}^{h_i} Z_{ji} / h_i \quad \text{para } i = 1, \dots, n$$

donde las variables Z_{ji} son independientes e idénticamente distribuidas como $Z_{ji} \sim \text{Bernoulli}(p_i)$, con p_i la verdadera proporción (poblacional). Entonces

$\sum_{j=1}^{h_i} Z_{ji} \sim \text{Binomial}(h_i, p_i)$, así que

$$E(Y_i) = E\left(\sum_{j=1}^{h_i} Z_{ji} / h_i\right) = \frac{1}{h_i} E\left(\sum_{j=1}^{h_i} Z_{ji}\right) = \frac{h_i p_i}{h_i} = p_i$$

y

$$\text{Var}(Y_i) = \text{Var}\left(\sum_{j=1}^{h_i} Z_{ji} / h_i\right) = \frac{1}{h_i^2} \text{Var}\left(\sum_{j=1}^{h_i} Z_{ji}\right) = \frac{1}{h_i^2} h_i p_i (1 - p_i) = p_i (1 - p_i) / h_i.$$

Ahora, supóngase que

$$Y_i = p_i + \eta_i$$

con η_i un error aleatorio tal que

$$E(\eta_i) = E(Y_i - p_i) = E(Y_i) - p_i = 0 \quad ,$$

$$\text{Var}(\eta_i) = \text{Var}(Y_i - p_i) = \text{Var}(Y_i) = p_i(1-p_i)/h_i \quad \text{y} \quad \text{Cov}(\eta_i, \eta_j) = \text{Cov}(Y_i, Y_j) = 0 \quad , i \neq j$$

Por lo tanto, expandiendo en serie de Taylor alrededor de p_i , se llega a

$$\begin{aligned} \log\left(\frac{Y_i}{1-Y_i}\right) &\doteq \log\left(\frac{p_i}{1-p_i}\right) + \left. \frac{d \log[Y_i/(1-Y_i)]}{d Y_i} \right|_{Y_i=p_i} (Y_i - p_i) \\ &= \log\left(\frac{p_i}{1-p_i}\right) + \left. \left[\frac{1}{Y_i(1-Y_i)} \right] \right|_{Y_i=p_i} (\eta_i) \\ &= \log\left(\frac{p_i}{1-p_i}\right) + \frac{\eta_i}{p_i(1-p_i)} \quad . \end{aligned}$$

De esta manera, si se supone que

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_i$$

se obtiene

$$\log\left(\frac{Y_i}{1-Y_i}\right) = \beta_0 + \beta_1 X_i + \varepsilon_i$$

en donde ε_i es un error aleatorio tal que

$$E(\varepsilon_i) = E\left[\frac{\eta_i}{p_i(1-p_i)}\right] = \frac{1}{p_i(1-p_i)} E(\eta_i) = 0 \quad ,$$

$$\text{Cov}(\varepsilon_i, \varepsilon_j) = \text{Cov}\left(\frac{\eta_i}{p_i(1-p_i)}, \frac{\eta_j}{p_j(1-p_j)}\right) = 0$$

y

$$\text{Var}(\varepsilon_i) = \text{Var}\left[\frac{\eta_i}{p_i(1-p_i)}\right] = \frac{1}{[p_i(1-p_i)]^2} \text{Var}(\eta_i) = \frac{1}{h_i p_i(1-p_i)}.$$

Como la varianza del error no es constante, no se debería usar MCO sino MCG, con la matriz $\Omega = \left\{ \frac{1}{h_i p_i(1-p_i)} \right\}$ que se podría estimar como $\bar{\Omega} = \left\{ \frac{1}{h_i Y_i(1-Y_i)} \right\}$.

Otra razón que hay que tener en cuenta para no-linealizar, es que en ocasiones se sabe que los errores son aditivos en la escala original y que, por lo tanto, no serán aditivos en la escala transformada, si se aplica una transformación no-lineal a los datos.

En general, los modelos intrínsecamente linealizables se pueden expresar como

$$Y^{\lambda_0} = \beta_0 + \sum \beta_j X_j^{\lambda_j} \quad \text{para } Y \text{ y } X_j > 0.$$

Cuando se ha llegado a la conclusión de que es necesario aplicar una transformación, hay que considerar que existen transformaciones para la variable dependiente y para las independientes. La manera más frecuentemente usada para seleccionar una transformación es por ensayo y error.

3.11 No-normalidad de los errores.

El supuesto de normalidad permite realizar inferencia estadística a partir de los datos muestrales.

Consecuencias

Aun cuando la inferencia sobre los parámetros de la regresión (basada en distribuciones t y F) no se ve seriamente invalidada por la no-normalidad, los estimadores b_0 y b_1 pueden ser sensibles al efecto de distribuciones con asimetría o colas pesadas y, por consiguiente, podrían tomar valores que no correspondan a la realidad.

Detección

La manera más inmediata de detectar la no-normalidad en los errores es a través de un histograma de residuos (o diagrama de punto) en donde se pueda apreciar, a simple vista, si la distribución es Normal o no (de manera aproximada) por su asimetría o su kurtosis.



Pruebas más formales se basan en el uso de los estadísticos de asimetría (S_k) y kurtosis (K_u) de los residuos. Se sabe que, para n grande

$$S_k = \left[\sum_{i=1}^n (e_i - \bar{e})^3 / n \right] / \left[\sum_{i=1}^n (e_i - \bar{e})^2 / n \right]^{3/2} \approx N\left(0, \frac{6}{n}\right)$$

y

$$K_u = \left[\sum_{i=1}^n (e_i - \bar{e})^4 / n \right] / \left[\sum_{i=1}^n (e_i - \bar{e})^2 / n \right]^2 \approx N\left(3, \frac{24}{n}\right)$$

El uso combinado de estos estadísticos es usado en la prueba de Jarque-Bera para construir el estadístico de normalidad

$$\left(\frac{Sk}{\sqrt{6/n}}\right)^2 + \left(\frac{Ku-3}{\sqrt{24/n}}\right)^2 = \frac{n}{24} [4Sk^2 + (Ku-3)^2] \sim \chi^2_{(2)}$$

Corrección

Mediante el uso de transformaciones potencia (por ensayo y error, ya que así se aprovecha el orden de la familia de transformaciones).

3.12 Media del error distinta de cero

Este supuesto muchas veces se pasa de largo, ya que se verifica cuando $\bar{e} = 0$, lo cual queda garantizado cuando existe la ordenada al origen en el modelo y se usa MCO para estimar los parámetros.

Consecuencias

En caso de que no exista ordenada al origen y $\bar{e} \neq 0$, se podría provocar un sesgo en los pronósticos del modelo.

Detección

Una manera de verificar el supuesto consiste en probar la hipótesis

$$H_0: E(\epsilon_i) = 0 \text{ vs. } H_A: E(\epsilon_i) \neq 0$$

y usar como estadístico de prueba la media de los residuos estandarizada, que sigue una distribución t de Student con n-2 grados de libertad, o sea,

$$\bar{e}/S \sim t_{(n-2)}.$$

Corrección

En caso de que se detecte media del error diferente de cero, la corrección se logra al introducir la constante en el modelo y volver a estimar los parámetros.

3.13 Observaciones atípicas

El supuesto está implícito en el hecho de que el modelo que se postula es válido para todas y cada una de las observaciones, o sea, para $i = 1, \dots, n$.

Consecuencias

Las consecuencias de que haya observaciones atípicas es muy diverso, pues dependen del número de ellas y de su tipo. Algunas pueden simplemente afectar de manera aditiva a la observación, sin que haya más efectos. En otros casos pueden ser observaciones que estén asociadas con cambios en la estructura del fenómeno en estudio y la presencia de dichas observaciones deja el modelo inútil para hacer análisis.

Detección

Una manera de verificar este supuesto es mediante las gráficas de los residuos usadas para detectar no-normalidad. La presencia de observaciones aberrantes se detecta cuando aparecen residuos con valores mayores que un cierto número de desviaciones estándar. Comúnmente se eligen tres desviaciones estándar, debido a que en una distribución Normal este tipo de datos se presentan con probabilidad de 0.02% (2 por cada mil observaciones) de manera que su presencia se considera un evento raro y, por ende, digno de ser estudiado en detalle.

Corrección

La manera más sencilla de corregir esta violación al supuesto consiste en cancelar la observación aberrante correspondiente. Sin embargo, esta solución no es del todo recomendable y sería preferible averiguar primero, a qué se debe que dicha observación tenga el valor atípico. Si es debido a error de copiado, la solución evidente es corregir el dato, pero si no es el caso, debería indagarse primero qué tanto afecta dicha observación a los resultados del modelo. Para esto conviene sustituir el valor aberrante por el valor estimado por el modelo y observar cómo cambian los resultados. Si no cambian radicalmente, de manera que las conclusiones que surjan del modelo en esencia se mantienen, no conviene eliminar el dato. En caso contrario, cuando sí cambian notoriamente, lo indicado es presentar los resultados del análisis con y sin la corrección por el dato atípico.

3.14 Colinealidad

Este problema surge en ocasiones cuando se construyen modelos de regresión lineal múltiple, con datos no planeados.

Cabe recordar que en MCO se hace uso de

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} \quad \text{y de } \text{Var}(\mathbf{b}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}.$$

Entonces, si $|\mathbf{X}'\mathbf{X}| = 0$ no se puede obtener $(\mathbf{X}'\mathbf{X})^{-1}$, y si $|\mathbf{X}'\mathbf{X}| \approx 0$ (existen dependencias lineales en las columnas de \mathbf{X}) se dice que existe colinealidad (una variable explica a otra), lo cual es de hecho un defecto de la matriz \mathbf{X} .

La forma en que afecta este problema, se refleja en que las varianzas de los estimadores mínimo-cuadráticos se verán infladas y puede ser que no se obtengan coeficientes significativamente distintos de cero, aun cuando sí exista la relación lineal entre las variables explicativas y la respuesta. Por ejemplo considérese el modelo

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

que, expresado en términos de las variables en forma de desviación con respecto a la media $y = Y - \bar{Y}$, $x_1 = X_1 - \bar{X}_1$ y $x_2 = X_2 - \bar{X}_2$, se convierte en

$$y = \beta_1 x_1 + \beta_2 x_2 + \varepsilon.$$

Supóngase ahora que existe una relación lineal entre X_1 y X_2 , o sea que $X_1 = \alpha + \beta X_2$ y $x_1 = \beta x_2$, de tal forma que

$$X'X = \begin{pmatrix} \sum x_{1i}^2 & \sum x_{1i}x_{2i} \\ \sum x_{1i}x_{2i} & \sum x_{2i}^2 \end{pmatrix} = \begin{pmatrix} \beta^2 \sum x_{2i}^2 & \beta \sum x_{2i}^2 \\ \beta \sum x_{2i}^2 & \sum x_{2i}^2 \end{pmatrix} = \sum x_{2i}^2 \begin{pmatrix} \beta^2 & \beta \\ \beta & 1 \end{pmatrix}.$$

Por consiguiente $|X'X| = 0$ y no se pueden estimar los parámetros β_1 y β_2 del modelo (en consecuencia, tampoco β_0).

Debido a que la colinealidad depende directamente de la matriz X , se tienen muy pocos elementos para resolver el problema.

Si se cuenta con información adicional sobre el fenómeno que se estudia, dicha información permitirá, quizá, cancelar alguna o algunas de las variables sin perder explicación, o bien, para incluir dicha información en el modelo. Por ejemplo, si alguno de los coeficientes se sabe que está en función del coeficiente de otra variable, las dos variables respectivas pueden combinarse en una sola para poder estimar los coeficientes con la muestra disponible. Por ejemplo, supóngase que $\beta_2 = \frac{\beta_1}{2}$ en la ecuación previa. Entonces se tiene que

$$Y = \beta_0 + \beta_1 X_1 + \frac{\beta_1}{2} X_2 + \varepsilon$$

$$= \beta_0 + \beta_1 (X_1 + X_2/2) + \varepsilon,$$

ecuación esta última, que puede estimarse como regresión lineal simple.

Una variante de esta situación se usa frecuentemente en estudios de demanda y consumo, en los cuales los intentos por estimar los efectos de precios y de ingreso sobre la demanda, a partir de datos de series de tiempo, son generalmente infructuosos porque los precios y el ingreso se mueven de forma simultánea. Así pues, para estimar la función de demanda

$$q = \beta_1 y + \beta_2 p + \varepsilon$$

en donde q es la cantidad demandada del producto, y es el ingreso y p el precio, se puede usar una combinación de datos de corte transversal con datos de series de tiempo. El coeficiente estimado del ingreso b_1 , se obtiene con datos de corte transversal y se construye entonces la demanda ajustada

$$q^* = q - b_1 y.$$

Después, la regresión de q^* sobre p con datos de series de tiempo, sirve sólo para estimar b_2 .

La forma más fácil de observar si la multicolinealidad causa problemas, es mediante el examen de los errores estándar de los coeficientes. Si varios coeficientes tienen errores estándar grandes y al quitar una o más variables de la ecuación, disminuyen los errores estándar de las variables restantes, seguramente la colinealidad es la causa del problema.

Una regla simplista comúnmente utilizada, indica que la colinealidad es probablemente un problema si la correlación simple entre dos variables explicativas, es mayor que la correlación de cualquiera de estas variables con la dependiente. Esta regla resulta ser bastante razonable si hay solamente dos variables independientes en el modelo, pero no es

del todo apropiada en el caso de regresión lineal múltiple, y esto es debido a que las correlaciones simples no brindan información acerca de posibles dependencias lineales múltiples entre las variables independientes.

En la búsqueda de extensiones de la noción de correlación simple, para tener en cuenta el hecho de que otras variables de hecho permanecen constantes cuando se examina a un determinado parámetro de regresión, se llega al concepto de *correlación parcial*.

Para introducir este concepto considérese el modelo

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon.$$

La idea es de que el coeficiente de correlación parcial entre Y y X_1 mida el efecto de X_1 sobre Y, independientemente del efecto de X_2 sobre Y. Para obtener esto, se sobreentiende que se debe eliminar la influencia de X_2 sobre ambas Y y X_1 , lo cual se logra al calcular las desviaciones residuales

$$e_Y = Y - \hat{Y} = Y - \hat{\alpha}_Y - \hat{\beta}_Y X_2 \quad \text{y} \quad e_{X_1} = X_1 - \hat{X}_1 = X_1 - \hat{\alpha}_{X_1} - \hat{\beta}_{X_1} X_2,$$

que son las partes de Y y de X_1 , no-explicadas por X_2 . Entonces, el coeficiente de correlación parcial se obtiene como la correlación simple entre e_Y y e_{X_1} , o sea, $r_{YX_1 \cdot X_2} = r_{e_Y e_{X_1}}$. Después de algunos pasos de álgebra se llega a la expresión

$$r_{YX_1 \cdot X_2} = \frac{r_{YX_1} - r_{YX_2} r_{X_1 X_2}}{\sqrt{1 - r_{X_1 X_2}^2} \sqrt{1 - r_{YX_2}^2}}$$

fórmula que muestra explícitamente, que no tiene porqué existir una correspondencia muy estrecha entre correlación simple y correlación parcial. Sin embargo, cuando Y y X_1 no

están correlacionados con X_2 (de manera que $r_{YX_2} = 0 = r_{X_1X_2}$) la fórmula se reduce a la de correlación simple.

De manera similar se define

$$r_{YX_2 \cdot X_1} = \frac{r_{YX_2} - r_{YX_1} r_{X_2X_1}}{\sqrt{1 - r_{X_2X_1}^2} \sqrt{1 - r_{YX_1}^2}}$$

Es de esperar que, en general $-1 \leq r_{YX_1 \cdot X_2} = r_{e_Y e_{X_1}} \leq 1$, al igual que ocurre con las correlaciones simples (ya que $r_{e_Y e_{X_1}}$ es lo que se calcula, como se mencionó antes) sin embargo, no es tan claro con la fórmula, en particular cuando $r_{X_2X_1}$ o r_{YX_1} tiende a ± 1 . De hecho, si X_1 y X_2 están perfectamente correlacionadas, $r_{X_1X_2} = 1$ y $r_{YX_1 \cdot X_2}$ no puede ser calculado. Además, si $r_{YX_1 \cdot X_2} = r_{e_Y e_{X_1}} = 0$, se concluye que no hay relación lineal entre Y y X_1 después de que se ha tenido en cuenta el efecto lineal de X_2 .