El modelo de Regresión Lineal Múltiple (RLM)

Erika R. Badillo
erika.badilloen@unaula.edu.co

Facultad de Economía
Universidad Autónoma Latinoamericana

- El modelo
- El modelo en matrices
- Hipótesis de partida en matrices
- Estimación MCO
- Propiedades de los estimadores MCO
- Inferencia estadística

Lecturas

- Wooldridge, Jeffrey (2013). *Introducción a la econometría*. 5a edición, Cengage Learning. Cap. 3, 4 y 5
- Gujarati, D. y Porter, D. (2010). *Econometría*. 5a edición, Mc Graw Hill. Cap. 7 y 8
- Judge, G., Hill, R., Griffiths, W., Lütkepohl, H. y Lee, T. (1988). *Introduction to the theory and practice of econometrics*. 2a edición, John Wiley & Sons. Cap. 5

El modelo

Un modelo de regresión lineal múltiple tiene la siguiente estructura:

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + ... + \beta_k X_{ik} + u_i$$

El modelo entonces contiene k parámetros poblacionales (desconocidos)

Se trabaja, entonces, con $n - k$ gdl

La pregunta que surge es: ¿será suficiente decir que el modelo de RLM es una extensión desde: simple$\Longrightarrow$múltiple?

El modelo

Recordando las hipótesis de partida (supuestos iniciales):

- Los coeficientes β_j con $j = 1, 2, 3, ..., k$ son fijos y desconocidos
- Las X_{ij} son estocásticamente fijas para $j = 2, 3, 4, ..., k$. Este es un supuesto de partida propio del laboratorio. Un supuesto más real en economía y que lleva a resultados similares es: las variables explicatorias son exógenas. Esto implica que (hipótesis de exogeneidad):

$$\begin{gathered} Cov(X_{i2}, u_i) = 0 \\ Cov(X_{i3}, u_i) = 0 \\ . \\ . \\ Cov(X_{ik}, u_i) = 0 \end{gathered}$$

- El modelo esta completo: $E(u_i) = 0, \forall i = 1...n$
- Homocedasticidad: $Var(u_i) = E(u_i - E(u_i))^2 = E(u_i^2) = \sigma_u^2$
- No autocorrelación: $Cov(u_i, u_j) = E[(u_i - E(u_i))(u_j - E(u_j))] = E(u_i u_j) = 0, \forall i \neq j$
- Normalidad: $u_i \sim NID(0, \sigma_u^2)$

El modelo

Se puede intentar aplicar MCO para obtener los estimadores del modelo. Partiendo de un modelo estimado con intercepto y tres variables explicativas:

$$Y_i = \widehat{\beta}_1 + \widehat{\beta}_2 X_{i2} + \widehat{\beta}_3 X_{i3} + \widehat{\beta}_4 X_{i4} + \widehat{u}_i$$

u_i es el residuo de la regresión

Se construye la SCR:

$$\sum \widehat{u}_i^2 = \sum (Y_i - \widehat{\beta}_1 - \widehat{\beta}_2 X_{i2} - \widehat{\beta}_3 X_{i3} - \widehat{\beta}_4 X_{i4})^2$$

Se minimiza la SCR respecto a los $\beta_j, j = 1, 2, 3, 4$

$$\frac{\partial \sum \widehat{u}_i^2}{\partial \widehat{\beta}_1} = -2 \sum (Y_i - \widehat{\beta}_1 - \widehat{\beta}_2 X_{i2} - \widehat{\beta}_3 X_{i3} - \widehat{\beta}_4 X_{i4}) = 0$$

$$\frac{\partial \sum \widehat{u}_i^2}{\partial \widehat{\beta}_2} = -2 \sum (Y_i - \widehat{\beta}_1 - \widehat{\beta}_2 X_{i2} - \widehat{\beta}_3 X_{i3} - \widehat{\beta}_4 X_{i4}) X_{i2} = 0$$

$$\frac{\partial \sum \widehat{u}_i^2}{\partial \widehat{\beta}_3} = -2 \sum (Y_i - \widehat{\beta}_1 - \widehat{\beta}_2 X_{i2} - \widehat{\beta}_3 X_{i3} - \widehat{\beta}_4 X_{i4}) X_{i3} = 0$$

$$\frac{\partial \sum \widehat{u}_i^2}{\partial \widehat{\beta}_4} = -2 \sum (Y_i - \widehat{\beta}_1 - \widehat{\beta}_2 X_{i2} - \widehat{\beta}_3 X_{i3} - \widehat{\beta}_4 X_{i4}) X_{i4} = 0$$

El modelo

Reordenando términos se obtienen las 4 ecuaciones normales:

$$\sum Y_i = n\widehat{\beta}_1 + \widehat{\beta}_2 \sum X_{i2} + \widehat{\beta}_3 \sum X_{i3} + \widehat{\beta}_4 \sum X_{i4}$$

$$\sum Y_i X_{i2} = \widehat{\beta}_1 \sum X_{i2} + \widehat{\beta}_2 \sum X_{i2}^2 + \widehat{\beta}_3 \sum X_{i3} X_{i2} + \widehat{\beta}_4 \sum X_{i4} X_{i2}$$

$$\sum Y_i X_{i3} = \widehat{\beta}_1 \sum X_{i3} + \widehat{\beta}_2 \sum X_{i2} X_{i3} + \widehat{\beta}_3 \sum X_{i3}^2 + \widehat{\beta}_4 \sum X_{i4} X_{i3}$$

$$\sum Y_i X_{i4} = \widehat{\beta}_1 \sum X_{i4} + \widehat{\beta}_2 \sum X_{i2} X_{i4} + \widehat{\beta}_3 \sum X_{i3} X_{i4} + \widehat{\beta}_4 \sum X_{i4}^2$$

¿Cómo resolver este sistema?

Al intentar resolver el sistema para despejar los β_j se encuentra que es "costoso" hacerlo con la notación que se tiene. Lo que se puede hacer es una agrupación en matrices.

El modelo en matrices

El modelo general es un polinomio de regresión de la forma:

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + ... + \beta_k X_{ik} + u_i$$

Lo que el modelo dice es:

$$\begin{gathered}
Y_1 = \beta_1 + \beta_2 X_{12} + \beta_3 X_{13} + ... + \beta_k X_{1k} + u_1 \\
Y_2 = \beta_1 + \beta_1 X_{22} + \beta_3 X_{23} + ... + \beta_k X_{2k} + u_2 \\
\vdots \\
Y_n = \beta_1 + \beta_2 X_{n2} + \beta_3 X_{n3} + ... + \beta_k X_{nk} + u_n
\end{gathered}$$

En matrices se tiene

$$\mathbf{Y}_{n\text{X}1} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} \qquad \mathbf{u}_{n\text{X}1} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_n \end{bmatrix} \qquad \mathbf{B}_{k\text{X}1} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_k \end{bmatrix} \qquad \mathbf{X}_{n\text{X}k} = \begin{bmatrix} 1 & X_{12} & X_{23} & X_{1k} \\ 1 & X_{22} & X_{23} & X_{2k} \\ 1 & X_{32} & X_{33} & X_{3k} \\ \vdots & & & \\ 1 & X_{n2} & X_{n3} & X_{nk} \end{bmatrix}$$

En esta notación X_{ij} indica: fila i (observación), columna j (variable explicatoria)

El polinomio de regresión en álgebra matricial se puede escribir como:

$$\mathbf{Y}_{n\text{X}1} = \mathbf{X}_{n\text{X}k} \mathbf{B}_{k\text{X}1} + \mathbf{u}_{n\text{X}1}$$

El modelo en matrices

En la expresión general $\mathbf{Y} = \mathbf{XB} + \mathbf{u}$ deben estar todos los modelos de regresión lineal hasta ahora conocidos y por conocer

- El ingenuo $\mathbf{X}_{n\text{x}1} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$ $\mathbf{B}_{1\text{x}1} = \begin{bmatrix} \beta_1 \end{bmatrix}$
- RLS $\mathbf{X}_{n\text{x}2} = \begin{bmatrix} 1 & X_{12} \\ 1 & X_{22} \\ 1 & X_{32} \\ \vdots & \\ 1 & X_{n2} \end{bmatrix}$ $\mathbf{B}_{2\text{x}1} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$
- RLS sin intercepto $\mathbf{X}_{n\text{x}1} = \begin{bmatrix} X_{12} \\ X_{22} \\ X_{32} \\ \vdots \\ X_{n2} \end{bmatrix}$ $\mathbf{B}_{1\text{x}1} = \begin{bmatrix} \beta_2 \end{bmatrix}$

Hipótesis de partida en matrices

Interesa ahora reconstruir las hipótesis acerca de u en álgebra lineal. Se necesita el concepto de esperanza de un vector o matriz, que es el vector o matriz de las esperanzas de cada elemento. Si $\mathbf{W}_{n\text{x}1}$ es un vector aleatorio entonces:

$$E(\mathbf{W}_{n\text{x}1}) = E\begin{bmatrix} W_1 \\ W_2 \\ W_3 \\ \vdots \\ W_n \end{bmatrix} = \begin{bmatrix} E(W_1) \\ E(W_2) \\ E(W_3) \\ \vdots \\ E(W_n) \end{bmatrix}$$

Las hipótesis sobre u en notación matricial se pueden expresar como:

- Modelo completo: $\begin{bmatrix} E(u_1) \\ E(u_2) \\ E(u_3) \\ \vdots \\ E(u_n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \Longrightarrow E\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \Longrightarrow E\left(\mathbf{u}_{n\text{x}1}\right) = \mathbf{0}_{n\text{x}1}$

Hipótesis de partida en matrices

Se requiere el concepto de matriz de varianzas de un vector

$$Var - Cov(\mathbf{W}_{n\text{x}1}) = E[(\mathbf{W}_{n\text{x}1} - E(\mathbf{W}_{n\text{x}1}))(\mathbf{W}_{n\text{x}1} - E(\mathbf{W}_{n\text{x}1}))']$$

Desarrollando esto, se obtiene que: $\text{Var-Cov}(\mathbf{W}_{n\text{x}1})_{n\text{x}n} = \begin{bmatrix} Var(W_1) & Cov(W_1,W_2) & Cov(W_1,W_3) & \dots & Cov(W_1,W_n) \\ Cov(W_2,W_1) & Var(W_2) & Cov(W_2,W_3) & \dots & Cov(W_2,W_n) \\ Cov(W_3,W_1) & Cov(W_3,W_2) & Var(W_3) & \dots & Cov(W_3,W_n) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ Cov(W_n,W_1) & Cov(W_n,W_2) & Cov(W_n,W_3) & \dots & Var(W_n) \end{bmatrix}$

La matriz $Var - Cov(\mathbf{W}_{n\text{x}1})$ contiene:

- sobre la diagonal principal las varianzas de cada una de los elementos del vector **W**
- el resto de elementos son las covarianzas entre los elementos $Cov(W_i, W_j), \forall i \neq j$

Hay que recordar que hay dos supuestos acerca de u_i que se refiere a varianzas y covarianzas y que por ende en el mundo matricial deben ir juntos:

- Homocedasticidad: $Var(u_i) = E(u_i^2) = \sigma_u^2$
- No autocorrelación: $Cov(u_i, u_j) = E(u_i u_j) = 0, \forall i \neq j$

Si se quiere organizar una matriz con esta información se procedería así

$$\text{Var-Cov}(\mathbf{u}) = \begin{bmatrix} E(u_1^2) & E(u_1u_2) & E(u_1u_3) & \dots & E(u_1u_n) \\ E(u_1u_2) & E(u_2^2) & E(u_2u_3) & \dots & E(u_2u_n) \\ E(u_1u_3) & E(u_3u_2) & E(u_3^2) & \dots & E(u_3u_n) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ E(u_1u_n) & E(u_2u_n) & E(u_3u_n) & \dots & E(u_n^2) \end{bmatrix} = \begin{bmatrix} \sigma_u^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_u^2 & 0 & \dots & 0 \\ 0 & 0 & \sigma_u^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \sigma_u^2 \end{bmatrix}$$

Es decir, que se tiene la matriz de varianzas-covarianzas de u_i:

$$\text{Var-Cov}(\mathbf{u}) = \text{E}[(\mathbf{u} - \text{E}(\mathbf{u}))(\mathbf{u} - \text{E}(\mathbf{u}))'] = \text{E}[\mathbf{uu}'] = \begin{bmatrix} \sigma_u^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_u^2 & 0 & \dots & 0 \\ 0 & 0 & \sigma_u^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \sigma_u^2 \end{bmatrix} = \sigma_u^2 \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} = \sigma_u^2 \mathbf{I}_{n\text{x}n}$$

Por lo tanto, los dos supuestos, homocedasticidad y ausencia de autocorrelación, se pueden expresar como:

$$\text{Var-Cov}(\mathbf{u}) = \text{E}[(\mathbf{u} - \text{E}(\mathbf{u}))(\mathbf{u} - \text{E}(\mathbf{u}))'] = \text{E}[\mathbf{u}\mathbf{u}'] = \begin{bmatrix} \sigma_u^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_u^2 & 0 & \dots & 0 \\ 0 & 0 & \sigma_u^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \sigma_u^2 \end{bmatrix} = \sigma_u^2 \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} = \sigma_u^2 \mathbf{I}_n$$

En este mundo se conoce como

Perturbaciones esféricas

Cuando el término de perturbación del modelo presenta heteroscedasticidad o autocorrelación, o ambas cosas a la vez, se dice que es no esférico.

A veces se representa como:

$$Cov(u_i, u_j) = E(u_i u_j) = \begin{cases} \sigma_u^2 & i = j \\ 0 & i \neq j \end{cases}$$

Hipótesis de partida en matrices

- Supuesto de normalidad. En el cálculo matricial, la generalización es una distribución de probabilidad multivariante, es decir, para el conjunto de variables aleatorias $(u_1 ... u_n)$:

$$\mathbf{u}_{n \times 1} \sim \mathbf{N}(\mathbf{0}_{n \times 1}, \sigma_u^2 \mathbf{I}_n)$$

- Exogeneidad: $E(\mathbf{X}'\mathbf{u}) = \mathbf{0}$

Hipótesis de partida en matrices

En resumen, en la especificación del modelo de Regresión Lineal Múltiple (RLM) haciendo uso del álgebra matricial se tiene

$$\mathbf{Y} = \mathbf{XB} + \mathbf{u}$$

- Modelo completo: $E(\mathbf{u}) = \mathbf{0}$
- Exogeneidad: $E(\mathbf{X}'\mathbf{u}) = \mathbf{0}$
- Perturbaciones esféricas: $Cov(\mathbf{u}) = E(\mathbf{uu}') = \sigma_u^2 \mathbf{I}_n$ (homocedasticidad y no autocorrelación)
- Normalidad: $\mathbf{u}_{n\text{x}1} \sim \mathbf{N}(\mathbf{0}_{n\text{x}1}, \sigma_u^2 \mathbf{I}_n)$

Estimación MCO

Se debe estar en condiciones de construir matricialmente la SCR y derivarla respecto a $\hat{\mathbf{B}}$

El modelo estimado es $\mathbf{Y} = \mathbf{X}\hat{\mathbf{B}} + \hat{\mathbf{u}}$ y sin necesidad de supuestos se sabe que

$$\hat{\mathbf{u}} = \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \vdots \\ \hat{u}_n \end{bmatrix}$$

y por ende la SCR $= \sum \hat{u}_i^2 = \hat{u}_1^2 + \hat{u}_2^2 + ... + \hat{u}_n^2 = \hat{\mathbf{u}}'\hat{\mathbf{u}}$

Si se transpone un vector y se premultiplica por el vector original se obtiene la suma de cuadrados de los elementos del vector

$$\begin{aligned} \text{SCR} = \hat{\mathbf{u}}'\hat{\mathbf{u}} &= (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})'(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}) \\ &= (\mathbf{Y}' - \hat{\mathbf{B}}'\mathbf{X}')(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}) \\ &= \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\hat{\mathbf{B}} - \hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y} + \hat{\mathbf{B}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{B}} \end{aligned}$$

Se tiene que $(\hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y})' = \mathbf{Y}'\mathbf{X}\hat{\mathbf{B}}$ y que $\hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y}$ y $\mathbf{Y}'\mathbf{X}\hat{\mathbf{B}}$ son dos escalares iguales

Estimación MCO

$$\text{SCR} = \hat{\mathbf{u}}'\hat{\mathbf{u}} = \mathbf{Y}'\mathbf{Y} - 2\hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y} + \hat{\mathbf{B}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{B}}$$

Derivando

$$\begin{aligned} \frac{\partial \hat{\mathbf{u}}'\hat{\mathbf{u}}}{\partial \hat{\mathbf{B}}} &\Longrightarrow \frac{\partial \mathbf{Y}'\mathbf{Y}}{\partial \hat{\mathbf{B}}} = \mathbf{0} \\ &\Longrightarrow \frac{\partial 2\hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y}}{\partial \hat{\mathbf{B}}} = 2\mathbf{X}'\mathbf{Y} \\ &\Longrightarrow \frac{\partial \hat{\mathbf{B}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{B}}}{\partial \hat{\mathbf{B}}} = 2(\mathbf{X}'\mathbf{X})\hat{\mathbf{B}} \end{aligned}$$

$$\frac{\partial \hat{\mathbf{u}}'\hat{\mathbf{u}}}{\partial \hat{\mathbf{B}}} = -2\mathbf{X}'\mathbf{Y} + 2(\mathbf{X}'\mathbf{X})\hat{\mathbf{B}} = \mathbf{0}$$

$$(\mathbf{X}'\mathbf{X})\hat{\mathbf{B}} = \mathbf{X}'\mathbf{Y}\text{: Ecuaciones normales}$$

$$\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

Estimador insesgado de la varianza de **u**:

$$\widehat{\sigma}_u^2 = \frac{SCR}{n-k} = \frac{\widehat{\mathbf{u}}'\widehat{\mathbf{u}}}{n-k} = \frac{(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}})'(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}})}{n-k}$$

Hipótesis de partida en matrices: Supuesto adicional

$$(\mathbf{X'X})^{-1}(\mathbf{X'Y}) = \widehat{\mathbf{B}}$$

Para que $\widehat{\mathbf{B}}$ se pueda calcular se requiere que $(\mathbf{X'X})^{-1}$ exista, esto implica que:

- $|\mathbf{X'X}| \neq 0$
- que todas las filas y columnas de **X'X** sean linealmente independientes entre si

Todo lo anterior conlleva a que el rango de **X'X** debe ser cuatro en este ejemplo:

$$\rho(\mathbf{X'X}) = 4$$

Así pues, aparece una nueva hipótesis de partida: condición de no multicolinealidad perfecta, esto es que $\rho(\mathbf{X'X}) =$completo

Queda claro que con el álgebra de sumatorias es muy engorroso obtener $\widehat{\mathbf{B}}$, además, no deja ver el supuesto de multicolinealidad. Se usa el álgebra matricial
En resumen: No multicolinealidad perfecta: $\rho(\mathbf{X}_{n\mathrm{x}k}) = k < n$

Hipótesis de partida en matrices

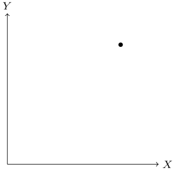
- Supondremos que no existe relación lineal exacta entre las variables explicativas del modelo, es decir, que no hay ninguna combinación lineal exacta entre las columnas de la matriz **X**. Es decir, no hay multicolinealidad perfecta entre las variables explicativas:

  $$\rho(\mathbf{X}_{n\times k}) = k\text{: La matrix } \mathbf{X} \text{ es de rango completo}$$

  Para esto tenga sentido se requiere que k sea menor que n. La intuición es que si voy a estimar k parámetros a partir de n observaciones, el número de incógnitas debe ser menor al número de observaciones

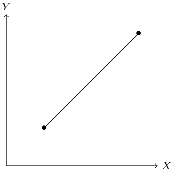
  En el caso de RLS se ve así ($k = 2$):

n=1, estima la recta con un punto



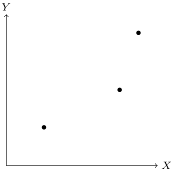
Infinitas rectas pasan por el punto

n=2, solución única



No hay error, no hay estadística

n=3



Habrá error de estimación

  Esta intuición la confirma los gdl del modelo $(n-k)$, en el RLS $(n-2)$. Sólo para $n \geq 3$ habrá gdl positivos

Hipótesis de partida en matrices

Muchas veces el supuesto de no multicolinealidad perfecta se escribe como:

$$\underbrace{\rho(\mathbf{X}_{n \times k})}_{\text{No multico. perfecta}} = \underbrace{k < n}_{\text{Sentido de la estimación}}$$

Ejemplo - Expresando el modelo en matrices

Se tiene un modelo de regresión con intercepto y dos variables explicativas, también se supone que $n = 10$, de la siguiente forma:

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + u_i$$

El modelo en forma matricial sería

$$\mathbf{Y}_{10\text{X}1} = \mathbf{X}_{10\text{X}3}\mathbf{B}_{3\text{X}1} + \mathbf{u}_{10\text{X}1}$$

Donde

$$\mathbf{Y}_{10\text{X}1} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_{10} \end{bmatrix} \quad \mathbf{X}_{10\text{X}3} = \begin{bmatrix} 1 & X_{1,2} & X_{1,3} \\ 1 & X_{2,2} & X_{2,3} \\ 1 & X_{3,2} & X_{3,3} \\ \vdots & \vdots & \vdots \\ 1 & X_{10,2} & X_{10,3} \end{bmatrix} \quad \mathbf{B}_{3\text{X}1} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \quad \mathbf{u}_{10\text{X}1} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_{10} \end{bmatrix}$$

Estimación

Ejemplo - Expresando el modelo en matrices

Ecuaciones normales

$$\frac{\partial \hat{\mathbf{u}}'\hat{\mathbf{u}}}{\partial \hat{\mathbf{B}}} = -2\mathbf{X}'\mathbf{Y} + 2(\mathbf{X}'\mathbf{X})\hat{\mathbf{B}} = \mathbf{0} \Longrightarrow (\mathbf{X}'\mathbf{X})\hat{\mathbf{B}} = \mathbf{X}'\mathbf{Y}$$

Donde

$$\mathbf{X}'\mathbf{X}_{3X3} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_{12} & X_{22} & \cdots & X_{102} \\ X_{13} & X_{23} & \cdots & X_{103} \end{bmatrix} \begin{bmatrix} 1 & X_{12} & X_{13} \\ 1 & X_{22} & X_{23} \\ 1 & X_{32} & X_{33} \\ \vdots & \vdots & \vdots \\ 1 & X_{102} & X_{103} \end{bmatrix} = \begin{bmatrix} n & \sum X_{i2} & \sum X_{i3} \\ \sum X_{i2} & \sum X_{i2}^2 & \sum X_{i2}X_{i3} \\ \sum X_{i3} & \sum X_{i2}X_{i3} & \sum X_{i3}^2 \end{bmatrix}$$

$$\mathbf{X}'\mathbf{Y}_{3X1} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_{12} & X_{22} & \cdots & X_{102} \\ X_{13} & X_{23} & \cdots & X_{103} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_{10} \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_{i2}Y_i \\ \sum X_{i3}Y_i \end{bmatrix}$$

$$(\mathbf{X}'\mathbf{X})\widehat{\mathbf{B}} = \mathbf{X}'\mathbf{Y} = \begin{bmatrix} n & \sum X_{i2} & \sum X_{i3} \\ \sum X_{i2} & \sum X_{i2}^2 & \sum X_{i2}X_{i3} \\ \sum X_{i3} & \sum X_{i2}X_{i3} & \sum X_{i3}^2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_{i2}Y_i \\ \sum X_{i3}Y_i \end{bmatrix}$$

Estimación MCO $\Longrightarrow \hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$

Estimación MCO

Ejemplo - Stata

Se tiene datos semanales por un año (52 semanas) de una cadena de hamburguesas sobre los ingresos (tr, en miles de dólares), el precio promedio (p, en dólares) y el gasto en publicidad (adv, en miles de dólares). El objetivo es determinar cómo los ingresos de esta cadena de hamburguesas responde ante el precio y el gasto en publicidad

```
* Cargando los datos desde excel
clear
import excel "C:\...\Table 9.1 GCJ.xlsx", ///
sheet("Hoja1") firstrow

* Corriendo el modelo
reg tr p adv
```

```
      Source |       SS       df       MS              Number of obs =      52
-------------+------------------------------           F(  2,    49) =  159.83
       Model |  11776.1839     2  5888.09194           Prob > F      =  0.0000
    Residual |  1805.16843    49   36.840172           R-squared     =  0.8671
-------------+------------------------------           Adj R-squared =  0.8617
       Total |  13581.3523    51  266.301026           Root MSE      =  6.0696

------------------------------------------------------------------------------
          tr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           p |   -6.64193   3.191193    -2.08   0.043    -13.05487   -.2289878
         adv |   2.984299   .1669361    17.88   0.000     2.648828     3.31977
       _cons |   104.7855   6.482719    16.16   0.000       91.758     117.813
------------------------------------------------------------------------------
```

```
* Calculando el modelo vía matrices
* Primero generamos la constante = 1
gen con = 1

* Ahora construimos la matriz X
mkmat con p adv, matrix(X)
matrix list X
```

```
X[52,3]
    con     p   adv
r1    1  1.92  12.4
r2    1  2.15   9.9
r3    1  1.67   2.4
r4    1  1.68  13.8
```

```
* Matriz Y
mkmat tr, matrix(Y)
matrix list Y
```

```
Y[52,1]
       tr
r1  123.1
r2  124.3
r3   89.3
r4  141.3
```

Estimación MCO

Ejemplo - Stata

```
      Source |       SS       df       MS              Number of obs =      52
-------------+------------------------------           F(  2,    49) =  159.83
       Model |  11776.1839     2  5888.09194           Prob > F      =  0.0000
    Residual |  1805.16843    49   36.840172           R-squared     =  0.8671
-------------+------------------------------           Adj R-squared =  0.8617
       Total |  13581.3523    51  266.301026           Root MSE      =  6.0696

------------------------------------------------------------------------------
          tr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           p |   -6.64193   3.191193    -2.08   0.043    -13.05487   -.2289878
         adv |   2.984299   .1669361    17.88   0.000     2.648828     3.31977
       _cons |   104.7855   6.482719    16.16   0.000       91.758     117.813
------------------------------------------------------------------------------
```

```
* Sabemos que B = (X'X)^-1(X'Y), entonces construimos
* cada parte de este vector

* Construyendo (X'X)
matrix XtX = X'*X
matrix list XtX
```

```
symmetric XtX[3,3]
           con          p        adv
con         52
  p     104.09   212.0153
adv      502.4   1012.755    6189.66
```

```
* Construyendo (X'Y)
matrix XtY = X'*Y
matrix list XtY
```

```
XtY[3,1]
           tr
con    6256.8
  p 12521.297
adv  64389.39
```

```
* Finalmente construimos B
matrix B = inv(XtX)*(XtY)
matrix list B
```

```
B[3,1]
             tr
con   104.78551
  p  -6.6419301
adv    2.984299
```

```
* Calculando Var(u) = SCR/(n-k)
                    = u'u/(n-k)
                    = (Y-XB)'(Y-XB)/(n-k)
matrix u_e = Y - X*B
matrix list u_e
```

```
U_e[52,1]
            tr
r1  -5.9383149
r2   4.2500764
r3  -11.555808
r4   6.4896034
```

```
matrix u_etu_e = u_e'*u_e
matrix list u_etu_e
```

```
symmetric U_etU_e[1,1]
           tr
tr  1805.1684
```

```
matrix varu = u_etu_e/(52-3)
matrix list varu
```

```
symmetric varU[1,1]
          tr
tr 36.840172
```

- Linealidad
  $\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ es lineal en Y y en u.
  A partir de la expresión de $\hat{\mathbf{B}}$ puede verse que son una combinación lineal del vector $\mathbf{Y}$
  Así mismo:

$$\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{XB}+\mathbf{u}) =$$
$$= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{XB} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} = \mathbf{B} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$$

  Los estimadores MCO quedan expresados como una combinación lineal del término de perturbación.

- Insesgadez
  Partiendo de la expresión aleatoria

$$E[\hat{\mathbf{B}}] = E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}] = E[\mathbf{B} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}] =$$
$$= E[\mathbf{B}] + E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}] = \mathbf{B} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\mathbf{u}] = \mathbf{B}$$

  $\hat{\mathbf{B}}$ es un estimador insesgado

Propiedades de los estimadores MCO

- Inclusión de una variable (o más) irrelevantes en un modelo de regresión o sobreespecificación del modelo

  Suponga que se especifica un modelo como:

  $$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + u_i$$

  u_i satisface los supuestos estándar, pero supongamos que X_{i4} no tiene ningún efecto sobre Y_i una vez que X_{i2} y X_{i3} se han controlado, lo que significa que $\beta_4 = 0$

  Como no se sabe que $\beta_4 = 0$, se estima el modelo con X_{i4}:

  $$Y_i = \widehat{\beta}_1 + \widehat{\beta}_2 X_{i2} + \widehat{\beta}_3 X_{i3} + \widehat{\beta}_4 X_{i4} + \widehat{u}_i$$

  ¿Qué efecto tiene incluir un regresor irrelevante?:

  - incluir una o más variables irrelevantes en un modelo de RLM, no afecta el insesgamiento de los estimadores MCO
  - incluir variables irrelevantes puede tener efectos indeseables en la varianza de los estimadores MCO

- Sesgo de la variable omitida

  Suponga ahora un modelo con dos variables explicativas (que cumple los supuestos estándar) y por error en la especificación (desconocimiento o falta de datos) se omite X_{i3} y el modelo se estima como:

  $$Y_i = \widetilde{\beta}_1 + \widetilde{\beta}_2 X_{i2} + \widetilde{u}_i$$

  Suponga que X_{i2} y X_{i3} están correlacionadas. Ya que X_{i3} se va para el término de error y dada la correlación con X_{i2}, el error y X_{i2} estarán correlacionados y, por tanto, nuestros estimadores MCO estarán sesgados

Propiedades de los estimadores MCO

- Mínima varianza (Eficiencia)
  El Teorema de Gauss-Markov garantiza que (bajo los supuestos estándar) dentro todos los estimadores lineales e insesgados, los estimadores MCO son los más eficientes, es decir, tienen mínima varianza (esto es, su dispersión más baja alrededor de su valor esperado)

  Calculamos la matriz de varianzas-covarianzas de $\hat{\mathbf{B}}$

$$Var - Cov(\hat{\mathbf{B}}) = E[(\hat{\mathbf{B}} - E(\hat{\mathbf{B}}))(\hat{\mathbf{B}} - E(\hat{\mathbf{B}}))'] = E[(\hat{\mathbf{B}} - \mathbf{B})(\hat{\mathbf{B}} - \mathbf{B})']$$

  De la expresión aleatoria tenemos que

$$\begin{aligned} \hat{\mathbf{B}} &= \mathbf{B} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \\ \hat{\mathbf{B}} - \mathbf{B} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \\ (\hat{\mathbf{B}} - \mathbf{B})' &= \mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

  Entonces

$$\begin{aligned} Var - Cov(\hat{\mathbf{B}}) &= E[(\hat{\mathbf{B}} - \mathbf{B})(\hat{\mathbf{B}} - \mathbf{B})'] \\ &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{u}\mathbf{u}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

  Por el supuesto de perturbaciones esféricas $E(\mathbf{u}\mathbf{u}') = \sigma_u^2\mathbf{I}_n$

$$Cov(\hat{\mathbf{B}}) = \sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}\underbrace{\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}}_{\mathbf{I}_k}$$

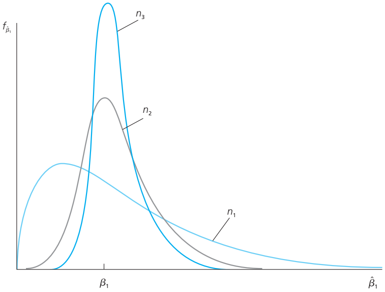
$$Cov(\hat{\mathbf{B}}) = \sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}$$

Propiedades de los estimadores MCO

$$Var{-}Cov(\hat{\mathbf{B}}) = \sigma_u^2(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} var(\beta_1) & Cov(\beta_1,\beta_2) & Cov(\beta_1,\beta_3) & \cdots & Cov(\beta_1,\beta_k) \\ Cov(\beta_2,\beta_1) & var(\beta_2) & Cov(\beta_2,\beta_3) & \cdots & Cov(\beta_2,\beta_k) \\ Cov(\beta_3,\beta_1) & Cov(\beta_3,\beta_2) & var(\beta_3) & \cdots & Cov(\beta_3,\beta_k) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ Cov(\beta_k,\beta_1) & Cov(\beta_k,\beta_2) & Cov(\beta_k,\beta_3) & \cdots & Var(\beta_k) \end{bmatrix}$$

Propiedades de los estimadores MCO

- Consistencia
  - Aunque la insesgadez de los estimadores es importante, no siempre puede lograrse: por ejemplo $\widehat{\sigma}_u^2$ es sesgado
  - Aunque no todos los estimadores útiles son insesgados, casi todos los economistas están de acuerdo en que la **consistencia** es un requisito mínimo para un estimador
  - $\widehat{\beta}_j$ es un estimador de β_j y como es insesgado la distribución de probabilidad de $\widehat{\beta}_j$ tiene una media de β_j. Como estimador consistente, entonces a medida que el tamaño de la muestra aumenta, la distribución de $\widehat{\beta}_j$ se estrechará cada vez más entorno a β_j
  - Esto significa que si es posible recolectar tantos datos como se desee, entonces puede hacerse que el estimador esté arbitrariamente cerca de β_j



Propiedades de los estimadores MCO

- Normalidad asintótica
  - Se sabe que la normalidad no juega ningún papel en la insesgadez de los estimadores MCO y tampoco afecta las conclusiones de que MCO es el mejor estimador lineal insesgado bajo los supuestos estándar
  - Pero la inferencia exacta basada en los estadísticos t y F requiere el supuesto que los errores se distribuyen normal
  - Aunque las Y_i no provienen de una distribución normal, puede emplearse el teorema central del límite para concluir que los estimadores de MCO satisfacen la **normalidad asintótica**, lo cual significa que están distribuidos de manera aproximadamente normal cuando se tienen muestras de tamaño suficientemente grande
  - Este teorema indica que, sin importar la distribución de la población de u, los estimadores de los MCO, cuando se estandarizan de manera apropiada, tienen distribuciones normales estándar aproximadas

$$\hat{\mathbf{B}} \sim N[\mathbf{B}, \sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}]$$

Ejemplo - Stata

Source	SS	df	MS		
				Number of obs =	52
				F(2, 49) =	159.83
Model	11776.1839	2	5888.09194	Prob > F =	0.0000
Residual	1805.16843	49	36.840172	R-squared =	0.8671
				Adj R-squared =	0.8617
Total	13581.3523	51	266.301026	Root MSE =	6.0696

tr	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
p	-6.64193	3.191193	-2.08	0.043	-13.05487	-.2289878
adv	2.984299	.1669361	17.88	0.000	2.648828	3.31977
_cons	104.7855	6.482719	16.16	0.000	91.758	117.813

```
* Calculando la matriz de var-cov de B
matrix varcovB = varu*inv(XtX)
matrix list varcovB
```

symmetric varcovB[3,3]

	con	p	adv
con	42.025645		
p	-19.863122	10.183712	
adv	-.16110867	-.05402121	.02786767

```
* En esta matriz var-cov de B la diagonal principal
* tiene las varianzas de los coeficientes
* mientras que por encima o debajo de la diagonal
* están las covarianzas
* Calculando los errores estándar

scalar es_b1 = sqrt(varcovB[1,1])
dis es_b1
6.482719

scalar es_b2 = sqrt(varcovB[2,2])
dis es_b2
3.1911929

scalar es_b3 = sqrt(varcovB[3,3])
dis es_b3
.16693614
```

Para saber cuál es el estimador insesgado de σ_u^2 se requiere calcular la siguiente esperanza

$$E(SCR) = E(\sum \hat{u}_i^2)$$

Sabemos que $\sum \hat{u}_i^2 = \hat{\mathbf{u}}'\hat{\mathbf{u}}$, ahora se tiene que

$$\hat{\mathbf{u}} = \mathbf{Y} - \mathbf{X}\hat{\mathbf{B}} = \mathbf{Y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{Y}$$

Sea $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, se dice que tiene buenas propiedades:

- Es simétrica: $\mathbf{M}' = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{M}$
- Es idempotente: $$\begin{aligned}\mathbf{MM} &= (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') \\ &= \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{X}\underbrace{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}}_{\mathbf{I}}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ &= \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ &= \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ &= \mathbf{M}\end{aligned}$$
- Es ortogonal a $\mathbf{X}$: $\mathbf{MX} = \mathbf{0}_{n \times k}$
$$\mathbf{MX} = (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{X} = \mathbf{X} - \mathbf{X}\underbrace{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}}_{\mathbf{I}} = \mathbf{X} - \mathbf{X} = \mathbf{0}$$
- Es semidefinida positiva: puesto que $\mathbf{M}_{n \times n}$ es simétrica, si $\mathbf{W}_{n \times 1}$ es un vector, entonces
$$\mathbf{W}'\mathbf{M}\mathbf{W} \geq \mathbf{0}$$

La estimación insesgada de σ_u^2

Volviendo a $\hat{\mathbf{u}}$

$$\hat{\mathbf{u}} = \mathbf{MY} = \mathbf{M}(\mathbf{XB} + \mathbf{u}) = \mathbf{MXB} + \mathbf{Mu} = \mathbf{Mu}$$

$$\Longrightarrow \hat{\mathbf{u}} = \mathbf{Mu} \Longrightarrow \hat{\mathbf{u}}' = \mathbf{u}'\mathbf{M}$$

$$\Longrightarrow \hat{\mathbf{u}}'\hat{\mathbf{u}} = \mathbf{u}'\mathbf{MMu} = \mathbf{u}'\mathbf{Mu}$$

Ahora se esta en condiciones de encontrar $E(\hat{\mathbf{u}}'\hat{\mathbf{u}}) = E(\mathbf{u}'\mathbf{Mu})$

Para ello se requiere tener presente las siguientes propiedades de la traza de una matriz

- La traza de una matriz cuadrada es la suma de los elementos de la diagonal principal. Sea $\mathbf{A}_{n\text{x}n}$, entonces tr($\mathbf{A}$)=$\sum_{i=1}^{n} a_{ij}$
- La traza de un escalar es el mismo escalar
- tr$(\alpha A) = \alpha tr(\mathbf{A})$
- tr($\mathbf{AB}$)=tr($\mathbf{BA}$) y tr($\mathbf{ABC}$)=tr($\mathbf{CAB}$)=tr($\mathbf{BCA}$)
- $tr(\mathbf{I}_n) = n$

La estimación insesgada de σ_u^2

Ahora, la SCR es un escalar, con lo cual

$$SCR = \mathbf{u}'\mathbf{M}\mathbf{u} = \text{tr}(\mathbf{u}'\mathbf{M}\mathbf{u})$$

$$\begin{aligned}\Longrightarrow E(SCR) &= E(\text{tr}(\mathbf{u}'\mathbf{M}\mathbf{u}))\\ &= E(\text{tr}(\mathbf{u}\mathbf{u}'\mathbf{M}))\\ &= \text{tr}(E(\mathbf{u}\mathbf{u}')\mathbf{M})\\ &= \text{tr}(\sigma_u^2\mathbf{I}\mathbf{M})\\ &= \sigma_u^2\ \text{tr}(\mathbf{M})\end{aligned}$$

$$\Longrightarrow \text{tr}(\mathbf{M}) = \text{tr}(\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = \text{tr}(\mathbf{I}_n)\ \text{-}\ \text{tr}(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')$$

$$\Longrightarrow \text{tr}(\mathbf{M}) = \text{tr}(\mathbf{I}_n)\ \text{-}\ \text{tr}(\mathbf{I}_k)$$

$$\Longrightarrow \text{tr}(\mathbf{M}) = N-k$$

$$\Longrightarrow E(SCR) = \sigma_u^2\ \text{tr}(\mathbf{M}) = \sigma_u^2(N-k)$$

De lo anterior se deduce que un estimador insesgado de σ_u^2 será

$$\hat{\sigma}_u^2 = \frac{SCR}{N-k} = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{N-k}$$

Ya que $E(\hat{\sigma}_u^2) = \frac{E(SCR)}{N-k} = \frac{\sigma_u^2(N-k)}{N-k} = \sigma_u^2$

Ejercicio: Demostrar que $SCR = \hat{\mathbf{u}}'\hat{\mathbf{u}} = \mathbf{Y}'\mathbf{Y} - \hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y}$ o $\hat{\mathbf{u}}'\hat{\mathbf{u}} = \mathbf{Y}'\mathbf{Y} - \hat{\mathbf{Y}}'\hat{\mathbf{Y}}$

El coeficiente de determinación R^2

Se parte de la identidad del análisis de varianza

$$\text{SCT} = \text{SEC} + \text{SCR}$$

que en el mundo de las sumatorias es

$$\sum(Y_i - \bar{Y})^2 = \sum(\hat{Y}_i - \bar{Y})^2 + \sum \hat{u}_i^2$$

En el mundo matricial sería

$$\begin{aligned}\sum(Y_i - \bar{Y})^2 = \sum\left(Y_i^2 + \bar{Y}^2 - 2\bar{Y}Y_i\right) = \sum Y_i^2 + N\bar{Y}^2 - 2\bar{Y}\sum Y_i = \\ = \mathbf{Y}'\mathbf{Y} + N\bar{Y}^2 - 2N\bar{Y}^2 = \mathbf{Y}'\mathbf{Y} - N\bar{Y}^2\end{aligned}$$

$$\begin{aligned}\sum(\hat{Y}_i - \bar{\hat{Y}})^2 = \sum(\hat{Y}_i - \bar{Y})^2 = \sum\left(\hat{Y}_i^2 + \bar{Y}^2 - 2\bar{Y}\hat{Y}_i\right) = \sum \hat{Y}_i^2 + N\bar{Y}^2 - 2\bar{Y}\sum \hat{Y}_i = \\ = \hat{\mathbf{Y}}'\hat{\mathbf{Y}} + N\bar{Y}^2 - 2N\bar{Y}^2 = \hat{\mathbf{Y}}'\hat{\mathbf{Y}} - N\bar{Y}^2\end{aligned}$$

Así,

$$\mathbf{Y}'\mathbf{Y} - N\bar{Y}^2 = \hat{\mathbf{Y}}'\hat{\mathbf{Y}} - N\bar{Y}^2 + \hat{\mathbf{u}}'\hat{\mathbf{u}}$$

El coeficiente de determinación R^2

Coeficiente de determinación o medida de bondad de ajuste del modelo

$$R^2 = \frac{SEC}{SCT} = 1 - \frac{SCR}{SCT} = 1 - \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{\mathbf{Y}'\mathbf{Y} - N\bar{Y}^2}$$

Algunas consideraciones

- Si $R^2 = 0 \Longrightarrow SCR = SCT$ Quiere decir que los X no agregan explicación al modelo
- Si $R^2 = 1 \Longrightarrow SEC = SCT$ No hay residuos. Todos los puntos están sobre el plano de regresión, esto sucede con identidades o tautologías

Recuerde la interpretación del R^2 o coeficiente de determinación

El coeficiente de determinación R^2

- Si el modelo no tiene intercepto el R^2 pierde su significado y puede dar negativo. No es comparable con el R^2 de modelos con intercepto
- Si son datos de series de tiempo (crecientes o decrecientes) que se mueven en el mismo sentido, el R^2 tiende a 1. En datos de corte transversal el R^2 es bajo
- Una limitación del R^2 en el modelo de RLM es que al incluir regresores el R^2 aumenta. Si el modelo tiene k regresores y se le agregan s adicionales, el R^2 de este segundo modelo es mayor que el del primero

  En este contexto si se usa el R^2 para comparar dos modelos de distinto número de regresores, el que tenga menos estará en desventaja. Por lo tanto, el R^2 debe ser ajustado por los grados de libertad asociados a las sumas de cuadrados. Sea $\overline{R}^2$ el coeficiente de determinación ajustado:

  $$\overline{R}^2 = 1 - \frac{SCR/(N-k)}{SCT/(N-1)} = 1 - \frac{N-1}{N-k}\frac{SCR}{SCT}$$

  Ya que $\frac{SCR}{SCT} = 1 - R^2$, entonces

  $$\overline{R}^2 = 1 - \frac{N-1}{N-k}(1 - R^2)$$

  El $\overline{R}^2$ no esta acotado entre cero y uno, por lo que puede dar negativo (en ese caso se asumiría como cero)