Estadística para la Economía y la Gestión IN 3401

3 de junio de 2010

De la sección anterior teníamos que el error estimado era:

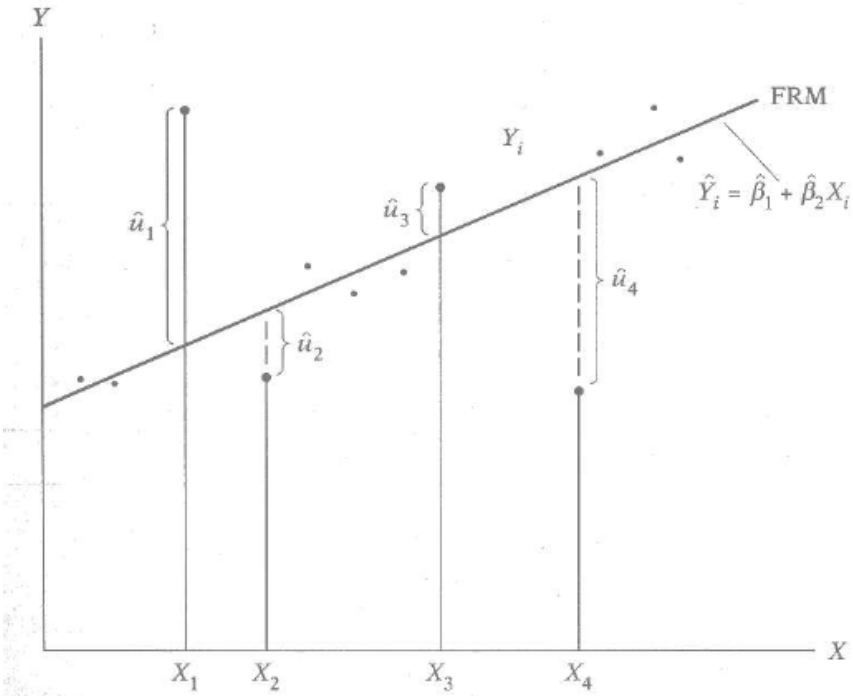
$$\begin{aligned} \hat{\epsilon}_i &= y_i - \hat{y}_i \\ &= y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i \end{aligned}$$

es decir, los residuos son simplemente la diferencia entre los valores verdaderos y estimados de y_i.

Si queremos que la función de regresión muestral sea lo más cercana posible a la poblacional, debemos tratar de escoger los coeficientes de regresión (los βs) de forma tal que los errores sean lo más pequeños posible.

De acuerdo a esto un criterio para escoger la función de regresión muestral podría ser minimizar la suma de los los errores: $\sum \hat{\epsilon}_i = \sum (y_i - \hat{y}_i)$, sin embargo este criterio no es muy bueno.

De acuerdo al gráfico, existe una gran diferencia en la magnitud de los errores, sin embargo en la suma de los errores todos reciben el mismo peso...



... es posible que la suma de los errores sea muy pequeña cercana a 0, incluso cuando la dispersión de los errores en torno a la función de regresión muestral es alta.

Este problema puede ser solucionado al considerar la suma de los errores al cuadrado como criterio a minimizar, en este caso los errores más lejos reciben un mayor peso:

$$\begin{aligned} \sum \hat{\epsilon}_i^2 &= \sum (y_i - \hat{y}_i)^2 \\ &= \sum (y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2 \end{aligned}$$

El **Método de Mínimos Cuadrados Ordinarios (MCO)** escoge $\hat{\beta}_1$ y $\hat{\beta}_2$ de forma tal que para una muestra dada, $\sum \hat{\epsilon}_i^2$ sea lo más pequeño posible.

Entonces el problema que este método propone resolver es el siguiente:

$$\min_{\hat{\beta}_1,\hat{\beta}_2} \sum (y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2$$

las condiciones de primer orden de este problema son:

$$\begin{aligned}
\frac{\partial \sum \hat{\epsilon}_i^2}{\partial \hat{\beta}_1} &= -2\sum (y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i) = -2\sum \hat{\epsilon}_i = 0 \\
\frac{\partial \sum \hat{\epsilon}_i^2}{\partial \hat{\beta}_2} &= -2\sum (y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i) X_i = -2\sum \hat{\epsilon}_i X_i = 0
\end{aligned}$$

Reordenando obtenemos las **ecuaciones normales**:

$$\begin{aligned} \sum y_i &= n\hat{\beta}_1 + \hat{\beta}_2 \sum X_i \\ \sum y_i X_i &= \hat{\beta}_1 \sum X_i + \hat{\beta}_2 \sum X_i^2 \end{aligned}$$

Debemos resolver un sistema con dos ecuaciones y dos incógnitas. De la ecuación anterior podemos despejar $\hat{\beta}_1$:

$$\hat{\beta}_1 = \frac{\sum y_i - \hat{\beta}_2 \sum X_i}{n}$$

reemplazando obtenemos...

obtenemos...

$$\hat{\beta}_2 = \frac{n\sum y_i X_i - \sum X_i \sum y_i}{n\sum X_i^2 - (\sum X_i)^2}$$

El que puede ser escrito de la siguiente forma:

$$\hat{\beta}_2 = \frac{\sum \widetilde{x}_i \widetilde{y}_i}{\sum \widetilde{x}_i^2}$$

donde $\widetilde{x}_i = X_i - \bar{X}$ e $\widetilde{y}_i = y_i - \bar{y}$, con $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ e $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

Ejemplo: Tenemos los siguientes datos de portales de internet, con los cuales queremos ver el impacto promedio del número de visitas en el valor de la empresa:

	VEMPRESA	VISITAS	$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$	$\hat{u}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i$
AOL	134844	50	98976.5	35867.5
Yahoo	55526	38	70403.7	-14877.7
Lycos	5533	28	46593.1	-41060.1
CNet	4067	8	-1028.3	5095.3
Juno Web	611	8	-1028.3	1639.3
NBC Internet	4450	16	18020.3	-13570.3
Earthlink	2195	5	-8171.5	10366.5
El Sitio	1225	2	-15314.7	16539.7
PROMEDIO	26056.4	19.4	26056.4	0

Utilizando los datos obtenemos:

$$\begin{aligned} \sum (X_i - \bar{X}_i)^2 &= 2{,}137,9 \\ \sum (y_i - \bar{y}_i)(X_i - \bar{X}_i) &= 5{,}090{,}422,9 \\ \hat{\beta}_2 &= \frac{5{,}090{,}422,9}{2{,}137,9} = 2{,}381,1 \\ \hat{\beta}_1 &= 26{,}056,4 - 2{,}381,1 \cdot 19,4 = 20{,}076,8 \end{aligned}$$

En el análisis de regresión nuestro objetivo no es sólo obtener los valores de $\hat{\beta}_1$ y $\hat{\beta}_2$ sino también hacer inferencia sobre los verdaderos β_1 y β_2:

- Nos interesa saber que tan cerca están $\hat{\beta}_1$ y $\hat{\beta}_2$ de sus contraparte poblacional
- O, que tan cerca esta $\hat{y}_i$ de la verdadera $E(y|X_i)$.

Los supuestos hechos sobre X_i y ϵ_i son fundamentales para lograr una interpretación válida de los valores estimados de la regresión. Mientras no se especifique la forma como se generan X_i y ϵ_i, no hay forma de hacer inferencia estadística sobre y_i ni sobre β_1 y β_2.

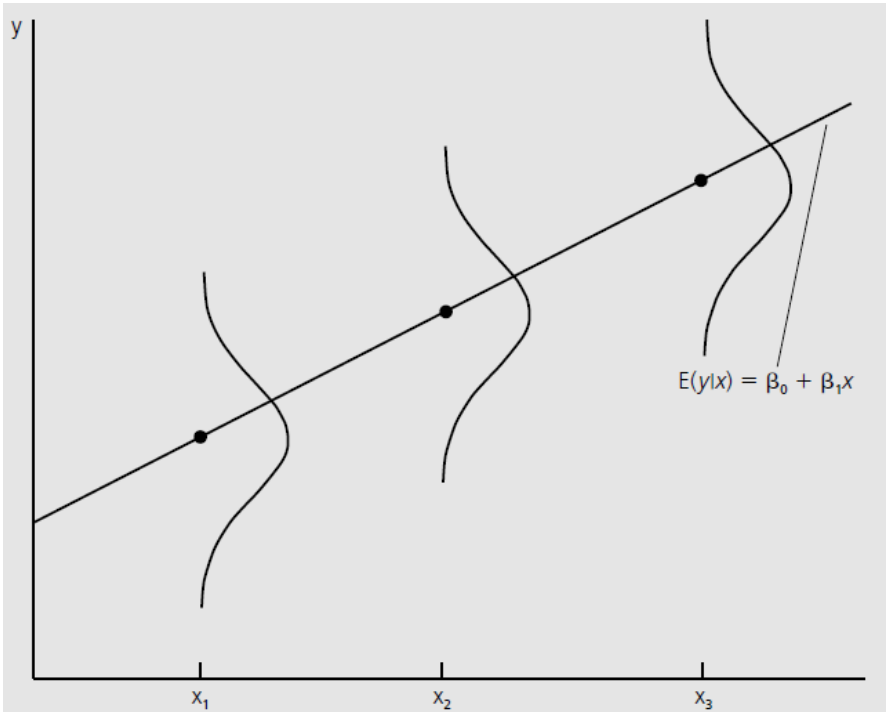
- **Supuesto 1:** Modelo de regresión lineal, el modelo de regresión es lineal en parámetros:

$$y_i = \beta_1 + \beta_2 X_i$$

- **Supuesto 2:** Los valores de X son fijos, X se supone no estocástica. Esto implica que el análisis de regresión es un análisis de regresión condicional, condicionado a los valores dados del regresor X.
- **Supuesto 3:** El valor medio del error ϵ_i es igual a cero. Dado el valor de X, el valor esperado del término de error ϵ_i es cero:

$$E(\epsilon_i|X) = 0$$

Lo que nos dice el supuesto 3 es que los factores que no están considerados en el modelo y que están representados a través de ϵ_i, no afectan sistemáticamente el valor de la media de y.

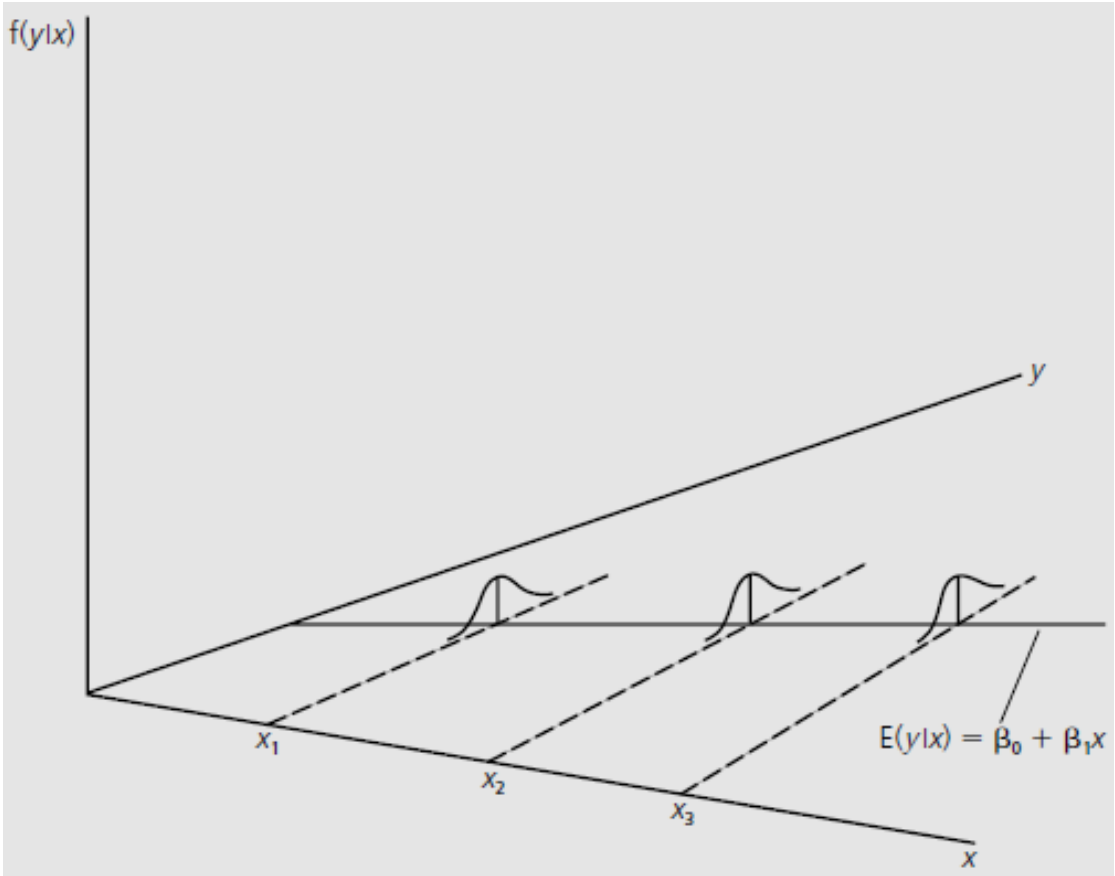


Es decir, los valores positivos de ϵ_i se cancelan con los valores negativos de ϵ_i. De esta forma, el efecto promedio de ϵ_i sobre y es cero.

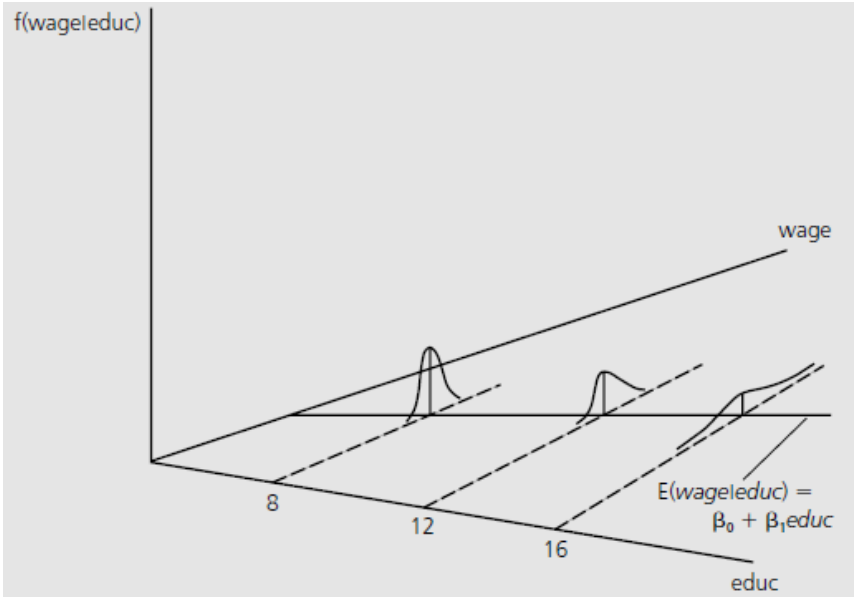
Supuesto 4: Homocedasticidad o igual varianza de ϵ_i. Dado el valor de X, la varianza de ϵ_i es la misma para todas las observaciones:

$$\begin{aligned} var(\epsilon_i|X_i) &= E[\epsilon_i - E(\epsilon_i)|X_i]^2 \\ &= E(\epsilon_i^2|X_i) \quad \text{por supuesto 3} \\ &= \sigma^2 \end{aligned}$$

El supuesto 4 implica la variación alrededor de la recta de regresión es la misma para todos los valores de X.



Por el contrario, el el siguiente grafo observamos el caso cuando la varianza del término de error varia para cada X_i, en este caso particular la varianza del error aumenta en la medida que la eduación auementa.



Esto se conoce como Heterocedasticidad o varianza desigual, lo que se expresa de la siguiente manera:

$$var(\epsilon_i|X_i) = \sigma_i^2$$

Supuesto 5: No existe autocorrelación entre los errores. Dado dos valores de X, X_i y X_j , con $i \neq j$, la correlación entre ϵ_i y ϵ_j es cero:

$$\begin{aligned} cov(\epsilon_i, \epsilon_j | X_i, X_j) &= E\{\epsilon_i - E(\epsilon_i)|X_i\}\{\epsilon_j - E(\epsilon_j)|X_j\} \\ &= E(\epsilon_i|X_i)E)(\epsilon_j|X_j) \\ &= 0 \end{aligned}$$

Si en la Función de regresión poblacional $Y_i = \beta_1 + \beta_2 X_i + \epsilon_i$, ϵ_i esta correlacionado con ϵ_j , entonces Y_i no depende solamente de X_i sino también de ϵ_j .

Al imponer le supuesto 5 estamos diciendo que solo se considerará el efecto sistemático de X_i sobre Y_i sin preocuparse de otros factores que pueden estar afectando a Y, como la correlación entre los errores.

Supuesto 6: La covarianza entre ϵ_i y X_i es cero $E(\epsilon_i X_i) = 0$:

$$\begin{aligned} cov(\epsilon_i, X_i) &= E[\epsilon_i - E(\epsilon_i)]E[X_i - E(X_i)] \\ &= E[\epsilon_i X_i - E(X_i)] \\ &= E(\epsilon_i X_i) - E(\epsilon_i)E(X_i) \\ &= E(\epsilon_i X_i) \\ &= 0 \end{aligned}$$

Se supone que X y ϵ tienen una influencia separada sobre Y (determinística y estocástica, respectivamente), ahora si X y ϵu están correlacionadas, no es posible determinar los efectos individuales sobre Y.

Supuesto 7: El número de observaciones n debe ser mayor que el número de parámetros por estimar. El número de observaciones tiene que ser mayor que el número de variables explicativas, de otra forma no se puede resolver el sistema de ecuaciones.

Supuesto 8: Variabilidad en los valores de X. No todos los valores de X en una muestra deben ser iguales, $var(X)$ debe ser un número finito positivo. Si las X son las mismas $\Rightarrow X_i = X$, de esta forma ni β_2 ni β_1 pueden ser estimados.

Supuesto 9: El modelo de regresión esta correctamente especificado.

La medida que utilizaremos para medir la precisión del estimador es el error estándar, que es la desviación estándar de la distribución muestral del estimador, la que a su vez es la distribución del conjunto de valores del estimador obtenidos de todas las muestras posibles de igual tamaño de una población dada.

Recordemos el estimador MCO de β_2:

$$\hat{\beta}_2 = \frac{\sum \widetilde{x}_i \widetilde{y}_i}{\sum \widetilde{x}_i^2}$$

donde $\widetilde{y}_i = \beta_2 \widetilde{x}_i + \epsilon_i$ (modelo poblacional en desviaciones con respecto a la media).

De esta forma reemplazando y_i en el estimador de β_2:

$$\begin{aligned}\hat{\beta}_2 &= \frac{\sum \widetilde{x}_i(\beta_2 \widetilde{x}_i + \epsilon_i)}{\sum \widetilde{x}_i^2} \\ &= \beta_2 \frac{\sum \widetilde{x}_i^2}{\sum \widetilde{x}_i^2} + \frac{\sum \widetilde{x}_i \epsilon_i}{\sum \widetilde{x}_i^2} \\ &= \beta_2 + \frac{\sum \widetilde{x}_i \epsilon_i}{\sum \widetilde{x}_i^2}\end{aligned}$$

Aplicando valor esperado a la expresión anterior:

$$\begin{aligned} E(\hat{\beta}_2) &= \beta_2 + E\left(\frac{\sum \epsilon_i \widetilde{x}_i}{\sum \widetilde{x}_i^2}\right) \\ &= \beta_2 + \left(\frac{\sum E(\epsilon_i) \widetilde{x}_i}{\sum \widetilde{x}_i^2}\right) \\ &= \beta_2 \end{aligned}$$

La ecuación anterior nos dice que en valor esperado el estimador MCO de β_2 es igual a su verdadero valor. Esta propiedad del estimador MCO se conoce como **insesgamiento**.

Ahora procedamos a calcular la varianza de el estimador MCO de β_2:

$$\begin{aligned} var(\hat{\beta}_2) &= E[\hat{\beta}_2 - E(\hat{\beta}_2)]^2 \\ &= E(\hat{\beta}_2 - \beta_2)^2 \\ &= E\left(\frac{[\sum \widetilde{x}_i \epsilon_i]^2}{[\sum \widetilde{x}_i^2]^2}\right) \end{aligned}$$

Por supuesto 4, $E(\epsilon_i^2) = \sigma^2$ y por supuesto 6 $E(\epsilon_i \epsilon_j) = 0$, esto implica que:

$$var(\hat{\beta}_2) = \frac{\sigma^2}{\sum \widetilde{x}_i^2}$$

Ahora debemos estimar el parámetro poblacional σ^2, como este corresponde al valor esperado de ϵ_i^2 y $\hat{\epsilon}_i$ es una estimación de ϵ_i, por analogía:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{\epsilon}_i^2}{n}$$

pareciera ser un estimador razonable. Pero los errores de MCO, están estimados imperfectamente si los comparamos con los errores poblacionales, ya que dependen de una estimación de β_1 y β_2.

Partiendo del Regresión poblacional expresado en desviaciones con respecto a la media:

$$\widetilde{y}_i = \beta_2 \widetilde{x}_i + (\epsilon_i - \bar{\epsilon})$$

y recordando también que:

$$\hat{\epsilon}_i = \widetilde{y}_i - \hat{\beta}_2 \widetilde{x}_i$$

sustituyendo se obtiene...

$$\hat{\epsilon}_i = \beta_2 \widetilde{x}_i + (\epsilon_i - \bar{\epsilon}) - \hat{\beta}_2 \widetilde{x}_i$$

Elevando al cuadrado la expresión anterior, aplicando sumatoria y tomando valor esperado:

$$\begin{aligned}
E(\sum \hat{\epsilon}_i^2) &= E(\hat{\beta}_2 - \beta)^2 \sum x_i^2 + \underbrace{E\left[\sum (\epsilon_i - \bar{\epsilon})^2\right]}_{(i)} \\
&- 2\underbrace{E\left[(\hat{\beta}_2 - \beta_2)\sum x_i(\epsilon_i - \bar{\epsilon})\right]}_{(ii)} \\
&= var(\hat{\beta}_2)\sum x_i^2 + (n-1)var(\epsilon_i) \\
&- 2E\left[\frac{\sum x_i\epsilon_i}{\sum x_i^2}\sum x_i(\epsilon_i - \bar{\epsilon})\right] \\
&= \sigma^2 + (n-1)\sigma^2 - 2\sigma^2 \\
&= (n-2)\sigma^2
\end{aligned}$$

Por lo tanto se define el estimador de la varianza $\hat{\sigma}^2$ como:

$$\hat{\sigma}^2 = \frac{\sum \hat{\epsilon}_i^2}{n-2}$$

De forma tal que, $\hat{\sigma}^2$ es un estimador insesgado de σ^2:

$$\hat{\sigma}^2 = \frac{1}{n-2} E\left(\sum \hat{\epsilon}_i^2\right) = \sigma^2$$

Ahora abandonemos la simplificación de solo usar dos variables, de ahora en adelante generalizaremos el modelo de regresión lineal para que pueda tener hasta k variables explicativas.

El Modelo de Regresión Poblacional en este caso es:

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \ldots + \beta_k x_{ik} + \epsilon_i, \qquad i = 1, ..., n$$

El modelo con k variables explicativas puede ser expresado en notación matricial:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{12} & \dots & x_{1k} \\ 1 & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n2} & \dots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

$$Y = X\beta + \epsilon$$

donde Y es un vector de dimensión $n \times 1$, X es la matriz de variables explicativas de dimensión $n \times k$ y ϵ es un vector correspondiente al término de error con dimensión $n \times 1$.

Ahora debemos expresar la distribución del término de error en términos matriciales:

$$E(\epsilon) = \begin{pmatrix} E(\epsilon_1) \\ E(\epsilon_2) \\ \vdots \\ E(\epsilon_n) \end{pmatrix} = 0_{n\times 1}$$

$$E(\epsilon\epsilon') = \begin{pmatrix} E(\epsilon_1\epsilon_1) & E(\epsilon_1\epsilon_2) & \dots & E(\epsilon_1\epsilon_n) \\ E(\epsilon_2\epsilon_1) & E(\epsilon_2\epsilon_2) & \dots & E(\epsilon_2\epsilon_n) \\ \vdots & \vdots & \ddots & \vdots \\ E(\epsilon_n\epsilon_1) & E(\epsilon_n\epsilon_2) & \dots & E(\epsilon_n\epsilon_n) \end{pmatrix} = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix}$$

Por lo tanto:

$$E(\epsilon\epsilon') = \sigma^2 I_{n\times n}$$

De los supuestos 3, 4 y 5, tenemos entonces que el término de error tiene la siguiente distribución:

$$\epsilon \sim (\mathbf{0}_{n\times 1}, \sigma^2 \mathbf{I}_{n\times n})$$

El método de MCO, plantea que los parámetros del modelo pueden ser estimados minimizando la suma de los errores al cuadrado ($S_E(\hat{\beta})$), la que en términos matriciales equivale a:

$$S_E(\hat{\beta}) = \sum \hat{\epsilon}_i^2 = \hat{\epsilon}'\hat{\epsilon}$$

donde $\hat{\epsilon} = Y - X\hat{\beta}$. Entonces el problema de minimizar la suma de los errores al cuadrado se expresa de la siguiente forma:

$$\begin{aligned} \min_{\hat{\beta}} S_E(\hat{\beta}) &= \min_{\hat{\beta}} [(Y - X\hat{\beta})'(Y - X\hat{\beta})] \\ &= \min_{\hat{\beta}} [Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}] \end{aligned}$$

Las CPO del problema son...

$$\frac{\partial S_E(\hat{\beta})}{\partial \hat{\beta}} = -2X'Y + 2X'X\hat{\beta} = 0$$

el estimador MCO de $\hat{\beta}$ es:

$$\hat{\beta}_{MCO} = (X'X)^{-1}X'Y$$

de la ecuación anterior podemos ver que:

$$X'(Y - X\hat{\beta}) = 0 \Rightarrow X'\hat{\epsilon} = 0$$

que es la condición de ortogonalidad.

Notemos que el vector $\hat{\beta}_{MCO}$ es un vector aleatorio, ya que depende del vector de errores:

$$\hat{\beta}_{MCO} = (X'X)^{-1}X'Y = (X'X)^{-1}X'(X\beta + \epsilon) = \beta + (X'X)^{-1}X'\epsilon$$

por lo tanto:

$$\begin{aligned} E(\hat{\beta}_{MCO}) &= E(\beta) + E[(X'X)^{-1}X'\epsilon] \\ &= \beta + (X'X)^{-1}X'E[\epsilon] \end{aligned}$$

La esperanza de β es el mismo parámetro, ya que este es un constante (valor poblacional), y por supuestos 2 y 3 el segundo término de la expresión anterior es cero,

$$E(\hat{\beta}_{MCO}) = \beta$$

Es decir, el estimador MCO es **insesgado**.

Podemos definir el error de estimación o sesgo como

$$\hat{\beta} - \beta = (X'X)^{-1}X'\epsilon$$

Ahora calculamos la varianza de $\hat{\beta}_{MCO}$

$$\begin{aligned} var(\hat{\beta}) &= E[(\hat{\beta} - E(\hat{\beta})) \cdot (\hat{\beta} - E(\hat{\beta}))'] \\ &= E[(\hat{\beta} - \beta) \cdot (\hat{\beta} - \beta)'] \\ &= E[(X'X)^{-1}X'\epsilon\epsilon'X(X'X)^{-1}] \\ &= (X'X)^{-1}X'E[\epsilon\epsilon']X(X'X)^{-1} \\ &= (X'X)^{-1}X'\sigma^2 I_n X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1} \end{aligned}$$

Para poder estimar la varianza de $\hat{\beta}_{MCO}$ necesitamos reemplazar σ^2 por su estimador insesgado:

$$\hat{\sigma}^2 = \frac{\hat{\epsilon}'\hat{\epsilon}}{n-k}$$

Se dice que $\hat{\beta}$, es el mejor estimador lineal insesgado (MELI) de β si se cumple lo siguiente:

- El **lineal**, es decir, es una función lineal de una variable aleatoria, como la variable **y** en el modelo de regresión.
- Es **insesgado**, es decir, su valor esperado, $E(\hat{\beta})$, es igual a el verdadero valor, β.
- Tiene varianza mínima dentro de la clase de todos los estimadores lineales insesgados; un estimador insesgado como varianza mínima es conocido como un estimador **eficiente**.

Proposición: El estimador MCO es el estimador lineal insesgado óptimo, en el sentido de que cualquier otro estimador lineal e insesgado tiene una matriz de covarianza mayor que la del estimador MCO. Es decir, el estimador MCO es MELI.

Demostración: Sea $\widetilde{\beta} = \widetilde{A}y$ un estimador lineal de β, donde $\widetilde{A}$ es una matriz de $k \times n$. Denotemos $A = \widetilde{A} - (X'X)^{-1}X'$, de modo que:

$$\begin{aligned} \widetilde{\beta} &= [A + (X'X)^{-1}X']Y \\ &= [A + (X'X)^{-1}X'](X\beta + \epsilon) \\ &= AX\beta + \beta + [A + (X'X)^{-1}X']\epsilon \end{aligned}$$

Aplicando esperanza a la expresión anterior:

$$\begin{aligned} E(\widetilde{\beta}) &= AX\beta + \beta + [A + (X'X)^{-1}X']E(\epsilon) \\ &= AX\beta + \beta \end{aligned}$$

El estimador $\widetilde{\beta}$ será insesgado sólo si la matriz A es tal que $AX = 0_{k\times k}$. De esta forma:

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

y su matriz de covarianza será:

$$\begin{aligned} cov(\widetilde{\beta}) &= E[(\hat{\beta} - E(\hat{\beta})) \cdot (\hat{\beta} - E(\hat{\beta}))'] \\ &= \sigma^2 AA' + \underbrace{\sigma^2(X'X)^{-1}}_{cov(\hat{\beta}_{MCO})} \end{aligned}$$

Como la matriz AA' es semidefinida positiva, se concluye que la diferencia entre la covarianza de $\widetilde{\beta}$ y $\hat{\beta}$ es una matriz semidefinida positiva, con lo que la covarianza de $\widetilde{\beta}$ es mayor o igual a la covarianza de $\hat{\beta}$

Sea el modelo poblacional usual con k variables:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + ... + \beta_k x_{ki} + \epsilon_i$$

donde $i = 1, ..., n$ y cuya contraparte estimada es:

$$y_i = \hat{\beta}_1 + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + ... + \hat{\beta}_k x_{ki} + \hat{\epsilon}_i$$

donde podemos reescribir el modelo estimado en desviaciones respecto a la media:

$$\widetilde{y}_i = \hat{\beta}_2 \widetilde{x}_{2i} + \hat{\beta}_3 \widetilde{x}_{3i} + ... + \hat{\beta}_k \widetilde{x}_{ki} + \hat{\epsilon}_i$$

Esta última expresión es similar a la anterior, excepto por dos importantes diferencias. Primero, el modelo no posee constante y segundo, las variables se encuentran expresadas en desvíos con respecto a la media. A pesar de ello, note que los coeficientes y los residuos son los mismos en ambos modelos.

De lo anterior surge un importante corolario respecto del término constante de nuestro modelo.

En general, el interés del investigador se centra en el impacto de los regresores sobre la variable dependiente, por lo cual, el término constante no es más que una corrección que garantiza que los promedios muestrales de ambos miembros del modelo econométrico coincidan.

Para transformar en desvíos con respecto a la media un modelo en términos matriciales, introduciremos una matriz fundamental para el análisis de esta sección. Denotaremos por M^0 una matriz de $n \times n$, definida como:

$$M^0 = I_n - \frac{ii'}{n} = \begin{pmatrix} 1 - \frac{1}{n} & -\frac{1}{n} & \dots & -\frac{1}{n} \\ -\frac{1}{n} & 1 - \frac{1}{n} & \dots & -\frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & \dots & 1 - \frac{1}{n} \end{pmatrix}$$

donde I_n es la matriz identidad de orden n e i corresponde al vector unitario de dimensión n. M^0 es singular, simétrica e idempotente. M^0 se conoce como la *matríz de desvíos*.

Suponga entonces el siguiente modelo poblacional:

$$Y = X\beta + \epsilon$$

Buscamos entonces definir la variación de la variable dependiente (Suma de los cuadrados totales = TSS) como:

$$TSS = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

Para encontrar entonces una expresión para la ecuación anterior, de la ecuación poblacional tenemos que nuestro modelo estimado en desvíos con respecto a la media es:

$$M^0 Y = M^0 X\hat{\beta} + M^0\hat{\epsilon}$$

Con lo cual, al particionar nuestra matriz X en $X = [i \;\; X_2]$, nuestro vector de parámetros en $\beta = [\beta_1 \;\; \beta_2]$ y considerando que $M^0 i = 0$ y que $M^0 \hat{\epsilon} = \hat{\epsilon}$, tenemos que:

$$\begin{aligned} M^0 Y &= M^0 i \hat{\beta}_1 + M^0 X_2 \hat{\beta}_2 + M^0 \hat{\epsilon} \\ &= M^0 X_2 \hat{\beta}_2 + \hat{\epsilon} \end{aligned}$$

Por lo tanto, para formar la TSS, multiplicamos por Y' la ecuación anterior:

$$\begin{aligned} Y' M^0 Y &= Y'(M^0 X_2 \hat{\beta}_2 + \hat{\epsilon}) \\ &= (X\hat{\beta} + \hat{\epsilon})'(M^0 X_2 \hat{\beta}_2 + \hat{\epsilon}) \\ &= \hat{\beta}_2' X_2' M^0 X_2 \hat{\beta}_2 + \hat{\epsilon}'\hat{\epsilon} \\ TSS &= ESS + RSS \end{aligned}$$

donde el segundo y el tercer término desaparecen gracias a que los residuos estimados son, por construcción, ortogonales a las variables explicativas.

Definimos entonces la bondad de ajuste del modelo a través del siguiente estadígrafo llamado también coeficiente de determinación:

$$R^2 = \frac{ESS}{TSS}$$

es decir, como la proporción de la varianza de Y que es explicada por la varianza de la regresión. Alternativamente:

$$R^2 = 1 - \frac{RSS}{TSS}$$

Note que:

1. El coeficiente de determinación es siempre menor a 1. Ello porque $RSS \leq TSS$ y por lo tanto $\frac{RSS}{TSS} \leq 1$.

2. El análisis de varianza anterior fue derivado bajo el supuesto que el modelo incluía una constante (por ello utilizábamos la matriz M^0). En dicho caso, necesariamente $R^2 \geq 0$. Qué sucede si el modelo no posee constante?

3. Al agregar regresores al modelo, el R^2 nunca decrecerá (se mantendrá constante o aumentará).

4. No es claro cuán bueno sea como predictor de ajuste.

Según el punto 3, el coeficiente de determinación probablemente crecerá al incluir regresores. Ello plantea incentivos a incluir regresores no relevantes para nuestro modelo, con el fin de obtener un mejor ajuste.

Por esta razón se creó el coeficiente de determinación ajustado, el cual corrige el R^2 original por los grados de libertad del numerador y el denominador. Entonces, definimos el R^2 ajustado ($\widetilde{R}^2$) como:

$$\widetilde{R}^2 = 1 - \frac{\hat{\epsilon}'\hat{\epsilon}/(n-k)}{Y'M^0Y/(n-1)}$$

o equivalentemente:

$$\widetilde{R}^2 = 1 - (1-R^2)\frac{(n-1)}{(n-k)}$$