Econometría I
Ayudantía 0

Profesora: Giselli Castillo

Ayudante: Agustín Sanhueza

Repaso Matemático

SE RECOMIENDA LEER LOS APÉNDICES A, B, C y D DEL LIBRO DE WOOLDRIDGE.

1. Sumatorias

Respuesta

Sean "a", "b" y "c" constantes.

$\sum_{i=1}^{T} x_i = x_1 + x_2 + ... + x_T$

$\frac{\sum_{i=1}^{T} x_i}{T} = \overline{x}$

$\sum_{i=1}^{T} c = T * c$

$\sum_{i=1}^{T} cx_i = c\sum_{i=1}^{T} x_i$

$\sum_{i=1}^{T}(ax_i + by_i) = a\sum_{i=1}^{T} x_i + b\sum_{i=1}^{T} y_i$

$\sum_{i=1}^{T}\left(\frac{x_i}{y_i}\right) \neq \frac{\sum_{i=1}^{T} x_i}{\sum_{i=1}^{T} y_i} \quad \forall y_i \neq 0$

$\sum_{i=1}^{T}(x_i - \overline{x})^2 = \sum_{i=1}^{T} x_i^2 - n(\overline{x})^2$

$\sum_{i=1}^{T}(x_i - \overline{x}) = 0$

$\sum_{i=1}^{T}(x_i - \overline{x})(y_i - \overline{y}) = \sum_{i=1}^{T} y_i(x_i - \overline{x}) = \sum_{i=1}^{T} x_i(y_i - \overline{y})$

2. Esperanzas

Respuesta

Sean "a_i", "b" y "c" constantes. X_i variables aleatorias.

$\mathbb{E}(c) = c$

$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$

$\mathbb{E}\left(\sum_{i=1}^{T} a_i X_i\right) = a_i \sum_{i=1}^{T} \mathbb{E}(X_i)$

3. Varianzas

Respuesta

Sean "a", "b" y "c" constantes. X e Y variables aleatorias. Además $\mathbb{E}(X) = \mu$

$Var(X) = \mathbb{E}[(X - \mu)^2] = \sigma^2$

$Var(X) = E[X^2] - \mu^2$

$Var(c) = 0$

$Var(aX + b) = a^2 Var(X)$

$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X, Y)$

4. Propiedades de Matrices

Respuesta

Sean A y B Matrices, α un escalar e Y una variable aleatoria.

$(A')' = A$

$AB \neq BA$ (Este es el caso más frecuente)

$(AB)' = B'A'$

$(AB)^{-1} = B^{-1}A^{-1}$ (Suponiendo A y B invertibles)

$(A')^{-1} = (A^{-1})'$ (Suponiendo A invertible)

$Var(AY) = AVar(Y)A'$

$A\alpha B = \alpha AB = AB\alpha$

$tr(A+B) = tr(A) + tr(B)$

$\alpha' = \alpha$

5. Derivadas de Matrices

Respuesta

Sean X una matriz cuadrada y "v" un vector columna

$$v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{bmatrix}_{m \times 1} \qquad X = \begin{bmatrix} x_{1,1} & \cdots & x_{1,m} \\ \vdots & \ddots & \vdots \\ x_{m,1} & \cdots & x_{m,m} \end{bmatrix}_{m \times m}$$

$\frac{\partial(v'X)}{\partial v} = X$

$\frac{\partial(v'X)}{\partial v'} = X'$

$\frac{\partial(v'Xv)}{\partial v} = (X + X')v$

Econometría I
Pauta Ayudantía 1

Profesora: Giselli Castillo

Ayudante: Agustín Sanhueza

Comentes

1. Mencione y explique los supuestos de la estimación por Mínimos Cuadrados Ordinarios.

Respuesta

1) La forma funcional es lineal en parámetros.

2) Los datos provienen de una muestra aleatoria.

3) Existe variación en las variables explicativas. $\sum_{i=1}^{n}(x_i - \overline{x})^2 > 0$

4) Independencia de media condicional: $E[u/x] = 0$. Luego, por ley de expectativas iteradas $E[u] = E[E[u/x]] = 0$

5) La varianza del error es constante (homocedasticidad): $Var(u/x) = \sigma^2$. Luego, el estimador será eficiente.

6) No existe multicolinealidad perfecta. Luego, $(X'X)^{-1}$ es invertible

7) Los errores distribuyen normal. (útil para hacer inferencia en muestras pequeñas)

8) No existe correlación entre las variables explicativas y el término de error (Exogenedidad): $Cov(x, u) = 0$. Luego, el estimador será insesgado y consistente.

9) No existe autocorrelación en el término de error. $Cov(u_t, u_{t-j}) = 0 \ \forall j > 0$ (útil en datos de series de tiempo y panel)

Matemático 1

Suponga que tiene el siguiente modelo de regresión.

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

a) Escriba el problema a minimizar para calcular los estimadores de mínimos cuadrados ordinarios (MCO), y luego obténgalos.

Respuesta

$$\min_{\{\beta_0,\beta_1\}} \sum_{i=1}^{T} u_i^2 = \sum_{i=1}^{T} (y_i - \beta_0 - \beta_1 x_i)^2$$

$$\frac{\partial \sum_{i=1}^{T} u_i^2}{\partial \beta_0}|_{\beta=\hat{\beta}} = -2\sum_{i=1}^{T}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\sum_{i=1}^{T}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\sum_{i=1}^{T} y_i - \sum_{i=1}^{T} \hat{\beta}_0 - \sum_{i=1}^{T} \hat{\beta}_1 x_i = 0$$

$$\sum_{i=1}^{T} y_i - T * \hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{T} x_i = 0$$

$$T * \hat{\beta}_0 = \sum_{i=1}^{T} y_i - \hat{\beta}_1 \sum_{i=1}^{T} x_i$$

$$\hat{\beta}_o = \sum_{i=1}^{T} \frac{y_i}{T} - \hat{\beta}_1 \sum_{i=1}^{T} \frac{x_i}{T}$$

$$\boxed{\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}} \tag{1}$$

$$\frac{\partial \sum_{i=1}^{T} u_i^2}{\partial \beta_1}|_{\beta=\hat{\beta}} = -2\sum_{i=1}^{T}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)x_i = 0$$

$$\sum_{i=1}^{T} x_i(y_i - (\overline{y} - \hat{\beta}_1 \overline{x}) - \hat{\beta}_1 x_i) = 0$$

$$\sum_{i=1}^{T} x_i[(y_i - \overline{y}) - \hat{\beta}_1 (x_i - \overline{x})] = 0$$

$$\sum_{i=1}^{T} x_i(y_i - \overline{y}) - \hat{\beta}_1 \sum_{i=1}^{T} x_i(x_i - \overline{x})] = 0$$

$$\sum_{i=1}^{T} (y_i - \overline{y})(x_i - \overline{x}) = \hat{\beta}_1 \sum_{i=1}^{T} (x_i - \overline{x})^2$$

$$\boxed{\hat{\beta}_1 = \frac{\sum_{i=1}^{T}(y_i - \overline{y})(x_i - \overline{x})}{\sum_{i=1}^{T}(x_i - \overline{x})^2}} \tag{2}$$

b) Demuestre que ambos estimadores son insesgados.

Respuesta

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{T}(x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^{T}(x_i-\overline{x})^2}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{T}(x_i-\overline{x})[(\beta_0+\beta_1 x_i+u_i)-(\beta_0+\beta_1\overline{x}+\overline{u})]}{\sum_{i=1}^{T}(x_i-\overline{x})^2}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{T}(x_i-\overline{x})(\beta_1(x_i-\overline{x})+u_i)}{\sum_{i=1}^{T}(x_i-\overline{x})^2}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{T}\beta_1(x_i-\overline{x})^2+(x_i-\overline{x})u_i}{\sum_{i=1}^{T}(x_i-\overline{x})^2}$$

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{T}(x_i-\overline{x})u_i}{\sum_{i=1}^{T}(x_i-\overline{x})^2}$$

$$\mathbb{E}[\hat{\beta}_1] = \beta_1 + \mathbb{E}\left[\frac{\sum_{i=1}^{T}(x_i-\overline{x})u_i}{\sum_{i=1}^{T}(x_i-\overline{x})^2}\right]$$

$$\mathbb{E}[\hat{\beta}_1] = \beta_1 + \frac{\sum_{i=1}^{T}(x_i-\overline{x})\mathbb{E}(u_i)}{\sum_{i=1}^{T}(x_i-\overline{x})^2}$$

$$\boxed{\mathbb{E}[\hat{\beta}_1] = \beta_1} \tag{3}$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x}$$

$$\hat{\beta}_0 = \beta_0 + \beta_1\overline{x} + \overline{u} - \hat{\beta}_1\overline{x}$$

$$\mathbb{E}[\hat{\beta}_0] = \beta_0 + \beta_1\overline{x} - \mathbb{E}[\hat{\beta}_1]\overline{x}$$

$$\mathbb{E}[\hat{\beta}_0] = \beta_0 + \beta_1\overline{x} - \beta_1\overline{x}$$

$$\boxed{\mathbb{E}[\hat{\beta}_0] = \beta_0} \tag{4}$$

c) Calcule la varianza de $\hat{\beta}_1$.

Respuesta

$$Var(\hat{\beta}_1) = \mathbb{E}[(\hat{\beta}_1 - \mathbb{E}(\hat{\beta}_1))^2]$$

$$Var(\hat{\beta}_1) = \mathbb{E}\left[\frac{\sum_{i=1}^{T}(x_i - \overline{x})u_i}{\sum_{i=1}^{T}(x_i - \overline{x})^2}\right]^2$$

$$Var(\hat{\beta}_1) = \mathbb{E}\left[\frac{\sum_{i=1}^{T}(x_i - \overline{x})^2 u_i^2}{\left(\sum_{i=1}^{T}(x_i - \overline{x})^2\right)^2}\right]$$

$$Var(\hat{\beta}_1) = \frac{\sum_{i=1}^{T}(x_i - \overline{x})^2 \mathbb{E}(u_i^2)}{\left(\sum_{i=1}^{T}(x_i - \overline{x})^2\right)^2}$$

$$\boxed{Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{T}(x_i - \overline{x})^2}} \tag{5}$$

Matemático 2

Suponga que tiene el siguiente modelo de regresión con k variables y T observaciones. $\forall k < T$

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_k x_{ki} + u_i$$

a) Escriba el problema a minimizar matricialmente y resuelva cuál debiese ser el estimador de MCO.

Respuesta

Matricialmente el modelo se podría escribir de la siguiente manera:

$$Y_{T\times 1} = X_{T\times(k+1)}\beta_{(k+1)\times 1} + u_{T\times 1}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix}_{T\times 1} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,k} \\ 1 & x_{2,1} & x_{2,2} & \cdots & x_{2,k} \\ \vdots & \vdots & \ddots & \vdots & \\ 1 & x_{T,1} & x_{T,2} & \cdots & x_{T,k} \end{bmatrix}_{T\times(k+1)} * \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}_{(k+1)\times 1} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{bmatrix}_{T\times 1}$$

$$\min_{\{\beta\}} \ S_T = u'u = (Y - X\beta)'(Y - X\beta)$$

$$u'u = Y'Y - \beta'X'Y - Y'X\beta + \beta X'X\beta$$

$$u'u = Y'Y - 2Y'X\beta + \beta X'X\beta$$

$$\frac{\partial u'u}{\partial \beta}|_{\beta=\hat{\beta}} = -2X'Y + 2X'X\hat{\beta} = 0$$

$$X'X\hat{\beta} = X'Y$$

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

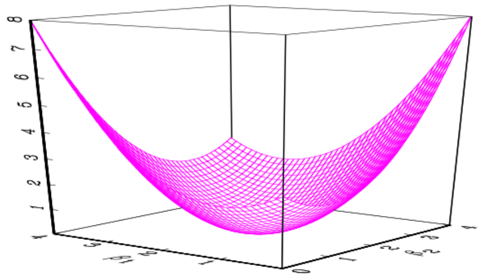


Figure 3.1: Sum-of-Squared Errors Function

b) Demuestre que el estimador MCO es insesgado.

Respuesta

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

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

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

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

$$\mathbb{E}[\hat{\beta}] = \beta + (X'X)^{-1}X'\mathbb{E}[u]$$

$$\boxed{\mathbb{E}[\hat{\beta}] = \beta}$$

c) Calcule la varianza del estimador MCO.

Respuesta

$$V(\hat{\beta}) = \mathbb{E}\left[(\hat{\beta} - \mathbb{E}[\hat{\beta}])(\hat{\beta} - \mathbb{E}[\hat{\beta}])'\right]$$

$$V(\hat{\beta}) = \mathbb{E}\left[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'\right]$$

$$V(\hat{\beta}) = \mathbb{E}\left[((X'X)^{-1}X'u)((X'X)^{-1}X'u)'\right]$$

$$V(\hat{\beta}) = \mathbb{E}\left[(X'X)^{-1}X'uu'X(X'X)^{-1}\right]$$

$$V(\hat{\beta}) = (X'X)^{-1}X'\mathbb{E}[uu']X(X'X)^{-1}$$

$$V(\hat{\beta}) = \sigma^2(X'X)^{-1}X'X(X'X)^{-1}$$

$$\boxed{V(\hat{\beta}) = \sigma^2(X'X)^{-1}}$$

Matemático 3

Usted tiene los siguientes datos

Estudiante	GPA	ACT
1	2.8	21
2	3.4	24
3	3	26
4	3.5	27
5	3.6	29
6	3	25
7	2.7	25
8	3.7	30

1. Obtenga los estimadores de MCO para escribir el siguiente modelo de regresión $G\hat{P}A_i = \hat{\beta}_0 + \hat{\beta}_1 ACT_i$. Luego calcule cada $G\hat{P}A_i$ y $\hat{u}_i$

Respuesta

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{8}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{8}(x_i - \overline{x})^2}$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x}$$

Viendo los datos tendremos que:
$\overline{y} = 3{,}2125$
$\overline{x} = 25{,}875$
$\sum_{i=1}^{8}(x_i - \overline{x})(y_i - \overline{y}) = 5{,}8125$
$\sum_{i=1}^{8}(x_i - \overline{x})^2 = 56{,}875$

Con esto llegamos a que:

$$\boxed{\hat{\beta}_1 = \frac{5{,}8125}{56{,}875} \approx 0{,}1022}$$

$$\boxed{\hat{\beta}_0 = 3{,}2125 - 0{,}1022 * 25{,}875 \approx 0{,}5681}$$

$$\boxed{G\hat{P}A_i = 0{,}5681 + 0{,}1022 ACT_i}$$

Reemplazando los datos que tenemos de la tabla en el modelo llegaremos a lo siguiente:

Estudiante	GPA	$G\hat{P}A$	$\hat{u}$
1	2.8	2.7143	0.0857
2	3.4	3.0209	0.3791
3	3	3.2253	-0.2253
4	3.5	3.3275	0.1725
5	3.6	3.5319	0.0681
6	3	3.1231	-0.1231
7	2.7	3.1231	-0.4231
8	3.7	3.6341	0.0659

2. Exprese el modelo anterior de forma matricial y encuentre los estimadores MCO.

Respuesta

$$Y = X\beta + U$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix}_{T\times 1} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,k} \\ 1 & x_{2,1} & x_{2,2} & \cdots & x_{2,k} \\ \vdots & \vdots & \ddots & \vdots & \\ 1 & x_{T,1} & x_{T,2} & \cdots & x_{T,k} \end{bmatrix}_{T\times(k+1)} * \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}_{(k+1)\times 1} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{bmatrix}_{T\times 1}$$

$$\begin{bmatrix} 2{,}8 \\ 3{,}4 \\ 3 \\ 3{,}5 \\ 3{,}6 \\ 3 \\ 2{,}7 \\ 3{,}7 \end{bmatrix}_{8\times 1} = \begin{bmatrix} 1 & 21 \\ 1 & 24 \\ 1 & 26 \\ 1 & 27 \\ 1 & 29 \\ 1 & 25 \\ 1 & 25 \\ 1 & 30 \end{bmatrix}_{8\times 2} * \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}_{2\times 1} + \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \\ u_7 \\ u_8 \end{bmatrix}_{8\times 1}$$

Recordando que

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

En este caso tendremos que

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} T & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \left[\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 21 & 24 & 26 & 27 & 29 & 25 & 25 & 30 \end{bmatrix} \begin{bmatrix} 1 & 21 \\ 1 & 24 \\ 1 & 26 \\ 1 & 27 \\ 1 & 29 \\ 1 & 25 \\ 1 & 25 \\ 1 & 30 \end{bmatrix} \right]^{-1} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 21 & 24 & 26 & 27 & 29 & 25 & 25 & 30 \end{bmatrix} \begin{bmatrix} 2{,}8 \\ 3{,}4 \\ 3 \\ 3{,}5 \\ 3{,}6 \\ 3 \\ 2{,}7 \\ 3{,}7 \end{bmatrix}$$

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} 8 & 207 \\ 207 & 5413 \end{bmatrix}^{-1} \begin{bmatrix} 25{,}7 \\ 670{,}8 \end{bmatrix}$$

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} \approx \begin{bmatrix} 11{,}8967 & -0{,}4549 \\ -0{,}4549 & 0{,}01758 \end{bmatrix} \begin{bmatrix} 25{,}7 \\ 670{,}8 \end{bmatrix}$$

$$\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\end{bmatrix} \approx \begin{bmatrix}0{,}5681\\ 0{,}1022\end{bmatrix}$$

Conociendo R

Paquetes y Datos

1. Si aún no lo tiene, vaya al siguiente link descargue e instale R[1]. Se recomienda bajar la interfaz "Rstudio" del siguiente link. Esta interfaz es bastante amigable, y la usaremos en las ayudantías (se requiere la instalación de R previamente).

2. Programe el cálculo de su nota promedio en el curso.

3. Instale y cargue los paquetes *wooldridge, tidyverse, dplyr, ggplot2, ggdark, readxl y writexl.*

4. Importe la base "wage1" del paquete *wooldridge.*

5. Seleccione las variables *lwage, wage, educ, exper y female*. Cambie el nombre de estas variables por *lnsalrio, salario, educación, experiencia y mujer* respectivamente.

6. Calcule el salario, años de escolaridad, y años de experiencia promedios de la base (aproxime los resultados a 1 decimal). También calcule el número de mujeres y hombres en la base.

7. Elimine la variable "lnsalario" y genere una nueva variable llamada "sueldos", que tome valores aleatorios entre el salario mínimo chileno y los 10 millones.

8. Genere un dataframe auxiliar llamado "df_aux" en la que estén las mismas variables que en "wage1", solo que con las personas con menos de 6 años de escolaridad.

Gráficos, regresión y exportación

9. Genere un histograma del salario.

10. Usando la base "wage1" Calcule los betas, residuos y valores predichos del siguiente modelo de regresión.
$$salario_i = \beta_0 + \beta_1 escolaridad_i + \mu_i$$
Guarde los resultados en el objeto "Regresión1" e interprete los resultados.

11. En un gráfico de dispersión, reproduzca los salarios vs los años de escolaridad. Muestre los valores predichos de la recta de regresión y su intervalo de confianza.

12. Haga lo mismo que en el ítem anterior solo que ahora separando un gráfico para hombres y otro para mujeres.

13. Exporte la base "wage1" en formato excel a alguna carpeta de su computador. Guárde el archivo con el nombre "sueldos1".

[1] Video guía para Windows. Video guía para Mac

Econometría I
Pauta Ayudantía 2

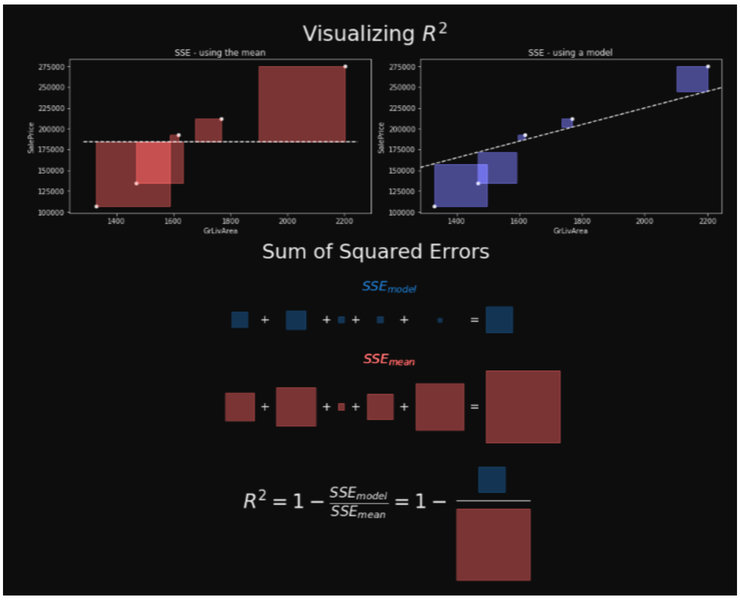
Profesora: Giselli Castillo
Ayudante: Agustín Sanhueza.

Comentes

1. Explique qué mide el R^2, apóyese en un análisis gráfico. Explique también qué es lo que NO mide el R^2, apoyese en un análisis teórico.

Respuesta

El R^2 hace un análisis de descomposición de varianzas (ANOVA en inglés), en donde se divide la variación del "componente sistemático" con respecto al total. Alternativamente se calcula como 1 - la variación del "componente no sistemático" con respecto al total. (En pizarra cambiar nombres de descomposición por SSE, SSR y SST para que alumnos lo relacionen con notación de clases y no se confundan)



Se dice mucho que el R^2 "explica" las variaciones (o comportamiento) de nuestra variable dependiente usando las variables regresoras (en un rango de 0 - 100 %). En realidad ese "explica" es simplemente una manera de hacer ver la descomposición de varianza.

La estadística y la econometría son **herramientas** que nos permiten estimar relaciones causales, hacer predicciones, ver asociaciones (no) lineales entre variables, etc. Estas herramientas **no explican nada de lo que estamos analizando**. Lo que realmente "explica" los fenómenos que observamos, es una teoría económica/psicológica/sociológica que nos entregan un marco teórico de cómo y por qué una variable de interés respondería a otras variables.

Por lo tanto, lo que NO hace el R^2 es indicarnos si el modelo que estamos usando es bueno o apropiado (algo que lamentablemente está muy internalizado cuando se estiman y/o comparan modelos).

Ejemplo 1, Aumento en N° de variables (basura):
Supongamos que se obtuvo un $R^2 = 0{,}6$ para el siguiente modelo de regresión

$$y_i = \beta_0 + \beta_1 x_{1,i} + u_i$$

Donde y_i es el ahorro de las persona i, y $x_{1,i}$ es el ingreso de la persona i.

Ahora agregamos 2 variables al modelo anterior obteniendo lo siguiente.

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} + u_i$$

Donde $x_{2,i}$ es el número de pelos en la cabeza que tiene la persona i, y $x_{3,i}$ es la altura de la persona i. Dado que el R^2 es creciente al número de regresores este modelo nos entregará un $R^2 > 0{,}6$ digamos un $R^2 = 0{,}67$. ¿Quiere decir esto que este modelo ”explica” de mejor manera el comportamiento del ahorro? NO!

Ejemplo 2, Reparametrización:
Supongamos que dos personas corren cada uno por su cuenta alguno de los siguientes modelos de regresión

$$y_i = \beta_0 + \beta_1 x_{1,i} + u_i \tag{1}$$

$$y_i - x_{1,i} = \beta_0 + \gamma_1 x_{1,i} + u_i \tag{2}$$

con $\gamma_1 = (\beta_1 - 1)$
Como podemos ver, ambos modelos son exactamente iguales solo que escritos de una manera distinta, por lo que sus implicancias económicas deberán ser las mismas.

Supongamos que por mala suerte el verdadero valor de $\beta_1 = 1$ o muy cercano a ese valor. Si ese es el caso, entonces tendremos que en el segundo modelo el valor de $\gamma_1 = 0$ o tiende a ese valor. Esto significaría que la persona que corrió el modelo 2 tendrá que su variable dependiente solo se ”explica” por el componente no sistemático, un R^2 que tiende a 0.

Por lo tanto, podemos tener dos modelos exactamente iguales solo que reparametrizados y obtener R^2 muy distintos. ¿Eso significa que la persona que corrió el modelo 1 ”explica” de mejor forma al fenómeno de estudio? NO!

Ejemplo 3, Observaciones y Parámetros a estimar:
Supongamos que tenemos el mismo modelo de regresión que en el ejemplo 1.

$$y_i = \beta_0 + \beta_1 x_{1,i} + u_i$$

Y que tenemos dos muestras. Una con 1000 observaciones y otra con 2 observaciones (caso extremo, pero que ilustra bien la idea). En el primer caso digamos que tendremos un $R^2 = 0{,}6$ El tema es que en el segundo caso tendremos por definición un $R^2 = 1$ (las dos observaciones me permiten calcular intercepto y pendiente de tal manera que no haya error posible en la estimación). Dos magnitudes totalmente distintas para el mismo modelo (este mismo argumento se puede extender para modelos con número de observaciones iguales a número de parámetros a estimar, consumo total de grados de libertad).

Ejemplo 4, Estimación de tautologías vs modelos:
Una tautología es por definición correcta (PIB=C+I+G+XN). Un modelo por definición es incorrecto (es una abstracción simplificada, pero poderosa, de la realidad para entenderla mejor). Si corremos una regresión de una Tautología tendremos siempre un $R^2 = 1$ (se "ajusta" perfectamente a los datos). Si corremos un modelo tendremos típicamente un $R^2 < 1$.

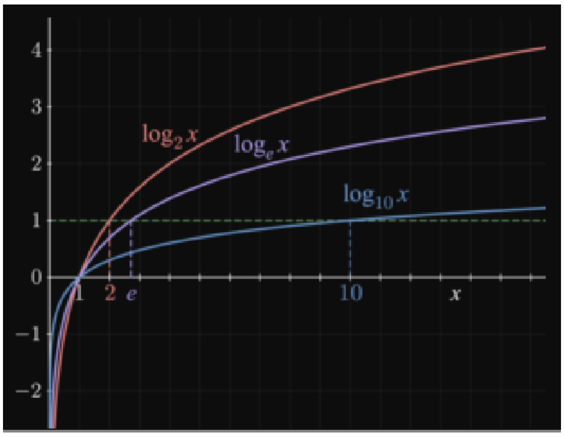
Conclusión: El estadístico R^2 es sensible a:
Número de variables.
Forma de parametrizar el modelo.
Observaciones y parámetros a estimar.
Estimación de tautologías o modelos.

Por todo lo argumentado anteriormente, jamás rankear estimaciones en función de este estadístico.

¿Qué podemos hacer entonces con el R^2?
Usarlo como un indicador más de muchos otros (que se verán más adelante en el curso) sobre un modelo pará así tener una imagen completa de este. El R^2 ("parcial") si es útil para cuantificar la colinealidad de las variables y el modelo en general (algo que se ve más adelante en el test VIF).

2. Señale en qué caso no se podría usar logaritmos en un modelo de regresión. Luego explique por qué el uso de logaritmos sirve para evaluar cambios porcentuales en modelos de regresión.

Respuesta



Como podemos ver de la imagen, independiente de la base que tenga el logaritmo, su dominio es $]0, \infty[$. Por lo tanto, no podemos usar un modelo de regresión en donde existan observaciones con valores igual a cero o negativos. Por ejemplo, si tenemos que y_i es el precio del pétroleo, y x_i es una medida de costos del petróleo en el año 2020 (donde hubo precios negativos). El modelo

$$log(y_i) = \beta_0 + \beta_1 log(x_i) + u_i$$

no podrá ser estimado.

Podemos escribir la variación porcentual de una variable Y de la siguiente manera

$$\Delta \% Y = \frac{y_t - y_{t-1}}{y_{t-1}}$$

Paralelamente tenemos que

$$\Delta log(Y) = log(y_t) - log(y_{t-1})$$

$$\Delta log(Y) = log\left(\frac{y_t}{y_{t-1}}\right)$$

$$\Delta log(Y) = log\left(\frac{y_t}{y_{t-1}} + 1 - 1\right)$$

$$\Delta log(Y) = log\left(\frac{y_t}{y_{t-1}} + 1 - \frac{y_{t-1}}{y_{t-1}}\right)$$

$$\Delta log(Y) = log\left(1 + \frac{y_t - y_{t-1}}{y_{t-1}}\right)$$

Si tenemos que $\Delta \% Y = \frac{y_t - y_{t-1}}{y_{t-1}}$ es cercano a 0 entonces.

$$\boxed{\Delta log(Y) = log\left(1 + \frac{y_t - y_{t-1}}{y_{t-1}}\right) \approx \frac{y_t - y_{t-1}}{y_{t-1}}}$$

Por ejemplo, si hay un cambio porcentual de $3\,\%$

$$ln(1 + 0{,}03) \approx 0{,}03$$

$$ln(1{,}03) \approx 0{,}03$$

$$0{,}0295588 \approx 0{,}03$$

Una aproximación bastante razonable

Por lo tanto, si queremos estimar el cambio porcentual de la variable Y ante cambios porcentuales de la variable X, podemos estimar lo siguiente

$$log(y_t) = \beta_0 + \beta_1 log(x_t) + u_t$$

¿Por qué?

$$\beta_1 = \frac{\Delta log(y_t)}{\Delta log(x_t)} \approx \frac{\frac{y_t - y_{t-1}}{y_{t-1}}}{\frac{x_t - x_{t-1}}{x_{t-1}}}$$

Mencionar que típicamente, usamos el logaritmo natural (logaritmo en base e) para estimar regresiones dado que este tiene propiedades que facilitan mucho la vida al hacer operaciones.

DISTINCIÓN A REFORZAR:
"Aumento de 1 punto porcentual" $\neq$ "Aumento de 1 %"

Pasar de 1 % a 2 % $\Rightarrow$ "Aumento de 1 punto porcentual" (100 puntos base)

Pasar de 100 a 101 $\Rightarrow$ "Aumento de 1 %"

3. Suponga que un modelo de regresión lineal simple con constante y una variable regresora. Explique la interpretación del coeficiente β_1 en modelos nivel-nivel, nivel-log, log-nivel y log-log.

Respuesta

Podemos resumir los cuatro tipos de modelos con la siguiente tabla (usaremos logaritmo natural):

Modelos	Regresión	β_1	Pendiente = $\frac{\partial y}{\partial x}$	Interpretación
Nivel-Nivel	$y_i = \beta_0 + \beta_1 x_i + u_i$	$\frac{\Delta y_i}{\Delta x_i}$	β_1	Cambio en nivel
Nivel-Log	$y_i = \beta_0 + \beta_1 ln(x_i) + u_i$	$\frac{\Delta y_i}{\Delta \% x_i}$	$\beta_1 \frac{1}{x}$	Semi-elasticidad
Log-Nivel	$ln(y_i) = \beta_0 + \beta_1 x_i + u_i$	$\frac{\Delta \% y_i}{\Delta x_i}$	$\beta_1 y$	Semi-elasticidad
Log-Log	$ln(y_i) = \beta_0 + \beta_1 ln(x_i) + u_i$	$\frac{\Delta \% y_i}{\Delta \% x_i}$	$\beta_1 \frac{y}{x}$	Elasticidad

El coeficiente β_1 mide en:
Nivel-Nivel $\Rightarrow$ "El cambio promedio de la variable Y cuando aumenta la variable X en una unidad dejando lo demás constante"

Nivel-Log $\Rightarrow$ "El cambio promedio de la variable Y cuando aumenta la variable X en un 1 % dejando lo demás constante"

Log-Nivel $\Rightarrow$ "El cambio promedio porcentual de la variable Y cuando aumenta la variable X en una unidad dejando lo demás constante"

Log-Log $\Rightarrow$ "El cambio promedio porcentual de la variable Y cuando aumenta la variable X en un 1 % dejando lo demás constante"

Considerar que a diferencia del cálculo, donde el cambio marginal era ϵ (una cantidad extremadamente pequeña mayor a 0), en contextos económicos el cambio marginal es una unidad.

Para no enredar a alumnos, no es necesario mostrar la penúltima columna ni mencionar lo que viene:
A excepción del primer tipo de modelos ahora existen diferencias entre el coeficiente β_1 y la pendiente $\frac{\partial y}{\partial x}$. Esto se explica por la " Desigualdad de Jensen". En el contexto de estos modelos de regresión, la desigualdad indicaría que: $\mathbb{E}[exp(ln(y_t)/x_t)] \neq exp[\mathbb{E}(ln(y_t)/x_t)]$ (el valor esperado de una función no lineal, es distinto a la función no lineal del valor esperado)

4. De un ejemplo de un modelo lineal y no lineal, tanto en parámetros como en regresores (4 ejemplos en total).

Respuesta

	Parámetros	Regresores
Líneal	$y_i = \beta_0 + \beta_1 x_{1,i} + u_i$	$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + u_i$
No Lineal	$\mathrm{y}_i = \beta_0 + \beta_1 exp(\beta_2 x_{1,i}) + u_i$	$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{1,i}^2 + u_i$

El punto aquí es dejar claro que un modelo puede ser lineal en parámetros y no lineal en regresores (cuadrante IV).

Mencionar que salvo casos muy excepcionales siempre debemos colocar constante β_0 en los modelos. Si no lo hacemos, forzaremos que la regresión pase por el origen ($\beta_0 = 0$) lo que muy seguramente no generará la minimización de SSR y hará que se pierda la propiedad de que la regresión pasa por el promedio de las variables. Casos excepcionales como la estimación de modelos CAPM (suponiendo que se cumplen) sería un ejemplo donde no habría problema.

Dado que no se ve en el curso, solo mencionar que modelos que no son lineales en parámetros como en el cuadrante III no tienen solución análitíca, por lo que habría que usar métodos númericos/iterativos (Gauss-Newton, Newton-Raphson, Método de Secante, etc).

Matemático

De la ayudantía 1 teníamos los siguientes datos

Estudiante	GPA	$\hat{GPA}$	$\hat{u}$	ACT
1	2.8	2.7143	0.0857	21
2	3.4	3.0209	0.3791	24
3	3	3.2253	-0.2253	26
4	3.5	3.3275	0.1725	27
5	3.6	3.5319	0.0681	29
6	3	3.1231	-0.1231	25
7	2.7	3.1231	-0.4231	25
8	3.7	3.6341	0.0659	30

a) Calcule el R^2 del siguiente modelo de regresión.

$$GPA_i = \beta_0 + \beta_1 ACT_i + u_i$$

Respuesta

Recordemos que

$$SST = SSE + SSR$$

$$\boxed{\sum_{i=1}^{T}(y_i - \overline{y})^2 = \sum_{i=1}^{T}(\hat{y}_i - \overline{y})^2 + \sum_{i=1}^{T}\hat{u}_i{}^2}$$

$$\boxed{R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}}$$

Viendo los datos tendremos que:
$\overline{y} = 3{,}2125$
$SST = \sum_{i=1}^{8}(y_i - \overline{y})^2 = 1{,}02875$
$SSE = \sum_{i=1}^{8}(\hat{y}_i - \overline{y})^2 = 0{,}59405028$
$SSR = \sum_{i=1}^{8}\hat{u}_i{}^2 = 0{,}43472528$

Por lo tanto

$$\boxed{R^2 = \frac{0{,}59405028}{1{,}02875} = 1 - \frac{0{,}43472528}{1{,}02875} \approx 0{,}58}$$

Sigamos con R

1. Revisión y preguntas del script de la ayudantía anterior.

Econometría I
Pauta Ayudantía 3

Profesora: Giselli Castillo

Ayudante: Agustín Sanhueza

Comentes

1. Usted tiene los siguientes modelos de regresión.

$$Salario_i = \beta_0 + \beta_1 Escolaridad_i + u_i \tag{1}$$

$$Salario_i = \beta_0 + \beta_1 Escolaridad_i + \beta_2 Escolaridad_i^2 + \beta_3 Escolaridad_i^3 + u_i \tag{2}$$

a) Calcule para cada modelo el efecto que tiene un año más de escolaridad en el salario dejando lo demás constante.
b) Identifique para cada uno de ellos si hay linealidad en parámetros y en regresores. ¿Cómo podemos saber si hay linealidad en cada elemento?

Respuesta

a)
Para (1) tenemos que

$$\frac{\partial\ Salario_i}{\partial\ Escolaridad_i} = \beta_1$$

Para (2) tenemos que

$$\frac{\partial\ Salario_i}{\partial\ Escolaridad_i} = \beta_1 + 2\beta_2 Escolaridad_i + 3\beta_3 Escolaridad_i^2$$

b)
Tenemos que ambos modelos son lineales en parámetros, pero solo el segundo es no lineal en regresores.

¿Cómo sabemos si un modelo es lineal en parámetros? Derivando el modelo respecto a cada uno de los parámetros, si alguna derivada me da como resultado un término con algún parámetro, el modelo será no lineal en parámetros.

¿Cómo sabemos si un modelo es lineal en regresores? Derivando el modelo respecto a cada uno de los regresores, si alguna derivada me da como resultado un término con algún regresor, el modelo será no lineal en regresores.

Matemático 1

Suponga que tiene el siguiente modelo de regresión.

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

Escriba el problema a minimizar para calcular los estimadores de mínimos cuadrados ordinarios (MCO), y luego obténgalos.

Respuesta

Propiedades Importantes de Sumatorias:

$\sum_{i=1}^{T} Constante = T \cdot constante \Rightarrow (1)$

$\sum_{i=1}^{T}(x_i - \overline{x})(y_i - \overline{y}) = \sum_{i=1}^{T}(x_i - \overline{x})y_i = \sum_{i=1}^{T}(y_i - \overline{y})x_i \Rightarrow (2)$

$$\min_{\{\beta_0, \beta_1\}} \sum_{i=1}^{T} u_i^2 = \sum_{i=1}^{T}(y_i - \beta_0 - \beta_1 x_i)^2$$

$$\frac{\partial \sum_{i=1}^{T} u_i^2}{\partial \beta_0}|_{\beta=\hat{\beta}} = -2\sum_{i=1}^{T}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\sum_{i=1}^{T}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\sum_{i=1}^{T} y_i - \sum_{i=1}^{T} \hat{\beta}_0 - \sum_{i=1}^{T} \hat{\beta}_1 x_i = 0$$

$$\sum_{i=1}^{T} y_i - T * \hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{T} x_i = 0$$

$$T * \hat{\beta}_0 = \sum_{i=1}^{T} y_i - \hat{\beta}_1 \sum_{i=1}^{T} x_i$$

$$\hat{\beta}_o = \sum_{i=1}^{T} \frac{y_i}{T} - \hat{\beta}_1 \sum_{i=1}^{T} \frac{x_i}{T}$$

$$\boxed{\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}} \qquad (3)$$

$$\frac{\partial \sum_{i=1}^{T} u_i^2}{\partial \beta_1}|_{\beta=\hat{\beta}} = -2\sum_{i=1}^{T}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)x_i = 0$$

$$\sum_{i=1}^{T} x_i(y_i - (\overline{y} - \hat{\beta}_1 \overline{x}) - \hat{\beta}_1 x_i) = 0$$

$$\sum_{i=1}^{T} x_i[(y_i - \overline{y}) - \hat{\beta}_1(x_i - \overline{x})] = 0$$

$$\sum_{i=1}^{T} x_i(y_i - \overline{y}) - \hat{\beta}_1 \sum_{i=1}^{T} x_i(x_i - \overline{x})] = 0$$

$$\sum_{i=1}^{T}(y_i - \overline{y})(x_i - \overline{x}) = \hat{\beta}_1 \sum_{i=1}^{T}(x_i - \overline{x})^2$$

$$\boxed{\hat{\beta}_1 = \frac{\sum_{i=1}^{T}(y_i - \overline{y})(x_i - \overline{x})}{\sum_{i=1}^{T}(x_i - \overline{x})^2}} \tag{4}$$

Matemático 2

Suponga que usted quiere estimar la relación que hay entre el peso y el tamaño que tienen los autos mediante el siguiente modelo de regresión.

$$Peso_i = \beta_0 + \beta_1 Centimetros_i + \beta_2 Metros_i + u_i$$

Para el cálculo de este modelo usted cuenta solo con 3 observaciones los que se detallan a continuación.
$Peso \Rightarrow (10, 15, 30)$
$Centímetros \Rightarrow (200, 250, 300)$
$Metros \Rightarrow (2, 2{,}5, 3)$

a) ¿Por qué el modelo no podrá ser estimado. ¿Cuál es el supuesto de regresión lineal que no se está cumpliendo?

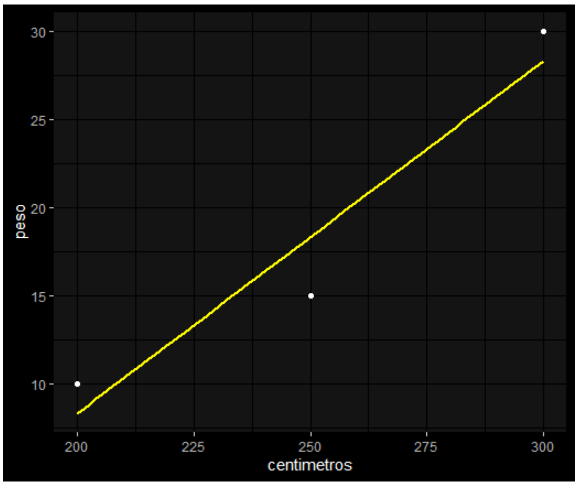
Respuesta

Este modelo no cumple con el supuesto de que no exista multicolinealidad perfecta. Tenemos que existe una combinación lineal perfecta entre la variable centímetros y la variable metros. Donde una unidad de la variable metros, corresponde a 100 unidades de la variable centímetros, para todas las observaciones. Esto hace que el $R_j^2 = 1$ (R cuadrado parcial). En otras palabras, hay una variable que no entrega ninguna información adicional al modelo dada las otras, existe una correlación perfecta entre centímetros y metros que hace irrelevante una de ellas. Esto más que un problema es un descuido, si nosotros tratamos de estimar el modelo nos saldrá un mensaje de alerta señalando que hay colinealidad perfecta, por lo que el computador procederá a eliminar una variable (típicamente la útima ingresada que tiene correlación perfecta).

b) El computador internamente eliminó la variable $Metros$ al tratar de estimar la regresión anterior. Ahora estimó el siguiente modelo

$$Peso_i = \beta_0 + \beta_1 Centímetros_i + u_i$$

Llegando a la siguiente gráfica



Calcule los estimadores de MCO.

Respuesta

$$\boxed{\hat{\beta}_1 = \frac{\sum_{i=1}^{T}(x_i - \overline{x})y_i}{\sum_{i=1}^{T}(x_i - \overline{x})^2}}$$

$$\boxed{\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x}}$$

$$\hat{\beta}_1 = \frac{(200-250)10 + (250-250)15 + (300-250)30}{(200-250)^2 + (250-250)^2 + (300-250)^2}$$

$$\boxed{\hat{\beta}_1 \approx 0{,}2}$$

$$\hat{\beta}_0 = 18{,}33 - 0{,}2 * 250$$

$$\boxed{\hat{\beta}_0 = -31{,}7}$$

c) Usando los estimadores encontrados anteriormente, se pudo escribir la siguiente tabla.

Observación	$Peso$	$\hat{Peso}$	$\hat{u}$
1	10	8.3	1.7
2	15	18.3	-3.3
3	30	28.3	1.7

Calcule el R^2 de este modelo.

Respuesta

Recordando que

$$\boxed{R^2 = \frac{\sum_{i=1}^{T}(\hat{y}_i - \overline{y})^2}{\sum_{i=1}^{T}(y_i - \overline{y})^2} = 1 - \frac{\sum_{i=1}^{T}\hat{u_i}^2}{\sum_{i=1}^{T}(y_i - \overline{y})^2}}$$

$$R^2 = \frac{(8{,}33 - 18{,}33)^2 + (18{,}33 - 18{,}33)^2 + (28{,}33 - 18{,}33)^2}{(10 - 18{,}33)^2 + (15 - 18{,}33)^2 + (30 - 18{,}33)^2}$$

$$R^2 = 1 - \frac{(1{,}67)^2 + (-3{,}33)^2 + (1{,}67)^2}{(10 - 18{,}33)^2 + (15 - 18{,}33)^2 + (30 - 18{,}33)^2}$$

$$\boxed{R^2 \approx 0{,}9231}$$

Econometría I
Pauta Ayudantía 4

Profesora: Giselli Castillo

Ayudante: Agustín Sanhueza.

Matemático 1

Suponga que usted quiere estimar la relación que hay entre el peso y el tamaño que tienen los autos mediante el siguiente modelo de regresión.

$$Peso_i = \beta_0 + \beta_1 Centimetros_i + \beta_2 Metros_i + u_i$$

Para el cálculo de este modelo usted cuenta solo con 3 observaciones los que se detallan a continuación.
$Peso \Rightarrow (10, 15, 30)$
$Centímetros \Rightarrow (200, 250, 300)$
$Metros \Rightarrow (2, 2{,}5, 3)$

a) ¿Por qué el modelo no podrá ser estimado. ¿Cuál es el supuesto de regresión lineal que no se está cumpliendo?

Respuesta

Este modelo no cumple con el supuesto de que no exista multicolinealidad perfecta. Tenemos que existe una combinación lineal perfecta entre la variable centímetros y la variable metros. Donde una unidad de la variable metros, corresponde a 100 unidades de la variable centímetros, para todas las observaciones.
En otras palabras, hay una variable que no entrega ninguna información adicional al modelo dada las otras. Existe una correlación perfecta entre centímetros y metros que hace irrelevante una de ellas. Algebráicamente significa que la varianza de los estimadores de β_1 y β_2 serán infinitas, por lo que no podrán ser estimados.
Esto más que un problema es un descuido, si tratamos de estimar el modelo por computador nos saldrá un mensaje de alerta señalando que hay colinealidad perfecta, por lo que el computador procederá a eliminar una variable (típicamente la útima ingresada que tiene correlación perfecta).

En general habrá un problema de multicolinealidad perfecta cuando usamos como variables lo mismo con distinta escala (distancias, pesos, tipo de cambio), cuando tenemos en el modelo algo que es el complemento de lo otro (género, proporciones), o cuando ocupamos una variable en logaritmo con distintos grados.
Ejemplo de esto último:

$$y_i = \beta_0 + \beta_1 ln(x_i) + \beta_2 ln(x_i^2) + u_i$$

Usando propiedad de logaritmos llegamos a que

$$y_i = \beta_0 + \beta_1 ln(x_i) + \beta_2 2ln(x_i) + u_i$$

Lo que nos lleva al problema de colinealidad perfecta.

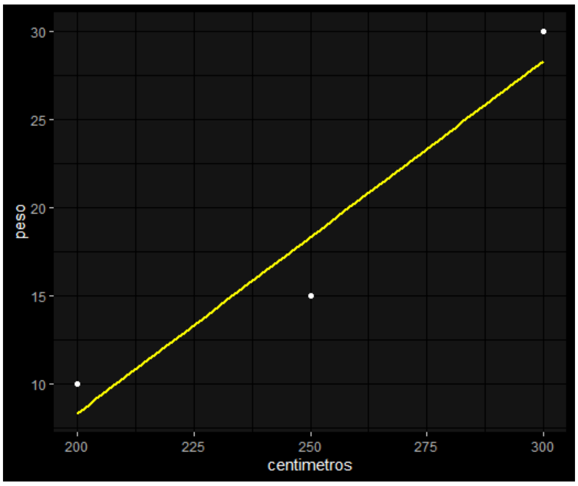
Lo que hay que hacer en estos casos es estimar

$$y_i = \beta_0 + \beta_1 ln(x_i) + \beta_2 ln(x_i)^2 + u_i$$

b) El computador internamente eliminó la variable $Metros$ al tratar de estimar la regresión anterior. Ahora estimó el siguiente modelo

$$Peso_i = \beta_0 + \beta_1 Centímetros_i + u_i$$

Llegando a la siguiente gráfica



Calcule los estimadores de MCO.

Respuesta

$$\boxed{\hat{\beta}_1 = \frac{\sum_{i=1}^{T}(x_i - \overline{x})y_i}{\sum_{i=1}^{T}(x_i - \overline{x})^2}}$$

$$\boxed{\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}}$$

$$\hat{\beta}_1 = \frac{(200-250)10 + (250-250)15 + (300-250)30}{(200-250)^2 + (250-250)^2 + (300-250)^2}$$

$$\boxed{\hat{\beta}_1 \approx 0{,}2}$$

$$\hat{\beta}_0 = 18{,}33 - 0{,}2 * 250$$

$$\boxed{\hat{\beta}_0 = -31{,}7}$$

c) Usando los estimadores encontrados anteriormente, se pudo escribir la siguiente tabla.

Observación	$Peso$	$\hat{Peso}$	$\hat{u}$
1	10	8.3	1.7
2	15	18.3	-3.3
3	30	28.3	1.7

Calcule el R^2 de este modelo.

Respuesta

Recordando que

$$\boxed{R^2 = \frac{\sum_{i=1}^{T}(\hat{y}_i - \overline{y})^2}{\sum_{i=1}^{T}(y_i - \overline{y})^2} = 1 - \frac{\sum_{i=1}^{T}\hat{u_i}^2}{\sum_{i=1}^{T}(y_i - \overline{y})^2}}$$

$$R^2 = \frac{(8{,}33 - 18{,}33)^2 + (18{,}33 - 18{,}33)^2 + (28{,}33 - 18{,}33)^2}{(10 - 18{,}33)^2 + (15 - 18{,}33)^2 + (30 - 18{,}33)^2}$$

$$R^2 = 1 - \frac{(1{,}67)^2 + (-3{,}33)^2 + (1{,}67)^2}{(10 - 18{,}33)^2 + (15 - 18{,}33)^2 + (30 - 18{,}33)^2}$$

$$\boxed{R^2 \approx 0{,}9231}$$

Matemático 2

Supongamos que el modelo real es:

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + u_i \tag{1}$$

Pero usted estimó el siguiente

$$y_i = \gamma_0 + \gamma_1 x_{1,i} + e_i \tag{2}$$

Por lo tanto $e_i = \beta_2 x_{2,i} + u_i$ (la variable x_2 estará en el término de error del modelo estimado)

1. ¿Cuál es la $\mathbb{E}[\hat{\gamma}_1]$?

Respuesta

De ayudantías pasadas sabemos que al estimar el modelo

$$y_i = \gamma_0 + \gamma_1 x_{1,i} + e_i$$

Llegaremos a que:
$\hat{\gamma}_0 = \overline{y} - \hat{\gamma}_1 \overline{x}$

$\hat{\gamma}_1 = \frac{Cov(X_1,Y)}{Var(X_1)}$

Lo que nos interesa es analizar $\hat{\gamma}_1$, por lo que trabajaremos con este término.

$$\hat{\gamma}_1 = \frac{\sum_{i=1}^{T}(x_{1,i} - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{T}(x_{1,i} - \overline{x})^2}$$

Reemplazamos y_i y el $\overline{y}$ del modelo real.

$$\hat{\gamma}_1 = \frac{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})[(\beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + u_i) - (\beta_0 + \beta_1\overline{x_1} + \beta_2\overline{x_2} + \overline{u})]}{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2}$$

$$\hat{\gamma}_1 = \frac{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})[\beta_1 x_{1,i} - \beta_1\overline{x_1} + \beta_2 x_{2,i} - \beta_2\overline{x_2} + u_i - \overline{u}]}{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2}$$

$$\hat{\gamma}_1 = \frac{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})[\beta_1(x_{1,i} - \overline{x_1}) + \beta_2(x_{2,i} - \overline{x_2}) + u_i]}{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2}$$

$$\hat{\gamma}_1 = \frac{\beta_1\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2 + \beta_2\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})(x_{2,i} - \overline{x_2}) + \sum_{i=1}^{T}(x_{1,i} - \overline{x_1})u_i]}{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2}$$

$$\hat{\gamma}_1 = \beta_1 + \frac{\beta_2\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})(x_{2,i} - \overline{x_2})}{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2} + \frac{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})u_i]}{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2}$$

Aplicamos esperanza

$$\mathbb{E}[\hat{\gamma}_1] = \mathbb{E}[\beta_1] + \mathbb{E}\left[\frac{\beta_2\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})(x_{2,i} - \overline{x_2})}{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2}\right] + \mathbb{E}\left[\frac{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})u_i]}{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2}\right]$$

$$\mathbb{E}[\hat{\gamma}_1] = \beta_1 + \frac{\beta_2\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})(x_{2,i} - \overline{x_2})}{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2} + \frac{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})\mathbb{E}[u_i]}{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2}$$

$$\boxed{\mathbb{E}[\hat{\gamma}_1] = \beta_1 + \underbrace{\beta_2\frac{Cov(X_1, X_2)}{Var(X_1)}}_{Sesgo}}$$

2. ¿Cuáles son las condiciones para tener un problema de "variable relevante omitida"?

Respuesta

Para tener un problema de "variable relevante omitida", es necesario que se cumplan las siguientes condiciones.
- $\beta_2 \neq 0$

- $Cov(X_1, X_2) \neq 0$

La primera condición significa que la variable X_2 si afecta a la variable Y (sería una variable "relevante").
La segunda condición significa que existe una correlación entre la variable X_1 y X_2. Esto rompe con el supuesto de la media condicional nula. Al tener una correlación entre la variable incluida y omitida del modelo, se tendrá que $E[e_i/x_i] \neq 0$.
$\therefore$ Si se cumplen estas dos condiciones, la estimación de $\hat{\gamma_1}$ sufrirá de un sesgo.

La dirección del sesgo depende del signo que tomen ambos elementos. Se dice que el estimador $\hat{\gamma_1}$ está "sobrevalorado" si el sesgo es positivo, y está "subvalorado" si el sesgo es negativo. Esto se resume en la siguiente tabla:

	$Cov(X_1, X_2) > 0$	$Cov(X_1, X_2) < 0$
$\beta_2 > 0$	Sobrevalorado	Subvalorado
$\beta_2 < 0$	Subvalorado	Sobrevalorado

Suponga que el modelo real es

$$Salario_i = \beta_0 + \beta_1 Escolaridad_i + \beta_2 Habilidad_i + u_i$$

Usted estimó este:

$$Salario_i = \gamma_0 + \gamma_1 Escolaridad_i + e_i$$

3. Comente si cree que $\hat{\gamma_1}$ está sobrevalorando o subvalorando el efecto que tiene la escolaridad en los salarios.

Respuesta

En este caso tendremos que

$$\boxed{\mathbb{E}[\hat{\gamma_1}] = \beta_1 + \beta_2 \frac{Cov(Escolaridad, Habilidad)}{Var(Escolaridad)}}$$

Podemos decir que las personas más hábiles tienen mayor productividad para el trabajo, lo que se ve reflejado en mayores salarios. Por lo que podemos esperar que $\beta_2 > 0$. Por otro lado, hay razones para pensar que las personas con más habilidad eligen niveles de educación más altos. Por lo que podemos esperar $Cov(Escolaridad, Habilidad) > 0$.

Por lo tanto tendríamos que el estimador $\hat{\gamma_1}$ está sobrevalorando el efecto que tiene la escolaridad en los salarios, dado que estamos omitiendo el efecto de la habilidad.
En otras palabras, no podemos distinguir cuánto de la escolaridad efectivamente afecta a los salarios, y cuánto de la habilidad, dado que estamos atribuyéndole erróneamente todo a lo primero.
Si estimásemos el modelo correcto, veríamos reflejado en los coeficientes $\hat{\beta}_1$ y $\hat{\beta}_2$ el impacto real de cada elemento en el salario.

Econometría I
Pauta Ayudantía 5

Profesora: Giselli Castillo

Ayudante: Agustín Sanhueza

Matemático

Supongamos que el modelo real es:

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + u_i \tag{1}$$

Pero usted estimó el siguiente

$$y_i = \gamma_0 + \gamma_1 x_{1,i} + e_i \tag{2}$$

Por lo tanto $e_i = \beta_2 x_{2,i} + u_i$ (la variable x_2 estará en el término de error del modelo estimado)

1. ¿Cuál es la $\mathbb{E}[\hat{\gamma_1}]$?

Respuesta

De ayudantías pasadas sabemos que al estimar el modelo

$$y_i = \gamma_0 + \gamma_1 x_{1,i} + e_i$$

Llegaremos a que:
$\hat{\gamma_0} = \overline{y} - \hat{\gamma_1}\overline{x}$

$\hat{\gamma_1} = \frac{Cov(X_1,Y)}{Var(X_1)}$

Lo que nos interesa es analizar $\hat{\gamma_1}$, por lo que trabajaremos con este término.

$$\hat{\gamma_1} = \frac{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})(y_i - \overline{y})}{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2}$$

Reemplazamos y_i y el $\overline{y}$ del modelo real.

$$\hat{\gamma_1} = \frac{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})[(\beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + u_i) - (\beta_0 + \beta_1\overline{x_1} + \beta_2\overline{x_2} + \overline{u})]}{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2}$$

$$\hat{\gamma_1} = \frac{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})[\beta_1 x_{1,i} - \beta_1\overline{x_1} + \beta_2 x_{2,i} - \beta_2\overline{x_2} + u_i - \overline{u}]}{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2}$$

$$\hat{\gamma_1} = \frac{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})[\beta_1(x_{1,i} - \overline{x_1}) + \beta_2(x_{2,i} - \overline{x_2}) + u_i]}{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2}$$

$$\hat{\gamma_1} = \frac{\beta_1\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2 + \beta_2\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})(x_{2,i} - \overline{x_2}) + \sum_{i=1}^{T}(x_{1,i} - \overline{x_1})u_i]}{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2}$$

$$\hat{\gamma}_1 = \beta_1 + \frac{\beta_2 \sum_{i=1}^{T}(x_{1,i} - \overline{x_1})(x_{2,i} - \overline{x_2})}{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2} + \frac{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})u_i]}{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2}$$

Aplicamos esperanza

$$\mathbb{E}[\hat{\gamma}_1] = \mathbb{E}[\beta_1] + \mathbb{E}\left[\frac{\beta_2 \sum_{i=1}^{T}(x_{1,i} - \overline{x_1})(x_{2,i} - \overline{x_2})}{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2}\right] + \mathbb{E}\left[\frac{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})u_i]}{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2}\right]$$

$$\mathbb{E}[\hat{\gamma}_1] = \beta_1 + \frac{\beta_2 \sum_{i=1}^{T}(x_{1,i} - \overline{x_1})(x_{2,i} - \overline{x_2})}{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2} + \frac{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})\mathbb{E}[u_i]}{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2}$$

$$\boxed{\mathbb{E}[\hat{\gamma}_1] = \beta_1 + \underbrace{\beta_2 \frac{Cov(X_1, X_2)}{Var(X_1)}}_{Sesgo}}$$

2. ¿Cuáles son las condiciones para tener un problema de "variable relevante omitida"?

Respuesta

Para tener un problema de "variable relevante omitida", es necesario que se cumplan las siguientes condiciones.
- $\beta_2 \neq 0$
- $Cov(X_1, X_2) \neq 0$

La primera condición significa que la variable X_2 si afecta a la variable Y (sería una variable "relevante").
La segunda condición significa que existe una correlación entre la variable X_1 y X_2. Esto rompe con el supuesto de la media condicional nula. Al tener una correlación entre la variable incluida y omitida del modelo, se tendrá que $E[e_i/x_i] \neq 0$.
$\therefore$ Si se cumplen estas dos condiciones, la estimación de $\hat{\gamma}_1$ sufrirá de un sesgo.

La dirección del sesgo depende del signo que tomen ambos elementos. Se dice que el estimador $\hat{\gamma}_1$ está "sobrevalorado" si el sesgo es positivo, y está "subvalorado" si el sesgo es negativo. Esto se resume en la siguiente tabla:

	$Cov(X_1, X_2) > 0$	$Cov(X_1, X_2) < 0$
$\beta_2 > 0$	Sobrevalorado	Subvalorado
$\beta_2 < 0$	Subvalorado	Sobrevalorado

3. Compare las varianzas de los estimadores $\hat{\beta}_1$ y $\hat{\gamma}_1$

Respuesta

Para el caso univariado

$$y_i = \gamma_0 + \gamma_1 x_{1,i} + e_i$$

La varianza del estimador $\hat{\gamma_1}$ es

$$Var(\hat{\gamma_1}) = \frac{\sigma^2}{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2}$$

Notar que

$$\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2 = SST_1$$

Por lo que podemos reescribir el término de la varianza como

$$\boxed{Var(\hat{\gamma_1}) = \frac{\sigma^2}{SST_1}}$$

Para el caso multivariado

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + u_i$$

La varianza del estimador $\hat{\beta_1}$ será

$$\boxed{Var(\hat{\beta_1}) = \frac{\sigma^2}{SST_1(1 - R_1^2)}}$$

Donde R_1^2 es el "R^2 parcial" de la variable 1. Este se obtiene dejando la variable 1 como variable dependiente, y como regresoras las otras variables explicativas. En este caso, R_1^2 es el R^2 del siguiente modelo.

$$x_{1,i} = \delta_0 + \delta_1 x_{2,i} + \varepsilon_i$$

Esto es análogo en un modelo con más variables explicativas. En el límite, cuando el R^2 parcial es 1, entonces la varianza del estimador será infinito, y el modelo no podrá ser estimado (multicolinealidad perfecta).

Si comparamos las varianzas de los estimadores llegaremos a que:

$$\boxed{Var(\hat{\beta_1}) \geq Var(\hat{\gamma_1})}$$

$$\boxed{\frac{\sigma^2}{SST_1(1 - R_1^2)} \geq \frac{\sigma^2}{SST_1}}$$

Notar que ambas varianzas serán iguales, solo si no existe una relación entre las variables X_1 y X_2, en cualquier otro caso, la varianza del estimador que tiene el modelo con más variables será mayor.

El punto es que el agregar más variables a nuestro modelo, genera menor precisión en la estimación de los coeficientes, por lo que el agregar una variable más "no sale gratis".

4. Señale cuál es el factor de inflación de varianza (VIF). Explique qué captura este indicador y desde qué valor se dice que hay un problema de multicolinealidad.

Respuesta

Para un modelo multivariado

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + ... + \beta_k x_{k,i} + u_i$$

Podemos escribir la varianza del estimador $\hat{\beta}_j$ como

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)}$$

Reescribimos

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j} \cdot \frac{1}{(1 - R_j^2)}$$

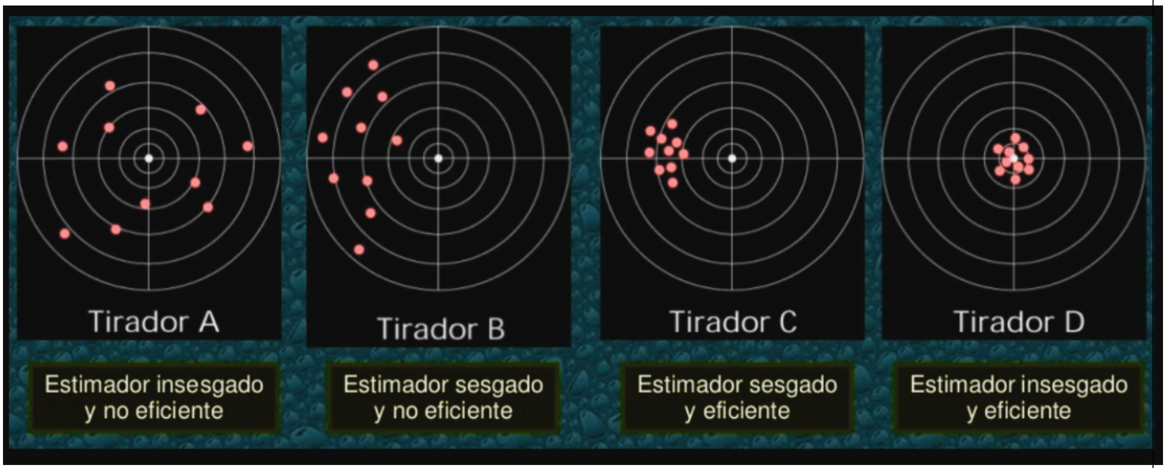
$$\boxed{Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j} \cdot VIF_j}$$

VIF_j captura la colinealidad que hay entre la variable j, y todas las otras del modelo. Se dice que una variable sufre de un gran problema de colinealidad cuando $VIF_j > 10$ (corte de convención), lo que es equivalente a un $R_j^2 > 0{,}9$. Esto no significa que variables con un $VIF_j > 10$ deban ser eliminados del modelo, solo hay que tener presente que esta variable correlaciona mucho con las otras haciendo menos eficiente la estimación. ¿Opciones?
- Omitir esta variable del modelo (el peligro es generar sesgo si esta variable era relevante)
- Aumentar el tamaño muestral. Así tendremos un mayor SST_j y por lo tanto una menor $Var(\hat{\beta}_j)$, lo que podría hacer el problema de multicolinealidad algo no tan grave.
- Aceptar el hecho de que la muestra obtenida no es buena para estimar el modelo. No hay necesidad de llorar sobre la leche derramada, la vida a veces es dura y no hay nada que hacerle.
- Usar estimadores alternativos como el de Ridge el cual por construcción es sesgado, pero disminuye la varianza (No mencionar este último a los alumnos).

5. Explique el trade-off que hay entre la precisión y la eficiencia en un estimador. ¿Cuál métrica podríamos usar para elegir entre estimadores tomando en cuenta estos elementos?

Respuesta

Podemos pensar la precisión y la eficiencia de un estimador con la siguiente imagen



El punto blanco representa el parámetro poblacional. Cada punto rojo representa los estimadores de este parámetro obtenidos en ejercicios de remuestreo (con repetición). Los 4 escenarios que se presentan son con con estimadores que (no) tienen sesgo y/o eficiencia. En el mundo ideal quisieramos tener un estimador insesgado y eficiente (último cuadro). En la realidad los estimadores presentan un atributo en desmedro del otro.

Si queremos rankear estimadores tomando en cuenta la precisión y la eficiencia, podemos usar el Error Cuadrático Medio (ECM).

$$ECM(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2]$$

Se podría demostrar que se llega al siguiente término (no demostrar, solo dejarles expresado)

$$\boxed{ECM(\hat{\theta}) = Var(\hat{\theta}) + Sesgo(\hat{\theta})^2}$$

Bajo esta métrica, los mejores estimadores serán aquellos que presenten un error cuadrático medio menor.

Econometría I
Pauta Ayudantía 6

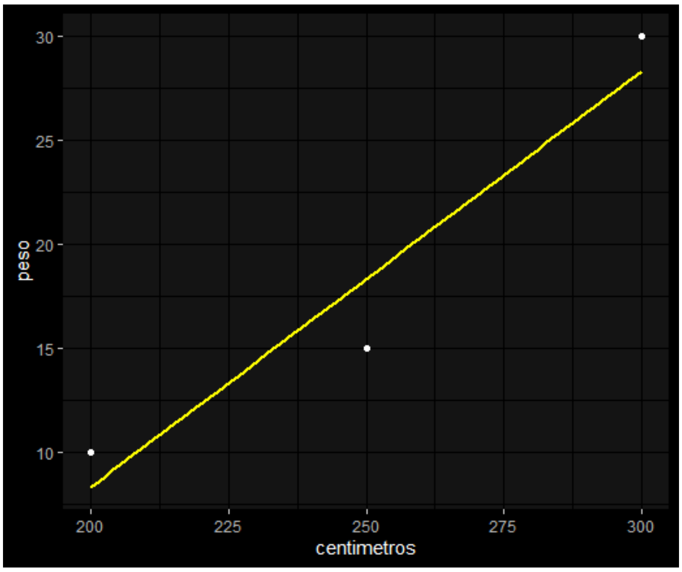
Profesora: Giselli Castillo
Ayudante: Agustín Sanhueza.

Matemático

De ayudantías anteriores estimamos el siguiente modelo

$$Peso_i = \beta_0 + \beta_1 Centímetros_i + u_i$$

Llegando a la siguiente gráfica y resultados



Observación	Centímetros	$Peso$	$\hat{Peso}$	$\hat{u}$
1	200	10	8.3	1.7
2	250	15	18.3	-3.3
3	300	30	28.3	1.7

$$\hat{Peso}_i = -31{,}7 + 0{,}2 \cdot Centímetros_i$$
$$R^2 = 0{,}9231$$

a) Calcule la desviación estándar de los estimadores $\hat{\beta}_0$ y $\hat{\beta}_1$

Respuesta

La varianza de los estimadores en este modelo univariado son:

$$Var(\hat{\beta}_0) = \frac{\sigma^2 \sum_{i=1}^{T} x_i^2}{T \cdot \sum_{i=1}^{T}(x_i - \overline{x})^2}$$

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{T}(x_i - \overline{x})^2}$$

(No demostrar esto, son un corcho enorme)

Notar que no conocemos cuánto es la varianza del error σ^2, por lo tanto debemos estimarla de la siguiente manera.

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{T} \hat{u_i}^2}{T-K-1}$$

Donde T es el número de observaciones, K es el número de regresores y el "1" cuenta a la constante. El denominador entonces nos señala los "grados de libertad" que tiene el modelo. Si estimamos la varianza del error dividiendo solo por el número de observaciones llegaremos a un estimador sesgado (solo mencionar, no demostrar).

Notar que a medida que la muestra se hace arbitrariamente grande, el ajuste por grados de libertad se hace cada vez menos relevante.

$$\hat{\sigma}^2 = \frac{1{,}7^2 + (-3{,}3)^2 + 1{,}7^2}{3-1-1}$$

$$\boxed{\hat{\sigma}^2 \approx 16{,}67}$$

Calculamos las varianzas de los estimadores.

$$Var(\hat{\beta}_0) = \frac{16{,}67 \cdot [200^2 + 250^2 + 300^2]}{3 \cdot [(200-250)^2 + (250-250)^2 + (300-250)^2]}$$

$$Var(\hat{\beta}_0) = \frac{3208975}{15000}$$

$$Var(\hat{\beta}_0) \approx 213{,}93$$

Por lo tanto

$$\boxed{sd(\hat{\beta}_0) \approx 14{,}63}$$

$$Var(\hat{\beta}_1) = \frac{16{,}67}{(200-250)^2 + (250-250)^2 + (300-250)^2}$$

$$Var(\hat{\beta}_1) = \frac{16{,}67}{5000}$$

$$Var(\hat{\beta}_1) = 0{,}003334$$

$$\boxed{sd(\hat{\beta}_1) \approx 0{,}0577}$$

b) Calcule el t_{obs} de cada estimador, con una $H_0 : \beta_j = 0$, señale como se distribuye este estadístico.

Respuesta

Para el caso general tendremos que

$$\boxed{t_{\hat{\beta}_j} = \frac{\hat{\beta}_j - \beta_j}{sd(\hat{\beta}_j)} \sim t_{t-k-1}}$$

Para $\hat{\beta}_0$

$$t_{\hat{\beta}_0} = \frac{-31{,}7}{14{,}63} \approx -2{,}17$$

$$\boxed{t_{\hat{\beta}_0} = -2{,}17 \sim t_1}$$

Para $\hat{\beta}_1$

$$t_{\hat{\beta}_1} = \frac{0{,}2}{0{,}0577} \approx 3{,}46$$

$$\boxed{t_{\hat{\beta}_1} = 3{,}46 \sim t_1}$$

c) Testee $H_0 : \beta_j = 0$ y $H_1 : \beta_j \neq 0$ Use la tabla que se encuentra en Anexo para responder esta pregunta considerando un nivel de significancia (α) del 5 %

Respuesta

Recordemos que los estadísticos t obtenidos del item anterior distribuyen como una t de Student con solo un grado de libertad, por lo que debemos mirar la primera fila de la tabla. Luego, nos dicen que debemos considerar un nivel de significancia del 5 %. Por lo tanto debemos ver la columna con un $\alpha = 0{,}025$ dado que estamos aplicando un test de dos colas. Entonces tendremos que $t_{crítico=12{,}7062}$

Se dice que el parámetro j es "estadísticamente significativo" si

$$\boxed{|t_{\hat{\beta}_j}| > t_{crítico}}$$

Tenenemos que

$$|t_{\hat{\beta}_0}| = |-2{,}17| = 2{,}17$$

$$|t_{\hat{\beta}_1}| = |3{,}46| = 3{,}46$$

Dado que ninguno de los valores es mayor a 12.7062, se dice que $\hat{\beta}_0$ y $\hat{\beta}_1$ no son estadísticamente significativos.

d) Responda el item anterior pero ahora tomando como $H_0 : \beta_j \leq 0$ y $H_1 : \beta_j > 0$

Respuesta

Lo único que cambió respecto al item anterior es el $t_{crítico}$ ya que no estamos en un test de dos colas. Viendo la tabla en Anexo ahora debemos quedarnos con la columna de $\alpha = 0{,}05$. Ahora tenemos que el $t_{crítico=6{,}3138}$

Dado que ninguno de los $|t_{\hat{\beta}_j}|$ es mayor que el valor crítico, tampoco podemos rechazar la hipótesis nula de que $\beta_j \leq 0$

e) Calcule los intervalos de confianza de cada estimador, considerando un nivel de significancia del 5 %

Respuesta

Para el caso general tendremos que:

$$\boxed{\hat{\beta}_j - c_\alpha \cdot se(\hat{\beta}_j) \leq \beta_j \leq \hat{\beta}_j + c_\alpha \cdot se(\hat{\beta}_j)}$$

$$\boxed{Pr[\hat{\beta}_j - c_\alpha \cdot se(\hat{\beta}_j) \leq \beta_j \leq \hat{\beta}_j + c_\alpha \cdot se(\hat{\beta}_j)] = 1\text{-}\alpha}$$

Dado que estamos usando un nivel de significancia del 5 % y tenemos solo 1 grado de libertad tendremos que $c_{0{,}05} = 12{,}7062$

Por lo tanto

$$-31{,}7 - 12{,}7062 \cdot 14{,}63 \leq \beta_0 \leq -31{,}7 + 12{,}7062 \cdot 14{,}63$$

$$\boxed{-217,59 \leq \beta_0 \leq 154,19}$$

$$0{,}2 - 12{,}7062 \cdot 0{,}0577 \leq \beta_1 \leq 0{,}2 + 12{,}7062 \cdot 0{,}0577$$

$$\boxed{-0{,}53 \leq \beta_1 \leq 0,93}$$

Notar que ambos intervalos pasan por el valor de 0. Debido a esto es que no podemos rechazar la hipótesis de que los parámetros poblacionales β_0 y β_1 sean 0 con una confianza del 95 %

f) Muestre gráficamente los p_{value} de $\hat{\beta}_0$ y $\hat{\beta}_1$ (donde $H_0 : \beta_j = 0$ y $H_1 : \beta_j \neq 0$), apóyese en el siguiente Link para su cálculo.

Respuesta

En el caso general para $H_0 : \beta_j = 0$ y $H_1 : \beta_j \neq 0$ tendremos

$$\boxed{P_{value} = Pr\left[|t_{obs}| > |t_{\hat{\beta}_j}|\right]}$$

Para $\hat{\beta}_0$ tendremos

$$P_{value} = Pr\left[|t_{obs}| > |-2{,}17|\right]$$

$$P_{value} = Pr\left[|t_{obs}| > 2{,}17\right]$$
$$P_{value} = Pr\left[t_{obs} > 2{,}17\right] + Pr\left[t_{obs} < -2{,}17\right]$$

Por simetría de la distribución de probabilidad tendremos

$$P_{value} = 2 \cdot Pr\left(t_{obs} > 2{,}17\right)$$
$$P_{value} = 2\left[1 - Pr\left(t_{obs} < 2{,}17\right)\right]$$
$$\boxed{P_{value} = 0{,}275}$$

Para $\hat{\beta}_1$ tendremos

$$P_{value} = Pr\left[|t_{obs}| > 3{,}46\right]$$
$$P_{value} = Pr\left[t_{obs} > 3{,}46\right] + Pr\left[t_{obs} < -3{,}46\right]$$

Por simetría de la distribución de probabilidad tendremos

$$P_{value} = 2 \cdot Pr\left(t_{obs} > 3{,}46\right)$$
$$P_{value} = 2\left[1 - Pr\left(t_{obs} < 3{,}46\right)\right]$$
$$\boxed{P_{value} = 0{,}179}$$

Dado que ambos valores son mayores a 0.05, se dice que ambos parámetros no son estadísticamente significativos.

Analítico

Considere el siguiente modelo que busca predecir el salario de los jugadores de la liga de baseball

$$log(salario_i) = \beta_0 + \beta_1 años_i + \beta_2 juegos_i + \beta_3 bateo_i + \beta_4 hruns_i + \beta_5 carreras_i + u_i$$

Donde
Años: Número de años en la liga
Juegos : Número de juegos hasta ahora
Bateo: Número de bateos hasta ahora
hruns: Número de Homeruns hechos hasta ahora
Carreras: Número de carreras hechas de una base a otra hasta ahora.

La recta de regresión encontrada es la siguiente

$$log(\hat{salario_i}) = \underset{(0{,}29)}{11{,}19} + \underset{(0{,}0121)}{0{,}0689} \cdot años_i + \underset{(0{,}0026)}{0{,}0126} \cdot juegos_i + \underset{(0{,}00110)}{0{,}00098} \cdot bateo_i + \underset{(0{,}0161)}{0{,}0144} \cdot hruns_i + \underset{(0{,}0072)}{0{,}0108} \cdot carreras_i$$

Con:
T=353
SSR=183,186
$R^2 = 0{,}6278$
Los números en paréntesis debajo de los estimadores corresponden a la desviación estándar de estos.

1. Señale si los estimadores son estadísticamente significativos (distintos de 0), considere un nivel de significancia del 5 %

> **Respuesta**
>
> Dado que tenemos 353 observaciones, podemos apoyarnos en la teoría asintótica y usar el valor crítico de 1.96 (dado que es un test de 2 colas para un nivel de significancia del 5 %)
>
> $$t_{\hat{\beta}_0} = \frac{11{,}19}{0{,}29} = 38{,}59$$
> $$t_{\hat{\beta}_1} = \frac{0{,}0689}{0{,}0121} = 5{,}69$$
> $$t_{\hat{\beta}_2} = \frac{0{,}0126}{0{,}0026} = 4{,}85$$
> $$t_{\hat{\beta}_3} = \frac{0{,}00098}{0{,}00110} = 0{,}89$$
> $$t_{\hat{\beta}_4} = \frac{0{,}0144}{0{,}0161} = 0{,}89$$
> $$t_{\hat{\beta}_5} = \frac{0{,}0108}{0{,}0072} = 1{,}5$$
>
> Tenemos entonces que $\hat{\beta}_0$, $\hat{\beta}_1$ y $\hat{\beta}_2$ son estadísticamente significativos.

Con el fin de testear $H_0 : \beta_3 = \beta_4 = \beta_5 = 0$ Ahora usted tiene la siguiente recta de regresión (este sería el modelo restringido)

$$\widehat{log(salario_i)} = \underset{(0{,}11)}{11{,}22} + \underset{(0{,}0125)}{0{,}0713} \cdot años_i + \underset{(0{,}0013)}{0{,}0202} \cdot juegos_i$$

Con:
T=353
SSR=198,311
$R^2 = 0{,}5971$
Los números en paréntesis debajo de los estimadores corresponden a la desviación estándar de estos.

2. Testee la hipótesis conjunta de que $H_0 : \beta_3 = \beta_4 = \beta_5 = 0$. Calcule el F_{obs} y muestre gráficamente el P_{value}.

> **Respuesta**
>
> Debemos usar el estadístico F
>
> $$\boxed{F_{obs} = \frac{(SSR_r - SSR_{ur})/q}{(SSR_{ur})/(T-K-1)} = \frac{(R^2_{ur} - R^2_r)/q}{(1-R^2_{ur})/(T-K-1)} \sim F_{q,t-k-1}}$$

$$F_{obs} = \frac{(198{,}311 - 183{,}186)/3}{183{,}186/(353-5-1)} = \frac{(0{,}6278 - 0{,}5971)/3}{(1-0{,}6278)/(353-5-1)} \sim F_{3,353-5-1}$$

$$F_{obs} = \frac{5{,}0416}{0{,}5279} = \frac{0{,}01023}{0{,}00107} \sim F_{3,347}$$

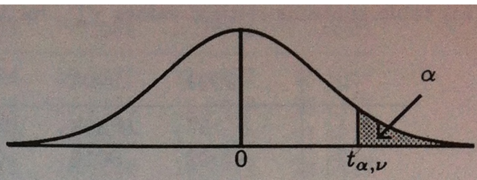
$$\boxed{F_{obs} = 9{,}55 \sim F_{3,347}}$$

Dado que 9.55 es mayor que el valor crítico al 1 % del nivel de significancia para $F_{3,347}$ rechazamos la hipótesis nula de que $H_0 : \beta_3 = \beta_4 = \beta_5 = 0$

Mostrar gráficamente la distribución F y recalcar la no negatividad de su distribución $SSR_r \geq SSR_{ur} \Leftrightarrow R^2_{ur} \geq R^2_r$

Anexo

This table contains critical values $t_{\alpha,\nu}$ for the t distribution defined by $P(T \geq t_{\alpha,\nu}) = \alpha$.



					α				
ν	.20	.10	.05	.025	.01	.005	.001	.0005	.0001
1	1.3764	3.0777	6.3138	12.7062	31.8205	63.6567	318.3088	636.6192	3183.0988
2	1.0607	1.8856	2.9200	4.3027	6.9646	9.9248	22.3271	31.5991	70.7001
3	.9785	1.6377	2.3534	3.1824	4.5407	5.8409	10.2145	12.9240	22.2037
4	.9410	1.5332	2.1318	2.7764	3.7469	4.6041	7.1732	8.6103	13.0337
5	.9195	1.4759	2.0150	2.5706	3.3649	4.0321	5.8934	6.8688	9.6776
6	.9057	1.4398	1.9432	2.4469	3.1427	3.7074	5.2076	5.9588	8.0248
7	.8960	1.4149	1.8946	2.3646	2.9980	3.4995	4.7853	5.4079	7.0634
8	.8889	1.3968	1.8595	2.3060	2.8965	3.3554	4.5008	5.0413	6.4420
9	.8834	1.3830	1.8331	2.2622	2.8214	3.2498	4.2968	4.7809	6.0101
10	.8791	1.3722	1.8125	2.2281	2.7638	3.1693	4.1437	4.5869	5.6938
11	.8755	1.3634	1.7959	2.2010	2.7181	3.1058	4.0247	4.4370	5.4528
12	.8726	1.3562	1.7823	2.1788	2.6810	3.0545	3.9296	4.3178	5.2633
13	.8702	1.3502	1.7709	2.1604	2.6503	3.0123	3.8520	4.2208	5.1106
14	.8681	1.3450	1.7613	2.1448	2.6245	2.9768	3.7874	4.1405	4.9850
15	.8662	1.3406	1.7531	2.1314	2.6025	2.9467	3.7328	4.0728	4.8800
16	.8647	1.3368	1.7459	2.1199	2.5835	2.9208	3.6862	4.0150	4.7909
17	.8633	1.3334	1.7396	2.1098	2.5669	2.8982	3.6458	3.9651	4.7144
18	.8620	1.3304	1.7341	2.1009	2.5524	2.8784	3.6105	3.9216	4.6480
19	.8610	1.3277	1.7291	2.0930	2.5395	2.8609	3.5794	3.8834	4.5899
20	.8600	1.3253	1.7247	2.0860	2.5280	2.8453	3.5518	3.8495	4.5385
21	.8591	1.3232	1.7207	2.0796	2.5176	2.8314	3.5271	3.8192	4.4929
22	.8583	1.3212	1.7171	2.0739	2.5083	2.8187	3.5050	3.7921	4.4520
23	.8575	1.3195	1.7139	2.0687	2.4999	2.8073	3.4850	3.7676	4.4152
24	.8569	1.3178	1.7109	2.0639	2.4922	2.7969	3.4668	3.7454	4.3819
25	.8562	1.3163	1.7081	2.0595	2.4851	2.7874	3.4502	3.7251	4.3517
26	.8557	1.3150	1.7056	2.0555	2.4786	2.7787	3.4350	3.7066	4.3240
27	.8551	1.3137	1.7033	2.0518	2.4727	2.7707	3.4210	3.6896	4.2987
28	.8546	1.3125	1.7011	2.0484	2.4671	2.7633	3.4081	3.6739	4.2754
29	.8542	1.3114	1.6991	2.0452	2.4620	2.7564	3.3962	3.6594	4.2539
30	.8538	1.3104	1.6973	2.0423	2.4573	2.7500	3.3852	3.6460	4.2340
40	.8507	1.3031	1.6839	2.0211	2.4233	2.7045	3.3069	3.5510	4.0942
50	.8489	1.2987	1.6759	2.0086	2.4033	2.6778	3.2614	3.4960	4.0140
60	.8477	1.2958	1.6706	2.0003	2.3901	2.6603	3.2317	3.4602	3.9621
120	.8446	1.2886	1.6577	1.9799	2.3578	2.6174	3.1595	3.3735	3.8372
∞	.8416	1.2816	1.6449	1.9600	2.3263	2.5758	3.0902	3.2905	3.7190

Econometría I
Ayudantía 7

Profesora: Giselli Castillo
Ayudante: Agustín Sanhueza.

Instalación

Vaya al siguiente link descargue e instale R[1]. Se recomienda bajar la interfaz "Rstudio" del siguiente link.

R como calculadora

1. Programe el cálculo de su nota promedio en el curso.
2. Calcule el valor presente de un flujo de efectivo de $1000 por un total de diez períodos. Asuma una tasa de descuento de 10 %.

$$VP = \sum_{t=1}^{10} \frac{1000}{(1+0{,}1)^t} = \frac{1000}{0{,}1}\left[1 - \frac{1}{(1+0{,}1)^{10}}\right]$$

Importación y Eliminación de bases

1. Instale y cargue el paquete *readxl*. Luego importe la base de datos "Sub-Casen" de formato excel, guardando este dataframe con el nombre "datos1".
2. Instale y cargue el paquete *wooldridge*. Luego importe la base de datos "wage1", guardando este dataframe con el nombre "datos2".
3. Instale y cargue el paquete *gapminder*. Luego importe la base de datos "gapminder", guardando este dataframe con el nombre "datos3".
4. Elimine el dataframe "datos1"
5. Cambie los nombres de los dataframes "datos2" y "datos3" por "salarios" y "mundo" respectivamente.

Manipulación de datos y Estadísticas descriptivas

1. Instale y cargue los paquetes *tidyverse* y *dplyr*.
2. De la base "salarios" seleccione las variables *lwage, wage, educ, exper y female*. Cambie el nombre de estas variables por *lnsalario, salario, educación, experiencia y mujer* respectivamente.
3. Calcule el salario, años de escolaridad, y años de experiencia promedios de la base (aproxime los resultados a 1 decimal). También calcule el número de mujeres y hombres en la base.

[1] Video guía para Windows. Video guía para Mac

4. Elimine la variable ”lnsalario” y genere una nueva variable llamada ”sueldos”, que tome valores aleatorios entre el salario mínimo chileno y los 10 millones
5. Genere un dataframe auxiliar llamado ”df_aux” en la que estén las mismas variables que en ”salarios”, solo que con las personas con menos de 6 años de escolaridad.

Gráficos, Regresión y Exportación

1. Instale y cargue los paquetes *ggplot2, gganimate, av, gifski, ggdark, y writexl*
2. Genere un histograma del salario.
3. Usando la base ”salarios” Calcule los betas, residuos y valores predichos del siguiente modelo de regresión.
$$salario_i = \beta_0 + \beta_1 escolaridad_i + \mu_i$$
Guarde los resultados en el objeto ”reg” e interprete los resultados.
4. En un gráfico de dispersión, reproduzca los salarios vs los años de escolaridad. Muestre los valores predichos de la recta de regresión y su intervalo de confianza.
5. Haga lo mismo que en el ítem anterior solo que ahora separando un gráfico para hombres y otro para mujeres.
6. Usando la base ”mundo” genere un gráfico dinámico que muestre la evolución de la esperanza de vida vs el Pib per cápita de los países a lo largo de los años. Use el color de las observaciones para marcar los continentes, y el tamaño para la población.
7. Exporte la base ”salarios” en formato excel a alguna carpeta de su computador. Guárde el archivo con el nombre ”sueldos”.

Econometría I
Pauta Ayudantía 8

Profesora: Giselli Castillo
Ayudante: Agustín Sanhueza

Pendiente

De ayudantías pasadas teníamos el siguiente modelo y recta de regresión

$$log(salario_i) = \beta_0 + \beta_1 años_i + \beta_2 juegos_i + \beta_3 bateo_i + \beta_4 hruns_i + \beta_5 carreras_i + u_i$$

$$log(\hat{salario_i}) = \underset{(0{,}29)}{11{,}19} + \underset{(0{,}0121)}{0{,}0689} \cdot años_i + \underset{(0{,}0026)}{0{,}0126} \cdot juegos_i + \underset{(0{,}00110)}{0{,}00098} \cdot bateo_i + \underset{(0{,}0161)}{0{,}0144} \cdot hruns_i + \underset{(0{,}0072)}{0{,}0108} \cdot carreras_i$$

T=353
SSR=183,186
$R^2 = 0{,}6278$
Los números en paréntesis debajo de los estimadores corresponden a la desviación estándar de estos.

Con el fin de testear $H_0 : \beta_3 = \beta_4 = \beta_5 = 0$ Ahora usted tiene la siguiente recta de regresión

$$log(\hat{salario_i}) = \underset{(0{,}11)}{11{,}22} + \underset{(0{,}0125)}{0{,}0713} \cdot años_i + \underset{(0{,}0013)}{0{,}0202} \cdot juegos_i$$

Con:
T=353
SSR=198,311
$R^2 = 0{,}5971$
Los números en paréntesis debajo de los estimadores corresponden a la desviación estándar de estos.

1. Testee la hipótesis conjunta de que $H_0 : \beta_3 = \beta_4 = \beta_5 = 0$. Calcule el F_{obs} y muestre gráficamente el P_{value}.

Respuesta

Debemos usar el estadístico F

$$\boxed{F_{obs} = \frac{(SSR_r - SSR_{ur})/q}{(SSR_{ur})/(T-K-1)} = \frac{(R^2_{ur} - R^2_r)/q}{(1 - R^2_{ur})/(T-K-1)} \sim F_{q,t-k-1}}$$

$$F_{obs} = \frac{(198{,}311 - 183{,}186)/3}{183{,}186/(353-5-1)} = \frac{(0{,}6278 - 0{,}5971)/3}{(1-0{,}6278)/(353-5-1)} \sim F_{3,353-5-1}$$

$$F_{obs} = \frac{5{,}0416}{0{,}5279} = \frac{0{,}01023}{0{,}00107} \sim F_{3,347}$$

$$\boxed{F_{obs} = 9{,}55 \sim F_{3,347}}$$

Dado que 9.55 es mayor que el valor crítico al 1 % del nivel de significancia para $F_{3,347}$ rechazamos la hipótesis nula de que $H_0 : \beta_3 = \beta_4 = \beta_5 = 0$

Mostrar gráficamente la distribución F y recalcar la no negatividad de su distribución $SSR_r \geq SSR_{ur} \Leftrightarrow R^2_{ur} \geq R^2_r$

Forma Funcional

Considere los siguientes modelos

$$Nota_i = \beta_0 + \beta_1 estudio_i + u_i$$

$$Nota_i = \gamma_0 + \gamma_1 estudio_i + \gamma_2 estudio_i^2 + \varepsilon_i$$

Donde la variable dependiente son las notas que tienen los alumnos de este curso, y la variable independiente son las horas de estudio que usan a la semana.

1. Calcule cuál es el efecto marginal de una hora más de estudio para cada modelo y resalte la diferencia que habría entre ellos.

Respuesta

Para calcular el efecto en cada modelo, debemos derivar la variable dependiente respecto a la independiente.

Para el primer modelo

$$\boxed{\frac{\partial Nota_i}{\partial estudio_i} = \beta_1}$$

Para el segundo modelo

$$\boxed{\frac{\partial Nota_i}{\partial estudio_i} = \gamma_1 + 2\gamma_2 estudio_i}$$

La principal diferencia es que en el modelo 1 el efecto de una hora más de estudio es siempre la misma, β_1. En cambio en el modelo 2, el efecto marginal está en función de la hora de estudio en la que estoy parado, por lo que es un efecto no lineal.

2. Si estimara el segundo modelo, ¿qué signos esperaría para los valores de γ_1 y γ_2?

Respuesta

Uno podría esperar que $\gamma_1 > 0$ y $\gamma_2 < 0$, obteniendo una recta de regresión cóncava.

En otras palabras, uno podría esperar que una hora más de estudio aumenta la nota promedio que uno podría esperar, pero a un ritmo decreciente. La primera hora de estudio es más 'productiva" que la décima. Sin embargo, podría ocurrir que pasado un punto una hora más de estudio sea contraproducente bajando la nota esperada.

3. ¿Cuál sería el "punto de giro" del segundo modelo?

Respuesta

Para encontrar ese punto de giro podemos derivar e igualar a 0 considerando al modelo 2 como una función cuadrática cualquiera.

$$Notas_i = \gamma_0 + \gamma_1 estudio_i + \gamma_2 estudio_i^2 + \varepsilon_i$$

$$f(x) = \gamma_0 + \gamma_1 x + \gamma_2 x^2 + \varepsilon_i$$

$$f(x)^{'} \Rightarrow \gamma_1 + 2\gamma_2 x = 0$$

$$\boxed{x^* = -\frac{\gamma_1}{2\gamma_2}}$$

Para saber si este punto crítico es un mínimo o máximo sacamos segunda derivada

$$f(x)^{''} = 2\gamma_2$$

El punto de giro será un mínimo si $\gamma_2 > 0$ y un máximo si $\gamma_2 < 0$

Suponga que al modelo lo modificamos y le agregamos la variable "motivacion" (suponga que es medible en alguna unidad) llegando a lo siguiente

$$Nota_i = \delta_0 + \delta_1 motivación_i + \delta_2 estudio_i + \delta_3 estudio \cdot motivación_i + \mu_i$$

4. ¿Como interpretaría cada coeficiente?

Respuesta

En este modelo tenemos un término de interacción entre estudio y motivación. Por lo tanto el efecto marginal de una hora más de estudio se verá influenciada por la motivación de la persona esa semana y viceversa. Esto se ilustra en los efectos marginales de cada variable.

$$\frac{\partial Nota_i}{\partial estudio_i} = \delta_2 + \delta_3 motivacion_i$$

$$\frac{\partial Nota_i}{\partial motivacion_i} = \delta_1 + \delta_3 estudio_i$$

$\therefore$ δ_2 captura el cambio promedio de la nota cuando aumentamos en una hora el estudio, y la motivación es 0, dejando lo demás constante.

$\therefore$ δ_1 captura el cambio promedio de la nota cuando aumentamos en una unidad la motivación, y el estudio es 0, dejando lo demás constante.

$\therefore$ δ_0 Nos dice cuánto será la nota promedio cuando la motivación y estudio es 0

$\therefore$ δ_3 captura el efecto interacción de las horas de estudio y la motivación que tenga la persona sobre las notas.

Un $\delta_3 > 0$ en el primer efecto marginal implica que en personas que han estudidado más, una unidad más de motivación produce un aumento mayor en las notas.

Un $\delta_3 > 0$ en el segundo efecto marginal implica que en personas que tienen una mayor motivación, una hora más de estudio produce un aumento mayor en las notas.

Variables Dummy

Usted está interesado en ver las brechas salariales por sexo a partir del siguiente modelo

$$w_i = \beta_0 + \beta_1 d_{i,1} + u_i \tag{1}$$

Donde w_i es el salario de la persona i y $d_{i,1}$ es una variable que toma el valor 1 si la persona es mujer y 0 si es hombre.
Un colega plantea que es posible capturar la brecha salarial re-escribiendo el modelo anterior de las siguientes formas:

$$w_i = \beta_0 + \beta_1 d_{i,2} + u_i \tag{2}$$

$$w_i = \beta_0 + \beta_1 d_{i,1} + \beta_2 d_{i,2} + u_i \tag{3}$$

$$w_i = \beta_1 d_{i,1} + \beta_2 d_{i,2} + u_i \tag{4}$$

Donde $d_{i,2}$ toma el valor de 1 si la persona es hombre y 0 si es mujer.

1. ¿Está en lo correcto el colega?

Respuesta

En primer lugar analizamos el modelo (1):

$$E(w_i/d_{i,1} = 1) = E(w_i/Mujer) = \beta_0 + \beta_1$$
$$E(w_i/d_{i,1} = 0) = E(w_i/Hombre) = \beta_0$$

$$E(w_i/Mujer) - E(w_i/Hombre) = \beta_0 + \beta_1 - \beta_0$$

$$\therefore \boxed{Brecha = \beta_1}$$

Si $\beta_1 > 0$ implica que las mujeres ganan en promedio más que los hombres.

Para el modelo (2) se observa que esta sí es una forma válida de re-escribir el modelo (1), porque $d_{i,2} = (1 - d_{i,1})$ (la dummy aporta la misma información). Lo que cambia bajo este modelo es la interpretación de los betas, tendremos:

$$E(w_i/d_{i,2} = 0) = E(w_i/Mujer) = \beta_0$$
$$E(w_i/d_{i,2} = 1) = E(w_i/Hombre) = \beta_0 + \beta_1$$

$$E(w_i/Hombre) - E(w_i/Mujer) = \beta_0 + \beta_1 - \beta_0$$
$$\therefore \boxed{Brecha = \beta_1}$$

Si $\beta_1 > 0$ en este modelo implica que los hombres ganan en promedio más que las mujeres.

Para el modelo (3) tendremos un problema de multicolinealidad perfecta por lo que no podrá ser estimado. Sabemos que independientemente de la observación que estemos analizando, se cumplirá por construcción que $d_{i,1} + d_{i,2} = 1$ (una de las dummies tiene que ser 1 y la otra 0 siempre). El problema de perfecta colinealidad se presenta al incluir ambas variables dummy y el intercepto. (Ilustrar con la matriz X pero no profundizar mucho en explicación)

$$X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \end{bmatrix}$$

Finalmente, analizando el modelo (4), vemos que esta sí es una forma válida de re-escribir el modelo (1). Si bien se incluyen ambas dummy en la ecuación, se omite el intercepto, lo que evita el problema de perfecta colinealidad. Eso si, bajo este modelo cambian los β y la interpretación con respecto al modelo (1), tendremos:

$$E[w_i/d_{i,1} = 1, d_{i,2} = 0] = E[w_i/Mujer] = \beta_1$$
$$E[w_i/d_{i,1} = 0, d_{i,2} = 1] = E[w_i/Hombre] = \beta_2$$
$$E[w_i/Mujer] - E[w_i/Hombre] = \beta_1 - \beta_2$$
$$\therefore \boxed{Brecha = \beta_1 - \beta_2}$$

En conclusión, los modelos (1), (2), y (4) permiten calcular la brecha salarial entre hombres y mujeres. La diferencia entre modelos radica en su forma funcional, lo que cambia la interpretación de los coeficientes. El colega está en lo correcto menos en el modelo (3).

Suponga que usted tiene la siguiente base de datos.

Salario	Sexo
400	Hombre
200	Mujer
650	Hombre
400	Mujer
600	Hombre
450	Mujer

2. Indique cuál sería la recta de regresión en los modelos (1), (2), y (4)

Respuesta

Tenemos que

$$E[w_i/mujer] = \frac{200 + 400 + 450}{3} = 350$$

$$E[w_i/hombre] = \frac{400 + 650 + 600}{3} = 550$$

$$\therefore |Brecha| = 200$$

Para el modelo (1)

$$w_i = \beta_0 + \beta_1 d_{1,i} + u_i$$

La recta de regresión sería

$$\boxed{\hat{w}_i = 550 + -200 d_{1,i}}$$

Para el modelo (2)

$$w_i = \beta_0 + \beta_1 d_{2,i} + u_i$$

La recta de regresión sería

$$\boxed{\hat{w}_i = 350 + 200 d_{2,i}}$$

Para el modelo (4)

$$w_i = \beta_1 d_{1,i} + \beta_2 d_{2,i} + u_i$$

La recta de regresión sería

$$\boxed{\hat{w}_i = 350 d_{1,i} + 550 d_{2,i}}$$

Econometría I
Pauta Ayudantía 9

Profesora: Giselli Castillo

Ayudante: Agustín Sanhueza

Variables Dummy y Categorías

Usted está interesado en ver las brechas salariales por sexo a partir del siguiente modelo

$$w_i = \beta_0 + \beta_1 d_{i,1} + \beta_2 esc_i + u_i \tag{1}$$

Donde w_i es el salario de la persona i y $d_{i,1}$ es una variable que toma el valor 1 si la persona es mujer y 0 si es hombre. La variable esc es el número de años de educación de la persona i. Un colega plantea que es posible capturar la brecha salarial re-escribiendo el modelo anterior de las siguientes formas:

$$w_i = \beta_0 + \beta_1 d_{i,2} + \beta_2 esc_i + u_i \tag{2}$$

$$w_i = \beta_0 + \beta_1 d_{i,1} + \beta_2 d_{i,2} + \beta_3 esc_i + u_i \tag{3}$$

$$w_i = \beta_1 d_{i,1} + \beta_2 d_{i,2} + \beta_3 esc_i + u_i \tag{4}$$

Donde $d_{i,2}$ toma el valor de 1 si la persona es hombre y 0 si es mujer.

1. ¿Está en lo correcto el colega?

Respuesta

En primer lugar analizamos el modelo (1):

$$E(w_i/d_{i,1} = 1, esc) = E(w_i/Mujer, esc) = \beta_0 + \beta_1 + \beta_2 esc_i$$
$$E(w_i/d_{i,1} = 0, esc) = E(w_i/Hombre, esc) = \beta_0 + \beta_2 esc_i$$

$$E(w_i/Mujer, esc) - E(w_i/Hombre, esc) = \beta_0 + \beta_1 + \beta_2 esc_i - \beta_0 - \beta_2 esc_i$$
$$\therefore \boxed{Brecha = \beta_1}$$

Si $\beta_1 > 0$ implica que las mujeres ganan en promedio más que los hombres.

Para el modelo (2) se observa que esta sí es una forma válida de re-escribir el modelo (1), porque $d_{i,2} = (1 - d_{i,1})$ (la dummy aporta la misma información). Lo que cambia bajo este modelo es la interpretación de los betas, tendremos:

$$E(w_i/d_{i,2} = 0) = E(w_i/Mujer) = \beta_0 + \beta_2 esc_i$$
$$E(w_i/d_{i,2} = 1) = E(w_i/Hombre) = \beta_0 + \beta_1 + \beta_2 esc_i$$

$$E(w_i/Hombre) - E(w_i/Mujer) = \beta_0 + \beta_1 + \beta_2 esc_i - \beta_0 - \beta_2 esc_i$$

$$\therefore \boxed{Brecha = \beta_1}$$

Si $\beta_1 > 0$ en este modelo implica que los hombres ganan en promedio más que las mujeres.

Para el modelo (3) tendremos un problema de multicolinealidad perfecta por lo que no podrá ser estimado. Sabemos que independientemente de la observación que estemos analizando, se cumplirá por construcción que $d_{i,1} + d_{i,2} = 1$ (una de las dummies tiene que ser 1 y la otra 0 siempre). El problema de perfecta colinealidad se presenta al incluir ambas variables dummy y el intercepto. (Ilustrar con la matriz X pero no profundizar mucho en explicación)

$$X = \begin{bmatrix} 1 & 1 & 0 & 4 \\ 1 & 0 & 1 & 8 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & 7 \end{bmatrix}$$

Finalmente, analizando el modelo (4), vemos que esta sí es una forma válida de re-escribir el modelo (1). Si bien se incluyen ambas dummy en la ecuación, se omite el intercepto, lo que evita el problema de perfecta colinealidad. Eso si, bajo este modelo cambian los β y la interpretación con respecto al modelo (1), tendremos:

$$E[w_i/d_{i,1} = 1, d_{i,2} = 0, esc] = E[w_i/Mujer, esc] = \beta_1 + \beta_2 esc_i$$

$$E[w_i/d_{i,1} = 0, d_{i,2} = 1, esc] = E[w_i/Hombre, esc] = \beta_2 + \beta_2 esc_i$$

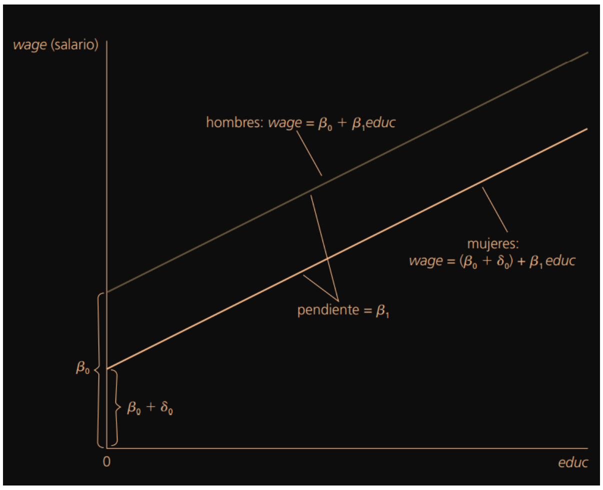
$$E[w_i/Mujer, esc] - E[w_i/Hombre, esc] = \beta_1 + \beta_2 esc_i - \beta_2 - \beta_2 esc_i$$

$$\therefore \boxed{Brecha = \beta_1 - \beta_2}$$

En conclusión, los modelos (1), (2), y (4) permiten calcular la brecha salarial entre hombres y mujeres. La diferencia entre modelos radica en su forma funcional, lo que cambia la interpretación de los coeficientes. El colega está en lo correcto menos en el modelo (3).

2. Usando el modelo (1) y suponiendo $\beta_1 < 0$, $\beta_2 > 0$, deje expresada las rectas de regresión para mujeres y hombres, luego grafiquelas e inteprete cada uno de los coeficientes.

Respuesta



La recta de regresión para las mujeres y hombres serían:

$$\boxed{E[w_i/Mujer, esc] = \beta_0 + \beta_1 + \beta_2 esc_i}$$

$$\boxed{E[w_i/Hombre, esc] = \beta_0 + \beta_2 esc_i}$$

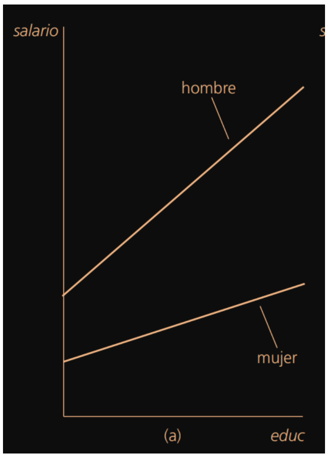
La interpretación de cada coeficiente es la siguiente:
β_0 : Salario promedio de los hombres con cero años de escolaridad
$(\beta_0 + \beta_1)$: Salario promedio de las mujeres con cero años de escolaridad
β_1 : Brecha salarial, con los mismos años de escolaridad.
β_2 : Retorno de un año más de escolaridad en el salario (este modelo asume que es el mismo para hombres y mujeres).

Ahora al modelo (1) le agregamos una variable interacción, llegando al siguiente modelo.

$$w_i = \beta_0 + \beta_1 d_{i,1} + \beta_2 esc_i + \beta_3 esc_i \cdot d_{i,1} + u_i$$

3. Suponiendo $\beta_1 < 0$, $\beta_2 > 0$, $\beta_3 < 0$, $|\beta_2| > |\beta_3|$ deje expresada las rectas de regresión para mujeres y hombres, luego grafiquelas e inteprete cada uno de los coeficientes.

Respuesta



La recta de regresión para las mujeres y hombres serían:

$$\boxed{E[w_i/Mujer, esc] = \beta_0 + \beta_1 + (\beta_2 + \beta_3)esc_i}$$

$$\boxed{E[w_i/Hombre, esc] = \beta_0 + \beta_2 esc_i}$$

La interpretación de cada coeficiente es la siguiente:
β_0 : Salario promedio de los hombres con cero años de escolaridad
$(\beta_0 + \beta_1)$: Salario promedio de las mujeres con cero años de escolaridad
β_1 : Brecha salarial, con los mismos años de escolaridad.
β_2 : Retorno de un año más de escolaridad en el salario para los hombres.
$(\beta_2 + \beta_3)$: Retorno de un año más de escolaridad en el salario para las mujeres.
β_3 : Brecha en el retorno de un año más de escolaridad.

Considere el siguiente modelo

$$w_i = \beta_0 + \beta_1 media_i + \beta_2 superior_i + u_i$$

Donde "media" es una dummy que toma el valor 1 si la persona i tiene entre 9 y 12 años de educación (inclusive). "superior" es una dummy que toma el valor 1 si la persona i tiene 13 o más años de educación.

4. Deje expresada las rectas de regresión para las personas con educación básica, media y superior, e inteprete cada uno de los coeficientes.

Respuesta

$$\boxed{E[w_i/básica, esc] = \beta_0}$$

$$\boxed{E[w_i/media, esc] = \beta_0 + \beta_1}$$

$$\boxed{E[w_i/superior, esc] = \beta_0 + \beta_2}$$

La interpretación de cada coeficiente es la siguiente:

β_0 : Salario promedio de las personas que tienen educación básica.
$(\beta_0 + \beta_1)$: Salario promedio de las personas que tienen educación media.
$(\beta_0 + \beta_2)$: Salario promedio de las personas que tienen educación superior.
β_1 : Brecha salarial entre las personas con educación básica y media.
β_2 : Brecha salarial entre las personas con educación básica y superior.

Modelo de Probabilidad Lineal

1. Mencione los problemas de estimar modelos de variable dependiente binaria de manera lineal.

Respuesta

Algunos valores predichos del modelo tendrán una probabilidad mayor a 1 o menor a 0, lo que no tiene sentido.

Por construcción el modelo será heterocedástico (dist bernoulli)

Para este tipo de modelos lo que se hace típicamente es estimar un probit/logit. Esto no se ve en el curso.

Usted está interesado en estudiar la probabilidad de tener deudas con el siguiente modelo

$$D_i = \beta_0 + \beta_1 ocupado_i + \beta_2 Edad_i + u_i$$

Donde D_i es 1 si la persona i tiene deudas y 0 sino. Ocupado es una dummy que toma el valor 1 si la persona i tiene empleo y 0 sino. $Edad_i$ la edad de la persona i.

2. Deje expresada las rectas de regresión del modelo para los ocupados y desocupados. Interprete cada coeficiente de este modelo

Respuesta

$$Pr[D_i = 1/ocupado = 1, Edad] = \beta_0 + \beta_1 + \beta_2 Edad_i$$
$$Pr[D_i = 1/ocupado = 0, Edad] = \beta_0 + \beta_2 Edad_i$$

La interpretación de los coeficientes es la siguiente:

β_0 : Probabilidad de tener deuda si la persona está desocupada y tiene 0 años.
$(\beta_0 + \beta_1)$: Probabilidad de tener deuda si la persona está ocupada y tiene 0 años.
β_1 : Brecha en la probabilidad de tener deudas entre ocupados y desocupados.
β_2 : El cambio en la probabilidad promedio de tener deudas ante un año más de edad (independiente de si la persona está ocupada o no).

Econometría I
Pauta Ayudantía 10

Profesora: Giselli Castillo
Ayudante: Agustín Sanhueza

Aplicando Dummies, Categorías e Interacciones

Solo Dummies

Considere los siguientes modelos de regresión

$$w_i = \beta_0 + \beta_1 d_{i,1} + u_i \tag{1}$$

$$w_i = \beta_0 + \beta_1 d_{i,2} + u_i \tag{2}$$

$$w_i = \beta_1 d_{i,1} + \beta_2 d_{i,2} + u_i \tag{3}$$

$$w_i = \beta_0 + \beta_1 d_{i,1} + \beta_2 d_{i,2} + u_i \tag{4}$$

$$d_{i,1} = \begin{cases} 1 & \text{si observación i es Mujer} \\ 0 & \text{sino} \end{cases}$$

$$d_{i,2} = \begin{cases} 1 & \text{si observación i es Hombre} \\ 0 & \text{sino} \end{cases}$$

Y suponga que usted tiene la siguiente base de datos.

Salario	Sexo
400	Hombre
200	Mujer
650	Hombre
400	Mujer
600	Hombre
450	Mujer

1. Indique cuál sería la recta de regresión en los modelos (1), (2) y (3), interpretando sus coeficientes. Compute las estimaciones de los modelos en R. Comente qué pasa computacionalmente con el modelo (4).

Respuesta

Tenemos que

$$E[w_i/mujer] = \frac{200 + 400 + 450}{3} = 350$$

$$E[w_i/hombre] = \frac{400 + 650 + 600}{3} = 550$$

$$\therefore |Brecha| = 200$$

Para el modelo (1)

$$w_i = \beta_0 + \beta_1 d_{1,i} + u_i$$

La recta de regresión sería

$$\boxed{\hat{w}_i = 550 - 200 d_{1,i}}$$

Para el modelo (2)

$$w_i = \beta_0 + \beta_1 d_{2,i} + u_i$$

La recta de regresión sería

$$\boxed{\hat{w}_i = 350 + 200 d_{2,i}}$$

Para el modelo (3)

$$w_i = \beta_1 d_{1,i} + \beta_2 d_{2,i} + u_i$$

La recta de regresión sería

$$\boxed{\hat{w}_i = 350 d_{1,i} + 550 d_{2,i}}$$

Solo categorías

Considere los siguientes modelos de regresión

$$w_i = \beta_0 + \beta_1 media_i + \beta_2 superior_i + u_i \tag{5}$$

$$w_i = \beta_0 + \beta_1 básica_i + \beta_2 superior_i + u_i \tag{6}$$

Con

$$básica_i = \begin{cases} 1 & \text{Si observación i está en educación básica} \\ 0 & \text{sino} \end{cases}$$

$$media_i = \begin{cases} 1 & \text{Si observación i está en educación media} \\ 0 & \text{sino} \end{cases}$$

$$superior_i = \begin{cases} 1 & \text{Si observación i está en educación superior} \\ 0 & \text{sino} \end{cases}$$

Y suponga que usted tiene la siguiente base de datos.

Salario	Educación
400	Básica
200	Básica
650	Superior
400	Media
600	Media
450	Superior

2. Indique cuál sería la recta de regresión de los modelos, interpretando sus coeficientes. Compute las estimaciones en R.

Respuesta

Tenemos que

$$E[w_i/básica] = \frac{400+200}{2} = 300$$
$$E[w_i/media] = \frac{400+600}{2} = 500$$
$$E[w_i/superior] = \frac{650+450}{2} = 550$$

Para el modelo (5)

$$w_i = \beta_0 + \beta_1 media_i + \beta_2 superior_i + u_i$$

La recta de regresión sería:

$$\boxed{\hat{w}_i = 300 + 200 media_i + 250 superior_i}$$

Para el modelo (6)

$$w_i = \beta_0 + \beta_1 básica_i + \beta_2 superior_i + u_i$$

La recta de regresión sería:

$$\boxed{\hat{w}_i = 500 - 200 básica_i + 50 superior_i}$$

Dummy + Interacción

Considere el siguiente modelo de regresión

$$w_i = \beta_0 + \beta_1 ubicación_i + \beta_2 educación_i + \beta_3 educación_i \cdot ubicación_i + u_i \tag{7}$$

Con

$$ubicación_i = \begin{cases} 1 & \text{Si observación i vive en la región metropolitana} \\ 0 & \text{Si observación i vive en regiones} \end{cases}$$

Y suponga que usted tiene la siguiente base de datos.

Salario	Educación	Lugar
400	17	Metropolitana
200	12	Araucanía
650	22	Metropolitana
400	18	Magallanes
600	20	Metropolitana
450	16	Valparaíso

3. Calcule la recta de regresión e interprete los coeficientes

Respuesta

$$E[w_i/Metropolitana, educacion] = \beta_0 + \beta_1 + (\beta_2 + \beta_3)educacion_i$$
$$E[w_i/Region, educacion] = \beta_0 + \beta_2 educacion_i$$

La interpretación de los coeficientes será:
β_0 : Salario promedio de las personas de región con cero años de escolaridad
$(\beta_0 + \beta_1)$: Salario promedio de las personas de la región metropolitana con cero años de escolaridad
β_1 : Brecha salarial promedio de las personas con cero años de escolaridad
$(\beta_1 + \beta_3 educacion_i)$: Brecha salarial promedio de las personas con algun año de escolaridad distinto de 0.
β_2 : Retorno de un año más de educación para las personas de regiones.
$(\beta_2 + \beta_3)$: Retorno de un año más de educación para las personas de la región metropolitana.
β_3 : Brecha en el retorno de un año más de educación.

La recta de regresión será

$$\boxed{\hat{w}_i = -225 - 234{,}21 ubicacion_i + 37{,}25 educacion_i + 37{,}5 educacion_i \cdot ubicacion_i}$$

Modelo de probabilidad lineal

Usted está interesado en estudiar la probabilidad de tener deudas con el siguiente modelo

$$D_i = \beta_0 + \beta_1 ocupado_i + \beta_2 ingreso_i + u_i$$

Con

$$D_i = \begin{cases} 1 & \text{Si persona i tiene deudas} \\ 0 & \text{sino} \end{cases}$$

$$ocupado_i = \begin{cases} 1 & \text{Si persona i tiene empleo} \\ 0 & \text{sino} \end{cases}$$

Y usted tiene los siguientes datos

Deuda	Estado	Ingreso
1	Empleado	600
0	Empleado	1000
0	Empleado	800
0	Desempleado	500
1	Desempleado	300
1	Desempleado	200

1. Calcule la recta de regresión en R graficándola e interprete los coeficientes.

Respuesta

$$Pr[D_i = 1/ocupado = 1, ingreso_i] = \beta_0 + \beta_1 + \beta_2 ingreso_i$$

$$Pr[D_i = 1/ocupado = 0, ingreso_i] = \beta_0 + \beta_2 ingreso_i$$

La interpretación de los coeficientes es la siguiente:

β_0 : Probabilidad promedio de tener deuda si la persona está desocupada y no tiene ingreso.
$(\beta_0 + \beta_1)$: Probabilidad promedio de tener deuda si la persona está ocupada y no tiene ingreso.
β_1 : Brecha en la probabilidad de tener deudas entre ocupados y desocupados con cualquier nivel de ingreso.
β_2 : El cambio en la probabilidad promedio de tener deudas ante una unidad adicional de ingreso (independiente de si la persona está ocupada o no).

La recta de regresión será

$$\boxed{Pr[D_i = 1/ocupado_i, ingreso_i] = 1{,}63 + 1{,}02 ocupado_i - 0{,}003 ingreso_i}$$

Econometría I
Pauta Ayudantía 11

Profesora: Giselli Castillo

Ayudante: Agustín Sanhueza

Heterocedasticidad

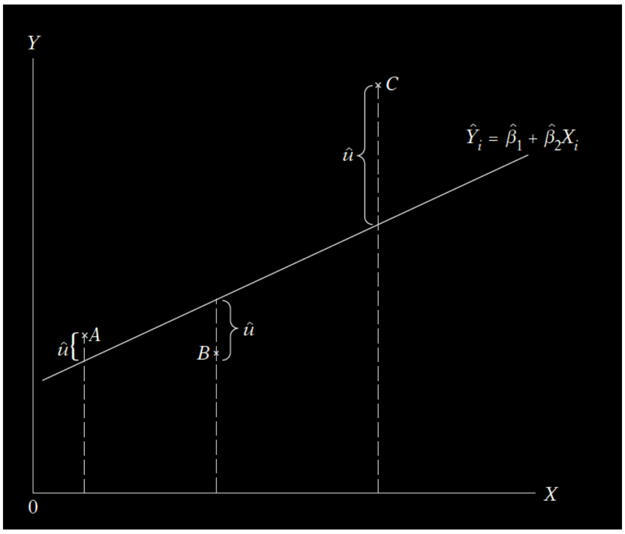
1. Mencione las consecuencias que trae la heterocedasticidad en el sesgo y eficiencia de los estimadores. En la validez de los test t y F. Y en los errores estándar.

> **Respuesta**
>
> La heterocedasticidad genera ineficiencia en la estimación, pero no sesgo. En consecuencia el estimador de mínimos cuadrados ordinarios ya no será MELI.
>
> La heterocedasticidad invalida los test t y F.
>
> La heterocedasticidad puede sobreestimar o subestimar los errores estándar de los estimadores.

2. Mencione las soluciones de estimación que existen para la heterocedasticidad

> **Respuesta**
>
> Seguir estimando por mínimos cuadrados ordinarios, pero con errores estándares robustos (matriz sandwich de White). Bajo muestras grandes el test t vuelve a ser válido, y el test F puede realizarse en una variante ajustada.
>
> Estimar por Mínimos Cuadrados Generalizados. Aquí es necesario conocer el patrón de la heterocedasticidad para ponderarla adecuadamente.
>
> Estimar por Mínimos Cuadrados Generalizados Factibles. Aquí es necesario usar un supuesto en el patrón de la heterocedasticidad. Típicamente se supone que se sigue una distribución exponencial para garantizar una varianza positiva.

3. Mencione la diferencia entre mínimos cuadrados ordinarios, y mínimos cuadrados ponderados (generalizados), a partir de la siguiente gráfica.



Respuesta

En MCO (sin ponderar), cada $\hat{u_i^2}$ asociada con los puntos A, B y C recibirá el mismo peso al reducir la SSR. Obviamente, en este caso la $\hat{u_i^2}$ asociada al punto C dominará la SSR. Pero en MCG la observación extrema C obtendrá relativamente un peso menor que las otras dos observaciones. Esto es lo que genera una estimación más eficiente de los coeficientes.

Transformando un modelo

Usted quiere estimar la siguiente regresión

$$Cerveza_i = \beta_0 + \beta_1 ingreso_i + \beta_2 precio_i + \beta_3 educación_i + \beta_4 mujer_i + u_i$$

Donde:
$E[u_i/ingreso_i, precio_i, educación_i, mujer_i] = 0$
$Var[u_i/ingreso_i, precio_i, educación_i, mujer_i] = \sigma_i^2 = \sigma^2 \cdot ingreso_i^2$

1. Escriba la ecuación transformada que tenga un término de error homocedástico

Respuesta

$$Cerveza_i = \beta_0 + \beta_1 ingreso_i + \beta_2 precio_i + \beta_3 educación_i + \beta_4 mujer_i + u_i$$

$$Cerveza_i \frac{1}{ingreso_i} = \beta_0 \frac{1}{ingreso_i} + \beta_1 + \beta_2 precio_i \frac{1}{ingreso_i} + \beta_3 educación_i \frac{1}{ingreso_i} + \beta_4 mujer_i \frac{1}{ingreso_i} + u_i \frac{1}{ingreso_i}$$

De esta manera tendremos

$$\frac{1}{ingreso_i} Var[u_i/X] = \frac{1}{ingreso_i} (\sigma^2 \cdot ingreso_i^2)$$

$$Var[u_i^*/X] = (\sigma^2 \cdot ingreso_i^2 \cdot \frac{1}{ingreso_i^2})$$

$$\boxed{Var[u_i^*/X] = \sigma^2}$$

Tests

1. Escriba los pasos del test de Breusch-Pagan

Respuesta

1) Estimar los residuos al cuadrado ${\hat{u}_i}^2$ del modelo de regresión

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + ... + \beta_k x_{i,k} + u_i$$

2) Guardar el $R_{{\hat{u}_i}^2}$ del modelo de regresión

$${\hat{u}_i}^2 = \gamma_0 + \gamma_1 x_{i,1} + \gamma_2 x_{i,2} + ... + \gamma_k x_{i,k} + \varepsilon_i$$

3) Testear la hipótesis nula conjunta $H_0 : \gamma_1 = \gamma_2 = ... = \gamma_k = 0$ con el estadístico F, o LM.

$$\boxed{F = \frac{R^2_{\hat{u}^2}/K}{1 - R^2_{\hat{u}^2}/(T-K-1)} \sim F_{K,(T-K-1)}}$$

$$\boxed{LM = T \cdot R^2_{\hat{u}^2} \sim \chi^2_K}$$

Si el valor P es lo suficientemente pequeño, se rechaza la hipótesis de homocedasticidad.

2. Escriba los pasos del test de White

Respuesta

1) Estimar los residuos al cuadrado ${\hat{u}_i}^2$ del modelo de regresión

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + ... + \beta_k x_{i,k} + u_i$$

2) Guardar el $R_{{\hat{u}_i}^2}$ del modelo de regresión

$${\hat{u}_i}^2 = \gamma_0 + \gamma_1 x_{i,1} + \gamma_2 x_{i,2} + ... + \gamma_k x_{i,k} + \gamma_{k+1} x^2_{i,1} + \gamma_{k+2} x^2_{i,2} + ... + \gamma_{2k} x^2_{i,k} + \gamma_{2k+1} x_{i,1} x_{i,2} + \gamma_{2k+2} x_{i,1} x_{i,3} + ... + \varepsilon_i$$

3) Testear la hipótesis nula conjunta $H_0 : \gamma_1 = \gamma_2 = ... = \gamma_k = 0$ LM.

$$\boxed{LM = T \cdot R^2_{\hat{u}^2} \sim \chi^2_K}$$

Si el valor P es lo suficientemente pequeño, se rechaza la hipótesis de homocedasticidad.

Alternativamente:
1) Estimar los residuos al cuadrado $\hat{u_i}^2$, y los valores predichos $\hat{y}_i$ del modelo de regresión

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + ... + \beta_k x_{i,k} + u_i$$

2) Guardar el $R_{\hat{u_i}^2}$ del modelo de regresión

$$\hat{u_i}^2 = \delta_0 + \delta_1 \hat{y}_i + \delta_2 \hat{y_i}^2 + \varepsilon_i$$

3) Testear la hipótesis nula conjunta $H_0 : \delta_1 = \delta_2 = 0$ con el estadístico F, o LM

$$\boxed{F = \frac{R^2_{\hat{u}^2}/K}{1 - R^2_{\hat{u}^2}/(T-3)} \sim F_{2,(T-3)}}$$

$$\boxed{LM = T \cdot R^2_{\hat{u}^2} \sim \chi^2_2}$$

Si el valor P es lo suficientemente pequeño, se rechaza la hipótesis de homocedasticidad.

Minímos Cuadrados Ponderados

Con el objetivo de estudiar el efecto de los gastos de defensa sobre otros gastos en la economía, usted estima el siguiente modelo.

$$C_t = \beta_0 + \beta_1 PNB_t + \beta_2 D_t + u_t \quad (1)$$

Con:
C_t : Consumo agregado del año t
PNB_t : Producto Nacional Bruto del año t
D_t : Gasto en defensa nacional del año t

Usted cree que $\sigma_t^2 = \sigma^2 (PNB_t)^2$ por lo que también estima el siguiente modelo

$$\frac{C_t}{PNB_t} = \frac{\beta_0}{PNB_t} + \beta_1 + \beta_2 \frac{D_t}{PNB_t} + \frac{u_t}{PNB_t} \quad (2)$$

Con datos desde 1946 hasta 1975 usted llegó a los siguientes resultados.

$$\hat{C}_t = \underset{(2{,}73)}{26{,}19} + \underset{(0{,}006)}{0{,}6248} \cdot PNB_t - \underset{(0{,}0736)}{0{,}4398} \cdot D_t \qquad R^2{=}0{,}999$$

$$\frac{\hat{C}_t}{PNB_t} = \underset{(2{,}22)}{25{,}92} \frac{1}{PNB_t} + \underset{(0{,}0068)}{0{,}6246} - \underset{(0{,}0597)}{0{,}4315} \cdot \frac{D_t}{PNB_t} \qquad R^2{=}0{,}875$$

1. ¿Qué supuesto se hace sobre la naturaleza de la heteroscedasticidad? ¿Puede justificarlo?

Respuesta

El supuesto es que la varianza del error crece proporcionalmente con el PNB al cuadrado (países más ricos generan mayor volatilidad en el consumo).

2. Compare los resultados de las dos regresiones. ¿La transformación del modelo original mejora los resultados, es decir, reduce los errores estándar estimados? ¿Por qué?

Respuesta

Los resultados son escensialmente los mismos, solo que en el segundo los errores estándar son menores. Por lo que es conveniente estimar el modelo de la segunda forma para ganar eficiencia.

3. ¿Puede comparar los dos valores de R^2? ¿Por qué?

Respuesta

No, la variable dependiente es distinta en ambos modelos, por lo que los R^2 no son comparables.