Estimación de Regresiones Lineales por OLS

$$\underset{\text{V.D}}{y} = \beta_0 + \beta_1 \underset{\text{V.I}}{x_1} + \dots + \beta_K x_K + u \qquad K+1 \text{ parámetros}$$

¿Cómo calcular parámetros?

$$\hat{\beta}_1 = \frac{\sum_{i=0}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2} = \frac{\mathrm{Cov}(x,y)}{\mathrm{Var}(x)}$$

$$= \frac{\sum x^2 - n\bar{x}^2}{\sum x_i(y_i - \bar{y})} = \frac{\sum x_i(x_i - \bar{x})}{\sum x_i(y_i - \bar{y})}$$

$$\hat{\beta}_0 = \frac{\sum y_i - \hat{\beta}_1 \sum x_i}{n} \qquad \hat{\beta}_2 = \frac{\sum x_1^2 \sum x_2 y - \sum x_1 x_2 \sum x_1 y}{\sum x_1^2 \sum x_2^2 - \left(\sum x_1 x_2\right)^2}$$

En forma matricial

$$Y = X\hat{\beta} + u$$

$$\underset{\text{ingreso}}{\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1K} \\ 1 & \vdots & & \\ 1 & x_{1n} & \cdots & x_{nK} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_K \end{bmatrix} + \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}$$

edad sexo owp

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

$$X'y = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_{i2} y_{i2} \end{bmatrix} \qquad y'y = \sum_{i=1}^{n} y^2 \qquad u'u = \sum_{i=1}^{n} u^2 = (y - x\beta)'(y - x\beta)$$

$$y = x\beta + u \qquad u = y - x\beta \qquad \hat{\beta} = (x'x)^{-1}x'y$$

Ajuste del modelo

$$SSR = \sum (y_i - \hat{y})^2$$

$$SEC = \sum (\hat{y} - \bar{y})^2$$

$$STC = \sum (y_i - \bar{y})^2$$

$$STC = SEC + SSR$$

$$R^2 = \frac{SEC}{STC} = 1 - \frac{SSR}{STC} \longrightarrow \text{porcentaje de var. explicada del modelo}$$

$$\bar{R}^2 = 1 - \left(\frac{n-1}{n-K-1}\right)(1 - R^2)$$

Pruebas de hipótesis

$$t_{n-K-1} = \frac{\hat{\beta} - \beta_j}{se(\hat{\beta})} \qquad se(\hat{\beta}_j) = \sqrt{var(\hat{\beta})} = \frac{\hat{\sigma}^2}{STC(1-R^2)} \qquad \hat{\sigma} = \frac{SSR}{n-K-1}$$

$$t_{n-K-1} = \frac{\hat{\beta}_1 - \hat{\beta}_2}{se(\hat{\beta}_1 - \hat{\beta}_2)} = \frac{\hat{\beta}_1 - \hat{\beta}_2}{\sqrt{v(\beta_1) + v(\beta_2) - 2\,cov(\beta_1\beta_2)}}$$

AYUDANTIA N°1 - ECONOMETRÍA

Profesor: Juan Urquiza

Ayudante: Valentina Andrade (vandrade@uc.cl)

TEMA I

Imagine que desea estimar la relación entre la salud infantil y el fumar durante el embarazo. Para ello, decide considerar como medida de salud infantil al peso al nacer, ya que un peso demasiado bajo se asocia con un mayor riesgo de contraer distintas enfermedades. Usted dispone de los siguientes antecedentes:

$$n = 694\,; \quad \sum_{i=1}^{n} Y_i = \overset{\bar{Y}}{\frac{2339.9}{n}}\,; \quad \sum_{i=1}^{n} X_i = \frac{83.7}{n}\,; = \bar{X}$$

$$\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X}) = -24.5\,; \quad \leftarrow \text{cov}(y, x)$$

$$\overset{\text{STC}}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} = 236.6\,; \quad \sum_{i=1}^{n}(X_i - \bar{X})^2 = 71.8\,; \quad \leftarrow \text{var}(x)$$

donde *Y* es el peso al nacer (en kilogramos) y *X* es el promedio de cajetillas (o *packs*) al día que la madre fumó durante el embarazo.

a. Estime la regresión simple de *Y* sobre *X* por MCO, y luego interprete los coeficientes.

Consciente de las limitaciones del modelo simple, decide controlar por el orden de nacimiento (x_2) y por el ingreso familiar mensual (x_3) (en miles de dólares). Utilice las salidas de STATA que se presentan a continuación para responder las siguientes preguntas:

Source	SS	df	MS		
SEC Model		3		Number of obs =	694 = n
SSR Residual	225.534408	690	.326861461	F(· · · ·) =	
STC Total		693		Root MSE =	.57172

peso	Coef. β	Std. Err. se(β)	t	P>|t|	[95% Conf. Interval]
x_1 packs	-.3363439				
x_2 orden	.0549438	.0242513	2.27		
x_3 ingreso	.0269743	.0143328	1.88		
_cons	3.25703	.0593687	54.86		

$Var(\beta_1) = \sqrt{var(\beta_1)}$

e(V)	packs	orden	ingreso	_cons
packs	.00476939			
orden	-.00018197	.00058812		
ingreso	.00017281	.00002231	.00020543	
_cons	-.00069549	-.00099472	-.00055443	.00352464

b. Calcule el R^2 y el R^2 ajustado de este modelo.

c. Evalúe formalmente la hipótesis nula que $\beta_1 = 0$ frente a la alternativa que $\beta_1 < 0$. Considere un nivel de significancia del 1%.

d. Evalúe formalmente la significancia conjunta de *orden* e *ingreso*, sabiendo que la *SCR* del modelo que omite ambas variables es igual a 228.25. Considere un nivel de significancia del 5%.

a. y = peso al nacer (en kg)

x = promedio de cajetillas (o packs)

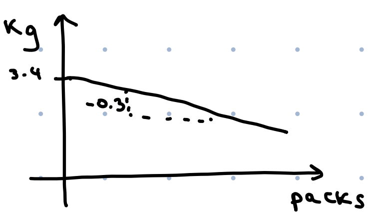
$y = \beta_0 + \beta_1 x_1 + u$ modelo original

$\downarrow$

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{u}$

$$\hat{\beta}_1 = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} = \frac{-24.5}{71.8} \cong -0.3141$$

$$\beta_0 = \underset{\frac{\sum y_i}{n}}{\bar{y}} - \underset{\beta_1}{\hat{\beta}_1} \cdot \underset{\frac{\sum x_i}{n}}{\bar{x}} = \left(\frac{2339.9}{694}\right) - (-0.341) \cdot \left(\frac{83.7}{694}\right) \approx 3.413$$



¿Interpretación?

$\hat{\beta}_1 = -0.3141$

Por cada cajetilla que fume la madre (x), el hijo disminuye su peso en 0.3141 kg, $\underbrace{\text{controlando por el resto de las v. del modelo}}_{\text{ceteris paribus}}$

$\hat{\beta}_0$ = En promedio los hijos tienen un peso de 3.413 kg.

b. Calcule el R^2 y el R^2 ajustado de este modelo. + interpretación

$STC = SEC + SSR$

$$R^2 = \frac{SEC}{STC} = \frac{STC - SSR}{STC} = 1 - \frac{SSR}{STC}$$

$$= 1 - \frac{225.5}{236.6} = 0.047 \cdot 100\%$$

4,6% de la varianza del modelo es explicada por las V.A.

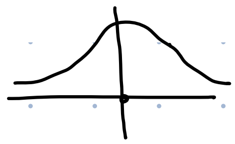
$$\bar{R}^2 = 1 - \left(\frac{n-1}{n-K-1}\right) \times (1 - R^2) = 1 - \left(\frac{694-1}{694-3-1}\right) \times (1 - 0.047)$$

$K = 3$

$$= 0.043$$

c. Evalúe formalmente la hipótesis nula que $\beta_1 = 0$ frente a la alternativa que $\beta_1 < 0$. Considere un nivel de significancia del 1%.

$Y = \beta_0 + \cancel{\beta_1 x}$



Pasos 1. Formalizar la hipótesis

$H_0 : \beta_1 = 0 \qquad H_1 : \beta_1 < 0$

Paso 2. Prueba de contraste (estadístico "crítico")

$$\hat{t}^{\alpha/2}_{n-K-1} = \frac{\hat{\beta}_j - \beta}{se(\hat{\beta}_j)} = \frac{-0.336 - 0}{\sqrt{0.0048}} = -4.85$$

Paso 3. Estadístico teórico

$$\hat{t}^{1-\alpha}_{n-K-1} = t^{1-0.01}_{694-3-1} = t^{0.99}_{690} \approx t^{0.99}_{1000} = -2.33$$

$n = 694$

$K = 3$

$\alpha = 0.01$

Paso 4. Análisis

Si tomamos $\hat{t}^{\alpha/2}_{n-K-1}$ y comparamos con t_{n-K-1} (obtenido de tabla t student) podemos decir con un 99% de confianza que se rechaza la H_0, y con ello, existe evidencia sustantiva del efecto parcial de X_1 sobre $\hat{y}$ ($\hat{t} < t$), controlando por el resto de las variables (ceteris paribus)

d. Evalúe formalmente la significancia conjunta de *orden* e *ingreso*, sabiendo que la *SCR* del modelo que omite ambas variables es igual a 228.25. Considere un nivel de significancia del 5%.

x_2 x_3 SSR 95%

Cuando evaluamos hipótesis conjuntas podemos hacer tres pruebas

1. Prueba t

2. Matrices

$\rightarrow \hat{F} = (R\hat{\beta} - r)'(\hat{\sigma} R(X'X)^{-1}(R\hat{\beta} - r)/q \sim F(q, n-k-1)$

3. Modelo restringido

① $Y = \beta_0 + \beta_1 x_1$ $\quad SSR_R = 228.25$

$\neq$ $\quad \neq$

② $Y = \beta_0 + \beta_1 x_1 + \underline{\beta_2 x_2 + \beta_3 x_3}$ $\quad SSR_{NR}$

3. Modelo restringido

Paso 1. Formalizar hipótesis

$H_0: \beta_2 = \beta_3 = 0$ $\quad$ H_1: Al menos uno de ellos no es cero.

Modelo restringido $\quad$ Modelo sin restringir

Paso 2. Prueba contraste

$$F = \frac{SSR_R - SSR_{NR}}{SSR_R} \cdot \frac{(n-k-1)}{q} \xrightarrow{SST} \frac{R^2_{NR} - R^2_R}{1 - R^2_R} \cdot \frac{(n-k-1)}{q}$$

$$\hat{F}_{(q, n-k-1)} = \frac{228.5 - 225.5}{225.5} \cdot \frac{(694-3-1)}{2} \approx 4.15$$

Paso 3. Estadístico teórico $\quad F_{q, n-k-1}$

$F_{q,n-k-1} = F_{(2, 694-3-1)} \approx F_{(2,120)} = 3.07$

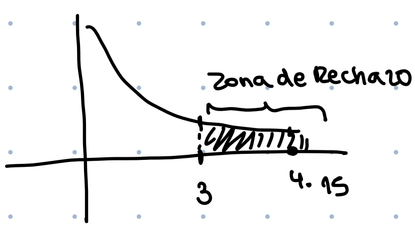
$n = 694$ $\quad \approx F_{(2, 100000)} = 3$

$k = 3$

$q = 2$

$gl = n-k-1$
grado de libertad del modelo NR

$q = 2$
restricciones de k parámetro



Paso 4. Análisis

Si tomamos $\hat{F}_{(q,n-k-1)}$, y comparamos con $F_{(q,n-k-1)}$ obtenido de la tabla podemos decir con un 95% de confianza que se rechaza la H_0, y con ello, existen evidencia concluyente del efecto de las variables incoporadas en las 2 restricciones de exclusión, ceteris paribus.

TEMA II

$\beta = (x'x)^{-1} x'y$

Considere el siguiente modelo de regresión lineal, donde se cumple con todos los supuestos desarrollados en clase:

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & \vdots & \ddots & \\ 1 & x_{1n} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}$$

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

$\sigma^2 (x'x)^{-1}$

$$\hat{\sigma}\ (X'X)^{-1} = \begin{bmatrix} 0{,}6477 & -0{,}0410 & -0{,}0639 \\ -0{,}0410 & 0{,}0071 & -0{,}0011 \\ -0{,}0639 & -0{,}0011 & 0{,}0152 \end{bmatrix} \quad ; \quad X'y = \begin{bmatrix} 91 \\ 699 \\ 448 \end{bmatrix}$$

$$n = 12 \quad ; \quad SCT = 104{,}9167.$$

A partir de estos antecedentes, se pide que:

a. Estime la regresión por MCO y compruebe que:

$$\overset{1}{\hat{\beta}} = (X'X)^{-1}X'y \cong \underbrace{\begin{bmatrix} 1{,}6545 \\ 0{,}7391 \\ 0{,}2258 \end{bmatrix}}, \quad \overset{2}{SCE} \cong \underbrace{78{,}2655}, \quad \overset{3}{R^2} \cong \underbrace{0{,}7459}$$

Recordemos la fórmula $\hat{\beta} = (x'x)^{-1} x'y$

$$x'y = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_{12} y_{i2} \end{bmatrix}$$

1 $$\hat{\beta} = \overset{(x'x)^{-1}}{\begin{bmatrix} 0{,}6477 & -0{,}0410 & -0{,}0639 \\ -0{,}0410 & 0{,}0071 & -0{,}0011 \\ -0{,}0639 & -0{,}0011 & 0{,}0152 \end{bmatrix}} \cdot \overset{x'y}{\begin{bmatrix} 91 \\ 699 \\ 448 \end{bmatrix}} = \begin{bmatrix} 1.6545 \to \beta_0 \\ 0.7391 \to \beta_1 \\ 0.2258 \to \beta_2 \end{bmatrix}$$

[ ]

2 $$SCE = \boxed{\sum_{i=0}^{n} (\hat{y}_i - \bar{y})^2} = \sum \hat{y}_i^2 - n\bar{y}^2 = \hat{y}'\hat{y} - n\bar{y}^2$$

$y = x\hat{\beta}$

$(AB)' = B'A'$

$A^{-1} \cdot A = I$

$$= [(x\hat{\beta})'(x\hat{\beta})] - n\bar{y}^2$$

$$= \hat{\beta}' x' x \hat{\beta} - n\bar{y}^2$$

$$= \hat{\beta}' \underbrace{x'x [(x'x)^{-1}}_{I} x'y] - n\bar{y}^2$$

$$= \hat{\beta}' x' y - n\bar{y}^2$$

$\sum y_i$

$$= [1.6545 \quad 0.7391 \quad 0.2258] \begin{bmatrix} 91 \\ 699 \\ 448 \end{bmatrix} - \underbrace{12 \cdot \left(\frac{91}{12}\right)^2}_{*}$$

$$= \boxed{768{,}35} - \underset{*}{690{,}08} \approx 78.26$$

3 $$R^2 = \frac{SCE}{SCT} = \frac{78{,}26}{104{,}9167} \approx 0.7459$$

b. Evalúe formalmente la hipótesis nula que $\beta_1 = 2 \times \beta_2$ frente a la alternativa que $\beta_1 \neq 2 \times \beta_2$. Considere un nivel de significancia del 5%. 95% de confianza

1. Prueba t

Pasos 1. Formalizar la hipótesis

$H_0: \beta_1 = 2 \times \beta_2$ $\quad$ $H_1: \beta_1 \neq 2 \times \beta_2$

$\beta = \boxed{\beta_1 - 2\beta_2 = 0}$

Paso 2. Prueba de contraste (estadístico "crítico")

$$\hat{t}^{\alpha/2}_{n-k-1} = \frac{\hat\beta_j - \beta}{se(\hat\beta_j)}$$

$$= \frac{\hat\beta_1 - 2\cdot\hat\beta_2 - 0}{\sqrt{var(\hat\beta_1 - 2\cdot\hat\beta_2)}} = \frac{\hat\beta_1 - 2\cdot\hat\beta_2}{\sqrt{V(\beta_1) - V(2\beta_2) - 2Cov(\beta_1, -2\beta_2)}}$$

$$= \frac{\hat\beta_1 - 2\cdot\hat\beta_2}{\sqrt{V(\beta_1) - 4V(-\beta_2) - 2\cdot(-2)\,cov(\beta_1, \beta_2)}}$$

$$= \frac{\hat\beta_1 - 2\hat\beta_2}{\sqrt{var(\hat\beta_1) + 4V(\hat\beta_2) + 4\,cov(\beta_1, \beta_2)}}$$

$$= \frac{0.7391 + 2\cdot 0.2258}{\sqrt{0.021 + 4\cdot 0.045 + 4\cdot(-0.0033}} \approx 0.6213$$

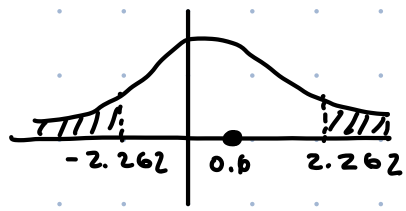
Paso 3. Estadístico teórico

$$t^{1-\frac{\alpha}{2}}_{n-k-1} = t^{1-\frac{0.05}{2}}_{12-2-1} = t^{0.975}_{9} = 2.262$$

$n = 12$

$K = 2$

$\alpha = 0.05$



Paso 4. Análisis

Si tomamos $\hat{t}^{\alpha/2}_{n-k-1}$ y comparamos con t_{n-k-1} (obtenido de tabla t student) podemos decir con un 95% de confianza que se rechaza la Ho, y con ello, existe evidencia sustantiva del efecto parcial de X_1 sobre $\hat{y}$ ($\hat{t} > t$), controlando por el resto de las variables (ceteris paribus)

c. Implemente un contraste de regresión (o prueba de significancia global) e interprete su resultado.

$M_R \quad y = \beta_0 \qquad SSR_R \quad R^2_R$

$M_{NR} = y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \qquad SSR_{NR} \quad R^2_{NR}$

3 Modelo restringido

Paso 1. Formalizar hipótesis

$H_0 : \beta_1 = \beta_2 = 0$ Modelo restringido

$H_1 : \beta_1 \neq 0$ o $\beta_2 \neq 0$ Modelo sin restringir

$$\frac{\cancel{SEC}}{SST} = 0$$

Paso 2. Prueba contraste

$$F = \frac{SSR_R - SSR_{NR}}{SSR_R} \cdot \frac{(n-k-1)}{q} \xrightarrow{SST} \frac{R^2_{NE} - R^2_R}{1 - R^2_{NE}} \cdot \frac{(n-k-1)}{q}$$

$$\hat{F}_{(q,n-k-1)} = \frac{0.7459 - 0}{1 - 0.7459} \cdot \frac{(12-2-1)}{2} \approx 13.21$$

Paso 3. Estadístico teórico $F_{q,n-k-1}$

$$F_{q,n-k-1} = F_{2,12-2-1} = F_{2,9} = 4.26$$

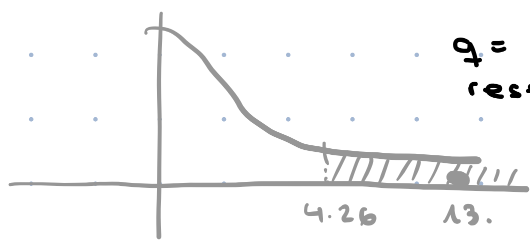
$n = 12$

$k = 2$

$q = 2$

$gl = n-k-1$
grado de libertad del modelo NR

$q =$
restricciones de k parámetro



Paso 4. Análisis

Si tomamos $\hat{F}_{(q,n-k-1)}$, y comparamos con $F_{(q,n-k-1)}$ obtenido de la tabla podemos decir con un 95 % de confianza que se rechaza la Ho, y con ello, existen evidencia concluyente del efecto de las variables incoporadas en las 2 restricciones de exclusión, ceteris paribus

TEMA III (TAREA 1)

Considere el siguiente modelo de regresión lineal, donde se cumplen todos los supuestos desarrollados en clase:

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

En base a una muestra aleatoria de 5 observaciones, sabemos que:

$$\boldsymbol{y} = \begin{bmatrix} 3 \\ 1 \\ 8 \\ 3 \\ 5 \end{bmatrix}, \qquad \boldsymbol{X} = \begin{bmatrix} 1 & 3 & 5 \\ 1 & 1 & 4 \\ 1 & 5 & 6 \\ 1 & 2 & 4 \\ 1 & 4 & 6 \end{bmatrix}, \qquad (\boldsymbol{X}'\boldsymbol{X})^{-\mathbf{1}} = \begin{bmatrix} 26.7 & 4.5 & -8.0 \\ 4.5 & 1.0 & -1.5 \\ -8.0 & -1.5 & 2.5 \end{bmatrix}.$$

A partir de estos antecedentes, se pide que:

a. Estime la regresión por MCO y luego calcule el vector de residuos ($\hat{\boldsymbol{u}}$).

b. Verifique que los residuos suman 0, que son ortogonales a las variables explicativas, y que la suma de cuadrados residuales es igual a 1.5.

c. Evalúe formalmente la hipótesis nula que $\beta_1 = \beta_2$. Considere un nivel de significancia del 5%.