ECONOMETRÍA I

Ayudantía 7

Profesor: *Víctor Macías E.*
Ayudante: *Juan Felipe Ly*
07 de Octubre del 2021

Comentes I

1. Si en el siguiente modelo de regresión lineal simple la varianza de X es cero,
$$y_i = \beta_0 + \beta_1 Xi + u_i$$
entonces $\hat{\beta}_1$ es igual al parámetro poblacional. Comente.

R: Al estar diviendo por 0, se indetermina.

2. Una de las principales ventajas del test t es que es insensible al tamaño de la muestra, y por lo tanto, sus resultados son siempre confiables. Comente.

R: Mientras mayor es el tamaño muestral, menor es la varianza de $\hat{\beta}$ y, por ende, mayor es el valor del test t. En consecuencia, es más probable que el coeficiente resulte ser estadísticamente signifcativo.

3. Es equivalente realizar un test F de significacia global del modelo y que este resulte significativo y realizar tests de significacia individual para todos los parámetros y que resulten significativos. Comente.

R: Falso, aunque realicemos todos los test de signifcancia individual para las pendientes del modelo este resultado no va a ser equivalente al test de signifcancia global, ya que este último considera la correlación entre las variables explicativas. Puede existir un caso extremo en que las variables tengan una correlación muy alta y esta varianza conjunta explique el comportamiento de la variable independiente, aunque cada variable por separado no sea capaz de explicar el comportamiento de Y.

4. Si en el siguiente modelo de regresión lineal simple,
$$y_i = \beta_0 + \beta_1 X_i + u_i$$
$\hat{\beta}_1 = 13$, entonces el test de significancia global del modelo llevará a rechazar la hipótesis nula. Comente.

R: Falso. Falta información respecto al nivel de significancia y el R^2 , sin embargo este último resulta ser al cuadrado de la correlación entre X e Y.

Comentes II

1. Una tienda de ropa usada está interesada en determinar cómo el género de un cliente afecta el monto total que gastará una vez que compra en la tienda. Para ello definió un modelo de la forma $Y_i = \beta_0 + \beta_1 \cdot X_i + u_i$, donde Y representa el monto total gastado en la tienda, mientras que X es el género del cliente, por lo que, al estimar usando Mínimo Cuadrado Ordinarios, se encontró que $\hat{\beta}_1 = 3500$, lo que representa la diferencia en el gasto entre un hombre y una mujer cualquiera. Comente.

R: Se tiene el problema de la interpretación de la pendiente, que en este caso sería una diferencia promedio y no de cualquier hombre o mujer de la muestra.

2. Usted está interesado en testear que el retorno a la educación es diferente para hombres y mujeres, pero no está seguro respecto al modelo a utilizar. Un amigo le recomienda estimar el siguiente modelo:

$$ln(salario_i) = \beta_0 + \beta_1 educ_i + \beta_2 sexo + u_i$$

¿Está de acuerdo con la recomendación de su amigo? Comente.

R: Para testear que el retorno a la educación es diferente para hombres y mujeres se debe agregar la variable interactiva en el modelo:

$$ln(salario_i) = \beta_0 + \beta_1 educ_i + \beta_2 sexo + \beta_3 (educ_i \times sexo_i) + u_i$$

3. Usted está interesada en estimar la relación entre horas de estudio y rendimiento académico de los estudiantes de la UDP, para lo que deberá recolectar datos. Un amigo le recomienda generar una encuesta, subirla al grupo de Facebook de los estudiantes de Ingeniería Comercial de la UDP y sortear un iPad para incentivar la participación. ¿Está usted de acuerdo con la recomendación de su amigo? Comente.

R: Claramente no se cumple el supuesto de muestra aleatoria. Además las estimaciones estarán sesgadas.

4. Usted está interesada testear que el retorno (en términos de la nota final de econometría) de una hora adicional de estudio es diferente para hombres y mujeres, pero no está segura respecto al modelo a utilizar. Un amigo le recomienda estimar el siguiente modelo:

$$nota_i = \beta_0 + \beta_1 horasestudio_i + \beta_2 sexo_i + u_i$$

¿Está de acuerdo con la recomendación de su amigo? Comente.

R: Para testear que el retorno (en términos de la nota final de econometría) de una hora adicional de estudio es diferente para hombres y mujeres se debe agregar la variable interactiva en el modelo:

$$nota_i = \beta_0 + \beta_1 horasestudio_i + \beta_2 sexo_i + \beta_3 (horasestudio_i \times sexo_i) + u_i$$

Analíticos

1. En la tabla 1 se muestran los resultados de la estimación del siguiente modelo:

$$lsalary_i = \beta_0 + \beta_1 \cdot roe_i + \beta_2 \cdot lsales_i + u_i$$

donde `lsalary` representa el logaritmo natural del salario, `roe` es el retorno al capital anual, y `lsales` representa el logaritmo natural de las ventas anuales.

Table 1:

	Dependent variable:
	lsalary
roe	0.018***
	(0.004)
lsales	0.275***
	(0.033)
Constant	4.362***
	(0.294)
Observations	209
R^2	0.282
Adjusted R^2	0.275
Residual Std. Error	0.482 (df = 206)
F Statistic	40.452*** (df = 2; 206)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 2: Matriz de Varianzas y Covarianzas

	(Intercept)	roe	lsales
(Intercept)	0.086	-0.0004	-0.009
roe	-0.0004	0.00002	0.00002
lsales	-0.009	0.00002	0.001

a) Usando un nivel de significancia de un 5%, testee la siguiente hipótesis:

$$H_0 : \beta_1 = \beta_2, \quad H_1 : \beta_1 \neq \beta_2$$

R: Para testear esta hipótesis se utilizará el método de t-test para coeficientes múltiples:

$$t_c = \frac{\hat{\beta}_1 - \hat{\beta}_2}{e.e.(\hat{\beta}_1 - \hat{\beta}_2)}$$

Primero estimamos la varianza:

$$V(\hat{\beta}_1 - \hat{\beta}_2) = V(\hat{\beta}_1) + V(\hat{\beta}_2) - 2COV(\hat{\beta}_1, \hat{\beta}_2)$$

$$V(\hat{\beta}_1 - \hat{\beta}_2) = 0.00002 + 0.001 - 2 \cdot 0.00002$$

$$V(\hat{\beta}_1 - \hat{\beta}_2) = 0.00098$$

$$e.e(\hat{\beta}_1 - \hat{\beta}_2) = \sqrt{V(\hat{\beta}_1 - \hat{\beta}_2)} = \sqrt{0.00098} = 0.0313$$

Por lo tanto:

$$t^c = \frac{0.018 - 0.275}{0.0313} = -8.209$$

El valor crítico al 5% de una distribución t con 206 ($n - k - 1$) grados de libertad es 1.96. Por lo tanto, dado que $|t^c| > 1.96$, se rechaza la hipótesis nula.

2. En la tabla 3 se muestran los resultados de la estimación del siguiente modelo:

$$salary_i = \beta_0 + \beta_1 \cdot roe_i + \beta_2 \cdot lsales_i + u_i$$

donde `salary` representa el salario en miles de dólares, `roe` es el retorno al capital anual, y `lsales` representa el logaritmo natural de las ventas anuales en millones de unidades.

Table 3:

	Dependent variable:
	salary
roe	22.674**
	(10.982)
lsales	286.265***
	(92.332)
Constant	−1,482.294*
	(815.971)
Observations	209
R^2	0.057
Adjusted R^2	0.048
Residual Std. Error	1,338.984 (df = 206)
F Statistic	6.247*** (df = 2; 206)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 4: Matriz de Varianzas y Covarianzas

	(Intercept)	roe	lsales
(Intercept)	665, 808.500	-3, 102.766	-72, 828.300
roe	-3, 102.766	120.596	124.263
lsales	-72, 828.300	124.263	8, 525.166

a) Usando un nivel de significancia de un 5%, testee la siguiente hipótesis:

$$H_0 : \beta_1 = \beta_2, \quad H_1 : \beta_1 \neq \beta_2$$

R: Para testear esta hipótesis se utilizará el método de t-test para coeficientes múltiples:

$$t_c = \frac{\hat{\beta}_1 - \hat{\beta}_2}{e.e.(\hat{\beta}_1 - \hat{\beta}_2)}$$

Primero estimamos la varianza:

$$V(\hat{\beta}_1 - \hat{\beta}_2) = V(\hat{\beta}_1) + V(\hat{\beta}_2) - 2COV(\hat{\beta}_1, \hat{\beta}_2)$$

$$V(\hat{\beta}_1 - \hat{\beta}_2) = 120.596 + 8525.166 - 2 \cdot 124.263$$

$$V(\hat{\beta}_1 - \hat{\beta}_2) = 8397.236$$
$$e.e(\hat{\beta}_1 - \hat{\beta}_2) = \sqrt{V(\hat{\beta}_1 - \hat{\beta}_2)} = \sqrt{8397.236} = 91.636$$

Por lo tanto:

$$t^c = \frac{22.674 - 286.265}{91.636} = -2.876$$

El valor crítico al 5% de una distribución t con 206 $(n - k - 1)$ grados de libertad es 1.96. Por lo tanto, dado que $|t^c| > 1.96$, se rechaza la hipótesis nula.

3. En la tabla 5 se muestran los resultados de la estimación del siguiente modelo:

$$lwage_i = \beta_0 + \beta_1 \cdot exper_i + \beta_2 \cdot exper_i^2 + \beta_3 \cdot hours + u_i$$

donde `lwage` representa el logarítmo natural del salario, `roe` es el retorno al capital anual, y `lsales` representa el logaritmo natural de las ventas anuales en millones de unidades.

Table 5:

	Dependent variable:
	lwage
exper	0.145***
	(0.011)
expersq	−0.008***
	(0.001)
hours	−0.00005***
	(0.00001)
Constant	1.205***
	(0.042)
Observations	4,360
R^2	0.057
Adjusted R^2	0.056
Residual Std. Error	0.517 (df = 4356)
F Statistic	87.784*** (df = 3; 4356)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 6: Matriz de Varianzas y Covarianzas

	(Intercept)	exper	expersq	hours
(Intercept)	0.002	-0.0003	0.00002	-0.00000
exper	-0.0003	0.0001	-0.00001	-0.00000
expersq	0.00002	-0.00001	0.00000	0
hours	-0.00000	-0.00000	0	0

a) Usando un nivel de significancia de un 5%, testee la siguiente hipótesis:

$$H_0 : \beta_1 = \beta_2, \quad H_1 : \beta_1 \neq \beta_2$$

R: Para testear esta hipótesis se utilizará el método de t-test para coeficientes múltiples:

$$t_c = \frac{\hat{\beta}_1 - \hat{\beta}_2}{e.e.(\hat{\beta}_1 - \hat{\beta}_2)}$$

Primero estimamos la varianza:

$$V(\hat{\beta}_1 - \hat{\beta}_2) = V(\hat{\beta}_1) + V(\hat{\beta}_2) - 2COV(\hat{\beta}_1, \hat{\beta}_2)$$

$$V(\hat{\beta}_1 - \hat{\beta}_2) = 0.0001 + 0.00000 + 2 \cdot 0.00001$$

$$V(\hat{\beta}_1 - \hat{\beta}_2) = 0.00012$$

$$e.e(\hat{\beta}_1 - \hat{\beta}_2) = \sqrt{V(\hat{\beta}_1 - \hat{\beta}_2)} = \sqrt{0.00012} = 0.0109$$

Por lo tanto:

$$t^c = \frac{0.145 + 0.008}{0.0109} = 14.036$$

El valor crítico al 5% de una distribución t con 4356 ($n - k - 1$) grados de libertad es 1.96. Por lo tanto, dado que $|t^c| > 1.96$, se rechaza la hipótesis nula.

4. En la tabla 7 se muestran los resultados de la estimación del siguiente modelo:

$$lprice_i = \beta_0 + \beta_1 \cdot rooms_i + \beta_2 \cdot crime_i + u_i$$

donde `lprice` representa el logarítmo natural del precio de las casas, `rooms` representa la cantidad de habitaciones que tiene la casa, y `crime` es la cantidad de crimenes cometidos per cápita.

Table 7:

	Dependent variable:
	lprice
rooms	0.317^{***}
	(0.018)
crime	-0.019^{***}
	(0.001)
Constant	8.022^{***}
	(0.113)
Observations	506
R^2	0.560
Adjusted R^2	0.558
Residual Std. Error	0.272 (df = 503)
F Statistic	319.453^{***} (df = 2; 503)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 8: Matriz de Varianzas y Covarianzas

	(Intercept)	rooms	crime
(Intercept)	0.013	-0.002	-0.00004
rooms	-0.002	0.0003	0.00001
crime	-0.00004	0.00001	0.00000

a) Usando un nivel de significancia de un 5%, testee la siguiente hipótesis:

$$H_0 : \beta_1 = \beta_2, \quad H_1 : \beta_1 \neq \beta_2$$

R: Para testear esta hipótesis se utilizará el método de t-test para coeficientes múltiples:

$$t_c = \frac{\hat{\beta}_1 - \hat{\beta}_2}{e.e.(\hat{\beta}_1 - \hat{\beta}_2)}$$

Primero estimamos la varianza:

$$V(\hat{\beta}_1 - \hat{\beta}_2) = V(\hat{\beta}_1) + V(\hat{\beta}_2) - 2COV(\hat{\beta}_1, \hat{\beta}_2)$$

$$V(\hat{\beta}_1 - \hat{\beta}_2) = 0.0003 + 0.00000 - 2 \cdot 0.00001$$

$$V(\hat{\beta}_1 - \hat{\beta}_2) = 0.00028$$

$$e.e(\hat{\beta}_1 - \hat{\beta}_2) = \sqrt{V(\hat{\beta}_1 - \hat{\beta}_2)} = \sqrt{0.00028} = 0.0167$$

Por lo tanto:

$$t^c = \frac{0.317 + 0.019}{0.0167} = 20.08$$

El valor crítico al 5% de una distribución t con 503 ($n - k - 1$) grados de libertad es 1.96. Por lo tanto, dado que $|t^c| > 1.96$, se rechaza la hipótesis nula.