Solemne de Econometría I

Tiempo: 120 minutos

Semestre Otoño 2019

Profesores: *Fernando Díaz*
Víctor Macías
Ayudantes: *Sebastián González*
Juan Felipe Ly
Mauricio Vásquez

Pregunta 1 (30 puntos)

Considere el siguiente modelo:

$$Y_i = \alpha + \beta X_i + u_i$$

donde u_i es independiente e idénticamente distribuido con media 0 y varianza σ^2

La variable X tiene las siguientes realizaciones: $X_1 = 1$, $X_2 = 2$, $X_3 = 3$, $X_4 = 4$, $X_5 = 5$, $X_6 = 6$.

Un econometrista estima la pendiente de esta regresión usando la siguiente expresión:

$$\hat{\beta} = \frac{1}{8}(Y_6 + Y_5 - Y_2 - Y_1)$$

(a) (15 puntos) Muestre que este estimador es insesgado

(b) (15 puntos) Derive su varianza y compare su varianza respecto al estimador de mínimos cuadrados ordinarios. ¿Cuál estimador es más eficiente?

Solución:

(a) $\widehat{\beta} = \frac{1}{8}(\alpha + 6\beta + u_6 + \alpha + 5\beta + u_5 - \alpha - 2\beta - u_2 - \alpha - \beta - u_1)$

$\widehat{\beta} = \frac{1}{8}(8\beta + u_6 + u_5 - u_2 - u_1)$

$\widehat{\beta} = \beta + \frac{1}{8}(u_6 + u_5 - u_2 - u_1)$

$E(\widehat{\beta}) = \beta$, porque $E(u_i) = 0$, $\forall\ i$

Por lo tanto, el estimador es insesgado.

(b) La varianza del estimador $\hat{\beta} = \frac{1}{8}(Y_6 + Y_5 - Y_2 - Y_1)$ es:

$$Var(\widehat{\beta}) = \frac{1}{64} Var(u_6 + u_5 - u_2 - u_1) = \frac{4\sigma^2}{64} = \frac{\sigma^2}{16}$$

donde $cov(u_i, u_j) = 0$, $\forall\ i \neq j$. La varianza del estimador de MCO es:

$$Var(\widehat{\beta}^{MCO}) = \frac{\sigma^2}{\sum_{i=1}^{N}(x_i - \overline{x})^2} = \frac{2\sigma^2}{35}$$

donde $\sum_{i=1}^{N}(x_i - \overline{x})^2 = \frac{35}{2}$.

Dado que $\frac{\sigma^2}{16} > \frac{2\sigma^2}{35}$, entonces $\widehat{\beta}^{MCO}$ es más eficiente.

Pregunta 2 (15 puntos)

Un investigador considera que la relación entre consumo (C_t) e ingreso (I_t) debe ser estrictamente proporcional. Por ello, plantea el siguiente modelo:

$$C_t = \beta_1 I_t + u_t$$

(a) (10 puntos) Determine el estimador de MCO de β_1

(b) (5 puntos) ¿Se cumple $\sum_{i=1}^{T} \hat{u}_t = 0$? Justifique su respuesta.

Solución:

(a) El problema a resolver es:

$$\min_{\widehat{\beta}_1} \sum_{i=1}^{T} (C_t - \widehat{\beta}_1 I_t)^2$$

y la condición de primer orden es:

$$-2 \sum_{i=1}^{T} (C_t - \widehat{\beta}_1 I_t) I_t = 0$$

Despejando $\widehat{\beta}_1$, se tiene:

$$\widehat{\beta}_1^{MCO} = \frac{\sum_{i=1}^{T} C_t I_t}{\sum_{i=1}^{T} I_t^2}$$

(b) $\sum_{i=1}^{T} \hat{u}_t = 0$ no necesariamente se cumple, ya que en este caso no se tiene la ecuación normal $\sum_{i=1}^{T} \hat{u}_t = 0$ como ocurre en el modelo que se estima con un intercepto.

Pregunta 3 (25 puntos)

Considere el siguiente modelo de regresión lineal simple:

$$y = \alpha_{yx} + \beta_{yx} x + u$$

donde α_{yx} y β_{yx} son parámetros (constantes) y u es un término de error. El siguiente modelo invierte los roles de las variables x e y:

$$x = \alpha_{xy} + \beta_{xy} y + v$$

donde α_{xy} y β_{xy} son parámetros (constantes) y v es un término de error.

Suponga que ambos modelos se estiman por Mínimos Cuadrados Ordinarios usando una muestra aleatoria de tamaño n.

(a) (10 puntos) Muestre que el estadígrafo R^2 obtenido de una regresión en que y es la variable dependiente (o explicada) y x es la variable independiente (o explicativa) puede expresarse como:

$$R^2_{yx} = \frac{[\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})]^2}{\sum_{i=1}^n (x_i - \overline{x})^2 \sum_{i=1}^n (y_i - \overline{y})^2}$$

(b) (5 puntos) Muestre formalmente que:

$$\widehat{\beta}_{yx} \widehat{\beta}_{xy} = R^2$$

donde $\widehat{\beta}_{yx}$ es el estimador MCO de una regresión de y sobre x, $\widehat{\beta}_{xy}$ es el estimador MCO de una regresión de x sobre y y $R^2 = R^2_{yx} = R^2_{xy}$ (*Ayuda: No necesita haber contestado (a) para responder esta pregunta. Puede usar la expresión para* R^2_{yx} *de la pregunta anterior*).

(c) (10 puntos) Las siguientes cantidades fueron calculadas a partir de una muestra de 200 observaciones:

$$\sum_{i=1}^n x_i = 11,34 \qquad \sum_{i=1}^n y_i = 20,72 \qquad \sum_{i=1}^n x_i^2 = 12,16 \qquad \sum_{i=1}^n y_i^2 = 84,96 \qquad \sum_{i=1}^n x_i y_i = 22,13$$

Estime $\widehat{\beta}_{yx}$, $\widehat{\beta}_{xy}$ para esta muestra en particular y obtenga el R^2 usando la relación demostrada en (b).

Solución:

(a) Por definición,

$$R^2_{yx} = \frac{\sum_{i=1}^n (\widehat{y}_i - \overline{y})^2}{\sum_{i=1}^n (y_i - \overline{y})^2}$$

donde

$$
\begin{aligned}
\widehat{y}_i &= \widehat{\beta}_0^{yx} + \widehat{\beta}_{yx} x_i \\
&= \overline{y} - \widehat{\beta}_{yx}\overline{x} + \widehat{\beta}_{yx} x_i \\
&= \overline{y} + \widehat{\beta}_{yx}\left(x_i - \overline{x}\right)
\end{aligned}
$$

Reemplazando en la definición de R^2_{yx},

$$
R^2_{yx} = \widehat{\beta}^2_{yx} \frac{\sum_{i=1}^n \left(x_i - \overline{x}\right)^2}{\sum_{i=1}^n \left(y_i - \overline{y}\right)^2}
$$

El estimador $\widehat{\beta}_{yx}$ es por definición

$$
\widehat{\beta}_{yx} = \frac{\sum_{i=1}^n \left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{\sum_{i=1}^n \left(x_i - \overline{x}\right)^2}
$$

Reemplazando arriba, obtenemos

$$
\begin{aligned}
R^2_{yx} &= \left[\frac{\sum_{i=1}^n \left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{\sum_{i=1}^n \left(x_i - \overline{x}\right)^2}\right]^2 \frac{\sum_{i=1}^n \left(x_i - \overline{x}\right)^2}{\sum_{i=1}^n \left(y_i - \overline{y}\right)^2} \\
R^2_{yx} &= \frac{\left[\sum_{i=1}^n \left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)\right]^2}{\sum_{i=1}^n \left(x_i - \overline{x}\right)^2 \sum_{i=1}^n \left(y_i - \overline{y}\right)^2}
\end{aligned}
$$

(b) Usando las definiciones de $\widehat{\beta}_{yx}$ y $\widehat{\beta}_{xy}$,

$$
\begin{aligned}
\widehat{\beta}_{yx} \cdot \widehat{\beta}_{xy} &= \left[\frac{\sum_{i=1}^n \left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{\sum_{i=1}^n \left(x_i - \overline{x}\right)^2}\right]\left[\frac{\sum_{i=1}^n \left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{\sum_{i=1}^n \left(y_i - \overline{y}\right)^2}\right] \\
&= \frac{\left[\sum_{i=1}^n \left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)\right]^2}{\sum_{i=1}^n \left(x_i - \overline{x}\right)^2 \sum_{i=1}^n \left(y_i - \overline{y}\right)^2}
\end{aligned}
$$

(c)

$$
\begin{aligned}
\widehat{\beta}_{yx} &= \frac{\sum_{i=1}^n \left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{\sum_{i=1}^n \left(x_i - \overline{x}\right)^2} \\
&= \frac{\left(\sum_{i=1}^n x_i y_i\right) - n\overline{x}\overline{y}}{\left(\sum_{i=1}^n x_i^2\right) - n\overline{x}^2}
\end{aligned}
$$

Usando los datos proporcionados,

$$\begin{aligned}
\overline{x} &= \frac{\sum_{i=1}^{n} x_i}{n} = \frac{11,34}{200} \\
&= 0,0567 \\
\overline{y} &= \frac{\sum_{i=1}^{n} y_i}{n} = \frac{20,72}{200} \\
&= 0,1036
\end{aligned}$$

Sustituyendo arriba,

$$\begin{aligned}
\widehat{\beta}_{yx} &= \frac{22,13 - 200\,(0,0567)\,(0,1036)}{12,16 - 200\,(0,0567)^2} \\
&= \frac{20,955176}{11,517022} \\
&= 1,820
\end{aligned}$$

Procedemos del mismo modo para obtener $\widehat{\beta}_{xy}$

$$\begin{aligned}
\widehat{\beta}_{xy} &= \frac{\sum_{i=1}^{n} (x_i - \overline{x})\,(y_i - \overline{y})}{\sum_{i=1}^{n} (y_i - \overline{y})^2} \\
&= \frac{\left(\sum_{i=1}^{n} x_i y_i\right) - n\overline{xy}}{\left(\sum_{i=1}^{n} y_i^2\right) - n\overline{y}^2}
\end{aligned}$$

Usando los datos proporcionados,

$$\begin{aligned}
\widehat{\beta}_{xy} &= \frac{22,13 - 200\,(0,0567)\,(0,1036)}{84,96 - 200\,(0,1036)^2} \\
&= \frac{20,955176}{82,813408} \\
&= 0,253
\end{aligned}$$

El R^2 es por definición igual a

$$\begin{aligned}
R^2 &= \frac{\left[\sum_{i=1}^{n} (x_i - \overline{x})\,(y_i - \overline{y})\right]^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2 \sum_{i=1}^{n} (y_i - \overline{y})^2} \\
&= \frac{\left[\left(\sum_{i=1}^{n} x_i y_i\right) - n\overline{xy}\right]^2}{\left[\left(\sum_{i=1}^{n} x_i^2\right) - n\overline{x}^2\right]\left[\left(\sum_{i=1}^{n} y_i^2\right) - n\overline{y}^2\right]}
\end{aligned}$$

Usando los datos proporcionados,

$$
\begin{aligned}
R^2 &= \frac{20,955176^2}{11,517022 \cdot 82,813408} \\
&= 0,4604
\end{aligned}
$$

Puede verificarse que

$$\widehat{\beta}_{yx} \cdot \widehat{\beta}_{xy} = 0,4604$$

Pregunta 4 (30 puntos)

Un analista político de la oposición está estudiando si los gastos de campaña afectan los resultados de las elecciones. Dado que desde el gobierno se trabaja en un proyecto de ley que establece que todos los gastos de campaña deben ser financiados con fondos públicos, los resultados del estudio pueden ser muy relevantes para la aprobación de este proyecto. El modelo general que el analista ha estimado es el siguiente:

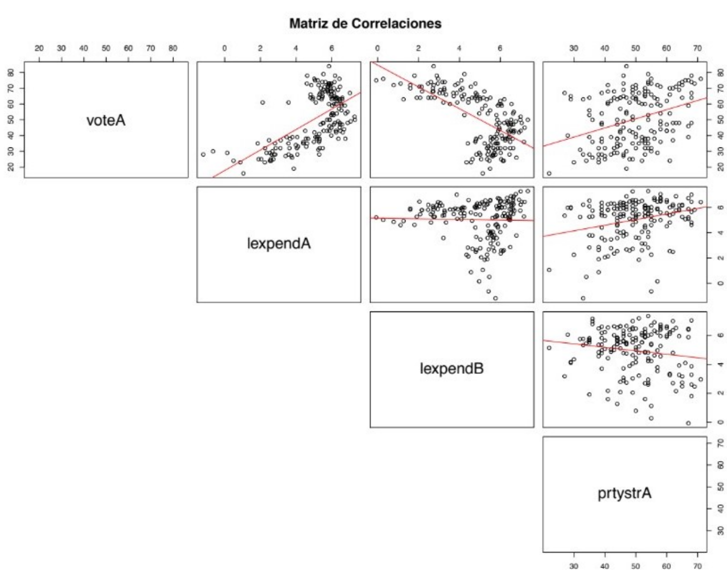
$$voteA = \beta_1 + \beta_2 lexpendA + \beta_3 lexpendB + \beta_4 prtystrA + \mu \tag{1}$$

donde *voteA* es el porcentaje de votos recibidos por el candidato A , *expendA* y *expendB* son los gastos del campaña del candidato A y del candidato B, respectivamente y *prtystrA* es una medida de la popularidad del partido del candidato A, medida como el porcentaje de votos que obtuvo el partido del candidato A en la elección presidencial más reciente. En la ecuación (1), *lexpendA* y *lexpendB* corresponden al logaritmo natural de los gastos del campaña del candidato A y B, respectivamente. El Cuadro 1 presenta la estadística descriptiva para las variables del estudio. En el Gráfico 1, se presentan los gráficos de dispersión entre las mismas.

Cuadro 1: Estadìstica Descriptiva

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
voteA	173	50.503	16.785	16	36	65	84
expendA	173	310.611	280.986	0.300	81.630	457.410	1,470.670
expendB	173	305.088	306.278	0.930	60.050	450.720	1,548.190
prtystrA	173	49.757	9.984	22	44	56	71
lexpendA	173	5.026	1.602	−1.197	4.402	6.126	7.293
lexpendB	173	4.944	1.571	−0.073	4.095	6.111	7.345

Grafico 1: Scatter Plots entre Variables



Nota: La línea corresponde a la FRM para un modelo de regresión simple entre las variables correspondientes.

Los resultados de las estimaciones preliminares que ha llevado a cabo el analista son los siguientes:

Cuadro 2: Estimación por MCO

	Dependent variable: vote A				
	1	2	3	4	5
lexpendA	6.506***			6.342***	6.083***
	(0.628)			(0.373)	(0.382)
lexpendB		-6.917***		-6.757***	-6.615***
		(0.623)		(0.380)	(0.379)
prtystrA			0.581***		0.152**
			(0.121)		(0.062)
Constant	17.806***	84.703***	21.575***	52.039***	45.079***
	(3.313)	(3.229)	(6.121)	(2.750)	(3.926)
Observations	173	173	173	173	173
R2	0.385	0.419	0.120	0.785	0.793
Adjusted R2	0.382	0.416	0.114	0.783	0.789
Residual Std. Error	13.197	12.829	15.795	7.825	7.712
F Statistic	107.238***	123.425***	23.226***	310.694***	215.227***
Note:	*p <0.1 **p <0.05 ***p <0.01 - Std. Errors in ()				

Cuadro 3: Matriz de Varianza Covarianza: Modelo 5

	Intercept	lexpendA	lexpendB	prtystrA
Intercept	15,416	-0,395	-0,874	-0,176
lexpendA	-0,395	0,146	-0,003	-0,007
lexpendB	-0,874	-0,003	0,144	0,004
prtystrA	-0,176	-0,007	0,004	0,004

El analista ha estimado un modelo de regresión simple para cada una de las variables explicativas contra $voteA$ (modelos (1) al (3)), una especificación en que $voteA$ se estima utilizando como variables explicativa los gastos de campaña de ambos candidatos (modelo (4)) y finalmente una especificación que incluye, además, la medida de la popularidad del partido del candidato A (modelo 5) y que corresponde al modelo general presentado en la ecuación (1). Los resultados de estas estimaciones se presentan en el Cuadro 2. La matriz de varianzas y covarianzas para la estimación del modelo (5) se presenta en el Cuadro 3.

En base a estos resultados:

(a) (5 puntos) Comente respecto de los resultados obtenidos por el analista: ¿Tienen los coeficientes obtenidos los signos esperados? En particular, ¿cómo se relacionan los resultados de las estimaciones de los modelos (1), (2) y (3) con los gráficos de dispersión presentados en el gráfico 1? ¿Son consistentes? Explique sus respuestas.

(b) (5 puntos) Comparando los resultados obtenidos para los modelos (4) y (5), ¿le parece a usted que la popularidad del partido del candidato A es relevante para explicar el porcentaje de votos recibidos por ese candidato? ¿Le parece a usted que esta variable debe ser considerada? ¿Puede testear formalmente su respuesta? Si puede, establezca la forma del test, incluyendo el estadígrafo de prueba, las hipótesis nula y alternativa, la distribución muestral del estadígrafo de prueba bajo la nula, sus grados de libertad y los criterios de no rechazo o rechazo.

(c) (5 puntos) ¿Cuál es la interpretación del coeficiente β_1 (intercepto o constante)? Al analista le preocupa el hecho de que el estimador de la constante es muy diferente entre el modelo (1) y los modelos (4) y (5). ¿Puede explicar por qué sucede esto?

(d) (5 puntos) En base a los resultados obtenidos para el modelo (5), lleve a cabo un test de significancia global del modelo utilizando el estadístico R^2. En su respuesta, establezca la forma del test, incluyendo el estadígrafo de prueba, las hipótesis nula y alternativa, la distribución muestral del estadígrafo de prueba bajo la nula, sus grados de libertad y los criterios de no rechazo o rechazo.

(e) (5 puntos) Construya un intervalo de confianza al 10% para el coeficiente de la variable *lexpendA* obtenida en el modelo (5).

(f) (5 puntos) Utilizando los resultados del modelo (5), testee la siguiente afirmación: "Un aumento del 1% en los gastos del candidato A es compensado por un incremento del 1% en los gastos del candidato B". Establezca la forma del test, incluyendo el estadígrafo de prueba, las hipótesis nula y alternativa, la distribución muestral del estadígrafo de prueba bajo la nula, sus grados de libertad y los criterios de no rechazo o rechazo.

Solución:

(a) En efecto, todos los coeficientes obtenidos tienen los signos esperados. Además, los coeficientes de *lexpendA*, *lexpendB* y *prtystrA* , en los modelos 1, 2 y 3, respctivamente, corresponden a las pendientes de la recta estimada (en rojo) de los gráficos de la primera fila del Gráfico 1.

(b) El coeficiente de *prtystrA* en el modelo 5 es positivo y significativo al 5%. Formalmente, un test t es de la forma:

$H_0 = 0$, $H_1 > 0$

$\hat{t} = \frac{\hat{\beta}_{prtystrA} - 0}{\sqrt{\hat{\sigma}^2_{prtystrA}}} \underset{H_0}{\sim} t_{173-4} \Rightarrow \hat{t} = \frac{0.152}{0.062} = 2.45$

$Pr\left[t_{169} < 2.25\right] = 0.9923$

Por ende, $\hat{\beta}_{prtystrA}$ es significativo al 1% en un test unilateral y signficativo al 5% en un test bilateral. En rigor, el test debería ser unilateral, ya que la popularidad de un partido no debería afectar negativamente el resultado de su candidato. Uno puede tener presunciones fundadas que el efecto es positivo.

(c) $\beta_1 = E\left[votaA\,|X_2 = X_3 = X_4 = 0\right)]$ El intercepto es una estimación de la media poblacional la variable dependiente -*voteA* - cuando las variables explicativas toman un valor de cero. La diferencia de valor para $\hat{\beta}_1$ entre el modelo 1 y los modelos 4 y 5 se debe a que el primer modelo adolece de un sesgo por omisión de variables relevantes.

(d) $H_0 : \beta_2 = \beta_3 = \beta_4 = 0$, H_1 : Al menos un β es distinto de cero.

$\hat{F} = \frac{\frac{R^2}{k-1}}{\frac{1-R^2}{n-k}} \underset{H_0}{\sim} F_{k-1,n-k} \Rightarrow \hat{F} = \frac{\frac{0.793}{3}}{\frac{1-0.793}{169}} = 215.8$

$Pr\left[F_{3,169} < 215.8\right] = 0.0000$

Por ende, se rechaza la nula a cualquier nivel estándar de significancia.

(e) $IC^{(1-\alpha)}_{\beta_2} : \hat{\beta}_2 \pm t^{\frac{\alpha}{2}}_{n-k} \cdot \sqrt{v\hat{a}r\left(\hat{\beta}_2\right)} \Rightarrow IC^{(1-10\%)}_{\beta_2} : 6.083 \pm 1.654 \cdot 0.382$

donde $t^{5\%}_{169} = 1.654$.

(f) $H_0 : \beta_2 + \beta_3 = 0$, $H_1 : \beta_2 + \beta_3 \neq 0$

$\hat{t} = \frac{\hat{\beta}_2 + \hat{\beta}_3 - 0}{\sqrt{v\hat{a}r\left(\hat{\beta}_2 + \hat{\beta}_3\right)}} = \frac{\hat{\beta}_2 + \hat{\beta}_3}{\sqrt{\widehat{var}\left(\hat{\beta}_2\right) + \widehat{var}\left(\hat{\beta}_3\right) + 2\widehat{cov}\left(\hat{\beta}_2, \hat{\beta}_3\right)}} \underset{H_0}{\sim} t_{173-4}$

$\Rightarrow \hat{t} = \frac{6.083 - 6.615}{\sqrt{0.146 + 0.144 - 2 \cdot 0.003}} = -0.9983$

$Pr\left[t_{169} < -0.9983\right] = 0.1598$, con un valor p cercano al 32%. Por ende, no puede rechazarse la veracidad de la afirmación.