



$\hat{\beta}_1^2 (x - \bar{x})$

## Prueba 1

Econometría – EAE 2510  
Primer Semestre 2023

### Lea cuidadosamente las siguientes instrucciones:

1. La evaluación es individual; es decir, está absolutamente prohibido comunicarse entre alumnos o usar ayuda de una persona externa al curso.
2. Se les entregará un cuadernillo para sus respuestas. Deberán contestar cada pregunta en el espacio asignado, y podrán escribir por ambos lados. Además, deberán indicar su nombre, sección y número de lista en cada página.
3. No se podrá usar apuntes de ningún tipo. Pueden encontrar una tabla estadística en la última hoja del examen.
4. Pueden utilizar una calculadora científica (sin pantalla grande). Los profesores pueden revisar la calculadora del alumno en cualquier momento.
5. El teléfono y los relojes inteligentes deberán guardarse en la mochila al frente de la sala.
6. Deberán presentar su TUC o carnet de identidad para poder rendir el control.
7. Deben firmar al lado de su nombre en la planilla tanto al comenzar el examen como al entregarlo.  
Es decir, es obligatorio que usted firme dos veces.
8. Es responsabilidad del alumno verificar que su entrega es correcta. Páginas faltantes descubiertas después del final de la evaluación no serán corregidas.

1. (23 puntos) Usted quiere estudiar la relación entre estudiar (medido en horas de estudio por semana) y la nota del individuo al finalizar el semestre. Para ello, cuenta con la siguiente base de datos para una muestra aleatoria de diez individuos:

individuo	horas_estudio	inteligencia_index (IQ)	nota
1	2	150	1
2	3	150	1
3	4	150	2
4	5	100	4
5	5	200	5
6	6	100	6
7	8	200	6
8	10	200	7
9	15	200	7
10	20	200	7

La definición de las variables es la siguiente:

- *horas\_estudio: horas semanales que cada individuo le dedicó a estudiar en el semestre.*
- *nota: nota obtenida por el alumno en el semestre.*
- *inteligencia\_index (IQ): índice que mide la inteligencia del alumno.*

Usted estima el siguiente modelo de regresión lineal múltiple:

$$nota_i = \beta_0 + \beta_1 \times horas\_estudio_i + \beta_2 \times inteligencia\_index + u_i.$$

Además, usted sabe los siguientes datos:

$$\sum_{i=1}^{10} y_i = 46, \quad \sum_{i=1}^{10} horas\_estudio_i \times nota_i = 457,$$

$$\sum_{i=10}^{10} inteligencia\_index_i \times nota_i = 8000, \quad \sum_{i=1}^{10} (\hat{y}_i - \bar{y})^2 = 32.6533,$$

$$(X'X)^{-1} = \begin{bmatrix} 2.00602 & 0.02431 & -0.0127 \\ 0.02431 & 0.0049 & -0.00038 \\ -0.0127 & -0.00038 & 0.0001 \end{bmatrix}$$

Y tiene acceso a la siguiente salida incompleta de STATA al hacer la estimación:

. reg nota horas\_estudio inteligencia\_index

Source	SS
Model	32.6533301
Residual	g
Total	54.4

Number of obs = 10  
F(2, 7) = 5.26  
Prob > F = 0.0404  
R-squared = F  
Adj R-squared = 0.4860  
Root MSE = 1.7626

	nota	Coef.	Std. Err.	t
$\beta_1$	horas_estudio	a	d	e
$\beta_2$	inteligencia_index	b		
$\beta_0$	_cons	c		

Se le pide que en base a los datos anteriores responda las siguientes preguntas. A lo largo del ejercicio, puede redondear a 5 decimales:

- (5 puntos) Use la información del ejercicio y la contenida en la salida de STATA para encontrar los estimadores de MCO faltantes. Es decir, complete (a), (b) y (c) en la tabla.
  - (2 puntos) Encuentre el R2 y la suma de cuadrados residuales. Es decir, (F) y (g) en la tabla.
  - (6 puntos) Encuentre los errores estándar del estimador MCO de  $\beta_1$  y luego obtenga el t-estadístico para la  $H_0$  de que  $\beta_1 = 0$ . Es decir, encuentre los elementos (e) y (d) de la tabla.
  - (2 puntos) Interprete el  $\widehat{\beta}_1$  obtenido por MCO.
  - (4 puntos) Evalúe la significancia global de todas las variables del modelo. Considere un nivel de significancia del 5%. Sea claro en plantear la hipótesis nula y la alternativa, luego construir el estadístico de prueba correspondiente, y finalmente haga e interprete el resultado del test.
  - (4 puntos) En base al modelo anterior, usted estima la predicción  $\widehat{nota}_i$  (o los valores ajustados) para cada individuo. Sin embargo, al llegar a su casa se da cuenta que al preparar el informe se olvidó de copiar el número para uno de los diez individuos. Usted suma las notas predichas para los 9 individuos que sí anotó y esa suma es igual a 37.5. En base a los datos del ejercicio, encuentre el valor ajustado  $\widehat{nota}_i$  para el individuo que falta. Explique su procedimiento.
- (14 puntos) Usted está trabajando para la Universidad Católica de Chile, y su jefe le dice que quieren entender cuáles son los factores que podrían estar relacionados con la asistencia a clase de los alumnos. Para eso, la universidad realizó una encuesta a 1200 estudiantes.

La misma incluye información sobre:

$$\begin{aligned} T &= E + R \\ T &= \sum (Y_i - \bar{Y})^2 \\ E &= \sum (\hat{Y}_i - \bar{Y})^2 \\ R &= \sum (Y_i - \hat{Y})^2 \end{aligned}$$

- Número de estudiante (NUM): número único con el que se identifica cada estudiante.
- Horas semanales que asiste a clases de cátedra de forma presencial (CATEDRA\_PRES).
- Horas semanales que asiste a ayudantías de forma presencial (AYUDANTIAS\_PRES).
- Horas semanales que asiste a clases de cátedra de forma virtual (CATEDRA\_VIRTUAL).
- Horas semanales que asiste a ayudantías de forma ~~virtual~~<sup>presencial</sup> (AYUDANTIAS\_VIRTUAL).
- Número de veces en la semana que el metro estuvo cerrado (METRO).
- Número de eventos semanales diurnos organizados por el centro de alumno (EVENTOS\_DIA).
- Número de eventos semanales nocturnos organizados por el centro de alumno (EVENTOS\_NOCHE).

Todos los datos vienen guardados en una base llamada *asistencia.dta*.

- a. (2 puntos) Escriba como cargaría la base de datos en Stata para su posterior uso cerciorándose de que la información cargada previamente es borrada.
- b. (2 puntos) Le comenta su compañero que la variable METRO tiene muchas entradas sin valor (missing value). Escriba cómo cambiaría estas entradas vacías por cero en Stata.  
*Ayuda: una entrada sin valor asignado en una variable numérica en Stata aparece como un punto (.).*
- c. (2 punto) Escriba en lenguaje de Stata cómo crearía una variable que contenga la información sobre las horas totales que alumno asistió a clase (de forma presencial y virtual, incluyendo cátedras y ayudantías). Llámela HORAS\_TOTALES.
- d. (2 puntos) Ahora se le pide que estime un modelo para las horas totales de asistencia a los cursos (variable creada en el punto 1), considerando como variables explicativas el número de veces que el metro estuvo cerrado, el número de eventos sociales (día y noche), incluyendo una constante. Escriba cómo lo estimaría en Stata.
- e. (2 puntos) Le pregunta su jefe si los eventos sociales tienen en su conjunto un efecto significativo en la asistencia a clases por parte de los alumnos. Escriba cómo implementaría dicha prueba en Stata.
- f. (2 puntos) Llega su jefe y le dice que en la base que le dio anteriormente faltaba la información de los estudiantes de agronomía. La base tiene exactamente las mismas variables que *asistencia.dta* y se llama *asistencia\_agronomia.dta*. Escriba cómo incluiría esta información en su base de datos.
- g. (2 puntos) Finalmente, un compañero le pasa información adicional sobre los estudiantes encuestados. Más precisamente, le pasa un archivo con el número de estudiante y la distancia en kilómetros de la vivienda del estudiante al campus (DIST). La base se llama *distancia.dta*. Escriba cómo incorporaría esta información a la base de datos.

3. (19 puntos) Suponga que en uno de sus cursos le piden estudiar la relación entre las ventas de una firma y su gasto en publicidad. Para eso, cuenta con información de 200 locales comerciales.

En primera instancia, usted estima la siguiente regresión por Mínimos Cuadrados Ordinarios (MCO):

$$ventas = \beta_0 + \beta_1 \times gastopublicidad + u,$$

donde las ventas se miden en miles de unidades y el gasto en publicidad en miles de dólares, y obtiene los siguientes resultados:

Source	SS	df	MS	Number of obs	=	200
Model	4078.70471	1	4078.70471	F(1, 198)	=	603.37
Residual	1338.44412	198	6.75981878	Prob > F	=	
Total	5417.14883	199	27.2218534	R-squared	=	0.7529
				Adj R-squared	=	0.7517
				Root MSE	=	2.6

	ventas	Coefficient	Std. err.	t	P> t	[95% conf. interval]
$\beta_1$	gastopublicidad	.0486879				
$\beta_0$	_cons	4.243028	.4385252	9.68	0.000	3.378249 5.107808

- (5 puntos) Interprete la estimación de  $\beta_1$ , y luego evalúe la hipótesis nula de  $\beta_1$  es igual a 0 frente a la alternativa que es distinto de 0. Considere un nivel de significancia del 5%.
- (6 puntos) Ahora decide aprovechar que también cuenta con el gasto en publicidad desagregado. En particular, cuenta con información respecto del gasto en publicidad televisiva, radial y en medios gráficos, todo medido en miles de dólares.

Source	SS	df	MS	Number of obs	=	200
Model	4860.32353	3	1620.10784	F(3, 196)	=	570.27
Residual	556.825297	196	2.84094539	Prob > F	=	0.0000
Total	5417.14883	199	27.2218534	R-squared	=	0.8972
				Adj R-squared	=	0.8956
				Root MSE	=	1.6855

	ventas	Coefficient	Std. err.	t	P> t	[95% conf. interval]
tv		.18853	.0086112	21.89	0.000	.1715475 .2055126
radio		.0457646	.0013949	32.81	0.000	.0430137 .0485156
mediosgraficos		-.0010375	.005871	-0.18	0.860	-.012616 .010541
_cons		2.938889	.3119082	9.42	0.000	2.323762 3.554016

Covariance matrix of coefficients of regress model

e(V)	tv	radio	mediosgraf~s	_cons
$\beta_1$ tv	.00007415			
$\beta_2$ radio	-4.470e-07	1.946e-06		
mediosgraf~s	-.0000178	-3.266e-07	.00003447	
_cons	-.00111549	-.00026573	-.00059102	.09728675

donde  $1.946e-06$  equivale a  $1.946 \times 10^{-6}$ .

A partir de esta información, evalúe la hipótesis nula que los gastos en publicidad televisiva, radial y en medios gráficos tienen la misma correlación con las ventas; es decir, que  $\beta_1 = \beta_2 = \beta_3$ . Considere un nivel de significancia del 5%. Ayuda: no se pide un contraste de regresión o una prueba de significancia global. Piense en cuál sería el modelo restringido bajo la hipótesis nula, y luego construya el estadístico de contraste a partir de la información disponible. Recuerde además que  $gastopublicidad = (tv + radio + mediosgraficos)$ .

- c. (8 puntos) Finalmente, evalúe la hipótesis nula que  $\beta_1 = 3 \times \beta_2$  frente a la alternativa que  $\beta_1 > 3 \times \beta_2$ . Considere un nivel de significancia del 5%.

ANEXO

**t Table**

df	cum. prob.											
	one-tail		0.50		0.25		0.20		0.15		0.10	
	two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.025	0.01	0.005	0.001
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62	
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599	
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924	
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610	
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869	
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959	
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408	
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041	
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781	
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587	
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437	
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318	
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221	
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140	
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073	
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.688	4.015	
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965	
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922	
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883	
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850	
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819	
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792	
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768	
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745	
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725	
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707	
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690	
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674	
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659	
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646	
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551	
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460	
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416	
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390	
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300	

Función de Distribución de la variable F, percentiles 95

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

$$X' Y = \begin{bmatrix} \sum y_i \\ \sum y_i x_1 \\ \sum y_i x_2 \end{bmatrix} = \begin{bmatrix} 46 \\ 457 \\ 8000 \end{bmatrix}$$

$$\begin{bmatrix} 2,00602 & 0,02431 & -0,0127 \\ 0,02431 & 0,0049 & -0,00038 \\ -0,0127 & -0,00038 & 0,0001 \end{bmatrix} \begin{bmatrix} -46 \\ 457 \\ 8000 \end{bmatrix} = \begin{bmatrix} 178,659 \\ 0,31756 \\ 0,04214 \end{bmatrix} \rightarrow \begin{array}{l} c \\ a \\ b \end{array}$$

$$b) SCR = SCT - SCE$$

$$= 54,4 - 32,6533301$$

$$= 21,7466699$$

$$R^2 = \frac{SCE}{SCT} = \frac{32,6533301}{54,4} = 0,60245$$

$$c) \underline{(X' X)^{-1} \cdot SSR} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 6,232991 & 0,075535 & -0,03946 \\ 0,075535 & 0,015225 & -0,00118 \\ -0,03946 & -0,001181 & 0,00031 \end{bmatrix}$$

$\downarrow \downarrow \downarrow$   
 $\frac{205}{10} = 20,5$

varianza  $\beta_1$

$$\text{Error estándar} = \sqrt{0,015225} = 0,12339$$

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

$$t_{\beta_j} = \frac{\hat{\beta}_j - \alpha_j}{s.e \hat{\beta}_j} = \frac{0,31756 - 0}{0,12339} = 2,57363 \rightarrow E(\text{tabla})$$

d. Un  $\Delta^+$  en 1 unidad de hrz semanal de estudio está asociado a un  $\Delta^+$  en prom de 0,31 ~~156~~ unidades en la nota, manteniendo lo demás cte.

$$e. H_0: \beta_1 = \beta_2 = 0$$

$$H_1: \beta_1 \neq 0 \quad o \quad \beta_2 \neq 0$$

$$F = \frac{R^2}{1-R^2} \cdot \frac{(n-k-1)}{K} = \frac{0,6}{1-0,6} \cdot \frac{10-2-1}{2} = 5,25$$

$$F(2,7) = 4,74$$

$F_{\text{OBS}} > F_{\text{TABLA}}$  se rechaza  $H_0$

$$f. \sum_{i=1}^{10} \text{nota}_i - \sum_{i=1}^9 \widehat{\text{nota}}_i = \widehat{\text{nota}}_{10}$$

$$46 - 37,5 = 8,5$$

③ a) Por cada 1000 dólares de gasto de publicidad, en promedio las ventas aumentan 49 unidades.

$$H_0: \beta_1 = 0 \quad t = \sqrt{F} = \sqrt{603,37} = 24,56$$

$$H_1: \beta_1 \neq 0 \quad F = \frac{R^2}{(1-R^2)} \cdot \frac{n-k-1}{K}$$

$$= \frac{0,7529}{1-0,7529} \cdot \frac{200-1-1}{1} = 603,37$$

$$t_{n-k-1}^{1-\alpha/2} = t_{198}^{0,975} \approx t_{100}^{0,975} = 1,984 < 24,56 \rightarrow \text{rechaza}$$

$$F(1,198) \approx F(1,120) = 3,92 < 603,37 \rightarrow \text{rechazo}$$

- b)  $H_0: \beta_1 = \beta_2 = \beta_3$   $\Rightarrow$  cant. restricciones lineales  
 $H_1: \text{Al menos 1 no sea igual}$

$$F = \frac{\overbrace{SCR_r - SCR_{Sr}}^{\frac{R_{Sr}^2 - R_r^2}{1 - R_{Sr}^2}}}{SCR_{Sr}} \cdot \frac{n - k - 1}{q} = \frac{1338,44 - 556,83}{556,83} \cdot \frac{197}{2} = 137,56$$

$$F(2, 196) \approx F(2, 120) = 3,07 < 137,56 \quad \text{Rechazo}$$

c)  $H_0: \beta_1 = 3\beta_2$   $t = \frac{(\beta_1 - 3\beta_2)}{\sqrt{V(\hat{\beta}_1 - 3\hat{\beta}_2)}} = \frac{0,18853 - 3 \cdot 0,0457646}{0,0097131869} \approx 5,27$   
 $H_1: \beta_1 > 3\beta_2$

$$\begin{aligned} V(\hat{\beta}_1 - 3\hat{\beta}_2) &= V(\hat{\beta}_1) + 9V(\hat{\beta}_2) + 2 \cdot 1 \cdot -3 \text{ COV}(\beta_1, \beta_2) \\ &= 0,00007415 + 9 \cdot 0,00001946 - 6 \cdot (-0,000000447) \end{aligned}$$

$$t_{n-k-1}^{1-\alpha} = t_{196}^{0,95} \approx t_{100}^{0,95} = 1,66 < 5,27 \quad \text{Rechazo}$$

### STATA

- use asistencia.dta, clear
- replace METRO=0 if METRO ==.
- gen HORAS\_TOTALES = CATEDRA\_PRES + CATEDRA\_VIRTUAL + AYUDANTIAS\_PRES + AYUDANTIAS\_VIRTUAL
- reg HORAS\_TOTALES METRO EVENTOS\_DIA EVENTOS\_NOCHE
- test EVENTOS\_DIA EVENTOS\_NOCHE
- append using asistencia\_agronomia.dta
- merge 1:1 NUM using distancia.dta