

Econometría 1
Examen Parcial 1
Dr. Francisco J. Cabrera Hernández
Centro de Investigación y Docencia Económicas A.C. (CIDE)

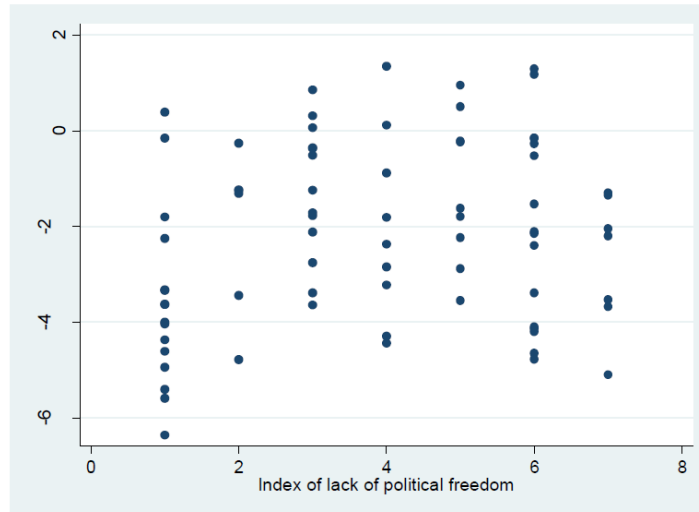
Instrucciones: coloque su nombre en todas las hojas de respuesta; responda a todas las preguntas; puede utilizar calculadora; no se permiten celulares o laptops; no responda sobre las hojas de preguntas. Tiene 1 hora y 50 minutos para responder.

¿Cuáles son las causas del terrorismo? ¿Pobreza? ¿Regímenes Políticos Represivos? ¿Conflictos étnicos que surgen de poblaciones heterogéneas? En la siguiente base de datos de cohorte transversal, se encuentra información sobre fuentes de terrorismo en distintos países donde existen muertes por ataques terroristas. Las variables se definen en la siguiente Tabla:

Variable	Definition
lnftmpop	Logaritmo natural del número de muertes por terrorismo en el país. Entre 1998-2004 y por millón de la población
evmpop	Número de eventos terroristas en el país, entre 1998-2004 y por millón de la población
lngdppc	Logaritmo natural del GDP per cápita en el país
lackpf	Índice de Falta de Libertad Política. En escala de 1 a 7, donde 7 = libertad política extremadamente limitada.
language	Índice de variedad lingüística en el país (escala 0 a 1, donde 0 = no hay variedad lingüística)
ethnic	Índice de variedad étnica en el país (escala 0 a 1, donde 0 = no variedad étnica)
religion	Índice de variedad religiosa en el país (0 to 1 scale, donde 0 = no variedad)
mideast, latinam, easteurope, africa, eastasia	Set de <i>dummies</i> donde = 1 si el país se encuentra en dicha región, = 0 si no se encuentra en una de esas regiones.

1. Interpretación de estimadores y su significancia: (50%)

- a. Abajo está el *scatterplot* entre lnftmpop y lackpf. Dado éste, escriba la ecuación teórica que estimaría si quisiera relacionar estas dos variables en un modelo de regresión lineal usando OLS. Incluya todos los controles que crea necesarios.



- ¿Es la relación entre *lngdppc* y *lnftmpop* endógena? Describa todas las razones posibles.
- En una regresión estimada de $\lnftmpop = \beta_0 + \beta_1 \ln gdppc + v$, demuestre matemáticamente si existe un sesgo por omitir la variable *lackpf* y la dirección de éste. Suponga una relación $lackpf = \gamma_0 - \gamma_1 \ln gdppc + \varepsilon$.
- Dada la tabla de resultados, Interprete la constante y el efecto de *lngdppc* en *lnftmpop*.

Source	SS	df	MS	Number of obs	=	76
Model		1		F(1, 74)	=	4.88
Residual	251.384892	74	3.39709313	Prob > F	=	0.0303
				R-squared	=	
				Adj R-squared	=	0.0492
Total	267.965127	75	3.57286836	Root MSE	=	1.8431

lnftmpop	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
lngdppc	-.3124901	.1414473			-.59433	-.0306502
_cons	-2.047654	.2215154	-9.24	0.000	-2.489033	-1.606274

- ¿Es la variable *lngdppc* estadísticamente significativa? Plantee H_0 y concluya (valor crítico: $\alpha_{0.05}=1.993$; $\alpha_{0.01}=2.644$).
- ¿cuáles son todos los supuestos estándar clave para estimar el estadístico t relacionado con *lngdppc*?
- Obtenga el intervalo de confianza al 99% de confianza estadística.
- Obtenga R^2 e interprétela (SS = “sum of squares” en la Tabla de resultados)
- Suponga que quiere obtener el efecto de *lngdppc* en *lnftmpop* en países con mayor o menor *variedad religiosa*, escriba la ecuación del modelo que estimaría
- ¿cómo interpretaría cada uno de los estimadores de la ecuación planteada en *i*. incluyendo la constante?

- Con la misma base de datos creamos una variable dummy *higdppc* que toma el valor de 1 si el país está arriba de la mediana de GDP y cero si está debajo de la mediana. Y se crea la interacción $hi_lack = higdppc * lackpf$ (30%)

	(6)	(7)	(8)	(9)
Dependent variable:	$\ln(ftmpop)$	$\ln(ftmpop)$	$\ln(ftmpop)$	$\ln(ftmpop)$
Regressor:				
<i>higdppc</i>	-2.84** (1.00)	-3.84 (2.41)	-.38 (1.04)	-.59 (1.08)
<i>lackpf</i>	-.248 (.165)	.131 (1.060)	1.622** (.525)	1.977** (.528)
<i>lackpf</i> ²	—	-.040 (.114)	-.189** (.062)	-.228** (.060)
<i>higdppc</i> × <i>lackpf</i>	.645** (.238)	1.889 (1.216)	—	—
<i>higdppc</i> × <i>lackpf</i> ²	—	-.196 (.141)	—	—
<i>ethnic</i>	—	—	1.462 (1.045)	1.237 (.996)
<i>religion</i>	—	—	-3.007 (1.418)	-3.406* (1.649)
<i>higdppc</i> × <i>ethnic</i>	—	—	-1.297 (1.698)	-.501 (1.801)
<i>higdppc</i> × <i>religion</i>	—	—	2.655 (1.900)	2.511 (2.018)
<i>Mideast</i>	—	—	—	-1.253 (1.312)
Other regional dummies (<i>latinam</i> , <i>easteurope</i> , <i>africa</i> , <i>eastasia</i>)?	No	No	No	Yes
Intercept	-.72 (.84)	-1.52 (2.24)	-4.38** (1.26)	-3.78* (1.55)
F-statistics testing the hypothesis that the population coefficients on the indicated regressors are all zero:				
<i>higdppc</i> × <i>lackpf</i> , <i>higdppc</i> × <i>lackpf</i> ²	—	1.34 (.269)	—	—
<i>lackpf</i> ² , <i>higdppc</i> × <i>lackpf</i> ²	—	4.12 (.020)	—	—
<i>lackpf</i> , <i>lackpf</i> ²	—	1.10 (.339)	4.81 (.011)	7.21 (.002)
<i>higdppc</i> × <i>ethnic</i> , <i>higdppc</i> × <i>religion</i>	—	—	1.24 (.297)	0.79 (.461)
<i>ethnic</i> , <i>religion</i> , <i>higdppc</i> × <i>ethnic</i> , <i>higdppc</i> × <i>religion</i>	—	—	1.45 (.227)	1.66 (.17)
Other regional dummies	—	—	—	0.63 (.642)
Regression summary statistics:				
\bar{R}^2	0.085	0.131	0.117	0.077
R^2	0.122	0.189	0.202	0.229
SER	1.808	1.762	1.769	1.809
<i>n</i>	76	76	74	74

Notas: Errores estándar robustos a heteroscedasticidad entre paréntesis. Debajo de los estadísticos F, entre paréntesis, se encuentran sus respectivos p-values, los cuales también son robustos a heteroscedasticidad. Niveles de significancia dados por +10%, *5%, **1%. Las “other regional dummies” incluidas en la regresión (5) son *latinam*, *easteurope*, *africa*, and *eastasia* (Categoría omitida: Europa del oeste y EE.UU.).

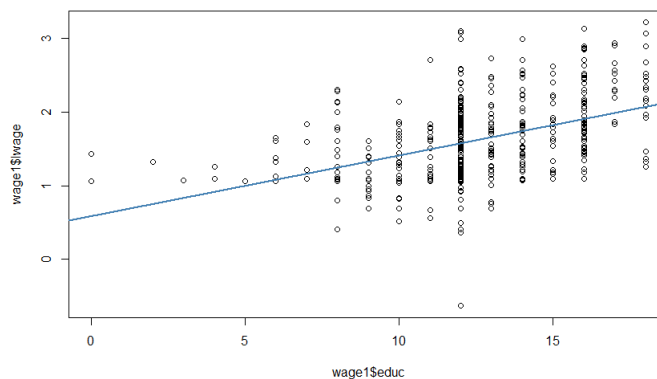
- Escriba la ecuación estimada en la regresión de la columna (6).
- Interprete la constante en la regresión estimada en la columna (6)
- Utilizando el modelo en Columna (6) ¿Es estadísticamente distinto el efecto de *lackpf* en fatalidades por terrorismo en países con GDP por arriba de la mediana respecto a aquellos por debajo de ésta? (valor crítico: $\alpha_{0.05}=1.993$; $\alpha_{0.01}=2.644$).

- d. ¿Usando el modelo de la columna (6) ¿cuál es el efecto *ceteris paribus* de incrementar el índice *lackpf* en países debajo de la mediana de GDP? ¿Es este efecto significativamente distinto de cero al 90%, 95% o 99% de confianza?
- e. El modelo en la Columna (7) incluye la prueba F que plantea la siguiente H_0 : $lackpf^2 = higdppc * lackpf^2 = 0$, con $F = 4.12$ (0.020) ¿Qué puede concluir con esta información respecto a la relación funcional entre *lackpf* y *lnftmpob*?
- f. Dibuje en dos dimensiones la relación entre *lackpf* y *lnftmpob* considerando la información obtenida en d. (hint: concéntrese en dibujar adecuadamente las constantes, el punto máximo/mínimo, si existe, y la forma funcional).
- g. En la regresión de la columna (7) interprete la variable *lackpf* en países por debajo de la mediana de *gdppc*, cuando *lackpf* pasa de 1 a 2 y cuando *lackpf* pasa de 6 a 7.
- h. En la regresión de la columna (7) interprete la variable *lackpf* en países por encima de la mediana de *gdppc*, cuando *lackpf* pasa de 1 a 2 y cuando *lackpf* pasa de 6 a 7.
- i. ¿En conjunto, son significativos los coeficientes de “other regional dummies”? Justifique su respuesta.
- j. ¿Si tuviera que escoger entre los cuatro modelos, con cuál se quedaría y cómo lo justifica? *Hint: al compararlos piense si los modelos son versiones “anidadas”*.

3. Tenemos las siguientes variables en una base de datos llamada “wage1” que contiene 526 observaciones (20%)

wage	Sueldo en dólares por hora
lwage	Logaritmo de wage
educ	Años de educación
exper	Años de experiencia

La línea de regresión y los residuos del sueldo y la educación dados por $lwage = \beta_0 + \beta_1 educ + u$ se grafica abajo:



Partiendo de $var(\hat{\beta}_1) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 var(u_i)}{SST_x^2}$ y dado el *scatterplot* arriba:

- a. ¿Cuál es el supuesto de Gauss-Markov que potencialmente se está violando y cuál es la consecuencia en el sesgo de $\hat{\beta}_1$ y la varianza de $\hat{\beta}_1$?

- b. Realizamos dos regresiones en R Studio con la base descrita:

```
#REGRESION A.
reg <- feols(wage ~ educ + exper, data = wage1)
SSTj <- sum(wage1$educ - mean(wage1$educ))^2
SSR <- sum((fitted(reg) - wage1$wage)^2)
```

La variación de educ (SSTj) es = 4025.43, con SSR = 5548.16 y n = 526. La R^2 de esta regresión es= 0.2251

```
#REGRESION B.
reg1 <- feols(educ ~ exper, data = wage1)
```

La R^2 de esta regresión es = 0.0897

- i. Obtenga SE de $\hat{\beta}_1$ en la REGRESION A considerando:

The estimated sampling variation of the estimated β_j → $se(\hat{\beta}_j) = \sqrt{\widehat{Var}(\hat{\beta}_j)} = \sqrt{\hat{\sigma}^2 / [SST_j(1 - R_j^2)]}$

- ii. ¿Es este error estándar válido? ¿por qué razón?

- c. Suponga que R^2 en REGRESION B. es 0.8972 ¿En cuánto aumenta el error estándar de $\hat{\beta}_1$ en la REGRESION A? ¿Por qué razón y qué supuesto se pone en entredicho?
- d. Dada la nueva varianza de $\hat{\beta}_1$ en REGRESION A, luego de c., y bajo MLR4 ¿es posible que $\hat{\beta}_1$ no sea consistente? ¿Es más eficiente? Explique.
- e. Si no se cumple MLR4 en la REGRESION A. Dibuje dos histogramas que reflejen el cambio en la distribución de $\hat{\beta}_1$ en m muestras repetidas de tamaño $n=526$ y $n=5260$ ¿es $\hat{\beta}_1$ consistente ante el aumento de n ?