

**Serie I (58 puntos):** Resuelva los problemas que se le presentan a continuación, deje constancia de todo su procedimiento. No olvide escribir su respuesta final en lapicero para tener derecho a revisión.

**1. (24 puntos)** Si el salario mensual de un ejecutivo se estima por el siguiente modelo:

$$(\text{Salario})_i = \alpha + \beta_1 (\text{Promedio})_i + \beta_2 (\text{Experiencia})_i + \beta_3 (\text{Edad})_i + \varepsilon_i$$

Donde el **Salario** está medido en Quetzales, el **Promedio** es el promedio universitario con que se graduó, en valores de 1 a 100, la **Experiencia** y la **Edad** se miden en años. La estimación del modelo utilizando OLS da el siguiente resultado:

$$(\text{Salario})_i = 11,583 + 51.9(\text{Promedio})_i + 70.64 (\text{Experiencia})_i + 6.1(\text{Edad})_i + \varepsilon_i$$

(2,000)                      (30.5)                      (50.13)                      (4.8)

En paréntesis se muestran errores estándar.

**a. A partir del contexto del problema, interprete el significado de cada coeficiente, incluyendo la constante.**

Coficiente "promedio": significa que por cada punto más con el que el estudiante se graduó el salario aumenta en proporción de 51.9 quetzales.

Coficiente "experiencia": significa que por cada año adicional de experiencia el salario aumenta en proporciones de 70.64Q.

Coficiente "edad": significa que por cada año adicional su salario aumenta en proporción de 6.1Q.

Constante: significa que si todas las variables fueran cero el salario en quetzales sería de 11,583Q.

**b. ¿Son significativos los coeficientes de forma individual?, utilice un alfa del 10% (n=15)**

	coeficiente	error	t estadístico	valor-p	significativos	
promedio	51.9	30.5	1.701639344	0.116881373	Acepto H0	Acepto H0
experiencia	70.64	50.13	1.409136246	0.186432011	Rechazo H0	Acepto H0
edad	6.1	4.8	1.270833333	0.230006027	Acepto H0	Acepto H0
			gl	11		
			alpha	10%		
			t crit	1.795884819		

\*Ver el excel "Proce 1" para evidencia de procedimiento.

Los coeficientes no son significativos de forma individual puesto a que ninguno pasa la prueba de hipótesis con significancia 10%.

**c. Utilizando la información que se le presenta a continuación estime el R<sup>2</sup> e interprételo:**

$$\sum \widehat{\varepsilon}_t^2 = 48,750$$

$$S_y^2 = 5,000$$

$R^2$  significa qué tan bien mi modelo explica los datos por lo que tener un  $R^2$  de aproximadamente 30% realmente significa que mi modelo no hace un muy buen trabajo en explicar los datos y por lo tanto no es significativo.

e_i^2	var.no.exp	48,750
S_y^2	var.total	70000
R^2	0.3035714	

\*Ver el excel "Proce 1" para evidencia de procedimiento.

2. (34 puntos) Cargue la base de datos que fue enviada a su correo en Gretl y estime un modelo en donde explique la tasa de criminalidad en función de las variables independientes proporcionadas (es posible construir distintos modelos, presente aquel que según su conocimiento es el “mejor”). No olvide explicar en términos generales los pasos que siguió para llegar al mejor modelo. A partir de los resultados obtenidos complete la siguiente tabla (vea la hoja en Excel en donde se provee la explicación de las variables para interpretar con precisión los coeficientes estimados):

Variable Independiente (nombre)	Coefficiente	Significativo al 5%	Interprete el valor del coeficiente
<b>Constante</b>	-492.119	1	Significa la cantidad de tasa de criminalidad cuando todas las demás variables son cero.
Edad = número de hombres entre 14 y 24 años de edad por cada 1 000.	1.06436	1	Significa que por cada unidad adicional de edad de numero de hombres entre 14-24 incrementa la tasa de criminalidad por 1.06
EP = número promedio de años de escolaridad multiplicado por 10 para personas de 25 años o mayores.	1.92255	1	Significa que por cada incremento en el numero promedio de años de escolaridad incrementa 1.92 la tasa de criminalidad.
GX0 = gastos de la policía per cápita a cargo del estado y el gobierno local en 1960.	1.17841	1	Significa que por cada incremento en los gastos de policía incrementa 1.17 la tasa de criminalidad.
GX1 = gastos de la policía per cápita a cargo del estado y el gobierno local en 1959.	N/A	0	La removimos del modelo.
FT = tasa de participación de la fuerza de trabajo por cada 1 000 hombres civiles urbanos entre 14 y 24 años de edad.	N/A	0	La removimos del modelo.

H = número de hombres por cada 1 000 mujeres.	-0.630377	0	Significa que por cada incremento en el número de hombres respecto de cada 1000 mujeres la tasa de criminalidad disminuye por 0.63.
N = tamaño de la población estatal en cientos de miles.	N/A	0	La removimos del modelo.
NB = número de no blancos por cada 1 000 personas.	N/A	0	La removimos del modelo.
U1 = tasa de desempleo de hombres urbanos por cada 100 habitantes de entre 14 y 24 años de edad.	N/A	0	La removimos del modelo
U2 = tasa de desempleo de hombres urbanos por cada 100 habitantes de entre 35 y 39 años de edad.	0.863030	1	Significa que por cada incremento en la tasa de desempleo de hombres incrementa la tasa de criminalidad por 0.86.
W = valor medio de bienes y valores transferibles, o ingreso familiar en decenas de dólares.	N/A	0	La removimos del modelo.
X = número de familias por cada mil con salarios iguales a la mediana del ingreso.	0.583293		Significa que por cada incremento en numero de familias por cada mil salarios la tasa de criminalidad incrementa por 0.58.
<b>Escriba el resultado del siguiente estadístico</b>	<b>Interprete el significado del R<sup>2</sup> ajustado</b>		
<b>R<sup>2</sup> ajustado</b>			<b>0.720806</b>
			El modelo explica 72% de los datos correctamente lo cual lo hace significativo para explicar los datos.

PASOS:

gretl: model 2

File Edit Tests Save Graphs Analysis LaTeX

Model 2: OLS, using observations 1-47  
Dependent variable: R

	coefficient	std. error	t-ratio	p-value	
const	-589.606	111.392	-5.293	7.15e-06	***
Edad	1.23860	0.378836	3.269	0.0025	***
EP	1.55223	0.610774	2.541	0.0158	**
GX0	1.80570	0.985830	1.832	0.0758	*
GX1	-0.810129	1.07172	-0.7559	0.4549	
FT	0.0546557	0.107164	0.5100	0.6133	
H	-0.692519	0.311790	-2.221	0.0331	**
N	-0.0599350	0.104996	-0.5708	0.5719	
NB	-0.0386349	0.0523758	-0.7376	0.4658	
U1	-0.220731	0.324534	-0.6801	0.5010	
U2	1.27133	0.784336	1.621	0.1143	
W	0.139491	0.0963537	1.448	0.1569	
X	0.761286	0.195731	3.889	0.0004	***

  

Mean dependent var	90.50851	S.D. dependent var	38.67627
Sum squared resid	14194.40	S.E. of regression	20.43238
R-squared	0.793714	Adjusted R-squared	0.720907
F(12, 34)	10.90164	P-value(F)	1.98e-08
Log-likelihood	-200.8858	Akaike criterion	427.7716
Schwarz criterion	451.8235	Hannan-Quinn	436.8225

Excluding the constant, p-value was highest for variable 7 (FT)

Retiro las variables con valor-p más grande. En este caso FT

gretl: model 3

File Edit Tests Save Graphs Analysis LaTeX

Model 3: OLS, using observations 1-47  
Dependent variable: R

	coefficient	std. error	t-ratio	p-value	
const	-576.001	107.002	-5.383	5.03e-06	***
Edad	1.22836	0.374283	3.282	0.0023	***
EP	1.69206	0.539983	3.134	0.0035	***
GX0	1.95605	0.930728	2.102	0.0429	**
GX1	-0.983814	1.00537	-0.9786	0.3345	
H	-0.687587	0.308328	-2.230	0.0323	**
N	-0.0625546	0.103756	-0.6029	0.5505	
NB	-0.0377215	0.0517890	-0.7284	0.4712	
U1	-0.230663	0.320507	-0.7197	0.4765	
U2	1.20659	0.765772	1.576	0.1241	
W	0.146503	0.0943545	1.553	0.1295	
X	0.783219	0.188920	4.146	0.0002	***

  

Mean dependent var	90.50851	S.D. dependent var	38.67627
Sum squared resid	14302.99	S.E. of regression	20.21526
R-squared	0.792136	Adjusted R-squared	0.726807
F(11, 35)	12.12537	P-value(F)	6.17e-09
Log-likelihood	-201.0649	Akaike criterion	426.1298
Schwarz criterion	448.3316	Hannan-Quinn	434.4845

Excluding the constant, p-value was highest for variable 9 (N)

$R^2$ ajustado no cayó mucho, por lo tanto procedo a retirar más variables no significativas.

gretl: model 4

File Edit Tests Save Graphs Analysis LaTeX

Model 4: OLS, using observations 1-47  
Dependent variable: R

	coefficient	std. error	t-ratio	p-value	
const	-579.701	105.877	-5.475	3.50e-06	***
Edad	1.27429	0.363194	3.509	0.0012	***
EP	1.73656	0.530165	3.276	0.0023	***
GX0	1.90668	0.918887	2.075	0.0452	**
GX1	-0.975310	0.996348	-0.9789	0.3342	
H	-0.702791	0.304567	-2.308	0.0269	**
NB	-0.0401220	0.0511772	-0.7840	0.4382	
U1	-0.207470	0.315365	-0.6579	0.5148	
U2	1.15766	0.754698	1.534	0.1338	
W	0.141207	0.0931105	1.517	0.1381	
X	0.765626	0.184995	4.139	0.0002	***

  

Mean dependent var	90.50851	S.D. dependent var	38.67627
Sum squared resid	14451.53	S.E. of regression	20.03576
R-squared	0.789977	Adjusted R-squared	0.731637
F(10, 36)	13.54098	P-value(F)	1.91e-09
Log-likelihood	-201.3077	Akaike criterion	424.6154
Schwarz criterion	444.9670	Hannan-Quinn	432.2739

Excluding the constant, p-value was highest for variable 11 (U1)

Remuevo N, el R cuadrado ajustado mejora. A 0.73. Sigo removiendo variables no significativas. U1.

gretl: model 5

File Edit Tests Save Graphs Analysis LaTeX

Model 5: OLS, using observations 1-47  
Dependent variable: R

	coefficient	std. error	t-ratio	p-value	
const	-580.345	105.058	-5.524	2.77e-06	***
Edad	1.26524	0.360141	3.513	0.0012	***
EP	1.59663	0.481894	3.313	0.0021	***
GX0	1.92665	0.911320	2.114	0.0413	**
GX1	-0.960087	0.988415	-0.9713	0.3377	
H	-0.739983	0.296971	-2.492	0.0173	**
NB	-0.0369907	0.0505633	-0.7316	0.4690	
U2	0.742464	0.410652	1.808	0.0787	*
W	0.152130	0.0909130	1.673	0.1027	
X	0.779197	0.182427	4.271	0.0001	***

  

Mean dependent var	90.50851	S.D. dependent var	38.67627
Sum squared resid	14625.27	S.E. of regression	19.88159
R-squared	0.787452	Adjusted R-squared	0.735751
F(9, 37)	15.23093	P-value(F)	5.73e-10
Log-likelihood	-201.5885	Akaike criterion	423.1771
Schwarz criterion	441.6786	Hannan-Quinn	430.1393

Excluding the constant, p-value was highest for variable 10 (NB)

Mejora el R cuadrado ajustado, sigo removiendo variables no significativas: NB.

gretl: model 6

File Edit Tests Save Graphs Analysis LaTeX

Model 6: OLS, using observations 1-47  
Dependent variable: R

	coefficient	std. error	t-ratio	p-value	
const	-575.635	104.217	-5.523	2.56e-06	***
Edad	1.16990	0.333672	3.506	0.0012	***
EP	1.69785	0.458774	3.701	0.0007	***
GX0	1.98738	0.901963	2.203	0.0337	**
GX1	-1.09183	0.965909	-1.130	0.2654	
H	-0.695538	0.288907	-2.407	0.0210	**
U2	0.731200	0.407846	1.793	0.0810	*
W	0.162594	0.0892300	1.822	0.0763	*
X	0.746286	0.175709	4.247	0.0001	***

  

Mean dependent var	90.50851	S.D. dependent var	38.67627
Sum squared resid	14836.83	S.E. of regression	19.75962
R-squared	0.784378	Adjusted R-squared	0.738983
F(8, 38)	17.27925	P-value(F)	1.70e-10
Log-likelihood	-201.9260	Akaike criterion	421.8521
Schwarz criterion	438.5034	Hannan-Quinn	428.1181

Excluding the constant, p-value was highest for variable 6 (GX1)

Mejora el R cuadrado ajustado. Sigo removiendo variables no significativas: GX1.

gretl: model 7

File Edit Tests Save Graphs Analysis LaTeX

Model 7: OLS, using observations 1-47  
Dependent variable: R

	coefficient	std. error	t-ratio	p-value	
const	-589.611	103.849	-5.678	1.45e-06	***
Edad	1.17599	0.334814	3.512	0.0011	***
EP	1.69641	0.460403	3.685	0.0007	***
GX0	0.985816	0.169275	5.824	9.10e-07	***
H	-0.649953	0.287095	-2.264	0.0292	**
U2	0.772059	0.407685	1.894	0.0657	*
W	0.166965	0.0894630	1.866	0.0695	*
X	0.779083	0.173912	4.480	6.36e-05	***

  

Mean dependent var	90.50851	S.D. dependent var	38.67627
Sum squared resid	15335.71	S.E. of regression	19.82986
R-squared	0.777127	Adjusted R-squared	0.737125
F(7, 39)	19.42683	P-value(F)	6.68e-11
Log-likelihood	-202.7032	Akaike criterion	421.4064
Schwarz criterion	436.2076	Hannan-Quinn	426.9762

R cuadrado ajustado disminuye diminutamente, sigo removiendo variables no significativas: W.

gretl: model 8

File Edit Tests Save Graphs Analysis LaTeX

Model 8: OLS, using observations 1-47  
Dependent variable: R

	coefficient	std. error	t-ratio	p-value	
const	-492.119	92.4976	-5.320	4.25e-06	***
Edad	1.06436	0.339499	3.135	0.0032	***
EP	1.92255	0.457750	4.200	0.0001	***
GX0	1.17841	0.138283	8.522	1.57e-010	***
H	-0.630377	0.295674	-2.132	0.0392	**
U2	0.863030	0.417134	2.069	0.0451	**
X	0.583293	0.142947	4.080	0.0002	***

  

Mean dependent var	90.50851	S.D. dependent var	38.67627
Sum squared resid	16705.33	S.E. of regression	20.43608
R-squared	0.757223	Adjusted R-squared	0.720806
F(6, 40)	20.79335	P-value(F)	6.91e-11
Log-likelihood	-204.7135	Akaike criterion	423.4270
Schwarz criterion	436.3780	Hannan-Quinn	428.3005

Todas son significativas al 0.05%.

**Serie II (42 puntos):** Responda de forma clara y concisa las preguntas que se le presentan a continuación.

**1. ¿A qué conclusión se llega cuando no se rechaza la hipótesis nula sobre un coeficiente de un modelo de regresión?**

La conclusión a la que se llega es que la variable no es significativa para el modelo y se debe de considerar removerla del mismo, al mismo tiempo hay que tener precaución al remover dicha variable porque a pesar de no ser significativa al removerse se puede bajar mucho el  $R^2$  ajustado, por lo que siempre es asunto de ver en qué proporción baja el  $R^2$  si se remueve la variable no significativa del modelo.

**2. ¿Por qué es necesario elaborar un marco teórico antes de correr un modelo de regresión?**

Nos sirve para poner en contexto el estudio estadístico o científico que se hará, pone en contexto cómo se obtuvieron los resultados y que se intenta probar o qué se descubrió además de establecer posibles correlaciones altas como causalidad (con argumentos).

**3. ¿Qué diferencia existe entre el término de error en un modelo de regresión y el residuo?**

Ninguna, el error de un modelo se calcula como  $Y - Y$  estimado, por lo que el residuo de esa resta y el error son sinónimos, es qué tanto le pegó al dato real el modelo y el promedio de los residuos siempre será igual a 0.

**4. ¿Es posible lograr un modelo de regresión con una variación no explicada igual a cero?, justifique claramente su respuesta.**

No, los modelos siempre contendrán errores por lo que siempre habrá un porcentaje que no se podrá explicar.

**5. Explique con sus propias palabras en qué consiste el método de mínimos cuadrados ordinarios.**

El método de mínimos cuadrados ordinarios consiste en un método que permite predecir parámetros poblacionales, minimizando los errores que se encuentran entre lo predicho y los datos de tal manera que se minimizan los errores entre el modelo y los datos reales haciéndolo responder a predicciones de una manera precisa con mínimos errores.

**6. Cuando una variable no es estadísticamente significativa ¿se debe dejar fuera del modelo de regresión?, justifique claramente su respuesta.**

La mayoría del tiempo se remueve del modelo, sin embargo, hay veces en las que remover una variable no significativa del modelo hace que las demás pierdan lo significativo y en esos casos es mejor quitarla, por lo tanto no se puede generalizar que se deben de quitar siempre sino hay que probar quitarla y observar en qué proporción baja en  $R^2$  ajustado, si baja demasiado se debe considerar dejarla pero si baja diminutamente es mejor quitar la variable.