

Universidad Panamericana
Maestría en Ciencia de Datos
Econometría

Actividad RLM

Enrique Ulises Báez Gómez Tagle

7 de septiembre de 2025

Índice

1	Introducción	2
2	Pregunta 1	2
3	Pregunta 2	3
4	Pregunta 3	4
5	Pregunta 4	4
6	Pregunta 5	5
7	Link al repositorio con código fuente	5

1. Introducción

Se presentan y analizan el siguiente juego de datos cuyas variables son:

Variable	Descripción
1 Educación	Gasto per cápita en educación pública (en dólares)
2 Ingreso	Ingreso per cápita anual (en dólares)
3 Menores	Porcentaje de menores de 18 años de edad (por cada mil)
4 Urbano	Proporción de la población que reside en áreas urbanas

Cuadro 1: Variables del conjunto de datos

A continuación se presentan las primeras filas del conjunto de datos:

Cuadro 2: Primeras filas del conjunto de datos utilizado en el análisis

Estado	educacion	Ingreso	Menores	Urbano
ME	189	2824	350.7000	508
NH	169	3259	345.9000	564
VT	230	3072	348.5000	322
MA	168	3835	335.3000	846
RI	180	3549	327.1000	871

2. Pregunta 1

Utilizando los datos, considere el modelo de regresión lineal múltiple

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

donde y representa la respuesta educación, x_1 el ingreso per cápita, x_2 el porcentaje de menores de 18 años y x_3 la proporción de habitantes que reside en áreas urbanas. Realice el ajuste del modelo (1).

Ajuste por Mínimos Cuadrados Ordinarios (MCO): El estimador viene dado por:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \quad \hat{\mathbf{y}} = \mathbf{X}\hat{\beta}.$$

Usando los datos, el modelo estimado es:

$$\hat{y} = -286.8388 + 0.080653 x_1 + 0.817338 x_2 - 0.105806 x_3,$$

donde x_1 es ingreso, x_2 menores y x_3 urbano.

Coefficientes y errores estándar:

Cuadro 3: Coeficientes del modelo OLS: Educación vs Ingreso, Menores y Urbano

Parámetro	Coef.	Err. Std.	t	p	IC 2.5 %	IC 97.5 %
Intercepto	-286.8388	64.9199	-4.4183	0.0001	-417.4408	-156.2367
ingreso	0.0807	0.0093	8.6738	0.0000	0.0619	0.0994
menores	0.8173	0.1598	5.1151	0.0000	0.4959	1.1388
urbano	-0.1058	0.0343	-3.0863	0.0034	-0.1748	-0.0368

Métricas del ajuste: $R^2 = 0.690$, $R^2_{aj} = 0.670$, $F(3, 47) = 34.81$, $p\text{-valor} < 10^{-10}$.

Cuadro 4: Métricas globales del modelo de regresión lineal múltiple

Métrica	Valor
R^2	0.6896
R^2 ajustado	0.6698
Estadístico F	34.8105
p-valor (F)	0.0000
AIC	483.5767
BIC	491.3040
Observaciones	51.0000

3. Pregunta 2

Encuentre una estimación de la varianza de los errores $S^2 = e'e/n$, la matriz de covarianzas del vector de parámetros y los errores estándar de los coeficientes individuales.

Estimación de la varianza del error: Sea $\hat{e} = \mathbf{y} - \mathbf{X}\hat{\beta}$. En este reporte usaremos la estimación insesgada de la varianza del error:

$$\hat{\sigma}^2 = \frac{\hat{e}'\hat{e}}{n-p}$$

Con los datos del ejercicio se obtiene:

Cuadro 5: Estimación de la varianza del error S^2 del modelo OLS.

Métrica	Valor
S^2 (Varianza del error)	712.5394

Matriz de covarianzas del vector de parámetros $\hat{\beta}$. Por teoría MCO,

$$\widehat{\text{Var}}(\hat{\beta}) = S^2(\mathbf{X}'\mathbf{X})^{-1}.$$

Su estimación numérica es:

Cuadro 6: Matriz de covarianzas de los estimadores del modelo OLS.

Parámetro	Intercepto	ingreso	menores	urbano
Intercepto	4214.5975	-0.1850	-9.7494	-0.1583
ingreso	-0.1850	0.0001	0.0001	-0.0002
menores	-9.7494	0.0001	0.0255	0.0002
urbano	-0.1583	-0.0002	0.0002	0.0012

Errores estándar de los coeficientes individuales. Los errores estándar son las raíces cuadradas de la diagonal de $\widehat{\text{Var}}(\hat{\beta})$. A continuación se reportan junto con los coeficientes:

Cuadro 7: Estimaciones, errores estándar y pruebas t de los coeficientes del modelo

Coefficiente	Error Estándar	t	p
-286.8388	64.9199	-4.4183	0.0001
0.0807	0.0093	8.6738	0.0000
0.8173	0.1598	5.1151	0.0000
-0.1058	0.0343	-3.0863	0.0034

4. Pregunta 3

Construya un intervalo del 90 % de confianza para el coeficiente β_2 .

IC del 90 % para β_2 (Menores): Sea $\hat{\beta}_2$ el estimador de β_2 y $EE(\hat{\beta}_2)$ su error estándar. El intervalo de confianza bilateral al 90 % está dado por

$$\hat{\beta}_2 \pm t_{1-\alpha/2, n-p} EE(\hat{\beta}_2), \quad \alpha = 0.10, \quad n = 51, \quad p = 4 \Rightarrow gl = n - p = 47.$$

A partir del ajuste MCO del inciso anterior (ver tabla de coeficientes) tenemos $\hat{\beta}_2 = 0.817338$ y $EE(\hat{\beta}_2) = 0.159790$. Usando $t_{0.95, 47} \approx 1.677$ se obtiene el intervalo numérico siguiente:

Cuadro 8: Intervalo de confianza del 90 % para el coeficiente β_2 (Menores).

Parámetro	Límite inferior	Límite superior
menores	0.5492	1.0855

Para referencia, volvemos a listar los coeficientes estimados y sus errores estándar:

Cuadro 9: Coeficientes del modelo OLS: Educación vs Ingreso, Menores y Urbano

Parámetro	Coef.	Err. Std.	t	p	IC 2.5 %	IC 97.5 %
Intercepto	-286.8388	64.9199	-4.4183	0.0001	-417.4408	-156.2367
ingreso	0.0807	0.0093	8.6738	0.0000	0.0619	0.0994
menores	0.8173	0.1598	5.1151	0.0000	0.4959	1.1388
urbano	-0.1058	0.0343	-3.0863	0.0034	-0.1748	-0.0368

5. Pregunta 4

Calcule el gasto en educación pública que se esperaría a un nivel “promedio” de los regresores, esto es $(1, \bar{x})$.

Predicción en el punto promedio $(1, \bar{x})$: Dado que la predicción en un punto específico $\mathbf{x}_0 = (1, \bar{x}_1, \bar{x}_2, \bar{x}_3)'$ viene dada por:

$$\hat{y}(\mathbf{x}_0) = \mathbf{x}_0' \hat{\boldsymbol{\beta}} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \hat{\beta}_2 \bar{x}_2 + \hat{\beta}_3 \bar{x}_3.$$

Los promedios muestrales de los regresores son:

Cuadro 10: Valores promedio de los regresores y gasto en educación esperado en el punto promedio

Variable	Valor promedio
Ingreso	3225.2900
Menores	358.8900
Urbano	664.5100
Educación esperada	196.3100

Sustituyendo estos valores en la ecuación estimada se obtiene el gasto en educación pública esperado en el nivel promedio de los regresores: $\hat{y}(\bar{x}) = 196.31$ dólares per cápita.

6. Pregunta 5

Realice la prueba de significancia del modelo de regresión (1), indicando claramente la hipótesis, estadístico de prueba, región de rechazo y conclusión.

Prueba de significancia global del modelo (Prueba F). Hipótesis:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0 \quad \text{vs.} \quad H_1 : \text{al menos uno } \beta_j \neq 0.$$

Estadístico de prueba: bajo H_0 , el estadístico

$$F_{obs} = \frac{SSR/k}{SSE/(n-p)} = \frac{(R^2/k)}{((1-R^2)/(n-p))} \sim F_{k, n-p},$$

donde $k = 3$ es el número de regresores, $p = k + 1 = 4$ el número de parámetros (incluye intercepto) y $n = 51$ el tamaño de muestra. Con los datos del ajuste se obtuvo:

$$F_{obs} = 34.8105, \quad (gl_1, gl_2) = (3, 47).$$

Región de rechazo (nivel $\alpha = 0.05$): Rechazar H_0 si

$$F_{obs} > F_{0.95; 3, 47} = 2.8024.$$

Decisión por valor crítico: como $34.8105 > 2.8024$, **se rechaza H_0 .**

Decisión por p -valor: $p\text{-valor} = 5.34 \times 10^{-12} < 0.05 \Rightarrow$ **se rechaza H_0 .**

Conclusión: el modelo de regresión es *globalmente significativo*; al menos una de las variables explicativas (Ingreso, Menores, Urbano) contribuye de forma estadísticamente significativa a explicar el gasto en educación.

Resumen numérico:

Cuadro 11: Prueba de significancia global del modelo de regresión (estadístico F).

Estadístico	Valor
F observado	34.8110
F crítico ($\alpha = 0.05$)	2.8020
p-valor	5.34e-12
gl1	3
gl2	47

Cuadro 12: Hipótesis y conclusión de la prueba F global del modelo de regresión

Elemento	Descripción
Hipótesis nula (H_0)	$\beta_1 = \beta_2 = \beta_3 = 0$ (el modelo no es significativo)
Hipótesis alternativa (H_1)	Al menos un $\beta_j \neq 0$
Conclusión	Se RECHAZA H_0 : el modelo es globalmente significativo.

7. Link al repositorio con código fuente

<https://github.com/enriquemomeztagle/MCD-Econometria/tree/main/HWs/MLR-practice>