

## Taller #4

**Respuestas Sugeridas**  
**Regresión Múltiple**  
**Econometría 06216**

**Profesor: Julio César Alonso**

**Monitores: Paúl Semaán**

**Francisco Quevedo**

**Notas:**

- o Recuerde que sólo tres preguntas, seleccionadas al azar, serán calificadas.
- o Este taller es para ser entregado en los primeros 10 minutos de la clase del próximo **20 de febrero**.

**INSTRUCCIONES:**

- Este taller puede ser escrito a mano, pero con letra legible.
- Cuando sea posible, debe mostrar el procedimiento efectuado para llegar a sus resultados.

FENALCO ha decidido analizar el efecto que tienen sobre el volumen de ventas, el tamaño de la empresa y el volumen de recursos propios disponibles. Para ello se dispone de información de la Encuesta Anual Manufacturera de 20 empresas con el mayor número de empleados en el año 2005 (los datos están disponibles en el archivo T4-01-06.xls). Donde el volumen de ventas ( $Vtas_i$ ) y los recursos propios ( $rec_i$ ) están medidos en millones de pesos y el tamaño de la empresa ( $tam_i$ ), por medio del número de empleados, un investigador estima el siguiente modelo.

$$\ln(Vtas_i) = \beta_0 + \beta_1 \ln(rec_i) + \beta_2 \ln(tam_i) + \varepsilon_i$$

1. Estime los parámetros  $\beta^T = (\beta_0 \quad \beta_1 \quad \beta_2)$  del modelo por medio del método de MCO (muestre todo el procedimiento y resultados parciales) (no emplee EasyReg, haga los cálculos matriciales en Excel y muestre claramente la matriz  $X^T X$  y  $X^T y$ )
2. De acuerdo a sus resultados en el numeral anterior:
  - a. Interprete el significado de los coeficientes estimados.
  - b. Estime  $\sigma^2$  y la matriz de varianzas y covarianzas. (no emplee EasyReg, haga los cálculos matriciales en Excel y muestre claramente los resultados parciales)

3. Continuando con la pregunta anterior,
  - a. Construya la tabla ANOVA del nuevo modelo. Encuentre el  $R^2$  e interprételo y calcule el  $R^2$  ajustado. (no emplee EasyReg, haga los cálculos matriciales en Excel y muestre claramente los resultados parciales)

El gobierno de los Estados Unidos desea determinar la relación existente entre la calidad de educación y el gasto en educación ejecutado local y estatalmente. Para ello, ha encargado al departamento de investigaciones del Ministerio de Educación determinar las variables que se encuentran involucradas. En el archivo T4-01-06.xls se encuentran los datos de 684 distritos escolares que recoge la información sobre:

- La calidad de la educación del distrito  $i$  (medido como el porcentaje de alumnos que aprueban los exámenes estandarizados de matemáticas e inglés ( $MAT_i$  y  $ENG_i$  respectivamente),
- GLOC y GSTA es el nivel de gasto por estudiante realizado por el gobierno local y estatal respectivamente. Ambos niveles medidos en dólares.

4. Determine los modelos a emplear de acuerdo a la información disponible, estímelos y reporte sus resultados en una tabla. Tenga en cuenta que la calidad de la educación en los Estados Unidos se mide de acuerdo a los resultados de las pruebas de inglés y de matemáticas de manera separada y no parece existir un consenso de cuál de las dos variables es mejor para medir calidad. (Escriba claramente el modelo a estimar y emplee EasyReg para efectuar los cálculos)
5. Interprete los coeficientes estimados y comente sobre su significancia individual y los signos de estos.
6. Con base en la información anterior
  - a. ¿Existe alguna diferencia o no en que los insumos educativos sean adquiridos con recursos estatales o locales?
  - b. Determine cuál de los dos modelos es el mejor para explicar la calidad de la educación.

Taller #4

Respuestas Sugeridas  
Regresión Múltiple  
Econometría 06216

Profesor: Julio César Alonso

Monitores: Paúl Semaán

Francisco Quevedo

Notas:

- o Recuerde que sólo tres preguntas, seleccionadas al azar, serán calificadas.
- o Este taller es para ser entregado en los primeros 10 minutos de la clase del próximo 20 de febrero.

INSTRUCCIONES:

- Este taller puede ser escrito a mano, pero con letra legible.
- Cuando sea posible, debe mostrar el procedimiento efectuado para llegar a sus resultados.

FENALCO ha decidido analizar el efecto que tienen sobre el volumen de ventas, el tamaño de la empresa y el volumen de recursos propios disponibles. Para ello se dispone de información de la Encuesta Anual Manufacturera de 20 empresas con el mayor número de empleados en el año 2005 (los datos están disponibles en el archivo T4-01-06.xls). Donde el volumen de ventas ( $Vtas_i$ ) y los recursos propios ( $rec_i$ ) están medidos en millones de pesos y el tamaño de la empresa ( $tam_i$ ), por medio del número de empleados, un investigador estima el siguiente modelo.

$$\ln(Vtas_i) = \beta_0 + \beta_1 \ln(rec_i) + \beta_2 \ln(tam_i) + \varepsilon_i$$

1. Estime los parámetros  $\beta^T = (\beta_0 \ \beta_1 \ \beta_2)$  del modelo por medio del método de MCO (muestre todo el procedimiento y resultados parciales) (no emplee EasyReg, haga los cálculos matriciales en Excel y muestre claramente la matriz  $X^T X$  y  $X^T y$ )

Inicialmente es necesario construir la nueva matriz  $X^T X$ .

$$X^T X = \begin{bmatrix} n & \sum(\ln rec_i) & \sum(\ln tam_i) \\ - & \sum(\ln rec_i)^2 & \sum(\ln rec_i) \cdot (\ln tam_i) \\ - & - & \sum(\ln tam_i)^2 \end{bmatrix}$$

Todas las entradas de la matriz pueden ser calculadas de la siguiente forma:

$$\sum(\ln rec_i) = 218.77$$

$$\sum(\ln tam_i) = 187.62$$

$$\sum(\ln rec_i)^2 = 2422.76$$

$$\sum(\ln rec_i) \cdot (\ln tam_i) = 2054.55$$

$$\sum(\ln tam_i)^2 = 1764.59$$

Entonces se tiene que  $X^T X = \begin{bmatrix} 20 & 218.77 & 187.62 \\ - & 2422.76 & 2054.55 \\ - & - & 1764.59 \end{bmatrix}$ .

Ahora es necesario construir la matriz  $X^T y = \begin{bmatrix} \sum(\ln Vtas_i) \\ \sum(\ln Vtas_i) \cdot (\ln rec_i) \\ \sum(\ln Vtas_i) \cdot (\ln tam_i) \end{bmatrix}$ , cada una de las

entradas viene dada por las siguientes sumatorias:

$$\sum(\ln Vtas_i) = 218.82$$

$$\sum(\ln Vtas_i) \cdot (\ln rec_i) = 2734.38$$

$$\sum(\ln Vtas_i) \cdot (\ln tam_i) = 2337.16$$

Por lo tanto  $X^T y = \begin{bmatrix} 248.82 \\ 2734.38 \\ 2337.16 \end{bmatrix}$ .

Ahora ya se puede calcular los coeficientes utilizando la siguiente fórmula:

$$\hat{\beta} = (X^T X)^{-1} X^T y, \text{ donde}$$

$$(X^T X)^{-1} = \begin{bmatrix} 20.46 & -0.22 & -1.92 \\ - & 0.04 & -0.02 \\ - & - & 0.23 \end{bmatrix}$$

Reemplazando en la fórmula se tiene que:

$$\hat{\beta} = (X^T X)^{-1} X^T y = \begin{bmatrix} 20.46 & -0.22 & -1.92 \\ - & 0.04 & -0.02 \\ - & - & 0.23 \end{bmatrix} \cdot \begin{bmatrix} 248.82 \\ 2734.38 \\ 2337.16 \end{bmatrix} = \begin{bmatrix} 3.91 \\ 0.39 \\ 0.46 \end{bmatrix}$$

2. De acuerdo a sus resultados en el numeral anterior:
  - a. Interprete el significado de los coeficientes estimados.

$\hat{\beta}_0 = 3.91$ , carece de significado económico

$\hat{\beta}_1 = 0.39$ , el efecto de un aumento del 1% de los recursos de la compañía implicara un aumento del 0.39% en el volumen de ventas. Es la elasticidad del volumen de ventas con respecto a los activos de la compañía.

$\hat{\beta}_2 = 0.46$ , el efecto de un aumento del 1% en el numero de trabajadores implicara un aumento del 0.46% en el volumen de ventas de la compañía. Es la elasticidad del volumen de ventas con respecto al tamaño de la compañía.

- b. Estime  $\sigma^2$  y la matriz de varianzas y covarianzas. (no emplee EasyReg, haga los cálculos matriciales en Excel y muestre claramente los resultados parciales)

Para estimar la varianza es necesario emplear la siguiente fórmula  $s^2 = \frac{y^T y - \hat{\beta}^T X^T y}{n - k}$ .

$$y^T y = \sum (\ln Ventas_t)^2 = 3104.14$$

Entonces

$$s^2 = \frac{3104.14 - [3.9128 \quad 0.3889 \quad 0.4556] \cdot \begin{bmatrix} 248.82 \\ 2734.38 \\ 2337.16 \end{bmatrix}}{20 - 3} = 0.130075$$

Y la matriz de varianzas y covarianzas será:

$$s^2 (X^T X)^{-1} = 0.130075 \begin{bmatrix} 20.46 & -0.22 & -1.92 \\ - & 0.04 & -0.02 \\ - & - & 0.23 \end{bmatrix}$$

- 3. Continuando con la pregunta anterior,
  - a. Construya la tabla ANOVA del nuevo modelo. Encuentre el  $R^2$  e interprételo y calcule el  $R^2$  ajustado. (no emplee EasyReg, haga los cálculos matriciales en Excel y muestre claramente los resultados parciales)

La tabla ANOVA está formada por las siguientes ecuaciones:

Fuente de la Variación	SS	Grados de libertad	MS
Regresión	$SSR = \hat{\beta}^T X^T y - n\bar{Y}^2$	$k - 1$	$MSR = \frac{SSR}{k - 1}$
Error	$SSE = y^T y - \hat{\beta}^T X^T y$	$n - k$	$MSE = s^2 = \frac{SSE}{n - k}$
Total	$SST = y^T y - n\bar{Y}^2$	$n - 1$	

Fuente de la Variación	SS	Grados de libertad	MS
Regresión	$SSR = 3101.93 - 3095.66 = 6.28$	$3 - 1 = 2$	$MSR = 3.138$
Error	$SSE = 3104.14 - 3101.93 = 2.21$	$20 - 3 = 17$	$MSE = 0.130075$
Total	$SST = 3104.14 - 3095.66 = 8.49$	$20 - 1 = 19$	

Dado que  $R^2 = \frac{SSR}{SST}$ , para este modelo sería igual a  $R^2 = \frac{6.28}{8.49} = 0.7395$ . Por lo tanto el modelo empleado explica el 73.95% de la variabilidad de la variable del volumen de ventas.

Por otro lado el  $\bar{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-k}$ , entonces sería igual a 0.7088.

El gobierno de los Estados Unidos desea determinar la relación existente entre la calidad de educación y el gasto en educación ejecutado local y estatalmente. Para ello, ha encargado al departamento de investigaciones del Ministerio de Educación determinar las variables que se encuentran involucradas. En el archivo T4-01-06.xls se encuentran los datos de 684 distritos escolares que recoge la información sobre:

- La calidad de la educación del distrito i (medido como el porcentaje de alumnos que aprueban los exámenes estandarizados de matemáticas e inglés (MAT<sub>i</sub> y ENG<sub>i</sub> respectivamente),
- GLOC y GSTA es el nivel de gasto por estudiante realizado por el gobierno local y estatal respectivamente. Ambos niveles medidos en dólares.

- Determine los modelos a emplear de acuerdo a la información disponible, estímelos y reporte sus resultados en una tabla. Tenga en cuenta que la calidad de la educación en los Estados Unidos se mide de acuerdo a los resultados de las pruebas de inglés y de matemáticas de manera separada y no parece existir un consenso de cuál de las dos variables es mejor para medir calidad. (Escriba claramente el modelo a estimar y emplee EasyReg para efectuar los cálculos)

Los modelos a estimar son:

$$MAT_i = \beta_0 + \beta_1 GLOC_i + \beta_2 GSTA_i \quad (1)$$

$$ENG_i = \gamma_0 + \gamma_1 GLOC_i + \gamma_2 GSTA_i \quad (2)$$

Tabla 1. Estimación de los modelos (1) y (2).

	Variable Dependiente:	
	MAT <sub>i</sub>	ENG <sub>i</sub>
	Estadísticos t entre paréntesis	
	Ecuación (1)	Ecuación (2)
	MCO	MCO
Constante	73.0035 (20.888) ***	82.5145 (28.233) ***
GLOC <sub>i</sub>	0.00002 (0.093)	-0.0003 (-1.782) *
GSTA <sub>i</sub>	-0.0009 (-1.692) *	-0.0006 (-1.357)
R <sup>2</sup>	0.0066	0.0049
R <sup>2</sup> Ajustado	0.0037	0.002
F	2.26	1.67
No. De Obs.	684	684

(\*): Nivel de significancia 10%

(\*\*): Nivel de significancia 5%

(\*\*\*): Nivel de significancia 1%

MCO: Mínimos Cuadrados Ordinarios

- Interprete los coeficientes estimados y comente sobre su significancia individual y los signos de estos.

$\hat{\beta}_0 = 73.0035$  Se refiere al puntaje obtenido en la prueba estandarizada de matemáticas que no depende del gasto gubernamental en educación, es significativo al 1%.

$\hat{\beta}_1 = 0.00002$  Es el aumento en puntos porcentuales de los estudiantes que aprueban el examen estandarizado de matemáticas que resulta de un aumento de un dólar del gasto local en educación por estudiante en el año escolar, no es estadísticamente significativo.

$\hat{\beta}_2 = -0.0009$  Es la disminución en puntos porcentuales de los estudiantes que aprueban el examen estandarizado de matemáticas que resulta de un aumento en un dólar del gasto local en educación por estudiante en el año escolar, es estadísticamente significativo al 10%.

$\hat{\gamma}_0 = 82.5145$  Se refiere al porcentaje de estudiantes que aprueban el examen estandarizado de inglés que no depende del gasto gubernamental en educación, es significativo al 1%.

$\hat{\gamma}_1 = -0.0003$  Es la disminución en puntos porcentuales de los estudiantes que aprueban el examen estandarizado de inglés que resulta de un aumento en un dólar del gasto local en educación por estudiante en el año escolar es estadísticamente significativo al 10%.

$\hat{\gamma}_2 = -0.0006$  Es la disminución en puntos porcentuales de los estudiantes que aprueban el examen estandarizado de inglés que resulta de un aumento en un dólar del gasto local en educación por estudiante en el año escolar no es estadísticamente significativo.

Si el modelo esta correctamente especificado se esperaría que el signo de los coeficientes sea positivo en todos los casos, si se considera al nivel de la educación como un output y al gasto como un input, es decir si se considera a la educación como el producto de un proceso de transformación.

- Con base en la información anterior
  - ¿Existe alguna diferencia o no en que los insumos educativos sean adquiridos con recursos estatales o locales?

Noten que para comprobar si existe o no relevancia en si los insumo usados en la educación de los estudiantes de bachillerato son adquiridos con recursos del estado o locales se debe comprobar las siguientes hipótesis nulas:

$$H_0 : \beta_1 - \beta_2 = 0$$

$$H_0 : \gamma_1 - \gamma_2 = 0$$

Versus las alternas de no Ho. Para el caso en el que se desea comprobar para la prueba estandarizada de la prueba de matemáticas se puede probar por medio de una prueba de Wald. En este caso el estadístico de Wald corresponde a 4.25 el cual con un p-valor de 0.039 permite rechazar la hipótesis nula de que no es relevante el origen de los recursos con un nivel de significancia del 5%. Por el otro lado para la prueba estandarizada de la prueba de inglés el resultado es diferente ya que con un estadístico de Wald igual a 0.59, con un p-valor de 0.4421 no se puede rechazar la hipótesis nula, por lo tanto el origen de los recursos no es relevante en la calidad de la educación si esta es medida por los resultados de las pruebas de inglés.

- Determine cuál de los dos modelos es el mejor para explicar la calidad de la educación.

En este caso no es posible emplear el  $R^2$  para decidir cuál es el mejor modelo, puesto que los dos modelos no poseen la misma variable explicativa. Por lo tanto es necesario analizar la bondad del modelo empleando la significancia individual y global (estadístico F). Noten que en el modelo (1) el coeficiente asociado al gasto local no es significativo, mientras que en el modelo (2) este si es significativo al 10%, lo contrario ocurre para el caso del coeficiente asociado a la variable gasto estatal. Por otro lado, el estadístico F de los dos modelos muestra que los coeficientes no son conjuntamente significativos a un nivel de confianza del 99%.

Por lo tanto no es posible decidir entre el modelo (1) y (2). Con el mismo criterio. Así, no podemos concluir cual es el mejor modelo.