

Taller 2: Regresión lineal simple
Econometría 06216
24-01-2011

Profesor: Carlos Giovanni González Espítia

Monitores: Adriana Caicedo – Jessica Echeverry – Samir Aristizábal

Notas:

- Recuerde que únicamente tres preguntas, seleccionadas al azar, serán calificadas.
- Este taller deberá subirse a la plataforma Moodle hasta las 7:10 del 31 de enero de 2011. **Sólo se calificaran talleres en formato pdf. Cualquier otro formato no será tenido en cuenta.** (no se recibirán talleres después de esa hora y fecha límite).

Instrucciones:

- Este taller debe ser escrito en computador. Cuando sea posible, debe mostrar el procedimiento efectuado para llegar a sus resultados.
- Este taller es un trabajo en pareja. Por tanto el taller debe reflejar **únicamente** el trabajo de la pareja.
- Si bien no es necesario reportar todos los números decimales, sí lo es hacer los cálculos **con todos ellos**.

Usted es un reconocido economista y ha sido contratado por WIKILINKS para analizar el comportamiento del salario de los trabajadores al interior de la compañía. El presidente de la empresa le informa que en la actualidad cuenta con 20 trabajadores y le propone la siguiente expresión para realizar su trabajo:

$$\frac{1}{W} = \varphi \left(\frac{1}{\ln Exp} \right)^{\gamma} \quad (1.1)$$

donde W representa el salario (medido en dólares) y Exp la experiencia laboral (medida en meses). Para ello usted cuenta con los datos consignados en el archivo de Excel adjunto (T2-11-01.xls).

Pregunta 1

- a. De acuerdo al enunciado anterior y a la muestra de datos suministrada, determine el tipo de datos y tipo de estructura de datos que se trabajará (y por tanto el subíndice de las variables del modelo a estimar).
- b. Antes de comenzar con el estudio, el vicepresidente de la compañía afirma que el modelo (1.1) no es lineal en sus parámetros y que por lo tanto el estudio contratado por el presidente no tiene sentido. Explique si el vicepresidente tiene o no razón mostrando todo su procedimiento.

- c. Realice un gráfico de dispersión y discuta la relación de la variable dependiente con la variable independiente.

Pregunta 2

- a. Determine el modelo econométrico a estimar. Teniendo en cuenta la forma funcional encontrada en el apartado anterior, los subíndices y el tamaño de la muestra.
- b. Intérprete los coeficientes a priori, teniendo en cuenta sus signos esperados.

Pregunta 3

- a. Estime los coeficientes del modelo (1.1) mostrando todo su procedimiento. Muestre las fórmulas que utilizó y los valores que se reemplazaron.
- b. Teniendo en cuenta los coeficientes estimados, escriba la ecuación estimada.
- c. Reporte sus resultados en una tabla teniendo en cuenta todos los resultados del taller (ver talleres de semestres anteriores para ejemplos de cómo presentar los resultados en una tabla).

Pregunta 4

- a. Interprete los coeficientes estimados y diga si los signos fueron los esperados. Discuta si los coeficientes son significativos. Para esto, realice todo el procedimiento que sea necesario y muestre las formulas que ha utilizado.
- b. Construya la tabla ANOVA correspondiente mostrando los valores reemplazados en cada casilla.
- c. Discuta qué tan bueno es el modelo explicando el por qué de su respuesta. Recuerde mostrar todos sus cálculos para obtener cualquier resultado.

Pregunta 5

- a. No contento con su arduo trabajo, el Presidente le pide que realice una proyección y un intervalo de confianza al 95% para el nivel de salario esperado del año 2011, si el nivel de experiencia de los trabajadores fuera de 82 meses. Muestre todos sus cálculos e interprete el resultado.
- b. Además, el Vicepresidente le pide que realice otra proyección con un intervalo de confianza al 95% para el nivel de salario del año 2011, si el nivel de experiencia de los trabajadores fuera de 105 meses. Muestre todos sus cálculos e interprete el resultado.

Pregunta 6

El asesor financiero del Presidente de la Compañía, le comenta a usted que el modelo que empleó en el estudio no es el único que sirve para analizar el comportamiento del salario, y por ello le propone evaluar las siguientes expresiones, pero no sabe si todas pueden ser estimadas por MCO. Demuestre en cualquier caso, utilizando transformaciones apropiadas, si dichos modelos no lineales pueden ser linealizados o no para poder ser estimados por este método:

a. $W_i = AExp_{1i}^{\alpha_1} e^{\varepsilon_i}$

b. $W_i = (\alpha_0 + \alpha_1 Exp_i + \varepsilon_i)^{-1}$

c. $W_i = \alpha_0 + \alpha_1^2 Exp_i + \varepsilon_i$

d. $\frac{1}{W_i} = \theta_1 \left(\frac{1}{\sin(\theta_3) + \ln(Exp_{1i})} \right)^{\theta_2} \mu_i$

Taller 2: Regresión lineal simple
Econometría 06216
24-01-2011

Profesor: Carlos Giovanni González Espítia

Monitores: Adriana Caicedo – Jessica Echeverry – Samir Aristizábal

Notas:

- Recuerde que únicamente tres preguntas, seleccionadas al azar, serán calificadas.
- Este taller deberá subirse a la plataforma Moodle hasta las 7:10 del 31 de enero de 2011. **Sólo se calificaran talleres en formato pdf. Cualquier otro formato no será tenido en cuenta.** (no se recibirán talleres después de esa hora y fecha límite).

Instrucciones:

- Este taller debe ser escrito en computador. Cuando sea posible, debe mostrar el procedimiento efectuado para llegar a sus resultados.
- Este taller es un trabajo en pareja. Por tanto el taller debe reflejar **únicamente** el trabajo de la pareja.
- Si bien no es necesario reportar todos los números decimales, sí lo es hacer los cálculos **con todos ellos.**

Usted es un reconocido economista y ha sido contratado por WIKILINKS para analizar el comportamiento del salario de los trabajadores al interior de la compañía. El presidente de la empresa le informa que en el año 2011 cuenta con 20 trabajadores y le propone la siguiente expresión para realizar su trabajo:

$$\frac{1}{W} = \varphi \left(\frac{1}{\ln Exp} \right)^{\gamma} \quad (1.1)$$

donde W representa el salario (medido en dólares) y Exp la experiencia laboral (medida en meses). Para ello usted cuenta con los datos consignados en el archivo de Excel adjunto (T2-11-01.xls).

Pregunta 1

- a. De acuerdo al enunciado anterior y a la muestra de datos suministrada, determine el tipo de datos y tipo de estructura de datos que se trabajará (y por tanto el subíndice de las variables del modelo a estimar).

Por el enunciado y el tipo de datos suministrados podemos afirmar que la muestra corresponde a datos de corte transversal. Ya que es información recogida en un momento puntual del tiempo para varios individuos (trabajadores de WIKILINKS) esta muestra corresponde a una estructura de datos de corte transversal. Y el subíndice del modelo debe ser (i), donde $i = 1, 2, \dots, 20$.

- b.** Antes de comenzar con el estudio, el vicepresidente de la compañía afirma que el modelo (1.1) no es lineal en sus parámetros y que por lo tanto el estudio contratado por el presidente no tiene sentido. Explique si el vicepresidente tiene o no razón mostrando todo su procedimiento.

Realizando las siguientes transformaciones podemos linealizar el modelo

$$\frac{1}{W} = \varphi \left(\frac{1}{\ln(\text{Exp})} \right)^\gamma$$

$$\ln\left(\frac{1}{W}\right) = \ln\left[\varphi \left(\frac{1}{\ln(\text{Exp})}\right)^\gamma\right]$$

$$-\ln(W) = \ln(\varphi) + \ln\left(\left(\frac{1}{\ln(\text{Exp})}\right)^\gamma\right)$$

$$-\ln(W) = \ln(\varphi) - \gamma \ln(\ln(\text{Exp}))$$

$$\ln(W) = -\ln(\varphi) + \gamma \ln(\ln(\text{Exp}))$$

$$\ln(W) = \alpha_0 + \alpha_1 \ln(\text{Exp})$$

Donde: $\alpha_0 = -\ln(\varphi)$; $\alpha_1 = \gamma$

De esta forma, el vicepresidente no tiene razón puesto que el modelo ahora es lineal.

- c.** Realice un gráfico de dispersión y discuta la relación de la variable dependiente con la variable independiente.



De acuerdo al gráfico de dispersión se evidencia una relación directa o positiva entre las variables dependiente e independiente, donde a medida que aumenta los meses de experiencia laboral, aumenta también el salario del trabajador i .

Pregunta 2

- a. Determine el modelo econométrico a estimar. Teniendo en cuenta la forma funcional encontrada en el apartado anterior, los subíndices y el tamaño de la muestra.

El modelo estadístico tiene tanto una parte determinística como una parte aleatoria, la forma funcional encontrada en la pregunta 1a es un modelo matemático, para ser un modelo de regresión lineal se le debe incluir un término error. Es así como nuestro modelo econométrico es:

$$\ln(W_i) = \alpha_0 + \alpha_1 \text{Exp}_i + \varepsilon_i$$

$$i = 1, 2, \dots, 20$$

- b. Intérprete los coeficientes a priori, teniendo en cuenta sus signos esperados.

α_0 = No tiene interpretación económica.

Derivada para interpretar α_1 :

$$\ln(W_i) = \alpha_0 + \alpha_1 \text{Exp}_i + \varepsilon_i$$

$$\frac{d \ln W_i}{d \text{Exp}_i} = \alpha_1$$

$$\frac{1}{W_i} dW_i = \alpha_1$$

$$\frac{\frac{dW_i}{W_i} * 100}{d \text{Exp}_i} = \alpha_1 * 100$$

$$\frac{\Delta \% W_i}{d \text{Exp}_i} = (\alpha_1 * 100) \%$$

α_1 = Una variación de 1 mes en la experiencia laboral del trabajador i genera, en promedio, una variación de $(\alpha_1 * 100)$ % dólares en el salario. Se espera signo positivo.

Pregunta 3

- a. Estime los coeficientes del modelo (1.1) mostrando todo su procedimiento. Muestre las fórmulas que utilizó y los valores que se reemplazaron.

Los coeficientes se estiman de la siguiente manera:

$$\hat{\alpha}_1 = \frac{\sum_{i=1}^{20} (\text{Exp}_i - \bar{\text{Exp}}) (\ln W_i - \overline{\ln W})}{\sum_{i=1}^{20} (\text{Exp}_i - \bar{\text{Exp}})^2}$$

$$\hat{\alpha}_1 = \frac{\text{Var}(x) \text{Var}(y)}{\text{Cov}(x)}$$

$$\hat{\alpha}_1 = \frac{57,73}{5194,95}$$

$$\hat{\alpha}_1 = 0,011$$

$$\hat{\alpha}_0 = \overline{\ln W} - \hat{\alpha}_1 \bar{\text{Exp}}$$

$$\hat{\alpha}_0 = 4,19 - 0,011(82,95)$$

$$\hat{\alpha}_0 = 3,27$$

b. Teniendo en cuenta los coeficientes estimados, escriba la ecuación estimada.

Por tanto la ecuación estimada es:

$$\ln \widehat{W}_i = 3,27 + 0,011 \text{Exp}_i$$

c. Reporte sus resultados en una tabla teniendo en cuenta todos los resultados del taller (ver talleres de semestres anteriores para ejemplos de cómo presentar los resultados en una tabla).

Tabla 1

Variable dependiente: Ln W_i

MCO

Variable	Parámetro	Significancia
Constante	3,27 (22,7)	***
Exp _i	0,011 (6,52)	***
R ²	0,7025	
n	20	

***Nivel de significancia 1%

**Nivel de significancia 5%

*Nivel de significancia 10%

Fuente: Calculos propios en EasyReg

Pregunta 4

a. Interprete los coeficientes estimados y diga si los signos fueron los esperados. Discuta si los coeficientes son significativos. Para esto, realice todo el procedimiento que sea necesario y muestre las formulas que ha utilizado.

Como dijimos, el intercepto no tiene interpretación económica. Por su parte, el valor estimado de la pendiente nos dice que ante un aumento de 1 mes en la experiencia laboral del trabajador i , en promedio, el salario aumentaría en 1,1% *ceteris paribus*. Por su parte, el signo resultó ser el esperado puesto que un aumento en la experiencia aumenta el salario de los individuos de acuerdo a la teoría económica.

Para conocer la significancia de los parámetros se debe realizar una prueba t de significancia individual planteando la siguiente hipótesis nula vs la siguiente hipótesis alterna:

$$H_0: \alpha_i = 0$$

$$H_a: \alpha_i \neq 0$$

Luego se calcula el estadístico t así

$$t = \frac{\hat{\alpha}_i}{s_{\hat{\alpha}_i}}$$

Después se compara con el valor crítico t con n-2 grados de libertad y $\alpha/2$ nivel de significancia, donde si éste último resulta ser menor, podemos rechazar la hipótesis nula a favor de la significancia del parámetro.

La desviación estándar para cada parámetro estimado y la varianza estimada del error se calculan a partir de las siguientes formulas,

$$s^2 = \frac{\sum_{i=1}^{20} (\ln W_i - \widehat{\ln W_i})^2}{n - 2} = 0,0151$$

$$s_{\alpha_1}^2 = \frac{s^2}{\sum_{i=1}^{20} (\text{Exp}_i - \overline{\text{Exp}_i})^2} = 0,002$$

$$s_{\alpha_0}^2 = \frac{s^2 \sum_{i=1}^{20} (\text{Exp}_i)^2}{n \sum_{i=1}^{20} (\text{Exp}_i - \overline{\text{Exp}_i})^2} = 0,14$$

Así, para determinar la significancia del intercepto, planteamos la siguiente hipótesis =

$$H_0 = \alpha_0 = 0$$

$$H_a = \alpha_0 \neq 0$$

Luego obtenemos que con un nivel de significancia del 99% ($\alpha = 0,01$), el $t_{\text{crítico}}$ es =

$t_{\text{crítico}} = t_{\alpha/2} = t_{(0,01/2)} = t_{0,005} = 2.878 \rightarrow$ Valor t encontrado en su respectiva tabla con $gl = n-2 = 18$ y un $\alpha/2 = 0,005$

- Ahora tenemos que el criterio de rechazo de H_0 y aceptar la H_a es =

$|t| > t_{\alpha/2} \rightarrow \frac{\hat{\alpha}_0}{s_{\hat{\alpha}_0}} > t_{\alpha/2} \rightarrow 22,7 > 2.878 \rightarrow$ Efectivamente el estadístico t es mayor que el valor crítico de t.

De este modo hay evidencia suficiente para rechazar la hipótesis nula, por lo tanto, es posible afirmar que α_0 es diferente de cero y por ello podemos inferir, con un nivel de confianza del 99%, que este coeficiente es significativo dentro del modelo.

Por otro lado, para determinar la significancia del coeficiente asociado a la pendiente, planteamos la siguiente hipótesis:

$$H_0 = \alpha_1 = 0$$

$$H_a = \alpha_1 \neq 0$$

Luego obtenemos que con un nivel de significancia del 99% ($\alpha = 0,01$), el $t_{\text{crítico}}$ es =

$$t_{\text{crítico}} = t_{\alpha/2} = t_{(0,01/2)} = t_{0,005} = 2.878$$

- Ahora tenemos que el criterio de rechazo de H_0 y aceptar la H_a es =

$|t| > t_{\alpha/2} \rightarrow \frac{\hat{\alpha}_1}{s_{\hat{\alpha}_1}} > t_{\alpha/2} \rightarrow 6,52 > 2.878 \rightarrow$ Efectivamente el estadístico t es mayor que el valor crítico de t.

De este modo hay evidencia suficiente para rechazar la hipótesis nula, por lo tanto, es posible afirmar que α_1 es diferente de cero y por ello podemos inferir, con un nivel de confianza del 99%, que este coeficiente es significativo dentro del modelo.

b. Construya la tabla ANOVA correspondiente mostrando los valores reemplazados en cada casilla.

TABLA ANOVA			
Fuente de Δ	SS	GdeL	MS
SSR	0,641	1	0,641
SSE	0,272	18	0,015
SST	0,913	19	

c. Discuta qué tan bueno es el modelo explicando el por qué de su respuesta. Recuerde mostrar todos sus cálculos para obtener cualquier resultado.

El R^2 se obtiene a partir de la tabla ANOVA calculando $\frac{SSR}{SST}$ o lo que es lo mismo $1 - \frac{SSE}{SST}$, cuyo resultado es 0,7025, es decir, el 70,25% de las variaciones en $\ln(W_i)$ son explicadas por la experiencia laboral de los trabajadores. Lo cual revela que en general este modelo es bueno.

Pregunta 5

- a. No contento con su arduo trabajo, el Presidente le pide que realice una proyección y un intervalo de confianza al 95% para el nivel de salario esperado del año 2011, si el nivel de experiencia de los trabajadores fuera de 82 meses. Muestre todos sus cálculos e interprete el resultado.

Tenemos que para $Exp = 82$ el nivel de salario será:

$$\hat{W} = 3,27 + 0,011 (82)$$

$$\hat{W} = 4,18$$

La proyección en la media de la variable dependiente se calcula a partir de:

$$4,18 \pm 2,101 * 0,12 = \sqrt{\frac{1}{20} + \left(\frac{(82 - 82,95)^2}{0,913}\right)}$$

Así, el intervalo de confianza para el valor esperado es [3,917 ; 4,443]. Se puede decir con un 95% de confiabilidad que dada una experiencia laboral de 82 meses, el salario estará entre 3,917 y 4,443 dólares.

- b. Además, el Vicepresidente le pide que realice otra proyección con un intervalo de confianza al 95% para el nivel de salario del año 2011, si el nivel de experiencia de los trabajadores fuera de 105 meses. Muestre todos sus cálculos e interprete el resultado.

Tenemos que para $Exp = 105$ el nivel de salario será:

$$\hat{W} = 3,27 + 0,011 (105)$$

$$\hat{W} = 4,43$$

La proyección individual de la variable dependiente se calcula a partir de:

$$4,18 \pm 2,101 * 0,12 = \sqrt{1 + \frac{1}{20} + \left(\frac{(82 - 82,95)^2}{0,913}\right)}$$

Así, el intervalo de confianza para el valor esperado es [3,812 ; 4,549]. Se puede decir con un 95% de confiabilidad que dada una experiencia laboral de 105 meses, el salario estará entre 3,812 y 4,549 dólares.

Se penalizará al estudiante que no reconozca la diferencia entre una proyección individual y en la media.

Pregunta 6

El asesor financiero del Presidente de la Compañía, le comenta a usted que el modelo que empleó en el estudio no es el único que sirve para analizar el comportamiento del salario y por ello, le propone las siguientes expresiones, pero no sabe si todas pueden ser estimadas por MCO. Demuestre en cualquier caso, utilizando transformaciones apropiadas, si dichos modelos no lineales pueden ser linealizados o no para poder ser estimados por dicho método:

a. $W_i = AExp_{1i}^{\alpha_1} e^{\varepsilon_i}$

Este modelo es linealizable calculando \ln a ambos lados de la ecuación para obtener

$$\ln W_i = \ln(A) + \alpha_1 \ln Exp_{1i} + \varepsilon_i$$

Reparametrizando $\ln(A) = \alpha_0$

$$\ln W_i = \alpha_0 + \alpha_1 \ln Exp_{1i} + \varepsilon_i$$

b. $W_i = (\alpha_0 + \alpha_1 Exp_i + \varepsilon_i)^{-1}$

Este modelo es linealizable calculando:

$$W_i = \frac{1}{\alpha_0 + \alpha_1 Exp_i + \varepsilon_i}$$

$$\frac{1}{W_i} = \alpha_0 + \alpha_1 Exp_i + \varepsilon_i$$

Reparametrizando $\frac{1}{W_i} = Z_i$

$$Z_i = \alpha_0 + \alpha_1 Exp_i + \varepsilon_i$$

Finalmente, este modelo puede estimarse por MCO.

c. $W_i = \alpha_0 + \alpha_1^2 Exp_i + \varepsilon_i$

El modelo no es lineal en el parámetro α_1 . Aunque se podría reparametrizar para ser estimado por MCO, no se recuperaría el parámetro original del modelo, es decir, conocer tanto su signo como su valor.

d. $\frac{1}{W_i} = \theta_1 \left(\frac{1}{\sin(\theta_3) + \ln(Exp_{1i})} \right)^{\theta_2} \mu_i$

Este no es lineal en los parámetros lo cual se puede mostrar de la siguiente manera:

$$\ln\left(\frac{1}{W_i}\right) = \ln\left(\theta_1 \left(\frac{1}{\text{sen}(\theta_3) + \text{Ln}(\text{Exp}_{1i})}\right)^{\theta_2} \mu_i\right)$$

$$-\ln(W_i) = \ln(\theta_1) + \ln\left(\left(\frac{1}{\text{sen}(\theta_3) + \text{Ln}(\text{Exp}_{1i})}\right)^{\theta_2}\right) + \ln(\mu_i)$$

$$-\ln(W_i) = \ln(\theta_1) - \theta_2 \ln(\text{sen}(\theta_3) + \ln(\text{Exp}_{1i})) + \ln(\mu_i)$$