Taller #5

Regresión Múltiple
Econometría 06216

Profesor: Julio César Alonso
Monitora: Stephanie Vergara
Mauricio A. Arcos

Notas:

- Recuerde que sólo dos preguntas, seleccionadas al azar, serán calificadas.
- Este taller es para ser entregado en los primeros 10 minutos de la clase del próximo 14 de febrero.

INSTRUCCIONES:

- Este taller debe ser escrito en computador y entregado en papel.
- Cuando sea posible, debe mostrar el procedimiento efectuado para llegar a sus resultados.

El Departamento de Planeación Nacional desea estimar la función de producción (P) del sector público utilizando como variables explicativas el trabajo (L) y el stock de capital (K). Usted, dado su renombrada experiencia ha sido encargado para realizar este estudio a partir de la información disponible en el archivo T5-01-05.xls.

1. A partir de esta información se desea:
   a) Estimar una función de producción lineal. Escriba el modelo, estímelo y repórtelo en una tabla.
   b) Verificar la igualdad entre las productividades marginales del trabajo y capital.
2. Continuando con la pregunta anterior, se desea
   a) Estimar una función de producción tipo Cobb-Douglas. Escriba el modelo, estímelo y repórtelo en una tabla.
   b) Verificar la existencia de rendimientos constantes a escala.

Un investigador supone para cada empresa t de la industria minera la siguiente función de costos:

$$y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \varepsilon_t \qquad t = 1, 2, \ldots.. \qquad (*)$$

donde y_t representa los costos totales de la empresa t, X_{2t} representa el precio del factor de producción A pagado por la empresa t, X_{3t} representa el logaritmo del precio del factor de producción B pagado por la empresa t y X_{4t} es la cantidad de toneladas producidas por la firma t. Además, el investigador supone que los errores (ε_t) son independientes entre si y tienen media cero y varianza constante. Una muestra de 20 empresas de esta industria produjo los siguientes resultados:

$$X^T y = \begin{pmatrix} 292 \\ 908 \\ 1232 \\ 976 \end{pmatrix} \qquad X^T X = \begin{pmatrix} 20 & 60 & 80 & 70 \\ 60 & 200 & 280 & 180 \\ 80 & 280 & 400 & 220 \\ 70 & 180 & 220 & 300 \end{pmatrix} \qquad y^T y = 4314.97$$

Por razones teóricas se tiene la siguiente relación:

$$\beta_2 + \beta_3 = 1 \qquad (**)$$

3. A partir de la información anterior,
   a) Muestre que el modelo (*) bajo la restricción (**), se puede trasformar en,

$$z_t = \beta_1 + \beta_2 W_t + \beta_4 X_{4t} + \varepsilon_t \qquad t = 1, 2, \ldots.. \qquad (***)$$

con $z_t = y_t - X_{3t}$ y $W_t = X_{2t} - X_{3t}$.

   b) Interprete el significado de los coeficientes del modelo (***)
4. Continuando con la pregunta anterior,
   a) Estime los parámetros $\beta^T = \begin{pmatrix}\beta_1 & \beta_2 & \beta_4\end{pmatrix}$ del modelo (***) por medio del método MCO.
5. Continuando con la pregunta anterior,
   a) Estime σ^2 y la matriz de varianzas y covarianzas de los estimadores MCO de β.
   b) Obtenga estimadores insesgados para β_3, $\left[\sigma_{(\beta_3)}\right]^2$ y $\sigma_{(\beta_3, \beta_2)}$.
6. Continuando con la pregunta anterior,
   a) Construya la tabla ANOVA para el modelo (***).
   b) Encuentre el R^2 (interprételo) y pruebe la hipótesis que todas las pendientes en el modelo (***) son cero. Explique sus resultados.

Taller #5

Respuestas Sugeridas
Regresión Múltiple
Econometría 06216

Profesor: Julio César Alonso
Monitora: Stephanie Vergara
Mauricio A. Arcos

Notas:

- Recuerde que sólo dos preguntas, seleccionadas al azar, serán calificadas.
- Este taller es para ser entregado en los primeros 10 minutos de la clase del próximo 14 de febrero.

INSTRUCCIONES:

- Este taller debe ser escrito en computador y entregado en papel.
- Cuando sea posible, debe mostrar el procedimiento efectuado para llegar a sus resultados.

El Departamento de Planeación Nacional desea estimar la función de producción (P) del sector público utilizando como variables explicativas el trabajo (L) y el stock de capital (K). Usted, dado su renombrada experiencia ha sido encargado para realizar este estudio a partir de la información disponible en el archivo T5-01-05.xls.

1. A partir de esta información se desea:
   a) Estimar una función de producción lineal. Escriba el modelo, estímelo y repórtelo en una tabla.

La función de producción de tipo lineal sigue la siguiente expresión:

$$P_t = \alpha_0 + \alpha_1 L_t + \alpha_2 K_t + \varepsilon_t \quad (1)$$

Las estimaciones se encuentran registradas en la tabla 1.

Tabla 1. Estimaciones de diferentes modelos

	Estadísticos t entre paréntesis	
	Ecuación 1 Niveles MCO	Ecuación 2 Logaritmos + MCO
constante	-1832783,11 (-2,53) **	4,3929 (2,56) **
L_t	2999,48206 *** (6,42) ***	0,6178 (1,74)
K_t	0,25526 (14,44) ***	0,4123 (6,53) ***
R^2	0,99630	0,99240
F	(1.218,62) ***	(588,73) ***
# de Obs.	12	12

(*) nivel de significancia: 10%
(**) nivel de significancia: 5%
(***) nivel de significancia: 1%
MCO: Mínimos Cuadrados Ordinarios
+ tanto la variable dependiente como las independientes se encuentran en logaritmos

b) Verificar la igualdad entre las productividades marginales del trabajo y capital.

Para verificar la igualdad entre las productividades marginales, se procede a realizar una contraste de hipótesis

$$H_0 : \alpha_1 - \alpha_2 = 0$$
$$H_1 : \alpha_1 - \alpha_2 \neq 0$$

Al realizar el test de Wald, el valor crítico es de 41.18. Se aprecia claramente que el p-valor es muy asociado (0.000) a éste es menor que cualquier nivel de significancia; por lo tanto, se rechaza la hipótesis nula de que las productividades marginales son iguales.

2. Continuando con la pregunta anterior, se desea
   a) Estimar una función de producción tipo Cobb-Douglas. Escriba el modelo, estímelo y repórtelo en una tabla.

La función de producción tipo Cobb-Douglas corresponde a la siguiente expresión:

$$P_t = AL_t^{\beta_1} K_t^{\beta_2} e^{\varepsilon_t}$$

Con el fin de estimar el modelo se procede a linealizarlo por medio de logaritmos naturales

$$LnP_t = \beta_0 + \beta_1 LnL_t + \beta_2 LnK_t + \varepsilon_t \quad (2)$$

Donde $Ln\ A = \beta_0$

Las estimaciones se encuentran registradas en la tabla 1.

b) Verificar la existencia de rendimientos constantes a escala.

Para verificar la existencia de rendimientos constantes a escala , se procede a realizar un contraste ente las siguientes hipótesis

$$H_0 : \beta_1 + \beta_2 = 1$$
$$H_1 : \beta_1 + \beta_2 \neq 1$$

Puesto que el p-valor asociado al valor crítico de 0.01 con un grado de libertad (91.835%) es mayor que el nivel todo nivel de significancia, no se puede rechazar la hipótesis nula; por lo tanto, se considera la existencia de rendimientos constantes a escala.

Un investigador supone para cada empresa t de la industria minera la siguiente función de costos:

$$y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \varepsilon_t \qquad t = 1, 2, ... \qquad (*)$$

donde y_t representa los costos totales de la empresa t, X_{2t} representa el precio del factor de producción A pagado por la empresa t, X_{3t} representa el logaritmo del precio del factor de producción B pagado por la empresa t y X_{4t} es la cantidad de toneladas producidas por la firma t. Además, el investigador supone que los errores (ε_t) son independientes entre si y tienen media cero y varianza constante. Una muestra de 20 empresas de esta industria produjo los siguientes resultados:

$$X^T y = \begin{pmatrix} 292 \\ 908 \\ 1232 \\ 976 \end{pmatrix} \qquad X^T X = \begin{pmatrix} 20 & 60 & 80 & 70 \\ 60 & 200 & 280 & 180 \\ 80 & 280 & 400 & 220 \\ 70 & 180 & 220 & 300 \end{pmatrix} \qquad y^T y = 4314.97$$

Por razones teóricas se tiene la siguiente relación:

$$\beta_2 + \beta_3 = 1 \qquad (**)$$

3. A partir de la información anterior,
   a) Muestre que el modelo (*) bajo la restricción (**), se puede trasformar en,

$$z_t = \beta_1 + \beta_2 W_t + \beta_4 X_{4t} + \varepsilon_t \qquad t = 1, 2, ... \qquad (***)$$

con $z_t = y_t - X_{3t}$ y $W_t = X_{2t} - X_{3t}$.

Noten que la restricción (**) se puede rescribir como $\beta_3 = 1 - \beta_2$. Remplazando esta restricción en (*) tenemos:

$$y_t = \beta_1 + \beta_2 X_{2t} + (1 - \beta_2) X_{3t} + \beta_4 X_{4t} + \varepsilon_t$$
$$y_t = \beta_1 + \beta_2 \cdot X_{2 \cdot t} + X_{3t} - \beta_2 \cdot X_{3t} + \beta_4 \cdot X_{4 \cdot t} + \varepsilon_t$$
$$y_t - X_{3t} = \beta_1 - \beta_2 \cdot \left(X_{2 \cdot t} - X_{3 \cdot t}\right) + \beta_4 \cdot X_{4 \cdot t} + \varepsilon_t$$

Ahora, definamos $z_t = y_t - X_{3t}$ y $W_t = X_{2t} - X_{3t}$. Entonces tenemos

$$z_t = \beta_1 + \beta_2 W_t + \beta_4 X_{4t} + \varepsilon_t \qquad t = 1, 2, ...$$

b) Interprete el significado de los coeficientes del modelo (***)

Noten que el modelo no ha sufrido transformaciones lineales, por lo que podemos interpretar sus coeficientes de la misma manera que el modelo (*), así:

$\hat{\beta}_1$ Son los costos de la empresa que no dependen ni del precio del factor de producción A pagado por la empresa, ni del logaritmo del precio del factor de producción B pagado por la empresa, ni de las toneladas producidas por la empresa.

$\hat{\beta}_2$ Es el monto en el que incrementan los costos totales de la empresa cuando hay un incremento de una unidad en el precio del factor de producción A pagado por la empresa.

$\hat{\beta}_4$ Cuando incrementan la producción de la empresa en una tonelada, los costos totales de la empresa se incrementan en $\hat{\beta}_4$ unidades.

4. Continuando con la pregunta anterior,
   a) Estime los parámetros $\beta^T = (\beta_1 \ \ \beta_2 \ \ \beta_4)$ del modelo (***) por medio del método MCO.

Noten que para este caso tenemos

$$z_t = \beta_1 + \beta_2 W_{2t} + \beta_4 X_{4t} + \varepsilon_t \qquad t = 1, 2, ...$$

entonces,

$$X^T X = \begin{bmatrix} n & \sum_{t=1}^{n} W_t & \sum_{t=1}^{n} X_{4t} \\ \blacksquare & \sum_{t=1}^{n} (W_t)^2 & \sum_{t=1}^{n} W_t \cdot X_{4t} \\ \blacksquare & \blacksquare & \sum_{t=1}^{n} (X_{4t})^2 \end{bmatrix}$$

Además noten que $\sum_{t=1}^{n} W_t = \sum_{t=1}^{n} (X_{2t} - X_{3t}) = \sum_{t=1}^{n} X_{2t} - \sum_{t=1}^{n} X_{3t} = 60 - 80 = -20$,

$$\sum_{t=1}^{n} W_t \cdot X_{4t} = \sum_{t=1}^{n} (X_{2t} - X_{3t}) \cdot X_{4t} = \sum_{t=1}^{n} X_{2t} \cdot X_{4t} - \sum_{t=1}^{n} X_{3t} \cdot X_{4t} = 180 - 220 = -40, \quad \text{y}$$

$$\sum_{t=1}^{n} (W_t)^2 = \sum_{t=1}^{n} (X_{2t} - X_{3t})^2 = \sum_{t=1}^{n} \left[(X_{2 \cdot t})^2 - (2 \cdot X_{2 \cdot t} \cdot X_{3 \cdot t}) + (X_{3 \cdot t})^2 \right] = \sum_{t=1}^{n} (X_{2 \cdot t})^2 - \blacksquare$$

$$2\cdot\sum_{t=1}^{n}\left(X_{2\cdot t}\cdot X_{3\cdot t}\right)+\sum_{t=1}^{n}\left(X_{3\cdot t}\right)^2 = 200-2\cdot 280+400 = 40$$

. Entonces tenemos que $X^T X = \begin{pmatrix} 20 & -20 & 70 \\ -20 & 40 & -40 \\ 70 & -40 & 300 \end{pmatrix}$.

Similarmente necesitamos encontrar $X^T z = \begin{pmatrix} \sum_{t=1}^{n} z_t \\ \sum_{t=1}^{n} W_t z_t \\ \sum_{t=1}^{n} X_{4t} z_t \end{pmatrix}$. Así

$$\sum_{t=1}^{n} z_t = 292-80 = 212$$

$$\sum_{t=1}^{n} W_t z_t = \sum_{t=1}^{n}\left(X_{2t}-X_{3t}\right)\left(y_t-X_{3t}\right) = 908-1232-280+400 = -204$$

$$\sum_{t=1}^{n} X_{4t} z_t = \sum_{t=1}^{n} X_{4t}\left(y_t-X_{3t}\right) = 976-220 = 756$$

Entonces tenemos que $X^T z = \begin{pmatrix} 212 \\ -204 \\ 756 \end{pmatrix}$

Ahora si podemos calcular nuestro estimadores MCO ($\beta hat = \left(X^T X\right)^{-1} X^T z$).Pero antes neceisatamos calcular la inversa de $X^T X$.. En este caso la inversa es:

$$\left(X^T X\right)^{-1} = \begin{pmatrix} \frac{13}{5} & \frac{4}{5} & \frac{-1}{2} \\ \frac{4}{5} & \frac{11}{40} & \frac{-3}{20} \\ \frac{-1}{2} & \frac{-3}{20} & \frac{1}{10} \end{pmatrix}$$

Entonces

$$\beta hat = \left(X^T X\right)^{-1} X^T z = \begin{pmatrix} \frac{13}{5} & \frac{4}{5} & \frac{-1}{2} \\ \frac{4}{5} & \frac{11}{40} & \frac{-3}{20} \\ \frac{-1}{2} & \frac{-3}{20} & \frac{1}{10} \end{pmatrix}\cdot\begin{pmatrix} 212 \\ -204 \\ 756 \end{pmatrix} = \begin{pmatrix} 10 \\ \frac{1}{10} \\ \frac{1}{5} \end{pmatrix}$$

5. Continuando con la pregunta anterior,

a) Estime σ^2 y la matriz de varianzas y covarianzas de los estimadores MCO de β.

Recuerden que

$$s^2 = \frac{(z)^T\cdot z - (\beta hat)^T\cdot (X)^T\cdot z}{n-k}$$

Es muy fácil mostrar que $z^T z = 2250.97$, entonces

$$s^2 = \frac{2250.97 - \begin{pmatrix} 10 & \frac{1}{10} & \frac{1}{5} \end{pmatrix}\cdot\begin{pmatrix} 212 \\ -204 \\ 756 \end{pmatrix}}{17} = \frac{2250.97-\frac{11254}{5}}{17} = \frac{1}{100}$$

Y la matriz de varianzas y covarianzas delos estimadores MCO es

$$s^2\cdot\left(X^T X\right)^{-1} = \frac{1}{100}\cdot\begin{pmatrix} \frac{13}{5} & \frac{4}{5} & \frac{-1}{2} \\ \frac{4}{5} & \frac{11}{40} & \frac{-3}{20} \\ \frac{-1}{2} & \frac{-3}{20} & \frac{1}{10} \end{pmatrix} = \begin{pmatrix} \frac{13}{500} & \frac{1}{125} & \frac{-1}{200} \\ \frac{1}{125} & \frac{11}{4000} & \frac{-3}{2000} \\ \frac{-1}{200} & \frac{-3}{2000} & \frac{1}{1000} \end{pmatrix}$$

b) Obtenga estimadores insesgados para β_3, $\left[\sigma_{(\beta_3)}\right]^2$ y $\sigma_{(\beta_3,\beta_2)}$.

Como $\beta_3 = 1-\beta_2$, entonces $\beta hat_3 = 1-\beta hat_2 = 1-\frac{1}{10} = \frac{9}{10}$ y además noten que $Var(\beta hat_3) = Var(1-\beta hat_2) = Var(\beta hat_2)$. Por tanto $\left[s_{(\beta_3)}\right]^2 = \frac{11}{4000}$ y entonces tenemos que Cov(βhat3,βhat2)= Cov(1-βhat2,βhat2)= Cov(-βhat2,βhat2)=-Var(,βhat2)=-11/4000.

6. Continuando con la pregunta anterior,
   a) Construya la tabla ANOVA para el modelo (***).

Sabemos que la tabla ANOVA viene dada por:

Tabla ANOVA

Fuente de variación	SS	G de L	MS
Regresión	$SSR = \hat{\beta}^T X^T z - n\bar{z}^2$	$k-1$	$\frac{SSR}{k-1}$
Error (Residuos)	$SSE = z^T z - \hat{\beta}^T X^T z$	$n-k$	$\frac{SSE}{n-k}$
Total	$z^T z - n\bar{z}^2$	$n-1$	

En este caso tenemos que $k := 3$, $n := 20$, $z^T z = 2250.97$, $(\beta hat)^T\cdot(X)^T\cdot z = \frac{11254}{5}$ y $zbar = \frac{212}{20}$. Asi, la tabla Anova es

Tabla ANOVA

Fuente de variación	SS	G de L	MS
Regresión	$\frac{11254}{5} - \left[20 \cdot \left(\frac{212}{20}\right)^2\right] = \frac{18}{5}$	$3 - 1 = \blacksquare$	$\frac{9}{5} = \blacksquare$
Error (Residuos)	$2250.97 - \frac{11254}{5} = \frac{17}{100} = 0.17$	$20 - 3 = \blacksquare$	$\frac{1}{100} = \blacksquare$
Total	$\frac{18}{5} + \frac{17}{100} = \frac{377}{100} = 3.77$	19	

b) Encuentre el R^2 (interpretelo) y pruebe la hipótesis que todas las pendientes en el modelo (***) son cero. Explique sus resultados.

Recuerde que $R^2 = \frac{SSR}{SST}$, en este caso $R^2 = \frac{SSR}{SST} = \frac{\frac{18}{5}}{\frac{377}{100}} = \frac{360}{377} = 0.955$. Este R^2 implica que nuestro modelo explica el 95.5% de la variabilidad de la variable z.

El F calculado par probar la hipótesis nula que todas las pendientes son iguales a cero es dado por

$$F_c = \frac{MSR}{MSE} = \frac{\frac{9}{5}}{\frac{1}{100}} = 180$$

Este F calculado se debe comparar con el F de la tabla con 2 grados de libertad en el númerador y 17 grados de libertad en el denominador. Par un nivel de signicficancia del 5% este F de la tabla es 6.11. Entonces, podemos rechazar la hipótesis nula que todas las pendientes son al mismo tiempo iguales a cero.