Universidad Carlos III de Madrid

Estadística II: Introduccion a la Econometría

Examen Final, Convocatoria Extaordinaria, Curso 2007-2008.

22 de Septiembre de 2008

PREGUNTA 1

Suponga que se han estimado dos modelos con una muestra de 750 trabajadores de una industria americana a nivel nacional

$$\begin{aligned}\widehat{w} &= \underset{(0.2)}{10} + \underset{(0.7)}{5}\exp + \underset{(0.9)}{4}\,male, \\ n &= 750,\ R^2 = 0.3,\end{aligned} \tag{1}$$

$$\begin{aligned}\widehat{w} &= \underset{(0.2)}{10} + \underset{(0.7)}{5}\exp + \underset{(0.9)}{4}\,male + \underset{(0.6)}{0.3}region + \underset{(0.5)}{0.2}(male * region) \\ n &= 750,\ R^2 = 0.528\end{aligned} \tag{2}$$

donde w es el salario del individuo, exp es su experiencia laboral, $male$ es una variable dummy que toma el valor 0 si el individuo es una mujer y 1 si es un hombre, $region$ es una dummy que toma el valor 1 para el norte y 0 para el sur, y $male * region$ es la interacción entre la dummy de región y de sexo. Entre paréntesis aparecen los errores estándard de cada coeficiente. El estimador de la matriz de varianzas y covarianzas de los estimadores en el modelo (2) es

$$\begin{pmatrix} 0.04 & 2.1E-05 & -1.6E-08 & 3.2E-05 & 7.6E-04 \\ 2.1E-05 & 0.49 & 0.09 & 0.01 & 6.4E-06 \\ -1.6E-08 & 0.09 & 0.81 & 0.05 & 0.07 \\ 3.2E-05 & 0.01 & 0.05 & 0.36 & 0.08 \\ 7.6E-04 & 6.4E-06 & 0.07 & 0.08 & 0.25 \end{pmatrix}$$

a. (0.75 puntos) Si se estimara el modelo (1) redefiniendo la dummy como 1 para mujeres y 0 para hombres, ¿qué resultados obtendría con esta especificación? Escriba una ecuación similar a la anterior detallando cuáles serían los coeficientes estimados, los errores estándard de los mismos y el R^2 de la regresión.

Respuesta: $female = 1 - male$

$$\begin{aligned} \widehat{w} &= \underset{(0.2)}{14} + \underset{(0.7)}{5} \exp - \underset{(0.9)}{4} female \\ n &= 750,\ R^2 = 0.3 \end{aligned}$$

El coeficiente estimado de $female$ sería igual a -4 con un error estándar de 0.9 y el estimador de la constante sería 14 con un error estándar de 0.2. El R^2 no varía.

b. (0.75 puntos) Contraste la hipótesis conjunta de la falta de significatividad de que un trabajador pertenezca a una región u otra. (Escriba la hipótesis nula y la alternativa, el estadístico del contraste y la regla de decisión).

Respuesta: En el modelo:

$$w = \beta_0 + \beta_1 \exp + \delta_1 male + \delta_2 region + \delta_3 (male * region) + error$$

$H_0 : \delta_2 = \delta_3 = 0$ vs $H_1 : \delta_2 \neq 0$ or $\delta_3 \neq 0$. Estadístico:

$$F = \frac{(0.528 - 0.3)}{(1 - 0.528)/750} = 362.29$$

El estadístico se distribuye aproximadamente, bajo la hipótesis nula, como una chi-cuadrado con dos grados de libertad. Se rechaza la hipótesis nula a cualquier nivel de significación razonable. Se podría haber utilizado la aproximación de la $F_{(2,750-4-1)}$, utilizando como estadístico $F = 179.94$. La conclusión, obviamente sería la misma.

c. (0.75 puntos) Contraste la hipótesis nula de que las mujeres del norte tienen un salario igual al de las mujeres del sur, frente a la alternativa de que es mayor. (Escriba la hipótesis nula y la alternativa, el estadístico del contraste y la regla de decisión).

Respuesta: $H_0 : \delta_2 = 0$ vs $H_1 : \delta_2 \neq 0$. Estadístico: $t = 0.3/0.6 = 0.5$. No podemos rechazar la hipótesis nula a ningún nivel de significación razonable..

d. (0.75 puntos) Contraste la hipótesis nula de que el salario de las mujeres del norte no es diferente al de los hombres del sur. (Escriba la hipótesis nula y la alternativa, el estadístico del contraste y la regla de decisión).

Respuesta: $H_0 : \delta_1 = \delta_2$ vs $H_1 : \delta_1 \neq \delta_2$. Estadístico: $t = (4-0.3)/\sqrt{0.81 + 0.36 - 2 \cdot 0.05.} = 3.57$. Rechazamos H_0 al 1% de significación.

PREGUNTA 2

La variable *rdintens* representa el gasto en Investigación y Desarrollo (I+D) como un porcentaje de las ventas facturadas, medidas en millones de dólares. La variable *profmarg* denota los beneficios como un porcentaje de las ventas. Para explicar los gastos en I+D, consideramos el siguiente modelo poblacional:

$$rdintens = \beta_0 + \beta_1 \log(sales) + \beta_2 profmarg + u$$

Utilizando datos de 32 empresas en al industria química, la ecuación estimada es:

$$\widehat{rdintens} = \underset{(1.369)}{0.472} + \underset{(0.216)}{0.321} \log(sales) + \underset{(0.046)}{0.050} profmarg$$

$$n = 32, \quad R^2 = 0.099$$

a. (1 punto) Interprete el coeficiente asociado a $\log(sales)$. Si las ventas se incrementan en un 10%, ¿cuál es la estimación del cambio que experimentará *rdintens*?

Respuesta: Cuando las ventas aumentan en un 1%, permaneciendo fijo *profmarg*, estimamos que el porcentaje de beneficios sobre las ventas variará en 0.00321 puntos. Si se incrementa en un 10%, el cambio en el porcentaje será de 0.0321 puntos.

b. (1 punto) ¿Bajo que supuesto es correcta la interpretación que se ha dado en a.?

Respuesta: Bajo el supuesto de que la esperanza condicional de los errores dadas todas las variables explicativas es igual a cero.

c. (1 punto) Utilizar un contraste de la t de que la intensidad en I+D no es afectada por el nivel de beneficios sobre ventas (escriba la hipótesis nula y la alternativa, el estadístico del contraste y la regla de decisión). Contrate al 5% y al 10%.

Respuesta: $H_0 : \beta_2 = 0$ vs $H_1 : \beta_2 \neq 0$. Estadístico: $t = 0.05/0.046 \simeq 1.087$. no se puede rechazar al 5% de significación.

PREGUNTA 3

Considere el modelo de regresión:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + U.$$

a. (1 punto) ¿Cuál de los siguientes casos puede provocar sesgo en los estimadores de mínimos cuadrados ordinarios?

(i) Que la varianza condicional de U dadas X_1 y X_2 varíe con X_1 y/o X_2.

(ii) Un coeficiente de correlación muestral entre X_1 y X_2 de 0.95.

(iii) Que la esperanza condicional de U dado X_1 y X_2 sea diferente de cero.

Respuesta: El (iii).

b. (1 punto) Suponiendo que U cumple los supuestos clásicos, que $\beta_2 \neq 0$, y dada una muestra aleatoria $(Y_i, X_{1i})\,, i = 1, ..., n$ de (Y, X_1), proporcione una fórmula para el sesgo del estimador de β_1

$$\bar{\beta}_1 = \frac{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)X_{1i}}{\sum_{i=1}^{n}\left(X_{1i} - \bar{X}_1\right)^2},$$

donde $\bar{Y}$ y $\bar{X}_1$ son las medias muestrales de Y y X_1, respectivamente. Indique bajo que circunstancias el estimador es insesgado cuando n es muy grande.

Respuesta: Sesgo:

$$sesgo\left(\bar{\beta}_1\right) = \beta_2 \frac{Cov\left(X_1, X_2\right)}{Var\left(X_1\right)},$$

donde Cov denota covarianza y Var varianza. El sesgo es igual a cero cuando la covarianza entre X_1 y X_2 es igual a cero. (Hemos supuesto $\beta_2 \neq 0$).

c. (1 punto) Suponga que para cualquier realización de $(X_1, X_2)\,,\ X_1 + X_2 = 0.27$. ¿Qué supuesto clásico se está incumpliendo? ¿Cuáles serán las consecuencias sobre el estimador de mínimos cuadrados ordinarios de β_1 y β_2?.

Respuesta: El supuesto de multicolinealidad (colinealidad perfecta). Como consecuencia inmediata el estimador de mínimos cuadrados ordinarios no se puede computar.

d. (1 punto) Suponiendo que se cumplen los supuestos clásicos y que $\beta_2 = 0$, considere el estimador:

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n} Y_i X_{1i}}{\sum_{i=1}^{n} X_{1i}^2}.$$

¿Cuándo es este estimador insesgado, aún cuando n es pequeña?

Respuesta: Cuando $\beta_0 = 0$.