

Soluciones Hoja de Ejercicios 4

Econometría I

1. Considera la siguiente ecuación para explicar los salarios de los directores generales en términos de ventas anuales de la compañía (*sales*), del rendimiento de la acción (*roe*, en porcentaje) y rendimiento del capital (*ros*, en porcentaje),

$$\log(\text{salary}) = \beta_0 + \beta_1 \log(\text{sales}) + \beta_2 \text{roe} + \beta_3 \text{ros} + u.$$

- a) En términos de los parámetros del modelo, establece la hipótesis nula de que, después de controlar por ventas y *roe*, *ros* no tiene efecto sobre el salario. Establece la alternativa de que un mejor comportamiento del rendimiento del capital mejora el salario.

$$H_0 : \beta_3 = 0, \quad H_1 : \beta_3 > 0.$$

- b) Estima el modelo anterior usando los datos de CEOSAL1. ¿Cuál es el porcentaje de incremento de *salary* si *ros* aumenta en 50 puntos porcentuales?

$50 * (100 * 0,000241655) = 1.2083\%$ ¿Tiene *ros* un efecto práctico grande sobre *salary*? No.

$$\begin{aligned} \log(\widehat{\text{salary}}) &= 4,31171 + 0,280315 \log(\text{sales}) + 0,0174168 \text{roe} + 0,000241655 \text{ros} \\ &\quad (0,31543) \quad (0,03532) \quad (0,0040923) \quad (0,0005418) \\ n &= 209 \quad R^2 = 0,2827 \end{aligned}$$

- c) Contrasta la hipótesis de que *ros* no tiene efecto sobre *salary*, contra la alternativa de que *ros* tiene un efecto positivo. Realiza el contraste al nivel del 10%.

Para un contraste unilateral, el valor crítico de una $t_{209-3-1} = t_{205} \approx 1,282$. Usando las hipótesis en (a),

$$t = \frac{\hat{\beta}_3}{se(\hat{\beta}_3)} = \frac{0,000241655}{0,0005418} = 0,44602,$$

que no es significativo: no podemos rechazar H_0 en favor de H_1 .

- d) ¿Incluirías *ros* es un modelo final para estimar la compensación al director general en términos de los resultados de la compañía? Explica tu respuesta.

No.

2. ¿Influye la población de estudiantes sobre los alquileres en una ciudad universitaria? Sea *rent* es el alquiler medio mensual pagado por apartamento en una ciudad universitaria en los Estados Unidos. Sea *pop* la población total en la ciudad, *avginc* la renta media de la ciudad y *pctstu* la población estudiantil como porcentaje de la población total. Un model para contrastar la existencia de relación es

$$\log(\text{rent}) = \beta_0 + \beta_1 \log(\text{pop}) + \beta_2 \log(\text{avginc}) + \beta_3 \text{pctstu} + u.$$

- a) Establece la hipótesis nula de que el tamaño relativo de la población de estudiantes respecto a la población total no tiene efecto *ceteris paribus* sobre el alquiler mensual. Establezca la alternativa de que hay un efecto.

$$H_0 : \beta_3 = 0, \quad H_1 : \beta_3 \neq 0.$$

b) ¿Cuáles son los signos esperados de β_1 y β_2 ?

En principio esperaríamos que $\beta_1 > 0$ y $\beta_2 > 0$.

c) Estima la ecuación usando los datos de RENTAL para 128 ciudades universitarias.

¿Qué es lo que está mal en la afirmación "un incremento del 10% en población está asociado con un incremento aproximado del 3,1% en los alquileres"?

$$\widehat{\log(\text{rent})} = -3,36831 + 0,0313455 \log(\text{pop}) + 0,877139 \log(\text{avginc}) + 0,00658486 \text{pctstu}$$

$$\begin{matrix} (0,46394) & (0,027079) & (0,041325) & (0,0012027) \end{matrix}$$

$$n = 128 \quad R^2 = 0,7971$$

El incremento del 3.1% es un efecto ceteris paribus, por lo que habría que añadir todos los otros factores constantes.

d) Contrasta la hipótesis establecida en (a) al nivel 1%.

$$t = \frac{\hat{\beta}_3}{\text{se}(\hat{\beta}_3)} = \frac{0,00658486}{0,0012027} = 5.4751,$$

que es muy significativo.

3. En clase estudiamos el ejemplo del contraste de la racionalidad de las tasaciones de los precios de las casas mediante un modelo log-log en price and assess. Ahora usaremos un modelo nivel-nivel.

a) En el modelo de regresión simple

$$\text{price} = \beta_0 + \beta_1 \text{assess} + u,$$

la tasación es racional si $\beta_1 = 1$ y $\beta_0 = 0$. La ecuación estimada es

$$\text{price} = -14,47 + 0,976 \text{assess}$$

$$\begin{matrix} (16,27) & (0,049) \end{matrix}$$

$$n = 88, SSR = 165644,51; R^2 = 0,820.$$

Primero, contrasta la hipótesis de que $H_0: \beta_0 = 0$ en contra de una alternativa bilateral.

$$t_{\beta_0} = \frac{-14,47}{16,27} = -0,88937,$$

que no es significativo, por lo que no rechazamos H_0 .

Entonces contrasta $\beta_1 = 1$ en contra de una alternativa bilateral.

$$t_{\beta_1} = \frac{0,976 - 1}{0,049} = -0,48980,$$

que tampoco es significativo a los niveles habituales.

¿Cuál es tu conclusión?

En principio parece que no podemos rechazar la hipótesis de que $\beta_1 = 1$ y $\beta_0 = 0$, pero los contrastes de la t individuales sólo proporcionan evidencia parcial, debemos realizar un contraste conjunto.

- b) Para contrastar la hipótesis conjunta de que $\beta_1 = 1$ y $\beta_0 = 0$, se necesita la SSR de un modelo restringido. Esto supone computar

$$\sum_{i=1}^n (\text{price}_i - \text{assess}_i)^2$$

cuando $n = 88$, ya que los residuos del modelo restringido son $\text{price}_i - \text{assess}_i$ (no hay que estimar ningún parámetro en este caso, ya que todos están especificados bajo H_0). Esto resulta en $SSR = 209448,99$. Efectúa el contraste de la F para hipótesis conjuntas.

$$F = \frac{SSR_r - SSR_{nr}}{SSR_{nr}} \frac{n - k - 1}{q} = \frac{209448,99 - 165644,51}{165644,51} \frac{88 - 2 - 1}{2} = 11,239.$$

El valor crítico al 5% de significación de la $F_{2,85}$ es aproximadamente $c \approx 3,10$, por lo que la F es muy significativa y rechazamos claramente la hipótesis conjunta de que $\beta_1 = 1$ y $\beta_0 = 0$.

- c) Ahora contrasta $H_0 : \beta_2 = 0, \beta_3 = 0, \beta_4 = 0$ en el modelo

$$\text{price} = \beta_0 + \beta_1 \text{assess} + \beta_2 \text{sqrft} + \beta_3 \text{lotsize} + \beta_4 \text{bdrms} + u.$$

El R^2 de la estimación de este modelo usando las mismas 88 viviendas es 0,829.

Como la variable dependiente es la misma porque sólo contrastamos restricciones de exclusión podemos usar la forma con R^2 de la F ,

$$F = \frac{R_{nr}^2 - R_r^2}{1 - R_{nr}^2} \frac{n - k - 1}{q} = \frac{0,829 - 0,820}{1 - 0,829} \frac{88 - 4 - 1}{3} = 1,4561,$$

que no es significativo ($c \approx 2,71$ para la $F_{3,83}$ al 5%), por lo que no rechazamos H_0 y preferimos el modelo corto.

- d) Si la varianza de price cambia con assess , sqrft , lotsize , o bdrms , ¿qué puedes decir sobre el contraste de la F realizado en (c)?

No es válido, habría que emplear otra formulación del estadístico F .

4. Considera el modelo de regresión múltiple con tres variables independientes, bajo los supuestos clásicos del modelo de regresión lineal, LRM.1-6,

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

Se pide calcular la hipótesis nula $H_0 : \beta_1 - 2\beta_2 = 1$.

- a) Sean $\hat{\beta}_1$ y $\hat{\beta}_2$ los estimadores MCO de β_1 y β_2 . Encuentra $\text{Var}(\hat{\beta}_1 - 3\hat{\beta}_2)$ en función de las varianzas de $\hat{\beta}_1$ y $\hat{\beta}_2$ y de la covarianza entre ellos. ¿Cuál es el error estándar de $\hat{\beta}_1 - 3\hat{\beta}_2$?

$$\begin{aligned} \text{Var}(\hat{\beta}_1 - 3\hat{\beta}_2) &= \text{Var}(\hat{\beta}_1) + \text{Var}(-3\hat{\beta}_2) + 2\text{Cov}(\hat{\beta}_1, -3\hat{\beta}_2) \\ &= \text{Var}(\hat{\beta}_1) + 9\text{Var}(\hat{\beta}_2) - 6\text{Cov}(\hat{\beta}_1, \hat{\beta}_2) \end{aligned}$$

$$\begin{aligned} \text{se}(\hat{\beta}_1 - 3\hat{\beta}_2) &= \left(\widehat{\text{Var}}(\hat{\beta}_1) + 9\widehat{\text{Var}}(\hat{\beta}_2) - 6\widehat{\text{Cov}}(\hat{\beta}_1, \hat{\beta}_2) \right)^{1/2} \\ &= \left(\text{se}(\hat{\beta}_1)^2 + 9\text{se}(\hat{\beta}_2)^2 - 6\widehat{\text{Cov}}(\hat{\beta}_1, \hat{\beta}_2) \right)^{1/2}. \end{aligned}$$

b) ¿Cuál es el estadístico t para contrastar $H_0 : \beta_1 - 3\beta_2 = 0$?

$$t = \frac{\hat{\beta}_1 - 3\hat{\beta}_2}{se(\hat{\beta}_1 - 3\hat{\beta}_2)}$$

c) Define $\theta = \beta_1 - 3\beta_2$ y $\hat{\theta} = \hat{\beta}_1 - 3\hat{\beta}_2$. Escribe un modelo de regresión que incluya $\beta_0, \theta, \beta_2, \beta_3$ que te permita obtener directamente $\hat{\theta}_1$ y su error estándar.

Hay que eliminar β_1 . Usando $\beta_1 = \theta + 3\beta_2$,

$$y = \beta_0 + (\theta + 3\beta_2)x_1 + \beta_2x_2 + \beta_3x_3 + u.$$

y por tanto

$$y = \beta_0 + \theta x_1 + \beta_2(x_2 + 3x_1) + \beta_3x_3 + u.$$

5. Ver Ejercicio 3.7. Ahora usamos el logaritmo de los precios de la vivienda como variable dependiente

$$\log(\text{price}) = \beta_0 + \beta_1 \text{sqrft} + \beta_2 \text{bdrms} + u.$$

a) Estamos interesados en estimar y obtener un intervalo de confianza para el cambio porcentual en price cuando se añade un dormitorio de 150 pies cuadrados. En forma decimal esto es $\theta_1 = 150\beta_1 + \beta_2$. Usa los datos en *HPRICE1* para estimar θ_1 .

Estimando la ecuación anterior,

$$\begin{aligned} \widehat{\log(\text{price})} &= 4,76603 + 0,000379446 \text{sqrft} + 0,0288844 \text{bdrms} \\ &\quad (0,097044) \quad (4,3212e-005) \quad (0,029643) \\ n &= 88 \quad \bar{R}^2 = 0,5883 \end{aligned}$$

obtenemos

$$\hat{\theta}_1 = 150\hat{\beta}_1 + \hat{\beta}_2 = 150 * 0,000379446 + 0,0288844 = 0,0858,$$

es decir, la vivienda subiría un 8,58 % aproximadamente.

b) Escribe β_2 en términos de θ_1 y β_1 y sustituye en la ecuación de $\log(\text{price})$.

Tenemos que $\beta_2 = \theta_1 - 150\beta_1$,

$$\begin{aligned} \log(\text{price}) &= \beta_0 + \beta_1 \text{sqrft} + (\theta_1 - 150\beta_1) \text{bdrms} + u \\ &= \beta_0 + \beta_1 (\text{sqrft} - 150 \text{bdrms}) + \theta_1 \text{bdrms} + u \end{aligned}$$

c) Usa (b) para obtener el error estándar de $\hat{\theta}_1$ y usa este error estándar para construir el intervalo de confianza al 95 %.

$$\begin{aligned} \widehat{\log(\text{price})} &= 4,76603 + 0,00037944 (\text{sqrft} - 150 \text{bdrms}) + 0,0858 \text{bdrms} \\ &\quad (0,097044) \quad (4,3212e-005) \quad (0,02677) \\ n &= 88 \quad R^2 = 0,5883 \end{aligned}$$

y el intervalo de confianza es

$$\hat{\theta}_1 \pm c \cdot se(\hat{\theta}_1) \equiv 0,0858 \pm 1,987 \cdot 0,02677 \equiv (0,0326, 0,139).$$

usando el valor crítico de la t_{85} .

6. Usa los datos *MLB1* para este ejercicio sobre salarios de una liga de baseball.

a) Re-estima el modelo obtenido en clase,

$$\begin{aligned}\widehat{\log(\text{salary})} &= 11,20 + 0,0689\text{years} + 0,0126\text{gamsyr} + 0,00098\text{bavg} + 0,0144\text{hrunsyr} + 0,0108\text{rbisy} \\ &\quad (0,289) \quad (0,012) \quad (0,0026) \quad (0,0011) \quad (0,016) \quad (0,0072) \\ n &= 353, SSR = 183,186, R^2 = 0,6278\end{aligned}$$

eliminando la variable *rbisy*. ¿Qué ocurre con la significación estadística de *hrunsyr*? ¿Y con el tamaño del coeficiente de *hrunsyr*?

$$\begin{aligned}\widehat{\log(\text{salary})} &= 11,021 + 0,06773\text{years} + 0,01576\text{gamsyr} + 0,001419\text{bavg} + 0,0359\text{hrunsyr} \\ &\quad (0,26572) \quad (0,012) \quad (0,00157) \quad (0,001066) \quad (0,00724) \\ n &= 353 \quad R^2 = 0,6254\end{aligned}$$

Ahora *hrunsyr* es muy significativa,

$$\beta_{\text{hrunsyr}} = \frac{0,0359}{0,00724} = 4,9586.$$

y su tamaño es ahora más del doble.

b) Añade las variables *runsyr*, *fldperc* y *sbasesyr* al modelo de (a). ¿Cuál de estos factores es significativo individualmente?

Ahora obtenemos

$$\begin{aligned}\widehat{\log(\text{salary})} &= 10,4083 + 0,069985\text{years} + 0,0079\text{gamsyr} + 0,00053\text{bavg} + 0,0232\text{hrunsyr} \\ &\quad (2,0033) \quad (0,012) \quad (0,00268) \quad (0,0011) \quad (0,00864) \\ &\quad + 0,01739\text{runsyr} + 0,001035\text{fldperc} - 0,00642\text{sbasesyr} \\ &\quad (0,00506) \quad (0,002) \quad (0,00518) \\ n &= 353 \quad R^2 = 0,6390\end{aligned}$$

bavg, *fldperc* y *sbasesyr* no son significativos individualmente.

c) En el modelo de (b), contrasta la significación conjunta de *bavg*, *fldperc* y *sbasesyr*.

El modelo restringido estimado es

$$\begin{aligned}\widehat{\log(\text{salary})} &= 11,5309 + 0,06965\text{years} + 0,00865\text{gamsyr} + 0,028\text{hrunsyr} + 0,01405\text{runsyr}, \\ &\quad (0,11798) \quad (0,01193) \quad (0,00259) \quad (0,00746) \quad (0,0039) \\ n &= 353 \quad R^2 = 0,6368\end{aligned}$$

y el estadístico *F* es

$$F = \frac{0,6390 - 0,6368}{1 - 0,6390} \frac{353 - 1 - 7}{3} = 0,70083$$

que no es significativa, no son significativas tampoco conjuntamente.

7. Usa los datos de *WAGE2*.

a) Considera la ecuación salarial estándar,

$$\log(\text{wage}) = \beta_0 + \beta_1\text{educ} + \beta_2\text{exper} + \beta_3\text{tenure} + u.$$

Establece la hipótesis de que otro año de experiencia en la población activa tiene el mismo efecto sobre $\log(\text{wage})$ que otro año de antigüedad en el empleo actual.

$$H_0 : \beta_2 = \beta_3$$

b) *Contrasta la hipótesis nula de (a) en contra de una alternativa bilateral al nivel de significación del 5 % mediante un intervalo de confianza al 95 %. ¿Cuál es tu conclusión?*

Para conseguir el error estándar de $\theta = \beta_2 - \beta_3$ hay que estimar un modelo que incluya θ , sustituyendo $\beta_2 = \theta + \beta_3$

$$\log(\text{wage}) = \beta_0 + \beta_1 \text{educ} + \theta \text{exper} + \beta_3 (\text{tenure} + \text{exper}) + u.$$

8. *Las empresas pueden financiarse mediante la emisión de acciones (fondos propios) o mediante la emisión de bonos (deuda). Una teoría económica afirma que la ratio de deuda sobre fondos propios, $100 * \text{Deuda} / (\text{Fondos Propios})$, dependerá de los tipos impositivos que afronte la empresa. En concreto,*

- (i) *las empresas preferirán utilizar el endeudamiento cuanto mayores sean los impuestos sobre los beneficios de las empresas, ceteris paribus*
- (ii) *cuantos mayores sean los impuestos sobre las ganancias de capital (ceteris paribus), mayor será el endeudamiento de las empresas*
- (iii) *cuantos mayores sean los impuestos de la renta (ceteris paribus), menor será el endeudamiento de las empresas*

Se dispone de datos para 51 estados americanos con información sobre los tipos impositivos de los impuestos sobre el beneficio empresarial, $tprof$, sobre las ganancias de capital, $tcap$, y sobre la renta, $tinc$; además, se conoce la ratio (media) de deuda sobre fondos propios de las empresas ubicadas en cada estado, $debrat$. Todas las variables están medidas en puntos porcentuales. Con la muestra, disponible se han estimado por Mínimos Cuadrados Ordinarios los siguientes modelos:

$$\widehat{debrat}_i = 69,3662 + 1,6223 tprof_i - 1,1214 tinc_i + 1,0777 tcap_i \quad (1)$$

(9,8700) (1,0607) (0,7028) (1,3427)

$n = 51 \quad R^2 = 0,350553 \quad SSR = 7328,5176$

$$(\widehat{debrat}_i - tcap_i) = 81,7726 + 0,9886 (tprof_i - tcap_i) - 0,3085 (tinc_i - tcap_i) \quad (2)$$

(4,8050) (0,9749) (0,4201)

$n = 51 \quad R^2 = 0,063316 \quad SSR = 7649,2574$

$$\widehat{debrat}_i = 69,4269 + 1,5778 tprof_i - 1,1302 (tinc_i - tcap_i) \quad (3)$$

(9,6765) (0,3989) (0,6680)

$n = 51 \quad R^2 = 0,350524 \quad SSR = 7328,8376$

$$\widehat{debrat}_i = 69,3662 + 1,6223 tprof_i - 1,1214 (tinc_i - tcap_i) - 0,0437 tcap_i \quad (4)$$

(9,8700) (1,0607) (0,7028) (0,9651)

$n = 51 \quad R^2 = 0,350553 \quad SSR = 7328,5176$

- a) Fíjese en los resultados del Modelo (1). Si no existieran impuestos, ¿tendrían las empresas en promedio más Deuda que Fondos Propios? ¿Cuál sería su respuesta si un estado fija un tipo impositivo de 15 puntos porcentuales para las ganancias de capital, de 20 puntos porcentuales en el impuesto sobre la renta y de 30 puntos porcentuales en el impuesto sobre los beneficios empresariales?
- b) Contraste con un nivel de significación del 1% la hipótesis de que conjuntamente las cuestiones fiscales son relevantes en las decisiones de endeudamiento de las empresas. ¿Sería la conclusión diferente si hubiera hecho tres contrastes de significatividad individual con el mismo nivel de significación cada uno? ¿Por qué?
- c) Contraste, con un nivel de significación del 1%, la hipótesis de que un aumento simultáneo de todos los impuestos en un punto porcentual provocará un aumento de un punto en la ratio de deuda sobre fondos propios. Ofrezca el valor tanto del estadístico F como el del estadístico t que se pueden utilizar para realizar este contraste.

a. Si no existieran impuestos, $tprof_i = tinc_i = tcap_i = 0$, la ratio de deuda sobre fondos propios estimado sería, en promedio, $\widehat{debrat}_i = 69,3662$; puesto que es menor que 100, habría más fondos propios que deuda. Por otro lado, con los tipos impositivos del enunciado, la ratio estimada sería de $\widehat{debrat}_i = 69,3662 + 1,6223 * 30 - 1,1214 * 20 + 1,0777 * 15 = 111,17$; es decir, habría más deuda que fondos propios.

b. En el modelo $debrat_i = \beta_0 + \beta_1 tprof_i + \beta_2 tinc_i + \beta_3 tcap_i + \varepsilon_i$, la hipótesis de que conjuntamente los impuestos no son relevantes y el correspondiente estadístico de contraste son:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0 \quad \text{vs.} \quad H_1 : H_0 \text{ no es cierta}$$

$$W = qF = \frac{R^2}{(1 - R^2)/n} \underset{H_0}{\overset{a}{\sim}} \chi^2_{(q)}$$

En este caso, $q = 3$ y sólo necesitamos el R^2 del modelo sin restringir porque para el modelo restringido es cero. Como $W = qF = \frac{0,350553}{(1-0,350553)/51} = 27,528$ y el valor crítico de una $\chi^2_{(3)}$ al 1% es $c_{0,01}^* = 11,37$, se rechaza la hipótesis nula al 1% ($W > c_{0,01}^*$). Para los contrastes de significatividad individual de cada parámetro se tendrían los siguientes valores de los estadísticos

$$t_{\widehat{\beta}_1} = \frac{1,6223}{1,0607} = 1,5294$$

$$t_{\widehat{\beta}_2} = \frac{-1,1214}{0,7028} = -1,5955$$

$$t_{\widehat{\beta}_3} = \frac{1,0777}{1,3427} = 0,8026$$

Comparándolos con el valor crítico de la distribución normal $N(0,1)$ al 1% para una alternativa bilateral: $c_{0,01}^* = 2,576$, resulta claro que en todos los casos $|t| < 2,576$; por tanto, no se puede rechazar individualmente que los coeficientes sean cero. Esta diferencia entre el resultado del contraste individual y conjunto sería un indicio de multicolinealidad: los estados con un tipo impositivo alto en la renta también suelen tener altos los tipos del resto de impuestos.

c. Puesto que $\Delta debrat_i = \beta_1 \Delta tprof_i + \beta_2 \Delta tinc_i + \beta_3 \Delta tcap_i$, si todos los impuestos suben en un punto ($\Delta tprof_i = \Delta tinc_i = \Delta tcap_i = 1$) el aumento de la ratio será $\Delta debrat_i = \beta_1 + \beta_2 + \beta_3$ que según

el enunciado debe ser igual a uno

$$H_0 : \beta_1 + \beta_2 + \beta_3 = 1 \quad \text{vs} \quad H_1 : \beta_1 + \beta_2 + \beta_3 \neq 1$$

Imponemos la restricción (por ejemplo, $\beta_3 = 1 - \beta_1 - \beta_2$), se tiene un modelo restringido como el (2). El estadístico de contraste es $W = qF = \frac{SSR_R - SSR_{SR}}{(SSR_{SR})/n} \stackrel{a}{H_0} \chi^2_{(q)}$. En este caso $q = 1$ y el estadístico tiene valor

$$W = F = \frac{7649,2574 - 7328,5176}{7328,5176/51} = 2,2321$$

El valor crítico de una $\chi^2_{(1)}$ al 1% es $c_{0,01}^* = 6,63$. Puesto que $W < c_{0,01}^*$ no evidencia empírica para rechazar la hipótesis nula al 1%. El valor del estadístico t se obtiene sin más cálculos usando que $t^2 = F$; por tanto, $t = \sqrt{2,2321} = 1,494$.