SOLUCIÓN Examen de Introducción a la Econometría

Universidad Carlos III de Madrid

2^a Convocatoria

Curso 2006/2007

Conteste las preguntas siguientes en cuadernillos separados en 2 horas

1. [3,5 puntos] *La variable rdintens son los gastos en investigación y desarrollo (R&D) en porcentaje sobre las ventas. Las ventas, sales, se miden en millones de dólares. La variable profmarg son los beneficios en porcentaje sobre las ventas. Usando 32 empresas del sector químico se ha obtenido la siguiente ecuación*

$$\begin{aligned} \widehat{rdintens} &= .472 + \underset{(0.216)}{.321} \log(sales) + \underset{(0.046)}{\hat{\beta}_2}\ profmarg \\ n &= 32, \quad R^2 = 0.098 \end{aligned} \tag{1}$$

(a) *Interpreta el coeficiente de* $\log(sales)$ *: si las ventas se incrementan en un 10%, ¿cuál es el cambio estimado ceteris paribus de rdintens? ¿Es estadísticamente significativo?*

Si las ventas aumentan en un 10%, entonces $rdintens$ aumenta aproximadamente en $10*(0.321/100) = 0.0321$ unidades, cp.

Significatividad:

$$\begin{aligned} H_0 &: \beta_{\log(sales)} = 0 \\ H_0 &: \beta_{\log(sales)} \neq 0 \end{aligned}$$

usando el contraste de la t,

$$t_{\log(sales)} = \frac{\hat{\beta}_{\log(sales)}}{se\left(\hat{\beta}_{\log(sales)}\right)} = \frac{0.321}{0.216} = 1.4861$$

que no es significativo al 10% comparado con una $N(0,1)$ ni con una t_{29}.

(b) *Se ha estimado este otro modelo usando la misma base de datos:*

$$\begin{aligned} \widehat{rdintens} &= 1.104 + \underset{(0.216)}{.302} \log(sales) \\ n &= 32, \quad R^2 = 0.061 \end{aligned} \tag{2}$$

¿Es el coeficiente de profmarg en el modelo del apartado (a) significativo? Si sabemos que $\hat{\beta}_2$ *es positivo, ¿cuál es su valor?*

Como no disponemos de su valor, sólo de su error estándar, no podemos calcular su estadístico t para contrastar

$$\begin{aligned} H_0 &: \beta_{\text{ profmarg}} = 0 \\ H_0 &: \beta_{\text{ profmarg}} \neq 0, \end{aligned}$$

pero sí podemos hacer un test de la F,

$$F = \frac{R^2_{nr} - R^2_r}{1 - R^2_{nr}} \frac{n-k-1}{q} = \frac{0.098 - 0.061}{1 - 0.098} \frac{32-2-1}{1} = 1.1896$$

comprobando que no es significativo a los niveles habituales al comparar con una $F_{1,29}$ o una χ^2_1. Usando que

$$t^2 = \left(\frac{\beta_{\text{ profmarg}}}{se\left(\hat{\beta}_{\text{ profmarg}}\right)} \right)^2 = \left(\frac{\beta_{\text{ profmarg}}}{0.046} \right)^2 = F = 1.1896$$

tenemos que

$$\beta_{\text{ profmarg}} = 0.046\,(1.1896)^{1/2} = 0.05017.$$

(c) *Considera este modelo que relaciona los beneficios con las ventas,*

$$\begin{aligned}\widehat{profmarg} &= 7.34 + \underset{se(\hat{\gamma}_1)}{\hat{\gamma}_1} \log(sales). \\ n &= 32, \quad R^2 = 0.0069\end{aligned}$$

¿Cuál es el signo y el valor de $\hat{\gamma}_1$? ¿Es significativo? Obtiene $se(\hat{\gamma}_1)$.
Tenemos que la relación entre los coeficientes estimados de estos dos modelos,

$$\begin{aligned}rdintens &= \beta_0 + \beta_1 \log(sales) + \beta_2 profmarg + u \\ rdintens &= \delta_0 + \delta_1 \log(sales) + v\end{aligned}$$

es

$$\hat{\delta}_1 = \hat{\beta}_1 + \hat{\beta}_2 \hat{\gamma}_1$$

por lo que, como $\hat{\delta}_1 < \hat{\beta}_1$ y $\hat{\beta}_2 > 0$, tenemos que $\hat{\gamma}_1 < 0$ y por tanto

$$\hat{\gamma}_1 = \frac{\hat{\delta}_1 - \hat{\beta}_1}{\hat{\beta}_2} = \frac{0.302 - 0.321}{0.05} = -0.38$$

y para calcular su error estándar tenemos que $F = t^2$ en el contraste de significatividad global del tercer modelo,

$$H_0 : \gamma_1 = 0,$$

por lo que si

$$F = \frac{R^2}{1 - R^2} \frac{n-k-1}{k} = \frac{0.0069}{1 - 0.0069} \frac{32 - 1 - 1}{1} = 0.208\,44,$$

tenemos que

$$se(\hat{\gamma}_1) = \left(\frac{\hat{\gamma}_1^2}{F}\right)^{1/2} = \left(\frac{(-0.38)^2}{0.208\,44}\right)^{1/2} = 0.832\,33.$$

2. [3 puntos] *Para estudiar el efecto de la asistencia a clase sobre las notas finales de una asignatura, se han utilizando datos de 680 alumnos de Introducción a Microeconomía en una universidad americana para ajustar el siguiente modelo de regresión:*

$$\begin{aligned}stndfnl = {} & \beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT \\ & + \beta_4 priGPA^2 + \beta_5 ACT^2 + \beta_6 priGPA \cdot atndrte + U,\end{aligned} \tag{3}$$

los resultados de la estimación son los siguientes

$$\widehat{stndfnl} = \underset{(1,36)}{2,05} - \underset{(0,0102)}{0,0067} atndrte - \underset{(0,48)}{1,63} priGPA - \underset{(0,098)}{0,128} ACT$$
$$+ \underset{(0,101)}{0,296} priGPA^2 + \underset{(0,0022)}{0,0045} ACT^2 + \underset{(0,0022)}{0,0045} priGPA \cdot atndrte, \; R^2 = 0,222,$$

donde $stndfnl$ es el resultado de un examen final estandarizado, $atndrte$ es el porcentaje de asistencia a clase, $priGPA$ es la nota media obtenida en los cursos anteriores y ACT es la nota de acceso a la Universidad.

(a) *¿Cuál es el efecto parcial estimado de ACT sobre el examen final? ¿Cuál es el punto crítico a partir del cuál cambia el signo del efecto de ACT sobre $stndfnl$?*
El efecto parcial estimado es $\hat{\beta}_3 + 2\hat{\beta}_5 ACT$. El punto crítico para el que ese efecto es cero es

$$ACT^* = -\frac{\hat{\beta}_3}{2\hat{\beta}_5} = -\frac{0,128}{2 * 0,0045} = 14,22.$$

(b) *Si añadimos el término $\beta_7 ACT * atndrte$ a la ecuación (3), ¿Cuál será el efecto parcial de la atención a clase en el modelo en términos de los parámetros desconocidos?*
Será $\beta_1 + \beta_6 priGPA + \beta_7 ACT$.

(c) *Explique como contrastaría que el modelo es lineal en todas las variables. Esto es, que $\beta_4 = \beta_5 = \beta_6 = 0$. Exprese con claridad todos los pasos que ha de seguir para realizar el contraste: 1) Estadístico a utilizar, explicando sus componentes con claridad y 2) Regla de decisión.*
1) Al ajuste del modelo (3) habría que añadir el ajuste del modelo restringido,

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT + V$$

que impone la restricción bajo la nula,

$$H_0 : \beta_4 = \beta_5 = \beta_6 = 0,$$

obteniéndose la Suma de Cuadrados de Residuos restringida, SCE_r. El estadístico a utilizar sería el F correspondiente,

$$\begin{aligned} F &= \frac{SCE_r - SCE_{nr}}{SCR_{nr}} \frac{n-k-1}{q} = \frac{R^2_{nr} - R^2_r}{1 - R^2_{nr}} \frac{n-k-1}{q} \\ &= \frac{SCE_r - SCE_{nr}}{SCR_{nr}} \frac{680-6-1}{3} = \frac{R^2_{nr} - R^2_r}{1 - R^2_{nr}} \frac{680-6-1}{3} \end{aligned}$$

donde SCE_{nr} se obtiene del ajuste modelo (1), $k = 6$ es el número de regresores y $q = 3$ es el número de restricciones en H_0. En este caso la variable dependiente no cambia y se puede utilizar la versión en términos de R^2.
2) Regla de decisión. El valor de F se ha de comparar con el valor crítico correspondiente de una distribución $F_{3,674} \approx F_{3,\infty} = \chi_3/3$, que para el 5% es igual a 3,60.

3. [3,5 puntos] *Considera los siguientes modelos para explicar el peso de un recién nacido, bwght,*

$$\begin{aligned} \widehat{\log(bwght)} &= \underset{(.019)}{4.69} - \underset{(.00085)}{0.0042} cigs + \underset{(.0059)}{0.0084} \log(faminc) + \underset{(.01)}{0.026} male + \underset{(.014)}{0.053} white \\ R^2 &= 0.0416, \quad n = 1388 \\ \widehat{\log(bwght)} &= 4.687 - 0.0042 cigs + 0.0083 \log(faminc) + 0.028 male + 0.054 white \\ &\quad -0.002 white * male \\ R^2 &= 0.0417, \quad n = 1388 \\ \widehat{\log(bwght)} &= 4.689 - 0.0042 cigs + 0.0077 \log(faminc) + 0.028 male * nowhite + 0.0677 white \\ R^2 &= 0.0381, \quad n = 1388 \end{aligned}$$

donde cigs es el número medio de cigarrillos fumados al día por la madre, faminc es la renta familiar, male es una variable ficticia que indica si el recién nacido es niño (male = 1) o niña (male = 0), white es otra variable ficticia que indica si es blanco (= 1) o no (= 0) y nowhite = 1 − white.

(a) *En la primera ecuación, interpreta el coeficiente de la variable cigs. Proporciona un intervalo de confianza al 95% para el efecto sobre el peso del recién nacido de fumar 10 cigarrillos más, todos los demás factores constantes.*
El coeficiente de *cigs* indica que si se aumenta el consumo de tabaco en un cigarrillo, el peso del recién nacido disminuye en media un $100 * 0.0042\% = 0.42\%$ aproximadamente.
El intervalo de confianza para β_1 será, usando la aproximación por la normal,

$$\hat{\beta}_1 \pm 1.96 se\left(\hat{\beta}_1\right)$$

es decir

$$-0.0042 \pm 1.96 * 0.00085 \to \left(-5.866 \times 10^{-3}, -2.534 \times 10^{-3}\right)$$

y por tanto el IC al 95% para el efecto de aumentar el consumo en 10 cigarrillos será

$$(-5.866\%, -2.534\%).$$

(b) *Considera ahora las dos primeras ecuaciones ¿Cuánto más predice cada modelo que pesará un niño (male = 1) recién nacido blanco que otro no blanco, manteniendo los demás factores constantes? ¿Es la diferencia entre las dos predicciones significativa?*

Primer modelo:

$$\Delta \widehat{\log(bwght)} = \hat{\beta}_{white} = 0.053.$$

Segundo modelo:

$$\Delta \widehat{\log(bwght)} = \hat{\beta}_{white} + \hat{\beta}_{whie*male} = 0.054 - 0.002 = 0.052.$$

Para contrastar si la diferencia entre ambas predicciones es significativa debemos realizar el contraste

$$\begin{aligned} H_0 &: \beta_{white*male} = 0 \\ H_0 &: \beta_{white*male} \neq 0. \end{aligned}$$

Como no disponemos del s.e. del estimador de $\beta_{white*male}$, debemos usar un contraste de la F que compare ambos modelos,

$$F = \frac{R^2_{nr} - R^2_r}{1 - R^2_{nr}} \frac{n-k-1}{q} = \frac{0.0417 - 0.0416}{1 - 0.0417} \frac{1388 - 5 - 1}{1} = 0.144$$

que claramente no es significativo comparado con una $F_{1,1382}$ o una χ^2_1.

(c) *Usando el segundo modelo, estima la diferencia de peso entre una niña y un niño al nacer, ambos blancos, todos los demás factores igual. ¿Es esa diferencia significativa?*

La diferencia estimada entre una niña y un niño blancos es

$$-\hat{\beta}_{male} - \hat{\beta}_{while*male} = -0.028 - (-0.002) = -0.026,$$

las niñas pesan un 2.6% menos en media que los niños. Para ver si es significativa la diferencia,

$$H_0 : \beta_{male} + \beta_{while*male} = 0$$

vemos que sustituyendo $\beta_{while*male} = -\beta_{male}$, es decir, $\beta_5 = -\beta_3$ en el segundo modelo obtenemos que

$$\begin{aligned} \widehat{\log(bwght)} &= \beta_0 + \beta_1 cigs + \beta_2 \log(faminc) + \beta_3 male + \beta_4 white + \beta_5 while * male + u \\ &= \beta_0 + \beta_1 cigs + \beta_2 \log(faminc) + \beta_3 male + \beta_4 white - \beta_3 while * male + u \\ &= \beta_0 + \beta_1 cigs + \beta_2 \log(faminc) + \beta_3 (male - male * white) + \beta_4 white + u \\ &= \beta_0 + \beta_1 cigs + \beta_2 \log(faminc) + \beta_3 male \underbrace{(1 - white)}_{=nowhite} + \beta_4 white + u \end{aligned}$$

por lo que el tercer modelo es la correspondiente versión restringida del segundo bajo H_0. Planteando el contraste de la F,

$$F = \frac{R^2_{nr} - R^2_r}{1 - R^2_{nr}} \frac{n-k-1}{q} = \frac{0.0417 - 0.0381}{1 - 0.0417} \frac{1388 - 5 - 1}{1} = 5.192$$

que es significativo al 1% comparado con el valor crítico de la $F_{1,\infty}$ o de la χ^2_1, por lo que la diferencia entre niños y niñas es significativa.