

Hoja de Ejercicios 4

Análisis de regresión con información cualitativa

Nota: En aquellos ejercicios en los que se incluyen estimaciones y referencia al archivo de datos utilizado, el estudiante debería comprobar los resultados obtenidos en Gretl.

1. Para explicar el salario de un director general, *salary*, se ha estimado la siguiente ecuación con los datos del archivo `CEOSAL1.GDT`:

$$\begin{aligned} \log(\widehat{salary}) &= 4,362 + 0,275 \log(sales) + 0,0179roe \\ &\quad (0,294) \quad (0,033) \quad (0,0039) \\ n &= 209, \quad R^2 = 0,282 \end{aligned}$$

donde *sales* son las ventas anuales y *roe* es el rendimiento sobre el valor nominal de una acción (return-on-equity).

- a) Interprete el coeficiente de $\log(sales)$ y contraste si es significativamente positivo.
 b) Se decide incluir una variable dummy, *rosneg*, que es igual a 1 cuando *ros* es negativa y 0 si *ros* es cero o positivo, donde *ros* es el rendimiento sobre el valor real de la acción,

$$\begin{aligned} \log(\widehat{salary}) &= 4,074 + 0,314 \log(sales) + 0,017roe \\ &\quad (0,307) \quad (0,035) \quad (0,004) \\ &\quad + 2,094rosneg - 0,258 \log(sales) * rosneg - 0,00353roe * rosneg \\ &\quad (1,009) \quad (0,112) \quad (0,178) \\ n &= 209, \quad R^2 = 0,315, \end{aligned}$$

obteniéndose las siguientes varianzas estimadas para los coeficientes de $\log(sales)$, *roe*, *rosneg*, $\log(sales) * rosneg$ y *roe * rosneg*, respectivamente,

$$\widehat{V} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \\ \hat{\beta}_5 \end{pmatrix} = \begin{pmatrix} 0,001236 & 1,47E-05 & 0,010414 & -0,001236 & -1,47E-05 \\ 1,47E-05 & 1,64E-05 & 0,000410 & -1,47E-05 & -1,64E-05 \\ 0,010414 & 0,000410 & 1,018975 & -0,109247 & -0,003193 \\ -0,001236 & -1,47E-05 & -0,109247 & 0,012544 & -0,000116 \\ -1,47E-05 & -1,64E-05 & -0,003193 & -0,000116 & 0,000318 \end{pmatrix}.$$

Contraste si es necesario distinguir en el modelo las empresas en función del signo de *ros*.

- c) Contraste si, para las empresas que tienen *ros* negativo, un aumento de las ventas implica necesariamente un aumento del salario del director general, manteniéndose los demás factores constantes.
 d) Explique cómo contrastaría la hipótesis de que para un director general de una empresa con *ros* negativo, $\log(sales) = 10$ y *roe* = 20, es igual de interesante (a) aumentar las ventas en un 1%, o (b) conseguir que el *ros* pase a ser positivo (en ambos casos, sin cambiar las otras variables).

2. Considere los siguientes modelos para explicar el peso de un recién nacido, $bwght$, con los datos para EE.UU. del archivo **BWGHT.GDT**, donde $bwght$ es el peso (en onzas) del niño al nacer, $cigs$ es el número diario de cigarrillos fumados por la madre durante el embarazo, $faminc$ es la renta familiar anual, en miles de dólares, $male$ es una variable ficticia que indica si el recién nacido es niño ($male = 1$) o niña ($male = 0$), y $white$ es otra variable ficticia que indica si es blanco ($= 1$) o no ($= 0$) y $nowhite = 1 - white$.

$$\log(\widehat{bwght}) = 4,69 - 0,0042cigs + 0,0084 \log(faminc) + 0,026male + 0,053white$$

(,019) (,00085) (,0059) (,01) (,014)

$$R^2 = 0,0416, \quad n = 1388$$

$$\log(\widehat{bwght}) = 4,687 - 0,0042cigs + 0,0083 \log(faminc) + 0,028male + 0,054white - 0,002white * male$$

$$R^2 = 0,0417, \quad n = 1388$$

$$\log(\widehat{bwght}) = 4,689 - 0,0042cigs + 0,0077 \log(faminc) + 0,028male * nowhite + 0,0677white$$

$$R^2 = 0,0381, \quad n = 1388$$

- a) En la primera ecuación, interprete el coeficiente de la variable $cigs$. Proporcione un intervalo de confianza al 95 % para el efecto sobre el peso del recién nacido de fumar 10 cigarrillos más, manteniendo todos los demás factores constantes.
- b) Considere ahora las dos primeras ecuaciones ¿En cuánto predice cada modelo el peso adicional de un niño ($male = 1$) recién nacido blanco respecto a otro que no sea blanco, manteniendo los demás factores constantes? ¿Es la diferencia entre las dos predicciones significativa?
- c) Usando el segundo modelo, estime la diferencia de peso entre una niña y un niño al nacer, ambos blancos, manteniéndose constantes todos los restantes factores. ¿Es dicha diferencia significativa?

3. Las siguientes ecuaciones de salarios han sido estimadas utilizando datos correspondientes a trabajadores en Bangladesh:

$$\log(\widehat{salarario}) = 1,25 + 0,15hombre + 0,02 experiencia, \quad (1)$$

(0,35) (0,03) (0,004)

$$\log(\widehat{salarario}) = 1,55 + 0,10hombre + 0,015experiencia - 0,005hombre*experiencia, \quad (2)$$

(0,48) (0,05) (0,005) (0,002)

donde $salarario$ está medido en dólares de EEUU, y $hombre$ es una variable binaria que toma el valor 1 si el trabajador es hombre y 0 si es mujer; $experiencia$ mide los años de experiencia laboral. Los números entre paréntesis son los errores estándar.

- a) ¿Cuál es la diferencia media estimada entre el salario de un hombre con 5 años de experiencia y el de una mujer con 10 años de experiencia utilizando la ecuación (1)?
- b) ¿Cuál es la diferencia media estimada entre el salario de un hombre con 5 años de experiencia y el de una mujer con 10 años de experiencia utilizando la ecuación (2)?
- c) Contraste que la diferencia salarial entre hombres y mujeres no depende de la experiencia.

4. Suponga que se reúne información sobre salarios, educación, experiencia y sexo a partir de una encuesta. Además, se pregunta sobre el consumo de marihuana. La pregunta se formula así: "¿En cuántas ocasiones fumaste marihuana el mes pasado?"
- Escriba una ecuación que permita estimar los efectos de su consumo en el salario, tomando en cuenta los efectos de otros factores. El objetivo es poder realizar afirmaciones del tipo "si consume cinco veces más marihuana al mes se prevé un cambio en el salario de x %".
 - Especifique un modelo que permita contrastar si el consumo de drogas tiene distintos efectos en los salarios de hombres y mujeres. ¿Cómo se contrastaría que no existen diferencias entre hombres y mujeres?
 - Suponga que se considera preferible medir el consumo de marihuana clasificando a la gente en cuatro categorías: no consumidor, consumidor ocasional (de una a cinco veces al mes), consumo moderado (de seis a diez) y consumidor habitual (más de diez veces al mes). Escriba un modelo que permita estimar los efectos de esta droga sobre el salario.
 - Usando el modelo propuesto en el apartado (c), explique con detalle cómo contrastaría la hipótesis nula de que el consumo de marihuana no afecta al salario. La respuesta debe ser muy específica e incluir una lista detallada de los grados de libertad.
 - ¿Cuáles son los problemas potenciales para realizar inferencia causal con estos datos de encuesta?
5. Supongamos que estamos interesados en analizar las posibles diferencias en el consumo de cerveza según el sexo. Para ello especificamos el modelo de regresión lineal

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_i + \beta_3 (D_i X_i) + u_i$$

donde Y_i es el gasto en cerveza del individuo i , X_i es su renta y D_i es una variable artificial que vale 1 si el individuo es mujer y 0 si es hombre. En base a una muestra de tamaño $n = 34$ se ha obtenido el siguiente resultado:

$$\hat{Y}_i = 186.47 - 126.00D_i + 2.33X_i - 1.29(D_i X_i) \quad R^2 = 0.5055$$

(45.67) (57.01) (0.86) (1.02)

Los números entre paréntesis son errores estándar. Además, utilizando la misma muestra se ha estimado el modelo $Y_i = \alpha_0 + \alpha_1 X_i + u_i$, obteniéndose un coeficiente de determinación de 0.1355.

- ¿Cuál será la diferencia en el consumo de cerveza entre un hombre y una mujer con el mismo nivel de renta?.
- Contraste al 5% las siguientes afirmaciones
 - No existen diferencias en el consumo de cerveza según el sexo.
 - No existen diferencias en la propensión marginal al consumo de cerveza según el sexo.