| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $\frac{\phantom{x}}{6}$ | + | $\frac{\phantom{x}}{4}$ | + | $\frac{\phantom{x}}{5}$ | = | $\frac{\phantom{x}}{15}$ | | CALIFICACION: |

VACACION.XLS

Este fichero está adaptado del fichero `vacation.gdt` del libro de Hill et al.(2008) que se puede descargar en:

http://gretl.sourceforge.net/gretl_data.html

**PARTE 1** (6 puntos)

Se ha tomado una muestra de 200 familias de Chicago con el fin de investigar los hábitos de vacaciones de la población de EE.UU., en particular, si se alejan mucho de sus lugares de origen. Se ha especificado el siguiente modelo de regresión lineal:

$$(1) \qquad Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \qquad i = 1, 2, \ldots, 200$$

donde $Y_i$ es la distancia recorrida al año en millas, $X_{1i}$ es la renta anual medida en miles de dólares y $X_{2i}$ es la edad media del cabeza de familia.

1. Estima el modelo por MCO y escribe la recta de regresión muestral y la matriz de covarianzas estimada de los estimadores MCO.

$$\hat{Y}_i = \underset{(173,01)}{-365,60} + \underset{(1,84)}{14,29}\, X_{1i} + \underset{(3,60)}{11,86}\, X_{2i} \quad SCR = 41960107 \quad SCT = 28952473,8$$

`Matriz de covarianzas de los estimadores`

| | C | Renta | Edad |
|---|---|---|---|
| C | 29931,25 | -148,816 | -453,503 |
| Renta | | 3,373 | -1,566 |
| Edad | | | 12,972 |

2. Interpreta los coeficientes estimados del modelo (1). ¿Presentan los signos esperados?

   $\hat{\beta}_0 = -365,60:$ cuando las variables explicativas renta y edad toman el valor cero, la distancia estimada es de -365,60 millas.

   $\hat{\beta}_1 = 14,30:$ al aumentar mil dólares la renta anual de la familia, el incremento estimado de la distancia recorrida es de 14,30 millas, manteniéndose fija la edad.

   Su signo positivo es el esperado ya que al aumentar la renta es de esperar que se puedan realizar vacaciones de más distancia que, en principio, resultarían más caras.

   $\hat{\beta}_2 = 11,86:$ al aumentar un año la edad del cabeza de familia, se estima que la distancia recorrida aumenta en 11,86 millas, manteniéndose fija la renta.

   Era de esperar que fuera positivo, con un aumento de la edad se puede tender a visitar lugares más lejanos antes no visitados.

3. Da una medida de la bondad de ajuste e interpreta el resultado.

   Una medida de la bondad de ajuste es el coeficiente de determinación. Para el modelo (1) estimado, se tiene que

$$R^2 = 1 - \frac{SCR}{SCT} = 0,31$$

   lo que indica que el 31 % de la variabilidad muestral de la variable distancia recorrida está explicada por el modelo, es decir, por las variaciones de las variables explicativas renta y edad en términos lineales.

4. ¿Son las variables explicativas conjuntamente significativas? ($\alpha = 5\,\%$).

$$H_0: \quad \beta_1 = \beta_2 = 0$$
$$H_a: \quad \beta_1 \neq 0 \text{ y / o } \beta_2 \neq 0$$

$$\text{F} = \frac{R^2/q}{(1-R^2)/(T-(k+1))} \sim F(q, T-(k+1))$$

$$\text{F} = \frac{0,31/2}{(1-0,31)/(200-3)} = 44,25 \quad > \quad F_{0,05}(2,197) = 3,00$$

Se rechaza la hipótesis nula con un nivel de significación del 5 %. Por lo tanto, las dos variables explicativas, edad y renta, son conjuntamente significativas.

5. Contrasta a un nivel de significación del 5 % la hipótesis de que la renta hace aumentar la distancia recorrida por una familia en vacaciones.

$$H_0: \quad \beta_1 = 0$$
$$H_a: \quad \beta_1 > 0$$

$$\text{t} = \frac{\hat{\beta}_1 - \beta_1^0}{\hat{\sigma}_{\hat{\beta}_1}} \sim t(T-(k+1))$$

$$\text{t} = \frac{14,30 - 0}{1,84} = 7,78 \quad > \quad t_{0,05}(200-3) = 1,645$$

Se rechaza la hipótesis nula con un nivel de significación del 5 %. Por lo tanto, la renta hace aumentar la distancia recorrida por una familia en vacaciones.

6. ¿Te parece aceptable la idea de que una familia con una renta anual de 45000 dólares y cuyo cabeza de familia tiene 40 años recorra al año 700 millas?

Si 700 millas fuera la distancia esperada o promedio para una familia con 45 mil dólares de renta anual y un cabeza de familia de 40 años, entonces debería responder al comportamiento promedio de la variable distancia en función de la renta y de la edad recogido en el modelo (1):

$$E(Y_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} \qquad \longrightarrow \qquad 700 = \beta_0 + \beta_1\, 45 + \beta_2\, 40$$

Como no se conocen los verdaderos valores de los coeficientes y solo se dispone de una estimación de los mismos, es preciso realizar un contraste sobre la plausibilidad de la hipótesis. El contraste que hay que hacer es el siguiente:

$$H_0: \quad 700 = \beta_0 + \beta_1\, 45 + \beta_2\, 40 \quad \rightarrow \quad \beta_0 + \beta_1\, 45 + \beta_2\, 40 - 700 = 0$$
$$H_a: \quad 700 \neq \beta_0 + \beta_1\, 45 + \beta_2\, 40 \quad \rightarrow \quad \beta_0 + \beta_1\, 45 + \beta_2\, 40 - 700 \neq 0$$

La hipótesis nula impone 1 sola restricción sobre los coeficientes del modelo, por lo el estadístico de contraste es:

$$\text{t} = \frac{\hat{\beta}_0 + \hat{\beta}_1\, 45 + \hat{\beta}_2\, 40 - 700 - 0}{\hat{\sigma}_{(\hat{\beta}_0 + \hat{\beta}_1\, 45 + \hat{\beta}_2\, 40 - 700)}} \sim t(T-(k+1))$$

$$\begin{aligned}
V(\hat{\beta}_0 + \hat{\beta}_1\, 45 + \hat{\beta}_2\, 40 - 700) &= V(\hat{\beta}_0) + 45^2\, V(\hat{\beta}_1) + 40^2\, V(\hat{\beta}_2) + 2\; 45\, cov(\hat{\beta}_0, \hat{\beta}_1) + \\
&+ 2\; 40\, cov(\hat{\beta}_0, \hat{\beta}_2) + 2\; 45\, 40\, cov(\hat{\beta}_1, \hat{\beta}_2) = 29931,25 + 2052(3,37) + \\
&+ 1600(12,97) + 90(-148,82) + 80(-453,50) + 3600(-1,57) = \\
&= 29931,25 + 6915,24 + 20752 - 13393,8 - 36280 - 5652 = \\
&= 2272,69
\end{aligned}$$

$$t = \frac{51,97 - 0}{\sqrt{2272,69}} = \frac{51,97}{47,67} = 1,09 \quad < \quad t_{0,025}(200-3) = 1,96$$

No se rechaza la hipótesis nula con un nivel de significación del 5 %. Por lo tanto, no se rechaza la idea de que una familia con una renta anual de 45000 dólares y cuyo cabeza de familia tiene 40 años recorra al año 700 millas.

*Nota: este apartado se puede resolver también construyendo un intervalo de predicción.*

**PARTE 2** (4 puntos)

Otro analista cree que la distancia recorrida puede depender de si la familia tiene hijos menores de 16 años o no.

1. Explica detalladamente cómo incluirías esta variable $X_3$, Tener hijos menores de 16 años, en el modelo (1). Especifica el modelo adecuado en este caso.

   La variable "Tener hijos menores de 16 años" es un factor cualitativo con dos opciones SI o NO. Para introducirlo en el modelo hay que utilizar variables ficticias:

   $$\text{Hijos}_i = \begin{cases} 1 & \text{si i tiene hijos menores} \\ 0 & \text{en otro caso} \end{cases} \qquad \text{NoHijos}_i = \begin{cases} 1 & \text{si i no tiene hijos menores} \\ 0 & \text{en otro caso} \end{cases}$$

   Existen diferentes alternativas para introducir en el modelo (1) la variable "Tener hijos menores de 16 años" a través de las variables ficticias así definidas. Eligiendo incluir una la variable ficticia $\text{Hijos}_i$ y el término independiente, el modelo queda como sigue:

   $$(2) \quad Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 \text{Hijos}_i + u_i \qquad i = 1, 2, \ldots, 200$$

   $$\begin{aligned} E[Y_i|\text{con hijos}] &= \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 \\ E[Y_i|\text{sin hijos}] &= \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} \end{aligned}$$

   Así, el coeficiente $\beta_3$ recoge la diferencia esperada en la distancia recorrida por una familia con hijos respecto a una familia sin hijos, manteniendo fijas las variables edad y renta.

2. A partir de los datos sobre el número de hijos menores de 16 años de las familias del fichero `Vacacion.xls` construye la variable definida en el apartado anterior. Estima por MCO un modelo de regresión con las tres variables explicativas, renta, edad, tener hijos menores de 16 años y escribe la renta de regresión muestral.

   $$\hat{Y}_i = \underset{(167,3)}{-245,87} + \underset{(1,76)}{13,43}\, X_{1i} + \underset{(3,52)}{15,25}\, X_{2i} - \underset{(68,50)}{307,42}\, \text{Hijos}_i \qquad i = 1, 2, \ldots, 200$$

3. ¿Qué significa -307,42?

   Dada una renta y una edad del cabeza de familia, se estima que una familia con hijos menores de 16 años recorre en vacaciones 307,42 millas menos que una familia sin hijos menores.

4. ¿Es significativa la variable "Tener hijos menores de 16 años" ($\alpha = 5\,\%$)?

   La variable "Tener hijos menores de 16 años" no será significativa cuando la diferencia en la distancia recorrida en promedio por familias con hijos y familias sin hijos sea nula, tomadas como fijas las variables renta y edad, es decir,

   $$\begin{aligned} H_0 &: \quad \beta_3 = 0 \\ H_a &: \quad \beta_3 \neq 0 \end{aligned}$$

$$\text{t} = \frac{\hat{\beta}_3 - \beta_3^0}{\hat{\sigma}_{\hat{\beta}_3}} \sim t(T-(k+1))$$

$$\text{t} = \frac{-307,42 - 0}{68,50} = -4,48 \qquad |t| = |-4,48| = 4,48 > t_{0,025}(200-4) = 1,96$$

Se rechaza la hipótesis nula con un nivel de significación del 5 %. Por lo tanto, se rechaza la hipótesis de que tener o no hijos menores de 16 años no influye en la distancia recorrida.

**PARTE 3** (5 puntos)

Un tercer analista opina que el efecto de un incremento de la renta sobre la distancia recorrida depende de si la familia tiene hijos menores de 16 años o no.

1. ¿Cómo modificarías el modelo propuesto en la Parte 2 para que recoja este tipo de comportamiento? Explícalo detalladamente.

   En el modelo de la Parte 2 el efecto de un incremento de la renta sobre la distancia recorrida, ceteris paribus, viene dado por $\beta_1$, es decir, es constante. Esto es debido a que se ha supuesto una relación lineal entre la variable dependiente y la variables explicativa renta. El supuesto de que el efecto de un incremento de la renta sobre la distancia recorrida depende de si la familia tiene hijos menores de 16 años o no implica que la relación entre la variable dependiente distancia y la variable explicativa renta no es lineal. Se ha de especificar el modelo de forma que el efecto de un incremento de la renta sobre la distancia no sea constante, es decir, que sea diferente para las familias con hijos y para las familias sin hijos.

   Este objetivo se puede conseguir introduciendo en el modelo un elemento de interacción entre la variable renta, $X_1$, y la variable "Tener hijos menores de 16 años", $X_3$ $(X_1 \times X_3)$. Como la variable "Tener hijos menores de 16 años" es cualitativa y se cuantifica mediante variables ficticias, para introducir la interacción en el modelo lo haremos también multiplicando la variable cuantitativa renta, $X_1$, con tantas variables ficticias como número de categorías tiene la variable $X_3$ menos 1. El modelo que incluye este efecto queda como sigue:

   $$(3) \quad Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 \text{Hijos}_i + \beta_4 (\text{Hijos} \times X_1)_i + u_i \quad i = 1, 2, \ldots, 200$$

   $$\frac{\partial E[Y_i|\text{con hijos}]}{\partial X_{1i}} = \beta_1 + \beta_4 \qquad \frac{\partial E[Y_i|\text{sin hijos}]}{\partial X_{1i}} = \beta_1$$

   El coeficiente $\beta_4$ recoge la interacción entre las variables renta y "Tener hijos menores de 16 años" y mide la diferencia en el incremento esperado en la distancia por incremento unitario de la renta entre las familias con hijos y las familias sin hijos, tomadas como fija la variable edad. Osea, un incremento unitario de la renta en una familia con hijos supone un incremento de la distancia recorrida de $\beta_4$ unidades más que en una familia sin hijos, ceteris paribus.

2. Estima el modelo por MCO y escribe la recta de regresión muestral.

   $$\hat{Y}_i = \underset{(223,52)}{-632,55} + \underset{(2,86)}{19,26}\, X_{1i} + \underset{(3,47)}{15,37}\, X_{2i} + \underset{(237,41)}{276,2}\ \text{Hijos}_i - \underset{(3,52)}{9,03}\ (\text{Hijos} \times X_1)_i$$

3. Contrasta la hipótesis del tercer analista a un nivel de significación del 5 %.

   La hipótesis del analista de que el efecto de un incremento de la renta sobre la distancia recorrida depende de si la familia tiene hijos menores de 16 años o no, viene recogida en el parámetro $\beta_4$. Si este coeficiente es cero, entonces un incremento de la renta tendría el mismo efecto sobre la distancia recorrida para todas las familias independientemente de si tienen hijos de menores de 16 años o no. Por lo tanto,

$$H_0: \quad \beta_4 = 0$$
$$H_a: \quad \beta_4 \neq 0$$

$$\mathrm{t} = \frac{\hat{\beta}_4 - \beta_4^0}{\hat{\sigma}_{\hat{\beta}_4}} \sim t(T-(k+1))$$

$$\mathrm{t} = \frac{-9,03-0}{3,52} = -2,56 \qquad |t| = |-2,56| = 2,56 > t_{0,025}(200-5) = 1,96$$

   Se rechaza la hipótesis nula con un nivel de significación del 5 %. Por lo tanto, se rechaza la hipótesis de que tener o no hijos menores de 16 años no influye en el incremento esperado en la distancia por incremento unitario de renta.

4. Dada la estimación del modelo, ¿cuál sería el incremento estimado de la distancia recorrida ante incrementos unitarios en la renta para una familia con hijos? ¿Y para una familia sin hijos?.

$$\frac{\partial \hat{E}[Y_i|\text{con hijos}]}{\partial X_{1i}} = \hat{\beta}_1 + \hat{\beta}_4 = 19,25 - 9,03 = 10,22 \quad \text{millas}$$
$$\frac{\partial \hat{E}[Y_i|\text{sin hijos}]}{\partial X_{1i}} = \hat{\beta}_1 = 19,25 \quad \text{millas}$$

5. Contrasta a un nivel de significación del 5 % la hipótesis de que la variable "Tener hijos menores de 16 es significativa" en el modelo que propusiste en el apartado 1.

   La variable "Tener hijos menores de 16 años" no será significativa cuando la diferencia en la distancia recorrida en promedio por familias con hijos y familias sin hijos sea nula y, además, cuando la variable "Tener hijos menores de 16 años" tampoco influya en el efecto de un incremento de la renta sobre la distancia. Por lo tanto:

$$H_0: \quad \beta_3 = \beta_4 = 0$$
$$H_a: \quad \beta_3 \neq 0 \quad \text{y/o} \quad \beta_4 \neq 0$$

   El estadístico de contraste es:

$$\mathrm{F} = \frac{(SCR_R - SCR_{NR})/q}{SCR_{NR}/(T-(k+1))} \sim F(q, T-(k+1))$$

   El modelo restringido es el modelo (1) y el modelo no restringido es el modelo (3).

$$\mathrm{F} = \frac{(41960107 - 36809300)/2}{36809300/195} = \frac{2575403,5}{188765,641} = 13,64 \quad > \quad F_{0,05}(2, 200-5) = 3,00$$

   Se rechaza la hipótesis nula con un nivel de significación del 5 %. Por lo tanto, se rechaza la hipótesis de que la variable "Tener hijos menores de 16 años" no influye en la distancia recorrida.