# 5 Análisis de regresión múltiple con información cualitativa

**Ezequiel Uriel**
**Universidad de Valencia**
**09-2013**



## 5.1 Introducción de información cualitativa en los modelos econométricos

Hasta ahora las variables que hemos utilizado para explicar la variable endógena tenían un carácter cuantitativo. Sin embargo, hay otras variables de carácter cualitativo que pueden ser importantes para explicar el comportamiento de la variable endógena, como el sexo, raza, religión, nacionalidad, región geográfica, etc. Por ejemplo, manteniendo todos los demás factores constantes, se ha constatado que las mujeres trabajadoras tienen unos salarios inferiores que sus homólogos masculinos. Este resultado puede ser consecuencia de la discriminación por género, pero cualquiera que sea la razón, las variables cualitativas como el género parece que influyen en la variable endógena y deberían incluirse en muchos casos entre las variables explicativas. Los factores cualitativos, a menudo pero no siempre, se presentan en forma de información binaria, es decir, una persona es hombre o mujer, está casada o no, etc. Cuando los factores cualitativos se presentan en forma dicotómica la información relevante puede mostrarse como una variable binaria o una variable de cero-uno. En econometría, las variables binarias que se utilizan como regresores son comúnmente llamadas variables ficticias. En la definición de una variable dicotómica, debemos decidir a qué caso se le asigna el valor 1 y a cual se le asigna el valor 0.

En el caso del género podemos definir

$$mujer = \begin{cases} 1 & \text{si la persona es una mujer} \\ 0 & \text{si la persona es un hombre} \end{cases}$$

Pero, por supuesto, también podemos definir

$$hombre = \begin{cases} 1 & \text{si la persona es un hombre} \\ 0 & \text{si la persona es una mujer} \end{cases}$$

Es importante señalar que ambas variables, mujer y hombre, contienen la misma información. Utilizar las variables cero-uno para captar información cualitativa es una decisión arbitraria, pero con esta elección los parámetros tienen una interpretación natural.

## 5.2 Una sola variable ficticia independiente.

Vamos a analizar cómo se puede incorporar la información dicotómica en los modelos de regresión. Considere el siguiente modelo para la determinación del *salario* por hora, en función de los años de educación (*educ*):

$$salario = \beta_1 + \beta_2 educ + u \tag{5-1}$$

Para medir la discriminación salarial debida al género se introduce una variable ficticia (*mujer*) como variable independiente en el modelo definido anteriormente,

$$salario = \beta_1 + \delta_1 mujer + \beta_2 educ + u \tag{5-2}$$

El *atributo* género tiene dos *categorías: mujer* y *hombre*. La categoría *femenina* ha sido incluida en el modelo; mientras que la categoría hombre, que ha sido omitida, es la *categoría de referencia*. El modelo (5-2) se muestra en la figura 5.1, tomando $\delta_1$<0. La interpretación de $\delta_1$ es la siguiente: $\delta_1$ es la diferencia en el salario por hora entre mujeres y hombres, dado el mismo nivel de educación (y el mismo término de perturbación, $u$). Así, el coeficiente $\delta_1$ determina si existe una discriminación contra las mujeres o no. Si $\delta_1$<0, entonces, para el mismo nivel de otros factores (educación, en este caso), las mujeres ganan menos que los hombres en promedio. Suponiendo que la esperanza de la perturbación es cero, si se toman esperanzas en ambas categorías se obtiene:

$$\begin{aligned} \mu_{salario|mujer} &= E(salario \mid mujer = 1, educ) = \beta_1 + \delta_1 + \beta_2 educ \\ \mu_{salario|hombre} &= E(salario \mid mujer = 0, educ) = \beta_1 + \beta_2 educ \end{aligned} \tag{5-3}$$

Como puede verse en (5-3), el término independiente para los hombres es $\beta_1$, y $\beta_1$+$\delta_1$ para las mujeres. Gráficamente, como puede verse en la figura 5.1, hay un desplazamiento del término independiente, pero las líneas para hombres y mujeres son paralelas.

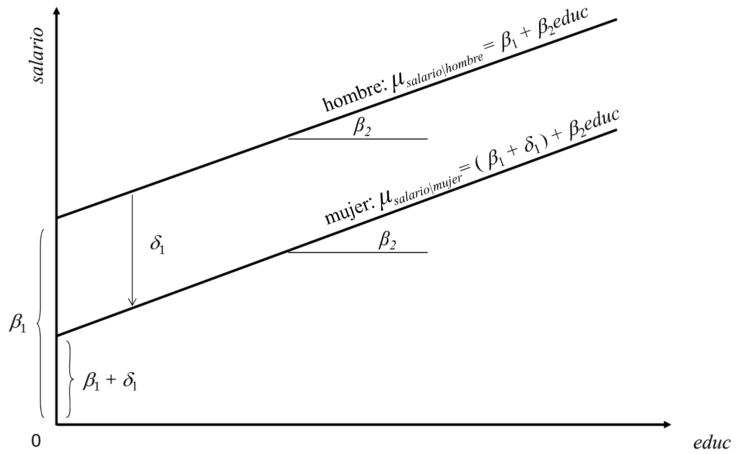


**FIGURA 5.1. Misma pendiente, término independiente diferente.**

En (5-2) hemos incluido una variable ficticia para las mujeres, pero no para los hombres porque incluir las dos variables ficticias habría sido redundante. De hecho, todo lo que necesitamos es dos términos independientes, uno para mujeres y otro para los hombres. Como hemos visto, con la introducción de la variable ficticia *mujer*, nos permite obtener un término independiente para cada género. La introducción de dos variables ficticias causaría multicolinealidad perfecta, ya que *mujer*+*hombre*=1, lo que significa que *hombre* es una función lineal exacta de *mujer* y del término independiente. La inclusión de variables ficticias para ambos sexos, además el término independiente, es el ejemplo más sencillo de la llamada trampa de las variables ficticias, como veremos más adelante.

Si usamos *hombre* en lugar de *mujer*, la ecuación de salarios sería la siguiente:

$$salario = \alpha_1 + \gamma_1 hombre + \beta_2 educ + u \tag{5-4}$$

Nada ha cambiado con la nueva ecuación, con excepción de la interpretación de $\alpha_1$ y $\gamma_1$: $\alpha_1$ es el término independiente para las mujeres, que ahora es la *categoría de referencia*, y $\alpha_1+\gamma_1$ es el término independiente para los hombres. Esto implica la siguiente relación entre los coeficientes:

$$\alpha_1=\beta_1+\delta_1 \text{ y } \alpha_1+\gamma_1=\beta_1 \Rightarrow \gamma_1=-\delta_1$$

En cualquier aplicación, no importa cómo elijamos la categoría de referencia, ya que sólo afecta a la interpretación de los coeficientes asociados a las variables ficticias, pero es importante tener presente qué categoría es la categoría de referencia. La elección de una categoría de referencia es, generalmente, una cuestión de conveniencia. También es posible eliminar el término independiente e incluir una variable ficticia para cada categoría. La ecuación será entonces

$$salario = \mu_1 hombre + \nu_1 mujer + \beta_2 educ + u \tag{5-5}$$

donde el término independiente es $\mu_1$ para los hombres y $\nu_1$ para las mujeres.

El contraste de hipótesis se realiza como de costumbre. En el modelo (5-2) la hipótesis nula de no discriminación entre hombres y mujeres es $H_0: \delta_1 = 0$, mientras que la hipótesis alternativa de que existe discriminación contra la mujer es $H_1: \delta_1 < 0$. Por lo tanto, en este caso, debemos aplicar un contraste *t* de una sola cola (la izquierda).

En las especificaciones formuladas en el trabajo aplicado es usual transformar la variable dependiente tomando logaritmos, ln(*y*), en modelos de este tipo. Por ejemplo:

$$\ln(salario) = \beta_1 + \delta_1 mujer + \beta_2 educ + u \tag{5-6}$$

Veamos la interpretación del coeficiente de la variable ficticia en un modelo con logaritmos. En el modelo (5-6), tomando $u$=0, el salario para una mujer y para un hombre son los siguientes:

$$\ln(salario_M) = \beta_1 + \delta_1 + \beta_2 educ \tag{5-7}$$

$$\ln(salario_H) = \beta_1 + \beta_2 educ \tag{5-8}$$

Dado el mismo nivel de educación, si restamos (5-7) de (5-8), tenemos

$$\ln(salario_M) - \ln(salario_H) = \delta_1 \tag{5-9}$$

Tomando antilogaritmos en (5-9) y restando 1 de ambos miembros de (5-9), obtenemos

$$\frac{salario_M}{salario_H} - 1 = e^{\delta_1} - 1 \tag{5-10}$$

es decir

$$\frac{salario_M - salario_H}{salario_H} = e^{\delta_1} - 1 \tag{5-11}$$

De acuerdo con (5-11), la tasa de variación entre el salario femenino y el salario de los hombres, para un mismo nivel de educación, es igual a $e^{\delta_1} - 1$. Por lo tanto, la tasa exacta de variación porcentual entre el salario por hora de hombres y mujeres es de $100 \times (e^{\delta_1} - 1)$. Como una aproximación a este cambio se puede utilizar $100\times\delta_1$, pero si la magnitud del porcentaje es alta esta aproximación no es tan buena.

***EJEMPLO 5.1 ¿Existe discriminación salarial para la mujer en España?***

Utilizando datos de la *Encuesta de Estructura Salarial de Españ*a para 2002 (fichero *wage02sp*), se ha estimado el modelo (5-6) y se han obtenido los siguientes resultados:

$$\widehat{\ln(wage)} = \underset{(0.026)}{1.731} - \underset{(0.022)}{0.307}\, female + \underset{(0.0025)}{0.055}\, educ$$

$$SCR=393 \quad R^2=0.243 \quad n=2000$$

donde *wage* es el salario hora en euros, *female* es una variable ficticia que toma el valor 1 si es mujer, y *educ* son los años de educación. (Los números entre paréntesis son los errores estándar de los estimadores.)

Para responder a la pregunta planteada más arriba, tenemos que contrastar $H_0 : \delta_1 = 0$ contra de $H_1 : \delta_1 < 0$. Dado que el estadístico *t* es igual a -14.27 se rechaza la hipótesis nula para $\alpha$=0.01. Es decir, hay evidencias de una discriminación en España contra la mujer en el año 2002. De hecho, la diferencia porcentual en el salario por hora entre hombres y mujeres es $100 \times (e^{0.307} - 1) = 35.9\%$, dados unos mismos años en educación.

***EJEMPLO 5.2 Análisis de la relación entre la capitalización de mercado y el valor contable: el papel del IBEX35***

Un investigador desea estudiar la relación entre la capitalización de mercado y el valor contable de las acciones cotizadas en el mercado continuo de la Bolsa de Madrid. En este mercado algunas empresas están incluidas en el Ibex35, que es un índice selectivo. El investigador también quiere saber si acciones incluidas en el Ibex 35 tienen una mayor capitalización en promedio. Con este propósito en mente, el investigador formula el siguiente modelo:

$$\ln(marktval) = \beta_1 + \delta_1 ibex35 + \beta_2 \ln(bookval) + u \tag{5-12}$$

- *marktval* es el valor de mercado de una compañía. Se calcula multiplicando el precio de la acción por el número de acciones emitidas.
- *bookval* es el valor contable de una compañía. También se conoce como valor neto de la compañía. El valor contable se calcula como la diferencia entre los activos de una compañía y sus pasivos.
- *ibex35* es una variable ficticia que toma el valor 1 si la compañía está incluida en el selectivo Ibex 35.

Utilizando las 92 compañías que cotizaron el 15 de noviembre 2011, y que suministraron información sobre el valor contable (fichero *bolmad11*), se obtuvieron los siguientes resultados:

$$\widehat{\ln(marktval)} = \underset{(0.243)}{1.784} + \underset{(0.179)}{0.690}\, ibex35 + \underset{(0.037)}{0.675} \ln(bookval)$$

$$SCR=35.672 \quad R^2=0.893 \quad n=92$$

La elasticidad *marktval*/*bookval* es igual a 0.690, es decir, si el valor contable se incrementa en 1%, la capitalización bursátil de las acciones que cotizan aumentará en un 0.675%.

Contrastar si las acciones incluidas en el Ibex35 tienen en promedio una mayor capitalización implica contrastar $H_0 : \delta_1 = 0$ contra $H_1 : \delta_1 > 0$. Dado que el estadístico *t* es (0.690/0.179)=3.85, entonces rechazamos la hipótesis nula para los niveles habituales de significación. Por otro lado, vemos que las acciones incluidas en el Ibex35 cotizan un 99.4% más elevado que las acciones no incluidas. El porcentaje se obtiene como sigue: $100 \times (e^{0.690} - 1) = 99.4\%$

***EJEMPLO 5.3 ¿Gastan más en pescado las personas que viven en zonas urbanas que las que viven en zonas rurales?***

Para ver si las personas que viven en zonas urbanas gastan más en pescado que las personas que viven en zonas rurales, se ha propuesto el modelo siguiente:

$$\ln(fish) = \beta_1 + \delta_1 urban + \beta_2 \ln(inc) + u \qquad (5\text{-}13)$$

donde *fish* es gasto en pescado, *urban* es una variable ficticia que toma el valor 1 si la persona vive en una zona urbana e *inc* es la renta disponible.

Utilizando una muestra de tamaño 40 (fichero *demand*), se estimó el modelo (5-13)

$$\widehat{\ln(fish)} = \underset{(0.511)}{-6.375} + \underset{(0.055)}{0.140}\,urban + \underset{(0.070)}{1.313}\ln(inc)$$

$$SCR=1.131 \qquad R^2=0.904 \qquad n=40$$

De acuerdo con estos resultados, las personas que viven en zonas urbanas gastan en pescado aproximadamente un 14% más que las personas que viven en zonas rurales. Si se contrasta $H_0: \delta_1 = 0$ contra $H_1: \delta_1 > 0$, constatamos que el estadístico $t$ es (0.140/0.055)=2.55. Teniendo en cuenta que $t_{37}^{0.01} \approx t_{35}^{0.01}$ =2.44, se rechaza la hipótesis nula en favor de la alternativa para los niveles habituales de significación. Es decir, hay evidencia empírica de que las personas que viven en las zonas urbanas gastan más en pescado que las personas que viven en las zonas rurales.

## 5.3 Categorías múltiples para un atributo

En el epígrafe anterior consideramos un atributo (género) que tiene dos categorías (mujer y hombre). Ahora vamos a considerar atributos con más de dos categorías. En concreto, vamos a examinar un atributo con 3 categorías

Para medir el impacto del tamaño de la empresa sobre el salario, podemos utilizar variables dicotómicas. Supongamos que las empresas se clasifican en tres grupos según su tamaño: pequeñas (hasta 49 trabajadores), medianas (de 50 a 199 trabajadores) y grandes (más de 199 trabajadores). Con esta información podemos construir 3 variables ficticias:

$$pequeña = \begin{cases} 1 & \text{hasta 49 trabajadores} \\ 0 & \text{en otros casos} \end{cases}$$

$$mediana = \begin{cases} 1 & \text{de 50 a 199 trabajadores} \\ 0 & \text{en otros casos} \end{cases}$$

$$grande = \begin{cases} 1 & \text{mas de 199 trabajadores} \\ 0 & \text{en otros casos} \end{cases}$$

Si queremos explicar el salario por hora introduciendo en el modelo el tamaño de la empresa, es necesario omitir una de las categorías. En el siguiente modelo la categoría omitida son las empresas pequeñas:

$$salario = \beta_1 + \theta_1 mediana + \theta_2 grande + \beta_2 educ + u \qquad (5\text{-}14)$$

La interpretación de los coeficientes $\theta_j$ es la siguiente: $\theta_1$ ($\theta_2$) es la diferencia en el salario por hora entre las empresas medianas (grandes) y las pequeñas, dado un mismo nivel de educación (y un mismo término de perturbación, $u$).

Vamos a ver qué pasa si también incluimos en (5-14) la categoría *pequeña*. En ese caso, tendríamos el siguiente modelo:

$$salario = \beta_1 + \theta_0 pequeña + \theta_1 mediana + \theta_2 grande + \beta_2 educ + u \qquad (5\text{-}15)$$

Ahora, consideremos que tenemos una muestra de seis observaciones: las observaciones 1 y 2 corresponden a empresas pequeñas, la 3 y la 4 a medianas, y la 5 y 6 a grandes. En este caso, la matriz **X** de regresores tendría la siguiente configuración:

$$\mathbf{X}=\begin{bmatrix} 1 & 1 & 0 & 0 & educ_1 \\ 1 & 1 & 0 & 0 & educ_2 \\ 1 & 0 & 1 & 0 & educ_3 \\ 1 & 0 & 1 & 0 & educ_4 \\ 1 & 0 & 0 & 1 & educ_5 \\ 1 & 0 & 0 & 1 & educ_6 \end{bmatrix}$$

Como puede verse en la matriz **X**, la columna 1 de esta matriz es igual a la suma de las columnas 2, 3 y 4. Por lo tanto, existe multicolinealidad perfecta, debido a la llamada *trampa de las variables ficticias*. Generalizando, si un atributo tiene $g$ categorías, en el modelo únicamente tenemos que incluir $g$-1 variables ficticias, junto con el término independiente. El término independiente para la categoría de referencia es el término independiente general del modelo, y el coeficiente de la variable ficticia de un grupo particular representa la diferencia estimada entre los términos independientes entre esa categoría y la categoría de referencia. Si incluimos $g$ variables ficticias, junto con un término independiente se caerá en la trampa de las variables ficticias. Una alternativa es incluir $g$ variables ficticias, y excluir el término independiente general. En el caso que nos ocupa, el modelo sería el siguiente:

$$salario=\theta_0 pequeña+\theta_1 mediana+\theta_2 grande+\beta_1 educ+u \qquad (5\text{-}16)$$

Esta solución no es aconsejable por dos razones. Con esta configuración del modelo es más difícil contrastar las diferencias con respecto a una categoría de referencia. En segundo lugar, esta solución sólo funciona en el caso de un modelo con sólo un atributo.

***EJEMPLO 5.4 ¿Influye el tamaño de la empresa en la determinación de los salarios?***

Utilizando la muestra del ejemplo 5.1 (fichero *wage02sp*), se estimó el modelo (5-14), tomando log en *wage*:

$$\widehat{\ln(wage)}=\underset{(0.027)}{1.566}+\underset{(0.025)}{0.281}medium+\underset{(0.024)}{0.162}large+\underset{(0.003)}{0.048}educ$$

$$SCR=406 \quad R^2=0.218 \quad n=2000$$

donde *medium* y *large* son dos variables dicotómicas para designar a empresas de tamaño mediano y grande respectivamente

Para responder a la pregunta inicial no haremos un contraste *individual* de $\theta_1$ o $\theta_2$. En vez de ello, contrastaremos conjuntamente si el tamaño de las empresas tiene una influencia significativa sobre el salario. Es decir, debemos contrastar si las medianas y grandes empresas tomadas conjuntamente tienen una influencia significativa en la determinación del salario. En este caso, las hipótesis nula y alternativa, tomando a (5-14) como el modelo no restringido, serán las siguientes:

$$H_0:\theta_1=\theta_2=0$$
$$H_1:H_0 \text{ no es cierta}$$

El modelo restringido en este caso es el siguiente:

$$\ln(wage)=\beta_1+\beta_2 educ+u \qquad (5\text{-}17)$$

La estimación de este modelo es la siguiente:

$$\widehat{\ln(wage)}=\underset{(0.026)}{1.657}+\underset{(0.003)}{0.053}educ$$

$$SCR=433 \quad R^2=0.166 \quad n=2000$$

Por lo tanto, el estadístico $F$ es

$$F=\frac{\left[SCR_R-SCR_{NR}\right]/q}{SCR_{NR}/(n-k)}=\frac{\left[433-406\right]/2}{406/(2000-4)}=66.4$$

Así, de acuerdo con el valor del estadístico $F$, se puede concluir que el tamaño de la firma tiene una influencia significativa en la determinación de los salarios para los niveles usuales de significación.

***Ejemplo 5.5 En el caso de Lydia E. Pinkham, ¿son significativas las variables temporales ficticias de forma individual y conjunta?***

En el ejemplo 3.4 vimos el caso de Lydia E. Pinkham en el que las ventas, *sales,* de un extracto de hierbas de esa empresa (en miles de dólares) se explicaba en términos del gasto en publicidad en miles de dólares (*advexp*) así como las ventas del año anterior ($sales_{t-1}$). Sin embargo, su autor, además de estas dos variables, incluyó tres variables temporales ficticias: $d1$, $d2$ y $d3$. Estas variables ficticias abarcan las distintas situaciones por las que pasó la compañía. Así, $d1$ toma el valor 1 en el periodo de 1907-1914 y 0 en los periodos restantes, $d2$ toma el valor 1 en el periodo de 1915-1925 y 0 en otros periodos, y finalmente, $d3$ toma el valor 1 en el periodo de 1926 a 1940 y 0 en los restantes periodos. Por lo tanto, la categoría de referencia es el periodo 1941-1960. En consecuencia, la formulación final del modelo fue la siguiente:

$$sales_t=\beta_1+\beta_2 advexp_t+\beta_3\, sales_{t-1}+\beta_4 d1_t+\beta_5 d2_t+\beta_6 d3_t+u_t \qquad (5\text{-}18)$$

Los resultados obtenidos en la regresión, utilizando el fichero *pinkham*, fueron los siguientes:

$$\widehat{sales_t}=\underset{(96.3)}{254.6}+\underset{(0.136)}{0.5345}advexp_t+\underset{(0.0814)}{0.6073}sales_{t-1}-\underset{(89)}{133.35}d1_t+\underset{(67)}{216.84}d2_t-\underset{(67)}{202.50}d3_t$$

$$R^2=0.929 \qquad n=53$$

Para contrastar si las variables ficticias de forma individual tienen un efecto significativo en las ventas, las hipótesis nula y alternativa son:

$$\begin{cases}H_0:\theta_i=0\\ H_1:\theta_i\neq 0\end{cases} \qquad i=1,2,3$$

Los correspondientes estadísticos $t$ son los siguientes:

$$t_{\hat{\theta}_1}=\frac{-133.35}{89}=-1.50 \qquad t_{\hat{\theta}_2}=\frac{216.84}{67}=3.22 \qquad t_{\hat{\theta}_3}=\frac{-202.50}{67}=-3.02$$

Como puede verse, el regresor $d1$ no es significativo para los niveles habituales de significación, mientras que por el contrario los regresores $d2$ y $d3$ son significativos para cualquiera de los niveles habituales.

La interpretación del coeficiente del regresor $d2$, por ejemplo, es la siguiente: manteniendo fijo el gasto en publicidad y dadas las ventas del año anterior, las ventas para un año del periodo 1915-1920 son 21.684 dólares mayores, en promedio, que las de un año cualquiera del periodo 1941-1960.

Para estimar el efecto conjunto de las variables temporales ficticias, las hipótesis nula y alternativa son

$$\begin{cases}H_0:\theta_1=\theta_2=\theta_3=0\\ H_1:H_0 \text{ no es cierto}\end{cases}$$

y el contraste estadístico correspondiente es

$$F=\frac{(R_{NR}^2-R_R^2)/q}{(1-R_{NR}^2)/(n-k)}=\frac{(0.9290-0.8770)/3}{(1-0.9290)/(53-6)}=11.47$$

Para cualquiera de los niveles habituales de significación la hipótesis nula es rechazada. Por lo tanto, las variables temporales ficticias tienen un efecto significativo sobre las ventas

## 5.4 Varios atributos

Ahora vamos a considerar la posibilidad de tener en cuenta dos atributos para explicar la determinación del salario: el género y duración de la jornada de trabajo (a tiempo parcial y a tiempo completo). La variable ficticia *tiempar,* va ser una variable binaria que toma el valor 1 cuando el tipo de contrato es a tiempo parcial y 0 si es a tiempo completo. En el siguiente modelo se introducen las dos variables ficticias: *mujer* y *tiempar*:

$$salario = \beta_1 + \delta_1 mujer + \phi_1 tiempar + \beta_2 educ + u \qquad (5\text{-}19)$$

En este modelo, $\phi_1$ es la diferencia en el salario por hora entre las personas que trabajan a tiempo parcial, para un género dado y para el mismo nivel de educación (y también el mismo término de perturbación, *u*).

Cada uno de estos dos atributos tiene una categoría de referencia, que es la categoría omitida. En este caso, *hombre* es la categoría de referencia para el género y tiempo completo para el tipo de contrato. Si tomamos las esperanzas para las cuatro categorías implicadas, se obtiene:

$$\begin{aligned}
\mu_{wage|mujer,tiempar} &= E\left[wage \mid mujer, tiempar, educ\right] = \beta_1 + \delta_1 + \phi_1 + \beta_2 educ \\
\mu_{wage|mujer,tiemcom} &= E\left[wage \mid mujer, tiemcom, educ\right] = \beta_1 + \delta_1 + \beta_2 educ \\
\mu_{wage|hombre,tiempar} &= E\left[wage \mid hombre, tiempar, educ\right] = \beta_1 + \phi_1 + \beta_2 educ \\
\mu_{wage|hombre,tiemcom} &= E\left[wage \mid hombre, tiemcom, educ\right] = \beta_1 + \beta_2 educ
\end{aligned} \qquad (5\text{-}20)$$

El término independiente general en la ecuación refleja el efecto de ambas categorías de referencia, hombre y tiempo completo; es decir, la categoría de referencia es hombre con jornada a tiempo completo. En (5-20) puede verse el término independiente para cada combinación de categorías.

***EJEMPLO 5.6 La influencia del género y duración de la jornada de trabajo en la determinación de los salarios***

El modelo (5-19), tomando log en *wage*, se estimó utilizando datos de la Encuesta de Estructura Salarial de España para el año 2006 (fichero *wage06sp*):

$$\widehat{\ln(wage)} = \underset{(0.026)}{2.005} - \underset{(0.021)}{0.233} female - \underset{(0.027)}{0.087} partime + \underset{(0.002)}{0.053} educ$$

$$SCR=365 \quad R^2=0.235 \quad n=2000$$

donde *partime* es un contrato a tiempo parcial.

De acuerdo con los valores de los coeficientes y los correspondientes errores estándar, es evidente que cada una de las dos variables ficticias, *female* y *partime*, son estadísticamente significativas para los niveles habituales de significación.

***EJEMPLO 5.7 Análisis del absentismo laboral en la empresa Buenosaires***

Buenosaires es una empresa dedicada a la fabricación de ventiladores, habiendo tenido resultados relativamente aceptables en los últimos años. Los directivos consideran que éstos habrían sido mejores si el absentismo en la empresa no fuera tan alto. Con el propósito de analiza los factores que determinan el absentismo, se propone el siguiente modelo:

$$absent = \beta_1 + \delta_1 bluecoll + \phi_1 male + \beta_2 age + \beta_3 tenure + \beta_4 wage + u \qquad (5\text{-}21)$$

donde *bluecollr* es una variable ficticia que indica que la persona es un trabajador manual (la categoría de referencia es cuello blanco), male es una variable dicotómica que toma el valor 1 si el trabajador es hombre. Las variables *tenure* y *age* son continuas que reflejan los años trabajando en la empresa y la edad respectivamente.

Utilizando una muestra de tamaño 48 (fichero *absent*), se ha estimado la siguiente ecuación

$$\widehat{absent} = \underset{(1.640)}{12.444} + \underset{(0.669)}{0.968} bluecoll + \underset{(0.712)}{2.049} male - \underset{(0.047)}{0.037} age - \underset{(0.065)}{0.151} tenure - \underset{(0.007)}{0.044} wage$$

$$SCR=161.95 \quad R^2=0.760 \quad n=48$$

Ahora vamos a ver si *bluecoll* es significativa. Contrastando $H_0: \delta_1 = 0$ contra $H_1: \delta_1 \neq 0$, el estadístico *t* es (0.968/0.669)=1.45. Como $t_{40}^{0.10/2}$ =1.68, fracasamos en rechazar la hipótesis nula para $\alpha$=0.10. Entonces no hay evidencia empírica para afirmar que el absentismo de los trabajadores manuales (cuello azul) es diferente del de los trabajadores de oficina (cuello blanco). Pero si se contrasta $H_0:\delta_1=0$ contra $H_1:\delta_1>0$, como $t_{40}^{0.10}$=1.30 para $\alpha$=0.10, no se puede rechazar que el absentismo de los trabajadores de cuello azul sea mayor que el de los trabajadores de cuello blanco.

Por el contrario, en el caso de la variable ficticia *male*, contrastando $H_0:\varphi_1=0$ contra $H_1:\varphi_1\neq 0$, dado que el estadístico *t* es (2.049/0.712)=2.88 y $t_{40}^{0.01/2}$ =2.70, rechazamos que el absentismo sea igual en hombres y mujeres para los niveles habituales de significación.

***EJEMPLO 5.8 Tamaño de la empresa y género en la determinación del salario***

Para conocer si el tamaño de la empresa y el género, de forma conjunta, son dos factores relevantes en la determinación del salario, se formula el siguiente modelo:

$$\ln(wage)=\beta_1+\delta_1 female+\theta_1 medium+\theta_2 large+\beta_2 educ+u \qquad (5\text{-}22)$$

En este caso tenemos que hacer un contraste conjunto, donde las hipótesis nula y alternativa son,

$$H_0:\delta_1=\theta_1=\theta_2=0$$
$$H_1:H_0 \text{ no es cierta}$$

El modelo restringido en este caso es el modelo de (5-17), que se estimó en el ejemplo 5.4 (fichero *wage02sp*). La estimación del modelo no restringido es la siguiente:

$$\widehat{\ln(wage)}=\underset{(0.026)}{1.639}-\underset{(0.021)}{0.327}\,female+\underset{(0.023)}{0.308}\,medium+\underset{(0.023)}{0.168}\,large+\underset{(0.0024)}{0.050}\,educ$$

$$SCR=361 \quad R^2=0.305 \quad n=2000$$

El estadístico *F* es

$$F=\frac{\left[SCR_R-SCR_{NR}\right]/q}{SCR_{NR}/(n-k)}=\frac{\left[433-361\right]/3}{361/(2000-5)}=133$$

Por lo tanto, de acuerdo con el valor de *F*, se puede concluir que el tamaño de la firma y el género tienen conjuntamente una influencia significativa en la determinación del salario.

# 5.5 Las interacciones que implican variables ficticias.

## 5.5.1 Interacciones entre dos variables ficticias

Para permitir la posibilidad de que exista una interacción entre el género y duración de la jornada de trabajo en la determinación salarial podemos añadir al modelo (5-19) un término de interacción entre *mujer* y *tiempar*, con lo que el modelo a estimar será el siguiente:

$$salario=\beta_1+\delta_1 mujer+\phi_1 tiempar+\varphi_1 mujer\times tiempar+\beta_2 educ+u \qquad (5\text{-}23)$$

Esto permite determinar si el efecto de la duración de la jornada de trabajo en el salario depende, o no, del género. Análogamente, también permite si la influencia del género en el salario depende, o no, de la duración de la jornada de trabajo.

***EJEMPLO 5.9 ¿Es la interacción entre las mujeres y el trabajo a tiempo parcial significativa?***

El modelo (5-23) se estimó utilizando los datos de la Encuesta de Estructura Salarial de España para 2006 (fichero *wage06sp*):

$$\widehat{\ln(wage)}=\underset{(0.026)}{2.007}-\underset{(0.022)}{0.259}\,female-\underset{(0.047)}{0.198}\,partime+\underset{(0.058)}{0.167}\,female\times partime+\underset{(0.002)}{0.054}\,educ$$

$$SCR=363 \quad R^2=0.238 \quad n=2000$$

Para responder a la pregunta planteada, tenemos que contrastar $H_0:\varphi_1=0$ contra $H_0:\varphi_1\neq 0$. Dado que el estadístico *t* es (0.167/0.058)=2.89, y teniendo en cuenta que $t_{60}^{0.01/2}$ =2.66 se rechaza la hipótesis nula en favor de la hipótesis alternativa. Por lo tanto, existe evidencia empírica de que la interacción entre *female* y *partime* es estadísticamente significativa.

***EJEMPLO 5.10 ¿Discriminan las empresas pequeñas a las mujeres más, o menos, que las empresas grandes?***

Para responder a esta pregunta se formula el siguiente modelo:

$$\begin{aligned}\ln(wage) &= \beta_1 + \delta_1 female + \theta_1 medium + \theta_2 large \\ &+ \varphi_1 female \times medium + \varphi_2 female \times large + \beta_2 educ + u\end{aligned} \qquad (5\text{-}24)$$

Utilizando la muestra del ejemplo 5.1 (fichero *wage02sp*), fue estimado el modelo (5-24):

$$\begin{aligned}\widehat{\ln(wage)} &= \underset{(0.027)}{1.624} - \underset{(0.034)}{0.262}\, female + \underset{(0.028)}{0.361} medium + \underset{(0.027)}{0.179} large \\ &- \underset{(0.050)}{0.159}\, female \times medium - \underset{(0.051)}{0.043}\, female \times large + \underset{(0.0024)}{0.050} educ\end{aligned}$$

$SCR$=359 $R^2$=0.308 $n$=2000

Si en (5-24) los parámetros $\varphi_1$ y $\varphi_2$ son igual a 0, esto implica que, en la ecuación para la determinación del salario, no hay interacción entre género y tamaño de la empresa. Así para responder a la pregunta planteada tomamos (5-24) como el modelo no restringido. Las hipótesis nula y alternativa serán las siguientes:

$$H_0 : \varphi_1 = \varphi_2 = 0$$
$$H_1 : H_0 \text{ no es cierta}$$

Por lo tanto, el modelo restringido es, en este caso, el modelo (5-22), que se estimó en el ejemplo 5.7. El estadístico $F$ toma el valor

$$F = \frac{\left[SCR_R - SCR_{NR}\right]/q}{SCR_{NR}/(n-k)} = \frac{\left[361-359\right]/2}{359/(2000-7)} = 5.55$$

Para $\alpha$=0.01, resulta que $F^{0.01}_{2,1993} \simeq F^{0.01}_{2,60} = 4.98$. Como $F$>5.61, rechazamos $H_0$ en favor de $H_1$. Si se rechaza $H_0$ para $\alpha$=0.01, también será rechazada para los niveles de 5% y 10%. Por tanto, para los niveles usuales de significación, la interacción entre género y tamaño de empresa es relevante en la determinación del salario.

## 5.5.2 Interacciones entre una variable ficticia y una variable cuantitativa

Hasta ahora, en los ejemplos sobre determinación del salario se ha utilizado una variable ficticia para desplazar el término independiente o para estudiar su interacción con otra variable ficticia, pero manteniendo la pendiente de *educ* constante. Ahora bien, también se pueden utilizar las variables ficticias para desplazar pendientes si interactúan con cualquier variable explicativa continua. Por ejemplo, en el siguiente modelo la variable ficticia *mujer* interactúa con la variable continua *educ*:

$$salario = \beta_1 + \beta_2 educ + \delta_1 mujer \times educ + u \qquad (5\text{-}25)$$

En este modelo, como puede verse en la figura 5.2, el término independiente es el mismo para hombres y para mujeres, pero la pendiente es mayor en hombres que en mujeres, porque $\delta_1$ es negativa.

En el modelo de (5-25), los rendimientos de un año adicional en educación dependen del género del individuo. De hecho,

$$\frac{\partial salario}{\partial educ} = \begin{cases} \beta_2 + \delta_1 & \text{para mujeres} \\ \beta_2 & \text{para hombres} \end{cases} \qquad (5\text{-}26)$$

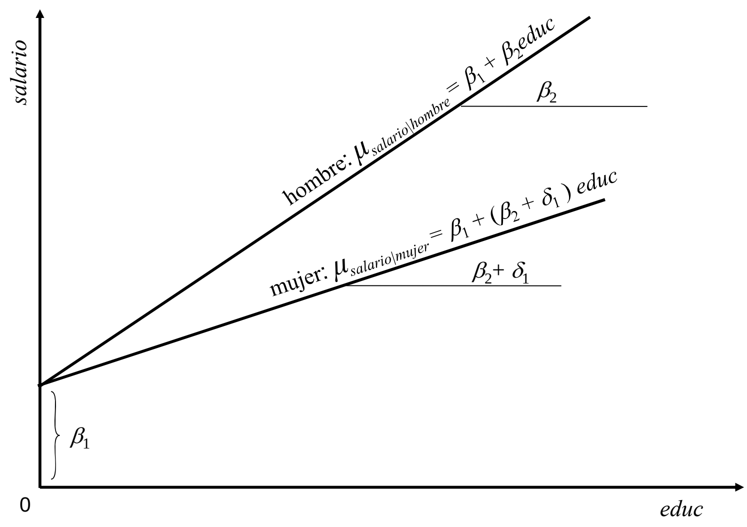


**FIGURA 5.2. Diferente pendiente, mismo término independiente.**

***EJEMPLO 5.11 ¿Es el rendimiento de la educación para los hombres mayor que para las mujeres?***

Utilizando la muestra del ejemplo 5.1 (fichero *wage02sp*), tomando logaritmos en *wage*, se ha estimado el modelo (5-25):

$$\widehat{\ln(wage)} = \underset{(0.025)}{1.640} + \underset{(0.0026)}{0.063}\, educ - \underset{(0.0021)}{0.027}\, educ \times female$$

$$SCR=400 \quad R^2=0.229 \quad n=2000$$

En este caso necesitamos contrastar $H_0 : \delta_1 = 0$ contra $H_1 : \delta_1 < 0$ . Dado que el estadístico *t* es (-0.028/0.0002) =-12.81, se rechaza la hipótesis nula en favor de la hipótesis alternativa para cualquier nivel de significación. Es decir, existe evidencia empírica de que el rendimiento de un año adicional de educación es mayor para hombres que para mujeres.

## 5.6 Contraste de cambio estructural

Hasta ahora hemos contrastado las hipótesis de que un parámetro, o un subconjunto de parámetros del modelo, son diferentes para dos grupos (mujeres y hombres, por ejemplo). Pero a veces, queremos contrastar la hipótesis nula de que dos grupos tienen la misma función de regresión poblacional, frente a la alternativa de que no es la misma. En otras palabras, queremos contrastar si la misma ecuación es válida para los dos grupos. Existen dos procedimientos para realizar este contraste, denominado *contraste de cambio estructural*: utilizando variables ficticias y realizando regresiones separadas mediante el contraste de Chow.

### 5.6.1 Utilizando variables ficticias

En este procedimiento, contrastar si hay diferencias entre grupos consiste en realizar un contraste de significación conjunto de la variable ficticia que diferencia entre los dos grupos y de sus interacciones con todas los otros regresores. Por lo tanto, estimamos el modelo con (*modelo no restringido*) y sin (*modelo restringido*) la variable ficticia y todas sus interacciones.

De la estimación de ambas ecuaciones se obtiene el estadístico *F*, ya sea a través de la *SCR* o del $R^2$. En el siguiente modelo, para la determinación del salario, tanto el término independiente como la pendiente son diferentes para hombres y mujeres:

$$salario = \beta_1 + \delta_1 mujer + \beta_2 educ + \delta_2 mujer \times educ + u \qquad (5\text{-}27)$$

En la figura 5.3, ha sido representado la función de regresión poblacional de este modelo. Como puede verse, si *mujer*=1, se obtiene que

$$salario = (\beta_1 + \delta_1) + (\beta_2 + \delta_2)educ + u \qquad (5\text{-}28)$$

Entonces, para las mujeres el término independiente es $\beta_1 + \delta_1$ y la pendiente $\beta_2 + \delta_2$. Para *mujer*=0, obtenemos la ecuación (5-1) En este caso, para los hombres el término independiente es $\beta_1$ y la pendiente $\beta_2$. Por lo tanto, $\delta_1$ mide la diferencia entre los términos independientes para mujeres y hombres y $\delta_2$ mide a su vez la diferencia en el rendimiento la educación entre mujeres y hombres. La figura 5.3 muestra un término independiente y una pendiente menores para mujeres que para hombres. Esto significa que las mujeres ganan menos que los hombres en todos los niveles de la educación, y que la brecha aumenta a medida que *educ* se hace más grande, es decir, un año adicional de educación tiene un rendimiento inferior para mujeres que para hombres.

La estimación (5-27) es equivalente a la estimación de dos ecuaciones de salarios, uno para hombres y otro para las mujeres, por separado. La única diferencia es que (5-27) impone la misma varianza a los dos grupos, mientras que las regresiones por separado no lo hacen. Esta especificación del modelo es ideal, como veremos más adelante, para contrastar la igualdad de pendientes, la igualdad de términos independientes, o la igualdad tanto de términos independientes como de pendientes en los dos grupos.

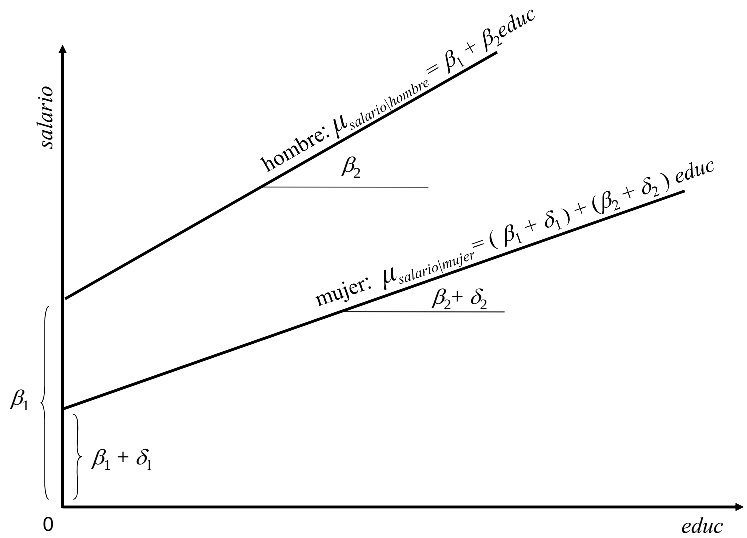


**FIGURA 5.3. Pendiente diferente, diferente término independiente.**

***EJEMPLO 5.12 ¿Es la ecuación de salarios válida tanto para hombres como para mujeres?***

Si los parámetros $\delta_1$ y $\delta_2$ son iguales a 0 en el modelo de (5-27), implica que la ecuación para la determinación de los salarios es la misma para hombres y mujeres. Entonces para responder a la cuestión planteada, tomamos (5-27), pero expresando el salario en logaritmos, como el modelo el modelo no restringido. Las hipótesis nula y alternativa serán las siguientes:

$$H_0 : \delta_1 = \delta_2 = 0$$

$$H_1 : H_0 \text{ no es cierto}$$

Por lo tanto, el modelo restringido es el modelo de (5-17). Utilizando la misma muestra que en el ejemplo 5.1 (fichero *wage02sp*), hemos obtenido la siguiente estimación de los modelos (5-27) y (5-17):

$$\widehat{\ln(wage)} = \underset{(0.030)}{1.739} - \underset{(0.055)}{0.332}\, female + \underset{(0.0030)}{0.054}\, educ - \underset{(0.0054)}{0.003}\, educ \times female$$

$$SCR=393 \qquad R^2=0.243 \qquad n=2000$$

$$\widehat{\ln(wage)} = \underset{(0.026)}{1.657} + \underset{(0.0026)}{0.0525} educ$$

$$SCR=433 \quad R^2=0.166 \quad n=2000$$

El estadístico $F$ toma el valor

$$F = \frac{\left[SCR_R - SCR_{NR}\right]/q}{SCR_{NR}/(n-k)} = \frac{\left[433-393\right]/2}{393/(2000-4)} = 102$$

Está claro que para cualquier nivel de significancia, las ecuaciones para hombres y mujeres son diferentes.

Cuando contrastamos en el ejemplo 5.1 si había discriminación contra la mujer en España ($H_0: \delta_1 = 0$ contra $H_1: \delta_1 < 0$), se asumió que la pendiente de *educ* (modelo (5-6)) era la misma para hombres y mujeres. Ahora también es posible utilizar el modelo (5-27) para contrastar la misma hipótesis nula, pero asumiendo que la pendiente es diferente. Dado que el estadístico $t$ es (-0.332/0.0546)=-6.06, entonces se rechaza la hipótesis nula mediante el uso de este modelo más general que el del ejemplo 5.1.

En el ejemplo 5.11 se contrastó si el coeficiente de $\delta_2$ en el modelo de (5-25), tomando log en *wage*, era 0, suponiendo que el término independiente es el mismo para hombres y mujeres. Ahora bien, si tomamos (5-27), tomando log en *wage*, como modelo no restringido, podemos contrastar la misma hipótesis nula, pero asumiendo que el término independiente es diferente para hombres y mujeres. Dado que el estadístico $t$ es (0.0027/0.0054)=0.493, entonces no se puede rechazar la hipótesis nula de que no existe interacción entre género y educación.

***EJEMPLO 5.13 ¿Tienen los consumidores urbanos el mismo patrón de comportamiento que los rurales con respecto al gasto en pescado?***

Para responder a esta pregunta se formula el siguiente modelo, que se tomará como modelo *no restringidos*:

$$\ln(fish) = \beta_1 + \delta_1 urban + \beta_2 \ln(inc) + \delta_2 \ln(inc) \times urban + u \tag{5-29}$$

Las hipótesis nula y alternativa serán las siguientes:

$$H_0: \delta_1 = \delta_2 = 0$$
$$H_1: H_0 \text{ no es cierta}$$

El modelo restringido correspondiente a esta $H_0$ es

$$\ln(fish) = \beta_1 + \beta_2 \ln(inc) + u \tag{5-30}$$

Utilizando la muestra del ejemplo 5.3 (fichero *demand*), se han estimado los modelos (5-29) y (5-30):

$$\widehat{\ln(fish)} = -\underset{(0.627)}{6.551} + \underset{(1.095)}{0.678} urban + \underset{(0.087)}{1.337} \ln(inc) - \underset{(0.152)}{0.075} \ln(inc) \times urban$$

$$SCR=1.123 \quad R^2=0.904 \quad n=40$$

$$\widehat{\ln(fish)} = -\underset{(0.542)}{6.224} + \underset{(0.075)}{1.302} \ln(inc)$$

$$SCR=1.325 \quad R^2=0.887 \quad n=40$$

El estadístico $F$ toma el valor

$$F = \frac{\left[SCR_R - SCR_{NR}\right]/q}{SCR_{NR}/(n-k)} = \frac{\left[1.325-1.123\right]/2}{1.123/(40-4)} = 3.24$$

Si miramos en la tabla estadística de la $F$ para 2 *grados de libertad* en el numerador y 35 *gl* en el denominador para $\alpha$=0.10, vemos que $F_{2,36}^{0.10} \simeq F_{2,35}^{0.10} = 2.46$. Como $F$>2.46 rechazamos la $H_0$. Sin embargo, como $F_{2,36}^{0.05} \simeq F_{2,35}^{0.05} = 3.27$, fracasamos en rechazar $H_0$ a favor de $H_1$ para $\alpha$=0.05 y, por tanto, para $\alpha$=0.01. Conclusión: no hay una evidencia fuerte de que las familias que viven en las zonas rurales tengan un patrón de consumo diferente de pescado con respecto a las familias que viven en zonas urbanas.

***Ejemplo 5.14 ¿Ha cambiado la estructura productiva de las regiones españolas?***

La pregunta que debe responderse es específicamente la siguiente: ¿ha cambiado la estructura productiva de las regiones españolas entre 1995 y 2008? El problema que se plantea es un problema de estabilidad estructural. Para especificar el modelo que se toma como referencia en la estimación, vamos a definir la variable ficticia $y2008$, que toma el valor 1 si el año es 2008 y 0 si el año es 1995.

El modelo de referencia es un modelo de Cobb-Douglas, que introduce parámetros adicionales para recoger los cambios estructurales que puedan haber ocurrido. Su expresión es la siguiente:

$$\ln(q) = \gamma_1 + \alpha_1 \ln(k) + \beta_1 \ln(l) + \gamma_2 y2008 + \alpha_2 y2008 \times \ln(k) + \beta_2 y2008 \times \ln(l) + u \qquad (5\text{-}31)$$

Es fácil ver, de acuerdo con la definición de la variable ficticia $y2008$ que las elasticidades de producción/capital en 1995 y 2008 son diferentes. En concreto, toman los siguientes valores:

$$\varepsilon_{Q/K(1995)} = \frac{\partial \ln(Q)}{\partial \ln(K)} = \alpha_1 \qquad \varepsilon_{Q/K(2008)} = \frac{\partial \ln(Q)}{\partial \ln(K)} = \alpha_1 + \alpha_2$$

En el caso de que $\alpha_2$ sea igual a 0, entonces la elasticidad de la producción/capital es la misma en ambos periodos.

Del mismo modo, las elasticidades de producción/trabajo para los dos periodos vienen dadas por

$$\varepsilon_{Q/K(1995)} = \frac{\partial \ln(L)}{\partial \ln(K)} = \beta_1 \qquad \varepsilon_{Q/K(2008)} = \frac{\partial \ln(L)}{\partial \ln(K)} = \beta_1 + \beta_2$$

El término independiente de la función Cobb-Douglas es un parámetro que mide la eficiencia. En el modelo de (5-31) se considera la posibilidad de que el parámetro de eficiencia (*PEF*) sea diferente en los dos periodos. Así,

$$PEF(1995) = \gamma_1 \qquad PEF(2008) = \gamma_1 + \gamma_2$$

Si los parámetros $\alpha_2$, $\beta_2$ y $\gamma_2$ son cero en el modelo (5-31), la función de producción es la misma en ambos periodos. Por lo tanto, en la estimación de estabilidad estructural de la función de producción, las hipótesis nula y alternativa son:

$$\begin{aligned} &H_0 : \gamma_2 = \alpha_2 = \beta_2 \\ &H_1 : H_0 \text{ no es cierta} \end{aligned} \qquad (5\text{-}32)$$

Bajo la hipótesis nula, las restricciones dadas en (5-32) conducen al modelo restringido siguiente:

$$\ln(q) = \gamma_1 + \alpha_1 \ln(k) + \beta_1 \ln(l) + u \qquad (5\text{-}33)$$

El fichero *prodsp* contiene información para cada una de las regiones españolas en 1995 y 2008 sobre el valor añadido bruto en millones de euros (*gdp*), la ocupación en miles de puestos de trabajo (*labor*), y el capital productivo en millones de euros (*captot*). También en ese archivo se puede encontrar la variable ficticia $y2008$.

A continuación se muestran los resultados del modelo de regresión no restringido (5-31). Es evidente que no podemos rechazar la hipótesis nula de que cada uno de los coeficientes $\alpha_2$, $\beta_2$ y $\gamma_2$, considerados individualmente, sea 0, ya que ninguno de los estadísticos $t$ llega a 0.1 en valor absoluto.

$$\widehat{\ln(gva)} = \underset{(0.916)}{0.0559} + \underset{(0.185)}{0.6743}\ln(captot) + \underset{(0.185)}{0.3291}\ln(labor)$$

$$-\underset{(2.32)}{0.1088}\,y2008 + \underset{(0.419)}{0.0154}\,y2008 \times \ln(captot) - \underset{(0.418)}{0.0094}\,y2008 \times \ln(labor)$$

$$R^2=0.99394 \qquad n=34$$

Los resultados del modelo restringido (5-33) son los siguientes:

$$\widehat{\ln(gva)} = -\underset{(0.200)}{0.0690} + \underset{(0.036)}{0.6959}\ln(captot) + \underset{(0.042)}{0.311}\ln(labor)$$

$$R^2=0.99392 \qquad n=34$$

Como puede verse, las $R^2$ de los dos modelos son prácticamente idénticas ya que difieren sólo a partir del quinto decimal. No es de extrañar, por tanto, que el estadístico $F$ para el contraste de la hipótesis nula (5-32) tenga un valor cercano a 0:

$$F = \frac{(R^2_{UR} - R^2_R)/q}{(1-R^2_{UR})/(n-k)} = \frac{(0.99394 - 0.99392)/3}{(1-0.99394)/(34-6)} = 0.0308$$

Así pues, la hipótesis alternativa de que exista cambio estructural en la economía productiva de las regiones españolas entre 1995 y 2008 se rechaza para cualquier nivel de significación.

### 5.6.2 Utilizando regresiones separadas: el contraste de Chow

Este contraste fue introducido por el económetra Chow (1960). Este autor consideró el problema de contrastar la igualdad de dos conjuntos de coeficientes de regresión. En contraste de Chow el modelo restringido es el mismo que en el caso de uso de variables ficticias para diferenciar entre grupos. Ahora, sin embargo, el modelo no restringido, en lugar de distinguir el comportamiento de dos grupos mediante variables ficticias, consiste simplemente en regresiones separadas. Así, en el ejemplo determinación de los salarios, el modelo no restringido consta de dos ecuaciones:

$$\begin{aligned} mujer: &\quad salario = \beta_{11} + \beta_{21} educ + u \\ hombre: &\quad salario = \beta_{12} + \beta_{22} educ + u \end{aligned} \tag{5-34}$$

Si estimamos ambas ecuaciones por *MCO*, se puede demostrar que la *SCR* del modelo no restringido, $SCR_{NR}$, es igual a la suma de la *SCR* obtenida de la estimación para las mujeres, $SCR_1$, y para los hombres, $SCR_2$. Es decir,

$$SCR_{NR}=SCR_1+SCR_2$$

La hipótesis nula establece que los parámetros de las dos ecuaciones en (5-34) son iguales. Entonces,

$$H_0: \begin{cases} \beta_{11} = \beta_{12} \\ \beta_{21} = \beta_{22} \end{cases}$$

$$H_1: \text{No } H_0$$

Aplicando la hipótesis nula al modelo (5-34), se obtiene el modelo (5-17), que es el modelo restringido. La estimación de este modelo para toda la muestra se suele denominar *regresión agrupada* o *pooled regression* (*P*). Por lo tanto, vamos a considerar que el $SCR_R$ y $SCR_P$ son expresiones equivalentes.

Por lo tanto, el estadístico *F* será la siguiente:

$$F = \frac{\left[SCR_P - \left(SCR_1 + SCR_2\right)\right]/k}{\left[SCR_1 + SCR_2\right]/\left[n-2k\right]} \tag{5-35}$$

Es importante señalar que, bajo la hipótesis nula, deben ser iguales las varianzas de la perturbación para los grupos. Observe que tenemos *k* restricciones: los *k*-1 coeficientes de pendiente (interacciones), más el coeficiente del término independiente. Nótese también que en el modelo no restringido estimamos 2 términos independientes diferentes y 2 coeficientes de pendiente diferentes, por lo que los *gl* del modelo son *n*-2*k*.

Una limitación importante del contraste de Chow es que bajo la hipótesis nula no hay diferencias en absoluto entre los grupos. En la mayoría de los casos, es más interesante permitir diferencias parciales entre ambos grupos, como hemos hecho mediante la utilización de variables ficticias.

El contraste de Chow se puede generalizar a más de dos grupos de un modo natural. Desde el punto de vista práctico, es probablemente más fácil estimar regresiones separadas para cada grupo que utilizar el procedimiento basado en la introducción de variables ficticias en el modelo.

En el caso de tres grupos el estadístico *F* en el contraste de Chow tiene la siguiente configuración:

$$F = \frac{\left[SCR_P - (SCR_1 + SCR_2 + SCR_3)\right]/2\times k}{(SCR_1 + SCR_2 + SCR_3)/(n-3k)} \qquad (5\text{-}36)$$

Observe que, como regla general, el número de *gl* del numerador es igual al (número de grupos-1)$\times k$, mientras que el número de *gl* del denominador es igual a $n$ menos (número de grupos)$\times k$.

***EJEMPLO 5.15 Otra forma de abordar la cuestión de la determinación de los salarios por criterio de género***

Utilizando la misma muestra que en el ejemplo 5.1 (fichero *wage02sp*), hemos obtenido la estimación de las ecuaciones en (5-34), tomando log en *wage*, para hombres y mujeres, las cuales tomadas conjuntamente dan lugar a la estimación del modelo *no restringido*:

Ecuación para la mujer $\qquad \widehat{\ln(wage)} = \underset{(0.042)}{1.407} + \underset{(0.0041)}{0.057}\,educ$

$SCR$=104 $\quad R^2$=0.236 $\quad n$=617

Ecuación para el hombre $\qquad \widehat{\ln(wage)} = \underset{(0.031)}{1.739} + \underset{(0.0032)}{0.054}\,educ$

$SCR$=289 $\quad R^2$=0.175 $\quad n$=1383

El modelo restringido, que se estima en el ejemplo 5.4, tiene la misma configuración que las ecuaciones (5-34), pero referido en este caso para toda la muestra. Por lo tanto, es la regresión agrupada ($P$) correspondiente al modelo restringido. El estadístico $F$ toma el valor

$$F = \frac{\left[SCR_P - (SCR_F + SCR_M)\right]/k}{SCR_F + SCR_M)/(n-2k)} = \frac{\left[433-(104+289)\right]/2}{(104+289)/(2000-2\times 2)} = 102$$

El estadístico $F$ tiene que ser, y lo es, igual al del ejemplo 5.12. En consecuencia, las conclusiones son las mismas.

***EJEMPLO 5.16 ¿El modelo de determinación de los salarios es el mismo para diferentes tamaños de empresa?***

En otros ejemplos bien el término independiente o bien la pendiente correspondiente a la variable educación, fue diferente para tres diferentes tamaños de empresa (pequeña, mediana y grande). Ahora consideramos una ecuación completamente diferente para cada tamaño de la empresa. Por lo tanto, el modelo no restringido estará compuesto por tres ecuaciones:

$$\begin{aligned} pequeña &: \ln(wage) = \beta_{11} + \delta_{11} female + \beta_{21} edu + u \\ mediana &: \ln(wage) = \beta_{12} + \delta_{12} female + \beta_{22} edu + u \\ grande &: \ln(wage) = \beta_{13} + \delta_{13} female + \beta_{23} edu + u \end{aligned} \qquad (5\text{-}37)$$

Las hipótesis nula y alternativa serán las siguientes:

$$H_0 : \begin{cases} \beta_{11} = \beta_{12} = \beta_{13} \\ \delta_{11} = \delta_{12} = \delta_{13} \\ \beta_{21} = \beta_{22} = \beta_{23} \end{cases}$$

$$H_1 : \text{No } H_0$$

Dada esta hipótesis nula, el modelo restringido es el modelo de (5-2).

Las estimaciones de las tres ecuaciones de (5-37), utilizando el fichero *wage02sp*, son las siguientes:

*pequeña* $\qquad \widehat{\ln(wage)} = \underset{(0.0346)}{1.706} - \underset{(0.0312)}{0.249}\,female + \underset{(0.0038)}{0.040}\,educ$

$SCR$=121 $\quad R^2$=0.160 $\quad n$=801

*mediana* $\qquad \widehat{\ln(wage)} = \underset{(0.0514)}{1.934} - \underset{(0.0390)}{0.422}\,female + \underset{(0.0046)}{0.055}\,educ$

$SCR$ =123 $\quad R^2$=0.302 $\quad n$=590

*grande* $\qquad \widehat{\ln(wage)} = \underset{(0.0462)}{1.749} - \underset{(0.0385)}{0.303}\,female + \underset{(0.0044)}{0.055}\,educ$

$SCR$ =114 $\quad R^2$=0.273 $\quad n$=609

La regresión agrupada ($P$) ya ha sido estimada en el ejemplo 5.1. El estadístico $F$ toma el valor

$$F=\frac{\left[SCR_P-(SCR_S+SCR_M+SCR_L)\right]/2\times k}{(SCR_S+SCR_M+SCR_L)/(n-3k)}$$
$$=\frac{\left[393-(121+123+114)\right]/6}{(121+123+114)/(2000-3\times 3)}=32.4$$

Para cualquier nivel de significación, rechazamos que las ecuaciones para la determinación de los salarios sean las mismas para los tres tamaños de empresa considerados.

***EJEMPLO 5.17 ¿Es el modelo Pinkham válido para los cuatro periodos?***

En el ejemplo 5.5 se introdujeron variables ficticias temporales y se contrastó si el término independiente era diferente para cada periodo. Ahora, vamos a contrastar si el modelo en su conjunto es válido para los cuatro periodos considerados. Por lo tanto, el modelo restringido estará compuesto por cuatro ecuaciones:

$$\begin{array}{ll} 1907\text{-}1914 & sales_t=\beta_{11}+\beta_{21}advexp_t+\beta_{31}\,sales_{t-1}+u_t \\ 1915\text{-}1925 & sales_t=\beta_{12}+\beta_{22}advexp_t+\beta_{32}\,sales_{t-1}+u_t \\ 1926\text{-}1940 & sales_t=\beta_{13}+\beta_{23}advexp_t+\beta_{33}\,sales_{t-1}+u_t \\ 1941\text{-}1960 & sales_t=\beta_{14}+\beta_{24}advexp_t+\beta_{34}\,sales_{t-1}+u_t \end{array} \tag{5-38}$$

Las hipótesis nula y alternativa serán las siguientes:

$$H_0:\begin{cases}\beta_{11}=\beta_{12}=\beta_{13}=\beta_{14}\\ \beta_{21}=\beta_{22}=\beta_{23}=\beta_{24}\\ \beta_{31}=\beta_{32}=\beta_{33}=\beta_{34}\end{cases}$$
$$H_1:\text{No } H_0$$

Dada esta hipótesis nula, el modelo restringido es el siguiente

$$sales_t=\beta_1+\beta_2advexp_t+\beta_3\,sales_{t-1}+u_t \tag{5-39}$$

Las estimaciones de las cuatro ecuaciones (5-38) son las siguientes:

$$1907\text{-}1914\quad \widehat{sales_t}=\underset{(603)}{64.84}+\underset{(1.025)}{0.9149}\,advexp+\underset{(0.425)}{0.4630}\,sales_{t-1}\qquad SCR=36017\quad n=7$$

$$1915\text{-}1925\quad \widehat{sales_t}=\underset{(190)}{221.5}+\underset{(0.557)}{0.1279}\,advexp+\underset{(0.425)}{0.9319}\,sales_{t-1}\qquad SCR=400605\quad n=11$$

$$1926\text{-}1940\quad \widehat{sales_t}=\underset{(112)}{446.8}+\underset{(0.115)}{0.4638}\,advexp+\underset{(0.0827)}{0.4445}\,sales_{t-1}\qquad SCR=201614\quad n=15$$

$$1941\text{-}1960\quad \widehat{sales_t}=\underset{(134)}{-182.4}+\underset{(0.241)}{1.6753}\,advexp+\underset{(0.111)}{0.3042}\,sales_{t-1}\qquad SCR=187332\quad n=20$$

La regresión agrupada, estimada en el ejemplo 3.4, es la siguiente:

$$\widehat{sales_t}=\underset{(95.7)}{138.7}+\underset{(0.156)}{0.3288}\,advexp+\underset{(0.0915)}{0.7593}\,sales_{t-1}\qquad SCR=2527215\quad n=53$$

El estadístico $F$ toma el valor

$$F=\frac{\left[SCR_P-(SCR_1+SCR_2+SCR_3+SCR_4)\right]/3\times k}{(SCR_1+SCR_2+SCR_3+SCR_4)/(n-4k)}$$
$$=\frac{\left[2527215-(36017+400605+201614+187332)\right]/9}{(36017+400605+201614+187332)/(53-4\times 3)}=9.16$$

Para cualquier nivel de significación, rechazamos que el modelo (5-39) sea el mismo para los cuatro periodos considerados.

# Ejercicios

**Ejercicio 5.1** Responda a las dos siguientes cuestiones relativas a un modelo con variables explicativas ficticias:

*a*) ¿Cuál es la interpretación de los coeficientes de las variables ficticias?

*b*) ¿Por qué no se deben incluir el mismo número de variables ficticias que categorías?

**Ejercicio 5.2** Se han obtenido las siguientes estimaciones de demanda de viviendas para alquiler con una muestra de 560 familias.

$$\hat{q}_i = \underset{(0.11)}{4.17} - \underset{(0.017)}{0.247}\, p_i + \underset{(0.026)}{0.960}\, y_i$$

$$R^2 = 0.371 \qquad n = 560$$

$$\hat{q}_i = \underset{(0.13)}{5.27} - \underset{(0.030)}{0.221}\, p_i + \underset{(0.031)}{0.920}\, y_i + \underset{(0.120)}{0.341}\, d_i y_i$$

$$R^2 = 0.380$$

donde $q_i$ es el logaritmo del gasto en alquiler de vivienda de la familia *i*-esima, $p_i$ es el logaritmo del precio de alquiler por $m^2$ en el área que vive la familia *i*-esima, $y_i$ es el logaritmo de la renta familiar disponible *i*-esima y $d_i$ es una variable ficticia que toma el valor uno si la familia reside en un municipio urbano y cero en uno rural.

(Los números entre paréntesis son los errores estándar de los estimadores.)

*a)* Contraste, en el primer modelo ajustado, la hipótesis de que la elasticidad del gasto en alquiler de vivienda con respecto a la renta es 1.

*b)* Contraste si la interacción entre la variable ficticia y la renta es significativa. ¿Existe una diferencia significativa de la elasticidad gasto en alquiler renta entre las áreas rurales y urbanas?

**Ejercicio 5.3** En un modelo de regresión lineal con variables ficticias conteste a las siguientes preguntas:

*a)* Significado e interpretación de los coeficientes de las variables ficticias en modelos con distintas formas funcionales de la variable endógena.

*b)* ¿Por qué no es conveniente incluir el mismo número de ficticias que de categorías existentes en la variable cualitativa?

*c)* Exprese cómo se ve afectado un modelo en el que se han introducido variables ficticias en forma aditiva y otro en el que sólo se introducen en forma multiplicativa con respecto a una variable cuantitativa.

**Ejercicio 5.4** En el contexto del modelo de regresión lineal múltiple,

*a)* ¿Qué es una variable ficticia? Ponga un ejemplo de especificación de un modelo econométrico con variables ficticias. Interprete los coeficientes, razonando la respuesta.

*b)* ¿Qué relación puede existir entre el problema de multicolinealidad y las variables ficticias?

**Ejercicio 5.5** Con datos correspondientes a los trabajadores de un departamento de una cierta empresa se ha obtenido la siguiente estimación:

$$\widehat{salario_i} = 500 + 50 antigüedad_i + 200 niveldeestudios_i + 100 hombre_i$$

donde *salario* es el salario en euros mensuales, *antigüedad* es la antigüedad laboral medida en años, *nivelestudios* es una variable ficticia que toma valor 1 si el trabajador tiene estudios superiores y 0 en caso contrario y hombre es una variable ficticia que toma el valor 1 si el trabajador es hombre y 0 en caso contrario.

*a)* ¿Qué salario predeciría para una trabajadora con 6 años de antigüedad laboral y con estudios superiores?

*b)* Suponiendo que todas las mujeres trabajadoras tienen estudios superiores y ninguno de los hombres trabajadores tienen estudios superiores, escriba una hipotética matriz de regresores (**X**) para seis observaciones. En este caso, ¿se plantearía algún problema en la estimación del modelo? Explique su respuesta.

*c)* Plantee un nuevo modelo econométrico que permita dilucidar si existen diferencias salariales entre los trabajadores con estudios primarios, con estudios medios y con estudios superiores.

**Ejercicio 5.6** Considere el siguiente modelo de regresión lineal:

$$y_i = \alpha + \beta x_i + \gamma_1 d_{1i} + \gamma_2 d_{2i} + u_i \qquad (1)$$

donde $y$ es el salario mensual de un profesor, $x$ es el número de años de experiencia docente y $d_1$ y $d_2$ son dos variables ficticias que toman los siguientes valores

$$d_{1i} = \begin{cases} 1 \text{ si el profesor es hombre} \\ 0 \text{ en todos los demás casos} \end{cases} \qquad d_{2i} = \begin{cases} 1 \text{ si el profesor es de raza blanca} \\ 0 \text{ en todos los demás casos} \end{cases}$$

*a)* ¿Cuál es la categoría de referencia en el modelo?

*b)* Interprete el significado de $\gamma_1$ y $\gamma_2$. ¿Cuál es el salario esperado para todas las categorías posibles?

*c)* Para mejorar la capacidad explicativa del modelo se consideró la siguiente especificación alternativa

$$y_i = \alpha + \beta x_i + \gamma_1 d_{1i} + \gamma_2 d_{2i} + \gamma_3 (d_{1i} d_{2i}) + u_i \qquad (2)$$

*d)* ¿Cuál es el significado del término $(d_{1i} d_{2i})$? Interprete el significado de $\gamma_3$.

*e)* ¿Cuál es el salario esperado para todas las categorías posibles en el modelo (2)?

**Ejercicio 5.7** Se ha obtenido la siguiente ecuación estimada por mínimos cuadrados ordinarios con una muestra de 36 observaciones:

$$\hat{y}_t = \underset{(0.12)}{1.10} - \underset{(0.34)}{0.96} x_{t1} - \underset{(3.35)}{4.56} x_{t2} + \underset{(0.07)}{0.34} x_{t3}$$

$$\sum_{t=1}^{n} \left( \hat{y}_t - \bar{y} \right)^2 = 109.24 \quad \sum_{t=1}^{n} \hat{u}_t^2 = 20.22$$

(Los números entre paréntesis son los errores estándar de los estimadores.)

*a)* Contraste la significatividad individual del coeficiente asociado a $x_2$.

*b)* Calcule el coeficiente de determinación, $R^2$, y dé una interpretación del mismo.

*c)* Contraste la significatividad conjunta del modelo.

*d)* Dos regresiones adicionales, con la misma especificación, fueron realizadas para los dos grupos, A y B, incluidos en la muestra ($n_1$=21 y $n_2$=15). En dichas estimaciones se obtuvieron las siguientes *SCR*, 11.09 y 2.17, respectivamente. Contraste si los grupos A y B tienen un distinto comportamiento.

**Ejercicio 5.8** Para explicar el tiempo dedicado a actividades deportivas (*depor*) se ha formulado el siguiente modelo:

$$depor = \beta_1 + \delta_1 mujer + \varphi_1 fumador + \beta_2 edad + u \qquad (1)$$

donde *depor* son los minutos dedicados al día, en promedio, a actividades deportivas en minutos; *mujer* y *fumador* son variables ficticias que toman el valor 1 si la persona es una mujer o si fuma al menos 5 cigarrillos diarios, respectivamente. La variable *edad* está expresada en años.

*a)* Interprete el significado de $\delta_1$, $\varphi_1$ y $\beta_2$.

*b)* ¿Cuál es el tiempo esperado dedicado a actividades deportivas para todas las categorías posibles?

*c)* Para mejorar la capacidad explicativa del modelo se consideró la siguiente especificación alternativa:

$$\begin{aligned} depor &= \beta_1 + \delta_1 mujer + \varphi_1 fumador + \gamma_1 mujer \times fumador \\ &+ \delta_2 mujer \times edad + \varphi_2 fumador \times edad + \beta_2 edad + u \end{aligned} \qquad (2)$$

En el modelo (2), ¿cuál es el significado de $\gamma_1$?¿Cuál es el significado de $\delta_2$ y $\varphi_2$?

*d)* ¿Cuáles son los posibles efectos marginales de *depor* con respecto a la *edad* en el modelo (2)? Detállelos.

**Ejercicio 5.9** Utilizando información de las regiones españolas en los años 1995 y 2000 se han estimado varias funciones de producción.

Para el conjunto de los dos periodos se obtuvieron los siguientes resultados

$$\widehat{\ln(q)} = 5.72 + 0.26\ln(k) + 0.75\ln(l) - 1.14f + 0.11f \times \ln(k) - 0.05f \times \ln(l) \qquad (1)$$

$$R^2 = 0.9594 \quad \bar{R}^2 = 0.9510 \quad SCR = 0.9380 \quad n = 34$$

$$\widehat{\ln(q)} = 3.91 + 0.45\ln(k) + 0.60(l) \qquad (2)$$

$$R^2 = 0.9567 \quad \bar{R}^2 = 0.9525 \quad SCR = 1.0007$$

Por otra parte, para cada uno de los años se estimaron separadamente los siguientes modelos:

1995 $$\widehat{\ln(q)} = 5.72 + 0.26\ln(k) + 0.75l \qquad (3)$$

$$R^2 = 0.9527 \quad \bar{R}^2 = 0.9459 \quad SCR = 0.6052$$

2000 $$\widehat{\ln(q)} = 4.58 + 0.37\ln(k) + 0.70l \qquad (4)$$

$$R^2 = 0.9629 \quad \bar{R}^2 = 0.9555 \quad SCR = 0.3331$$

donde $q$ es producción, $k$ es capital, $l$ es empleo y $f$ es una variable ficticia que toma el valor 1 para los datos de 1995 y 0 para los del año 2000.

*a*) Contraste si se produce un cambio estructural entre 1995 y 2000.

*b*) Compare los resultados de las estimaciones (3) y (4) con la estimación (1).

*c)* Contraste la significatividad global del modelo (1).

**Ejercicio 5.10** Con una muestra de 300 empresas del sector de servicios, se estimó la siguiente función de coste (*cost*):

$$\widehat{cost_i} = 0.847 + \underset{(0.025)}{0.899} qty_i \quad SCR = 901.074 \quad n = 300$$

donde $qty_i$ es la cantidad producida.

Las 300 empresas están distribuidas en tres grandes áreas (100 en cada una). Los resultados obtenidos fueron los siguientes:

$$\text{Área 1: } \widehat{cost}_i = 1.053 + \underset{(0.038)}{0.876}\, qty_i \qquad \hat{\sigma}^2 = 0.457$$

$$\text{Área 2: } \widehat{cost}_i = 3.279 + \underset{(0.096)}{0.835}\, qty_i \qquad \hat{\sigma}^2 = 3.154$$

$$\text{Área 3: } \widehat{cost}_i = 5.279 + \underset{(0.10)}{0.984}\, qty_i \qquad \hat{\sigma}^2 = 4.255$$

*a)* Calcule una estimación insesgada $\sigma^2$ de la función de costes para el conjunto de las 300 empresas.

*b)* ¿Es la misma función de coste válida para las tres áreas?

**Ejercicio 5.11** Para el estudio del gasto en revistas (*rev*) se han formulado los siguientes modelos:

$$\ln(rev) = \beta_1 + \beta_2 \ln(renta) + \beta_3 edad + \beta_4 hombre + u \qquad (1)$$

$$\ln(rev) = \beta_1 + \beta_2 \ln(renta) + \beta_3 edad + \beta_4 hombre + \beta_5 prim + \beta_6 sec + u \qquad (2)$$

donde *renta* es la renta disponible, *edad* es la edad en años, *hombre* es una variable dicotómica que toma el valor 1 si es hombre, *prim* y *sec* son variables ficticias que toman el valor 1 cuando el individuo ha alcanzado, a lo sumo, los niveles primarios y secundarios de estudios, respectivamente.

Con una muestra de 100 observaciones, se han obtenido los siguientes resultados

$$\widehat{\ln(rev)_i} = \underset{(0.124)}{1.27} + \underset{(0040)}{0.756} \ln(renta_i) + \underset{(0.001)}{0.031} edad_i - \underset{(0.022)}{0.017}\, hombre_i$$

$$SCR{=}1.1575 \qquad R^2{=}0.9286$$

$$\widehat{\ln(rev_i)} = \underset{(0.020)}{1.26} + \underset{(0.007)}{0.811} \ln(renta_i) + \underset{(0.0002)}{0.030}\, edad_i + \underset{(0.003)}{0.003} hombre_i$$

$$-\underset{(0.004)}{0.250}\, prim_i + \underset{(0.005)}{0.108} sec_i$$

$$SCR{=}0.0306 \qquad R^2{=}0.9981$$

*a)* ¿Es la educación un factor relevante para explicar el gasto en revistas? ¿Cuál es la categoría de referencia para la educación?

*b)* En el primer modelo, ¿es mayor el gasto en revistas para hombres que para mujeres? Justifica tu respuesta.

*c)* Interprete el coeficiente de la variable *hombre* en el segundo modelo. ¿Es mayor el gasto en revistas para hombres que para mujeres? Compare con el resultado obtenido en la parte *a)*.

**Ejercicio 5.12** Consideremos que *fruit* es el gasto en frutas en un año, expresado en euros, realizado por un hogar y $r_1$, $r_2$, $r_3$, y $r_4$ son variables dicotómicas que reflejan las cuatro regiones de un país.

*a)* Si se realiza una regresión de *fruit* sobre $r_1$, $r_2$, $r_3$, y $r_4$ sin término independiente, ¿cuál es la interpretación de los coeficientes?

*b)* Si se realiza una regresión de *fruit* sobre $r_1$, $r_2$, $r_3$, y $r_4$ y con un término independiente, ¿qué sucedería? ¿Por qué?

*c)* Si se realiza una regresión de *fruit* sobre $r_2$, $r_3$, y $r_4$ sin término independiente, ¿cuál es la interpretación de los coeficientes?

*d)* Si se realiza una regresión de *fruit* sobre $r_1$- $r_2$, $r_2$, $r_4$ - $r_3$, $r_4$ sin término independiente, ¿cuál es la interpretación de los coeficientes?

**Ejercicio 5.13** Considere el siguiente modelo

$$salario = \beta_1 + \delta_1 mujer + \beta_2 educ + u$$

Ahora, vamos a considerar tres posibilidades de definir la variable ficticia *female*.

$$1)\ mujer = \begin{cases} 1 & \text{para mujer} \\ 0 & \text{para hombre} \end{cases} \quad 2)\ mujer = \begin{cases} 2 & \text{para mujer} \\ 1 & \text{para hombre} \end{cases} \quad 3)\ mujer = \begin{cases} 2 & \text{para mujer} \\ 0 & \text{para hombre} \end{cases}$$

*a)* Interprete el coeficiente de la variable ficticia para cada definición.
*b)* ¿Es alguna definición preferible a las otras? Justifique la respuesta.

**Ejercicio 5.14** Se considera el siguiente modelo de regresión:

$$salario = \beta_1 + \delta_1 mujer + u$$

donde *mujer* es una variable dicotómica que toma el valor 1 para las mujeres y el valor 0 para los hombres.

Demuestre que aplicando las fórmulas de *MCO* para la regresión simple se obtiene que

$$\hat{\beta}_1 = \overline{wage_H}$$
$$\hat{\delta}_1 = \overline{wage_M} - \overline{wage_H}$$

donde *M* indica mujer y *H* hombre.

Con el fin de facilitar la obtención de la solución, considere que en la muestra hay $n_1$ mujeres y $n_2$ hombres: la muestra total es $n= n_1+n_2$.

**Ejercicio 5.15** Los datos de este ejercicio se obtuvieron de un experimento de marketing controlado en las tiendas en París sobre el gasto en café, publicado por CA Bemmaor y Mouchoux D., “Measuring the Short-Term Effect of In-Store Promotion and Retail Advertising on Brand Sales: A Factorial Experiment”, *Journal of Marketing Research*, 28 (1991), 202-14. En este experimento se formuló el siguiente modelo para explicar la cantidad vendida de café por semana:

$$\ln(coffqty) = \beta_1 + \delta_1 advert + \beta_2 \ln(coffpric) + \delta_2 advert \times \ln(coffpric) + u$$

donde *coffpric* toma tres valores: 1, que es el precio habitual, 0.95 y 0.85; *advert* es una variable dicotómica que toma valor 1 si se hace publicidad en esa semana, y 0 si no se hace. El experimento duró 18 semanas. El modelo original y otros tres modelos más fueron estimados, utilizando el fichero *coffee2*:

1)
$$\widehat{\ln(coffqty_i)} = \underset{(0.04)}{5.85} + \underset{(0.099)}{0.2565}\, advert_i - \underset{(0.450)}{3.9760} \ln(coffpric_i) - \underset{(0.883)}{1.069}\, advert_i \times \ln(coffpric_i)$$
$$R^2 = 0.9468 \qquad n = 18$$

2)
$$\widehat{\ln(coffqty_i)} = \underset{(0.04)}{5.83} + \underset{(0.057)}{0.3559}\, advert_i - \underset{(0.393)}{4.2539} \ln(coffpric_i)$$
$$R^2 = 0.9412 \qquad n = 18$$

3)
$$\widehat{\ln(coffqty_i)} = \underset{(0.04)}{5.88} - \underset{(0.513)}{3.6939} \ln(coffpric_i) - \underset{(0.582)}{2.9575}\, advert_i \times \ln(coffpric_i)$$
$$R^2 = 0.9214 \qquad n = 18$$

4)
$$\widehat{\ln(coffqty_i)} = \underset{(0.07)}{5.89} - \underset{(0.674)}{5.1727} \ln(coffpric)$$
$$R^2 = 0.7863 \qquad n = 18$$

*a)* En el modelo (2), ¿cuál es la interpretación del coeficiente de *advert*?
*b)* En el modelo (3), ¿cuál es la interpretación del coeficiente de *advert*×ln*(coffpric*?
*c)* En el modelo (2), ¿tiene el coeficiente de *advert* un efecto positivo significativo al 5% y al 1%?
*d)* ¿Es el modelo (4) válido para semanas con publicidad y para semanas sin publicidad?
*e)* En el modelo (1), ¿es el término independiente el mismo para semanas con publicidad y para semanas sin publicidad?
*f)* En el modelo (3), ¿es la elasticidad de demanda de café/precio diferente en semanas con publicidad y en semanas sin publicidad?
*g)* En el modelo (4), ¿es la elasticidad de demanda de café/precio inferior a -4?

**Ejercicio 5.16** (Continuación del ejercicio 4.39). Utilizando el fichero *timuse03*, se han estimado los siguientes modelos:

$$\widehat{houswork_i} = \underset{(23)}{132} + \underset{(1.497)}{2.787}\,educ_i + \underset{(0.308)}{1.847}\,age_i - \underset{(0.023)}{0.2337}\,paidwork_i \qquad (1)$$

$$R^2 = 0.142 \quad n = 1000$$

$$\widehat{houswork_i} = -\underset{(22.29)}{3.02} + \underset{(1.356)}{3.641}\,educ_i + \underset{(0.279)}{1.775}\,age_i - \underset{(0.021)}{0.1568}\,paidwork_i + \underset{(2.16)}{32.11}\,female_i \qquad (2)$$

$$R^2 = 0.298 \quad n = 1000$$

$$\widehat{houswork_i} = -\underset{(35.18)}{8.04} + \underset{(2.352)}{4.847}\,educ_i + \underset{(0.502)}{1.333}\,age_i - \underset{(0.032)}{0.0871}\,paidwork_i + \underset{(8.15)}{32.75}\,female_i$$
$$-\underset{(0.546)}{0.1650}\,educ_i \times female_i + \underset{(0.112)}{0.1019}\,age_i \times female_i - \underset{(0.009)}{0.02625}\,paidwork_i \times female_i \qquad (3)$$

$$R^2 = 0.306 \quad n = 1000$$

*a)* En el modelo (1), ¿existe una compensación estadísticamente significativa entre el tiempo dedicado a trabajo remunerado y el tiempo dedicado a trabajo doméstico?
*b)* Manteniendo igual todos los demás factores y tomando como modelo de referencia al (2), ¿existe evidencia de que las mujeres dedican más tiempo al trabajo doméstico que los hombres?
*c)* Compare el $R^2$ de los modelos (1) y (2). ¿Cuál es su conclusión?
*d)* En el modelo (3), ¿cuál es el efecto marginal del tiempo dedicado al trabajo doméstico con respecto al tiempo dedicado al trabajo remunerado?
*e)* ¿Es significativa la interacción entre *paidwork* y género?
*f)* ¿Son las interacciones entre género y las variables cuantitativas del modelo conjuntamente significativas?

**Ejercicio 5.17** Utilizando datos de la Bolsa de Madrid del 19 de noviembre de 2011 (fichero *bolmad11*), se han estimado los siguientes modelos:

$$\widehat{\ln(marktval_i)} = \underset{(0.243)}{1.784} + \underset{(0.179)}{0.6998}\,ibex35_i + \underset{(0.0369)}{0.6749}\ln(bookval_i) \qquad (1)$$

$$SCR=35.69 \quad R^2=0.8931 \quad n=92$$

$$\widehat{\ln(marktval_i)} = \underset{(0.275)}{1.828} + \underset{(0.778)}{0.4236}\,ibex35_i + \underset{(0.0423)}{0.6678}\ln(bookval_i)$$
$$+\underset{(0.088)}{0.0310}\,ibex35_i \times \ln(bookval_i) \qquad (2)$$

$$SCR=35.622 \quad R^2=0.8933 \quad n=92$$

$$\widehat{\ln(marktval_i)} = \underset{(0.310)}{2.323} + \underset{(0.785)}{0.1987}\, ibex35_i + \underset{(0.0405)}{0.6688} \ln(bookval_i)$$

$$+ \underset{(0.089)}{0.0369}\, ibex35_i \times \ln(bookval_i) - \underset{(0.236)}{0.6613}\, services_i - \underset{(0.221)}{0.6698}\, consump_i \qquad (3)$$

$$- \underset{(0.263)}{0.1931} energy_i - \underset{(0.207)}{0.3895}\, industry_i - \underset{(0.324)}{0.7020}\, itt_i$$

$$SCR=30.781 \quad R^2=0.9078 \quad n=92$$

$$\widehat{\ln(marktval_i)} = \underset{(0.234)}{1.366} + \underset{(0.0305)}{0.7658} \ln(bookval_i) \qquad (4)$$

$$SCR=41.625 \quad R^2=0.8753 \quad n=92$$

$$\text{For } finance=1 \qquad \widehat{\ln(markval_i)} = \underset{(0.560)}{0.558} + \underset{(0.0702)}{0.9346} \ln(bookval_i) \qquad (5)$$

$$SCR=2.7241 \quad R^2=0.9415 \quad n=13$$

donde

- *marktval* es el valor de mercado de una compañía.
- *bookval* es el valor contable de una compañia.
- *ibex35* es una variable ficticia que toma el valor 1 si la compañía está incluida en el selectivo Ibex 35.
- *services*, *consump* (*consumo*), *energy*, *industry* e *itc* (tecnologías de la información y la comunicación) son variables ficticias. Cada uno de ellas toma el valor 1 si la compañía está clasificada en ese sector en la Bolsa de Madrid. La categoría de referencia es el sector financiero (*finance*).

  *a)* En el modelo (1), ¿cuál es la interpretación del coeficiente de *ibex35*?
  *b)* En el modelo (1), ¿es la elasticidad *marktval/bookval* igual a 1?
  *c)* En el modelo (2), ¿es la elasticidad *marktval/bookval* la misma para todas las compañías incluidas en la muestra?
  *d)* ¿Es el modelo (4) válido tanto para las compañías incluidas en el Ibex 35 y para las compañías excluidas?
  *e)* En el modelo (3), ¿cuál es la interpretación del coeficiente de *consump*?
  *f)* ¿Es el coeficiente de *consump* significativamente negativo??
  *g)* ¿Está justificada estadísticamente la introducción de variables ficticias para los diferentes sectores?
  *h)* ¿Es la elasticidad *marktval/bookval* para el sector financiero igual a 1?

**Ejercicio 5.18** (Continuación del ejercicio 4.37). Utilizando el fichero *rdspain*, se han estimado las ecuaciones que aparecen en el cuadro adjunto

Las variables que aparecen en el cuadro son las siguientes:

- *rdintens* es el gasto en investigación y desarrollo (I+D) medido como porcentaje de las ventas,
- *sales*, ventas medidas en millones de euros,
- *exponsal* son las exportaciones medidas como porcentaje de las ventas,
- *medtech* e *hightech* son dos variables ficticias que reflejan si la empresa pertenece a un sector de media o de alta tecnología. La categoría de referencia corresponde a las empresas de baja tecnología.
- *workers* es el número de trabajadores de la empresa.

| | (1) *rdintens* | (2) *rdintens* | (3) *rdintens* | (4) *rdintens* para *hightech*=1 | (5) *rdintens* para *medtech*=1 | (6) *rdintens* para *lowtech*=1 |
|---|---|---|---|---|---|---|
| *exponsal* | 0.0136<br>(0.00195) | 0.0101<br>(0.00193) | 0.00968<br>(0.00189) | 0.00584<br>(0.00792) | 0.0116<br>(0.00300) | 0.00977<br>(0.00169) |
| *workers* | 0.000433<br>(0.0000740) | 0.000392<br>(0.0000725) | 0.000394<br>(0.000208) | 0.00196<br>(0.000338) | 0.0000563<br>(0.0000815) | 0.000393<br>(0.000121) |
| *hightech* | | 1.448<br>(0.141) | 0.976<br>(0.151) | | | |
| *medtech* | | 0.361<br>(0.109) | 0.472<br>(0.112) | | | |
| *hightech*× *workers* | | | 0.00153<br>(0.000271) | | | |
| *medtech*× *workers* | | | -0.000326<br>(0.000222) | | | |
| *término independiente* | 0.394<br>(0.0598) | 0.137<br>(0.0691) | 0.143<br>(0.0722) | 1.211<br>(0.313) | 0.577<br>(0.103) | 0.142<br>(0.0443) |
| *n* | 1983 | 1983 | 1983 | 296 | 616 | 1071 |
| $R^2$ | 0.0507 | 0.0986 | 0.138 | 0.113 | 0.0278 | 0.0459 |
| *SCR* | 9282.7 | 8815.0 | 8425.3 | 4409.0 | 2483.6 | 1527.5 |
| *F* | 52.90 | 54.06 | 52.90 | 18.71 | 8.776 | 25.72 |
| *df_n* | 2 | 4 | 6 | 2 | 2 | 2 |
| *df_d* | 1980 | 1978 | 1976 | 293 | 613 | 1068 |

Errores estándar entre paréntesis

*a)* En el modelo (2), manteniéndose igual todos los demás factores, ¿hay evidencia de que el gasto en investigación y desarrollo (expresado como un porcentaje de las ventas) en empresas de alta tecnología sea mayor que en empresas de baja tecnología? ¿Es fuerte la evidencia empírica?

*b)* En el modelo (2), manteniéndose igual todos los demás factores, ¿hay evidencia de que el gasto en I+D, *rdintens*, en las empresas de tecnología media sea igual al de empresas de baja tecnología? ¿Es fuerte la evidencia empírica?

*c)* Tomando como modelo de referencia (2), si tuviera que contrastar la hipótesis de que *rdintens* en las empresas de alta tecnología es igual a las empresas de tecnología media, formule un modelo que le permita contrastar esta hipótesis sin necesidad de utilizar la información sobre la matriz de covarianzas de los estimadores.

*d)* ¿Hay influencia de los trabajadores asociados en *rdintens* con el nivel de tecnología en las empresas?

*e)* ¿Es el modelo (1) válido para todas las empresas independientemente de su nivel tecnológico?

**Ejercicio 5.19** Para explicar la satisfacción general de las personas (*stsfglo*) se estimaron los siguientes modelos utilizando datos del fichero *hdr2010*:

$$\widehat{stsfglo_i} = \underset{(0.584)}{-0.375} + \underset{(0.00000617)}{0.0000207}\, gnipc_i + \underset{(0.009)}{0.0858}\, lifexpec_i \qquad (1)$$
$$R^2 = 0.642 \quad n = 144$$

$$\widehat{stsfglo_i} = \underset{(0.897)}{2.911} + \underset{(0.00000572)}{0.0000381}\, gnipc_i + \underset{(0.18)}{1.215}\, lifexpec_i$$
$$+ \underset{(0.179)}{1.215}\, dlatam_i - \underset{(0.259)}{0.7901}\, dafrica_i \qquad (2)$$
$$R^2 = 0.748 \quad n = 144$$

$$\widehat{stsfglo_i} = \underset{(1.146)}{0.6984} + \underset{(0.000006)}{0.0000198}\, gnipc_i + \underset{(0.0164)}{0.0724}\, lifexpec_i + \underset{(1.950)}{4.099}\, dafrica_i$$
$$+ \underset{(0.000052)}{0.0000801}\, gnipc_i \times dafrica_i - \underset{(0.0336)}{0.0896}\, lifexpec_i \times dafrica_i \qquad (3)$$
$$R^2 = 0.6840 \quad n = 144$$

donde

- *gnipc* es el producto nacional bruto per cápita expresado en *PPA* (paridad de poder adquisitivo) en dólares americanos de 2008,
- *lifexpec* es la esperanza de vida al nacer, es decir, el número de años que un recién nacido puede esperar vivir,
- *dafrica* es una variable dicotómica que toma el valor 1 si el país se encuentra en África,
- *dlatam* es una variable dicotómica que toma el valor 1 si el país está en América Latina.

  *a)* En el modelo (2), ¿cuál es la interpretación de los coeficientes de *dlatam* y *dafrica*?

  *b)* En el modelo (2), *dlatam* y *dafrica*, individualmente, ¿tienen una influencia significativamente positiva sobre la satisfacción global?

  *c)* En el modelo (2), *dlatam* y *dafrica* ¿tienen una influencia conjunta sobre la satisfacción global?

  *d)* ¿Es la influencia de la esperanza de vida sobre la satisfacción global menor en África que en otras regiones del mundo?

  *e)* ¿Es la influencia de la variable de *gnipc* mayor en África que en otras regiones del mundo en un 10%?

  *f)* ¿Son las interacciones de las personas que viven en África y las variables *gnipc* y *lifexpec* conjuntamente significativas?

**Ejercicio 5.20** Las ecuaciones que aparecen en el cuadro adjunto se han estimado utilizando los datos del fichero *timuse03*. Este archivo contiene 1000 observaciones correspondientes a una submuestra aleatoria extraída de la encuesta de uso del tiempo en España que se llevó a cabo en el periodo 2002-2003.

Las variables que aparecen en el cuadro son:

- *educ* son los años de educación alcanzados,
- *sleep* (dormir), *paidwork* (trabajo remunerado) and *unpaidwrk* (trabajo no remunerado se miden en minutos por día,
- *female* (mujer), *workday* (*lunes a viernes*), *spaniard* (español) y *houswife* (ama de casa) son variables ficticias.

*a)* En el modelo (1), ¿existe una compensación estadísticamente significativa entre el tiempo dedicado al trabajo remunerado y el tiempo dedicado a dormir?

*b)* En el modelo (1), ¿es el coeficiente de *unpaidwk* estadísticamente significativo?

*c)* ¿Existe evidencia de que las mujeres duermen más que los hombres?

*d)* En el modelo (2), ¿son *workday* y *spaniard* individualmente significativas? ¿Son conjuntamente significativas?

*e)* ¿Es el coeficiente de *housewife* estadísticamente significativo?

*f)* ¿Son las interacciones de *female* con *educ*, *paidwork* y *unpaidwk* conjuntamente significativas?

| | (1) Sleep | (2) Sleep | (3) Sleep | (4) Sleep | (5) Sleep | (6) sleep |
|---|---|---|---|---|---|---|
| *educ* | -4.669 | -4.787 | -4.805 | -4.754 | -4.782 | -4.792 |
| | (0.916) | (0.912) | (0.912) | (0.913) | (0.917) | (0.917) |
| *persinc* | 0.0238 | 0.0207 | 0.0195 | 0.0210 | 0.0208 | 0.0208 |
| | (0.00587) | (0.00600) | (0.00607) | (0.00601) | (0.00601) | (0.00601) |
| *age* | 0.854 | 0.879 | 0.895 | 0.884 | 0.879 | 0.891 |
| | (0.174) | (0.174) | (0.174) | (0.174) | (0.174) | (0.302) |
| *paidwork* | -0.258 | -0.247 | -0.246 | -0.248 | -0.246 | -0.247 |
| | (0.0150) | (0.0159) | (0.0159) | (0.0160) | (0.0210) | (0.0159) |
| *unpaidwk* | -0.205 | -0.198 | -0.188 | -0.224 | -0.198 | -0.198 |
| | (0.0184) | (0.0184) | (0.0196) | (0.0365) | (0.0185) | (0.0184) |
| *female* | 4.161 | 3.588 | 3.981 | 2.485 | 3.638 | 3.727 |
| | (1.465) | (1.467) | (1.493) | (1.975) | (1.691) | (3.287) |
| *workday* | | -19.31 | -19.46 | -19.47 | -19.30 | -19.30 |
| | | (7.168) | (7.165) | (7.171) | (7.173) | (7.172) |
| *spaniard* | | -47.50 | -46.88 | -47.90 | -47.63 | -47.51 |
| | | (19.99) | (19.98) | (20.00) | (20.10) | (20.00) |
| *houswife* | | | -14.71 | | | |
| | | | (10.42) | | | |
| *unpaidwk ×female* | | | | 0.00607 | | |
| | | | | (0.00726) | | |
| *paidwork* ***×female*** | | | | | -0.000324 | |
| | | | | | **(0.00540)** | |
| *age× female* | | | | | | -0.00308 |
| | | | | | | (0.0652) |
| *término independiente* | 588.9 | 648.3 | 646.6 | 651.9 | 648.2 | 647.8 |
| | (13.62) | (24.34) | (24.36) | (24.73) | (24.39) | (26.40) |
| *N* | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 |
| $R^2$ | 0.316 | 0.325 | 0.326 | 0.325 | 0.325 | 0.325 |
| *SCR* | 9913901.3 | 9789312.3 | 9769648.2 | 9782424.0 | 9789276.9 | 9789290.3 |
| *F* | 76.58 | 59.62 | 53.27 | 53.06 | 52.95 | 52.95 |
| *df_n* | 6 | 8 | 9 | 9 | 9 | 9 |
| *df_d* | 993 | 991 | 990 | 990 | 990 | 990 |

Errores estándar entre paréntesis

**Ejercicio 5.21** Para el estudio de la mortalidad infantil en el mundo se han estimado los siguientes modelos a partir de los datos del fichero *hdr2010*:

$$\widehat{deathinf_i} = \underset{(4.58)}{93.02} - \underset{(0.0002)}{0.00037}\, gnipc_i - \underset{(0.1866)}{0.6046}\, physicn_i - \underset{(0.003)}{0.003} contrcep_i \qquad (1)$$

$$SCR{=}40285 \quad R^2{=}0.6598 \quad n{=}108$$

$$\widehat{deathinf_i} = \underset{(5.96)}{78.55} - \underset{(0.0002)}{0.00042}\,gnipc - \underset{(0.1879)}{0.3809}\,physicn_i - \underset{(0.1042)}{0.6989}\,contrcep_i + \underset{(5.05)}{17.92}\,dafrica \quad (2)$$

$$SCR=35893 \quad R^2=0.6851 \quad n=108$$

$$\widehat{deathinf_i} = \underset{(6.76)}{72.58} - \underset{(0.0002)}{0.00044}\,gnipc - \underset{(0.1879)}{0.3994}\,physicn_i - \underset{(0.1234)}{0.5857}\,contrcep_i$$
$$+\underset{(5.05)}{17.92}\,dafrica - \underset{(0.000826)}{0.0000914}\,gnipc \times dafrica - \underset{(2.2351)}{2.0013}\,physicn \times dafrica \quad (3)$$
$$-\underset{(0.2716)}{0.2172}\,contrcep_i \times dafrica$$

$$SCR=34309 \quad R^2=0.7109 \quad n=108$$

donde

- *deatinf* es el número de muertes infantiles (de una año o menos) por cada 1000 nacidos vivos en 2008,
- *gnipc* es el producto nacional bruto per cápita expresado en *PPA* en dólares americanos de 2008,
- *physicn* son los médicos por cada 10000 habitantes en el periodo 2000-2009,
- *contrcep* es la tasa de uso de anticonceptivos de cualquier tipo, expresada como % de las mujeres casadas de 15-49 años para el periodo 1990-2008,
- *dafrica* es una variable dicotómica que toma el valor 1 si el país se encuentra en África.

  *a)* En el modelo (1), ¿cuál es la interpretación de los coeficientes de *gnipc*, *physicn* y *contrcep*?
  *b)* En el modelo (2), ¿cuál es la interpretación del coeficiente de *dafrica*?
  *c)* En el modelo (2), manteniendo igual todos los demás factores, ¿tienen los países de África una mortalidad infantil mayor que los países de otras regiones del mundo?
  *d)* ¿Cuál es el efecto marginal de la variable *gnipc* sobre la mortalidad infantil en el modelo (3)?
  *e)* ¿Es la pendiente correspondiente al regresor *contrcep* significativamente mayor para los países de África?
  *f)* ¿Son las pendientes correspondientes a los regresores *gnipc*, *physicn* y *contrcep* conjuntamente diferentes para los países de África?
  *g)* ¿Es el modelo (1) válido para todos los países del mundo?

**Ejercicio 5.22** Utilizando una submuestra aleatoria de 2000 observaciones extraídas de las encuestas de uso del tiempo para España llevadas a cabo en los periodos 2002-2003 y 2009-2010 (fichero *timus309*), se han estimado los siguientes modelos para explicar el tiempo que se pasa viendo la televisión:

$$watchtv = \underset{(9.46)}{114} - \underset{(0.620)}{3.523}\,educ + \underset{(0.129)}{1.330}\,age - \underset{(0.010)}{0.1111}\,paidwork \quad (1)$$
$$R^2 = 0.169 \quad n = 2000$$

$$watchtv = \underset{(9.92)}{127} - \underset{(0.615)}{3.653}\,educ + \underset{(0.129)}{1.291}\,age - \underset{(0.010)}{0.120}\,paidwork - \underset{(4.903)}{25.15}\,female$$
$$+\underset{(5.25)}{17.14}\,y2009 \qquad R^2 = 0.184 \quad n = 2000 \quad (2)$$

$$watchtv = \underset{(10.01)}{123} - \underset{(0.615)}{3.583} educ + \underset{(0.129)}{1.302} age - \underset{(0.012)}{0.1053} paidwork - \underset{(4.90)}{24.87} female$$
$$+ \underset{(6.115)}{24.54} y2009 - \underset{(0.021)}{0.0501} y2009 \times paidwork \qquad R^2 = 0.186 \quad n = 2000 \qquad (3)$$

donde

- *educ* son los años de educación alcanzados,
- *watchtv* (ver televisión) y *paidwork* (trabajo remunerado) se miden en minutos por día,
- *female* (mujer) es una variable ficticia que toma valor 1 si el entrevistado es una mujer
- *y2009* es una variable ficticia que toma valor 1 si la encuesta se llevó a cabo en el bienio 2008-2009

*a)* En el modelo (1), ¿cuál es la interpretación del coeficiente de *educ*?

*b)* En el modelo (1), ¿hay una compensación estadísticamente significativa entre el tiempo dedicado al trabajo y el tiempo dedicado a ver la televisión?

*c)* Manteniendo igual todos los demás factores y tomando como modelo (2) como referencia, ¿existe evidencia de que los hombres ven la televisión más que las mujeres? ¿Es fuerte esa posible evidencia?

*d)* En el modelo (2), ¿cuál es la diferencia estimada del tiempo dedicado a ver la televisión entre las mujeres encuestadas en 2008-2009 y los hombres encuestados en el periodo 2002-2003? ¿Es esta diferencia estadísticamente significativa?

*e)* En el modelo (3), ¿cuál es el efecto marginal del tiempo dedicado al trabajo remunerado sobre el tiempo dedicado a ver televisión?

*f)* ¿Existe una interacción significativa entre el año de la encuesta y el tiempo dedicado al trabajo remunerado?

**Ejercicio 5.23** Utilizando el fichero *consumsp*, se estimaron los siguientes modelos para analizar si la entrada de España en la Comunidad Europea en 1986 tuvo algún impacto en el comportamiento de los consumidores españoles:

$$\widehat{conspc_t} = -\underset{(84.88)}{7.156} + \underset{(0.0857)}{0.3965} incpc_t + \underset{(0.0903)}{0.5771} conspc_{t-1} \qquad (1)$$
$$R^2 = 0.9967 \quad SCR = 1891320 \quad n = 56$$

$$\widehat{conspc_t} = -\underset{(108)}{102.4} + \underset{(0.0879)}{0.3573} incpc_t + \underset{(0.0901)}{0.5992} conspc_{t-1} + \underset{(92.56)}{148.60} y1986_t \qquad (2)$$
$$R^2 = 0.9968 \quad SCR = 1802007 \quad n = 56$$

$$\widehat{conspc_t} = \underset{(114)}{78.18} + \underset{(0.1100)}{0.5181} incpc_t + \underset{(0.1199)}{0.4186} conspc_{t-1} + \underset{(456.3)}{819.82} y1986_t$$
$$- \underset{(0.2338)}{0.5403} incpc_t \times y1986_t + \underset{(0.2182)}{0.5424} conspc_{t-1} \times y1986_t \qquad (3)$$
$$R^2 = 0.9972 \quad SCR = 1600714 \quad n = 56$$

$$\widehat{conspc_t} = \underset{(118)}{117.03} + \underset{(0.0968)}{0.3697} incpc_t + \underset{(0.1051)}{0.5823} conspc_{t-1} + \underset{(348)}{41.62} y1986_t$$
$$+ \underset{(0.0326)}{0.0104} incpc_t \times y1986_t \qquad (4)$$
$$R^2 = 0.9968 \quad SCR = 1798423 \quad n = 56$$

$$\widehat{conspc_t} = \underset{(114)}{120.1} + \underset{(0.0854)}{0.3750}\,incpc_t + \underset{(0.0890)}{0.5758}\,conspc_{t-1} + \underset{(0.0087)}{0.0141}\,incpc_t \times y1986_t \qquad (5)$$

$$R^2=0.9968 \quad SCR=1798927 \quad n=56$$

donde el consumo (*conspc*) y la renta disponible (*incpc*) se expresan en euros constantes per cápita, tomando 2008 como año de referencia.

(Los números entre paréntesis son los errores estándar de los estimadores.)

*a)* Compruebe en el modelo (6) si la propensión marginal al consumo a corto plazo se redujo en 1986 y años sucesivos.

*b)* ¿Son las interacciones de *y1986* con las variables cuantitativas del modelo significativas en forma conjunta?

*c)* Estime si hubo un cambio estructural en la función de consumo en 1986 y siguientes años.

*d)* Comprobar si el coeficiente de $conspc_{t-1}$ cambió en 1986.

*e)* ¿Existe una brecha entre el consumo que se realizaba antes de 1986 con respecto al consumo en 1986 y años sucesivos?