

# Apéndice A

## Uso de variables ficticias

Existe la posibilidad de introducir variables cualitativas en el modelo de regresión. Ese tipo de variables se consideran dicotómicas, tomando el valor 1 si la variable en estudio esta presente y el valor 0 si no lo esta. Por ejemplo, si analizamos el salario de distintos trabajadores se puede introducir la variable “poseer grado universitario” considerando que toma el valor 1 si lo posee y el valor 0 en caso contrario.

En este caso, dado el modelo:

$$Y_t = \beta_0 + \beta_1 D_t + u$$

donde  $Y_t$  es la variable dependiente y la variable  $D_t$  es la variable independiente de tipo cualitativo.

De esta manera si:

- Si  $D_t = 1$  el modelo queda expresado como:

$$E(Y_t) = \beta_0 + \beta_1$$

- Si  $D_t = 0$  el modelo queda expresado como:

$$E(Y_t) = \beta_0$$

Por lo tanto, la interpretación de los coeficientes sería la siguiente:  $\beta_0$  sería el valor medio de la variable explicada cuando no se satisface la categoría ( $D_t = 0$ ) y  $\beta_1$  sería la diferencia en el valor medio de la variable explicada en el caso de que se satisfaga la categoría ( $D_t = 1$ ).

Si incluimos una variable cuantitativa en el modelo, por ejemplo, los años de experiencia ( $X_2$ ):

$$Y_t = \beta_0 + \beta_1 D_t + \beta_2 X_t + u$$

De manera que:

- Si  $D_t = 1$  el modelo quedaría expresado como:

$$E(Y_t) = \beta_0 + \beta_1 + \beta_2 X_t$$

- Si  $D_t = 0$  el modelo quedaría expresado como:

$$E(Y_t) = \beta_0 + \beta_2 X_t$$

Entonces, la interpretación de los coeficientes sería la siguiente:  $\beta_0$  es el valor medio de la variable explicada cuando no se satisface la categoría ( $D_t = 0$ ) y la variable cuantitativa toma el valor 0,  $\beta_1$  es la diferencia en el valor medio de la variable explicada en el caso de que se satisfaga la categoría ( $D_t = 1$ ) y la variable cuantitativa se mantenga constante.

**Ejemplo 18** Se tiene el siguiente modelo obtenido a partir de datos de 41 trabajadores en relación a su retribución mensual ( $Y$ ), si poseen o no grado universitario ( $X_1$ ) (toma el valor 1 si poseen) y sus años de experiencia ( $X_2$ ):

$$Y = 845,48 + 7122,08X_1 + 1301,31X_2$$

$$R^2 = 0,857$$

La estimación obtenida para los parámetros se interpreta de la siguiente forma:

- $\hat{\beta}_0 = 845,48$  se interpreta como la retribución de un empleado sin grado universitario y cero años de experiencia.
- $\hat{\beta}_1 = 7122,08$  se interpretar como la diferencia entre la retribución anual de un empleado con grado universitario y otro que no lo tenga, a igualdad del resto de variables explicativas.
- $\hat{\beta}_2 = 1301,31$  se interpreta como lo que aumenta la retribución anual de un empleado por cada año de experiencia, a igualdad del resto de variables explicativas.

### Trampa de las variables ficticias

Al introducir la variable cualitativa en el modelo hay que tener presente el numero de categorías que incluye, ya que una de las categorías debe quedarse como variable de control y, por tanto, se introducen tantas variables dicotómicas en el modelo como categorías tenga la variable menos una. De hecho, al describir la variable “poseer grado universitario” hemos introducido una única variable porque si incluyéramos ambas, la suma de las dos variables coincidiría con el termino independiente dando lugar a un problema de colinealidad exacta que impediría la estimación del modelo. Este problema se conoce como trampa de las variables ficticias. Por

ejemplo, si queremos incorporar la variable “titulacion cursada en la FCCEE” que hemos descrito en cuatro categorias (GMIM, GADE, GFICO, GECO), se crearían tres variables dicotómicas:

- $D_1$  toma el valor 1 si la titulacion cursada es GMIM y cero si no lo es.
- $D_2$  toma el valor 1 si la titulacion cursada es GADE y cero si no lo es.
- $D_3$  toma el valor 1 si la titulacion cursada es GFICO y cero si no lo es.

quedando la variable GECO como variable de control.