

Ejercicios de Wooldridge

Diana Suarez Hernandez, Melannie Machado Barreto, Paula Peña Lozano

2024-03-02

```
library(wooldridge)

## Warning: package 'wooldridge' was built under R version 4.2.3

library(modelsummary)

## Warning: package 'modelsummary' was built under R version 4.2.3

## Version 2.0.0 of `modelsummary`, to be released soon, will introduce a
## breaking change: The default table-drawing package will be
## `tinytable`
## instead of `kableExtra`. All currently supported table-drawing
## packages
## will continue to be supported for the foreseeable future, including
## `kableExtra`, `gt`, `huxtable`, `flextable`, and `DT`.
##
## You can always call the `config_modelsummary()` function to change
## the
## default table-drawing package in persistent fashion. To try
## `tinytable`
## now:
##
## config_modelsummary(factory_default = 'tinytable')
##
## To set the default back to `kableExtra`:
##
## config_modelsummary(factory_default = 'kableExtra')
```

C.3.1 Por ejemplo, un nivel de ingresos más alto en general da como resultado el acceso a mejores cuidados prenatales y a una mejor nutrición de la madre. Una ecuación que reconoce estos factores es:

$$bwght = \beta_0 + \beta_1 cigs + \beta_2 faminc + u$$

```
View(bwght)
```

¿Cual es el signo más probable para B2?

El B2 es probablemente significativo ya que indica que el signo mas probable de B2 es positivo, esto quiere decir que si existe una relación entre la variable “faminc” y la variable dependiente, por tanto a medidas que el ingreso familiar aumenta, es probable que resulte una atención prenatal y una mejor nutrición para la madre.

ii ¿Cree que cigs y faminc estén correlacionados? Explique por qué la correlación puede ser positiva o negativa?

```
cor(bwght$cigs, bwght$faminc )  
## [1] -0.1730449
```

la correlacion entre los datos de cigs y faminc da un resultado negativo debido a que las madres con ingresos mas altos presentan mas consumismo a fumar debido a varios factores.

iii Ahora, calcule la ecuación con y sin faminc utilizando los datos del archivo BWGHT, Dé los resultados en forma de ecuación incluyendo el tamaño de la muestra y la R cuadrada. Explique sus resultados enfocándose en si el añadir faminc modifica de manera.

```
reg <- lm(bwght~cigs, data=bwght)  
summary(reg)  
  
##  
## Call:  
## lm(formula = bwght ~ cigs, data = bwght)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -96.772 -11.772   0.297  13.228 151.228   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)  119.77190    0.57234  209.267  < 2e-16 ***  
## cigs         -0.51377    0.09049   -5.678  1.66e-08 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 20.13 on 1386 degrees of freedom  
## Multiple R-squared:  0.02273,    Adjusted R-squared:  0.02202   
## F-statistic: 32.24 on 1 and 1386 DF,  p-value: 1.662e-08  
  
reg <- lm(bwght~cigs+faminc, data=bwght)  
summary(reg)  
  
##  
## Call:  
## lm(formula = bwght ~ cigs + faminc, data = bwght)  
##  
## Residuals:
```

```
##      Min      1Q  Median      3Q      Max
## -96.061 -11.543   0.638  13.126 150.083
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 116.97413    1.04898 111.512 < 2e-16 ***
##      cigs      -0.46341    0.09158  -5.060 4.75e-07 ***
##      faminc     0.09276    0.02919   3.178 0.00151 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.06 on 1385 degrees of freedom
## Multiple R-squared:  0.0298, Adjusted R-squared:  0.0284
## F-statistic: 21.27 on 2 and 1385 DF, p-value: 7.942e-10
```

**En la ecuación que incluye faminc, el coeficiente para cigs es -0.46, lo que indica que por cada cigarrillo adicional fumado por día, el peso al nacer disminuye en promedio en 0.46 unidades. Sin embargo, el coeficiente para faminc es positivo (0.093), lo que indica que un aumento en el ingreso familiar se asocia con un aumento en el peso al nacer. Por otro lado, en la ecuación que no incluye faminc, el coeficiente para cigs es un poco mayor en magnitud (-0.51), lo que indica que sin tener en cuenta el efecto del ingreso familiar, el efecto negativo de fumar sobre el peso al nacer parece ser un poco más fuerte. Por lo tanto, al comparar los resultados de las dos ecuaciones, con y sin faminc, podemos decir que al añadir faminc a la ecuación modifica de manera sustancial el efecto esperado de cigs sobre bwght al disminuir el impacto negativo del tabaquismo sobre el peso al nacer con el efecto positivo del ingreso familiar.

C3.2 Utilice los datos del archivo HPRICE1.RAW para estimar el modelo

$$price = \beta_0 + \beta_1sqrft + \beta_2bdrms + u$$

i Escriba los resultados en forma de ecuación.

$$price = \beta_0 + \beta_1sqrft + \beta_2bdrms + u_2$$

```
View(hprice1)

modelo2 <- lm(formula = price ~ sqrft + bdrms, data = hprice1)

summary(modelo2)

##
## Call:
## lm(formula = price ~ sqrft + bdrms, data = hprice1)
##
## Residuals:
##      Min      1Q  Median      3Q      Max
## -127.627  -42.876   -7.051   32.589  229.003
```

```
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -19.31500   31.04662  -0.622   0.536
## sqrft       0.12844    0.01382   9.291 1.39e-14 ***
## bdrms       15.19819    9.48352   1.603   0.113
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 63.04 on 85 degrees of freedom
## Multiple R-squared:  0.6319, Adjusted R-squared:  0.6233
## F-statistic: 72.96 on 2 and 85 DF, p-value: < 2.2e-16
```

Por cada habitación adicional, el precio de venta aumenta en promedio 15,19, mientras que por cada pie cuadrado adicional, aumenta en promedio 0,12 y por lo tanto decimos que la variable `sqrft` es estadísticamente significativa.

ii ¿Cual es el incremento en precio estimado para una casa con una habitación (`bdrms`) más, manteniendo constante la superficie en pies cuadrados (`sqrft`)?

```
#Definir los coeficientes
intercepto <- -19.31500
coef_sqrft <- 0.12844
coef_bdrms <- 15.19819

precio2 <- coef_bdrms*1
summary(precio2)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      15.2   15.2   15.2   15.2   15.2   15.2
```

El incremento en el precio estimado para una casa con una habitación (`bdrms`) más, manteniendo constante la superficie en pies cuadrados (`sqrft`), es de \$15,198.2. Lo que significa que por cada habitación adicional, el precio estimado de la casa aumenta en promedio \$15,198.2, manteniendo constante el área en pies cuadrados

iii ¿Cual es el incremento en precio estimado para una casa con una habitación adicional de 140 pies cuadrados? Compare esto con su respuesta al inciso (ii).

```
precio1 <- coef_sqrft*140
summary(precio1)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.1284  0.1284  0.1284  0.1284  0.1284  0.1284

habitacion_adicional <- precio1*140
summary(habitacion_adicional)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      17.98  17.98  17.98  17.98  17.98  17.98
```

El incremento en el precio estimado para una casa con una habitación adicional de 140 pies cuadrados es de \$17,981.6. En este caso, el precio se incrementa en \$128.44 por cada pie cuadrado adicional, manteniendo constante el número de habitaciones. Entonces, para 140 pies cuadrados adicionales, el precio sería de \$17,981.6.

iv ¿Qué porcentaje de la variación en el precio se explica por la extensión en pies cuadrados y el número de habitaciones?

El resultado de R-cuadrado nos muestra que alrededor del 63.19% de la variabilidad en el precio de las casas se puede explicar por la extensión en pies cuadrados y el número de habitaciones. Lo que indica que el modelo, al incorporar las variables de `sqrft` y `bdrms`, explica una parte significativa de la variación observada en los precios de las propiedades.

v La primera casa en la muestra tiene `sqrft`=2,438 y `bdrms`=4. Determine el precio de venta estimado para esta casa con la línea de regresión de MCO.

```
# Definir Los valores de sqrft y bdrms para La primera casa
sqrft_primera_casa <- 2438
bdrms_primera_casa <- 4

# Calcular el precio de venta estimado utilizando La ecuación de
regresión
precio_estimado_primera_casa <- intercepto + coef_sqrft *
sqrft_primera_casa + coef_bdrms * bdrms_primera_casa
summary(precio_estimado_primera_casa)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    354.6   354.6   354.6   354.6   354.6   354.6
```

Utilizando la ecuación de regresión lineal múltiple, el precio de venta estimado para la primera casa en la muestra, que tiene una superficie de 2,438 pies cuadrados y 4 habitaciones, es de \$354,605.

vi El precio de venta de la primera casa en la muestra fue \$300,000 (así que `price`=300). Determine el residual para esta casa. ¿Sugiere esto que el comprador pagó de más o de menos por la casa.

```
# Definir el precio de venta observado para La primera casa
precio_observado_primera_casa <- 300000

# Calcular el residual para La primera casa
residual_primera_casa <- precio_observado_primera_casa -
precio_estimado_primera_casa
summary(residual_primera_casa)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   299645   299645   299645   299645   299645   299645
```

En este caso, vemos que el precio pagado por la casa fue significativamente menos de lo que el modelo sugirió, lo que implica que el comprador obtuvo un buen trato y pagó menos de lo que el modelo predijo que valía la casa.