Chapter 3 - Multiple Regression Analysis - Estimation

Thalles Quinaglia Liduares

15/02/2022

Exercise 3.1

Upload package

```
library(lmreg)
library(wooldridge)
```

Upload database

```
data<-wooldridge::bwght
```

A problem of interest to health officials (and others) is to determine the effects of smoking during pregnancy on infant health. One measure of infant health is birth weight; a birth weight that is too low can put an infant at risk for contracting various illnesses. Since factors other than cigarette smoking that affect birth weight are likely to be correlated with smoking, we should take those factors into account. For example, higher income generally results in access to better prenatal care, as well as better nutrition for the mother. An equation that recognizes this is

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

(i) What is the most likely sign for β_2?

As family income increase, the birth weight of children is expected to increase as well, ceteris paribus. Hence, the expected sign of β_2 is positive.

(ii) Do you think cigs and faminc are likely to be correlated? Explain why the correlation might be positive or negative.

There's a variety of factors that influence a people to smoke. The correlation between quantity of cigarrestes smoked and family's income might be positive or negative. Given that correlation does not imply causality, there's much more observable and unobservable factors that influence a people behavior in smoke or not smoke.

(iii) Now, estimate the equation with and without faminc, using the data in BWGHT .RAW. Report the results in equation form, including the sample size and R-squared. Discuss your results, focusing on whether adding faminc substantially changes the estimated effect of cigs on bwght.

Regression with `faminc`

```
est1<-lm(data$bwght~data$cigs+data$faminc, data)

summary(est1)
```

```
##
## Call:
## lm(formula = data$bwght ~ data$cigs + data$faminc, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -96.061 -11.543   0.638  13.126 150.083
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 116.97413    1.04898 111.512  < 2e-16 ***
## data$cigs    -0.46341    0.09158  -5.060 4.75e-07 ***
## data$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
```

The estimated equation is given by $bwgth = 116.97 - 0.46cigs + 0.09faminc$.

The sample size and R-Squared are 1387 and 2.9%, respectively.

The coefficients present the expected sign and are both statistically significant.

For each smoked cigar, the weight of children lowers 0.46 pounds, *ceteris paribus*.

For each $1 increase in family income, the weight increases 0.09 pounds, *ceteris paribus*.

Now, the estimate without `faminc`.

```
est2<-lm(data$bwght~data$cigs)

summary(est2)
```

```
##
## Call:
## lm(formula = data$bwght ~ data$cigs)
##
## 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 ***
## data$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
```

In this case, the coefficient associate to number of smoked cigs, is equal to -0.51. Impliyng in a little difference in relation to the first estimative.