Chapter 3 - The Multiple Regression Model - Estimation

Thalles Quinaglia Liduares

01/03/2022

Exercise 3.7

Upload packages

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

Upload database

```
data<-wooldridge::meap93

attach(data)
```

Use the data in MEAP93.RAW to answer this question.

(i) Estimate the model

$$math10 = \beta_0 + \beta_1 log(expend) + \beta_2 lnchprg + u$$

and report the results in the usual form, including the sample size and R-squared. Are the signs of the slope coefficients what you expected? Explain.

```
lm1<-lm(math10~lexpend+lnchprg)

summary(lm1)
```

```
##
## Call:
## lm(formula = math10 ~ lexpend + lnchprg)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -24.294  -6.172  -1.293   4.855  43.203
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -20.36075   25.07288  -0.812   0.4172
## lexpend       6.22969    2.97263   2.096   0.0367 *
## lnchprg      -0.30459    0.03536  -8.614   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.526 on 405 degrees of freedom
## Multiple R-squared:  0.1799, Adjusted R-squared:  0.1759
## F-statistic: 44.43 on 2 and 405 DF,  p-value: < 2.2e-16
```

The estimated equation is given by

$$math10 = -20.36 + 6.22lexpend - 0.30lnchprg$$

The coefficient of `lexpend` has the expected sign, but the coefficient of `lnchprg` , similary to the example 2.12 still have negative sign. Hence, the problem of ommited variables might persists in this case.

The variability in `math10` is just around 18% explaneid by this specific model. The sample size is equal to 408 observations.

(ii) What do you make of the intercept you estimated in part (i)? In particular, does it make sense to set the two explanatory variables to zero? [Hint: Recall that log(1)=0.]

The intercept represents the value of math exam if `lexpend` and `lnchprg` were equal to 0. In this case, the estimated intercept is not statistically significant, cause a negative value for exam grade do not make sense.

(iii) Now run the simple regression of math10 on log(expend), and compare the slope coefficient with the estimate obtained in part (i). Is the estimated spending effect now larger or smaller than in part (i)?

```
lm2<-lm(math10~lexpend)

summary(lm2)
```

```
##
## Call:
## lm(formula = math10 ~ lexpend)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -22.343  -7.100  -0.914   6.148  39.093
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -69.341     26.530  -2.614 0.009290 **
## lexpend       11.164      3.169   3.523 0.000475 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.35 on 406 degrees of freedom
## Multiple R-squared:  0.02966,    Adjusted R-squared:  0.02727
## F-statistic: 12.41 on 1 and 406 DF,  p-value: 0.0004752
```

Now, the coefficient is quite larger in relation to the multiple regression model. A difference of 4.94

(iv) Find the correlation between lexpend = log(expend) and lnchprg. Does its sign make sense to you?

```
cor(lexpend, lnchprg)
```

```
## [1] -0.1927042
```

The correlation between these two variables is negative. Hence, if one increase, other decreases, what might explains the obtained values in item (iii).