Chapter 4 - Inference

Thalles Quinaglia Liduares

24/03/2022

Exercise 4.10

Upload packages

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

Upload database

```
data<-wooldridge::elem94_95

attach(data)
```

Use the data in ELEM94_95 to answer this question. The findings can be compared with those in Table 4.1. The dependent variable `lavgsal` is the log of average teacher salary and `bs` is the ratio of average benefits to average salary (by school).

(i) Run the simple regression of `lavgsal` on `bs`. Is the estimated slope statistically different from zero? Is it statistically different from -1?

Regression

```
summary(lm1<-lm(lavgsal~bs))
```

```
##
## Call:
## lm(formula = lavgsal ~ bs)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.19563 -0.14294  0.00525  0.14699  0.77352
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.74790    0.05166 208.042  < 2e-16 ***
## bs          -0.79512    0.14965  -5.313 1.21e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2312 on 1846 degrees of freedom
## Multiple R-squared:  0.01506,    Adjusted R-squared:  0.01453
## F-statistic: 28.23 on 1 and 1846 DF,  p-value: 1.208e-07
```

Hypothesis test: different from zero

```
linearHypothesis(lm1, c("bs=0"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## bs = 0
##
## Model 1: restricted model
## Model 2: lavgsal ~ bs
##
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1   1847 100.181
## 2   1846  98.672  1    1.5089 28.229 1.208e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

It's possible to reject the null hypothesis that the coefficient is equal to 0 at the 1% significance level.

Hypothesis test: different from -1

```
linearHypothesis(lm1, c("bs=-1"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## bs = - 1
##
## Model 1: restricted model
## Model 2: lavgsal ~ bs
##
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1   1847 98.773
## 2   1846 98.672  1   0.10018 1.8741 0.1712
```

Isn't possible to reject the null hypothesis that the coefficient is equal to -1, because the p-value=0.17>0.05.

(ii) Add the variables lenrol and lstaff to the regression from part (i). What happens to the coefficient on bs? How does the situation compare with that in Table 4.1?

```
summary(lm2<-lm(lavgsal~bs+lenrol+lstaff))
```

```
##
## Call:
## lm(formula = lavgsal ~ bs + lenrol + lstaff)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.88375 -0.10941 -0.01043  0.10294  0.61615
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.953049   0.107234 130.118  < 2e-16 ***
## bs          -0.605061   0.108743  -5.564 3.02e-08 ***
## lenrol      -0.031585   0.008477  -3.726    2e-04 ***
## lstaff      -0.713719   0.017790 -40.119  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1678 on 1844 degrees of freedom
## Multiple R-squared:  0.482,  Adjusted R-squared:  0.4812
## F-statistic:   572 on 3 and 1844 DF,  p-value: < 2.2e-16
```

The coefficient on `bs` changes from -0.79 to -0.60.

(iii) How come the standard error on the `bs` coefficient is smaller in part (ii) than in part (i)? (Hint: What happens to the error variance versus multicollinearity when `lenrol` and `lstaff` are added?)

In progress..

(v) Now add the variable lunch to the regression. Holding other factors fixed, are teachers being compensated for teaching students from disadvantaged backgrounds? Explain.

```
summary(lm3<-lm(lavgsal~bs+lenrol+lstaff+lunch))
```

```
##
## Call:
## lm(formula = lavgsal ~ bs + lenrol + lstaff + lunch)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.92076 -0.11018 -0.00995  0.10614  0.58991
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.8314939  0.1097259 126.055  < 2e-16 ***
## bs          -0.5161290  0.1097747  -4.702 2.77e-06 ***
## lenrol      -0.0284092  0.0084560  -3.360 0.000796 ***
## lstaff      -0.6906322  0.0183604 -37.615  < 2e-16 ***
## lunch       -0.0007581  0.0001615  -4.695 2.87e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1668 on 1843 degrees of freedom
## Multiple R-squared:  0.4882, Adjusted R-squared:  0.487
## F-statistic: 439.4 on 4 and 1843 DF,  p-value: < 2.2e-16
```

No, holding other factors fixed, the effect is negative, besides it’s small magnitude.