Chapter 4 - Inference

Thalles Quinaglia Liduares

29/03/2022

Exercise 4.11

Upload packages

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

Upload database

```
data<-wooldridge::htv

attach(data)
```

Use the data in HTV.RAW to answer this question. See also Computer Exercise C10 in Chapter 3.

(i) Estimate the regression model

$$educ = \beta_0 + \beta_1 motheduc + \beta_2 fatheduc + \beta_3 abil + \beta_4 abil^2 + u$$

by OLS and report the results in the usual form. Test the null hypothesis that `educ` is linearly related to `abil` against the alternative that the relationship is quadratic.

```
summary(lm1<-lm(educ~motheduc+fatheduc+abil+I(abil^2)))
```

```
##
## Call:
## lm(formula = educ ~ motheduc + fatheduc + abil + I(abil^2))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.2506 -1.1274 -0.1355  1.0223  7.0482
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 8.240226   0.287410  28.671  < 2e-16 ***
## motheduc    0.190126   0.028096   6.767 2.03e-11 ***
## fatheduc    0.108939   0.019601   5.558 3.35e-08 ***
## abil        0.401462   0.030288  13.255  < 2e-16 ***
## I(abil^2)   0.050599   0.008304   6.093 1.48e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.758 on 1225 degrees of freedom
## Multiple R-squared:  0.4444, Adjusted R-squared:  0.4425
## F-statistic: 244.9 on 4 and 1225 DF,  p-value: < 2.2e-16
```

The estimated equation is expressed as follows

$$\widehat{educ} = 8.2 + 0.19motheduc + 0.10fatheduc + 0.40abil + 0.05abil^2$$

```
linearHypothesis(lm1, c("I(abil^2)=0"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## I(abil^2) = 0
##
## Model 1: restricted model
## Model 2: educ ~ motheduc + fatheduc + abil + I(abil^2)
##
##   Res.Df    RSS Df Sum of Sq     F    Pr(>F)
## 1   1226 3900.0
## 2   1225 3785.2  1    114.73 37.13 1.478e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
linearHypothesis(lm1, c("abil=0","I(abil^2)=0"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## abil = 0
## I(abil^2) = 0
##
## Model 1: restricted model
## Model 2: educ ~ motheduc + fatheduc + abil + I(abil^2)
##
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1   1227 5114.3
## 2   1225 3785.2  2    1329.1 215.06 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

```
## Linear hypothesis test
##
## Hypothesis:
## abil = 0
##
## Model 1: restricted model
## Model 2: educ ~ motheduc + fatheduc + abil + I(abil^2)
##
##   Res.Df    RSS Df Sum of Sq     F    Pr(>F)
## 1   1226 4328.1
## 2   1225 3785.2  1     542.9 175.7 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on all combinations of hypothesis tests, its possible to reject the null that `abil^2` is equal to 0 in the estimated model.

(ii) Using the equation in part (i), test $H_0: _1 = _2 $ against a two-sided alternative. What is the p-value of the test?

```
linearHypothesis(lm1, c("motheduc-fatheduc=0"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## motheduc - fatheduc = 0
##
## Model 1: restricted model
## Model 2: educ ~ motheduc + fatheduc + abil + I(abil^2)
##
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1   1226 3796.8
## 2   1225 3785.2  1    11.578 3.7468 0.05314 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In this case, its not possible to reject the null hypothesis, because the p-value associated to F-Test is equal to 0.053>0.05.

(iii) Add the two college tuition variables to the regression from part (i) and determine whether they are jointly statistically significant.

```
summary(lm2<-lm(educ ~ motheduc + fatheduc + abil + I(abil^2)+tuit17+tuit18))
```

```
##
## Call:
## lm(formula = educ ~ motheduc + fatheduc + abil + I(abil^2) +
##     tuit17 + tuit18)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.148 -1.161 -0.114  1.032  7.071
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 8.082e+00  3.128e-01  25.840  < 2e-16 ***
## motheduc    1.929e-01  2.818e-02   6.845 1.21e-11 ***
## fatheduc    1.084e-01  1.962e-02   5.528 3.96e-08 ***
## abil        3.990e-01  3.035e-02  13.148  < 2e-16 ***
## I(abil^2)   5.055e-02  8.313e-03   6.082 1.59e-09 ***
## tuit17      1.576e-02  6.250e-02   0.252    0.801
## tuit18      6.033e-05  6.365e-02   0.001    0.999
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.758 on 1223 degrees of freedom
## Multiple R-squared:  0.4451, Adjusted R-squared:  0.4424
## F-statistic: 163.5 on 6 and 1223 DF,  p-value: < 2.2e-16
```

```
linearHypothesis(lm2, c("tuit17=0","tuit18=0"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## tuit17 = 0
## tuit18 = 0
##
## Model 1: restricted model
## Model 2: educ ~ motheduc + fatheduc + abil + I(abil^2) + tuit17 + tuit18
##
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1   1225 3785.2
## 2   1223 3780.1  2    5.1884 0.8393 0.4322
```

Based on the estimated coefficients and the hypothesis test its possible conclude that both `tuit17` and `tuit18` do not influence the dependent variable `educ`.

(iv) What is the correlation between `tuit17` and `tuit18`? Explain why using the average of the tuition over the two years might be preferred to adding each separately. What happens when you do use the average?

Correlation

```
cor(tuit17, tuit18)
```

```
## [1] 0.9808333
```

The extreme high correlation between `tuit17` and `tuit18` equal to 98% might be causing multicollinearity problems in the estimation. Using, one of the two, or the median between them, might be a possible solution to this problem.

(v) Do the findings for the average tuition variable in part (iv) make sense when interpreted causally? What might be going on?

```
data$tuit_mean<-(tuit17+tuit18)/2
```

```
summary(lm3<-lm(educ ~ motheduc + fatheduc + abil + I(abil^2) + data$tuit_mean))
```

```
##
## Call:
## lm(formula = educ ~ motheduc + fatheduc + abil + I(abil^2) +
##     data$tuit_mean)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.1469 -1.1591 -0.1132  1.0312  7.0709
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    8.081339   0.312614  25.851  < 2e-16 ***
## motheduc       0.192863   0.028168   6.847 1.19e-11 ***
## fatheduc       0.108368   0.019601   5.529 3.94e-08 ***
## abil           0.399081   0.030336  13.156  < 2e-16 ***
## I(abil^2)      0.050599   0.008302   6.095 1.46e-09 ***
## data$tuit_mean 0.015963   0.012373   1.290    0.197
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.757 on 1224 degrees of freedom
## Multiple R-squared:  0.4451, Adjusted R-squared:  0.4428
## F-statistic: 196.4 on 5 and 1224 DF,  p-value: < 2.2e-16
```

The inclusion of `tuit_mean` do not show individual significance.