Chapter 4 - Inference

Thalles Quinaglia Liduares

18/03/2022

Exercise 4.7

Upload packages

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

Upload database

```
data<-wooldridge::twoyear

str(data)
```

```
## 'data.frame':    6763 obs. of  23 variables:
##  $ female  : int  1 1 1 1 1 0 0 0 0 0 ...
##  $ phsrank : int  65 97 44 34 80 59 81 50 8 56 ...
##  $ BA      : int  0 0 0 0 0 0 1 0 0 1 ...
##  $ AA      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ black   : int  0 0 0 0 0 0 0 1 0 1 ...
##  $ hispanic: int  0 0 0 1 0 0 0 0 0 0 ...
##  $ id      : num  19 93 96 119 132 156 163 188 199 200 ...
##  $ exper   : int  161 119 81 39 141 165 127 161 138 64 ...
##  $ jc      : num  0 0 0 0.267 0 ...
##  $ univ    : num  0 7.03 0 0 0 ...
##  $ lwage   : num  1.93 2.8 1.63 2.22 1.64 ...
##  $ stotal  : num  -0.442 0 -1.357 -0.19 0 ...
##  $ smcity  : int  0 1 0 1 0 1 1 0 1 0 ...
##  $ medcity : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ submed  : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ lgcity  : int  0 0 0 0 0 0 0 1 0 0 ...
##  $ sublg   : int  1 0 1 0 0 0 0 0 0 0 ...
##  $ vlgcity : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ subvlg  : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ ne      : int  1 0 1 0 0 0 0 0 0 0 ...
##  $ nc      : int  0 1 0 0 0 0 1 0 0 0 ...
##  $ south   : int  0 0 0 0 1 1 0 1 0 1 ...
##  $ totcoll : num  0 7.033 0 0.267 0 ...
##  - attr(*, "time.stamp")= chr "25 Jun 2011 23:03"
```

Refer to the example used in Section 4.4. You will use the data set TWOYEAR.RAW.

(i) The variable `phsrank` is the person's high school percentile. (A higher number is better. For example, 90 means you are ranked better than 90 percent of your graduating class.) Find the smallest, largest, and average `phsrank` in the sample.

```
summary(data$phsrank)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00   44.00   50.00   56.16   76.00   99.00
```

The smallest, largest and average value of `phsrank` are 0, 99.0 and 56.16 respectively.

(ii) Add `phsrank` to equation (4.26) and report the OLS estimates in the usual form. Is `phsrank` statistically significant? How much is 10 percentage points of high school rank worth in terms of wage?

```
lm1<-lm(lwage~jc+totcoll+exper+phsrank, data)

summary(lm1)
```

```
##
## Call:
## lm(formula = lwage ~ jc + totcoll + exper + phsrank, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.09049 -0.28135  0.00538  0.28543  1.79060
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.4587472  0.0236211  61.756   <2e-16 ***
## jc          -0.0093108  0.0069693  -1.336    0.182
## totcoll      0.0754756  0.0025588  29.496   <2e-16 ***
## exper        0.0049396  0.0001575  31.360   <2e-16 ***
## phsrank      0.0003032  0.0002389   1.269    0.204
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4301 on 6758 degrees of freedom
## Multiple R-squared:  0.2226, Adjusted R-squared:  0.2222
## F-statistic: 483.8 on 4 and 6758 DF,  p-value: < 2.2e-16
```

The estimated equation is expressed as follows

$$\widehat{log(wage)} = 1.45 - 0.009jc + 0.075totcoll + 0.004exper + 0.0003phsrank$$

The variable `phsrank` do not show statistical significance at any level.

In terms of return to wage, an increase of 10% in `phsrank` , implies in 0.003% increase wage.

(iii) Does adding `phsrank` to (4.26) substantively change the conclusions on the returns to two- and four-year colleges? Explain.

In progress…

(iv) The data set contains a variable called id. Explain why if you add id to equation (4.17) or (4.26) you expect it to be statistically insignificant. What is the two-sided p-value?

In progress…