Chapter 4 - Inference

Thalles Quinaglia Liduares

21/03/2022

Exercise 4.8

Upload packages

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

Upload database

```
data1<-wooldridge::k401ksubs
```

The data set 401KSUBS.RAW contains information on net financial wealth (nettfa), age of the survey respondent (age), annual family income (inc), family size (fsize), and participation in certain pension plans for people in the United States. The wealth and income variables are both recorded in thousands of dollars. For this question, use only the data for single-person households (so fsize = 1).

(i) How many single-person households are there in the data set?

```
single_person<- data1 %>%
  filter(marr == 0) %>%
  summarise(n=n())

single_person
```

```
##      n
## 1 3445
```

Hence, there's 3445 single persons in the sample.

(ii) Use OLS to estimate the model

$$nettfa = \beta_0 + \beta_1 inc + \beta_2 age + u$$

and report the results using the usual format. Be sure to use only the single-person households in the sample. Interpret the slope coefficients. Are there any surprises in the slope estimates?

```
lm1<-lm(nettfa~age+inc, data1, subset= fsize==1)

summary(lm1)
```

```
##
## Call:
## lm(formula = nettfa ~ age + inc, data = data1, subset = fsize ==
##     1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -179.95  -14.16   -3.42    6.03 1113.94
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -43.03981    4.08039 -10.548   <2e-16 ***
## age           0.84266    0.09202   9.158   <2e-16 ***
## inc           0.79932    0.05973  13.382   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 44.68 on 2014 degrees of freedom
## Multiple R-squared:  0.1193, Adjusted R-squared:  0.1185
## F-statistic: 136.5 on 2 and 2014 DF,  p-value: < 2.2e-16
```

The estimated equation is expressed as follows

$$\widehat{nettfa} = -43.03 + 0.84age + 0.79inc$$

The estimated coefficients have the expected sign. For an unit increase in age of individuals, the net familiar financial assets, increases by \$0.84 (thousands). For an increase of \$1 in income, the `nettfa` increases by \$0.8 (thousands).

(iv) Find the p-value for the test $H_0 : \beta_2 = 1$ against $H_1 : \beta_2 < 1$. Do you reject H_0 at the 1% significance level?

```
linearHypothesis(lm1, c("inc=1"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## inc = 1
##
## Model 1: restricted model
## Model 2: nettfa ~ age + inc
##
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1   2015 4043586
## 2   2014 4021048  1     22538 11.288 0.0007945 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Yes. It's possible to reject the null hypothesis at the 1% significance level.

(v) If you do a simple regression of nettfa on inc, is the estimated coefficient on inc much different from the estimate in part (ii)? Why or why not?

```
lm2<-lm(nettfa~inc, data1, subset=fsize==1)

summary(lm2)
```

```
##
## Call:
## lm(formula = nettfa ~ inc, data = data1, subset = fsize == 1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -185.12  -12.85   -4.85    1.78 1112.66
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -10.5709     2.0607   -5.13 3.18e-07 ***
## inc           0.8207     0.0609   13.48  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 45.59 on 2015 degrees of freedom
## Multiple R-squared:  0.08267,    Adjusted R-squared:  0.08222
## F-statistic: 181.6 on 1 and 2015 DF,  p-value: < 2.2e-16
```

In this case, the estimated equation is expressed as follows

$$\widehat{nettfa} = -10.5 + 0.82inc$$

This represent a sligth difference in terms of magnitude of coefficients.