#Chapter 1 - Introduction Econometrics - J. Wooldridge

Thalles Quinaglia Liduares

04/02/2022

Upload package for data base

```
library(wooldridge)
```

Package for data analysis

```
library(dplyr)
```

Upload the Database

```
data<-wooldridge::wage1

attach(data)

View(data)
```

Exercise 1 - Use the data in WAGE1.RAW for this exercise

(i) Find the average education level in the sample. What are the lowest and highest years of education?

```
mean_educ<-round(mean(educ),2)

mean_educ
```

```
## [1] 12.56
```

The medium value for years of education are 12.5 years.

```
min_educ<-min(educ)

min_educ
```

```
## [1] 0
```

The minimum value for years of education are 0.

```
max_educ<-max(educ)

max_educ
```

```
## [1] 18
```

The maximum value for years of education are 18 years.

(ii) Find the average hourly wage in the sample. Does it seem high or low?

```
mean_wage<-round(mean(wage),2)

mean_wage
```

```
## [1] 5.9
```

The medium salary/hour are $5.89.

(v) How many women are in the sample? How many men?

```
num_woman <- data %>%
  filter(female ==1) %>%
  summarise(n=n())

num_woman
```

```
##     n
## 1 252
```

There are 252 woman in the sample.

Now, counting the number of mens in the sample.

```
num_man <- data %>%
  filter(female ==0) %>%
  summarise(n=n())

num_man
```

```
##     n
## 1 274
```

There are 274 man in the sample.

Exercise 2

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

```
data2<-wooldridge::bwght

attach(data2)
```

(i) How many women are in the sample, and how many report smoking during pregnancy?

```
num_woman<-length(motheduc)

num_woman
```

```
## [1] 1388
```

There are 1388 woman in the sample.

```
cig0<-data2 %>%
      filter(cigs == 0)

length(cig0)
```

```
## [1] 14
```

```
cig1<-1388-14

cig1
```

```
## [1] 1374
```

1374 woman reported that they smoked at least 1 cigar during pregnancy.

(ii) What is the average number of cigarettes smoked per day? Is the average a good measure of the "typical" woman in this case? Explain.

```
mean_cigs<-round(mean(cigs),2)

mean_cigs
```

```
## [1] 2.09
```

The mean value for smoked cigarretes is 2.09.

(iii) Among women who smoked during pregnancy, what is the average number of cigarettes smoked per day? How does this compare with your answer from part (ii), and why?

```
mean_smoked<-round(mean(packs),2)

mean_smoked
```

```
## [1] 0.1
```

(iv) Find the average of fatheduc in the sample. Why are only 1,192 observations used to compute this average?

```
mean_fatheduc<-round(mean(fatheduc, na.rm=T),2)

mean_fatheduc
```

```
## [1] 13.19
```

The mean value for father's education is 13.18 years. There's only 1,192 observations in the sample because there's presence of NA's.

(v) Report the average family income and its standard deviation in dollars.

```
mean_income<-round(mean(faminc),2)

mean_income
```

```
## [1] 29.03
```

```
sd_income<-sd(faminc)

sd_income
```

```
## [1] 18.73928
```

The mean income value is $29,026 and the sd is $18.730.

#Exercise 3

EX 3 data in MEAP01.RAW are for the state of Michigan in the year 2001. Use these data to answer the following questions.

Upload of database

```
data3<-wooldridge::meap01

attach(data3)
```

(i) Find the largest and smallest values of math4. Does the range make sense? Explain.

```
larg<-max(math4)

larg
```

```
## [1] 100
```

```
small<-min(math4)

small
```

```
## [1] 0
```

How many schools have a perfect pass rate on the math test? What percentage is this of the total sample?

```
perf_pass_rate<- data3 %>%
  filter(math4==100) %>%
  summarise(n=n())

perf_pass_rate
```

```
##    n
## 1 38
```

There are 38 students with perfect pass rate in the sample.

```
share_perf_pass_rate<-((38)/(1823)*100)

share_perf_pass_rate
```

```
## [1] 2.084476
```

The percentual share of perfect pass rate is 2.08%.

(iii) How many schools have math pass rates of exactly 50%?

```
pass_rate50<- data3 %>%
  filter(math4==50) %>%
  summarise(n=n())

pass_rate50
```

```
##    n
## 1 17
```

17 students passed in the exame with grade of 50%.

Exercise 4

Upload the database

```
data4<-wooldridge::jtrain2

attach(data4)
```

The data in JTRAIN2.RAW come from a job training experiment conducted for lowincome men during 1976–1977

(i) Use the indicator variable train to determine the fraction of men receiving job training.

```
men_jobtraining<- data4 %>%
                  filter(train == 1) %>%
                  summarise(n=n())

share_menjobtraining<-(men_jobtraining/length(train))*100

share_menjobtraining
```

```
##          n
## 1 41.57303
```

The percentage of men who received job training was 41.5%

(ii) The variable re78 is earnings from 1978, measured in thousands of 1982 dollars. Find the averages of re78 for the sample of men receiving job training and the sample not receiving job training. Is the difference economically large?

```
men_jobtraining1<-data4 %>%
                filter(train == 1) %>%
                summarise(mean(re78))
men_jobtraining1
```

```
##   mean(re78)
## 1   6.349145
```

```
men_notjobtraining<-data4 %>%
                    filter(train == 0) %>%
                    summarise(mean(re78))

men_notjobtraining
```

```
##   mean(re78)
## 1   4.554802
```

The salary of men who received job training was $6,349.
The salary of men who not received job training was $ 4,544.

The salarial gap is about $2,400, a quite large difference.

Exercise 5

Upload the database

```
data5<-wooldridge::fertil2
attach(data5)
```

The data in FERTIL2.DTA were collected on women living in the Republic of Botswana in 1988. The variable children refers to the number of living children. The variable electric is a binary indicator equal to one if the woman's home has electricity, and zero if not.

(i) Find the smallest and largest values of children in the sample. What is the average of children?

```
larg_chi<-max(children)

larg_chi
```

```
## [1] 13
```

```
small_chi<-min(children)

small_chi
```

```
## [1] 0
```

```
average_chi<-mean(children)

average_chi
```

```
## [1] 2.267828
```

The largest value are 13, the minimum 0 and the average value of children is 2.26.

(ii) What percentage of women have electricity in the home?

```
n_elec <- data5 %>%
  filter(electric == 1) %>%
  summarise(n=n())

n_elec
```

```
##     n
## 1 611
```

```
share_elec<-round((n_elec/length(electric))*100,2)

share_elec
```

```
##       n
## 1 14.01
```

Only 14.0% of women has electricity in home.

(iii) Compute the average of children for those without electricity and do the same for those with electricity.

```
chi_with_elec<- data5 %>%
          filter(electric == 1 & children > 0) %>%
          summarise(n=n())

num_children<- data5 %>%
              filter(children > 0) %>%
              summarise(n=n())

num_children
```

```
##      n
## 1 3229
```

```
share_chi_with_elec<-(chi_with_elec/num_children)*100

share_chi_with_elec
```

```
##         n
## 1 13.9362
```

Only 13.93% of childrens have electricity in home.

```
chi_without_elec<-data5 %>%
            filter(electric == 0 & children > 0) %>%
            summarise(n=n())



share_chi_without_elec<-(chi_without_elec/num_children)*100

share_chi_without_elec
```

```
##          n
## 1 85.97089
```

85.97% of children do not have electricity in home.

(iv) From part (iii), can you infer that having electricity “causes” women to have fewer children? Explain.