Homework Assignment 1

Tue Sep 04 11:50:43 2018

library(TeachingDemos)  
txtStart("Homework assignment 1.txt")

## Output being copied to text file,  
## use txtStop to end

# Homework Assignment 1  
# Name: Borná Djavdan   
# ANR: 749817  
  
# Exercise 1  
load("~/Study/Semester 3/Econometrics for ECO/R Files/R Data/bwght.RData")  
attach(data)

#1a)   
#There are as many women in the sample as there are observations in the data.   
sum(cigs>0) #There are 212 women who report smoking at least 1 cigarette per day during pregnancy.

## [1] 212

#1b)  
mean(cigs)

## [1] 2.087176

# The average number of cigarettes a woman smokes per day during pregnancy equals about 2.09. Before I answer the next question, I would like to describe what I define as a 'typical' woman. According to the sample, the typical woman doesn't smoke.  
# Therefore I conclude that this average is not a good measure for the average of the 'typical' woman. Some women smoke during pregnancy and some do not, this mean makes it look like all women smoke.

# 1c)   
mean\_smoke\_during\_pregnancy <- mean(cigs[cigs>0])  
mean\_smoke\_during\_pregnancy

## [1] 13.66509

# Among women who smoke, the average cigarettes smoked per day during pregnancy is about 13.67.   
#This average is higher than the answer in b, because now the average does not include the women who do not smoke, which would skew the average lower.   
  
# 1d)   
mean(na.omit(motheduc))

## [1] 12.93583

# The average of 'motheduc' is about 12.94, I made sure that the missing value was not used in the mean calculation, because otherwise the mean function gave 'NA' as output.   
sum(is.na(motheduc))

## [1] 1

# The amount of observations that are missing from this variable is only one (1).   
  
#1e)   
mean(faminc)

## [1] 29.02666

sd(faminc)

## [1] 18.73928

# The average family income in 1988 is $29026.66. The standard deviation of family income in 1988 is $18739.28. The unit in the data sample is in 1988 U.S. Dollars.   
  
#1f)  
rich <- as.numeric(faminc>=30)   
sum(rich)

## [1] 574

# There are 574 families (observations) categorized as 'rich' (meaning family income is above or equal to $30.000).   
  
#1g)  
table(rich,cigs)

## cigs  
## rich 0 1 2 3 4 5 6 7 8 9 10 12 15 20 30 40 46  
## 0 655 0 3 6 6 11 3 2 4 1 44 2 14 51 4 6 1  
## 1 521 3 1 1 3 8 3 2 1 0 11 3 5 11 1 0 0  
## cigs  
## rich 50  
## 0 1  
## 1 0

#There are 53 'rich' women who smoke during pregnancy and 159 women classified as 'poor' who smoke during pregnancy inside the sample.  
# Yes, I would say there is a striking difference, the number of women who smoke differs by 106 more for the 'poor’ women.   
mean(cigs[cigs>0&rich==0])

## [1] 14.62893

mean(cigs[cigs>0&rich==1])

## [1] 10.77358

# Also, the 'poor' women who smoke during pregnancy, smoke about 4 more cigarettes per day than the 'rich' women who smoke during pregnancy.   
detach(data)  
  
# Exercise 2  
load("~/Study/Semester 3/Econometrics for ECO/R Files/R Data/jtrain2.RData")  
attach(data)  
  
#2a)  
sum(train) / 445

## [1] 0.4157303

# The fraction of men receiving low-income job training is calculated using the summation of train values divided by the number of total observations in this dataset.  
#2b)   
mean(re78[train==1])

## [1] 6.349145

mean(re78[train==0])

## [1] 4.554802

# The average income of men receiving job training is about $6349.15, and the average income of men receiving no job training is about $4554.80.   
# Yes, I would say this difference would constitute as economically large.   
# This is my conclusion because the average income differs by almost 40% more for men who have received training.   
  
#2c)   
sum(unem78&train==1)/sum(unem78)

## [1] 0.3284672

sum(unem78&train==0)/sum(unem78)

## [1] 0.6715328

# The fraction of the men who received job training and are unemployed is about 0.3285.   
# The fraction of the men who did not receive job training and are unemployed is about 0.6715.   
# The difference is significant, because the fraction of men who did not receive job training and are unemployed is more than twice the fraction of the men who received job training and are unemployed.   
# Therefore, I can say that receiving job training is positively effective in gaining employment.   
  
#2d)Yes, it does appear that job training was effective, because the earnings of people receiving job training are on average significantly higher than the people receiving no job training.   
# And when comparing the unemployment rate of people who have received job training and those who have not received job training, the job training was positively effective in combatting unemployment.   
# Therefor I would conclude that it does appear that the job training program was effective.   
# What would make our conclusions more convincing would be to use panel data, a.k.a. to add the dimension of time to the data and to be able to see if the findings would be consistent over time.   
  
detach(data)  
txtStop()