ECOM20001 Econometrics 1

Week 4

Zheng Fan

The University of Melbourne

Semester 1, 2023

Introduction

Zheng Fan

- Ph.D student in Economics at Unimelb
  - ↪ Research interest: Bayesian and Financial Econometrics
- My email: fan.z@unimelb.edu.au

Seek for help:

- Ed discussion board
- Consultations: refer to Canvas for details
- Admin, assign, Covid, please reach Richard Hayes
- You may access my tutorial materials at https://github.com/zhengf1/2023ECOM1

Pre-tute

Make sure you have went through the pre-tute materials!!!

Pre-tute question 1

Suppose you collected 'ahe' from a random sample $n = 5,000$ of Victorians. You compute the sample mean of 28.25 and sample standard deviation of 10.66.

a) Conduct a two-sided hypothesis test of the null that the population mean is 28.

b) Using both the p-value and critical value approach to hypothesis testing, what do you conclude?

Pre-tute question 1

Suppose you collected 'ahe' from a random sample $n = 5,000$ of Victorians. You compute the sample mean of 28.25 and sample standard deviation of 10.66.

c) Construct a 95% confidence interval for the population mean.

Pre-tute question 1

Suppose you collected 'ahe' from a random sample $n = 5,000$ of Victorians. You compute the sample mean of 28.25 and sample standard deviation of 10.66.

d) Report the p-value for the two-sided hypothesis test of the null that the population mean is 28 as well as the 95% CI for the following sample sizes:
- $n = 2,500$ - $n = 5,000$ - $n = 10,000$ - $n = 20,000$

Pre-tute question 2

Suppose you also randomly sampled $n = 3,000$ individuals from NSW and obtained a sample mean ahe of \$30.88 and sample standard deviation \$11.22.

b) Conduct a two-sample t -test of the null that the difference in means ahe for individuals in Victoria and NSW is 0. Conduct the test, at the 5% level of significance, using both the p-value and critical value approaches.

In-tute question 1

Suppose you have a random sample of data with a population mean, μ, and you conduct the following hypothesis test:

$$H_0 : \mu = 10 \qquad H_1 : \mu \neq 10$$

Having performed the test, you obtain a p-value of 0.07.

a. Does a 90% CI for the population mean contain $\mu = 10$? Please explain.

b. With only the information provided in the question, can you determine if $\mu = 8$ is contained in the 90% CI? Explain.

In-tute question 2

Using R, what is the sample mean and standard deviation of *ahe* for males and females?

```
setwd("~/Library/CloudStorage/Dropbox/01 UoM-Teaching/2023-S1-Ecom1/Tut
data=read.csv("tute4_cps.csv")
## Mean and standard deviation of earnings for females
mean(data$ahe[data$female==1])
## [1] 17.80898

sd(data$ahe[data$female==1])
## [1] 8.873493

## Mean and standard deviation of earnings for males
mean(data$ahe[data$female==0])
## [1] 20.57906

sd(data$ahe[data$female==0])
## [1] 10.5533
```

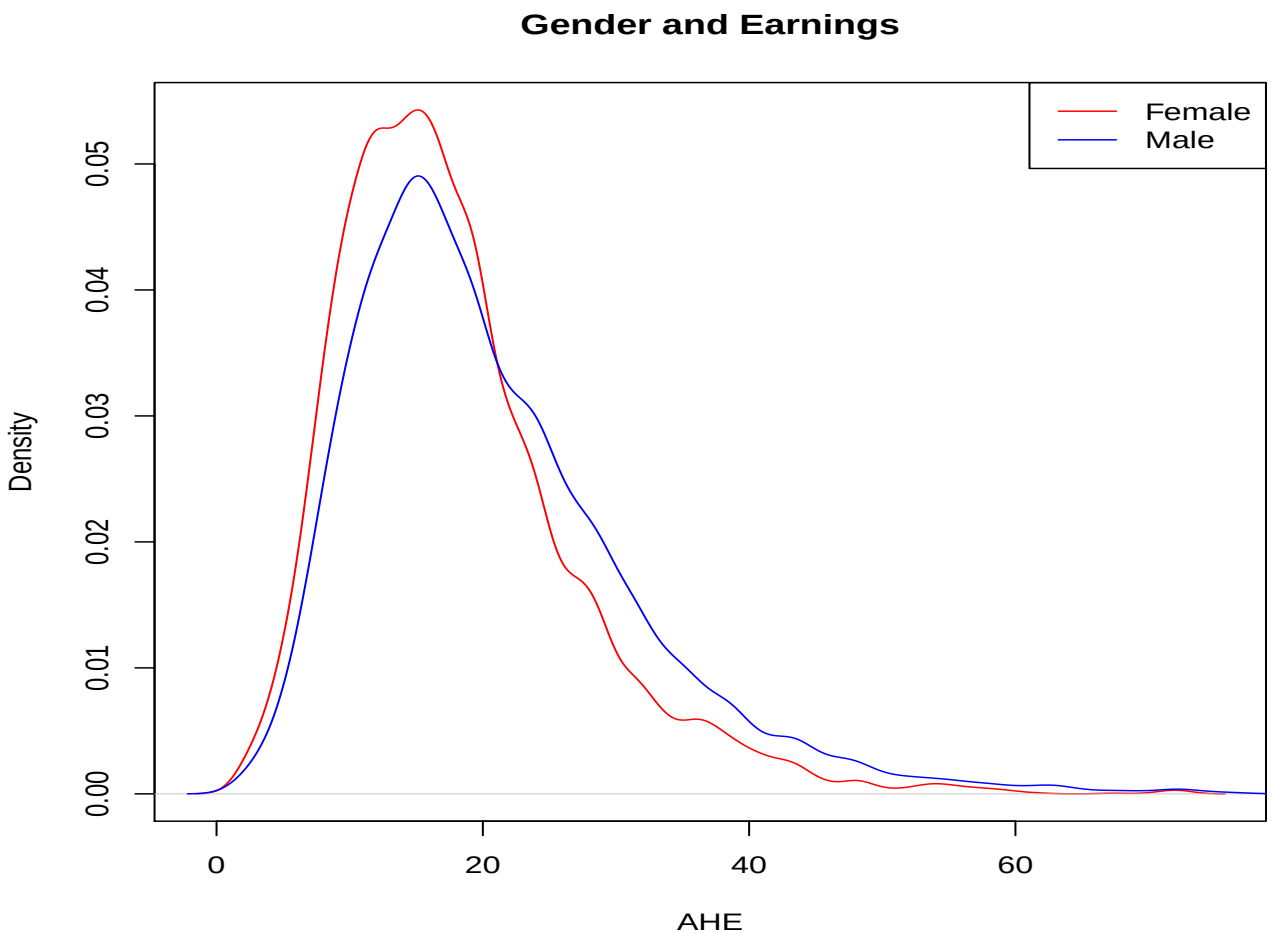
In-tute question 2

Discuss these numbers and the density plots produced for *ahe* for males and females (reproduced below), which reveals what is known as the gender wage gap.

- Provide economic explanations for your results. (Recall from tutorial 2 that an economic explanation focuses on the costs and benefits of a behaviour for explaining empirical patterns).
- In this example, what are the different economic costs and benefits among males and females in generating household earnings?

In-tute question 2

```
plot(density(data$ahe[data$female==1]), col="red",lty=1,xlab="AHE",
     main="Gender and Earnings")
lines(density(data$ahe[data$female==0]), col="blue",lty=1)
legend("topright", legend=c("Female", "Male"), col=c("red", "blue"),
       lty=c(1,1))
```



In-tute question 3

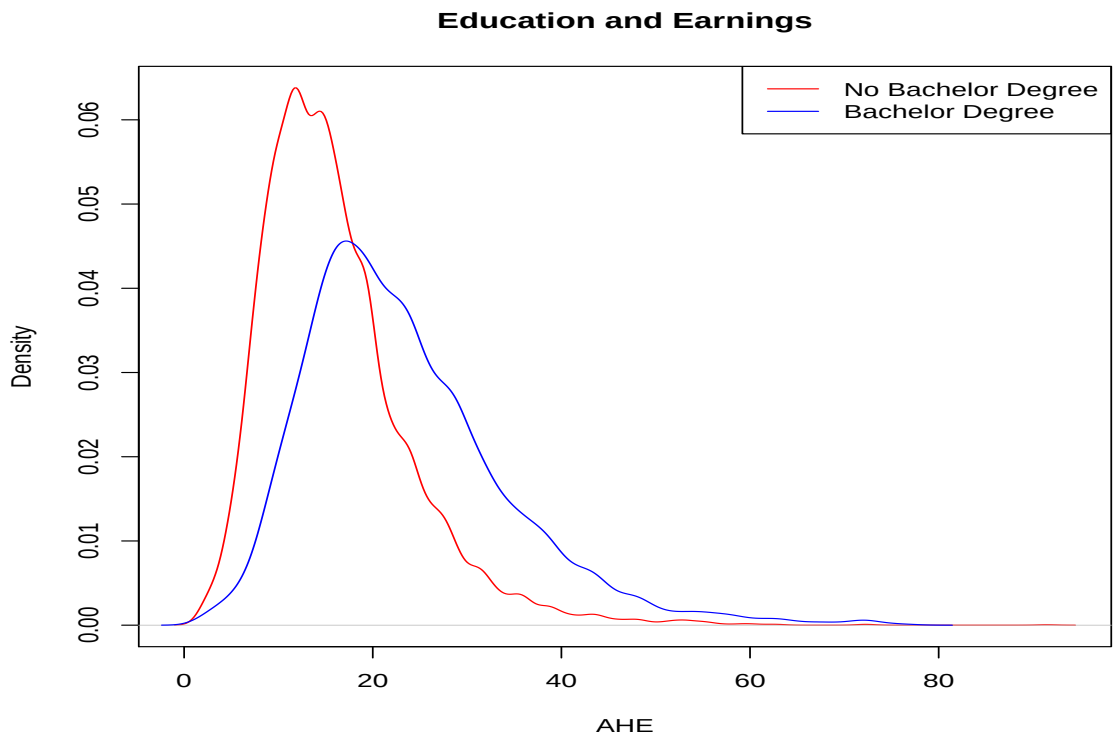
Using R, what is the sample mean and standard deviation for individuals of *ahe* for individuals with and without bachelor's degrees?

```
## Mean and standard deviation of earnings for individuals with bachelo
mean(data$ahe[data$bachelor==1])
## [1] 23.34672

sd(data$ahe[data$bachelor==1])
## [1] 10.71684

## Mean and standard deviation of earnings for individuals without bach
mean(data$ahe[data$bachelor==0])
## [1] 16.04614

sd(data$ahe[data$bachelor==0])
## [1] 7.855756
```

In-tute question 3

Discuss these numbers and the density plots produced for *ahe* for individuals with and without bachelor's degrees.

```
plot(density(data$ahe[data$bachelor == 0]), col = "red",
     lty = 1, xlab = "AHE", main = "Education and Earnings")
lines(density(data$ahe[data$bachelor == 1]), col = "blue", lty = 1)
legend("topright", legend = c("No Bachelor Degree", "Bachelor Degree"),
       col = c("red", "blue"), lty = c(1,1))
```



In-tute question 4

There does seem to be a difference in the average ahe between males and females who have degrees, and without degrees in 2012.

```
female_nodegree2012 = data$ahe[data$female==1 & data$year==2012 &
                                 data$bachelor==0]
male_nodegree2012 = data$ahe[data$female==0 & data$year==2012 &
                               data$bachelor==0]
diff1 = mean(female_nodegree2012) - mean(male_nodegree2012)
print(diff1)
## [1] -3.924525

se = sqrt(sd(female_nodegree2012)^2/length(female_nodegree2012) +
            sd(male_nodegree2012)^2/length(male_nodegree2012))
(t = diff1/se)
## [1] -15.36113
```

In-tute question 4

Alternatively, we may just simply

```
t.test(data$ahe[data$female==1 & data$year==2012 & data$bachelor==0],
       data$ahe[data$female==0 & data$year==2012 & data$bachelor==0])
##
##  Welch Two Sample t-test
##
## data:  data$ahe[data$female == 1 & data$year == 2012 & data$bachelor
## t = -15.361, df = 3269.9, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4.425451 -3.423600
## sample estimates:
## mean of x mean of y
##  13.11905  17.04357
```

In-tute question 5

The dataset consumption.csv contains a population of 60 families. The variables are:

- Consumption: family consumption in $/week
- Income: family disposable income in $/week

a). What is the population mean of consumption , i.e. E(Consumption)

b). What is the conditional mean E(Consumption|Income$\leq$100)?

```
data1 = read.csv(file="consumption.csv")

(ymean = mean(data1$Consumption))
## [1] 121.2

(ycondmean = mean(data1[data1$Income <= 100, "Consumption"], na.rm = TR
## [1] 71.54545
```

In-tute question 5

c). Run the following in the population and confirm the Population Regression Line (PRL) is Consumption=17+0.6Income

d). Using the R code below, construct a random sample of 13 families for the population. Call it Sample A. Run the following regression and also create a scatterplot.