Empirical Exercise 1, Stock and Watson Chapter 3

Econ 440 - Introduction to Econometrics

Patrick Toche, ptoche@fullerton.edu

21 March 2022

(a)

(i)

Compute the sample mean for average hourly earnings (AHE) in 1996 and 2015.

```
mu.1996 = mean(df[df$year == 1996,]$ahe)
mu.1996
```

```
## [1] 12.693
```

```
mu.2015 = mean(df[df$year == 2015,]$ahe)
mu.2015
```

```
## [1] 21.237
```

(ii)

Compute the sample standard deviation for AHE in 1996 and 2015.

```
sd.1996 = sd(df[df$year == 1996,]$ahe)
sd.1996
```

```
## [1] 6.359
```

```
sd.2015 = sd(df[df$year == 2015,]$ahe)
sd.2015
```

```
## [1] 12.125
```

(iii)

Construct a 95% confidence interval for the population means of AHE in 1996 and 2015.

Let's make a function, so we don't have to copy-paste the same code twice:

```
make_ci <- function(x, alpha=0.05){
    mu = mean(x)
    sd = sd(x)
    n = length(x)
    t = qt(1-alpha/2, n-1)  # for n large, qnorm(1-alpha/2) will do
    se = sd/sqrt(n)
    me = t*se
    ci = c(mu-me, mu+me)
    ci
}
```

```
ci.1996 = make_ci(df[df$year == "1996",]$ahe)
ci.1996
```

```
## [1] 12.534 12.853
```

```
ci.2015 = make_ci(df[df$year == "2015",]$ahe)
ci.2015
```

```
## [1] 20.955 21.520
```

(iv)

Construct a 95% confidence interval for the change in the population means of AHE between 1996 and 2015.

Sample sizes:

```
n.1996 = length(df[df$year == 1996,]$ahe)
n.1996
```

```
## [1] 6103
```

```
n.2015 = length(df[df$year == 2015,]$ahe)
n.2015
```

```
## [1] 7098
```

Sample variances:

```
var.1996 = var(df[df$year == 1996,]$ahe)
var.1996
```

```
## [1] 40.437
```

```
var.2015 = var(df[df$year == 2015,]$ahe)
var.2015
```

```
## [1] 147
```

Standard error:

```
se = sqrt((var.1996/n.1996 + var.2015/n.2015))
se
```

```
## [1] 0.16534
```

Critical z-value: Since the sample sizes are large, we can use the standard normal distribution.

```
alpha = 0.05
z = qnorm(1-alpha/2)
z
```

```
## [1] 1.96
```

Margin of error:

```
me = z*se
me
```

```
## [1] 0.32405
```

Confidence interval for the difference in means:

```
ci = c(mu.2015-mu.1996-me, mu.2015-mu.1996+me)
ci
```

```
## [1] 8.2201 8.8682
```

(b)

In 2015, the value of the Consumer Price Index (CPI) was 237.0. In 1996, the value of the CPI was 156.9. Repeat (a), but use AHE measured in real 2015 dollars ($2015); that is, adjust the 1996 data for the price inflation that occurred between 1996 and 2015.

We create a new variable for the CPI-adjusted AHE, using 2015 prices, denoted *ahe.real* (for the year 2015, we keep the existing data).

```
df$ahe.real <- df$ahe
df[df$year == 1996,]$ahe.real <- df[df$year == 1996,]$ahe.real * 237.0/156.9

# 2015 values are unchanged
mu.2015.real = mu.2015
mu.2015.real
```

```
## [1] 21.237
```

```
mu.1996.real = mean(df[df$year == 1996,]$ahe.real)
mu.1996.real
```

```
## [1] 19.173
```

```
sd.2015.real = sd.2015
sd.2015.real
```

```
## [1] 12.125
```

```
sd.1996.real = sd(df[df$year == 1996,]$ahe.real)
sd.1996.real
```

```
## [1] 9.6054
```

```
make_ci(df[df$year == "1996",]$ahe.real)
```

```
## [1] 18.932 19.414
```

Or simply change the units of the already computed values:

```
sd.1996 * 237.0/156.9
```

```
## [1] 9.6054
```

```
ci.1996 * 237.0/156.9
```

```
## [1] 18.932 19.414
```

Confidence interval for difference in means:

```
make_ci_diff <- function(x1, x2, alpha=0.05){
    # means
    m1 = mean(x1)
    m2 = mean(x2)
    # sample variances
    v1 = var(x1)
    v2 = var(x2)
    # sample sizes:
    n1 = length(x1)
    n2 = length(x2)
    # standard error:
    se = sqrt(v1/n1+v2/n2)
    # critical value:
```

```
    t = qt(1-alpha/2, df=n1+n2-2)  # or just qnorm(1-alpha/2)
    # confidence interval
    ci = c(m2-m1-t*se, m2-m1+t*se)
    ci
}
# variables
x1 = df[df$year == "1996",]$ahe.real
x2 = df[df$year == "2015",]$ahe.real
make_ci_diff(x1, x2)
```

```
## [1] 1.6930 2.4351
```

Or by leveraging R's built-in *t.test* function:

```
t.test(df[df$year == 2015,]$ahe.real, df[df$year == 1996,]$ahe.real)
```

```
##
##  Welch Two Sample t-test
##
## data:  df[df$year == 2015, ]$ahe.real and df[df$year == 1996, ]$ahe.real
## t = 10.9, df = 13113, p-value <2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.6930 2.4351
## sample estimates:
## mean of x mean of y
##    21.237    19.173
```

(c)

If you were interested in the change in workers' purchasing power from 1996 to 2015, would you use the results from (a) or (b)? Explain.

Purchasing power is measured in real terms, so we would use the results from (b).

(d)

Using the data for 2015:

(i)

Construct a 95% confidence interval for the mean of AHE for high school graduates.

```
make_ci(df[df$year == "2015" & df$bachelor == 0,]$ahe)
```

```
## [1] 16.093 16.670
```

(ii)

Construct a 95% confidence interval for the mean of AHE for workers with a college degree.

```
make_ci(df[df$year == "2015" & df$bachelor == 1,]$ahe)
```

```
## [1] 25.192 26.038
```

(iii)

Construct a 95% confidence interval for the difference between the two means.

```r
x1 = df[df$year == "2015" & df$bachelor == 0,]$ahe.real
x2 = df[df$year == "2015" & df$bachelor == 1,]$ahe.real
make_ci_diff(x1, x2)
```

```
## [1] 8.7223 9.7455
```

(e)

Repeat (d) using the 1996 data expressed in \$2015.

(i)

```r
make_ci(df[df$year == "1996" & df$bachelor == 0,]$ahe.real)
```

```
## [1] 16.013 16.523
```

(ii)

```r
make_ci(df[df$year == "1996" & df$bachelor == 1,]$ahe.real)
```

```
## [1] 22.635 23.441
```

(iii)

```r
x1 = df[df$year == "1996" & df$bachelor == 0,]$ahe.real
x2 = df[df$year == "1996" & df$bachelor == 1,]$ahe.real
make_ci_diff(x1, x2)
```

```
## [1] 6.2932 7.2464
```

(f)

Using appropriate estimates, confidence intervals, and test statistics, answer the following questions:

(i)

Did real (inflation-adjusted) wages of high school graduates increase from 1996 to 2015?

```r
x1 = df[df$year == "1996" & df$bachelor == 0,]$ahe.real
x2 = df[df$year == "2015" & df$bachelor == 0,]$ahe.real
make_ci_diff(x1, x2)
```

```
## [1] -0.27200  0.49776
```

The confidence interval associated with the mean difference in the real wages of high-school graduates is $(-0.27, 0.50)$. Since 0 belongs to the interval, there is no statistical evidence of an increase.

(ii)

Did real wages of college graduates increase?

```r
x1 = df[df$year == "1996" & df$bachelor == 1,]$ahe.real
x2 = df[df$year == "2015" & df$bachelor == 1,]$ahe.real
make_ci_diff(x1, x2)
```

```
## [1] 1.9932 3.1608
```

The confidence interval associated with the mean difference in the real wages of college graduates is $(1.99, 3.16)$, well above 0 and thus clearly pointing to a statistically significant increase.

(iii)

Did the gap between earnings of college and high school graduates increase? Explain.

```
gap1 = mean(df[df$year == "1996" & df$bachelor == 1,]$ahe.real)
       -mean(df[df$year == "1996" & df$bachelor == 0,]$ahe.real)
```

```
## [1] -16.268
```

```
gap2 = mean(df[df$year == "2015" & df$bachelor == 1,]$ahe.real)
       -mean(df[df$year == "2015" & df$bachelor == 0,]$ahe.real)
```

```
## [1] -16.381
```

```
gap2-gap1
```

```
## [1] 2.577
```

The gap between the mean earnings of college and high school graduates has increased by about 2.46.