HW05_Sampathirao_A

Anvita Sampathirao

6/10/2019

R Markdown

#1a.

$$H_0 : \mu \geq 2.03$$
$$H_1 : \mu < 2.03$$

#1b.

$$t - stat = \frac{\bar{x} - \mu_0}{se}$$

```
serr<-0.07/10
t1<-(1.75-2.03)/serr
t1
```

```
## [1] -40
```

#1c.

```
lowertail<-qt(0.05,99,lower.tail = TRUE)
lowertail
```

```
## [1] -1.660391
```

i.e.,

$$RR : (-\infty, -1.66)$$

Because t-statistic of our sample set, -40 lies in the rejection region of t distribution, we can thus reject the null hypothesis.

#1d.

```
lower_lim<-(1.75-(1.64*serr))
lower_lim
```

```
## [1] 1.73852
```

```
upper_lim<-(1.75+(1.64*serr))
upper_lim
```

```
## [1] 1.76148
```

The 95% confidence interval thus is:

$$CI = [1.73852, 1.76148]$$

#1e.

```r
pt(t1,99,lower.tail=TRUE)
```

```
## [1] 3.192543e-63
```

$$Since, p - value \leq 0.05$$

Thus, null hypothesis can be rejected at 95% significance level

```r
#t.test to verify
set.seed(1)
test_sample<-rnorm(100, mean = 1.75, sd = 0.07)
t.test(test_sample,mu=2.03)
```

```
##
##  One Sample t-test
##
## data:  test_sample
## t = -43.321, df = 99, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 2.03
## 95 percent confidence interval:
##  1.745147 1.770098
## sample estimates:
## mean of x
##  1.757622
```

#2a.

$$H_0 : \mu_{\text{men}} - \mu_{\text{women}} = 0$$

$$H_0 : \mu_{\text{men}} - \mu_{\text{women}} \neq 0$$

#2b.

$$df = 2 * n - 2$$

$$S.E. = s/\sqrt{n}$$

$$t - stat = \frac{\bar{x_1} - \bar{x_2}}{\sqrt{se_1^2 + se_2^2}} = \frac{x_1 - x_2}{\sqrt{2} * S.E.}$$

```r
serr2<-200/sqrt(50)
serr2
```

```
## [1] 28.28427
```

```r
alpha<-0.05
t2<-(1124-1245)/(sqrt(2)*serr2)
t2
```

```
## [1] -3.025
```

#2c.

```
df<-(2*50)-2
df
```

```
## [1] 98
```

```
v<-c(qt(c(alpha/2,1-(alpha/2)),df))
v
```

```
## [1] -1.984467  1.984467
```

i.e.,

$$RR : (-\infty, -1.984467) \cap (1.984467, \infty)$$

Because t-statistic of our sample sets, -3.025 lies in the rejection region of t distribution, we can thus reject the null hypothesis.

#2d.

$$CI_(\mu_1 - \mu_2) : [(\bar{x_1} - \bar{x_2}) \pm 1.98 * se]$$

$$\bar{x_1} - \bar{x_2} = 1124 - 1245 = -121$$

```
lowerlimit<-(-121-(1.98*serr2))
lowerlimit
```

```
## [1] -177.0029
```

```
upperlimit<-(-121+(1.98*serr2))
upperlimit
```

```
## [1] -64.99714
```

#2e.

```
pt(t2,df)*2
```

```
## [1] 0.003174979
```

$$Since, p - value \leq 0.05$$

Thus, null hypothesis can be rejected at 95% significance level

```
#t.test to verify
set.seed(1)
men<- rnorm(50, 1124,200)
women<-rnorm(50,1245,200)
t.test(men,women)
```

```
##
##  Welch Two Sample t-test
##
## data:  men and women
## t = -3.4444, df = 95.793, p-value = 0.0008507
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -196.05382  -52.69745
## sample estimates:
## mean of x mean of y
##  1144.090  1268.465
```

#3a.

$$H_0 : \mu_t - \mu_c = 0$$
$$H_1 : \mu_t - \mu_c > 0$$

Therefore we use a right tailed t test

```
x_t<-78
x_c<-75
sd_t<-20
sd_c<-5
n<-50
x_diff<-x_t-x_c
x_diff #mean difference between the 2 groups
```

```
## [1] 3
```

```
se_t<-sd_t/sqrt(n)
se_t #standard error for treatment group
```

```
## [1] 2.828427
```

```
se_c<-sd_c/sqrt(n)
se_c #standard error for control group
```

```
## [1] 0.7071068
```

```
se_diff<-sqrt((se_c)^2+(se_t)^2)
se_diff #standard error difference between the 2 groups
```

```
## [1] 2.915476
```

```
df1<-(se_diff^4)/((se_t^4/(n-1))+(se_c^4/(n-1)))
df1 #degree of freedom
```

```
## [1] 55.10117
```

```r
t3<-x_diff/se_diff
t3 #test statistic
```

```
## [1] 1.028992
```

```r
uppertail<-function(alpha1)
#function to determine if we can reject null hypothesis at the given significance level
{
c<-qt(alpha1,df1,lower.tail = FALSE)
if(t3>=c)
{
  CI<-c(x_diff-(c*se_diff),x_diff+(c*se_diff))
  print("Reject Null hypothesis")
}
else
{
  CI<-c(x_diff-(c*se_diff),x_diff+(c*se_diff))
  print("Fail to Reject Null Hypothesis, because mu_0=0 lies in CI, which is")
  return(CI)
}
}
uppertail(0.10) #significance level 90%
```

```
## [1] "Fail to Reject Null Hypothesis, because mu_0=0 lies in CI, which is"

## [1] -0.7816794  6.7816794
```

```r
uppertail(0.05) #significance level 95%
```

```
## [1] "Fail to Reject Null Hypothesis, because mu_0=0 lies in CI, which is"

## [1] -1.877537  7.877537
```

```r
uppertail(0.01) #significance level 99%
```

```
## [1] "Fail to Reject Null Hypothesis, because mu_0=0 lies in CI, which is"

## [1] -3.985333  9.985333
```

i.e., We do not have sufficient evidence to test our null hypothesis.

```r
set.seed(1)
treatment<- rnorm(50, 78,20)
control<-rnorm(50,75,5)
t.test(treatment,control)
```

```
##
##  Welch Two Sample t-test
##
```

```
## data:  treatment and control
## t = 1.8056, df = 57.258, p-value = 0.07625
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.4818144  9.3264811
## sample estimates:
## mean of x mean of y
##  80.00897  75.58663
```

#4a.

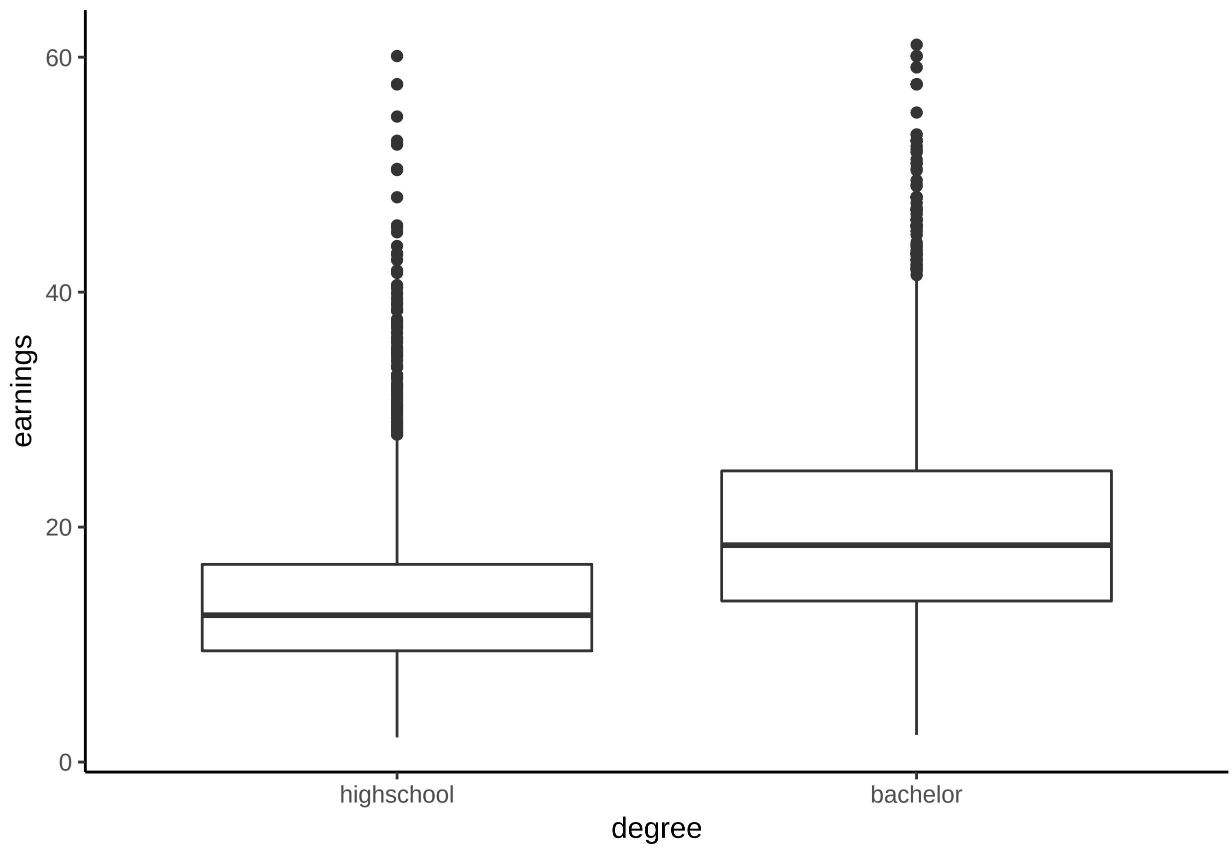
```r
library(AER)
```

```
## Loading required package: car

## Loading required package: carData

## Loading required package: lmtest

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

## Loading required package: sandwich

## Loading required package: survival
```

```r
data("CPSSW04")
library(ggplot2)
```

```
## Registered S3 methods overwritten by 'ggplot2':
##   method         from
##   [.quosures     rlang
##   c.quosures     rlang
##   print.quosures rlang
```

```r
ggplot(CPSSW04, aes(x = degree, y = earnings)) + geom_boxplot() + theme_classic()
```



#4b.

```
t.test(earnings ~degree, data = CPSSW04)
```

```
##
##  Welch Two Sample t-test
##
## data:  earnings by degree
## t = -34.486, df = 6369.9, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -6.866823 -6.128134
## sample estimates:
## mean in group highschool   mean in group bachelor
##                 13.80961                 20.30709
```

#4c.

```
t.test(CPSSW04$earnings[CPSSW04$age>=25 & CPSSW04$age<=29],
       CPSSW04$earnings[CPSSW04$age>=30 & CPSSW04$age<=34])
```

```
##
##  Welch Two Sample t-test
##
## data:  CPSSW04$earnings[CPSSW04$age >= 25 & CPSSW04$age <= 29] and CPSSW04$earnings[CPSSW04$age >= 30
```

```
## t = -12.59, df = 7967.3, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.795689 -2.042393
## sample estimates:
## mean of x mean of y
##  15.47681  17.89586
```