#amulya geereddy

Preprocessing Commands:

library(rio)

library(moments)

original=import("6304 Module 3 Assignment Data.xlsx")

colnames(original)=tolower(make.names(colnames(original)))

attach(original)

set.seed(62067273)

set1=subset(original,original$authority.name =="Capital District Transportation Authority")

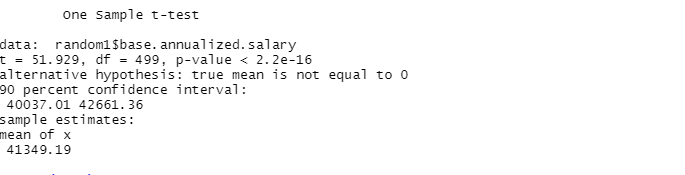
random1=set1[sample(1:nrow(set1),500),]

set2=subset(original,original$authority.name =="Central New York Regional Transportation Authority")

random2=set2[sample(1:nrow(set2),500),]

Question1:

result1=t.test(random1$base.annualized.salary,conf.level = 0.9)



From the result we can say that there is 90% chance that the true population mean for base annualized salary lies somewhere between 42661.36 and 40037.01.

Question2:

mean(set1$base.annualized.salary)

> mean(set1$base.annualized.salary)

[1] 41748.75

From the above result it can be said that, we have 90% certainty that the mean of base annualized salary for Capital District Transportation Authority dataset is between the confidence intervals of 40037.01 and 42661.36. Here, 90% is the confidence we have oon the true mean lying in the intervals.

Question3:

result1=t.test(random1$base.annualized.salary,conf.level = 0.90)

result1

result2=t.test(random1$base.annualized.salary,conf.level = 0.95)

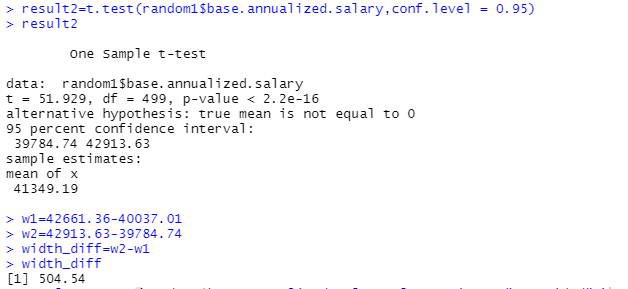
result2

w1=42661.36-40037.01

w2=42913.63-39784.74

width\_diff=w2-w1

width\_diff



From Comparing Q1 and Q2 we can say that the 95% confidence interval is slightly wider than 90% confidence interval.

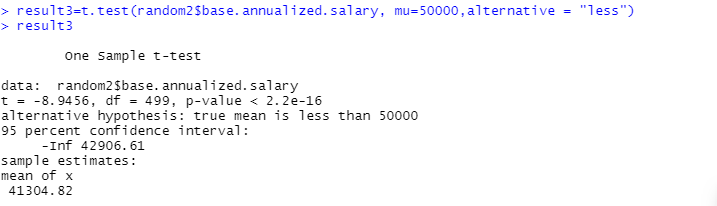
Question4:

result3=t.test(random2$base.annualized.salary, mu=50000,alternative = "less")

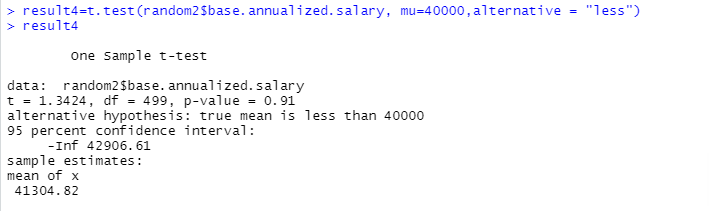
result3

result4=t.test(random2$base.annualized.salary, mu=40000,alternative = "less")

result4



For As p-value is much lesser than alpha i.e. 0.05 we reject the null hypothesis so we have evidence to show that there could be population mean less than $50000.



As p-value is much higher than the alpha i.e. 0.05 we fail to reject the null hypothesis so we don’t have proper evidence to show that there could be population mean less than $40000

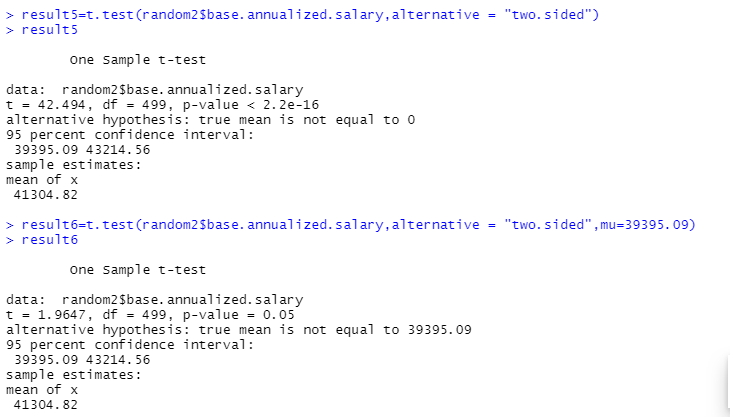
Question5:

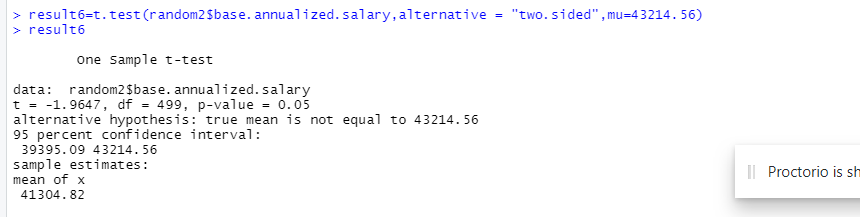
result5=t.test(random2$base.annualized.salary,alternative = "two.sided")

result5

result6=t.test(random2$base.annualized.salary,alternative = "two.sided", mu=39395.09)

result6



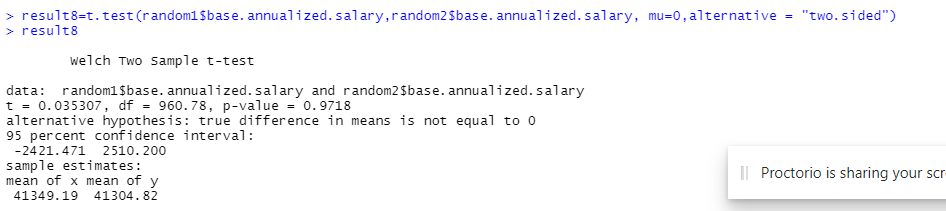


From the above data of Central New York Regional Transportation Authority dataset we can say that the mu value of 39395.09(lowerlimit) or mu value of 43214.56(upperlimit) would result in the pvalue of 0.05.

Question6:

result8=t.test(random1$base.annualized.salary,random2$base.annualized.salary, mu=0,alternative = "two.sided")

result8



As p-value is much higher than alpha i.e 0.05. we fail to reject null hypothesis. So we don’t have proper evidence to show that there could be some differences in population mean for base annualized salary for employees.