**#U54500765**

**#Name – Manish Sanjay Talekar**

**#QMB Assignment 3**

**#Preprocessing**

**Q1**.

> master\_data <- read\_xlsx("6304 Module 3 Assignment Data.xlsx")

**Q2**.

> dist\_transp <- subset(master\_data,

+ authority.name=="Capital District Transportation Authority")

> regional\_transp <- subset(master\_data,

+ authority.name=="Central New York Regional Transportation Authority")

> set.seed(54500765)

> dist\_samp <- sample\_n(dist\_transp, 500)

> set.seed(54500765)

> regional\_samp <- sample\_n(regional\_transp, 500)

**#Analysis**

**Q1.**

> a1 <- t.test(dist\_samp$base.annualized.salary, conf.level = 0.9)

> a1

One Sample t-test

data: dist\_samp$base.annualized.salary

t = 50.759, df = 499, p-value < 2.2e-16

alternative hypothesis: true mean is not equal to 0

90 percent confidence interval:

40333.09 43039.82

sample estimates:

mean of x

41686.46

INTERPRETATION – The true mean of the population lies with in the confidence interval of (40333.09, 43039.82).

**Q2.**

> mean(dist\_transp$base.annualized.salary)

[1] 41748.75

INTERPRETATION – The true mean of the “Capital District Transportation Authority” lies within the confidence interval of (40333.09, 43039.82).

**Q3.**

> a2 <- t.test(dist\_samp$base.annualized.salary)

> a2

One Sample t-test

data: dist\_samp$base.annualized.salary

t = 50.759, df = 499, p-value < 2.2e-16

alternative hypothesis: true mean is not equal to 0

95 percent confidence interval:

40072.90 43300.01

sample estimates:

mean of x

41686.46

> width1 <- a1$conf.int[2] - a1$conf.int[1]

> width1

[1] 2706.732

> width2 <- a2$conf.int[2] - a2$conf.int[1]

> width2

[1] 3227.109

> width2/width1 \* 100

[1] 119.2253

INTERPRETATION – 95% confidence interval is approximately 119% wider than the 90% confidence interval.

**Q4.**

> t.test(regional\_samp$base.annualized.salary, mu=50000,

+ alternative = "less")

One Sample t-test

data: regional\_samp$base.annualized.salary

t = -8.9196, df = 499, p-value < 2.2e-16

alternative hypothesis: true mean is less than 50000

95 percent confidence interval:

-Inf 43213.03

sample estimates:

mean of x

41674.96

#Reject null hypothesis

> t.test(regional\_samp$base.annualized.salary, mu=40000,

+ alternative = "less")

One Sample t-test

data: regional\_samp$base.annualized.salary

t = 1.7946, df = 499, p-value = 0.9633

alternative hypothesis: true mean is less than 40000

95 percent confidence interval:

-Inf 43213.03

sample estimates:

mean of x

41674.96

#Fail to reject null hypotheses

INTERPRETATION – Based on the above tests, I can say that population mean Base Annualized Salary of “Central New York Regional Transportation Authority” is less than $50,000 but greater than $40000.

**Q5.**

> t.test(regional\_samp$base.annualized.salary, mu=40000,

+ alternative = "two.sided")

One Sample t-test

data: regional\_samp$base.annualized.salary

t = 1.7946, df = 499, p-value = 0.07333

alternative hypothesis: true mean is not equal to 40000

95 percent confidence interval:

39841.19 43508.72

sample estimates:

mean of x

41674.96

> t.test(regional\_samp$base.annualized.salary, mu=39841.2,

+ alternative = "two.sided")

One Sample t-test

data: regional\_samp$base.annualized.salary

t = 1.9647, df = 499, p-value = 0.05

alternative hypothesis: true mean is not equal to 39841.2

95 percent confidence interval:

39841.19 43508.72

sample estimates:

mean of x

41674.96

INTERPRETATION – mu value of “39841.2” gives the pvalue of 0.05.

**Q6.**

> t.test(dist\_samp$base.annualized.salary,

+ regional\_samp$base.annualized.salary)

Welch Two Sample t-test

data: dist\_samp$base.annualized.salary and regional\_samp$base.annualized.salary

t = 0.0092489, df = 982.1, p-value = 0.9926

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-2428.175 2451.172

sample estimates:

mean of x mean of y

41686.46 41674.96

#Fail to reject the null hypotheses.

INTERPRETATION – Based on the sample test above, there is a very small difference between the mean of Base Annualized Salary for employees of the two agencies which is insignificant. Therefore, I can conclude that the difference between the means is nearly zero and fail to reject the null hypotheses.