```
library(car)
library(tidyverse)
X1662617767_data = read_excel("1662617767_data.xlsx")
data= X1662617767_data

############ Step-1 ################
head(data)
tail(data)

############ Step-2 #################
```

(a)  Calculating Means of Pre and Post Variables

```
mean(data$Pre)
mean(data$Post)
```

(b)  Calculating Median of Pre and Post Varibles

```
median(data$Pre)
median(data$Post)
```

(c) Calculating Mode of Pre and Post Varibales

```
install.packages("statip")
library(statip)

mfv(data$Pre)
mfv(data$Pre,5)
head(mfv(data$Pre,5))
head(mfv(data$Post,5))
```

(d) Calculating First and Third Quantile for Pre and Post Variables

```
quantile(data$Pre, 0.25)
quantile(data$Pre, 0.75)
quantile(data$Post, 0.25)
quantile(data$Post, 0.75)
```

(e) Calculating range of pre and post Variables

```
range(data$Pre)
range(data$Post)
```

(f) Calculating Variance and SD of Pre and Post Variables

```
var(data$Pre)
var(data$Post)
sd(data$Pre)
sd(data$Post)
```

(g) Calculating coefficient of Variable and MAD of Pre and Post Variabbles

```
cv= sd(data$Pre) / mean(data$Pre)*100
cv
cv<-sd(data$Post) / mean(data$Post)*100
```

```
cv
mad(data$Pre)
mad(data$Post)
```

(h) Calculating InterQuartile Range of Pre and Post Variables

```
summary(data$Pre)
IQR(data$Pre)
summary(data$Post)
IQR(data$Post)
```

######################## Step-3 #####################

3. Measuring skewness of Pre and Post Variables and
applying the Agostino Test to check the skewness

```
Norm=rnorm(1000)
skewness(Norm)
skewness(data$Pre)
skewness(data$Post)
agostino.test(Norm)
agostino.test(data$Pre)
agostino.test(data$Post)
```

###################### Step - 4 ########################

4.Checking Kurtosis and using Anscombe Test on Pre and Post Variables

```
kurtosis(Norm)
kurtosis(data$Pre)
kurtosis(data$Post)
anscombe.test(Norm)
anscombe.test(data$Pre)
```

#################### Step-5 #############################

5.Plotting graph to check the skewness and peakedness of Pre and Post Variables

```
plot(data$Pre)
plot(density(data$Pre))
plot(density(data$Pre),col="blue",lwd=1,main="Density-Graph",xlab="data")
plot(density(data$Post),col="red",lwd=1,main="Density-Graph",xlab="data")
plot(density(data$Pre),col="blue",lwd=1,main="Density-Graph",xlab="data",
    xlim=c(1,7), ylim=c(0,0.8))
lines(density(data$Post), col="red",lwd=1)
legend("topleft", c("Post","Pre"), fill=c("red","blue"))
```

#################### Step-6 ############################

6. Calculate Frequency and Relative Frequency of Cold Drink Variable

```
SoftDrink=data$'Cold-Drink'
table(SoftDrink)
SoftDrink=cbind.data.frame(table(SoftDrink))
```

```
SoftDrink$RelativeFreq=SoftDrink$Freq/sum(SoftDrink$Freq)

################### Step - 7 ###########################
```

7. Creating a Pie chart to show Soft Drink Preferences

```
pie(SoftDrink$RelativeFreq, labels = paste0(round(100*SoftDrink$RelativeFreq,2),"%"),
   main="Soft_Drink",
   col=c("brown","blue","green","pink","yellow","orange"))
legend("topright",c("CC","CD","DTC","Pep","PEP","PSI","Sprite"),cex=0.5,
      fill=c("brown","blue","green","pink","yellow","orange"))
```

Creating a Bar Graph to show Soft Drink Preferences

```
barplot(SoftDrink$RelativeFreq, names.arg=SoftDrink$SoftDrink, col =
         c("brown","blue","green","pink","yellow","orange"))

################### Step- 8 ################################
```

8. Creating a Density graph of Cold Drink

```
plot(density(SoftDrink$Freq), main="Density-Graph", xlab="Data",col=c("blue"), lwd=2)
kurtosis(SoftDrink$Freq)
anscombe.test((SoftDrink$Freq))
skewness(SoftDrink$Freq)

##################### Step-9 ##############################
```

9. Converting the ‘Status’, ‘Rating’, and ‘Outlook’ variables into factor types and summarize them

```
data$Status=as.factor(data$Status)
data$Rating=as.factor(data$Rating)
data$Outlook=as.factor(data$Outlook)
summary(data$Status)
summary(data$Rating)
summary(data$Outlook)
str(data)

##################### Step- 10 ###########################
```

10. Calculating the difference in the average pre-training satisfaction ratings of member and observer status and for the post-training member and observer status

```
Member=subset(data, data$Status=="Member")
Observer=subset(data, data$Status=="Observer")
MemPre=mean(Member$Pre)
ObsPre=mean(Observer$Pre)
DiffPre=MemPre-ObsPre
DiffPre
MemPost=mean(Member$Post)
ObsPost=mean(Observer$Post)
```

```
DiffPost=MemPost-ObsPost
DiffPost
```

########################### Step-11 #########################

11. Computing the average pre-satisfaction and post-satisfaction ratings of employees with a ‘Stable’ Outlook

```
Stab=subset(data,data$Outlook=="Stable")
MeanPre=mean(Stab$Pre, na.rm=TRUE)
MeanPre
MeanPost=mean(Stab$Post, na.rm=TRUE)
MeanPost
```

########################## Step-12 ###########################

12. Constructing a confidence interval at a 2.5%, 5%, and 1% level of significance for the salary variable

```
Salary=data$Salary; n=30
m=mean(Salary)
s=sd(Salary)
n=length(Salary); dof=n-1
t=qt(0.95, dof); t=qt(0.95, dof); m-t*s/sqrt(n); m+t*s/sqrt(n)
t=qt(0.975,dof); t=qt(0.975,dof); m-t*s/sqrt(n); m+t*s/sqrt(n)
t=qt(0.995,dof); t=qt(0.995,dof); m-t*s/sqrt(n); m+t*s/sqrt(n)
```

######################## Step-13 ############################

13. Constructing a 99%, 95%, and 90% confidence interval estimate for the Pre and Post variable

```
m=mean(data$Pre)
s=sd(data$Pre)
n=length(data$Pre); dof=n-1
t=qt(0.95,dof);t=qt(0.95, dof); m-t*s/sqrt(n); m+t*s/sqrt(n)
t=qt(0.995,dof);t=qt(0.995, dof); m-t*s/sqrt(n); m+t*s/sqrt(n)
t=qt(0.90,dof);t=qt(0.90, dof); m-t*s/sqrt(n); m+t*s/sqrt(n)
m=mean(data$Post)
s=sd(data$Post)
n=length(data$Post); df=n-1
t=qt(0.95,dof);t=qt(0.95, dof); m-t*s/sqrt(n); m+t*s/sqrt(n)
t=qt(0.995,dof);t=qt(0.995, dof); m-t*s/sqrt(n); m+t*s/sqrt(n)
t=qt(0.90,dof);t=qt(0.90, dof); m-t*s/sqrt(n); m+t*s/sqrt(n)
```

#################### Step 14 #####################3

14. (a) Taking a sample of 50 from Pre and Post Variables

```
Pre=sample(data$Pre, 50)
Pre
Post=sample(data$Post, 50)
Post
```

14. (b) Stating null and alternate Hypothesis

H0:Sample Mean = Population Mean
H1:Sample Mean!= Population Mean

For Pre

```
Pre=sample(data$Pre, 50)
n=length(Pre)
m=mean(data$Pre)
s=sd(data$Pre)/sqrt(n)
z=(mean(Pre)-mean(data$Pre))/s
```

14. (c) Calculating z values for Pre and Post variables

For Post

```
Post=sample(data$Post, 50)
n1=length(Post)
m1=mean(data$Post)
s1=sd(data$Post)/sqrt(n1)
z1=(mean(Post)-mean(data$Post))/s1
```

14. (c) Calculating z values for Pre and Post variables

z1

################ Step-15 ############################

Using the p-value method, determining whether the sample
mean for the pre and post variables differs significantly
from the population mean at the 10% significance level

```
pnorm(z)
1-pnorm(z)*2
```

For pre the this value is less than the level of significance,Hence, we can reject the null hypothesis at 0.10

```
pnorm(z1)
1-pnorm(z1)*2
```

For post the this value is more than the level of significance,
Hence, we cannot reject the null hypothesis at 0.10

################ Step-16 #####################

Calculating the critical Z value for the 10% level of significance
and the decision rule using the critical value approach

```
z=qnorm(0.975)
```

z

-Decision Rule for Pre: The resulting z value is the Z score beyond which we would reject
-The null hypothesis at the 10% significance level
-Decision Rule for Post: The resulting z value is the Z score which is less than the critical value
-so, we fail to reject the null hypothesis at the 10% significance level

######################## Step-17 ##########################

17.Computing the T-statistics value for the pre and post variables

```
u=mean(data$Pre)
x=mean(Pre)
s=sd(Pre)
n=length(Pre)
dof=n-1
SE=s/sqrt(n)
t=(x-u)/SE;
t

u1=mean(data$Post)
x1=mean(Post)
s1=sd(Post)
n1=length(Post)
dof1=n1-1
SE1=s1/sqrt(n1)
t1=(x1-u1)/SE1;
t1
```

############################## Step-18 #################################3

18.Calculating the p-value and the decision using the p-value approach for pre and post variables at a 10% level of significance

```
p=(1-pt(t,dof))*2
p= 1.449148
```

The p value here is much lower than the significance p level. Hence, we reject the null hypothesis

```
p1=(1-pt(t1, dof1))*2
p1= 1.198156
```

The p value here is higher than the significance p1 level. Hence, we cannot reject the null hypothesis

19. Calculating the critical T value for the level of significance of 10% and the decision rule using the critical value approach

```
tv=qt(0.95, dof)
tv


tv1=qt(0.95, dof1)
tv
```