**A F&B manager wants to determine whether there is any significant difference in the diameter of the cutlet between two units. A randomly selected sample of cutlets was collected from both units and measured? Analyze the data and draw inferences at 5% significance level. Please state the assumptions and tests that you carried out to check validity of the assumptions.**







> library(readxl)

> dimeter <- read\_excel("E:/data science r studio/Assignment code 1/hypotisis/dimeter.xlsx")

> View(dimeter)

> ###################Normaity test#################

> library(nortest)

> shapiro.test(dimeter$`Unit A`)

Shapiro-Wilk normality test

data: dimeter$`Unit A`

W = 0.96495, p-value = 0.32

> shapiro.test(dimeter$`Unit B`)

Shapiro-Wilk normality test

data: dimeter$`Unit B`

W = 0.97273, p-value = 0.5225

> #############################check external condition##############

> #################external condition not same##################

> #################check variance ###############

> library(moments)

> var(dimeter$`Unit A`)

[1] 0.08317945

> var(dimeter$`Unit B`)

[1] 0.117924

> t.test(dimeter$`Unit A`,dimeter$`Unit B`,alternative = "two.side",conf.level = 0.95,correct=TRUE)

Welch Two Sample t-test

data: dimeter$`Unit A` and dimeter$`Unit B`

t = 0.72287, df = 66.029, p-value = 0.4723

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

95 percent confidence interval:

-0.09654633 0.20613490

sample estimates:

mean of x mean of y

7.019091 6.964297

> t.test(dimeter$`Unit A`,dimeter$`Unit B`,alternative = "greater",var.equal = T)

Two Sample t-test

data: dimeter$`Unit A` and dimeter$`Unit B`

t = 0.72287, df = 68, p-value = 0.2361

alternative hypothesis: true difference in means is greater than 0

95 percent confidence interval:

-0.0716096 Inf

sample estimates:

mean of x mean of y

7.019091 6.964297

**Ho=p>0.005 hence,null fly, no action**

**2 A hospital wants to determine whether there is any difference in the average Turn Around Time (TAT) of reports of the laboratories on their preferred list. They collected a random sample and recorded TAT for reports of 4 laboratories. TAT is defined as sample collected to report dispatch. Analyze the data and determine whether there is any difference in average TAT among the different laboratories at 5% significance level.**

> library(readxl)

> loboratry <- read\_excel("E:/data science r studio/Assignment code 1/hypotisis/ques 2/loboratry.xlsx")

> View(loboratry)

> ####################check Normaity test################

> library(nortest)

> shapiro.test(loboratry$`Laboratory 1`)

Shapiro-Wilk normality test

data: loboratry$`Laboratory 1`

W = 0.99018, p-value = 0.5508

> shapiro.test(loboratry$`Laboratory 2`)

Shapiro-Wilk normality test

data: loboratry$`Laboratory 2`

W = 0.99363, p-value = 0.8637

> shapiro.test(loboratry$`Laboratory 3`)

Shapiro-Wilk normality test

data: loboratry$`Laboratory 3`

W = 0.98863, p-value = 0.4205

> shapiro.test(loboratry$`Laboratory 4`)

Shapiro-Wilk normality test

data: loboratry$`Laboratory 4`

W = 0.99138, p-value = 0.6619

> #Ho=data is normal

> #H1= data is not norma

> ########################check variance##################

> library(moments)

> ##Ho=variance of LAB1is equal LAB2

> ##H1=variance of lab1 is not equal

> var.test(loboratry$`Laboratory 1`,loboratry$`Laboratory 2`)

F test to compare two variances

data: loboratry$`Laboratory 1` and loboratry$`Laboratory 2`

F = 0.77573, num df = 119, denom df = 119, p-value = 0.1675

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.5406345 1.1130690

sample estimates:

ratio of variances

0.7757342

> ##Ho=variance of Lab3 is eual lab 4

> ##H1=variance of Lab3 is not equal to lab4

> var.test(loboratry$`Laboratory 3`,loboratry$`Laboratory 4` )

F test to compare two variances

data: loboratry$`Laboratory 3` and loboratry$`Laboratory 4`

F = 1.2021, num df = 119, denom df = 119, p-value = 0.3168

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.8377527 1.7247817

sample estimates:

ratio of variances

1.202057

> var.test(loboratry$`Laboratory 1`,loboratry$`Laboratory 4`)

F test to compare two variances

data: loboratry$`Laboratory 1` and loboratry$`Laboratory 4`

F = 0.76263, num df = 119, denom df = 119, p-value = 0.1408

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.5315011 1.0942648

sample estimates:

ratio of variances

0.762629

> # variance same because var>0.005

> #then go to one way ANOVA test

> ?anova

> library(car)

Loading required package: carData

> stacked\_data<-stack(loboratry)

> View(stacked\_data)

> attach(stacked\_data)

> Anova\_results <- aov(values~ind ,data=stacked\_data)

> summary(Anova\_results)

Df Sum Sq Mean Sq F value Pr(>F)

ind 3 79979 26660 118.7 <2e-16 \*\*\*

Residuals 476 106905 225

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

>

>

>

**3 Sales of products in four different regions is tabulated for males and females. Find if male-female buyer rations are similar across regions.**

###############################question 3 hypoti################

#######################campare the value#####################

library(readxl)

buyratio <- read\_excel("ques3/buyer.xlsx")

View(buyer)

##############################check camparition############3

library(nortest)

library(readxl)

shapiro.test(buyer$East)

shapiro.test(buy$West)

ad.test(buyer$East)

ad.test(buyer)

shapiro.test(buy$North)

shapiro.test(buy$South)

ad.test(buyer$East)

####################chi-square test#####################

library(readxl)

attach(buyer)

chisq.test(data$East,data$West,data$North,data$south,correct = FALSE)

> shapiro.test(buyer$East)

Error in shapiro.test(buyer$East) :

sample size must be between 3 and 5000

> shapiro.test(buy$West)

Error: object 'buy' not found

> ad.test(buyer$East)

Error in ad.test(buyer$East) : sample size must be greater than 7

> ad.test(buyer)

Error: Length of logical index vector for `[` must equal number of columns (or 1):

\* `.data` has 5 columns

hence not run this question

**4 TeleCall uses 4 centers around the globe to process customer order forms. They audit a certain % of the customer order forms. Any error in order form renders it defective and has to be reworked before processing. The manager wants to check whether the defective % varies by centre. Please analyze the data at *5%* significance level and help the manager draw appropriate inferences**

* > chisq.test(data$Phillippines,data$Indonesia,data$Malta,data$India,correct=FALSE)
* Pearson's Chi-squared test
* data: data$Phillippines and data$Indonesia
* X-squared = 0.55216, df = 1, p-value = 0.4574
* P-value 0.4574>0.05.Hence we fail to reject Null.

**Hence training requires, if proportions of defective is equal in all regions**

**5 Fantaloons Sales managers commented that *%* of males versus females walking in to the store differ based on day of the week. Analyze the data and determine whether there is evidence at *5 %* significance level to support this hypothesis.**

Create hypothesis

Ho= Proportions of Male and Female are same

Ha= Proportions of Male and Female are not same

> prop.test(x=c(58,152),n=c(480,740),conf.level = 0.95,correct = FALSE,alternative = "two.sided")

* 2-sample test for equality of proportions without continuity correction
* data: c(58, 152) out of c(480, 740)
* X-squared = 14.613, df = 1, p-value = 0.000132
* alternative hypothesis: two.sided
* 95 percent confidence interval:
* -0.1257722 -0.0433719
* sample estimates:
* prop 1 prop 2
* 0.1208333 0.2054054
* P-value is 0.000132< 0.05 and hence we fail to reject Null.Hence proportions of Male and Female are not same
* Now we will try to find out whose proportion is higher.

We create another hypothesis

Ho= Proportions of Male is less than or equal to Female

Ha= Proportions of Male is greater than Female

> prop.test(x=c(58,152),n=c(480,740),conf.level = 0.95,correct = FALSE,alternative = "less")

2-sample test for equality of proportions without continuity correction

data: c(58, 152) out of c(480, 740)

X-squared = 14.613, df = 1, p-value = 6.599e-05

alternative hypothesis: less

95 percent confidence interval:

-1.0000000 -0.0499958

sample estimates:

prop 1 prop 2

0.1208333 0.2054054

* P-value 6.599e-05<0.05 and hence we reject null.**Hence proportion of Male is greater than Female.**