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.

Assumption

**Step 1 Check for normality of diameter of cutlet from UnitA and Unit B**

*print(stats.shapiro(cutlet.UnitA)) # Shapiro Test*

*(0.9649458527565002, 0.3199819028377533) #P>0.05 – P high null fly*

*shapiro.test(UnitA) #W = 0.96495, p-value = 0.32*

*print(stats.shapiro(cutlet.UnitB))*

*(0.9727300405502319, 0.5224985480308533)#P>0.05 – P high null fly*

*shapiro.test(UnitB) #W = 0.97273, p-value = 0.5225*

**Step 2 : Check for variances of diameter of cutlet from UnitA and Unit B**

*scipy.stats.levene(cutlet.UnitA, cutlet.UnitB)*

*Out[15]: LeveneResult(statistic=0.665089763863238, pvalue=0.4176162212502553) #– P high null fly*

*var.test(UnitA,UnitB)*

*# p-value = 0.3136 > 0.05 so p high null fly => Equal variances*

**Step 3 : Ho: Unit A and Unit B have same diameter of the shape**

scipy.stats.ttest\_ind(cutlet.UnitA, cutlet.UnitB)

Out[16]: Ttest\_indResult(statistic=0.7228688704678061, pvalue=0.47223947245995)- *#–* ***P high null fly***

***t.test(UnitA, UnitB, alternative = "two.sided", conf.level = 0.95)***

***Welch Two Sample t-test***

*data: UnitA and UnitB*

*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*

Hence, **Accept Ho => Unit A and Unit B have same diameter of the shape**

File : **Cutlets.csv**

**Hypothesis Testing Exercise**

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.

**Step 1: Check for normality of Average TAT of different laboratories**

print(stats.shapiro(labTAT.LabA)) # Shapiro Test

(0.9901824593544006, 0.5506953597068787)

print(stats.shapiro(labTAT.LabB))

(0.9936322569847107, 0.8637524843215942)

print(stats.shapiro(labTAT.LabC))

(0.9886345267295837, 0.4205053448677063)

print(stats.shapiro(labTAT.LabD))

(0.9913753271102905, 0.6618951559066772)

**Step 2: Check for variances among 4 labs**

scipy.stats.levene(labTAT.LabA, labTAT.LabB)

Out[27]: LeveneResult(statistic=3.5495027780905763, pvalue=0.06078228171776711)

scipy.stats.levene(labTAT.LabB, labTAT.LabC)

Out[28]: LeveneResult(statistic=0.9441465124387124, pvalue=0.33220021420602397)

scipy.stats.levene(labTAT.LabC, labTAT.LabD)

Out[29]: LeveneResult(statistic=2.037958464521512, pvalue=0.15472618294425391)

scipy.stats.levene(labTAT.LabB, labTAT.LabD)

Out[30]: LeveneResult(statistic=0.2889202799636133, pvalue=0.5914154837597723)

scipy.stats.levene(labTAT.LabC, labTAT.LabA)

Out[31]: LeveneResult(statistic=7.547664894290509, pvalue=0.006468575869839467)

scipy.stats.levene(labTAT.LabD, labTAT.LabA)

Out[32]: LeveneResult(statistic=1.5000140718506723, pvalue=0.22188001348277267)

**Step 3: Ho: Average TAT among the different laboratories at 5% significance level are equal.**

mod = ols('LabA ~ LabB + LabC + LabD',data = labTAT).fit()

aov\_table = sm.stats.anova\_lm(mod, type=2)

aov\_table

Out[101]:

df sum\_sq mean\_sq F PR(>F)

LabB 1.0 332.030416 332.030416 1.940311 0.166299

LabC 1.0 203.853111 203.853111 1.191271 0.277335

LabD 1.0 265.614707 265.614707 1.552192 0.215323

Residual 116.0 19850.186366 171.122296 NaN NaN

**R code**

# Normality test

shapiro.test(`LabA`)

shapiro.test(`LabB`)

shapiro.test(`LabC`)

shapiro.test(`LabD`)

# Variance test

var.test(`LabA`,`LabB`)

var.test(`LabB`,`LabC`)

var.test(`LabC`,`LabA`)

var.test(`LabA`,`LabD`)

var.test(`LabC`,`LabD`)

var.test(`LabB`,`LabD`)

Anova\_results <- aov(LabA~LabB+LabC+LabD, data = LabTAT)

summary(Anova\_results)

# p-value = 0.166 > 0.05 Accept null hypothesis

# p-value = 0.277 > 0.05 Accept null hypothesis

# p-value = 0.215 > 0.05 Accept null hypothesis

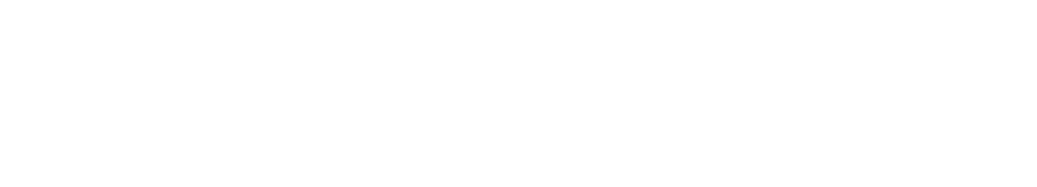
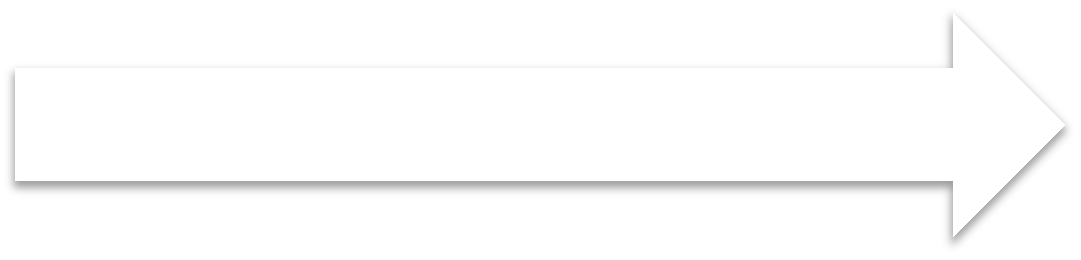
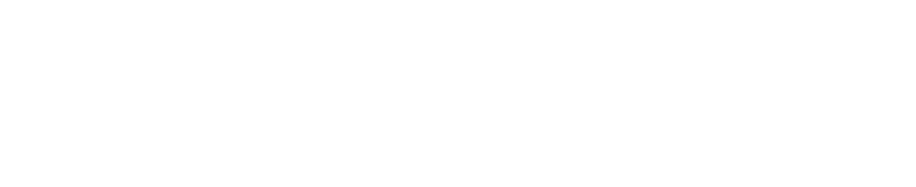
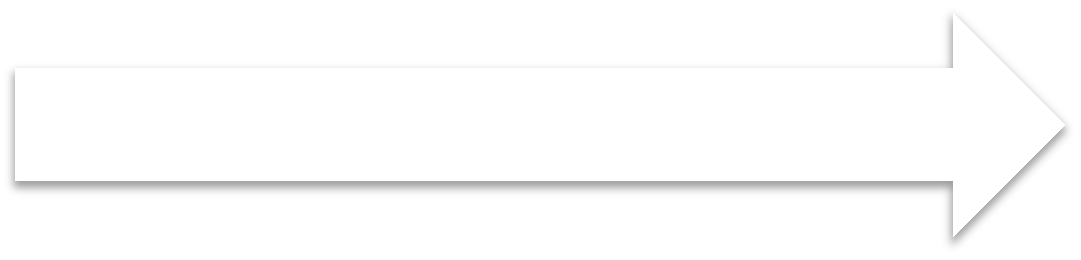
# 4 Labs TAT times are equal

**P value is high, hence accept Ho => average TAT among the different laboratories at 5% significance level are equal**

File: **LabTAT.csv**

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **East** | **West** | **North** | **South** |
| Males | 50 | 142 | 131 | 70 |
| Females | 550 | 351 | 480 | 350 |



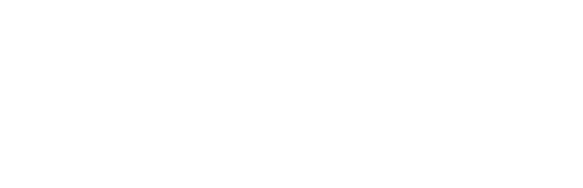
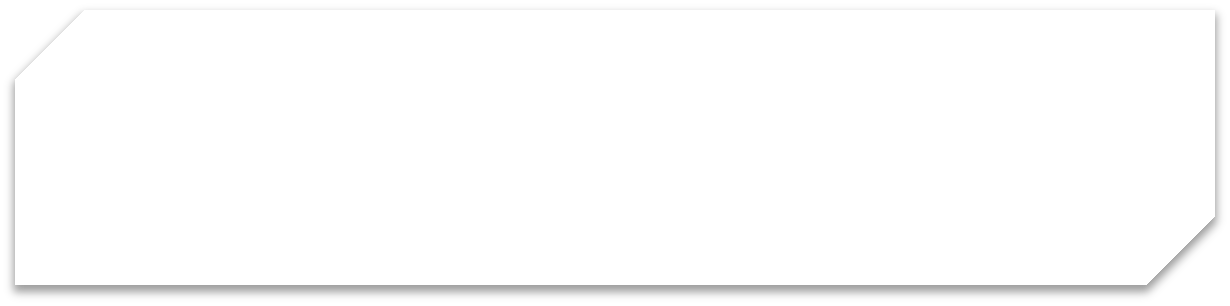
H0

* All proportions are equal

Ha

* Not all Proportions are equal

1. Check p-value
2. If p-Value < alpha, we reject Null Hypothesis



Buyer Ratio.csv

**Region East**

stats, pval = proportions\_ztest(count\_East, nobs, alternative='larger')

print(pval) # Pvalue 0.999

0.9999999999999818

stats, pval = proportions\_ztest(count\_East, nobs, alternative='larger')

print(pval) # Pvalue 0.999

0.9999999999999818

Buyer Ratio for region East using two sided test larger has p value – 0.999 =>

**(p1>p2) proportion for east is larger than the observations**

**Region West**

stats, pval = proportions\_ztest(count\_West, nobs, alternative='two-sided')

print(pval) # Pvalue- 0.000

1.8039532503031428e-11

stats, pval = proportions\_ztest(count\_West, nobs, alternative='larger')

print(pval) # Pvalue 0.000

9.019766251515714e-12

stats, pval = proportions\_ztest(count\_West, nobs, alternative='smaller')

print(pval) # Pvalue 0.999

0.9999999999909802

Buyer Ratio for region West using two sided test smaller has p value – 0.999 =>

**(p1<p2) proportion for west is smaller than the observations**

**Region North**

stats, pval = proportions\_ztest(count\_North, nobs, alternative='two-sided')

print(pval) # Pvalue < 0.05

0.026731043242145935

stats, pval = proportions\_ztest(count\_North, nobs, alternative='larger')

print(pval) # Pvalue 0.0.05

0.013365521621072967

stats, pval = proportions\_ztest(count\_North, nobs, alternative='smaller')

print(pval) # Pvalue 0.999

0.986634478378927

Buyer Ratio for region North using two sided test smaller has p value – 0.999 =>

**(p1<p2) proportion for North is smaller than the observations**

**Region South**

stats, pval = proportions\_ztest(count\_South, nobs, alternative='two-sided')

print(pval) # Pvalue- 0.279

0.27929720478250486

stats, pval = proportions\_ztest(count\_South, nobs, alternative='larger')

print(pval) # Pvalue 0.8603

0.8603513976087476

stats, pval = proportions\_ztest(count\_South, nobs, alternative='smaller')

print(pval) # Pvalue 0.139

0.13964860239125243

Buyer Ratio for region south using two sided test smaller has p value – 0.86 =>

**(p1<p2) proportion for North is smaller than the observations**

**R code**

M\_F\_Prop\_Test<-read.csv("D:/Data science/Module 5-HT/Hypothesis\_Testing\_Assignment/Buyer Ratio.csv")

View(M\_F\_Prop\_Test)

attach(M\_F\_Prop\_Test)

#East

prop.test(x=c(50,550),n=c(393,1731), conf.level = 0.95, alternative = "two.sided")

# two.sided -> means checking for equal proportions of Male and Female under purchased

# p-value = 5.869e-14 < 0.05 accept alternate hypothesis i.e.

# Unequal proportions

prop.test(x=c(50,550),n=c(393,1731), conf.level = 0.95, alternative = "greater")

# Ha -> Proportions of East > Proportions of observations

#West

prop.test(x=c(142,351),n=c(393,1731), conf.level = 0.95, alternative = "two.sided")

# two.sided -> means checking for equal proportions of Male and Female under purchased

# p-value = 2.835e-11 < 0.05 accept alternate hypothesis i.e.

# Unequal proportions

prop.test(x=c(142,351),n=c(393,1731), conf.level = 0.95, alternative = "greater")

#p-value = 1.418e-11 < 0.05 check for lesser

prop.test(x=c(142,351),n=c(393,1731), conf.level = 0.95, alternative = "less")

# Ha -> Proportions of West < Proportions of observations

#North

prop.test(x=c(131,480),n=c(393,1731), conf.level = 0.95, alternative = "two.sided")

# two.sided -> means checking for equal proportions of Male and Female under purchased

# p-value = p-value = 0.03126 < 0.05 accept alternate hypothesis i.e.

# Unequal proportions

prop.test(x=c(131,480),n=c(393,1731), conf.level = 0.95, alternative = "greater")

#p-value = 0.01563 <0.05 - check for less

prop.test(x=c(131,480),n=c(393,1731), conf.level = 0.95, alternative = "less")

#p-value = 0.9844 > 0.05 - Accept Ha

#i.e Proportion of North is less than proportion of observations

prop.test(x=c(70,350),n=c(393,1731), conf.level = 0.95, alternative = "two.sided")

# two.sided -> means checking for equal proportions of Male and Female under purchased

# p-value = p-value = 0.3117 > 0.05 accept null hypothesis i.e.

# Unequal proportions

prop.test(x=c(70,350),n=c(393,1731), conf.level = 0.95, alternative = "greater")

#p-value = 0.8442

##i.e Proportion of South is less than proportion of observations

prop.test(x=c(70,350),n=c(393,1731), conf.level = 0.95, alternative = "less")

#p-value = 0.1558

**Hence all the proportions are not equal**

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 must be reworked before processing. The manager wants to check whether the defective % varies by center. Please analyze the data at *5%* significance level and help the manager draw appropriate inferences

File: **Customer OrderForm.csv**

Ho: Defect percentage does not vary by center

Out[201]:

Country India Indonesia Malta Phillippines

Defective

Defective 20 33 31 29

Error Free 280 267 269 271

Chisquares\_results=scipy.stats.chi2\_contingency(count)

Chi\_square=[['', 'Test Statistic', 'p-value'],['Sample Data', Chisquares\_results[0], Chisquares\_results[1]]]

Chi\_square

Out[204]:

[['', 'Test Statistic', 'p-value'],

['Sample Data', 3.858960685820355, 0.2771020991233135]]

p-value is greater than 0.05 . Hence, Reject Ho

**Defect percentage varies in all the 4 centers**

Fantaloons Sales managers commented that *%* of males versus females walking into 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.

File: **Fantaloons.csv**

**Ho – Percentage of males vs Females walking into store differ based on weekday or weekend**

count\_Fantaloons = pd.crosstab(Fantaloons["Weekdays"], Fantaloons["Weekend"])

count\_Fantaloons

Chisquares\_results\_Fantaloons = scipy.stats.chi2\_contingency(count\_Fantaloons)

Chi\_square\_Fantaloons = [['', 'Test Statistic', 'p-value'],['Sample Data', Chisquares\_results\_Fantaloons[0], Chisquares\_results\_Fantaloons[1]]]

Chi\_square\_Fantaloons

Out[220]:

[['', 'Test Statistic', 'p-value'],

['Sample Data', 0.005274808283592733, 0.9421022439386241]]

p-value greater than 0.05, hence accept Ho

**R code**

Fantaloons <- read.csv("D:/Data science/Module 5-HT/Hypothesis\_Testing\_Assignment/Faltoons.csv")

attach(Fantaloons)

table(Weekdays,Weekend)

chisq.test(table(Weekdays,Weekend))

X-squared = 2.2781e-30, df = 1, p-value = 1

**Percentage of males vs females walking into store differ based on weekday or weekend**