**Hypothesis Testing Exercises (Module - 5)**

1.) 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.

# File: Cutlets.csv

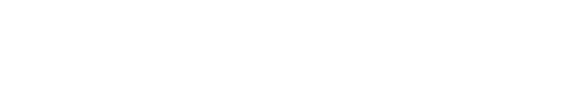
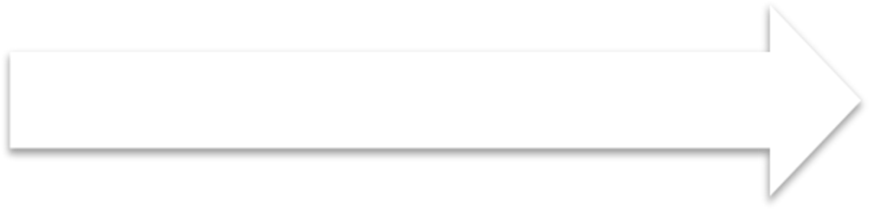
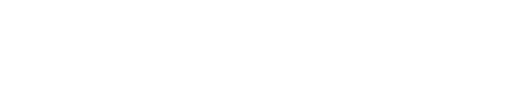
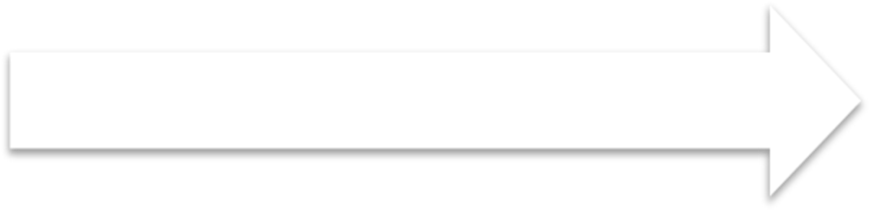
* Ans: - Here ,
  + Ha: diameter of the both cutlets is equal
  + Ho: diameter of the both cutlets is unequal
  + In cutlets data we have to columns and Y is continuous and X factor is discrete. So I am doing hypothesis testing for continuous variable. Both columns are in normal distribution, now after doing variance test with different external conditions we find both have equal variance. So finally now we
* are doing two sample T test for equal variance test (mean test ) at 5% significance level ( 95% confidence level ) we found p=0.4723 (p high null fly ,mean fail to reject Ho:) then go with Ho :
  + we can say that average of unit\_a is equal to average unit\_b with 95% confidence level.

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. File: **LabTAT.csv**

* Ans - Inputs are 4 lab reports. So **Input** is **Discrete** in **more than 2 categories**. **Output**is **continuous**as we are trying to see the difference in average TAT. we proceed with **ANOVA one-way test. Here all columns are in normal distribution and the variance of columns are equal at 95% signification level .** As there are more than 2 discrete variables and output variable TAT is a continuous variable. Hence we will go with **Anova one way test**. we can see that P -value is < 0.05. P Low so Ho go. Hence there is difference in the average TAT for all the labs.

3.) 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



H

o

•

All proportions are

H

a

•

Not all Proportions

are

1

.

Check

p

-

value

2

.

If

p

-

Value

<

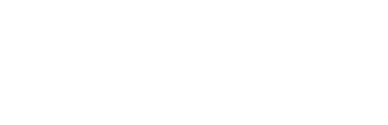
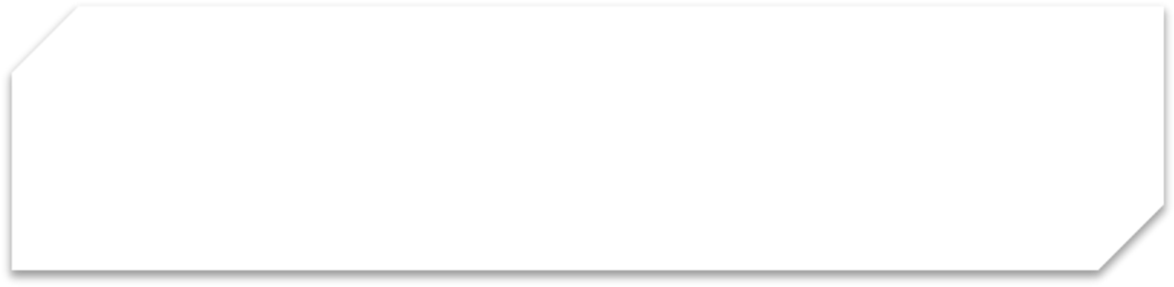
alpha,

we

reject Null

Hypothesis

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Males | 50 | 142 | 131 | 70 |
| Females | 550 | 351 | 480 | 350 |



Buyer Ratio.csv

* Ans - **Inputs**are **4 discrete variables** (east, west, north, south ). **Output**is also **discrete**. We are trying to find out if proportions of male and female are similar or not across the regions .We proceed with **chi-square test.** P-value>0.05.Hence we fail to reject Null. **Hence proportion of male and female across regions is same.**

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 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

Ans - **Inputs**are **4 discrete variables** (east, west, north, south ). **Output**is also **discrete**. We are trying to find out if proportions of male and female are similar or not across the regions .We proceed with **chi-square test.** **We can see that P -value is < 0.05. P high so Ho fly. Hence there is no significance difference in the average defective % varies by centre at 5% significance level.**

5.) 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

Ans - **Inputs**are **2 discrete variables**.**Output**is **Discrete**as we are trying to find out if proportions of male and female walking in to the store is same or not. We proceed with **2-proportion test.** P-value is less than 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. P-value <0.05 and hence we reject null.**Hence proportion of Male is greater than Female.**

**Hints:**

1. Business Problem
   1. Objective
   2. Constraints (if any)
2. Data Pre-processing
   1. Data cleaning, Feature Engineering, EDA etc.
3. Model Building
   1. Partition the dataset
   2. Model(s) - Reasons to choose any algorithm
   3. Model(s) Improvement steps
   4. Model Evaluation
   5. Python and R codes
4. Deployment
   1. Deploy solutions using R shiny and Python Flask.
5. Result Share the benefits/impact of the solution - how or in what way the business (client) gets benefit from the solution provided.

**Note:**

1. For each assignment the solution should be submitted in the format 2. For Hypothesis Testing Assignments, explanation of the solutions along with Business Objectives & Business Constraints should be documented in black and white along with the codes.

1. All the codes (executable programs) are running without errors
2. From Hypothesis module assignment onwards, along with R & Python code, Documentation

must be submitted in the same order as mentioned above.

* 1. For Hypothesis Testing Assignments, explanation of the solutions Business Objectives & Business Constraints should be documented in black and white along with the codes (R & Python).

* 1. All the test should be explained well in documentation (Normality test, Variance test etc.)