

Hypothesis Testing Exercises (Module - 5)

Name: Kavali Rakesh

Batch Id: DSWDMCON 18012022

- 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:

Business Objective: To determine the difference of diameter in two cutlets.

Defining Null Hypothesis H_0 and Alternative Hypothesis H_a

Case1: a) Mean of cutlet A = mean of cutlet B

b) Mean of cutlet A \neq mean of cutlet B

Case 2: a) Mean of cutlet A > mean of cutlet B

b) Mean of cutlet A < mean of cutlet B

Normalization Test: both the variables are normalized and the same is tested using Python.

T-Test: Applying T-test to the given variables and selecting the condition.

Conclusion: After analysis the data I suggest Cutlet B gives the good dimensions.

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:

Business Objective: To determine the difference in average turn around time of the laboratories.

Defining Null Hypothesis H_0 and Alternative Hypothesis H_a

Case1: a) Mean of lab A = mean of lab B = mean of lab C = mean of lab D

b) Mean of lab A \neq mean of lab B = mean of lab C = mean of lab D

Case 2: a) Mean of lab A > mean of lab B > mean of lab C > mean of lab D

b) Mean of lab A < mean of lab B < mean of lab C < mean of lab D

Normalization Test: all the variables are normalized and the same is tested using Python.


Variance Test : applying variance test for all the variables of the data set to know whether they are equal or not to apply Anova test

Conclusion: After applying the Anova Test it concludes that accept the null hypothesis means the mean time taken by all the laboratories are equal.


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

East West North South

Males	50	142	131	70
Females	550	351	480	350




H_0




- All proportions are

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

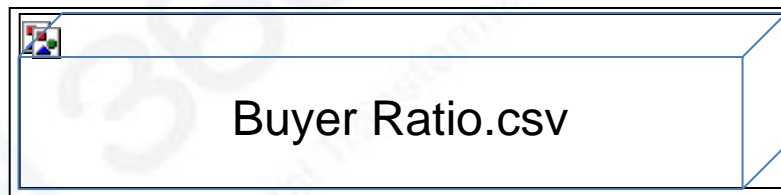


H_a

are



- Not all Proportions



Ans:

Ans:

Business Objective: To determine the buyer ratio of males and females in different regions

Defining Null Hypothesis H_0 and Alternative Hypothesis H_a

Case1: a) Avg. Buyer ratio of males = Avg. buyer ratio of females

b) Avg. Buyer ratio of males \neq Avg. buyer ratio of females

Case 2: a) Avg. Buyer ratio of males > Avg. buyer ratio of females

b) Avg. Buyer ratio of males < Avg. buyer ratio of females

Normalization Test: all the variables are normalized and the same is tested using Python.

Conclusion: It is observed that the average buyer ration of the females are more when compared to males.

- 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:

Business Objective: To determine the difference in average defective percentage of all the calling center's.

Defining Null Hypothesis H_0 and Alternative Hypothesis H_a

Case1: a) Avg. Defective % of Philippines = Avg. Defective % of Indonesia = Avg. Defective % Maltia = Avg. Defective % India

b) Avg. Defective % of Philippines \neq Avg. Defective % of Indonesia = Avg. Defective % Maltia = Avg. Defective % India

Case 2: a) Avg. Defective % of Philippines > Avg. Defective % of Indonesia > Avg. Defective % Maltia > Avg. Defective % India

b) Avg. Defective % of Philippines < Avg. Defective % of Indonesia < Avg. Defective % Maltia < Avg. Defective % India

Normalization Test: all the variables are normalized and the same is tested using Python.

Chi Square Test : applying Chi Square test for all the variables of the data set to know the variation of average Defective%

Conclusion: After applying Chi Square it is concluded that the India is leading country which is maintaining less defective percentage after that Philippines then Indonesia and Maltia respectively.

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:

Business Objective: To determine the difference % of males and females during week days and weekend

Defining Null Hypothesis H_0 and Alternative Hypothesis H_a

Case1: a) Avg. % males and females in week days = Avg. % males and females in week ends

b) Avg. % males and females in week days \neq Avg. % males and females in week ends

Case 2: a) Avg. % males and females in week days $>$ Avg. % males and females in week ends

b) Avg. % males and females in week days $<$ Avg. % males and females in week ends

Normalization Test: all the variables are normalized and the same is tested using Python.

Chi Square Test: applying Chi Square test for all the variables of the data set to know the variation of average % males and females.

Conclusion: After applying Chi Square it is concluded that the Avg. % of males and females are equal during week days and the weekends also.

Hints:

1. Business Problem

1.1. Objective

1.2. Constraints (if any)

2. Data Pre-processing

2.1 Data cleaning, Feature Engineering, EDA etc.

3. Model Building

3.1 Partition the dataset

3.2 Model(s) - Reasons to choose any algorithm

3.3 Model(s) Improvement steps

3.4 Model Evaluation

3.5 Python and R codes

4. Deployment

4.1 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.
4. All the codes (executable programs) are running without errors
5. From Hypothesis module assignment onwards, along with R & Python code, Documentation must be submitted in the same order as mentioned above.
 - 5.1. 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).
 - 5.2. All the test should be explained well in documentation (Normality test, Variance test etc.)