

Hypothesis Testing Exercise

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**

Hypothesis Testing Exercise

Assumption:

H_0 : Diameter are equal

H_1 : Diameter are not equal

Since 2 data are continuous, use 2 sample t-test as well as variation test for collinearity between variables.

1. Normality test (Shapiro test)

Unit A: p-value = 0.32 > 0.05 => It follows normal distribution

Unit B: p-value = 0.5225 > 0.05 => It follows normal distribution

2. Variance test

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

denom df = 34 (high degrees of freedom follow normal distribution)

3. Welch 2 Sample t-Test

null Hypothesis -> Equal means

p-value = 0.4723 > 0.05 accept null Hypothesis

4. 2 Sample t-Test

p-value = 0.2361 > 0.05 accept null Hypothesis

Conclusion: Diameter are equal

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.

File: **LabTAT.csv**

Assumption:

H_0 : Average Turn Around Time (TAT) of reports are equal

H_1 : Average Turn Around Time (TAT) of reports are not equal

Since 4 data are continuous, use ANOVA test.

1. Normality test (Shapiro test)

Lab 1: p-value = 0.5508 > 0.05 => It follows normal distribution

Lab 2: p-value = 0.8637 > 0.05 => It follows normal distribution

Lab 3: p-value = 0.4205 > 0.05 => It follows normal distribution

Lab 4: p-value = 0.6619 > 0.05 => It follows normal distribution

2. Variance test

Between Lab 1 and Lab 2: p-value = 0.1675 > 0.05 so p high null fly => Equal variances

Between Lab 2 and Lab 3: p-value = 0.2742 > 0.05 so p high null fly => Equal variances

Between Lab 3 and Lab 4: p-value = 0.3168 > 0.05 so p high null fly => Equal variances

Between Lab 4 and Lab 1: p-value = 0.1408 > 0.05 so p high null fly => Equal variances

df = 119 (high degrees of freedom follow normal distribution)

3. ANOVA test

p-value = $2e-16$ < 0.05 reject null Hypothesis, accept Alternative Hypothesis

Conclusion: Average Turn Around Time (TAT) of reports are not equal between 4 labs

Hypothesis Testing Exercise

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 equal

H_a

• Not all Proportions are equal

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

Buyer Ratio.csv

Hypothesis Testing Exercise

Assumption:

H_0 : Male-female buyer rations are similar across regions

H_1 : Male-female buyer rations are not similar across regions

Since 2 data are categorical, use chi-squared test.

1. Chi-squared test

p-value = 0.631524 > 0.05 accept null Hypothesis, reject Alternative Hypothesis

Conclusion: Male-female buyer rations are similar across regions

Hypothesis Testing Exercise

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**

Hypothesis Testing Exercise

Assumption:

H_0 : Defective percentages are similar across all centers

H_1 : Defective percentages varies by center

Since 2 data are categorical, use chi-squared test.

1. Chi-squared test

p-value = $1.146e-07 < 0.05$ reject null Hypothesis, accept Alternative Hypothesis

Conclusion: Defective percentages varies by center

Hypothesis Testing Exercise

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**

Hypothesis Testing Exercise

Assumption:

H_0 : Percentage of males versus females walking into the store is equal on any day of the week

H_1 : Percentage of males versus females walking into the store differ based on day of the week

Since 2 data are categorical (binary), use proportion test.

1. Chi-squared test

p-value = $8.543e-05 < 0.05$ reject null Hypothesis, accept Alternative Hypothesis

Conclusion: Percentage of males versus females walking into the store differ based on day of the week