**CHI-SQUARE TEST**

1. **State the Hypotheses:**

The **Chi-square test of independence** is used to find out whether two categorical variables are related or independent. Like all hypothesis tests, it has two possible hypotheses:

* Null Hypotheses (H0): The variables are independent, there is no relationship between the two categorical variables. Knowing the value of one variable does not help to predict the value of the other variable.
* Alternative Hypotheses (H1): The variables are dependent, there is a relationship between the two categorical variables. Knowing the value of one variable helps to predict the value of the other variable.

1. **Compute the Chi-Square Statistics:**

Formula for chi-square statistics.

χ2=∑​(Oi​−Ei​)2​/Ei

Where:

* χ2 is the Chi-square statistic.
* Oi​ is the observed frequency (the actual count you have from the data).
* Ei​ is the expected frequency (the count you would expect if there were no relationship between the variables).
* The sum∑ is taken over all possible categories.

**Steps to Calculate Chi-square Statistic:**

1. **Find the Observed Frequencies (O):** These are the actual values you collected from your data.
2. **Calculate the Expected Frequencies (E):** These are what you expect if the two variables are independent. The formula for the expected frequency for each cell is:

**Ei=(Row Total× Column Total)/Grand  Total**

1. **Apply the Chi-square formula:** For each cell, subtract the expected frequency from the observed frequency, square the result, and then divide by the expected frequency.
2. **Sum the results** for all cells to get the Chi-square statistic.
3. **Determine the critical value:**

The **critical value** in a Chi-square test helps you determine whether to reject or fail to reject the null hypothesis.

**1. Determine the Significance Level (α)**: alpha=0.05

**2. Calculate Degrees of Freedom (df)**: df=(r−1)(c−1)

Number of Rows = 5 (satisfaction levels)

Number of Columns = 2 (device types)

Degrees of Freedom = (5 - 1) \* (2 - 1) = 4

**3 .Use a Chi-square Distribution Table**:

The table provides critical values for various degrees of freedom and significance levels.

1. **Make a Decision:**

**Compare the Chi-square Statistic to the Critical Value**:

* If χ2 (calculated statistic) > Critical Value:

**Reject the Null Hypothesis (H₀)**: There is a significant association between the variables.

* If χ2(calculated statistic) ≤ Critical Value:

**Fail to Reject the Null Hypothesis (H₀)**: There is no significant association between the variables.

1. **Python Code:**

import numpy as np

from scipy.stats import chi2

# observed frequency table

observed=np.array([[50,70],

[80,100],

[60,90],

[30,50],

[20,50]])

# Totals

row\_totals=observed.sum(axis=1)

col\_totals = observed.sum(axis=0)

grand\_total = observed.sum()

# Expected frequency table

expected = np.outer(row\_totals, col\_totals) / grand\_total

# Chi-Square statistic calculation

chi2\_stat = ((observed - expected) 2 / expected).sum()

chi2\_stat

# Critical value for alpha = 0.05 and df = 4

alpha=0.05

df=4

critical\_value=chi2.ppf(1-alpha,df)

critical\_value

#Decision

if chi2\_stat > critical\_value:

decision = "Reject the null hypothesis"

else:

decision = "Fail to reject the null hypothesis"

decision

**Output:**

The Chi-Square statistic is 5.64, while the critical value is 9.49. Since the Chi-Square statistic is lower than the critical value, we fail to **reject the null hypothesis.** This indicates that there isn’t enough evidence to show that the two variables are related.

**HYPOTHESIS TESTING**

**01.Hypothesis Statement:**

1. **Null Hypothesis (H₀)**:  
   The restaurant owners' observed weekly operating costs are consistent with the theoretical cost model. That is, the mean weekly operating cost is not higher than what the theoretical model predicts.
2. **Alternative Hypothesis (H₁)**:  
   The restaurant owners' observed weekly operating costs are higher than the theoretical cost model. That is, the mean weekly operating cost is greater than what the theoretical model predicts.

**02. Calculate the Test Statistics:**

### Theoretical Mean Weekly Cost:

According to the cost model W=1,000+X where X represents the number of units produced:

* Given that X=600 units, we can calculate the weekly cost as follows:

W=1,000+5×600=1,000+3,000=4,000W

Therefore, the theoretical mean weekly cost is **$4,000**.

**Sample Mean Weekly Cost**:

- The average weekly cost from the sample is Rs. 3,050.

**Theoretical Mean Cost:**

- For producing 600 units, the calculated theoretical mean weekly cost is $4,000 (as determined earlier).

**Standard Deviation of Weekly Cost:**

- The standard deviation of the number of units produced is 25 units.

- Since the cost per unit is $5, the standard deviation of the weekly cost can be calculated as:

σW​=5×25=125

- Therefore, the standard deviation of the weekly cost is $125.

**Standard Error of the Mean (SEM):**

- The SEM is calculated using the formula:

SEM=σW​/sqt(n)​=125/sqt(25)​=125/5​=25

- Thus, the standard error of the mean is $25.

**Test Statistic:**

To calculate the test statistic T, we use the formula:

T = xˉ−μ/SEM

Where:

* xˉ=3050 Rs. (the sample mean)
* μ=4000 USD (the theoretical mean)
* SEM=25 USD (the standard error of the mean)

Now, we can substitute the values into the formula:

T = 3050−4000/25

Calculating this gives:

T = −950/25=−38t

So, the test statistic T is **-38**.

**03. Determine the Critical Value:**

For a one-tailed test with a significance level of α=0.05, we find the critical value using the Z-distribution table.

The critical value for α=0.05 in a one-tailed test is **Z = 1.645**.

**04. Make a Decision:**

**Comparison of Test Statistic and Critical Value**

* **Test Statistic**: T = −38
* **Critical Value**: Z\_{0.05}=−1.645Z
* Since−38 is much lower than −1.645, we are well into the rejection area. This means the data strongly suggests that the weekly costs are significantly lower than expected.

**05. CONCLUSION**

Based on the analysis, the test statistic of T=−38 is significantly lower than the critical value of Z\_{0.05}=−1.645. This indicates that there is strong evidence to reject the null hypothesis. Therefore, we conclude that the weekly operating costs for the franchises are significantly lower than the expected cost of $4,000 based on the theoretical model.

**Python Code:**

import numpy as np

from scipy.stats import norm

# Given data

sample\_mean = 3050 # Rs.

theoretical\_mean = 4000 # Rs.

std\_dev\_units = 25 # units

cost\_per\_unit = 5 # Rs per unit

sample\_size = 25

# Step 1: Calculate the standard deviation of the weekly cost

cost\_std\_dev = cost\_per\_unit \* std\_dev\_units # Rs.

cost\_std\_dev

# Step 3: Calculate the test statistic (z-score)

test\_statistic = (sample\_mean - theoretical\_mean) / standard\_error

test\_statistic

# Step 4: Determine the critical value for a one-tailed test at alpha = 0.05

alpha = 0.05

critical\_value = norm.ppf(1 - alpha)

critical\_value

# Step 5: Make a decision based on the test statistic and critical value

if test\_statistic > critical\_value:

decision = "Reject the null hypothesis"

else:

decision = "Fail to reject the null hypothesis"

decision

# Output the results

print(f"Test Statistic: {test\_statistic:.2f}")

print(f"Critical Value: {critical\_value:.2f}")

print(f"Decision: {decision}")

This Python code calculates the test statistic and critical value, then determines whether to reject the null hypothesis based on the provided data and significance level.

**OUTPUT:**

Test Statistic: -38.00

Critical Value: 1.64

Decision: Fail to reject the null hypothesis