Unit 1
EXPERIMENT 3

CO2: Utilize a modeler to create, manage, and run data mining streams.

Experiment 1.3: To perform a Chi-Square Goodness of Fit test using OriginPro to determine if the observed frequencies of a categorical variable differ significantly from the expected frequencies.

Objective

To perform and understand the Chi-Square Test, analyze whether observed frequencies differ from expected frequencies.

Theory

Chi-Square Test Overview

The Chi-Square Test is a statistical method used to compare observed data with data expected to be obtained according to a specific hypothesis. There are two main types of Chi-Square Tests:

1. **Chi-Square Goodness of Fit Test (Parametric)**: Determines if sample data matches a population with a specific distribution.

2. **Chi-Square Test of Independence (Non-Parametric)**: Assesses whether two categorical variables are independent.

Formula

- Goodness of Fit Test:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

where O_i is the observed frequency and E_i is the expected frequency.

- Test of Independence:

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

where O_{ij} is the observed frequency for the ith row and jth column, and E_{ij} is the expected frequency for the ith row and jth column.

Sample Data

Sample Table for Goodness of Fit (Observed Frequencies)

Category	Observed Frequency
A	15
B	25
C	30
D	20
E	10

Sample Table for Test of Independence

	Category 1	Category 2	Total
Group A	15	10	25
Group B	20	30	50
Total	35	40	75

Procedure

Goodness of Fit Test (Parametric)
1. **Input Data**:
   o Open OriginPro 8.5.
   o Enter the observed frequencies into a worksheet column.
2. **Calculate Expected Frequencies**:
   o Assume a theoretical distribution (e.g., equal distribution across categories).
3. **Perform Chi-Square Test**:
   o Go to Statistics > Descriptive Statistics > Chi-Square Test for Variance.
   o Select the column containing observed frequencies.
   o Enter expected frequencies if not calculated by OriginPro.
4. **Analyze Results**:
   o Check the Chi-Square statistic and p-value.
   o Determine if the observed frequencies significantly differ from expected frequencies.

Test of Independence (Non-Parametric)
1. **Input Data**:
   o Enter the observed frequencies into a contingency table format in the worksheet.
2. **Perform Chi-Square Test**:
   o Go to Statistics > Descriptive Statistics > Chi-Square Test for Variance.
   o Select the table with observed frequencies.
   o OriginPro will calculate expected frequencies based on row and column totals.
3. **Analyze Results**:
   o Check the Chi-Square statistic and p-value.
   o Determine if the variables are independent.

Results
- **Goodness of Fit Test**:
   o Chi-Square statistic (e.g., $\chi^2 = 5.67$)
   o p-value (e.g., $p=0.34$p = 0.34p=0.34)
- **Test of Independence**:

- Chi-Square statistic (e.g., $\chi^2 = 8.49$)
- p-value (e.g., $p = 0.04$)

Learning Outcomes
1. **Understanding of Chi-Square Tests**:
    - Differentiate between Goodness of Fit and Test of Independence.
    - Interpret the Chi-Square statistic and p-value.
2. **Proficiency with OriginPro 8.5**:
    - Input data and perform Chi-Square Tests using the software.
    - Analyze and interpret results.
3. **Application of Statistical Concepts**:
    - Use the Chi-Square Test to analyze real-world data.
    - Assess the independence of categorical variables and the fit of observed data to theoretical distributions.