The **Chi-Square test** is a statistical test used to determine if there is a significant association between two categorical variables or to assess how well a set of observed data fits an expected distribution. It is commonly used in hypothesis testing for categorical data.

There are two main types of Chi-Square tests:

1. **Chi-Square Test of Independence**: Used to determine if there is a relationship between two categorical variables.
2. **Chi-Square Goodness-of-Fit Test**: Used to determine if observed data fits a particular theoretical distribution.

1. Chi-Square Test of Independence

This test helps you determine whether two categorical variables are independent or associated with each other in a population.

Example Scenario:

Suppose you're trying to determine if gender (male/female) is associated with preference for a product (like/dislike). This test would tell you whether the preference for the product is independent of gender or not.

Steps for Chi-Square Test of Independence:

- **Step 1: Set up Hypotheses**
  - **Null Hypothesis (H_0)**: The two variables are independent (no association).
  - **Alternative Hypothesis (H_1)**: The two variables are dependent (there is an association).
- **Step 2: Create a Contingency Table**
  - A contingency table is a table that shows the frequency distribution of the variables.
  - Example of a contingency table:

	Like	**Dislike**	**Total**
Male	20	30	50
Female	40	10	50
Total	60	40	100

- **Step 3: Calculate Expected Counts**
  - The expected count for each cell (combination of variables) is calculated assuming the null hypothesis is true (that the variables are independent).

    - The formula to calculate the expected value for a cell is: Expected Value=Row Total×Column TotalGrand Total\text{Expected Value} = \frac{\text{Row Total} \times \text{Column Total}}{\text{Grand Total}}Expected Value=Grand TotalRow Total×Column Total
- For example, the expected count for the "Male-Like" cell is: E=50×60100=30E = \frac{50 \times 60}{100} = 30E=10050×60=30 Similarly, you calculate expected values for all cells.
- **Step 4: Calculate the Chi-Square Statistic**
    - The Chi-Square statistic compares the observed and expected counts. The formula is: χ2=∑(O−E)2E\chi^2 = \sum \frac{(O - E)^2}{E}χ2=∑E(O−E)2 Where:
        - OOO = Observed value
        - EEE = Expected value
- This is done for each cell, and the sum gives the Chi-Square statistic.
- **Step 5: Compare the Statistic to the Critical Value or p-value**
    - The Chi-Square statistic is compared to a critical value from the Chi-Square distribution table or converted into a p-value.
    - The critical value depends on the degrees of freedom (df) and the chosen significance level (usually 0.05).
    - The degrees of freedom for this test is: df=(number of rows−1)×(number of columns−1)\text{df} = (\text{number of rows} - 1) \times (\text{number of columns} - 1)df=(number of rows−1)×(number of columns−1)
- **Step 6: Decision**
    - If the calculated Chi-Square statistic is greater than the critical value (or if the p-value is less than the significance level), you reject the null hypothesis, meaning the variables are associated.
    - If not, you fail to reject the null hypothesis, meaning the variables are independent.

2. Chi-Square Goodness-of-Fit Test

This test is used to see if an observed distribution fits a theoretical distribution.

Example Scenario:

Suppose you roll a die 60 times, and you expect each face (1 to 6) to appear 10 times. The goodness-of-fit test would check whether the observed frequencies of the dice rolls match the expected frequencies.

Steps for Chi-Square Goodness-of-Fit Test:

- **Step 1: Set up Hypotheses**
    - **Null Hypothesis (H_0)**: The observed frequencies match the expected frequencies.
    - **Alternative Hypothesis (H_1)**: The observed frequencies do not match the expected frequencies.

- **Step 2: Calculate Expected Counts**
    - If each face of the die should appear with equal frequency, and you rolled 60 times, the expected count for each face is 10.
- **Step 3: Calculate the Chi-Square Statistic**
    - Similar to the Chi-Square test of independence, you calculate the Chi-Square statistic using the formula: $\chi^2 = \sum \frac{(O - E)^2}{E}$ $\chi^2 = \sum \frac{(O - E)^2}{E}$
- **Step 4: Compare the Statistic to the Critical Value or p-value**
    - The degrees of freedom here is: $\text{df} = \text{number of categories} - 1$ $\text{df} = \text{number of categories} - 1$
    - Compare the Chi-Square statistic to the critical value, or use the p-value approach.
- **Step 5: Decision**
    - If the calculated Chi-Square statistic is greater than the critical value or the p-value is less than the significance level, reject the null hypothesis (meaning the observed distribution does not match the expected one).
    - If not, fail to reject the null hypothesis (the observed distribution fits the expected distribution).

Assumptions of the Chi-Square Test:

- The data must be categorical (e.g., gender, preference).
- The observations must be independent (no repeated measures).
- The expected frequency in each category should be 5 or more (for reliable results).

Conclusion:

- **Chi-Square Test of Independence** checks if two categorical variables are related.
- **Chi-Square Goodness-of-Fit Test** checks if observed frequencies match expected frequencies in a single categorical variable.