Unit IV

Analysis of Time Series - Measurement of Trend and Relational :-

Introduction to Time Series:

- Time series data involves observations taken on a variable or several variables over time.
- Time series analysis is essential for understanding patterns, trends, and relationships within the data.

Components of Time Series:

1. **Trend:**

- Trend represents the long-term movement or direction in a time series.
- It can be upward, downward, or stable.
- Trends can be linear or nonlinear.

2. **Seasonal Variation:**

- Seasonal variation refers to regular fluctuations in the data that occur at specific intervals, such as seasons or months.
- It is often associated with calendar effects.

3. **Cyclical Variation:**

- Cyclical variation involves oscillations or waves that are longer than the seasonal cycles and are not necessarily regular.

4. **Irregular Fluctuations:**

- Irregular fluctuations, also known as residual or random variations, are unpredictable and erratic movements in the time series.

Measurement of Trend:

1. **Moving Averages:**

- Moving averages are used to smooth out short-term fluctuations and highlight long-term trends.
- Simple Moving Average (SMA) calculates the average of a fixed number of data points over a specified period.
- Weighted Moving Average (WMA) assigns different weights to different data points.

2. **Exponential Smoothing:**

- Exponential smoothing is a weighted moving average method that gives more weight to recent observations.
- It is based on the principle that recent observations are more relevant for predicting the future.

3. **Least Squares Method:**

- The least squares method is used to fit a line to the time series data.
- It minimizes the sum of the squared differences between the observed and predicted values.

Relational Analysis:

1. Correlation Coefficient:

- The correlation coefficient measures the strength and direction of a linear relationship between two variables.
- It ranges from -1 to 1, where -1 indicates a perfect negative linear relationship, 0 indicates no linear relationship, and 1 indicates a perfect positive linear relationship.

2. Covariance:

- Covariance measures how two variables vary together.
- Positive covariance indicates a direct relationship, while negative covariance indicates an inverse relationship.

3. Cross-Correlation:

- Cross-correlation measures the similarity between two time series as a function of the displacement of one relative to the other.
- It helps identify lagged relationships between variables.

Statistical Tests in Time Series Analysis:

1. Augmented Dickey-Fuller Test:

- ADF test is used to test the stationarity of a time series.
- It helps determine if differencing is necessary to achieve stationarity.

2. Granger Causality Test:

- Granger causality test assesses whether one time series can be used to predict another.
- It helps identify the direction of causality between variables.

Unit IV

Conclusion:

- Time series analysis is crucial for extracting meaningful information from temporal data.
- Proper measurement of trend and relational analysis helps in understanding the underlying patterns and dependencies within the time series.

Chi-Square Test:-

Introduction: The Chi-Square Test is a statistical test used to determine whether there is a significant association between two categorical variables. It comes in two main forms: the Chi-Square Test for Independence and the Chi-Square Test for Goodness of Fit.

1. Chi-Square Test for Independence of Attributes:

Objective: To determine whether there is a significant association between two categorical variables.

Hypotheses:

- Null Hypothesis (H_0): The two categorical variables are independent.
- Alternative Hypothesis (H_1): The two categorical variables are not independent; there is an association.

Test Statistic:

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

where:

- O_i= observed frequency for category
- E_i= expected frequency for category
-

Degrees of Freedom:

$$df = (rows - 1) \times (columns - 1)$$

Decision Rule:

Compare the calculated χ^2 statistic with the critical χ^2 value. If the calculated χ^2 is greater than the critical value, reject the null hypothesis.

2. Chi-Square Test for Goodness of Fit:

Objective:

To test whether an observed frequency distribution differs from a theoretical (expected) distribution.

Hypotheses:

- Null Hypothesis (H_0): The observed and expected frequencies are consistent.

Unit IV

- Alternative Hypothesis (H_1): There is a significant difference between observed and expected frequencies.

Test Statistic:

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

where:

- O_i = observed frequency for category
- E_i= expected frequency for category

Degrees of Freedom:
df=(number of categories or groups) - 1

Decision Rule:
Compare the calculated χ2 statistic with the critical χ2 value. If the calculated χ2 is greater than the critical value, reject the null hypothesis.

Procedure for both Tests:

1. Formulate Hypotheses:

- State the null and alternative hypotheses.

2. Collect and Organize Data:

- Create a contingency table for the Chi-Square Test for Independence.
- Create observed and expected frequency distributions for the Chi-Square Test for Goodness of Fit.

3. Calculate Expected Frequencies:

- For independence, calculate expected frequencies for each cell.
- For goodness of fit, determine the expected frequencies based on the theoretical distribution.

4. Calculate the Chi-Square Statistic:

- Use the formula to compute the test statistic.

5. Determine Degrees of Freedom:

- Based on the number of categories or groups involved.

6. Find Critical Value:

- Refer to the Chi-Square distribution table.

7. Make a Decision:

- Compare the calculated χ2 statistic with the critical value.
- If $\chi 2_{calculated} > \chi 2_{critical}$, reject the null hypothesis.

8. Draw Conclusions:

- Based on the decision made in step 7, interpret the results in the context of the specific problem.

Conclusion:

- The Chi-Square Test is a powerful tool for analyzing the association between categorical variables and assessing the goodness of fit between observed and expected frequency distributions. It is widely used in various fields, including biology, sociology, and market research.

Probability:

Probability is a measure of the likelihood that a given event will occur. It is expressed as a number between 0 and 1, where 0 indicates impossibility and 1 indicates certainty.

Addition Rule:

The addition rule states that the probability of either of two mutually exclusive events occurring is the sum of their individual probabilities. Mathematically, for events A and B:

$P(A \cup B)=P(A)+P(B)$

Multiplication Rule:

The multiplication rule calculates the probability of the joint occurrence of two or more independent events. For independent events A and B:

$P(A \cap B)=P(A) \times P(B)$

Conditional Probability:

Conditional probability is the probability of an event occurring given that another event has already occurred. It is denoted as

$P(A|B)$ and is calculated as:

$P(A|B)= P(A \cap B)/ P(B)$

Unit IV

Bayes' Theorem:

Bayes' Theorem is a formula that relates the conditional and marginal probabilities of random events. For events A and B:

P(A|B)= P(B|A)×P(A)/ P(B)

Theoretical Distribution:

Theoretical distributions describe the probability distribution of a random variable. Common types include discrete distributions (e.g., binomial and Poisson) and continuous distributions (e.g., normal).

Binomial Distribution:

The binomial distribution describes the number of successes in a fixed number of independent Bernoulli trials, each with the same probability of success (p). The probability mass function is given by:

$$P(X = k) = \binom{n}{k} \times p^k \times (1 - p)^{n-k}$$

Where n is the number of trials and k is the number of successes.

Poisson Distribution:

The Poisson distribution models the number of events occurring in a fixed interval of time or space. The probability mass function is given by:

where λ is the average rate of events.

$$P(X = k) = \frac{e^{-\lambda} \times \lambda^k}{k!}$$

Normal Distribution:

The normal distribution, or Gaussian distribution, is a continuous probability distribution that is symmetric and bell-shaped. It is fully characterized by its mean (μ) and standard deviation (σ).

The probability density function for the normal distribution is:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$