Unit V: Time Based Data

A. Time Series as Chronological Data

A **Time Series** is a set of statistical observations arranged in **chronological order**. Mathematically, if we represent the time of observation as t and the value of the variable as y_t, a time series is defined as a functional relationship: $y_t = f(t)$ where t represents time (years, months, days, etc.) and y is the dependent variable (sales, temperature, population, etc.).

Identification

- **Chronology:** Data must be recorded at successive, usually equal, intervals of time.
- **Univariate Nature:** Analysis typically focuses on the variations of a single variable over time.
- **Purpose:** The primary goal is to understand past behavior and forecast future tendencies.

B. Components of Time Series

Variations in a time series are caused by a combination of factors. The sources describe how complex systems (like matrices or functions) can be broken down into constituent parts. In time series, we use the **Additive Model** or the **Multiplicative Model**.

1. Secular Trend (T)

The **long-term tendency** of the data to increase, decrease, or remain stagnant over an extended period. (Detailed in Section D).

2. Seasonal Variations (S)

These are rhythmic fluctuations that occur regularly within a period of **one year or less**.

- **Causes:** Weather conditions (e.g., ice cream sales in summer) or social customs (e.g., gift shopping in December).

3. Cyclical Variations (C)

Fluctuations that recur over a period of **more than one year**. These are usually associated with the "Business Cycle" consisting of four phases: Prosperity, Recession, Depression, and Recovery.

4. Irregular Variations (I)

Also known as **random** or **residual** variations. These are unpredictable and caused by unforeseen events like floods, strikes, or pandemics.

The Mathematical Models

- **Additive Model:** $y_t = T + S + C + I$
- **Multiplicative Model:** $y_t = T \times S \times C \times I$

C. Time Series Analysis for Univariate Data

Univariate analysis involves interpreting the data based solely on its historical values.

Solving Practical Problems

When presented with statistical data, the student must:

1. **Plot the data:** Use a "Historigram" (Time Plot) to visualize fluctuations.
2. **Smoothing:** Apply methods (like Moving Averages) to eliminate short-term fluctuations and reveal the underlying trend.
3. **Interpretation:** If the smoothed curve shows a steady rise, it indicates a positive secular trend.

D. Secular Trend: The Long-Term Tendency

The **Secular Trend** is the most significant component for long-term planning. It ignores short-term zig-zags (seasonal or irregular) to focus on the "big picture."

- **Linear Trend:** When the data changes by a constant amount per unit of time. Represented by a straight line: $y = a + bt$
- **Non-linear (Curvilinear) Trend:** When the rate of change is not constant. Represented by a parabolic or exponential function.

E. Methods of Measuring Trend

There are four primary techniques used to determine the trend in a time series.

1. Freehand Graphic Method

The simplest method where a trend line is drawn through the data points on a graph by visual inspection.

- **Advantage:** Extremely simple and flexible.
- **Disadvantage:** Highly subjective; different analysts will draw different lines.

2. Method of Semi-Averages

1. Divide the data into two equal parts (if n is odd, omit the middle value).
2. Calculate the arithmetic mean for each part.
3. Plot these two points against the mid-points of their respective time periods.
4. Join the points to form the trend line.

3. Method of Moving Averages

This method "smooths" the data by calculating a series of averages over a fixed period (k).

- **3-Year Moving Average:**
  $$\text{Trend}t = \frac{yt - 1 + y_t + y_{t+1}}{3}$$
- **Effect:** It eliminates periodic fluctuations if the period of the moving average coincides with the period of the cycle.

4. Method of Least Squares

This is the most accurate mathematical method. It uses **linear algebra** and **calculus** principles found in the sources to minimize the sum of the squares of the vertical deviations between the observations and the trend line.

For a linear trend $y = a + bt$, we solve the **Normal Equations**:

1. $\sum y = na + b\sum t$
2. $\sum ty = a\sum t + b\sum t^2$

Simplification Shortcut: If we shift the origin such that $\sum t = 0$ (by letting $x = t - \text{middle year}$), the equations simplify to: $a = \frac{\sum y}{n}$ $b = \frac{\sum xy}{\sum x^2}$

Summary Table: Comparing Trend Methods

Method	Accuracy	Basis	Complexity
Freehand	Low	Visual Inspection	Very Low

Semi-Averages	Medium	Means of two halves	Low
Moving Averages	High	Successive Averages	Medium
Least Squares	Very High	Mathematical Optimization	High

Key Formulas for Calculation

Purpose	Formula
Linear Trend Equation	$y = a + bx$
Normal Equation 1	$\sum y = na + b\sum x$
Normal Equation 2	$\sum xy = a\sum x + b\sum x^2$
Moving Average (Period k)	$\frac{\sum_{i=1}^{k} y_i}{k}$