WELCOME

BUSINESS STATISTICS

PREPARED BY :
G.KAMALAM, B.COM.,M.COM.,M.PHIL.,SET,PGDCA,PGDPM
ASSISTANT PROFESSOR (SF)
DEPARTMENT OF COMMERCE CS
ST.JOHN'S COLLEGE
PALAYAMKOTTAI

INTRODUCTION TO STATISTICS

Importance :

- **Statistical** knowledge helps you use the proper methods to collect the data, employ the correct analyses, and effectively present the results.
- **Statistics** is a crucial process behind how we make discoveries in science, make decisions based on data, and make predictions

Limitations

- Statistical methods are best applicable to quantitative data.
- Statistics cannot be applied to heterogeneous data.
- If sufficient care is not exercised in collecting, analyzing and interpreting the data, statistical results might be misleading.
- Only a person who has an expert knowledge of statistics can handle statistical data efficiently.

Sources Of Data

- Primary, or "statistical" sources are data that are collected primarily for creating official statistics, and include statistical surveys and **censuses**. Secondary, or "non-statistical" sources, are data that have been primarily collected for some other purpose (administrative data, private sector data etc).

Techniques

- Two main statistical methods are used in data **analysis**: **descriptive statistics**, which summarize data from a sample using indexes such as the mean or standard deviation, and **inferential statistics**, which draw conclusions from data that are subject to random variation (e.g., observational errors, sampling variation).

Sampling

In statistics, quality assurance, and survey methodology, sampling is the selection of a subset of individuals from within a statistical population to estimate characteristics of the whole population. Statisticians attempt for the samples to represent the population in question

Classification And Tabulation Of Data

- A process of condensing **data** and presenting it in a compact form, by putting **data** into **statistical** table, is called **tabulation**
- **Data classification** is based on similar attributes and variables of the observations. Conversely, in **tabulation** the **data** is arranged in rows and columns

Diagrammatic Representation Of Data

- **Diagrammatic presentation** is a technique of presenting numeric **data** through Pictograms, Cartograms, Bar Diagrams & Pie Diagrams etc. It is the most attractive and appealing way to represent **statistical data**

Measures Of Central Tendency

- A **measure** of central **tendency** is a summary **statistic** that represents the center point or typical value of a dataset. ... In **statistics**, the three most common **measures** of central **tendency** are the mean, median, and mode. Each of these **measures** calculates the location of the central point using a different method.

Mean

- The **statistical mean** refers to the **mean** or average that is used to derive the central tendency of the data in question. It is determined by adding all the data points in a population and then dividing the total by the number of points. The resulting number is known as the **mean** or the average.

Median

- The **median** is a simple measure of central tendency. To find the **median**, we arrange the observations in order from smallest to largest value. If there is an odd number of observations, the **median** is the middle value. If there is an even number of observations, the **median** is the average of the two middle values

Mode

- The mode of a set of data values is the value that appears most often. If X is a discrete random variable, the mode is the value x at which the probability mass function takes its maximum value. In other words, it is the value that is most likely to be sampled

Formula for Mean

1) Arithmetic Mean:

- $\bar{x} = (\Sigma x_i) / n$
- $\bar{x} = \sum fx / \sum f$

2) Geometric Mean:

- $Gm = \sum \log x / n$
- $Gm = \sum f \log x / n$
- $Gm = \sum f \log m / n$

3) Harmonic Mean:

- $Hm = N / \sum (1/x)$
- $Hm = N / \sum f(1/x)$

Formulae

Median:

- Median = $(n+1/2)^{th\ item}$
- Median = L + (n/2) – cf / f x c

Mode:

- Z = L + **Δ** 1 / **Δ** 1 + **Δ** 2 x c

Measures Of Dispersion

Range:

- The range of a set of data is the difference between the largest and smallest values.
- Formula is **Range = L-S**

Mean Deviation

- **Definition** of **mean deviation**. : the **mean** of the absolute values of the numerical differences between the numbers of a set (such as statistical data) and their **mean** or median.
- Formula is $MD = \sum |D| / N$

  $MD = \sum F|D| / N$, this is used for both discrete and continuous series

Standard Deviation

- The **standard deviation** is a statistic that measures the dispersion of a dataset relative to its mean and is calculated as the square root of the variance
- Formula is $\sigma = \sqrt{\sum x2 / N}$

$$\sigma = \sqrt{\sum Fx2 / N}$$

$$\sigma = \sqrt{\sum Fm2 / N}$$

Skewness

- In probability theory and statistics, skewness is a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean. The skewness value can be positive or negative, or undefined.

Formula of Skewness

- skewness = (3 * (mean - median)) / standard deviation

Coefficient of Skewness:

- The **coefficient** compares the sample distribution with a normal distribution. The larger the value, the larger the distribution differs from a normal distribution. A value of zero means no **skewness** at all. A large negative value means the distribution is negatively skewed.

CORRELATION

- **Correlation** is a **statistical** technique that can show whether and how strongly pairs of variables are related. For example, height and weight are related; taller people tend to be heavier than shorter people . An intelligent **correlation** analysis can lead to a greater understanding of your data.

Types of correlation

- Usually, in **statistics**, we **measure** four **types of correlations**:
- Pearson **correlation**
- Kendall rank correlation
- Spearman **correlation**
- Point-Biserial **correlation**.

Regression

In statistical modeling, regression analysis is a set of statistical processes for estimating the relationships between a dependent variable and one or more independent variables.

Analysis:

- **Regression analysis** is a setof **statistical** methods used for the estimation of relationships between a dependent variable and one or more independent variables

Regression Equations

- The **Regression Equation**. A **regression equation** is a **statistical** model that determined the specific relationship between the predictor variable and the outcome variable.

Regression Formula:

- $Y = a + bX$

Index Numbers

- An **index number** is the measure of change in a variable (or group of variables) over time. **Index numbers** are one of the most used **statistical** tools in economics. **Index numbers** are not directly measurable, but represent general, relative changes. They are typically expressed as percents

Types of Index Numbers

- One such very important tool are **index numbers**. They help reveal the trends and tendencies of the economy and also help in the formulation of economic policies and laws. There are broadly three **types of index numbers** – **price index numbers**, **value index numbers**, and **quantity index numbers**.

Analysis of Time Series

- **Time series analysis** is a **statistical** technique that deals with **time series** data, or trend **analysis**. ... **Time series** data: A set of observations on the values that a variable takes at different **times**. Cross-sectional data: Data of one or more variables, collected at the same point in **time**.

Importance

- **Time series analysis** is use in order to understand the underlying structure and function that produce the observations. Understanding the mechanisms of a **time series** allows a model to be developed that explains the data in such a way that prediction, monitoring, or control can occur.

Components of Time series

- An observed **time series** can be decomposed into three **components**:
- the trend (long term direction)
- the seasonal (systematic, calendar related movements)
- irregular (unsystematic, short term fluctuations).

Methods of Least Square

- The **least squares method** is a **statistical** procedure to find the best fit for a set of data points by minimizing the sum of the offsets or residuals of points from the plotted curve. **Least squares regression** is used to predict the behavior of dependent variables

Measurement of Trend

- **Measurement of Trend** by the Method of Moving Average
- It **measures** the **trend** by eliminating the changes or the variations by means of a moving average. The simplest of the mean used for the **measurement** of a **trend** is the arithmetic means (averages).

THANK YOU