

MAR GREGORIOS COLLEGE OF ARTS & SCIENCE

Block No.8, College Road, Mogappair West, Chennai – 37

Affiliated to the University of Madras
Approved by the Government of Tamil Nadu
An ISO 9001:2015 Certified Institution



DEPARTMENT OF COMMERCE

SUBJECT NAME: BUSINESS STATISTICS

SUBJECT CODE: CDZ3A

SEMESTER: III

PREPARED BY: PROF. B. HARISWARAN

BUSINESS STATISTICS SYLLABUS

Objectives

- □ To facilitate the understanding of the relevance and need of the Statistics in the Current Scenario.
- □ To Customize the importance of Business Statistics for the Commerce Students.

UNIT - I Introduction

Meaning and Definition of Statistics - Collection and Tabulation of Statistical Data - Presentation of Statistical Data - Graphs and Diagrams

UNIT- II Measures of Central Tendency and Measures of Variation

Measures of Central Tendency - Arithmetic Mean, Median, Mode, Harmonic Mean and Geometric Mean. Measures of Variation - Standard deviation - Mean Deviation - Quartile Deviation - Skewness and Kurtosis - Lorenz Curve

UNIT- III Correlation and Regression Analysis

Simple Correlation - Scatter Diagram - Karl Pearson's Correlation - Spearman's Rank Correlation - Regression - Meaning - Linear Regression.

UNIT - IV Time Series

Analysis of Time Series - Causes of variation in Time Series Data - Components of Time series; Additive and multiplicative models - Determination of Trend by Semi average, Moving average and Least squares (Linear, Second degree and Exponential) Methods - Computation of Seasonal indices by Simple average, Ratio-to-moving average, Ratio-to Trend and Link relative methods

UNIT - V Index Numbers

Meaning and Types of Index numbers - Problems in Construction of Index numbers - Methods of Construction of Price and Quantity indices - Tests of adequacy - Errors in Index numbers - Chain Base Index numbers - Base shifting - splicing - deflating - Consumer Price index and its uses - Statistical Quality Control

UNIT-1
Introduction of Statistics

INTRODUCTION

The word ‘Statistics ‘ is derived from a Latin term “Status’ or Italian term ‘Statistics’ or the German term ‘Statistick’ is the French term ‘Statistique’ each of which means a political state. The term statistics was applied to mean facts and figures and figures which were needed the state in respect of the division of the state, their respective population birth rate, income and the like.

Statistics -Meaning:

The term ‘Statistics’ is conveyed to two different things, In the plural use, statistics means some systematic collection of numerical data about some particular topic.

In the singular use, it means the science of statistics. In the general practice, statistics is used to mean the science of statistics and data or statistical data used for the numerical variables.

Statistics – Definition

“Statistics are numerical statement of facts in any department of enquiry placed in relation to each other”. – **A.I.Bowley**.

“Statistics may be defined as the science of collections presentation, analysis and interpretation of numerical data”.- **Croxton and Cowden**.

Characteristics of statistics

Aggregate of fact’s

Statistical enquiry is to get information from a mass of observation with regards to the group behavior of individual items. For example, the aggregate of figures related to production, sale and profit over different times is called statistics.

Numerically expressed

Numerical expression of the observed fact in terms of quantitative standards of particular scores could be regarded as statistics.

Estimated

The numerical data pertaining by field of enquiry can be observed either by enumerating or by estimation. Enumeration is used for sell field of enquiry while estimation is used for wide and large field of enquiry.

Standard of Accuracy

In case of enumeration and estimation, it is essential to fix the desired standard of accuracy beforehand.

Predetermined purpose

The purpose of enquiry is specifically stated, and then the data should be collected in a systematic manner through some suitable plan, so as to make the figures free from bias and errors.

Comparability

The ultimate aim of statistical data is, for the purpose of the comparative or relative study. Therefore, it is homogeneous to make valid comparison.

Objective of statistics

- To improve the unknown and to cast light upon the statistics out of facts and figures
- To enable comparison to be made between past and present
- To throw light on the reasons of changes, effects of changes and plans for future
- To handle analyze and draw valid inferences.
- To help to drawing conclusion from facts effected by a multiplicity of causes

Importance / Scope of Statistics / Application of Statistics in various fields in States

Statistics was regarded as the “Science of Kings”. It supplies the essential information to run the government; Policies are adopted by the government with the help of statistics.

In economics

In economics, the problems are studied by the use of statistical methods economic loss is based on the study of collected statistical data. The loss economics refer to statistics to prove their accuracy. Statistics in economics as given birth to a new discipline called econometrics.

In Business

In the competitive business, the business people face some like shortage is overstocking, uneconomic crisis etc., which can be solved through statistical analysis. To a greater extent statistics help the businessman maximize their profit.

In Education

Statistics is widely used in education for research purpose. It is used to test the past knowledge and evolved new knowledge.

In Astronomy

Astronomers study the eclipse and astronomical issues by applying statistics. They rely on estimation in many cases and it was corrected with the help of statistics.

In accounting

In accounting correlation analysis between profit and sales is widely used. In auditing, Sampling techniques are commonly followed.

In Banking

In this past developing technology, the banking sector needs a lot of information about the present and future business development.

In Investment Decision

Statistics helps an investors in selecting securities, which are safe, yielding a good return an appreciation in the market price.

In Insurance

Statistics is extensively used in the field of Insurance. Actuarial statistics is must of the insurance company through fix the premium relates which is based on the mortality tables.

Market researchers largely depends upon statistical methods in drawing conclusion

In management

Statistical tools are used widely by business enterprises for the promotion of new business.

It also helps in the assessment of quantum of product to be manufactured, the amount of raw material, labor needed, marketing avenues for the product and the competitive products in the market and so on.

In Industry

In Industry statics is used in quality control through control charge which has its basis on the theory of probability, normal distribution and inspection, which are based on sampling techniques

In Medical sciences

In medical sciences, the test of significance by student T –test for testing the efficiency of new drug, injection for controlling and curing specific ailments is done carried out by statistics. Comparative study for the effectiveness of different medicine by different concerns can also be also be made by statistical techniques of T & F test of significance

In War

The theory of decision functions propounded by A.Wald can be of great assistance to the military and technical personnel to plan maximum destruction with minimum effort. Moreover, the statistical data obtained in the post war period reveal some useful information for planning future military strategies.

Functions of Statistics

- It prevents facts in a definite numerical form
- It simplifies the complexity of the data
- It provides a technique of comparison
- It helps in formulation and testing hypothesis
- It helps in forecasting of future trends and tendencies
- It studies relationship
- It helps the government

Limitation of statistics

- Statistics cannot be applied to individual term
- Statistical study qualitative phenomena in indirect form
- Statistical law are not exact
- Statistical results are uncertain
- Statistics is not simple
- Statistical data may be incomparable
- Statistics is liable to be misused
-

Collection of Data Meaning

Data Collection means the assembling for the purpose of a particular investigation of entirely new data, presumably not available in published sources

Data: meaning

Data refer to the facts, figures or information collected for a specific purpose

Types of Data

Primary data & Secondary data

Choice between Primary and Secondary Data

- Nature and scope of the enquiry
- Availability of financial resources
- Availability of time
- Degree of accuracy designed
- Collecting agency
- Primary Data
- Primary data are new and original in nature which are firsthand information generated to achieve the purpose of the research

Advantages of Primary data

- First and new information
- More reliable
- Formulated in such a manner, which best suits the purpose

Methods of collection of Primary Data Experiment Method

Here the researcher examines the truth contained in his hypotheses by conducting experiments, through which the data are collected.

Survey Method

Under this method, data can be collected by any one or more of the following ways:

A) Observation method

This method refers to the collection of information by way of investigator's own

observation without interviewing the respondents.

B) Interview Method

In the interview method, a lot of questions relation to the proposed study is prepared and the answer for these questionnaires obtained from the respondents.

C) Mailed Questionnaire method

Under this method, the questionnaire is sent to the respondents with a covering letter to fill up the questionnaire and send back within a specified time.

D) Through Schedules

Under this method, enumerators are appointed and trained .Who will take the questionnaire to the respondents and fill the answer to the questions, obtained from the respondents

Secondary data

Secondary data are not new and original in nature which are obtained from published and or unpublished sources

Sources of Secondary Data

- **Published Sources**
- **Un published Sources**

Classification and tabulation

Meaning

Classification is the process of arranging the data under various understandable homogeneous groups for the purpose of convenient interpretation. The grouping of data is making on the basis of common characteristics

Definition

The process of grouping a large number of individual facts or observations on the basis of similarity among the items is called classification. - Stactor and Clark

Characteristics of classification

- All facts can be arranged into homogeneous groups
- Classification may be according to their resemblances and affinities
- Classification may be made on either actuality or nationality
- Going expression to the unity of attributes
- It should be flexible to accommodate adjustment

Objectives of classification

- To facilitate comparison
- To study the relationship
- To trace location of important facts at a glance
- To eliminate unnecessary details
- To effect statistical treatment of the collected data

- To facilitate easy interpretation

Significance of classification

- It is helpful to tabulation
- It leads to a valid result
- It makes interpretation clear and meaningful

Types of Classification Geographical Classification

In this type the data are classified on the basis of geographical locational differences among various items on the basis of states districts, cities, regions, and the like

Chronological Classification

Under this type data are classified in them basis of differences in time or period such as rainfall for 12 months.

Qualitative Classification

In this classification, data are classified on the basis of some attributes or qualitative phenomena such as religion, sex, marital status, literacy, occupation and the like.

Quantitative Classification

Under this type data are classified according to some quantitative phenomena capable of quantitative measurement such as age, experience, income, prices, production, sales and the like

Frequency Distribution

Frequency distribution is the process or method in simplify mass of data into grouped form of classes and the member of items in such class is recorded

- Univariate Frequency Distribution
- Bivariate Frequency Distribution

a. Univariate Frequency distribution

It is one way frequency single variable distribution and further classification into

- Individual Observation
- Discrete Frequency Distribution
- Continuous Frequency Distribution

b. Bivariate Frequency Distribution

Bivariate Frequency Distribution is a two way Frequency distribution, where two variables are measured in the same set of items through cross distribution

Tabulation

Tabulation is a systematic arrangement of raw data in a compact form of horizontal Rows and vertical column

Uses of tables

- It simplifies the presentation
- It facilitates comparison
- It is easier to distend the required information
- It reflects the trends and tendencies

Parts of tables

- Table Number
- Title of the table
- Head Note
- Caption
- Body of the table
- Source Note
- Foot Note

Diagrammatic and Graphic Presentation Diagrams and Graphs

Diagrams and graphs are easy methods of understanding of data as they are a visual form of presentation of presentation of statistical data.

Diagrams and attractive and useful to find out the result. Data should be simplified before presenting in the diagram. Two or more sets of data can be compared with the help of diagrams. Diagrams provide moiré information than the table.

Methods of Diagrams

Points, lines, bars, squares, rectangles, circles cube and so on.

Types of Charts

- Charts, pictures Maps and the like
- Advantages of diagrams
- Visual form of presentation Provide attractive and Impressive view
- Save time and labour
- Made Comparison Easy
- Useful for production
- Provide more information

Limitations of Diagrams and Charts

- Further analysis is not possible
- They show only approximate values
- All details cannot be presented diagrammatically and or graphically
- Construction of diagrams and graphs require some skill
- It is complementary in the table but not an alternative to it

Types of Diagrams

- One dimensional diagram
- Two dimensional diagram
- Three dimensional diagram
- Pictogram
- Cartograms

Bar Diagrams

Bar is a thick wide line. Statistical data presented in the form of bar is called bar diagram. Simple bar diagram is commonly used in business

Types of bar diagram

- Simple bar diagram
- Percentage bar diagram
- Bilateral deviation bar diagram
- Multiple bar diagram
- Sub divided bar Diagram

Unit-2

Measures of Central Tendency and Measures of Variation

MEASURES OF CENTRAL TENDENCY

Average Meaning

Average is a single value that represents group of values

Definition

An Average is a value which is typical or representative of a set of data

Characteristics of a Good Average

- It should be defined clear and unambiguous so that it leads to one and only one interpretation by different persons
- It should be easy to understand and simple to compute and should not involve heavy arithmetical calculations
- It should be based on all the items of the given set of data to compute the average.
- It should be suitable for further algebraic mathematical treatment and capable of being used in further statistical computations

Uses of Average

- It is useful to describe the distribution in a concise manner
- It is useful to compare different distributions

- It is useful to compare various statistical measures such as dispersion, skewness, kurtosis and so on

Functions of An average

- To facilitate Quick understanding of complex data
- To facilitate Comparison
- It establishes mathematical relationship
- Capable of further statistical comparison

Types of Average

- Mathematical Average
- Location Average
- Commercial Average

Objectives of an Average

- To get a single value that describe the features of the entire group
- To provide ground for better comparison
- To provide ground for further statistical computation and analysis

Arithmetic Mean

The arithmetic mean of a series of items is the sum of the values of all items divided by that total number. It is a multinational average and it is the most popular measure of central tendency

Merits of Arithmetic Mean

- Easy to calculate and understand
- It is a perfect average, affect by the value of every item in the series
- It is calculated value and not based on position in the series
- It is determined by a rigid formula. Hence, everyone who computes the average gets the same answer
- It is used in further calculation
- It gives a good base for comparison

Demerits of Arithmetic Mean

- The mean is unduly affected by the extreme items
- It is unreliable It may lead to a false conclusion
- It is not useful for the study of qualities
- It cannot be located by the graphic method

Arithmetic Mean Individual Series

Find our mean from the following data

Roll No	1	2	3	4	5	6	7	8	9	10
Marks	21	30	28	40	26	34	40	9	15	17

Solution:

Roll No	Marks (X)
1	21
2	30
3	28
4	40
5	26
6	34
7	40
8	9
9	15
10	17
N=10	$\sum X = 300$

Formula = $X = \frac{\sum X}{N}$ $X = \frac{300}{10} = 30$.

The mean marks = 30

Discrete Series

Calculate the arithmetic mean for the wages of workers in a Factory

Wages in Rs.	4	6	8	10	15	16
Workers	5	15	6	7	8	2

Solution

Wages in Rs.	Workers f	fx
4	5	$4 \times 5 = 20$
6	15	$6 \times 15 = 96$
8	6	$8 \times 6 = 48$
10	7	$10 \times 7 = 70$
15	8	$15 \times 8 = 120$
16	2	$16 \times 2 = 32$
	$N = \sum f = 43$	$\sum fx = 380$

$$X = \frac{\sum fx}{N} = \frac{380}{43}, N=43$$

$$= \frac{380}{43} = 8.837$$

The average wage of workers = Rs.8.84

Continuous Series Calculate Arithmetic Mean

Class Intervals	0-10	10-20	20-30	30-40	40-50
Frequency	6	5	8	15	7

Class Intervals	Mid-point	Frequency	fm
0-10	5	6	30
10-20	15	5	75
20-30	25	8	200
30-40	35	15	525
40-50	45	7	315
		$N = \sum f = 41$	$N = \sum fm = 1145$

Arithmetic Mean = $X = \frac{\sum fm}{N}$

The Arithmetic mean = 27.92

Median

Median is the value of the middle item of a series arranged in ascending or descending order of magnitude. Hence it is the “Middle most” or “Most central” value of a set of number. It divide the series into two equal part, one part containing values greater and the other with values less than the median.

Meaning

The number is that value of the variable which divides the group into two equal parts, one part comprising all values greater and the other, all values less than median.

Merits of Median:

- It is easy to compute and understand
- It eliminates the effect of extreme item
- The value of median can be located graphically
- Demerits of Median
- The calculating media, it is necessary to arrange the data other averages do not need an arrangement
- It is affected more by fluctuation of sampling than the arithmetic mean.
- It is not based on all the items of the series

Individual Series

Arrange the data either ascending or descending order

Median – Size of $N + 1$
th
 Item2

Find out the median from the following

57	58	61	42	38	65	72	66	80
----	----	----	----	----	----	----	----	----

Solution

Sl.No	Data arranged in ascending order
1	38
2	42
3	57
4	58
5	62
6	65
7	66
8	72
9	80

Median= – Size of $N + 1$
th Item
 2

- = Size of $9 + 1$
 - ----- the item

- 2
 - = $10 / 2 = 5^{\text{th}}$ item

Median = 62

Discrete Series

Compute the median for the following distribution of weeks of wagers of 65 employees of the xyz company

Weekly wages in Rs	55	65	785	85	95	105	115
Number of employees	8	10	16	14	10	5	2

Solution

Weekly wages in Rs	No of Employees	Cumulative frequency (cf)
55	8	8
65	10	18
75	16	34
85	14	48
95	10	58
105	5	63
115	2	65

$$\text{Median} = \frac{\text{Size of } N + 1}{2} \text{th Item}$$

$$= \frac{65 + 1}{2} \text{th Item}$$

$$= 33 \text{ which is nearer to } 34$$

$$\text{Cf of } 34 = 75$$

$$\text{Median weekly wages} = 75$$

Continuous Series

Calculate the median form the following data

Marks	0-20	20-40	40-60	60-80	80-100
No of Students	5	15	30	8	2

Solution

Marks	No of students	Cumulative frequency
0-20	5	5
20-40	15	20
40-60	30	50
60-80	8	58
80-100	2	60

$$\text{Median} = \frac{\text{Size of } N}{2} \text{th Item c size of } \frac{60}{2} \text{th Item}$$

$$= \frac{N}{2} - cf$$

$$= \text{size of } L + \frac{\frac{N}{2} - cf}{f} \times C$$

$$= 40 + \frac{30 - 20}{2} \times 20 = 46.47$$

Median marks = 46.676

Mode

Mode is the modal value in the value of the variable which occurs more number of times or most frequently is a distribution. Mode is the value which occurs with the greatest number of frequency in a series

Types of modal

i. Uni-model

If there is only one mode in series is called uni-model

ii. Bi-Modal

If there are two modes in the series, it is called bi-model

iii. Tri-Modal

If they are three modes in the series, it is Relationship between different Averages Symmetrical is called Tri-model.

iv. Multimodal

If there are more than three modes in the series it is called multi-mode.

Relationship among mean, median and mode

The three averages are identical, when the distribution is symmetrical. In an asymmetrical distribution, the values of mean, median and mode are not equal.

Median = $\frac{1}{3}$ (Mean - mode)

Mode = 2 median - 2 mode

Median = Mode * $\frac{2}{3}$ (Mean – Mode)

Individual Series

Calculate the mode form the following data of the marks obtain by 10 students

Serial No	1	2	3	4	5	6	7	8	9	10
Marks obtained	60	77	74	62	77	77	70	68	65	80

Solution

Marks obtained by 10 students is here 77 is repeated three times

Therefore the Mode mark is 77

Discrete Series

Calculate the mode from the following data of the wages of workers of an establishment. Find the modal wages

Daily wages in Rs	3	4	6	7	9	10	12	13	15
No of wage earners	2	3	2	6	10	11	12	5	1

Solution Grouping Table

Daily Wages is Rs.	Frequency of Wages Earners					
	1	2	3	4	5	6
3	2	5		7		
4	3		5		11	
6	2	8				18
7	6		16	27		
9	10	21			33	
10	11		23			28
12	12	17		18		
13	5		6			
15	1					

Analysis Table:

Column	Size of Item						
	4	6	7	9	10	12	15
1						I	
2				I	I	I	
3					I	I	
4			I	I	I		

5				I	I	I	
6					I	I	I
			I	3	5	4	1

From the analysis table it is known that size 10 has been repeated the maximum number of times, thus is, so the modal wages Rs10

Continuous series

Find out the mode from the following series

X	0-5	5-10	10-15	15-20	20-25	25-30	30-35
frequency	1	2	5	14	10	9	2

Modal value lies in 15-20 as it occurs most frequently

$$f_1 - f_0$$

$$\text{Mode (Z)} = L + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times C$$

$$14 - 5$$

$$\text{Mode (z)} = 15 + \frac{14 - 5}{2(14) - 5 - 10} \times 5$$

$$= 15 + \frac{9}{13} \times 5 = 15 + \frac{45}{13}$$

$$= 15 + 3.46$$

$$\text{Mode} = 18.46$$

Geometric Mean

Merits of geometric Mean

- Every item in the distribution is included in the calculation
- It can be calculated with mathematical exactness, provided that all the qualities are greater than zero and positive
- Large items have less effect on it than in the arithmetic average.
- It is amenable to further algebraic manipulation

Demerits of Geometric mean

- It is very difficult to calculate
- It is impossible to use it when any item is zero or negative
- The value of the geometric mean may not correspond with any actual value in the distribution

Frequency	2	3	3	1	1	7	4	2	8
-----------	---	---	---	---	---	---	---	---	---

Solution:

Size of Item (X)	Frequency (f)	Log X	F log x
120	2	2.0792	4.1584
125	3	2.0969	6.2907
1360	3	2.1139	6.3417
135	1	2.1303	2.1303
136	2	2.1335	4.2670
138	7	2.1399	14.9793
139	4	2.1430	8.5720
140	2	2.1461	4.2922
147	8	2.1673	17.3384
	N=∑f=32		N=∑f log x = 68.3700

$$G.M = \text{Anti log of } \frac{\sum \log X}{N} = \text{Anti log of } \frac{68.375}{32}$$

$$= \text{Antilog of } 2.1366 = 137 \text{ Therefore } G.M = 137$$

Continuous series

Geometric mean from the following data

Yield of wheat	7.54-10.5	10.5-13.5	13.5 -16.5	16.5-19.5	19.5-22.5	22.5-25.5	25.5-28.5
No of forms	5	9	19	23	7	4	1

Solution:

Yield of wheat	Mid Value	Log M	No. of Forms	f log m
7.54-10.5	9	0.9542	5	4.7710
10.5-13.5	12	1.0792	9	9.7128
13.5 -16.5	15	1.1761	19	22.3459
16.5-19.5	18	1.2553	23	28.8719
19.5-22.5	21	1.3222	7	9.2554
22.5-25.5	24	1.3802	4	5.5208
25.5-28.5	27	1.4314	1	1.4314
			N=68	$\sum f \log m = 81.9092$

G.M = Anti log of $\frac{\sum f \log m}{N}$

$$= \frac{81.9092}{68} = 1.204547$$

$$= \text{Antilog of } 1.204547 = 16.02 \text{ G.M} = 16.02$$

Harmonic Mean

Meaning

Harmonic Mean is the reciprocal of the arithmetic average of the reciprocal of values of various item in the invariable

Merits of Harmonic Mean

- It utilizes all values of a variable
- It is very important to small values
- It is amenable to further algebraic manipulation
- It provides consistent results in problems relating to time and rates than similar averages

Demerits of Harmonic Mean

- It is not very easy to understand
- The method of calculation is difficult

- The presence of both positive and negative items in a series makes it impossible to compute its value. The same difficulty is felt if one or more items are zero
- It is only a summary figure and may not be the actual item in the series.

Individual Series

N

$$\text{H.M} = \frac{N}{\sum 1/x}$$

Find out the Harmonic mean

Family	1	2	3	4	5	6	7	8	9	10
Income	85	70	10	75	500	8	42	250	40	36

Solution

Computation of Harmonic Mean

Famil y	Income (X)	1 / x
1	85	0.01176
2	70	0.01429
3	10	0.10000
4	75	0.01333
5	500	0.00200
6	8	0.12500
7	42	0.02381
8	250	0.00400
9	40	0.02500
10	36	0.02778
N =10		$\sum 1/x =$ 0.34697

$$\text{H.M} = \frac{N}{\sum 1/x}$$

$$= 10 / 0.34697 = 28.82 \quad \text{H.M} = 28.82$$

Discrete Series

Size of Item	6	7	8	9	10	11
Frequency	4	6	9	5	2	8

Solution:

Size of Item X	Frequency f	1/x	F 1/x
6	4	0.1667	0.6668
7	6	0.1429	0.8574
8	9	0.1250	1.1250
9	5	0.1111	0.5555
10	2	0.1000	0.20000.7272
11	8	0.0909	
	$N = \sum f = 34$		$\sum f 1/x = 4.1319$

$$H.M = \frac{N}{\sum f1/x} = \frac{34}{4.1319} = 8.23$$

Continuous Series

Compute Harmonic Mean

Size	0-10	10-20	20-30	30-40	40-50
Frequency	5	8	12	6	4

Solution

Size	Frequency f	Mid value	reciprocal	F(1/m)
0-10	5	5	0.20000	1.00000
10-20	8	15	0.06667	0.53336
20-30	12	25	0.04000	0.48000
30-40	6	35	0.02857	0.17142
40-50	4	45	0.02222	0.08888
	$\sum f = 35$			$\sum f 1/m = 2.27366$

$$H.M = \frac{N}{\sum f1/m} = \frac{35}{2.27366} = 15.393682$$

Measures of Variation or Dispersion Meaning

Dispersion is the study of scatterness around an average

Definition

Dispersion is the measures of the variation of the items ---A.L.Bowley

Dispersion is a measure of extent to which the individual items vary ----L.R.Connor

Importance of measuring variation or dispersion

- Testing the Reliability of the Measures of Central Tendency
- Comparing two or more series on the basis of their variability
- Enabling to control the variability
- Facilitating as a Basis for further statistical Analysis

Characteristics of a Measure of Variation

- It is easy to understand and simple to calculate
- It should be rigidly defined
- It should be based on all observations and it should not be affected by extreme observations
- It should be amenable to further algebraic treatment
- It should have sampling stability

Methods of Measuring Dispersion

- Range
- Inter Quartile range
- Quartile Deviation
- Mean Deviation
- Standard Deviation
- Lorenz Curve

Range

Range is the difference between the largest and the smallest value in the distribution. It is the simplest and crudest measure of dispersion

Uses of Range

- It is used in industries for the statistical quality control of the m infected product
- It is used to study the variations such as stock, shares and other commodities
- It facilitates the use of other statistical measures

Advantages of Range

- It is the simplest method of studying variation
- It is easy to understand and the easiest to compute
- It takes minimum time to calculate
- It is accurate

Disadvantages of Range

- Range is completely depended on the two extreme values
- It is subject to fluctuations of considerable magnitude from sample to sample
- It is not suitable for mathematical treatment
- It cannot be applied to open and classes
- Range cannot tell us anything about the character of the distribution

Quartile deviation

Quartile deviation is an absolute measure of dispersion. It is calculated on the basis of the difference of upper quartile and the lower Quartile divided by 2.

In the series, four quartiles are there. By eliminating the lowest (25%) items and the highest (25%) items of a series, we can obtain a measure of dispersion and can find out half the distance between the first and the third quartiles.

$$\text{Quartile Deviation (Q.D)} = \frac{Q_3 - Q_1}{2}$$

$$\text{Co-efficient of Q.D} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Merits of Quartile Deviation

- It is simple to calculate and easy to understand
- Risk of extreme item variance is eliminated, as it depend upon the central 50 per cent items
- It can be applied to open and classes

Demerits of quartile Deviation

- Items below Q1 and above Q3 are ignored
- It is not capable of further mathematical treatment
- It is affected much by the fluctuations of sampling
- It is not calculated from a computed average, but from a positional average.

Mean deviation

Mean deviation is the average difference between the items in a distribution computed from the mean, median or mode of that series counting all such deviation as positive. The mean deviation is also known as the average deviation

$$\text{Mean deviation} = \frac{\sum I D I}{N}$$

$$\text{Co-efficient of Mean Deviation (M.D)} = \frac{MD}{\text{Mean or Median}}$$

Merits of Mean Deviation

- It is clear and easy to understand
- It is based on each and every item of the data It can be calculated from any measure of central tendency and as such as flexible too.

Demerits of mean Deviation

- It is not suitable for further mathematical processing
- It is rarely used in sociological studies
- It is mathematically unsound and illogical, because the signs are ignored in the calculation of mean deviation

Standard deviation

Standard deviation is the square root of the means of the stranded deviation from the Arithmetic mean. So, it is also known as Root Mean Square Deviation an Average of Second order. Standard deviation is denoted by the small Greek letter 'σ' the concept of standard deviation is introduced by Karl Pearson in 1893.

Uses of Standard deviation

- It is used in statistics because it possesses must of the characteristics of an ideal measure of dispersion.
- It is widely used in sampling theory and by biologists.
- It is applied in co-efficient of correlation and in the study of symmetrical frequency distribution

Advantages of standard deviation

- It is rigidly defined determinate
- It is based on all the observations of a series
- It is less affected by fluctuations of sampling and hence stable
- It is amenable to algebraic treatment and is less affected by fluctuations of sampling most other measures of dispersion
- The standard deviation is more appropriate mathematically than the mean deviation, since the negative signs are removed by squaring the deviations rather than by ignoring

SKEWNESS

Introduction

The term 'Skewness' refers to lack of symmetry, that is, when a distribution is not symmetrical it is called a skewed distribution. If the curve is normal or the data distributed symmetrically or uniformly. Spread will be the same on both sides of the central point and the means, median and mode will all have the same value.

Definition

'Skewness or asymmetry is the attribute of a frequency distribution that extends further on one side of the class with the highest frequency on the other--- **Simpson and Kafka**

When a series is not symmetrical it is said to be asymmetrical or skewed -**Croxtan and Cowden**

Skewness of a Distribution

When a distribution is not symmetrical it is called a skewed Distribution.

The analysis of presence of skewness in a distribution implies two main tasks. They are

- I. Determination of the sign of skewness and testing of skewness and
- II. Determination of the extent of skewness

Absolute measures of skewness

- i) The Karl Pearson's Coefficient of Skewness
- ii) The Bowley's Co efficient of Skewness
- iii) The Kelly's Coefficient of Skewness
- iv) Measure of Skewness based on moments

Karl Pearson's Co-efficient of Skewness

This method is based upon the difference between mean and mode and the difference is divided by standard deviation to give a relative measures.

Bowley's Coefficient of Skewness

Bowley's measure is based on quartiles, in a symmetrical distribution first and third quartiles are equidistant from the median

Objectives of Skewness

- I) To find out the direction and extent of asymmetry in a series.
- II) To compare two or more series with regards to skewness.
- III) To study the nature of variation of the items about the central value.

Graphic method of dispersion Lorenz Curve

Lorenz Curve is a device used to show the measurement of economic inequalities as in the distribution of income and wealth. It can also be used in business to study the disparities of distribution of profit, wages, turnover, production and the like.

Unit-3

Correlation and Regression Analysis

Correlation and Regression Analysis

Meaning:

Correlation is the study of the natural relationship between two or more variables. Hence, it should be noted that the detection and analysis of correlation between two statistical variables requires relationship of some sort which associates the observation in pairs each of which is a value of the two variables

Definition

The relationship that exists between two variables

---Smith

Correlation analysis deals with the association between two or more variables ---- Tuite

Uses of Correlation

- I) Correlation is very useful in physical and social sciences. Business and economics
- II) Correlation analysis is very useful in economics to study the relationship between price and demand
- III) It is also useful in business to estimate costs, value, price and other related variables
- IV) Correlation is the basis of the concept of regression
- V) Correlation analysis helps in calculating the sampling error.

Types of Correlation

- Positive correlation
- Negative Correlation
- Simple Correlation
- Multiple Correlations
- Partial Correlation
- Linear Correlation
- Non-Linear Correlation

Positive Correlation

Correlation is said to be positive when the values of two variables move in the same direction, so that an increase in the value of one variable is accompanied by an increase in the value of the other variable or a decrease in the value of one variable is followed by a decrease in the value of the other variable.

Negative Correlation

Correlation is said to be negative when the values of two variables move in opposite directions, so that an increase in the values of one variable is followed by a decrease in the value of the other and vice-versa.

Simple Correlation

When only two variables are stated, it is said to be simple correlation.

Multiple Correlations

When more than two variables are stated simultaneously, the correlation is said to be multiple.

Partial Correlation

Partial correlation coefficient provides a measure of relationship between a dependent variable and a particular independent variable when all other variables involved are kept constant. Analysis of yield and rainfall; it becomes a problem relating to simple correlation.

Linear Correlation
The correlation is said to be linear, if the amount of change in one variable tends to bear a constant ratio to the amount of change in the other.

Non Linear Correlation

The correlation is non linear, if the amount of change in one variable does not bear a constant ratio to the amount of change in the other related variable.

Methods of studying correlation

Graphical method

- Scatter diagram
- Simple graph method

Mathematical Methods

- Karl Pearson's Co-efficient of correlation
- Spearman's Rank Correlation coefficient
- Concurrent deviation method
- Method of least square

Scatter diagram method

It is a method of studying correlation between two related variables. The two variables X and Y will be taken upon the X and Y axes of a graph paper. For each part of X and Y values, we mark a dot and we go as many points as the numbers of observation.

Graphical method

In this method curves are drawn for separate series on a graph paper. By examining the direction and closeness of the two curves we can offer whether prompt variances are related. If both the curves are moving in the same direction correlation is said to be positive. On the contrary, if the curves are moving in the opposite directions is said to be negative.

Karl Pearson's Co-efficient of correlation

Karl Pearson, a great statistician introduced a mathematical method for measuring the magnitude of relationship between two variables. This method. Known as Pearson Coefficient of correlation is widely used. It is denoted by the symbol "r"

Spearman's Rank Correlation Co-efficient

In 1904, a famous British psychologist Charles Edward Spearman found out the method of Co-efficient of correlation of rank. Rank correlation is applicable to individual observation. This measure is useful in dealing with qualitative characteristics. The result, by using ranking method, is only approximate.

Regression

Analysis

Meaning

The statistical method employed to estimate the unknown value of one variable from the known value of the related variables is called regression

Definition

Regression is the measure of the average relationship between two or more variables in terms of the original units of the data - Blair

Regression analysis Meaning regression analysis is statistical device with which we estimate or predict the unknown values of one variable from known value of another variable

Regression analysis definition

One of the most frequently used techniques in economics and business research, to find a relation between two or more variables that are related causally, is regression analysis.

- Taro Farnane

Uses of regression analysis

- It is useful to estimate the relationship between two variables
- It is useful for production of unknown value
- It is widely used in social sciences like economics, Natural and physical sciences
- It is useful to forecast the business situation
- It is useful to calculate correlation coefficient and coefficient of determination

Methods of studying Regression

- Graphic method
- Algebraic method

Graphic method

Under the method the dots are plotted on a graph paper representing pair of values of the given variables having a linear relationship the independent variable is taken in the X axis and the dependent variable taken on Y axis. The regression line of X on Y provides the most probable value of X given the most probable value of Y when the exact value of X is known. Thus we get two regression lines.

Regression lines

- I) Regression of X on Y
- II) Regression of Y on X

Unit-4
Time Series

Time series Analysis

- Time series analysis is the analysis of identifying different components such as trend, seasonal, cyclical and irregular in a given time series data.

Definition

A time series is a set of observation arranged in chronological order. Morris Hamberg requirement of a time series Data must be available for a long period of time. Data must consist of a homogeneous set of values belonging to different time periods. The time gap between the variables or composite of variables must be as For Possible equal.

Causes of variation in Time Series Data

- Social customs, festivals etc. Seasons
- The four phase of business: prosperity, decline, depression, recovery
- Natural calamities: earthquake, epidemic, flood, drought etc. Political movements/changes, war etc.

Components of Time series; Additive and multiplicative models

Components of Time Series

1. Secular Trend
2. Seasonal Variation
3. Cyclical Variations
4. Irregular variation

1. Secular Trend

A secular trend or long-term trend refers to the movements of the series reflecting continuous growth or decline over a long period of time. There are many types of trend. Some trends rise upward and some fall downward

2. Seasonal Variation

Is that periodic investment in business activities within the year recurring periodically year after year?

Generally, seasonal variation appear at weekly, monthly or quarterly intervals

3. Cyclical Variation

Up and down movements are different from seasonal fluctuations, in that they extend over longer period of time – usually two or more years. Business time series is influenced by the wave-like changes of prosperity and depression.

4. Irregular Variation

Irregular variations or random variations constitute one of four components of a time series. They correspond to the movements that appear irregularly and generally during short periods. Irregular variations do not follow a particular model and are not predictable.

Mathematical Model for a Time Series

In classical analysis, it is assumed that some types of relationship exist among the four components of time series

Additive Model

According to this model, the time series is expressed as $Y = T + S + C + I$

Y = the value of original time series
T = Time

Value

S =

Seasonal variation

C = Cyclical Variation

I = Irregular fluctuation

Multiplicative Model

According this model, the time series is expressed as $Y = T \times S \times C \times I$

Determination of Trend

Measurements of Trends

Following are the methods by which we can measure the trend.

- (i) Freehand or Graphic Method.
- (ii) Method of Semi-Averages.
- (iii) Method of Moving Averages.
- (iv) Method of Least Squares.

Free hand Graphic Method

In this method we must plot the original data on the graph. Draw a smooth curve carefully which will show the direction of the trend. The time is taken on the horizontal axis (X) and the value of the variable on the vertical axis (Y)

Semi – Average Method

In this method the original data are divided into two equal parts and average are calculated for both the parts. These averages are called semi average. Trend line is drawn with the help of the semi averages

Fit a trend line by the method of semi- averages for the given data.

	2000	2001	2002	2003	2004	2005	2006
	105	115	120	100	110	125	135

Solution:

Since the number of years is odd (seven), we will leave the middle year's production value and obtain the averages of first three years and last three years.

Year	Production	Average
2000	105	
2001	115	$\frac{105 + 115 + 120}{3} = 113.33$
2002	120	
2003	100 (left out)	
2004	110	
		$\frac{110 + 125 + 135}{3} = 123.33$

Moving average method

In this method, the average value of a number of years or months or weeks is taken into account and placed it at the Centre of the time span and it is the normal or trend value for the middle period.

Calculate three-yearly moving averages of number of students studying in a higher secondary school in a particular village from the following data.

Year	Number of students
1995	332
1996	317
1997	357
1998	392
1999	402
2000	405
2001	410
2002	427
2003	435
2004	438

Solution:

Computation of three-yearly moving averages.

Year	Number of students	3- yearly moving Total	3- yearly moving Averages
1995	332	---	---
1996	317	1006	335.33
1997	357	1066	355.33
1998	392	1151	383.67
1999	402	1199	399.67
2000	405	1217	405.67
2001	410	1242	414.00
2002	427	1272	424.00
2003	435	1300	433.33
2004	438	---	---

Method of Least Squares

The line of best fit is a line from which the sum of the deviations of various points is zero. This is the best method for obtaining the trend values. It gives a convenient basis for calculating the line of best fit for the time series. It is a mathematical method for measuring trend. Further the sum of the squares of these deviations would be least when compared with other fitting methods.

Fit a straight line trend by the method of least squares and tabulate the trend values.

Year	2000	2001	2002	2003	2004	2005	2006
Prod. of Sugarcane	40	45	46	42	47	50	46

Solution:

Computation of trend values by the method of least squares (ODD Years).

Year (x)	Production of Sugarcane (Y)	$X = (x - 2003)$	X^2	XY	Trend values (Yt)
2000	40	-3	9	-120	42.04
2001	45	-2	4	-90	43.07
2002	46	-1	1	-46	44.11
2003	42	0	0	0	45.14
2004	47	1	1	47	46.18
2005	50	2	4	100	47.22
2006	46	3	9	138	48.25
N=7	$\Sigma Y = 316$	$\Sigma X = 0$	$\Sigma X^2 = 28$	$\Sigma XY = 29$	$\Sigma Yt = 316$

Therefore, the required equation of the straight line trend is given by

$$Y = a + bX$$

$$Y = 45.143 + 1.036 (x - 2003)$$

The trend values can be obtained by

When $X = 2000$,

$$Y_t = 45.143 + 1.036 (2000 - 2003) = 42.035$$

$$Y_t = 45.143 + 1.036 (2001 - 2003) = 43.071,$$

similarly other values can be obtained.

Computation of Seasonal indices by Simple average

Seasonal Variations can be measured by the method of simple average. The data should be available in season wise likely weeks, months, quarters.

Method of Simple Averages:

This is the simplest and easiest method for studying Seasonal Variations.

Ratios-to-moving-average method

Ordinarily does not fluctuate so much as the index based on straight-line trends. This is because the 12-month **moving average** follows the cyclical course of the actual data quite closely.

Ratio-to-trend Method

- The ratio-to-trend method is similar to ratio-to-moving-average method
- The only difference is the way of obtaining the trend values
- Whereas in the ratio-to-moving-average method, the trend values are obtained by the method of moving averages, in the ratio-to-trend method

Link relative's method:

Link relatives method are calculated by dividing the figure of each season* by the figure of immediately preceding season and multiplying it by 100. These percentages are called link relatives since they link each month (or quarter or other time period) to the preceding one.

Unit-5

Index Numbers

Meaning and of Index number

As index number is a specialized average designed to measure the change in a group of related variable over a period of time. It was first constructed in the year

Concept

In its simplest form on Index number is a Ratio of two numbers expressed as percent.

Definition

Index number devices for measuring difference in the magnitude of a group of related variables

---- Croxtonand Cowden

Types of Index Numbers

- Price Index
- Quantity Index
- Value Index

Problems in Construction of Index numbers**Simple Aggregative Method:**

In this method, the index number is equal to the sum of prices for the year for which index number is to be found divided by the sum of actual prices for the base year.

$$P_{01} = \frac{\sum P_1}{\sum P_0} \times 100$$

Where P_{01} Stands for the index number

$\sum P_1$ Stands for the sum of the prices for the year for which index number is to be found :

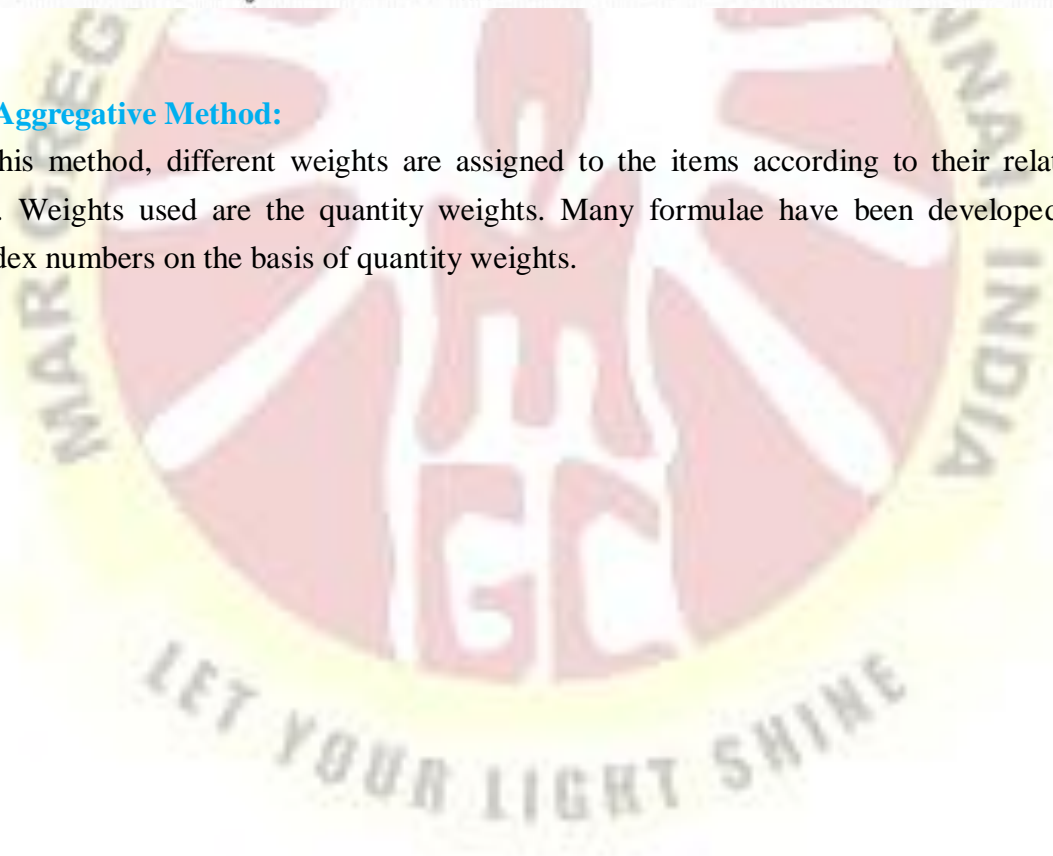
$\sum P_0$ Stands for the sum of prices for the base year.

Commodity	Prices in Base Year 1980 (in Rs.) P_0	Prices in current Year 1988 (in Rs.) P_1
A	10	20
B	15	25
C	40	60
D	25	40
Total	$\sum P_0 = 90$	$\sum P_1 = 145$

$$\text{Index Number } (P_{01}) = \frac{\sum P_1}{\sum P_0} \times 100 ; P_{01} = \frac{145}{90} \times 100 ; P_{01} = 161.11$$

Weighted Aggregative Method:

In this method, different weights are assigned to the items according to their relative importance. Weights used are the quantity weights. Many formulae have been developed to estimate index numbers on the basis of quantity weights.



- (i) **Laspeyre's Formula.** In this formula, the quantities of base year are accepted as weights.

$$P_{01} = \frac{\sum P_1 q_0}{\sum P_0 q_0} \times 100$$

Where P_1 is the price in the current year ; P_0 is the price in the base year ; and q_0 is the quantity in the base year.

- (ii) **Paasche's Formula.** In this formula, the quantities of the current year are accepted as weights.

$$P_{01} = \frac{\sum P_1 q_1}{\sum P_0 q_1} \times 100$$

Where q_1 is the quantity in the current year.

- (iii) **Dorbish and Bowley's Formula.** Dorbish and Bowley's formula for estimating weighted index number is as follows :

$$P_{01} = \frac{\frac{\sum P_1 q_0}{\sum P_0 q_0} + \frac{\sum P_1 q_1}{\sum P_0 q_1}}{2} \times 100 \quad \text{or} \quad P_{01} = \frac{L + P}{2}$$

Where L is Laspeyre's index and P is paasche's Index.

- (iv) **Fisher's Ideal Formula.** In this formula, the geometric mean of two indices (i.e., Laspeyre's Index and paasche's Index) is taken :

$$P_{01} = \sqrt{\frac{\sum P_1 q_0}{\sum P_0 q_0} \times \frac{\sum P_1 q_1}{\sum P_0 q_1}} \times 100 \quad \text{or} \quad P_{01} = \sqrt{L \times P} \times 100$$

where L is Lespeyre's Index and P is paasche's Index.

Example

Comm- odity	Base Year		Current Year		$P_0 q_0$	$P_1 q_0$	$P_0 q_1$	$P_1 q_1$
	P_0	q_0	P_1	q_1				
A	10	5	20	2	50	100	20	40
B	15	4	25	8	60	100	120	200
C	40	2	60	6	80	120	240	360
D	25	3	40	4	75	120	100	160
Total					265 $\sum P_0 q_0$	440 $\sum P_1 q_0$	480 $\sum P_0 q_1$	760 $\sum P_1 q_1$

- (i) Laspeyre's Formula :

$$P_{01} = \frac{\sum P_1 q_0}{\sum P_0 q_0} \times 100$$

$$P_{01} = \frac{440}{265} \times 100 = 166.04$$

Test of adequacy for an Index Number

- Index numbers are studied to know the relative changes in price and quantity for any two years compared.

- Factor Reversal **Test**.
- The criterion for a good index number is to satisfy the above two **tests**.
- Fisher's index number formula satisfies the above relationship.

Chain Index Numbers

Under this method, firstly we express the figures for each year as a percentage of the preceding year. These are known as Link Relatives. We then need to chain them together by successive multiplication to form a chain index.

Steps in the construction of Chain Index Numbers

1. Calculate the link relatives by expressing the figures as the percentage of the preceding year. Thus,

$$\text{Link Relatives of current year} = \text{price of current year} / \text{price of previous year} \times 100$$

2. Calculate the chain index by applying the following formula:
Chain Index = Current year relative \times Previous year link relative/100

Base Shifting

For a variety of reasons, it frequently becomes necessary to change the reference base of an index number series from one time to another without returning to the original raw data and recomposing the entire series. This change of reference base period is usually referred to as “shifting the base”.

Splicing

The process of combining two or more index numbers covering different bases into a single series is called splicing. Example the following are two series A and B of the index numbers of a commodity taking 1991 and 1994 as the base years.

Deflating

Deflating means making allowances for the effect of changing price levels. The process of adjusting a series of salary or wages or income according to current price changes to find out the level of real salary wages or income is called deflating of index numbers.

Consumer Price Index

The Consumer Price Index (CPI) measures the average price change of a set of consumer goods and services. CPIs can be calculated for single items or a predetermined group of items. All of these items are defined as "household goods and services."

Uses of Consumer Price Index

The consumer price index is mainly used to measure inflation over a given period of time. It can also be leveraged to determine the cost of living.

CPI is mainly used to determine the efficacy of economic policies. Inflation indicates the health (or lack thereof) of an economy, so tracking it and responding to it appropriately is important for policymakers. When inflation sharply increases or decreases, the CPI provides economists and policymakers insight into how a government's economic policy affects the market.

Statistical quality control

The use of statistical methods in the monitoring and maintaining of the quality of products and services. One method, referred to as acceptance sampling, can be used when a decision must be made to accept or reject a group of parts or items based on the quality found in a sample.

LET YOUR LIGHT SHINE