BUSINESS STATISTICS

UNIT I
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 make 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 over stocking, 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

Astronauts study the eclipse and astronocal 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.

In Research

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

Distrust of statistics

- Manipulation of information by dishonest and unscrupulous persons
- Deliberate twisting of facts for personal and selfish motives
- Selection of non-representative sample (or) statistical unit
- Incomparable data
- Lack of technical knowledge of statistics

Collection of
Data Meaning

Data Collection means the assembling for the purpose of a particular investigation of entirely new data, presumably nor 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 date 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 questionnaire 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**

Statistical Investigation

Census method

Each and every unit or item of the field of survey os studied and conclusions are derived by computing the sum of all units. This type of survey is called Census Survey

Sample Method

A unit or a part in the field of study is studied and the conclusions of these sample units are extended to the whole field. This type of survey is called sample method

Sample Meaning

A sample is the part of the population or universe selected for the purpose of investigation

Sample Definition

- A sample is a smaller representation of the later whole.
- Elements in the process of sampling
- Selecting the sample
- Collecting the information
- Making an inference about the population

Essentials of sampling

- Economy
- Less time consuming
- Reliability
- Detailed study

Objectives of samples

- To make an inference about an unknown parameter from a measurable sample statics
- To test the hypothesis relating to populations
- To avoid the vase study about the entire population
- To obtain quick result

Characteristics of a Good Sample

- Representativeness
- Independence
- Adequacy
- Homogeneity

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 make 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

a. Univarite Frequency Distribution
b. 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

Terms Used in the Frequency Distribution Class Limits

The class limits are the lowest and the highest values but can be included in the class

Class Frequency

The number of items included or counted in each of the classes is called class Frequency

Mid-point

Mid-point in their value lying half way between the lower and upper limits of a class interval

Methods of constructing class intervals
Exclusive method (or) Overlapping Class Limits

Under this method, the upper limit of one class would be the lower limit of the next class inclusive method under this method the upper limit of each class is not repeated as the lower limit of the next class.

Open End Class Interval

We wish to include a person whose age is 72. The inclusion necessitates three additional classes. The frequency of each of the first two classes would be and the third class 1. Insisted of adding three new intervals to account be for once case an alternative would be to add the case with the last class interval marked ’50 and above’

A class which does not give the upper and lower limits, but unsteady is defined as more than or less than the specified limit, is called as open class or open ended interval

Cumulative Frequency Distribution

Cumulative Frequency Distribution is obtained by recessively adding the frequencies of the values of the variable or classes. They are

- Less than cumulative Frequency
- More than cumulative frequency

Steps to construct a frequency distribution

Step -1 in the first column write the value

Step-2 In the next column marks a vertical bar to denote the number of repetition of a particular value. If, it is repeated fifth time cross the four bars. It is called Tally mark.
Step-3 do the same for all values
Step -4 now count then total number of frequencies will be help of tally marks

Tabulation

Tribulation is a systematic arrangement of raw data in a compact form of historical
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

UNIT –II DIAGRAMMATIC AND GRAPHIC REPRESENTATION

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 labor
- 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

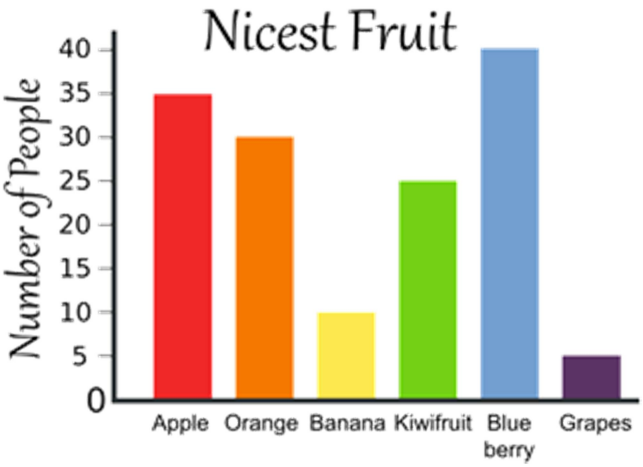
Bar Diagrams

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

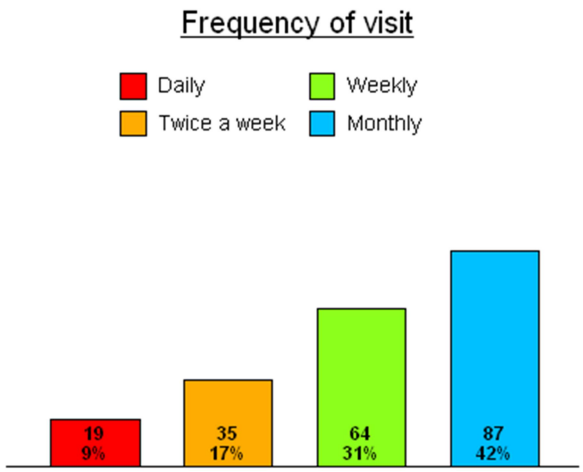
Types of bar diagram

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

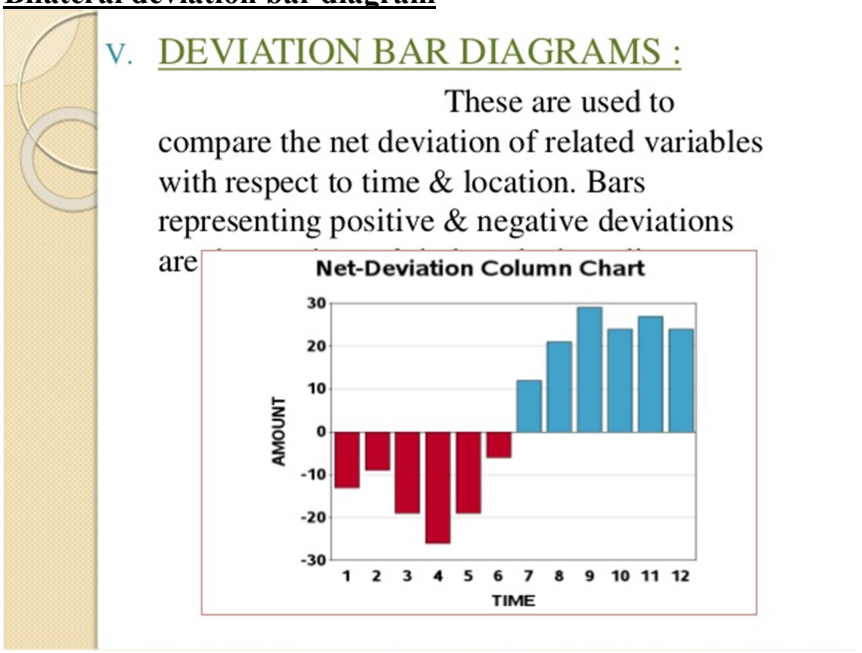
Simple bar diagram



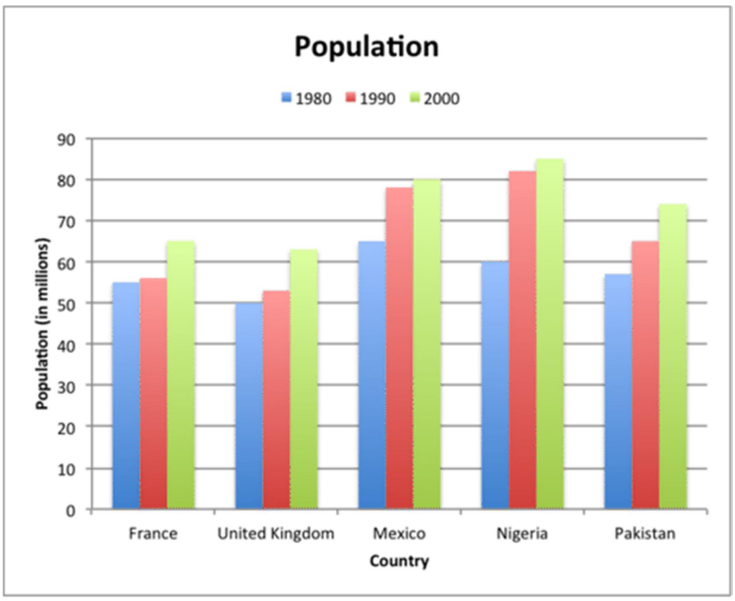
Percentage bar diagram



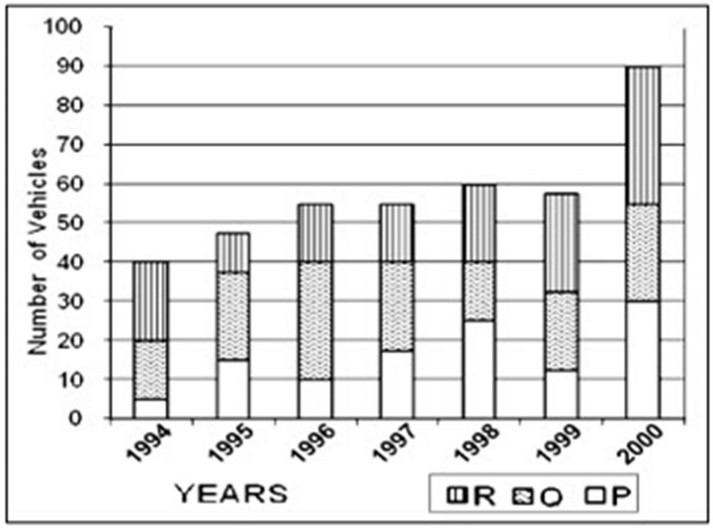
Bilateral deviation bar diagram



Multiple bar diagram



Sub divided bar Diagram



II
MEASURES OF CENTRAL TENDENCY

U
N
I
T

Average is a single value that represents group of values

Definition
An Average is a value which a typical or representative of a net 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 is compute the average.
- It should be suitable for further algebras mathematical treatment and capable of being used is further statistical computations

Uses of Average
- It is useful to describe the distribution in a concise manner
- It is useful to com pare different distributions
- It is useful to compare various statistical measures such as dispersion, skewness, kurtosis and so on

Functions or 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 Antiemetic 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
- I t 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	∑X =300

Formula = **X**
=∑X/N X=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	4x5=20
6	15	6x15=96
8	6	8x6=48
10	7	10x7=70
15	8	15x8=120
16	2	16x2=32
	∑f=43	∑fx=380

X = ∑fx /N

∑f= 380, N=43

= 380 / 43 = 8.837

The average wage of workers = Rs.8.837

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
		∑f=41	N=∑fm=1145

Arithmetic Mean = X = ∑fm / N

=1145/41

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
- ates 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 $(N + 1)/2$

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 Item (N + 1)/2
= Size of Item (9 + 1)/2
= 10 / 2 = 5^{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

Median = Size of Item (N + 1)/2
= size of Item (65 + 1)/2
= 33' which is nearer to 34
Cf of 34 = 75
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	20CF
40-60(L1-L2)	30F	50
60-80	8	58
80-100	2	60

Median = size of item (N/2) =60/2=30
L1=Lower limit =40, cumulative frequency =20 Frequency=30 Class interval=20

$$\text{=size of item } [L1 + \frac{N/2 - CF}{f} X\ C]$$

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

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 = 1/3 (Mean -mode)
Mode = 2 median -2mode
Median = 2/3 (Mean – Mode)

Individual Series

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

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 form the following data of the wages of workers of are 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					

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 size10 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**

Grouping Table

X	Frequency					
	1	2	3	4	5	6
0-5	1	3		8		
5-10	2		7		21	
10-15	5	19				29
15-20	14		24	**33**		
20-25	10	**19**			**21**	
25-30	9		**11**			
30-35	**2**					

Analysis Table

Column	Size of Item						
	0-5	5-10	10-15	15-20	20-25	25-30	30-35
1				I			
2			I	I	I		
3				I	I	I	
4				I	I	I	
5		I	I	I	I	I	I
6			I				
		1	3	6	5	3	1

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

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

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

$$= 15 + \frac{(9)}{28 - 5 - 10} \times 5$$

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

$$= 15 + 3.461$$

$$= 18.461$$

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 Geometirc 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

Uses of Geometric mean

- This average is often used to construct index numbers, where we are chiefly concerned with relative changes over a period of time
- It is the only useful average that can be employed to indicate rate of have

Individual series

G.M = Anti log of $\sum \mathbf{logX} / \mathbf{N}$

Calculate Geometric Mean

50	72	54	82	93

Solution

X	Logt X
50	1.6990
72	1.8573
54	1.7324
82	1.9238
93	1.9685
	∑logX=9.181

G.M = Anti log of $\sum \mathbf{logX} / \mathbf{N}$ = 9.181/5

=**1.8362**

= Antilog of 1.8342 = 68.26

Discrete Series
Calculate
Geometric mean from the following data

Size of Item	**120**	**125**	**130**	**135**	**136**	**138**	**139**	**140**	**147**
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		**∑f log x = 68.3700**

G.M = Anti log of ∑ log X/N = Anti log of 68.3700/32

= Antilog of 2.1366

= 137

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**

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	**∑flog m=81.9092**

G.M = Anti log of ∑f log m/N

=81.9092/68

=1.204547

= Antilog of 1.204547

=16.02 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

H.M = N/∑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

Family	**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		**∑1/x = 0.34697**

H.M = N/∑1/x

= 10 /0.34697

= 28.82 H.M

= 28.82

Discrete Series

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

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=∑f =34		**∑f 1/x = 4.1319**

H.M = N/∑f1/x
=34/4.1319
=8.23

Continuous Series Compute Harmonic Mean

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
	∑f = 35			**∑f 1/m = 2.27366**

H.M = N/∑ F (1/m)
=35/2.27366
=15.393682

UNIT - IV
MEASURES OF DISPERSION,SKEWNESS & KURTOSIS:

Dispersion Meaning

Dispersion is the study of scatterness around an average

Definition

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

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.

$$\text{Quartile Deviation (Q.D)} = \frac{Q3 - Q1}{2}$$

$$\text{Co-efficient of Q.D} = \frac{Q3 - Q1}{Q3 + Q1}$$

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

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

$$\textbf{Co – efficient of Mean Deviation (M.D)} = \frac{MD}{X \text{ or } Z \text{ or } M}$$

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 byignoring

Co efficient of Variance

Standard deviation is an absolute measure of dispersion. The corresponding relative measure is known as the co=-efficient of variation. It is used to compare the variability of two or more series

$$\text{Co –efficient of Standard deviation} = \frac{\sigma}{\bar{X}}$$

$$\textit{Co-efficient of Variance (C.V)} = \frac{\sigma}{\bar{x}} \times 100$$

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.

Range

Range = L-S

$$\text{Co-efficient of range} = \frac{L-S}{L+S}$$

Solved Problems

Find the range and co-efficient of range for the heights of 8 students of a class 158,160,165,168,170,173,

Solution

Range = L – S

Given Series the largest value of the series=173

Smallest value of the series = 158

Range = 178 – 158 = 15

$$\text{Co-efficient of range} = \frac{L-S}{L+S} =$$

=173-158/173+158

=15/331

= 0.045

Quartile Deviation

$$\textbf{Quartile Deviation} = \frac{\textbf{Q3 - Q1}}{\textbf{2}}$$

Individual Series

Find out the value of quartile deviation and it's co-efficient from the following data

Roll No	1	2	3	4	5	6	7
Marks	20	28	40	30	50	60	52

Solution

Marks arranged in ascending order

20,28,30,40,50,52,60

Q1 = size of (N + 1)/4 the item

= size of (7 + 1)/4= 8/4

= size of 2nd item = 28

Q3 = size of 3(N+1)/4th item

=Size of 3(7+1)/4th item

= 24/4

=6th item

Size of the 6thitem = 52

$$Q.D = \frac{Q3 - Q1}{2} = \frac{52 - 28}{2}$$

$=24/2=12$

$$\text{Co efficient of } Q.D = \frac{Q3 - Q1}{Q3 + Q1} = \frac{52 - 28}{52 + 28}$$

$=24/80$

$=0.3$

Continuous series

From the following table. Compute the quartile deviation as well as it's co-efficient

Size	4-8	8-12	12-16	16-20	20-24	24-28	28-32	32-36	36-40
Frequency	6	10	18	30	15	12	10	6	2

Solution

Weekly wages (x)	**No of workers f**	**cf**
4-8	6	4
8-12	10	16
12-16	18	34
16-20	30	64
20-24	15	79
24-28	12	91
28-32	10	101
32-34	6	107
34-40	2	109
	N=∑f=109	

Q1 = N/4 109/4 = 27.25

Q1 is lies between the classes 12-16

$$Q1 = L = \frac{N/4 - cf}{F} \times C$$

$$12 + \frac{(25 - 16)}{18} \times 4$$

$$= 12 + \frac{11.25}{18} 4$$

= 124 + 5/ 18

= 121 + 2.5

= 14.5

Q1 = 14.5

Q3 = 3N /4 = 3(109)/4 = 81.75

Q3 lies between the class intervals 24 – 28

$$= 24 + \frac{5-79}{12} X\ 4$$

$$= 24 + 0.916$$

$$Q3 = 24.92$$

$$3\quad Q.D = \frac{Q3-Q1}{2} = \frac{24.92 - 14.5}{2} = \frac{10.42}{2} = 5.21$$

$$\frac{N}{4} - cf$$

$$Q_3 = L + \left(\frac{}{F}\right) x\ C$$

81.7