Kamaraj College

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

Aggregatre 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.

Kamaraj College

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 todrawingconclusion fromfacts 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

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 sectorneeds 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

Kamaraj College

Statistics is extensively used in the field of Insurance. Actuarial statistics is must of theinsurancecompany 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

Kamaraj College

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.

Kamaraj College

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**

Sampling

Meaning:

Population or universe is a complete set of all possible observation of the type which is to be investigated.

Types of Population

- Finite Population
- Infinite population
- Hypothetical population
- Existent population

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

Kamaraj College

- 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

Methods of Sampling

Probability or Random Sampling

The method is based on the theory of probability, Under this method, every item in the universe has a known chance of being chosen

Classification of Probability sampling

Unrestricted (Or) Simple Random Sampling

Lottery Method

Lottery method of sampling refers to the process of drawing a lot among the population or universe. Under this method, the required number of samples is selected from the total population by blind fold form the drum or urn.

Random Number Method

This method is an alternative to lottery method under which samples are drawn by using the table of random numbers

Restricted Random Sampling

As the size of population or universe is not restricted under unrestricted method of sampling, it consumes much expenses and time

These samples are usually obtained by using the following methods such as:

a. Stratified random Sampling
b. Systematical Sampling
c. Cluster and Area Sampling
d. Multi-stage Sampling
e. Sequential Sampling

a. Stratified random Sampling

Under this method random sampling, the population is divided into some groups or classes based on their homogeneity. Samples are drawn from each stratum at random It is a method used for increasing the precisions of sampling.

Kamaraj College

b. Systematical Sampling

Under this method the universe or population is arranged on the basis of some systems like alphabetical,. Numerical, geographical etc.,

c. Cluster and Area Sampling

A cluster may refer to anything, a school, a company, an industry or a society. Cluster sampling refers to the procedure of dividing the [population into groups called clusters andsamplesdrawn firm theses clusters. A cluster may consist ofeither the primary or secondary sample units.

d. Multi stage sampling

Multi stage sampling is a type of sample design in which some information is collected from the whole sample and additional information is also collected from sub-sample of the full sample

e. Sequential sampling

Under this sequentialsampling method, the size of the sample units is nor determined in advance, but fixed according to mathematical division rules based on the survey.

Non Probability or Non Random sampling

This method is not based on theirs of probability in which the researcher cannot assum that every elements has an equal chance of being chosen

I. Convenience sampling
II. Purposive or Judgment sampling
III. Quota sampling

I. Convenience sampling

The method is also called chunk method. A chunk refers to the fraction of the population to be investigated. This chunk is not selected by probability but by judgment or convenience.

II. Purposive orJudgment sampling

It is that method of sampling in which the samples are drawn on the basis of personal judgment of a person. Generally the researcher uses his judgment in the choice of the samples which he thinks most suitable for his study. While choosing the samples only the average items are selected and extreme items are omitted

Selection of the samples is adjusted in accordance with the object of the survey. This method is suitable only when small number of samples is required

III. Quota sampling

It is one of the commonly used methods of sampling in markets surveys and opinion polls. Though it is a non-random sampling it combines the technique of probability sampling and purposive selection. This method is convenient and economical.

Kamaraj College

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

Kamaraj College

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. Univaraite 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 isdefined as more than or less than the specified limit, is called as open class or open ended interval

CumulativeFrequencyDistribution

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

- Less than cumulative Frequency
- More than cumulative frequency

Steps to construct a frequency distribution

Step -1 In the first column write the value

Kamaraj College

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

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 presentationProvide 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

Kamaraj College

- 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

Kamaraj College

UNIT - II
MEASURES OF CENTRAL TENDENCY

Average – Meaning

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 besuitable for further algebras mathematical treatment and capable of being used is further statistical computations

Uses of Average

- It is useful todescribe the distribution in a concise manner
- It is useful to com pare different distributions
- It is useful to comparevarious 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

Kamaraj College

- 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 Itmay 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

Kamaraj College

Discrete Series

Calculate the arithmetic mean for the wages of workersin 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
	N=∑f=43	∑fx=380

X = ∑fx /N = 380, N=43
= 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=∑f=41	N=∑fm=1145

Arithmetic Mean = X = =∑fm / N
The Arithmetic mean = 27.92

Kamaraj College

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 variablewhich 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 samplingthan 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 $\frac{N+1}{2}$ th Item

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

Kamaraj College

Median= – Size of N + 1
---------- th Item
2

- = Size of 9 + 1
- ---------- the item
-
- 2
- = 10 / 2 = 5th 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 N + 1
---------- th Item
2

= size of 65 + 1
---------- th Item
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

Kamaraj College

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

Median = –ize of N
---------- th Item c size of 60/2th Item
2

N / 2 - cf

=size of L + ---------- X C
f

30 - 20

= 40 + ----------x 20 = 46.47

2

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 - 2 mode

Median = Mode * 2/3 (Mean – Mode)

Individual Series

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

Kamaraj College

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					

Kamaraj College

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 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**					

Kamaraj College

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 frequently

$$\text{Mode } (Z) = L + \frac{f1 - f0}{2f1 - f0 - f2} \times C$$

$$\text{Mode } (z) = 154 + \frac{14 - 5}{2(14) - 5 - 10} \times 5$$

$$= 15 + 9/13 \times 5 = 15 + 45/13$$

$$= 15 + 3.46$$

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 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

Kamaraj College

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

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

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

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

$$= 9.1710$$

$$\frac{}{} = 1.8342$$

$$\frac{}{5}$$

= 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**

Kamaraj College

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**

∑log X **68.375**

G.M = Anti ling of ---------- = Anti log of ----------

N 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**

Kamaraj College

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	**∑flog m=81.9092**

G.M = Anti logof ∑f logm

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 samedifficulty 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
H.M = -------

Kamaraj College

∑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**

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

= 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

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

Kamaraj College

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
	∑f = 35			**∑f 1/m = 2.27366**

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

Kamaraj College

UNIT - III
MEASURES OF DISPERSION AND VARIABILITY

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 theirvariability
- 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

Kamaraj College

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

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

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 tendencyand as such as flexible too.

Kamaraj College

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

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

Kamaraj College

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

$$\text{Co-efficient of range} = \frac{L-S}{L+S} = 0.045$$

Quartile Deviation

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

Individual Series

Find out the value of quartile deviation and its 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

$$\textbf{Q1} = \textbf{size of } \frac{\textbf{N + 1}}{4} \textbf{ the item}$$

$$= \textbf{size of } \frac{7 + 1}{4} \textbf{ th item} = \textbf{8/4 th Item}$$

$$= \textbf{size of } 2^{nd} \textbf{ item} = \textbf{28}$$

$$\textbf{Q3} = \textbf{size of } \frac{\textbf{3(N + 1)}}{4} \textbf{ th item}$$

3(7+1)

= ----------

4

= 3 x 8 / 4 = 24/4 = 6 th item

Size of the 6thitem = 52

Q3 – Q1 52 -28

Q.D = ---------- = ---------- =

2 2

Q3 - Q1 52-28

Co-efficient of Q.D = ---------- = -------- = 24/80 =0.3

Q3 + Q1 52+28

Discrete Series

From the following data calculate Quartile deviation and it's co-efficient

x	26	28	32	35	29	24
f	6	7	9	10	7	4

Solution

X	f	cf
24	6	6
26	6	12 Q1
28	7	19
29	7	26
32	9	35 Q3
35	10	45
	N=∑f=45	

N + 1

Q1 = size of --------- the item

4

45 + 1

= size of ---------- th item = 46 /4

4

= size of 11.5 th item = 26

Kamaraj College

3(N + 1)

Q3 = size of --------- th item

4

= 3(45 +1)

---------- = 34.5 th Item =32

4

Q3 - Q1 52-28

Co-efficient of Q.D = ---------- = -------- = 24/80 =0.3

Q3 + Q1 52+28

Continuous series

From the following table. Compute the quartile deviation as well as its 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 class 12-16

N/4 cf

Q1 = L = ---------- x C

F

Kamaraj College

27.25 - 16

$$= 12 + \frac{\text{------------------}}{18} \times 4$$

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

$$= 12 + 45/18 = 121 + 2.5 = 14.5$$

Q1 = 14.5

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

Q3 lies between the class interval 24 – 28

$$Q3 = L + \frac{3N/4 - cf}{f} \times C$$

$$= 24 + \frac{81.75-79}{12} \times 4$$

$$= 24 + 0.916$$

Q3 = 24.92

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

Mean Deviation
Individual Series

$$M.D = \frac{\sum I\, DI}{N}$$

$$\text{Co efficient of M.D} = \frac{M.D}{\text{Mean or Median or Mode}}$$

Calculate mean deviation a d its coefficient from Arthimetic mean for the following the data

100	150	200	250	300

Kamaraj College

Solution

X	**IDI = X-X**
100	100
150	50
200	0
250	50
300	100
∑x= 1000	**∑I DI**

Mean = ∑x / N = 1000/5 = 200

$$M.D = \frac{\sum I\ DI}{N} = 300/5 = 60$$

$$\text{Co-efficient of M.D} = \frac{\text{Mean Deviation}}{\text{Mean}} = \frac{60}{200} = 0.3$$

Discern Series

Find mean deviationfrom mediation and itsco-efficientfrom the following data

X	10	11	12	13	14
f	3	12	18	12	3

Solution

X	**f**	**cf**
10	3	3
11	12	15
12	18	33
13	12	45
14	3	48

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

$$= \text{Size of } \frac{48+1}{2} \text{ th item}$$

$$= \text{Size of } 24.5^{th} \text{ item} = 12$$

Kamaraj College

X	f	IDI (X –Median)= x-12	fIDI
10	3	2	6
11	12	1	12
12	18	0	0
13	12	1	12
14	3	2	6
			∑fI DI = 36

Mean Deviation = ∑fI DI / N = 36 / 48 = 0.75

M.D 0.75

Co – efficient of M.D = M.D / Median = 0.75 / 12 = 0.0625

Continuous Series

Find the co-efficient of mean deviation fromk mean for the following data

Age in years	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No of persons	20	25	32	40	42	35	10	8

Solution

Age in years	**m**	**No of persons f**	**D=m-A**	**fd**	**IDI=m-X**	**fIDI**
0-10	5	20	-30	-600	31.5	630.0
10-20	15	25	-20	-500	21.5	537.5
20-30	25	32	-10	-320	11.5	368.0
30-40	35	40	0	0	1.5	60.0
40-50	45	42	10	420	8.5	357.0
50-60	55	35	20	700	18.5	647.5
60-70	65	10	30	300	28.5	285.0
70-80	75	8	40	320	38.5	308.0
		N = 212		**∑fd=320**		**∑fI DI = 3193.0**

Kamaraj College

$$\bar{X} = A + \frac{\sum fd}{N} = 35 + \frac{320}{212} = 35n + 1.5 = 36.5$$

$\bar{X}$ = 36.5 Years

$$M.D = \frac{\sum fd}{N} = \frac{3193}{212} = 15.06$$

$$\text{Co-efficient of M.D} = \frac{M.D}{\text{Mean}} = \frac{15.06}{36.5} = 0.41\text{years}$$

Standard Deviation

Individual Series

Compute standard Deviation form the following data of the income of 10 employees of a firm

Monthly income	600	620	640	620	680	670	680	640	700	650

Solution

X	$X=X-\bar{X}$	X^2
600	-50	2500
620	-30	900
640	-10	100
620	-30	900
680	30	900
670	20	400
680	30	900
640	-10	100
700	50	2500
650	0	0
$\sum X = 6500$		$\sum X^2 = 9200$

$$\bar{X} = \frac{\sum X}{N} = \frac{6500}{10} = 650$$

$$\sigma = \sqrt{\frac{\sum X^2}{N}} = \sqrt{9200 / 10} = 30.3 \qquad \sigma = 30.3$$

Discrete Series

Calculate standard deviation from the following data

Marks (X)	10	20	30	40	50	60
No of Students(f)	8	12	20	10	7	3

Kamaraj College

X	f	fx	$x = X-\bar{X} = X-30.8$	X^2	Fx^2
10	8	80	-20.8	432.64	3461.12
20	12	240	-10.8	116.64	1399.68
30	20	600	-0.8	0.64	12.80
40	10	400	9.2	84.64	846.40
50	7	350	19.2	368.64	2580.48
60	3	180	29.2	852.64	2557.92
	N=60	**∑fX =1850**			**$\sum Fx^2$ =10858.40**

$$\bar{X} = \frac{\sum fX}{N} = \frac{1850}{60} = 30.5$$

$$\sigma = \sqrt{\frac{\sum fX^2}{N}} = \sqrt{10858.40 / 60} = 13.5 \quad \sigma = 13.5$$

Continuous series

Calculate standard deviation form the following data

Class	0-10	10-20	20-30	30-40	40-50
Frequency	5	8	15	16	6

Solution

Class	Mid point	frequency	$d = \frac{X-A}{C}$	d^2	fd	Fd^2
0-10	5	5	-2	4	-10	20
10-20	15	8	-1	1	-8	8
20-30	25	15	0	0	0	0
30-40	35	16	1	1	16	16
40-50	45	6	2	4	12	24
		N=50			**∑fd=10**	**$\sum fd^2$ =68**

Assumed mean A= 25

Class interval C = 10

Standard Deviation (σ) = $\sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2} \times C$

$$= \sqrt{68/50 - (10/50)^2} \times 10$$

$$= \sqrt{1.36 - (0.2)^2} \times 10$$

$$= 11.49$$

Kamaraj College

SKEWNESS

Introduction

The term 'Skewness' refers to lack of symmetry, that is, when a distributionis to symmetrical it is called a skewed distribution. It the curve us normal or the data distributed symmetrically or uniformly. Spread will be the same on both sides of the cent repoint and the means median and mode will all have the same value.

Definition

'Skewness or symmetry 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 -**Croxton and cowden**

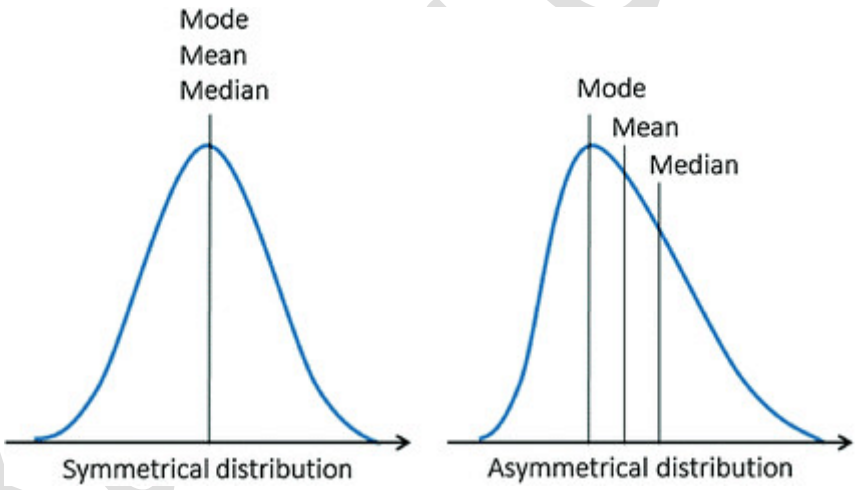
Skewness of a Distribution

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

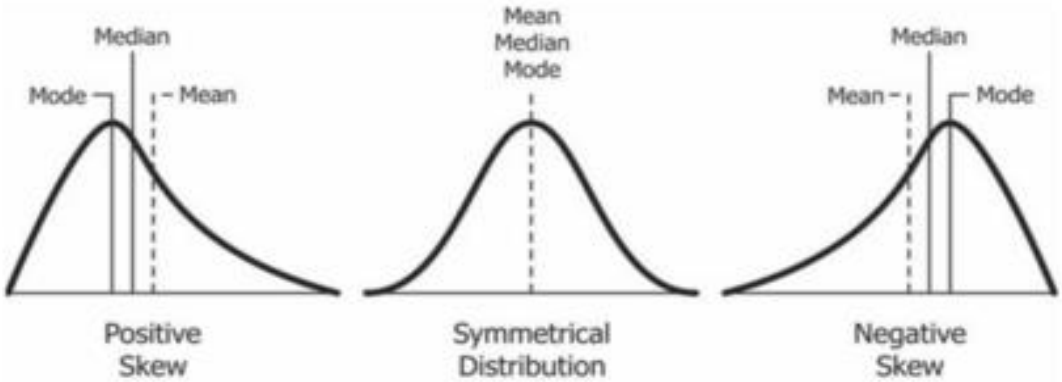
I. Determination of the sign of skewness and testing of skewness and
II. Determination of the extent of skewness

Symmetrical Distribution

In a symmetrical distribution, the values of mean, median and mode are coinciding. The spread of the frequencies is the same on both sides of the centre point of the curve.



Skewed Distribution



A distribution which is not symmetrical is called a skewed distribution it is called skewed distribution. It may be either positivelySkewed or negatively skewed Distribution

I) Positively Skewed Distribution

In a frequency distribution positively skewed distribution the curve has longer tail to the rights and its value of the mean is highest and the made is least. The median lies in between the two. That is $\bar{X} > M > Z$

II) Negatively Skewed Distribution

In a frequency distribution if the curve has long tail to the left then it is negatively skewed distribution in which value of mode is higher and mean is the least. The median lies in between the two. That is $\bar{X} < M < Z$

Various measures of Skewness

Skewness can be measured absolutely or relatively. Absolutely measures are called measures of skewness and relative measures are called the co-efficient of skewness

Absolute measures of skewness

i) The Karl Peason'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

Bowelys 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.

Karl Pearsons Coefficient of Skewness

Calculate Karl Pearson's coefficient of skewness for the following data

25	15	23	40	27	25	23	25	20

Solution

Size of the Item	**Deviation d=X-A**	**D^2**
25	-2	4
15	--12	144
23	-4	16
40	13	169

Kamaraj College

27	0	0
25	-2	4
23	-4	16
25	-2	4
20	-7	49
	$\sum d = -20$	**$\sum d^2 = 406$**

$$\bar{X} = A + \frac{\sum d}{N} = 27 + \frac{-29}{9} = 27 - 2.22 = 24.78$$

$$(\sigma) = \sqrt{\frac{\sum fd^2}{N} - \frac{(\sum fd)^2}{N}} = \sqrt{406 / 9 - (-20/9)^2}$$

$$= \sqrt{45.11 - (2.2)^2} = 6.3$$

In the given series, 25 is repeated three times

Mode is 25

$$SK_p = \frac{\bar{X} - \text{Mode}}{\sigma} = \frac{24.78 - 25}{6.3} = 0.03$$

Bowley's Co-efficient of Skewness

Calculate the coefficient of Skewness

Age	0-10	10-20	20-30	30-40	40-50
No of Persons	8	11	26	9	6

Solution

Age	**No of Persons (f)**	**cf**
0-10	8	8
10-20	11	19
20-30	26	45
30-40	9	54
40-50	6	60
	N = 60	

N/4 = 60/4 = 15 lies between the class interval 10 – 20

$$Q_1 = L + \frac{N/4 - cf}{F} \times C$$

$$= 10 + \frac{15 - 8}{11} \times 10 = 10 + \frac{7}{11} \times 10 = 10 + 6.36 = 16.36$$

Kamaraj College

Q_3 = 3N /4 = 3 (60)/4 = 45 lies between the CL 20 -30

$$Q_3 = L + \frac{3N/4 - cf}{F} \times C = 20 + \frac{45 - 19}{26} \times 10 = 20 + \frac{28}{26} \times 10$$

=20 + 10 = 30

N/2 = 60 /2 = 30 lies between the CL 20 - 30

$$\text{Median} = L + \frac{N/2 - cf}{F} \times C$$

$$= 20 + \frac{30 - 19}{26} \times 10 = 20 + \frac{11}{26} \times 10 = 20 + \frac{110}{26}$$

= 20 + 4.23 = 24.23

Bowley's Co – efficient of Skewness

$$SK_p = \frac{Q_3 + Q_1 - 2\text{Median}}{Q_3 - Q_1} = \frac{30 + 16.36 - 2(24.23)}{30 - 16.3613.64} = \frac{48.38 - 48.46}{13.64} = \frac{-2.1}{}$$

= - 0.15

Kelly's Co efficient of Skewness

From the data given below calculate Kelly's co-efficient of skewness

Median =130

P_{20} = 27

P_{90} =242

Solution

$$\mathbf{SK_p = \frac{P_{90} + P_{10} - 2\ Median}{P_{90} - P_{10}}}$$

$$= \frac{242 + 27 - (2 \times 130)}{242 - 27}$$

$$= \frac{269 - 260}{215}$$

$$= \frac{9}{215} = 0.042$$

Kamaraj College

UNIT - IV
CORRELATION

Correlation 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 estimates costs, value, price and other related variables
IV) Correlation is the basis of the concept of regression
V) Correlation analysis help in calculation the sampling once.

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 direction, 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

Kamaraj College

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 to 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 is 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 canofferwhether 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.

Concurrent deviation method

Kamaraj College

Under this method, the direction of change in X variable and y variable is taken into account to find out the deviation for each term the change in the value of the variable form its preceding value which may be + or –

Co-efficient of correlation

Find the Karl Pearson's Coefficient of Correlation

X	6	2	10	4	8
Y	9	11	5	8	7

Solution

X	**Y**	**X^2**	**Y^2**	**XY**
6	9	36	81	54
2	11	4	121	22
10	5	100	25	50
4	8	16	64	32
8	7	64	49	56
30	40	220	340	214

$$\textbf{Coefficient of Correlation} = \frac{N\sum XY - (\sum X)(\sum y)}{\sqrt{N\sum X^2 - (\sum x)^2}\,.\sqrt{N\sum Y^2 - (\sum Y)^2}}$$

$$r = \frac{(5 \times 214) - (30 \times 40)}{\sqrt{5 \times 220 - (30)^2}\ \sqrt{5 \times 340 - (40)^2}}$$

$$= \frac{1070 - 1200}{\sqrt{1100-900}\ \sqrt{1700-1600}} = \frac{-130}{\sqrt{200}\ \sqrt{100}} = \frac{-130}{14.14(10)} = -0.9194$$

Rank Correlation

Two judges ina beauty contest rank the 12 entries as follows

X	1	6	5	10	3	2	4	9	7	8
Y	6	4	9	8	1	2	3	10	5	7

Rank X	Rank Y	D=R(X)-(Y)	D^2
1	6	-5	25
6	4	2	4
5	9	-4	16
10	8	2	4
3	1	2	4
2	2	0	0
4	3	1	1
9	10	-1	1

Kamaraj College

7 8	5 7	2 1	4 1
N=10			$\sum D^2 = 60$

$$R = 1 - \frac{6\sum D^2}{N^3 - N}$$

$$= 1 - \frac{6(60)}{(10^3 - 10)} = 1 - \frac{360}{990} = 1 - 0.36 = 0.64$$

Calculate the co-efficient of concurrent deviation from the following data

X	65	50	45	60	40	55	30	40	60	60	55
Y	70	55	50	70	60	85	50	40	70	55	5

Solution

X	Direction of change (dx)	Y	Direction of change (dy)	dxdy
65		70		
50	-	55	-	+
45	-	50	-	+
60	+	70	+	+
40	-	60	-	+
55	+	85	+	+
30	-	50	-	+
40	+	40	-	-
60	+	70	+	+
60	0	55	-	0
55	-	55	0	0
				C = 7

$$C = \frac{+\sqrt{2C - N}}{N}$$

$$= \frac{\sqrt{2X\ 7 - 10}}{10}$$

$$= \sqrt{4} / 10 = 0.6333$$

Regression Analysis

Meaning

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

Definition

Kamaraj College

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 estimator 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, top find a relation between two or more variables that are related causally, is regression analysis. - Taro Famane

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 co-efficient and co-efficient of determinations

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

Algebraic Method

Regression equation is an algebraic method. It is an algebraicexpression of the regression line.

Regression Equations

A) Regression Equation of X on Y

X_c = a+ by

B) Regression Equation of Y on X

Y_c = a + bx

Solved Problems

Find regression lines by using actual Mean

X	3	5	6	8	9	11
Y	2	3	4	6	5	10

Kamaraj College

Solution

X	X = (X-$\bar{X}$)	X^2	Y	Y = (Y-$\bar{Y}$)	Y^2	xy
3	-4	16	2	-3	9	12
5	-2	4	3	-2	4	4
6	-1	1	4	-1	1	1
8	1	1	6	1	1	1
9	2	4	5	0	0	0
11	4	16	10	5	25	20
∑X = 42	∑x = 0	$\sum x^2$ = 42	∑Y = 30	∑y = 0	$\sum y^2$ =40	∑xy = 38

$$\bar{X} = \frac{\sum x}{N} = 42 / 6 = 7$$

Regression equation of X on Y

$$X - \bar{X} = r \frac{\sigma x}{\sigma y} (Y - \bar{Y})$$

$$r \frac{\sigma x}{\sigma y} = \sum xy / \sum y^2$$

X-7 =).95 (Y- 5)
X = 0.95Y – 4.75 + 7
X =).95y + 2.25
Regressionbe equation of Y on X

$$(Y - \bar{Y}) = r \frac{\sigma y}{\sigma x} (X - \bar{X})$$

Y – 5 0.904 (X-7)

Y = 0.904X – 6.328 + 5

Y = 0.904X – 1.328

Kamaraj College

UNIT - V
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

Characteristics of Index number

- They are specialized average
- They measure the net change in a group of related variables
- They measure the effect of changes over a period of time
- They help comparisonofgroups of variables directly

Uses of Index Number

- Index number is most widely used statistical devises
- Index numbers are used to measure the relative changes
- They are widely used in the evaluation of business and economic conditions
- It is useful for better comparison
- It is a good guide for the progress of every country
- It is useful for better comparison
- It is useful to know trends and techniques
- For forecasting future activities

Types of Index Numbers

- Price Index
- Quantity Index
- Value Index

Methods of Index Number

- Unweighted Index Number
- Simple Aggregative method
- Simple average of price Relative method

Weighted Index Number

1.Laapeyre's index number $P_{01} = \frac{\sum p_1 q_0}{\sum P_0 q_0} \times 100$

2.PAASCHEY'S Method

Kamaraj College

$$P_{01} = \frac{\sum p_1 q_1}{\sum P_0 q_0} \times 100$$

3.Bowley and Dorfish method

$$P_{01} = \frac{\frac{\sum p_1 q_0}{\sum P_0 q_0} + \frac{\sum p_1 q_0}{\sum P_0 q_0}}{2}$$

4.Fisher's Ideal method or Fishers Price Index Number

$P_{01} = \sqrt{L \times P}$

Consumer price Index Number (or) Cost of living index

Consumer price Index is designed to measure the change in the cost of living of workers because of change in the retail price. A change in the price level affects the cost of living of the people. People con some different types of commodities. So there is need to construct consumer's price index. Consumer price index can be used in different places for many purposes.

Uses of cost of living index

- It is useful in fixing the wages
- It is useful to know the purchasing power of money
- By using the cost of living index the Government determines the price and other variables
- It is useful the analysis of price situations

Limitations of Index numbers

- If the chosen base year is not a normal one, the purpose is list
- Every index number has its own purpose. No index number can serve all purpose
- These are only appropriate indications of the relative level.

Solved Problems

Construct an index number for 2014 taking2013 as base from the following data

Commodity	price in 2013(Rs)	Price in 2014 (Rs)
A	50	60
B	40	80
C	70	110
D	90	70
E	50	40

Solution

Commodity	price in 2013(Rs)	Price in 2014 (Rs)
A	50	60
B	40	80

Kamaraj College

C	70	110
D	90	70
E	50	40
	$\sum p_0 = 300$	**$\sum p_1 = 360$**

$\sum p_1 360$

Price Index P_{01} = ---------- X 100 = ---------- X 100 = 120

$\sum p_0$ 300

This means that we compared to 2013, in 2014 there is a net increase in the prices of commodities to the extent of 20 %

Compute Fisher's Ideal Index from the following data

Item	2000		2003	
	Price	Quantity	Price	Quantity
A	6	50	10	56
B	2	200	2	120
C	4	60	6	60
D	10	30	12	24
E	8	40	12	36

Solution

COMMODITY	P_0	Q_0	P_1	Q_1	p_1q_0	P_0q_0	p_1q_1	P_0q_1
A	6	50	10	56	500	300	560	336
B	2	100	2	120	200	200	240	240
C	4	60	6	60	360	240	360	240
D	10	30	12	24	360	300	288	240
E	8	40	12	36	480	320	432	288
					$\sum p_1q_0$=1900	$\sum P_0q_0$ =1360	$\sum p_1q_1$ =1880	$\sum P_0q_1$ =1344

P_{01} = √ L X P

$$P_{01} = \sqrt{\frac{\sum p_1q_0}{\sum P_0q_0} \times \frac{\sum p_1q_0}{\sum P_0q_0}}$$

= √ 1,3971 X 1.3988 X 100

= 1.39796 X 100 = 139.80

Analysis of time Series

An arrangement of statistical data in accordance with time of occurrence or in chronological order is called a time series.

Kamaraj College

Definition

A time series is a set of observation arranged in chronological order. MorrisHamberg requirement of a time seriesData 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 Passible equal.

Uses of Time series

i. It helps in understanding the past behaviors and in establishing the future behavior
ii. It helps in planning and forecasting the future operation
iii. It facilitates comparison between data of one period with those of another period
iv. It helps in evaluating current accomplishment
v. It is useful in forecasting the trade cycles

Time Series Models

Mathematical Models andMultiplicative method

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

1. 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

Irregular fluctuation

Multiplicative Model

According this model, the time series is expressed as

Y = Y X S X C X I

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.
- Components of time series
- Time series data containvariations of the following types
    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

Kamaraj College

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.

Causes

Changes of property and depression

Uses

I) Useful to study the character of business fluctuation
II) Useful to take timely decision in maintaining the business during different stages
III) Helps in facing recession and utilizing the booms

Un-Secular Variation

Irregular variation refers to such variation in business activity which do not repeat in a definite pattern. They are also called 'erratic 'accidental or random variations which are generally non-recurring and unpredictable

Causes

War, food, revolution, strike, lockouts and the like

Measurement of Secular of secular trend

a) Free hand Graphic Method
b) 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 I(X) and the value of the variable on the vertical axis (Y)

Merits

a. It is the simplest and easiest method
b. It can be applied to all types of trends
c. It is useful to understand the character of time series

Demerits

a. It is subject to personal bias
b. Its results depend upon the judgments of the person who draw the time
c. It does not help to measures trend

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 thesemi averages

Merits

I) It is simple and easier to understand
II) Everyone will get the same trend like
III) We can predict the future values based on the intermediate values

Demerits

I) It is affected by the limitations of arithmetic mean

Kamaraj College

II) It is not enough for forecasting the future trend

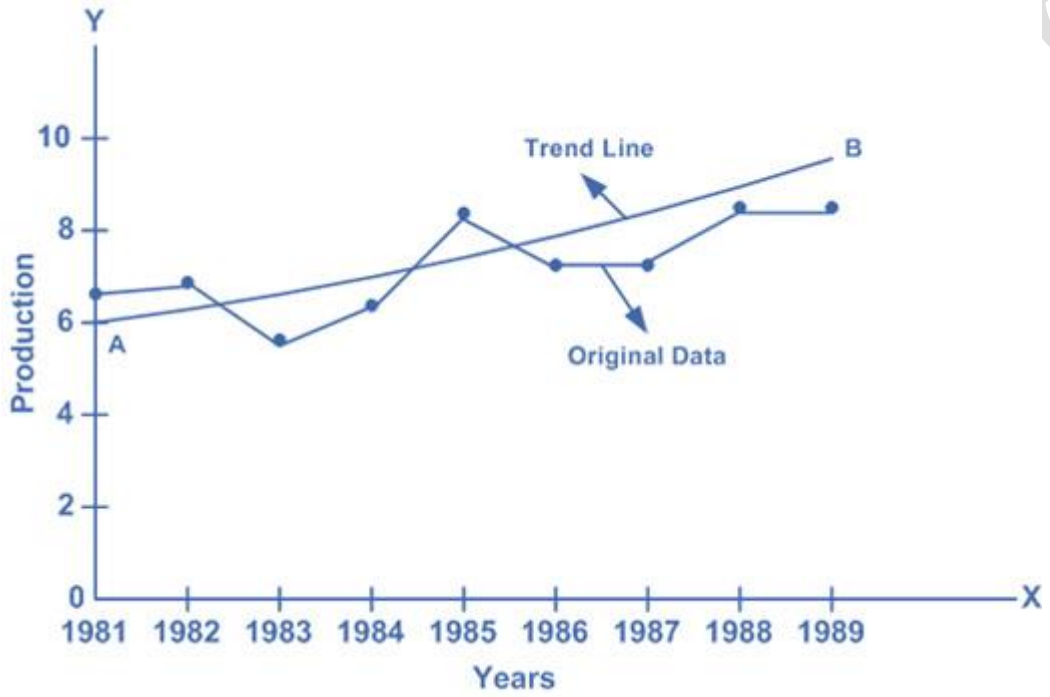
Moving average method

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 is the normal or trend value for the middle period.

Solved Problems

Free-Hand Trend line1985

Year	**1981**	**1982**	**1983**	**1984**	**1985**	**1986**	**1987**	**1988**	**1989**
Production tons	**0**	**2**	**40**	**6**	**8**	**10**	**6**	**8**	**8**



Moving Average

1.Find the 3 yearly moving average from the following time series data

Year	1998	1999	2000	2001	2002	2003	2004	2005
Sales(In tones)	30.1	45.4	39.3	41.4	42.2	46.4	46.6	49.2

Solution

Year	**Sales (in tons)**	**3yearly Moving Total**	**3 Yearly moving value**
1998	30.1	-------	-------

1999	45.4	114.8	38.27
2000	39.3	126.1	42.03
2001	41.4	122.9	40.97
2002	42.2	130.0	43.33
2003	46.4	135.2	45.07
2004	46.6	142.2	47.40
2005	49.2	-------	--------

2. Calculate the 5 yearly moving average from the following data

Year	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Number of students	705	685	703	687	705	689	715	685	725	730

Solution

Year	No of students	5 yearly moving total	Moving value
1998	**705**	--	--
1999	**685**	--	--
2000	**703**	**3485**	**697.0**
2001	**687**	**3469**	**693.8**
2002	**705**	**3499**	**699.8**
2003	**689**	**3481**	**696.2**
2004	**715**	**3519**	**703.8**
2005	**685**	**3544**	**708.8**
2006	**725**	---	--
2007	**730**	--	--

3. Calculate the Four – Yearly moving average for the following data

Year	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Production (in '000Tons)	464	515	518	467	502	540	557	570	586	612

Kamaraj College

Solution

Year	Production in '000 Tons	4yearly Moving	Combined Total	Moving Average
1998	464			
1999	515			
		1964		
2000	518		3966	495.75
		2002		
2001	467		4029	503.63
		2027		
2002	502		4093	511.63
		2066		
2003	540		4236	529.50
		2170		
2004	557		4424	553.00
		2254		
2005	570		4580	572.50
		2326		
2006	586			
2007	612			

4. Compute the trend from the following by the method of least square method

Year	2000	2001	2002	2003	2004
Production in Lakhs	830	920	710	900	1690

Solution
Computation of trend Values

Year	Production in Lakhs (Y)	Deviation from 2002(X)	XY	X^2
2000	830	-2	-1600	4
2001	920	-1	-920	1
2002	710	0	0	0
2003	900	1	900	1
2004	1690	2	3380	4
	∑y = 5050	**∑x = 0**	**∑xy=1700**	**$\sum x^2$ =10**

Since ∑x = 0

Kamaraj College

$$a = \frac{\sum y}{N} = 5050 / 5 = 1010$$

$$b = \frac{\sum xy}{\sum x^2}$$

$Y = a + bx$ ==1010 + 170x

$Y_c = 1010 + 170\,x$

When x= -2

$Y_{2000} = 1010 + 170\,(-2) = 1010 - 340 = 670$

When X = -1

$Y_{2001} = 1010 + 170\,(-1) = 1010 - 170 = 840$

When X = 0

$Y_{2002} = 1010 + 170\,(0) = 1010 - 0 = 1010$

When X =1

$Y_{2003} = 1010 + 170\,(-1) = 1010 + 170 = 1180$

When X=2

$Y_{2003} = 1010 + 170\,(2) = 1010 + 340 = 1350$

Kamaraj College