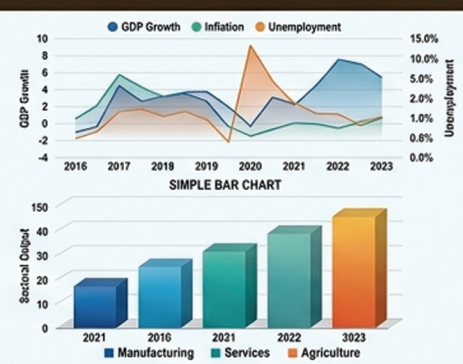
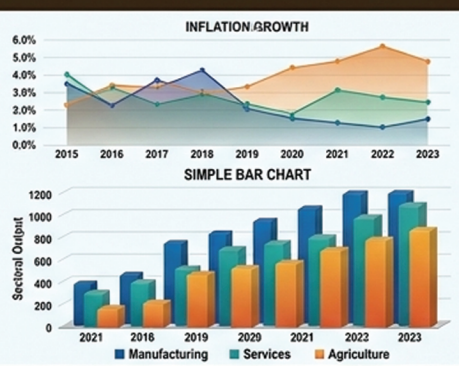
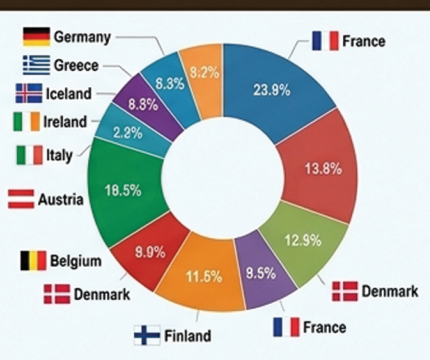
Statistics for Economics

B.A.- 2nd Yr.CBCS- III Semester









Nalgonda Economics Forum

NALGONDA

A teacher par excellence

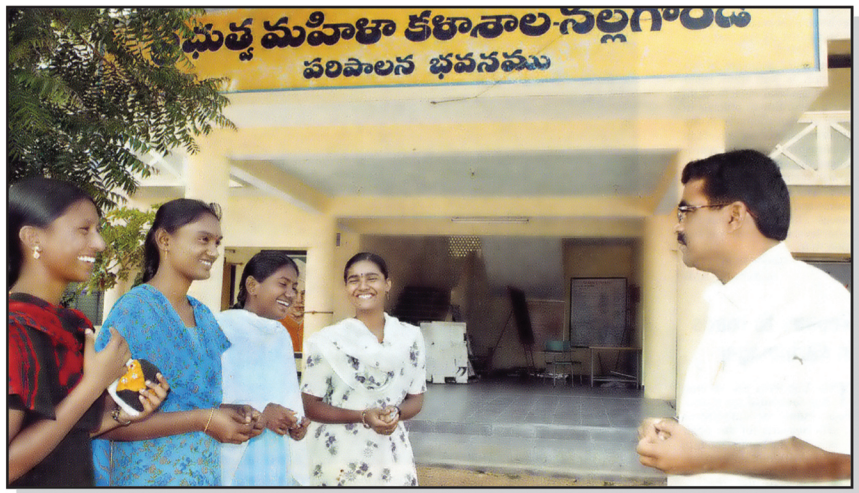
Thanks to the free coaching imparted by Akkenapalli Meenaiah, students from rural background can now hope to secure a seat in Osmania University

Getting a seat in any post graduation course in Osmania university campus is always a tough task for rural graduates. Those who are keen to get a seat there have to spend money for tailor-made coaching classes.

Realising the need to impart free coaching, Akkenapalli Meenaiah, lecturer in Economics, Government Women's Degree College, Nalgonda, offered special training for students of Economics in the district. Indeed the effort yielded good results and seven out of 25 students got seats in PG courses.

While Maroju Kavita, a bright student of the Women's Degree College could secured eighth rank, three of her friends — Koorella Himavathi, Chintapalli Sailaja and Cherukuri Seshu Kumari — got seats in the OU campus, Hyderabad. They are all eligible to get a hostel facility thus getting a wonderful opportunity to prepare for competitive exams. Five other students are all set to get seats in other institutions. "But for Mr. Meenaiah coaching, I would not fared well in the entrance test," Ka-vitha said. The convener of the Nalgonda Economics Forum, Mr. Meenaiah said that a small initiative taken by him yielded very good result. He opted out of spot valuation to train his students to prepare for the entrance test. "I might have lost about Rs.10,000 for spending my time with the students but the ranks they got gave me an immense satisfaction," he maintained. Asked for the secret of success of his mission, Mr. Meenaiah, a native of Gopulaipallypalli in Narketpally mandal, said: "Every teacher must strive hard to - create interest on the subject among the students. Motivation works wonders among the rural students." It is remains to be seen whether other teachers emulate this concerned lecturer or not.

- S. RAMU, in Nalgonda



ALL SMILES : A. Meenaiah, lecturer in economics, with his students, all rank holders in Osmania University M.A.Economics entrance test in Nalgonda. - PHOTO: NAGARA GOPAL

STATISTICS FOR ECONOMICS

B.A. 2 - Year (3rd Semester)

Revised Syllabus : 2025 - 26

: Chief Editor :

Dr. Akkenapally Meenaiah, M.A. M.Phil, Ph.D.
President : Nalgonda Economics Forum
Retired HOD Economics : N.G College Nalgonda (A)
Economy Columnist : Velugu, Sakshi, Namaste Telangana & Nava Telangana Daily News Papers.

Sub Editors :

T. Bhsker Reddy, M.A., SET
Asst Professor of Economics
GDC (W), Begumpet, Hyderabad

Dr.Dubba Ranjith M.A., Ph.D
Asst Professor of Economics (PT)
Nizam College (A), Hyderabad

: Authors :

Dr. K. Laxman Goud, M.A., Ph.D — **1st Module**
Associate Professor of Economics,
Govt City College (A) Hyderabad

Dr. S.P Krishnaveni, M.A., SLET, B.Ed. Ph.D. — **2nd Module**
Associate Professor of Economics
Govt City College (A) Hyderabad

Dr. Chanda Appa Rao, M.A., Ph.D. B.Ed, — **3rd Module**
Associate Professor of Economics
K.R.R. Govt. Arts & Science College (A), Kodad

T. Venkataiah, M.A, M.Phil, SET (Ph.D.) — **4th Module**
Asst Professor of Economics,
Govt. Degree College Kollapur

Shaik Hasrath Begum, (Ph.D), — **5th Module**
Asst Professor of Economics (PT)
Nagarjuna Govt College (A) Nalgonda

Acknowledgement Slip

Raja Rammohun Roy National Agency for ISBN (RRRNA),
Department of Higher Education,
Ministry of Education,
Government of India.

Email: isbn-mhrd[at]gov.in
Website: https://isbn.gov.in

Date [dd-mm-yyyy]: 08-05-2026

Allotted Date [dd-mm-yyyy] : 07-05-2026
Reference No.:(Application No.) : 48320|ISBN|2025|P
This is to acknowledge that the following ISBN has been officially issued by the **Indian ISBN Agency (RRRNA)** to:
Publisher Name : Nalgonda Economics Forum
Imprint :
Against the metadata provided by the publisher as per the standard format.
Allotted ISBN : 978-81-996558-6-7
Book Title : Statistics for Economics
Author(s) : Dr.K.Laxman Goud and Shaik Hasrath Begum, Dr.S.P. Krishnaveni, Dr.Chanda Appa Rao, T.Venkataiah
Edition :
Language : English
Format : Single-component retail product (Book)
Year of Publication : 2026

STATISTICS FOR ECONOMICS

B.A. 2 - Year (3rd Semester)

Revised Syllabus 2025 - 26

Published by

NALGONDA ECONOMICS FORUM

First Edition : **July, 2026** Copies : **1000**

ISBN Number : 978-81-996558-6-7

Cover Page by

Kathula Giri Babu

Publishers :

NALGONDA ECONOMICS FORUM PUBLICATIONS

(Regd. No. 297/13)

Nalgonda - 508001. Telangana

Copies Available at

Dr. Akkenapally Meenaiah, MA, M.Phil, Ph.D

President, Nalgonda Economics Forum,

Regd office: Plot No.33, Sri Raghavendra Colony, Nalgonda - 508001

Cell : 9490138118, 7013774141, 08682-295581

DTP & Printed at

SHREERAMA Graphics

Prakasham Bazar, Nalgonda - 508 001 (TS)

Phone : 94906 70002, 950 561 8383

Price **Rs. 175/-**

Foreword



The Nalgonda Economics Forum initiated its activity on the 12th of January 2007 and was subsequently registered on 10th May, 2013 with the registered number 297/13. This purely service oriented free coaching institution primarily aimed at training up students belonging to the rurally poor and socially and economically backward classes for admission through entrance into MA Economics course, but subsequently expanded to imparting training in a variety of areas of competitive examinations. When it came into existence some 18 years ago, almost every one of not more than ten students got selected to MA Economics course in different universities. And the Forum has now gradually, year after year, grown into fetching admission into PG Economics course to a bulky strength of as many as 800 candidates over these years. It is also to be noted that candidates trained at this Forum stood in the first rank in the PG Economics entrance examinations at Osmania University in 2009 and both at the Osmania and the Kakatiya universities subsequently in 2016, 2017 and 2018. The students of Nalgonda Economics Forum brought in the top three1st, 2nd and 3rd ranks followed by 7th, 8th, 9th and 10th ranks in CPGET -2020. In the year 2022 our students bagged state 1st, 7th and 10th ranks. We are happy to announce our forum student 10th ranker of CPGET-2022 Chanda Pranitha awarded Osmania University Gold Medal in Economics for the year 2023-24 The CPGET 2025 results released on 8th September, 2025, our forum got excellent results, first rank secured by S.Bhuvana Laxmi, Second Rank Secured by N.Ajay, Third Rank secured by A. Manasa and 6th Rank secured by Y. Lokesh. And our forum students also secured 12th, 13th, 14th, 17th & 20th Ranks.

The outstanding results achieved by our students in the UGC-NET Economics Examination (February 2026) and TGSET Economics Examination (March 2026) stand as a proud testimony to this sustained commitment. In UGC-NET Economics, five (5) candidates qualified for Assistant Professor eligibility and ten (10) candidates secured Ph.D. eligibility, while Sixteen (16) students qualified in TGSET Economics, Altogether, thirty-one (31) students qualified in the NET and SET examinations, reflecting the effectiveness of our structured academic programme, dedicated mentoring, and the sincere efforts of our students."These achievements are not merely examination results; they represent the collective perseverance of our students, the tireless dedication of our faculty, and the unwavering support of parents and well-wishers. They reaffirm our belief that quality guidance, disciplined preparation, and continuous assessment are the pillars of academic success. The remarkable accomplishments of our students inspire us to strive for even greater heights in the years to come.

As part of the extended areas of training programmes in different areas of competitive examinations, in addition to the basically mooted out MA Economics entrance training, the following are the achievements to our credit. Larger number of students trained here were qualified to SET and NET and **10 members of forum students were awarded Ph.D. in Economics.** In the year 2004, **41 candidates of our forum were selected as Govt Junior Lecturers, 10 candidates were selected as Degree Lecturers and 12 candidates wre selected as Junior Lecturers at Residential Degree & Junior Colleges** respectively in the state. A notable number of 100 above

candidates trained here are now working as Economics faculty in different private unaided junior and degree colleges.

The Forum so far brought out Six Text Books in English and Three Text Books in Telugu for B.A Economics with ISBN number for the undergraduate students in the state, in tune with the choice based credit system that has come into vogue with the academic year 2019-20. We already published Micro Economics 1st semester and Macro Economics Second Semister Text Book with ISBN number, as per new syllabus (2025-26) which is designed incorporating Artificial Intelligence concepts which are ruling present world Economy. Now we are brought Statistics for Economics 3rd Semister Text Books as per new syllabus along with ISBN Number.

It is also to be mentioned here that the Forum, with a view to be in tune with the changing examination patterns year after year, has been keeping on updating itself and catering to the newest needs of the examination goers. It is particularly in this process that, apart from imparting training directly in classroom situations, Forum created a website of its own to help the students face the computer-based tests, besides providing study material on-line, bringing to the doorsteps the latest relevant information that boosts up the students' morale and self-confidence.As has already been mentioned, Forum created a website of its own to help the students in ways more than one and started providing study material on-line. The Forum started YouTube channel on 17th April 2020 and as yet uploaded 500 video lessons on Economics subject. This Forum channel was monetized on 21st November 2020. As on 14th July 2026, the channel bagged 5877 subscribers.

Another innovative achievement of the Nalgonda Economics Forum is facilitating Google Meet & Zoom meeting platforms, especially at a situation of disrupted communication on account of the prevalence of Covid -19 pandemic havoc. Thus Forum successfully conducted under this roof CPGET-2020, CPGET-2021, CPGET-2022, CPGET-2023, CPGET-2024, CPGET-2025 and Junior Lecturers & Degree Lecturers coaching classes for Economics.

As mentioned in one of the columns above, there has been a sustained growth of the student potentiality since the inception of the Forum. And, besides the unhindered growth in the size of the student strength, there has been a proportionately parallel increase of the achievements of the goals the Forum has been relentlessly endeavouring to meet year after year. More specifically, the training activity at the Forum has been continuing over so many years in no way uninterrupted at any moment of time. That there has always been an affirmatively evinced confidence in the reliability of our commitment to the interests' and aspirations of the students as well as to a scientifically reasoned approach to the curriculum activity stands proof positive to the continued scalability and sustainability of the motto before the Forum and its organizers.

I sincerely congratulate all the successful candidates and express my heartfelt appreciation to the faculty members, staff, and every individual associated with M.V.N. Vignana Kendra and the Nalgonda Economics Forum for their invaluable contributions. I hope this publication will serve as a source of inspiration for future aspirants and further strengthen our mission of fostering academic excellence in Economics.

- Dr. Akkenapally Meenaiah, M.A, M.Phil, Ph.D.
President Nalgonda Economics Forum

B.A. Statistics for Economics : Semester - III

Revised Syllabus 2025-2026

Module-I : INTRODUCTION TO STATISTICS

Definition and scope of Statistics.Basic Concepts: Primary and Secondary Data, Cross-sectional and Time Series Data, Population, Census and Sample. Methods of Collecting Samples: Random and Non-Random Sampling Methods. Big Data Analysis: Use of SPSS, R & Role of AI in EconomicData Collection, Real-time Data analysis.

Module - II : MEASURES OF CENTRAL TENDENCY AND DISPERSION

Measures of Central Tendency: Mean and Types of Mean, Median, and Mode.Properties of Good Average – Comparison of Different Averages.Measures of Dispersion – Absolute and Relative Measures of Dispersion: Range, Quartile Deviation, Mean Deviation, Standard Deviation, Coefficient of Variation and Variance.Use of Excel & Python in Statistical Applications.

Module - III : CORRELATION AND REGRESSION

Correlation: Meaning and Types – Karl Pearson's Correlation Co-efficient – Spearman's Rank Correlation – Regression: Meaning of Simple & Multiple Regression.

Module - IV : INDEX NUMBERS

Meaning, Uses and Difficulties in the Construction of Index Numbers.Types of Index Numbers. Methods of Index Numbers:Laspeyres', Paasche's and Fisher's. Concept of CPI(Consumer Price Index), WPI(Wholesale Price Index), IIP(Index of Industrial Production).

Module - V : ANALYSIS OF TIME SERIES

Meaning,Uses and Components of Time Series Analysis: Secular, Seasonal, Cyclical and Irregular Variations. Methods of Measurement of Secular Trends: Graphic, Semi-Averages, and Moving Averages. Impact of Covid-19 on the Growth of Service sector in India.

Reference Books :

Allen, R. G. D. (2015). Mathematical analysis for economists (Reprint ed.). Andesite Press. (Original work published 1938).

Bhardwaj, R. S. (2006). Mathematics for economics and business (1st ed.). Excel Books.

Chiang, A. C., & Wainwright, K. (2005).Fundamental methods of mathematical economics (4th ed.). McGraw-Hill.

Gupta, S.P. (2025) Statistical Methods (4th ed.), Sultan Chand & Sons

INDEX

Module-1 : Introduction to Statistics **01 - 25**

1. Introduction to Statistics 01
1.2. Definitions of Statistics 02
1.3. Scope of Statistics 03
1.4. Basic Concepts of Statistics 05
1.5. Data - Meaning and Types of Data 07
Primary Data - Methods of Collection 07
Secondary Data - Methods of Collection 08
Cross - Sectional Data 10
Time Series Data 11
1.6. Population Method 14
1.7. Census Method 15
1.8. Sample Method 16
1.9. Sample Method - Random 17
1.10. Sample Method - Non Random 17
1.11. Big Data Analysis 18
1.12. Use of SPSS, R 21
1.13. Role of AI in Economic Data Collections 22
1.14. Real Time Data Analysis 24
1.15 Model Examination Questions 25

Module-2 : Measures of Central Tendency and Dispersion **26-61**

2.1. Measures of Central Tendency 26
2.2. Types of Mean 26
2.3. Arithmetic Mean - Calculation 28
2.4. Arithmetic Mean - Merits and Demerits 31
2.5. Median - Calculations 32
2.6. Median -Merits and Demerits 36
2.7. Mode - Calculations 37
2.8 Mode - Merits and Demerits 41
2.9. Properties of Good Average 42
2.10. Comparison of Different Averages 43
2.11. Measures of Dispersion 44
2.12 Absolute and Relative Measures of Dispersion 45
2.13. Range - Calculation 45
2.14. Quartile Deviation - Calculation 48
2.15. Mean Deviation - Calculation 53
2.16. Standard Deviation - Calculation 56
2.17. Variance. and Coefficient of Variation 59
2.18. Use of Excel & Python in Statistical Applications 60
2.19. Model Examination Questions 61

Module - 3 : Correlation and Regression **63 - 77**

3.1. Meaning of Correlation 63
3.2. Types of Correlation 64
3.3. Importance of Correlation 65
3.4. Methods of Calculating Correlation 66
3.5. Karl Pearson's Correlation Co-efficient 67
3.6. Spearman's Rank Correlation 69
3.7. The Concurrent Deviation Method 70
3.8. The Least Squares Method 72
3.9. Meaning of Regression 73
3.10. Uses of Regression Analysis 74
3.11. Simple Regression 75
3.12. Multiple Regressions 76
3.13 Model Examination Questions 77

Module - 4 : INDEX NUMBERS **79 - 91**

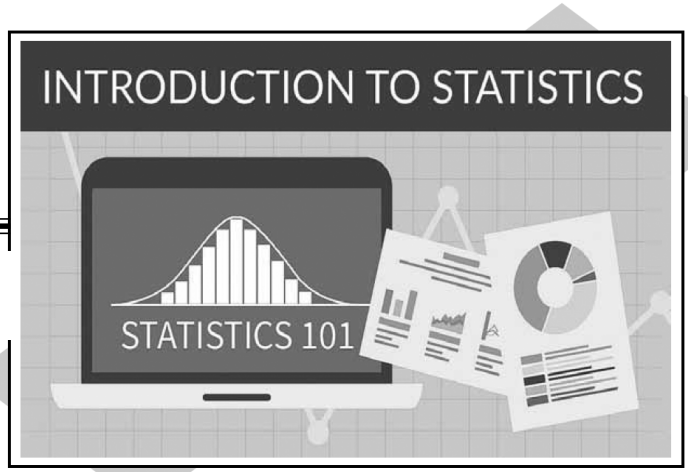
4.1. Meaning of Index numbers 79
4.2. Features of Index Numbers. 79
4.3. Construction of Price Index Numbers 80
4.4. Difficulties in Constructing of Price Index 82
4.5. Uses of Index Numbers 83
4.6. Types of Index Numbers 84
4.7. Laspayer's Price Index Method 85
4.8. Paasche's Price Index Method 86
4.9. Fisher Ideal index number 87
4.10. Concept of Consumer Price Index (CPI) 88
4.11. Wholesale Price Index (WPI) 89
4.12. Index of Industrial Production (IIP) 90
4.13 Model Examination Questions 91

Module-5: ANALYSIS OF TIME SERIES **69 - 92**

5.1. Meaning of Time Series Analysis 93
5.2. Uses of Time Series Analysis 93
5.3. Components of Time Series Analysis 94
5.4. Secular Trend or Long Run 94
5.5. Seasonal Variations 95
5.6. Cyclical Variations 96
5.7. Irregular Variations or Random Variations 97
5.8. Measurement of Secular Trend 98
5.9. Graphical method or Free Hand Drawing Method 98
5.10. Semi Averages Method 99
5.11. Moving Averages Method 100
5.12. Impact of Covid-19 on the Growth of Service sector in India ... 101
5.13 Model Examination Questions 103

Module

1



Introduction to Statistics

1.	Introduction to Statistics	01
1.2.	Definitions of Statistics	02
1.3.	Scope of Statistics	03
1.4.	Basic Concepts of Statistics	05
1.5.	Data - Meaning and Types of Data	07
	1.5.1. Primary Data - Methods of Collection	07
	1.5.2. Secondary Data - Methods of Collection	08
	1.5.3. Cross - Sectional Data	10
	1.5.4. Time Series Data	11
1.6.	Population Method	14
1.7.	Census Method	15
1.8.	Sample Method	16
1.9.	Sample Method - Random	17
1.10.	Sample Method - Non Random	17
1.11.	Big Data Analysis	18
1.12.	Use of SPSS, R	21
1.13.	Role of AI in Economic Data Collections	22
1.14.	Real Time Data Analysis	24
1.15	Model Examination Questions	25

1.1. Introduction

The word 'Statistics' is derived from the Latin word 'status' or the Italian word 'statista' or German 'statistic'. All these words mean a political state. In the olden days statistics was necessary for the proper functioning of the affairs of state. Thus, in those days Statistics was called as 'science of state' or 'science of king' as it was mainly used by the state or kings. Today statistics is defined as a field of study relating to the collection analysis, interpretation and presentation of data.

"In the general sense, Statistics means numerical facts and figures relating to a particular phenomenon. In the traditional period the scope of statistics was very much limited to the collection of facts and figures pertaining to the age-wise and sex-wise distribution of population wealth etc. But now-a-days we can say that Statistics constitutes an integral part of very scientific and economic inquiry: Social and economic studies without Statistics are useless. Statistics thus play a vital role and as Tippet has rightly remarked, 'It affects everybody and touches life at many points".

1.2. Definitions of Statistics

Statistics May be defined as the techniques by which data are collected, organized, summarized, presented and analyzed for the purpose of making reasonably valid decisions and especially under conditions of uncertainty. Some definitions of statistics is as follows; "Statistics can be defined as the collection, presentation, interpretation of numerical data" - Croxton and Cowden.

"Statistics are numerical statements of facts in any department of enquiry placed interrelation to each other." - Bowley.

"Statistics are measurement, enumerations or estimates of natural or social phenomena, systematically arrangement to exhibit their inner relation." - Conner.

"By Statistics we mean quantitative data affected to a marked extend by a multiplicity of causes."- Yule and Kendall.

"The science of Statistics is essentially a branch of applied mathematics and can be regarded as a mathematics applied to observation data." - R.A Fisher.

Tuttle defined Statistics as "measurement, enumerations or estimates of natural or social phenomena usually systematically arranged analyzed and presented as to exhibit important relationships among them'.

In summary statistics is the science of collecting, organizing, summarizing and analyzing information in order to draw conclusion. Now-a-Days Statisticians use statistics both in singular and plural sense. In the singular sense the term Statistics is associated with "A Body of methods for making decisions when there is uncertainty arising from incompleteness or the instability of the information available for making such decisions". In its plural sense Statistics refer to numerical Statement of facts such as per capita income, population etc.

1.3. Scope of Statistics

Statistics is the science of collecting, classifying, presenting, analyzing, and interpreting numerical data. The scope of statistics refers to the wide range of fields where statistical methods are applied for decision-making and analysis.

1. In Economics: Statistics is essential in economics for:

1. Measuring national income 2. Studying demand and supply
3. Preparing price index numbers 4. Economic planning and policy formulation

2. In Business and Commerce

1. Market research and consumer analysis 2. Sales forecasting
3. Quality control 4. Financial decision-making

3. In Government and Administration

1. Population census and surveys 2. Policy formulation and evaluation
3. Budget preparation 4. Welfare program assessment

4. In Medicine and Public Health

1. Clinical trials and testing medicines 2. Study of diseases and epidemics
3. Health statistics and life expectancy

5. In Agriculture

1. Crop production analysis 2. Rainfall and weather forecasting
3. Soil and irrigation studies

6. In Education

1. Student performance evaluation 2. Research and academic analysis
3. Examination systems

7. In Social Sciences

1. Study of population, poverty, and unemployment 2.Social surveys and opinion polls

8. In Scientific Research

1. Formulation and testing of hypotheses 2. Experimental design and data analysis

9. In Banking and Finance

1. Risk and credit analysis 2. Stock market studies
3. Insurance calculations

10. In Daily Life :

1. Weather forecasting, 2. Sports statistics,
3. Election predictions

1.4. Basic Concepts of Statistics

1. **Populations and samples:** A population is any specific collection of objects of interest. A sample is any subset or sub collection of the population, including the case that the sample consists of the whole population, in which case it is termed a census.
2. **Measurements and Sample Data:** Measurements and Sample Data's measurement is a number or attribute computed for each member of a population or of a sample. The measurements of sample elements are collectively called the sample data.
3. **Parameters:** A parameter is a number that summarizes some aspect of the population as a whole. A statistic is a number computed from the sample data.
4. **Statistics:** Statistics is a collection of methods for collecting, displaying, analyzing, and drawing conclusion from data.
5. **Descriptive statistics:** Descriptive statistics is the branch of statistics that involves organizing, displaying, and describing data.
6. **Inferential statistics:** Inferential statistics is the branch of statistics that involves drawing conclusions about a population based on information contained in a sample taken from that population. The measurement made on each element of a sample need not be numerical. In the case of automobiles, what is noted about each car could be its color, its make its body type and so on. Such data are categorical or qualitative, as opposed to numerical or quantitative data such as value or age. This is a general distinction.
7. **Qualitative data:** Qualitative data are measurements for which there is no natural numerical scale, but which consist of attributes, labels, or other non-numerical characteristics.
8. **Quantitative data:** Quantitative data are numerical measurements that arise from a natural numerical scale.
9. **Frequency Distribution:** A table in which the frequencies and the associated values of a variable are written side by side is known as a frequency distribution. According to Croxton and Cowden, *"Frequency distribution is a statistical table which shows the set of all distinct values of the variable arranged in order of magnitude, either individually or in a group with their corresponding frequencies side by side."* A frequency distribution can be discrete or continuous depending upon whether the variable is discrete or continuous.
10. **Individual series:** In an individual series, the names of the individuals are written against their corresponding values. When individual units are arranged in an order (ascending or descending) is called individual series. The individual series may be arranged in following two orders.

1). **Ascending order:** when data are arranged in ascending order i.e. a small value to a big value it is known as arranging them in ascending order.

2). **Descending order:** when data are arranged serially starting from a big value to a small value.

11. **Discrete series:** On the other hand discrete gives data for each specific value of variable. The important thing in the discrete series is that the variable involved in the series has integer values. Data are presented in a manner that exact measurement of items is clearly shown.

12. **Continuous series:** Frequency distribution in continuous series give a better picture if the number of items is large enough continuous series is a statistical series in which all the class intervals along with their corresponding frequency are listed out in the form of a table. All the class intervals may not have the same frequency. When data can take any value without any gap the series becomes a continuous series. In the continuous series, classes are formed in two ways exclusive series and inclusive series.

1) **Exclusive Series:** In such a series, for a particular class interval all the data items having values ranging from its lower limit to just below the upper limit are counted in the class interval. In other words, we do not include the items that have values less than the lower limit, equal to the upper limit and greater than the upper limit. This way of forming classes is known as exclusive series .In this series upper limit of one class always appears as the lower limit for next class.

2) **Inclusive Series:** On the contrary to exclusive series, an inclusive series includes both its upper and lower limit. Of course this means that we do not include the items with values less than the lower limit and greater than the upper limit.

13. **Cumulative Frequency:** Statistical series can be either simple or Cumulative. Frequency (f) tells how many are there in each class i.e. within the lower and upper limit. In order to know how many items are there as less than or more than a particular number, we need to construct a cumulative frequency table. In cumulative series the frequencies are progressively totaled and aggregates are shown. They can be shown by less than method or more than method.

1) **Less than cumulative frequency Distribution:** To prepare a less than' frequency table the frequencies are added up from the lowest class to the higher class. In this table cumulative frequencies are arranged as less than the upper limit of each class.

2) **More than cumulative frequency:** In case of more than method we start with lower limit of the classes. As against less than frequency table more than frequency table is useful to know how many students have more than particular marks.

1.5. Data - Meaning and Types of Data

A sequence of observations, made on a set of objects included in the sample drawn from population, is known as statistical data. The first step in any enquiry (investigation) is the collection of data. The data may be collected for the whole population or for a sample only. It is mostly collected on a sample basis. Collecting data is very difficult job. The enumerator or investigator is the well trained individual who collects the statistical data.

Definition: "Data are facts, figures and other relevant materials past and present serving as bases for study and analysis".

Meaning of Data: The search for answers to research questions calls collection of Data. Data are facts, figures and other relevant materials, past and present, serving as bases for study and analysis

Importance of Data: The data serve as the bases or raw materials for analysis without data no specific inferences can be drawn on our study. The reliability of data determines the quality of research. Data form the basis of testing hypothesis data provides the facts and figures for constructing measuring scale. The scientific process of research can be carried out only through accurate data. Data may be classified as: (1) Primary data and (2) Secondary data.

1.5.1. Primary Data- Methods of Collection :

Primary data are original facts collected for the first time for the fulfillment of the objectives. They are called as internal source of data as the data are collected directly from the respondents fields of the study. They are obtained from living persons directly related to the problem and objectives of the research study. Primary sources are original sources from which the researcher directly collects data that have not been previously collected. Primary Data are first hand information collected through various methods such as observation, interview etc.

Methods of Collecting Primary Data: The data which are originally collected for the first time for the purpose of the survey are called primary data.

I. Observation Method: Observation means viewing or seeing. We observe things while we are awake. Observation is classical method of scientific enquiry Definition Observation may be defined as "a systematic viewing of a specific phenomenon in its proper setting Observation as a method includes both seeing and hearing"

2. Direct Personal Interview: Under this method the investigator collects the data personally from the respondent. The person who collects the information is called the investigator and the person who gives the responses/answers the questions asked by the investigator are called a respondent. The data collected in this manner are therefore most reliable.

3. Through correspondents: Under this method correspondents or agents are appointed by the investigator to obtain data from various places. These correspondents are required to collect and pass the transmit information to the investigator or the central office. This method is widely used by newspaper offices.

4. Questionnaire: Collection of data through questionnaires is the most popular method for collecting primary data. A questionnaire is 'a List of questions pertaining to the enquiry. Under this method a questionnaire is sent to various informants with a request to answer the questions and return the completed questionnaires to the investigative.

5. Through schedules: Under this method the field workers are asked to go to the respondent with questions contained in the schedule. They collect the answers in their own hand writing and provide data to the investigator. This method is useful when the respondent is illiterate. This method of data collection is similar to that of the questionnaire. The schedule is also a proforma containing a set of questions. The 'difference be1'veen the questionnaire and the schedule is that the schedule is being filled in by the enumerators who are specially appointed for the purpose. These enumerators go to respondents with the schedules and ask them the questions from the schedule in the order they are listed. The enumerator records the replies in the space meant for the same in the schedule itself. This method is usually adopted in investigations conducted by governmental agencies or by some big organisations.

1.5.2. Secondary Data- Methods of Collection :

Secondary or external source of data are related to that information which are obtained from outside as either public source or someone else, who have already worked or encountered on the subject. This type of data can help to save time money and energy. But a researcher should verify these data with the help of other sources.. These are sources containing data which have been collected and compiled for another purpose. The secondary sources consist of readily available and compiled statistical statements and reports. Secondary sources consist of not only published but also unpublished records. They consist

of Data over which a researcher has no original control.

Secondary data are opposite to primary data. They are collected and published already (by some organization for instance). They can be used as a source of data and used by surveyors to collect data from and conduct the analysis. Secondary data are impure in the sense that they have undergone statistical treatment at least once.

Methods of Secondary data collection

1. Official publications such as the Ministry of Finance, Statistical Departments of the government, Federal Bureaus Agricultural Statistical boards etc. Semi-official sources include State Bank, Boards of Economic Enquiry etc.
2. Data published by Chambers of Commerce and trade associations and boards.
3. Articles in the newspaper, from journals and technical publications. This type of source can be obtained in the form of published and unpublished documents.

   When we use the data, which have already been collected by others, the data are called secondary data. This data is said to be primary for the agency which collects it first and it becomes secondary for all the other users.

Sources of secondary data: Secondary data may exist in the form of published or unpublished form. In its published form secondary data may be obtained from

1. Published reports of newspapers RBI and periodicals.
2. Publication from trade associations.
3. Financial data reported in annual reports.
4. Data available in SEBI publication.
5. Information from official publications.
6. Publication of international bodies such as UNO World Bank etc.
7. Others in its unpublished form secondary data may exist as (1) Internal reports of the government departments (2) Records maintained by the institutions (3) Research reports prepared by students in the universities.
8. Census and Sample Method: In any statistical investigation one is interested in studying the population characteristics. This can be done either by studying the entire items in the population or on a part drawn from it. If we are studying each and every element of the population the process is called census method and if we are studying only a sample the process is called sample survey sample method or sampling. For example the Indian population census or a socio economic survey of a whole village by a college planning forum is examples of census studies. The national sample survey enquiries are examples of sample studies.

1.5.3. Cross - Sectional Data :

Statistics deals with different types of data used for analysis and decision-making. One of the most important types of data in economics is cross-sectional data. It helps in understanding variations among different individuals, groups, or regions at a particular point in time. It is widely used in economic surveys, censuses, and research studies. In simple words, it is a snapshot of many units at one specific time.

1. Features / Characteristics :

1. **Single Time Period:** Data is collected only once.
2. **Large Number of Units:** Covers individuals, firms, households, etc.
3. **Comparative Nature:** Enables comparison between units.
4. **No Time Dimension:** Does not show changes over time.
5. **Heterogeneity:** Data varies across different units.

Examples :

1. Income of households in a district in 2025
2. Marks obtained by students in a particular examination
3. Prices of goods in different markets on the same day
4. Population of different states in a given year

2. Uses / Importance :

1. **Economic Analysis:** Helps study income distribution, poverty, and inequality
2. **Policy Formulation:** Governments use it for planning welfare programs
3. **Market Research:** Firms analyze consumer preferences and demand
4. **Social Studies:** Used to study education, health, and employment patterns
5. **Comparison:** Useful in comparing different groups or regions

3. Advantages :

1. **Simple and Easy to Collect**
2. **Less Costly and Time-Saving**
3. **Useful for Comparative Analysis**
4. **Provides Quick Results**

4. Disadvantages :

1. **No Time Analysis:** Cannot study trends or changes over time

2. **Limited Explanation:** Does not show cause-and-effect relationships clearly
3. **Static Nature:** Only gives a snapshot, not a dynamic picture
4. **Possibility of Bias:** Data may be affected by sampling errors

Cross-sectional data plays a crucial role in statistical and economic analysis. It provides a clear picture of different units at a specific point in time and is highly useful for comparison and policy-making. However, it has limitations in analyzing trends and dynamic changes. Therefore, it is often used along with time series data for better analysis.

1.5.4. Time Series Data

"Time series is a sequence of statistical data arranged in chronological order." Time Series Data refers to a set of observations collected **over a period of time**, usually at regular intervals such as daily, monthly, quarterly, or yearly. In statistics and economics, time series data helps us understand how a variable changes over time.

Examples: Daily stock prices :

1) Monthly sales of a company
2) Annual GDP of a country
3) Yearly rainfall data

Features of Time Series Data

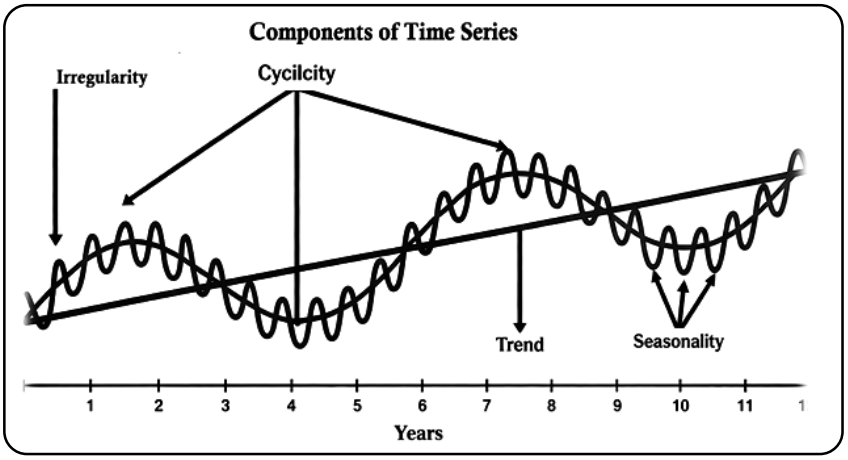
1) **Chronological Order** – Data is arranged according to time

2) **Equal Intervals** – Observations are taken at uniform intervals

3) **Dynamic Nature** – Values change over time

4) **Dependence** – Present values may depend on past values

Components of Time Series: A time series is generally composed of four components:



1. Trend (T)

1) Long-term movement in data
2) Shows general direction (upward or downward)

   Example: Increase in population over years

2. Seasonal Variation (S)

1) Regular fluctuations within a year
2) Occurs due to seasons, festivals, climate

   Example: Ice cream sales increase in summer

3. Cyclical Variation (C)

1) Long-term oscillations (more than one year)
2) Associated with business cycles

   Example: Boom and recession phases

4. Irregular Variation (I)

1) Unpredictable changes
2) Due to sudden events (wars, pandemics)

   Example: COVID-19 impact on economy

Types of Time Series

1) **Annual Data** – Data collected yearly

2) **Quarterly Data** – Data collected every 3 months

3) **Monthly Data** – Data collected every month

4) **Daily/Hourly Data** – High-frequency data

Methods of Measuring Trend: Trend is the most important component of time series. It can be measured by:

1) **Graphical Method (Freehand Curve Method)**

2) **Semi-Average Method**

3) **Moving Average Method**

4) **Least Squares Method** (Most scientific method)

Uses of Time Series Analysis: Time series data is very useful in economics, business, and planning:

1) **Forecasting** future values (sales, demand, prices)
2) **Policy making** by government
3) **Business planning** and budgeting
4) **Identifying patterns and trends**
5) **Economic analysis** (GDP, inflation, unemployment)

Importance of Time Series in Economics

1) Helps in understanding **economic growth**
2) Useful in studying **business cycles**
3) Aids in **demand and supply analysis**
4) Supports **planning and development programs**

Limitations of Time Series :

- Requires large and reliable data
- Past trends may not continue in future
- Affected by sudden disturbances
- Complex methods may require technical knowledge

Time Series Data is an essential tool in statistics and economics. It helps in analyzing past behavior, understanding present conditions, and predicting future trends. By studying its components and patterns, economists and planners can make informed decisions and develop effective policies.

1.6. Population Method

Under the census or complete enumeration method, the statistician collects the data for each and every unit of the population or universe. This universe is a complete set of items which are of interest in any situation. The population census is another example of a census investigation. Usually, this method is recommended in cases where the area of investigation is limited and requires intensive examination of the population. The census method can be applied in a situation where the separate data for every unit in the population is to be collected such that the separate action for each is taken. For example, the preparation of the voters list for election purposes income tax assessment, recruitment of personnel etc. are some of the areas where the census method is adopted. This method can be used where the population is comprised of heterogeneous items, i.e. different characteristics. Through the census method provides a complete data of the population under study, it is very costly and time-consuming. Often, this method is dropped down because of these constraints and the **sampling method,** where certain items representative of the larger group, is selected to draw the conclusions.

Merits of Population Method :

1. **Intensive Study** - Under census investigation you must obtain data from each and every unit of the population. Further, it enables the statistician to study more than one aspect of all items of the population.
2. **Reliable Data** - The data that a statistician collects through a census investigation is more reliable, representative and accurate. This is because; in a census the statistician observes every item personally.
3. **Suitable Choice** - It is a great choice in situations where the different items of the population are not homogeneous.
4. **The basis of various surveys** - Data from a census investigation is used as a basis in various surveys.

Demerits of Population Method :

1. **Costly:** Since the statistician closely observes each and every item of the population before collecting the data, it makes a census investigation a very costly method of investigation.
2. **Time-consuming** - A census investigation is time-consuming and also requires manpower to collect original data.
3. **Possibilities of Errors** - There are many possibilities of errors in the census investigation method due to non-response, measurement, lack of preciseness of the definition of statistical units or even the personal bias of the investigators.

1.7. Census Method

The Census Method, also known as the Complete Enumeration Method, is a statistical technique in which data is collected from each and every unit of the population. Unlike sampling, no part of the population is left out. This method provides comprehensive and accurate information about the entire population. The Census Method refers to a method of data collection in which every item or individual in the population is investigated. For example:

The Population Census of India collects data from every household in the country.

Features of Census Method :

1) **Complete Coverage** – All units of the population are included.
2) **High Accuracy** – Results are more reliable as no sampling error occurs.
3) **Detailed Information** – Provides in-depth and comprehensive data.
4) **Time-consuming** – Requires a large amount of time.
5) **Costly Method** – Needs more resources and manpower.

1. Steps in Census Method :

1) Defining the Population
2) Planning Data Collection
3) Preparation of Questionnaires/Schedules
4) Training Enumerators
5) Collection of Data from Every Unit
6) Classification and Tabulation
7) Analysis and Interpretation

2. Merits (Advantages) :

1) **Maximum Accuracy** – Since all units are studied, results are highly precise.
2) **Reliable Data** – No sampling bias or sampling error.
3) **Detailed Study** – Useful for policy-making and planning.
4) **Suitable for Small Populations** – Best when population size is manageable.

3. Demerits (Disadvantages) :

1. **Very Costly** – Requires huge financial resources.
2. **Time-Consuming** – Takes a long time to complete.

3. **Difficult to Manage** – Needs large administrative setup.

4. **Not Suitable for Large Populations (frequent studies)** – Difficult to repeat often The Census Method is a comprehensive and reliable statistical technique that provides accurate and detailed information about a population. However, due to its high cost and time requirements, it is not always practical. Therefore, while it is ideal for national-level surveys and critical decision-making, sampling methods are often preferred for routine studies.

1.8. Sample Method

The sampling method is the one in which only some of the representative items of the population are selected and data are collected from these. Instead of collecting information for and from all the units of population, we select a sample i.e.only a few items of the population Conclusions derived from the small sample are generalized for the whole population. Sampling is a statistical procedure that is concerned with the selection of the individual observation; it helps us to make statistical inferences about the population.

In sampling, we assume that samples are drawn from the population and sample means and population means are equal. A population can be defined as a whole that includes all items and characteristics of the research taken into study. However, gathering all this information is time consuming and costly. We therefore make inferences about the population with the help of samples.

Advantages of Sample Method :

1. **Reduce Cost:** It is cheaper to collect data from a part of the whole population and is economically advantageous.
2. **Greater Speed:** Sampling gives more time to researcher for data collection, so it is quickly and has a lot of time for collection of information
3. **Detailed Information:** Investigator during studying a small universe provides a detail and comprehensive information's.
4. **Practical Method.** Sampling is the only practical method when the population is infinite.
5. **Much Easier:** it is much easier to collect information from many individuals in a universe.

Disadvantages of Sampling Method :

1. Careful sampling selection is difficult.
2. Experts are required for careful study of the universe.
3. If the information's is required for each and every unit in the study, then it is difficult to interview each and every person in sampling method.
4. In the absence of a thorough knowledge, sampling methods the result option may be incorrect or misleading.
5. A complicated sampling may require more labour than a complete coverage.
6. A pure representation is impossible in most cases.

1.9. Random Sample Methods

1. Simple random sampling: Here each unit in the population gets an equal chance of being represented in a sample, and such a sample is called 'random sample' and the technique of selecting a random sample is called simple random sampling. This is the common method of sampling. The sample so obtained is called a random sample. By a random sample we mean, we should give equal chance or opportunity to every unit in the population to be included in the sample. A random sample can be selected by two methods **(i) lottery method (ii) random number method.**

i. **Lottery Method:** The lottery method is practicable when the population size is comparatively small.

ii. **Random Number Method:** This method is adopted with the aid of random number tables. These are table of numbers in which digits selected by a mechanical process of randomization are tabulated. This method can be used to select a random sample even when the population is large.

2. **Stratified Random Sampling:** When the population is heterogeneous with respect to some characteristic, but can be divided into a number of homogeneous subgroups known as 'strata' this method can be applied. Here, we divide the population into different strata and from each stratum select the items proportionally using lottery method.

3. **Systematic Sampling:** If we can arrange the sampling units is a definite order, say, alphabetical, chronological, geographical etc., this method can be employed.

4. **Cluster Sampling:** Cluster sampling is the selection of sample units in two stages. In the first stage, certain groups or clusters called primary sampling units are selected from the population. These units, say, might be companies selected from an industry. In the second stage, individual items called elementary sampling units are drawn from each of these clusters. These units might be employees-either a sample or all of the employees in a company. This second step in the sampling process is called sub-sampling. When each cluster is contained in a compact geographic area, cluster sampling is also called area sampling.

1.10. Non Random Sampling Methods

1. Judgment Sampling: Under this method, the choice of sample items depends exclusively on the judgment of the investigator. On the basis of his own choice, he tries to select the best representative of the whole population. It is also known as Purposive and Deliberate Sampling.

2. Quota Sampling: Under this method, the items of the population are subdivided into various groups and then a quota (number of items to be selected from each sub-group) is fixed. But, within the given quota, selection of sample units depends upon the personal judgment of the investigator. So, this is a kind of Judgment Sampling only.

3. Convenience Sampling: Under this method, while selecting the sample units, the investigator gives special attention to his convenience. For example: To estimate the average height of an Indian, the investigator (belonging to Delhi) can take a convenience sample from the Delhi State only and estimate the average height of an Indian. This method of selecting the sample is also called 'Chunk'.

1.11. Big data Analysis

Big data analytics is the process of analyzing large and complex datasets to uncover meaningful patterns, trends and insights that support data driven decision making. By combining traditional statistical techniques with modern computing tools it enables organizations to extract value from rapidly growing and diverse data sources. It analyzes massive volumes of structured, semi-structured and unstructured data generated from data sources. Advanced tools and frameworks like Hadoop, Spark and NoSQL databases support distributed storage and large scale data processing. Big data analytics applies methods such as descriptive, predictive and prescriptive analysis to understand past behavior and forecast future outcomes. By integrating machine learning and advanced analytics big data analytics helps uncover complex relationships and generate scalable, actionable intelligence.

1. How Big Data Analytics Works :

Big data analytics follows a systematic process that transforms raw, complex data into actionable insights for decision-making.

1) **Data Collection:** Gather data from sources such as cloud platforms, IoT devices, social media, mobile apps and transactional systems. Data is stored in data lakes or warehouses.
2) **Data Storage and Management:** Organize, index and manage large datasets to ensure easy accessibility and efficient querying.
3) **Data Processing:** Transform and structure data using batch or stream processing techniques (e.g., Hadoop, Spark, Flink) to prepare it for analysis.

4) **Data Cleaning:** Refine data by removing duplicates, handling missing or inconsistent values and ensuring overall quality and reliability.
5) **Data Analysis:** Apply techniques like data mining, predictive analytics, machine learning, deep learning and NLP to identify patterns, trends and relationships.
6) **Visualization and Insights:** Present results through interactive dashboards, charts and reports to support informed, data-driven decision-making.
7) **Decision and Action:** Implement insights in business processes, generating new data and completing the analytics feedback loop.

2. Types of Big Data Analytics :

1) Descriptive Analytics: Summarizes historical data to understand what has happened by using metrics, dashboards, and reports. It helps identify trends, patterns, and performance in past events such as sales or website traffic

2) Diagnostic Analytics: Analyzes data to uncover reasons behind past outcomes and events. Useful for identifying root causes, like reasons for customer churn or high patient readmissions.

3) Predictive Analytics: Uses historical data and statistical/machine learning models to forecast future events. Applied in scenarios like predicting customer behavior, stock prices or weather trends

4) Prescriptive Analytics: Provides actionable recommendations to optimize future outcomes based on insights. Used in pricing strategies, marketing campaigns and resource allocation decisions.

5) Real-time Analytics: Processes and analyzes data instantly as it is generated for immediate insights. Enables quick decision-making in areas like fraud detection, stock trading and live traffic management.

6) Spatial Analytics: Examines location-based data to optimize operations and logistics. Applied in urban planning, route optimization and analyzing customer movement patterns.

7) Text Analytics: Extracts insights from unstructured text data using natural language processing (NLP). Used for sentiment analysis, trend detection and improving customer experience from reviews or social media.

3. Applications: Big Data Analytics has a significant impact in various sectors:

1) **Healthcare:** It aids in precise diagnoses and disease prediction, elevating patient care.
2) **Retail:** Amazon's use of Big Data Analytics offers personalized product recommendations based on your shopping history, creating a more tailored and enjoyable shopping experience.

3) **Finance:** Credit card companies such as Visa rely on Big Data Analytics to swiftly identify and prevent fraudulent transactions, ensuring the safety of your financial assets.
4) **Transportation:** Companies like Uber use Big Data Analytics to optimize drivers' routes and predict demand, reducing wait times and improving overall transportation experiences.
5) **Agriculture:** Farmers make informed decisions, boosting crop yields while conserving resources.
6) **Manufacturing:** Companies like General Electric (GE) use Big Data Analytics to predict machinery maintenance needs, reducing downtime and enhancing operational efficiency.

4. Benefits of Big Data Analytics :

1) **Real-time Intelligence:** Analyze data as it is generated to make quick decisions, respond to market changes and seize opportunities instantly.
2) **Better-informed Decisions:** Discover hidden patterns, trends and correlations to guide strategic planning and operational improvements.
3) **Cost Savings:** Identify inefficiencies, optimize processes and allocate resources effectively to reduce waste and expenses.
4) **Enhanced Customer Engagement:** Understand customer behavior and preferences to personalize experiences and improve satisfaction.
5) **Optimized Risk Management:** Detect, assess and mitigate risks proactively using predictive analytics and real-time monitoring.

5. Limitations :

1) **Data Overload:** Massive volumes of data from sources like social media make it challenging to extract meaningful insights efficiently.
2) **Data Quality Issues:** Inaccurate, incomplete or inconsistent data can lead to flawed analytics and poor decision-making.
3) **Privacy Concerns:** Handling personal and sensitive data requires balancing personalized insights with compliance and ethical considerations.
4) **Security Risks:** Large datasets are attractive targets for cyberattacks, demanding strong security measures and protocols.
5) **High Costs:** Implementing, maintaining and scaling big data analytics infrastructure can be expensive.

1.12. Use of SPSS and R

SPSS (Statistical Package for the Social Sciences) and R are two widely used tools for statistical analysis and data manipulation. Each has its strengths and limitations, making them suitable for different use cases depending on user expertise, project requirements, and budget.

1. Key Features of SPSS

SPSS is a commercial software package developed by IBM, known for its **user-friendly graphical interface**. It is particularly popular among social scientists and researchers who prefer minimal coding. SPSS excels in:

1) **Ease of Use: Its GUI (Graphical User Interface):** allows users to perform complex statistical analyses with minimal programming knowledge.
2) **Data Management**: SPSS provides robust tools for sorting, merging, and aggregating data, making it efficient for handling structured datasets.
3) **Decision Trees**: SPSS offers a user-friendly interface for implementing decision tree algorithms, making it accessible for non-programmers.
4) **Visualization**: While SPSS supports basic visualizations, its customization options are limited compared to R.

2. However, SPSS has notable drawbacks :

1) It is **expensive**, requiring a license for full functionality.
2) It struggles with **large datasets** and lacks advanced visualization capabilities.
3) Its **closed-source nature** limits flexibility and community-driven enhancements.

3. Key Features of R :

R is an **open-source programming language** designed for statistical computing and data visualization. It is widely used by data scientists, statisticians, and researchers for its flexibility and extensive package ecosystem. R stands out in:

1) **Cost**: R is free and open-source, making it accessible to individuals and organizations with budget constraints.
2) **Customization**: R offers unparalleled flexibility with packages like *ggplot2* and *shiny* for creating highly customizable and interactive visualizations.
3) **Community Support**: R has a strong community that regularly contributes new libraries and provides extensive documentation.

4) **Exploratory Data Analysis (EDA)**: R is ideal for EDA due to its interactive features and command-line interface.

4. However, R has its limitations:

1) It has a **steep learning curve**, requiring programming knowledge.
2) R loads all data into memory, which can be a bottleneck for **large datasets**.
3) It lacks built-in security features, making it less suitable for sensitive data.

5. Choosing Between SPSS and R:

1) **For Beginners**: SPSS is better suited for users with limited programming experience due to its intuitive GUI.
2) **For Advanced Users**: R is ideal for those who need flexibility, advanced visualizations, and the ability to handle diverse analytical tasks.
3) **Budget Considerations**: R is the preferred choice for cost-conscious users, while SPSS is more suitable for organizations that can afford its licensing fees.
4) **Visualization Needs**: R is superior for creating detailed and interactive visualizations, whereas SPSS is limited to basic charts.

SPSS is a great tool for quick and straightforward statistical analysis, especially for non-programmers. R, on the other hand, is a powerful and versatile tool for those willing to invest time in learning its programming capabilities. Both tools have their place in the analytics ecosystem, and the choice depends on the specific requirements of the project and the user's expertise.

1.13. Role of AI in Economic Data Collections

Artificial Intelligence (AI) refers to the use of machines and algorithms to perform tasks that normally require human intelligence. In economics, AI plays a crucial role in collecting, processing, and analyzing large volumes of data efficiently.

1. Sources of Economic Data using AI:

AI collects economic data from various sources:

1) Social media (consumer sentiment)
2) Online transactions (e-commerce data)
3) Satellite images (agriculture, urban growth)
4) Government databases
5) Sensors and IoT (Internet of Things) devices

2. Automation of Data Collection: AI automates data collection through web scraping, APIs, and machine learning tools. It reduces human effort and collects real-time data.

3. Big Data Handling: AI can process large datasets quickly and accurately, which is difficult using traditional statistical methods.

4. Real-Time Data Collection: AI enables continuous data collection, helping economists track inflation, employment, and market trends in real time.

5. Data Accuracy and Quality Improvement: AI reduces human errors and improves data accuracy through validation techniques and anomaly detection.

6. Cost and Time Efficiency: AI reduces the cost and time required for large-scale surveys and data collection processes.

7. Predictive Data Collection: AI predicts future trends and identifies what type of data needs to be collected for policy-making.

8. Use in Government and Policy: Governments use AI for:

1) Census data analysis 2) Tax data monitoring
3) Welfare scheme targeting

9. Challenges:

1) Data privacy issues 2) High initial cost
3) Lack of skilled workforce 4) Bias in algorithms

AI has revolutionized economic data collection by making it faster, more accurate, and cost-effective. However, ethical and technical challenges must be addressed for its effective use.

1.14. Real Time Data analysis

Real-time data analysis refers to the immediate processing and analysis of data as it is generated. It helps organizations and governments make quick decisions using live data. It involves data collection, processing, analysis, and action. It is widely used in stock markets, banking, e-commerce, and economic monitoring. Though it offers speed and accuracy, it requires high infrastructure and advanced technology.

1. Key Features :

1) Instant processing (no delay)
2) Continuous data flow (streaming data)
3) Quick decision-making
4) Automated systems
5) High-speed Computing

2. How It Works (Process) :

1. **Data Collection (Ingestion):** Sensors, apps, transactions generate data (e.g., stock market prices)
2. **Data Processing:** Data is cleaned and processed instantly
3. **Data Analysis:** Algorithms analyze patterns in real-time
4. **Output / Action:** Dashboards, alerts, or automated actions

3. Economic Examples :

1) Stock market price tracking
2) Online payment fraud detection
3) Inflation monitoring using digital price data
4) E-commerce demand tracking

4. Tools & Technologies :

1) Apache Kafka (data streaming)
2) Apache Spark (real-time processing)
3) Power BI / Tableau (dashboards)
4) Python / R (analysis)

5. Advantages :

1) Faster decision-making
2) Better accuracy with live data
3) Useful for forecasting
4) Improves efficiency

6. Limitations :

1) High cost of infrastructure
2) Requires advanced technology
3) Data privacy issues
   Complexity in handling big data

1.15 Model Examination Questions

Short Type Questions :

1. Population method
2. Sampling method
3. Frequency and cumulative frequency
4. Individual, discrete and continuous series
5. Primary data
6. Secondary data
7. Use of SPSS & R
8. Real Time Data Analysis
9. Random sampling methods
10. Non-random sampling methods

Essay Type Questions :

1. Explain the definitions and scope of statistics
2. Explain the basic concepts of statistics
3. Define population method and explain the merits and demerits of population method
4. Define sampling method and explain the merits and demerits of sampling method
5. Describe various types of sampling
6. Define data and explain the types of data
7. Explain the methods of collecting primary data
8. Explain the methods of collecting secondary data
9. Explain types of big data and how it works
10. Explain role of AI in Economic data collections

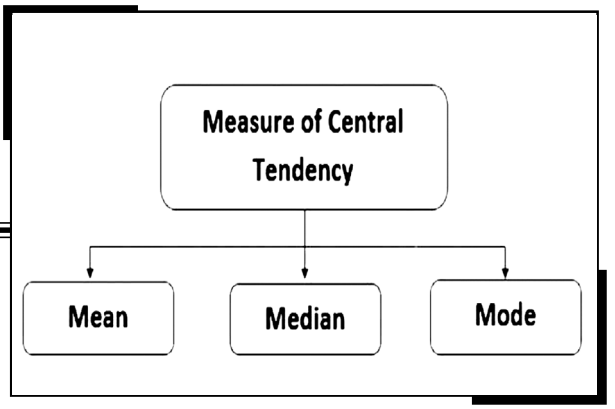
Author

Dr. K. Laxman Goud, M.A., Ph.D.

Associate Professor of Economics, Govt City College (A) Hyderabad.

Module

2



Measures of Central Tendency and Dispersion

2.1.	Measures of Central Tendency	26
2.2.	Types of Mean	26
2.3.	Arithmetic Mean - Calculation	28
2.4.	Arithmetic Mean - Merits and Demerits	31
2.5.	Median - Calculations	32
2.6.	Median -Merits and Demerits	36
2.7.	Mode - Calculations	37
2.8	Mode - Merits and Demerits	41
2.9.	Properties of Good Average	42
2.10.	Comparison of Different Averages	43
2.11.	Measures of Dispersion	44
2.12	Absolute and Relative Measures of Dispersion	45
2.13.	Range - Calculation	45
2.14.	Quartile Deviation - Calculation	48
2.15.	Mean Deviation - Calculation	53
2.16.	Standard Deviation - Calculation	56
2.17.	Variance. and Coefficient of Variation	59
2.18.	Use of Excel & Python in Statistical Applications	60
2.19.	Model Examination Questions	61

2.1. Measures of Central Tendency

Quantitative data in a population under study exhibit certain general characteristics. They show a tendency to concentrate at certain values, usually somewhere in the centre of the distribution. Measures of this tendency are called measures of central tendency or averages. The tendency of the observations to cluster near the central part of the distribution is called the central tendency. It can be measured statistically. This is nothing but an average which is a most representative value of the given data. An average of a data set is the value of the variable which is representative of the entire distribution.

The purpose of computing an average value for the set of observations is to obtain a single value which is representative of all the items and which the mind can grasp easily and quickly. Averages occupy a prominent place in statistical analysis. Most of the statistical techniques are based on averages. Averages are more widely used than any other measures because of their many applications.

2.2. Types of Mean

Classification of Averages or measures of central tendency: The various measures of central tendency or averages may be grouped in to two types.

I. **Mathematical Averages**

II. **Positional Averages**

I. Mathematical Averages: Mathematical average refers to all such average where a figure is taken out through mathematical methods from the given series that represents the whole series. Mathematical averages are three types.1.**Arithmetic Mean,**2. **Geometric Mean**, 3. **Harmonic Mean**

l. Arithmetic Mean: Arithmetic mean is most widely used measure for representing the entire data. To a layman it is average but for a statistician it is called arithmetic mean. Arithmetic mean is the number which is obtained by adding the values of all the items of a series and dividing the total by the number of items. Thus the arithmetic mean of a group of observations is defined as:

Mean= Sum of observations/number of observations

$\overline{X}$ is the symbol of the arithmetic mean $\bar{x} = \frac{\sum X}{N}$

2. Geometric Mean: The geometric mean is a mean or average which indicates the central tendency or typical value of a set of numbers by using the product of their values. The procedure to calculate geometric mean is the same as that of arithmetic mean. The only difference is that in geometric mean, we take the sum total of log values of all the items and divide it by the number of items.

$$GM = \sqrt[n]{X_1 \times X_2 \times X_3 \times .. \times X_n} \quad \text{or} \quad \log GM = \frac{\sum \log Xi}{N} \quad i = 1,2,3 \ldots.$$

3. Harmonic Mean: Harmonic mean is calculated using the reciprocals of the values rather than the sum. The harmonic mean can be expressed as the reciprocal of the arithmetic mean of the reciprocal of the series. Harmonic Mean is a type of average used when data values are expressed as **rates, ratios, or speeds.** It is defined as the **reciprocal of the arithmetic mean of reciprocals**

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

II. Positional Averages : Positional averages are those averages whose values are worked out on the basis of their position in the statistical series. Positional averages are two types. 1. Median, 2. Mode

1. **Median :** Median is the single value of the data set that measures the central item in the data. Median is the value of the middle items of a series arranged ascending or descending order of magnitude. The median is the middle value in a given set, or the value that separates the higher half from the lower half.
2. **Mode :** Mode is the value which occurs most often in the data. Mode is the common item of a series. It is the value which occurs more frequently. It is denoted by the symbol **"Z".**

2.3 Arithmetic Mean – Calculation

Arithmetic mean is a statistical measure which is widely used to represent the whole data with the help of a single value. To a layman it is average but for a statistician, it is called mean or arithmetic mean.The arithmetic mean of a group of observations is defined as:

Mean = Sum of observations/Number of observations Definitions:

"Arithmetic mean is the quotient that result when the sum of all the items in the series is divided by number of items"- **Simpson & Kafka.**

"Arithmetic mean is the sum of the observations in a sample divided by the number of observations in the sample"- **Ya-Lun Chou.**

Calculation of Arithmetic Mean: Arithmetic mean is computed for the data relating to individual, discrete and continuous series.

1. Individual series: The arithmetic mean in statistics is nothing but the ratio of all observations to the total number of observations in a data set.

Arithmetic Mean Formula: Arithmetic mean is the sum of all observations divided by the number of observations.

Arithmetic mean formula = Sum of observation ÷ Total numbers of observations

$$\bar{X} = \frac{\sum X}{N}$$

$\overline{X}$ = Arithmetic Mean

$\sum x$ = Sum of all observations

N = Number of observations

Steps to calculate Arithmetic mean in individual series: The following steps are required to calculate arithmetic mean.

1. Add up all the values and denote it by ... Σx
2. Divide the sum (Σx) by the number of observations (N) to obtain the arithmetic mean

Illustration : Calculate Arithmetic Mean from the following data

Variable (X) : 10, 20, 30, 40, 50

Solution:

Formula: Arithmetic mean $(\bar{x}) = \frac{\sum X}{N}$

Summation of observations $(\sum X) = 10+20+30+40+50 = 150$

Number of observations (N) = 5

$(\bar{x}) = \frac{\sum X}{N}$ $\quad \bar{x} = \frac{150}{5} = 30$ $\quad \bar{x} = 30$

Arithmetic mean $(\bar{x}) = 30.$

2. Discrete Series :

Formula : Arithmetic Mean $(\bar{x}) = \frac{\Sigma fx}{N}$

f = frequency, X = variable; N = total frequency (Σf)

Steps to calculate Arithmetic mean in discrete series : The following steps are required to compute arithmetic mean.

1. Multiply each value of the variable with the corresponding frequency and denote the column by 'fX'
2. Add all the values of fX and denote it by ΣfX.
3. Add all the frequencies of the distribution and denote it by N.
4. Apply the formula $\bar{x} = \frac{\Sigma fx}{N}$ and obtain the value of arithmetic mean **$(\bar{x})$.**

Illustration : Calculate Arithmetic Mean from the following data

Variable (X) : 20, 30, 40, 50, 60, 85

Frequency : 8, 12, 20, 10, 6, 4

Solution : Construct a table for the above given data

Variable (X)	Frequency (f)	f. X
20	08	8 x 20 = 160
30	12	12 x 30 = 360
40	20	20 x 40 = 800
50	10	10 x 50 = 500
60	06	06 x 60 = 360
85	04	04 x 85 = 340
	N = 60	Σ fX = 2520

$$\bar{x} = \frac{\Sigma fX}{N} \quad \bar{x} = \frac{2520}{60} \quad \bar{x} = 42$$

Arithmetic Mean $(\bar{x})$ = 42

3. Continuous Series :

Arithmetic mean formula : $\bar{x} = \frac{\Sigma fm}{N}$

f = frequency : m = mid value; N = total frequency (Σf)

Mid value formula : $m = \frac{L1+L2}{2}$

L1 = lower limit of the class; L2 = upper limit of the class.

Steps to calculate Arithmetic mean in continuous series : The following steps are required to calculate arithmetic mean.

1. Find out the mid value of each class by using the formula $m = \frac{L1+L2}{2}$ and denote the column by 'm'
2. Multiply the mid value of each class with their respective frequencies and denote the column by 'f.m'
3. Add all the values of f.m and denote it by Σfm.
4. Add all the frequencies of the distribution and denote it by N.
5. Substitute these values in the formula $\bar{x} = \frac{\Sigma fm}{N}$ and obtain the value of arithmetic

Illustration : Calculate Arithmetic Mean from the following data

Class : 0-10, 10-20, 20-30, 30-40, 40-50

Frequency : 5, 10, 15, 10, 10

Solution : Construct a table for the above given data.

Class	Frequency (f)	Mid Value m=	f. m
0-10	05	$\frac{0+10}{2} = 5$	5x5 = 25
10-20	10	$\frac{10+20}{2} = 15$	10 x 15 = 150
20-30	15	$\frac{20+30}{2} = 25$	15 x 25 = 375
30-40	10	$\frac{30+40}{2} = 35$	10 x 35 = 350
40-50	10	$\frac{40+50}{2} = 45$	10 x 45 = 450
	N = 50		Σfm = 1350

Arithmetic mean $(\bar{x}) = \frac{\Sigma fm}{N}$ $\bar{x} = \frac{1350}{50} = 27$

Arithmetic Mean $(\bar{x}) = 27$

2.4. Arithmetic Mean – Merits and Demerits

Arithmetic mean is a statistical measure which is widely used to represent the whole data with the help of a single value. To a layman it is average but for a statistician, it is called mean or arithmetic mean.

Merits or Advantages of Arithmetic Mean:

1. It is very easy to understand. A common man can also understand arithmetic mean easily.
2. It is very easy to calculate. It does not involve any tedious mathematical calculations.
3. Arithmetic mean is the good representative of the entire data, as it is computed on the basis of
   all the values of a variable.
4. It is rigidly defined.
5. Arithmetic mean is a computed measure. Hence, it is useful for further algebraic treatment.
6. It is better than other measures of central tendency
7. Arithmetic mean is least effected by fluctuations of sampling. Thus, it is the most stable measure of central tendency
8. Arithmetic mean is the centre of gravity, balancing the values on either side of it.
9. Arithmetic mean is a calculated value and is not based on position in the series.
10. Arithmetic mean is a stable statistic and does not vary too much when repeated samples are taken from the same population.
11. It provides good basis for comparison.
12. It is used to computation of various other statistical measures (Mean deviation Standard deviation Correlation etc.)
13. Since, the methods of computation of arithmetic mean are defined scientifically; anyone who computes the arithmetic mean by any method for a given series will get the same answer.

Demerits or Limitations or Disadvantages of Arithmetic Mean:

1. The value of mean depends on each and every item of the series. So it can be significantly impacted by extreme values, either very small or very large.
2. Arithmetic mean cannot be calculated if the distribution has open-ended classes
3. It cannot be calculated when some items of the series are missing.
4. It cannot be calculated even if a single observation is missing or lost or illegible.
5. The value of arithmetic mean cannot be located graphically.
6. Arithmetic mean cannot be determined by investigation or by observation.
7. Arithmetic mean is not useful in qualitative data such as beauty, intelligence, honesty etc.

Despite the above limitations arithmetic mean is popularly used as it is simple to calculate and easy to understand.

2.5. Median – Calculations

Computation of arithmetic mean is based on each and every item of the series. Due to this the value of mean is influenced by the extreme values of the distribution. Thus, the computed value of the mean may mislead the reader. In view of this an alternative average is needed to represent the distribution. Median is one such popular measure of central tendency.

Unlike arithmetic mean median is determined by the position. Hence it is called as positional average. Median is a single value of the data set that measures the central item in the data. Median is the value of the middle item of a series arranged ascending or descending order of magnitude. Median divides the observations in to two equal parts. The median is the value separating the higher half from the lower half of a data sample.

Calculation of median: Median can be computed from individual, discrete and continuous series.

1. Individual series:

Median formula: M = (the size of $(\frac{N+1}{2})$ th observation

M = Median; N = Number of observations

Steps to calculate median in individual series: The following steps are required to calculate median.

1. Arrange the data either in ascending order or in descending order.
2. Write the serial numbers to the left side of the variable values.
3. Apply the formula $(\frac{N+1}{2})$ th observation and obtain the value of median.

Illustration: Calculate Median from the following dataVariable(X): 2.5.3.7. 1.9.6

Solution: Construct a table for the above given data. (Arrange the data in ascending order)

Sl. No	Variable
1	1
2	2
3	3
4	5
5	6
6	7
7	9
N= 7	

M = the size of $(\frac{N+1}{2})$th item

M= the size of $(\frac{7+1}{2})$th item M= the size of $(\frac{8}{2})$th item

M = 4th item

4th item is 5. So Median = 5

M= = 5

2. Discrete Series:

Median formula : M = the size of $(\frac{N+1}{2})$ th observation

M = Median; N = Total frequency (Σf)

Steps to calculate median in discrete series: Median value is obtained with the help of the following steps.

1. Arrange the data either in ascending order or in descending order. Only the value of variable
2. Add all the frequencies of the distribution and denote it by N.
3. Find out the cumulative frequencies by adding the given frequency with the preceding frequency.
4. Apply the formula $(\frac{N+1}{2})$th item and obtain the size of the median.
5. Locate the size of the median in the cumulative frequency column and find the total which is either equal to the size of $(\frac{N+1}{2})$ th item or more than the size of $(\frac{N+1}{2})$th item and obtain the median

Illustration : Calculate Median from the following data

Variable(X) : 20, 50, 30, 70, 10, 40, 60

Frequency : 2, 3, 4, 7, 6, 5, 1

Solution : Construct a table for the above given data. (Arrange the data in ascending order)

Variable	Frequency (f)	Cumulative Frequency (Cf)
10	6	6
20	2	6+2 = 8
30	4	6+2+4 = 12
40	5	6+2+4+5 = 17
50	3	6+2+4+5+3 = 20
60	1	6+2+4+5+3+1 = 21
70	7	6+2+4+5+3+1+7 = 28
	N= 28	

M = the size of $\left(\frac{N+1}{2}\right)$th item

N = 28

M = the size of $(\frac{28+1}{2})$th item

M = the size of $(\frac{29}{2})$th item

M = the size of 14.5^{th} item

Size of 14.5^{th} item lies in the cumulative frequency is 17.

The corresponding value of cumulative frequency 17 is 40.

Hence, the median value is 40.

3. Continuous Series:

Formula: Median class = the size of th item.

Steps to calculate median in continuous series: The following steps are required to calculate median in continuous series.

1. Find out the cumulative frequencies by adding the given frequency with the preceding frequency.
2. Add all the frequencies of the distribution and denote it by N.
3. Determine the median class by using size of N/2th item.
4. Use the following formula and obtain the median value.

$$M = L1 + \left[\frac{\frac{N}{2} - Cf}{f}\right]c$$

M = Median

L1 - lower limit of the median class

N/2 = size of the median item

Cf = cumulative frequency of the class preceding the median class

f = frequency of the median class

c = class interval of the median class

Illustration : Calculate Median from the following data.

Class : 0-10, 10-20, 20-30, 30-40.40-50, 50-60. Frequency; 10. *5,* 6, 7,4, 6.

Solution : Construct a table for the above given data.

Class	Frequency (f)	Cumulative Frequency (Cf)
0-10	10	10
10-20	5	10+5 = 15
20-30	6	10+5+6 = 21
30-40	7	10+5+6+7 = 28
40-50	4	10+5+6+7+4 = 32
50-60	6	10+5+6+7+4+6 = 38
	N = 38	

Median class = the size of $\frac{N}{2}$ th item. = 38/2th item. = 19th item.

19th item lies in the cumulative frequency is 21 and its corresponding class is 20-30. Hence, the median class is 20-30.

$$\mathbf{M= L1+\left[\frac{\frac{N}{2}-Cf}{f}\right]c}$$

L1 = 20; N/2 = 19; Cf = 15; f = 6; c = 10

$$\mathbf{M= 20+\left[\frac{19-15}{6}\right]10}$$

$$\mathbf{M = 20 + \left[\frac{4}{6}\right]10}$$

$$\mathbf{M = 20 + \left[\frac{40}{6}\right]}$$

M = 20 + 6.67

Median (M) = 26.6 7

2.6 Median – Merits and Demerits

"Median 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"- **L. R. Connor.** "Median is the middle most or central value of the variable when the values are arranged in order of magnitude"- Yule & Kendall.

Merits or Advantages of Median:

1. It is very easy to calculate.
2. Like an ideal average median is rigidly defined.
3. It can be located by inspection or by observation in many cases.
4. It is easy to understand.
5. It can be calculated even if the value of the extreme items is not known, but the number of items should be known.
6. Median is not influenced by extreme values of the distribution.
7. Median is the best suited to those areas where direct quantitative measurement is not possible.
8. Median can be calculated for distribution with open end classes.
9. It is the most appropriate average in dealing with quartile data.
10. It can be located graphically.
11. Unlike arithmetic mean, median can be calculated from the data which are incomplete and irregular.
12. Median is used for computing further statistical measures such as dispersion and skewness.

Demerits or Limitations or Disadvantages of Median:

I. Data should be arranged in ascending or descending order for calculation of median. Thus its computation involves additional work.

2. Median is a positional average. Calculation of median does not consider all the items of the series. Thus, it is not fully representative of the entire data.

3. In case of even number series for an ungrouped data, median cannot be determined accurately.

4. Median is not amenable for further algebraic treatment.

5. Median has no sampling stability as it is significantly impacted by fluctuations of sampling. Hence, it is less reliable..

2.7 Mode – Calculations

In certain circumstances arithmetic mean and median fail to represent the mass data meaning full and convincingly as their computed values tend to be illogical. To overcome this limitation, mode has been introduced. Mode is one of the measures of central tendency. Like median mode also a positional average. The term mode has been derived from the French word 'La-Mode ' which means fashion or the most popular phenomenon.

The mode is the value which occurs most often in the data. Mode is the most common item of a series. It is the value which occurs more frequently. It is denoted by the symbol Z. Mode as a statistical average is the observation that occurs with the greatest frequency and thus it is the most fashionable value. The dictionary meaning of the mode is the value at which a relative or absolute maximum occurs more frequently or for a more number of times than any other value in the given frequency distribution.

Calculation of Mode : Mode is computed for the data relating to individual discrete and continuous series.

1. **Individual Series:** In the case of individual series mode is determined by counting the number of times each value repeats. The value of that item which repeats more number of times is takes as mode or modal value.
2. **Discrete Series:** In discrete series mode can be computed either by inspection method or by grouping and analysing method.

   1) Inspection Method: In this method, mode of a discrete series is determined by inspecting the value and their respective frequencies. The value which occurs maximum number of times in the series is treated as mode.

   Illustration: Calculate Mode from the following data

   Variable (X) : 40, 42, 46, 50, 60, 65, 70

   Frequency (f) : 8, 6, 13, 11, 9, 5, 4

   Solution : Construct a table for the above given data.

Variable (X)	Frequency (f)
40	8
42	6
46 (Mode)	13
50	11
60	9
65	5
70	4

By observing the data, it can be noticed that the variable 46 has occurred the maximum number of times i.e. 13 times. Hence, the modal value is 46. Mode (Z) = 46.

2. Grouping Method: The mode determined by inspection method may not be atypical value when the given data is irregular and there is no significant difference between the occurrences of two or more values. In such cases we shall adopt the grouping method to compute the mode. Calculation of mode by using grouping method involves the preparation of two tables. 1. Grouping table 2. Analysis table.

1. Preparation of Grouping Table: A grouping table has seven columns.

1. Write the values of variable vertically.
2. Write the frequency in 6 columns in different ways (Column.No. 1 to 6).
3. Write the frequencies given in the distribution and denote it by column-1.
4. Add two frequency values at a time of column-1, from the very first value and denote it by column-2.
5. Exclude the first frequency value and add two frequency values at a time of column-1, and denote it by column-3.
6. Add three frequency values at a time of column-1, from the very first value and denote it by column-4.
7. Exclude the first frequency value and add three frequency values at a time of column-1, and denote it by column-5.
8. Exclude the first two frequency values of column-1 and add three frequency values at a time of column-1- and denote it by column-6.
9. Now under line the highest frequencies in each column from column-1 to column-6.

2. Preparation of Analysis Table :

1. Write the values of variable horizontally on the right side of the table.
2. Write the column numbers vertically on left-hand side of the table.
3. Record the highest frequencies of each column corresponding their values with mark 1 in the table.
4. Add all 1 marks vertically for each variable and write the total number in the bottom.
5. Find out the mode value which has been repeated maximum number of times.

Illustration **:** Calculate Mode from the following data

Variable(X) **:** 40, 42, 46, 50, 60, 65, 70

Frequency (f) **:** 8, 6, 13, 11, 9, 5, 4

Solution:

Grouping Table:

Variable (X)	Column-1 Frequency (f)	Column-2	Column-3	Column-4	Column-5	Column-6
40 42 46 50 60 65 70	8 6 13 11 9 5 4	8+6=14 13+11=24 9+5=14	6+13=19 11+9=20 5+4=9	8+6+13 = 27 11+9+5 = 25	6+13+11=30 9+5+4 = 18	13+11+9 = 33

Analysis Table:

Column Number	Variable (X)						
	40	42	46	50	60	65	70
1	-	-	1	-	-	-	-
2	-	-	1	1	-	-	-
3	-	-	-	1	1	-	-
4	1	1	1	-	-	-	-
5	-	1	1	1	-	-	-
6	-	-	1	1	1	-	-
	1	2	5	4	2	0	0

As per the above analysis table, the variable 46 has repeated more number of times i.e. 5 times. Hence, the modal value is 46. Mode (Z) = 46.

3. Continuous Series :

Steps to calculate mode in continuous series: The following steps are required to calculate mode in a continuous series.

1. First prepare the grouping table.
2. Prepare the analysis table and find the modal class.
3. Use the following formula and obtain the mode value.

$$Z = L1 + \frac{\Delta 1}{\Delta 1 + \Delta 2} \times C$$

L_1 = Lower limit of the modal class : $\Delta 1 = f_1 - f_0$: $\Delta 2 = f_1 - f_2$: C = Class interval = $(L_2 - L_1)$

L_2 = Upper limit of the mode class : L1 = lower limit of the mode class

f_1 = Frequency of Modal class

f_0 = Frequency of class preceding to modal class

f_2 = Frequency of class succeeding to modal class

Illustration : Calculate Mode from the following

Data Class : 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80..

Frequency (f) : 2, 5, 11, 15, 9, 6, 2.

Solution :

Grouping Table:

Class	Column-1 Frequency (f)	Column-2	Column-3	Column-4	Column-5	Column-6
10-20 20-30 30-40 40-50 50-60 60-70 70-80	2 5 11 15 9 6 2	2+5=7 11+15=26 9+6=15	5+11=16 15+9=24 6+2=8	2+5+11=18 15+9+6=30	5+11+15 = 31 9+6+2 = 17	11+15+9 = 35

Analysis Table:

Column Number	Class						
	10-20	20-30	30-40	40-50	50-60	60-70	70-80
1	-	-	-	1	-	-	-
2	-	-	1	1	-	-	-
3	-	-	-	1	1	-	-
4	-	-	-	1	1	1	-
5	-	1	1	1	-	-	-
6	-	-	1	1	1	-	-
	0	1	3	6	3	1	0

As per the above analysis table, the class 40-50 has repeated more number of times i.e. 6 times.

Hence, the modal class is 40-50.

Formula: Mode (Z) = L1 + $(\frac{\Delta 1}{\Delta 1+\Delta 2})$ C

Class	Frequency (f)	
10-20	2	
20-30	5	
30-40	11	f_0
40-50	15	f_1
50-60	9	f_2
60-70	6	
70-80	2	

$L_1 = 40$

$\Delta 1 = f_1 - f_0 = 15 - 11 = 4$

$\Delta 2 = f_1 - f_2 = 15 - 9 = 6$

$C = L_2 - L_1 = 50 - 40 = 10$

$Z = L1 + \left(\frac{\Delta 1}{\Delta 1 + \Delta 2}\right) C \qquad Z = 40 + \left(\frac{4}{4+6}\right) 10$

$Z = 40 + \left(\frac{4}{10}\right) \times 10 \qquad Z = 40 + \frac{40}{10}$

$Z = 40 + 4 = 44$ **Mode (Z) = 44.**

2.8. Mode – Merits and Demerits

The mode is the value which occurs most often in the data. Mode is the most common item of a series. It is the value which occurs more frequently. It is denoted by the symbol Z. Mode as a statistical average is the observation that occurs with the greatest frequency and thus it is the most fashionable value. The dictionary meaning of the mode is the value at which a relative or absolute maximum occurs more frequently or for a more number of times than any other value in the given frequency distribution.

Merits or Advantages or Uses of Mode :

L. It is very easy to compute and easy to understand the mode as in many a case mode can be determined simply by observing the data.

2. Mode refers to the value that occur maximum number of times in the series. Hence mode is considered to be the typical value which represents the whole data.
3. Unlike arithmetic mean, mode is not affected by the extreme values of the variables.
4. It can be used to describe qualitative phenomenon.
5. It can be calculated for distributions which have open end classes.
6. Unlike arithmetic mean, model value of a distribution can be determined graphically.
7. Mode is the most meaningful measure of central tendency.
8. Mode has wide business application.
9. It is useful in market research.

Demerits or Limitations or Disadvantages of Mode:

l. Mode cannot be taken as the real representative of the entire data as its calculation is not based on all the observations of a given data.

2. It is not capable of further algebraic treatment.
3. Mode is also not rigidly defined.

4. It is not easy to compute a modal value as it involves length calculations both in grouping the items and analyzing the values and their occurrences.
5. Mode cannot be determined always. There can be bi-modal or tri-modal or multi-modal series as well.
6. It is also stated that mode is ill-defined in definite and indeterminate.
7. There are different methods of calculating mode, but all of them do not render the same results.
8. Mode is significantly impacted by fluctuations of sampling.
9. Mode is a very unstable average and its true value is difficult to determine.
10. Mode located graphically may give an approximate value but not an accurate and precise value.
11. The value of mode is affected significantly by the size of the class interval. A change in the size of the class interval will change the value of the mode.

2.9. Properties of Good Average

In statistics, an average (measure of central tendency) is a value that represents the entire dataset with a single figure. It helps in simplifying complex data and facilitates comparison and analysis. However, for an average to be useful and reliable, it must possess certain essential qualities known as the properties of a good average.

Properties of a Good Average

1. Rigidly Defined: A good average should be clearly and precisely defined. Its value should not vary according to the person calculating it. This ensures objectivity and consistency. For Example: Arithmetic Mean has a fixed formula.

2. Easy to Understand and Simple to Calculate: The average should be simple to comprehend and easy to compute. It should not involve complicated procedures. This makes it useful for practical applications.

3. Based on All Observations: A good average must consider all values in the dataset. This ensure it represents the entire data. For Example: Arithmetic Mean includes every observation.

4. Not Unduly Affected by Extreme Values: The average should not be highly influenced by extreme values (outliers). Otherwise, it may give misleading results. Example: Median is better in such situations.

5. Capable of Further Algebraic Treatment: A good average should allow mathematical manipulation. It should be useful in advanced statistical analysis.

6. Stable and Less Affected by Sampling Fluctuations: The value of the average should not change much with different samples.

7. Representative of the Data: The average should reflect the central tendency accurately.It should lie within the range of the data. It must represent the dataset as a whole.

8. Applicable to Different Types of Data: A good average should be suitable for various types of data such as individual, discrete, and continuous series.

In practice, no single average satisfies all the properties completely. However, the Arithmetic Mean is considered the most widely used average because it fulfills most of these characteristics. At the same time, Median and Mode are useful in specific situations, especially when dealing with extreme values or qualitative data.

2.10. Comparison of Different Averages

In statistics, an average (or measure of central tendency) represents a single value that summarizes a dataset. The three main averages are, 1. Arithmetic Mean and (Geometric Mean & Harmonic Mean.

1. Arithmetic Mean (A.M.): Sum of all observations divided by the number of observations.

X is the symbol of the arithmetic mean $\bar{x} = \frac{\sum X}{N}$

1) Geometric Mean: The nth root of the product of n observations.
2) Harmonic Mean: Reciprocal of the arithmetic mean of reciprocals.

2. Median: The middle value in a series when arranged in ascending or descending order.

3. Mode: The value that occurs most frequently.

Comparison Table

Basis	A.M.	Median	Mode	G.M.	H.M.
Definition	Sum ÷ N	Middle value	Most frequent	Root of product	Reciprocal mean
Based on all values	Yes	No	No	Yes	Yes
Effect of extremes	High	None	None	Low	Low
Ease of calculation	Easy	Moderate	Easy	Difficult	Difficult
Use	General purpose	Skewed data	Popular value	Growth rates	Rates
Use of calculation	Easy	Moderate	Easy	Difficult	Difficult

2.11. Measures of Dispersion

The Measures of central tendency are inadequate to describe a distribution completely. Two data sets can have the same mean but they can be entirely different. Thus to describe data, one needs to know, the extent of variability. This is given by the measures of dispersion.

The Measures of central tendency gives us a bird's eye view of the entire data they are called averages of the first order, it serve to locate the centre of the distribution but they do not reveal how the items are spread out on either side of the central value. Thus the measures of central tendency must be supported by other measures and one such measure is the measures of dispersion.

Dispersion is also called as variability, scatter or spread. The measures of dispersion are also called the average of second order, since they measure the average of the deviations taken from the central tendency of the distribution. Dispersion is the measure of variation of items. It measures the extent in which the items vary from central value. A measure of dispersion indicates the scattering of data. It explains the disparity of data from one another delivering a precise view of their distribution. Measures of dispersion play an important role in various statistical investigations. Therefore it is necessary to use an appropriate and satisfactory measure of dispersion. Measures of dispersion may be grouped in to two types.1.Absolute Measures of Dispersion, 2. Relative Measures of Dispersion.

2.12 Absolute and Relative Measures of Dispersion

Measures of dispersion may be grouped in to tv o types.

I. Absolute Measures of Dispersion

II. Relative Measures of Dispersion

I. Absolute Measures of Dispersion: The measures of dispersion which are expressed in terms of the original units of a series are called as absolute measures of dispersion. It is used to compare two or more distributions which are expressed in homogeneous units. These measures are not suitable for comparing the variability in two distributions having variables expressed in different units.

Absolute measures are four types.

1. Range
2. Quartile Deviation
3. Mean Deviation
4. Standard Deviation.

II. Relative Measures of Dispersion: Relative measures of dispersion also called as 'Coefficients'. These measures are considered to be pure numbers independent of the units of measurement. These measures are expressed as ratio of absolute measures of dispersion to the appropriate average used for the measurement of variation. They can also be expressed in terms of percentages. Relative measures are used to compare the distributions expressed in different statistical units of measurement and the relative accuracy of the data.

Relative measures are four types.

1. Co-efficient of Range
2. Co-efficient of Quartile Deviation
3. Co-efficient of Mean Deviation.
4. Co-efficient of Standard Deviation or Co-efficient of Variation

2.13. Range - Calculation

Range is the simplest method of studying dispersion. Range is an absolute measure of dispersion. The difference between highest value and lowest value of a series is known as range. It is defined as the difference between the two extreme items of the distribution.

Symbolically this can be expressed as :

Range = Largest value - Smallest value

Range = Highest value - Lowest value

Range = Maximum value - Minimum value

Calculation of Range: Range can be calculated from individual, discrete and continuous series.

1. Individual series: Steps to calculate range in individual series: The following steps are required to calculate range in individual series.

1. Ascertain largest value and smallest value
2. To calculate range, subtract the smallest value from the largest value (L - S)
3. To calculate co-efficient of range, divide the range by the sum of the largest and the smallest values in the distribution : $\frac{L - S}{L+S}$

Illustration: Calculate range and co-efficient of range from the following data.

Marks: 90, 36, 75, 84, 23, 66, 57, 21, 39

Solution:

Formula : R = L-S

R = Range; L = Largest Value; S = Smallest Value

Largest value = 90

Smallest value = 21

R = L-S

R = 90 - 21 = 69

Range = 69.

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

Co-efficient of Range = $\frac{90 - 21}{90+21}$

Co-efficient of Range = $\frac{69}{111}$

Co-efficient of Range = 0.622

2. Discrete Series: Steps to calculate range in discrete series: The following steps are required to calculate range in discrete series.

1. Ascertain largest value and smallest value
2. To calculate range, subtract the smallest value from the largest value (L- S).
3. To calculate co-efficient of range, divide the range by the sum of the largest and the smallest values in the distribution : $\frac{L - S}{L+S}$
4. Frequencies are not considered at all for computing range and co-efficient of range.

Illustration : Calculate range and co-efficient of range from the following data.

Marks : 10, 20, 30, 40, 50, 60, 70, 80, 90

Frequency : 12, 15, 8, 3, 7, 6, 14, 30, 10

Solution :

Formula: R = L - S

R = Range : L - Largest Value; S - Smallest Value

Largest value = 90

Smallest value = 10

R = L - S

R = 90 - 10 = 80 Range = 80

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

Co-efficient of Range = $\frac{90 - 10}{90+10}$

Co-efficient of Range = 0.8

3. Continuous Series : Steps to calculate range in continuous series: The following steps are required to calculate range in continuous series.

1. Ascertain the upper limit of the highest class (largest value) and lower limit of lower class (smallest value).
2. To calculate range, subtract the lower limit of lower class from the upper limit of the highest class (L - S).
3. To calculate co-efficient of range, divide the range by the sum of upper limit of the highest class and the lower limit of lower class (L - S) ÷ (L + S).
4. Frequencies are not considered at all for computing range and co-efficient of range.

Illustration : Calculate range and co-efficient of range from the following data.

Marks : 0-10, 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80

Frequency : 3, 1, 5, 2, 6, 8, 2, 9

Solution :

Formula: R = L - S

R = Range : L = Largest Value: S = Smallest Value

Largest value (upper limit of the highest class) = 80

Smallest value (lower limit of the lower class) = 0

R = L - S R = 80 - 0 = 80 Range = 80.

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

Co-efficient of Range = $\frac{80 - 0}{80 + 0}$

Co-efficient of Range = 1.

2.14. Quartile Deviation - Calculation

Quartile Deviation: To overcome certain limitations of range another method known as "Inter-Quartile Range" has been developed. Inter quartile range is the difference between the upper and lower quartiles (Q3 - Q1) of a given frequency distribution. It is the range for the middle 50% of the data.

Inter Quartile Range / Inter Quartile Deviation =

Upper Quartile - Lower Quartile (or) third quartile - first quartile

IQR / IQD = Q3-Q1

Semi-Inter Quartile Range/Quartile Deviation: The interquartile range is reduced to the form of the semi-inter quartile range or quartile deviation by dividing it by

2. Half the distance between upper quartile (Q3) and lower quartile (Q1) is called the semi-inter quartile range or quartile deviation. Thus it represents the average difference between two quartiles Symbolically this can be expressed as...

$$\text{Semi - Inter Quartile Range / Quartile Deviation} = \frac{\text{Upper Quartile - Lower Quartile}}{2}$$

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

Coefficient Of Quartile Deviation :

Quartile deviation is an absolute measure of dispersion. This is useful in comparing the distributions expressed in homogeneous statistical units only. To compare the degree of variation of different sets of data, relative measure of dispersion known as "coefficient of quartile deviation" is calculated.

Coefficient of quartile deviation can be calculated by dividing the quartile deviation by the average of the two quartiles.

$$\text{Co-efficient of Quartile Deviation} = \frac{\text{(Upper Quartile - Lower Quartile) / 2}}{\text{(Upper Quartile + Lower Quartile) / 2}}$$

$$\text{Co-efficient of Quartile Deviation} = \frac{\text{Upper Quartile - Lower Quartile}}{\text{Upper Quartile + Lower Quartile}}$$

$$\text{Co-efficient of Quartile Deviation} = \frac{(Q3 - Q1) / 2}{(Q3 + Q1) / 2}$$

$$\text{Co-efficient of Quartile Deviation} = \frac{Q3 - Q1}{Q3 + Q1}$$

Calculation of Quartile Deviation: Quartile deviation can be calculated from individual, discrete and continuous series.

1. Individual Series:

Steps to calculate Quartile Deviation in individual series: The following steps are required to calculate quartile deviation in individual series.

1. Arrange the values in ascending order or descending order.
2. Obtain the value of lower quartile (Q1) by using the formula $\mathbf{Q1 = size\ of\ \left(\frac{N+1}{4}\right)th\ item.}$

   Where, N is the number of observations.
3. Obtain the value of upper quartile (Q3) by using the formula $\mathbf{Q3 = size\ of\ 3\left(\frac{N+1}{4}\right)th\ item.}$
4. Calculate Quartile Deviation by applying the formula $\mathbf{Q.D = \frac{Q3-Q1}{2}}$
5. Calculate co-efficient of Quartile Deviation by applying the formula $\mathbf{CQD = \frac{Q3-Q1}{Q3+Q1}}$

Illustration: Calculate quartile deviation and co-efficient of quartile deviation from the following data.

Variable (X) : 24, 30, 40, 56, 60, 80, 90

Solution: Construct a table for the above given data.
(Arrange the data in ascending order)

Sl. No	Variable (X)
1	24
2	30
3	40
4	56
5	60
6	80
7	90
N=7	

Formula: $\mathbf{Q1 = size\ of\ \left(\frac{N+1}{4}\right)th\ item.}$ $\mathbf{Q3 = size\ of\ 3\left(\frac{N+1}{4}\right)th\ item.}$

$Q1 = size\ of\ \left(\frac{7+1}{4}\right)th\ item \quad Q1 = size\ of\ \left(\frac{8}{4}\right)th\ item$

Q1 = size of 2nd item. 2nd item is 30. Hence, **Q1 = 30.**

$\mathbf{Q3 = size\ of\ 3\left(\frac{N+1}{4}\right)th\ item.}$

$Q3 = size\ of\ 3\left(\frac{7+1}{4}\right)th\ item.\ Q3 = size\ of\ 3\left(\frac{8}{4}\right)th\ item.$

Q3 = 3x2 item = 6th item. 6th item is 80. Hence, **Q3 = 80.**

$Quartile\ Deviation = \frac{Q3-Q1}{2} \quad Q.D = \frac{80-30}{2} \quad Q.D = \frac{50}{2} \quad Q.D = 25.$

$Co\text{-}efficient\ of\ Quartile\ Deviation = \frac{Q3-Q1}{Q3+Q1} \quad \mathbf{CQD = \frac{80-30}{80+30}} \quad \mathbf{CQD = \frac{50}{110}} \quad \mathbf{CQD = 0.45.}$

2. Discrete Series:

Steps to calculate Quartile Deviation in Discrete series: The following steps arc required to calculate quartile deviation in discrete series.

1. Arrange the values in ascending order or descending order. Only the value of variable should be arranged in ascending or descending order.
2. Add all the frequencies of the distribution and denote it by N.
3. Find out the cumulative frequencies by adding the given frequency with the preceding frequency.
4. Apply the formula $\left(\frac{N+1}{4}\right)$th item and obtain the size of the lower quartile (Q1) Where, N is the total frequency.
5. Locate the size of the lower quartile (Q1) in the cumulative frequency column and find the total which is either equal to the size of $\left(\frac{N+1}{4}\right)$th item or more than the size of item and obtain the lower quartile (Q1).
6. Apply the formula $3\left(\frac{N+1}{4}\right)$th item and obtain the size of the upper quart ilc (Q3).
7. Locate the size of the upper quartile (Q3) in the cumulative frequency column and find the total which is either equal to the size of $3\left(\frac{N+1}{4}\right)$th item or more than the size of 3 item and obtain the upper quartile (Q3).
8. Calculate Quartile Deviation by applying the formula $\mathbf{Q.D = \frac{Q3-Q1}{2}}$
9. Calculate co-efficient of Quartile Deviation by applying the formula $\mathbf{CQD = \frac{Q3-Q1}{Q3+Q1}}$

Illustration: Calculate Q.D. from the following data

Variable (X): 30, 36, 40, 46, 50, 60, 70.

Frequency: 12, 6, 10, 20, 4, 8, 2

Variable (X)	Frequency (f)	Cumulative Frequency (Cf)
30	12	12
36	6	12+6 = 18
40	10	12+6+10 = 28
46	20	12+6+10+20 = 48
50	4	12+6+10+20+4 = 52
60	8	12+6+10+20+4+8 = 60
70	2	12+6+10+20+4+8+2 = 62
	N= 62	

Q1 = size of $\left(\frac{N+1}{4}\right)$th item Q1 = size of $\left(\frac{62+1}{4}\right)$th item

Q1 = size of $\left(\frac{63}{4}\right)$th item Q1 = size of 15.75^{th} item

Size of 15.75^{th} item lies in the cumulative frequency is 18. The corresponding value of cumulative frequency 18 is 36. Hence, the lower quartile is 36. **Q1 = 36.**

Q3 = size of $3\left(\frac{N+1}{4}\right)$th item Q3 = size of $3\left(\frac{62+1}{4}\right)$th item

Q3 = size of $3\left(\frac{63}{4}\right)$th item Q3 = size of 3 x 15.75^{th} item

Q3 = size of 47.25^{th} item.

Size of 47.25^{th} item lies in the cumulative frequency is 48.

The corresponding value of cumulative frequency 48 is 46.

Hence, the upper quartile is 46. **Q3 = 46**.

Q.D $= \frac{Q3-Q1}{2}$ Q.D $= \frac{46-36}{2}$ Q.D $= \frac{10}{2} = 5$

Quartile deviation = 5

Co-efficient of Quartile Deviation $= \frac{Q3-Q1}{Q3+Q1}$

CQD $= \frac{46-36}{46+36}$ CQD $= \frac{10}{82} = 0.122$

Co-efficient of Quartile Deviation = 0.122.

2. Continuous Series:

Steps to calculate Quartile Deviation in Continuous series: The following steps are required to calculate quartile deviation in continuous series.

1. Add all the frequencies of the distribution and denote it by N.
2. Find out the cumulative frequencies by adding the given frequency with the preceding frequency.
3. Apply the formula $\left(\frac{N}{4}\right)$th item and obtain the size of the lower quartile (Q1).
   Where, N is the total frequency.
4. Locate the size of the lower quartile (Q1) in the cumulative frequency column and obtain the lower quartile (Q1) class.
5. Calculate lower quartile (Q1) by applying the formula $Q1 = L1+\left[\frac{\frac{N}{4}-Cf}{f}\right]c$
   L1= lower limit of the Q1 class: Cf = cumulative frequency of class preceding to Q1 class; f = frequency of the Q1 class; c = class interval of the Q1 class (L2-L1).
6. Apply the formula $3\left(\frac{N}{4}\right)$th item and obtain the size of the upper quartile (Q3).
7. Locate the size of the upper quartile (Q3) in the cumulative frequency column and obtain the upper quartile (Q3).
8. Calculate upper quartile (Q3) by applying the formula $Q3 = L1+\left[\frac{3\left[\frac{N}{4}\right]-Cf}{f}\right]c$
9. Calculate Quartile Deviation by applying the formula $Q.D = \frac{Q3-Q1}{2}$
10. Calculate co-efficient of Quartile Deviation by applying the formula $CQD = \frac{Q3-Q1}{Q3+Q1}$

Illustration : Calculate quartile deviation from the following data

Class : 5-10, 10-15,15-20,20-25,25-30.

Frequency : 3, 10, 17, 6, 8

Solution : Construct a table for the above given data.

Class	Frequency (f)	Cumulative Frequency (Cf)
5-10	3	3
10-15	10	3+10 = 13
15-20	17	3+10+17 = 30
20-25	6	3+10+17+6 = 36
25-30	8	3+10+17+6+8 = 44
	N=44	

Q1 = size of $\left(\frac{N}{4}\right)$th item

Q1 = size of $\left(\frac{44}{4}\right)$th item

Q1 = size of 11th item.

Size of 11th item lies in the cumulative frequency is 13. The corresponding class of cumulative frequency 13 is 10-15. Hence, the lower quartile class is 10-15. **Q1class = 10-15.**

$$Q1 = L1 + \left[\frac{\left[\frac{N}{4}\right] - Cf}{f}\right] c$$

L1 = 10; N = 44; Cf =3; f = 10; c = L2 - L1 = 15 - 10 = 5.

$$Q1 = 10 + \left[\frac{\left[\frac{44}{4}\right] - 3}{10}\right] 5 \qquad Q1 = 10 + \left[\frac{11-3}{10}\right] 5 \qquad Q1 = 10 + \left[\frac{8}{10}\right] 5$$

$$Q1 = 10 + \left[\frac{40}{10}\right] \quad Q1 = 10 + 4 = 14$$

Q1 = 14

Q3 = size of 3 $\left(\frac{N}{4}\right)$th item; **Q3 = size of 3** $\left(\frac{44}{4}\right)$th item

Q3 = size of 3 x 11th item; **Q3 = size of 33th** item.

Size of 33th item lies in the cumulative frequency is 36. The corresponding class of cumulative frequency 36 is 20-25. Hence, the upper quartile class is 20-25. **Q3class =20-25.**

$$Q3 = L1 + \left[\frac{3\left[\frac{N}{4}\right] - Cf}{f}\right] c$$

L1 = 20; N = 44; Cf =30; f = 6; c = L2 - L1 = 25 - 20 = 5.

$$Q3 = 20 + \left[\frac{3\left[\frac{44}{4}\right] - 30}{6}\right] 5 \qquad Q3 = 20 + \left[\frac{3(11)-30}{6}\right] 5 \qquad Q3 = 20 + \left[\frac{33-30}{6}\right] 5$$

$$Q3 = 20 + \left[\frac{15}{6}\right] \quad Q3 = 20 + 2.5 = 22.5$$

Q3 = 22.5

$Q.D = \frac{Q3-Q1}{2}$ $Q.D = \frac{22.5-14}{2}$

$Q.D = \frac{8.5}{2}$ Quartile Deviation (Q.D) = 4.25.

Co-efficient of Quartile Deviation = $\frac{Q3-Q1}{Q3+Q1}$

$CQD = \frac{22.5-14}{22.5+14}$ $CQD = \frac{8.5}{36.5} = 0.23.$

Co-efficient of Quartile Deviation (CQD) = 0.23.

2.15. Mean Deviation - Calculation

Mean Deviation, also known as Average Deviation, is the average of the absolute deviations of a series of values from a measure of central tendency. Theoretically, it can be calculated from any average such as Arithmetic Mean, Median, Mode, Geometric Mean, or Harmonic Mean.

Though deviations can be taken from any measure of central tendency arithmetic mean is preferred because it is mathematically a sound measure of central tendency. This is the reason why the measure is called Mean Deviation. For calculation of mean deviation, the algebraic signs + or - are to be ignored because, if signs are taken into consideration, the sum of deviations from mean will be zero.

Mean Deviation Formula: Mean deviation is calculated by dividing the sum of absolute deviations taken from the average (mean, median or mode) by the number of observations. The formula of mean deviation, when deviations are taken from arithmetic mean

The formula of mean deviation, when deviations are taken from arithmetic mean ($\bar{x}$) is $MD\ \bar{x} = \frac{\sum D}{N}$ Deviations (D) $= |X - \bar{x}|$ N = numbers of observations

The formula of mean deviation, when deviations are taken from median is

$MDM = \frac{\sum D}{N}$ Deviations (D) $= |X - M|$(median) N = numbers of observations

The formula of mean deviation, when deviations are taken from mode is

$MDZ = \frac{\sum D}{N}$ Deviations (D) $= |X - Z|$(mode) N = numbers of observations

Coefficient of Mean Deviation : Mean deviation is an absolute measure of dispersion. This is useful in comparing the distributions expressed in homogeneous statistical units only. To compare the degree of variation of different sets of data, relative measure of dispersion is calculated. The relative measure of mean deviation is called coefficient of mean deviation. Coefficient of mean deviation is computed by dividing the mean deviation by the average (mean, median or mode) used in measuring such deviations. The formula of co efficient of mean deviation, when deviations are taken from mean is

Coefficient of Mean Deviation (CMDx) = **MD/$\overline{X}$**
MD = Mean Deviation x = Arithmetic mean
The formula of co efficient of mean deviation, when deviations are taken from median is Coefficient of Mean Deviation (CMDM) = **MD/M**
MD = Mean Deviation M = Median
The formula of co efficient of mean deviation, when deviations are taken from mode is Coefficient of Mean Deviation (CMDz) = MD/Z
MD = Mean Deviation Z = Mode.

Calculation of Mean Deviation : Mean deviation can be calculated from individual discrete and continuous series

Calculation of Mean Deviation: Mean deviation can be calculated from individual, discrete and continuous series. 1. Individual series:
Steps to calculate Mean Deviation in individual series: The following steps are required to calculate mean deviation in individual series.

1. Arrange the data in ascending or descending order, when median is used as average.
2. Calculate average (mean/median/mode) of the series.
3. Subtract the average (mean/median/mode) from variable (X - mean or median) and denote it by D.(D = X-Average)
4. Add all the deviations ignoring + or − signs and denote it by
5. Calculate mean deviation by dividing the total deviations (∑D) by total number of observations **(N). ∑D/N**
6. Divide the mean deviation by the average used to obtain co-efficient of mean deviation. **CMD = MD/Average**

Illustration : Calculate Mean deviation from the following data
Variable (X) : 34, 48, 41, 36, 44, 37, 43, 39, 40, 38.
Solution : Construct a table for the above given data. (Deviations are obtained from arithmetic mean)

SL.No	Variable (X)	D=|X- $\bar{X}$|
1	34	34-40=6
2	48	48-40=8
3	41	41-40=1
4	36	36-40=4
5	44	44-40=4
6	37	37-40=3
7	43	43-40=3
8	39	39-40=1
9	40	40-40=0
10	38	38-40=2
N=10	∑X =400	∑D =32

$$\bar{X} = \frac{\sum x}{N} = \frac{400}{10} = 40$$

$$MD = \frac{\sum D}{N} = \frac{32}{10} = 3.2$$

Mean Deviation (MD) = 3.2

$$CMD = \frac{MD}{\bar{X}} = \frac{3.2}{40} = 0.08$$

Co-efficient of Mean Deviation (CMD) = 0.08

Continuous Series :

Steps to calculate Mean Deviation in continuous series: The following steps are required to calculate mean deviation in continuous series.

1. Find out the mid values of each class by using the formula m = (L1 + L2)/2 and denote the column by 'm'.
2. Add all the frequencies of the distribution and denote it by N.
3. Calculate average (mean/median/mode) of the series.
4. Take deviations from the mid values of each class and denote it by D. (D = m-average)
5. Multiply each deviation by its corresponding frequencies (f.D) and obtain the total by ignoring + or - signs and denote it by $\sum fD$
6. Calculate mean deviation by using the formula = $\frac{\sum fD}{N}$
7. Calculate co-efficient of mean deviation by using the formula CMD = $\frac{MD}{Average}$

Illustration : Calculate mean deviation from the following data
Class: 0-10, 10-20, 20-30, 30-40, 40-50, 50-60

Frequency (f) : 5, 9, 11, 15, 6, 4

Solution : Construct a table for the above given data.
(Deviations are taken from arithmetic mean)

Class	Frequency(f)	$M = \frac{L1+L2}{2}$	fm	$D = m-\bar{X}$ (m-29)	f.D
0-10	5	(0+10)/2 = 5	5 x 5 = 25	5 – 29 = 24	5 x 24 = 120
10-20	9	(10+20)/2 = 15	9 x 15 = 135	15 – 29 = 14	9 x 14 = 126
20-30	11	(20+30)/2 = 25	11 x 25 = 275	25 – 29 = 4	11 x 4 = 44
30-40	15	(30+40)/2 = 35	15 x 35 = 525	35 – 29 = 6	15 x 6 = 90
40-50	6	(40+50)/2 = 45	6 x 45 = 270	45 – 29 = 16	6 x 16 = 96
50-60	4	(50+60)/2 = 55	4 x 55 = 220	55 – 29 = 26	4 x 26 = 104
	N = 50		∑fm=1450		∑fD=580

$$\bar{X} = \frac{\sum fm}{N} = \frac{1450}{50} = 29 \qquad \bar{X} = 29; \qquad MD = \frac{\sum fD}{N} = \frac{580}{50} = 11.6$$

Mean Deviation (MD) = 11.6

$$CMD = \frac{MD}{\bar{X}} = \frac{11.6}{29} = 0.40$$

Co. efficient of Mean Deviation (CMD) = 0.40

2.16. Standard Deviation - Calculation

The main defect of mean deviation is that it ignores algebraic signs of deviations. To remove the drawback of mean deviation standard deviation is used. The standard deviation is the average amount of variability of a dataset. It tells us on average how far each value lies from the mean. A high standard deviation means that values are generally far from the mean while a low standard deviation indicates that values are clustered close to the mean.

To overcome the limitation of ignoring the algebraic signs in mean deviation the concept of standard deviation was introduced by Karl Pearson in 1893. Standard deviation is an important member of the family of measures of dispersion. Standard deviation is the most important and widely used measure of dispersion. It satisfies most of the properties of a good measure of dispersion.

Calculation of Standard Deviation: Standard deviation can be calculated from individual, discrete and continuous series.

1. Individual Series: Steps to calculate Standard Deviation in individual series: The following steps are required to calculate standard deviation in individual series.

1. Calculate mean of the series by using the formula $\bar{X} = \frac{\sum X}{N}$ (N=number of observations).
2. Take deviations from the actual mean by taking out the difference between the value of variable (X) and the actual mean ($\bar{X}$) and denote it by d. ($d = X - \bar{X}$)
3. Square the deviations add all the squared deviations and denote it by Σd^2
4. Calculate standard deviation by using the formula $\sigma = \sqrt{\frac{\sum d^2}{N}}$ (N=number of observations)
5. Divide the standard deviation by the arithmetic mean to obtain co-efficient of variation. $CV = \frac{\sigma}{\bar{X}} \times 100$

Illustration: Calculate standard deviation from the following data **Variable (X):** 10.12. 14, 16, 18, 20, 22, 24, 26, 28. **Solution:** Construct a table for the above given data.

Sl. No	Variable (X)	$d = X - \bar{X}$ (X – 19)	d^2
1	10	10 - 19 = -9	9 x 9 = 81
2	12	12 - 19 = -7	7 x 7 = 49
3	14	14 - 19 = -5	5 x 5 = 25
4	16	16 - 19 = -3	3 x 3 = 9
5	18	18 - 19 = -1	1 x 1 = 1
6	20	20 – 19 = 1	1 x 1 = 1
7	22	22 – 19 = 3	3 x 3 = 9
8	24	24 – 19 = 5	5 x 5 = 25
9	26	26 - 19 = 7	7 x 7 = 49
10	28	28 - 19 = 9	9 x 9 = 81
N = 10	$\sum x = 190$		$\sum d^2 = 330$

$$\bar{X} = \frac{\sum x}{N} \quad \frac{190}{10} = 19$$

$$\sigma = \sqrt{\frac{\sum d^2}{N}} = \sqrt{\frac{330}{10}} = \sqrt{33} = 5.744$$

Standard Deviation (σ) = 5.744

Co-efficient of Variation (C.V) = $\frac{\sigma}{\bar{X}} \times 100$

CV = $\frac{5.744}{19} \times 100 = 30.23$

Co. efficient of Variation (C.V.) = 30.23%

2. Discrete series:

Steps to calculate Standard Deviation in Discrete series: The following steps are required to calculate standard deviation in discrete series.

1. Calculate arithmetic mean of the series by using the formula $\bar{X} = \frac{\Sigma fx}{N}$ (N=total frequency)
2. Obtain the deviations from the mean and denote it by d. $(d = X - \bar{X})$.
3. Square these deviations
4. Multiply the squared deviations with the respective frequencies (fd^2) and obtain the total i.e. $\sum fd^2$
5. Calculate standard deviation by using the formula $\sigma = \sqrt{\frac{\Sigma fd^2}{N}}$ (N=total frequency)
6. Divide the standard deviation by the arithmetic mean to obtain co-efficient of variation. CV $= \frac{\sigma}{\bar{X}} \times 100$

Illustration: Calculate standard deviation and co-efficient of variation from the following data

Variable (X): 10. 12, 15,18, 20 **Frequency (f) :** 5, 10, 6, 5, 8

Solution: Construct a table for the above given data.

Variable (X)	Frequency (f)	fx	$d=\|X-\bar{X}\|$ $(X-15)$	d^2	fd^2
10	5	10 x 5 = 50	10 - 15 = -5	5 x 5 = 25	5 x 25 = 125
12	10	12 x 10 = 120	12 - 15 = -3	3 x 3 = 9	10 x 9 = 90
15	6	15 x 6 = 90	15 - 15 = 0	0 x 0 = 0	6 x 0 = 0
18	5	18 x 5 = 90	18 - 15 = 3	3 x 3 = 9	5 x 9 = 45
20	8	20 x 8 = 160	20 - 15 = 5	5 x 5 = 25	8 x 25 = 200
	N= 34	$\sum fx = 510$			$\sum fd^2 = 460$

$$\bar{X} = \frac{\Sigma fx}{N} = \frac{510}{34} = 15$$

$$\sigma = \sqrt{\frac{\Sigma f d^2}{N}} = \sqrt{\frac{460}{34}} = \sqrt{13.529} = 3.678$$

Standard Deviation (σ) = 3.678

$$C.V = \frac{\sigma}{\bar{X}} \times 100 = \frac{3.678}{15} \times 100 = 24.52$$

Co. efficient of Variation (C.V.) = 24.52%

3. Continuous Series:

Steps to calculate Standard Deviation in Continuous series: The following steps are required to calculate standard deviation in continuous series.

1. Find out the mid values of each class by using the formula $m = \frac{L1 + L2}{2}$ and denote the column by "m'.
2. Multiply the mid value of each class with their respective frequencies and denote it by f.m and obtain the total i.e.
3. Add all the frequencies of the distribution and denote it by N.
4. Substitute these values in the formula $\bar{x} = \frac{\sum fm}{N}$ and obtain the value of arithmetic mean (x).
5. Take deviations from the mid values of each class and denote it by d. (d = m - X)
6. Square these deviations (d^2)
7. Multiply the squared deviations with the respective frequencies (fd^2) and obtain the total i.e.
8. Calculate standard deviation by using the formula $\sigma = \sqrt{\frac{\sum fd^2}{N}}$ (N=total frequency)
9. Divide the standard deviation by the arithmetic mean to obtain co-efficient of variation (× 100).

Illustration: Calculate standard deviation and co-efficient of variation from the following data

Class: 0-2, 2-4. 4-6, 6-8, 8-10. 10-12 **Frequency (f):** 3, 4, 3, 2, 1, 2.

Solution: Construct a table for the above given data.

Class	Frequency (f)	Mid value (m)	fm	Deviations $d = m - \bar{X}$ (m – 5)	d^2	fd^2
0-2	3	1	3 x 1 = 3	1 – 5 = –4	4 x 4 = 16	3 x 16 = 48
2-4	4	3	4 x 3 = 12	3 – 5 = -2	2 x 2 = 4	4 x 4 = 16
4-6	3	5	3 x 5 = 15	5 – 5 = 0	0 x0 = 0	3 x 0 = 0
6-8	2	7	2 x 7 = 14	7 – 5 = 2	2 x 2 = 4	2 x 4 = 8
8-10	1	9	1 x 9 = 9	9 -5 = 4	4 x 4 = 16	1 x 16 = 16
10-12	2	11	2 x 11 = 22	11 – 5 = 6	6 x 6 = 36	2 x 36 = 72
	N= 15		$\sum$ fm = 75			$\sum fd^2 = 160$

$\bar{x} = \frac{\sum fm}{N} = \frac{75}{15} = 5$

$\sigma = \sqrt{\frac{\sum f d^2}{N}} = \sqrt{\frac{160}{15}} = \sqrt{10.667} = 3.27$ **Standard Deviation (σ) = 3.27**

Co-efficient of Variation (C.V) $= \frac{\sigma}{\bar{X}} \times 100 = \frac{3.27}{5} \times 100 = 65.4\%$

2.17. Variance. and Coefficient of Variation

Variance: Unlike range and quartile deviation, the variance combines all the values in a data set to produce a measure of spread. The variance and standard deviation are the most commonly used measures of dispersion. Variance is a measure of how spread out a data set is. It is the square of standard deviation. The average of the sum of squared deviations from the mean is called Variance. Variance is calculated exactly as standard deviation, without square root. The Square of standard deviation is variance.

$$\mathbf{V = \sigma^2 \ (or) \ V = \frac{\sum d2}{N}}$$

V= Variance σ = Standard Deviation

Coefficient of Variation (C.V): The standard deviation is an absolute measure of dispersion. The relative measure of dispersion is known as the co-efficient of variation. This concept developed by Karl Pearson. Co-efficient of variation is denoted by C.V. It is used in such problems where variability of two or more series is to be compared. The series for which the co-efficient of variation is greater is said to be more variable or less consistent. The series for which co-efficient of variation is less, is said to be less variable or more consistent.

$CV = \frac{\sigma}{\overline{X}} \times 100$, where CV = Co-efficient of Variation

σ = Standard Deviation $\overline{X}$ = Arithmetic mean

2.18. Use of Excel & Python in Statistical Applications

Statistical analysis has become faster and more accurate with the help of modern tools like Microsoft Excel and Python. These tools help in organizing, analyzing, and visualizing data efficiently.

1. **Excel in Statistics: Excel in Statistics** refers to the use of **Microsoft Excel** software for **organizing, analyzing, and interpreting statistical data** using built-in formulas, tools, and charts.
   1) **Data Entry & Organization:** Easy to store large datasets in tabular form
   2) **Statistical Functions:** Built-in formulas like Average (Mean), Median, Mode and Standard Deviation
   3) **Data Analysis Tool Pak:** Helps in regression, correlation, ANOVA
   4) **Charts & Graphs:** Bar charts, histograms, pie charts for visualization
   5) **Pivot Tables:** Summarize and analyze data quickly
2. **Python in Statistics: Python in Statistics** refers to the use of the **Python programming language** for **data analysis, statistical calculations, and data visualization** using specialized libraries.
   1) **Data Analysis :** Using libraries like 'pandas' and 'numPy'
   2) **Statistical Computations :** Mean, variance, regression, hypothesis testing
   3) **Data Visualization :** Using 'matplotlib' and 'seaborn'
   4) **Handling Big Data :** Efficient for large datasets
   5) **Automation :** Repetitive tasks can be automated

Comparison Table

Basis	Excel	Python
Ease of Use	Very easy	Moderate (needs coding)
Cost	Paid (MS Office)	Free (Open-source)
Data Size	Small to medium	Very large datasets
Statistical Power	Limited	Very high
Automation	Limited	High
Visualization	Basic charts	Advanced visualization
Flexibility	Less	Highly flexible

2.19. Model Examination Questions

Short Type Questions :-

1. Average
2. Measures of central tendency
3. Mathematical averages
4. Positional averages
5. Arithmetic mean
6. Median
7. Mode
8. Harmonic mean
9. Geometric mean
10. Measures of dispersion
11. Absolute measures of dispersion
12. Relative measures of dispersion
13. Range and co-efficient of range
14. Quartile deviation and co-efficient of quartile deviation
15. Mean deviation and co-efficient of mean deviation
16. Standard deviation
17. Variance
18. Co-efficient of variation

Essay Type Questions :-

1. Define average and explain the properties of a good average
2. Define Measures of central tendency and explain various types of measures of central tendency
3. Define arithmetic mean and explain the merits and demerits of arithmetic mean
4. Define median and explain the calculation
5. Define mode and explain the calculation
6. Define measures of dispersion and explain the properties of a good measure of dispersion.
7. Define range and explain the calculation
8. Define quartile deviation and explain the calculation
9. Define standard deviation and explain the calculation
10. Explain Use of Excel & Python in Statistical Applications

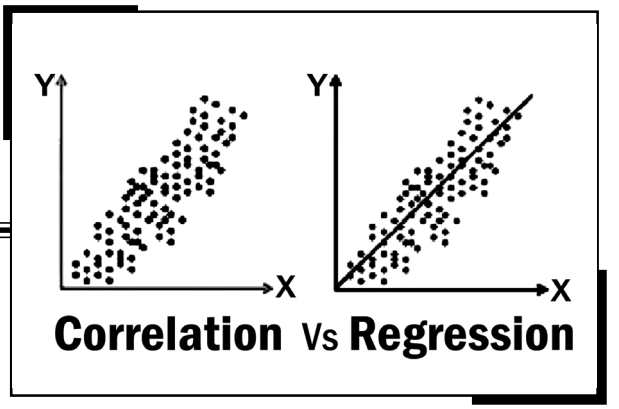
Author

Dr. S.P Krishnaveni, M.A., SLET, B.Ed., Ph.D.

Associate Professor of Economics Govt City College (A) Hyderabad.

Module

3



Correlation and Regression

3.1.	Meaning of Correlation	63
3.2.	Types of Correlation	64
3.3.	Importance of Correlation	65
3.4.	Methods of Calculating Correlation	66
3.5.	Karl Pearson's Correlation Co-efficient	67
3.6.	Spearman's Rank Correlation	69
3.7.	The Concurrent Deviation Method	70
3.8.	The Least Squares Method	72
3.9.	Meaning of Regression	73
3.10.	Uses of Regression Analysis	74
3.11.	Simple Regression	75
3.12.	Multiple Regressions	76
3.13	Model Examination Questions	77

3.1. Meaning of Correlation

Correlation means association - more precisely it is a measure of the extent to which two variables are related. There are three possible results of a correlation study: a positive correlation, a negative correlation and no correlation. A positive correlation is a relationship between two variables in which both variables move in the same direction. Therefore, when one variable increases as the other variable increases, or one variable decreases while the other decreases. An example of positive correlation would be height and weight. Taller people tend to be heavier. A negative correlation is a relationship between two variables in which an increase in one variable is associated with a decrease in the other. An example of negative correlation would be price and demand. As a price of a commodity increases its demand will be decreases. A zero correlation exists when there is no relationship between two variables. For example there is no relationship between the amount of tea drunk and level of intelligence.

Correlation shows the strength of a relationship between two variables and is expressed numerically by the correlation coefficient. To measure the degree of association or relationship between two variables quantitatively an index of relationship is used and is termed as co-efficient of correlation. Co-efficient of correlation is a numerical index that tells us to what extent the two variables are related and to what extent the variations in one variable changes with the variations in the other. The size of *'r'* indicates the amount (or degree or extent) of correlation-ship between two variables. If the correlation is positive the value of *r* is positive and if the correlation is negative the value of 'r' is negative. Thus, the signs of the coefficient indicate the kind of relationship. The value of' '*r'* varies from +1 to -1.

Correlation can vary in between perfect positive correlation and perfect negative correlation. The top of the scale will indicate perfect positive correlation and it will begin from +**l** and then **it** will pass through zero indicating entire absence of correlation. The bottom of the scale will end at -1 and it will indicate perfect negative correlation. Thus numerical measurement of the correlation is provided by the scale which runs from +1 to -1.

3.2. Types of Correlation

Depending upon the direction and proportion of changes in the variables and the number of data series, the correlation may be of the following types:

1. Positive,. Negative and Zero Correlation
2. Linear and curvilinear (Non-linear).
3. Simple, multiple and partial correlations.

1. Positive, Negative and Zero Correlation:

1) **Positive Correlation:** When increase in one variable (X) is followed by a corresponding increase in the other variable (Y), the correlation is said to be positive correlation. The positive correlation is range from 0 to l: the upper limit i.e. + 1 is the perfect positive coefficient of **correlation.** The perfect positive correlation specifies that, for every unit increase in one variable, there is proportional increase in the other. For example ··Heat" and '·Temperature' have a perfect positive correlation.

2) **Negative Correlation:** If an increase in one variable **(X),** results **in** a corresponding decrease in the other variable (Y), the correlation is said to be negative correlation. The negative correlation rang is from O to - l: the lower limit giving the perfect negative correlation. The perfect negative correlation indicates that for very unit increase in one variable, there is proportional unit decrease in the other.

3) **Zero Correlation:** Zero correlation means no relationship between the two variables X and Y; i.e. the change in one variable (X) is not associated with the change in the other variable (Y). For example, body weight and intelligence. Shoe size and monthly salary: etc. The zero correlation is the mid-point of the range - l to+ l. ie Zero.

2. Linear or Curvilinear Correlation:

1) **Linear Correlation:** Linear correlation is the ratio of change between the two variables either in the same direction or opposite direction and the graphical representation of the one variable with respect to other variable is straight line.

2) **Curvilinear** (Non**-linear-):** Consider another situation. first. with increase of one variable, the second variable increases proportionately up to some point: after that with an increase in the first variable the second variable starts decreasing. The graphical representation of the two variables will be a curved line. Such a relationship between the two variables is termed as the curvilinear correlation.

3. Simple, Multiple and Partial Correlations:

l) **Simple Correlation:** The correlation existing between two variables or data series is said to be simple correlation. Of the two series one which causes change in the other is called independent or subject series and the other which is affected is called dependent series.

2) **Multiple Correlations:** When the correlation involves three or more variables or data series that is called multiple correlation. Out of several correlated series, one is dependent series and the remaining two or more series are independent or subject series which jointly affect the dependent series.

3) **Partial Correlation:** When the correlation exists between two or more variables at a time, the correlation is considered between any two of them keeping the interest of other series constant, such a correlation is called partial correlation.

3.3. Importance of Correlation

Correlation is a fundamental concept in statistics that helps us understand the **relationship between two variables**. Its importance is widely seen in economics, business, and research. Correlation shows whether variables move **together or in opposite directions** and helps identify **positive, negative, or zero relationship**

Correlation is very important in the field of Psychology and Education as a measure of relationship between test scores and other measures of performance. With the help of correlation it is possible to have a correct idea of the working capacity of a person. With the help of it, it is also possible to have knowledge of the various qualities of an individual.

After finding the correlation between the two qualities or different qualities of an individual, it is also possible to provide his vocational guidance. In order to provide educational guidance to a student in selection of his subjects of study, correlation is also helpful and necessary.

The use of coefficient of correlation is very wide. In almost all social sciences this is now being increasingly used. It is useful whenever each number of a group of individuals has been measured on two or more traits. It expresses at one figure the average degree of resemblance etc. between the two traits. Its specific uses may be given as follows:

1. **Prognosis (Prediction) :** The coefficient of correlation is used quite profitably in Prediction. In a number of studies it is used to predict the success one will achieve in his further educational careers.
2. **Reliability :** The co-efficient of correlation has been used very often to test the reliability. Through calculation of this statistics it has been sought to be asserted whether or not a test measures on two successive occasions the same type of thing.
3. **Validity :** A test's width value can be obtained through correlation. Whenever a test is constructed the tests, not what it claims to test. This question is answered by the magnitudes of the coefficient with various criteria.
4. **Test Construction :** The coefficient of correlation is also being used in the test construction. Whenever a new test is constructed, there is always the question of whether each element of the test is related to other elements or to the test as a whole and as to whether each element is related to the criteria chosen. Those relationships are all examined through the technique of correlation.

3.4. Methods of Calculating Correlation

Correlation is a statistical concept that shows the **degree and direction of relationship between two variables**. Correlation shows the strength of a relationship between two variables and is expressed numerically by the correlation coefficient. To measure the degree of association or relationship between two variables quantitatively an index of relationship is used and is termed as co-efficient of correlation. Co-efficient of correlation is a numerical index that tells us to what extent the two variables are related and to what extent the variations in one variable changes with the variations in the other.

The following methods are generally used to determine simple correlation:1.Graphic method, 2. Scatter diagram or Dotogram method, 3. Karl Pearson's method, 4. Spearman's ranking method, 5.Concurrent Deviation Method. 6. The Least Square Method.

1. Graphic Method: When the values of independent series are plotted on O -X axis and dependent series are plotted on O-Y axis of graph paper, a linear or non-linear graph will be obtained which will simply indicate the direction of correlation and not the numerical magnitude. If the graph line of two independent series move in upward direction from left to right the correlation is positive but if the graph line of one series moves upward from left to right and that of the other independent series moves downward from left to right, they show negative correlations.

2. Scatter Diagram Method: This method is more or less similar to graphic method. In this method, the values of independent data series are plotted on **O** - X axis and those of dependent series on 0-Y axis and then the pairs of values are plotted on the graph paper. In this ways, graphs of dots are obtained. These dots are scattered in different forms. Therefore, the graphs are called scatter diagrams or dotograms. The patterns of scatter diagrams indicate the direction and magnitude of correlation.

3. Karl Pearson's Coefficient of Correlation: This is the best mathematical method of determining the correlation. Coefficient of correlation (r) is obtained by dividing the product of values of covariance of the two series by the product of their standard deviations. The value of (r) is between +1 and -1. If r = +1 then there is a perfectly positive correlation between 2 variables. If r = -1 then there is an absolute negative correlation between 2 variables. r = 0 so that there is no correlation between the two variables.

4. Spearman's Rank Correlation Coefficient: Prof. Charles Spearman has propounded a simple method for finding correlation in individual word units and this method is based on the serial position of the words. This method is known as rank difference or serial and absence method etc. This method is suitable for finding correlation in individual categories in such situations where it is possible to determine only order instead of numerical measurement. Example: Qualitative facts like intelligence, beauty etc. In such a situation, we give them order like first, second, third, fourth, fifth, sixth, seventh etc. and find out correlation coefficient on their basis and it is expressed with the symbol ρ (Rho).

5. The Concurrent Deviation Method: This method is a simple method of measuring correlation based on the direction of change (increase or decrease) in two variables over time. It does not use actual values, but only checks whether variables move: **Upward** or **Downward.**

6. The Least Square Method: This Method is a statistical technique used to find the best-fitting line (trend line or regression line) for a given set of data.The "best" line is the one that minimizes the sum of the squares of the errors (differences) between actual values and estimated values. The Least Squares Method is a fundamental statistical tool used to determine the most accurate relationship between variables, forming the basis of regression and forecasting techniques.

3.5. Karl Pearson's Correlation Co-efficient

Karl Pearson's Correlation Coefficient is a statistical measure that shows the degree (strength) and direction of the linear relationship between two variables. It is also called as Product Moment Correlation Coefficient, which is denoted by 'r'. This is the best mathematical method of determining the correlation. Coefficient of correlation (r) is obtained by dividing the product of values of covariance of the two series by the product of their standard deviations. The value of (r) is between +1 and -1. If r = +1 then there is a perfectly positive correlation between 2 variables. If r = -1 then there is an absolute negative correlation between 2 variables. r = 0 so that there is no correlation between the two variables. Karl Pearson Calculating the correlation coefficient.

$$r = \frac{\sum(dx,dy)}{N\sigma x \,.\sigma y}$$

Where

r = Karl Pearson correlation coefficient;

dx = deviations of X variable $(X-\bar{X})$

dy = deviations of Y variable $(Y-\bar{Y})$

$\sum dx.\,dy$ = Multiply the deviations of the X and Y variables $(d_x.d_y)$, and the sum of them

N = number of items

σx = Standard deviation of the X variable

σY = Standard deviation of the Y variable

Steps:

1) First calculate the Arithmetic mean of the variables X, Y.

$$\bar{X} = \frac{\sum X}{N} \quad , \bar{Y} = \frac{\sum Y}{N}$$

2) Subtract the mean value of X from the variable values of X, and find the deviations of X i.e. dx. $(dx = X-\bar{X})$

3) Subtract the mean value of Y from the variable values of Y, and find the deviations of Y i.e. dy.

$(dy = Y-\bar{Y})$

4) The deviations of the X and Y variables i.e. dx and dy should be squared separately. (dx^2, dy^2). Write them separately as $\sum dx^2$, $\sum dy^2$

5) Multiply the x and y deviations by each other and write the sum as $\sum dx.\,dy$

6) $\sqrt{\frac{\sum dx^2}{N}}$ Applying the formula, find out the standard deviation of X.

7) $\sqrt{\frac{\sum dy^2}{N}}$ Applying the formula, find out the standard deviation of Y.

8) $r = \frac{\sum dx.dy}{N.\sigma x\, \sigma y.}$ Now calculate Karl Pearson correlation coefficient applying the formula

Example: Calculate the Karl Pearson Correlation Coefficient based on the following data

X : 7, 6, 5, 4, 3, 2, 1

Y : 18, 16, 14, 12, 10, 8, 6

X	Y	dx $X-\bar{X}$ (X-4)	dx^2	dy $Y-\bar{Y}$ (Y-12)	dy^2	dx.dy
7	18	7-4=3	9	18-12=6	36	3x6=18
6	16	6-4=2	4	16-12=4	16	2x4=8
5	14	5-4=1	1	14-12=2	4	1x2=2
4	12	4-4=0	0	12-12=0	0	0x0=0
3	10	3-4=-1	1	10-12=-2	4	1x2=2
2	8	2-4=-2	4	8-12=-4	16	2x4=8
1	6	1-4=-3	9	6-12=-6	36	3x6=18
$\sum X = 28$	$\sum Y = 84$		$\sum dx^2 = 28$		$\sum dy^2 = 112$	$\sum dx.dy = 56$

The arithmetic mean of X = $(\bar{X}) = \frac{\sum X}{N} = \frac{28}{7} = 4$

The arithmetic mean of Y = $(\bar{Y}) = \frac{\sum Y}{N} = \frac{84}{7} = 12$

The standard deviation of the X variable = $(\sigma X) = \sqrt{\frac{\sum dx^2}{N}} = \sqrt{\frac{28}{7}} = \sqrt{4} = 2$

The standard deviation of the X variable $= (\sigma X) = 2$

The standard deviation of the Y variable = $(\sigma Y) = \sqrt{\frac{\sum dy^2}{N}} = \sqrt{\frac{112}{7}} = \sqrt{16} = 4$

The standard deviation of the Y variable $(\sigma Y) = 4$

$$r = \frac{\sum dx.dy}{N.\sigma x\, \sigma y.} = \frac{56}{7x2x4} = \frac{56}{56} = 1$$

r = 1 means that, there is a perfect positive correlation between the X and Y variables.

3.6. Spearman's Rank Correlation

Prof. Charles Spearman has propounded a simple method for finding correlation in individual units and this method is based on the serial position of the units. This method is known as rank difference or serial and absence method etc. This Correlation Coefficient is a method of measuring the degree and direction of relationship between two variables using ranks instead of actual values.This method is suitable for finding correlation in individual categories in such situations where it is possible to determine only order instead of numerical measurement. Example: Qualitative facts like intelligence, beauty etc. In such a situation, we give them order like first, second, third, fourth, fifth, sixth, seventh etc. and find out correlation coefficient on their basis and it is expressed with the symbol ρ (Rho).

$$p = 1 - \frac{6\sum d2}{N(N2-1)}$$

D = difference between the ranks of two variables ($D = R_1 - R_2$)

R_1 = the rank of the first variable

R_2 = the rank of the second variable

N = number of items

Steps :

1. Subtract the second variable ranks (R_2) from the first variable ranks (R_1). It should be written as D ($D=R_1-R_2$).
2. D should be squared (D_2)
3. $p = 1 - \frac{6\sum d2}{N(N2-1)}$ formula for the Spearman rank of the correlation coefficient

Example : Calculate the correlation coefficient for Spearman rank based on the following data.

Ranks in Economics

(R1) : 6, 1, 5, 2, 4, 3

Ranks in history

(R2) : 3, 1, 4, 2, 5, 6

R_1	R_2	$D = R_1 - R_2$	D^2
6	3	6-3=3	9
1	1	1-1=0	0
5	4	5-4=1	1
2	2	2-2=0	0
4	5	4-5=-1	1
3	6	3-6=-3	9
N = 6			$\sum D^2 = 20$

$\rho = 1 - \frac{6\sum D^2}{N^3-N}$ = Formula for the Spearman rank of the correlation coefficient

$\rho = 1 - \frac{6 \times 20}{6^3-6}$ $\rho = 1 - \frac{6 \times 20}{(6 \times 6 \times 6) - 6}$

$\rho = 1 - \frac{120}{216-6}$ $\rho = 1 - \frac{120}{210}$ = 1- 0.57 = 0.43

Spearman rank of the correlation coefficient (ρ) = 0.43

3.7 The Concurrent Deviation Method

The **Concurrent Deviation Method** is a **simple and rough method** of measuring correlation based on the **direction of change** (increase or decrease) in two variables over time**.** A very simple and casual method of finding correlation when we are not serious about the magnitude of the two variables is the application of concurrent deviations. This method involves in attaching a positive sign for x-value (except the first) if this value is more than the previous value and assigning a negative value if this value is less than the previous value. This is done for the y-series as well. The deviation in the x-value and the corresponding y-value is known to be concurrent if both the deviations have the same sign. Denoting the number of concurrent deviation by c and total number of deviations as n (which must be one less than the number of pairs of x and y values) the coefficient of concurrent-deviations is given by

$r = \pm\sqrt{\frac{2C - n}{n}}$

If (2c n) > 0, then we take the positive sign both inside and outside the radical sign and if (2c-n) < 0 we are to consider the negative sign both inside and outside the radical sign. Like Pearson's correlation coefficient and Spearman's rank correlation coefficient the coefficient of concurrent- deviations also lies between - I and 1, both inclusive.

That is, $-1 \leq r \leq 1$

Concurrent Deviation Method

If both variables move in the same direction → Concurrent deviation

Year	Price	Sign of deviation from the previous figure (a)	Demand	Sign of deviation from the previous figure (b)	Product of deviations (ab)
1990	25		35		
1991	28	+	34	-	-
1992	30	+	35	+	+
1993	23	-	30	-	+
1994	35	+	29	-	-
1995	38	+	28	-	-
1996	39	+	26	-	-
1997	42	+	23	-	-

If they move in **opposite directions** → **Non-concurrent deviation**
We count how many times movements are **same (C)**

$r = \pm\sqrt{\frac{2C - n}{n}}$ **Where:**

r = coefficient of

Correlation

C = number of concurrent deviations
n = number of comparisons (pairs of changes)
+ if general trend is same
− if general trend is opposite

Steps to Calculate

1. Arrange data in time order
2. Find direction of change (↑ or ↓)
3. Compare directions of both variables
4. Count concurrent cases (C)
5. Apply formula

Example:

Year	X	Y	Change in X	Change in Y	Concurrent
1	10	20	--------	--------	--------
2	15	25	+	+	Yes
3	12	22	-	-	Yes
4	18	20	+	-	No
5	20	25	+	+	Yes

Total comparisons (n = 4)
Concurrent deviations (C = 3)

$$r = \pm\sqrt{\frac{2C - n}{n}}$$ r $$r = \sqrt{\frac{2(3) - 4}{4}} = \sqrt{\frac{6 - 4}{4}} = \sqrt{\frac{2}{4}} = \sqrt{0.5} \approx 0.707$$

Concurrent Deviation Method - Advantages

1. Very simple and quick
2. No complex calculations
3. Useful for small or rough data

Concurrent Deviation Method - Limitations

1. Not accurate (approximate only)
2. Ignores actual values
3. Not suitable for detailed statistical analysis

The Concurrent Deviation Method is best for getting a quick idea of correlation direction, especially in time series data, but not for precise measurement.

3.8. The Least Squares Method

The least square method is the process of finding the best-fitting curve or line of best fit for a set of data points by reducing the sum of the squares of the offsets (residual part) of the points from the curve. During the process of finding the relation between two variables, the trend of outcomes is estimated quantitatively. This process is termed as regression analysis. The method of curve fitting is an approach to regression analysis. This method of fitting equations which approximates the curves to given raw data is the least squares. It is quite obvious that the fitting of curves for a particular data set are not always unique. Thus, it is required to find a curve having a minimal deviation from all the measured data points. This is known as the best-fitting curve and is found by using the least-squares method. The least-square method states that the curve that best fits a given set of observations, is said to be a curve having a minimum sum of the squared The equation of least square line is given by Y = a+ bX **Where:** 1) y = dependent variable, 2) x = independent variable, 3. a = intercept, 4.b = slope

Example :

X	Y	X^2	XY
1	4	1	4
2	2	4	4
3	5	9	15
4	7	16	28
$\Sigma X = 10$	$\Sigma Y = 18$	$\Sigma X^2 = 30$	$\Sigma XY = 51$

Solving these two normal equations we can get the required trend line equation. Thus we can get the line of best fit with formula Y = a + bX.

Normal equation for 'a'.

$\Sigma Y = na + b\Sigma X$ - 1

Normal equation for 'b':

$\Sigma XY = a\Sigma X + b\Sigma X^2$ - 2

From the given example the values are:

$\Sigma X = 10$, $\Sigma Y = 18$, $\Sigma X^2 = 30$, $\Sigma XY = 51$ and n=4

Substitute the values in the equations : $18 = 4a + 10b$ (1)

$51 = 10a + 30b$ (2)

Formula (1) multiply by 5 $\Rightarrow$ $4a + 10b = 18$

$20a + 50b = 90$

Formula (2) multiply by 2 $\Rightarrow$ $10a + 30b = 51$

$20a + 60b = 102$

Subtract equation 1 from 2 $\Rightarrow$ $(20a - 20a) + (60b - 50b) = 102 - 90$

$10b = 12$ then $b = 12/10 = 1.2$

To find out 'a', substitute b value in the 1 formula

$4a + 10 (1.2) = 18$

$4a + 12 = 18$

$4a = 18 - 12$ then $4a = 6$, $a = 3/2 = 1.5$

Substitute the values of a and b in the equation $Y = a + bX$

$Y = 1.5 + 1.2x$

3.9. Meaning and Types of Regression

Regression is a statistical technique used to study the **relationship between two or more variables.** It helps us understand how one variable changes when another variable changes. The variable we **predict or explain** is called the **dependent variable (Y).** The variable used to **predict** is called the **independent variable (X)** In simple words: **Regression tells us how Y depends on X.**

Regression is a statistical method used in finance, investing and other disciplines that attempts to determine the strength and character of the relationship between one dependent variable and a series of other variables, known as independent variables. This method was first used in 1877 by Sir Francis Galton, a British biologist. He used to estimate the relationship between father and son height.

Types of Regression :

Regression can be classified in different ways depending on the **number of variables** and the **nature of relationship**.

1. Based on Number of Variables :

(1) Simple Regression: This Regression involves one independent variable (X) and one dependent variable (Y). It shows how Y depends on a single X

Example: Marks (Y) based on Study Hours (X)

(2) Multiple Regression: This Regression **involves** two or more independent variables**. It helps in analyzing the effect of** many factors together

Example: Marks (Y) based on: Study Hours (X1), Attendance (X2), Sleep (X3)

2. Based on Nature of Relationship :

(1) Linear Regression: Relationship between variables is a straight line

Equation form: **Y = a + bX**(2) Non-Linear Regression: **Relationship is curved (not a straight line). It is used when data does not follow a straight-line pattern**

3. Based on Direction of Relationship:

(1) Positive Regression: Both variables move in the same direction**.**

If X increases → Y also increases

(2) Negative Regression: Variables move in opposite directions**.**

If X increases → Y decreases

Understanding the types of regression helps in choosing the **correct model** for analysis and improves the accuracy of **prediction and interpretation** in statistics.

3.10. Uses of Regression Analysis

Regression is one of the most important statistical tools used for **analysis, prediction, and decision-making** in economics, business, and research.

1) **Forecasting and Prediction:** Regression helps to predict future values based on past data. For Examples: 1.Predicting sales based on past trends, 2. Estimating demand for a product, 3. Forecasting GDP or income levels.

2) **Measuring Relationship between Variables:** Regression shows how one variable depends on another variable**.** For Examples: 1. Relationship between price and demand, 2. Relationship between income and consumption**.** It helps in understanding cause and effect relationships.

3) **Business Decision Making:** Managers use regression to make better decisions**.** For Examples: 1.Deciding pricing strategies, 2. Planning production levels, 3. Budget allocation

4) **Economic Analysis:** Regression is widely used in economics to study relationships between economic variables. For Examples: 1.Income verses Consumption, 2. Inflation verses Unemployment, 3. Investment verses Interest rates

5) **Estimation of Unknown Values:** Regression helps to estimate unknown or missing values For Example: If we know study hours, we can estimate expected marks or If we know rainfall, we can estimate crop yield.

6) **Trend Analysis:** Regression helps to identify long-term trends in data**. For** Examples: 1.Population growth trends, 2. Market trends, 3. Stock price movements.

7) **Scientific and Research Applications:** Regression is used in scientific studies and research work**.** For Examples: 1Medicine**:** Effect of dosage on recovery**,** 2. Agriculture: Fertilizer vs crop yield, 3. Psychology: Behavior analysis

8) **Risk Analysis and Control:** Regression helps in analyzing risk factors and taking preventive measures. For Examples: 1.Financial risk prediction**,** 2. Credit scoring in banks, 3. Insurance risk assessment

9) **Performance Evaluation:** Used to measure and evaluate performance. For Examples: 1. Employee productivity vs. training, 2. Student performance vs study time

10) **Policy Formulation:** Governments use regression for planning and policy decisions**.** Examples: 1.Tax policy decisions, 2.Employment programs, 3.Welfare schemes

Regression is a powerful and versatile statistical tool that plays a key role in understanding relationships, making predictions, and supporting decisions across various fields**.**

3.11. Simple Regression

Simple Regression studies the relationship between one independent variable (X) and one dependent variable (Y). It shows how Y changes when X changes.

X → Independent variable

Y → Dependent variable

The equation of simple linear regression is :

$Y = a + bX$

Where:

a = Intercept (value of Y when X = 0)

b = Slope (rate of change of Y with respect to X)

Numerical Example :

X (Hours)	Y (Marks)
2	25
4	45
6	65
8	85

Step 1: Calculate Means of X and Y variables

$$\bar{X} = \frac{2+4+6+8}{4} = \frac{20}{4} = 5$$

$$\bar{Y} = \frac{25+45+65+85}{4} = \frac{220}{4} = 55$$

Step 2 Calculate slop $\sum(X-\bar{X})(Y-\bar{Y})$. Use table and find out the values to know the slop value b and a interception.

X Hours	Y Marks	$X - \bar{X}$	$Y - \bar{Y}$	$(X-\bar{X})(Y-\bar{Y})$	$(X - \bar{X})^2$
2	25	-3	-30	90	9
4	45	-1	-10	10	1
6	65	1	10	10	1
8	85	3	30	90	9
$\sum X = 20$	$\sum Y = 220$			$\sum 200$	$\sum 20$

$\sum(X - \bar{X})(Y - \bar{Y}) = 200$, $\sum(X - \bar{X})^2 = 20$, Slope (b) = 200/20 = 10, b = 10

Interception (a) = $\bar{Y} - b\bar{X}$

Substitute the value of 'b' in the equation

$a = 55 - 10(5) = 55 - 50 = 5$

a=5

For example X is 7 hours Y =?

Y = 5 + 10 (7)

Y= 75 (Marks)

3.12. Multiple Regressions

Multiple Regression is a statistical method used to study the relationship between **one dependent variable (Y)** and **two or more independent variables (X_1 , X_2 , X_3 X_n)**. It explains how **several factors together influence one outcome**. For **example:** Marks depend not only on study hours but also on attendance, sleep, etc.

General Equation of Multiple Regression

$$Y = a + b_1 X_1 + b_2 X_2 + b_3 X_3 + ... + b_n X_n$$

Where:

Y = Dependent variable

X_1 , X_2 , X_3 .....X_n = Independent variables

a = Intercept

b_1, b_2, b_n .....X_n = Regression coefficients (impact of each variable). Each coefficient shows **how much Y changes when that variable changes**, keeping others constant.

A student's marks (Y) depend on: Study Hours (X_1) and Attendance (%) (X_2) Assume the estimated regression equation is:

$$Y=10 + 5X_1 + 0.3 X_2$$

Step 1: Given Values

Study Hours (X_1) = 6

Attendance (X_2) = 80

Step 2: Substitute Values

$Y = 10 + 5(6) + 0.3 (80)$

$Y = 10 + 30 + 24$

$Y = 64$

Predicted Marks = 64

Interpretation of Coefficients

b_1 = 5 → If study hours increase by 1, marks increase by **5 marks** (keeping attendance constant)

b_2 = 0.3 → If attendance increases by 1%, marks increase by **0.3 marks**

Uses of Multiple Regression :

1) **Business:** Sales based on price, advertising, income
2) **Economics:** Demand based on price and income
3) **Education:** Performance based on study, attendance, environment
4) **Agriculture:** Yield based on rainfall, fertilizer, temperature

Limitation :

1) Calculation is **more complex** than simple regression
2) Requires **more data**

Multiple regression is an advanced and powerful tool that helps in understanding how **several variables jointly affect a single outcome**, making it very useful for **prediction, planning, and analysis**.

3.13 Model Examination Questions

Short Type Questions :

1. Meaning of Correlation
2. Positive correlation
3. Negative correlation
4. Linear correlation
5. Non-linear correlation
6. Simple correlation
7. Multiple correlations
8. Partial correlation
9. Scatter diagram method
10. Karl Pearson co-efficient correlation
11. Spearman rank correlation co-efficient
12. Concurrent deviation method
13. Uses of Regression Analysis

Essay Type Questions :

1. Discuss concept of correlation in detail.
2. Define correlation and explain its importance
3. Describe various types of correlation
4. Explain the methods of studying or calculation correlation
5. Define regression and explain its importance and limitations

Author

Dr. Chanda Appa Rao, M.A., Ph.D. B.Ed,

Associate Professor of Economics, K.R.R. Govt. Arts & Science College (A), Kodad

Module

4



Index Numbers

4.1.	Meaning of Index numbers	79
4.2.	Features of Index Numbers.	79
4.3.	Construction of Price Index Numbers	80
4.4.	Difficulties in Constructing of Price Index	82
4.5.	Uses of Index Numbers	83
4.6.	Types of Index Numbers	84
4.7.	Laspeyres Price Index Method	85
4.8.	Paasche's Price Index Method	86
4.9.	Fisher Ideal index number	87
4.10.	Concept of Consumer Price Index (CPI)	88
4.11.	Wholesale Price Index (WPI)	89
4.12.	Index of Industrial Production (IIP)	90
4.13	Model Examination Questions	91

4.1. Meaning of Index Numbers

An index number is a method of evaluating variations in a variable or group of variables in regards to the geographical location, time and other features. The base value of the index number is usually l00 and indicates either to price, date, a level of production, etc. There are various kinds of index numbers however in present; the most relatable is price index numbers which particularly indicates the changes in overall price level for a particular time. Here, the value of money is not constant: even if it falls or rises it will affect and change the price level. An increase in the price level determines a decline in the value of money and a decline in the price level means an increase in the value of money. Therefore, the differences in the value of money are indicated by the differences in the overall price level for a particular time. Therefore changes in the overall prices can be evaluated by a statistical device known as 'index number'.

Price index number indicates the average of changes in the prices of representative commodities at one time in comparison with that at some other time taken as the base period. According to L.V. Lester, ·'An index number of prices is a figure showing the height of average prices at one time relative to their height at some other time which is taken as the base period."

4.2. Features of Index Numbers

1. Index numbers are a special type of average. Whereas mean, median and mode measure the absolute changes and are used to compare only those series which are expressed in the same units the technique of index numbers is used to measure the relative changes in the level of a phenomenon where the measurement of absolute change is not possible and the series are expressed in different types of items.

2. Index numbers are meant to study the changes in the effects of such factors which cannot be measured directly. For example the general price level is an imaginary concept and is not capable of direct measurement. But, through the technique of index numbers, it is possible to have an idea of relative changes in the general level of prices by measuring relative changes in the price level of different commodities.

3. The technique of index numbers measures changes in one variable or group of related variables. For example, one variable can be the price of wheat, and group of variables can be the price of sugar, the price of milk and the price of rice.

4. The technique of index numbers is used to compare the levels of a phenomenon on a certain date with its level on some previous date (e.g., the price level in 2025 as compared to that in 2020 taken as the base year) or the levels of a phenomenon at different places on the same date (e.g. the price level in India in 2025 in comparison with that in other countries in 2025.

4.3. Construction of Price Index Numbers

Construction of price index numbers through various methods can be understood with the help of the following examples:

1. Simple Aggregative Method: In this method, the index number is equal to the sum of prices for th year for which index number is to be found divided by the sum of actual prices for the base year.

The formula for finding the index number through this method is as follows:

Where $$\mathbf{P = \frac{\sum P1}{\sum P0} \times 100}$$

P: Stands for the Index Number

∑P1 : Stands for the sum of the prices for the year which index number is to be found

∑P0 : Stands for the sum of prices for the base year

Commodity Type	**Prices in base year** 2020 (P_0)	**Prices in current year** 2025 ($\mathbf{P_1}$)
A	10	20
B	15	25
C	40	60
D	25	40
	$\sum P_0 = 90$	$\sum P1 = 145$

Index Number = $\mathbf{P = \frac{\sum P1}{\sum P0} \times 100 = \frac{145}{90} \times 100 = 161.11}$

2. Simple Average of Price Relatives Method: In this method the index number is equal to the sum of price relatives divided by the number of items and is calculated by using the following formula:

$$P = \frac{\sum R}{N}$$

Where

∑R: Stands for the sum of price relatives i.e. $R = \frac{P1}{P0} \times 100$

N: Stands for the number of Items

Commodity Type	Price in the base year 2020 (P_0)	Price in the current year 2025 (P_1)	Price Relatives $R=\frac{P1}{P0} \times 100$
A	10	20	$\frac{20}{10} \times 100 = 200.00$
B	15	25	$\frac{25}{15} \times 100 = 166.67$
C	40	60	$\frac{60}{40} \times 100 = 150.00$
D	25	40	$\frac{40}{25} \times 100 = 160.00$
	$\sum P_0 = 90$	$\sum P_1 = 145$	$\sum = R = 676.7$

Price Index = $\mathbf{P = \frac{\sum R}{N} = \frac{676.7}{4} = 169.2}$

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

4. Weighted Average of Relatives Method: In this method also different weights are used for the items according to their relative importance. Weighted average is a calculation that takes into account the varying degrees of importance of the numbers in a data set. In calculating a weighted average, each number in the data set is multiplied by a predetermined weight before the final calculation is made. **A** weighted average can be more accurate than a simple average in which all numbers in a data set are assigned an identical weight.

The price index number is found out with the help of the following formula:

Where ∑W stands for the sum of weights of different commodities

∑R stands for the sum of price relatives

P_{01} = Price Index Number of the current year (1) with respect to the base year (0)

Commodity Type	Weights W	Price in the base year 2020 (P_0)	Price in the current year 2025(P_1)	Price Relatives $R=\frac{P1}{P0}$ x 100	RW
A	5	10	20	$\frac{20}{10}$ x 100 = 200.00	1000.0
B	4	15	25	$\frac{25}{15}$ x 100 = 166.67	666.7
C	2	40	60	$\frac{60}{40}$ x 100 = 150.00	300.0
D	3	25	40	$\frac{40}{25}$ x 100 = 160.00	480.0
	∑W=14	∑ P_0 = 90	∑P_1 = 145	∑ =R = 676.7	∑RW = 2446.7

$P_{01}= \frac{\sum RW}{\sum W} = \frac{2446.7}{14} = 174.8$

4.4 Difficulties in Constructing of Price Index

Measurement of changes in the value of money through price index number is not an easy and reliable technique. There are a number of theoretical as well as practical difficulties in the construction of price index numbers. Moreover, the index number technique itself has many limitations.

1. **1. Selection of Base Year:** While preparing the index number, first difficulty arises regarding the selection of base year. The base year should be a normal year. But, it is very difficult to find out a fully normal year free from any unusual happening. There is every possibility that the selected base year may be an abnormal year or a distant year, or may be selected by an immature or a biased person
2. **Selection of Items:** The selection of the representative commodities is the second difficulty in the construction of index numbers:
3. **Collection of Prices:** It is also difficult to obtain correct, adequate and representative data regarding prices. It is not an easy job to select representative places from which the information about prices to be collected and to select the experienced and unbiased individuals or institutions who will supply price quotations.
4. **Assigning Weights:** Another important difficulty that arises in preparing the index numbers is that of assigning proper weights to different items in order to arrive at correct and unbiased conclusions. As there are no hard and fast rules to weights for the commodities according to their relative importance, there is very likelihood that the weights are decided arbitrarily on the basis of personal judgment and involve biasness.
5. **Selection of Averages:** Another major problem is that which average should be employed to find out the price relatives. There are many types of averages such as arithmetic average, geometric average, mean, median, mode, etc. The use of different averages gives different results. Therefore, it is essential to select the method with great care. Alfred Marshall has advocated the use of chain index number to solve the problem of averaging and weighing.
6. **Problem of Dynamic Changes:** In the dynamic world the consumption pattern of the individuals and the number and varieties of goods undergo continuous changes. Since, in the course of time, old commodities may disappear and many new ones come into existence, the long-run comparison may become difficult. The quantity and quality of commodities may also change over the period of time thus making the choice of commodities for constructing index numbers difficult. A number of factors, like income, education, fashion, etc., bring changes in the consumption pattern of the people which render the index numbers incomparable.

4.5 Uses of Index Numbers

Index numbers are widely used in **economics, business, and government policy** to measure changes over time. These are used to measure all types of quantitative changes in different fields.

1. **Measure Price Changes:** Index numbers are used to measure changes in the value of money or the price level from time to time. Changes in the price level generally influence production and employment of the country as well as various sections of the society. The price index numbers also forewarn about the future inflationary tendencies and in this way, enable the government to take appropriate anti- inflationary measures.
2. **Changes in Cost of Living:** Index numbers highlight changes in the cost of living in the country. They indicate whether the cost of living of the people is rising or falling. On the basis of this information, the wages of the workers can be adjusted accordingly to save the wage earners from the hardships of inflation.
3. **Changes in Production:** Index numbers are also useful in providing information regarding production trends in different sectors of the economy. They help in assessing the actual condition of different industries, i.e., whether production in a particular industry is increasing or decreasing or is constant
4. **Importance in Trade:** Importance in trade with the help of index numbers, knowledge about the trade conditions and trade trends can be obtained. The import and export indices show whether foreign trade of the country is increasing or decreasing and whether the balance of trade is favourable or unfavourable.
5. **Formation of Economic Policy:** "Index numbers are very useful to the government in formulating as well as evaluating economic policies." Index numbers measure changes in the economic conditions and, with this information help the planners to formulate appropriate economic policies. Further, whether particular economic policy is good or bad is also judged by index numbers.
6. **Comparison over Time and Place:** Compare prices across different years, and prices between different regions.
7. **Useful in All Fields:** Index numbers are useful in almost all the fields. They are especially important in economic field in analyzing markets for specific commodities in the share market. With the help of index numbers, the Railways can get information about the changes in goods traffic. The bankers can get information about the changes in deposits by means of index numbers.

4.6. Types of Index Numbers

Index numbers are classified based on **what they measure**. The three main types are: 1. Price Index Numbers, Quantity Index Numbers and 3. Value Index Numbers, 4.Special Index Numbers

1. Price Index Numbers: To measure **inflation and cost of living Price Index Numbers are used. They are: 1.**Wholesale Price Index, 2. Consumer Price Index.3. Working Class cost of Living Index, 4. Wage Index

1) Wholesale Price Index: Wholesale price index numbers are constructed on the basis of the wholesale prices of certain important commodities. The commodities included in preparing these index numbers are mainly raw-materials and semi-finished goods. The wholesale price index numbers are generally used to measure changes in the value of money.

2) Consumer Price Index : The **Consumer Price Index (CPI)** is an index number that measures the **average change in prices of goods and services consumed by households** over time. It is also called **Cost of Living Index.** CPI is based on a basket of goods and services like: Food, Clothing, Housing, Transport and Education. It shows how much the cost of living changes over time.

2. Quantity Index Numbers: This index Measure changes in quantity (volume) of good produced, consumed, or sold. They are: 1.Index of Industrial Production (IIP) 2. Agricultural production index

1). Industrial Index Numbers: Industrial index numbers are constructed with an objective of measuring changes in the industrial production. The production data of various industries are included in preparing these index numbers.

2) Agricultural production index: The **Agricultural Production Index** is a **quantity index number** that measures the **change in agricultural output (production of crops)** over time compared to a base year. It shows whether agricultural production has **increased or decreased**.

3. Value Index Numbers: This index measure **changes in total value (Price × Quantity).**

It measure **total sales or revenue changes.**

Value Index $= \frac{\sum P_1 Q_1}{\sum P_0 Q_0} \times 100$

4. Special Index Numbers: This Index measures Living Expenses, Changes in Wages and Levels of Employment

1) Cost of Living Index: The working class cost-of-living index numbers aim at measuring changes in the cost of living of workers. These index numbers are consumed on the basis of only those goods and services which are generally consumed by the working class. The prices of these goods and index numbers are of great importance to the workers because their wages are adjusted according to these indices.

2) Wage Index: The purpose of these index numbers is to measure time to time changes in money wages. These index numbers, when compared with the working class cost-of-living index numbers, provide information regarding the changes in the real wages of the workers.

3) Employment Index: The **Employment Index is a quantity index number** that measures the **change in the level of employment (number of workers employed)** over time compared to a base year. It shows whether employment has **increased or decreased. It measures** number of people employed in: Industries, Agriculture and Services and helps understand job trends in the economy.

4.7. Laspeyres – Price Index Method

The Laspeyres Price Index is a consumer price index used to measure the change in the prices of a basket of goods and services relative to a specified base period weighting. Developed by German economist Etienne Laspeyres. The Laspeyres Price Index is also called the base year quantity weighted method. The Laspeyres Price Index is a price index used to measure the economy's general price level and cost of living, and to calculate inflation. The index commonly uses a base year figure of 100 with periods of higher price levels shown by an index greater than 100 and periods of lower price levels by indexes lower than 100.

The **Laspeyres Price Index** measures price changes using **base year quantities as weights**. It answers:

"How much would it cost today to buy the same basket as in the base year?"

The formula for the Laspeyres Price Index is as follows:

Where:

P_1 = Current year prices

P_0 = Base year prices

Q_0 = Base year quantities

Quantities ($\mathbf{Q_0}$) remain **fixed** (base year). Only prices change from $\mathbf{P_0} \rightarrow \mathbf{P_1}$•

Steps of Calculation :

1. Take **base year quantities ($\mathbf{Q_0}$)**
2. Collect **base year prices ($\mathbf{P_0}$) and current year prices ($\mathbf{P_1}$)**
3. Calculate : P_0Q_0 and P_1Q_0
4. Find totals : ΣP_0Q_0 and ΣP_1Q_0
5. Apply formula

$$P_{01} = \frac{\sum P_1Q_0}{\sum P_0Q_0} \times 100$$

Commodity	P_0	P_1	Q_0	$P_0.Q_0$	P_1Q_0
Rice	10	12	5	50	60
Wheat	20	25	3	60	75
Milk	30	33	2	60	66
				Σ 170	Σ 201

Substitute the values in the equation= $\frac{201}{170}$ x100 = 118.24

Index = 118.24

Prices increased by **18.24%** compared to base year

$$\frac{\text{Sum of Price at Observation Period X Base Quantity X 100}}{\text{Sum of the Price at Base Period X Base Quantity}}$$

4.8. Paasche's Price Index Method

The Paasche Price Index is a consumer price index used to measure the change in the price and quantity of a basket of goods and services relative to a base year price and observation year quantity. Developed by German economist Hermann Paasche, his Price Index is commonly referred to as the 'current weighted index' The Paasche Price Index is a price index used to measure the general price level and cost of living in the economy and calculate inflation. The index commonly uses a base year of 100 with periods of higher price levels shown by an index greater than 100 and periods of lower price levels by indexes lower than 100..

The **Paasche Price Index** measures price changes using **current year quantities (Q_1) as weights**.

Formula $P_{01} = \frac{\sum P_1 Q_1}{\sum P_0 Q_1} \times 100$

Where :

P_1 = Current year prices

P_0 = Base year prices

Q_1 = Current year quantities

Steps of Calculation

1. Take **current year quantities (Q_1)**
2. Collect **P_0 (base prices)** and **P_1 (current prices)**
3. Calculate: P_0Q_1 and P_1Q_1
4. Find Totals $\sum P_0Q_1$ and $\sum P_1Q_1$
5. Apply formula

Commodity	P0	P1	Q1	P0Q1	P1Q1
Rice	10	12	6	60	72
Wheat	20	25	4	80	100
Milk	30	33	3	90	99
				Σ230	Σ271

Substitute the values in the formula $\frac{271}{230}$ x 100 = 117.83

Index = 117.83

Prices increased by 17.83% compared to the base year

Merits

1) Reflects current consumption pattern
2) More realistic than Laspeyres
3) Useful for modern economic analysis

Demerits

1) Difficult to collect current quantity data
2) Time-consuming
3) May underestimate inflation

4.9. Fisher Ideal Index Number

The Fisher Price Index, also called the Fisher's Ideal Price Index, is a consumer price index (CPI) used to measure the price level of goods and services over a given period. The Fisher Price Index is a geometric average of the Laspeyres Price Index and the Paasche Price Index. It is deemed the "ideal price index' as it corrects the positive price bias in the Laspeyres Price Index and the negative price bias in the Paasche Price Index.

Similar to other consumer price indices, the Fisher Price index is used to measure the price level and cost of living in an economy and to calculate inflation. The index corrects for the upward bias of the Laspeyres Price Index and the down ward bias of the Paasche Price index by taking the geometric average of the two weighted indices. The Fisher Price Index,' is the geometric average of the Laspeyres and Paasche Price indices, and the formula is rendered as:

Fisher Price Index = (Laspeyres Price Index X Paasche Price Index) $^{0.5}$

Fisher's Equation $F = \sqrt{\left(\frac{\sum P_1Q_0}{\sum P_0Q_0}\right) \times \left(\frac{\sum P_1Q_1}{\sum P_0Q_1}\right)} \times 100$

Explanation of Fisher's Ideal Index Number with the help of values given below in the Table

Commodity	P_0	P_1	Q_0	Q_1	$P_0.Q_0$	$P_1.Q_0$	$P_0.Q_1$	$P_1.Q_1$
A	10	12	5	6	50	60	60	72
B	20	25	3	4	60	75	80	100
					∑110	∑135	∑140	∑172

To find out Fisher's Ideal Index substitute above values in the formula.

Find out Laspayer's Price Index = $P_{01} = \frac{\sum P_1Q_0}{\sum P_0Q_0} \times 100$ $\frac{135}{110}$ x 100 = 122.73

Find out Paasche Price Index = $P_{01} = \frac{\sum P_1Q_1}{\sum P_0Q_1} \times 100$ $\frac{172}{140}$ x 100 = 122.86

Fisher Price Index = $F = \sqrt{\left(\frac{\sum P_1Q_0}{\sum P_0Q_0}\right) \times \left(\frac{\sum P_1Q_1}{\sum P_0Q_1}\right)} \times 100$

Substitute the Index numbers values of Laspyer,s and Paasche in the Fisher's equation

Fishers Ideal Index Number = $\sqrt{122.73 \; x \; 122.86}$

= $\sqrt{15078.6}$

= 122.80

4.10. Concept of Consumer Price Index (CPI)

The Consumer Price Index (CPI) is a statistical measure used to track the average change in prices of a fixed basket of goods and services consumed by households over time. In simple words: CPI shows how much the cost of living has increased or decreased for consumers. CPI is an index number that measures the weighted average of prices of a basket of consumer goods and services such as food, clothing, housing, transport, and medical care.

The **Consumer Price Index (CPI)** is an important economic indicator that measures the average change in retail prices of a fixed basket of goods and services consumed by households, thereby reflecting the cost of living and inflation in an economy. In India, the CPI has recently been updated with a new base year of **2024 = 100**, replacing the earlier base year of 2012 = 100, in order to better capture current consumption patterns and economic conditions. The CPI basket consists of a wide range of commodities and services representing both rural and urban consumers, and the number of items has been revised to include more modern goods and services such as digital and communication-related expenses. The index is calculated using a weighted average method, where weights are assigned based on the relative importance of items in consumer expenditure. CPI is widely used for measuring inflation, adjusting wages and dearness allowance, framing government policies, and analyzing purchasing power. Thus, the revised CPI with the 2024 base year provides a more accurate and realistic measure of price changes and cost of living in the present economy.

Types of Consumer Price Index: In India, CPI is calculated for different groups:

1) CPI for Industrial Workers (CPI-IW)
2) CPI for Agricultural Labourers (CPI-AL)
3) CPI for Rural Labourers (CPI-RL)
4) CPI (Combined) – Rural + Urban (most widely used today)

$$CPI = \frac{\sum(P_1 \times Q_0)}{\sum(P_0 \times Q_0)} \times 100$$

Formula of CPI: CPI is generally calculated using the Laspeyres Index Method:

Where: P_1 = Current year prices
P_0 = Base year prices
Q_0 = Base year quantities
It uses base year quantities as weights

Uses of CPI

1) Measures inflation
2) Helps in wage adjustments (DA)
3) Used in government policy decisions
4) Indicates cost of living
5) Helps in economic planning
6) Basis for indexation (tax, pensions)

The Consumer Price Index (CPI) is one of the most important tools in economics for measuring inflation and understanding how price changes affect consumers. It plays a crucial role in policy-making, wage determination, and economic analysis.

4.11. Wholesale Price Index (WPI)

The Wholesale Price Index (WPI) is an important statistical tool used to measure the average change in prices of goods at the wholesale level over a period of time. It reflects price movements of goods that are traded in bulk between producers and wholesalers, before reaching the final consumers. Hence, WPI indicates inflation from the producer's or supply side.

WPI is defined as an index number that shows the weighted average of wholesale prices of a selected basket of commodities. In India, the present base year for WPI is 2011–12 = 100. The index is constructed using a fixed basket of goods, and weights are assigned based on the relative importance of each commodity in total production.

The WPI basket is broadly divided into three main groups: Primary Articles (such as food grains and minerals), Fuel and Power (such as petrol, diesel, and electricity), and Manufactured Products (such as cement, steel, and textiles). Among these, manufactured products carry the highest weight.

The Wholesale Price Index (WPI) is an important measure used to track changes in prices of goods at the wholesale level in an economy. In India, the current WPI series is based on the base year 2011–12 = 100 and includes a total of 697 commodities representing bulk transactions across different sectors. These commodities are classified into three major categories: Primary Articles, Fuel and Power, and Manufactured Products. Primary Articles consist of 117 items with a weight of 22.62%, covering food items, agricultural products, and minerals. Fuel and Power include 16 items with a weight of 13.15%, comprising petroleum products, coal, and electricity. Manufactured Products form the largest group with 564 items and the highest weight of 64.23%, including industrial goods such as cement, steel, textiles, and machinery. The higher weight of manufactured products indicates the dominant role of industrial goods in determining wholesale price movements. Thus, WPI provides a comprehensive picture of price changes in the production sector and serves as an important indicator of inflation in the economy.

Where $$WPI = \frac{\sum(P_1 \times Q_0)}{\sum(P_0 \times Q_0)} \times 100$$

(P_1) represents current prices, (P_0) represents base year prices, and (Q_0) represents base year quantities.

WPI is widely used for measuring **inflation at the wholesale level**, helping the government in **economic policy formulation**, and assisting businesses in planning production and pricing strategies. It also acts as an **early indicator of inflation trends** in the economy.

However, WPI has certain limitations. It excludes services, does not reflect the actual cost of living of consumers, and may not represent price changes experienced by the general public.

In conclusion, WPI is a useful measure of **price changes at the production level**, but it should be used along with CPI for a complete understanding of inflation in the economy.

4.12 Index of Industrial Production (IIP)

The Index of Industrial Production (IIP) is an important economic indicator that measures the level of industrial activity in an economy over a period of time. It reflects the performance and growth of the industrial sector, which includes mining, manufacturing, and electricity. IIP helps in understanding whether industrial production is increasing or decreasing and is widely used for analyzing economic trends. IIP is an index number that measures the relative change in the volume of production of industrial sectors such as mining, manufacturing, and electricity.

Base Year: Current Base Year (India): 2011–12

1) The base year is taken as a reference point for comparison
2) The index value for the base year is always 100

Components of IIP: IIP is classified into three main sectors:

1. **Mining:** Extraction of minerals such as coal, crude oil, iron ore
2. **Manufacturing:** Production of goods like textiles, machinery, cement
   This sector has the highest weight
3. **Electricity:** Generation and distribution of power

Weightage of Sectors: Mining-14.37%, Manufacturing-77.63% and Electricity-7.99%. Manufacturing dominates IIP due to its large contribution to industrial output.

Use-Based Classification: IIP is also classified based on use of goods:

1) Primary goods
2) Capital goods
3) Intermediate goods
4) Infrastructure/Construction goods
5) Consumer durables
6) Consumer non-durables

Method of Calculation: IIP is calculated using the Laspeyres Index Method, where base year weights are used.

Where : $$IIP = \frac{\sum(Q_1 \times W)}{\sum(Q_0 \times W)} \times 100$$

(Q_1) = Current year production
(Q_0) = Base year production
(W) = Weights

Uses of IIP

1) Measures industrial growth
2) Helps in economic planning
3) Used in policy formulation
4) Indicates business cycle trends
5) Helps investors and analyst

The Index of Industrial Production (IIP) is a crucial tool for measuring the growth and performance of the industrial sector. It helps policymakers, economists, and businesses understand economic trends and make informed decisions. However, for a complete picture of the economy, IIP should be used along with other indicators like GDP and CPI.

4.13 Model Examination Questions

Short Type Questions :-

1. Meaning of Index Numbers
2. Price Index Numbers
3. Whole Sale Price Index
4. Consumer Price Index
5. Index of Industrial Production
6. Fisher Ideal Index Number

Essay Type Questions :-

1. Discuss features of index numbers
2. Explain difficulties in constructing of price index
3. Explain types of index numbers
4. Explain Laspeyres price index method
5. Explain Paasche's price index method
6. Explain the concept of Consumer Price Index
7. Describe the Whole Sale Price Index
8. Explain Index of Industrial Production

Author

T. Venkataiah,, M.A, M.Phil, SET (Ph.D.)

Asst Professor of Economics, Govt. Degree College Kollapur

Module

5



Analysis of Time Series

5.1. Meaning of Time Series Analysis	93
5.2. Uses of Time Series Analysis	93
5.3. Components of Time Series Analysis	94
5.4. Secular Trend or Long Run	94
5.5. Seasonal Variations	95
5.6. Cyclical Variations	96
5.7. Irregular Variations or Random Variations	97
5.8. Measurement of Secular Trend	98
5.9. Graphical method or Free Hand Drawing Method	98
5.10. Semi Averages Method	99
5.11. Moving Averages Method	100
5.12. Impact of Covid-19 on the Growth of Service sector in India	101
5.13 Model Examination Questions	103

5.1. Meaning of Time Series Analysis

A Time Series refers to a sequence of data points that are collected, recorded, or observed at successive points of time, usually at equal intervals. These time intervals may be daily, weekly, monthly, quarterly, or yearly. In simple terms, a time series shows how a variable changes over time. For example, data on population growth, price levels, production, sales, or income recorded over different years form a time series. Time series analysis helps us to understand the past behavior of a variable and to predict future trends, making it highly useful in economics, business, and statistics.

Features of Time Series

1) **Time-ordered data –** Observations are arranged chronologically.
2) **Equal intervals –** Data is generally collected at uniform time gaps.
3) **Dynamic nature –** It shows changes and fluctuations over time.
4) **Useful for forecasting –** Helps in predicting future values.

Definitions of Time Series

According to Croxton and Cowden: "A time series is a set of statistical observations arranged in chronological order."

According to Ya-Lun Chou: "A time series is a sequence of observations taken sequentially in time."

According to Spiegel: "A time series is a set of data pertaining to different time periods."

Explanation of Definitions: All the above definitions emphasize that:

1) Data must be collected over time 2) It must be systematically arranged
3) It helps in analyzing patterns such as trend, seasonal, cyclical, and irregular variations

Examples of Time Series :

1) Monthly prices of commodities 2) Annual national income of a country
3) Daily stock market prices 4) Yearly production of a factory

In conclusion, a time series is an essential statistical tool that helps in studying the behavior of variables over time. It plays a crucial role in economic planning, business decision-making, and forecasting future trends. By analyzing time series data, we can identify patterns and make informed predictions, thereby reducing uncertainty about the future.

5.2. Uses of Time Series Analysis

1. It helps in understanding past behaviour: By observing data over a period of time one can easily understand what changes have taken place in the past. Such analysis will be extremely helpful in predicting the future behaviour.

2. It helps in planning future operations: Plans for the future cannot be made without forecasting events and relationship they have. Statistical techniques have been evolved which enable time series to be analysed in such a way that the influences which have determined the form of that series may be ascertained. If the regularity of occurrence of any feature over a sufficiently long period could be clearly established then within limits, prediction of probable future variations would become possible.

3. Facilitate the comparison: The time series analysis is helpful in making a comparison between of the behaviour of different time series when the data are recorded systematically. Over a period of time, comparisons become easy and conclusions are drawn easily between two time series.

4. To predict future behaviour: The probable future behaviour of a particular variable can be estimated easily through the analysis of time series. It helps in predicting the short ranges as well as seasonal variations in the phenomenon.

5. Forecasting the trade cycles: The basic objective of the time series analysis is to forecasting the trade cycles. This provides a safeguard to business and trading community against depression.

6. Isolation of effects of different forces: The time series analysis helps in isolation of the effects of various forces affecting the time series.

5.3. Components of Time Series Analysis

The components of time series are identified on the analysis of time series data over a period of time. The changes that have taken place in the trends, patterns, movements as they often called components or elements of a time series data are:

1. **Secular Trend or Long run:** The long run or the secular trend refers to the general tendency of the time series data to increase decrease or remain constant during long time. The secular trend is the general, smooth, long term average tendency.
2. **Seasonal Variations:** The periodic movements in business activity are those which occur at regular intervals i.e. every year. These variations repeat during a period of 12 months they can be predicted fairly accurately.
3. **Cyclic Variations:** These variations are taking minimum 2 years maximum 8-10 years to complete the cycle. The cyclical variations are of a long duration than a year.
4. **Irregular Variations:** Irregular variations, also called episode', 'erratic', random, accidental fluctuations. Refer to such variations in business activities which do not repeat in a definite pattern.

5.4. Secular Trend or Long Run

In time series analysis, data observed over a long period shows various types of movements. Among these, the most important component is the Trend, also known as the Secular Trend (T). It represents the general direction in which a variable moves over a long period of time, ignoring short-term fluctuations. The study of trend is essential because it provides a broad picture of growth or decline, helping economists and business analysts understand long-term tendencies. It is called "secular" because it relates to long-duration changes rather than short-term or temporary variations.

Types of Trend :

1) **Upward Trend (Rising Trend):** When values show a continuous increase over time. Example: Growth of population, rise in prices.
2) **Downward Trend (Falling Trend):** When values show a continuous decrease. Example: Decline in demand for outdated technology.
3) **Horizontal Trend (Stable Trend):** When there is no significant increase or decrease. Example: Stable production levels.

Characteristics of Secular Trend :

1) Long-term movement of data
2) Smooth and regular pattern
3) Unaffected by short-term fluctuations
4) Represents overall growth or decline
5) Can be linear or non-linear (curvilinear)

Causes of Trend: The trend in a time series may arise due to several underlying factors:

1) **Population Changes** – Increase in population leads to higher demand.
2) **Technological Progress** – Innovations improve productivity.
3) **Economic Development** – Industrialization and growth.
4) **Inflation or Deflation** – Changes in price levels.
5) **Social and Cultural Changes** – Changing lifestyles and preferences.

Importance of Secular Trend :

1) Helps in **long-term forecasting**
2) Useful for **economic planning and policy-making**
3) Assists in **business decision-making**
4) Provides insight into **growth patterns**
5) Helps in separating other components like seasonal and cyclical variations

The **Secular Trend (T)** is a fundamental component of time series analysis that reflects the **long-term direction of change** in data. By understanding and measuring trend, analysts can make **accurate forecasts and informed decisions**. It acts as the foundation upon which other components of time series are studied and interpreted.

5.5. Seasonal Variation

In time series analysis, data often exhibits short-term fluctuations that repeat at regular intervals within a year. These recurring patterns are known as Seasonal Variations. They are one of the key components of a time series and play a vital role in understanding periodic changes in economic and business activities. Seasonal variation helps identify patterns that occur due to seasonal factors such as climate, customs, and festivals, enabling better planning and forecasting. These variations are short-term in nature and occur at fixed intervals such as monthly, quarterly, or weekly. According to statistical interpretation, Seasonal variation is the regular and recurring fluctuation in a time series that occurs within a year due to seasonal factors.

Characteristics of Seasonal Variation

1) Repeats regularly every year
2) Short-term fluctuations (within a year)
3) Predictable pattern
4) Influenced by natural and social factors
5) Can be measured and adjusted statistically

Causes of Seasonal Variation: Seasonal changes arise due to various factors:

1) **Climatic Conditions:** Weather changes like summer, winter, and rainy seasons affect demand. Example*:* Increased demand for cool drinks in summer.
2) **Customs and Traditions:** Social habits and cultural practices influence consumption. Example: Increased purchase of clothes during festivals.
3) **Festivals and Holidays:** Special occasions create temporary demand spikes. Example: High sales during Diwali or Christmas.
4) **Business Practices:** Seasonal production and marketing strategies. Example: Discounts during year-end sales.

Examples of Seasonal Variation :

1) Ice cream sales increase in summer and decrease in winter
2) Woolen clothes demand rises in winter
3) Agricultural output varies by seasons
4) Retail sales peak during festive seasons

Measurement of Seasonal Variation: Seasonal variations can be measured using the following methods:

1) Simple Average Method: Calculates seasonal indices using averages.
2) Ratio to Trend Method: Compares actual values with trend values.
3) Ratio to Moving Average Method : Uses moving averages to eliminate trend and cyclical effects.
4) Link Relative Method: Measures changes relative to previous periods.

Importance of Seasonal Variation :

1) Helps in business planning and forecasting
2) Assists in inventory and production control
3) Useful for marketing strategies
4) Aids in policy-making and economic analysis
5) Improves accuracy of predictions

Seasonal variation is an essential component of time series analysis that reflects regular and predictable short-term fluctuations. Understanding these variations enables businesses and economists to anticipate demand, allocate resources efficiently, and make informed decisions. Proper measurement and adjustment of seasonal effects lead to more accurate analysis and forecasting.

5.6. Cyclical Variations

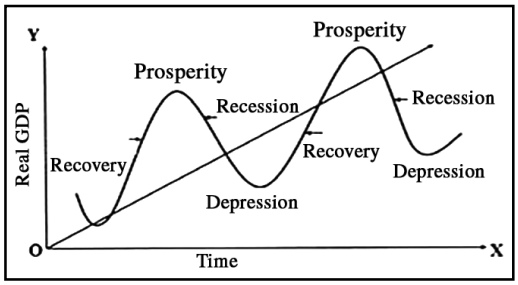
Cyclical variation refers to long-term oscillations or fluctuations in a time series that occur due to changes in overall economic activity. These variations are associated with the business cycle, and they usually extend over more than one year (often 3–10 years or more). In simple terms: Cyclical variations show the rise and fall of economic activities like production, employment, income, and prices over a period of years. Cyclical variation may be defined as: "The recurrent upward and downward movements in a time series around the trend line, occurring over a period of several years due to economic forces.

1. Features of Cyclical Variations :

1) **Long-term movement** (lasts several years)
2) **Wave-like pattern** around the trend line
3) **Not regular or fixed** like seasonal variation
4) Influenced by **economic conditions** (boom, recession, depression)
5) Hard to predict accurately

2. Phases of Cyclical Variation (Business Cycle)

1. **Expansion (Boom)** :
   a) High production, employment, and income
   b) Rising prices and profits
2. **Peak** :
   a) Highest point of economic activity
   b) Economy becomes unstable
3. **Recession** :
   a) Decline in output, income, and employment
   b) Falling demand
4. **Depression (Trough)** :
   a) Lowest level of economic activity
   b) High unemployment and low income



3. **Recovery** :
   a) Gradual improvement in economic conditions
   b) Increase in investment and employment

Diagrammatic Explanation: A time series graph of cyclical variation shows:

1) A trend line (long-term direction)
2) Waves above and below the trend line (cycles)

Causes of Cyclical Variations :

1) Changes in investment and business expectations
2) Fluctuations in demand and supply
3) Monetary and fiscal policies
4) Technological changes
5) External shocks (wars, pandemics, global crises)

Importance of Cyclical Variations :

1) Helps in **economic planning and policy-making**
2) Useful for **business forecasting**
3) Assists investors in decision-making
4) Helps understand **economic stability or instability**

Cyclical variations are an important component of time series analysis, reflecting the **natural ups and downs of an economy over time**. Though difficult to predict precisely, understanding these cycles helps governments and businesses **prepare for future economic changes**.

5.7. Irregular Variations or Random Variations

Irregular variations are **random, unpredictable, and short-term fluctuations** in a time series caused by unforeseen or accidental events. In simple words: They are **sudden disturbances** in data that do not follow any regular pattern. "Irregular variations refer to those movements in a time series which are caused by accidental, unforeseen, or non-recurring factors and do not show any regularity."

Features of Irregular variations :

1) Highly unpredictable in nature
2) Occur due to **unexpected events**
3) No regular pattern or periodicity
4) Usually **short-term effects**
5) May cause **sharp rises or falls** in data
6) Cannot be measured with precision

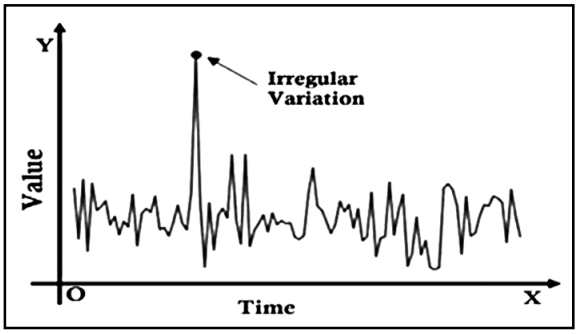
Causes of Irregular Variations:

Irregular fluctuations arise due to sudden and abnormal situations such as :

1) Natural disasters (floods, earthquakes, cyclones)
2) Wars and political instability
3) Strikes, lockouts, or riots
4) Pandemics (e.g., COVID -19)
5) Sudden government policy changes
6) Technological accidents or failures

Graphical Explanation :

1. Appears as **sudden spikes or dips** in a time series graph
2. Does not follow smooth or regular movement
3. Deviates randomly from the trend line



Importance of Irregular Variations :

1) Helps identify **unexpected shocks in data**
2) Useful for **risk management and forecasting limitations**
3) Assists policymakers during **crisis situations**
4) Helps businesses prepare for **uncertain events**

Irregular variations represent the **uncertain and accidental component** of a time series. Though they cannot be predicted or controlled, their study is important to understand the **real-world fluctuations** affecting data.

5.8. Measurement of Secular Trend

Trend is measured in long term series data and after the measurement of trend, we can only present that whether it is growing or declining? There are two important reasons for trend measurement. First is to find out trend characteristics in and of them. In studying trend in and of it. we ascertain the growth factor. For example. we can compare the growth in software industry with the growth in the economy as a whole. The growth factor also helps us in predicting the future behaviour of the data. Second is to enable us to eliminate trend in order to study other elements. The elimination of trend leaves us with seasonal, cyclical and irregular factors. We can then, in two or more series, compare or use the impact of these three relatively short term elements divorced from long term factors. The various methods that can be used for determining the trend are:

1. Graphic Method or Free hand Method.
2. Semi - Average Method.
3. Moving Average Method.
4. Method of Least Square

5.9 Graphical Method or Free Hand Method

It is simplest and the most flexible method of estimating the trend. It consist of first obtaining a graph by plotting the time series values on a graph paper and then drawing a free hand smooth curve through the points of so that it correctly shows the long term tendency of the data. The smoothness of the curve removes other components i.e. seasonal, cyclical and irregular variations. The procedure of getting a straight line trend by this method is given below.

1. Plot the time series on a graph
2. Examine carefully the direction of the trend based on the plotted information (dots).
3. Draw a straight line which will best fit to the data according to personal judgment. The line now shows a direction of the trend. Graphical Method Explained in Table No. 5.1 and figure No.5.5

Table No. 5.1

Years	2006	2007	2008	2009	2010	2011	2012	2013	2014
Production (In thousands)	40	44	48	42	46	50	46	52	50

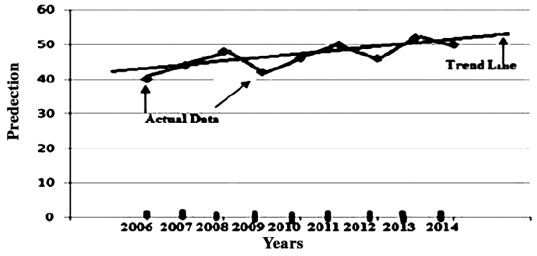


Figure 5.5

When a trend line is fitted by the free hand method, an attempt should be made to make it conforms as much as possible to following conditions:

1. It should be smooth, either a straight line or a combination of long gradual curves
2. The sum of vertical deviations from the trend of the annual observations above the trend should equal the sum of the vertical deviations from the trend of the observations below the trend.
3. The sum of the squares of the vertical deviation of the observations from the trend should be as small as possible.

The trend line drawn by the free hand method can be extended to predict future values. However, since the free hand curve fitting is too subjective. Therefore, this method should not be used for prediction.

5.10 Semi Averages Method

After the data have been divided into two parts, an average (arithmetic mean) of each part is obtained. Then we get two points. Each point is plotted at the mid-point of the class interval covered by the respective parts and then the two points are joined by a straight line which gives us the required trend line. The line can be extended downwards or upwards to get intermediate values or to predict future values. Table No 5.2 Fit a trend line to the following data by the method of semi-averages

Table No. 5.2

S. No	Year	Production	Average
1	1990	15	
2	1991	11	
3	1992	20	
4	1993	10	$15+11+20+10=\frac{56}{4}=14$
5	1994	15	
6	1995	25	
7	1996	35	
8	1997	30	$15+25+35+30=\frac{105}{4}=26.25$

Solution : Since eight years (an even number) are given, the data are divided into two equal halves and the average of the first four years and the last four years is obtained.
The averages of the two halves are :

$$\frac{15+11+20+10}{4}=\frac{56}{4}=14$$

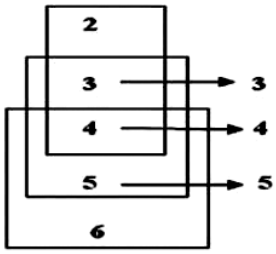
$$\frac{15+25+35+30}{4}=\frac{105}{4}=26.25$$

5.11 Moving Averages Method

It consists in measurement of trend by smoothing out the fluctuations of the data means of a moving average. In this method the average value of a number of year, month, week or days is taken as the trend value for the middle part of the period of moving average. While applying this method, it is necessary to select a period for moving average such as three yearly or six yearly moving average etc. Given the values for the period from 2001 to 2005, in the table No.5.3 We can find out the moving average for three years as follows:

Table No. 5.3

Year	2001	2002	2003	2004	2005
Values	2	3	4	5	6



$= \frac{2+3+4}{3} = 3$

$= \frac{3+4+5}{3} = 4$

$= \frac{4+5+6}{3} = 5$

The 3-yearly moving average shall be computed

$$\frac{a+b+c}{3}, \frac{b+c+d}{3}, \frac{c+d+e}{3}, \frac{d+e+f}{3}$$

The 5-yearly moving average shall be computed

$$\frac{a+b+c+d+e}{5}, \frac{b+c+d+e+f}{5}, \frac{c+d+e+f+g}{5}, \frac{d+e+f+g+h}{5}$$

While applying this method, it is necessary to select a period for moving average such as 3-yearly moving average, 5-yearly moving average, 8-yearly moving average, etc. The period of moving average is to be decided in the light of the length of the cycle. Since, moving average method is most commonly applied to the data which are characterised by cyclical movements, it is necessary to select a period for moving average which consider with the length of the cycle, other-wise the cycle will not be entirely removed. The danger is more severe, the shorter the time period represents by the average.

5.12. Impact of Covid-19 on the Growth of Service sector in India

The service sector is the backbone of the Indian economy, contributing around **50–55% of GDP** and employing a large share of the workforce. It includes sectors such as **trade, transport, tourism, banking, education, health, IT, and communication services.** The outbreak of the **COVID-19 pandemic in 2020** created an unprecedented disruption in this sector due to lockdowns, travel restrictions, and reduced economic activity.

Immediate Impact of COVID-19 :

1. Severe Contraction in Service Activities: *During nationwide lockdowns, many service activities came to a standstill: As a result, the service sector experienced a sharp fall in output and revenue.*

1) Tourism, hospitality, aviation, and retail sectors were **worst affected**
2) Hotels, restaurants, malls, and entertainment centers were closed
3) Transport services (railways, airlines, taxis) faced drastic decline

2. **Decline in Employment:** This created income insecurity and reduced consumer demand

1) Millions of workers in informal services lost jobs
2) Contractual workers and daily wage earners were heavily affected
3) Urban unemployment increased significantly

3. Disruption of Supply Chains :

1) Restrictions on movement interrupted logistics and trade services
2) Banking and financial services faced operational challenges
3) Small service providers struggled to survive

Sector-wise Impact :

1. Tourism and Hospitality :

1) One of the most severely hit sectors
2) International and domestic travel restrictions led to **near-zero activity**
3) Hotels and travel agencies suffered heavy losses

2. Aviation and Transport :

1) Flights were suspended for months
2) Passenger traffic declined drastically
3) Transport companies incurred huge financial losses

3. Retail and Trade :

1) Physical retail shops faced closures
2) Consumer spending declined due to uncertainty
3) Shift towards **online shopping and e-commerce**

4. Education Services :

1) Schools and colleges were closed
2) Rapid transition to **online education platforms**
3) Digital divide affected rural students

5. Health Services :

1) Healthcare sector faced immense pressure
2) Demand for medical services, PPE kits, and infrastructure increased
3) Highlighted the importance of public health investment

6. IT and Digital Services (Positive Impact) :

1) IT sector showed resilience and growth
2) Work-from-home model increased productivity in some cases
3) Expansion of digital services like cloud computing, fintech, and e-learning

Positive Outcomes and Structural Changes :

1. ***Digital Transformation:*** *Accelerated adoption of digital technologies. Growth of online services such as : 1) E-commerce, 2) Online banking, 3) Telemedicine, 4) Online education.*
2. ***Rise of Remote Work:*** *1) Work-from-home became a norm, 2) Reduced operational costs for firms, 3). Increased flexibility in employment*
3. ***Innovation and Startups:*** *1.) Emergence of new business models, 2) Growth in delivery services, digital payments, and health-tech startups*

Government Measures:

The Government of India introduced several measures to support the service sector:

1) **Atmanirbhar Bharat package** to boost economic recovery
2) Financial support to MSMEs and service providers
3) Credit guarantees and liquidity measures
4) Promotion of digital economy and infrastructure

Recovery Phase: By 2022–23, the service sector showed strong recovery and growth momentum

1) Gradual reopening led to recovery in services
2) IT, banking, and communication sectors recovered faster
3) Tourism and hospitality took longer to revive
4) Vaccination drives improved consumer confidence

The COVID-19 pandemic had a **dual impact** on India's service sector. While traditional services like tourism, transport, and hospitality suffered severe losses, digital and IT-enabled services experienced rapid growth. The crisis acted as a **catalyst for structural transformation**, pushing the sector towards digitization and innovation. In the long run, the service sector in India has emerged more resilient, adaptive, and technologically advanced.

5.13 Model Examination Questions

Short Type Questions :-

1. Time series analysis
2. Secular trend
3. Linear trend
4. Non-linear trend
5. Seasonal variations
6. Cyclical variations
7. Irregular variations
8. Semi averages method
9. Moving averages method

Essay Type Questions :-

1. Define time series analysis and explain its importance
2. Describe the components of time series analysis
3. Explain the methods of measurement of secular trend
4. Discuss the Impact of Covid-19 on the Growth of Service sector in India
5. Explain Graphical method or Free Hand Drawing Method

Author

Shaik Hasrath Begum, M.A. SET (Ph.D),

Asst Professor of Economics (PT) Nagarjuna Govt College (A) Nalgonda



'పాస్ మార్కులొస్తే సాలు' అనుకుంట కాలేజీకి వచ్చే ఆర్ట్స్ విద్యార్థుల్ని మెరికల్లా తీర్చిదిద్దడం ఎలాగో ఈ సార్‌ను చూసి నేర్చుకోవాలె. గ్రామీణ, సామాజిక, ఆర్థిక వెనుకబాటుతనంలో ఉన్న విద్యార్థుల్ని చేరదీసి.. బువ్వ పెట్టి.. పాఠం చెప్పి తీర్చిదిద్దిండు. ఎకనామిక్స్ ఎంట్రన్స్ జరిగితే ఆయన స్టూడెంట్సే టాపర్స్. అర్థశాస్త్రంలో విజేతలకు అడ్డా 'నల్లగొండ ఎకనామిక్ ఫోరం'కు కర్త, కర్మ, క్రియ డాక్టర్ అక్కెనపల్లి మీనయ్య సార్! ఒక దీపంతో వేల దీపాలు వెలిగించడమంటే ఏమిటో.. ఈ సార్ జీవితం చెప్తది.

మనసున్న మీనయ్య సార్

ఎకనామిక్స్ జూనియర్ లెక్చరర్‌గా పనిచేస్తున్న డాక్టర్ అక్కెనపల్లి మీనయ్య సార్‌కి ఓ ఇరవై ఏండ్ల కిందట డిగ్రీ కాలేజ్ లెక్చరర్‌గా ప్రమోషన్ వచ్చింది. నల్లగొండ పట్టణంల గవర్నమెంట్ ఉమెన్స్ కాలేజ్‌ల పోస్టింగ్ ఇచ్చినరు. సార్ ఆ కాలేజీకి పోయే నాటికి విద్యార్థినులకు చదువంటే పెద్దగ ఆసక్తి లేకుండెనట. అంతకుముందు ఏడాది చూస్తె.. ఆ కాలేజీ విద్యార్థినులకు ఎవరికీ డిగ్రీ తర్వాత పీజీలో సీటు రాలె. అట్లనే మొదటి ఏడాదిలో చేరినవాళ్లల నలుగురైదుగురు రెండో ఏడాది కల్లా మానేసేటోళ్లట. 'ఈ డ్రాప్ అవుట్స్‌ని తగ్గించాలె. ఆడపిల్లల్ని ఉన్నత చదువులవైపు మళ్లించాలె. బతకనీకి వాళ్లకో దారి చూపించాలె' అనుకున్నడు మీనయ్య సార్!

పీజీ గేట్ వే

డిగ్రీ తర్వాత ఉన్న చదువేంది? ఏది చదివితే ఏ జాబొస్తదో స్టూడెంట్స్‌కి చెప్పి 'బాగా చదవాలె. మంచిగ బతకాలె' అనే ఆశపుట్టించిండు సార్. తన పాఠాలతో స్టూడెంట్స్‌కి ఆర్థిక శాస్త్రం మీద ఇష్టం పెంచుకుంట పోయిండు. స్టూడెంట్స్‌తోనే పాఠాలు చెప్పించడం, ప్రాజెక్ట్‌లు చేయించడం, క్విజ్ పోటీలు పెట్టడం షురూ చేశిండు. అటెండెన్స్ పెరిగింది. డ్రాప్ అవుట్స్ తగ్గినయ్. పిల్లలు మంచిగ చదువుడు మొదలైంది. పరీక్షలు అయిపోయినయ్. ఎండాకాలం వచ్చింది. అయినా కాలేజీకి వస్తున్నరు. ఎందుకో తెలుసా? డిగ్రీ అయిపోంగనే పీజీకి పోవాలె గదా. ఎంట్రన్స్ రాయాల్నంటె మంచిగా ప్రిపేర్ కావాలె. పేదింటి పిల్లలకు కాలేజీలనే ఫ్రీగ కోచింగ్ చెప్పిండు సార్. అయినె కష్టం ఫలించింది. ఉస్మానియా యూనివర్సిటీ పీజీ ఎంట్రన్స్ 2005ల మంచి ర్యాంకులు తెచ్చుకొని.. అయిదుగురు అమ్మాయిలు ఎంఏ ఎకనామిక్స్‌ల చేరినరు.

కష్టపడితే ఫలితమొస్తదని మీనయ్య సార్‌కే గాదు స్టూడెంట్స్‌కీ ఎరుకైంది. 'మన బాగు కోసం సార్ అంత గనం కష్టపడుతుంటె.. మన కోసం మనం కష్టపడలేమా' అని ఉమెన్స్ కాలేజీ అమ్మాయిలు పోటీలు పడి సదివినరు. అట్ల సదువబట్టి ఆ స్టూడెంట్స్ ఏటా కోచింగ్ వస్తున్నరు. యూనివర్సిటీలకు పోతున్నరు. మీనయ్య సార్ 2008ల ముప్పై మంది స్టూడెంట్స్‌కి కోచింగ్ ఇస్తె అందుల ఇరవై మంది ఉస్మానియా యూనివర్సిటీల ఎం.ఎ. ఎకనామిక్స్‌ల సీటు తెచ్చుకున్నరు. అప్పటి సంది మీనయ్య సార్ స్టూడెంట్స్ ఏటా పీజీ సీట్లు తెచ్చుకుంటనే ఉన్నరు. ఉస్మానియా యూనివర్సిటీ ఆర్ట్స్ కాలేజ్, ఉమెన్స్ కాలేజ్, నిజాం కాలేజ్‌ల ఆయన స్టూడెంట్స్‌లేని బ్యాచ్ లేదనుకో!

సీట్ గ్యారెంటీ!

ఉమెన్స్ కాలేజీ నుంచి నాగార్జున గవర్నమెంట్ డిగ్రీ కాలేజ్‌కి మీనయ్య సార్ ట్రాన్స్‌ఫర్ అయిండు. ఆడగూడ ఇదే పని మీదేసుకున్నడు. 'నా స్టూడెంట్స్ బాగా సదువాలె. పీజీలోకి పోవాలె' అన్న పట్టుదలతోని పీజీ కోచింగ్ పెట్టిండు. అమ్మాయిలు ఆడిదాంక రాలేరనుకొని ఉమెన్స్ కాలజీలనే కోచింగ్ క్లాసులు నడిపిండు. ఆ బ్యాచ్‌ల ఉన్న షేక్ సుల్తానా ఉస్మానియా యూనివర్సిటీ ఎంట్రన్స్ 2009ల స్టేట్ ఫస్ట్ ర్యాంక్ సాధించింది. అప్పటి దాంక నల్లగొండ సుట్టుపక్కల ఊళ్లకే ఎరుకైన మీనయ్య సారు కోచింగ్ గురించి రాష్ట్రమంతా ఎరుకైంది. 'సారు దగ్గర సదువాలే. ఎకనామిక్స్ సీటు కొట్టాలె' అనే టార్గెట్‌తోని వరంగల్, ఖమ్మం, మహబూబాబాద్ జిల్లాల నుంచి గూడ సమ్మర్‌లో కోచింగ్‌కి వచ్చేది. వాళ్లందరికీ కలిపి ఎన్‌జీ కాలేజీలనే ఎకనామిక్స్ (సమ్మర్) కోచింగ్ ఇచ్చేది. ఆ ఏడాదిలనే కొత్త యూనివర్సిటీలొచ్చినయ్. ఎకనామిక్స్ సబ్జెక్ట్ పీజీ సీట్లు పెరిగినయ్. ఇగ మీనయ్య సార్ పాఠాలు విన్నోళ్లందరికీ సీటు గ్యారెంటీ అన్న పేరు పడ్డది. అంత పేరొచ్చినా ఏమన్నా సొమ్ము చేసుకున్నడా!? అంటే.. లేదనే చెప్పాలి! పైసా తీసుకోకుండా పాఠాలు చెబుతడు. చెప్పిన పాఠాలు గుర్తుండేటట్టు కోచింగ్ మెటీరియల్ తయారు చేస్తడు.



అభిమానమే నా ఆస్తి

నేను పేద కుటుంబంలో పుట్టిన. గోపులాయిపల్లి నుంచి ఆరు కిలోమీటర్లు నడుకుంటవోయి నార్కెట్‌పల్లి బడిల చదివిన. నేను డిగ్రీ చదివిన ఎన్‌జీ కాలేజీల పాఠాలు చెప్పిన. మొదటి నుంచి నాకు సమాజానికి ఏదో ఒకటి చేయాలనే ఆలోచన ఉండేది. ఆచరించి చూపిన. నా సహచరి పార్వతి నా ఆలోచనకు గౌరవమిచ్చి, ఆచరణకు సహకారం అందించింది. పైసా ఆశించకుండా పాఠాలు చెప్పిన. ఆస్తులు పెంచుకోకున్నా అభిమానించే మనుషుల్ని సంపాదించుకున్న.

- డాక్టర్ అక్కెనపల్లి మీనయ్య

ఎకనామిక్స్ టాపర్స్ అడ్డా!

ఎక్కడెక్కడి నుంచో వచ్చినోళ్లకు పాఠాలు చెప్పేందుకు మీనయ్య సార్ మరికొంత మందితో కలిసి 'నల్లగొండ ఎకనామిక్స్ ఫోరం' పెట్టిండు. పెద్ద సదువులు సదివిన ఆయన స్టూడెంట్స్ నరేశ్, శోభ, కవిత వీళ్లంతా సార్ తీరుగ నలుగురికి సాయపడాల్నని ముందుకొచ్చినరు. 2013ల నల్లగొండ ఎకనామిక్స్ ఫోరంని రిజిస్ట్రేషన్ చేసినరు. మీనయ్య సార్ రిటైర్ అయినా.. స్టూడెంట్స్‌ని, ఎకనామిక్ పాఠాలని వదులుకోలే. నల్లగొండల ఉన్న మల్లు వెంకటనర్సింహా రెడ్డి విజ్ఞాన కేంద్రంల చేరిండు. ఆ విజ్ఞాన కేంద్రానికి కన్వీనర్‌గా ఉండుకుంటు.. ఎంఏ కోవింగే కాదు లెక్చరర్ జాబ్‌లకు, నెట్, స్లెట్ అర్హత పరీక్షలకూ కోచింగ్ ఇస్తున్నడు. విజ్ఞాన కేంద్రంల ఉచితంగ వసతి కల్పించి, అన్నం పెట్టి సదువు చెప్తున్నడు. ఏడాది కింద లెక్చరర్ పోస్టుల నియామకాలు జరిగినయి. మొత్తం 81 పోస్టుల్లో 42 ఫోరం స్టూడెంట్స్‌కే వచ్చినయ్. సగం కొలువులు సార్ స్టూడెంట్సే దక్కించుకున్నరన్నట్టు.

స్మార్ట్ కోచింగ్

పీజీ ఎంట్రన్స్‌కి ప్రతి బ్యాచ్‌కి యాభై నుంచి అరవై మంది విద్యార్థులు ఉంటుండె. కొవిడ్ వచ్చినంక పిల్లగాళ్ల సదువు పాడుకాగూడదని ఆన్‌లైన్ క్లాసులు మొదలు పెట్టినరు. ఒక ఎన్ఆర్ఐ ఆన్‌లైన్ క్లాసులు, ఆన్‌లైన్ కంప్యూటర్ బేస్డ్ ఎగ్జామ్స్ కోసమని సాఫ్ట్‌వేర్ ఇచ్చిండు. జూమ్‌ల క్లాసులు చెబుతున్నరు. ఆన్‌లైన్‌ల పరీక్షలు పెడుతున్నరు. ఆన్‌లైన్ క్లాస్‌ల వీడియోలని యూట్యూబ్‌ల అప్‌లోడ్ చేస్తున్నరు. పీజీ ఎంట్రన్స్, స్లెట్, నెట్ అర్హత పరీక్షలకు పాఠాలన్నీ https://www.youtube.com/@nalgondaeconomicsforum33యూట్యూబ్‌ల ఉన్నయ్.

...✍ నాగవర్ధన్ రాయల

Determined she moves forward



A ROLE MODEL : Shaik Sultana, who secured the first rank in the MA (Economics) common entrance examination conducted by the Osmania University, engrossed in studies in Nidmanur.
- PHOTO: SINGAM VENKATARAMANA

Secures first rank

NALGONOA: Sheik Sulthana, a native of Nidamanur, secured the first rank in the common entrance examination-2009 conducted by the Osmania University for admission to MA (Economics) course. Nalgonda Economics Forum Convenor A. Meenaiah said that Sulthana enrolled for the free coaching conducted by the forum.

- Staff Reporter

Undeterred by poverty and domestic work, Shaik Sultana, a married woman of Nidmanur, is excelling in her studies with rock solid determination.

Sultana secured first rank in the highly competitive PG common entrance examination-2009 conducted by the Osmania University for admission into MA (economics) course.

A daughter of an agriculturist and wife of a driver, Sultana surpassed all the hurdles posed by poverty with a strong will and completed her degree in first class from the Government Women's Degree College in Nalgonda.

She resumed her studies few years after her marriage and joined Intermediate. Sultana bagged the Prathibha award by securing the top marks in the Intermediate (HEC) group in the entire district.

Later, she pursued her degree after giving birth to a baby boy. She proved her calibre by securing top rank in the MA (economics) common entrance test of Osma-nia University.

"I owe a lot to the constant encouragement from my lecturers and principal", Sultana remarked and said "I credit my success in the PGCET-2009 to the free coaching given by the Economics Forum". I would like to write Group-I and other competitive examinations to become an officer of high repute, she said. I will secure admission in MA (Economics) in the Arts College in Osmania University by virtue of my first ran)?", she said.

I will take coaching for the competitive examinations during my stay in Hyderabad to realise my goal, Sultana said with confidence.

- P. SRIDHAR, Nalgonda

CPGET - 2025 ఫలితాల్లో ఘోరం విద్యార్థులు 1వ, 2వ, 3వ, 6వ, 12వ, 13వ, 14వ, 17వ మరియు 20వ ర్యాంకులు పొంది అద్భుత ఫలితాలు సాధించినారు.





S. Bhuvana Laxmi

1st Rank



N. Ajay

2nd Rank



A. Manasa

3rd Rank



Y. Lokesh

6th Rank

MVN VIGNANA KENDRAM

NALGONDA ECONOMICS FORUM (Regd. No.297/13)

డా॥ అక్కెనపల్లి మీనయ్య గారి ఆధ్వర్యంలో 18 సం.ల ఉచిత అర్థశాస్త్ర ప్రవేశ పరీక్ష శిక్షణ లో 2009, 2016, 2017, 2018, 2020, 2022 మొదటి ర్యాంకులతో పాటు, 2024 లో 41 మంది ప్రభుత్వ జూనియర్ ఎకనామిక్స్ లెక్చరర్స్‌గా, 10 మంది గురుకుల ఎకనామిక్స్ డిగ్రీ లెక్చరర్స్‌గా మరియు 12 మంది గురుకుల ఎకనామిక్స్ జూనియర్ లెక్చరర్స్‌గా ఎంపిక కావడం

రాష్ట్ర చరిత్రలో అఖండ విజయం



1st Rank
Sk. Sulthana
OU CET-2009
Present: Degree Lecturer in Economics (Gurukul)



1st Rank
K.Shravan Kumar
OU CET-2016



1st Rank
B. Bhavana
KU CET-2016



1st Rank
G. Krishna
OU CET-2017



1st Rank
Bugga Narender
JRF (Ph.D)
OU CET - 2018
Present: Govt. JL in Economics



1st Rank
B. Guru Prasad
KU CET-2018
Present: S.A. Social



1st Rank
K. Sai Chandra
CPGET-2020
Present : Gr-IV Junior Assistant



1st Rank
G. Prabhakar Reddy
OU CET-2022, Present: Enterprize Development Cell Manager (Industries), Nalgonda



డా॥ అక్కెనపల్లి మీనయ్య
M.A, M.Phil, Ph.D.
PRESIDENT
Nalgonda Economics Forum

978-81-996558-6-7

Rs. **175/-**

SHREERAMA_Nlg @ 950 561 8383



Regd.No. 297/13

Nalgonda Economics Forum

M.V.N. Vignana Kendram, Doddi Komaraiah Bhavan, NALGONDA-508001 (TS)
Cell:9490138118,7013774141, 08682-295581
Website : nalgondaeconomicsforum.org