KAMARAJ COLLEGE

SELF FINANCING COURSES
(Reaccredited with "A+" Grade by NAAC)
(Affiliated to Manonmaniam Sundaranar University, Tirunelveli.)
THOOTHUKUDI – 628003.

STUDY MATERIAL FOR BA (ECONOMICS)

STATISTICS FOR ECONOMICS-I

SEMESTER – I

Academic Year 2022-23

Prepared by

ECONOMICS DEPARTMENT

SYLLABUS

UNIT-I Introduction of Statistics and Collection of Data

Definition of statistics- Functions – Importance- Limitations-collection of data-primary and secondary data-census and sampling methods.

UNIT-II Classification, Tabulation, Diagrams and Graphs

Meaning and objectives of classification-Tabulation-Rules of Drawing Diagrams- Merits and demerits of tables-Graphs- merits and demerits of graphs

UNIT –III Measures of Central Tendency

Definition-Characteristics of good average – Types of average – a) Arithmetic mean b) Median c) Mode d) Geometric mean e) Harmonic mean– merits and demerits of different types of averages

UNIT-IV Dispersion

Measures of Dispersion — Types of dispersion – a) Range b) Quartile deviation c) Mean deviation –d) Standard deviation e) Lorenz curve f) Co-efficient of variation– merits and demerits of dispersion

UNIT-V Skewness and Kurtosis

Skewness – Types – Measures of Skewness – Kurtosis –a) Karl Pearson's co-efficient of skewness b) Bowley's co-efficient of skewness- Kurtosis –Measures of kurtosis.

Reference Book:

1. Mathematics and Statistics in Economics - D. R. Agarwal
2. Business statistics and mathematics - Joseph A. Mangala doss
3. Business statistics - D.R .M .Wilson
4. Statistics - K. Pazhani

Table of Contents

UNIT-I

INTRODUCTION OF STATISTICS AND COLLECTION OF DATA

Meaning of statistics and what are the functions of statistics?

Meaning of Statistics:

Statistics is a science of facts and figures. It is used in singular as well as plural sense. When it is used in singular sense it refers to 'analytical statistics. When it is used in plural sense it refers to descriptive statistics. It refers to numerical facts and figures collected on subject.

Functions of statistics:

1. Statistics simplifies complex data.

It provides us the ways of condensing a mass of information's into a single figure. For example, from the individuals income of the entire populations, we can calculate the per capita income. This single figure helps us to know about the standard of living of the people.

2. It presents facts in a definite form.

Figures are precise and more convincing. They convey definite information to the readers. For example, the rate of growth of population was 2.2 per cent during 1971-1981. This statistical statement gives definite information.

3. It helps in making comparisons.

Figures become useful when they are compared with others. Statistics provides us with suitable device for comparison of data over time or between regions.

4. It helps in formulating and testing hypotheses.

For example, hypothesis like whether a particular coin is biased or not, whether a particular medicine would be effective in checking a disease or not.

5. It helps in making predictions:

Looking at past data predictions can be made about the future. For example, looking at the current rate of growth of population, the government can Predict how many schools or hospitals would be required in the future.

6. It helps to formulate suitable policies:

Statistics related to production and demand is useful to frame suitable export-import policies.

7. It helps to measure the effects of a change:

For example, the effect of a change in the bank rate and tax-rate can be studied with the help of statistics.

8. It provides tools for numerical measurement:

For example, index number helps to measure complex variables which can be easily understood.

2) What are the limitations of statistics?

LIMITATIONS OF STATISTICS:

Statistics does have several limitations. They are:

1. Statistics does not deal with individuals.

It deals with aggregates of facts.

2. Statistical studies only quantitative aspect of a problem.

It does not study in qualitative aspects.

3. Statistical results are valid only on an average.

For example, the average level of income does not imply that each individual is getting same amount of income. There can be variations on either side.

4. Statistical results are only approximately correct.

Unlike mathematicians or accountancy, statistics is a science of estimates. It fails in the field where perfect accuracy is desired.

5. Statistical results can be misused.

It can be manipulated to get the desired results. Whenever we come across any statistical data, we must check up on the reliability.

6. Statistical data can be handled by experts.

It cannot be used properly by unskilled persons.

3) Explain the definition of Horace secrist.

Horace Secrist's Definition.

"By statistic we mean aggregates of facts, affected to a marked extent by municipality of causes, numerically expressed, enumerated or estimated according to reasonable standards of accuracy, collected in a systematic manner, for a pre-determined purpose and placed in relation to each other." This is the most satisfactory definition of statistics. On the basis of this definition, we can identify the following as the characteristics of statistics.

1. Statistics should be aggregates of facts.

A single figure or isolated figures are not statistics. Only mass of data or facts can be compared. Statistical methods extract meaningful information's from a mass of data.

2. Statistics is affected to a marked extent by multiplicity of causes.

Business and economic phenomena are very complex. They are influenced by a number of factors. For example, the statistics on production of paddy is affected by number of factors like soil condition, rainfall, use of manure, fertilizers etc.

3. Statistics should be numerically expressed.

It deals with quantitative data. They give meaningful relationship between variables. Qualitative characteristics such as intelligence, beauty etc. cannot be included in statistics unless they are quantified.

4. Statistics should be enumerated or estimated according to a reasonable standard of accuracy.

The degree of accuracy depends on circumstances. For example, time series data

analysis gives a general trend in the change of a given variable. A high degree of accuracy is not insisted upon in statistics as in the case of Mathematics and Accountancy.

5. Statistics should be collected in a systematic manner.

There should be a definite method and purpose in the collection of data. A good deal of caution is necessary in the collection of data. A haphazard collection of figures may lead to erroneous conclusions.

6. Statistics should be collected for a pre-determined purpose.

Vagueness and ambiguity of purpose may lead to collection of useless information. This may pave the way for wastage of time, energy and money.

7. Statistics should be placed in relation to each other.

It facilitates comparison. Relationship and comparison of variables are important parts of the study of statistics

4) Explain the importance of statistics in economics and what the Relations between statistics and Economics are.

IMPORTANCE OF STATISTICS

Statistics in Economics

Statistics is of utmost importance to economics. In fact, it is an indispensable part of the work of an economist. Therefore Marshall observes: "Statistics are straw out of which I like every other economists have to make the bricks."

1. Statistics helps in solving the basic economic problems of what to produce? How to produce? And from whom to produce?

2. Statistics plays a very important role in the formulation of economic theory. A hypothesis becomes an economic theory only when it is supported by adequate statistical data. For example, a hypothesis that there exits an inverse relationship between rate of interest and amount borrowed becomes a theory when it is proved with statistical facts.

3. National income is the joint effort of the four factors of production. It can be equitably distributed if we have enough statistics related to national income and the amount of different factors employed to produce it. Thus, statistics paves the way for proper distribution of national income.

4. Statistics is used in economic planning. Planning involves estimation of resources present in the economy, setting up of targets and allocation of resources. All these three steps require statistical data and statistical methods so as to set things in the right direction to achieve the targets.

5. Statistical data are of immense use in framing suitable economic policies. It is the census of population that gives an idea about the demand for schools, hospitals, transport services etc. in the future. It is followed by allocation of resources in these areas. Statistic related to export and import of our country helps to make changes in our trade policy. Similarly, the government can bring about changes in its fiscal and monetary policy based on statistical data.

Thus, statistics has become an integral part of economics. This situation induced Universities all over the world to prescribe statistics as one of the papers for the students of economics. Students doing research in Economics can do wonders if they are well-versed in the application of statistical methods in Economic research.

5) What is meant by secondary data? What are its sources?

SECONDARY DATA

Data that had already been collected made available are called 'secondary data'. Primary data becomes secondary data if it is used for the second time for any purpose. They can be collected from two sources. They are:

(i) Published sources: and
(ii) Unpublished sources

(i) Published Sources

Governmental and non-governmental agencies publish statistics on different subjects periodically or ado basis. They are much useful for research and policy making. The chief sources of published statistics are:

a) Government publications

The Central Statistical Organization (CEO) collects information on various subjects and publishes them. For example, the CSO publishes national accounts statistics which gives information on the national income of our country. The other publications are:

- Annual Survey of Industries
- Agricultural Statistics of India
- Indian Trade Journal
- RBI Bulletin
- Statistical Abstract of India (Annual)
- Report of the enquiry Commissions
- Monthly Abstract of Statistics, etc.

b) Semi-official publications

Municipal Corporations collect information about births and deaths of different areas.

c) Reports of Committees and Commissions

The Central as well as the State Governments set up Committees and Commission to study a specific problem. They have to submit a detailed report on it. For example, the National Institute of Public Finance made a survey about black money in India. Other commissions are:

- The Reports of National Labour Commission,
- Monopolies Commission.
- Finance Commission.
- National Agricultural Commission

d) Publications of International Organizations

The following are some of the important international organizations publishing statistical information relating to different countries:

IMF - International Monetary Fund
IBRD - International Bank for Reconstruction and Development
ILO - International Labour Organization
UNO - United Nations Organization

WHO - World Health Organization
WTO - World Trade Organization.

e) Journal and news Papers

Journals like Commerce, Eastern Economist, Southern Economist, Journal of Industry and Trade etc. and newspapers like 'Economic Times' and Financial Express collect and publish statistics on different aspect.

f) Research Works

Research institutes like the National Council of Applied Economics Research (NCAER), National Institute for Rural Development (NIRD), Institute for Social and Economic change (ISEC) publish the research works undertaken in different aspects in economics, various research scholars doing research in various universities publish their research papers or doctoral thesis from where we get secondary data.

g) Reports

The federation of Indian Chamber of Commerce, The Institute of Chartered Accountants, Stock Exchanges, Trade Union etc. which would publish information relating to trade and commerce from the annual reports of joint stock companies. We can get information about the profits or losses of individual companies.

ii) Unpublished Sources

All statistical information and survey reports are not always published. Such sources are

a) Unpublished survey reports
b) Unpublished thesis kept in libraries
c) Government records and files
d) Records of private companies.

6) Explain the random sampling method.

MEANING OF RANDOM SAMPLING

All the items in the population have an equal chance of being chosen in the sample. Random Sampling is of two types. They are:

(i) Simple Random Sampling
(ii) Restricted Random Sampling

(i) Simple Random Sampling

This method can be of two types. They are

(a) Lottery Method
(b) Table of Random Numbers

a) Lottery Method

In this method, list of population is made. A number is assigned to each and every unit of population. Units are written on slips of paper. They are put into a box. They are shuffled together and the required numbers of slips is taken out of the box. The numbers which appear on slips that are chosen will constitute the sample.

b) Table of Random Numbers

In this method, there is a list of random numbers. One can start from any point and can go either vertically, horizontally or even diagonally. You can keep on noting down the numbers. These numbers will form our sample.

(ii) Restricted Random Sampling

The restricted random sampling can be of three types. They are:

(a) Stratified Sampling
(b) Systematic Sampling
(c) Cluster Sampling

(a) Stratified Sampling

In this method, the population is divided into groups on the basis of some characteristics. This method is used when information about some parts of the population is wanted. At the time of constructing strata the following points should be kept in mind:

(i) There should be perfect homogeneity in the different units of the strata.
(ii) Stratification must be clear, well-defined and free from overlapping.
(iii) The size of stratified sample must not be too small.
(iv) Different variables involved in the study should be taken into account.

c) Systematic Sampling

In this method, the sample is chosen in a systematic manner. This method is used when a complete list of population is available.

d) Cluster Sampling

In this method, samples are selected in different stages at random. This method is also called as 'multi-stage sampling'. The selection of the sample is done in two stages or more. First, certain groups or clusters are selected from the population. They are primary sample units. From each group elementary sample units are drawn. This method is applicable where the universe is large.

7) Distinguish between Census and Sampling.

Distinction between Census and Sample Method

Census Method	Sample Method
1. Each and every unit of the population is included in the study.	1. Only parts of the population instead of all units are included in the study
2. No sampling error is present because all the individuals are considered	2. Sampling error may be present because items are chosen is a random order.
3. It is more costly	3. It is less expensive
4. It is time consuming	4. It saves time
5. A large team of investigators is required for carrying out the survey.	5. It can be carried out by a minimum number of investigators.
6. It cannot be used for an intensive survey	6. Intensive survey can be done by using this method

7. As each and every unit is included this method is more reliable and Accurate.	7. It is not much reliable as a portion of population is chosen at Random
8. It is not scientific method. It is not based on any law	8. It is a scientific method. It is based on (a) the law of inertia of large numbers and (b) the law of statistical regularities.
9. This error is higher as we are dealing with large number	9. Magnitude of non-sampling error is less.
10. It can be carried out by a large organization only	10.It can be carried out by small organization and even individuals
11. For example, in case of testing the life of bulbs, this method can't be used.	11.This can be used even for special blood testing this method can be used

8) Distinguish between Primary and Secondary data.

Distinction between Primary and Secondary data.

Primary Data	Secondary Data
1. Data collected for the first time for a statistical investigation	1. Data already collected and available
2. They are original in nature	2. They are already processed data
3. They are first hand information's	3. They are second hand information's
4. They are in the form of raw data	4. They are available in published or unpublished form
5. They require appoint of more number of investigators or enumerators	5. A single person is enough to collect
6. They are accurate and reliable	6. They are inaccurate and less reliable as compared to primary data
7. Personal bias may creep in	7. No chance for personal bias
8. They are expensive. They consume more time, money and labour	8. They are less expensive. They save time, money and labour.

8 Mark

1) Describe the scope and importance of statistics.

Scope of Statistics

- Statistics was used in olden days by the state to collect information for public administration. Now its scope was not only in collection of information, but also analyses of the data collected, to draw inference or formulate a theory.
- Hence, it helps for planning and formulation of policies. The methods and techniques available in statistics helps people to solve various problems presented in statistical data.

- Hence, it has a good application and scope in the field of commerce, economics, physics, chemistry, botany, zoology, psychology etc.

Statistics in Business

- Statistics is most commonly used in Business. It helps to take decision regarding whether a company can start a new business.
- The statistical data regarding the demand and the supply of products can be collected and analyzed to take a decision regarding the new business.
- The company can also calculate the cost of production and then the selling price. The selling price can be compared with the selling price of competitive products to make a final decision for starting a new business.

Statistics in Management

- Most of the managerial decisions are taken with the help of statistics. The data regarding the performance of a company for the inception period will facilitate to take decision regarding future.
- Statistical techniques like correlation analysis, regression analysis and time series technique can be used in this regard.
- The important managerial activities like planning, directing and controlling are properly executed with the help of statistical data and statistical analysis.

Statistics in Banking and Finance

- Banking and Financial activities use statistical most commonly. In banks statistical data regarding the customer deposit, loans etc. are represented in statistical data.
- Banks also applies the statistical techniques in calculating interests.
- Stock exchange, financial institutions like Industrial Development Bank of India, State Financial Corporation of India also uses statistics in projecting the future and to solve various statistical problems.

IMPORTANCE OF STATISTICS

1. Statistics helps in solving the basic economic problems of what to produce? How to produce? And from whom to produce?
2. Statistics plays a very important role in the formulation of economic theory. A hypothesis becomes an economic theory only when it is supported by adequate statistical data. For example, a hypothesis that exits an inverse relationship between rate of interest and amount borrowed becomes a theory when it is proved with statistical facts.
3. National income is the joint effort of the four factors of production. It can be equitably distributed if we have enough statistics related to national income and the amount of different factors employed to produce it. Thus, statistics paves the way for proper distribution of national income.
4. Statistics is used in economic planning. Planning involves estimation of resources present in the economy, setting up of targets and allocation of resources. All these three steps require statistical data and statistical methods so as to set things in the right direction to achieve the targets.
5. Statistical data are of immense use in framing suitable economic policies. It is the census of

population that gives an idea about the demand for schools, hospitals, transport services etc. in the future. It is followed by allocation of resources in these areas. Statistic related to export and import of our country helps to make changes in our trade policy. Similarly, the government can bring about changes in its fiscal and monetary policy based on statistical data.

2) Bring out the characteristics features of Statistics.

FEATURES OF STATISTICS

1. Statistics should be aggregates of facts.

A single figure or isolated figures are not statistics. Only mass of data or facts can be compared. Statistical methods extract meaningful information from a mass of data.

2. Statistics is affected to a marked extent by multiplicity of causes.

Business and economic phenomena are very complex. They are influenced by a number of factors. For example, the statistics on production of paddy is affected by number of factors like soil condition, rainfall, use of manure, fertilizers etc.

3. Statistics should be numerically expressed.

It deals with quantitative data. They give meaningful relationship between variables. Qualitative characteristics such as intelligence, beauty etc. cannot be included in statistics unless they are quantified.

4.Statistics should be enumerated or estimated according to a reasonable standard of accuracy.

The degree of accuracy depends on circumstances. For example, time series data analysis gives a general trend in the change of a given variable. A high degree of accuracy is not insisted upon in statistics as in the case of Mathematics and Accountancy.

5. Statistics should be collected in a systematic manner.

There should be a definite method and purpose in the collection of data. A good deal of caution is necessary in the collection of data. A haphazard collection of figures may lead to erroneous conclusions.

6. Statistics should be collected for a pre-determined purpose.

Vagueness and ambiguity of purpose may lead to collection of useless information. This may pave the way for wastage of time, energy and money.

7. Statistics should be placed in relation to each other.

It facilitates comparison. Relationship and comparison of variables are important parts of the study of statistics

3) Explain the various methods of collection of data.

COLLECTION OF DATA

Any statistical enquiry is conducted to collect required data to suit the purpose of the enquiry. Statistical data may be collected by an individual researcher, a research institute, a business firm, a state or central government or a newspaper or a magazine.

SOURCES OF DATA

Statistical data can be collected from two sources. They are:

1. Primary Data 2. Secondary Data

PRIMARY DATA

Data collected for the first time by the investigator himself or by his agents are called 'primary data'. They are collected for a pre-determined purpose. They are original in nature. They are first-hand information collected in order to satisfy the purpose of a particular statistical enquiry.

Methods of Collecting Primary Data

The following are the methods of collecting primary data.

(i) Direct personal interview
(ii) Indirect oral interview
(iii) Information through correspondents (Local reports)
(iv) Mailed questionnaires
(v) Schedules

(i)Direct Personal interview

In this method the investigator personally contacts the informants. Informants are persons who provide answers to the investigator's question. The investigator notes down the information as data for his survey. Thus, if a person wants to study the socio-economic conditions of workers in an industry, he has to go to the industry, contact the workers and obtain the desired data. This method can be used for intensive studies if adequate time and money is available.

(ii)Indirect Oral Law

In this method, the investigator contacts friends, relatives and neighbors of the person about whom we require information. This method is mostly used when informants are reluctant to give information. This method is generally used by enquiry committees or commission appointed by the government and also police department.

(iii)Information through correspondents

Under this method the investigator appoints agents or correspondents in different places of the field of enquiry. This method is used by newspapers and periodicals which require certain information regularly. They appoint correspondents permanently in different areas. They collect information from their area and supply it to the head office. All agricultural statistics are gathered by using this method.

(iv)Mailed Questionnaires

A questionnaire contains a list of questions related to an investigation. It is sent by post to the informants. Prepaid postages stamps are fixed. A covering letter should be attached to the questionnaire. It will contain the ways and means for filling and returning the filled-up questionnaires.

(v)Schedules

A schedule is a set of questions arranged so as to obtain information on a given subject from the informants. It is sent through trained enumerators. A set of enumerators are selected under this method. The enumerators personally contact the informants and fill the schedule in their own handwriting.

SECONDARY DATA

Data that had already been collected made available are called 'secondary data'. Primary data becomes secondary data if it is used for the second time for any purpose. They can be collected from two sources. They are:

(iii) Published sources: and
(iv) Unpublished sources

(i) Published Sources

Governmental and non-governmental agencies publish statistics on different subjects periodically or ado basis. They are much useful for research and policy making. The chief sources of published statistics are:

a) Government publications

The Central Statistical Organization (CEO) collects information on various subjects and publishes them. For example, the CSO publishes national accounts statistics which gives information on the national income of our country. The other publications are:

- Annual Survey of Industries
- Agricultural Statistics of India
- Indian Trade Journal
- RBI Bulletin
- Statistical Abstract of India (Annual)
- Report of the enquiry Commissions
- Monthly Abstract of Statistics, etc.

b) Semi-official publications

Municipal Corporations collect information about births and deaths of different areas.

c) Reports of Committees and Commissions

The Central as well as the State Governments set up Committees and Commission to study a specific problem. They have to submit a detailed report on it. For example, the National Institute of Public Finance made a survey about black money in India. Other commissions are:

- The Reports of National Labour Commission,
- Monopolies Commission.
- Finance Commission.
- National Agricultural Commission

d) Publications of International Organizations

The following are some of the important international organizations publishing statistical information relating to different countries:

IMF - International Monetary Fund
IBRD - International Bank for Reconstruction and Development

ILO - International Labour Organization
UNO - United Nations Organization
WHO - World Health Organization
WTO - World Trade Organization.

e) Journal and news Papers

Journals like Commerce, Eastern Economist, Southern Economist, Journal of Industry and Trade etc. and newspapers like 'Economic Times' and Financial Express collect and publish statistics on different aspect.

f) Research Works

Research institutes like the National Council of Applied Economics Research (NCAER), National Institute for Rural Development (NIRD), Institute for Social and Economic change (ISEC) publish the research works undertaken in different aspects in economics, various research scholars doing research in various universities publish their research papers or doctoral thesis from where we get secondary data.

g) Reports

The federation of Indian Chamber of Commerce, The Institute of Chartered Accountants, Stock Exchanges, Trade Union etc. publish information relating to trade and commerce from the annual reports of joint stock companies. We can get information about the profits or losses of individual companies.

ii)Unpublished Sources

All statistical information and survey reports are not always published. Such sources are

a) Unpublished survey reports
b) Unpublished thesis kept in libraries
c) Government records and files
d) Records of private companies.

4) Discuss the various methods of collecting primary data.

PRIMARY DATA

Data collected for the first time by the investigator himself or by his agents are called 'primary data'. They are collected for a pre-determined purpose. They are original in nature. They are first-hand information collected in order to satisfy the purpose of a particular statistical enquiry.

Methods of Collecting Primary Data

The following are the methods of collecting primary data.

i. Direct personal interview
ii. Indirect oral interview
iii. Information through correspondents (Local reports)
iv. Mailed questionnaires
v. Schedules

(i)Direct Personal interview

In this method the investigator personally contacts the informants. Informants are persons who provide answers to the investigator's question. The investigator notes down the information as data for his survey. Thus, if a person wants to study the socio-economic conditions of workers in an industry, he has to go to the industry, contact the workers and obtain the desired data. This method can be used for intensive studies if adequate time and money is available.

(ii)Indirect Oral Law

In this method, the investigator contacts friends, relatives and neighbors of the person about whom we require information. This method is mostly used when informants are reluctant to give information. This method is generally used by enquiry committees or commission appointed by the government and also police department.

(iii)Information through correspondents

Under this method the investigator appoints agents or correspondents in different places of the field of enquiry. This method is used by newspapers and periodicals which require certain information regularly. They appoint correspondents permanently in different areas. They collect information from their area and supply it to the head office. All agricultural statistics are gathered by using this method.

(iv)Mailed Questionnaires

A questionnaire contains a list of questions related to an investigation. It is sent by post to the informants. Prepaid postages stamps are fixed. A covering letter should be attached to the questionnaire. It will contain the ways and means for filling and returning the filled-up questionnaires.

(v)Schedules

A schedule is a set of questions arranged so as to obtain information on a given subject from the informants. It is sent through trained enumerators. A set of enumerators are selected under this method. The enumerators personally contact the informants and fill the schedule in their own handwriting.

5) Explain the various methods of Sampling.

SAMPLING METHOD

These are two methods of selecting samples from population or universe. They are:

(i) Random sampling
(ii) Non-random sampling

A. MEANING OF RANDOM SAMPLING

All the items in the population have an equal chance of being chosen in the sample. Random Sampling is of two types. They are:

i. Simple Random Sampling
ii. Restricted Random Sampling

(i)Simple Random Sampling

This method can be of two types. They are

a) Lottery Method
b) Table of Random Numbers

a) Lottery Method

In this method, list of population is made. A number is assigned to each and every unit of population. Units are written on slips of paper. They are put into a box. They are shuffled together and the required numbers of slips is taken out of the box. The numbers which appear on slips that are chosen will constitute the sample.

b) Table of Random Numbers

In this method, there is a list of random numbers. One can start from any point and can go either vertically, horizontally or even diagonally. You can keep on noting down the numbers. These numbers will form our sample.

(ii)Restricted Random Sampling

The restricted random sampling can be of three types. They are:

(a) Stratified Sampling
(b) Systematic Sampling
(c) Cluster Sampling

(a) Stratified Sampling

In this method, the population is divided into groups on the basis of some characteristics. This method is used when information about some parts of the population is wanted. At the time of constructing strata the following points should be kept in mind:

i. There should be perfect homogeneity in the different units of the strata.
ii. Stratification must be clear, well-defined and free from overlapping.
iii. The size of stratified sample must not be too small.
iv. Different variables involved in the study should be taken into account.

c) Systematic Sampling

In this method, the sample is chosen in a systematic manner. This method is used when a complete list of population is available.

d) Cluster Sampling

In this method, samples are selected in different stages at random. This method is also called as 'multi-stage sampling'. The selection of the sample is done in two stages or more. First, certain groups or clusters are selected from the population. They are primary sample units. From each group elementary sample units are drawn. This method is applicable where the universe is large.

B.NON-RANDOM SAMPLING

There are three methods of non-random sampling. They are

(i) Judgment sampling
(ii) Quota sampling and
(iii) Convenient Sampling

(i)Judgment Sampling

In this method, a sample is chosen at the discretion of the investigator. He decides which individuals should be chosen as sample. He selects the items which he considers as the most representative of the universe. For example, if an investigator wants to select 20 students from a batch of 500 students to know their spending habits, he selects the students who in his opinion are

representative of the college. This method is more useful to businessmen and government officials. This is also called as 'purposive sampling' or 'deliberative sampling'.

(ii)Quota Sampling

In this method, the investigator gets instructions to collect information from an assigned number of individuals (Quota). The individuals are left to the personal choice of the investigator. Each interviewer's workload has been based on a fixed quota of individuals. Quotas are affected by investigators bias.

(iii)Convenient Sampling

In this method, a sample is chosen according to the convenience of the investigator. This method is also called the 'chunk'. If the universe is not well defined or source list is not available this method is applicable. A sample selected from readily available list such as birth-death registers, telephone directories etc. is a convenient sample. For example, in a survey regarding village life, the surveyor may choose the village which is nearer or more accessible to him.

6) Discuss the principles on which the theory of Sampling based.

Theories of Sampling

There are two important principles on which the theory of sampling is based. They are

(i) Law of Statistical Regularity

(ii) Law of Inertia of Large Numbers

1. Law of Statistical Regularity

The law states that if we choose a few units at random from the population, it is likely to possess the characteristics of the entire population. By random selection we mean a selection where each and every unit of the population has an equal chance of being selected in the sample. This law is derived from the mathematical theory of probability. For example, to know the impact of Self-Help Groups (SHGs) on the socio-economic uplift of women we have to select sample of women and collect necessary information. The average values we can get will reveal the facts.

2. Law of Inertia of Large Numbers

The law states the other things being equal, as the sample size increases, the accuracy increases. It implies that when a large number of units are taken as sample, the results will be more accurate. For example, if we toss a coin 1000 times, there are 50 per cent chances for heads and 50 per cent tails. But if we toss a coin only four times, we may not get exactly 2 heads and 2 tails.

Merits

1. Sampling method is easy to organize and administer a sample survey
2. It is possible to collect more detailed information in a sample survey
3. The results obtained from sampling method are more reliable than those obtained from census method.
4. It is more economical. It requires less number of enumerators and saves money and man-hours.
5. It is less time-consuming. Since sampling is a study of a part of the population considerable time and labour are saved.

6. It is more adaptable than census enquiry. We can increase or decrease the number of samples to suit the purpose of the enquiry.
7. The size and methods of sampling are more scientific.
8. It is the only method that can be used in certain cases. Enquiries related to strength of bullets, the life of a lamp and quality control.
9. This method is often used to judge the accuracy of the information obtained by using census method.
10. The is also the possibility of determining the extent of sampling error and take necessary steps to reduce it.

Demerits

1. Careless planning and execution of sample survey may lead to inaccurate and misleading results.
2. Accuracy and reliability of the results obtained depends upon the representatives of the sample taken. But it is difficult to make a representative sample.
3. Sampling requires the services of trained and qualified personnel. It also requires sophisticated equipment for its planning, execution and analysis.
4. If the size of sample is a large proportion of the total population, the sampling method requires more time, labour and money.
5. If the size of the sample is inadequate, sampling may fail to indicate the true characteristics of the population.
6. There is every possibility for bias in the process of selection of sample.
7. The magnitude of sampling errors may become quite high if samples are not chosen in a proper way.

7) Explain the importance of sampling.

Meaning of Sampling

Sampling may be defined as the selection of some part of an aggregate or totally on the basis of which a judgment or inference about the aggregate or totally is made. It is a process of obtaining information about an entire population by examining only a part of it.

Importance or Role of Sampling in Statistics

Sampling plays a significant role in statistics. The details of which are stated below.

1. Sampling minimizes the time of the study: The sample is a representative of the population. Hence the number of variables to be studied is very minimum. So it requires minimum time for collection and analysis data.
2. Sampling reduces the cost of the study: Since the sample study reduces the number of variables the expenses in collection and analyse of data are also reduced to a considerable extent. Expenses toward visiting various places, printing questionnaires etc. are minimized due to sampling technique.
3. Sampling helps in collecting detailed information: Detailed information can be collected in sampling technique, since the number of respondents are very small. For example, a study is to be conducted with regard to the income and expenses of 5000 workers working in a village. Detailed information could not be collected if the study is conducted for all the 5000 workers.

But the information would be in detailed if a sample study is conducted by randomly selecting 100 workers.

4. Sample study gives more accurate results: Sample study may give more accurate result than a population study. This is because of various reasons. Population study requires lot of time and money. Hence detailed information may not be collected. In addition to this data once collected may become outmoded when the entire process of collection is over. In the sample study quality of data may be good since experts can be appointed to collect and process data.
5. Population study may not be possible in all areas: In some areas the population study may not be possible. Only sampling study could be conducted in these areas. When the population is infinite like the fishes in a sea, birds in a locality, number of trees in forest etc., it is impossible to conduct a population study. The only remedy is to select few samples of the population and to study the required character. In some areas though the population is finite it is difficult to conduct the study on Census Method. Study about the food habit of the people above 60 years old, in India.

UNIT-II

CLASSIFICATION, TABULATION, DIAGRAMS AND GRAPHS

1. Explain the meaning and objective of classification?

Meaning of objective:

Classification is grouping of items on the basis of some common characteristics. The data collected in any investigation is known as raw data. The process of arranging the data into groups or classes according to similarities and resemblances is called 'Classification'.

Objectives of classification:

The following are the important objectives of classification:

1. Classification helps to present the facts in a simple form;
2. To bring out clearly points of similarities and dissimilarities;
3. To facilitate comparison;
4. To give due weightage to various items;
5. To present a mental picture for easy understanding and remembrance;
6. To prepare the basis for tabulation;
7. To bring out relationship.

2. Briefly explain the different kinds [types] of classification?

Types of classification:

There are four types of classification. They are

1. Quantitative Classification
2. Qualitative Classification
3. Chronological Classification
4. Geographical Classification

1. Quantitative Classification:

When data are classified on the basis of some characteristics which can be measured it is 'quantitative classification'. For example, data collected for a statistical enquiry can be classified on the basis of variables like age, income, height, weight, hours of employment etc.

2. Qualitative Classification:

Classification of data on the basis of some attributes, qualities or characteristics, which cannot be measured, is called 'qualitative classification'. Qualitative classifications are of two kinds.

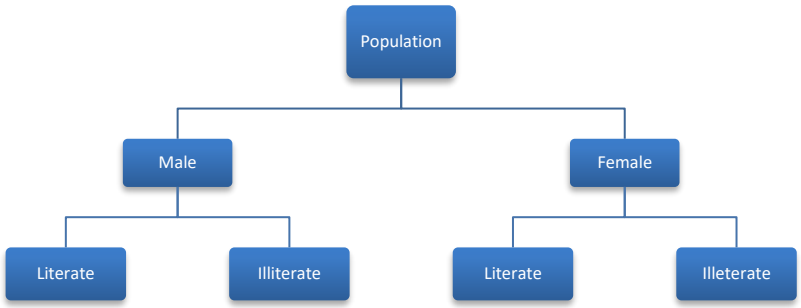
(i) Simple Classification
(ii) Manifold Classification

(i)Simple Classification

When classification is made on the basis of presence and absence of an attribute, it is 'simple classification'. Under this, the data are divided into two parts. For example, population may be divided into male and female on the basis of sex.

(ii)Manifold Classification

When classification is made on the basis of more than one attribute it is 'manifold classification'. For example, the population of a country can be classified on the basis of many attributes.



3. Chronological Classification

When classification is made on the basis of time (Chronology) it is 'chronological classification'. For example, population of India since 1981 can be classified as follows.

Year	Population (million)
1981	683.3
1991	846.3
2001	1027.0

4. Geographical Classification

When data are classified on the basis of place, area or locality it is 'geographical classification'.

Area	Population (million)	Percentage
Rural	74.20	72.20
urban	28.50	27.90
Total	102.70	100.00

3. State the rules [characteristics] of classification.

Data are classified according to some common features or objects in view. The classification should possess the following characteristics:

1. The system of classification must be exhaustive. There must be a class for each item of data in one of the classes.
2. Classified data must be mutually exclusive. Each item of data must find its place in one class only.
3. Classification must be proceeded in accordance with one principle. If a classification is not stable then data would not be fit for comparison.
4. A good classification should be flexible. It should be adjusted according to changing new circumstances.
5. The items included in the class should be homogeneous.
6. Classification of data should be made to suit the objects of the enquiry.
7. The total of the items included in different classes should tally with the total of the universe.

4. What are the meaning and objectives of tabulation?

Meaning of tabulation:

After classification the next step is tabulation of data. A systematic and logical arrangement of data in the form of rows and columns is called tabulation. Rows are horizontal arrangement whereas columns are vertical arrangement data. Tabulation presents maximum information in the minimum possible spece without sacrificing quality and usefulness of the data.

Objectives of data:

The following are the important objectives of tabulation.

1. To present mass of data in a simple form.
2. To facilitate comparison.
3. To help further analysis of data.
4. To create lasting impression in human mind.
5. To present facts in the minimum space.
6. To help in clarifying the characteristics of data.

5. Explain the different parts of a table.

Parts of a table:

The following are the important parts of a table.

1. Table number:

On the top of the table, we should put the table number. It is useful for reference and identification.

2. Title:

There must be a title of a table placed below the table number. It should be brief, clear and self-explanatory. It should answer the questions `what'? `Where'? and `When?'

3. Head note:

Any additional information which is not in the title can be written head note. For example, to denote the units we can write million or Rupees notes etc. Just below the title on the right-hand side we put the head note.

4. Caption:

The column heading is called `caption'. Columns should be numbered. The wording of caption should be brief. Each caption can be sub-divided into sub-caption.

5. Stub:

The row heading is called `stub'. It shows what is in each row of line of the body of table.

6. Body:

The body of the table contains the numerical information. They should be logically arranged. Intersection of the row and column of the body is called a `cell'. All numerical information is put in the cell. Item of special significance must be underlined or written in bold letters.

7. Footnotes:

At the bottom of a table, we have a footnote. It is used to call attention to some figure which, without note, might be misunderstood.

8. Source:

Below the footnote, we must give the source of the data. In source, we must give the name of the agency which has collected the data, name of the publication, exact page number etc.

STRUCTURE OF A TABLE

Table Number

Title

(Head Note)

Stub	Caption	Caption	Total
Stub entries	**BODY**		
Total			

Food Note

Source

6. What are the general rules of tabulation?

The following are the rules for the preparation of a table.

Rues of tabulation

The following are the rules for the preparation of table.

(i) Every table must have a table number.
(ii) Every table should have a title. Title should be brief, clear and self-explanatory.
(iii) The table should suit the size of the paper with more rows than columns.
(iv) The table should not be overloaded with information.
(v) Use of abbreviations should be avoided
(vi) Rounding of numbers should be done to avoid unnecessary details.
(vii) If any information is not available, we should write N.A.
(viii) The units of measurement should be given in the table.
(ix) The arrangement of the table should be logical and related items should be placed near each other
(x) A table should be neat, attractive and well balanced.

7) What is the importance of graph in statistics?

IMPORTANCE OF GRAPH:

1. Graphs give attractive, interesting and impressive view about the data.
2. They render complex data simple and easily understandable.
3. They are the simplest method of presenting statistical data. They save time and labour.
4. They make comparison between two or more data very easy.
5. They are also helpful to locate median, mode and quartiles graphically.

8) What are the meaning and limitations of Graph?

MEANING OF GRAPH:

When statistical data are presented in the form of lines and curves it is known as graphic presentation of data. Graphs are used to explain the relationship between different variables. They are mostly used for representing time series and frequency distributions are that the total area under a graph will be set to unity.

LIMITATIONS OF GRAPHS:

1. Graphs shows tendency and fluctuations, actual values are not known.
2. They can provide only limited amount of information.
3. They are subjective in nature. They may be interpreted differently by different people.
4. Graphs are not easy to construct

8 MARK

1) Briefly explain the different methods of classification.

Methods of Classification:

There are three methods of classification of data. They are:

(i) Inclusive method
(ii) Exclusive method
(iii) Open-end class method

(i)Inclusive method

When the upper limit of one class is included in the same class it is 'inclusive method'. For example,

Wage (Rs.)	No.of. Workers
50-59	10
60-69	25
70-79	30
80-89	20
90-99	15
Total	100

Under this method, a worker getting a wage of Rs.60/- will be included in the class 60-69. He will not be included in the first class (50-59). We have to convert this distribution by converting the inclusive classes into exclusive classes

(ii)Exclusive Method

When the upper limit of one class is used as lower limit of the next class, it is known as the 'exclusive method' of classification.

Marks	No.of. Students
0-25	7
25-50	18

50-75	10
75-100	5
Total	40

In this method, if a student gets 25 marks, he should be excluded from the first class and included in the second class. This method is useful for the construction of continuous frequency distribution. In the method, there is no need for conversion of class intervals.

(iii)Open and class method

If the lower limit of the first class and the upper limit of the last class are unknown it is known as 'Open-end class method'.

Marks	No.of. Students
Below 10	2
10-20	5
20-30	10
30-40	7
40 and above	36
Total	60

In this method, there is a need for finding the two unknown class limits. If the class intervals are in uniform size, we can fix the lower limit of the first class by looking into the class interval of the second class.

2) Explain the various types of diagrams and state their merits and demerits.

Types of Diagrams

The following are the common types of diagrams.

1. One-dimensional diagram (line and bar diagram)
2. Two-dimensional diagram (Rectangles, Squares and pie or circle diagrams)
3. Three-dimensional diagram (cubes, spheres, prisms, cylinders and blocks)
4. Pictograms
5. Cartograms

1. One dimensional Diagrams

When diagrams are drawn by taking into account only height or length, it is one-dimensional diagram. It is the most commonly used diagram. The height or length is proportional to the given frequency. One dimensional diagram may be of the following types,

(i) Line diagram
(ii) Simple-Bar diagram
(iii) Multi-Bar diagram
(iv) Sub-divided Bar
(v) Percentage Bar diagram
(vi) Other Bar diagram

2. Two Dimensional Diagrams

Diagrams are drawn by using both length as well as width are called two-dimensional diagrams. The magnitude of given values are represented by the 'area of the diagrams' or 'surface diagrams. The

following are the commonly used two-dimensional diagrams,

(i) Rectangles
(ii) Squares
(iii) Circles
(iv) Pie diagrams
(v) Pictograms
(vi) Cartograms

3. Three-dimensional diagram

When three dimensional, i.e., length, breadth and height are taken into account to draw a diagram, it is called 'Three-dimensional diagram'. It is also called as 'volume diagram'. The common forms of such diagrams are cubes, sphere, cylinders, blocks, etc. These diagrams are especially useful if there are very wide variations between the smallest and the largest magnitudes to be represented. Of these three-dimensional diagrams, 'cubes' are the simplest and most commonly used device.

4. Pictograms

When statistical data are presented through appropriate picture it is called, 'pictogram'. It is one of the popular attention of masses is to be drawn such as in exhibitions, fairs, pictograms are mostly used. They have more attraction value than other types of diagrams. They create a lasting picture in the minds of the viewers. But they are very difficult to construct. They give only an overall picture. They do not give minute details about the given data.

5. Cartograms

A simple type of diagram is not suitable for all purposes. Choice of diagrams requires great skill, intelligence and expertise. It mainly depends on the nature of data and the object of presentation. Utmost care should be given while choosing diagrams. Inappropriate selection might lead to very wrong and misleading interpretation.

Merits of Diagrams

1. Diagrams are attractive and impressive. They create lasting impression in the minds of the people.
2. They make data simple and intelligible. The mass of complex data presented in the form of diagram can be understood easily.
3. They make comparison between two sets of data possible.
4. They save time and labour. They give the basic characteristics clearly in minutes.
5. They have universal utility. They are widely used in economic, business, social and other fields.
6. They are important means of detecting mistakes in the computation of data.
7. They give more information. They reveal the existing trend and also how the trend changes.
8. Interpolation of values of the variables can be made easily

Demerits of Diagrams

Diagrammatic presentation of data has the following limitations.

1. Diagrams are useful to common man. Its utility to an expert is limited
2. They are poor substitutes for data.
3. They can show only appropriate values
4. They show only a limited amount of information

5. They cannot be used for further analysis
6. They can be drawn only by people having skill. It takes time to draw a diagram.
7. They are only means of drawing conclusion. One can make use of tables for detailed reference and diagrams for rapid understanding.
8. They do not prove or disapprove any fact.
9. There is every possibility of misusing diagrams.

3) Illustrate the advantages and disadvantages of a graph.

ADVANTAGES OF GRAPH:

1. Graphs give attractive, interesting and impressive view about the data.
2. They render complex data simple and easily understandable.
3. They are the simplest method of presenting statistical data. They save time and labour.
4. They make comparison between two or more data very easy.
5. They are also helpful to locate median, mode and quartiles graphically.

DISADVANTAGES OF GRAPHS:

1. Graphs shows tendency and fluctuations, actual values are not known.
2. They can provide only limited amount of information.
3. They are subjective in nature. They may be interpreted differently by different people.
4. Graphs are not easy to construct.

4) What are the meaning and General rules for making a diagram.

Meaning Of Diagrams:

The visual representation of statistical data is called ‘diagram’. Diagrams and graphs are the devices used to present statistical data in a convincing, appealing and understandable manner. They are usually in the form of lines, bars, squares, circles, cubes, pictures, maps or charts.

General Rules for making a diagram

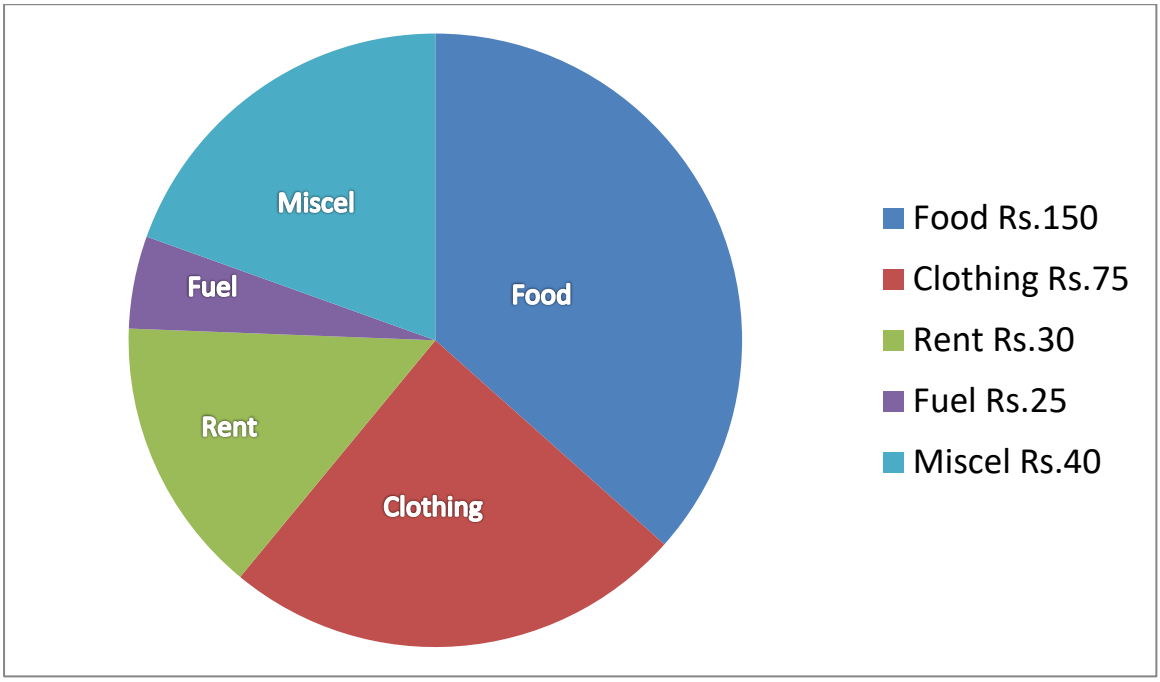
1. Diagrams should be neatly drawn and attractive. It should be appealing to the eyes.
2. Each diagram should be given a suitable title. The title should be brief, self-explanatory, clear and unambiguous.
3. The size of the diagram should match the size of the paper.
4. Diagrams should be drawn in a proper size to show clearly the necessary details. Its scale should be mentioned.
5. When many items are shown in a diagram, an index must be given for easy identification.
6. A footnote can be added (if necessary) so as to explain certain points clearly.
7. The measurement used for drawing a diagram should be accurate and proportional to the given data.
8. Each diagram should be numbered for ready reference
9. Statistical data represented in the form of diagrams should be homogenous
10. The original data on which diagram has been based should be given

11. A diagram drawn should be economical in time and energy.
12. If possible, provide enough space in all the sides of the diagram.

5) Draw a pie diagram for the following data.

Items	:	Food	Clothing	Rent	Fuel	Miscel
Monthly Expenses (Rs)	:	150	75	30	25	40

PIE DIAGRAM
MONTHLY EXPENDITURE



UNIT – III

MEASURES OF CENTRAL TENDENCY

5 MARK

1) Find Mean for the data given below

x:	45	55	60	65	75	80	90
f:	20	10	35	30	15	08	12

Solution:

X	f	d=x-a	Fd
45	20	-20	-400
55	10	-10	-100
60	35	-5	-175
A 65	30	0	0
75	15	10	150
80	8	15	120
90	12	25	300
	N=130		-675+570 ∑fd = -105

Arithmetic Mean $\overline{x}$ = A + $\frac{\sum fd}{N}$

Where,

A = Assumed Mean

N = Sum of frequencies

∑fd = Sum of the product of frequencies and the deviation of x

A = 65, N = 130, ∑fd = -105

$\bar{x}$ = 65 + $[\frac{-105}{130}]$

= 65 + (-0.80) = 65 – 0.80 = 64.2

Answer:

$\bar{x}$ = 64.2

2) Compute median from the following data

75, 52, 63, 45, 85, 69, 55, 30

Solution:

Arranging the given numbers in ascending order of magnitude.

30, 45, 52, 55, 63, 69, 74, 85

Where,

M = $\frac{n+1}{2}$ th item

M = median; n = number of items;

n=8

M = $\frac{8+1}{2}$ = $\frac{9}{2}$ = 4.5th item

Median lies between 4th and 5th item

M = $\frac{55+63}{2}$ = $\frac{118}{2}$ = 59

Answer:

Median = 59

3) From the following data, find out arithmetic mean and median.

Height (in cms)	**:**	**120**	**122**	**124**	**126**	**128**	**130**
No.of. students	**:**	**5**	**7**	**9**	**6**	**4**	**10**

Solution:

Height x	No.of.Students f	d=x-a	fd	Cf
120	5	-4	-20	5
122	7	-2	-14	12
A 124	9	0	0	21
126	6	2	12	27
128	4	4	16	31
130	10	6	60	41
	N = 41		-34 + 88 ∑fd = 54	

Arithmetic Mean $\overline{x}$ = A + $\frac{\sum fd}{N}$

Median:

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

A = 124; ∑fd = 54; N = 41

$\overline{x}$ = 124 + [$\frac{54}{41}$]= 124 + 1.31 = 125.31

$\overline{x}$ = 125. 31

M =the size of [$\frac{41+1}{2}$] th intem

= [$\frac{42}{2}$] th item

= 21th item is 124

M = 124

Answer:

$\overline{x}$ = 125.31

M = 124

4) Locate mode from the data given below.

x :	28	29	30	31	32	33
f :	10	20	40	65	50	15

Solution:

x	f
28	10
29	20
30	40
31	65
32	50
33	15

The mode (z) is equal to 31, since it has the maximum frequency of 65.

Answer:

Z = 31

5) Find the value of the mode from the data given below.

Weight (in kgs):	93-97	98-102	103-107	108-112	113-117	118-122	123-127	128-132
No.of. students:	2	5	12	17	14	6	3	1

Solution:

Weight (in kgs)	No.of. Students
93-97	2
98-102	5
103-107	12
108-112	17
113-117	14
118-122	6
123-127	3
128-132	1

Mode:

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

Where,

Z = Mode

L = lower limit of the model class

Δ_1 = different between frequency of the model class and frequency of the pre-model class

Δ_2 = difference between frequency of the model class and frequency of the post model class

C = Class interval

Model class is 108-112

L = 108; Δ_1 = 17-12 = 5; Δ_2 = 17-14 = 3; C = 4

Z = $108 + [\frac{5}{5+3} \times 4] = 108 + [\frac{5}{8} \times 4] = 108 + [0.625 \times 4] = 108 + 2.5 = 110.5$

Answer:

Z = 110.5

6. What are the qualities [character, requisites] of a good average?

A good average should possess the following characteristics or requisites:

1. An average should be rigidly defined.
2. It should be easy to calculate.
3. It should include all the items in the series.
4. It should not be affected by the values of the extreme items.
5. It should have sampling stability.
6. It should be helpful for further statistical calculations.

7. What are functions of an average?

An average is the precise and simple indicator of the central tendency of a group. The following are the functions of an average:

1. An average helps us to present the salient features of a mass of complex data.
2. Averages facilitate comparison.
3. The mean of a sample gives an idea about the mean of the population. Thus, it helps to know about the universe from a sample.
4. To know the mathematical relationship between groups or classes, an average becomes essential.
5. Averages are highly helpful in decision-making. They are valuable in setting standards, estimating and planning and other managerial decision-making.

8. State the characteristics of Arithmetic Mean?

Arithmetic Mean is the best measures of central tendency. It has the following merits:

31. Arithmetic mean is rigidly defined.
2. It is easy to understand.
3. It is very easy to calculate. An elementary knowledge of addition, multiplication and division is enough to calculate mean.
4. It takes into account all the items in the series. It gives them due weights.
5. It is a more stable measure of central tendency. It is least affected by fluctuations in sampling.
6. It is helpful for further statistical calculations. Correlation, regression, time series, index numbers, mean deviation, standard deviation, etc., rely upon arithmetic mean for their calculation.
7. It does not necessitate the arrangement or grouping of items as required for finding median and mode.

9. List down the advantages of mode?

The following are the advantages of mode:

1. Mode is easy to understand.

2. It is easy to calculate.
3. It is not affected by the value of the extreme values in the series.
4. It can be located by mere inspection in the case of raw data and discrete series.
5. It can be calculate even when there are open-end classes.
6. It can also be located graphically.
7. When mode is ill-defined, it can also be find out by using empirical formula.

8 Mark

1. Calculate Arithmetic Mean, Median and Mode

Class interval: 10-20 20-30 30-40 40-50 50-60 60-70 70-80
Frequency : 24 15 36 41 14 17 5
Solution:

Class Interval	f	Mid X	d'=x-A	fd' fxd'	cf
10-20	24	15	-3	-72	24
20-30	15	25	-2	-30	39
30-40	36	35	-1	-36	75
40-50	41	45	0	0	116
50-60	14	55	1	14	130
60-70	7	65	2	34	147
70-80	5	75	3	15	152
	N = 152			∑fd' = -75	

Arithmetic mean $\overline{x}$ = A + $\frac{\sum fd'}{N}$ x C

Median M = L + [$\frac{\frac{N}{2}-m}{f}$ x C]

Mode Z = l + [$\frac{\Delta 1}{\Delta 1+\Delta 2}$ x C]

$\overline{x}$ = A + $\frac{\sum fd'}{N}$ x C

A = 45; C = 10; N = 152; ∑fd' = -75

$\overline{x}$ = 45 + $\frac{-75}{152}$ x 10
= 45 + (-0.49) x 10
= 45- 0.49 x 10
= 45 – 4.9

$\overline{x}$ = 40.1

M = l + [$\frac{\frac{N}{2}-m}{f}$ x C]

$\frac{N}{2} = \frac{152}{2} = 76$

$\frac{N}{2}$ = 76, Median class 40-50

L = 40; m = 75; f = 41; c = 10

= [40 + $\frac{76-75}{41}$ x 10]

= [40 + $\frac{1}{41}$ x 10]

= 40 + (0.024 x 10)

= 40 + 0.24

M = 40.24

Z = L + [$\frac{\Delta 1}{\Delta 1 + \Delta 2}$ x C]

Model class = 40-50; l = 40; Δ_1 = 5; Δ_2 = 27; C = 10

= 40 + $\frac{5}{5+27}$ x 10

= 40 + $\frac{5}{32}$ x 10

= 40 + 0.15 x 10

= 40 + 1.5

Z = 41.5

Answer:

$\overline{x}$ = 40.1

M = 40.24

Z = 41.5

2) Calculate mean, median and mode from the following data.

Central size:	15	25	35	45	55	65	75	85
Frequency:		05	09	13	21	20	15	08	03

Solution:

X	f	d = x-A	fd	cf
15	5	-40	-200	5
25	9	-30	-270	14
35	13	-20	-260	27
45	21	-10	210	48
55	20	0	0	68
65	15	10	150	83
75	8	20	160	91
85	3	30	90	94
	N = 94	d = 40	-120	

Arithmetic Mean $\bar{x}$ = A + $\frac{\sum fd}{N}$

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

Mode Z = the value that corresponds to the maximum frequency is the modal value

$\bar{x}$ = A + $\frac{\sum fd}{N}$

A = 55; $\sum fd$ = -120; N = 4

$\bar{x}$ = 55 + $\frac{-120}{94}$

= 55 - 1.276

$\bar{x}$ = 53.72

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

= the size of $(\frac{94+1}{2})$th item

= the size of $(\frac{95}{2})$th item

= the size of 47.5^{th} item

M = 45

Z = 45

Answer:

$\bar{x}$ = 53.72

M = 45

Z = 45

3) Find model mark from the data given below.

Marks :	0-20	20-40	40-60	60-80	80-100
No.of.students:	6	12	18	10	4

Solution:

Marks Class Interval	No. of. Students
0-20	6
20-40	12
40-60	18
60-80	10
80-100	4

Mode:

To find mode, locate mode class which corresponds to highest frequency.

Z = L + $\frac{\Delta 1}{\Delta 1+\Delta 2}$ x C

L = 40; Δ_1 = 6; Δ_2 = 8; C = 20

= 40 + $\frac{6}{6+8}$ x 20 = 40 + $\frac{6}{14}$ x 20 = 40 + (0.428 x 20) = 40 + 8.56

Z = 48.56

UNIT-IV

DISPERSION

5 Marks

1) Calculate range and it's co-efficient from the data given below

60, 70 50, 45, 38, 58, 63

Solution:

R = L- S

L = Largest value

S = Smallest value

L = 70;S = 38

R = 70 – 38

= 32

Co-efficient of Range = $\frac{L-S}{L+S}$ = $\frac{70-38}{70+38}$= $\frac{32}{108}$ = 0.29

Answer:

Range = 32

Co-efficient of Range = 0.29

2)Find co-efficient of variation from the data given below.

N = 8; $\sum x = 112$; $\sum x^2 = 16$

Solution:

C.V = $\frac{\sigma}{\bar{x}}$ x 100

$\bar{x} = \frac{\sum x}{n}$ = $\frac{112}{8}$ = 14

σ (S.D) = $\sqrt{\frac{\sum x^2}{n}}$ = $\sqrt{\frac{16}{8}}$ = $\sqrt{2}$ = 1.41

$\bar{x}$ = 14; σ (S.D) = 1.41

C.V = $\frac{1.41}{14}$ x 100 = 0.1007 x 100 = 10.07

Answer:

C.V = 10.07

3)Compute co-efficient of quartile deviation from the following data.

x :	20	30	45	50	60	70	80
y :	03	61	132	153	140	51	03

Solution:

x	f	Cf
20	3	3
30	61	64

45	132	196
50	153	349
60	140	489
70	51	540
80	3	543
	543	

Quartile Deviation:

Q.D = $\frac{Q_3 - Q_1}{2}$

Q_3 = the size of 3 $(\frac{N+1}{4})$th item

= the size of 3 $(\frac{543+1}{4})$th item

= the size of 3 $(\frac{544}{4})$th item

= the size of 3 (136)th item

= the size of 408th item

408^{th} item = 60

Q_3 = 60

Q_1 = the size of $(\frac{N+1}{4})$th item

= the size of $(\frac{543+1}{4})$th item

= the size of $(\frac{544}{4})$th item

= the size of 136^{th} item

136^{th} item = 45

Q_1 = 45

Q.D = $\frac{60-45}{2}$

= $\frac{15}{2}$

= 7.5

Q.D = 7.5

Co-efficient of Quartile = $\frac{Q_3 - Q_1}{Q_3 + Q_1}$

= $\frac{60 - 45}{60 + 45}$

= $\frac{15}{105}$

= 0.142

Answer:

Co-efficient of Quartile = 0.142

4) Calculate co-efficient of quartile deviation for the data given below.

x	**:**	**30-32**	**32-34**	**34-36**	**36-38**	**38-40**	**40-42**	**42-44**
y	**:**	**12**	**18**	**16**	**14**	**12**	**8**	**6**

Solution:

X	f	Cf
30-32	12	12
32-34	18	30
34-36	16	46
36-38	14	60
38-40	12	72
40-42	8	80
42-44	6	86
	N = 86	

Quartile Deviation:

Q.D = $\frac{Q_3 - Q_1}{2}$

Q_1 = L +[$\frac{\frac{N}{4} - m}{f}$ x C]

$\frac{N}{4} = \frac{86}{4}$ = 21.5; L = 32;f = 12; m = 12; c = 2;

Q_1 = 32 +[$\frac{\frac{21.5}{4} - 12}{12}$ x 2]

= 32 +[$\frac{91.5}{12}$ x 2] = 32 + 0.79 x 2 = 32 + 1.58 = 33.58

Q_1 = 33.58

Q_3 = L + [$\frac{3\left(\frac{N}{4}\right) - m}{f}$ x C]

= 3 x $\frac{N}{4}$ = 3 x 21.5 = 64.5

$3(\frac{N}{4})$ = 64.5; m = 60; f = 14; c = 2; L = 38

Q_3 = 38+ [$\frac{64.5 - 60}{14}$ x 2]

= 38+ [$\frac{4.5}{14}$ x 2]

= 38 + 0.32 x 2

= 38 + 0.64

Q_3 = 38.64

Quartile Deviation = $\frac{38.64 - 33.58}{2}$ = $\frac{5.06}{2}$ = 2.53

Q.D = 2.53

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

$= \frac{38.64 - 33.58}{38.64 + 33.58} = \frac{5.06}{72.22}$ = 0.07

Answer:

Co-efficient of Q.D = 0.07

5)Compute Arithmetic mean and co-efficient of Mean Deviation from the following data.

Class	**:**	**0-10**	**10-20**	**20-30**	**30-40**	**40-50**	**50-60**	**60-70**	**70-80**
frequency	**:**	**5**	**8**	**12**	**15**	**20**	**14**	**12**	**6**

Solution:

Class Interval	Frequency f	Mid x	d' = $\frac{x-A}{c}$	Fd'	$\|d\|$ = x - $\bar{x}$	f$\|d\|$
0-10	5	5	-3	-15	37	185
10-20	8	15	-2	-16	27	216
20-30	12	25	-1	-12	17	204
30-40	15	35 A	0	0	7	105
40-50	20	45	1	20	3	60
50-60	14	55	2	28	13	182
60-70	12	65	3	36	23	276
70-80	6	75	4	24	33	198
	N = 92			∑fd' = 65		∑f$\|d\|$ = 1426

Mean:

$\bar{x}$ = A + $\frac{\sum fd'}{N}$ x C

A = 35; ∑fd' = 65; N = 92; c= 10

$\bar{x}$ = 35 + $\frac{65}{92}$ x 10 = 35 + [0.70 + 10] = 35 + 7 = 42

$\bar{x}$ = 42

Mean Deviation:

Mean deviation from mean = $\frac{\sum f|d|}{N}$

$|d|$ = x - $\bar{x}$

= $\frac{1426}{92}$ = 15.5

Mean deviation from Mean = 15.5

Co-efficient of Mean deviation from Mean:

Co-efficient of Mean deviation of Mean = $\frac{Mean\ deviation}{\bar{x}}$

$$= \frac{15.5}{42}$$
$$= 0.36$$

Answer:

Co-efficient of Mean deviation from Mean = 0.36

6. What are the demerits of range?

Demerits of range:

1. It is not based on all the items in the series.
2. It is very much affected by the presence of an extremely high or low value.
3. It cannot be calculated from grouped frequency distribution with open-end Classes.
4. It does not tell anything about the characteristics of the series.

7. What are the qualities of a good measure of dispersion?

A good measure of dispersion should possess the following characteristics:

1. It should be rigidly defined.
2. It should be easy to understand.
3. It should be easy to calculate.
4. It should include all the items in the series.
5. It should not be affected by the values of a extreme items.
6. It should have sampling stability.
7. It should be amenable for further mathematical treatment.

8 MARK

1. Explain the different methods used to measure dispersion.

(Or)

Explain the various types of measures of dispersion.

Meaning of Dispersion:

According to A.L. Bowley, "Dispersion is the measure of the variation of the items".

Measures of Dispersion:

The various measures of dispersion can be classified into two:

(a) Absolute Measures and
(b) Relative Measures

Absolute measures are expressed in terms of the original units of a series. They are not suitable for comparing the variability of two distributions which are expressed in different units of measurement. The absolute measures of dispersion are:

(i) Range
(ii) Quartile Deviation
(iii) Mean Deviation

(iv) Standard Deviation

Relative measures of dispersion are expressed as ratios or percentages. They are independent of units of measurement. They are useful for comparing the variability of two distributions. The relative measures of dispersion are:

(i) Co-efficient of Range
(ii) Co-efficient of Quartile Deviation
(iii) Co-efficient of Mean Deviation
(iv) Co-efficient of Standard Deviation
` (v) Co-efficient of Variation

Range:

Range is the simplest method of studying dispersion. It is the difference between the largest item and the smallest item in the series. In symbol, R = L – S

Co-efficient of Range:

The relative measure of Range is called the 'Co-efficient of Range'.

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

Quartile Deviation:

Quartile Deviation is an absolute measure of dispersion. It is defined by the formula

Q.D = $\frac{Q_3 - Q_1}{2}$

Co-efficient of Quartile Deviation:

The relative measure of quartile deviation is called as co-efficient of quartile deviation. In symbol,

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

Mean Deviation:

The mean deviation of a series of value is the arithmetic mean of their absolute deviations. The deviations may be taken from mean or median or mode. Usually, deviations are taken from median to find mean deviation. While computing mean deviation the algebraic sign of the deviations are ignored (|d|).

Co-efficient of Mean Deviation:

(i)Co-efficient of M.D from Mean = $\frac{M.D}{Mean}$

(ii)Co-efficient of M.D from Median = $\frac{M.D}{Median}$

Standard Deviation:

Standard deviation is the most important measure of dispersion. It is the square-root of the arithmetic mean of the squares of deviation. It is also called as 'root-mean-square deviation'. It is an absolute measure of dispersion. It is denoted by 'S.D' or $\sigma. The$ relative measure of standard deviation is called the 'co-efficient of standard deviation'.

Co-efficient of Standard Deviation:

Co-efficient of S.D = $\frac{\sigma}{\bar{x}}$

Co-efficient of Variation:

Co-efficient of variation is the ratio between Standard deviation and Arithmetic mean expressed in percentages. It can be obtained by dividing Standard deviations by Arithmetic mean. The quotient multiplied by 100 gives the value of C.V. It is a relative measure of dispersion. In symbol,

C.V = $\frac{\sigma}{\bar{X}}$ x 100

2. Explain the advantages and disadvantages of quartile deviation.

Or

What are the merits and demerits of quartile deviation?

Merits of Quartile Deviation:

1. Quartile Deviation is simple to understand.
2. It is easy to calculate.
3. It is not affected by the values of the extreme items.
4. It is useful to find the skewness of a distribution.
5. It can be calculated from a grouped frequency distribution with open-end

Classes.

Demerits of Quartile Deviation:

1.It is not rigidly defined.

2. It is not based on all the items in the series.

3. It does not take into account the variation of each item about the central value.

4. Its value is not stable. It is affected by fluctuations in sampling.

5. It is not useful for further mathematical treatment compared to standard deviation.

3) Calculate the Standard Deviation and co-efficient of standard deviation and co-efficient of variation from the following.

Marks :	**10**	**80**	**35**	**65**	**60**
Students:	**02**	**04**	**05**	**06**	**07**

Solution:

X	f	d = x - A	d^2	fd^2 fxd^2	fd
10	2	-25	625	1250	-50

80	4	45	2025	8100	180
A 35	5	0	0	0	0
65	6	30	900	5400	180
60	7	25	625	4375	175
	N = 24			$\sum fd^2 = 19125$	$\sum fd = 305$

Standard Deviation:

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

('A' will be taken at middle of x)

$\sum fd^2 = 19125$; N = 24; $\sum fd = 305$

$\sigma = \sqrt{\frac{19125}{24} - (\frac{305}{24})^2}$

$= \sqrt{796.875 - (12.708)^2}$

$= \sqrt{796.875 - 161.49}$

$= \sqrt{635.385}$

$\sigma = 25.20$

Co-efficient of Variation:

Co-efficient of Variation = $\frac{\sigma}{\bar{x}}$ x 100

$\bar{x}$ =A + $\frac{\sum fd}{N}$

A = 35; N = 24; $\sum fd = 305$

= 35 + $\frac{305}{24}$

= 35 + 12.70

$\bar{x} = 47.70$

$\sigma = 25.20$

Co-efficient of variation = $\frac{25.50}{47.70}$ x 100

= 0.528 x 100

Co-efficient of Variation = 52.8

Co-efficient of Standard deviation:

Co-efficient of $\sigma = \frac{\sigma}{\bar{x}}$

$\sigma = 25.50$; $\bar{x} = 47.70$;

$\sigma = \frac{25.20}{47.70}$

= 0.528

$\sigma = 0.58$

Answer:

σ = 0.58

Co-efficient of variation = 52.8

Co-efficient of Standard Deviation = 0.528

4) Find the standard deviation and co-efficient of variation for the following frequency distribution.

X: 10-20 20-30 30-40 40-50 50-60 60-70 70-80

F: 05 12 15 20 10 04 02

Solution:

x	F	Mid x'	d' = $\frac{x-A}{C}$	d'^{2} d' x d'	fd' f x d'	fd'^{2} f x d'^{2}
10-20	5	15	-3	9	-15	45
20-30	12	25	-2	4	-24	48
30-40	15	35	-1	1	-15	15
40-50	20	A 45	0	0	0	0
50-60	10	55	1	1	10	10
60-70	4	65	2	4	8	16
70-80	2	75	3	9	6	18
	N = 68				∑fd' = -30	∑fd'^{2} = 152

Standard Deviation:

$\sigma = \sqrt{\frac{\sum fd^2}{N} - (\frac{\sum fd}{N})^2} \text{x C}$

N = 68; C = 10 ∑fd' = -30; ∑fd'^{2} = 152;

$\sigma = \sqrt{\frac{152}{68} - (\frac{-30}{68})^2} \text{x } 10$

$= \sqrt{2.235 - (-0.44)^2} \text{x } 10$

$= \sqrt{2.235 - 0.1936} \text{ x } 10$

$= \sqrt{2.0414} \text{ x } 10$

= 1.428 x 10

σ = 14.28

Co-efficient of Variation:

Co-efficient of Variation = $\frac{\sigma}{\bar{x}}$ x 100

$\bar{x}$ =A + $\frac{\sum fd'}{N}$ x C

A = 45; N = 68; C = 10; ∑fd' = -30

= 45 + $\frac{-30}{68}$ x 10

= 45 + (-0.441) x 10

= 45 – (0.441 x 10)

= 45 – 4.41

= 40.59

$\bar{x}$ = 40.59

Co-efficient of Variation $= \frac{14.28}{40.59}$ x 100

= 0.3518 x 100

= 35.181

Co-efficient of Variation = 35.18

UNIT-V

SKEWNESS AND KURTOSIS

5 Marks

1) Give the meaning for skewness.

Skewness refers to lack of symmetry in a distribution. When a symmetric is not symmetrical it is said to be symmetrical it is said to be asymmetrical or skewed. Any deviation from symmetry is called skewness. A distribution can be either positively skewed or negatively skewed. When a distribution is skewed to the right it is a positively skewed distribution. A distribution is positively skewed if,

Mean > Median > Mode

When a distribution is skewed to the left it is negatively skewed distribution. A distribution is negatively skewed if,

Mean < Median < Mode

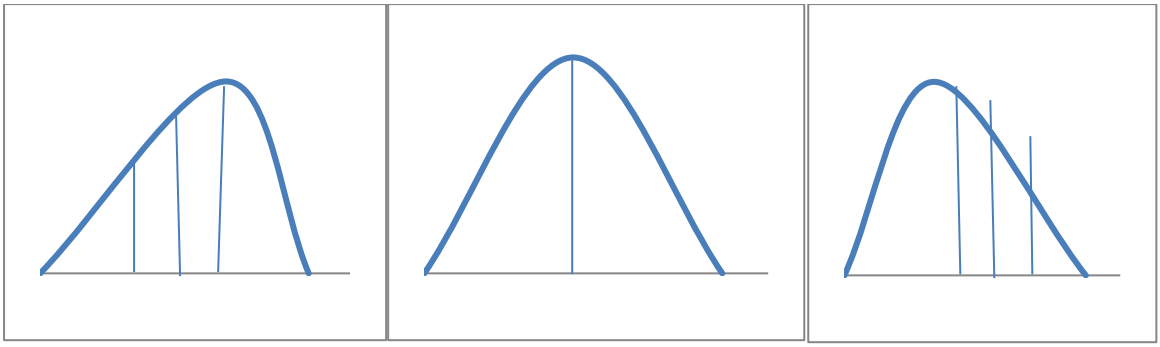
In a symmetrical distribution, the mean, median and mode coincide. A distribution is said to be symmetrical if

Mean = Median = Mode.

x	f
1	2
2	4
3	8
4	19
5	2

x	f
1	2
2	8
3	15
4	8
5	2

x	F
1	2
2	19
3	8
4	4
5	2



$\bar{x}$ M Z

Negative Skewness

$\bar{x}$ = M = Z

Symmetrical Distribution

$\bar{x}$ M Z

Positive Skewness

2) Explain the four types of moments

Moments:

The term 'Moment' in mechanic refers to the turning or the rotating effect of a force. In statistics it is popularly used to describe the characteristic of a distribution. It is useful in measuring the central tendency of a set of values, their scatterness (dispersion), their asymmetry and the peakness of the curve.

For a frequency distribution we imagine that at a distance 'x' from the origin 'o' is equal to fx. The total is ∑fx. To correct the number of items involved we divide ∑fx by ∑f. The moments about the actual mean are denoted by the Greek letter μ (mu). To find moments, deviations are taken from actual arithmetic mean.

There are four moments about mean. They are called as 'central moments'. They are defined as follows:

Moments	Raw data	Frequency Distribution
First Moment	$\mu_1 = \frac{\sum(X_i - \bar{x})}{N}$	$\mu_1 = \frac{\sum f(X_i - \bar{x})}{N}$ or $\frac{\sum fdx}{N}$
Second Moment	$\mu_2 = \frac{\sum(X_i - \bar{x})^2}{N}$	$\mu_2 = \frac{\sum f(X_i - \bar{x})^2}{N}$ or $\frac{\sum fdx^2}{N}$
Third Moment	$\mu_3 = \frac{\sum(X_i - \bar{x})^3}{N}$	$\mu_3 = \frac{\sum f(X_i - \bar{x})^3}{N}$ or $\frac{\sum fdx^3}{N}$
Fourth Moment	$\mu_4 = \frac{\sum(X_i - \bar{x})^4}{N}$	$\mu_4 = \frac{\sum f(X_i - \bar{x})^4}{N}$ or $\frac{\sum fdx^4}{N}$

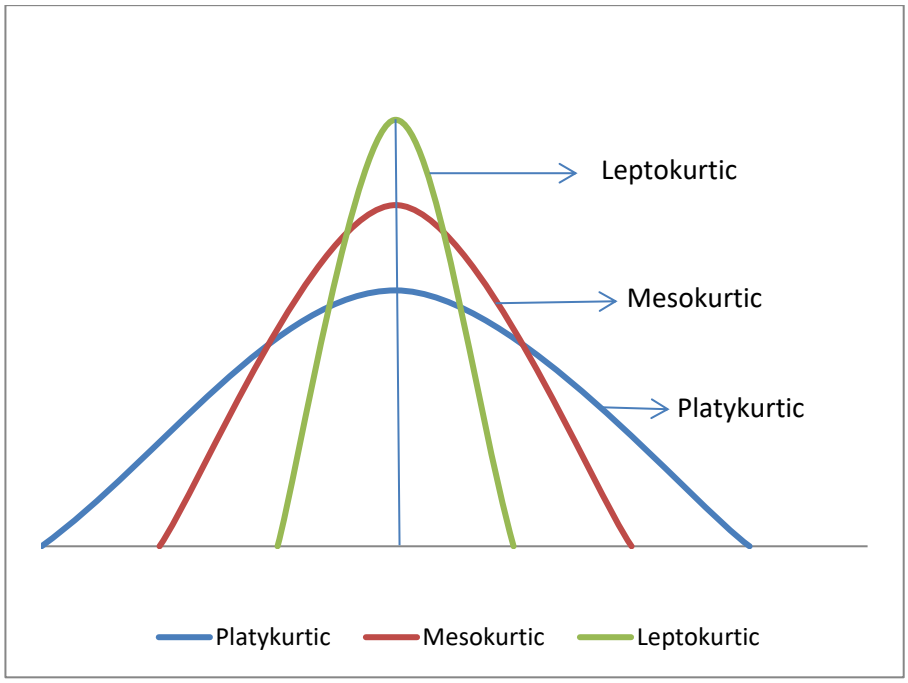
3) Write a note on Kurtosis.

Kurtosis

The term 'Kurtosis' is used to describe the peakness of a curve. Kurtosis refers to the degree of flatness or peakeness in the region about the mode of the frequency curve. Kurtosis points at the nature of distribution of items in the middle of the series. It is measured relative to the peakeness of a normal curve. If a curve is more peaked than a normal curve it is known as 'Leptokurtic'. If a curve is more flat than a normal curve it is known as 'Platykurtic'. The normal curve is known as 'Mesokurtic'.

In a Leptokurtic distribution, most of the frequencies are around the mean, median or mode. The sign of mean, median or mode is large. In a Mesokurtic distribution, most of the frequencies follow particular trend from mean, mean and mode. The mean, median and mode coincide ($\bar{x}$ = M = Z). This type of distribution is known as normal distribution. In a Platykurtic distribution most of the frequencies are spread for away from the mean, median or mode. The frequency curve will be much flatter than the other two curves and closer to OX – axis.

Three curves with the same central location but different Kurtosis.



Measurement of Kurtosis:

Kurtosis is measured by β_2 or its derivative γ_2

$\beta_2 = \frac{\mu_4}{\mu^2 2} = \frac{\mu_4}{\sigma^4}$ and $\gamma_2 = \beta_2 - 3$

For a normal curve $\beta_2 = 3$ (or $\gamma_2 = 0$). For a leptokurtic curve, $\beta_2 > 3$ (or $\gamma_2 > 0$) and for a platykurtic curve $\beta_2 < 3$ (or $\gamma_2 < 0$).

Kurtosis is mainly used in biological studies. It is of not much utility so far as sociological studies are concerned.

4) Find the Bowley's co-efficient of skewness

Division: **0-3** **3-6** **6-9** **9-12** **12-15**

Frequency: **11** **15** **12** **22** **10**

Solution:

Class Interval	f	Mid x	cf	d' = $\frac{x-A}{c}$	fd'	d'^{2}	Fd'^{2}
0-3	11	1.5	11	-2	-22	4	44
3-6	15	4.5	26	-1	-15	1	15
6-9	12	7.5	38	0	0	0	0
9-12	22	10.5	60	1	22	1	22
12-15	10	13.5	70	2	20	4	44
	N= 70				∑fd' = 5		∑fd'^{2} = 121

Karl Pearson's co-efficient of Skewness:

$\bar{x}$ = A + $\frac{\sum fd\prime}{N}$ x C

N =70; A = 7.5; C = 3; $\sum$fd' = 5

$\bar{x}$ = 7.5 + ($\frac{5}{70}$ x 3)

= 7.5 + (0.071 x 3)

= 7.5 + 0.214

$\bar{x}$ = 7.714

Median = L + [$\frac{\frac{N}{2}-m}{f}$ x c]

$\frac{N}{2}$ = $\frac{70}{2}$ = 35; L = 6; f = 12; c = 3

M = 60 + [$\frac{35-26}{12}$ x 3]

= 6 + $\frac{9}{12}$ x 3

= 6 + (0.75 x 3)

= 6 + 2.25

M = 8.25

σ = $\sqrt{\left(\frac{\sum fd'^2}{N}\right) - (\frac{\sum fd\prime}{N})^2}$ x 100

N = 70; $\sum$fd' = 5; $\sum$fd'^{2} = 121

= $\sqrt{\left(\frac{121}{5}\right) - (\frac{5}{5})^2}$ x 100

= $\sqrt{24.2 - 1}$ x 100

=$\sqrt{23.2}$ x 100

= 4.8166 x 100

σ = 481.66

Co-efficient of Skewness = $\frac{3\,(7.714-8.25)}{481.66}$

= $\frac{3\,(-0.536)}{481.66}$

= $\frac{1.608}{481.66}$

= -0.0033

Answer:

Co-efficient of Skewness = -0.0033

5) Find the value of β_2 for the data given below.

N = 100; $\sum fd = 25$; $\sum fd^2 = 195$; $\sum fd^3 = 296$; $\sum fd^4 = 865$;

Solution:

$\beta_2 = \frac{\mu_4}{\mu_2^2}$

$\mu_4 = \frac{\sum f dx^4}{N}$

$\mu_2 = \frac{\sum f dx^2}{N}$

$\sum fd^4 = 865$; $\sum fd^2 = 195$; N = 100

$\mu_4 = \frac{865}{100}$ = 8.65

$\mu_2^2 = \frac{(195)^2}{100}$ = $\frac{18025}{100}$ = 380.25

$\beta_2 = \frac{8.65}{380.25}$ = 0.022

Answer:

β_2 = 0.022

8 Marks

1) Explain the various types of skewness.

(or)

Explain the measures and tests of skewness.

Skewness refers to lack of symmetry in a distribution. When a symmetric is not symmetrical it is said to be symmetrical it is said to be asymmetrical or skewed. Any deviation from symmetry is called skewness. A distribution can be either positively skewed or negatively skewed. When a distribution is skewed to the right it is a positively skewed distribution. A distribution is positively skewed if,

Mean > Median > Mode

When a distribution is skewed to the left it is negatively skewed distribution. A distribution is negatively skewed if,

Mean < Median < Mode

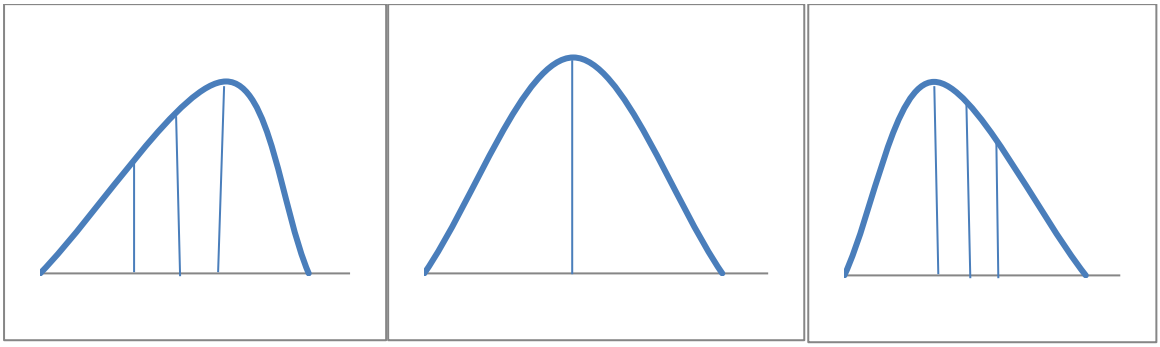
In a symmetrical distribution, the mean, median and mode coincide. A distribution is said to be symmetrical if

Mean = Median = Mode.

x	f
1	2
2	4
3	8
4	19
5	2

x	f
1	2
2	8
3	15
4	8
5	2

x	F
1	2
2	19
3	8
4	4
5	2



$\bar{x}$ M Z $\bar{x}$ = M = Z $\bar{x}$ M Z

Negative Skewness Symmetrical Distribution Positive Skewness

Features:

A skewed distribution will have the following features.

1. The value of mean, median and mode do not coincide in a skewed distribution.
2. The Quartiles are not equi-distant from median.
3. The sum of the positive deviations from median is not equal to the sum of the negative deviations.
4. When the data are not plotted on a graph, they do not give a normal bell-shaped curve.
5. The frequencies are not similarly distributed on either side of the mode.

Measures of Skewness:

A measure of skewness gives a numerical expression and the direction of asymmetry in a distribution. It gives information about the degree of variation on either side of mode.

Absolute Skewness = Mean – Mode or

$= \bar{X} - Z$

Relative Measures of Skewness

Karl Pearson's Co-efficient of Skewness

$Sk_P = \frac{\bar{X}-Z}{\sigma}$

Where,

Sk_P = Karl Pearson's Co-efficient of skewness

$\bar{X}$ = Arithmetic Mean

Z = Mode

σ = Standard Deviation

Bowley's Co-efficient of Skewness

Bowley's measure of skewness is based on positional measures like Q_1, Q_2, Q_3 (Median) and Q_3. So this method is also called as 'Quartile Co-efficient of Skewness'. This method can be better used when mode is ill-defined or in the case of distribution with open-end classes.

The formula is,

$$Sk_B = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1} \quad \text{or} \quad \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}$$

Where,

Sk_B = Bowley's co=efficient of skewness

Q_1, Q_2, Q_3 = First, second and third Quartile

M = Median (Q_3)

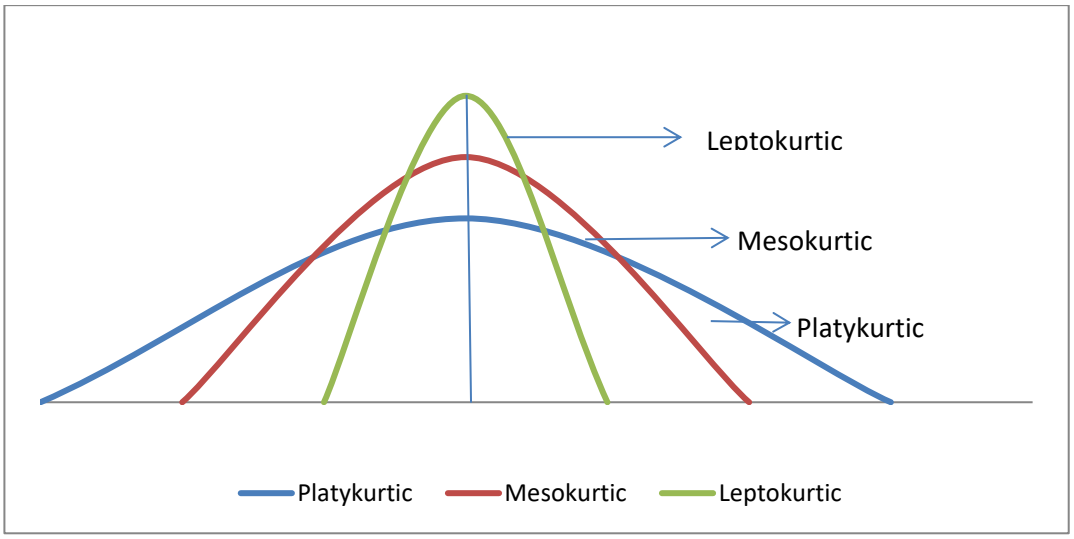
2) Explain the measures of kurtosis.

Kurtosis

The term 'Kurtosis' is used to describe the peakness of a curve. Kurtosis refers to the degree of flatness or peakeness in the region about the mode of the frequency curve. Kurtosis points at the nature of distribution of items in the middle of the series. It is measured relative to the peakedness of a normal curve. If a curve is more peaked than a normal curve it is known as 'Leptokurtic'. If a curve is more flat than a normal curve it is known as 'Platykurtic'. The normal curve is known as 'Mesokurtic'.

In a Leptokurtic distribution, most of the frequencies are around the mean, median or mode. The sign of mean, median or mode is large. In a Mesokurtic distribution, most of the frequencies follow particular trend from mean, mean and mode. The mean, median and mode coincide ($\bar{x}$ = M = Z). This type of distribution is known as normal distribution. In a Platykurtic distribution most of the frequencies are spread for away from the mean, median or mode. The frequency curve will be much flatter than the other two curves and closer to OX – axis.

Three curves with the same central location but different Kurtosis.



Measurement of Kurtosis:

Kurtosis is measured by β_2 or its derivative γ_2

$\beta_2 = \frac{\mu_4}{\mu^2 2} = \frac{\mu_4}{\sigma^4}$ and $\gamma_2 = \beta_2 - 3$

For a normal curve $\beta_2 = 3$ (or $\gamma_2 = 0$). For a leptokurtic curve, $\beta_2 > 3$ (or $\gamma_2 > 0$) and for a platykurtic curve $\beta_2 < 3$ (or $\gamma_2 < 0$).

Kurtosis is mainly used in biological studies. It is of not much utility so far as sociological studies are concerned.

3) Calculate Pearson's co-efficient of skewness from the data given below.

Size	**:**	**4**	**6**	**8**	**10**	**12**	**14**	**16**
Frequency	**:**	**2**	**4**	**5**	**03**	**02**	**01**	**04**

Solution:

x	f	d = x - A	fd	d^2	fd^2	cf
4	2	-6	-12	36	72	2
6	4	-4	-16	16	64	6
8	5	-2	-10	4	20	11
10	3	0	0	0	0	14
12	2	2	4	4	8	16
14	1	4	4	16	16	17
16	4	6	24	36	144	21
	N = 21		$\sum fd = -6$		$\sum fd^2 = 324$	

Pearson's Co-efficient of skewness:

Co-efficient of Skewness = $\frac{3\,(\bar{x}-M)}{\sigma}$

$\bar{x}$ = A + $\frac{\sum fd}{N}$

A = 10; N = 21; $\sum$fd = -6

= 10 + $\frac{-6}{21}$ = 10 – 0.28

$\bar{x}$ = 9.72

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

= the size of $\frac{21+1}{2}$ th item

= the size of $\frac{22}{2}$ th item

= the size of 11th item

11^{th} item is 8

M = 8

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

$= \sqrt{\frac{324}{21} - (\frac{-6}{21})^2}$

$= \sqrt{15.42 - (0.285)^2}$

$= \sqrt{15.42 + 0.081}$

$= \sqrt{15.501}$

σ = 3.937

Co-efficient of skewness = $\frac{3\,(9.72-8)}{3.937}$ = $\frac{3 \text{ x } 1.72}{3.937}$ = $\frac{5.16}{3.937}$

= 1.310

Co-efficient of skewness = 1.310

4) Find Bowley's co-efficient of skewness for the following data

Size	**: 15-20**	**20-25**	**25-30**	**30-35**	**35-40**	**40-45**	**45-50**	**50-55**
Frequency	**: 13**	**29**	**46**	**60**	**112**	**94**	**45**	**21**

Solution:

Size	Frequency	Cf
15-20	13	13
20-25	29	42
25-30	46	88
30-35	60	148
35-40	112	260
40-45	94	354
45-50	45	399
50-55	21	420
	N = 420	

$Sk_B = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}$

$Q_1 = L + [\frac{\frac{N}{4} - m}{f} \text{ x } c]$

$\frac{N}{4} = \frac{420}{4} = 105$

Class interval = 25-30

L = 25; $\frac{N}{4}$ = 105; m = 29; f = 46; c = 5

$= 25 + [\frac{105-29}{46} \text{ x } 5]$

$= 25 + [\frac{76}{46} \text{ x } 5]$

$= 25 + [1.65 \text{ x } 5]$

$= 25 + 8.25$

$= 33.25$

$Q_1 = 33.25$

$Q_3 = L + [\frac{3(\frac{N}{4}) - m}{f} \text{ x } c]$

$3(\frac{N}{4}) = 3 \text{ x } \frac{420}{4} = 3 \text{ x } 105$

$3(\frac{N}{4}) = 315$

Class Interval = 40 – 45

L = 40; f = 94; m = 112

$= 40 + [\frac{315-112}{94} \text{ x } 5]$

$= 40 + [\frac{203}{94} \text{ x } 5]$

$= 40 + [2.15 \text{ x } 5]$

= 40 + 10.75

= 50.75

Q_3 = 50.75

M = L + [$\frac{\frac{N}{2}-m}{f}$ x c]

$\frac{N}{2} = \frac{420}{2}$ = 210

Class Interval = 35 – 40

L = 35; $\frac{N}{2}$ = 210; m = 60; f = 112

= 35 + [$\frac{210-60}{112}$ x 5]

= 35 + [$\frac{150}{112}$ x 5]

= 35 + [1.33 x 5]

= 35 + 6.65

= 41.65

M = 41.65

Sk_B = $\frac{50.75+33.25-2(41.65)}{50.75-33.25}$

= $\frac{84-83.3}{17.5}$

= $\frac{0.7}{17.5}$ = 0.04

Answer:

Sk_B = 0.0