UNIT - I

1 Introduction to Statistics and Sampling Theory

Syllabus

Statistics : *Introduction, Origin and Development of Statistics, Definition, Importance and Scope, Limitations, Distrust of Statistics*

Population and Sample: *Sampling - Introduction, Types of Sampling, Purposive Sampling, Random Sampling, Simple Sampling, Stratified Sampling, Parameter and Statistic, Sampling Distribution, Statistical Inference, Sampling With and Without Replacement, Random Samples : Random Numbers, Population Parameters, Sample Statistics, Sampling Distributions*

Contents

Scanned by TapScanner

1.1 Origin and Development of Statistics

- The word statistics might be derived from the latin word 'status' or 'statista' in Italian language. The meaning of this word is same that is political state. In ancient era the kings of different states used to survey the data of population in their states. The purpose of doing this was to ensure the manpower and later to regain the taxes. In the historical era during the wars also the kings were collecting this information to meet loss in war and implementation of new takes to balance the loss.
- The theory of probability gives a new face to statistics. It has been developed by many researchers from India, England, France and Germany. Indian scientists also contributed for development of statistics. Sample survey, design of experiments in agricultural, multivariate analysis are some of remarkable contributions to statistics.
- A theory of probability being a part of modern statistics came into play in the mid seventeenth century with introduction of theory of probability and theory of games and chance. A gambling game and theory of mathematics contributed for development of statistics which were from France, Germany and England. A problem of points solved by french mathematician Pascal and P. Fermat which was raised by gambler chevalier de-Mere. It became a foundation for theory of probability which is backbone of modern theory of statistics.
- This development of statistics was carried by notable mathematicians in their theories and research work. Among them Jame Bernoulli (1654 - 1705) who wrote the first treatise on the theory of probability. Next De-Moivre (1667 - 1754) contributed for probabilities and annuities and published his work in 'The Doctrine of chances' in 1718. Laplace (1749 - 1827) published his work in 1782 which was on theory of probability. Gauss (1777 - 1855) a notable contributer in consistent development of statistics gave 'Principle of least squares' and the 'Normal Law of Errors'. A modern theory of probability touched by many in 18^{th}, 19^{th} and 20^{th} centuries such as euler, lagrange, bayes, etc.
- Karl Pearson, a pioneer of 'Correlation Analysis' gave the first and more significant test in statistics known as Chi-square Test (X^2-test) of goodness of fit. W.S. Gosset found a t-test for exact (small) sample tests.
- R. A. Fisher (1890-1962) who used statistics in different fields like agriculture, genetics, education etc. known as 'Father of statistics'. His contribution in theory of statistics became the foundation for research workers in statistics. He is pioneer in estimation theory, exact sampling distributions, analysis of variance and design of experiments.

Scanned by TapScanner

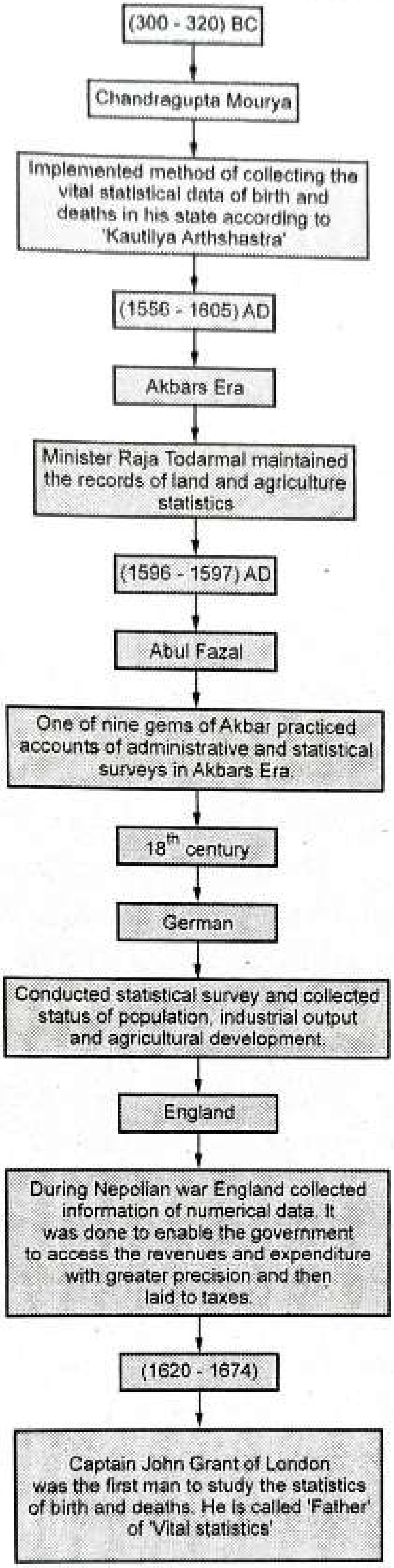

Fig. 1.1.1

Scanned by TapScanner

- In his honour his contribution is described in the following manner.

  "R. A. Fisher is the real giant in the development of the theory of statistics".

- Indian mathematicians also contributed notably for development of statistics in diversified fields. Among them a few are Parthsarathi (Theory of probability) R.C. Bose, Panse, J.N. Srivastava (Design of experiments in agriculture), C.R. Rao (Statistical inference), P.C. Mahalanobis and P.V. Sukhatme (Sample survey) etc. shared distinguishable work in statistics.

Definition

- Statistics is a branch of mathematics which is the collection, description, analysis and interpretation of data.
- Statistics mainly deals with the information or any data which is on a large scale and needs to be estimated properly.

  Examples : Data mining, data compression, artificial intelligence and network, traffic modeling. Examples given above are the integral part of A.I.D.S.

1.2 Importance and Scope of Statistics

- In various field now a days this type of analysis and estimation of data become mandatory. Without doing statistical treatment one can't reach to conclusion of any technique, process or up and down of factors mainly seasonal. Some misunderstandings of people etc. No field remained untouched from statistics methods such as business, economics, science and engineering and social sciences.
- A statistical approach day-by-day becoming dominant almost in every field such as business, management and marketing. If we expect the good quality products and processes within business organizations, then statistical treatment becomes unavoidable.

1.2.1 Statistics in Business and Economics

- The economist and managers have to face and analyze the data, data gathering, management and data interpretation daily. A good businessman has to be kin about the desire of common customer and their needs. They have to rely on statistical techniques. A success of business depends entirely on statistical correct forecast which lead to satisfaction of supply and demand of a product. It need accuracy and precision and it is possible with good statistical method. Example : To start manufacturing of electronic components. Then it must considered that what is the scope of device in industry ? How many gazettes require such a device on routine basis ?

Scanned by TapScanner

- How should one present this product in market ? To achieve this it must be prepared with well plan in advance which can not be done without quantitative facts.
- Economical trends are also based on statistical techniques. It is useful in finding the economical solution over price of product labor cost, transportations etc. The tools in statistics like data analysis, number indexing time series analysis demand and forecasting techniques are used to improve planning and economical development.
- Index number helps us to arrive at a single conclusion which represents general level of variation in factors like increase, decrease and rate of increase and rate of decrease. This method covers commodity prices, industrial production sales import and export, etc.

1.2.2 Statistics in Engineering and Technology

- The design, modeling and manufacturing of a product is an engineering. The process which is used for above mentioned is the use of technology. Which is suitable for design, modeling and manufacturing is the outcome of statistical study of principals, theories and mathematical treatment etc. statistical quality control techniques are based on theory of probability and sampling. It has been extensively used in industries like automobile textile, electronic component making, electric equipment, computer softwares, etc.
- The different updated skills are the most demanded part of development of engineering. It is not only to apply and design of product but also to keep balance between current technologies.
- Engineers job is to apply the knowledge and create a product statistical survey plays key role in this development.

1.2.3 Statistics and Medical or Biological Fields

- In medical field issues related to diseases are important aspects to study variation in growth of any virus. The conditions in surrounding like increase or decrease in temperature, moisture content of air also affect rate of growth or decrease in viruses. The collection and analysis of data on the basis of infections produced among people presented statistically always helps to reach to conclusion. On the basis of these conclusion medicines can be prepared for curing the patients. The statistics is not only a mathematical tool but also proved to be a useful tool for mankind. On the other side in the treatment of patients uses of drugs, its side effects and advantages are also studied statistically. This type of survey leads to finalize the drug and treatment of diseases.

Scanned by TapScanner

- In biological fields cells of plants, any disease to plant as per atmospheric variation are studied by collecting species of different kind. The proper use of particular pesticide depends on survey of different kind of plants, soil quality of water, etc.
- This was first studied by Francis Galton in his work in 'Regression'. According to Prof. Karl Pearson the whole "Theory of heredity" rests on statistical basis. The efficiency of a manufactured drug or injection or medicine is tested by using the test of significance (t - test).

1.3 Importance of Statistics

1) Statistics is important mainly in data analysis.
2) It provides a single solution to multiple variational parametric data.
3) Statistics sometime used find and establish the relations between physical quantities.
4) Statistics also provides the prediction and estimation of certain relevant data.
5) It also provides dimensional nature of data like survey, spread of data, divergence of data.
6) Statistics support to reach to conclusion of a processed data.
7) Statistics itself provides a tool of comparison.
8) Uncertainty in data is reduced in statistical estimation.
9) Statistics is a tool of comparison, investigation of data independent of type of filed.
10) Statistics is important in business, economics, industry, technology science and many more.

1.4 Limitations of Statistics

1.4.1 Statistics does not Study Individuals

- Statistics mainly deals with data in bulk form and does not treat any individual part in separate. Individual data treated statistically becomes meaningless statistics is used only where the data is in the form of group.
- W. I. King states, "Statistics from their very nature of subject cannot and will never be able to take into account individual cases. When these are important other means must be used for their study".

1.4.2 Statistics does not Study Qualitative Phenomena

- It is applicable to quantitative data which can be expressed in numerical form. It does not provide the base to study qulitative data. Qualitative phenomena poverty, wisdom, intelligence, etc. cannot be expressed in numerical form. However it can be applied when qualitative data will be reduced to numerical form.

Scanned by TapScanner

1.4.3 Statistical Laws are not Exact

- Probabilities and approximate values can be predicted by statistical method. It does not give exact nature of laws like scientific laws. There is always some uncertainty in predicted solution, but when it is averaged with certain details, uncertainty is reduced.

1.4.4 Statistical Results are True only on an Average

- It is always based on average values of any phenomena experiment or survey. It cannot produce a true factor for individual. It is in the form of average studied factor. It is not useful for substitution of any unit or event. Statistical average data when applied to an individual encounters error in observation. W. I. King states "Statistics largely deals with averages and these averages may be made up of individual items radically different from each other".

1.4.5 Statistical Relations do not Necessarily Bring Out the 'Cause and Effect' Relationship between Phenomena

- It is entirely depend on interpreters data. It reveals the association in relation of different quantities. It varies by the nature and type of quantities involved and depends on the view and expectation of interpreter.

1.5 Distrust of Statistics

- Many times applications and usefulness of statistics is ignored due to misunderstanding among some people. It may be due to their insufficient confidence in doing the proper analysis of data. It is often said that statistical analysis calculations and results are manipulated by the uses.

Causes of distrust :

1) Figures are always easily believed.
2) Statistics has also some limitations. These are ignored.
3) Figures are represented by manipulation. This is the misuse of figures.
4) Insufficient information or knowledge of the subject.

- Though the figures are accurate, some people manipulates the figures to hide practical data and try to represent expected data to achieve their selfish desires. The skilled speakers, writers force some wrong input in statistical data analysis which lead to lose faith of public in the subject.

Scanned by TapScanner

- It should be rightly understand that statistics does not prove anything or disprove the same. It is only the tool to achieve some single conclusion, relation, dependence of quality, etc.
- According to bowley "Statistics only furnishes a tool, necessary though imperfect, which is dangerous in the hands of those who do not know its use and its deficiencies".
- An example can be discussed here to illustrate the point. Medicines are used to cure the patient. If wrong medicine or excess dose is taken a person may die. We cannot blame the medicine for such a result.

1.6 Misinterpretation of Statistical Data

1) A survey report : Usually electronic gazettes are compared by customer on an online portal. Sometimes the opinions of customers if in favour of any particular company product, then it is retained. On the other side if the same data is believed to be true without comparison then it is wrong interpretation. It does not mean a product displayed is only the good.
2) The number of students taking the subject economics in institute increased twice as compared to last two years. Thus economics is most demanded subject among the students. This conclusion is faulty because the same time status of other subjects is not considered.

1.7 Population and Sample

- The word population is used in a wider sense. For example, In the study of industrial development all the industries under consideration is population. In titration experiment solution in breaker is a population. Thus population may be a group of employees, collection of books, a group of students, etc.

Definition :

- A set or group of observations relating to a phenomenon under statistical investigation called statistical population or population.
- The population is finite or infinite according to whether the set contains a finite or infinite number of observations.
- For any statistical analysis complete enumeration of the population is impracticable.
- If we want to have an idea of the average per capita (monthly) income of the people in India. We will have to enumerate all the earning individuals in the country which is a very difficult task.

Scanned by TapScanner

- In order to study a group of large number of items we require to draw sample. We use technique of sampling several times in everyday life.

Example :

i) Purchasing food grains we inspect only handful of grains and draw conclusion about the whole sack.

ii) While examining blood of an individual few drops are enough for diagnosis.

Sampling is well accepted means of collecting information. Sampling plays very important role in statistical inference.

Sample : Definition : A finite subset of statistical individuals in a population is called a sample and number of individuals in a sample is called the sample size.

Types of sampling :

- Some of commonly known types of sampling are :

1) Purposive sampling
2) Random sampling
3) Simple sampling
4) Stratified sampling

1.7.1 Purposive Sampling

- Purposive sampling is one in which the sample units are selected with definite purpose in view. It is less costly and less time consuming.

Example :

1) To select candidates for debating competition, certainty deliberate selection of suitable candidates will be done. A survey may be carried out by investigator interested in opinions and views on certain specific issues.
2) Financial institutions can ensure representative character in a purposive selection of sick units financed by them.

1.7.2 Random Sampling

- In this case sample units are selected at random and the drawback of purposive sampling completely overcome. A random sample is one in which each unit of population has an equal chance of being included in it.
- Suppose we take a sample of size n from N size of finite population. Possibility of samples are ${}^{N}C_{n}$. A sampling technique in which each of ${}^{N}C_{n}$ samples has an equal

Scanned by TapScanner

chance of being selected is known as random sampling and the technique is called as random sample.

- For example r candidates are selected from n candidates. Assigning 1 to n numbers to each candidate and make slips which are homogeneous and put all slips in a bag and thoroughly shuffled and then r slips are drawn one by one 'r' candidates corresponding to numbers on the slips drawn will constitute random sample.
- According to W. M. Harper "A random sample is one which is selected in such a way that every item in the population has an equal chance of being included".

1.7.3 Sample Sampling

- Under this sampling the whole population is taken as a single composite unit for purposes of sampling.
- This method is easiest and commonly used.
- If population consists of N elements then probability of selection of any element is $\frac{1}{N}$.

Examples :

1) To find diameter of a rod, we take reading at few points on a rod and then find the average of readings.
2) To conduct socio-economic survey of a certain village and find per capita income of a village.

- This method is widely used due to its simplicity and convenience. However, it suffers from some drawbacks such as, it may not be proper representative when population is heterogeneous, widely spread, etc.

1.7.4 Stratified Sampling

- If population is heterogeneous, simple sampling is not effective. Entire population is divided into many homogeneous groups called as strata (plural). The size of each stratum is same as a simple sample combining of all sampled observations formed stratified sampling.
- This method gives better results. This method finds suitability in administrative purposes.

Examples :

1) Estimation of annual income per family we divide population into groups such as families with yearly income below 20,000 between ₹ 20,000 to ₹ 50,000 between ₹ 50,000 to 1 Lakh and above.

Scanned by TapScanner

2) To conduct health survey in a college we can use stratified sampling by considering strata as the facilities or classes or sex etc.

1.8 Parameter and Statistics

- A parameter is a descriptive measure of characteristic of the population. It's frequently possible to have a good idea of the population from a few descriptive measures. A measure which is calculated from all the compliances on the population is a parameter of the population. For illustration, if an average is calculated from a set of all observations on the population, the average will be a parameter of the population. The average is population average. A corresponding descriptive measure can be obtained from the observations contained in a sample of population. A descriptive measure computed from the observations in a sample is called a statistic. An average computed from observation in a sample is a statistics and is called sample average.
- In order to avoid verbal confusion with statistical constants of the population viz. mean (μ) and variance (σ^2) etc. which are usually referred to as parameters : Statistical measures calculated from the sample observations alone, e.g. mean ($\bar{X}$), variance (S^2) etc. referred as statistics.
- It may be noted that a parameter is constant for population the corresponding statistics may vary from sample to sample.

Remark :

1) μ and σ^2 will refer to the population mean and variance respectively while the sample mean and variance will be denoted by $\bar{X}$ and S^2 respectively.
2) Unbiased estimate : A statistic $t = t(x_1, x_2, \ldots, x_n)$ a function of the sample values $x_1, x_2, \ldots, x_n$ is an unbiased estimate of the population parameter F, if E(t) = F, i.e. E(statistic) = Parameter then statistic is said to be unbiased estimate of the parameter.

1.9 Sample Distribution of a Statistic

To draw a sample of size n from a given finite population of size N then total number of possible sample is,

$$^{N}C_n = \frac{N_1}{n!(N-n)!} = r(\text{say})$$

r samples we can compute some statistic $t = t(x_1, x_2, \ldots, x_n)$, in particular mean $\bar{x}$ and variance S^2 etc. given below

Scanned by TapScanner

Sample number	Statistic		
	t	$\bar{x}$	S^2
1	t_1	$\bar{x}_1$	S_1^2
2	t_2	$\bar{x}_2$	S_2^2
3	t_3	$\bar{x}_3$	S_3^2
⋮	⋮	⋮	⋮
r	t_r	$\bar{x}_r$	S_r^2

Mean and variance of the sampling distribution of the statistic t are given by,

$$\bar{t} = \frac{1}{r}(t_1 + t_2 + \ldots + t_r) = \frac{1}{r}\sum_{i=1}^{r} t_i$$

$$\text{var}(t) = \frac{1}{r}[(t_1 - \bar{t})^2 + (t_2 - \bar{t})^2 + \ldots + (t_r - \bar{t})^2]$$

$$= \frac{1}{r}\sum_{i=1}^{r}(t_i - \bar{t})^2$$

1) **Standard error :** Standard deviation of sampling distribution of a statistics is known a its Standard Error (S.E.).

For large samples, where n is the sample size, σ^2 is population variance, p is populatio proportion and $q = 1 - p$, n_1 and n_2 represent sizes of two independent random sample reply drawn from given population (s).

Sr. No.	Statistics	Standard error
1.	Sample mean : $\bar{x}$	$\sigma/\sqrt{n}$
2.	Observed sample proportion 'P'	$\sqrt{pq/n}$
3.	Sample s.d : S	$\sqrt{\sigma^2/2n}$
4.	Sample variance : S^2	$\sigma^2\sqrt{2/n}$
5.	Sample correlation coefficient (r)	$(1-\rho^2)/\sqrt{n}$
6.	Sample moment : μ_3	$\sigma^3\sqrt{96/n}$
7.	Sample moment : μ_4	$\sigma^4\sqrt{96/n}$

Scanned by TapScanner

8.	Difference of two sample means : $(\bar{x}_1 - \bar{x}_2)$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
9.	Difference of two sample s.d : $(S_1 - S_2)$	$\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}$

1.10 Statistical Inference

- It is method to draw the conclusion or what do we summarize from given analysis of data. This is the main objective of the statistics. There are two important problems in statistical inference - 1) Estimation 2) Testing of hypothesis.

1.10.1 Methods of Estimation

- For obtaining estimation commonly used methods are as follows :

1) Method of maximum likelihood estimation : This is the most general method of estimation.
2) Method of minimum variance.
3) Method of moments.
4) Method of least squares.
5) Method of minimum chi-square.
6) Method of inverse probability.

1.10.2 Testing of Hypothesis

- A test of a statistical hypothesis is a two action decision problem after the experimental sample values have been obtained, the two actions being the acceptance or rejection of the hypothesis under consideration.

Null hypothesis :

- A statistician should work without thinking of profit or loss which affects the acceptance of calculations. It should not be partial and should not influence the decision at any cost.

  Example :

  A writing pen is provided by two methods of manufacturing. There are different views like old method is more reliable or new one. It provides comparative observation of two methods. The zero difference method i.e. both methods are sufficient are accepted.

Scanned by TapScanner

Alternative hypothesis :

- The acceptance of null hypothesis depends on any other testing method. Acceptance or rejection depends on the test against rival hypothesis.

1.11 Sampling with and without Replacement

- In statistical study of any data, samples are collected to reach to a single or average conclusion of any process. This population contains two types of sampling units.

  1) A finite number of sampling unit. 2) An infinite number of sampling units.

- But mostly population contains to found a finite number of sampling units. Sampling can be done by with replacement or without replacement. When the same unit of the population involves in each sample more than once, then it is sampling with replacement. In sampling without replacement same unit of the population may not be included in each sample more than once.
  - If the size of population is N and sample size in n then
  - If sampling is done with replacement, the total number of possible samples will amount to N^n.
  - If sampling is done without replacement the total number of possible samples will be ${}^NC_n$.
- In case of population of finite size sampling without replacement results in ${}^NC_n$ distinct ways where N is size of population and n is total number of possible samples.

Example 1.11.1 *If the population size is 5 and the sample size is 2. Find the number of possible samples 1) With replacement, 2) Without replacement.*

Solution :

1)

$$\text{Population size} = N = 5$$

$$\text{Sample size} = n = 2$$

$$\text{Samples} = N^n = 5^2 = 25$$

$$\text{Possible samples with replacement} = 25$$

2)

$$\text{Population size} = N = 5$$

$$\text{Sample size} = n = 2$$

$$\text{Samples} = {}^NC_n = {}^5C_2 = \frac{5!}{2!(5-2)!} = \frac{5!}{2!3!} = \frac{4 \times 5}{1 \times 2} = 10$$

$$\text{Possible samples without replacement} = 10$$

Scanned by TapScanner

1.12 Random Sample

- Random samples are characterized by the way in which they are selected.
- Random sample is one where each item in the universe has an equal or known opportunity of being selected.
- According to C. H. Mayer "A sample is said to be random when each unit drawn has a probability identical to the probability of all the other units which might have been drawn in its place".

1) Sample is selected to study the concerned population. Sample is a miniature of population.
2) Sampling units should be independent.
3) Samples can be achieved by dividing population in homogeneous subgroups and selecting a random sample from each sub-groups.

- In random selection equal chance of selection is given to each sampling unit. It can be achieved by two ways.

1.12.1 Lottery Method

- In this method n elements are drawn by using N different slips which are made of the same size and shape. All slips are mixed properly. With the help of handle slips are thoroughly mixed and n elements of slips are drawn one by one.

1.12.2 Random Numbers

- A sample is selected for smooth operation purpose by use of random numbers. This can be done by using availble random tables. These tables are prepared by using the numbers from 0 to 9 with equal chance of finding or equal frequency of its occurence. Two digit numbers are selecting by formation of two rows and two columns. We form the numbers from 00 to 99. For 3 digit number 3 rows and 3 columns are formed to get the numbers from 000 to 999.
- For selection of a random sample any part of random numbers selected and numbers are chosen serially in row or column. If the number selected in this manner is in between 1 to N, the corresponding element is taken in the sample. Hence n elements are chosen.

Tables of random numbers :

1) **Tippet's random number tables :** This table consist of 10400 four digited numbers, gives in all 10400×4 i.e. 41600 digits. These tables proved to be fairly random in character.

Scanned by TapScanner

2) **Fisher and yates tables :** Statistical tables for biological, agricultural and medical research. It is obtained by drawing numbers at random from the 10^{th} to 19^{th} digits of A.S.

3) **Kendall and babington smith table :** It consist of 1,00,000 digits grouped into 25000 sets of 4-digit random numbers.

4) **Rand corporation random table :** It consist of one million random digits consisting of 2,00,000 random numbers of 5 digit each.

1.13 Population Parameter

- It is a quantity or statistical measure that, for a given population is fixed and that is used as the value of a variable is same in general distribution or frequency function to make it descriptive of that population.
- These are not similar to parameters in mathematics which refers the constant value for a mathematical function.
- Parameters refer to the whole population whereas statistics refers a part of that population.

1.14 Sample Statistic

- Sample statistic is a piece of statistical information from the items under consideration and to be analyzed for the purpose of its characteristics.
- It is a fraction of data or a part of population which is small in comparison with population.
- Sample is a piece under the study taken from population which is on large scale.

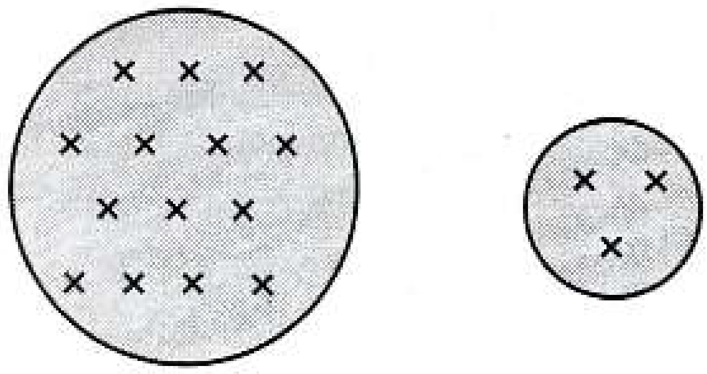

Fig. 1.14.1

1.15 Sampling Distribution

- Each and every unit of data collected by any means called as population is not considered for extracting the information. In short the whole population is not analyzed with its individual constituent elements.

Scanned by TapScanner

- A few units selected from among the total lot in accordance with some specified procedure constitutes a sample.
- The various values assumed by statistics distributed which are computed from various samples of same size and drawn randomly from the population is called sampling distribution.
- Population of constant nature contains a large number of samples of same size. Each of these sample may contain different population members though it has same size. Each sample goes under treatment of statistics like mean, standard deviation. Variance etc. and series of such a sample are brought together at one place.
- The various values produced by this kind of method may be arranged as a frequency distribution which is known as sampling distribution.
- Sampling is done with replacement and without replacement.
- The samples are analyzed by taking their mean first. A mean of sample will be different because it has been derived from different units of same sample. The sample mean taken as random variable. Each sample mean will then constitute an observed value of this new random variable. It is given by a new set $\bar{X}_1, \bar{X}_2, \bar{X}_3, \ldots, \bar{X}_k$. These mean values may be used to form a frequency distribution. The frequency distribution of the statistic $\bar{X}$ is known as the sampling distribution of sample mean.

Exercise

1. *What is statistics ? Elaborate on the statistics with reference to business and economics.*
2. *Explain the scope of statistics in engineering and technology.*
3. *Give a brief note on the statistics in medical or biological fields.*
4. *Define statistics. Give importance of statistics. (any 5)*
5. *What are the limitations of statistics ? Explain one limitation in brief.*
6. *What are the reasons of distrust of statistics ? Give suitable example.*
7. *How statistical data can be misinterpreted ? Explain it with suitable examples.*
8. *What is population and sample ? Give difference between them.*
9. *Define sampling. Give its types. Explain any one.*

Scanned by TapScanner

10. *Discuss the following in brief : 1) Purposive sampling, 2) Random sampling, 3) Simple sampling, 4) Stratified sampling.*
11. *What is parameter of population ? Discuss it in brief.*
12. *Give method of sample distribution of a statistics.*
13. *Mention methods of estimation. Discuss testing of hypothesis in short.*
14. *Give comparison between sampling with and without replacement.*
15. *What is random sample ? Explain lottery method and random numbers.*
16. *Explain random numbers and mention tables of random numbers.*
17. *What are the advantages and disadvantages of statistical analysis ?*

Multiple Choice Questions

Q.1 A pioneer of 'Corelation Analysis', Karl Pearson gave first and more significant test in statistics known as __________.

a Chi-square test
b Bi-square test
c Analytic test
d None

Q.2 Statistics involves which of the following process related to data __________.

a collection of data
b description of data
c analysis and interpretation of data
d all of the above

Q.3 The role of statistics in engineering and technology referred to as __________.

a to design
b modeling of product
c manufacturing of product
d all of the above

Q.4 In survey of a data some uncertainty is observed. The uncertainty can be reduced by __________.

a trignometric calculation
b estimation of data
c algebraic calculation
d none

Q.5 Statistical analysis requires a ________ data for analysis.

a individual data
b single parameter data
c form of a group
d very few no. of data

Scanned by TapScanner

Q.6 Which of the following statement is wrong ________.

[a] statistics studies individual data

[b] statistics refers qualitative phonomena

[c] statistical laws are exact

[d] all of the above

Q.7 The cause of distrust of statistics is due to ________.

[a] collection of data

[b] sampling of data

[c] interpretation of data

[d] manipulation in collected data

Q.8 The number of individual in a sample is called ________.

[a] population

[b] sample

[c] sample size

[d] none

Q.9 Which of the following is not a type of sampling ________.

[a] purposive sampling

[b] random sampling

[c] continuity in sampling

[d] stratified sampling

Q.10 Each unit of population has an equal chance of being included in sample is the example of ________.

[a] purposive sampling

[b] random sampling

[c] simple sampling

[d] stratified sampling

Q.11 For a given data, a sample of size 'r' from a given finite population of size N, then total number of possible sample is ________.

[a] $^{N}C_{r} = \frac{N!}{r!(N-r)!}$

[b] $^{r}C_{N} = \frac{r!}{r\,(r-N)!}$

[c] $^{N}C_{r} = 1$

[d] none

Q.12 In measurement of height of a wall, following readings are taken 5.26, 5.28, 5.32, 5.24 and 5.31 respectively. In statistical analysis the mean and variance of data will be ________.

[a] 5.28 and 9×10^{-4}

[b] 4.26 and 9×10^{-3}

[c] 3.21 and 9×10^{-2}

[d] 2.26 and 9

Scanned by TapScanner

Q.13 In sampling with replacement the total number of possible samples if the size of population is N and sample size is n will be ______.

a N^n

b n^N

c N^{n-1}

d n^{N-1}

Q.14 Formula for standard error of sample mean $\bar{x}$ based on sample of size n having variances σ^2 when population consisted of N items is ______.

a σ/n

b $\sigma/\sqrt{n-1}$

c $\sigma/\sqrt{N-1}$

d $\frac{\sigma}{\sqrt{n}}$

Q.15 Which of the following statement is not true ?

a Standard error cannot be zero

b Standard error cannot be 1

c Standard error can be negative

d All the above

Q.16 If the sample values are 1, 3, 5, 7, 9 the standard error of sample mean is : ______.

a $S.E = \sqrt{2}$

b $S.E = \frac{1}{\sqrt{2}}$

c $S.E = 2.0$

d $S.E = \frac{1}{2}$

Q.17 If we have a sample of size n from a population of N units, the finite population correction is ______.

a $\frac{N-1}{N}$

b $\frac{n-1}{N}$

c $\frac{N-n}{N}$

d $\frac{N-n}{n}$

Q.18 If n units are selected in a sample from N population units, the sampling fraction is given as : ______.

a $\frac{N}{n}$

b $\frac{1}{N}$

c $\frac{1}{n}$

d $\frac{n}{N}$

Scanned by TapScanner

Q.19 If sampling is done without replacement the total number of possible sample will be ________. If size of population is N and sampling size is n.

a $^{N}C_{n}$

b $^{n}C_{N}$

c $^{N-1}C_{n}$

d $^{n-1}C_{N}$

Q.20 If population size is 5 and the sample size is 2. The number of possible samples with replacement will be ________.

a 25

b 16

c 125

d 625

Q.21 If the sum of N observations is 630 and their mean is 42, then value of N is :

a 21

b 30

c 15

d 20

Q.22 If population size is 10 and the sample size is 4. The number of possible samples without replacement is ________.

a 210

b 215

c 110

d 115

Q.23 The method of sampling in which n elements are drawn by using N different slips made up of same size and shape is ________.

a lottery method

b random numbers

c both (a) and (b)

d can't say

Q.24 Which of the following is not a table of random numbers

a Tippet's Random Number Table

b Fisher and Yates Table

c Newton's Number Table

d Kendall and Bobington Smith Table

Q.25 The class interval of the continuous grouped data 10 - 19, 20 - 29, 30 - 39, 40 - 49, 50 - 59 is ________.

a 10

b 9

c 14.5

d 4.5

Scanned by TapScanner

Q.26 Chi-square distribution is used for the test of : ________.

a goodness of fit

b hypothetical value of population variance

c both (a) and (b)

d neither (a) nor (b)

Q.27 Stratified sampling comes under the category of ________.

a unrestricted sampling

b subjective sampling

c restricted sampling

d purposive sampling

Q.28 If the observations recorded on five sampled items are 3, 4, 5, 6, 7 the sample variance is ________.

a 2

b 1

c 0

d 2.5

Q.29 Stratified sampling belongs to the category of ________.

a judgement sampling

b subjective sampling

c controlled sampling

d non-random sampling

Q.30 The magnitude of the standard error of an estimate is an index of its :

a accuracy

b precision

c efficiency

d all the above

Answer Keys for Multiple Choice Questions :

Q.1	a	Q.2	d	Q.3	d	Q.4	b	Q.5	c
Q.6	d	Q.7	d	Q.8	c	Q.9	c	Q.10	b
Q.11	a	Q.12	a	Q.13	a	Q.14	d	Q.15	d
Q.16	a	Q.17	c	Q.18	d	Q.19	a	Q.20	a
Q.21	c	Q.22	a	Q.23	a	Q.24	c	Q.25	a
Q.26	c	Q.27	c	Q.28	a	Q.29	c	Q.30	b

□□□

Scanned by TapScanner

Unit II

2 Descriptive Statistics : Measures of Central Tendency

Syllabus

Frequency Distributions and Measures of central Tendency : *Frequency Distribution, Continuous Frequency Distribution, Graphic Representation of a Frequency Distribution, Histogram, Frequency Polygon, Averages or Measures of Central Tendency or Measures of Location, Requisites for an Ideal Measure of Central Tendency, Arithmetic Mean, Properties of Arithmetic Mean, Merits and Demerits of Arithmetic Mean, Weighted Mean, Median, Merits and Demerits of Median, Mode, Merits and Demerits of Mode, Geometric Mean, Merits and Demerits of Geometric Mean, Harmonic Mean, Merits and Demerits of Harmonic Mean, Selection of an Average.*

Contents

2.1 Introduction

- In statistical exploration, once data is collected and edited... "The first task of the statistician is to organize of the figures in such a type that their significance for the aim in hand could also be appreciated, that comparison with masses of similar data could also be facilitated and that further analysis could also be possible.
- This is done through classification and tabulation.
- Classification is necessary and always precedes tabulation but after the determination of class categories the mode of presentation may take any form.

2.2 Classification

- Classification refers to grouping of data into homogeneous classes and categories.
- "Classification is the process of arranging data into sequences and groups according to their common characteristics or separating them into different but related parts".
- Classification is the process of arrangement into groups or classes. A group or a class category has to be determined on the basis of the nature of data and purpose for which it is going to be used.
- For example : Income of 3000 individuals is given for analysis.
- It becomes essential to condense the data in suitable form classification that can be used as a tool.
- Entire process of making homogeneous and non-overlapping groups of observation according to similarities is called **classification.** The groups formed are called **classes** or **class intervals.**
- The objectives of classification can be summarized as follows :
  1) To condense the data.
  2) To prepare the data for tabulation.
  3) To provide ease in comparison with other data.
  4) To avoid unwanted details.
  5) To disclose leading characteristics of the data.
  6) To unlock further analysis like computation of averages, dispersion, etc.

2.3 Frequency Distribution

- A table containing class intervals with frequencies is called **frequency distribution,** 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 groups. With their corresponding frequencies side by side.

- To study how the observations are classified and a frequency distribution formed. Initially discrete variables will be considered for classification.

2.3.1 Frequency Distribution of Discrete Variable Procedure

1) Find smallest and largest observations.
2) Prepare a column of all possible values of variables from smallest to largest.
3) In the next column put a tally mark against values to which it relates.
4) Count the number of tally marks and place them in the next column in front of the corresponding value.

Example 2.3.1 *Row data for the survey of 100 families for studying the number of children in family are given as -*

3	2	5	2	3	1	2	4	6	4
3	1	2	2	4	1	1	3	2	3
2	4	3	6	0	2	5	3	4	2
3	3	3	2	5	0	4	3	4	3
4	3	2	1	2	3	2	2	3	4
2	2	2	3	4	2	3	8	5	1
0	2	4	0	3	3	4	2	4	6
2	4	2	5	2	3	4	3	3	3
4	6	3	1	0	5	4	3	3	3
3	2	3	1	5	0	2	3	4	3

Prepare a frequency distribution of the above data.

Solution : By observation we can locate the largest value as 8 and smallest as 0. We prepare the first column with entries from 0 to 8. We consider the first observation, it is 2, so we mark it as a vertical bar (|) called a **tally mark** and put a tally mark in the next column. If observations repeated three times we mark it as (|||) as a **tally mark** for fourth times repeated observation we mark it as (||||). But for fifth time repeated observations we mark as (/) on the earlier bunch of four tallies finally the structure looks like (卌). In this manner tally marks are arranged in a group of five. Finally, the number of tallies are countėd and placed in the last column procedure given in Table 2.3.1.

Frequency distribution of number of the children in 100 families.

Number of children	Tally marks	Frequency
0	卌 |	6
1	卌 |||	8
2	卌 卌 卌 卌 卌	25
3	卌 卌 卌 卌 卌 卌 |	31
4	卌 卌 卌 |||	18
5	卌 ||	7
6	||||	4
7		0
8	|	1
Total		100

Table 2.3.1

Example 2.3.2 *Marks secured by 50 students in mathematics are given below :*

62	45	46	67	23	53	58	38	26	43
36	63	26	48	76	45	66	72	28	84
56	38	56	92	34	42	38	68	48	52
38	57	65	46	38	46	65	56	48	75
56	84	66	68	58	76	46	76	52	54

Prepare the frequency distribution table. Frequency distribution of marks secured by 50 students.

Solution :

Marks obtained	Tally marks	Frequency
23	|	1
26	||	2
28	|	1
34	|	1
36	|	1
38	卌	5

42	|	1
43	|	1
45	||	2
46	||||	4
48	|||	3
52	||	2
53	|	1
54	|	1
56	||||	4
57	|	1
58	||	2
62	|	1
63	|	1
65	||	2
66	||	2
67	|	1
68	||	2
72	|	1
75	|	1
76	|||	3
84	||	2
92	|	1
	Total	**50**

2.3.2 Frequency Distribution of Continuous Variable

In this distribution the frequencies refer to groups of values.

The procedure of classification of continuous frequency distribution -

Procedure :

1) Find smallest and largest observations. Calculate the difference between them. Difference is called **'range'**.

2) Prepare a column of class intervals.
3) Classify observations one by one in appropriate class by putting tally marks in th second column.
4) Count tally marks and enter in the last column.

Example 2.3.3 *Height of 50 students to the nearest cm are given as below :*

151	147	145	153	156	152	159
153	157	152	144	151	157	147
150	157	153	151	149	147	151
147	155	156	151	158	149	147
153	152	149	149	153	150	152
154	150	152	151	151	151	154
155	152	154	152	156	155	154
150						

Construct a frequency distribution table.

Solution : In this problem the highest and lowest observations are 159 and 145 respectively So we form classes as : 145-146, 147-148, 149-150, 151-152, 153-154, 155-156, 157-158, 159-160 and construct a table as - frequency distribution of heights of 50 students.

Class interval (Height in cm)	Tally marks	Frequency (No. of students having height)
145-146	||	2
147-148	𝍸	5
149-150	𝍸 |||	8
151-152	𝍸 𝍸 𝍸	15
153-154	𝍸 ||||	9
155-156	𝍸 |	6
157-158	||||	4
159-160	|	1
Total		50

Example 2.3.4 *Number of tools produced by 50 workers in a factory are given below :*

43	18	25	18	39	44	19	20	20	26
40	45	38	25	13	14	27	41	42	17
34	31	32	27	33	37	25	26	32	25
33	34	35	46	29	24	31	34	35	24
28	30	41	32	29	28	30	31	30	31

Construct a frequency distribution table.

Solution : Here the smallest observation is 13 and the largest observation is 45. So the range is $45 - 13 = 32$. Since there are 50 observations we make classes each of size 5. Classes are from $13 - 17$, $18 - 22$, $23 - 27$, $28 - 32$, $33 - 37$, $38 - 42$, $43 - 47$.

Frequency distribution of number of tools produced by 50 workers.

Class interval	Tally marks	Number of tools produced
13-17	|||	3
18-22	~~||||~~	5
23-27	~~||||~~ ~~||||~~	10
28-32	~~||||~~ ~~||||~~ ||||	14
33-37	~~||||~~ |||	8
38-42	~~||||~~ |	6
43-47	||||	4
Total		50

2.4 Graphic Representation of a Frequency Distribution

- It is frequently useful to represent a frequency distribution via diagram which makes the unwieldy data and conveys to the eye the general run of the observations. It represents the comparison of two or more frequency distributions.

- There are two ways of graphical representations of frequency distributions are as follows :

  1) Histogram 2) Frequency polygon

2.4.1 Histogram

- It is one of the popularly used graphs for the representation of frequency distribution. Histogram distribution is drawn as follows :
  1) The class intervals are marked by taking a suitable scale along the X-axis.
  2) The rectangles with height depending on the frequency of corresponding class intervals are obtained such as to get the area of the rectangle proportioned to the frequency of the class.
  3) For unequal width the height of the rectangle depends on the ratio of frequencies to the width of the classes.
- The constituted construction of continuous rectangles called **histogram.**

Remark :

1) The histogram of an ungrouped frequency distribution of a variable is constructed by assuming the frequency with value of variable x between the interval $x-\frac{h}{2}$ to $x+\frac{h}{2}$ where h is the width.
2) In case of discontinuous grouped frequency distribution, it is converted into continuous distribution to construct the histogram.
3) There is no proportionality between the fractions of height and frequency of class. Hence histogram does not help to read the frequency over a fraction of class interval.
4) Histogram of the distribution of marks 100 students is obtained as follows :

Marks	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64
No. of students	9	18	12	15	21	05	10	7	2	1

- Since grouped frequency distribution is not continuous, we first convert it into a continuous distribution with exclusive type classes as given below :

Marks	14.5-19.5	19.5-24.5	24.5-29.5	29.5-34.5	34.5-39.5	39.5-44.5
No. of students	9	18	12	15	21	5
Marks	44.5-49.5	49.5-54.5	54.5-59.5	59.5-64.5		
No. of students	10	7	2	1		

Histogram for frequency distribution

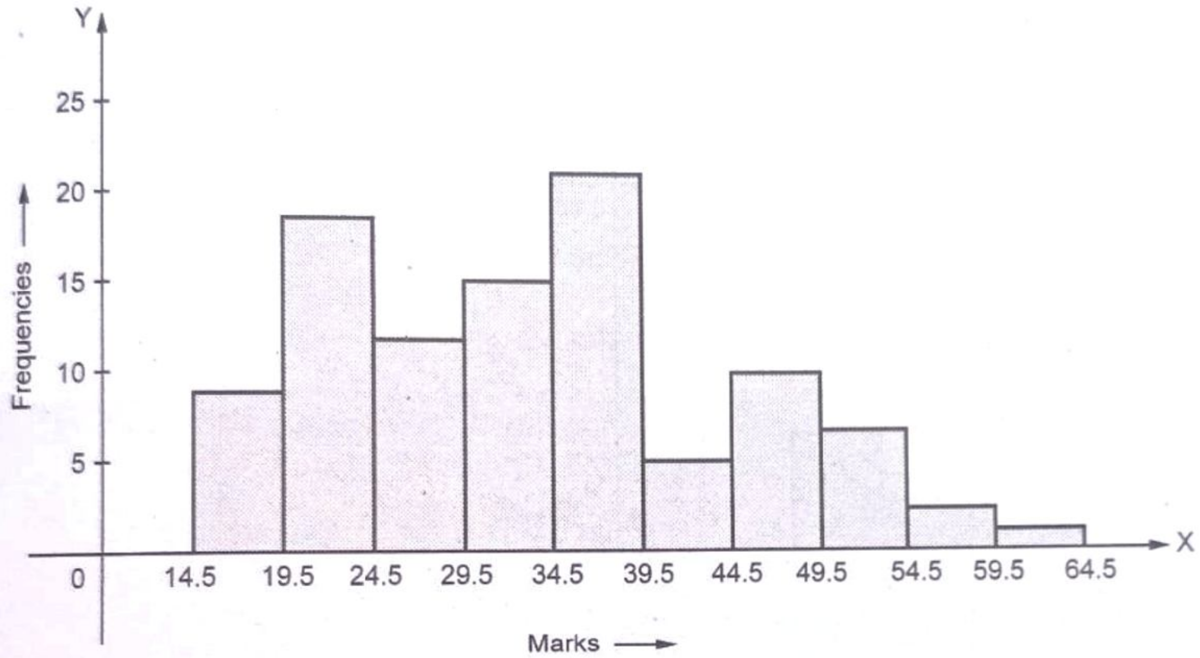

Fig. 2.4.1 Histogram

Note : Upper and lower class limits of the new exclusive types classes are known as class boundaries.

- Itch is gap between upper limit of any class and lower limit of the succeeding class, class boundaries for any class are given by,

$$\text{Upper class boundary} = \text{Upper class limit} + \frac{h}{2}$$

$$\text{Lower class boundary} = \text{Lower class limit} - \frac{h}{2}$$

Example 2.4.1 *Draw a histogram to represent the following frequency distribution.*

Marks obtained	0-10	10-20	20-30	30-40	40-50
No. of students	12	16	5	38	3

Solution :

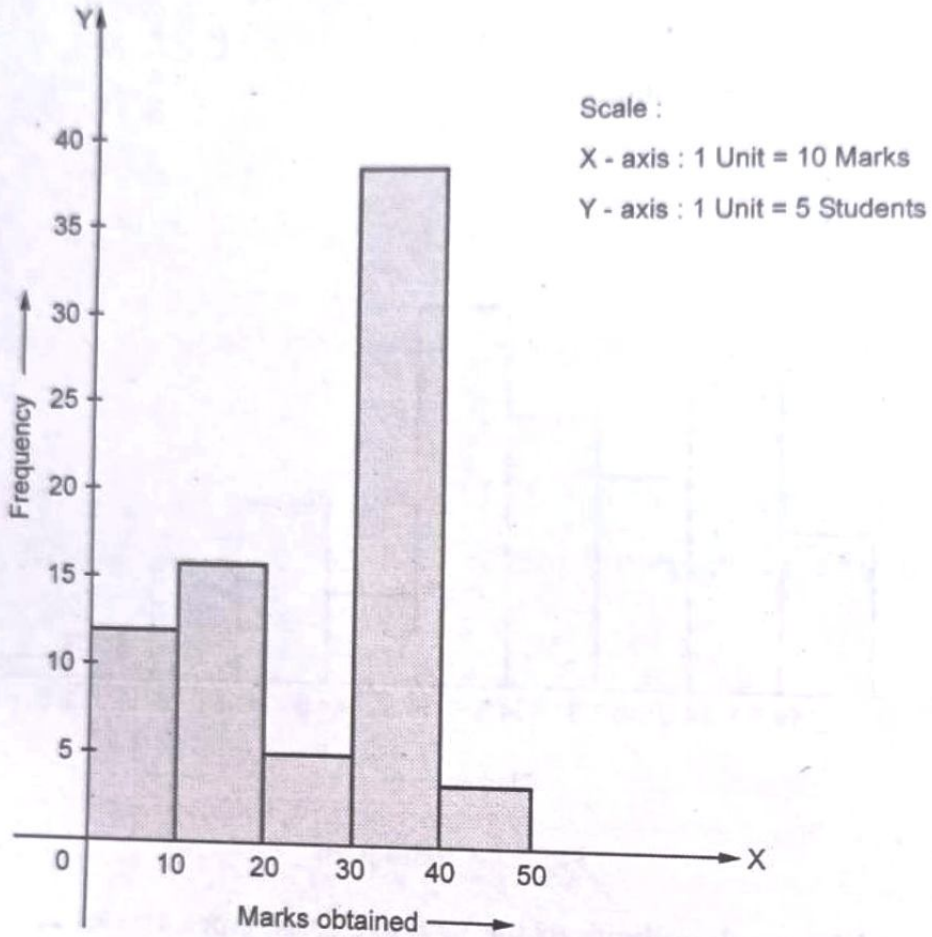

Fig. 2.4.2 Histogram

Example 2.4.2 *Draw a histogram to represent following frequency distribution -*

Age group (in years)	0-20	21-40	41-60	61-80	81-100	101-120
Population	500	2100	2200	2000	1600	400

Solution :

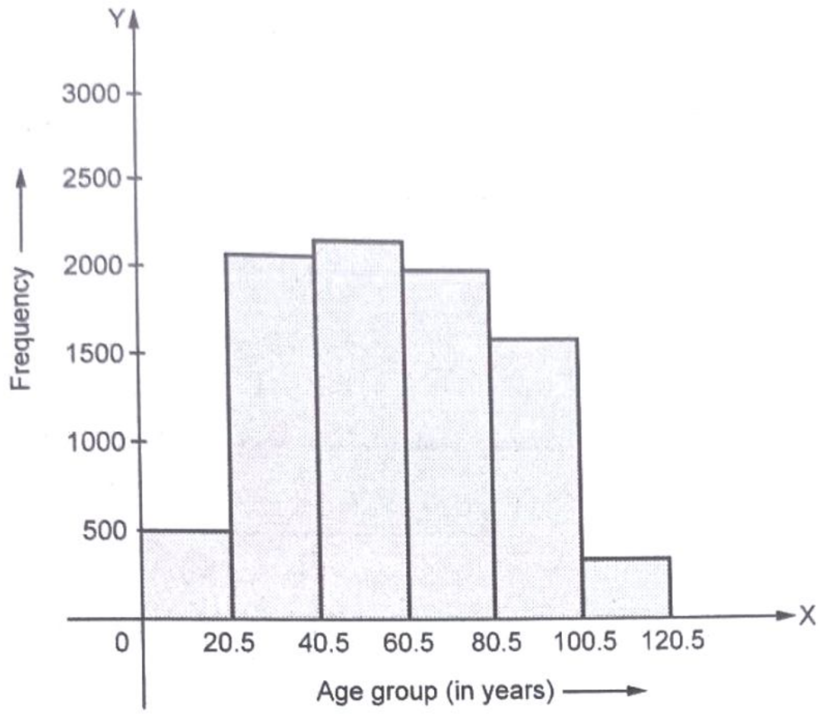

Fig. 2.4.3 Histogram

Example 2.4.3 *Draw a histogram for the following data -*

Age (in years)	2-5	5-11	11-12	12-14	14-15	15-16
No. of boys	6	6	2	5	1	3

Solution : Since classes are of unequal width, we need to compute frequency density.

Age (in years)	2-5	5-11	11-12	12-14	14-15	15-16
Frequency density	$\frac{6}{3}=2$	$\frac{6}{6}=1$	$\frac{2}{1}=2$	$\frac{5}{2}=2.5$	$\frac{1}{1}=1$	$\frac{3}{1}=3$

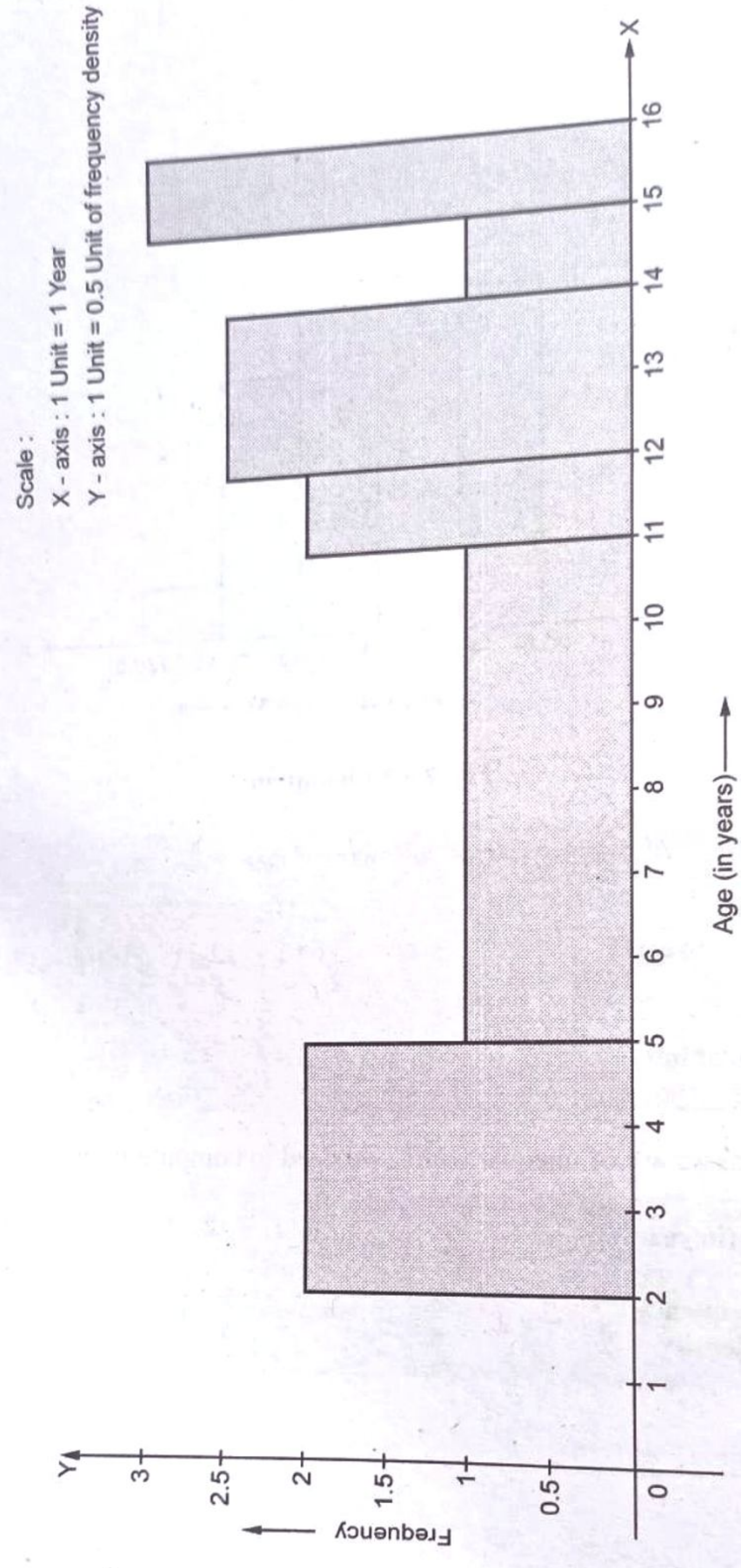

Fig. 2.4.4 Histogram

2.4.2 Frequency Polygon

- Graphical representation connects the variable values to corresponding base values. It must be a uniform and smooth curve. In histogram it does not satisfy these conditions. Therefore the alternate method of presentation of frequency distribution is frequency polygon. It provides clarity of understanding.
- Frequency polygon are constructed for two types of distribution,

  A) Ungrouped distribution B) Grouped distribution

A) Ungrouped distribution :

- Frequency polygon are constructed variate values along the X axis and corresponding frequencies along the Y axis. The nature of this plot is generally obtained by joining the points by a straight line.

B) Grouped distribution :

- In this case mid values of class intervals are along the X-axis. Polygon is obtained by joining the middle points by means of a straight line. Polygon can be drawn by a free hand curve for class intervals of small width.

Example 2.4.4 *Draw a frequency polygon for the following data :*

Monthly house rent	100-300	300-500	500-700	700-900	900-1100	1100-1300
No. of families	6	16	24	20	10	4

Solution : Along x-axis mid values of classes are taken and along y-axis frequency is taken. First point is (200, 6), second point will be (400, 16), third point will be (600, 24) and so on. Last point will be (1200, 4). To get a closed figure we take two more points (0, 0) and (1400, 0). Join all these points by line segments and we get frequency polygons. (Refer Fig. 2.4.5 on next page)

Example 2.4.5 *Draw a frequency polygon for following data :*

Marks obtained	0-10	10-20	20-30	30-40	40-50
No. of students	5	12	43	32	8

Solution : Mid values of classes are taken along X-axis and frequency along Y-axis. First point to plot is (5, 5), the second point will be (15, 12) and so on. Last point will be (45, 8). For a closed figure we take two more points (0, 0) and (50, 0). Join all points by line segments we and get frequency polygons. (Refer Fig. 2.4.6 on next page)

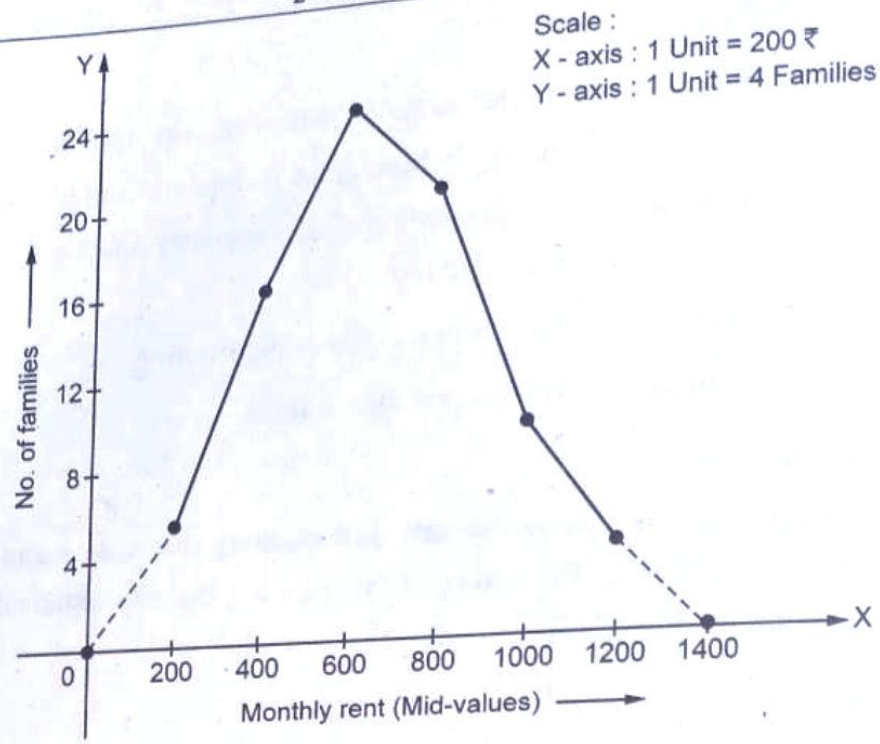

Fig. 2.4.5 Frequency polygon

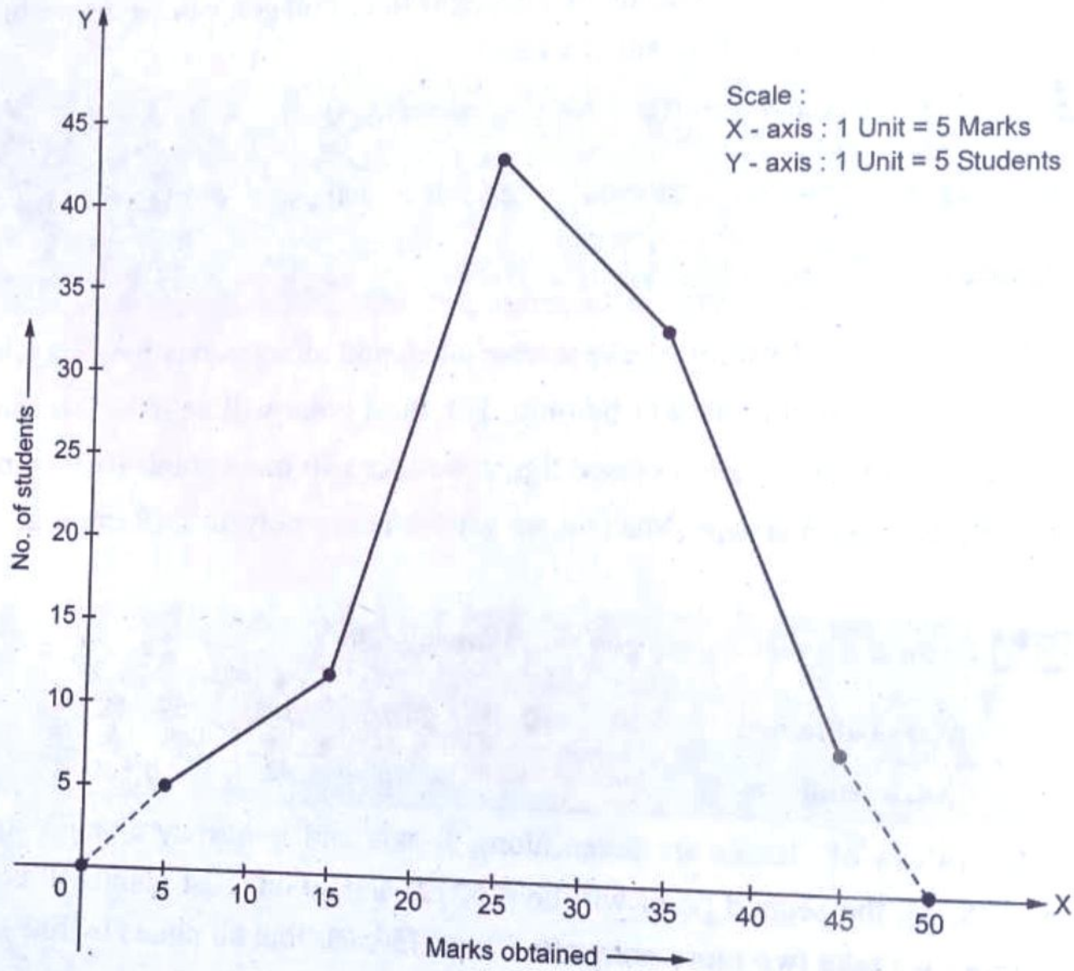

Fig. 2.4.6 Frequency polygon

2.5 Advantages and Limitations of Graphic Representation of Frequency Distribution

Advantages :

1) Graphical representation enables μs to understand the resultant conclusion after analysis of data.
2) It is a collective presentation of information.
3) Graphs are more effective and impressive facts of tabulated values.
4) Everybody can read, analyse and understand the graph.
5) It is the understanding of data, concept for a longer time.
6) It helps to compare.

Limitations :

1) Exact results and values are not provided by the graph but it is in the form of approximate values.
2) Insufficient for analysis of statistical data but gives a general nature of the phenomenon.

2.6 Central Tendency

- On account of classification and frequency polygon we get an idea about the shape of frequency distribution. We observe that all class-frequencies are not the same. Initially the frequency is small in magnitude and then increases, it reaches maximum in the middle part and again falls down. We conclude that the observations are not uniformly spread. Most of the observations get clustered in the central part of data. This property of observations is called the **central tendency.**
- We select a representative observation from the central part. This is referred to as an average or measure of central tendency.

2.7 Average or Measure of Central Tendency

2.7.1 Requisites for an Ideal Measure of Central Tendency

- According to professor Yule, the following are the characteristics which are satisfied by an ideal measure of central tendency.
  1) It must include all the values of observations.
  2) It must be open to perform mathematical treatment repeatedly.
  3) There must be the least effect of extreme values on an average.

4) It must ensure the stability of sampling.

5) It must be defined rigidly.

6) It must be easy to calculate and simple to understand.

2.7.2 Types of Averages or Measures of Central Tendency

- Following are the measures of central tendency :

  1) Arithmetic mean 2) Median

  3) Mode 4) Geometric mean

  5) Harmonic mean.

- Among these arithmetic mean, geometric mean and harmonic mean are called **mathematical averages** and median mode are called **positional averages.**
- The type of average depends on the nature of data. Each method of averages has different advantages and disadvantages.

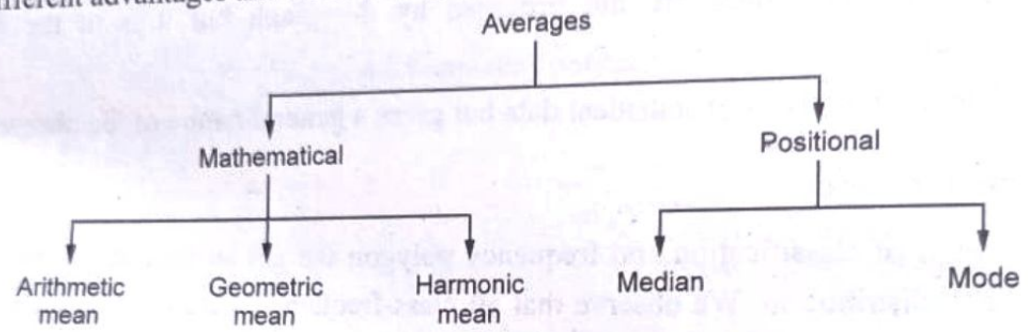

Fig. 2.7.1

2.8 Arithmetic Mean

- This is the method which is used in most computations.
- **Definition :** Sum of set of observations divided by the number of observations called as arithmetic mean

$$\text{A.M.} = \frac{\text{Sum of observations}}{\text{Number of observations}}$$

- There are different types of data analysis of A.M. different types are given as follows.

2.8.1 Row Data or Individual Observations

- Let $x_1, x_2, \ldots, x_n$ be n observations then arithmetic mean $\overline{X}$ is given by,

$$\text{A.M.} = \overline{X} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Example 2.8.1 *Calculate the arithmetic mean of marks scored by a student in 5 subjects given below : 45, 50, 40, 60, 55.*

Solution :

$$\text{A.M.} = \bar{X} = \frac{45+50+40+60+55}{5} = \mathbf{50}$$

Example 2.8.2 *Calculate arithmetic mean of weights of 10 students are 50, 46, 48, 51, 49, 52, 60, 32, 36, 42.*

Solution :

$$\text{A.M.} = \bar{X} = \frac{50+46+48+51+49+52+60+32+36+42}{10}$$

$$\bar{X} = \mathbf{46.6}$$

2.8.2 Ungrouped Data

- Let $x_1, x_2, \ldots, x_n$ be the observations and $f_1, f_2, \ldots, f_n$ be the frequencies then arithmetic mean is given by -

$$\text{A.M.} = \bar{X} = \frac{f_1x_1 + f_2x_2 + \ldots + f_nx_n}{f_1 + f_2 + \ldots + f_n}$$

$$\bar{X} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{\sum_{i=1}^{n} f_i x_i}{N} \qquad \left\{\because \sum_{i=1}^{n} f_i = N\right\}$$

Example 2.8.3 *Calculate arithmetic mean of the frequency distribution*

x	1	2	3	4	5	6	7
f	5	9	12	17	14	10	6

Solution :

$$\text{A.M.} = \frac{5\times1+9\times2+12\times3+17\times4+14\times5+10\times6+6\times7}{5+9+12+17+14+10+6}$$

$$\text{A.M.} = \frac{299}{73} = \mathbf{4.09}$$

OR

x_i	f_i	$f_i x_i$
1	5	5
2	9	18
3	12	36
4	17	68
5	14	70
6	10	60
7	6	42
Total	N = 73	299

$$\text{A.M.} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{\Sigma f_i x_i}{N}$$

$$= \frac{299}{73} = \mathbf{4.09}$$

Example 2.8.4 *Calculate arithmetic mean for the following frequency distribution :*

x	103	110	112	118	95
f	4	6	10	12	3

Solution :

x_i	f_i	$f_i x_i$
103	4	$103 \times 4 = 412$
110	6	$110 \times 6 = 660$
112	10	$112 \times 10 = 1120$
118	12	$118 \times 12 = 1416$
95	3	$95 \times 3 = 285$
Total	N = 35	$\Sigma f_i x_i = 3893$

$$\bar{X} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{3893}{35} = \mathbf{111.2286}$$

OR

- By using change of origin method we define $u = x - a$ where a is any value and prepare the table :

x_i	$u_i = x_i - a$ $u_i = x_i - 100$	f_i	$f_i u_i$
103	3	4	12
110	10	6	60
112	12	10	120
118	18	12	216
95	− 5	3	− 16
Total		$\Sigma f_i = 35$	$\Sigma f_i u_i = 393$

$$\bar{u} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{393}{35} = 11.2286$$

As $$u_i = x_i - 100$$

$$\bar{u} = \bar{x} - 100$$

$$11.2286 = \bar{x} - 100$$

$$\bar{x} = 11.2286 + 100$$

$$\bar{x} = \mathbf{111.2286}$$

Example 2.8.5 *Calculate A.M. for the following frequency distribution :*

C.I.	0-8	8-16	16-24	24-32	32-40	40-48
f	8	7	16	24	15	7

Solution : For calculation of arithmetic mean, we find mid point of each class interval.

C.I.	Mid-values (x_i)	f_i	$f_i x_i$
0-8	4	8	$8 \times 4 = 32$
8-16	12	7	$7 \times 12 = 84$
16-24	20	16	$16 \times 20 = 320$
24-32	28	24	$24 \times 28 = 672$
32-40	36	15	$15 \times 36 = 540$

40-48	44	7	$7 \times 44 = 308$
Total		$\Sigma f_i = 77$	$\Sigma f_i x_i = 195.6$

$$\text{A.M.} = \bar{X} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{1956}{77} = \mathbf{25.4025}$$

Example 2.8.6 *Following is a distribution of weekly salaries of the employees of the firm :*

Salary (in ₹)	0-600	600-1200	1200-1800	1800-2400	2400-3000
No. of employees	3	7	14	27	23

Solution : For calculation of A.M., we calculate midpoint of each class interval. In most of the problems we get midpoints with equally spaced. We choose 'a' as possible closer to $\bar{X}$. Any midpoint in the central part can be taken as 'a'. We define $u_i = \frac{x_i - a}{h}$ where h is class width or any suitable number. From this it is observed that calculation of $\bar{X}$ seemed to be difficult than $\bar{u}$.

Class	Mid-values (x_i)	$u_i = \frac{x_i - 1500}{600}$	f_i	$f_i u_i$
0-600	300	− 2	3	− 6
600-1200	900	− 1	7	− 7
1200-1800	1500	0	14	0
1800-2400	2100	1	27	27
2400-3000	2700	2	23	46
Total	-	-	**74**	**60**

Here $$u_i = \frac{x_i - 1500}{600}$$

Therefore $$\bar{u} = \frac{\bar{X} - 1500}{600}$$

$$600\,\bar{u} = \bar{X} - 1500$$

$$\bar{X} = 600\,\bar{u} + 1500$$

As
$$\bar{u} = \frac{\Sigma f_i u_i}{\Sigma f_i} = \frac{60}{74} = 0.81$$

$$\bar{X} = 600 \times 0.81 + 1500$$

∴
$$\bar{X} = 486 + 1500$$

$$\bar{X} = \mathbf{1986}\ ₹$$

2.8.3 Grouped Data

- In grouped or continuous frequency distribution arithmetic mean is given by,

$$\bar{X} = A + \frac{\Sigma fd}{N} \times h$$

where
$$A = \text{Assumed mean}$$

$$d = \text{Deviation} = \frac{x - A}{h}$$

$$h = \text{Class width}$$

$$N = \Sigma f$$

Example 2.8.7 *Calculation the mean for the following frequency distribution :*

C.I.	0-10	10-20	20-30	30-40	40-50	50-60
Frequency	12	18	27	20	17	6

Solution : Here we take A = 35 and h = 10

C.I.	Mid-values (x)	Frequency (f)	$d = \frac{x - A}{h}$	fd
0-10	5	12	− 3	− 36
10-20	15	18	− 2	− 36
20-30	25	27	− 1	− 27
30-40	35	20	0	0
40-50	45	17	1	17
50-60	55	6	2	12
Total	-	**100**	-	**− 70**

$$\bar{X} = A + \frac{\Sigma fd}{N} h$$

$$= 35 + \frac{(-70)}{100} \times 10$$

$$= 35 - 7$$

$$= \mathbf{28}$$

Example 2.8.8 *Calculate A.M. for the following frequency distribution :*

C.I.	0-10	10-20	20-30	30-40	40-50
f	1	4	6	5	10

Solution : Here we take A = 25 and h = 10

C.I.	Mid-values (x_i)	Frequency (f_i)	$d = \frac{x - A}{h}$	fd
0-10	5	1	− 2	− 2
10-20	15	4	− 1	− 4
20-30	25	6	0	0
30-40	35	5	1	5
40-50	45	10	2	20
Total	-	26	-	19

$$\bar{X} = A + \frac{\Sigma fd}{N} \times h$$

$$= 25 + \frac{19}{26} \times 10$$

$$= 25 + 7.3076$$

$$= \mathbf{32.3076}$$

Example 2.8.9 *Find missing frequency from the following data :*

Marks	0-5	5-10	10-15	15-20	20-25	25-30	30-35
Frequency	10	12	16	-	14	10	8

where the average mark is 16.82.

Solution : Here we consider missing frequency for the class 15-20 as f_4. We prepare a frequency distribution table.

Here we consider A = 17.5 and h = 5

C.I.	Mid-values (x_i)	Frequency (f_i)	$d = \frac{x-A}{h}$	fd
0-5	2.5	10	– 3	– 30
5-10	7.5	12	– 2	– 24
10-15	12.5	16	– 1	– 16
15-20	17.5	f_4	0	0
20-25	22.5	14	1	14
25-30	27.5	10	2	20
30-35	32.5	8	3	24
Total	-	**$N = 70 + f_4$**	-	**– 12**

As
$$\bar{X} = A + \frac{\Sigma fd}{N} \times h$$

$$16.82 = 17.5 + \frac{(-12)}{70 + f_4} \times 5$$

$$16.82 = \frac{17.5(70 + f_4) - 60}{70 + f_4}$$

$$16.82(70 + f_4) = 17.5(70 + f_4) - 60$$

$$1177.4 + 16.82f_4 = 1225 + 17.5f_4 - 60$$

$$1177.4 + 16.82f_4 = 1165 + 17.5f_4$$

$$1177.4 - 1165 = 17.5f_4 - 16.82f_4$$

$$12.4 = 0.68f_4$$

$$f_4 = \frac{12.4}{0.68} = 18.23$$

$\therefore$ Missing frequency = **18.23**

2.9 Properties of Arithmetic Mean

Example 2.9.1 *Sum of observations is equal to the product of arithmetic mean and number of observations i.e.* $\left(\sum_{i=1}^{n} x_i = n\bar{X}\right)$

Solution : Proof : Let $x_1, x_2, \ldots, x_n$ be n observations

By definition,

$$\bar{X} = \sum_{i=1}^{n} x_i$$

$$\therefore \quad n\bar{X} = \sum_{i=1}^{n} x_i$$

- **Note** : To calculate $\bar{X}$, n and $\sum_{i=1}^{n} x_i$ are sufficient instead of knowing individual observations.

Example 2.9.2 *Algebraic sum of deviations of observations from their arithmetic mean is zero i.e. $\Sigma(x_i - \bar{X}) = 0$*

Solution : Let $x_1, x_2, \ldots, x_n$ be n observations and deviations are $(x_1 - \bar{X}), (x_2 - \bar{X}), \ldots (x_n - \bar{X})$

$$\text{Sum of deviations} = \sum_{i=1}^{n} (x_i - \bar{X})$$

$$= \sum_{i=1}^{n} x_i - n\bar{X}$$

$$= n\bar{X} - n\bar{X} \qquad \ldots \{\text{Using property (1) } n\bar{X} = \Sigma x_i\}$$

$$= \mathbf{0}$$

Example 2.9.3 *Sum of squares of the deviations taken from arithmetic mean is minimum i.e. $\Sigma(x_i - \bar{X})^2 \leq \Sigma(x_i - a)^2$*

Solution : Let $Z = \Sigma(x_i - a)^2$ be the sum of the squares of the deviations of given values from any arbitrary point a.

- We have to prove Z is minimum when $a = \bar{X}$.
- By applying principle of maxima and minima, Z will be minimum for a if $\frac{\partial Z}{\partial a} = 0$ and $\frac{\partial^2 Z}{\partial a^2} > 0$

As

$$Z = \Sigma(x_i - a)^2 \qquad \ldots(1)$$

$$\frac{\partial Z}{\partial a} = -2\Sigma(x_i - a) = 0 \qquad \ldots(2)$$

$$\Rightarrow \quad \Sigma(x_i - a) = 0$$

$$\Rightarrow \quad \Sigma x_i - \Sigma a = 0$$

$$n\bar{X} - na = 0$$

$\Rightarrow$ ... {Using property (1) $\Sigma x_i = n\bar{X}$}

$$\bar{X} = a$$

$\Rightarrow$

Again $$\frac{\partial^2 Z}{\partial a^2} = -2\Sigma(-1) \quad ...(3)$$

Therefore, $$\frac{\partial Z}{\partial a} = 0 \text{ at } a = \bar{X} \text{ and}$$

$$\frac{\partial^2 Z}{\partial a^2} > 0 \text{ at } a = \bar{X}$$

Hence z is minimum at $a = \bar{X}$

$$= 2n > 0$$

Therefore, $$\Sigma(x_i - \bar{X})^2 \leq \Sigma(x_i - a)^2$$

OR

- Let 'a' be any arbitrary constant.
- Then $x_i - \bar{X}$ is deviation of x_i from $\bar{X}$
- $x_i - a$ is deviation of x_i from a

$$\begin{aligned}
\Sigma(x_i - a)^2 &= \Sigma(x_i - \bar{X} + \bar{X} - a)^2 \\
&= \Sigma[(x_i - \bar{X} + \bar{X} - a)]^2 \\
&= \Sigma[(x_i - \bar{X})^2 + 2(x_i - \bar{X})(\bar{X} - a) + (\bar{X} - a)^2] \\
&= \Sigma(x_i - \bar{X})^2 + 2\Sigma(x_i - a)(x_i - \bar{X}) + \Sigma(\bar{X} - a)^2 \\
&= \Sigma(x_i - \bar{X})^2 + 0 + n(\bar{X} - a)^2 \qquad ... \{\because \Sigma(x_i - \bar{X}) = 0\} \\
&= \Sigma(x_i - \bar{X})^2 + n(\bar{X} - a)^2 \\
&= \Sigma(x_i - \bar{X})^2 + \text{Non-negative quantity}
\end{aligned}$$

$\therefore$ $$\Sigma(x_i - a)^2 \geq \Sigma(x_i - \bar{X})^2$$

Example 2.9.4 *(Mean of combined groups) If $\bar{x}_1$ be the arithmetic mean of first group of size n_1 and $\bar{x}_2$ be the arithmetic mean of second group of size n_2 then,*

combines mean $= \bar{x}_c = \dfrac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1 + n_2}$

Solution :

Let $x_{11}, x_{12}, \ldots, x_{1n_1}$ be n_1 members of the first series.

$x_{21}, x_{22}, \ldots, x_2n_2$ be n_2 members of the second series.

Then by definition,

$$\bar{x}_1 = \frac{x_{11} + x_{12} + \ldots + x_1n_1}{n_1}$$

$$\bar{x}_1 = \frac{\text{Sum of observations in first group}}{n_1}$$

$$\bar{x}_2 = \frac{x_{21} + x_{22} + \ldots + x_{n2}}{n_2}$$

$$\bar{x}_2 = \frac{\text{Sum of observations in second group}}{n_2}$$

$\therefore$ Sum of observations in first group $= n_1\bar{x}_1$

Sum of observations in second group $= n_2\bar{x}_2$

$$\bar{x}_c = \frac{\text{(Sum of observations in first group) + (Sum of observations in second group)}}{\text{(Size of first group) + (Size of second group)}}$$

$$\bar{x}_c = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$

Remark : Above result can be generalised to k ($k \geq 2$) groups as follows :

Let k groups of size of i^{th} group as n_i and the arithmetic mean as $\bar{x}_i (i = 1,2,3,\ldots,k)$. Then $\bar{x}_c$ is the arithmetic mean of all k groups combined together is given by,

$$\bar{x}_c = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \ldots + n_k\bar{x}_k}{n_1 + n_2 + \ldots + n_k}$$

$$= \frac{\sum_{i=1}^{k} n_i\bar{x}_i}{\sum_{i=1}^{k} n_i}$$

Example 2.9.5 *Average weight of 70 boys is 55 kg and the average weight of 60 girls is 50 kg. Calculate the mean of a combined group of boys and girls.*

Solution : Let $\bar{x}_1$ be the average of boys with n_1 be the size of the group of boys.

Let $\bar{x}_2$ be the average of girls with n_2 be the size of the group of girls.

$\therefore$ $n_1 = 70$ $\quad \bar{x}_1 = 55$

$n_2 = 60$ $\quad \bar{x}_2 = 50$

$$\text{Combined mean} = \bar{x}_c = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} = \frac{70 \times 55 + 60 \times 50}{70 + 60}$$

$$= \frac{6850}{130} = \mathbf{52.6923}$$

Example 2.9.6 *Arithmetic mean of 40 items is 100. While checking, it was noticed that observation 76 was misread as 67. Find the correct value of mean.*

Solution :

$$\text{Incorrect mean} = 100 = \frac{\text{Incorrect sum}}{n}$$

$$\text{Incorrect sum} = 100 \times 40 = 4000$$

$$\text{Correct sum} = \text{Incorrect sum} + \text{Correct observation} - \text{Incorrect observation}$$

$$= 4000 + 76 - 67 = 4009$$

$$\text{Correct mean} = \frac{\text{Correct sum}}{n} = \frac{4009}{40} = \mathbf{100.225}$$

Example 2.9.7 *Average salary of male employees in a firm was ₹2500 and that of females was ₹2000. The mean salary of all employees was 2200. Find the percentage of male and female employees.*

Solution : Let n_1 and n_2 be the number of male and female employees in the firm respectively. And $\bar{x}_1$, $\bar{x}_2$ be the average salary of male and female employees respectively.

Let $\bar{x}$ be the average salary of all workers in the firm.

Given :

$$\bar{x}_1 = 2500, \; \bar{x}_2 = 2000, \; \bar{x} = 2200$$

As
$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$

$$2200 = \frac{n_1 \times 2500 + n_2 \times 2000}{n_1 + n_2}$$

$$2200(n_1 + n_2) = 2500n_1 + 2000n_2)$$

$$2200n_1 + 2200n_2 = 2500n_1 + 2000n_2$$

$$2200n_2 - 2000n_2 = 2500n_1 - 2200n_1$$

$$\frac{n_1}{n_2} = \frac{200}{300}$$

$$\frac{n_1}{n_2} = \frac{2}{3}$$

$\therefore$ Percentage of male employees in the firm,

$$= \frac{2}{2+3} \times 100 = \frac{2}{5} \times 100 = 40$$

and percentage of female employees in the firm

$$= \frac{3}{2+3} \times 100 = \frac{3}{5} \times 100 = 60$$

2.10 Merits and Demerits of Arithmetic Mean

Arithmetic means satisfies all the needs of a good average. Hence most commonly it is used in practice. Merits and demerits of arithmetic mean are as follows :

Merits :

1) It is a very simple and easy tool. It provides easy calculation.
2) Each and every reading is covered in this method.
3) It provides stability of sampling.
4) It is easy to understand at any level of analysis.
5) It is not limited to individual observation but also covers the mean of a group of observations.
6) It is defined rigidly.
7) It is independent of fluctuations in sampling. It does not affect the final result.

Demerits :

1) It cannot be located graphically nor can it be determined by inspection.
2) Arithmetic means can not deal with qualitative characteristics. It is limited to quantitative data only.
3) For this, continuity of observations is required. It is not applicable to observations with missing terms.
4) It is affected by extreme values.
5) Accuracy of arithmetic means depends on the source of data from which it is derived.
6) It is not suitable in extremely asymmetrical distribution.

2.11 Weighted Mean

- While determining the arithmetic mean of given values in a data, equal importance is given to all values contained in it. But some of the values may be more important as compared to other present values in a table. Average value is a common value for all readings in the data. According to the importance of that value in a data, some weights must be given to it. For example, if cost of living of a particular group is surveyed then simple means will not solve the purpose. It contains the cost of commodities are used by them. But all commodities are not equally important. Food items are important but tea, coffee, confectionery, etc. are not so important.

- Let w_i be the weight attached to the item x_i where $i = 1,2,3,\ldots,n$. Then we define.

$$\text{Weighted arithmetic mean or weighted mean} = \frac{\sum\limits_i w_i x_i}{\sum\limits_i w_i}$$

- Weighted arithmetic mean is the same as simple the mean with frequencies $f_i(i = 1,2,\ldots,n)$ replaced by $w_i(1 = 1,2,\ldots,n)$ the weights.
- Weighted mean is meaningful only if each value is equal. For the larger weight of a larger item, it will be a higher value. For smaller weights, it will be a small value. If larger items are given small weight and smaller items given larger weight then weighted mean results in smaller value than simple mean.

Example 2.11.1 *Find simple and weighted arithmetic mean of the first n numbers, the weights being the corresponding numbers.*

Solution : First n natural numbers are $1,2,\ldots,n$

We know that,

$$1 + 2 + 3 + \ldots + n = \frac{n(n+1)}{2}$$

$$1^2 + 2^2 + 3^2 + \ldots + n^2 = \frac{n(n+1)(2n+1)}{6}$$

$$\text{Simple A.M. } (\bar{X}) = \frac{\Sigma x}{n} = \frac{1 + 2 + 3 + \ldots + n}{n} = \frac{n(n+1)}{2} \times \frac{1}{n} = \frac{n+1}{2}$$

$$\text{Weighted A.M. } (\bar{X}_w) = \frac{\Sigma wx}{\Sigma w} = \frac{1^2 + 2^2 + 3^2 + \ldots + n^2}{1 + 2 + 3 + \ldots + n}$$

$$= \frac{n(n+1)(2n+1)}{6} \times \frac{2}{n(n+1)} = \frac{(2n+1)}{3}$$

2.12 Median

- Median of a distribution is the value of the variable which divides it into two equal parts. It is the central observation also called the **positional average.**

2.12.1 Ungrouped Data

- If the number of observations is odd then the median is the middle value after the values have been arranged in ascending or descending order of magnitude.

For example : The median of the values 25, 35, 10, 13, 5

- First of all arrange the observations either in ascending or descending order

  5, 10, 13, 25, 35

  Middle value is 13.

  ∴ Median is 13.

- If observations are even then there are two middle terms and median is obtained by taking the arithmetic mean of the middle terms.

  For example, median of the values 25, 20, 15, 35, 18, 50.

  Ascending order : 15, 18, 20, 25, 35, 50

$$\text{Median} = \frac{20+25}{2} = \frac{45}{2} = \mathbf{22.5}$$

2.12.2 Discrete Frequency Distribution

- For discrete frequency distribution median is obtained by considering the cumulative frequencies. Steps for calculating median are as follows :

  1) Find $\frac{N}{2}$ where $N = \Sigma f$.

  2) Find cumulative frequency equal to or just greater than $\frac{N}{2}$.

  3) The corresponding value of x is median.

Example 2.12.1 *Obtain median for the following frequency distribution :*

x	1	2	3	4	5	6	7	8	9
f	8	10	11	16	20	25	15	9	6

Solution : Here $N = 120 \Rightarrow \frac{N}{2} = 60$

x	f	c.f.
1	8	8
2	10	18
3	11	29
4	16	45
5	20	65
6	25	90
7	15	105
8	9	114
9	6	120

Cumulative frequency just greater than $\frac{N}{2}$ is 65.

Value of x corresponding to 65 is 5.

∴ Median = **5**

Example 2.12.2 *Obtain the median for the following frequency distribution :*

x	1	2	3	4	5	6	7	8	9
f	7	10	12	15	18	20	25	38	45

Solution : Here $N = 190 \Rightarrow \frac{N}{2} = 95$

x	f	c.f.
1	7	7
2	10	17
3	12	29
4	15	44
5	18	62
6	20	82
7	25	107
8	38	145
9	45	190

Cumulative frequency is just greater than $\frac{N}{2}$ is 107.

Value of x corresponding to 107 is 7.

$\therefore$ Median = 7

2.12.3 Continuous Frequency Distribution

For continuous frequency distribution, the class corresponding to the cumulative frequency just greater than $\frac{N}{2}$ is called **median class** and the value of median is obtained by using following formula -

$$\text{Median} = l + \frac{h}{f}\left[\frac{N}{2} - C\right]$$

where

l = Lower limit of median class

f = Frequency of the median class

h = Class interval of median class

c = c.f. of the class preceding to the median class

$N = \Sigma f$

Graphical Explanation of Formula :

Median is along x-axis and less than c.f. is along the y-axis with $\frac{N}{2}$ where N is the total frequency.

We locate $\frac{N}{2}$ and median class. Let l be the lower boundary of the median class. h is the width and $l + h$ be the upper boundary of class. f is the frequency of median class and c.f. be the less than type cumulative frequency of the preceding median class. Use less than the cumulative frequency curve.

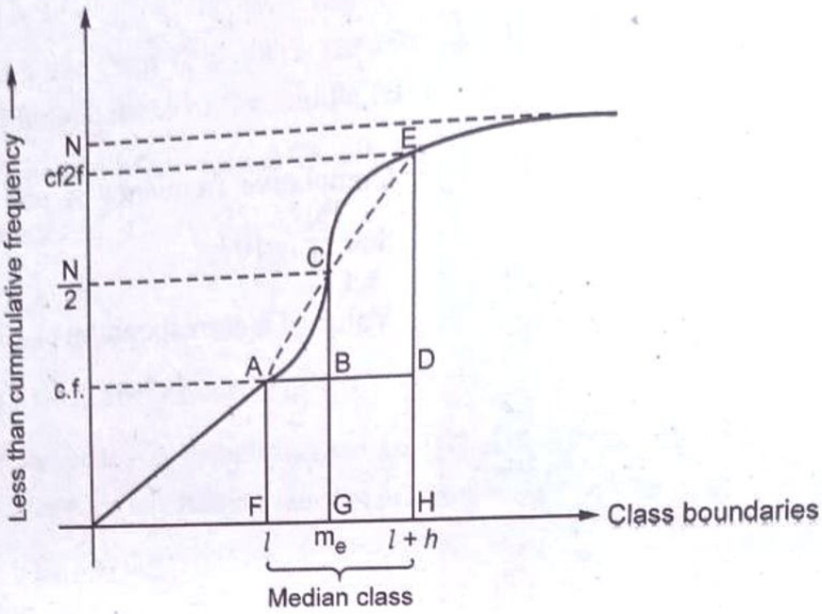

Fig. 2.12.1

Construction :

Let C (M_e, $\frac{N}{2}$) be point on less than c.f. with x-co-ordinate as median (M_e). Let CG ⊥ x-axis. Let f be the lower boundary and H be the upper boundary of the median class. Thus (F, H) is median class. Let AF ⊥ x-axis, EH ⊥ x-axis. Let AD ll^{el} to FH. Let us join ACE in a straight line. We assume over the median class less than c.f. is uniformly increasing.

$$\text{Median} = OG = OF + FG = l + FG$$

Now Δ ABC and Δ ADF are similar,

$$\frac{AB}{AD} = \frac{BC}{DE}$$

$$\frac{FG}{FH} = \frac{CG - BG}{HE - HD} \qquad \left\{\begin{array}{l} \because AB = FG \\ AD = FH \end{array}\right.$$

$$\frac{M_e - l}{h} = \frac{N/2 - c.f.}{f} \qquad \ldots FH = \text{Width OR median class}$$

$$M_e - l = h\left(\frac{\frac{N}{2} - c.f.}{f}\right)$$

$$M_e = l + h\left(\frac{\frac{N}{2} - c.f.}{f}\right)$$

Example 2.12.3 *Obtain median for the following frequency distribution :*

Marks obtain	0-20	21-40	41-60	61-80	81-100
No. of students	1	9	32	16	7

Solution :

C.I.	f	Less than c.f.
0-20.5	1	1
20.5-40.5	9	10
40.5-60.5	32	42
60.5-80.5	16	58
80.5-100	7	65

→ Median class (40.5-60.5)

Here N = 65

$\frac{N}{2} = 32.5$

$l = 40.5,\ h = 20,\ f = 32,\ c.f. = 10$

$$\text{Median} = l + \left(\frac{\frac{N}{2} - c.f.}{f}\right)h$$

$$= 40.5 + \left(\frac{32.5 - 10}{32}\right) \times 20$$

$$= 40.5 + \frac{22.5}{32} \times 20$$

$$= \mathbf{54.5625}$$

Example 2.12.4 *Obtain median for the following frequency distribution*

C.I.	0-10	10-20	20-30	30-40	40-50
f	5	14	29	21	25

Solution :

C.I.	f	Less than c.f.
0-10	5	5
10-20	14	19
20-30	29	48
30-40	21	69
40-50	25	94

→ Median class (20-30)

Here N = 94

$\frac{N}{2} = 47$

$l = 20,\ h = 10,\ f = 29,\ c.f. = 19$

$$\text{Median} = l + \left(\frac{\frac{N}{2} - c.f.}{f}\right) \times h$$

$$= 20 + \left(\frac{47 - 19}{29}\right) \times 10$$

$$= 20 + 9.6551$$

$$= \mathbf{29.6551}$$

2.13 Merits and Demerits of Median

Merits :

1) Median gives positional average. In odd numbers the median is the middle value when numbers are arranged in ascending or descending mode. For even numbers it means of middle terms.
2) It can compute the value for unknown extreme values or missing.
3) The method of calculation is easy and able to find graphically also.
4) The median depends on the status of middle values but not on extreme values.
5) It is very useful tool in calculation of grouped data such as income of group average of weight etc.
6) It provides ease to study certain attributes which cannot be directly measured.
7) In case of grouped data, it is determined with open end intervals or under data.

Demerits :

1) It is not comfortable with long series data
2) It is quite tedious and time consuming for large values.
3) It is not suitable for algebraic manipulation.
4) It is not assumed a representative value in many situations.
5) It is based on the assumption that all frequency data are distributed uniformly with a class interval.

2.14 Mode

It is one of the measures in central tendency. To overcome drawbacks of arithmetic mean mode is used.

Definition :

- The observation with maximum frequency or the most repeated observations is called as **mode.**
- General nature of the frequency curve is bell shaped. Initially frequency is small, it increases and reaches maximum and then it declines value on the x-axis at which the maxima or peak of the frequency curve appears as a mode.

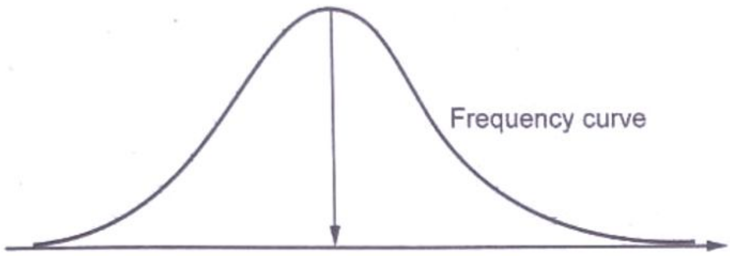

Fig. 2.14.1

- In the analysis of the voters percentages to different political parties are different. But in these results the political party having largest votes is assumed to be a representative. This is the mode and it is the appropriate average.

2.14.1 Discrete Frequency Distribution

In this case we can find observations with the largest frequency just by inspection.

Example 2.14.1 *Find mode of the following frequency distribution :*

x	5	6	7	8	9	10
f	15	7	9	21	18	13

Solution : Since maximum frequency is associated with observation 8. So mode is 8.

2.14.2 Continuous Frequency Distribution

In this case mode is given by the formula,

$$\text{Mode} = l + \frac{h(f_1 - f_0)}{2f_1 - f_0 - f_2}$$

where

l = Lower boundary of modal class

h = Class width of modal class

f_1 = Frequency of modal class

f_0 = Frequency of pre-modal class

f_2 = Frequency of post modal class

Graphical explanation of formula :

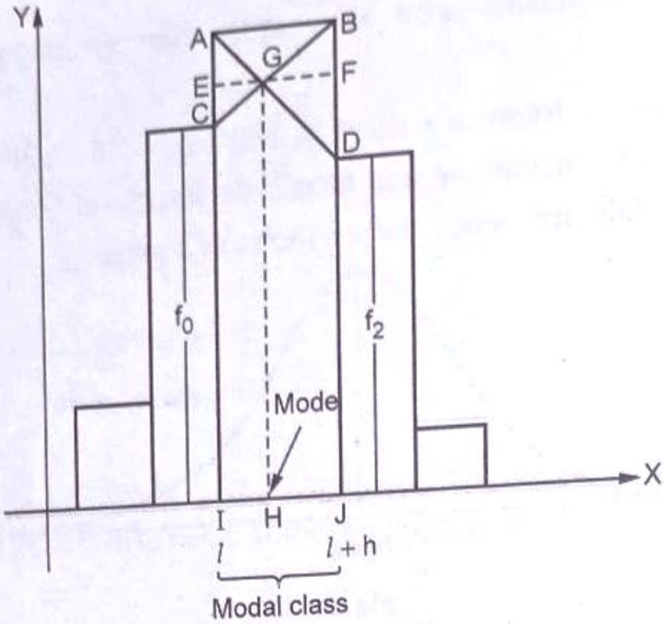

Fig. 2.14.2

- Let A and B be the vertices of modal class. Join the lines AD and CB. Let G be the point of intersection of AD and BC. Let GH $\perp^{ar}$ to X-axis, GF ll^{el} to X-axis.
- Δ CGA and Δ DGB are similar. EG and GF are the altitudes of similar triangles.

Hence,

$$\frac{EG}{GF} = \frac{CA}{DB}$$

$$\frac{IH}{HJ} = \frac{CA}{DB}$$

$$\frac{IH}{IH + HJ} = \frac{CA}{CA + DB}$$

$$\frac{IH}{IJ} = \frac{CA}{CA + DB} \qquad \ldots(1)$$

As

$$CA = IA - IC = f_1 - f_0$$

$$DB = JB - JD = f_1 - f_2$$

$$IH = \text{Mode} - l$$

$$IJ = h = \text{Width of modal class}$$

Hence, equation (1) becomes,

$$\frac{\text{Mode} - l}{h} = \frac{f_1 - f_0}{(f_1 - f_0) + (f_1 - f_2)}$$

$$\frac{\text{Mode} - l}{h} = \frac{f_1 - f_0}{2f_1 - f_0 - f_2}$$

$$\text{Mode} - l = h\left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right)$$

$$\text{Mode} = l + h\left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right)$$

Example 2.14.2 *Calculate mode for the following distribution :*

C.I.	0-10	10-20	20-30	30-40	40-50	50-60	60-70
f	10	7	15	28	25	10	15

Solution :

C.I.	f
0-10	10
10-20	7
20-30	15
30-40	28
40-50	25
50-60	10
60-70	15

Here the maximum frequency is 28.

Thus class 30-40 is modal class.

where,

$l = 30, f_1 = 28, f_0 = 15, f_2 = 25, h = 10$

$$\therefore \quad \text{Mode} = l + h\left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right)$$

$$= 30 + 10\left(\frac{28 - 15}{2 \times 28 - 15 - 25}\right)$$

$$= 30 + \frac{10 \times 13}{56 - 40}$$

$$= 30 + \frac{130}{16}$$

$$= \mathbf{38.125}$$

Example 2.14.3 *Calculate mode for the following distribution :*

Daily income	0-30	31-60	61-90	91-120	121-150	151-180
No. of persons	40	115	192	132	37	45

Solution :

C.I.	f
0-30.5	40
30.5-60.5	115
60.5-90.5	192
90.5-120.5	132
120.5-150.5	37
150.5-180	45

Here maximum frequency is 192.

So the modal class is 60.5 - 90.5.

where,

$l = 60.5$, $f_1 = 192$, $f_0 = 115$,

$f_2 = 132$, $h = 30$

$$\text{Mode} = l + h\left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right)$$

$$= 60.5 + 30\left(\frac{192-115}{2 \times 192 - 115 - 132}\right)$$

$$= 60.5 + \frac{30 \times 77}{384 - 247}$$

$$= 60.5 + \frac{2310}{137}$$

$$= 60.5 + 16.8613 = \mathbf{107.3613}$$

Remark : Sometimes, mode is estimated from the mean and the median. For a symmetric distribution mean, median and mode coincide. If the distribution is moderately asymmetrical mean, median and mode obey the following relationship,

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

2.15 Merits and Demerits of Mode

Merits :

1) As compared to other methods, it is very easy to understand.
2) The easiest tool for descriptive measure of average.
3) It is possible to locate in ungrouped data.
4) It does not depend on extreme values.
5) It is an important and useful average value.

Demerits :

1) Precise determination is missing.
2) Single mode does not exist in many calculations. Therefore it is not useful as an average in such calculation.
3) Algebraic manipulation is not possible.
4) It does not involve each value of a set.

Note : Mode is the average to be used to find the ideal size e.g. - In business forecasting, in the manufacture of ready-made garments, shoes etc.

2.16 Geometric Mean

Geometric mean of a set of n observations is the n^{th} root of their product. Thus the geometric mean G of n observations x_i , i = 1, 2, …, n is given by,

$$G = (x_1, x_2, \dots x_n)^{1/n} \qquad \dots (2.16.1)$$

Computation is facilitated by using logarithms,

Taking logarithm of both, sides

$$\log G = \frac{1}{n}[\log x_1 + \log x_2 + \dots + \log x_n]$$

$$\log G = \frac{1}{n} \sum_{i=1}^{n} \log x_i$$

$$G = \text{Antilog}\left[\frac{1}{n} \sum_{i=1}^{n} \log x_i\right]$$

In case of frequency distribution (x_i, f_i) i = 1, 2, …, n then geometric mean G is,

$$G = \left(x_1^{f1} \cdot x_2^{f2} \cdot x_3^{f3} \dots x_n^{fn}\right)^{1/N} \qquad \text{where } N = \sum_{i=1}^{n} f_i$$

Taking logarithm of both sides

$$\log G = \frac{1}{N}[f_1 \log x_1 + f_2 \log x_2 + \dots + f_m \log x_n]$$

$$\log G = \frac{1}{N} \sum_{i=1}^{n} f_i \log x_i$$

$$\therefore \quad G = \text{Antilog}\left[\frac{1}{N} \sum_{i=1}^{n} f_i \log x_i\right]$$

For grouped frequency distribution or continuous frequency distribution, x is taken to be the value corresponding to the midpoint of the class intervals.

Example 2.16.1 *Find the geometric mean of 24,18, 15, 60, 50*

Solution :

$$G.M. = (24 \cdot 18 \cdot 15 \cdot 60 \cdot 50)^{1/5} = 28.6905$$

OR

x_i	$\log x_i$
24	1.3802
18	1.2552
15	1.1761
60	1.7782
50	1.6989

$$G = \text{Antilog}\left(\frac{1}{n}\sum \log x_i\right)$$

$$G = \text{Antilog}\,(1.45772)$$

$$= \mathbf{28.6893}$$

Example 2.16.2 *Show that in finding the arithmetic mean of a set of readings on a thermometer it does not matter whether we measure temperature in centigrade or Fahrenheit, but that in finding the geometric mean it does matter which scale we use.*

Solution : Let, $C_1, C_2, \ldots C_n$ be n readings on the centigrade thermometer.

Then the arithmetic mean is given by,

$$\bar{C} = \frac{1}{n}(C_1 + C_2 + \ldots + C_n)$$

If F and C be the readings in Fahrenheit and Centigrade respectively, then we have the relation.

$$\frac{F-32}{180} = \frac{C}{100} \Rightarrow F = 32 + \frac{9}{5}C$$

Thus Fahrenheit equivalents of $C_1, C_2, \ldots C_n$ are ... $32 + \frac{9}{5}C_1,\ 32 + \frac{9}{5}C_2, \ldots, 32 + \frac{9}{5}C_n$ respectively.

Hence arithmetic mean of the reading in Fahrenheit is,

$$\bar{F} = \frac{1}{n}\left[(32 + \frac{9}{5}C_1) + (32 + \frac{9}{5}C_2) + \ldots + (32 + \frac{9}{5}C_n)\right]$$

$$= \frac{1}{n}\left[32\,n + \frac{9}{5}(C_1 + C_2 + \ldots + C_n)\right]$$

$$= 32 + \frac{9}{5}\left(\frac{C_1 + C_2 + \ldots + C_n}{n}\right)$$

$$= 32 + \frac{9}{5}\bar{C}, \text{ which is the Fahrenheit equivalent to } \bar{C}.$$

Hence in finding the arithmetic mean of a set of n readings on a thermometer, it is immaterial whether we measure temperature, it is centigrade or Fahrenheit.

Geometric mean G, of n readings in centigrade is,

$$G = (C_1 \cdot C_2, \ldots, C_n)^{1/n}$$

Geometric mean G, of Fahrenheit equivalent to $C_1, C_2, \ldots, C_n$ is

$$G = \left[(32 + \frac{9}{5}C_1) \cdot (32 + \frac{9}{5}C_2) \cdot (32 + \frac{9}{5}C_3) \ldots + (32 + \frac{9}{5}C_n)\right]^{1/n}$$

Which is not equal to Fahrenheit equivalent of G.

Hence in finding the geometric mean of the n reading on the thermometer, the scale (centigrade or equal to Fahrenheit) is important.

Note : Geometric mean of the combined group is given by,

$$\text{Log } G = \frac{n_1 \log G_1 + n_2 \log G_2}{n_1 + n_2}$$

where n_1 and n_1 are the sizes G_1 and G_2 be the geometric means of two series respectively.

2.17 Merits and Demerits of Geometric Mean

Merits

1) It involves all values contained in observations.
2) Geometric mean provides precise determination and each and every observations is considered.
3) It has algebraic properties and can be manipulated by a algebraic method.
4) It tries to keep the effect of extreme values to a minimum.
5) Geometric mean provides a representative average value in different large or small types of data.

Demerits

1) It refers to the positive value of a variable for a negative, geometric mean is undefined.
2) It uses logarithms.

2.18 Harmonic Mean

Harmonic mean is defined as the reciprocal of the arithmetic mean of the reciprocals of given values.

Let $x_1, x_2, \ldots x_n$ are observations then harmonic mean is given by,

$$H = \frac{1}{\left(\frac{1}{x_1}+\frac{1}{x_2}+\ldots+\frac{1}{x_n}\right)/n}$$

$$= \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{x_i}\right)}$$

In case of frequency distribution (x_i, f_i), $i = 1,2,\ldots n$, then harmonic mean is given by,

$$H = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{f_i}{x_i}\right)} \quad \text{where } N = \sum_{i=1}^{n} f_i$$

Example 2.18.1 *Compute harmonic mean of 72, 80, 28, 35, 58.*

Solution :

$$\text{H.M.} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{x_i}\right)}$$

$$= \frac{1}{\frac{1}{5}\left[\frac{1}{72}+\frac{1}{80}+\frac{1}{28}+\frac{1}{35}+\frac{1}{58}\right]}$$

$$= \frac{5}{0.1077} = \mathbf{46.4253}$$

Example 2.18.2 *Milk is sold at the rates of 8,10,12,15 rupees per liter in four different months. Assuming that equal amounts are spent on milk by a family in the four months, find the average price in rupees per month.*

Solution : Since equal amounts of money are spent by the family for each of the four months the average price of milk per month is given by the harmonic mean of 8,10,12 and 15.

$$\therefore \text{Average price of milk per month} = \frac{1}{\frac{1}{4}\left(\frac{1}{8}+\frac{1}{10}+\frac{1}{12}+\frac{1}{15}\right)}$$

$$= \frac{4\times 120}{15+12+8+10} = \mathbf{10.67}$$

Example 2.18.3 *Reena drives a car from her house to her college at a speed of 10 km/h and back from the college to her house at 15 km/h find the average speed.*

Solution : Let the distance from house to college be x km. In going from house to college, the Distance (x km) is covered in $\frac{x}{10}$ hours, while in coming from college to house the distance is covered in $\frac{x}{15}$ hours.

Thus a total distance of 2x km is covered in $\left(\frac{x}{10}+\frac{x}{15}\right)$ hours.

Hence

$$\text{Average speed} = \frac{\text{Total distance travelled}}{\text{Total time taken}}$$

$$= \frac{2x}{\left(\frac{x}{10}+\frac{x}{15}\right)}$$

$$= \frac{2x}{x\left(\frac{1}{10}+\frac{1}{15}\right)} = \frac{2}{\left(\frac{1}{10}+\frac{1}{15}\right)} = \frac{2}{\left(\frac{15+10}{150}\right)}$$

$$= \frac{2\times 150}{(15+10)} = \frac{2\times 150}{25} = \mathbf{12\ km/h}$$

In this case average speed is given by the harmonic mean of 10 and 15 not by the arithmetic mean.

Remark :

1) If equal distances are covered (traveled) per unit of time with speed equal to $v_1, v_2, \ldots, v_n$ then average speed is given by,

$$\text{Average speed} = \frac{1}{\frac{1}{n}\left(\frac{1}{v_1}+\frac{1}{v_2}+\ldots+\frac{1}{v_n}\right)} = \frac{n}{\sum_{i=1}^{n}\left(\frac{1}{v_i}\right)}$$

2) Instead of fixed distance being traveled with changing speed.

Let $S_1, S_2, \ldots, S_n$ are different distances and $v_1, v_3, \ldots, v_n$ are different speeds, the weights being the corresponding distances traveled is,

$$\text{Average speed} = \frac{S_1+S_2+\ldots+S_n}{\left(\frac{S_1}{v_1}+\frac{S_2}{v_2}+\ldots+\frac{S_n}{v_n}\right)} = \frac{\sum_{i=1}^{n} S_i}{\sum_{i=1}^{n}\left(\frac{S_i}{v_i}\right)}$$

3) If $x_1, x_2, \ldots x_n$ are n observations with weights $W_1, W_2, \ldots W_n$ respectively, then weighted harmonic mean is defined as -

$$H = \frac{\sum_{i=1}^{n} W_i}{\sum_{i=1}^{n} \left(\frac{W_i}{x_i}\right)}$$

2.19 Merits and Demerits of Harmonic Mean

Merits :

1) Harmonic means provide precise determination.
2) It is open to algebraic manipulation.
3) It depends on each and every value in observation.
4) It determines relative variation.

Demerits :

1) It is a complicated method to reach a final conclusion.
2) It gives importance to small values.
3) It is applicable in situations where more weightages are given to small values.

2.20 Selection of an Average

The selection of an average should be controlled by the ultimate purpose of investigation. The concrete conclusion and comparison with other series depends on proper choice. It can be justified by considering all the facts.

The selection of an average should be chosen by considering following point :

1) **Objective :** It depends on the nature of the data. Most arithmetic mean are preferred for keeping the importance of each value in a series. Mode is preferred for when items occur frequently. To indicate position of an average, median would be the choice. Geometric mean is considered when a small item is to be preferred and harmonic mean is used to give larger weights to smaller items.
2) **Representative :** An average must be representative of all values contained in a data.
3) **Nature and form of distribution :** The collected dada might be suitably distributed or non-uniform. The choice of an average depends on frequency distribution of data. If there is variation of items in collected data, then mode or median would be the right choice.
4) **Needs further analysis :** The proper selection of an average possesses many mathematical characteristics; hence it is suitable for further analysis.

5) **Sampling stability :** There must be stability in evaluated average values. It avoids sample fluctuations. The repeated samples will show minimum fluctuations in analysis.

6) **Weighting system :** It is applicable for assigning the importance to each value.

7) **Qualitative phenomena :** It is applicable to more stable data. It cannot be suitable where there is lack of stability.

Exercise

1. *What is the need of classification ?*
2. *Give procedure of frequency distribution of discrete variables.*
3. *Following are the scores in the unit test conducted for 50 students in a class.*

8	7	6	9	7	6	5	9	10	2
7	7	5	8	9	9	8	6	6	5
8	8	8	9	9	9	7	7	6	7
10	9	7	6	8	9	9	10	5	4
5	7	8	10	9	8	8	7	7	7

prepare a frequency distribution table.

4. *The data given below relates to the number of T.V. sets sold by a dealer on 25 working days of a certain month. Prepare a frequency distribution of the number of T.V. sets sold.*
2, 4, 3, 0, 2, 1, 5, 3, 2, 0, 3, 4, 5, 1, 1, 4, 3, 2, 5, 4, 2, 2, 1, 3, 0.
5. *Heights in cm of 50 students in a class are given below :*

167.7	168.2	169.1	166.5	161.5	157.3	168.9
170.1	165.8	168.2	158.7	159.6	168.0	162.6
179.0	170.2	169.3	159.2	171.7	163.7	162.3
171.9	172.6	157.7	158.0	165.2	165.8	167.4
170.1	166.7	160.8	161.3	161.5	168.9	166.3
162.6	162.0	166.7	158.0	167.7	170.1	160.8
163.1	161.5	157.5	167.1	168.9	159.6	172.6
164.0						

Classify the above exclusive method of classification. Take the first class interval as 157-160.

6. *Prepare a frequency distribution for each of the following :*

Class mark	4	8	12	16	20
Frequency	24	45	20	10	1

7. *Give procedure of classification of continuous frequency distribution.*

8. *Following is a frequency distribution of heights in cm*

Classes	150-154	155-159	160-164	165-169	170-174
Frequency	2	17	29	21	1

Prepare frequency distribution table.

9. *Prepare frequency distribution for the following*

Marks obtained	0-10	10-20	20-30	30-40	40-50	50-60
No. of students	4	6	20	10	7	3

10. *Discuss the importance of graphic representation of a frequency distribution.*

11. *Explain the following terms : 1) Histogram, 2) Frequency polygon.*

12. *Draw a histogram of the frequency distribution given below :*

Class interval	10-14	15-19	20-29	30-39	40-49	50-74	75-99
Frequency	4	12	20	18	14	25	10

13. *Draw the histogram for the following data :*

Monthly wages (₹)	10000-13000	13000-15000	15000-17000	17000-19000	19000-21000	21000-23000	23000-25000
No. of workers	6	53	85	56	21	16	8

14. *Draw a histogram for the following data -*

Age (in years)	2-5	5-11	11-12	12-14	14-15	15-16
No. of boys	6	6	2	5	1	3

15. *Draw a frequency polygon for the following data*

Marks	0-20	20-40	40-60	60-80	80-100
No. of students	2	18	42	28	5

16. *Draw a frequency polygon for the following data :*

Mid-values	25	35	45	55	65
Frequencies	5	12	33	13	7

17. *Draw a frequency polygon for the following data :*

I. Q.	60-69	70-79	80-89	90-99	100-109	110-119	120-129
Frequency	21	37	51	49	21	13	4

18. *State the advantages and limitations of graphical representation of data.*

19. *What is central the tendency ? What are the requisites for an ideal measure of central tendency ?*

20. *What are the measures of central tendency ? Define each with necessary formulas.*

21. *Given below the distribution of marks obtaining for 140 students.*

Marks obtained	10	20	30	40	50	60	70	80	90	100
No. of students	7	15	18	25	30	20	16	7	2	0

Calculate the mean of the distribution.

22. *Calculate mean of the following distribution :*

C.I.	2.5-7.5	7.5-12.5	12.5-17.5	17.5-22.5	22.5-27.5	27.5-32.5	32.5-37.5	37.5-42.5	42.5-47.5	47.5-52.5	52.5-57.5	57.5-62.5
Frequency	12	28	65	121	175	198	176	120	66	27	9	3

23. *Calculate arithmetic mean of the group of students with weights (in kg) given below : 51, 52, 53, 51, 54, 55, 50, 53, 54, 51.*

24. *If* $n = 10$ *and* $\Sigma(x - 5) = 90$ *find the mean.*

25. *Age distribution of hundred life insurance policy holders as follows :*

Age as on nearest birthday	17-19.5	20-25.5	26-35.5	36-40.5	41-50.5	51-55.5	56-60.5
Number	9	16	12	26	14	12	6

calculate arithmetic mean.

26. *For a certain frequency table which has only been partly reproduced here. The mean was found to be 1.46.*

No. of accidents	0	1	2	3	4	5
Frequency (No. of days)	46	-	-	25	10	5

$N = \Sigma f = 200$

Calculate missing frquecncy

27. *Mean daily salary of 50 employees in a firm is ₹ 88.40. Frequency distribution of salaries of these employees in which some frequencies are missing is given below :*

Salary	40-60	60-80	80-100	100-120	120-140
Frequency	6	-	17	-	5

Find missing frequency.

28. *Give properties of arithmetic mean.*

29. *State and prove any two properties of arithmetic mean.*

30. *Find the weighted arithmetic mean of first n natural numbers with the same numbers and weights.*

31. Scores of students along with weights are given below :

Test	Written	Viva-voce	Group discussion
Scores out of 100	75	60	65
Weights	2	1	2

Find the weighted arithmetic mean of scores.

32. Mean monthly salary of 77 workers in a certain factory is 1560/-. Mean salary of 32 of them is 1500/- and that of the next 25 of the remaining is 1640/-. What is the mean salary of the remaining 20 workers ?

33. Given :

Group 1	Group 2
$n_1 = 100$	$n_2 = 100$
$\Sigma(x - 19) = 68$	$\Sigma(y - 35) = 250$

Find $\overline{X}$, $\overline{Y}$ and combine arithmetic mean of two groups.

34. State the merit and demerits of A. M.

35. Compute the median of the following frequency distribution

x	1	2	3	4	5
y	2	7	15	5	2

36. Compute the median of the following frequency distribution :

Wages (in ₹)	above 30	above 40	above 50	above 60	above 70	above 80	above 90
No. of workers	520	470	399	210	105	45	7

37. Obtain median from the following table :

Class	0-100	100-200	200-300	300-400	400-500	500-600	600-700
Frequency	9	15	18	21	18	14	5

38. In a factory employing 3000 persons, in a day 5 percent work less than 3 hours, 580 work from 3.01 to 4.50 hours, 30 percent work from 4.5 to 6.00 hours. 500 work from 6.01 to 7.50 hours 20 percent work from 7.51 to 9.00 hours and the rest work 9.01 or more hours. What is the median hours of work ?

39. An incomplete frequency distribution is given as follows :

Variable	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Frequency	12	30	-	65	-	25	18

Total frequency = $\Sigma f = N = 229$

Given that the median value is 46. Determine the missing frequencies, using the median formula.

40. Given merits and demerits of median.

41. Daily expenditure of 100 families on transport is given below :

Expenditure	20-29	30-39	40-49	50-59	60-69
No. of families	14	-	27	-	15

If the mode of the distribution is 43.5, find the missing frequencies.

42. Obtain the mode from following frequency distribution.

Marks	0-4	4-8	8-12	12-14	14-18	18-20	20-24	24-28
No. of students	10	12	18	7	5	3	4	6

43. Age distribution of hundred life insurance policy holders is as follows :

Age	17-19	20-22	23-25	26-28	29-31	32-34	35-37	38-40
Number	9	16	12	26	14	12	6	5

Calculate mode.

44. Obtain mode from following frequency distribution :

x	1	2	3	4	5	6	7	8	9	10
f	3	15	45	57	50	36	25	12	9	15

45. What are the merits and demerits of mode.

46. Find the geometric mean 5, 10, 17, 0, 256.

47. Monthly consumption of electricity in units of a certain family in a year is given below : 210, 207, 315, 250, 240, 232, 216, 208, 209, 215, 300, 290. Calculate geometric mean.

48. Define geometric mean and harmonic mean and state the formula for each, in case of individual observations and frequency distributions.

49. A variable takes values $a, ar, ar^2, \ldots, ar^{n-1}$ find G.M.

50. Compute weighted A. M. of index numbers from the following table

Index number	300	200	250	150	250
Weight	62	4	6	12	16

51. Arithmetic mean and geometric mean of two items are 5 and 2.5 respectively. Find the harmonic mean.

52. What are the merits and demerits of G. M.

53. What are the merits and demerits of H. M.

54. Compare the method of G. M. and H. M. on the basis of their merits.

55. Calculate H.M. of the following series : 15, 250, 15.7, 157, 1.57, 105.7, 10.5, 1.06, 25.7 and 0.257.

56. Calculate H.M. of the following series

Values	2	6	10	14	18
Frequency	4	12	20	9	5

57. From the following data, calculate H. M.

Class-interval	10-20	20-30	30-40	40-50	50-60
Frequency	30	75	70	135	220

58. A train runs 25 kms at a speed of 30 km/h another 50 km at a speed of 40 km/h, due to repairs to the track travels for 6 minutes at a speed of 10 km/h and finally covers the remaining distance of 24 kms at a speed of 24 km/h. What is the average speed in km/h ?

Answers :

3

Marks scored	Tally marks	No. of students
1		0
2	|	1
3		0
4	|	1
5	~~||||~~	5
6	~~||||~~ |	6
7	~~||||~~ ~~||||~~ ||	12
8	~~||||~~ ~~||||~~	10
9	~~||||~~ ~~||||~~ |	11
10	||||	4
Total		50

4.

No. of T.V. sets sold	Tally marks	No. of days
0	|||	3
1	||||	4
2	~~||||~~ |	6
3	~~||||~~	5
4	||||	4
5	|||	3
Total		25

5.

Class	Tally marks	Frequency
157-160	~~||||~~ ||||	9
160-163	~~||||~~ ~~||||~~	10
163-166	~~||||~~ |	6
166-169	|||| ~~||||~~ ||||	14
169-172	~~||||~~ |||	8
172-175	||	2
175-178	-	0
178-181	|	1
Total		50

6.

Class Marks	Tally marks	Frequency
4	~~||||~~ ~~||||~~ ~~||||~~ ~~||||~~ ||||	24
8	~~||||~~ ~~||||~~ ~~||||~~ ~~||||~~ ~~||||~~ ~~||||~~ ~~||||~~ ~~||||~~ ~~||||~~	45
12	~~||||~~ ~~||||~~ ~~||||~~ ~~||||~~	20
16	~~||||~~ ~~||||~~	10
20	|	1
Total		100

8.

Classes	Tally marks	Frequency
150-154	||	2
155-159	~~||||~~ ~~||||~~ ~~||||~~ ||	17
160-164	~~||||~~ ~~||||~~ ~~||||~~ ~~||||~~ ~~||||~~ ||||	29
165-169	~~||||~~ ~~||||~~ ~~||||~~ ~~||||~~ |	21
170-174	|	1
Total		70

9.

Marks obtained	Tally marks	No. of students
0-10	||||	4
10-20	~~||||~~ |	6
20-30	~~||||~~ ~~||||~~ ~~||||~~ ~~||||~~	20
30-40	~~||||~~ ~~||||~~	10
40-50	~~||||~~ ||	7
50-60	|||	3
Total		50

21. A. M. = $\bar{X}$ = 46.2143

22. A. M. = $\bar{X}$ = 30.005

23. A. M. = 52.4

24. Mean = 14

25. A. M. = 40.7593

26. $f_1 = 76$, $f_2 = 38$

27. $f_2 = 10$, $f_4 = 12$

31. Weighted A. M. = 68

32. Mean salary of remaining 20 workers = ₹ 1556

33. $\bar{X} = 19.68$, $\bar{Y} = 37.5$, combined A. M. = 28.59

35. Median = 3

36. Median = 47.6170

37. Median = 338.0952

38. Median hours of work = 5.79

39. $f_3 = 33$, $f_5 = 46$

41. $f_2 = 27.77$, $f_4 = 28.23$

42. Mode = 9.41

43. Mode = 27.0769

44. Mode = 4

46. Geometric mean = 0

47. Geometric mean = 238.2609

49. Geometric mean = $ar^{(n-1)/2}$

50. Weighted A. M. = 267
51. Harmonic mean = 1.25
55. H. M. = 1.7374
56. H. M. = 7.2254
57. H. M. = 37.8571
58. Average speed = 31.4139 km/h

Multiple Choice Questions

Q.1 The class interval of the continuous grouped data

10 - 19

20 - 29

30 - 39

40 - 49

50 - 59 is

a 9 b 10

c 14.5 d 4.5

Q.2 Class interval is measured as -

a The sum of the upper and lower limit.

b Half of the sum of lower and upper limit.

c Half of the difference between upper and lower limit.

d The difference between upper and lower limit.

Q.3 A frequency distribution can be -

a Discrete b Continuous

c Both (a) and (b) d None of (a) and (b)

Q.4 Frequency of a variable is always _______.

a in percentage b a fraction

c an integer d none of above

Q.5 Following frequency distribution

x	12	17	24	36	45	48	52
f	2	5	3	8	9	6	1

is classified as ________.

a continuous distribution

b discrete distribution

c cumulative frequency distribution

d none of the above

Q.6 Following frequency distribution

Classes	Frequency
0 - 10	3
10 - 20	5
20 - 30	15
30 - 40	24
40 - 50	9

is of the type

a Discrete series

b Continuous series

c Individual series

d None of the above

Q.7 The data given is 12, 7, 25, 65, 87, 91 will be called as ________.

a continuous series

b discrete series

c individual series

d time series

Q.8 With the help of histogram we can prepare ________.

a frequency polygon

b frequency curve

c frequency distribution

d all of above

Q.9 Histogram can be used only when ________.

a class intervals are equal or unequal

b class intervals are all equal

c class intervals are unequal

d frequencies in class interval are equal

Q.10 Histogram is suitable for the data presented as ________.

a continuous grouped frequency distribution

b discrete grouped frequency distribution

c individual series

d all of above

Q.11 In a histogram with equal class intervals, height of bar are proportional to ________.

a mid-values of the classes
b frequencies of respective classes

c either (a) or (b)
d neither (a) nor (b)

Q.12 With the help of histogram, which of the following can be determine ________.

a median
b deciles

c percentiles
d mode

Q.13 Which of the following is not a measure of central tendency ?

a Mean deviation
b Mean

c Median
d Mode

Q.14 For n observations, harmonic mean is given by ________.

a $\frac{\Sigma 1/x}{n}$
b $\frac{n}{\Sigma 1/x}$

c $\frac{1/\Sigma x}{n}$
d $\frac{\Sigma 1/x}{1/n}$

Q.15 If $x_1, x_2, \ldots, x_n$ is a set of n observations, then harmonic mean of X is the reciprocal of ________.

a given observations and their arithmetic mean

b arithmetic mean of the given observations c arithmetic mean of the reciprocals of the given observations

d reciprocal of mean

Q.16 Formula for geometric mean G is ________.

a $\frac{1}{n}\Sigma \log x_i$

b $\log\left[\frac{1}{n}\Sigma x_i\right]$

c $\text{antilog}\left[\frac{1}{n}\Sigma \log x_i\right]$

d $\text{antilog}\left[\frac{1}{n}\log \Sigma x_i\right]$

Q.17 Arithmetic mean of first n natural numbers is ________.

a $n(n+1)/2$

b $(n^2+1)/2$

c $n(n^2+1)/2$

d $(n^2+n)/2n$

Q.18 Which of the following relation is true

a $\text{Mean} = \frac{1}{2}(3\text{ median} - \text{Mode})$

b $\text{Mean} - 3(\text{Mean} - \text{Median}) = 2\text{ mode}$

c $\text{Median} = \text{Mode} + \frac{2}{3}(\text{Mode} - \text{Mean})$

d $\text{Mode} = 2\text{ median} - 3\text{ mean}$

Q.19 Geometric mean of three numbers 7, 21, 63 is ________.

a 30.3

b $\sqrt{91}$

c $\sqrt{9251}$

d 21

Q.20 If n = 10, $\frac{\Sigma(x-5)}{5}$ = 18 then mean is ________.

a 12

b 14

c 13

d 15

Q.21 Arithmetic mean of 5, 15, 21, 30, 34 is ________.

a 20

b 21

c 18

d 30

Q.22 Arithmetic mean of two numbers is 6.5 and their geometric mean is 6. Then two numbers are ________.

a 9, 6

b 9, 5

c 7, 6

d 4, 9

Q.23 Geometric mean of two numbers $\frac{1}{16}$ and $\frac{4}{25}$ is ________.

a $\frac{1}{10}$

b $\frac{1}{100}$

c 10

d 100

Q.24 Median of the values 11, 7, 6, 9, 12, 15, 19 is ________.

a 9

b 12

c 15

d 11

Q.25 If the two observations are 10 and – 10, then their harmonic mean is ________.

a 10

b 0

c 5

d ∞

Q.26 Mean of the following frequency distribution

x	7	12	16	22	25
f	4	5	8	3	2

is ________.

a 16.40

b 15.09

c 20.80

d none of above

Q.27 Median of the variate values 48, 35, 36, 40, 42, 54, 58, 60 is ________.

a 40

b 41

c 44

d 45

Q.28 Given n = 100, $\Sigma(x - 19) = 68$, then arithmetic mean is ________.

a 19.86

b 19.50

c 19.68

d 19.48

Q.29 Weighted mean gives a higher value than unweighted mean if ________.

[a] all the items have equal weights

[b] larger items have higher weights and small items have lower weights

[c] larger items have lower weights and smaller items have higher weights

[d] none of the above

Q.30 If for values of X, A.M = 25, H.M = 9, then the G.M is ________.

[a] 17 [b] 15

[c] 5.83 [d] 16

Q.31 Given the following frequency distribution of income of employees,

Income ₹/Month	No. of employees
0 - 250	12
250 - 500	20
500 - 750	23
750 - 1000	15
1000 - 1250	10
1250 - 1500	20

Median income of employees is ________.

[a] 625.00 [b] 760.00

[c] 695.65 [d] none of the above

Q.32 Histogram is useful to determine graphically the value of :

[a] Mean [b] Median

[c] Mode [d] All of the above

Q.33 Which of the following is not correct ?

[a] Extreme values affect the median less strongly than they do affect the mean.

[b] A median can be calculated for qualitative descriptions.

c The median can be calculated for every set of data, even for all sets containing open ended classes.

d The median is adaptable for further mathematical manipulations.

Q.34 If we add 15 in each observation of a set, then arithmetic mean is :

a 15 times the original data

b not affected

c increased by 15

d decreased by 15

Q.35 The class intervals of the grouped data :

5 - 9	10 - 14	15 - 19	20 - 24

is of the type ________.

a inclusive class

b discrete class

c exclusive class

d none of above

Q.36 If 0.3, 0.5, 0.8, 0.7 and 1.5 are the respective weights of the values 10, 15, 20, 25, 30, then the weighted mean is ________.

a 20.0

b 23.42

c 16.58

d none of above

Q.37 When all the observations are same, then the relation between A.M, G.M and H.M is ________.

a A.M = G.M = H.M

b A.M < G.M < H.M

c A.M < G.M = H.M

d A.M > G.M > H.M

Answer Keys for Multiple Choice Questions :

Q.1	a	Q.2	d	Q.3	c	Q.4	c	Q.5	b
Q.6	b	Q.7	c	Q.8	d	Q.9	b	Q.10	a
Q.11	b	Q.12	d	Q.13	a	Q.14	b	Q.15	c
Q.16	b	Q.17	a	Q.18	d	Q.19	b	Q.20	b
Q.21	b	Q.22	d	Q.23	a	Q.24	d	Q.25	d
Q.26	b	Q.27	d	Q.28	c	Q.29	b	Q.30	b
Q.31	c	Q.32	c	Q.33	d	Q.34	b	Q.35	a
Q.36	b	Q.37	a						

SOLVED MODEL QUESTION PAPER (In Sem)

Statistics

S.E. (AI and DS) Semester - IV (As Per 2020 Pattern)

Time : 1 Hour] [Maximum Marks : 30

N. B. :

i) Attempt Q.1 or Q.2, Q.3 or Q.4.

ii) Neat diagrams must be drawn wherever necessary.

ii) Figures to the right side indicate full marks.

iv) Assume suitable data, if necessary.

Q.1 a) *What is statistics ? Give importance and limitations of statistics.* **(Refer sections 1.1, 1.3 and 1.4)** [5]

b) *Define sampling. Explain random sampling.* **(Refer sections 1.7 and 1.7.2)** [4]

c) *What is population and sample ? Give difference between them.* **(Refer section 1.7)** [4]

OR

Q.2 a) *What are the methods of estimation ? Give brief on testing of hypothesis.* **(Refer sections 1.10.1 and 1.10.2)** [5]

b) *Give advantages and disadvantages of statistical analysis.* **(Refer sections 1.3 and 1.4)** [4]

c) *Explain the scope of statistics in engineering and technology.* **(Refer section 1.2.2)** [4]

Q.3 a) *What is histogram ? Draw the histogram for the following data -* [5]

Age (in years)	2 - 5	5 - 11	11 - 12	12 - 14	14 - 15	15 - 16
No. of boys	6	6	2	5	1	3

Ans. :

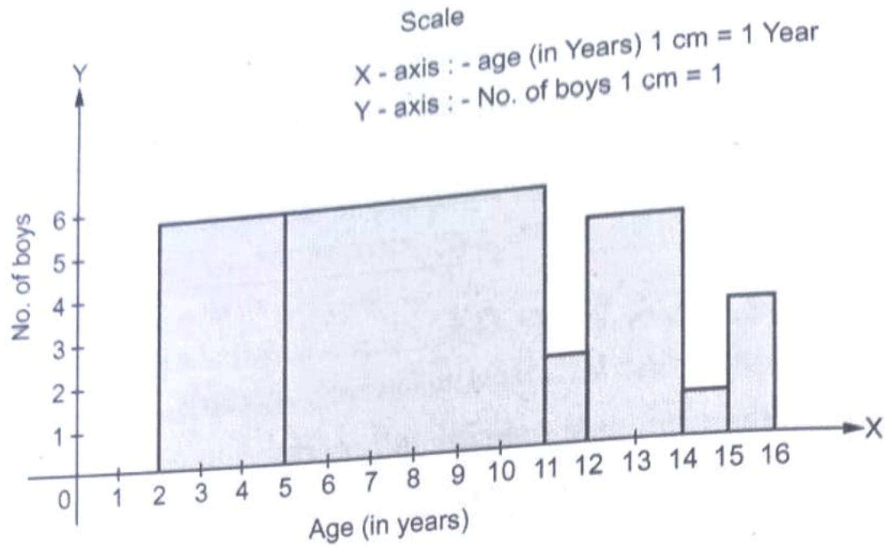

Fig. 1

b) *State merits and demerits of arithmetic mean (two each).* **(Refer section 2.10)** [4]

c) *Obtain the median from following table.* [4]

Class	*0 - 100*	*100 - 200*	*200 - 300*	*300 - 400*	*400 - 500*	*500 - 600*	*600 - 700*
Frequency	*9*	*15*	*18*	*21*	*18*	*14*	*5*

Ans. :

C.I.	f	Less than c.f.
0-100	9	9
100-200	15	24
200-300	18	42
300-400	21	63
400-500	18	81
500-600	14	95
600-700	5	100

N = 100

$\frac{N}{2} = 50$

Median class

$l = 300$, h = 100, f = 21, c.f. = 42

$$\text{Median} = l + \left(\frac{\frac{N}{2} - \text{c.f.}}{f}\right) \times h$$

$$= 300 + \left(\frac{50 - 42}{21}\right) \times 100 = 300 + 0.3809 \times 100$$

$$= 338.0952$$

d) *Define Geometric mean and harmonic mean. Compare them on the basis of merits.* **(Refer sections 2.17 and 2.19)** **[4]**

OR

Q.4 a) *Daily expenditure of 100 families on transport is given below :* **[5]**

Expenditure	20 - 29	30 - 39	40 - 49	50 - 59	60 - 69
No. of families	14	?	27	?	15

If mode of the distribution is 43.5 find the missing frequencies.

Ans. :

c.f.	f
20 - 29.5	14
29.5 - 39.5	f_2
39.5 - 49.5	27
49.5 - 59.5	f_3
59.5 - 69.5	15

$$\text{Mode} = 43.5$$

Here maximum frequency is 27

Thus class 39.5 - 49.5 is modal class.

where $l = 39.5,\ f_1 = 27,\ f_0 = f_2,\ f_2 = f_3,\ h = 10$

$$\text{Mode} = 39.5 + 10\left(\frac{27 - f_2}{54 - f_2 - f_3}\right)$$

$$43.5 = 39.5 + \frac{270 - 10f_2}{54 - f_2 - f_3}$$

$$43.5 = \frac{2133 - 39.5f_2 - 39.5f_3 + 270 - 10f_2}{54 - f_2 - f_3}$$

$$2349 - 43.5f_3 - 43.5f_2 = 2403 - 49.5f_2 - 39.5f_3$$

$$49.5f_2 + 39.5f_3 - 43.5f_3 - 43.5f_2 = 54$$

$$6f_2 - 4f_3 = 54 \qquad \text{... (1)}$$

As $N = \sum f = 100$

$$14 + f_2 + 27 + f_3 + 15 = 100$$

$$f_2 + f_3 = 100 - 56$$

$$f_2 + f_3 = 44 \qquad \text{... (2)}$$

Solving equation (1) and (2),

$\boxed{f_2 = 23}$ $\boxed{f_1 = 21}$

b) *Give merits and demerits of median.* **(Refer section 2.13)** **[4]**

c) *Calculate harmonic mean of the following series* **[4]**

Values	2	6	10	14	18
Frequency	4	12	20	9	5

Ans. :

w_i	x_i
2	4
6	12
10	20
14	9
18	5

$$\sum w_i = 50$$

$$\sum \frac{w_i}{x_i} = \frac{2}{4} + \frac{6}{12} + \frac{10}{20} + \frac{14}{9} + \frac{18}{5}$$

$$= \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + 1.5556 + 3.6$$

$$= 1.5 + 5.1556 = 6.6556$$

$$H = \frac{\sum w_i}{\sum\left(\frac{w_i}{x_i}\right)} = \frac{50}{6.6556} = 7.5124$$

d) *State the advantages and limitation of graphical representation of data.* **(Refer section 2.5)** **[4]**