

TUTOR MARKED ASSIGNMENT

Course Code	:	ECO - 07
Course Title	:	Elements of Statistics
Assignment Code	:	ECO – 07/TMA/2016-17
Coverage	:	All Blocks

Maximum Marks: 100

Attempt all the questions.

1. (a) Discuss the different types of enquiries that are used while organizing a statistical survey.

(b) Describe the factors affecting the type of enquiry.

(15+5)

2. Describe in detail various types of diagrams and graphics presentation of data.

(20)

3. (a) Calculate arithmetic mean from the data given below:

Marks	5	10	15	20	25	30	35	40
No. of students	5	7	9	10	8	6	3	2

(8)

- (b) Find out the median by drawing ogive and verify with the calculated result from the following distribution.

Class	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Frequency	5	10	14	20	35	15	1

(12)

4. What is the meaning of Standard Deviation? What are the major steps involved in computation of Standard Deviation. Explain with the help of hypothetical data. Also discuss its merits and limitations

(12+8)

5. Write short notes on the following:

- (a) Primary data and secondary data
- (b) Sampling and non sampling errors
- (c) Census and sampling
- (d) Schedule and questionnaire

(4×5)

ASSIGNMENT SOLUTIONS GUIDE (2016-2017)

E.C.O.-7

Elements of Statistics

Disclaimer/Special Note: These are just the sample of the Answers/Solutions to some of the Questions given in the Assignments. These Sample Answers/Solutions are prepared by Private Teachers/Tutors/Authors for the help and guidance of the student to get an idea of how he/she can answer the Questions given in the Assignments. We do not claim 100% accuracy of these sample answers as these are based on the knowledge and capability of Private Teacher/Tutor. Sample answers may be seen as the Guide/Help for the reference to prepare the answers of the Questions given in the assignments. As these solutions and answers are prepared by the private teacher/tutor so the chances of error or mistake cannot be denied. Any Omission or Error is highly regretted though every care has been taken while preparing these Sample Answers/Solutions. Please consult your own Teacher/Tutor before you prepare a Particular Answer and for up-to-date and exact information, data and solution. Student should must read and refer the official study material provided by the university.

Attempt all the questions.

Q. 1. (a) Discuss the different types of enquiries that are used while organizing a statistical survey.

Ans. Different types of Enquiries

1. Census or Sample Enquiry: Census enquiry means a complete enumeration, since the enquiry is done to all items constituting the population. Census enquiry involves lots of efforts in terms of time, money and energy. Still there are more chances of occurrence of error called “Bias”, as it is impossible to cover all items with perfect accuracy. The error becomes larger and larger with the increasing number of observation. Therefore, organising census enquiry on large scale becomes difficult because of the resources involved.

In sample enquiry, only a part of population is surveyed, which is assumed to represent the whole population. It is conducted by specifically trained and experienced investigators therefore sample method need better organisation than the census method. Money expenditure and effort is much less in sample enquiry than comparative to census enquiry. There is less chance of occurrence of error as the information is collected from only a small proportion of population. Generally, this method is preferable as it produce valid and reliable results.

2. Original or Repetitive Enquiry: When the enquiry is done for the first time, it is called Original enquiry. There is a choice in employing the statistical method for the collection of data with reference to the type of enquiry in case of original one.

When the enquiry is a repeated continuation of the original enquiry it is termed as repetitive enquiry. There is no choice in the statistical methods as it is based on the initial methods and the chances of wrong conclusions are more as the various terms are not defined according to the new pattern.

3. Confidential or Open Enquiry: As the name itself suggest that in a confidential enquiry, the details of the results of the survey conducted are not available in public domain. While in the case of open enquiry, the results are available to everyone.

4. Direct or Indirect Enquiry: When the data can be expressed quantitatively such as marks, weight, length, etc., it is called Direct enquiry.

In case of indirect enquiry, the qualitative data such as beauty, intelligence etc., is converted into quantitative form for the measurements.

Regular or Ad-hoc Enquiry

When the statistical data is collected regularly at known interval of time, it is called Regular enquiry.

When the collection of data depends on the need i.e. there is no dependence on time, it is referred as ad-hoc enquiry.

5. Non-official or Semi-official or Official Enquiry: In official type of enquiry, the urgent information is required and people are bound to do it. This type of Survey is conducted by government body.

Semi-official type of survey is conducted by semi-government bodies like railways and the information regarding statistical data is collected through some efforts.

Non-official enquiry is conducted by private bodies or individuals. In this type of enquiry, the investigator comes across a lot of difficulty in collecting data.

(b) Describe the factors affecting the type of enquiry.

Ans. Factors Affecting the type of Enquiry: The type of enquiry to be selected depends on the following factors:

1. Objective and scope of the survey: In planning a statistical enquiry, the objective and scope of the enquiry is decided with reference to space, time and population of items covered.

2. Who conducts the survey: The type of enquiry depends upon whether the survey is conducted by some individual or some organisation or by the state.

3. Financial implications: Availability of funds plays an important role in deciding the type of enquiry. It is obvious that large scale survey requires more money than the small scale survey.

4. Sources of data: Sources of data may be primary or secondary depending upon the type of enquiry. For example the data collected during census operation are primary to the census department of the Government of India, but secondary to an investigator, who uses it for further research.

Q. 2. Describe the factors in detail various types of diagrams and graphics presentations of data.

Ans. The main objective of statistical methods is to reduce the size of statistical data to render them easily accessible. To attain this object, the statistical method such as classification, tabulation, graphical presentation, diagrammatic representation etc., are employed. We have studied the first two methods i.e., the classification and the tabulation. Now, we will discuss about diagrams. In diagrams, data is presented visually to bring out the salient features of data. These create a lasting impression of relative magnitude and numerical values on the mind of an observer so that it can be easily remembered. Diagrammatic presentation is an aid to clarify the conclusions that have been obtained by a process of tabulation. Thus classification and tabulation precede diagrammatic representation.

Diagrammatic process can accommodate and explain complex issues upto a reasonable standard of accuracy.

Types Of Diagrams

Broadly, the diagrams are classified into:

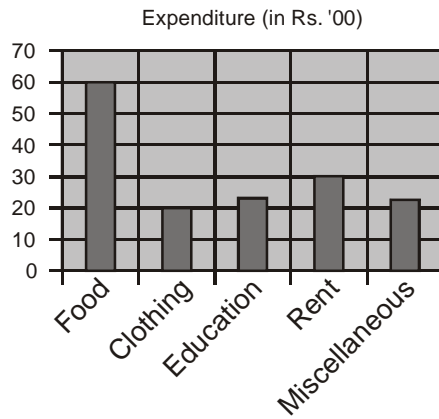
1. One dimensional diagram
2. Two dimensional diagrams
3. Three dimensional diagrams
4. Pictograms
5. Cartograms

One Dimensional Diagrams

As the name itself suggests, the one dimensional diagrams are prepared on the basis of only one dimension, which is length. They are also known as bar diagrams. These diagrams are quiet easy to understand, simple to construct and useful in comparison. These diagrams are further classified into simple, multiple, sub-divided and percentage sub-divided bar diagrams.

1. Simple Bar Diagrams: The term 'bar' is used for thick wide lines. The bars may be vertical as well as horizontal. Usually the various bars, whose length is in ratio of the magnitude of the data given, are drawn according to some suitable scale at equal intervals. Generally, width of all bars is same and both positive and negative values can be shown with the help of simple bar diagram:

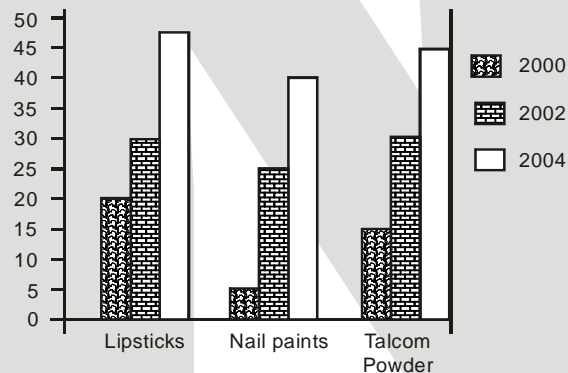
Items	Expenditure (in Rs'00)
Food	60
Clothing	20
Education	25
Rent	30
Miscellaneous	25



■ Expenditure (in Rs. '00)

2. Multiple Bar Diagrams: The following data relates to sales (in lakh number of padas) of cosmetic products during 2000, 2002 and 2004.

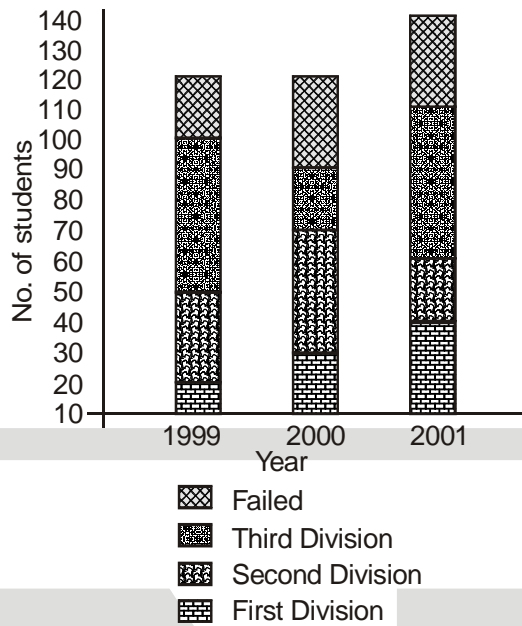
Cosmetics	Year		
	2000	2002	2004
Lipsticks	20.5	30.0	48.0
Nail paints	5.0	25.7	37.5
Talcom Powder	15.0	28.3	41.5



For representing different variables, multiple bar diagrams are used by constructing the bars adjoining to each other. The method for the construction of multiple bar is quiet similar to the simple bar, the only difference is that it shows the relationship among the two or more components of the same variable. It facilitates the comparison between the different values of same variables or different variables. The bars are shaded differently or may be shown in different colours.

Sub-divided Bar Diagrams: The following data depicts the result of MBA of students of a college for three years.

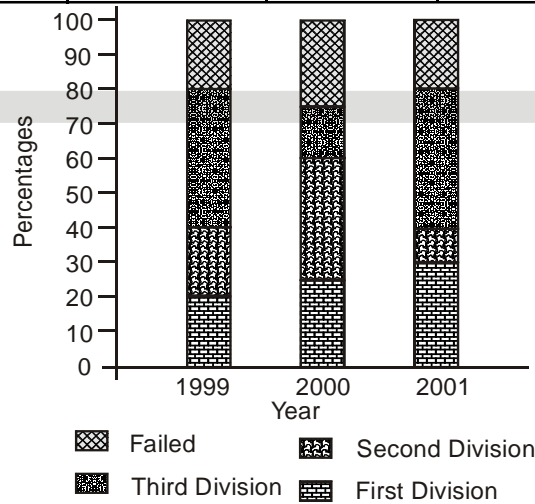
Year	First Division	Second Division	Third Division	Failed
1999	20	30	50	20
2000	30	40	20	30
2001	40	20	50	20



Under this method, the bar is used to exhibit the divisions as a whole into its components part. It is also known as component diagram. Simple bars are sub-divided in the ratio of the different components of the variable.

J4. Percentage Sub-divided Bar Diagrams: By expressing the last data (sub-divided bar) into percentage sub-divided Bar diagram, we have

Number age	1999		2000		2001	
	%	Number	% age	Number	% age	Number
First Division	20	16.7	30	25.0	40	28.6
Second Division	30	25.0	40	33.3	20	14.3
Third Division	50	41.7	20	16.7	50	35.7
Failed	20	16.7	30	25.0	30	21.4
Total	120	100.1	120	100	140	100



In the percentage sub-divided bar diagram, the distribution of given data into its parts can be done on the percentage basis. The single bar is sub-divided into two or more components used and the length of the bar is assumed as 100. The length of different bars is kept to find the relative change in data.

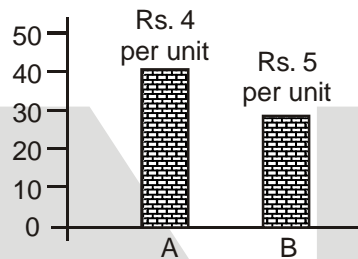
Two Dimensional Diagrams

The two dimensional diagrams are prepared on the basis of two dimensions i.e., length and breadth. Since the area is determined by the product of these two dimensions, they are also known as area diagrams. The two dimensional diagrams can be classified into rectangles, sub-divided rectangles, squares, circles and pie-diagrams.

1. Rectangles: A rectangle is a two-dimensional diagram having length and width. In statistical methods, it is used to find out the area by representing the variables as length and width. They are employed in those cases, when it is desired to give more detailed information than can be conveyed by using bar.

Following data is related to the cost of two commodities:

Price per unit	Rs. 4	Rs. 5
Quantity Sold	40	30



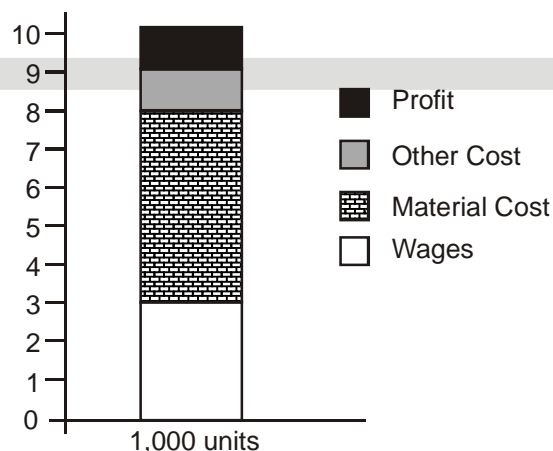
2. Sub- divided Rectangles: Rectangles can also be sub-divided to show the area with respect to different components. Under this method, the rectangle is subdivided on the basis of cumulative values of data.

Consider the following data

<i>Wages (Rs.)</i>	<i>Material (Rs.)</i>	<i>Other (Rs.)</i>	<i>Profit (Rs.)</i>	<i>No. of units (Rs.) produced</i>
3,000	5,000	1,000	1,000	1,000

Computation for sub-divided Rectangle

<i>Cost per unit (Rs.)</i>	<i>Cumulative</i>	<i>Cost per unit</i>		
Wage	3,000	3	3	Material
Cost				
Other	1,000	1	9	
Cost				
Profit	1,000	1	10	
Total	10,000	10		



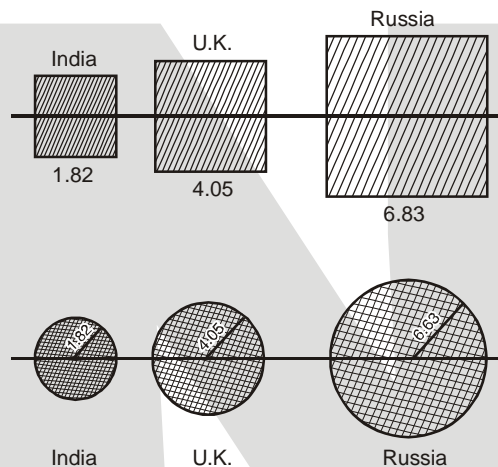
3. Squares and Circles: Square is also a two-dimensional figure having all the sides of same length. Squares are used, when we compare the quantities that differ widely in magnitude. The side of the square varies with the square root of its area. While constructing more than one square in a single diagram, the base line remains the same.

We know that the area of the circles is proportional to the squares of their radii. Therefore, to find out the ratio of various circles, we calculate the square root of the given in the questions. For the construction of more than one circle in a single diagram, the centers of the circles should lie on the same line.

Following data is related to coal production of three countries:

<i>Country</i>	<i>Production in crores of tons</i>	<i>Square roots</i>
Russia	44.0	6.63
U.K.	16.40	4.05
India	3.3	1.82

Square Root consider as the side of square or radius of the circle.



4. Pie Diagrams: A pie diagram consists of a circle divided into as many sectors to indicate the various components of a given variable. The area of each sector is in proportion to the angle, which is subtends at the centre of the circle and with the relative frequency of the class.

The sum of the sectors is taken equivalent to 360° . So, in order to get the angles of the desired sectors, we divide 360° in the proportion of the various relative frequencies.

Following data is related to result of different class in a school:

A \Rightarrow Number of students who passed in VIth standard = 315

B \Rightarrow Number of students who passed in VIIth standard = 180

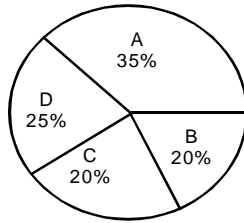
C \Rightarrow Number of students who passed in VIIIth standard = 180

D \Rightarrow Number of students who passed in IXth standard = 225

Total number of students who are passed = 900

	<i>Number os students who passed</i>	<i>% age</i>
VIth	315	35%
VIIth	180	20%
VIIIth	180	20%
IXth	225	25%

Representing the data in the pie chart, are have



Graphic presentation is also a technique of visual presentation of data other than diagrammatic presentation. In this method, the use of graph paper is compulsory for representing the data. Usually, the graph is represented by rectangular co-ordinates in two dimensions. Graphs are more reliable for presenting time series and frequency distribution. Presenting data in a frequency distribution makes inspection of the data set much more manageable than presenting the entire set of raw data. A frequency distribution can be considered a type of descriptive statistic. Graphic presentation makes comparison of data easier and it is also helpful in determining the trends of past performance.

Types of Historigrams

There are two methods of presenting historigrams:

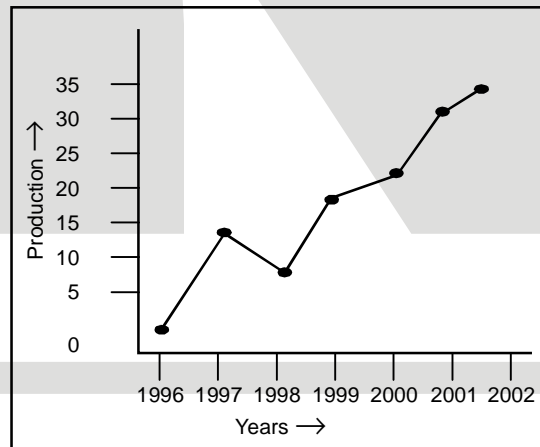
1. By means of natural graph method, in which the graph reflects in absolute values over a time period.
2. By means of ratio scale method, in which the graph reflects the relative changes over a time period.

Historigrams can be categorised into one dependent variable historigram, more than one dependent variable historigram, mixed graph and range graph.

1. One Dependent Variable Historigram: This is the most common type of graph, in which dependent variable is taken on the y-axis and the time period is taken on the x-axis. The points are plotted in accordance to the given data, and then the points can be joined together through the straight line.

One Dependent Variable Graph

Following graph represent the one variable:



Production of a company (in lakh unit) over the given years.

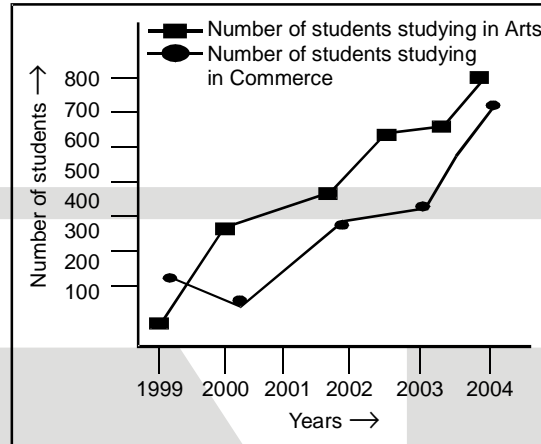
False Base Line

As discussed earlier, when there is a large difference between the zero and the lowest value of the given data to be taken on the y-axis, then the false base line is taken instead of showing the entire scale from zero to the highest value of the given data by breaking the line of the y-axis near the x-axis by drawing the line in a zigzag manner. It helps in presenting the data more significantly.

2. More than one Dependent Variable Histogram: In this type of graph, we present two or more variables exactly in the same manner as for one variable by making separate curve to show each variable in the same graph. The units for all variables remain same in such type of graphs.

Following graph represent the two variables in the same graph:

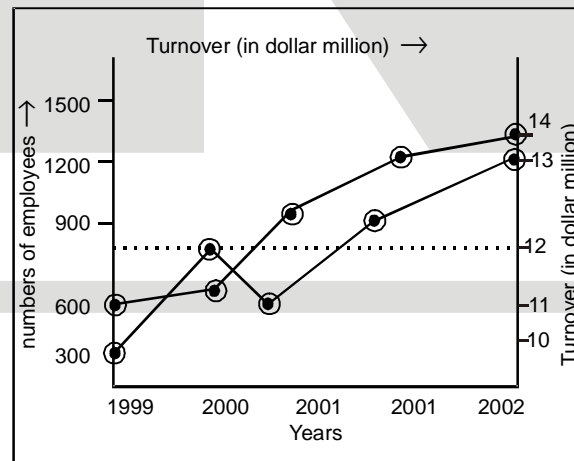
Number of Students studying Arts and Commerce in a College over the year.



3. Mixed Graph: When the units of two dependent variables are different, then the values of these variables are presented by using different scales. Such type of graph is known as mixed graph. One scale is taken on the usual y-axis and the other scale is taken on the right of the x-axis. But it is important to remember that the mid points of the two scales must lie on the same line. If required, false base line can be taken.

The following data relates to number of employer and turnover of a company:

Years:	1999	2000	2001	2002	2003
Number of employees:	500	750	1100	1250	1400
Turnover: (In dollar million)	10	12	11.2	12.5	13.5



4. Range Graph: When the maximum and minimum values are represented on the graph, then it is called the range graph as range is defined as the difference between the maximum and the minimum value. The two curves one for the maximum values and one for the minimum values are drawn on the same graph in the same manner as the two dependent variables are drawn.

Graphs Of Frequency Distribution

Graphs of frequency distribution helps in locating the positional measures of central tendency such as median, quartiles and mode therefore play a significant role in identifying characteristics and relationships.

Principles of Constructing Graphs of Frequency Distribution

1. The values of the variable are taken on x -axis.
2. The values obtained due to the frequency of the class size are presented on the y -axis.
3. The scale of x -axis not necessarily starts with zero but for the y -axis it is essential to start with zero i.e., no false base line can take in frequency graphs.
4. The scale for the two axes must be specified clearly.
5. The title for the graph must be concise and self explanatory.

Types of Frequency Distribution Graphs

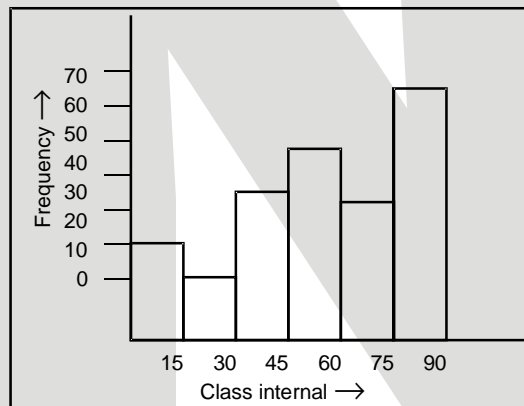
The frequency distribution graphs are categorised into histogram, frequency polygon, frequency curve and cumulative frequency graph, which are discussed below:

1. Histogram: Histograms are the graph chart, which are formed by using frequency of data. It is a series of vertical bars called rectangles, which are proportional in area to the frequency of the respective class intervals and the width of the each bar represents the corresponding class interval of the variable.

When the class intervals are of equal width

In this case, each class interval will have a width of one unit and frequency and frequency density will remain the same. Frequency density is determined by dividing the frequency by the width unit.

Following diagram present the histogram when the class intervals are equal:



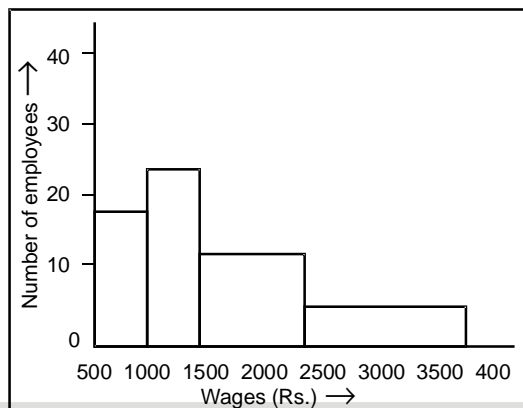
When the class intervals are of unequal width

In this case, the frequencies of the class intervals and frequency density will differ. The smallest class interval is taken as one unit width. Frequency density is determined as follows:

$$\text{Frequency density} = \frac{\text{Width of the smallest class}}{\text{Width of the given class}} \times \text{frequency of the given class}$$

Following data relates to the wages (in Rs.) given to the employees of a firm:

Wages (Rs.):	0-500	500-1000	1000-2000	2000-4000
Number of employees:	25	35	20	20



For class 1000-2000,

$$\text{Frequency density} = \frac{500}{1000} \times 20 = 10$$

for class 2000-4000, frequency density

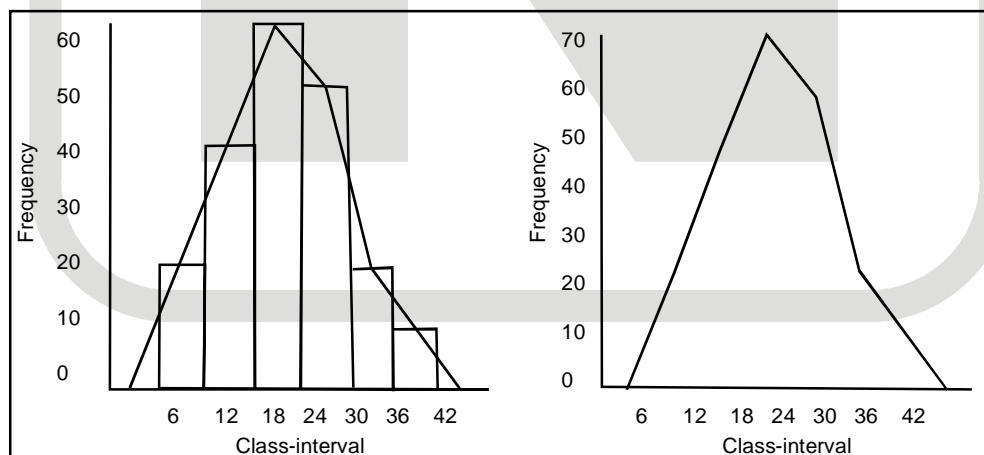
$$= \frac{500}{2000} \times 20 = 5.$$

NOTE: There should be no gap between two successive rectangles and area of the rectangle represents the frequency.

2. Frequency Polygon: Frequency Polygon is based on the assumption that the frequencies in a class interval are evenly distributed throughout the class and mid points are the representatives. Frequency polygon is obtained by using two methods:

- (i) By joining the middle points of the top of the adjacent rectangular bars of histogram with a straight line.
- (ii) By plotting the frequency densities against the mid-point of each class and by joining them with a straight line. This is called direct method.

These straight lines are extended at both the ends to join the x -axis thus indicating zero frequency.



NOTE: The area of the polygon is equal to the area of the histogram and represents the total frequency.

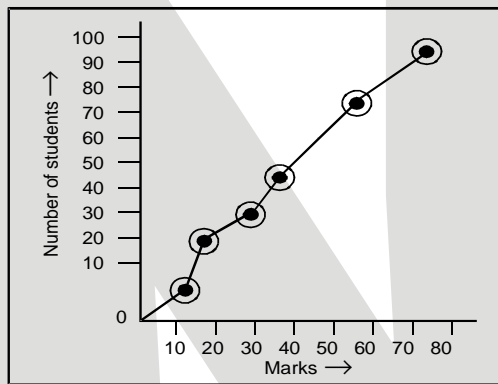
3. Frequency Curve: A frequency curve is prepared by free hand to smoothen the frequency polygon in such a way that the area under polygon and the area under curve drawn is the same.

4. Ogive or Cumulative Frequency Graph: As the name itself suggests that an ogive and a cumulative frequency graph depend on the cumulative frequency. Cumulative frequency graphs can be classified into “less than ogive” and “more than ogive”.

In “**less than ogive**”, the cumulative frequencies are in ascending order i.e. frequency obtained by successive addition of frequency given in the data. In this case, the frequencies of classes are plotted against its upper limits. The following graph present the marks obtained by students of class X in a school.

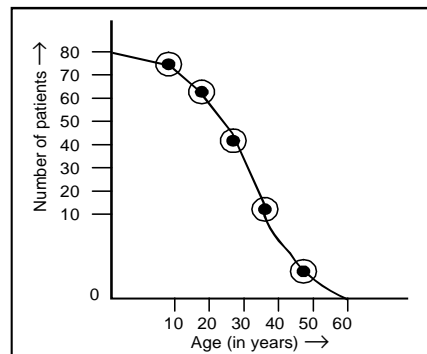
Marks	Number of Students
Less than 10	7
Less than 20	21
Less than 30	34
Less than 40	46
Less than 50	66
Less than 60	77
Less than 70	92
Less than 80	100

In “**more than ogive**”, the cumulative frequencies are in descending order. In this case, the frequencies of classes are plotted against its lower limits.



The following graph represents the age of patients admitted in a Hospital.

Age (in years)	Number of patients
More than 0	80
More than 10	74
More than 20	63
More than 30	42
More than 40	19
More than 50	5



NOTE : Ogives help us to represent positional averages such as median and other partition values such as quartiles, percentiles etc. Median is the point, where the less than ogive and more than ogive intersect.

Q. 3. (a) Calculate arithmetic mean from the data given below:

Marks	5	10	15	20	25	30	35	40
No. of students	5	7	9	10	8	6	3	2

Ans. Calculation of Arithmetic mean

Marks	No. of Students f_i	$u_i = \frac{x_i - A}{5}$	$F_i u_i$
5	5	-4	-20
10	7	-3	-21
15	9	-2	-18
20	10	-1	-10
-25	8	0	0
30	6	1	6
35	3	2	6
40	2	3	6
	$\Sigma f_i = 50$		$\Sigma F_i u_i = -51$

$$\begin{aligned} \bar{X} &= A + \frac{\Sigma f_i u_i}{\Sigma f_i} \cdot h \\ &= 25 + \frac{-51}{50} \cdot 5 \\ &= 25 - 5.1 \\ &= 19.9 \text{ Ans.} \end{aligned}$$

(b) Find out the median by drawing ogive and verify with the calculated result from the following distribution.

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No. of students	5	10	14	20	35	15	1

Class	F	C.f
0-10	5	5
10-20	10	15
20-30	14	29
30-40	20	49
40-50	35	84
50-60	15	99
60-70	100	100

$$\frac{n}{2} = \frac{100}{2} = 50$$

$$\text{Median} = L + \frac{\frac{n}{2} - c.f_0}{f} \cdot h$$

Where = Lower of the Median class

$C.F$ = Cumulative Frequency preceding median class

h = Class Size

f = Frequency of the median class

$$\text{median} = 40 + \frac{50 - 49}{35} \cdot 10$$

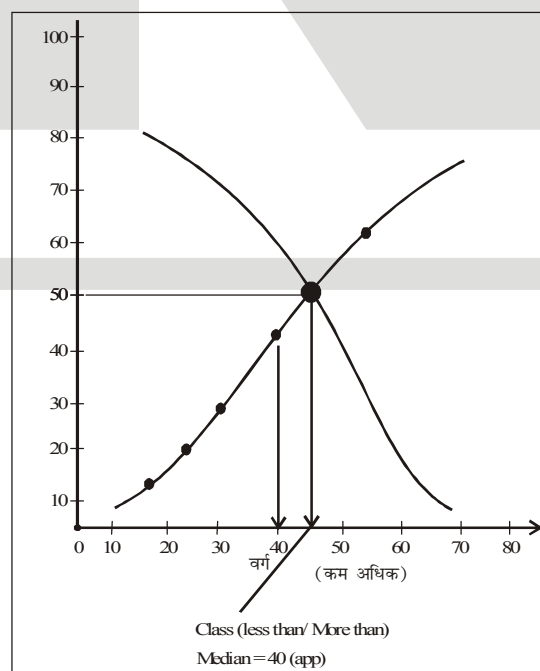
$$= 40 + \frac{1}{35} \cdot 10$$

$$= 40 + 0.29$$

$$= 40.29 \text{ (app)}$$

II. Drawing of ogive

Class	Cumulative frequent	Class	Cumaltive fegrement
Less than 10	5	more than 0	100
Less than 20	15	more than 10	95
Less than 30	29	more than 20	85
Less than 40	49	more than 30	71
Less than 50	84	more than 40	51
Less than 60	99	more than 50	16
Less than 70	100	more than 60	1



Q. 4. What is the meaning of Standard Deviation? What are the major steps involved in computation of Standard Deviation. Explain with the help of hypothetical data. Also discuss its merits and limitations.

Ans. Standard deviation is the square root of arithmetic average of the squares of deviations measured from the mean. Therefore, it is also known as Root Mean Square Deviation. Like the mean deviations standard deviations can be calculated by subtracting mean, median and mode. Mathematically, the standard deviation about arithmetic mean can be expressed as:

$$\text{Standard Deviation } (\sigma) = \sqrt{\frac{1}{n} \sum (X - \bar{X})^2}$$

Where, \bar{X} is the arithmetic mean and n is the total number of observations.

Computation

Standard deviation is calculated by direct and short cut methods, which are explained below:

1. Direct Method: For ungrouped data, the formula is given by:

$$\text{Standard Deviation } (\sigma) = \sqrt{\frac{1}{n} \sum (X - \bar{X})^2}$$

For grouped data, the formula is given by:

$$\text{Standard Deviation } (\sigma) = \sqrt{\frac{1}{n} \sum f(X - \bar{X})^2}$$

If the total number of observations are not very large, then formulas for both types of data can be modified as:

For ungrouped data, the formula is given by:

$$\text{Standard Deviation } (\sigma) = \sqrt{\frac{\sum X^2}{n} - \bar{X}^2} \quad \left\{ \bar{X} = \frac{\sum X}{n} \right\}$$

For grouped data, the formula is given by

$$\text{Standard Deviation } (\sigma) = \sqrt{\frac{\sum fX^2}{n} - \bar{X}^2} \quad \left\{ \bar{X} = \frac{\sum fx}{n} \right\}$$

2. Short-cut Method: For making calculation easy for large number of data, we adopt the step deviation method by taking assumed mean as follows:

For ungrouped data

$$\text{Standard Deviation } (\sigma) = \sqrt{\frac{\sum d^2}{n} - \frac{(\sum d)^2}{n}}$$

where: $d = X - A$

For grouped data

$$\text{Standard Deviation } (\sigma) = \sqrt{\frac{\sum fd^2}{n} - \frac{(\sum fd)^2}{n^2}}$$

where: $d = X - A$

Or

$$\text{Standard Deviation } (\sigma) = C \sqrt{\frac{\sum fd'^2}{n} - \frac{(\sum fd')^2}{n^2}}$$

Where $d = \frac{X - A}{C}$ and $C =$ common factor.

Properties

The important properties of standard deviations are as follows:

1. On adding or subtracting the constant value to each observation, the value of standard deviation remains the same.

- By multiplying or dividing the constant value to each observation, the value of standard deviation will also be affected similarly.
- For normal data, the standard deviation is approximately $\frac{5}{4}$ of mean deviation $\frac{3}{2}$ of quartile deviation.
- The standard deviation for the combined group is given by:

$$\sigma_{12} = \sqrt{\frac{(n_1 d_1^2 + n_2 d_2^2) + n_1 d_1^2 + n_2 d_2^2}{n_1 + n_2}}$$

Where n_1 and n_2 are observations with mean \bar{X}_1 and \bar{X}_2 and standard deviation σ_1 and σ_2 respectively.

$$d_1 = \bar{X}_1 - \bar{X}$$

$$d_2 = \bar{X}_2 - \bar{X}$$

\bar{X}_{12} = Combined arithmetic mean of the two groups.

- Root mean square deviation about a value other than arithmetic mean is greater than the standard deviation.
- In the interval extending to a distance on either side of the mean of a normal distribution 68.27% of the items are included; in the interval the mean $\pm 2\sigma$, 95.45% of the case will lie; and within the interval $\pm 3\sigma$, 99.73% or nearly all of the items are included.

Merits and Limitations

The merits and limitations of standard deviation are as follows:

Merits:

- Standard deviation is rigidly defined and provides most accurate results.
- It is based on all the observations of data.
- It is amenable to algebraic treatment and possesses many mathematical properties.
- It is least affected by the fluctuation of sampling.
- It is also known as the standard measure because of its quality to cover 68.27% values on either side of the mean in normal distribution.

Limitations:

- Standard deviation is not easy to calculate nor is it easily understood.
- If two or more series of data are given in different units, then standard deviation can not be used for comparing the two data unless its coefficient is not defined.
- It can not be applied in cases of open and distribution.

Q. 5. Write short notes on the following:

(a) Primary data and secondary data

Ans. Difference between Primary and Secondary Data:

- When the data pertaining to an investigation is originated by the investigator himself, then it is called primary data. But the data, which the investigator obtains from someone else's records, is termed as secondary data.
- Primary data are the raw materials to which statistical methods are applied for its analysis and secondary data are the finished products since they are already used by someone in some other form.
- To collect primary data is comparatively more expensive and more time consuming than the collection of secondary data.
- Primary data are more accurate in supplying information in comparison to the secondary data as primary data are collected as per requirements of the investigation but the secondary data should be utilised with due precautions.
- Primary data may be influenced by personal prejudice of the investigator, but this is not so in the case of secondary data.

The various factors that affect the choice for data between primary and secondary to be used in an investigation are:

1. Object of the Enquiry: The objective of the enquiry is the most important factor that determines the type of information required for the survey. Therefore, the purpose or object of the enquiry should be specific and well defined.

2. Scope of the Enquiry: The scope of the enquiry is decided with reference to the coverage of the survey. For example, the data collected during census operations are primary but to a person who makes use of these data for further research will said to be secondary.

3. Financial Resources: The availability of finance resources describes the type of data used because the collection of primary data requires more money than the secondary data.

4. Time Factor: Time factor also plays an important role in making a choice of data as the work of collection of data must be finished in reasonable period of time. Collection of primary data consumes relatively more time than the collection of secondary data.

5. Status of the Investigating Agency: If the investigating agency is a government or semi-government organisation, then we can collect primary data on a very large scale but for private (organisation or individual) body it is economical to use secondary data.

6. Human Resources: Large number of trained and efficient peoples is required for conducting the survey on a very large scale. Therefore, for collecting primary data we require more persons.

7. Availability of Secondary Data: Secondary data depends on the availability of primary data. If the available data is not adequate, then it cannot be used for further research.

8. Degree of Accuracy Desired: While planning a statistical investigation, the determination of the desired degree of accuracy is another major factor.

(b) Sampling and non sampling errors

Ans. Sampling Errors

The errors caused by drawing inference by observing the sample instead of the whole population are called sampling errors. Sampling errors occurs due to the fact that the selection of two random sample items may not be identical. These errors give us some idea of the precision of our statistical estimates. These errors arise from estimating a population characteristic by looking at only one portion. Sampling errors are evitable but can be minimised by designing the survey appropriately. Sampling errors can be classified as biased sampling errors and unbiased sampling errors.

Sampling Errors

The errors caused by drawing inference by observing the sample instead of the whole population are called sampling errors. Sampling errors occurs due to the fact that the selection of two random sample items may not be identical. These errors give us some idea of the precision of our statistical estimates. These errors arise from estimating a population characteristic by looking at only one portion. Sampling errors are evitable but can be minimised by designing the survey appropriately. Sampling errors can be classified as biased sampling errors and unbiased sampling errors.

Biased Sampling Errors

Biased sampling errors exist when the errors have a tendency to accumulate only at one direction. Bias in the selection of sample units and in its analysis, limitation of collection procedure, presence of consistent error in measuring instruments, selection through deliberate sampling methods and insufficiency in providing information for the statistical survey are the main source of occurrence of these types of errors.

Unbiased Sampling Errors

Errors which arise on the account of chance difference between the units of population included in the sample and those not included are called unbiased sampling errors. These errors have a tendency to cancel each other, thereby get neutralised with the increase in the number of observations.

Non-sampling Errors

Non-sampling errors includes biases as well as mistakes made by human and arises from inaccurate sample frames, wrong methods of data clarification or verification, improper sample allocation, poorly designed structure etc.

The sources of non-sampling errors are:

- (i) Inadequate or ambiguous data specification with respect to the objective of the survey.
- (ii) Imprecise definition of unit adopted while conducting survey.
- (iii) Inappropriate methods of interview, observation or measurement using ambiguous questionnaire, definition or instructions.
- (iv) Lack of trained and qualified investigators.
- (v) Errors in data processing operations such as coding, keying, verification, compilation and tabulation.
- (vi) Personal bias of the investigators or enumerators.

(c) Census and Sampling

Ans. Census and Sampling: Census Method: In this method each and every unit of the population is investigated for the characteristic study. This method is suitable where limited coverage of issues is required and intense examination for high degree of accuracy and reliability is an essential

Merits:

1. Information about every item on the population is obtained which is more reliable and accurate.
2. When items in a universe are of complex nature it is necessary to study each time in that situation census method is applicable.
3. Elements of bias is practically eliminated because each and every item in the population is to study.

Demerits:

1. It requires more money and time.
2. Where the population is complex practice it is difficult to contact each item during testing.
3. It could take lot of man power and proper training to the person delayed for collection of India.

Sampling method: A groups of items called sample is taken from the populations and studies from which conclusions are drawn. This method is suitable where size of population is large way high degree of accuracy is not needed.

Ques Difference between census and sampling method.

Census	Sample
1. Each and every unit of population is investigated	1. A group of items is called sampling.
2. It requires more money and time.	2. It requires speed Technical and Knowledge.
3. Needs a lot of will power and training.	3. Does not need any.
4. It is suitable where limited coverage of issues is required.	4. It is suitable where size of population is large.

(d) Schedule and questionnaire

Ans. Questionnaires: Questionnaires are one of the most popular methods for collecting data. A questionnaire contains a list of printed questions relevant to the subject under investigation. These questionnaires are sent to the persons concerned with a request to answer the questions and return back to the proper authority for data processing.

Merits:

- (i) A large field of investigation may be studied through this method.
- (ii) This method is comparatively cheaper and more reliable because the informants themselves supply the information. The information is free from the bias of the investigator.
- (iii) This method is approachable to remote areas also as the questionnaires can be sent through the mails.

Demerits:

- (i) Lack of interest from the part of informants can create a problem as they are not bothered to return the questionnaires.
- (ii) Questionnaires are circulated only to those informants who are educated and co-operative.
- (iii) Chances of obtaining wrong information enhance due to difficulty in questionnaires.
- (iv) Lack of flexibility as no modification can take place for the individual response after dispatching the questionnaires.
- (v) Personal prejudice of the informants affects the correctness of data.
- (vi) This method is quiet time consuming as the respondents take their own time to complete the questionnaires.

Features of a good questionnaires:

- (i) The questionnaire should be short, simple, attractive and free from ambiguity.
- (ii) The questions must be arranged logically and those questions should be excluded whose answers put the informants on the defensive or touch his pride.
- (iii) The questions must be framed in dichotomous manner or there must be multiple choice questions.
- (iv) Some cross checking questions should be present to check the capability of the respondents and to ensure the quality of the information.
- (v) Sufficient space must be provided for answering the questions.

Schedule: Schedule method is similar to the questionnaire method. The only difference is that in schedule method, the questionnaires are sent through enumerators. Enumerators are the agents, who are specifically trained and then employed to help the informants personally by explaining the objective of the investigation. They also remove the difficulties faced by the respondents while filling the answers. This method is adopted by the government organisation or some big agencies for conducting the survey on large scale.

Merits:

- (i) Even for the uneducated people, this method is applicable.
- (ii) As the numerators obtain the information personally from the informants, there are a less chances of non-response.
- (iii) This method provides reliable results and can be used for extensive enquiry.
- (iv) Identity of the informants is secured in this method.

Demerits:

- (i) Enumerators are paid for their work therefore this method is expensive and can be adopted by government or some big organisation.
- (ii) Enumerators should be capable and efficient as their minute mistakes can affect the quality of data compiled by them.
- (iii) The respondent may not give the answers freely of some personal questions in the presence of enumerators.