B. srividhya

Unit - 1

Statistics:

2m (The science which deals with the collection, analysis and interpretation of numerical (data.))

Sources of collecting data:

(i) Primary data
ii) Secondary data.

Primary Data:

2m (Data originally collected for an investigation are known as Primary data. They are original.)

Secondary Data

Data which are not originally collected but collected from either published or unpublished sources are called secondary data.

Methods of collecting Primary data.

1. Direct Personal Interviews.
2. Indirect Oral Interview
3. Information from Correspondents
4. Mailed Questionnaire method.
5. Schedules sent through enumerators.

I. Direct Personal Interviews:

The data are personally collected by the investigator.

Merit:

1. Data collected are original.
2. Data are true and reliable.
3. A high degree of accuracy canbe aimed.
4. The investigator can give correct information.
5. Uniformity and homogeneity canbe maintained.

Demerit.

1. It is not suitable if the area is large.
2. It is expensive and time-consuming. *costly money*
3. The chances of bias are more.
4. An untrained investigator will not bring good result.

II. Indirect Oral Interview.

The investigator contacts witness or neighbours or friends or some other third parties who are capable of giving the necessary information.

Merit.

1. It is simple and convenient. *easy*
2. It saves time, money and Labour.
3. It canbe used in the investigation of a large area.
4. The information is unbiased. *Error*

Demerits

1. Interview with an improper man will spoil the results.
2. The careless attitude of the informant will affect the degree of accuracy. *result analyse*
3. Witness may colour the information according to their interest.

IV. Information through Correspondants.

Local Agents are appointed to collect the information and submit it to the person or office. This system is adopted by newspapers, periodicals, agencies etc.

Merits.

1. Extansive information canbe had.
2. It is the most cheap and economical meth
3. Speedy information is possible
4. It is useful where information is needed regularly.

Demerits.

1. The information maybe biased. *error*
2. Degree of accuracy cannot be maintaine
3. Uniformity cannot be maintained.
4. Data maynot be original.

IV Mailed Questionaire

A Questionaire consisting of a list of necessary questions for the enquiry. There are blank spaces for answers. This Questionnaire is sent to the respondents. A covering letter is also sent, requesting the respondents to give their full co-operation by answering and return them within a specified time.

Merits.

1. It is the most <u>economical</u> method.
2. It canbe <u>widely used</u>, when the area of Investigation is large.
3. It saves money, Labour and time.
4. Error in the investigation is very small.

Demerits.

1. We cannot be sure about the accuracy and reliability of the data.
2. This method is suitable only for literate people
3. There is long delay in receiving the filled up Questionnaire/

V. Schedules sent through enumerators.

The <u>investigator</u> goes to the <u>respondents</u> Answer Person along with the questionnaires and get the replies and record their answers.

Merits.

1. It gives reliable and accurate results

2) There is less chance of non-response.

Demerits.

1. This is a very costly method.
2. It is not suited to all persons.

Sources of Secondary data.

i) Published Sources.
ii) Unpublished sources.

I. Published Sources.

Various governmental, international and local agencies publish statistical data.

a) International Publications.

International agencies and international bodies publish (regular) and (occasional reports) on economic and statistical matters.
They are the I.M.F, the I.B.R.D, the I.C.A.F.E and U.N.O etc.

b) Official Publications of Central and State Governments

Departments of the Union and State Governments regularly publish reports on a (number of subjects.) Some of the important publications are;

1) The Reserve Bank of India Bulletin,
2) Census of India.
3) Statistical Abstracts of States,
4) Agricultural Statistics of India.
5) Indian Trade Journal.

C. Semi- Official Publications.

Semi- Government institutions like Municipal Corporation, District Board, Panchayat etc Publish reports.

II Un Published Sources :

The records of some government and private institutions are not published for various reasons. Such unpublished data maybe furnished on request.

Distinction between primary and Secondary data

Primary Data	Secondary data
1. It is original, because the investigator himself collects the data	It is not original. The investigator makes use of the data collected by other agency
2. It involves large expenses interms of time, energy and money.	It is relatively a less costly method.
3. If the data hasbeen collected in a Systematic manner, its Suitability will be positive	It may or maynot suit the objective of the survey.
4. No extra precautions steps need be taken in making use of this data.	It should be used with care.

Series: ^{2M}

A series is defined as things or attribute of things arranged according to some logical order

Array: ^{2M}

A series of values arranged in either ascending or descending order are said to form an array.

Frequency distribution. ^{2M} - ^{5M}

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 frequencies side by side.^{2M}

① Discrete or Ungrouped frequency distribution.

In this the various values with their frequency are given.

X :	x_1	x_2	x_3	. . . . . .	x_n.
frequency	f_1	f_2	f_3	. . . . . .	f_n.

Continuous or Grouped frequency distribution.

In this, class intervals and their frequencies are given.

Class :	0-10	10-20	20-30	30-40	40-50	50-
frequency :	5	8	7	12	28	20)

Variables;

There are two types of Variables.
(i) Discrete Variable and (ii) Continuous Variable

Discrete Variable;

These are countable and have the property of skipping from one value to the next without taking intermediate Values.
Eg: Income, marks, Number of defective items

Continuous Variable.

These are measurable and have the property of taking all the values in an interval. Eg: Height, weight and rainfall.

Class Interval.

This is an interval of values which form a class or group.

Eg: 0-10 10-20 20-30 30-40 40-50

Exclusive Class Intervals.

When the class intervals are continuous they are called Exclusive class intervals.

Eg. 160-165, 165-170, 170-175, 175-180

Inclusive Class Intervals.

When the class intervals are not continuous, they are called inclusive class intervals.

Eg: 0-9, 10-19, 20-29, 30-39, 40-49.

Class limits:

When the class intervals are continuous or not, they are called class limits. The Least value of an interval is called the lower limit and the highest Value is called upper limit.

Eg In 10-20; 10 is the lower limit and 20 is the upper limit.

In 0-19; 0 is the lower limit and 19 is the upper limit. 2m

Class Frequency:

The Number of Values which belong to a class interval is called class frequency (or) frequency

Size of a class Interval

Size = upper limit − lower limit

$$\left(\text{Mid Value} = \frac{\text{lower limit} + \text{upper limit}}{2} \right)$$

Open - End class Intervals:

Either the lower limit of the first class interval (or) the upper limit of the last class interval (or) both are not given, they are called Open − end class intervals.

Example: 5-10 10-15 15-20 20-25 25 and above
(or)
Under 5 5-10 10-15 15-20 20-25 25-30.
(or)
Under 5 5-10 10-15 15-20 20-25 25 and above

Diagrams are various geometrical shapes like lines, bars, squares, rectangles circles, cubes, pictures maps or charts. They are visual aids for presenting the data. Thus they give a birds

Graph sheets are not needed for drawing diagrams. They are more attractive than graphs, and easier to understand. They are used for advertisement and publicity.

Rules for Construction.

1) Title:

A suitable title is given to every diagram It is to be placed at the top or bottom. It should be in capital letters.

2) Proportion between width and height

Proper proportion between width and height enhance the look. The proportion should be $1:\sqrt{2}$. Disproportion may make a diagram ugly.

3) Size.

The size of a diagram must match the sheet of paper. The diagram should not be too big or too small.

4) Scale:

Proper scale should be chosen for each axis. It should be written at the top right side corner of the diag

5) Suitable Diagram.

Suitable Diagram is chosen for the nature of the data.

6) Simplicity.

Diagrams should be as simple as possible, so that they could be understand easily. For a large and complex data, more than one diagram should be drawn.

7. Neatness.

Diagrams should be drawn neatly and accurately. Drawing instruments are to be used wherever necessary.

8. Foot Note and Source:

Foot note are to be given to clarify the details which are not available. Note is given for the source. These are placed at the bottom left corner.

9. Identification Number.

Each diagram is to be given an identification number. It is placed at the bottom, middle position.

Types of Diagrams

1. One - Dimensional diagrams - Bar diagram
2. Two - Dimensional diagrams - pie diagram, Rectangles, squares and Circles.
3. Three - dimensional diagrams - Cubes.
4. Pictograms and Cartograms.

Suitable Diagram is determined on the basis of the nature of the data and kind of people for whom it is meant.

One Dimensional Diagrams.

Bar diagram consists of rectangles of equal width. The height or lengths varys according to the values. Rectangles are drawn on a common base line with equal gaps in between.

Vertical bars are used for all types of data. Horizontal bars are used for geographically or qualitatively classified data.
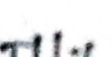
Advantages:

1) They are easy to draw and simple to understand.
2) When items are in large number all of them are represented in a bar chart.
3) Effective comparisons are possible in a bar diagram.

2(i) Simple Bar Diagram.

It is suitable to represent one - way classified data.

11) Multiple Bar Diagrams.

i) (Equal number of adjoint bars are draw
for each group.) (ii) Bars of equal width are
drawn on a common base line, with equal
gap between the groups of bars). Bars which
represent the same aspect in different groups
are given the same colour or design. (iv) (It is
suitable when components are tobe compared with
each group.

700
600
500
400
300
200
100

1961   1971   1981   X
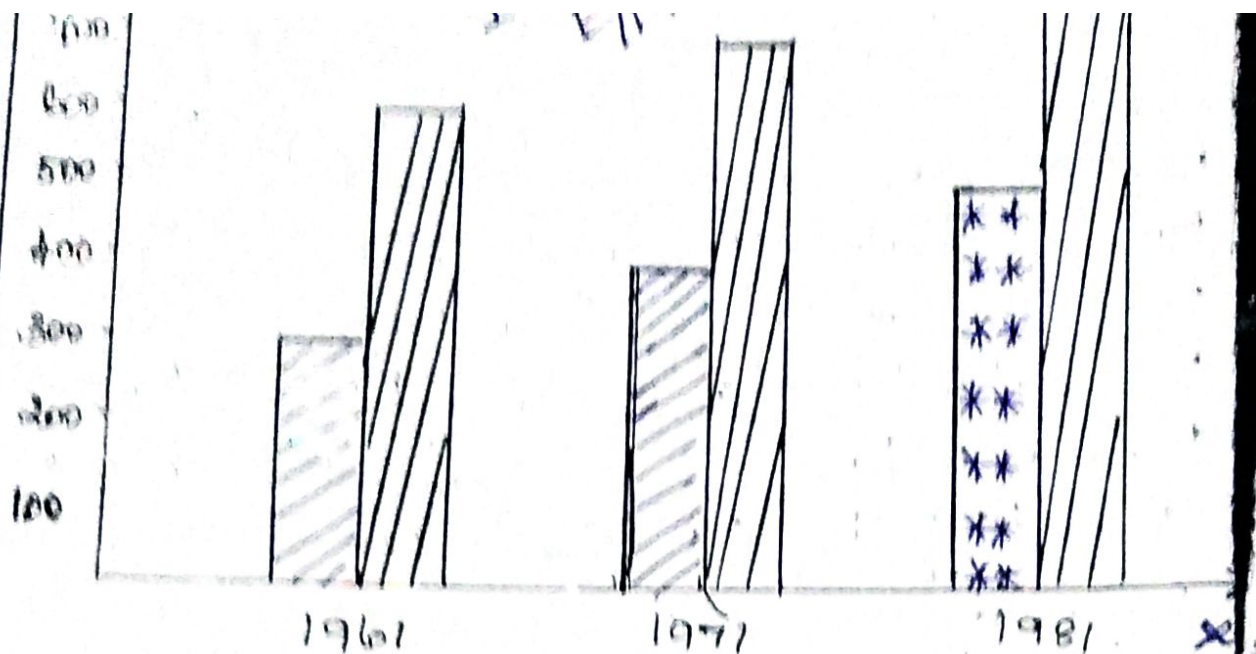
III Sub-divided Bar diagram.

(i) This is also called Component Bar diagram
(ii) Each bar is drawn for the total and it subdivided into parts. (iii) The bars are of equal width. (iv) Equal gaps between the bars. Different colours are used for the subdivisa
(v) It is suitable when the totals a well as components are to be compared

Percentage Bar diagram

It is nothing but a subdivided bar diagram representing the percentage of the components. The total percentage is 100. So the bars are of equal height.

$$Percentage = \frac{Actual\ Value}{Total} \times 100$$

It is suitable when the components are to be compared in percentage) 10M basis.

class	No. of St	%
BSc	320	$\frac{320}{800} \times 1 = 40\%$
Bcom	200	$\frac{200}{800} \times 1 =$
BA	160	$\frac{160}{800} \times 10 = :$
BBA	120	$\frac{120}{800} \times 101$
	800	1

4

Pie Diagram:

It consists of one or more circle whic[h] are divided into a number of sectos. [I] so called because the circle looks like a and the sector resembles slices cut from pie. It is more attractive. Pie diagram remains effective when there are more tha[n] four components.

Limitations.

(i) It is less effective than Bar diagra[m] for comparison and the interpretation.

[ii]) There should not be more than eight sect[o]

ii) More than three set of Values could not presented in a single diagram.

Steps For drawing one circle

(1) $\text{Angle} = \left\{ \dfrac{\text{Actual Value}}{\text{Total Value}} \right\} \times 360$

If percentages are given,

$\text{Angle} = \left\{ \dfrac{\text{Item's percentage}}{100} \right\} \times 360$

Angles are taken to be nearest integral value[s] The total must be 360°.

2) Draw a circle

3) Using the protector divide the circle into sectors whose angles are calculated

4) Different colours or designs are used for each sector

5) Write the Title and Index.

Graphs.

✱ Graphs are charts consisting of points, lines and curves. It is drawn on graph sheets. Suitable scales are chosen for X and Y axes. The area of the chart on the two axes of the graph sheet should be in the ratio 1:1.5 for a good look.

General Rules.

1. Title:

A suitable title is given to every graph. It should be in Capital letters. It is to be placed at the top or bottom.

2) Proportion.

For X and Y axis: 1:1½ - length of a gra

3) Size

The size of a graph must match the graph Sheet. The graph should not be too big or too small.

4) Scale:

Proper scale is to be chosen for each axes. It is written at the top-rightside corner.

5) Index:

Different kinds of lines may be drawn in a graph. They are clearly indicated.

6) Axis

Independent Variable or class interval is represented in X-axis. Dependent Variable or frequency is taken in Y axis.

7) Origin :

The X-co-ordinate of the origin maybe different from zero. But the Y-co-ordinate is taken as zero. If the least Y value is much away from zero "false base line" is used.

8) Simplicity

Graphs should be as simple as possible so that they could be understood easily. Too many graphs should not be over crowded in a Sheet.

9. Neatness

Graphs should be drawn neatly and Accurately. Drawing instruments are to be used Wherever necessary.

10. Foot Note and Source

Foot notes are to be given to clarify the details which are not available. Note is given for the source. These are placed at the bottom Left corner,

11. Identification Number.

Each graph is to be given an Identification number. It is placed at the bottom, middle position.

Histogram

It is suitable to represent a frequency distribution. It has adjoint bar erected on x-axis. Histogram is drawn for equal or unequal class intervals. Class intervals are taken in X axis, Frequencies in Y axis. The heights of the bars are equal to the corresponding frequences. Proper scales are selected for drawing Histogram.

Advantages

(i) It is the popular method of presenting a frequency distribution.

(ii) It helps to find mode for equal class intervals.

Disadvantages

It is not useful for finding Median, Quartiles, Deciles and percentiles.

How To find Mode from Histogram

Draw two lines diagonally inside the highest bar, starting from each upper corner to the nearby corner of the adjoint bars. From the intersecting points draw a dotted line perpendicular to x-axis. Find the Value. This is the Modal Value.

Note: Inclusive intervals are converted into true class intervals for drawing Histogram.

Frequency Polygon and Frequency Curve

Mid points are taken in X-axis. Frequency in y axis. After plotting the points join them by lines with the help of a scale. Thus we get a Frequency Polygon.

If a smooth curve is drawn along all the points. It is called a frequency curve.