Statistics:- R. A. Fisher defined statistics as "The science of statistics is essentially a branch of applied mathematics and may be regarded as mathematics applied to observational data.

Statistics is the science which deals with the collecting, classifying, presenting, comparing and interpreting numerical data.

Functions of statistics:- The following are the main functions of statistics.

1. Presents facts in numerical figures:-
The 1st function of statistics is to present a given problem interms of numerical figures. We know that the numerical presentation helps in having a better understanding of the nature of a problem.

2. Presents complex facts in a simplified form:-
Generally, a problem to be investigated is represented by a large mass of numerical figures which are very difficult to understand and remember. Using various statistical methods

this large data can be presented in a simplified form.

3. Studies relationship between two or more phenomena:
Statistics can be used to investigate whether two or more phenomena are related.

4. Provides techniques for the comparison of phenomena:-
Many times, the purpose of undertaking a statistical analysis is to compare various phenomena by computing one or more measures.

5. Helps in the formulation of policies:-
Statistical analysis of data is the starting point in the formulation of policies in various economic, business, and government activities.

6. Helps in forecasting:-
The success of planning by the govt. or of a business depends to a large extent upon the accuracy of their forecasts. Statistics provides a scientific basis for making such forecasts.

7. Provides techniques for testing a hypothesis:-

A hypothesis is a statement about some characteristics of a population. By using some statistical techniques, it is possible to test the validity of the statement.

8. Provides techniques for making decisions under uncertainty:-

Many times we face an uncertain situation. For ex, a person may face a situation of rain or no rain and he wants to decide whether to take his umbrella or not. The answer to such problems are provided by the statistical techniques of decision making under uncertainty.

Collection of data:-

Data collection is the process to gather information about the relevant topic of research. which is being done by researcher.

Sources and methods of collecting data:-

There are two sources in collecting data. They are,

Sources of collection of data:-

(1) primary data (2) Secondary data.

(1) primary data:- Data collected by investigator himself is called primary data.

methods in collecting primary data:- the following are the some of the methods to collect primary data.

(i) Direct personal interview:- In this method, data is personally collected by the interviewer.

(ii) Indirect oral investigation:- Data is collected from third parties who have information about subject of enquiry.

(iii) mailed questionnaire method:- Data is collected through questionnaire mailed to the informant. Questionnaire means a list of questions.

(iv) Telephonic interviews method:- Data is collected through an interview over the telephone with the interviewer.

(2) Secondary data:- the data, which have been collected by some individual or agency and statistically treated to draw certain conclusions. i.e, data collected by someone and used by the investigator. Secondary data is already existing and not original. secondary data has already been collected for some other purpose.

(3) presentation of data:- presentation of data includes classification and tabulation of data.

classification of data:- classification is the process of arranging data in groups according to their resemblance. Different modes of classificatin are, (i) Geographical classification (ii) chronological (iii) qualitative and (iv) quantitative classification.

Geographical classification is according to place, area, or region.

chronological classification is according to the time, i.e, monthly, yearly, daily etc.

Qualitative classification, according to the attributes of the subjects & items. i.e, honesty, beauty, colour, qualification etc.

Quantitative classification, according to the magnitude of the numerical values i.e, income, height, weight, marks, etc.

Tabulation of data:- It is the process of presenting data in rows and columns, so that it can more easily be understood and can be used for further statistical analysis.

Objectives of tabulation of data:-

(1) To reduce complexity of data.
(2) To economic space.
(3) To clarify the object of investigation.

Components of table:-

The main components of table are,
(1) Table number (2) Title (3) caption (column headings) (4) stubs (Row headings)
(5) Body of the table (6) source (7) unit of measurement (8) Head note (9) Foot note.

General rules for tabulation:-

(1) The table should suit the size of paper usually with more rows than columns. Space must be allowed for reference or any other matter which is to be included in the table.

(2) In all tables the captions and stubs should be arranged in some ~~sy~~systematic order. The arrangement of items basically depends upon the type of data.

(3) The point of measurement should be clearly defined and given in the table such as income in rupees or weight in pounds etc.

4. The table should not be overloaded with details

5. percentages and ratios should be computed and shown.

6. Abbreviations should be avoided especially in titles and headings. For example. "yr" should not be used for year.

⇒ (b) Difference between classification and Tabulation:

Ans:-

Classification	Tabulation
1. It is the basis for tabulation.	1. It is the basis for further analysis
2. It is the basis for simplification.	2. It is the basis for presentation.
3. Data is divided into groups and subgroups on the basis of similarities and dissimilarities	3. Data is listed according to a logical sequence of related characteristics.
4. Data are separated and grouped based on a property of the data common to all values	4. Data is arranged into columns and rows based on characteristics of properties

Graphical presentation:-

Graphical presentation refers to the way of presenting the data with the help of graphs.

Guidelines for construction of graphs:-

1) **Title**:- A heading depicting the contents of the data must be provided as title for all the graphical representation.

2) **Scale**:- The scale selected must satisfy by all the values to be plotted on the graph

3) **Index**:- The index must be provided to show the scale of x and y axis.

4) **Source of data**:- The source of data gives the information about the data and is mentioned at the bottom of the graph.

Functions of graphs:-

1) The shape of the graph offers easy answers to several questions.

2) The shape of the graph gives an exact idea of the variations of the distribution trends.

(3) Graphic presentation, therefore, serves as an easy technique for quick and effective comparison between two or more frequency distributions.

Different types of graphs:-

The graphical demonstration of statistical data in a chart is normally specified as statistical graph chart. There are many kinds of graphs and charts which are used to indicate a set of data. These graphs are very helpful to recognize the statistical data. The following are some of the graphs based on statistics.

(1) Line graph, (2) Bar graph (3) Histogram (4) Frequency polygon (5) Smoothed frequency curve and (6) ogive (7) cumulative frequency curves. (8) pie charts.

(3), (4), (5), and (6) are called graphs of frequency distribution.

(1) Line graph:- A line graph is a diagram that shows a line joining several points. A line graph can be taken as xy plane, where there will be an independent variable and a dependent variable. Mostly the independent variable is taken on the x-axis while the dependent variable on the y-axis.

(2) Bar graph:- Bar graph is drawn on an x-y graph and it has labelled horizontal or vertical bars that show different values. The size, length, and colour of the bars represent different values. Bar graph is very useful for non-continuous data.

(3) Histogram:- one of the most commonly used and easily understood methods for graphical representation of frequency distribution data is called histogram. It is also known as column diagram.

During the construction of histogram variable is taken on the x-axis and frequencies

on y-axis. If the difference between the class intervals are same then distance between the rectangles on the x-axis should be same. The frequencies of each class which is equivalent to its height of rectangle can be shown on y-axis.

(4) Frequency polygon:- A frequency polygon is a graph of frequency distribution. There are two ways in which a frequency polygon may be constructed.

(1) We may draw a histogram of the given data and then join by straight lines the mid points of the upper horizontal side of each rectangle with the adjacent ones. The figure so formed is called frequency polygon.

(2) Another method of constructing frequency polygon is to take the mid-points of the various class-intervals and then plot the frequencies corresponding to each point and to join all these points by straight lines. The figure obtained would exactly be the same as obtained by method 1. The only difference is that here we have not to construct histogram.

5) Ogive curve:- The ogive curve is also known as cumulative frequency curve. There are two techniques for constructing an ogive curve.

(a) more than ogive curve

(b) Less than ogive curve.

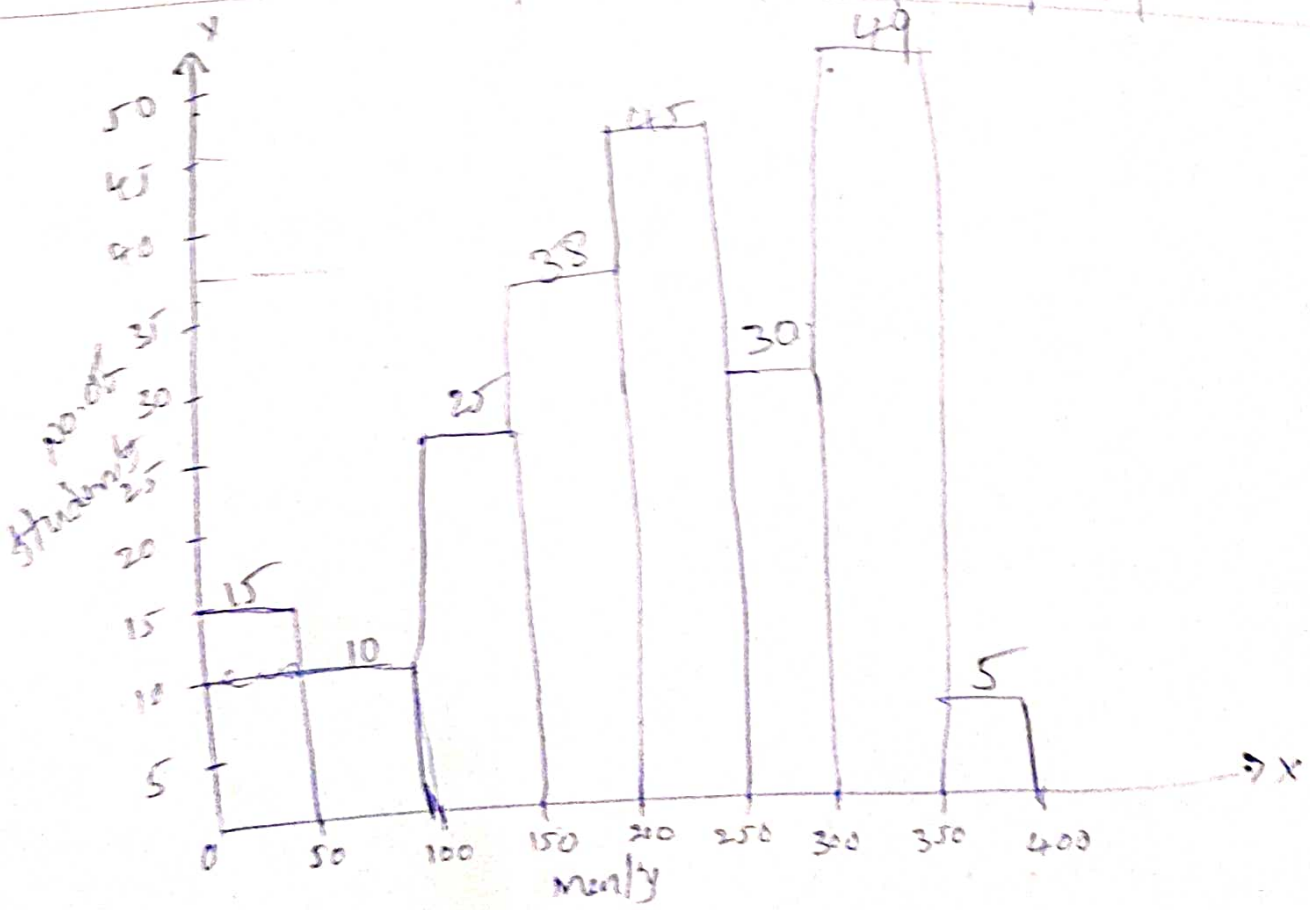
(a) More than ogive curve:- In this method, we start with the lower limits of the classes and from the frequencies we subtract the frequency of each class. when these frequencies are plotted we get a decline curve.

(b) Less than ogive curve:- In this method, we start adding the frequencies. when these frequencies are plotted we get a rising curve.

(6) Pie chart: A pie chart can be taken as a circular graph which is divided into different disjoint pieces, each displaying the size of some related information. Pie charts are best used with respect to categorical data which helps one understand what percentage each of these category constitutes.

(7) Draw the histogram for the following.

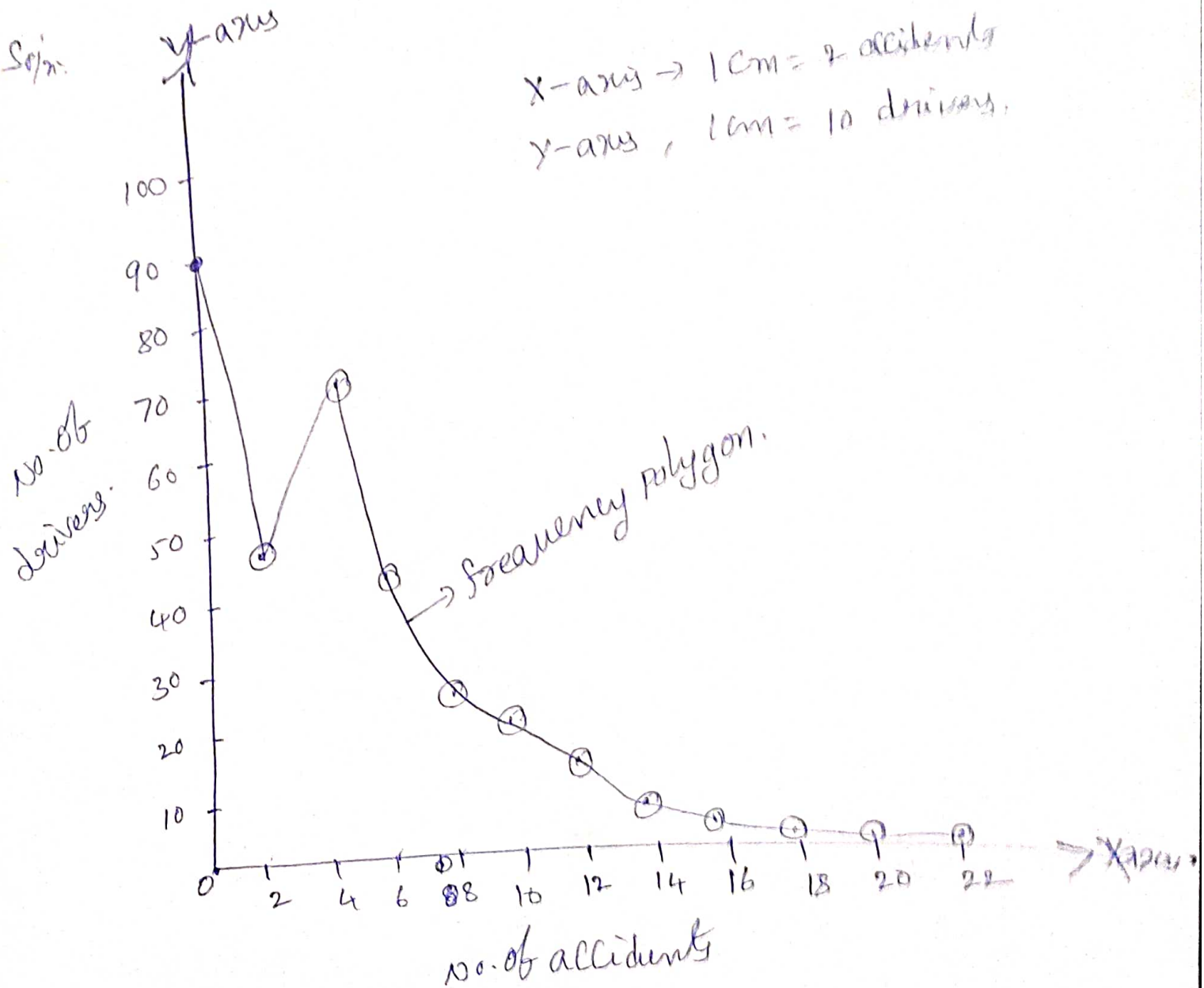
Marks	0-50	50-100	100-150	150-200	200-250	250-300	300-350	350-400
No. of students	15	10	25	38	45	30	49	5



(8) Draw the frequency polygon.

No. of accidents	0	2	4	6	8	10	12	14	16	18	20	22
No. of Drivers	90	46	70	45	28	25	15	9	4	3	2	1

Soln.



(3) Construct less than, more than ogive curves.

class :-	0-100	100-200	200-300	300-400	400-500	500-600	600-700
f :-	100	180	220	80	70	60	40

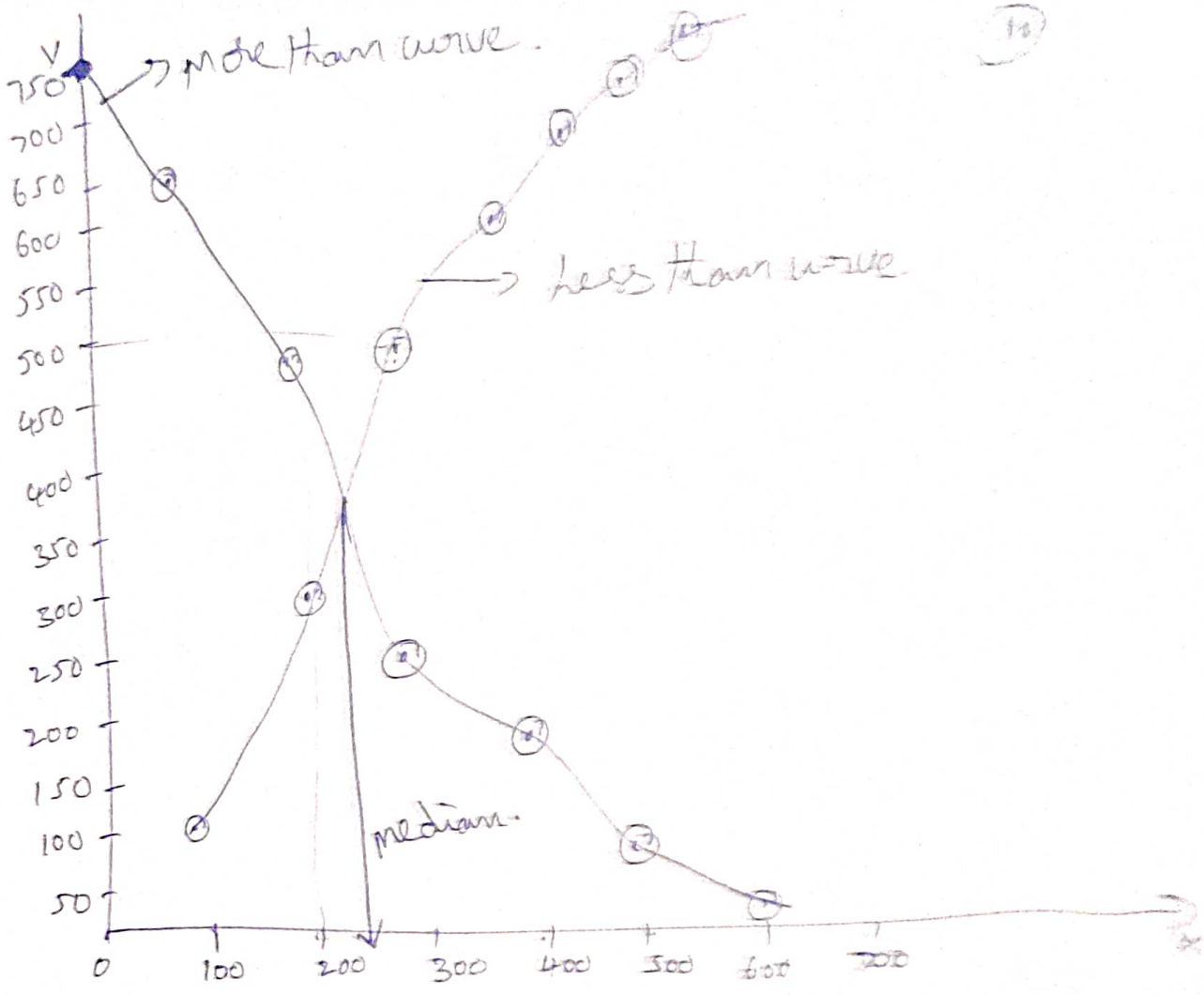
Sol:- ~~less than~~ ~~more than cf~~

Soln:- more than curve → we start with lower limits

Lower limits	f	more than cf
		750
0	100	650 (750-100)
100	180	470 (650-180)
200	220	250
300	80	170
400	70	100
500	60	40.
600	40	
	750	

less than curve → we start with upper limit.

upper limit	f	less than cf
		100
100	100	280
200	180	500
300	220	580
400	80	650
500	70	710
600	60	750.
700	40	
	750	



Median = 245 (approximately).