Kavitha
Date
Page

Module - 01
Introduction to Statistics

INTRODUCTION :-

The word statistics is derived from latin word 'status', Italian word 'statista', french word 'statistique' and german word 'statistik'. All these words means "A political state".

In earlier year statistics means collection of data facts about the people of state for the purpose of the formula-tion of policies.

MEANING :-

Statistics refers to quantitative data or, it is the method of dealing with the quantitative data.

Statistics is the science and art of collection, organisation, presentation, analysis and interpretation of numerical data. In Business statistics is used in making decision, estimations and comparision.

DEFINITION :-

According A.L. Bowley, Statistics is defined as "the science of counting",

later he redefined it as "The science of average".

- According to Boddington, "Statistics is the science of estimation and probabilities.

- According to Croxton and cowden, "Statistics is the science of collection, presentation, analysis and interpretation of numerical data.

CHARACTERISTICS OR NATURE :-

* Statistics means an aggrate of facts.
* Statistics are affected to marked extent by multiplicity cause [Probability].
* Statistics are numerically expressed.
* Statistics are collected in a systematic manner.
* Statistics are collected for predetermin-ed purpose.

USES / ADVANTAGES / FUNCTIONS :-

* It simplifies the complexity of the data.
* It helps in comparing different sets of figures.
* It helps in decision making.

* It helps in future forecasting.
* CAPIF → classification, Analysis, presentation, interpretation and forecasting.

LIMITATIONS / DISADVANTAGES:-

* It deals with only quantitative data.
* A common man cannot handle or cannot understand statistics.
* Statistical result are true only on average basis.
* It doesnot deal with individual.
* collected data may be misused or manipulated.

COLLECTION OF DATA:-

Data means the facts, figures, features, evidences, characteristics or any other material serving as a Basis for statistical investigation, once this data are produced, it becomes the information.

The person who is involved in the collection of data is called as investigators.

The person from whom the data is collected is called as respondent.

SOURCES OF DATA COLLECTION:-

(i) **PRIMARY DATA:-**

The data which are originally collected for the first time by an investigator for statistical investigations is known as primary data.

The sources of collecting primary data are as follows:-

a) By observation method
b) By direct personnel interview
c) By indirect interview
d) Mailed - questionaire method
e) Schedule sent through the enumerators etc.

(ii) **SECONDARY DATA:-**

The data which are already collected by someone else is called as secondary data.

The sources of collecting secondary data are as follows:-

a) **PUBLISHED SOURCES:-**

When data or sources are available in printed format is known as published sources.

Ex:- Newspapers, Magzines, Journals, Articles, printed notes, Text books, Govt gazettes, publication of central govt, state govt and international organisation.

Kavitha

UNPUBLISHED SOURCES:-

When data or sources are not available in printed format is known as unpublished sources.

Ex:- Records maintained by research institution and by researches and the records maintained by private and commercial organisations.

CLASSIFICATION OF DATA:-

MEANING:-

It is a process of arranging the data into the homogenous group, sequence, (or) classes according to their common characteristics (or) Attributes in the simple classification is the process of grouping of related facts into the classes.

DEFINITION:-

According to Horale Secrist, "classification is the process of arranaging the data into sequence or groups according to their characteristics or seperating them into different but related parts.

OBJECTIVES:-

* To condense the large data to make them easily understood.

* To facilitate comparision and highlights the significant aspects of data.

* To focus on the important information collected.

* Helps to draw inferences.

TYPES OR METHODS OF CLASSIFICATION OF DATA:-

classification of data can be done in four ways

(i) Geographical classification
(ii) chronological classification
(iii) Quantitative classification
(iv) Qualitative classification.

(i) GEOGRAPHICAL CLASSIFICATION:-

Here, the data are classified on the basis of geographical or location difference between the items or data like country, state, region, district wise etc.

Ex:- population of various states as per von census - state wise etc.

(ii) CHRONOLOGICAL CLASSIFICATION:-

Here, the data are classified on the basis of occurrance of the time or the difference in the time.

Ex:-

Year	Sales (₹ in crores)
2018	35.8
2019	49.2
2020	53.16
2021	68.5

(iii) **QUANTITATIVE CLASSIFICATION:-**

Here, the data are classified on the basis of some characteristics which can be measured numerically. (Or) the data classified on the basis of characteristics which are of quantitative nature like Age, Height, weight, income, marks, production etc.

(iv) **QUALITATIVE CLASSIFICATION:-**

Here, the data are classified on the basis of some qualities which cannot be measured numerically. (Or) the data classified on the basis of characteristics which are of qualitative nature like gender, beauty, intelligence, honesty, efficiency etc.

a) Simple classification:-

when only one attribute or character is used classification is called as simple classification.

bb Manifold classification:-

when more than one attribute or

character is used for classification is called manifold classification.

TYPES OF DISTRIBUTIONS:-

It can be divided into 3 types.

(i) INDIVIDUAL Observation method
(ii) Discrete series
(iii) Continous series

(i) INDIVIDUAL OBSERVATION METHOD:-

If the data consist of only the value of variable is known as individual observation method or ungrouped data.

Ex:-

x (height in cms) = 100, 110, 120, 130, 140, 150.

(ii) DISCRETE SERIES:-

If the data consist of value of variable and also the frequency such a distribution is called discrete series.

x (height in cms)	No. of students
100	5
110	10
120	20
130	7
140	4
150	1

Kavitha

(iii) **CONTINOUS SERIES :-**

If the data consist the value of variables with a class interval and a frequency such a distribution is called as a continous series.

x (height in cms)	100-110	110-120	120-130	130-140
No. of students	5	10	20	7

140-150
4

Under continous series the distribution can be made 2 ways :-

⇒ Exclusive distribution

⇒ Inclusive distribution

⇒ **EXCLUSIVE DISTRIBUTION :-**

when the value of lower limit will be included in the same class of the value of upper limit is excluded from the class and included in the succeeding classes the upper limit of the class will be the lower limit of the succeeding classes such as distribution is called as exclusive distribution or exclusive class limit.

x (heights in cms)	100-110	110-120	120-130	130-140	140-150
No of students	5	10	20	7	4

⇒ INCLUSIVE DISTRIBUTION:-

When the value of both upper limit and lower limit will be included in the same classes, such a distribution is called as inclusive distribution.

x (Height in cms)	100-109	110-119	120-129	130-139	140-149
No of students	5	10	20	7	4

TABULATION OF DATA:-

MEANING:-

Tabulation may be defined as the systematic presentation of numerical data in rows and columns.

According to certain characteristics. In simple it is a process of systematic arrangement of data classified into rows and columns is called as tabulation.

FORMAT OF TABLE:-

Title:-

Headnote:-

Table no:- Date:-

Row-↓ column→	caption (column heading)	Total
stub entries (row heading)	Body of the table	
Total		Grand total

Source :-

Foot note :-

PARTS OF THE TABLE :-

(i) **TABLE NUMBER :-**

A table should be numbered for identification and for future references especially when there are a large number of table is a study

(ii) **TITLE :-**

Every table must be given a suitable title which describes the content of the table. It should be clear and brief.

(iii) **DATE :-**

Date of preparing a table should be written to identify the chronology of the table prepared.

Kavitha

(iv) **STUBS:-**

Stubs are designation of rows (rows headings).

(v) **CAPTIONS:-**

Captions are designation of columns (column headings).

(vi) **BODY OF THE TABLE:-**

The actual data are arranged in the body part of the table. It is the most important part of the table.

(vii) **HEADNOTE:-**

It is the brief explanatory statement to all or a major part of the data table.
Ex:- ₹ in crores or units in 1000's etc.

(viii) **SOURCES:-**

A note at the bottom of the table indicating the sources from which the data contain in the table are collected should be shown.

(ix) **FOOT NOTE:-**

In case of irregularity occuring in a table or when any item thereoff as not been adiquatly explained, or any abbre-viations used, it is preferdable added as

an explanatory note at the bottom of the table as a foot note.

TYPES OF STATISTICAL TABLE:-

* **SIMPLE AND COMPLEX TABLE:-**

Simple table is also called as one way table showing only one characteristics of the data. On the other hand two or more characteristics are shown simultaneously in a table is called as complex or manifold table.

* **GENERAL PURPOSE AND SPECIFIC PURPOSE TABLES:-**

General purpose tables or reference tables are the tables which provides information for general purpose. Special purpose table or summary tables provides information for a particular purpose these tables are sometimes called derivative table since they are derived from the general table.

PROBLEMS ON TABULATION:-

1) In the house of lokshaba, there were 600 members, were present. During the discussion on a motion put to vote 400 voted

in favour of the resolution. The government members in the house were 380, 65 members belonging to the Opposition voted for the resolution. All the members were belonging to either of the two groups and there were no absentees. Tabulate the information.

Solu:-

Tittle:- Table showing voting Pattern in Lokshada and draw the table.

Headnote:-

Table no:- Date:-

Party. / Votes	Ruling Party	Opposition Party	Total
Infavour	335	65	400
Against	45	155	200
Total.	380	220	600

2) The number of workers in a large factory in 2006 was 540. Out of which 30% were females and the rest males. In 2008, the strength of the workers is increased by 100 females and 200 males. In 2010, the total no of workers increased by 25% on its value in 2008. The female workers were 340. Tabulate the above information.

Kavitha

Solu:- Table showing the no. of workers in a factory.

Year / Gender	2006	2008	2010
Male	378	578	710
females	162	262	340
Total	540	840	1050

3) In a sample study about coffee habits in two towns, the following informations was received.

Town A = females were 40%, Total coffee drinkers were 45% and Male, non-coffee drinkers were 20%.

Town B = female = 55%, Male non-coffee drinkers were - 30% & female coffee drinkers were 15%.

Represent the information in a form of tabulation.

Solu:- Table showing coffee drinking habits in town A & B (figures in %)

Town	A			B		
Habits / Gender	M	F	T	M	F	T
coffee drinkers	40	5	45	25	15	40
Non-CD	20	35	55	30	30	60
Total	60	40	100	55	45	100

Kavitha

4b In a state, there are 30 lakh people. Out of this, 10 lakh people live in urban areas and the rest in rural areas. In urban areas, there are 7 lakhs male people, out of which, 2.5 lakh are illitrates. In urban areas 2 lakh ladies are illitrates. In rural areas, there are 15 lakhs male people of which 5 lakhs are litrates. In rural areas litrates ladies are 3 lakhs. Tabulate the above information.

Solut Table showing litrates and illitrates in a state. [figures in lakhs]

Area / General literacy	Urban			Rural			Total		
	M	F	T	M	F	T	M	F	T
literates	4.5	1	5.5	5	3	8	9.5	4	13.5
illiterates	2.5	2	4.5	10	2	12	12.5	4	16.5
Total.	7	3	10	15	5	20	22	8	30

5b Present the following information in a statistical table.

In the year 1991, there were 100 Banks out of which 50 were private banks. In these 50 Banks 10 had ATM facility and none of the nationalised had this facility. In the year 2001 there was increase of 50 Banks in the private sector and total number of banks were 175. 20 nationalised banks had ATM

facility and 75 private sector banks had these facility.

Soln Table showing sectorwise banks having ATM's for the year 1991 and 2001.

Year / Sector	1991			2001		
ATM	Public	Private	Total	Public	Private	Total
with ATM	0	10	10	20	75	95
without ATM	150	40	190	155	25	180
Total	150	50	200	175	100	275

6) In 2006, Out of total 2000 workers in a factory, 1550 were members of a trade union. The no of women workers employed was 250, out of which 200 didnot belong to any trade union.

In 2010, the no of union workers was 1725 of which 1600 were men. The no of non-union workers was 380, among which 155 were women. Present the information in a suitable tabular form

Soln- Table showing gender and trade union classification of workers for the year 2006 and 2010

Year	2006			2010		
Gender / Union trade.	M	F	T	M	F	T
Members	1300	250	1550	1600	105	1705
Non-members	250	200	450	205	175	380
Total	1550	450	2000	1805	280	2185

Q6 In 2020, out of the total customers visiting the hotel, 750 were non-vegetarian and 1250 were vegetarian customers. In total there were 550 male non-veg customers and 300 female veg customers.

In 2021, the total no of customers increased by 25%, while Non-veg customers increased by 20%. In all there were 1700 male customers, among whom 650 were non-veg in 2021. Present the above information in a suitable statistical table.

Sol: Table showing genders & foods Habits of Hotel customers for the year 2020 and 2021.

Year	2020			2021		
Genders / Habits	M	F	T	M	F	T
Veg	950	300	1250	1050	550	1600
Non-veg	550	200	750	650	250	900
Total	1500	500	2000	1700	800	2500

Kavitha

8) Out of total number of 10000 candidates who applied for jobs and government department, 6854 were males, 3146 were graduates and others were non-graduates. The number of candidates with some experi-ence was 2623 of whom 1860 were males. The number of male graduates was 2012. The number of graduates with experience was 1093. That includes 323 females. Tabulate the information.

Solu: Table showing details of qualifications of candidates who applied for jobs and government department.

Gender	Male			female			Total		
Qualification / Experience	G	NG	T	G	NG	T	G	NG	T
Experience	770	1090	1860*	323*	440	763	1093*	1530	2623*
Inexperience	1242	3752	4994	811	1572	2383	2053	5324	7377
Total	2012*	4842	6854*	1134	2012	3146	3146*	6854	10000

9) Out of, total no of 2807 women who were interviewed for employment in a textile factory, 912 were from textile areas, and the rest from non-textile areas. Among the married women who belongs to textile areas, 347 were having some work experien-ce and 173 didnot have work experience.

Kavitha

while for non-textile areas the corresponding figures were 199 and 670 respectively. The total number of women having no experience was 1841 out of whom 311 resided in textile areas. Out of the total number of women 1418 were unmarried and of those the number of women having experience in the textile and non-textile areas was 254 and 166 respectively. Tabulate the above information.

Soln:- Table showing the details of women who were interviewed for employment in a textile factory.

Areas / Marital status / Experience	textile			Non-textile			Total		
	M	UM	T	M	UM	T	M	UM	T
Experience	347	254	601	199	166	365	546	400	966
Inexperience	173	138	311	670	860	1530	843	998	1841
Total.	520	392	912	869	1026	1895	1389	1418	2807

DIAGRAMATIC REPRESENTATION:-

The term diagram means a symbolic simplified and non-numerical resprentation of data. It is a usual form for presentation of statistical data which helps to compare the variables.

Kavitha

TYPES OF DIAGRAMS:-

(i) One dimensional diagrams
- Line diagrams
- Simple bar diagrams
- Multiple bar diagrams
- Sub-divided bar diagram.
- Sub-divided percentage bar diagram

(ii) Two dimensional diagrams.
- Rectangles
- Squares
- Pie charts
- Circles

ADVANTAGES OF DIAGRAMS:-

* They are attractive and impressive.
* They have universal applicability.
* They may comparision easy.
* They provide more information and less calculations.

(i) ONE DIMENSIONAL DIAGRAMS:-

In this kind of diagram only one dimensional measurement is taken into account like height or length of the bar is used and the width is not considered, it is also called as bar diagram.

(a) **LINE DIAGRAM:-**

It is the form of vertical lines drawn on ox-axis. The distance between lines is kept uniform. It is used to compare the variables when large number of values of variables are given.

(b) **SIMPLE BAR DIAGRAMS:-**

They are very popular in practice, they are also called as columnar diagrams.

A bar diagram may be either vertical or horizontal drawn to columns of equal width and vary with width.

(c) **MULTIPLE BAR DIAGRAMS:-**

It is used for comparing two or more sets of statistical data. Bars are constructed site to site to respresent the set of value for comparision. In order to distinguish bars they may either be differently colour, crossed or dotted.

(d) **SUB-DIVIDED BAR DIAGRAMS:-**

In a sub-divided bar diagram one bar is sub-divided into various bars. In proportion to values given in

the data and whole bar represents the total. It is also called as component bar diagram. The sub-divisions are distinguished by different colours, crossing or shades etc.

(e) **SUB-DIVIDED PERCENTAGE BAR DIAGRAMS:-**

This is an other form of compo-nent bar diagram, here the components are not the actual values but percenta-ges of the whole. The main difference between sub-divided bar diagram and sub-divided percentage bar diagram. The formal bars are of different heights since their totals may be different whereas in the lateral the bars are off equal heights, since each bar represents 100%.

(ii) **TWO-DIMENSIONAL DIAGRAMS:-**

In this both length and width are taken into account, this is also called as surface diagram.

(a) **PIE CHART:-**

A pie diagram is a circular diagram which is a circle divided into sectors. It is the sub division of sector or sections according to

different attributes of data as different segments of circle. It is also called as circular or Angular or Pie charts or pie diagrams.

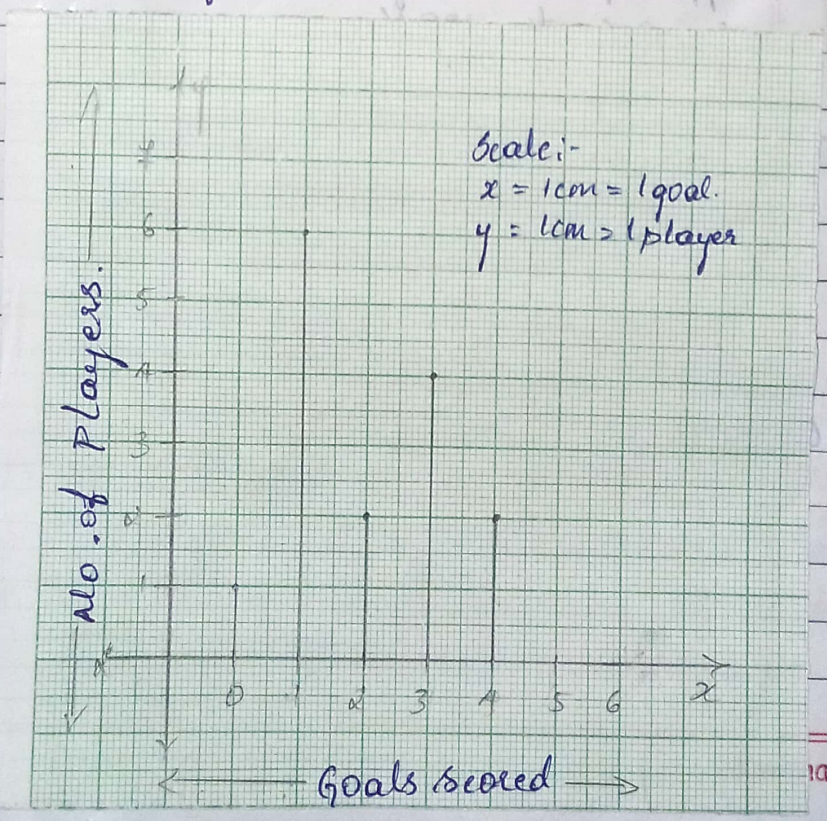
There are 360° at the centre of the circle.

PROBLEMS ON LINE DIAGRAM:-

1b There are 15 football players, the members of the star sports club of bengaluru who have scored goals in 5 matches in the year 2022 has under, Draw a line diagram showing the information.

Goals scored	0	1	2	3	4
No. of players	1	6	2	4	2

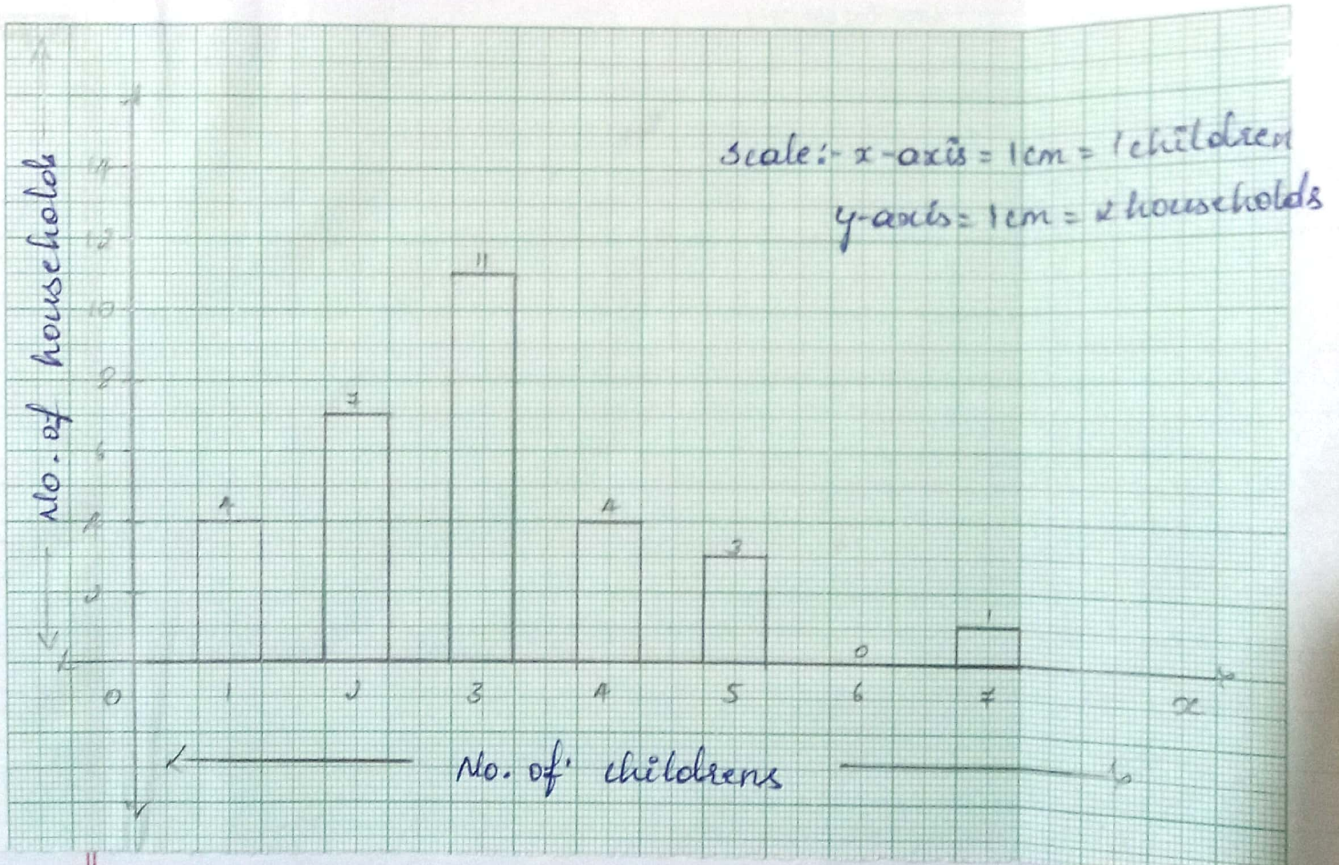
Solut



PROBLEMS ON SIMPLE BAR DIAGRAMS:-

Ex The following frequency table shows the number of children in household in a sample of 30 households during the year 2021. Draw diagram in the form of bar chart to illustrate the information.

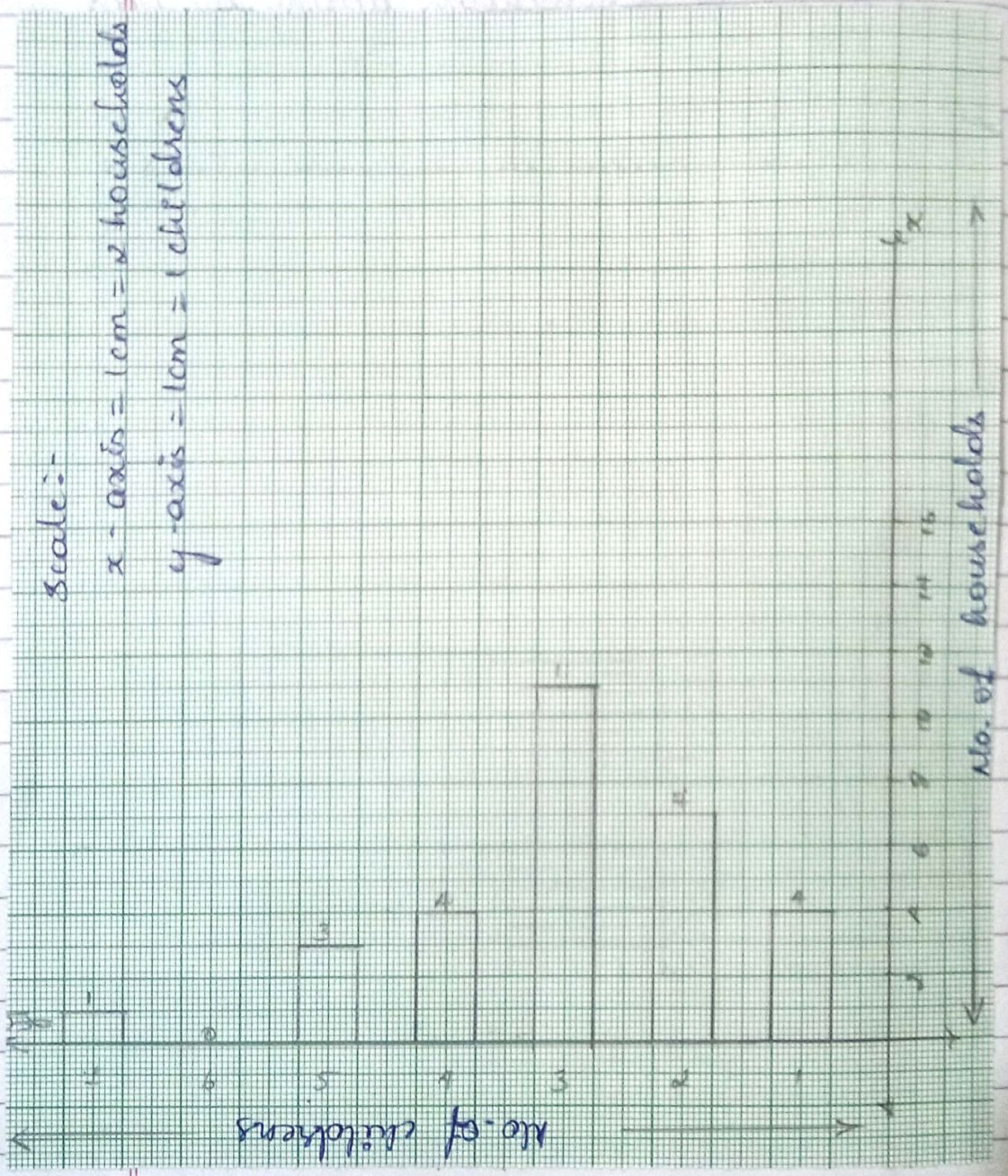
No. of children	1	2	3	4	5	6	7
No. of household	4	7	11	4	3	0	1



Kavitha

Date
Page

Scale:-
x - axis = 1 cm = 2 households
y - axis = 1 cm = 1 children

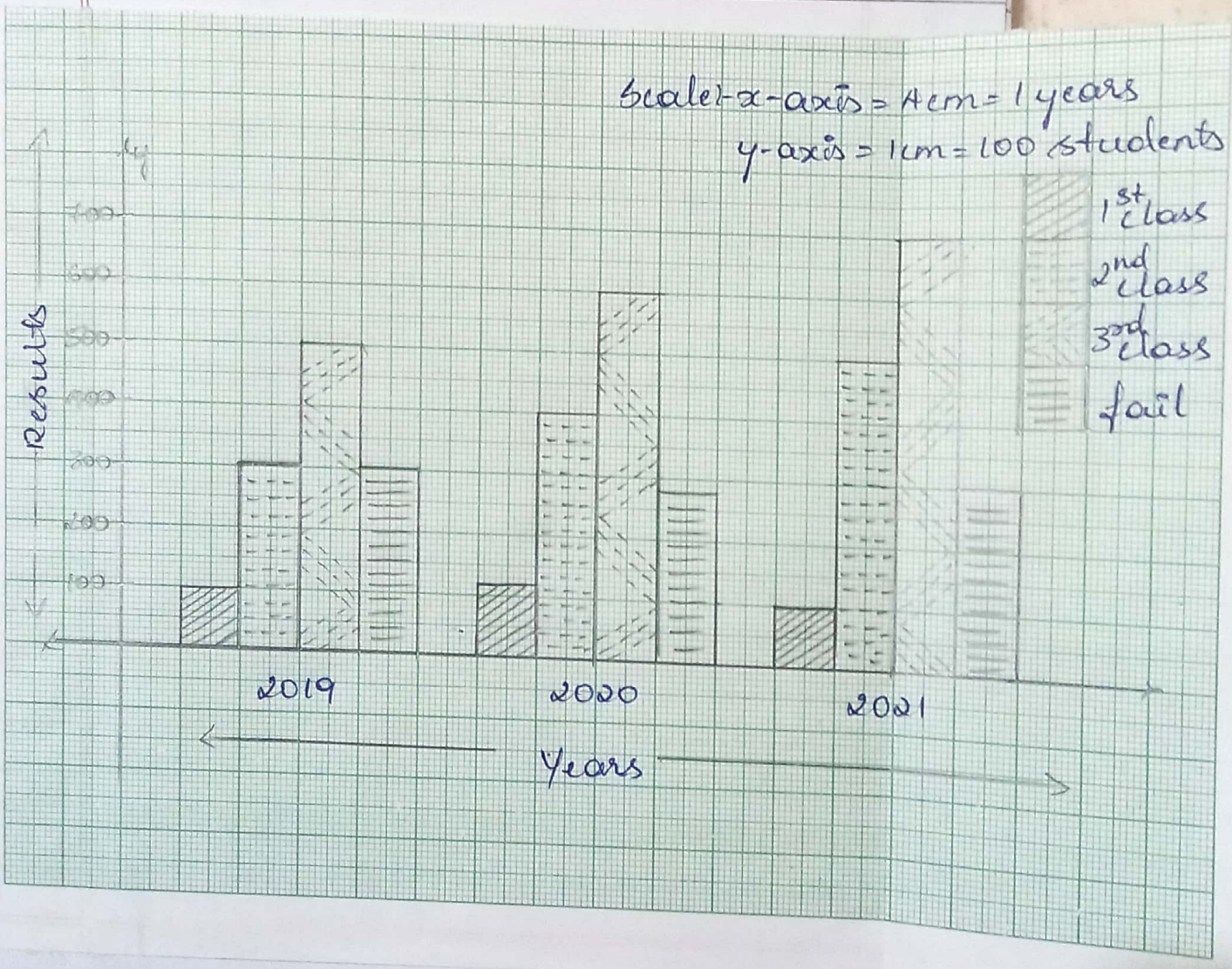


Kavitha

PROBLEMS ON MULTIPLE BAR DIAGRAMS:-

1) Represent the following data of results of Bcom students in an examination of BCU held in June 2019, 2020 and 2021.

Years	1st class	2nd class	3rd class	fail
2019	100	300	500	300
2020	120	400	600	280
2021	100	500	700	300

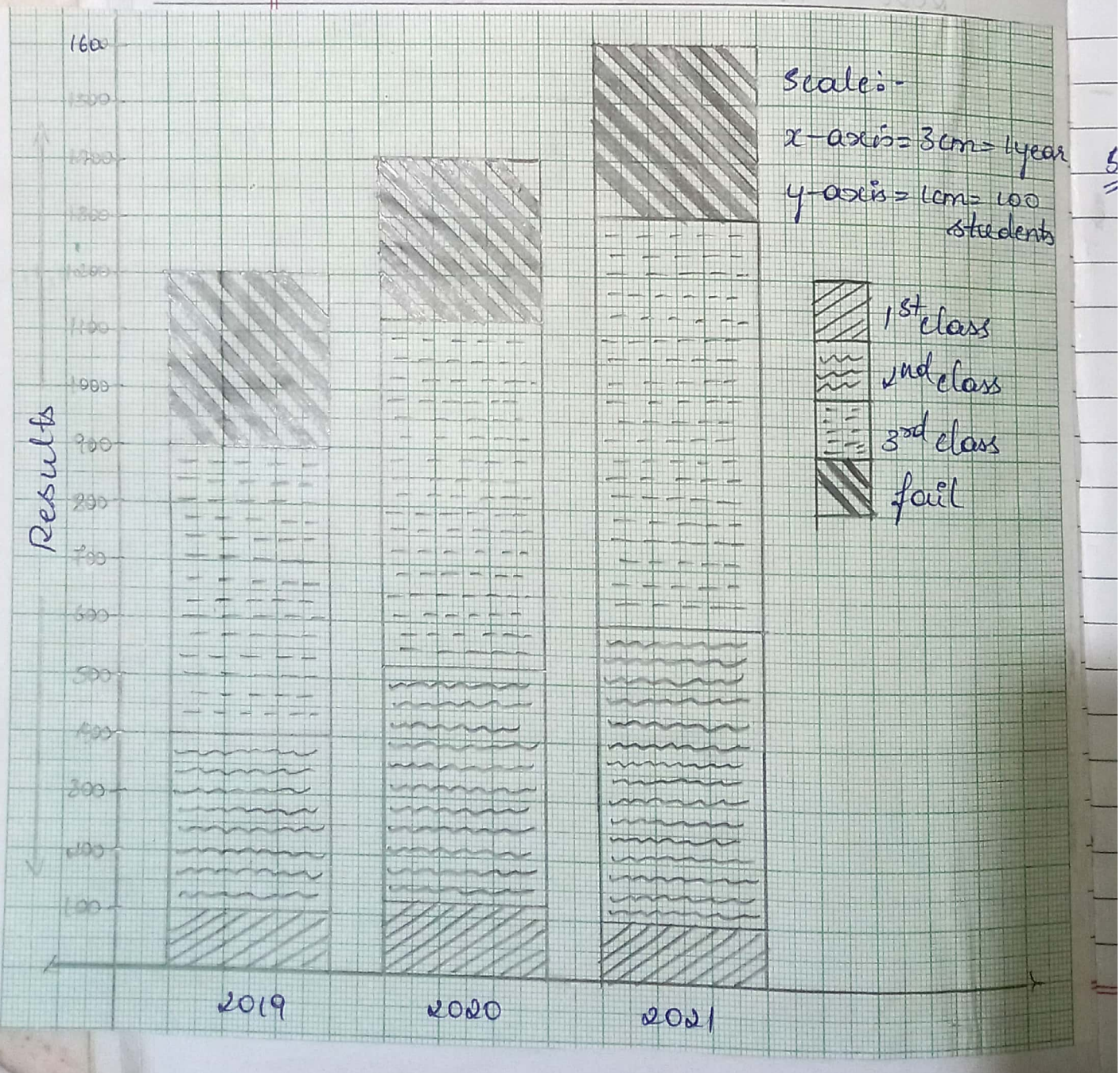


Karitha
Date
Page

PROBLEMS ON SUB-DIVIDED BAR DIAGRAMS:-

1) Represent the following data of results of Bcom students in an examination of BCU held in June 2019, 2020 and 2021

Years	1st class	2nd class	3rd class	fail	Total
2019	100	300	500	300	1200
2020	120	400	600	280	1400
2021	100	500	700	300	1600



PROBLEMS ON PERCENTAGE BAR DIAGRAM:-

* Represent the following data of results of Bcom students in an examination of BCU held in June 2019, 2020 and 2021.

Years	1st class	2nd class	3rd class	fail	Total
2019	100	300	500	300	1200
2020	120	400	600	280	1400
2021	100	500	700	300	1600

Solution:

Years	2019		2020		2021	
	value	%	value	%	value	%
2019	100	8.33	120	8.57	100	6.25
2020	300	25	400	28.57	500	31.25
2021	500	41.67	600	42.86	700	43.75
fail	300	25	280	20	300	18.75
Total	1200	100	1400	100	1600	100

Kavitha

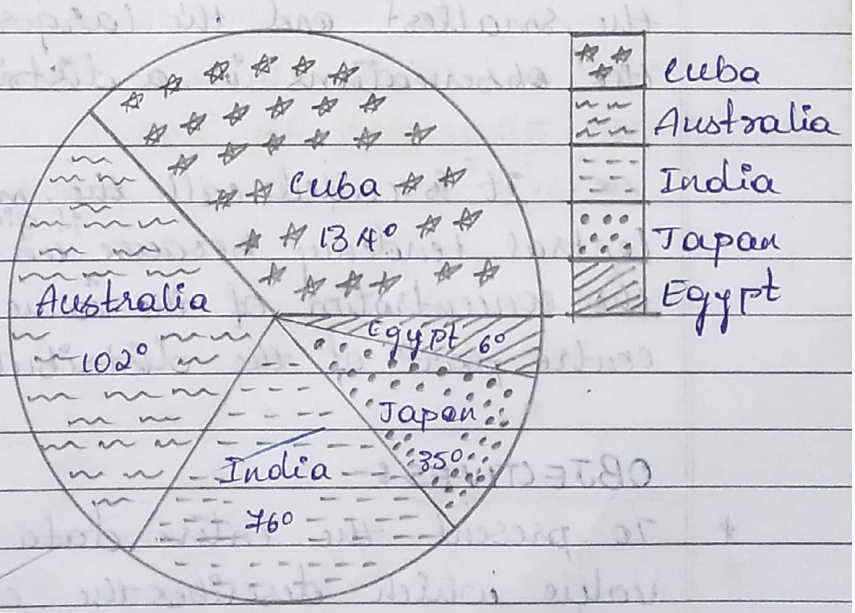
PROBLEMS ON PIE DIAGRAMS-

1. Draw a pie diagram for the following data of production of sugar in quantials of various countries.

Country	Production of Sugar (in quintals)	Degrees
Cuba	62	134°
Australia	47	102
India	35	76
Japan	16	35
Egypt	6	13
Total.	166	360.

Sol:-

Scale :- Radius 4 cm



Kavitha