3 Classification and Tabulation

Introduction

In the previous chapter we learnt how the data is collected from the source. The collected data is usually contained in schedules or questionnaires. It is called raw data. Arriving at a conclusion from the raw data is a difficult task, because they are always in an unorganised form. Therefore a proper organisation and presentation of data is required for the systematic and comprehensive statistical analysis.

From a hypermarket we can easily select the required items, because of the items of same kind are arranged together. Otherwise it would be very difficult for the customer to search where the required items are stored. Likewise the raw data should be arranged in an organised manner and systematically to simplify the further statistical procedures. This process is called classification of data.

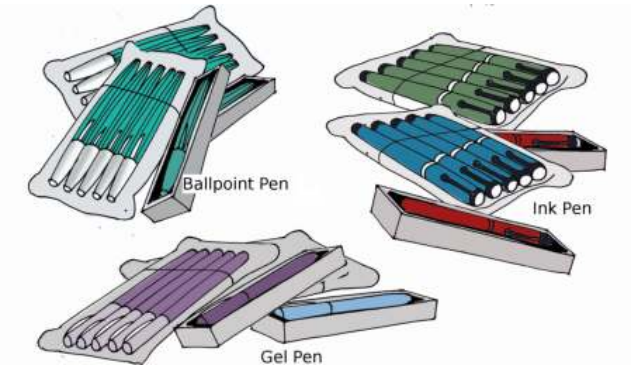
The process of grouping the data according to some characteristics is called Classification of Data.

3.1 Types of Classification

In some electronics hypermarkets we can see TVs, Refrigerators, Home Theatres, Washing Machines, Air Conditioners etc, arranged in separate section. In some other hypermarkets the item of same kind may be arranged in such a way that separate sections are allotted to each brand. In both cases the items are arranged according to some criterion or different types of classifications. In the first hypermarket the classification is according to the type of the item, whereas in the second hypermarket the classification is according to the brand of the product. Similarly the data is also classified in different ways.

Consider you are appointed as a salesman at a wholesale shop which deals in different kinds of pen. The shop manager shows a box containing packets of different types of pens, some packets contain a single pen and some others five. There are Ball Pens, Gel Pens and Ink Pens. Pens are made in various countries viz India, China, Thailand and Japan. The packing dates are also varied from February to June. You are asked to group same kind of pen together. How will you do the task?

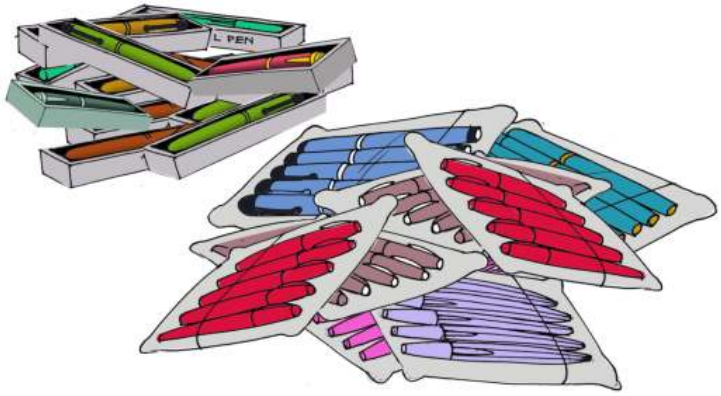
You can classify the pens according to their characteristic, such as ball point, gel or ink pens as shown in the picture. In other word you classified the pens according to their quality. This kind of classification is called Qualitative Classification.



Qualitative classification of pens

Classification based on the characteristics like Sex, Colour, Literacy, Religion,etc..., which cannot be numerically measured is called Qualitative Classification.

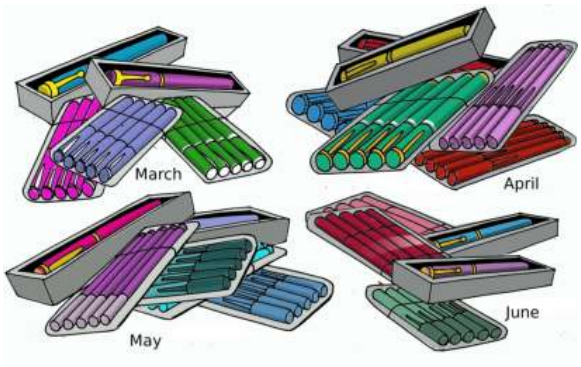
You can classify the pens with respect to the quantity of pens in the packet. If the pens are classified according to the quantity of pens contained in the packet, then it is known as Quantitative Classification.



Quantitative classification of pens

Classification based on the characteristics like Height, Weight, Thickness, Count, Area, Volume, etc.. which can be numerically measured is called Quantitative Classification

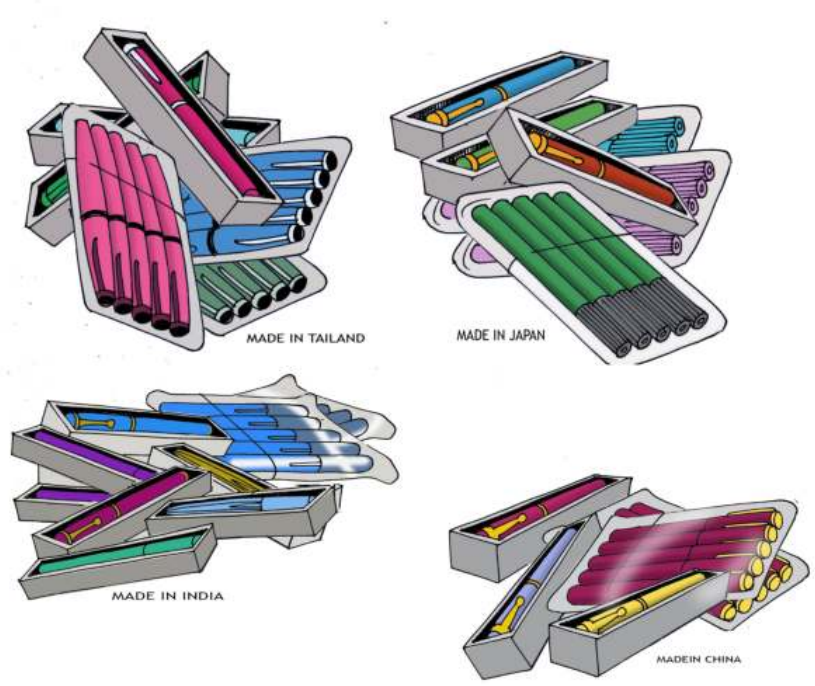
The pens may be classified according to the month of packing also. That is the pens packed in February, March, April, May and June are arranged separately. Then it is termed as Chronological Classification.



Chronological classification of pens

Classification based on some units of time like year, month, week, hour,etc, is called Chronological Classification

If you classify the pen according to the country where they are made, it is called Geographical Classification.



Geographical classification of pens

Classification based on place like Continents, Country, State, District, Village, etc is called Geographical Classification.

3.2 Tabulation of Data

We can present the classified items in different ways. Some of the methods are textual presentation, tabular presentation, diagrammatic presentation and graphical presentation.

In textual presentation, data is presented within the text. For example, you can present the data regarding the pens as shown below.

The box contains 260 packets of pens. 136 packets carry only one pen per pack.

124 packets contain 5 pens each. 101 packets of pen are made in India, 106 packets in China, 31 packets in Thailand and 22 packets in Japan. 81 packets contain ball pointed pens, 86 packets contain gel pens and 93 packets contain ink pens. 53 packets of pen are packed in February 50, 47, 47 and 63 packets were packed in March, April, May and June respectively. The box contains a total of 756 pens.

But it is not easy and suitable to present the data in this method when the quantity of data is too large or minute level classification is required. We cannot represent the data like, the number of packets containing 5 ink pens made in India and packed on June is 5, number of packets containing one ball pointed pen made in Thailand and packed on April is 2, etc.

Such details can be easily represented by a statistical table. The method of representing data with the help of a statistical table is called **Tabulation of Data.** In another words Tabulation is the way of systematic summarisation and presentation of information contained in the given data, in rows and columns. Tabular representation of data facilitates comparison by bringing related information close to each other and helps in further statistical analysis and interpretation. In fact tabulation is the final stage in collection and compilation of data and forms a gateway for further statistical analysis and interpretations as well as making the data suitable for further diagrammatic and graphic representation.

Type of Pen	Number of Packets
Ball Point	81
Gel Pen	86
Ink Pen	93
Total	**260**

Table 3.1: Qualitative Classification of data

The above table (Table - 3.1) represents the qualitative classification of data.Similarly the number of packets of pens can be presented in different tables according to different types of classification. See the following tables, Table - 3.2, Table -

3.3 and Table - 3.4.

Number of Pens per Packets	Number of Packets
1	136
5	124
Total	260

Table 3.2: Quantitative Classification of data

Made in	Number of Packets
India	101
China	106
Thailand	31
Japan	22
Total	**260**

Table 3.3: Geographical Classification of data

Month of Packing	Number of Packets
February	53
March	50
April	47
May	47
June	63
Total	**260**

Table 3.4: Chronological Classification of data

3.3 Objectives of Classification and Tabulation

Following are the objectives of classification and tabulation of data.

1. To simplify the complex data.
2. To facilitate comparison.
3. To facilitate statistical analysis.
4. To save time, space and energy.
5. To clarify similarity and dissimilarity.
6. To organise data logically and scientifically.
7. To grasp the information.

Activity

Collect tables published in journals, news papers and websites and identify the type of classifications applied in it.

3.4 One Way and Two Way Classification

In Table- 3.1, the classification is based on the characteristic, type of the pen. In Table - 3.2, Table - 3.3 and Table - 3.4 the classifications are based on number of pens per packet, month of packing and country respectively. In all these cases only one characteristic is considered for classification. So these classifications are called One-Way Classification of Data. The tables used to represent the one way classifications are called One Way Tables. If we consider two characteristics at a time for classification of data, it is termed as Two Way Classification of Data. The Table representing Two Way Classification is called **Two Way Table.** The following tables, Table - 3.5 and Table - 3.6 are examples of two way tables.

Month	Country				Total
	India	China	Thailand	Japan	
February	18	23	5	7	53
March	18	26	4	2	50
April	20	18	5	4	47
May	21	16	6	4	47
June	24	23	11	5	63
Total	101	106	31	22	260

Table 3.5: Two way table representing country and month of packing

No of Pens per Packet	Type of pen			Total
	Ball Point	Gel pen	Ink Pen	
ONE	38	49	49	136
FIVE	43	37	44	124
Total	81	86	93	260

Table 3.6: Two Way Table representing type of pen and number of pens per packet

Activity

Construct the skeleton of two way tables regarding the above problem. one example is given below.

Month	No. of Pens		Total
	ONE	FIVE	
February			
March			
April			
May			
June			
Total			

Similarly we can represent the data by considering more than two characteristics or attributes together for classification and tabulation of data. Table 3.7 is an example for such a classification.

Made in		India				China				Thailand				Japan				Total			
Packed on	Type	Ball Point	Jell Pen	Ink Pen	Total	Ball Point	Jell Pen	Ink Pen	Total	Ball Point	Jell Pen	Ink Pen	Total	Ball Point	Jell Pen	Ink Pen	Total	Ball Point	Jell Pen	Ink Pen	Total
Feb	1/Pack	2	3	3	**8**	3	5	7	**15**	0	0	1	**1**	0	3	1	**4**	5	11	12	**28**
	5/Pack	5	4	1	**10**	4	2	2	**8**	1	1	2	**4**	1	1	1	**3**	11	8	6	**25**
	Total	**7**	**7**	**4**	**18**	**7**	**7**	**9**	**23**	**1**	**1**	**3**	**5**	**1**	**4**	**2**	**7**	**16**	**19**	**18**	**53**
Mar	1/Pack	3	4	3	**10**	3	4	5	**12**	0	0	1	**1**	0	0	0	**0**	6	8	9	**23**
	5/Pack	2	2	4	**8**	4	4	6	**14**	0	1	2	**3**	1	0	1	**2**	7	7	13	**27**
	Total	**5**	**6**	**7**	**18**	**7**	**8**	**11**	**26**	**0**	**1**	**3**	**4**	**1**	**0**	**1**	**2**	**13**	**15**	**22**	**50**
Apr	1/Pack	3	4	5	**12**	2	2	5	**9**	2	0	0	**2**	2	0	0	**2**	9	6	10	**25**
	5/Pack	1	2	5	**8**	4	3	2	**9**	1	0	2	**3**	1	0	1	**2**	7	5	10	**22**
	Total	**4**	**6**	**10**	**20**	**6**	**5**	**7**	**18**	**3**	**0**	**2**	**5**	**3**	**0**	**1**	**4**	**16**	**11**	**20**	**47**
May	1/Pack	3	5	4	**12**	2	3	2	**7**	1	1	0	**2**	1	1	0	**2**	7	10	6	**23**
	5/Pack	3	3	3	**9**	3	4	2	**9**	2	0	2	**4**	2	0	0	**2**	10	7	7	**24**
	Total	**6**	**8**	**7**	**21**	**5**	**7**	**4**	**16**	**3**	**1**	**2**	**6**	**3**	**1**	**0**	**4**	**17**	**17**	**13**	**47**
Jun	1/Pack	5	4	5	**14**	5	6	4	**15**	1	3	2	**6**	0	1	1	**2**	11	14	12	**37**
	5/Pack	2	3	5	**10**	3	4	1	**8**	2	2	1	**5**	1	1	1	**3**	8	10	8	**26**
	Total	**7**	**7**	**10**	**24**	**8**	**10**	**5**	**23**	**3**	**5**	**3**	**11**	**1**	**2**	**2**	**5**	**19**	**24**	**20**	**63**
Total	1/Pack	16	20	20	**56**	15	20	23	**58**	4	4	4	**12**	3	5	2	**10**	38	49	49	**136**
	5/Pack	13	14	18	**45**	18	17	13	**48**	6	4	9	**19**	6	2	4	**12**	43	37	44	**124**
	Total	**29**	**34**	**38**	**101**	**33**	**37**	**36**	**106**	**10**	**8**	**13**	**31**	**9**	**7**	**6**	**22**	**81**	**86**	**93**	**260**

Table 3.7: Detailed classification of pens

3.5 Parts of a Table

The structure and parts of a table is depended on the nature of the data and purpose of our study. However in general a statistical table is divided into eight parts, which are explained below.

a) **Table Number**

Each table should have a table number to identify the table and for further references. The table number is usually given on the top of the table or on the left hand side along with the title of the table. Sometimes table number is given in the bottom centre of the table.

b) **Title of the table**

The title of a table is generally given at the top or bottom of the table in the centre, along with the table number or just below the table number. A good title will contain a brief statement about the nature, geographical

region, time span, etc. to which the data relate. This gives fairly good information about the content.

c) **Captions**

Caption refers to the headings of the vertical columns. A caption generally has a main heading and a number of small sub headings. A caption should be written in unambiguous terms and placed at the middle of the column. The unit of measurement may also be mentioned along with the caption.

d) **Stubs**

Stubs refer to the headings of the horizontal rows and they are usually written in the left hand side of the rows. The number of stubs is depending upon the nature of the data.

e) **Body**

The most vital part of the table is body of the table. It contains statistical data arranged according to captions and stubs. Usually the content of the body is numerical information in the cells, mostly frequencies or observations. The column totals, row totals and grand total should also be mentioned in the body of the table.

f) **Head Note**

Head notes are usually given on the right top corner of the table, just below the title. It refers to the data contained in the body of the table which is not mentioned in the title. For example the units of measurements are generally written as head notes like in lakhs or in kilogram, etc. Generally it is put in a bracket.

g) **Foot Note**

Foot notes are given below the table and are meant to clarify anything which is not clear from the title, Caption, Stubs, etc.

h) **Source Note**

The source of the data given in the table is to be disclosed as a rule. This enables one to verify all facts about the data. It is given below the foot note.

All the parts may not be compulsory for a table.

Table Number:

Title:

Head Note:

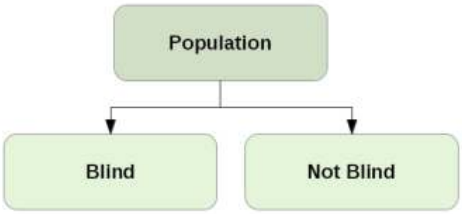
Stubs	Caption
	B_1 B_2 ..... B_n
A_1 A_2 .. .. .. A_m	BODY

Foot Note:

Source Note:

3.6 Classification according to Attributes

In qualitative analysis, data is classified on the basis of descriptive characteristics or on the basis of attributes like sex, literacy, religion, education etc. which cannot be numerically measured. We can only find out the presence or absence of a particular attribute in an individual. If we are dealing with the problem of blindness, we can only find out whether the individual is blind or not blind. We cannot measure the blindness. If we are dealing with the problem of deafness, we can only find out whether the individual is deaf or not. We cannot measure the deafness. In such cases the attribute divides the population into two parts. One in which the attribute is present and the other in which the attribute is not present. In the case of blindness there are only two classes, that is, those who are blind and those who are not blind. These classes are mutually exclusive (disjoint), so that those who are blind cannot come into the category of those who are not blind. Such a classification is called **Dichotomy**

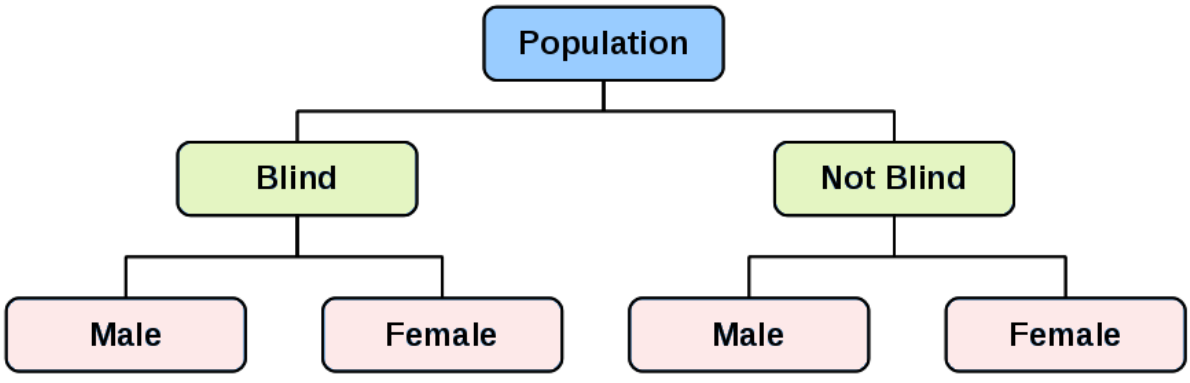


or **Two-fold Classification.**

The following table is an example of two-fold classification

Course	Number of students		
	Passed	Failed	Total
Humanities	29	18	47
Commerce	83	16	99
Science	96	4	100
Total	208	38	246

If we consider another attribute male or female together with the above attribute blindness there will be four classes. Males who are blind, males who are not blind, females who are blind and females who are not blind. The classification can be further extended if we have a third attribute, say nationality. Then we will get $2^3 = 8$ mutually exclusive classes. Such classification in which more than one attribute is taken into account is called **Manifold Classification.**



The following table is an example of manifold classification

Course		Number of students		
		Passed	Failed	Total
Humanities	Boys	11	16	47
	Girls	18	2	
	Total	29	18	
Commerce	Boys	32	11	99
	Girls	51	5	
	Total	83	16	
Science	Boys	42	3	100
	Girls	54	1	
	Total	96	4	
Total	Boys	85	30	246
	Girls	123	8	
	Total	208	38	

Illustration 3.1

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

Town X: Females were 40%, Total coffee drinkers were 45% and male non coffee drinkers were 20%.

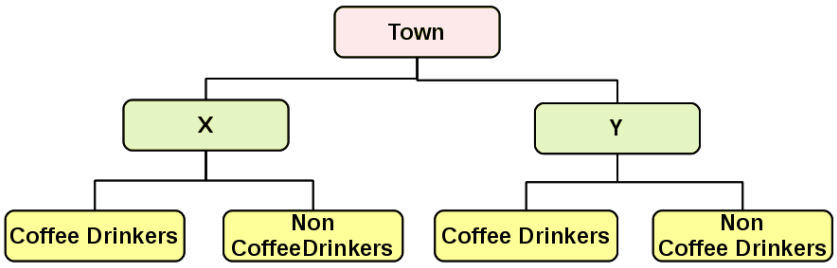
Town Y: Males were 55%, male non coffee drinkers were 30% and female non coffee drinkers were 15%.

a) Draw the classification sketch of the above data

b) Present the above data in tabular form and complete the table.

Solution:

a).



b).

Persons	*Town X*			*Town Y*		
	Male	*Female*	*Total*	*Male*	*Female*	*Total*
Coffee Drinkers	*40*	*5*	*45*	*25*	*30*	*55*
Non Coffee Drinkers	*20*	*35*	*55*	*30*	*15*	*45*
Total	*60*	*40*	*100*	*55*	*45*	*100*

Frequency

The number of repetitions of a particular observation in a series is called **Frequency of the observation.** For example if in a class 15 students scored 40 marks, then the frequency of the mark 40 is 15.

3.7 Frequency Tables

The series of observations in which items are listed individually is called **Raw data** or **Individual series.** Usually it will be in an ungrouped format. The number of family members of 60 students in a class is simply listed below. It is an example of individual series or ungrouped data.

4	3	3	2	3	4	5	7	10	6	3	9	6	3	4	5	5	5	8	7
8	6	9	7	2	7	6	5	11	4	4	3	5	6	6	7	6	5	7	6
4	5	6	10	8	6	5	4	5	6	4	5	5	6	5	4	6	7	5	8

The above series is arranged in the order of serial numbers. It means no organisation is made on the data. The first step in descriptive statistics is organisation of data. The above data can be arranged in ascending or descending order of the observations as you learnt in your lower classes. Try it yourself.

The arrangement of data in ascending or descending order is more helpful for a statistical study than the individual series as it is. Though it is not suitable for our statistical analysis and graphical presentation. Thus we can arrange the observations and its frequencies in a table to make the data more compact and

useful. Such tables are called frequency tables. The frequency table prepared based on the previous data of the number of family members is shown below.

Number of family members	Tally Mark	Frequency
2	II	2
3	𝍸 I	6
4	𝍸 IIII	9
5	𝍸 𝍸 IIII	14
6	𝍸 𝍸 III	13
7	𝍸 II	7
8	IIII	4
9	II	2
10	II	2
11	I	1
Total		60

Table 3.8:

Here the number of family members is a discrete variable and exact measurements of units are clearly shown in the first column. This type of frequency table is called **Discrete Series** , **Discrete Frequency Table** or **Frequency Array.**

A Discrete Frequency Table is that series in which data are presented in a way that exact measurements of units are clearly shown.

Know your progress

The marks obtained by 48 students of a class in a class test on statistics with maximum score 10 are given below. Construct a frequency table.

5 10 5 6 6 6 7 7 2 7 8 8 7 8 10 6 8 7
3 5 5 5 5 7 8 5 2 6 8 5 1 7 6 9 6 7
4 8 4 6 4 3 6 7 9 2 7 10

When the body weight of 60 students is collected, it will be very difficult to tabulate as before, since the body weight is a continuous variable and may take infinitely many possible values. Following are the body weights of the students studying in graduate courses. The measures are in kilograms .

51.1, 53, 66.6, 64.2, 63.9, 54.1, 69.7, 56.4, 52.9, 52.3, 59.5, 57.1, 45.5, 73.4, 41.4, 57.5, 62.4, 43.4, 38.4, 69.5, 48.5, 48.9, 45.2, 69.7, 36.4, 36.4, 58.8, 64.5, 66.4, 47.8, 54.7, 49, 49.9, 46.2, 53.2, 56.2, 32.8, 59.4, 31.8, 53.1, 50.2, 57.3, 30.1, 46.1, 41.8, 60.5, 40.2, 45.4, 47.1, 44.1, 52.6, 64.1, 38.3, 79.2, 48.5, 53.4, 51.6, 51.3, 63.5, 37.8 Clearly the body weight varies from 30.7 to 78.3. Here we cannot tabulate by taking the exact values in the first column and its corresponding frequencies in the second column. Instead of that we can find the frequency of observations falling in an interval of weights like 30-35, 35-40, 40-45, etc. It is shown in the table.

Weight	Tally Mark	Frequency
30-35	III	3
35-40	𝍸	5
40-45	𝍸	5
45-50	𝍸 𝍸 II	12
50-55	𝍸 𝍸 III	13
55-60	𝍸 III	8
60-65	𝍸 II	7
65-70	𝍸	5
70-75	I	1
75-80	I	1
Total		60

Table 3.9:

Here the intervals 30-35, 35-40, 40-45, etc. are usually called classes. 30 is the lower limit and 35 is the upper limit of the class 30-35. The difference between the lower limit of a class and lower limit of the next class or difference between the upper limit of a class and upper limit of the previous class is called **Class width.** Here 35-30=5 is the class width of first class. Frequency tables with classes and corresponding frequencies are known as **Continuous Frequency Distribution** or simply **Frequency Distribution.** Continuous Frequency Distribution are used for both discrete and continuous data by assuming the data is continuous.

Open Ended Classes:

If the lower limit of the first class or upper limit of the last class is not specified, then it is called open ended class. Sometimes we may not be able to determine the lower limit of the first class or upper limit of the last class or both. In such situations we can use open ended classes.

Illustration 3.2

Q) The marks obtained by 50 students in an examination with maximum score 100 are given below. Construct a frequency table to the data.

32	45	67	44	25	53	54	37	42	48	46	59	72
36	46	58	65	68	78	57	68	43	23	45	58	12
71	36	55	61	58	23	34	25	17	61	70	56	43
59	40	14	39	60	41	56	65	33	29	58		

Here the marks obtained by the students is a discrete data. The minimum and maximum marks are 12 and 78 respectively. So representing individual marks and its frequencies is not practical in this situation. So we consider the data is continuous and constructs a continuous frequency distribution as shown below.

Mark	Tally Mark	Frequency
10-19	III	3
20-29	𝍸	5
30-39	𝍸 II	7
40-49	𝍸 𝍸 I	11
50-59	𝍸 𝍸 II	12
60-69	𝍸 III	8
70-79	IIII	4
Total		50

Table 3.10:

Know your progress

The temperature in a village is observed at randomly selected days during a year. Tabulate the data

33.98, 29.07, 29.90, 34.15, 28.98, 36.02, 32.77, 33.11, 27.55, 34.77, 39.76, 30.65, 29.15, 27.82, 33.67, 34.31, 33.18, 34.65, 34.76, 24.03, 27.14, 37.43, 23.88, 34.27, 27.25, 25.57, 35.86, 34.98, 36.75, 30.28, 33.10, 25.67, 29.97, 34.17, 27.44, 27.61, 35.47, 29.28, 28.56, 27.46, 29.12, 34.69, 25.07, 34.05, 35.00, 30.28, 37.65, 35.08, 33.03, 33.84, 34.65, 33.86, 36.76, 26.04, 35.20, 29.89, 27.14, 29.54, 32.25, 28.69, 32.23, 25.41, 27.27, 27.44, 37.76, 29.32, 31.11, 27.02, 36.93, 32.54, 37.31, 34.87, 31.61, 37.01, 23.08, 34.21, 31.48, 37.75, 24.75, 24.19, 33.69, 32.86, 28.46, 27.77, 37.14, 30.24, 24.45, 31.46, 29.80, 40.42

Inclusive and Exclusive Classes

An inclusive class includes all items between the lower limit and upper limit, including the limits. But the upper limits are not included in the exclusive classes. In an exclusive series the upper limit of one class is the lower limit of the next class. But in an inclusive series the upper limit of a class does not repeat itself as the lower limit of the next class. Usually inclusive classes are used only for the discrete data. Exclusive classes can be used for both discrete and continuous data. Table - 3.9 is an example of exclusive classes and Table - 10 is an example of inclusive classes.

An inclusive class can be converted to exclusive class by modifying its class limits. Let us convert the class intervals given in Table - 10 to exclusive classes as shown in the table.

Mark	9.5 - 19.5	19.5 - 29.5	29.5 - 39.5	39.5 - 49.5	49.5 - 59.5	59.5 - 69.5	69.5 - 79.5	Total
Frequency	3	5	7	11	12	8	4	50

Table 3.11:

Unequal Class Intervals:
The class widths need not be equal for all the classes. But it is better to use equal class intervals.

Note:-

The total number of observations in an individual series is usually represented by n and the total frequency of a frequency table is usually represented by N.

Relative Frequency Tables

The ratio of frequency to the total frequency is called **Relative Frequency.**

$$\text{Relative Frequecy} = \frac{frequency}{\text{Total frequency}} = \frac{f}{N}$$

The table representing relative frequency of observations (or classes) is called **Relative frequency tables.** The relative frequency table prepared from Table - 3.8 is shown below.

Number of family members	Frequency	Relative Frequency
2	2	2/60=0.03
3	6	6/60=0.10
4	9	9/60=0.15
5	14	14/60=0.23
6	13	13/60=0.22
7	7	7/60=0.12
8	4	4/60=0.07
9	2	2/60=0.03
10	2	2/60=0.03
11	1	1/60=0.02
Total	60	60/60=1

Table 3.12:

Know your progress

- Prepare a relative frequency table from Table 3.9
- A special diet is practiced for three months by 35 working women in an IT company to reduce their body weight. The loss of weight (in Kgs)occured is given below. Prepare a relative frequency table to the data.

4	2	3	2	6	4	3	2	7	3
4	4	3	3	2	1	5	2	1	5
2	3	2	5	4	0	3	2	1	4
3	5	1	1	3					

Percentage Frequency Tables

In a percentage frequency table the percentage of the total frequency corresponding to the observations(or classes) are shown. The percentage frequencies are obtained by the following relationship

$$\text{Percentage Frequency} = \frac{frequency}{\text{Total frequency}} \times 100 = \frac{f}{N} \times 100$$

Note:

$$\text{Percentage Frequency} = \text{Relative Frequency} \times 100$$

We can prepare a percentage frequency table from Table -3.9 as shown below.

Weight	Frequency	Percentage Frequency
30 - 35	3	$\frac{3}{60} \times 100 = 5.0$
35 - 40	5	$\frac{5}{60} \times 100 = 8.33$
40 - 45	5	$\frac{5}{60} \times 100 = 8.33$
45 - 50	12	$\frac{12}{60} \times 100 = 20.0$
50 - 55	13	$\frac{13}{60} \times 100 = 21.67$
55 - 60	8	$\frac{18}{60} \times 100 = 13.33$
60 - 65	7	$\frac{17}{60} \times 100 = 11.67$
65 - 70	5	$\frac{3}{60} \times 100 = 5.0$
70 - 75	1	$\frac{1}{60} \times 100 = 1.67$
75 - 80	1	$\frac{1}{60} \times 100 = 1.67$
Total	60	$\frac{60}{60} \times 100 = 100$

Table 3.13:

Know your progress

1. Prepare a percentage frequency table to the data contained in Table - 3.8.
2. A researcher decided to study the impact of drug abuses on the road accidents. He collected the number of accidents reported per day at the police station, due to the vehicles driven by drunkards and drug abusers. Data is given below. Construct a percentage frequency table.

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

The sum of relative frequencies is always 1 and the sum of percentage frequencies is 100.

Cumulative Frequency Tables

The number of observations less than or equal to a particular value is called **Less than cumulative frequency** of that value. Similarly the number of observations greater than or equal to a particular value is called **Greater than cumulative frequency** or **More than cumulative frequency** of that value. Consider the Table-3.8 The less than cumulative frequency of the observation 3 is 8, since 8 families have less than or equal to 3 family members. The less than cumulative frequency of the observation 5 is 31. The greater than cumulative frequency of the observation 9 is 5, since only 5 families have 9 or more than 9 family members. The tables representing cumulative frequencies are called Cumulative Frequency Tables. The preparation of a Less than Cumulative Frequency Table is demonstrated below.

Number of family members	Frequency	Less than Cumulative Frequency
2	2	2 + 0=2
3	6	6 + 2=8
4	9	9 + 8=17
5	14	14 + 17=31
6	13	13 + 31=44
7	7	7 + 44=51
8	4	4 + 51=55
9	2	2 + 55=57
10	2	2 + 57=59
11	1	1 + 59=60
Total	60	

Table 3.14:

Illustration 3.3

Construct a less than cumulative frequency table to the Table - 3.9

Weight	Upper Bounds	Frequency	Less than Cumulative Frequency
30 - 35	35	3	3
35 - 40	40	5	8
40 - 45	45	5	13
45 - 50	50	12	25
50 - 55	55	13	38
55 - 60	60	8	46
60 - 65	65	7	53
65 - 70	70	5	58
70 - 75	75	1	59
75 - 80	80	1	60
Total	-	60	-

Table 3.15:

Illustration 3.4

Construct a greater than cumulative frequency table for Table - 3.8.

Number of family members	Frequency	Greater than Cumulative Frequency
2	2	60
3	6	60-2=58
4	9	58-6=52
5	14	52-9=43
6	13	43-14=29
7	7	29-13=16
8	4	16-7=9
9	2	9-4=5
10	2	5-2=3
11	1	3-2=1
Total	60	

Table 3.16:

Illustration 3.5

Construct a Greater than Cumulative Frequency table to the Table - 3.9

Weight	Lower Bounds	Frequency	Greater than Cumulative Frequency
30 - 35	30	3	60
35 - 40	35	5	57
40 - 45	40	5	52
45 - 50	45	12	47
50 - 55	50	13	35
55 - 60	55	8	22
60 - 65	60	7	14
65 - 70	65	5	7
70 - 75	70	1	2
75 - 80	75	1	1
Total	-	60	-

Table 3.17:

Know your progress

1. The departure time of the train Mangala-Lakshadweep Express at the Kozhikode Railway Station is observed for three months. The following frequency table shows the number of minutes late by the train on these days. Construct a less than cumulative frequency table to the data. Determine the number of days in which the train was not late for more than five minutes.

Late (in Minutes)	0	1	2	3	4	5	6	7	8	9	10
No. of days	44	13	9	5	5	3	4	2	3	1	1

2. Prepare a greater than cumulative frequency table to the frequency table given in the above table and obtain the number of days in which train was late by at least two minutes.
3. The Compact Fluorescent Lamps (CFL) produced by a company is inspected to study about the life span of the lamps and the following frequency table is prepared.

Life in Hours ('00)	0-5	5-10	10-20	20-25	25-30	30-35	35-40
No. of CFLs	3	16	160	323	80	17	1

Compute the less than cumulative frequencies and obtain the number of CFLs damaged within 2500 hours.

4. Calculate the greater than cumulative frequencies to the frequency table given in the above table. Estimate the number of CFLs illuminated for at least 2000 hours.

3.8 Bivariate Frequency Distribution

If only one characteristic of the sampling units is measured for the study, it is called **Univariate Data.** If two characteristics are measured simultaneously from each unit, it is known as **Bivariate Data.** Similarly data containing measurements of more than two characteristics of each unit is called **Multivariate Data.** For example if only the height of the students is measured for the study, it is Univariate Data. Usually we represent it by x, y ,z, etc.

Height (in inches) : 52, 51, 57, 62, 68

If we measure the height and weight of each student for a study, it is a Bivariate Data. We represent it by (x, y) or (x_1, y_1) where first variable is the height and the second variable is the weight.

Height in inches and weight in Kg:

(52, 45), (51, 62), (57, 58), (62, 70), (68, 73)

The same data can also be represented as,

Height (x)	:	52	51	57	62	68
Weight (y)	:	45	62	58	70	73

The frequency distribution table of a Bivariate Data is called **Bivariate Frequency Table.**

Illustration 3.6

The heights (in inches) and weights (in Kg) of 40 students in a class are given. Construct a frequency table to the data.

(59, 60)	(62, 70)	(67, 65)	(72, 80)	(71, 58)	(68, 73)	(54, 42)	(59, 55)
(55, 53)	(60, 55)	(54, 42)	(65, 72)	(62, 71)	(69, 82)	(65, 47)	(68, 64)
(65, 74)	(64, 84)	(67, 69)	(72, 75)	(64, 65)	(71, 78)	(70, 74)	(67, 62)
(60, 57)	(59, 48)	(67, 71)	(60, 65)	(56, 49)	(63, 62)	(71, 69)	(58, 53)
(67, 62)	(57, 62)	(57, 55)	(62, 64)	(66, 73)	(66, 53)	(69, 72)	(56, 44)

Solution:

Height / Weight	54 - 58	59 - 63	64 - 68	69 - 73	Total
36-45	3				3
46-55	3	4	2		9
56-65	1	5	5	1	12
66-75		2	6	4	12
76-85			1	3	4
Total	7	11	14	8	40

The frequencies given in the last row and last column are called marginal frequencies. Frequencies in the last row are the marginal frequency of heights and those in the last column are the marginal frequency of weights. The marginal frequency tables of heights and weights are shown below.

Marginal frequency table of Height

Height	54-58	59-63	64-68	69-73	Total
Total	7	11	14	8	40

Marginal frequency table of weight

Weight	36-45	46-55	56-65	66-75	76-85	Total
Frequency	3	9	12	12	4	40

Know your progress

The amount spent for advertisement (In Lakh Rupees) and the profits (in Crore Rupees) of a company is observed for different months and the following data is obtained. Prepare a bivariate frequency table to the data. Also obtain the marginal frequency

tables.

(1, 13)	(3, 15)	(2, 16)	(6, 17)	(13, 23)	(12, 24)	(17, 25)	(11, 19)
(15, 23)	(18, 27)	(19, 28)	(18, 24)	(16, 16)	(13, 27)	(7, 15)	(8, 19)
(3, 14)	(4, 17)	(7, 21)	(19, 33)	(16, 35)	(17, 27)	(10, 21)	(11, 17)
(10, 16)	(13, 19)	(11, 18)	(12, 19)	(14, 21)	(16, 25)	(19, 24)	(20, 23)

3.9 Advantages of Tabulation

Following are some of the advantages of tabulation of data

1. Tables consolidate the data
2. A table presents the data in such a simple a3333nd systematic form that one can understand the values associated with the variable and / or attributes.
3. They reveal the association of a variable or attribute with the other.
4. Comparison of one factor with the another becomes easy and reliable.
5. Diagramatic and graphical presentation of data becomes simple and accurate with the help of tables.
6. Tables portray more information in less space.
7. Tables are time savers. Information can easily be gathered from a table.
8. Tables are the basis of statistical calculations and analysis of data.
9. Tables make interpretation of data easier and better as compared to raw data.

Let us sum up

In this Chapter we have discussed the need, objectives and advantages of Classification and Tabulation of data in statistics. It is the process of making the data ready for the analysis. We can classify the data under different ways called Qualitative, Quantitative,

Chronological and Geographical classifications. The process of arranging the classified data in columns and rows is called Tabulation. The tables representing frequencies of observations are called Frequency Tables. Relative frequency table, percentage frequency table and cumulative frequency table are some of the different types of statistical tables.

Learning outcomes

After transaction of this unit, the learner:-

- identifies the need of Classification and Tabulation.
- recognises the different methods of Classification and Tabulation.
- classifies raw data to useful information
- constructs frequency tables.
- interprets the data.

Evaluation Items

Choose the correct answer

1. The process of grouping the data according to their characteristics is called .....
   a) Organisation of data b) Classification of data
   c) Tabulation of data d) Presentation of data

2. Classification based on time is called .....
   a) Qualitative Classification b) Quantitative Classification
   c) Chronological Classification d) Geographical Classification

3. Tabulation of data saves .....
   a) Time b) Energy c) Space d) All the above

4. Classification of data to two disjoint classes is called .....
   a) One - Fold Classification b) Two - Fold Classification
   c) Many - Fold Classification d) None of these

5. The number of observations greater than a particular value is the .....
   a) Cumulative frequency b) Less than Cumulative frequency
   c) Relative frequency d) More than Cumulative frequency

6. An exclusive class excludes .....
   a) Lower limit b) Upper limit
   c) Both the limits d) Either of the limits

7. The less than cumulative frequencies of 7^{th} and 8^{th} observations are 32 and 84 respectively. then the frequency of 8^{th} observation is .....
   a) 52 b) 116 c) 42 d) 64

8. Relative frequency = ...
   a) $f \times N$ b) $\frac{f}{N}$ c) $f - N$ d) $f + N$

9. Percentage frequency = ...
   a) $f \times 100$ b) $\frac{f}{100}$
   c) $RelativeFrequency \times 100$ d) $\frac{RelativeFrequency}{100}$

10. A relative frequency table is prepared for 120 observations. The relative frequency of 5^{th} item is 0.1. Then the frequency of 5^{th} item is .....
    a) 100 b) 50 c) 12 d)10

11. Explain the objectives of classification and tabulation of data?

12. Make a comparison between Classification and Tabulation?

13. What are the different types of Classifications?

14. State the advantages of Tabulation?

15. Explain the term dichotomy?

16. There are three streams in a Higher Secondary School, Humanities, Commerce and Science. Boys and Girls are studying with three different second languages, Malayalam, Hindi and Urdu.

a) Draw a Classification Sketch.

b) Draw the skeleton of the Table

17. 1200 employees are working in a company as Managers, Executives and Clerks. Draw the Classification Sketch of the data of the employees according to the attributes Designation, Nationality (Indians or Non Indians), Gender, Age Group, etc Also tabulate the data by giving arbitrary frequencies.

18. 110000 students are registered for different courses in a university during 2000 - 2002. 63000 students registered for Arts stream and 30000 students registered for Science stream. 17000 students registered for Arts, 10000 students for Science and 4000 for Law in the year 2000. 25000 students among the 43000 students registered at the university in the year 2002 were in the Arts Stream. The students registered for Law in 2002 was one thousand more than that in 2001. Tabulate the data.

19. The number of flowers obtained from different plants in a garden in a day is given below.

3	9	7	6	0	10	6	8	5	4
9	8	4	3	6	7	1	2	5	7
6	7	8	4	6	5	6	8	3	6
7	2	9	7	6	3	5	0	3	8
2	8	5	9	7	4	5	8	6	7
10	9	7	6	7	7	8	6	5	6
6	7	7	7	8	9	6	5	7	7

a) Construct a frequency table to the number of flowers per plant.

b) Prepare a less than cumulative frequency table.

c) Prepare a greater than cumulative frequency table.

d) Obtain a relative frequency table

e) Obtain the percentage frequency table.

20. The weights of six months old rabbits in a farm are measured in grams and the following data is obtained.

862 816 971 932 877 958 854 928 802 950 946 928
837 952 855 812 836 958 933 946 902 925 941 882
900 861 907 832 917 858 888 868 860 827 946 886
976 889 937 806 944 916 951 951 855 940 890 828
802 822 926 808 916 914 943 828 954 892 844 938
809 882 918 928 979 830 935 840 809 919 873 915
865 901 894 863 870 862 814 913 861 875 971 922
906 829 938 969 828 910 972 876 961 930 949 864
864 955 935 907 870 980 839 940 843 885 938 920
801 873 877 847 856 842 921 958 906 914 878 829
898 898 852 925 896 867 939 975 849 917 922 904
852 848 927 820 864 952 911 975 963 930 802 823
976 890 816 856 841 906 867 929 921 929 896 965
809 967 928 943 816 895 813 804 880 970 847 972

a) Construct a frequency table by taking a suitable class interval.
b) Prepare a less than cumulative frequency table.
c) Prepare a greater than cumulative frequency table.
d) Obtain a relative frequency table.
e) Obtain the percentage frequency table.

21. Following table shows the cumulative frequency table of profit and number of companies.

Profit less than (in 000s)	15	30	45	60	75	90	105	120	135
No. of Companies	3	10	28	53	73	85	91	96	98

Construct the frequency table.

22. The students in a Statistics class were trying to study the heights of participants in a sports meet. They collected the height of 20 participants, as displayed in the table.

Height (x)	49	53	54	55	66	70	80
No of participants (f)	1	2	4	5	3	2	1

Construct

a) Less than cumulative frequency table
b) Greater than cumulative frequency table
d) Relative frequency table
e) Percentage frequency table

23. The table below show the age of 55 patients selected to study the effectiveness of a particular medicine.

Age	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No of Patients	5	7	17	12	5	2	7

Construct

a) Less than cumulative frequency table
b) Greater than cumulative frequency table
d) Relative frequency table
e) Percentage frequency table

24. The frequency distribution below represents the weights in kg of parcels carried by a small logistic company.

Weight	No of Parcels
10.0 - 10.9	2
11.0 - 11.9	3
12.0 - 12.9	5
13.0 - 13.9	8
14.0 - 14.9	12
15.0 - 15.9	15
16.0 - 16.9	13
17.0 - 17.9	11
18.0 - 18.9	6
19.0 - 19.9	2

Prepare a frequency table with exclusive classes.

25. Thirty automobiles were tested for fuel efficiency and the following table is obtained.

Mileage (in KMs)	Below 13	Below 18	Below 23	Below 28	Below 33
No: of Vehicles	3	8	23	28	30

Construct a frequency table.

26. Following table shows the relative frequencies of 500 workers in a company with respect to their daily wages (in hundred rupees).

Daily wages	250 - 300	300 - 350	350 - 400	400 - 450	450 - 500	500 - 550	550 - 600
Relative freq.	0.05	0.1	0.3	0.225	0.2	0.1	0.025

Construct a frequency table.

Answers:

1) b 2) c 3) d 4) b 5) d
6) b 7) a 8) b 9) c 10) c