Statistics

CLASSIFICATION AND TABULATION OF DATA

Source: Business Statistics - 2006, R S Bhardwaj, Excel Books, New Delhi

CHAPTER 6

CLASSIFICATION AND TABULATION OF DATA

The collected data are a complex and unorganised mass of figures which is very difficult to analyse and interpret. Therefore, it becomes necessary to organise this so that it becomes easier to grasp its broad features. Further, in order to apply the tools of analysis and interpretation, it is essential that the data are arranged in a definite form. This task is accomplished by the process of classification and tabulation.

6.1 CLASSIFICATION

Classification is the process of arranging the available data into various homogeneous classes and subclasses according to some common characteristics or objective of investigation. In the words of L.R. Connor, *"Classification is the process of arranging things (either actually or notionally) in the groups or classes according to the unity of attributes that may subsist amongst a diversity of individuals."* The chief characteristics of any classification are :

1. The collected data are arranged into homogeneous groups.
2. The basis of classification is the similarity of characteristics or features inherent in the collected data.
3. Classification of data signifies unity in diversity.
4. Classification of data may be actual or notional.
5. Classification of data may be according to certain measurable or non-measurable characteristics or according to some combination of both.

Objectives of Classification

The main objectives of any classification are :

1. To present a mass of data in a condensed form.
2. To highlight the points of similarity and dissimilarity.
3. To bring out the relationship between variables.
4. To highlight the effect of one variable by eliminating the effect of others.
5. To facilitate comparison.
6. To prepare data for tabulation and analysis.

Requisites of a Good Classification

A good classification must possess the following features :

1. *Unambiguous* : The classification should not lead to any ambiguity or confusion.
2. *Exhaustive* : A classification is said to be exhaustive if there is no item that cannot be allotted a class.
3. *Mutually Exclusive* : Different classes are said to be mutually exclusive if they are non-overlapping. When a classification is mutually exclusive, each item of the data can be placed only in one of the classes.
4. *Flexibility* : A good classification should be capable of being adjusted according to the changed situations and conditions.
5. *Stability* : The principle of classification, once decided, should remain same throughout the analysis, otherwise it will not be possible to get meaningful

results. In the absence of stability, the results of the same type of investigation at different time periods may not be comparable.

6. *Suitability* : The classification should be suitable to the objective(s) of investigation.
7. *Homogeneity* : A classification is said to be homogeneous if similar items are placed in a class.
8. *Revealing* : A classification is said to be revealing if it brings out essential features of the collected data. This can be done by selecting a suitable number of classes. Making few classes means over summarisation while large number classes fail to reveal any pattern of behaviour of the variable.

6.1.1 Types of classification

The nature of classification depends upon the purpose and objective of investigation. The following are some very common types of classification :

1. Geographical (or spatial) classification
2. Chronological classification
3. Conditional classification
4. Qualitative classification
5. Quantitative classification

1. Geographical (or spatial) classification

When the data are classified according to geographical location or region, it is called a geographical classification. For example, the Statewise Net Domestic Product for 1984-85 at current prices can be shown as below :

Statewise Net Domestic Product, 1984-85 (In Rs million, at current Prices)

1. *Andhra Pradesh*	116945	13. *Manipur*	3462
2. *Arunachal Pradesh*	1605	14. *Meghalaya*	2702
3. *Assam*	45544	15. *Mizoram*	1026
4. *Bihar*	114140	16. *Nagaland*	2276
5. *Gujrat*	114434	17. *Orissa*	47413
6. *Haryana*	46121	18. *Punjab*	74392
7. *Himachal Pradesh*	10161	19. *Rajasthan*	72862
8. *Jammu and Kashmir*	13786	20. *Sikkim*	958
9. *Karnataka*	87906	21. *Tamil Nadu*	111625
10. *Kerala*	57136	22. *Tripura*	4376
11. *Madhya Pradesh*	96500	23. *Uttar Pradesh*	213266
12. *Maharashtra*	215545	24. *West Bengal*	150994

For the purpose of immediate location or comparison of the data, it is necessary that it should be presented either in alphabetical or in ascending (or descending) order of the figures.

2.Choronological classification

When the data are classified on the basis of its time of occurrence, it is called a chronological classification. Various time series such as ; National Income figures (annual), annual output of wheat, monthly expenditure of a household, daily consumption of milk, *etc.*, are some examples of chronological classification.

3. Conditional classification

When the data are classified according to certain conditions, other than geographical or chronological, it is called a conditional classification. An example of such classification is given below :

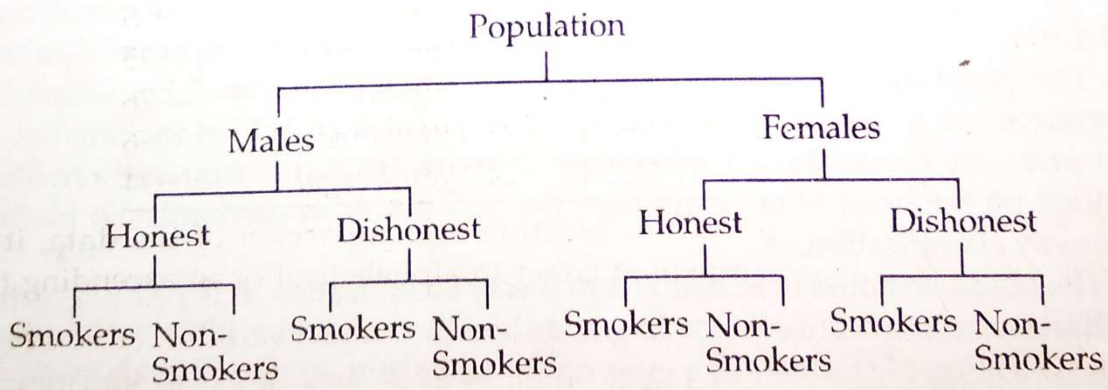
Categorised Classification of Public Expenditure, 1986-87.

Category	Amount (Rs Crores)	Share (Percentage)
1. General	9464	12.9
2. Defence	10854	14.8
3. Education	10202	13.9
4. Health	2536	3.5
5. Housing, Welfare and Social Security	7002	9.5
6. Economic Services	32411	44.2
7. Other Services	902	1.2
Total	73371	100.0

4. Qualitative classification or classification according to attributes

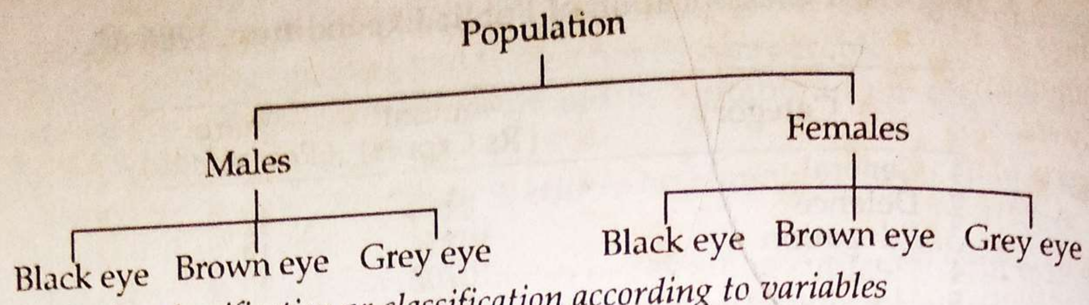
When the characteristics of the data are non-measurable, it is called a qualitative data. The examples of non-measurable characteristics are sex of a person, marital status, colour, honesty, intelligence, *etc.* These characteristics are also known as attributes. When qualitative data are given, various items can be classified into two or more groups according to a characteristic. If the data are classified only into two categories according to the presence or absence of an attribute, the classification is termed as dichotomous or twofold classification. On the other hand, if the data are classified into more than two categories according to an attribute, it is called a manifold classification. For example, classification of various students of a college according to the colour of their eyes like black, brown, grey, blue, *etc.* The conditional classification, given above, is also an example of a manifold classification.

If the classification is done according to a single attribute, it is known as a one-way classification. On the other hand, the classification done according to two or more attributes is known as a two-way classification. The example of a three-way classification, where population is dichotomised according to each attribute; sex, honesty and smoking habit, is given below :



We note that there will be eight subgroups of individuals like male, honest, smokers; male, honest, nonsmokers, *etc.*

In the classification, given above, the population is dichotomised with respect to each of the three attributes. There may be situations where classification with respect to one attribute is dichotomous while it is manifold with respect to the other. A two way classification of this type is shown as :



5. *Quantitative classification or classification according to variables*

In case of quantitative data, the characteristic is measurable in terms of numbers and is termed as variable, *e.g.*, weight, height, income, the number of children in a family, the number of crime cases in a city, life of an electric bulb of a company, *etc.* A variable can take a different value corresponding to a different item of the population or universe.

Variables can be of two types (a) Discrete and (b) Continuous.

(a) Discrete Variable

A discrete variable can assume only some specific values in a given interval. For example, the number of children in a family, the number of rooms on each floor of a multistoried building, *etc.*

(b) Continuous Variable

A continuous variable can assume any value in a given interval. For example, monthly income of a worker can take any value, say, between Rs 1,000 to 2,500. The income of a worker can be Rs 1,500.25, *etc.* Similarly, the life of an electric bulb is a continuous variable that can take any value from 0 to ∞.

It must be pointed out here that, in practice, data collected on a continuous variable also look like the data of a discrete variable. This is due to the fact that measurements, done even with the finest degree of accuracy, can only be expressed in a discrete form. For example, height measured even with accuracy upto three places after decimal gives discrete values like 167.645 cms, 167.646 cms, *etc.* Similarly age, income, time, *etc.*, are continuous variables but their actual measurements are expressed in terms of discrete numbers.

In the classification according to variables, the data are classified by the values of the variables for each item. As in the case of attributes, the classification on the basis of a single variable is termed as a one-way classification. Similarly, there can be a two-way and multi-way classification of the data. For example, if the students of a class are classified on the basis of their marks in statistics, we get a one-way classification. However, if these students are simultaneously classified on the basis of marks in statistics and marks in economics, it becomes a two-way classification.

It should be noted here that in a two-way classification, it is possible to have simultaneous classification according to an attribute and a variable. For example, the classification of students of a class on the basis of their marks in statistics and on the basis of the sex of the person.

6.2 STATISTICAL SERIES

The classified data when arranged in some logical order, *e.g.*, according to the size, according to the time of occurrence or according to some other measurable or non-measurable characteristics, is known as *Statistical Series*. H. Secrist defined a statistical series as, "*A series, as used statistically, may be defined* *as things or attributes of things arranged according to some logical order."* Another definition given by L. R. Connor as, *"If the two variable quantities can be arranged side by side so that the measurable differences in the one correspond to the measurable differences in the other, the result is said to form a statistical series."*

A statistical series can be one of the following four types :

(i) Spatial Series, (ii) Conditional Series, (iii) Time Series and (iv) Qualitative or Quantitative Series

The series formed by the geographical or spatial classification is termed as spatial series. Similarly, a series formed by the conditional classification is known as the conditional series. The examples of such series are already given under their respective classification category.

Time Series

A time series is the result of chronological classification of data. In this case, various figures are arranged with reference to the time of their occurrence. For example, the data on exports of India in various years is a time series.

Year	1980	1981	1982	1983	1984	1985	1986	1987	1988
Exports (in Rs cr.)	6591	7242	8309	8810	9981	10427	11490	15741	20295

Qualitative or Quantitative Series

This type of series is obtained when the classification of data is done on the basis of qualitative or quantitative characteristics. Accordingly, we can have two types of series, namely, qualitative and quantitative series.

(a) Qualitative Series

In case of qualitative series, the number of items in each group are shown against that group. These groups are either expressed in ascending order or in descending order of the number of items in each group. The example of such a series is given below.

Distribution of Students of a College according to Sex

Sex	*Males*	Females	*Total*
No. of Students	1700	500	2200

(b) Quantitative Series

In case of quantitative series, the number of items possessing a particular value are shown against that value.

A quantitative series can be of two types :

I. Individual Series, and II. Frequency distribution.

I. Individual series

In an individual series, the names of the individuals are written against their corresponding values. For example, the list of employees of a firm and their respective salary in a particular month.

II. Frequency Distribution

A table in which the frequencies and the associated values of a variable are written side by side, is known as a frequency distribution. According to Croxton and Cowden, *"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 a group with their corresponding frequencies side by side."* A frequency distribution can be discrete or continuous depending upon whether the variable is discrete or continuous.

6.3 CONSTRUCTION OF A FREQUENCY DISTRIBUTION

6.3.1 Construction of a Discrete Frequency Distribution

A discrete frequency distribution may be ungrouped or grouped. In an ungrouped frequency distribution, various values of the variable are shown along with their corresponding frequencies. If this distribution fails to reveal any pattern, grouping of various observations become necessary. The resulting distribution is known as grouped frequency distribution of a discrete variable. Furthermore, a grouped frequency distribution is also constructed when the possible values that a variable can take are large.

(a) Ungrouped Frequency Distribution of a Discrete variable

Suppose that a survey of 150 houses was conducted and number of rooms in each house was recorded as shown below :

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

Counting of frequency using Tally Marks

The method of tally marks is used to count the number of observations or the frequency of each value of the variable. Each possible value of the variable is written in a column. For every observation, a tally mark denoted by | is noted against its corresponding value. Five observations denoted as 𝍸, *i.e.*, the fifth tally mark crosses the earlier four marks and so on. The method of tally marks is used below to determine the frequencies of various values of the variable for the data given above.

Number of Rooms (X)	Tally Marks	Frequency
1	𝍸 𝍸 𝍸 𝍸 𝍸	25
2	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 IIII	34
3	𝍸 𝍸 𝍸 𝍸 𝍸 IIII	29
4	𝍸 𝍸 𝍸 𝍸 𝍸 I	26
5	𝍸 𝍸 𝍸 IIII	19
6	𝍸 𝍸 𝍸 II	17
Total		150

In the above frequency distribution, the number of rooms 'X ' is a discrete variable which can take integral values from 1 to 6. This distribution is also known as ungrouped frequency distribution. It should be noted here that, in case of ungrouped frequency distribution, the identity of various observations is not lost, *i.e.*, it is possible to get back the original observations from the given frequency distribution.

(b) Grouped Frequency Distribution of a Discrete Variable

Consider the data on marks obtained by 50 students in statistics. The variable 'X' denoting marks obtained is a discrete variable, let the ungrouped frequency distribution of this data be as given in the following table.

Marks	*Frequency*	*Marks*	*Frequency*	*Marks*	*Frequency*
33	1	57	1	76	1
35	2	59	1	77	2
39	1	60	2	78	1
41	2	61	1	80	1
42	1	64	1	81	1
45	1	65	3	84	1
48	2	66	2	85	2
50	1	67	1	88	1
52	1	69	2	89	1
53	1	71	1	91	1
54	1	73	2	94	2
55	2	74	2	98	1

This frequency distribution does not reveal any pattern of behaviour of the variable. In order to bring the behaviour of the variable into focus, it becomes necessary to convert this into a grouped frequency distribution.

Instead of above, if the individual marks are grouped like marks between and including 30 and 39, 40 and 49, *etc.* and the respective frequencies are written against them, we get a grouped frequency distribution as shown below :

Marks between and including	*Frequency*
30 - 39	4
40 - 49	6
50 - 59	8
60 - 69	12
70 - 79	9
80 - 89	7
90 - 99	4
Total	50

The above frequency distribution is more revealing than the earlier one. It is easy to understand the behaviour of marks on the basis of this distribution. It should, however, be pointed out here that the identity of observations is lost after grouping. For example, on the basis of the above distribution we can only say that 4 students have obtained marks between and including 30 - 39, etc. Thus, it is not possible to get back the original observations from a grouped frequency distribution.

6.3.2 Construction of a Continuous Frequency Distribution

As opposed to a discrete variable, a continuous variable can take any value in an interval. Measurements like height, age, income, time, *etc.*, are some examples of a continuous variable. As mentioned earlier, when data are collected regarding these variables, it will show discreteness, which depends upon the degree of precision of the measuring instrument. Therefore, in such a situation, even if the recorded data appear to be discrete, it should be treated as continuous. Since a continuous variable can take any value in a given interval, therefore, the frequency distribution of a continuous variable is always a grouped frequency distribution.

To construct a grouped frequency distribution, the whole interval of the continuous variable, given by the difference of its largest and the smallest possible values, is divided into various mutually exclusive and exhaustive sub-intervals. These sub-intervals are termed as *class intervals*. Then, the frequency of each class interval is determined by counting the number of observations falling under it. The construction of such a distribution is explianed below :

The figures, given below, are the 90 measurements of diameter (in mm.) of a wire.

1.86, 1.58, 1.13, 1.46, 1.53, 1.65, 1.49, 1.03, 1.10, 1.36, 1.37, 1.46, 1.44, 1.46, 1.95, 1.67, 1.59, 1.35, 1.44, 1.40, 1.50, 1.41, 1.19, 1.16, 1.27, 1.21, 1.82, 1.55, 1.52, 1.42, 1.17, 1.62, 1.42, 1.22, 1.56, 1.78, 1.98, 1.31, 1.29, 1.69, 1.32, 1.68, 1.36, 1.55, 1.54, 1.67, 1.81, 1.47, 1.30, 1.33, 1.38, 1.34, 1.40, 1.37, 1.27, 1.04, 1.87, 1.45, 1.47, 1.35, 1.24, 1.48, 1.41, 1.39, 1.38, 1.47, 1.73, 1.20, 1.77, 1.25, 1.62, 1.43, 1.51, 1.60, 1.15, 1.26, 1.76, 1.66, 1.12, 1.70, 1.57, 1.75, 1.28, 1.56, 1.42, 1.09, 1.07, 1.57, 1.92, 1.48.

The following decisions are required to be taken in the construction of any frequency distribution of a continuous variable.

1. *Number of Class Intervals*

Though there is no hard and fast rule regarding the number of classes to be formed, yet their number should be neither very large nor very small. If there are too many classes, the frequency distribution appears to be too fragmented to reveal the pattern of behaviour of characteristics. Fewer classes imply that the width of the class intervals will be broad and accordingly it would include a large number of observations. As will be obvious later that in any statistical analysis, the value of a class is represented by its mid-value and hence, a class interval with broader width will be representative of a large number of observations. Thus, the magnitude of loss of information due to grouping will be large when there are small number of classes. On the other hand, if the number of observations is small or the distribution of observations is irregular, *i.e.*, not unitorm, having more number of classes might result in zero or very small frequencies of some classes, thus, revealing no pattern of behaviour. Therefore, the number of classes depends upon the nature and the number of observations. If the number of observations is large or the distribution of observations is regular, one may have more number of classes. In practice, the minimum number of classes should not be less than 5 or 6 and in any case there should not be more than 20 classes.

The approximate number of classes can also be determined by Struge's formula : $n = 1 + 3.322 \log_{10} N$, where n (rounded to the next whole number) denotes the number of classes and N denotes the total number of observations.

Based on this formula, we have

$n = 1 + 3.322\ 2.000 = 7.644$ or 8, when $N = 100$

$n = 1 + 3.322\ 2.699 = 9.966$ or 10, when $N = 500$

$n = 1 + 3.322\ 4.000 = 14.288$ or 15, when $N = 10{,}000$

$n = 1 + 3.322\ 4.699 = 16.610$ or 17, when $N = 50{,}000$

From the above calculations we may note that even the formation of 20 class intervals is very rarely needed.

For the given data on the measurement of diameter, there are 90 observations. The number of classes by the Sturge's formula are

$n = 1 + 3.322.\log_{10}90 = 7.492$ or 8

2. *Width of a Class Interval*

After determining the number of class intervals, one has to determine their width. The problem of determining the width of a class interval is closely related to the number of class intervals.

As far as possible, all the class intervals should be of equal width. However, there can be situations where it may not be possible to have equal width of all the classes. Suppose that there is a frequency distribution, having all classes of equal width, in which the pattern of behaviour of the observations is not regular, *i.e.*, there are nil or very few observations in some classes while there is concentration of observations in other classes. In such a situation, one may be compelled to have unequal class intervals in order that the frequency distribution becomes regular.

The approximate size of a class interval can be decided by the use of the following formula :

$$\text{Class Interval} = \frac{\text{Largest observation} - \text{Smallest observation}}{\text{Number of class intervals}}$$

or using notations, Class Interval $= \frac{L-S}{n}$

In the example, given above, $L = 1.98$ and $S = 1.03$ and $n = 8$.

$\therefore$ Approximate size of a class interval

$$= \frac{1.98 - 1.03}{8} = 0.1188 \text{ or } 0.12 \text{ (approx.)}$$

Before taking a final decision on the width of various class intervals, it is worthwhile to consider the following points :

(a) Normally a class interval should be a multiple of 5, because it is easy to grasp numbers like 5, 10, 15, ..., *etc.*

(b) It should be convenient to find the mid-value of a class interval.

(c) Most of the observations in a class should be uniformly distributed or concentrated around its mid-value.

(d) As far as possible, all the classes should be of equal width. A frequency distribution of equal class width is convenient to be represented diagrammatically and easy to analyse.

On the basis of above considerations, it will be more appropriate to have classes, each, with interval of 0.10 rather than 0.12. Further, the number of classes should also be revised in the light of this decision.

$$n = \frac{L-S}{\text{Class Interval}} = \frac{1.98 - 1.03}{0.10} = \frac{0.95}{0.10} = 9.5 \text{ or } 10$$

(rounded to the next whole number)

3. *Designation of Class Limits*

The class limits are the smallest and the largest observation in a class. These are respectively known as the lower limit and the upper limit of a class. For a frequency distribution, it is necessary to designate these class limits very unambiguously, because the mid-value of a class is obtained by using these limits. As will be obvious later, this mid-value will be used in all the computations about a frequency distribution and the accuracy of these computations will depend upon the proper specification of class limits. The class limits should be designated keeping the following points in mind :

(a) It is not necessary to have lower limit of the first class exactly equal to the smallest observation of the data. In fact it can be less than or equal to the smallest observation. Similarly, the upper limit of the last class can be equal to or greater than the largest observation of the data.

(b) It is convenient to have lower limit of a class either equal to zero or some multiple of 5.

The chosen class limits should be such that the observations in a class tend to concentrate around its mid-value. This will be true if the observations are uniformly distributed in a class.

The designation of class limits for various class intervals can be done in two ways : (i) Exclusive Method and (ii) Inclusive Method.

(i) Exclusive Method

In this method the upper limit of a class is taken to be equal to the lower limit of the following class. To keep various class intervals as mutually exclusive, the observations with magnitude greater than or equal to lower limit but less than the upper limit of a class are included in it. For example, if the lower limit of a class is 10 and its upper limit is 20, then this class, written as 10 - 20, includes all the observations which are greater than or equal to 10 but less than 20. The observations with magnitude 20 will be included in the next class.

(ii) Inclusive Method

Here all observations with magnitude greater than or equal to the lower limit and less than or equal to the upper limit of a class are included in it.

The two types of class intervals, discussed above, are constructed for the data on the measurements of diameter of a wire as shown below:

Exclusive Type Class Intervals	*Inclusive Type Class Intervals*	*Class Boundaries*
1.00 - 1.10	1.00 - 1.09	0.995 – 1.095
1.10 - 1.20	1.10 - 1.19	1.095 – 1.195
1.20 - 1.30	1.20 - 1.29	1.195 – 1.295
1.30 - 1.40	1.30 - 1.39	1.295 – 1.395
1.40 - 1.50	1.40 - 1.49	1.395 – 1.495
1.50 - 1.60	1.50 - 1.59	1.495 – 1.595
1.60 - 1.70	1.60 - 1.69	1.595 – 1.695
1.70 - 1.80	1.70 - 1.79	1.695 – 1.795
1.80 - 1.90	1.80 - 1.89	1.795 – 1.895
1.90 - 2.00	1.90 - 1.99	1.895 – 1.995

Mid-Value of a Class

In exclusive types of class intervals, the mid-value of a class is defined as the arithmetic mean (to be discussed later) of its lower and upper limits. However, in the case of inclusive types of class intervals, there is a gap between the upper limit of a class and the lower limit of the following class which is eliminated by determining the *class boundaries*. Here, the mid-value of a class is defined as the arithmetic mean of its lower and upper boundaries. To find class boundaries, we note that the given data on the measurements of diameter of a wire is expressed in terms of milimetres, approximated upto two places after decimal. This implies that a value greater than or equal to 1.095 but less than 1.10 is approximated as 1.10 and, thus, included in the class interval 1.10 - 1.19. Similarly, an observation less than 1.095 but greater than 1.09 is approximated as 1.09 and is included in the interval 1.00 - 1.09. Keeping the precision of measurements in mind, various class boundaries, for the inclusive class intervals, given above, can be obtained by subtracting 0.005 from the lower limit and adding 0.005 to the upper limit of each class. These boundaries are given in the third column of the above table.

Construction of a Grouped Frequency Distribution for the Data on the Measurements of Diameter of a Wire

Taking class intervals as 1.00 - 1.10, 1.10 - 1.20, *etc.* and counting their respective frequencies, by the method of tally marks, we get the required frequency distribution as given below :

Class Intervals	Tally Marks	Frequency
1.00 - 1.10	||||	4
1.10 - 1.20	𝍸 ||	7
1.20 - 1.30	𝍸 𝍸	10
1.30 - 1.40	𝍸 𝍸 ||||	14
1.40 - 1.50	𝍸 𝍸 𝍸 𝍸	20
1.50 - 1.60	𝍸 𝍸 |||	13
1.60 - 1.70	𝍸 ||||	9
1.70 - 1.80	𝍸 |	6
1.80 - 1.90	||||	4
1.90 - 2.00	|||	3
Total		90

Example 1.

Given below are the weights (in pounds) of 70 students.

(i) Construct a frequency distribution when class intervals are inclusive, taking the lowest class as 60 - 69. Also construct class boundaries.

(ii) Construct a frequency distribution when class intervals are exclusive, taking the lowest class as 60 - 70.

61, 80, 91, 113, 100, 106, 109, 73, 88, 92, 101, 106, 107, 97, 93, 96, 102, 114, 87, 62, 74, 107, 109, 91, 72, 89, 94, 98, 112, 103, 101, 77, 92, 73, 67, 76, 84, 90, 118, 107, 108, 82, 78, 84, 77, 95, 111, 115, 104, 69, 106, 105, 63, 76, 85, 88, 96, 90, 95, 99, 83, 98, 88, 72, 75, 86, 82, 86, 93, 92.

Solution.

(i) Construction of frequency distribution using inclusive class intervals.

Class Intervals	Tally Marks	Frequency	Class Boundaries
60 - 69	𝍸	5	59.5 - 69.5
70 - 79	𝍸 𝍸 |	11	69.5 - 79.5
80 - 89	𝍸 𝍸 ||||	14	79.5 - 89.5
90 - 99	𝍸 𝍸 𝍸 |||	18	89.5 - 99.5
100 - 109	𝍸 𝍸 𝍸 |	16	99.5 -109.5
110 - 119	𝍸 |	6	109.5 -119.5
Total		70	

To determine the class boundaries, we note that measured weights are approximated to the nearest pound. Therefore, a measurement less than 69.5 is approximated as 69 and included in the class interval 60 - 69. Similarly, a measurement greater than or equal to 69.5 is approximated as 70 and is included in the class interval 70 - 79. Thus, the class boundaries are obtained by subtracting 0.5 from the lower limit and adding 0.5 to the upper limit of various classes. These boundaries are shown in the last column of the above table.

(ii) The frequency distribution of exclusive type of class intervals can be directly written from the above table as shown below :

Class Inervals	*Frequency*
60 - 70	5
70 - 80	11
80 - 90	14
90 - 100	18
100 - 110	16
110 - 120	6
Total	70

Example 2.

Determine the class boundaries for the following distribution of ages of 40 workers of a factory, where quoted age is the age completed on last birthday.

Class Intervals	20 - 29	30 - 39	40 - 49	50 - 59	*Total*
Frequency	8	15	10	7	40

Solution.

The determination of class boundaries depends upon the nature of approximation. Since the quoted age is the age completed on last birthday, therefore, a number greater than 29 but less than 30 is approximated as 29. Therefore, the boundaries of this class will be 20 - 30. Similarly, the boundaries of other classes will be 30 - 40, 40 - 50 and 50 - 60, respectively.

6.3.3 Relative or Percentage Frequency Distribution

If instead of frequencies of various classes their relative or percentage frequencies are written, we get a relative or percentage frequency distribution.

$$\text{Relative frequency of a class} = \frac{\text{Frequency of the class}}{\text{Total Frequency}}$$

Percentage frequency of a class = Relative frequency × 100

These frequencies are shown in the following table.

Class Intervals	*Frequency*	*Relative Frequency*	*Percentage Frequency*
1.00 - 1.10	4	0.044	4.4
1.10 - 1.20	7	0.079	7.9
1.20 - 1.30	10	0.111	11.1
1.30 - 1.40	14	0.156	15.6
1.40 - 1.50	20	0.222	22.2
1.50 - 1.60	13	0.144	14.4
1.60 - 1.70	9	0.100	10.0
1.70 - 1.80	6	0.067	6.7
1.80 - 1.90	4	0.044	4.4
1.90 - 2.00	3	0.033	3.3
Total	90	1.000	100.0

6.3.4 Cumulative Frequency Distribution

In order to answer the questions like; the measurements on diameter that are less than 1.70 or the number of measurements that are greater than 1.30, *etc.*, a cumulative frequency distribution is constructed. A cumulative frequency distribution can be of two types :

(i) Less than type cumulative frequency distribution

(ii) More than type cumulative frequency distribution

These frequency distributions, for the data on the measurements of diameter of a wire, are shown in Table I and Table II respectively.

Table I

Diameters	*Cumulative Frequency*
less than 1.10	4
less than 1.20	11
less than 1.30	21
less than 1.40	35
less than 1.50	55
less than 1.60	68
less than 1.70	77
less than 1.80	83
less than 1.90	87
less than 2.00	90

Table II

Diameters	*Cumulative Frequency*
More than 1.10	90
More than 1.20	86
More than 1.30	79
More than 1.40	69
More than 1.50	55
More than 1.60	35
More than 1.70	22
More than 1.80	13
More than 1.90	7
More than 2.00	3

6.3.5 Frequency Density

Frequency density in a class is defined as the number of observations per unit of its width. Frequency density gives the rate of concentration of observations in a class :

$$\text{Frequency Density} = \frac{\text{Frequency of the class}}{\text{Width of the class}}$$

Frequency density of various classes are shown in the following table :

Class Intervals	*Frequency*	*Frequency Density*
1.00 - 1.10	4	40
1.10 - 1.20	7	70
1.20 - 1.30	10	100
1.30 - 1.40	14	140
1.40 - 1.50	20	200
1.50 - 1.60	13	130
1.60 - 1.70	9	90
1.70 - 1.80	6	60
1.80 - 1.90	4	40
1.90 - 2.00	3	30
Total	90	

6.4 BIVARIATE AND MULTIVARIATE FREQUENCY DISTRIBUTIONS

Bivariate Frequency Distributions

In the frequency distributions, discussed so far, the data are classified according to only one characteristic. There may be a situation where it is necessary to classify data, simultaneously, according to two characteristics. A frequency distribution obtained by the simultaneous classification of data according to two characteristics, is known as a bivariate frequency distribution. An example of such a classification is given below, where 100 couples are classified according to the two characteristics, Age of Husband and Age of Wife.

Classification according to Age of Husband and Age of Wife

Age of Husband ↓	Age of wife →	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	*Total*
10 - 20		6	3	0	0	0	9
20 - 30		3	16	10	0	0	29
30 - 40		0	10	15	7	0	32
40 - 50		0	0	7	10	4	21
50 - 60		0	0	0	4	5	9
Total		9	29	32	21	9	100

It should be noted that in a bivariate classification either or both the variable can be discrete or continuous. Further, there may be a situation in which one characteristic is a variable and the other is an attribute.

Multivariate Frequency Distribution

If the classification is done, simultaneously, according to more than two characteristics, the resulting frequency distribution is known as a multivariate frequency distribution.

Example 3.

Find the lower and upper limits of the classes when their mid-values are given as 15, 25, 35, 45, 55, 65, 75, 85 and 95.

Solution.

Note that the difference between two successive mid-values is same, *i.e.,* 10. Half of this difference is subtracted and added to the mid value of a class in order to get lower limit and the upper limit respectively. Hence, the required class intervals are 10 - 20, 20 - 30, 30 - 40, 40 - 50, 50 - 60, 60 - 70, 70 - 80, 80 - 90, 90 - 100.

Example 4.

Find the lower and upper limits of the classes if their mid-values are 10, 20, 35, 55, 85.

Solution.

Here the difference of two successive mid-values are different. In order to find the limits of the first class, half of the difference between the second and first mid-value is subtracted and added. Therefore, the first class limits are 5 - 15. The lower limit of second class is taken as equal to upper limit of the first class.

The upper limit of a class = lower limit + width,

where width = 2(Mid-value – lower limit).

∴ The upper limit of the second class = 15 + 2(20 – 15) = 25.

Thus, second class interval will be 15 - 25. Similarly, we can find the limits of third, fourth and fifth classes as 25 - 45, 45 - 65 and 65 - 105, respectively.

6.5 TABULATION

Tabulation is a systematic presentation of numerical data in rows and columns. Tabulation of classified data make it more intelligible and fit for statistical analysis. According to Tuttle, "*A statistical table is the logical listing of related quantitative data in vertical columns and horizontal rowsh of numbers, with sufficient explanatory and qualifying words, phrases and statements in the form of titles, headings and footnotes to make clear the full meaning of the data and their origin.*" The classified data presented in tabular form helps to bring out their essential features.

Objectives of Tabulation

The main objectives of tabulation are :

(i) To simplify complex data.
(ii) To highlight chief characteristics of the data.
(iii) To clarify objective of investigation.
(iv) To present data in a minimum space.
(v) To detect errors and omissions in the data.
(vi) To facilitate comparison of data.
(vii) To facilitate reference.
(viii) To identify trend and tendencies of the given data.
(ix) To facilitate statistical analysis.

Difference between classification and tabulation

The basic points of difference between classification and tabulation, inspite of the fact that these are closely related, are as given below :

(i) Classification of data is basis for tabulation because first the data are classified and then tabulated.

(ii) Classification is a process of statistical analysis while tabulation is a process of presentation.

(iii)By classification the data are divided into various groups and subgroups on the basis of their similarities and dissimilarities while tabulation is a process of arranging the classified data in rows and columns with suitable heads and subheads.

6.2.1 Main Parts of a Table

The main parts of a table are as given below :

(i) Table Number

This number is helpful in the identification of a table. This is often indicated at the top of the table.

(ii) Title

Each table should have a title to indicate the scope, nature of contents of the table in an unambiguous and concise form.

(iii) Captions and stubs

A table is made up of rows and columns. Headings or subheadings used to designate columns are called captions while those used to designate rows are called stubs. A caption or a stub should be self explanatory. A provision of totals of each row or column should always be made in every table by providing an additional column or row respectively.

(iv) Main Body of the Table

This is the most important part of the table as it contains numerical information. The size and shape of the main body should be planned in view of the nature of figures and the objective of investigation. The arrangement of numerical data in main body is done from top to bottom in columns and from left to right in rows.

(v) Ruling and Spacing

Proper ruling and spacing is very important in the construction of a table. Vertical lines are drawn to separate various columns with the exception of sides of a table. Horizontal lines are normally not drawn in the body of a table, however, the totals are always separated from the main body by horizontal lines. Further, the horizontal lines are drawn at the top and the bottom of a table.

Spacing of various horizontal and vertical lines should be done depending on the available space. Major and minor items should be given space according to their relative importance.

(vi) Head-note

A head-note is often given below the title of a table to indicate the units of measurement of the data. This is often enclosed in brackets.

(vii) Foot note

Abbreviations, if any, used in the table or some other explanatory notes are given just below the last horizontal line in the form of footnotes.

(viii) Source-Note

This note is often required when secondary data are being tabulated. This note indicates the source from where the information has been obtained. Source note is also given as a footnote.

The main parts of a table can also be understood by looking at its broad structure given below :

Structure of a table

Table No : ....

Title : ....

Stub Heading	Captions		Captions			Total
	Captions	Captions	Captions	Captions	Captions	
↑ Stub Enteries ↓	M A I N		B O D Y			
Total						

Foot Note :

Source :

Rules for Tabulation

The rules for tabulation of data can be divided into two broad categori
(i) Rules regarding structure of a table, explained above, and (ii) General rules.

General Rules

(i) The table should be simple and compact which is not overloaded with de-tails.

(ii) Tabulation should be in accordance with the objective of investigation.

(iii) The unit of measurements must always be indicated in the table.

(iv) The captions and stubs must be arranged in a systematic manner so that it is easy to grasp the table.

(v) A table should be complete and self explanatory.

(vi) As far as possible the interpretative figures like totals, ratios and percent-ages must also be provided in a table.

(vii) The entries in a table should be accurate.

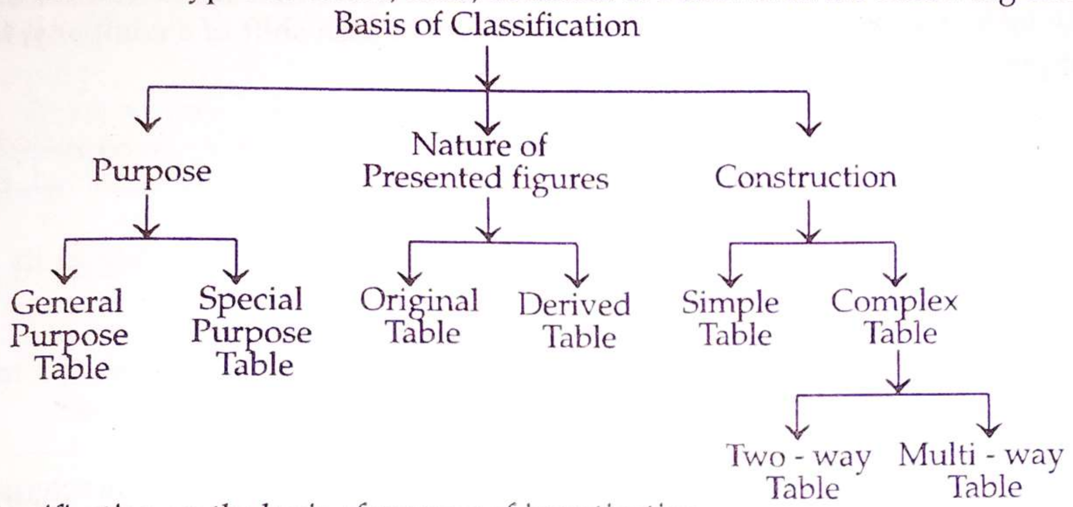
(viii) Table should be attractive to draw the attention of readers.

6.6.2 Type of Tables

Statistical tables can be classified into various categories depending upon the basis of their classification. Broadly speaking, the basis of classification can be any of the following :

(i) Purpose of investigation
(ii) Nature of presented figures
(iii) Construction

Different types of tables, thus, obtained are shown in the following chart.



1. Classification on the basis of purpose of investigation

These tables are of two types viz. (i) General purpose table and (ii) Special purpose table.

(i) General purpose table

A general purpose table is also called as a reference table. This table facilitates easy reference to the collected data. In the words of Croxton and Cowden, "*The primary and usually the sole purpose of a reference table is to present the data in such a manner that the individual items may be readily found by a reader.*" A general purpose table is formed without any specific objective, but can be used for a number of specific purposes. Such a table usually contains a large mass of data and are generally given in the appendix of a report.

(ii) Special purpose table

A special purpose table is also called a text table or a summary table or an analytical table. Such a table presents data relating to a specific problem. According to H. Secrist, " *These tables are those in which are recorded, not the detailed data which have been analysed, but rather the results of analysis.*" Such tables are usually of smaller size than the size of reference tables and are generally found to highlight relationship between various characteristics or to facilitate their comparisons.

2. Classification on the basis of the nature of presented figures

Tables, when classified on the basis of the nature of presented figures can be (i) Primary table (ii) Derivative table.

(i) Primary Table

Primary table is also known as original table and it contains data in the form in which it were originally collected.

(ii) Derivative Table

A table which presents figures like totals, averages, percentages, ratios, coefficients, *etc.*, derived from original data.

A table of time series data is an original table but a table of trend values computed from the time series data is known as a derivative table.

3. *Classification on the basis of construction*

Tables when classified on the basis of construction can be (i) Simple table, (ii) Complex table or (iii) Cross-classified table.

(i) Simple Table

In this table the data are presented according to one characteristic only. This is the simplest form of a table and is also known as table of first order. The following blank table, for showing the number of workers in each shift of a company, is an example of a simple table.

Table No. ....

Shifts	*No. of Workers*
I	
II	
III	
Total	

(ii) Complex Table

A complex table is used to present data according to two or more characteristics. Such a table can be two-way, three-way or multi-way, *etc.*

(a) Two-way table

Such a table presents data that is classified according to two characteristics. In such a table the columns of a table are further divided into sub-columns. The example of such a table is given below.

Table No. ....

Distribution of workers of a factory according to shift and sex

Shifts	*No. of Workers*		*Total*
	Males	*Females*	
I			
II			
III			
Total			

(b) Three-way table

When three characteristics of data are shown simultaneously, we get a three-way table as shown below.

Table No. ....

Distribution of workers of a factory according to shift, sex and training

Shift	*No. of Workers*						*Total No. of Workers*
	Males		*Total*	*Females*		*Total*	
	Skilled	*Unskilled*		*Skilled*	*Unskilled*		
I							
II							
III							

(c) Multi-way table

If each shift is further classified into three departments, say, manufacturing, packing and transportation, we shall get a four-way table, *etc.*

(iii) The Cross-Classified Table

Tables that classify entries in both directions, *i.e.*, row-wise and column-wise, are called cross-classified tables. The two ways of classification are such that each category of one classification can occur with any category of the other. The cross-classified tables can also be constructed for more than two characteristics also.

A cross-classification can also be used for analytical purpose, *e.g.*, it is possible to make certain comparisons while keeping the effect of other factors as constant.

Example 5.

Draw a blank table to show the population of a city according to age, sex and unemployment in various years.

Solution.

Table No. ....

Population of a city according to age, sex and unemployment in various years

Years	Age / Sex	Population (in thousands)							
		Employed				Unemployed			
		below 20	20 - 60	60 & above	Total	below 20	20 - 60	60 & above	Total
1991	Males								
	Females								
	Total								
1992	Males								
	Females								
	Total								

Note : The table can be extended for the years 1993, 94, 95, 96, *etc.*

Example 6.

In a sample study about coffee habit in two towns; the following information were received :

Town A : Females were 40%; total coffee drinkers were 45%; and male non-coffee drinkers were 20%.

Town B : Males were 55%; male non-coffee drinkers were 30%; and female coffee drinkers were 15%.

Represent the above data in a tabular form.

Solution.

Table No. ......

Distribution of population, according to sex and coffee habit, in two towns

Habit	Town A			Town B		
	Males	Females	Total	Males	Females	Total
Coffee Drinkers	40	5	45	25	15	40
Non - Coffee Drinkers	20	35	55	30	30	60
Total	60	40	100	55	45	100

Note : The figures are in percentage.

Example 7.

Prepare a blank table for showing the percentage of votes polled by various political parties in India according to states, during 1996 general elections.

Solution.

Table No. ......

Percentage distribution of votes polled by political parties according to States in India during the 1996 general elections

States	*No. of Votes Polled*							
	Congress	*B.J.P.*	*Janta Dal*	*C.P.M.*	*B.S.P.*	*C.P.I.*	*Others*	*Total*
Assam								
Andhra								
Arunachal								
Bengal								
- - - - -								
- - - - -								
- - - - -								
Total								

6.7 METHODS OF TABULATION

Tabulation of the collected data can be done in two ways : (i) By Manual Method, and (ii) By Mechanical Method.

(i) Manual Method

When field of investigation is not too large and the number of characteristics are few, the work of tabulation can be done by hand.

(ii) Mechanical Method

This method is used when the data are very large. The use of machines save considerable amount of labour and time. With the development of high speed computers, the work of tabulation and analysis of data can be done very quickly and with greater accuracy.

6.8 SUMMARY

Classification and tabulation of data are necessary to understand its broad features and to make it fit for statistical analysis.

Classification of data on the basis of one, two or more factors is termed as a one-way, two-way or multi-way classification, respectively.

Classified data, when arranged in some logical order such as, according to size or according to time of occurrence or according to some other criterion, is known as statistical series. A statistical series, in which data are arranged according to magnitude of one or more characteristics, is known as a frequency distribution. Data classified according to the magnitude of only one characteristic is known as uni-variate frequency distribution. Similarly, data classified, simultaneously, according to the magnitude of two or more characteristics are known as bivariate or multivariate frequency distributions respectively.

When a characteristics is an attribute, the data can be classified into two or more classes according to this attribute, known as dichotomous or manifold classification respectively. Similarly, when the data are simultaneously classified according to two or more attributes, the classifications are two-way or multi-way respectively. It is also possible to have a two-way or multi-way classification in which one or more characteristics are variables while others are attributes.

The process of classification is facilitated by writing the classified data in tabular form. Using tables, it is possible to write huge mass of data in a concise form. Further, it helps to highlight essential features of the data and make it fit for further analysis.

A table is made up of rows and columns. Various headings and subheadings used to designate columns and rows of a table are known as captions and stubs respectively. A table can be of general or special purpose. If it represents original data, it is called a primary table otherwise it is called a derivative table. Finally, a table can be simple, complex, cross-classified; one, two or multi-way, *etc.*

QUESTIONS

1. What do you mean by Classification and Tabulation? Explain their importance in statistical studies.
2. What are the different factors that should be kept in mind while classifying data?
3. Distinguish between classification and tabulation. Discuss the purpose and methods of classification.
4. What are objects of classification of data? Discuss different methods of classification.
5. Discuss the purpose, methods and importance of tabulation in any statistical investigation. Mention the types of tables generally used.
6. Define tabulation. Describe different parts of a table. What precautions should be taken into consideration while preparing a statistical table?
7. What do you understand by a statistical series?. Explain various types of statistical series.
8. Distinguish between an ungrouped and a grouped frequency distribution. What are the points that should be taken into consideration while determining the following :
   (a) Number of Groups
   (b) Magnitude of Class-Intervals
   (c) Class Limits.
9. Point out the correct answer from the following :
   (a) In an exclusive series :
      (i) Both limits of a class interval are considered.
      (ii) Upper limit is excluded.
      (iii) Lower limit is excluded.
      (iv) Both the limits are excluded.
   (b) In tabulation of the grouped data :
      (i) It is necessary to have frequencies.
      (ii) It is necessary to have frequencies in all the groups.
      (iii) Frequency can be negative.
      (iv) None of the above is correct.
   (c) In a discrete series :
      (i) Class intervals should be equal.
      (ii) Class intervals can be different.
      (iii) Class intervals can be zero.
      (iv) None of the above is correct.

10. Twenty students of a class appeared in an examination. Their marks out of 50 are as under :
5, 6, 17, 17, 20, 21, 22, 22, 22, 25, 25, 26, 26, 30, 31, 31, 34, 35, 42, 48.
Prepare a classified table by taking class intervals of 10 each, according to exclusive and inclusive methods.

11. Construct a frequency table for the following data by taking width of each class as 10. Use inclusive method of classification.
30, 38, 43, 59, 82, 40, 45, 39, 83, 85, 72, 66, 45, 33, 53, 67, 70, 72, 52, 50, 43, 44, 60, 89, 67, 66, 78, 32, 56, 47, 65, 56, 38, 84, 64, 52, 43, 33, 31, 35, 38, 39, 40, 37, 52, 53, 60.
If these figures represent the age of persons approximated to the nearest whole number, construct the class boundaries.

12. The number of children in 50 families of a locality are given below. Construct an appropriate discrete frequency distribution.
2, 2, 2, 3, 2, 4, 5, 4, 6, 8, 3, 3, 1, 4, 3, 1, 3, 3, 2, 1, 3, 3, 2, 4, 3, 5, 4, 3, 3, 2, 2, 5, 2, 5, 3, 3, 3, 4, 3, 5, 4, 4, 2, 6, 3, 6, 3, 3, 7, 3.

13. Construct a frequency distribution of the marks obtained by 50 students in economics as given below :
42, 53, 65, 63, 61, 47, 58, 60, 64, 45, 55, 57, 82, 42, 39, 51, 65, 55, 33, 70, 50, 52, 53, 45, 45, 25, 36, 59, 63, 39, 65, 30, 45, 35, 49, 15, 54, 48, 64, 26, 75, 20, 42, 40, 41, 55, 52, 46, 35, 18.
(Take the first class interval as 10 - 20)

14. The following figures give the ages, in years, of newly married husbands and their wives. Represent the data by an appropriate frequency distribution.

Age of Husband	:	24	26	27	25	28	24	27	28	25	26
Age of Wife	:	17	18	19	17	10	18	18	19	18	19
Age of Husband	:	25	26	27	25	27	26	25	26	26	26
Age of Wife	:	17	18	19	19	20	19	17	20	17	18

15. Tabulate the following information :
In a trip organised by a college, there were 80 persons each of whom paid Rs 15.50 on an average. There were 60 students, each of whom paid Rs 16. Members of the teaching staff were charged at a higher rate. The number of servants was 6, all males and they were not charged anything. The number of ladies was 20% of the total of which one was a lady staff member.

16. Prepare a blank table to represent the students of a college according to
(i) Percentage of marks obtained in an annual examination by taking class intervals 0 - 10, 10 - 20, ..., *etc.*
(ii) Sex-wise : males and females.
(iii) Faculty-wise : science, arts and commerce.

17. There were 850 union and 300 non union workers in a factory in 1988. Of these, 250 were females out of which 100 were non union workers. The number of union workers increased by 50 in 1989 out of which 40 were males. Of the 350 non union workers, 125 were females. In 1990, there were 1,000 workers in all and out of 400 non union workers there were only 100 females. There were only 400 male workers in the union.
Present the above information in a tabular form.

18. A super market divided into five main sections; grocery, vegetables, medicines, textiles and novelties, recorded the following sales in 1985, 1986 and 1987 :

    In 1985 the sales in groceries, vegetables, medicines and novelties were Rs 6,25,000, Rs 2,20,000, Rs 1,88,000 and Rs 94,000 respectively. Textiles accounted for 30% of the total sales during the year.

    In 1986 the total sales showed 10% increase over the previous year while grocery and vegetables registered 8% and 10% increase over the corresponding previous year, medicines dropped by Rs 13,000 and textiles increased by Rs 53,000 over their corresponding figure of 1985.

    In 1987, though total sales remained the same as in 1986, grocery fell by Rs 22,000, vegetables by Rs 32,000, medicines by Rs 10,000 and novelties by Rs 12,000.

    Tabulate the above data.

19. A survey of 370 students from the Commerce Faculty and 130 students from the Science Faculty revealed that 180 students were studying for only C.A. examinations, 140 for only Costing examinations and 80 for both C.A. and Costing examinations. The rest had offered part-time Management courses. Of those studying Costing only, 13 were girls and 90 boys belonged to the Commerce Faculty. Out of 80 students studying for both C.A. and Costing, 72 were from the Commerce Faculty amongst which 70 were boys. Amongst those who offered part-time Management courses, 50 boys were from the Science Faculty and 30 boys and 10 girls from the Commerce faculty. In all there were 110 boys in the Science Faculty.

    Present the above information in a tabular form. Find the number of students from the Science Faculty studying for part-time Management courses.

20. Construct a blank table in which could be shown, at two different dates and in five industries, the average wages of the four groups, males, females, eighteen years and over and under eighteen years. Suggest a suitable title.

Thank you!

DiligentEdu©

www.DiligentEdu.in

www.YouTube.com/@DiligentEdu