Note

Frequency Distribution Table

Classes	Frequencies (*f*)	Deviation from A.M. (*d*)	Frequencies × Deviation from (*fd*)	
45 – 49	||| = 3	3	9	
40 – 44	|| = 2	2	4	
35 – 39	|| = 2	1	2	+17
30 – 34	𝍸 ||| = 8	(A.M. = 32)0	0	
25 – 29	𝍸 || = 7	– 1	– 7	
20 – 24	|||| = 4	– 2	– 8	
15 – 19	|| = 2	– 3	– 6	– 29
10 – 14	|| = 2	– 4	– 8	
***Ci* = 5**	***N* = 30**		**$\Sigma fd = 14$**	

$$\therefore \quad \text{Mean} = A.M. + \left(\frac{\Sigma fd}{N}\right) \times Ci$$

$$= 32 + \left(\frac{-14}{30}\right) \times 5$$

$$= 32 + \left(\frac{-70}{30}\right)$$

$$= 32 - \frac{7}{3}$$

$$= 32 - 2.33 = 29.67$$

Precautions

1. The assumed mean should be considered in some middle class which has the highest number of frequencies.
2. As the mid point of the class is the assumed mean.
3. The deviation of the classes above the assumed mean in the ascending series is positive and that of the lower ones is negative.

Note

4.6.5 Characteristics of Mean

1. The mean is the mean value of the scores, it can be easily found by mathematical methods.
2. The mean value is such a focal value whose gravitational value (deviation) is the same on both sides.
3. The mean is very sensitive, changes in the value of any of the scores in the group. This changes the value of the mean. Hence it is correct for the scores of a group is considered a representative value.
4. Mean is more reliable than other centroids such as median and mode.

4.6.6 Limitations of Mean

1. The mean represents the scores of a group only if the distribution of scores is normal.
2. If the value of some scores of a group is very high or very less than the value of other scores, then in that case also the mean value does not represent the group correctly.

4.6.7 Meaning and Definition of Median

Median is the central value that divides the scores of a group into such two equal parts that all the scores of one part are greater than that and all the scores of the other part are less than that. We make it clear with an example. Suppose the marks of 5 students in an examination of 50 marks are as follows:

17 40 38 21 and 41

then the scores in rank

17 21 38 40 and 41

Now just look at these five scores. Of these, 38 is such a score, which is less than two marks and more than two marks. Hence the median of this group is 38.

Then the median can be defined as follows :

The median is the score of a group that, when placed in their rank order, falls exactly in the middle of them and on either side of which half (50-50 percent) of the group's scores are distributed.

Note

4.6.8 Methods of Calculating Median from Ungroup Data :

The most straightforward method of finding the median value is to arrange the scores of the students of the group in the rank order, *i.e.* to be arranged from minimum to maximum and then take the middle score from it.

Example : The marks of 7 students in an examination of 50 marks are 15, 8, 30, 21, 19 and 33 respectively. Find the median.

Calculation : On arranging the scores in rank order

8, 15, 19, 21, 24, 30, 33

The middle score in this is the median number (mdn) 21, above which there are three scores and there are three marks below.

Note When the number of students is even, then the average of the middle two marks is taken.

Example : In an examination of 50 marks, 8 students got 23, 15, 8, 12, 24, 26, 33 and 31 marks respectively. Find the median.

Calculation : On putting the scores in the rank order

8, 12, 15, 23, 24, 26, 31, 33

$\therefore$ In these the middle scores are both 23 and 24, they have 3-3 scores here and there.

$$\therefore \quad \text{Median} = \frac{23+24}{2} = \frac{47}{2} = 23.5$$

Note Look at the above two examples. In these, the median value is half of the middle score or the sum of the two middle scores. If we want to express it in formula form, we can do it as follows :

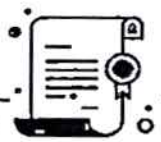
$$\text{Median,} \quad Mdn = \frac{N+1}{2}\text{th score}$$

where Mdn is the sign of the median and N is the number of students. In the first *example,*

$$\text{Median,} \quad Mdn = \frac{N+1}{2}\text{th score}$$

$$= \frac{7+1}{2} = \frac{8}{2} = 4\text{th score}$$

$$= 21$$



Note

In the Second *Example,*

Median, $Mdn = \frac{N+1}{2}$ th score

$$= \frac{8+1}{2} = \frac{9}{2} = 4.5\text{th score}$$

i.e., average of 4th and 5th scores

$$= \frac{23+24}{2} = \frac{47}{2}$$

$$= 23.5$$

4.6.9 Methods of Calculating Median from Group Data

When the number of students is very large, it becomes a bit difficult to organize their scores. In that case, we prepare the frequency distribution table and calculate the Median value with the help of frequency distribution table.

In this, we first calculate the cumulative frequencies in the frequency distribution table and then calculate with the help of this table we find out the class in which the $\frac{N}{2}$th score is given. After this, with the help of following formulas, the median of the marks obtained is determined.

1. Median, $Mdn = L + \frac{\left(\frac{N}{2} - Fb\right)}{f} \times Ci$

   Where,

   L the class with the median just below it indicates the lower limit.

   N denotes the total number of students.

   Fb is the sign of the sum of the frequencies before the class with the median.

   f denotes the frequency of the class with the median and

   Ci is an indication of the class interval, that is, the size of the classes.

   Or

Note

2. Median, $Mdn = U + \frac{\left(\frac{N}{2} - Fa\right)}{f} \times Ci$

Where,

U denotes the upper limit of the class with the median.

N denotes the number of students.

Fa is the sign of the sum of all the frequencies above the class of the median.

f denotes the frequency of the class with the median and

Ci is an indication of the class interval, that is, the size of the classes. We make this method clear with an *example.*

Example : Find the Median of the scores from the following frequency distribution table :

Classes	Frequencies	Cumulative Frequency *(cf)*
45-49	3	40
40-44	2	37
35-39	5	35
30-34	11	30 (Median value square)
25-29	8	19
20-24	5	11
15-19	4	6
10-14	2	2
***Ci* = 5**	***N* = 40**	

Calculation

According to the first formula,

$$Mdn = L + \frac{\left(\frac{N}{2} - Fb\right)}{f} \times Ci$$

Note

Before using this formula we should first find out in which class the median value. For this we have to find the value $\frac{N}{2}$ and the class it comes in is the $\frac{N}{2}$ median class.

Median score, $Mdn = \frac{n}{2}$th score

$$= \frac{N}{2} = 20\text{th score}$$

In this, 20th are falling 30-34 class-interval. Therefore, we consider this class as a median class.

Then,

L i.e., its lower limit = 29.5

N i.e., the number of students = 40

Fb i.e., the sum of the frequencies before the class with the median value = 19

f i.e., the frequency of the class with the median value = 11

Ci i.e., the class size = 5

$$\therefore \quad \text{Median, } Mdn = 29.5 + \frac{\left(\frac{40}{2} - 19\right)}{11} \times 5$$

$$= 29.5 + \frac{(20 - 19)}{11} \times 5$$

$$= 29.5 + \frac{1}{11} \times 5$$

$$= 29.5 + 0.45 = 29.95$$

According to another formula,

$$Mdn = U + \frac{\left(\frac{N}{2} - Fa\right)}{f} \times Ci$$

Where,

U means the upper limit of the class with the median value = 34.5

N means number of students = 40

Fa means upward frequencies of the class having median value = 11

f means the frequencies of the class with the median value = 5

Note

and Ci means class size = 5

$$\therefore \quad \text{Median, } Mdn = 34.5 - \frac{\left(\frac{40}{2} - 10\right)}{11} \times 5$$

$$= 34.5 - \frac{(20-10)}{11} \times 5$$

$$= 34.5 - \frac{10}{11} \times 5$$

$$= 34.5 - \frac{50}{11}$$

$$= 34.5 - 4.55$$

$$= 29.95$$

4.6.10 Characteristics of Median

1. The median is in the middle of the arrangement of the scores of a group in the order of rank, so this group is suitable for comparison and interpretation of other scores.
2. When the deviation of the scores in a group is abnormal or the value of some of its scores is very high or very less than the other scores, then in that case the median is more useful than the mean.
3. Median score is more useful for qualitative interpretation of scores of a group than other centroids such as mean and mode.
4. Interpretation of other scores of the group from the median score displayed by the graph of scores of a group can be done.

4.6.12 Limitations of Median

1. The median can be calculated for scores of the same group, not for the scores of two groups.
2. The median is the rank of the scores of a group, so to find the median, it is necessary to arrange the position of all the scores of the group. When there is a large group, it takes more time and energy to do all this.

4.6.13 Uses and Importance of Median in Field of Education

In the field of Education, the maximum use is of Mean but there are

Note

some situations in which Median is used such as;

1. When the distribution of scores of a group is not normal.
2. When all the scores of a group are not obtained.
3. When we have to find out the position of a particular person in a group that he is visually comes in the first 50% or comes in the last 50%.
4. When there is no time to calculate the mean.
5. When it is needed in the conclusion rendering of academic researches.

4.7 Meaning and Definition of Mode

The score which has the highest frequency in the scores of a group is called the score of that group called mode.

For example, look at the following scores :

5 4 3 4 6 4 5 7 3 and 6

Of these, 5 has a frequency of 2, 4 has a frequency of 3, 3 has a frequency of 2, 6 has a frequency of 2, and 7 has a frequency of 1.

So the mode of these scores will be 4 as it has the highest frequency in the group. Then the mode can be defined in the following way :

Mode is the score of a group which has the highest frequency among the scores of that group.

4.7.1 Methods of Calculating Mode from Ungroup Data

The most straightforward method of finding the mode value is to arrange the scores of the students of the group in the order of rank *i.e.,* from lowest score to highest score and then observe the frequency of scores. The score with the highest frequency will be the mode value.

Example : 10 students got the following marks in an examination of 50 marks. Find the Mode.

40 11 17 18 8 17 24 37 18 17

Calculation : Arranging the scores in rank order,

8 11 17 17 17 18 18 24 37 40

$\therefore$ Mode = 17

Note

Note Sometimes it happens that the frequencies of two or more scores are the same, in that case the following rules are there to find the mode value :

1. The group in which the frequencies of two adjacent scores are highest and mutually exclusive. is equal to, the average of these two scores in that group is the mode.

   For example : 8 11 17 17 17 19 19 19 24 and 37

   Mode, $Mo = \frac{17+19}{2} = \frac{36}{2} = 18$

2. The group in which the frequency of two such scores is highest and equal among themselves, which are not close to each other, then both the scores are mode in that group. Such groups are called Bi-Modal.

   For example : 8 11 17 17 17 18 24 24 24 and 26

   Mode, $Mo = 17$ and 24

3. In a group in which many or all of the scores have the same frequency, the question of finding the mode in that group does not arise.

   For example : There is no mode in 11 11 11 16 16 16 17 17 17 19 19 19 23 23 23.

4.7.2 Methods of Calculating Mode from Group Data

When the number of students is very large, it is difficult to arrange the scores in rank order and there is a possibility of error. In that case, the frequency distribution table is prepared and with the help of that the mode value is extracted. There are also two methods for this : one is to find the approximate mode and from the second the correct mode is taken.

4.6.3 Method of Calculating Crude Mode

To work out for the approximate mode, first let's make a frequency distribution table. Then find the mid-point of the class in which the frequencies of the scores are highest. This midpoint is the approximate mode of that group.

Example : Find the polynomial of the scores from the ahead frequency distribution table :

Note

Frequency Distribution Table

Classes	Frequencies
90 - 99	2
80 - 89	5
70 - 79	8
60 - 69	9
50 - 59	12
40 - 49	4
30 - 39	5
20 - 29	3
10 - 19	2

Calculation

Look at the frequency distribution table carefully. The highest frequency of scores in this is in the 50 - 59 class.

$\therefore$ Approximate mode, $Mo = \frac{50+59}{2} = \frac{109}{2} = 54.5$

4.7.4 Method of Calculating True Mode

To find the correct mode, first the mean is taken, then the median and then twice the mean is subtracted from three times the median. It is expressed in the formula as follows :

$$Mo = 3 \times Mdn - 2 \times M$$

We have already told the method of finding the mean and median. Here, after completing the above table, calculating them and extracting mode numbers :

Frequency Distribution Table

Classes	Frequencies *(f)*	Cumulative Frequencies	Deviation from A.M. *(d)*	Frequencies × Deviation
90 - 99	2	50	4	8
80 - 89	5	48	3	15

Note

70 - 79	8	49	2	16
60 - 69	9	35	1	9
50 - 59	12	26	0 (A.M.)	0
40 - 49	4	14	−1	−4
30 - 39	5	10	−2	−10
20 - 29	3	5	−3	−9
10 - 19	2	2	−4	−8
$Ci = 10$	$N = 50$			$\Sigma fd = 17$

$\therefore$ Mode, $M = A.M. + \frac{(\Sigma fd)}{N} \times Ci$

$= \frac{50+59}{2} + \left(\frac{17}{50}\right) \times 10$

$= \frac{109}{2} + \frac{17}{50} \times 10$

$= \frac{109}{2} + \frac{17}{5}$

$= 54.50 + 3.40 = 57.90$

Median, $Mdn = L + \frac{\left(\frac{N}{2} - Fb\right)}{f} \times Ci$

$= 49.5 + \frac{\left(\frac{50}{2} - 14\right)}{12} \times 10$

$= 49.5 + \frac{25-14}{12} \times 10$

$= 49.5 + \frac{11}{12} \times 10$

$= 49.5 + \frac{11}{6} \times 5$

$= 49.5 + \frac{55}{6}$

$= 49.5 + 9.17 = 58.67$

$\therefore$ Mean $Mo = 3 \times Mdn - 2 \times M$

$$= 3\times58.67 - 2\times57.90$$

$$= 176.01 - 115.80$$

$$= 60.21$$

Comment

1. The predicted mode of this frequency distribution is 54.5 and the correct mode is 60.21. Although there is a slight difference between these two, but this difference makes the decisions of evaluation different. Therefore, as far as possible, we should find the correct mode.
2. When the distribution forms an approximately normal probability curve, there is little difference between the mean value, median value and mode value. In the above example the mean value 57.90, median value 58.67 and the mode value is 60.21. There is very little difference between these three.
3. There are also separate formulas for finding the mode, by which the mode can be found directly without the mean and median.

   (i) Mode, $Mo = L + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times Ci$

   (ii) Mode, $Mo = L + \frac{f_2}{f_0 + f_2} \times Ci$

   Where,

   L = Net lower limit of modal class

   f_0 = Frequency of the class preceding the modal class

   f_1 = Frequency of the modal class

   f_2 = Frequency of the class immediately following the modal class

   In this $L = 49.5$, $f_0 = 4$ and $f_2 = 9$

   Mode, $Mo = L + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times Ci$

   $$= 49.5 + \frac{12 - 4}{2\times12 - 4 - 9} \times 10$$

Note

$$= 49.5 + \frac{8}{11} \times 10$$

$$= 49.5 + \frac{8}{11}$$

$$= 49.5 + 7.27 = 56.77$$

In second formula,

Mode, $$Mo = L + \frac{f_2}{f_0 + f_2} \times Ci$$

$$= 49.5 + \frac{9}{4+9} \times 10$$

$$= 49.5 + \frac{90}{13}$$

$$= 49.5 + 6.92 = 56.42$$

Note

1. If only mode has to be done, then it remains to use the above formula because in the first method, if by mistake, the mean or one of the middle values, then the mode will automatically be wrong.
2. The estimated mode from this iteration is 54.5 and the pure mode from the first and second methods is 60.21, 56.21 and 56.42 respectively. This difference is due to the use of different methods. Usually it is more useful to find the pure mode.

4.6.5 Characteristics of Mode

1. The mode of scores of a group can be found very easily, of non-general scores too.
2. The value of mode of the scores of a group can also be found by graphing the scores.
3. If the value of some scores of a group is much more or less than the value of other scores, then in that case also the mode of the scores of that group is not affected.
4. It is more useful for qualitative interpretation of scores of a group.

Note

4.7.6 Limitations of Mode

1. When none of the scores in a group are mode or more than two such scores with a large difference, come in the category of mode, then in that case the mode will be meaningless.
2. The mode is an approximate centroid value and therefore less useful for statistical calculations.

4.8 Standard Deviation (SD)

While finding the mean deviation, no attention is paid to the plus and minus signs of the deviation, and this is the biggest limit of mean deviation. Therefore, to remove this defect, calculate the standard deviation. In other words, we can say that standard deviation is an improved form of mean deviation.

To overcome the problem of markings under the standard deviation, the deviations are squared due to which the minus sign also turns into a plus sign, then the average of these squared deviations is taken and their square root is taken. This is called the standard deviation. According to **Reichmann**, "The standard deviation is also known as root mean square deviation. It is the square root of the mean value of all deviation squared taken from the distribution mean."

In simple words, the standard deviation can be defined as :

Very generally, the standard deviation is a measure of how the scores cluster or disperse around the mean and is conventionally represented by the Greek letter sigma.

Symbolically, $\sigma = \sqrt{\frac{\Sigma d^2}{N}}$

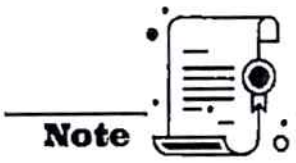
Calculation : The following three types of questions are used to calculate the standard deviation :

Case I : When only unclassified data is given.

Case II : When unclassified data along with their frequencies are also given.

Case III : When classified data is given.

Case I : When only unclassified data is given.



These type of questions are solved by two methods : Actual Mean Method and Assumed Mean Method. The formulas used in these methods are as follows :

$$S.D. = \sqrt{\frac{\Sigma d^2}{N}}$$ (Actual mean method)

and $$S.D. = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2}$$ (Assumed mean method)

Example : Find the standard deviation from the following data by the actual mean method.

Table

S.No.	X	$d=(X-M)$	d^2
1	10	$10-18=-8$	64
2	15	$15-18=-3$	9
3	10	$10-18=-8$	64
4	20	$20-18=2$	4
5	25	$25-18=7$	49
6	15	$15-18=-3$	9
7	25	$25-18=7$	49
8	20	$20-18=2$	4
9	17	$17-18=-1$	1
10	23	$23-18=5$	25
$N=10$	$\Sigma X=180$		$\Sigma d^2=278$

$$M = \frac{\Sigma X}{N} = \frac{180}{10} = 18$$

Now, $$\sigma = \sqrt{\frac{\Sigma d^2}{N}}$$

$$= \sqrt{\frac{278}{10}}$$

$$\sqrt{27.8} = 5.27$$

Example : Find the assumed deviation from the following data by the actual mean method.

Note

Table

S.No.	X	$d=(X-M)$	d^2
1	10	$10-17=-8$	49
2	15	$15-17=-2$	4
3	10	$10-17=-7$	49
4	20	$20-17=3$	9
5	25	$25-17=8$	64
6	15	$15-17=-2$	4
7	25	$25-17=8$	64
8	20	$20-17=3$	9
9	17	$17-17=0$	0
10	23	$23-17=6$	36
$N=10$		$\Sigma d=10$	$\Sigma d^2=288$

$$\sigma=\sqrt{\frac{\Sigma d^2}{N}-\left(\frac{\Sigma d}{N}\right)^2}$$

$$=\sqrt{\frac{288}{10}-\left(\frac{10}{10}\right)^2}$$

$$=\sqrt{28.8-1}$$

$$=\sqrt{27.8}$$

$$=5.27$$

Important Steps

1. Arrange points distribution.
2. Find the actual mean from the score distribution.
3. Find the deviation using the formula $d=X-M$.
4. Find the class d^2 of the deviation.
5. Find Σd^2 by adding up the deviation squares obtained.
6. Substituting all these values in the formula and calculate S.D.

Case II : When unclassified data along with their frequencies are given.

Note

Example : Find the standard deviation from the following data by assumed mean method.

Table

S.No.	X	f	$d=(X-M)$	fd	fd^2
1	65	3	$65-75=-10$	-30	300
2	68	4	$68-75=-7$	-28	196
3	70	4	$70-75=-5$	-20	100
4	72	5	$72-75=-3$	-15	45
5	74	6	$74-75=-1$	-6	6
6	75	7	$75-75=0$	0	0
7	78	6	$78-75=3$	18	54
8	80	5	$80-75=5$	25	125
9	82	4	$82-75=7$	28	196
10	84	6	$84-75=9$	54	486
		$N=50$		$\Sigma fd=26$	$\Sigma fd^2=1508$

Now, S.D. or $\sigma = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2}$

$$= \sqrt{\frac{1508}{50} - \left(\frac{26}{50}\right)^2}$$

$$= \sqrt{30.16 - (0.52)^2}$$

$$= \sqrt{3.16 - 0.27}$$

$$= \sqrt{29.89}$$

$$= 5.47$$

Note

1. If the data is in the range of 10-20, then the actual mean method is convenient.
2. If the actual mean comes in decimal point, then the assumed mean method should be used.
3. The assumed mean data should be selected carefully.

Note

Case III : When classified data is given.

The standard deviation of questions related to classification data is done by two methods : Long Method and Short-cut Method. Following are the formulas used in these methods :

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N}} \qquad \text{(long method)}$$

and

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2 \times Ci} \qquad \text{(short method)}$$

Example : Find the standard deviation from the following grouped data by long method.

S.No.	C.I.	f	X	$d(X-M)$	fd	fX	fd^2
1	40 – 44	3	42	14.3	42.9	126	613.47
2	35 – 39	4	37	9.3	37.2	148	345.96
3	30 – 34	4	32	4.3	17.2	128	73.96
4	25 – 29	7	27	0.7	4.9	189	3.43
5	20 – 24	3	22	5.7	28.5	110	162.45
6	15 – 19	3	17	10.7	32.1	51	343.47
7	10 – 15	2	12	15.7	31.4	24	492.98
		$N=28$	$\Sigma fX = 776$				$\Sigma fd^2 = 2035.72$

$$M = \frac{\Sigma fx}{N} = \frac{776}{28} = 27.7$$

Now,

$$S.D. = \sqrt{\frac{\Sigma fd^2}{N}}$$

$$= \sqrt{\frac{2035.72}{28}}$$

$$= \sqrt{72.71} = 8.52$$

Important Steps:

1. Find the midpoint *(X)* of each C.I. Related formula is :

$$X = \frac{x_1 + x_2}{2}$$

Note

2. Find the fX in the mid-points of all the class intervals by multiplying them by the respective frequency.
3. Find the sum of all the fX values thus obtained.
4. Calculate the mean using the formula ($M = \Sigma fX/N$).
5. Find the deviation *(d)* using the formula $d = X - M$.
6. Find the fd by multiplying the frequency of each class interval by the deviation of the same class.
7. Find fd^2 by multiplying fd and d.
8. Find the sum of all the fd^2 values thus obtained.
9. Divide the Σfd^2 value by N.
10. Find the square root of the obtained quotient. This will be the intended standard deviation.

Example : Find the standard deviation of the following data by the abbreviation method.

S.No.	C.I.	f	d	fd		fd^2
1	90 – 99	1	+4	4	+40	16
2	80 – 89	4	+3	12		36
3	70 – 79	7	+2	14		28
4	60 – 69	10	+1	10		10
5	50 – 59	9	0	0		0
6	40 – 49	9	–1	–9	–38	9
7	30 – 39	4	–2	–8		16
8	20 – 29	3	–3	–9		27
9	10 – 19	3	–4	–12		47
		$N = 50$		$\Sigma fd = 2$		$\Sigma fd^2 = 190$

Now, S.D. or $\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times C.I.$

Applying the formula, we get,

$$\sigma = \sqrt{\frac{190}{50} - \left(\frac{2}{50}\right)^2} \times 10$$

Note

$$= \sqrt{3.8 - (0.04)^2} \times 10$$

$$= \sqrt{3.8 - 0.016} \times 10$$

$$= \sqrt{3.7984} \times 10$$

$$= 1.95 \times 10$$

$$= 19.50$$

4.9 Graphical Representation of Data

In research, data can be presented in many ways. Dot graphical representation of data can show important facts easily and simply. In fact, it is a meditative method as well as an effective and economical method from the point of view of making the results clear, simple, understandable and alive. Line drawings also make complex and monotonous subject matter simple and alive. With the help of diagrams, even complex subjects can be presented in an easy way.

Benefit

1. It describes the correlations.
2. It helps in forecasting the future.
3. Helps in comparing two or more variables or situations.
4. Complex things can be explained easily.
5. It is easy for the learners.

4.10 Methods of Graphical Representation

For graphical representation, the researcher can make a graphical representation of the data compiled by him himself or with the help of MS Excel in the computer. One can use Chart Wizard for this. MS Excel has this nice feature that on inputting the data, it can display any type of chart or graph immediately. It has different types of charts as follows :

- Column
- Bar
- Line

Note

- Pie
- XY (Scatter)
- Area
- Doughnut
- Lader
- Surface
- Bubble
- Stock
- Cylinder
- Cone
- Pyramid

4.10.1 Line Graph

It is easy to make and present. See Figure 1 In this, the educational status of the readers has been displayed in a diagram.

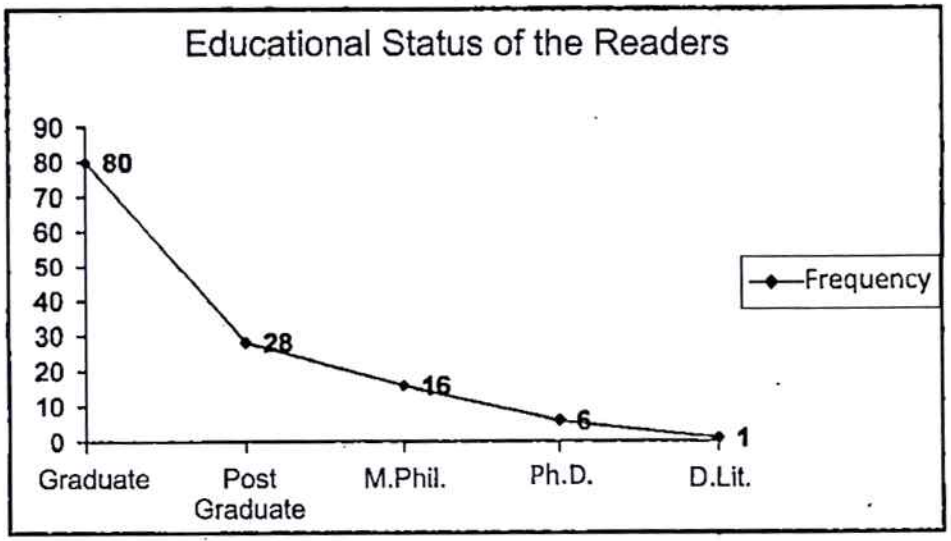


Figure1

4.10.2 Circle Graph or Pie Diagram

This type of diagram is used to display data where an information is displayed as a percentage. Its construction method is simple since the circle is 360 degrees. Therefore, as a percentage, the whole circle will be considered equal to 100.

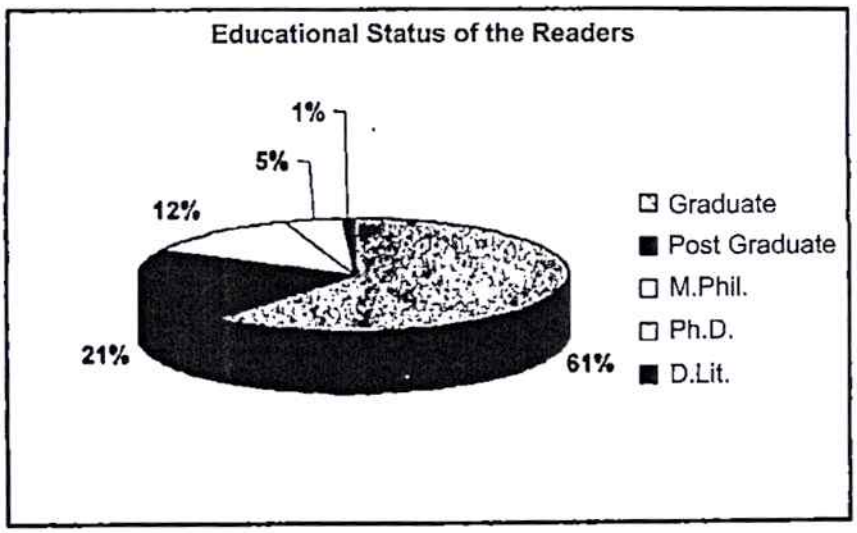


Figure 2

4.10.3 Column Chart

The data is also displayed through the bar. In this bar means bar or stripe, in which the stripes are drawn horizontally in proportion to the value of the variables. It should be noted that the width of the bars is the same. With the help of computer it becomes very easy.

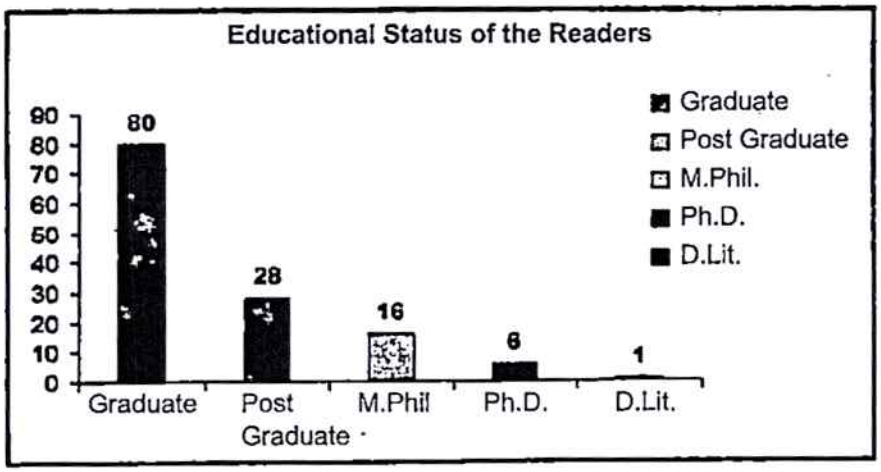


Figure 3

Testing of Hypothesis : T-test and chi square test have been expanded as follows :

T-test : T-test has been described under primary examinations. If only groups are formed at independent variable, *i.e.*, independent variable is manipulated in only two categories and only these two

Note

groups are to be compared at dependent variable then in that case T-test is used goes. The T-test is a systematic test and consists of a single independent variable that is handled in two or more categories. Such research designs are called "single variable designs". From this point of view there are two types of T-test :

(i) Independent group and

(ii) Co-related or reconstructed.

In case of free group the units of both the groups are different. In the case of correlative groups, units of the same group are subjected to two treatments. If the units are equalized and distributed in both the groups, then even though the units of the groups are different, they are considered to be correlated. The T-test can be used in both the inferior conditions, but the formula for both is different.

From a second point of view, both the above types of T-tests have two forms :

1. Micro-specimen and 2. Macro-specimen

Different formulas are used in these two situations as well. Thus the total of T-test are four types :

1. **Large-sample Independent Group T-test :** This test is used where the number of groups is two, the group is independent, the sample is large and the condition of the parameter tests is satisfied. In this, groups are compared on the basis of difference of their mean. The hypothesis that it is tested is as follows, "There is no significant difference in the mean between the groups." For this, the difference of the mean values is converted into T-value or C.R. Value. The obtained t or C.R. is compared with the value given in the T-table at the pre-selected significance level (.05 or 01). While looking at the value of T in the table, the degrees of freedom also have to be taken into account. If the obtained t value is equal to or greater than the table value at that significance level and at that degree of freedom, the non-substantiated hypothesis (H_0) is rejected and the alternative hypothesis (H_1) is accepted. If this obtained value is found to be less than the table value, then H_1, is rejected by not rejecting H_0. Thus, indirectly *i.e.,* by not rejecting H_0, conclusion is

Note

drawn about the research hypothesis. It will be clear by the *i.e.*, given below :

Analysis Option : 11 years old with high achievement motivation and low achievement motivation. By forming two independent groups of children, separate mean and deviation values of their academic achievement (test scores) are derived. The following *example* table presents the contents of :

Table
Relation between achievement motivation and educational achievement

	Group-I (low motivation)	Group-II (high motivation)
Mean	88.80	90.5
Mean scale	7.80	11.56
No. *(N)*	83	97

This is an *i.e.*, of a large sample independent group. In this, the process of finding the C-Ratio (C.R. or T) is as follows :

Gradation 1 : Find the S.E. of each mean. Its formula is :

$$\text{S.E.} = \frac{\text{S.D.}}{\sqrt{N}}$$

So $$\text{S.E.}_1 = \frac{7.81}{\sqrt{83}} = 0.857$$

Similarly, $$\text{S.E.}_2 = \frac{11.56}{\sqrt{95}} = 1.86$$

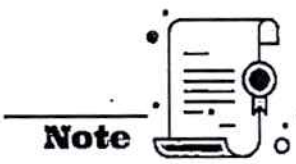
Gradation 2 : Now on the basis of these, find the S.E._D (S.E. of difference), its formula is :

$$\text{S.E.}_D = \sqrt{\left[(\text{S.E.}_1)^2 + (\text{S.E.}_2)^2\right]}$$

According to this formula,

$$\text{S.E.}_D = \sqrt{\left[(0.857)^2 + (1.86)^2\right]}$$

$$= 1.463$$



Gradation 3 : Now find the value of T. Its formula is :

$$t = \frac{D}{S.E._D} = \frac{1.70}{1.463} = 1.162$$

Gradation 4 :

Test of Significance : For this the t-value obtained at 196 *df* and 0.5 level has to be compared with the team value (1.97) given in Table (B). It is clear that the obtained T value of 1.162 is less than the tabular T value of 1.97. Hence, this difference is not significant at 0.05 level. If it is not meaningful at this level then it will not be meaningful even at 0.01 level. Hence H_0 cannot be rejected. As a result H_1, (there is a difference) is rejected.

Gradation 5 :

Conclusion : So it was concluded that there is no real difference between the mean values of both the groups. That is, there is probably no relation between achievement motivation and academic achievement.

2. **Large-sample Correlation Group T-test :** The method given in the example 1 is used where the groups are independent, but when the groups are correlated or matched, the following method is used. In this, the formula for finding the quantifier (S.E) remains the same as in *example* 1, but the formula for extracting $S.E._D$ is as follows :

$$S.E._D = \left(S.E._1^2 + S.E._2^2 - 2r_{12}\, S.E._1\, S.E._2\right) \quad \text{..... Formula (2)}$$

Or $$\sqrt{\left[\frac{1}{N}\left(S.D._2^2 - 2r_{12}\right)\left(S.D._1 \times S.D._2\right)\right]}$$

Using the second formula, the $S.E._D$ value will be the same.

$$t = \frac{D}{S.E._D}$$

This is made clear by the example given below :

Example 2 : A personality balance test was given to a group of 625 eleven year old boys. On this their mean and S.D. were found to be 96.7 and 8.0 respectively. After this he was given

Note

group counselling treatment for two months. After two months, he again gave the same balance test. This time the mean value of their balance was found to be 106.7 and S.D. 10.0. The correlation between the two time balance tests was 0.52. Was the group counselling treatment effective? To get an answer to this question, one has to test the significance of the difference between the mean values of the balance test for both the times, because both the groups are correlated. Therefore, formula-2 has to be used to find the $S.E._D$, of the difference of their mean values. The whole process will be like this :

(i) $S.E._1$ (Mean of the first time)

$$= \frac{S.D.}{\sqrt{N}} = \frac{8}{\sqrt{625}} = \frac{8}{25} = 0.32$$

(ii) $S.E._2$ (Mean of second time)

$$\frac{S.D.}{N} = \frac{10}{\sqrt{625}} = \frac{10}{25} = 0.40$$

(iii) $S.E._D$ $= \sqrt{\left(S.E._1^2 + S.E._2^2 - 2r_{12}\ S.E._1\ S.E._2\right)}$

$$= \sqrt{\left[(0.32)^2 + (0.04)^2 - 2\times 0.52\times 0.32\times 0.40\right]}$$

$$= 0.3596$$

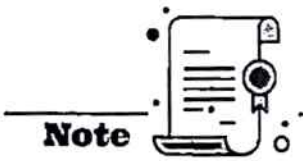
Using the second option formula, the $S.E._D$ value will be the same.

(iv) $$t = \frac{D}{S.E._D} = \frac{106.7 - 96.7}{0.3596}$$

$$= \frac{10}{0.3596} = 27.80$$

This t-value is much higher than the value given in the table at 624 $df\,(N-1)$ at each level. Hence H_0 is rejected and it is concluded that group counseling treatment has a definite effect on the balance.

Calculation of Independence : It is necessary to remember that in the case of independent groups, the calculation of independence is done on the basis of $n_1 + n_2 - 2$ formula. The



reason for this is that in the case of correlated and equalized groups, the same person is tested twice, so a correlation is established between their modulus. The freedom to change the scores obtained on the second test decreases because the group's second time achievement cannot be completely different, independent of the first time, as the group remains the same. Therefore, the modulus of the second time cannot be completely independent from the one which came for the first time. Therefore, the modulus of both the times of a person is given as one df instead of two df. Thus to find in the position of the correlated df groups only $N-1$ formula is used.

3. **Small-sample Independent Group T-test :** If the sum of the units of both the independent groups is less than 30, then it is considered as small sample. In this situation S.D. and $S.E._D$ extraction formulas vary. The rest of the process remains undone. S.D. and the formulas for extracting $S.E._D$, are as follows :

(i) $$S.D. = \sqrt{\left(\frac{\Sigma x^2 + \Sigma y^2}{N_1 + N_2 - 2}\right)}$$

In which Σx^2 and Σy^2 are taken from the respective mean of the scores of both the groups respectively is the sum of the squares of the differences.

(ii) $$S.E._D = S.D.\sqrt{\left(\frac{N_1 + N_2}{N_1 \times N_2}\right)}$$

The whole process of estimation is explained by the example given below :

Table

Group X			Group Y		
X	x	x^2	Y	y	y^2
110	4.5	20.25	115	2.0	4.0
112	6.5	42.25	112	– 1.0	1.0
95	– 10.5	110.25	109	– 4.0	16.0
105	– 0.5	0.25	112	– 1.0	1.0

Note

111	5.5	30.25	117	+4.0	16.0
97	−8.5	72.25			
112	6.5	42.25			
102	−3.5	12.25			
Sum = 844		330	565		38
N 8		Σx^2	5		Σy^2
Mean = 105.5			Mean = 113.0		

Steps:

(i) Pooled S.D. $= \sqrt{\left(\frac{\Sigma x^2 + \Sigma y^2}{N_1 + N_2 - 2}\right)}$

$= \sqrt{\left(\frac{330+38}{8+5-2}\right)} = 5.78$

(ii) S.E. = Pooled S.D. $\times \left(\frac{N_1 + N_2}{N_1 \times N_2}\right) = 5.78 \times \sqrt{\left(\frac{13}{40}\right)} = 3.302$

(iii) C.R. $(t) = \frac{\text{Diff.}}{3.302} = \frac{7.5}{3.302} = 2.271$ for df 11

This method of analysis is very important in the context of laboratory studies because in these studies the result of the sample is small, but in this situation it is necessary to check the homogeneity of the groups. This method can be used only if there is homogeneity of variance.

4. **Small Sample Correlated Group T-test :** In this case also the formula for finding $S.E._D$ remains same as that used in case of large sample correlation group *i.e.*, this situation,

$$S.E._D = \sqrt{S.E._1^2 + S.E._2^2 - 2r_{12} \times S.E._1 \times S.E._2}$$

But there is another way to get it. Its basis is the distribution of the difference of the modulus of both. The mean and S.D. of the differences are derived. These are named as D and S_D respectively. Based on these $S.E._D$ is derived whose formula is

$$S.E._D = \frac{S_D}{\sqrt{N}}$$

Note

This S.E.$_D$ is converted to T value by the formula $\frac{D}{\text{S.E.}_D}$. From the following example it will become clear :

Table

X	Y	D	D^2
100	105	– 5.0	25
90	96	– 5.0	25
106	100	+ 6.0	36
95	105	– 10.0	100
102	99	+ 3.0	9
104	98	+ 6.0	36
90	110	– 20.0	400
105	106	– 1.0	1
99	100	– 1.0	1
		$\Sigma D = -27.0$	$\Sigma D^2 = 633$

(i) $\overline{D} = \frac{-27}{9} = 3.0$

Let $d = (D - \overline{D})$, a deviation score in D. Then:

(ii) $\Sigma d^2 = \Sigma D^2 - \frac{(\Sigma d)^2}{n} = 633 - \frac{(-27)^2}{9} = 552.00$

(iii) $S_D = \sqrt{\left(\frac{\Sigma d^2}{n-1}\right)} = \sqrt{\left(\frac{552.0}{8}\right)} = 8.306$

(iv) $\text{S.E.}_D = \frac{S_D}{\sqrt{N}} = \frac{8.306}{\sqrt{9}} = 2.768$

(v) $Z(t) = \frac{D}{\text{S.E.}_D} = \frac{-3.0}{2.768} = 1.08$

If the value of t is calculated using formula 2, then the same will come as above obtained method.

The value of T at 8 $df(N-1)$ and 0.05 is found to be 2.31 in the table. The obtained T value is 1.08 which is less than the table value. Hence H_0 cannot be rejected. That is why the research hypothesis (H_1) *i.e.*, "there is a difference" is rejected and the

Note

conclusion is established that there is no significant difference between X and Y modulus.

This method can also be used in case of large sample co-related groups. This method is very simple from the point of view of estimation, but very little information is obtained on its basis. This does not give knowledge of the correlation between the modulus of the two groups, nor does it reveal the mean values and scores of S.D. the groups.

But it should be remembered that the methods of finding the value of T in both the case of large and small sample correlated groups are the same, but if the groups are made equal on the basis of mean and S.D., then the formula to calculate $S.E._D$ is as follows :

$$S.E._D = \sqrt{\left[S.E._1^2 + S.E._2^2 \times \left(1 - r_{xy}^2\right)\right]}$$

The rest of the process remains the same even in this situation.

Chi-Square Test :

Chi-square (x^2) is a nonparametric statistical method to test whether a distribution of frequencies obtained from a research is significantly different or net from the distribution of frequencies expected on the basis of a hypothesis. If not different then both of them are considered related. Thus, chi-square is a test method for the mutual relationship of two variables or between two frequency-variables. This analysis method is also used to find out whether the differences between the received and expected frequencies are just accidental or real. Also to find out, this Analysis Method is resorted/used. There are many types of this option. Mainly used in four different situations goes. These are :

1. Single group position.
2. Position of two independent groups.
3. Contingency Analysis
4. Position of more than two independent groups.

1. **Single Group Position :** This method is used where the obtained frequency distribution is compared with a distribution

Note

which is a standard distribution and which is based on a certain hypothesis such as hypotheses of an even distribution or similar is expected based on the hypothesis of distribution. It will be clear from the below example:

Example 1 : A researcher wants to know whether educated women really like to work. To find out, he selected 48 educated women and asked them whether they would like to work. They were asked to choose one option out of three options ('yes', 'no', 'neutral'). The frequencies of each answer he got 24, 12, 12. On the basis of the uniform distribution he tested whether the distribution obtained (24, 12, 12) differs significantly from the uniform distribution (16, 16, 16). His belief was that if women actually like to work, then this distribution should be different from the expected distribution (16, 16, 16) example in other words, if they want to work or not, their particular inclination towards someone. If not, then the distribution of 48 women should be equal (16, 16, 16) under all three option answers. He tested whether the distribution obtained from this distribution is significantly different or not by chi-square. For this, he first prepared the table as follows and based on that the value of chi-square was derived. The complete method of this calculation is given in the below :

Table
One Sample Chi-Square Test

	Preferred	Not-perferred	Indifferent	Total
Fo	24	12	12	48
Fe	16	16	16	48
$Fo - Fe$	8	4	4	
$(Fo - Fe)^2$				
$\frac{(Fo - Fe)}{Fe}$	64 / 4	16 / 1	16 / 1	
				$6 = X^2$ value

In the above table, *Fo* means received frequencies, *Fe* means expected frequencies. This single sample is an *i.e.*, of chi-square because research material has been collected from only one sample (48 females) *i.e.*,

Note

only one sample has been used. The formula for finding the chi-square is :

$$X^2 = \Sigma \frac{(Fo - Fe)^2}{Fe}$$

In which *Fe* means expected frequencies and *Fo* means received frequencies.

The chi-square value of each cell of the table has been derived by the above formula. Later the values of all the cells are added up. This whole chi-square came to be 6.0.

If the chi-value obtained is equal to or more than the value given in the table, then it is considered as a meaningful is believed. Rest of the problem procedure and logic remain exactly like T-test.

Calculation of Independence *(df)* : The concept of df is also used in the use of table of calculation of degrees of freedom *(df)*. The formula to find it is $(\rho - 1) \times (k - 1)$ *i.e.,* subtracting one each from the number of rows and columns multiply the remainder by each other. In the given example, there are 3 columns and rows. Thus, $df (3 - 1) \times (2 - 1) = 2$.

Significance : In table 2*df* and at 0.5 level the value of chi is 5.991. The value obtained is 6.0 which table value exceeds. Therefore, rejecting the hypothesis *Ho* that "there is no difference between the two distributions", it will be concluded that there is a significant difference between the distributions. In other words, women prefer to work because the frequency of "yes" answers is highest. If the frequency of "no" answers were high, then the conclusion would be that they did not like to do the job. So what exactly will be the conclusion, it depends on the observation of the frequencies of the cells. If a one-sided test is to be done, then the value of Chi in the table, for level 0.05 should be seen in column 0.01 and for 0.01 in column 0.02.

Hypotheses of Expected Frequencies : Hypothesis of "Equal Distribution" can be used to find the expected frequencies in case of a single sample as explained in the *example* given below. Apart from this "Even-Distribution" Hypothesis can also be used. Which one should be used depends on the research situation. *For example*, if we want to know whether the marks distribution of a class is even or not then

Note

the expected frequencies have to be worked out on the basis of even distribution.

Limitations of Chi-Square : The use of Chi-square test is prohibited in the following circumstances :

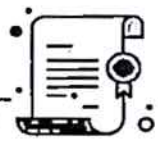
(i) When *df* is only 1 *example* k (group) = 2 and any expected frequency is less than 5.

(ii) When df is greater than 1, but more than 20% of the frequencies are less than 5 or any expected frequency is less than 1.

2. Two Independent Group Algorithm : In this case two independent groups are compared on the basis of chi-square, just like in t-test by taking the value of t. The difference is that in this, its significance is tested by taking the value of chi-square in place of t and the conclusion is reached whether the two groups are different from each other from the point of view of any characteristic or dependent variable. "The two groups are not different from each other" is the basis of the hypothesis *(Ho)* test. Therefore, the chi-squared parameter test is the only non-parametrical alternative to T. Where t-test conditions are not met, chi-squares can be used where they cannot be used. *For example*, suppose a researcher wants to know whether "there is a difference between men and women in their willingness to work". He asks such a question to some women and men and receives research material in the form of three answer options "yes", "no" and "neutral". In this situation the chi-square test can be used.

Method of Use : In this case of using chi-square also the same method as described earlier in case of single group will be followed. The difference between the two is that in this case instead of a group There are two groups. By preparing the table, the value of the chi-square of each cell has to be worked out and at the end all chi-squares will have to be obtained by adding them all. The formula for finding the chi-square of the square is also the same as before happens *i.e.*,

$$\text{Chi-square} = \Sigma \frac{(Fo - Fe)^2}{Fe}$$



The method of computation is explained by the example given below :

Example 2 : Suppose the above question was asked by the researcher 90 men and 100 women and got answers in the form of three types of options 'yes', 'no' and "neutral". A table was prepared by preparing the frequency distribution of those answers and to test the hypothesis that "there is no significant difference in the job desire of men and women." The following table shows the research material and the process of computing.

Table
The relationship between employment needs and sexuality

		Response		Mode	
		Yes	No	Indifferent	Total
Male	Fo	14.0	66.0	10.0	90
	Fe	19.42	62.52	8.05	
	$Fo - Fe$	5.42	3.48	1.95	
	$\frac{(Fo-Fe)^2}{Fe}$	1.51	0.20	0.47	
Female	Fo	27.0	66.0	7.0	100
	Fe	21.57	64.44	9.0	
	$Fo - Fe$	5.43	3.49	2.0	
	$\frac{(Fo-Fe)^2}{Fe}$	1.30	0.18	0.49	
Total		41	132	17	190

The solution to this problem is to calculate the frequency of each cell in the chi-square value and finally, all the obtained values will be solved by the following procedure :

(i) Fe (1) *i.e.*, cell-1 $= \frac{41}{190} \times \frac{90}{0} = 19.42$ (out of 190 total Fe for 'yes' response = 41; how many out of 90, total of first now).

Note

(ii) Fe (2) *i.e.*, for cell-2 $= \frac{132}{190} \times \frac{90}{1} - 62.52$ (out of 190 total Fe for 'No' response = 132; how many out of 90).

(iii) Fe (3) *i.e.*, for cell-3 $= \frac{17}{190} \times \frac{90}{1} - 8.05$ (out of 190 total Fe for total cell-3, indiff. response = 17; how many out of 90).

(iv) $Fe\,(4) = \frac{11}{190} \times \frac{100}{2} = 21.57$ (same argument).

(v) $Fe\,(5) = \frac{132}{190} \times \frac{100}{1} = 64.44$ (same argument).

(vi) $Fe\,(6) = \frac{17}{190} \times \frac{100}{1} = 9.00$ (same argument).

After this, we find out ($Fo - Fe$) for each cell. These values are also given in the table for each cell.

After this, take the square of each ($Fo - Fe$) and divide it by Fe and convert it into a chi-square. These come to 1.51 + 20 + 0.47 + 1.30 + 0.18 + 0.49 respectively. The sum of all these means $\Sigma X^2 = 4.51$.

Significance of Chi-Square : In the table, the value of chi at ($2 \times 1 = 2df$) and at level 0.05 is 4.991, but the value obtained is 4.15 which is less than the table value. Therefore, the hypothesis of *Ho*, *i.e.*,"no difference" at level 0.05, cannot be rejected. Consequently "there is no difference" has to be accepted because the value of the chi-square is not significant. It could also happen by chance.

Conclusion : So it was concluded that sex has no relation with job desire.

3. **Contingency Analysis (Table 2 × 2) :** If the distribution of frequencies is presented 2 × 2 in a table, there is another method of chi-squared calculation, which is much simpler and is used in most situations. In this, the number of rows (r) is also 2 and the number of columns (k) is also 2. So, in this $df = (r - 1)\,(k - 1)$ *i.e.*, $(2 - 1)\,(2 - 1) = 1$. The formula for finding chi-square is as follows :

Note

$$X^2 = \frac{N\left[|AD - BC| - \frac{N}{2}\right]^2}{(A+B)(C+D)(A+C)(B+D)}$$

In which $|AD - BC|$ means that AD can be subtracted from either BC or BC from AD. In the above formula $\frac{N}{2}$ is subtracted $(AD - BC)$ which is called Yates-Correction. The reason for this is to make the obtained chi-values according to their actual, that is, theorized and expected chi-square distribution. According to Morone (1954), a tabular distribution is a bionomial distribution that is not continuous, whereas a chi-square distribution is continuous. So this correction is done to make the binomial distribution similar to the chi-square distribution.

Example 3: A researcher wants to know whether the economic and social status of students affects their academic achievement. He selected 48 students from low level economic social level and collected their academic achievement by dividing them into successful or unsuccessful two categories. He arranges this research material in a table of 2 × 2 and tests the hypothesis by finding the value of chi-square on the basis of contingency formula. The following table and the corresponding computation are given below :

Table : Exam Results

	Failed	Passed	Total
High SES	36 (A)	12 (B)	48
Low SES	32 (C)	16 (D)	48
Total	68	28	96

To find the answer to the question present in this problem, a chi-square test can be adopted. In this, by making the data a contingency table of 2 × 2, then putting the values in the formula :

$$X^2 = \frac{N\left(|AD - BC| - \frac{N}{2}\right)^2}{(A+B)(C+D)(C+A)(B+D)}$$

Note

$$= \frac{[96|(36)(16)-(12)(32)|-96/2]^2}{48\times48\times68\times28}$$

$$= \frac{96[576-384-48]^2}{4386816}$$

$$= \frac{1990656}{4386816} = 0.4537 \text{ for } df$$

Significance and Conclusion : In the table, the value of chi square at 1 *df* and 0.05 level is 3.84 obtained.

The chi-square value is 0.4537 which is much less than the table value. Therefore, *Ho.* Hypothesis "There is no difference in academic achievement of both the groups (HSES and LSES)" cannot be rejected. Obtained chi-square is not significant *i.e.*, it could also happen by chance. The economic social level of the students did not affect their educational achievement. There does not seem to be any relationship between the economic social level and the educational achievement.

Limitations of the 2 × 2 Contingency Method : If this condition is not met, the frequencies of the 2 × 2 table can be reduced by adding nearby frequencies and reducing the number of cells can be increased, even then it is important to keep in mind the following points :

(i) When the size of *N* is more than 40, then purification must be used for integrity.

(ii) When *N* is between 20-40, use the above formula only if all the expected frequencies of the cells are 5 or more than 5.

Caution : Chi-square should not be used when N is less than 20.

4. **More than Two Independent Groups :** Even if more than two independent groups are to be compared Chi-square can be used. In this case the frequencies are arranged in the $K \times r$ table. The irreputable hypothesis (*Ho*) remains as before that the samples are not different from each other. The chi-square test is also done in the same way as described back, but the method of finding the expected frequencies of each cell is different. This will be described further by an *example.*

Note

This method is a nonparametric alternative to the one-way analysis of variance. If in a research situation more than two groups have to be compared but the conditions of the F-test are not fulfilled, then it can be used there.

Example 4 : A researcher measures the intelligence of 100 students in a class and divide them into three categories to find out whether there is a relationship between intelligence and academic achievement; Such as, more intelligent, simply intelligent and less intelligent. Similarly divides their academic achievement into four categories; *For example*, taking first, second, third and failing grades in which all the students come, prepares a table of 4 × 3 whose format and the calculation of the chi-squares are presented below it.

		Failed	III division	II division	I division	Total
Less intelligent	O	5	6	9	8	28
	E	5.88	8.40	7.84	5.88	
	O-E	0.88	2.40	1.16	3.88	
Ordinary intelligent	O	5	16	13	8	42
	E	8.82	12.00	11.76	8.82	
	O-E	3.82	3.40	1.24	0.82	
Intelligent Less intelligent	O	11	8	6	5	30
	E	6.30	9.0	8.40	6.30	
	O-E	4.70	1.0	2.40	1.30	
Total		21	30	28	21	100

In the table, the received frequencies of each cell are given against O and the expected frequencies against E. To find the expected frequencies, multiply the row sum and column sum of each cell and divide by the total number of frequencies (N). There is such a logic behind it. Take the first cell. In this, all the persons are those who come in the first row and first column. The probability of any one person being in this row will be that sum of the total number (N) of this row will be

Note

divided by 100 *i.e.,* it will be 28/100. Similarly the probability of that person being in the first column will be 21/100. So the probability of it being in the first row and column will be $\left(\frac{28}{100}\times\frac{21}{100}\right)$. This is the probability that happens out of one. The total frequencies are 100.

So to find out how many they will be out of 100, the probability is $\left(\frac{28}{100}\times\frac{21}{100}\right)$ multiply by 100.

Thus, the total expected frequencies of the first cell will be $\frac{28}{100}\times\frac{21}{100}\times\frac{100}{1}=5.88$. Similarly, the expected frequencies of all other cells can be multiplied by their row sum and column sum and divided by 100. These are listed in the table. Based on the following formula, we can find the chi-square of each cell :

$$X^2=\frac{(O-E)^2}{E} \text{ or}$$

$$(0.88)^2/5.88+(2.40)^2/8.40+(1.16)^2/7.84+(3.88)^2/5.88$$

$$+(3.82)^2/8.82+(3.40)^2/12.60+(1.24)^2/11.76+(0.82)^2/8.82$$

$$+(4.70)^2/6.30+(1.0)^2/9.0+(2.40)^2/8.40+(1.30)^2/6.30$$

Finding the value of the above chi-square =

$$0.774+0.686+0.172+2.560+1.654+0.917+1.538+0.76$$

$$+0.506+0.111+0.686+0.268=12.948.$$

Significance and Conclusion

In this example, $df=(4-1)(3-1)=6$|.

The chi-square at *df* 6 and 0.05 level in the table is 12.59. The value obtained in the *example* is 12.948. which is greater than the table-value. Therefore, I reject that "IQ does not affect academic achievement or that intelligence has no relation with academic achievement" and the conclusion is established that intelligence affects academic achievement or that both are related to each other.

Note

Exercises

Very Short Answer Questions

1. Write the types of analysis.
2. Write the types of measurements under statistical analysis.
3. How many types of measures are there in central tendency?
4. What is the mean?
5. Write the formula to find the mean by averaging method.
6. Write the precautions to be taken while calculating the mean.
7. Write the meaning of median.
8. Write the formula for finding the median.
9. Write the utility and importance of median in the field of education.
10. Write the definition of standard deviation.

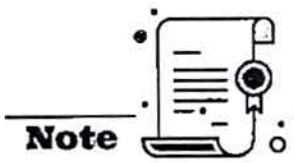
Short Answer Questions

1. Describe qualitative analysis.
2. Describe the analysis of data by descriptive statistic.
3. How to calculate the mean of unsorted scores? Explain.
4. Describe the graph, circle and bar method followed in dot-line representation.
5. What is T-test? How many types are these?

Long Answer Questions

1. Write the meaning of mode. Describe the method of finding the mode from the ungrouped scores.
2. Describe the types of T-test.
3. Describe in detail the chi-square test.

UNIT-5
Bibliometric Law



UNIT 5

Bibliometric Law

5.1 Introduction

Statistical or computational methods are being used continuously in many subject areas, mainly in the social sciences and natural sciences. The use of statistical methods in library and information science is also a link in this chain. Bibliometrics, Scientometrics, Informatics, Webmatrics, Cybermetrics, etc. are different names for the use of statistics in the fields of library and information science. Bibliometric acts as a guide for proper management of texts etc. The use of these statistical calculations has also become possible in computerized information technology. The related terms of bibliography are Scientometrics, Informatrics, Webmatrics, Cybermetry, Socimatrics.

5.2 Origin of the Bibliometrics Term

Origin of Bibliometrics Term with the development of library and information science as a subject, librarians started using mathematical or statistical methods in one way or the other and they gave many names. F. J. Cole and N. B. Eales analyzed books published between 1550 AD and 1860 AD in the development of subjects in 1917. They called it "Statistical Analysis". E.W. Hulme named statistical calculations in the library as Statistical Bibliography in 1922. Alan Pritchard first used the term Bibliometrics in 1969 in his article 'Statistical Bibliography or Bibliometrics'. In this way the word Granthamiti originated which is valid even today. Many new words have also been used as its branch. Scientometrics by Russians, Informatrics by FID and in addition Socimatrics, Webmatrics,

Note

Cybermetrics etc. are prominent. In India also, D. Ranganathan used librametry in 1948.

5.3 Definition of Bibliometrics

The term bibliometrics is derived from the combination of two words Biblio + Metrics. The word Biblio is derived from the Greek compound word Biblion which means text or paper and Metrics is derived from the Greek word Metrikos, which means the science of measurement or measurement. In a general sense, the term Granthamiti be used for any form of quantitative analysis related to the production, distribution and use of published and semi-published literature.

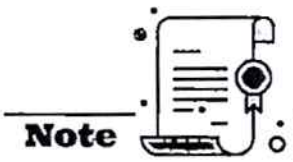
Many scientists have defined bibliometrics in different ways. According to **Alan Prichard,** Application of Mathematics and Statistical methods to books and media of communication. **Potter** has defined bibliography as "The study and measurement of the publication patterns of all forms of their authorship". According to **Sengupta,** bibliography is the study of the publication, organization, classification and quantitative evaluation of patterns of all types of long and subtle communications in which they are characterized by their creativity Arithmetic and Statistical calculations.

Thus we can summarize that the quantitative analysis of the characteristics, behavior and presentation of all aspects of written communication is called bibliography. Now electronic medium has also been included in the written communication medium, due to which it has come to be known by new names like Webmatrics and Cybermetrics etc.

5.4 Need and Objective of Bibliometrics

5.4.1 Need of Bibliometrics

A librarian, like a manager of an industry, is a manager of a library. He also needs objective data for decision making. Statistical references, Cost of library collection (Value distribution of documents), Cost and replacement of both documents and non-book material, Library performance indicators, Circulation statistics, Library timings, Number of readers and references, Date and place



of publication, Educational program, Publication Such basic bibliographic data, distribution etc. is useful for policy making. Therefore, bibliography is an essential quantitative data in libraries, which libraries are always in need of.

5.4.2 Objective of Bibliometrics

Bibliometrics includes the development of published literature in all subjects and the amount of literature publications contributed by individuals, groups, institutions or nations. In this, the volume and scattering of publications in different languages and subjects (types of documents, languages, journals) and out of date etc. are also studied. Along with this, the study of the sources referenced by the authors is also a part of the first initiative, from which information about the most cited authors and magazines etc. can also be obtained.

In bibliography, special emphasis is placed on quantitative calculation and not qualitative evaluation. On the objective of bibliometrics **Hume**, clearly said that this spot the light on the four written process and nature direction of development of any subject.

The objectives of bibliography are as follows :

1. To create a more economical and important system.
2. Improving the efficiency rate of information processing.
3. To identify and measure the defects in the functional services.
4. To discover and explain the Empirical Laws that of informatics to provide the basis for developing theories.

Qualitative experiments emphasize the practical use of study findings.

Some important findings useful for Library management are as follows :

1. Identification of major publications.
2. Sorting the publications in order of decreasing importance.
3. Establishing transition point between zones of higher and lower utility.
4. To find the impact value of documents etc.

5.5 Types of Bibliometrics

Bibliometric is mainly of two types :

5.5.1 Descriptive Studies

It includes measurement of characteristics of documents or publishers, such as nature, quantity, time and frequency of information, study of origin and geographical scattering or distribution, etc.

5.5.2 Behavioural Studies

Under this, the Authorship of the documents, the Title Statement, the bibliography history (version, change in the number, date etc.), the frequency of publication (in relation to the current publication), the Form, Place, Language, Subject, Related document, Content Analysis, Physical Make-up, Shape. Publication, Quality, Price, Circulation, Translation, Formats, Citation, etc. are studied.

5.6 Utility of Bibliometrics

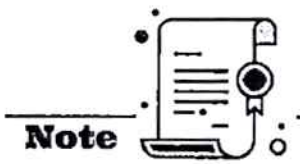
Bibliography is used for the following purposes :

1. To calculate the productivity of literature.
2. Area of study, quantity and type of research in different branches, different parts of the regions, different for comparative study of publications derived over time period.
3. For the type and amount of publications.
4. For scattering and pattern of publications.
5. For Impact of Publications.
6. To identify social change etc.

5.7 Uses of Bibliometrics

The use of bibliography can be cited as follows :

1. The aim of the study of bibliography is to improve the vocal control. Vocabulary analysis helps in knowing the size and characteristics of literature (publications) in different fields. The growth and quantity of primary literature directly affects the structure of secondary literature. Therefore, computed growth rate and change direction can provide important help



in determining the future approach and coverage of secondary service editors.

2. A major area of bibliography study is the statistical determination of the origin, nation, subject, form and language of documents as well as translations of related literature. These data provide useful information related to the determination of scope and may suggest improvements in the assessment of possible improvement of secondary services.
3. The study of bibliography derives the relatedness of the subjects, which are similar to those of common patterns can give the required suggestions for the secondary service area (Coverage). Such bibliographic studies can also help establish a roadmap for service in a specific subject area.
4. Use of citation analysis data and year wise publication volume to make Vocallist can be done in the plan.
5. Bibliographic analysis helps in comparative evaluation of secondary services. It can also shed light on the achievements of publishers.
6. Bibliographic data are also helpful in taking some managerial decisions. *For example,* it is helpful in the selection of specialized primary and secondary journals and helps in the improvement of future requirement, staff, building requirement and library services.
7. Citation data also determines the highly cited periodicals or books which are available in the library can also be useful in decision making for collection.
8. Bibliography provides subject relatedness through citation analysis which also helps in suggesting publications of journals by specific subject in a particular library. The study of bibliography also provides information about the structure of knowledge and the nature of communication.

   Thus, the analysis of the size and growth of bibliography publications can identify the growth in the Decling area and the growth of literature (publications) method can be known.

Note

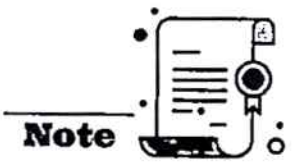
Bibliography also has the following other uses:

1. To provide quantification of research and enhancement of various fields of knowledge.
2. To estimate the prevalence of secondary journals.
3. To identify the authorship and user of documents in different subjects.
4. To measure the utility of Ad-hoc and Retrospective Selected Extension Service (SDI).
5. For need based acquisition policy within limited budget without any impact on the research interest of the parent institution.
6. To identify leading journals in various disciplines.
6. To usher in an effective multi-layered network mechanism.
8. To regulate the flow of information and communication
9. To develop norms for standardization etc.

5.8 Problems and Limitations of Bibliomatrics

Through bibliometrics we can evaluate the effectiveness of research and paper productivity etc., but bibliometrics judges have their own problems and limitations as follows :

1. Full address of the authors of the main article, details of institutional affiliations etc. Opportunity which makes analysis difficult.
2. Institutions do not give full details of the publications of their scientists in the annual progress report.
3. Research scientists do not deliberately give complete information about their research publications, due to which Productometric study becomes difficult.
4. Reference lists are not always accurate and complete, so there is a need to include many incorrect data there is a possibility.
5. Actually the citations are not used at all and are cited in the reference list as they are done in some other article. *For example*, old citations are given for definitions when in reality they are not quoted at all.



6. Often the paper from which the main part of the article is taken is deliberately hidden in the reference list, making the analysis misleading.
7. Often the name of the first author is given in the quotation and sometimes 'and others' is also not given, due to which the information of all the authors is incomplete, the correct analysis cannot be done.
8. Often the quotations get scattered due to the difference in the chromaticity of the authors or in the story.
9. Often the articles themselves are cited unnecessarily. Quotes from friendly authors are also given to oblige.

5.9 Laws of Bibliometrics and Their Use

In addition to Percentage, Mean, Mode and Median, some bibliographic rules have also been prepared. Two of these rules are used for the first count of publications, while the third rule is used for the analysis of references. Some other rules have also been formulated. The details of the three main rules are as follows :

5.9.1 Bradford's Law

Samuel Clement Bradford's law governs the scattering of articles of a specific type in different journals. Hence it is also called the Law of Scattering.

In his paper in Bradford, two rhetorical lists prepared in the Science Library, Britain, in the source of information in the subject-specific, Applied Geophysics Year 1928-31 and Lubrication Year 1931- 32 and arranged the referenced research journals in these hierarchical lists in the decreasing order of the number of references cited. By studying the list of referenced research journals, he found that only a few journals are mostly referenced, *i.e.*, very productive, whereas in most of the remaining journals this productivity is very low and they are referred very rarely. On this basis, Bradford also prepared a graph which he named as Bibliography. He prepared three zones by dividing the total references into three parts almost equally.

Bradford found a relationship between them which was called Bradford's law. 1 : n : n2 according to them.

Note

Here 1 means the number of research journals referred for 1/3 of the references in the first zone. Similarly, here n is the ratio obtained by dividing the number of periodicals referred to the first zone by the number of journals referred to the second zone. The number of magazines required in the third zone should be n 2 of the value of n thus obtained.

It is sometimes difficult to apply this rule in research. Actually it is very simple. First of all, arrange the reference numbers of all the referenced journals in descending order and add the cumulative references from these tables. Then take the value of 1/3 of some of the references. Then note down the number of magazines referred to in the three zones in the above table.

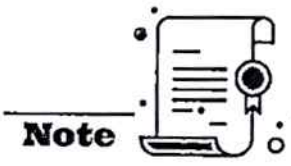
By following this rule, we can divide the research journals of any subject into three parts and the library can acquire research journals of the first group. If the budget is available, we can buy some magazines in the second zone also and decide not to buy the magazines in the third zone, we can ask for some photocopies of them if needed of papers.

Bradford showed with the help of graph which we call Bradford Bigliograph. For this, on line 'A' the log of the cumulative number of referenced magazines is shown and on line 'B' the log of the cumulative number of the number of magazines received in Zone 1, 2, 3 is shown.It should be obtained in a straight line. Each zone is represented by parallel and horizontal lines.

Similarly, other formulas such as the square root formula of Price's productivity, Smarfield's formula for concentration and Sen's formulas are also used in bibliography, which can be studied in detail from the reference texts.

5.9.2 Zipf's Law

This rule was propounded by George Kingsley Zipf. Hence it is called Zipf's rule. This rule is related to the frequency of words used in the text of publications and in the title. The frequency of occurrence of words in a text / title is counted as the number of times a word appears in a text / text. It is also called minimum effort (Law of least efforts).



According to Zipf's law, the words used in a long text content are arranged in decreasing order of frequency, then the position (r) of any given word in the text is opposite to the frequency of the word (f) (Frequency of occurrence) or will be proportional. Zipf's rule can be written mathematically in the following way :

$$r = 1/f$$

or $$rf = c$$

or $$\log f(r) + \log n = c$$

Where,

r = Ranks of word

f = frequency word and

c = is a constant.

We can easily find its value using log table. For this, the frequency of words used in a research publication is counted and arranged in decreasing order and the constants are obtained from the above bibliography formula.

5.9.3 Lotka's Inverse Square Rule

This rule was given by Alfred J. Lotka in 1926. It is also called Inverse Square Formula. According to this rule, the bibliographic calculation of the Productivity of Research Papers is done.

The scholarly productivity of scientists can be measured by the number of published papers according to the general prevailing rules of the research field. Initially, not much attention is paid to other factors. It has been concluded from the preliminary research by Lotka that the articles published in any research journal in this the contribution of maximum writers is very less and there are very few highly productive writers. Lotka propounded a formula for this. According to this formula, if the number of authors who have published only one paper is n, then the number of authors who have published two papers should be equal to the inverse square of this number n, *i.e.*, $1/n^2$, that is, the scientist who has published two papers. (n = 2), it will be only n = 3 of the number of scientists who publish a paper. Similarly, if $1/n^2$ is $1/3^2$ *i.e.*, 1/9 and similarly the value obtained from $1/4^2$, $1/5^2$, $1/6^2$ will be equal to the number of authors who have published

Note

three, four, five or six papers, *i.e.*, one or the number of scientists who have published a large number of papers will continue to decrease in proportion to the number of scientists who have published one paper. By this formula we can find out how much more or less the required number of high productivity scientists is than the actual number and thus we can know the latest trend of publications.

5.10 Some Other Laws

Other Laws

Apart from bibliometrics, some other rules have also emerged, which are described below. Cybermetrics is not in the syllabus but still it is given for the students to understand.

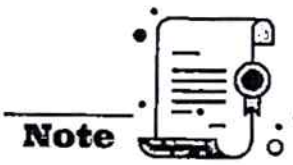
5.10.1 Scientometrics

The term was first used by T. Braunin in 1977 from a journal "Scientometrics" published in Hungary and presently in Amsterdam. The scope of this journal publishes all those studies of the quantitative aspects of science as a discipline or as an economic activity in order to measure the effectiveness of what is achievable from the expenditure incurred in developing it.

Scientometry is part of the sociology of science and 'science' is used for policy making. Scientometrics covers the quantitative study of scientific activities which includes publications. Therefore, to some extent it also includes the bibliometrics. Thus we can say that scientometrics is the branch of "science of science".

The main goal of scientometrics is to determine the state and prospect of a subject and its further development. For this, many Scientometric indicators are used. Out of this the most significant indicator is the publication number. This indicator can be a standard tool of evaluation and analysis in the research management of scientometrics making.

At present, special importance has been given to the measurement of scientific information so that information search can be regularized. Mathematical and statistical techniques have made it possible for science to quantify the social structure or process of quantifying the number of individuals and publishers interested in research.



5.10.2 Informatrics

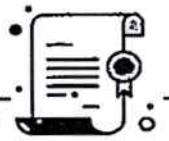
In 1979, Otto Nacke of Germany proposed the term informatrics for the first time. The FID Committee then gave this name. The term was not widely accepted until 1987. Theoretical aspects of information retrieval and bibliometrics but at the First Conference on Bibliometrics and Theoretical Aspects of Information Retrival, Diepenback Belgium it was suggested that the name of Informatrics be included in the name of the Second International Conference. In the proceedings of the conference it was said that the word Informatrics is more effective than the words bibliometrics and scientometrics. Therefore, for the promotion of a new name, it was decided that the use of the word informatrics with the name of bibliometrics should be used in the report of the next conference and in the report of the proceedings of the published conference. Thus the second conference was named International Conference on Bibliometrics, Informatrics and Scientometrics. Thus the word "informatrics" is being used continuously. Two of its conferences have also been organized in India.

Tagui Sutclife defined the term as "the study of the quantitative aspects of any form of information. It is not limited to the study of mere inscriptions or vandalism lists or only scientists in any social group. It also extends beyond the fields of scientometrics and bibliometrics to the quantitative study of information."

5.10.3 Webmetrics Webmatrics

The term Webmatrics was first used by Almind and Ingewerson in 1997. The web appeared in the 1990s. The term netometrics was also used in 1995. The University of Berkeley developed the term by studying the size of documents available on the web, the number and type of tags, the number of additional links, file extension, etc.

Thus, webmatrics is the study of quantitative aspects of the creation, use, structure and technology of information sources on the web (www). Webmatrics is used to study both the construction and use of the web. The main areas of study are webpage content analysis, weblink structure analysis (web usage analysis) and web technology analysis.



Note

5.10.4 Cybermatrics

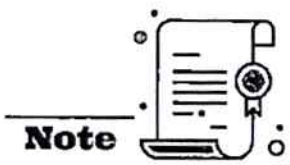
The word cyber was first used as a prefix to the word cybermetrics by Norbert Weiner in his book in 1948. The word cyber, originating from the Greek language, means a person operating or operating man in which the idea of control is also embodied. Thereafter, William Gibson used this word in a story published in 'Omni Magazine'. He also used the term again in his 1984 novel Neuromencer. In the novel this word meant electronic space. Over time the meaning of this word evolved and now it includes information technology, words like Internet, Virtual Reality etc. are also included.

The term cyberculture, cybernate, cybernation, cybermatrics, cybermetical, cybermetically, cybermetician, cybermeticist etc. came into use. Similarly Cyber crime, Cyber reference, Cyber Library, Cyber Librarian Cyber marriage, Cyber love etc. have also been originated. The word Metrics is derived from Meter. The word Cyber Matrix consists of two words Cyber and Matrix, which are respectively derived from the Latin word Metrum and the Greek word Meteron which means 'measure'. Combining this word with cyber means 'the science of measuring cyber objects'. It also involves the use of arithmetic and statistical methods. In this, instead of documents, statistical calculations are used in the website.

According to B.K. Sen, it is the branch of knowledge that uses mathematical and statistical techniques to measure websites or their elements and concepts and measures their growth, stability, propagation and usage and checks the authenticity of the content. To control these factors, it studies the rules, cyber information systems, services and products and studies the effects of cyber age on society.

Similar words to this word are also used such as cybermetric *i.e.*, based on cybermetrics and cybermetrician and cybermetricist which means-specific person in cybermetrics. The word wavemetrics is synonymous with the word cybermetrics.

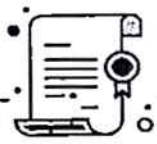
The geographical area in the field of study of cybermetrics is the entire cyber universe and time from the birth of the website to the end of the cyber world. All languages in which the website is available all the formats and all the subjects in which the website is made come under the domain of cyber studies.



5.10.5 Components of Cybermetrics

Cybermetrics has the following five components :

- **The First Component :** This component is Cyber Concepts. Cyber Space, Cyber Crime, e-Law, e-Books, e-Commerce, e-Government, e-Magazines, e-Journals, Digital Libraries Origin, Rate of Origin, Change in Format, Quantity of Use of Obsolescence Rates etc. are also part of cybermetrics.
- **The Second Component :** This component is websites. There is a website in cybermetrics. Everyday many new websites are created. Some change very quickly, some change very little, some do not change and some disappear. Websites of one subject vary in quality among themselves. Methods are being developed to rank these websites, so that the Web Impact Factor and the cited site can be studied.
- **The Third Component :** This component is Co-citation and Cyber Coupling. Co-citation and web coupling are also studied like bibliomatrices in web study.
- **The Fourth Component :** This component is the Search Engine. Investigator engines play a major role in the cyber space. Several exploratory engines have emerged over the years and have varied capabilities. The capabilities of these probe engines can be measured and thus the exploration engine is also a part of cybermetrics.
- **The Fifth Component:** This component is Cyber Information. Cyber universe is absolutely free. Anyone can enter and enter any information they want without any restrictions. There is no one to check the quality and authenticity of cyber information. Cyber information travels at a much faster rate than any other medium. Cyber information multiplies at a higher rate and can be retrieved at a higher speed and updated easily. These information can range from incorrect spelling to incorrect data and incorrect information. They can be used for free or even for a very small fee. In this way cyber information can be compared with printed information and statistical analysis can be done. Thus, cyber information is an important study area of cybermetrics.



Note

In the end we can say that all online information systems, services and products are in the cyber universe. Just as traditional systems, services and products are studied in bibliometrics, they are also studied in cybermetrics. For this, the structure of the system, the organization of the content, the data security, the period of update, the reproducibility and the ability to provide the service, etc. can be made parameters of comparison and compare with the traditional information system, services and products comparison can be made.

Exercises

Very Short Answer Questions

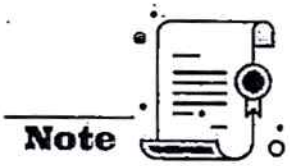
1. Who first used the term bibliometrics?
2. Who and when used the term librarianship in India?
3. The word Biblio is derived from which language word?
4. Write a definition of the term bibliometrics.
5. Write two purposes of bibliometrics.
6. Write the utility of bibliometrics.
7. What are the laws of bibliometrics?

Short Answer Questions

1. Describe the need and objectives of bibliometrics.
2. Describe the types of bibliometrics.
3. Write the experiments of bibliometrics.
4. Describe the breadboard rule.
5. Write the problems and limitations of bibliometrics.

Long Answer Questions

1. Describe the Zipf and Lotka rule of bibliometrics.
2. Description of Scientometrics, Informatrics and Webmetrics Rules Briefly.
3. Describe the components of Cybermetrices Rules.



Research and Report Writing

6.1 Introduction

The final stage of the entire research work is the preparation of the research report. After the research is completed, the researcher has to be inclined to give a proper description of the hypotheses tested under his research and the related results and the undertaking, methods and tools used to reach them. In this, applying the style of scientific writing, the whole description is presented in a systematic manner through appropriate headings and subheadings divided into appropriate chapters. This type of report has been named as thesis or dissertation. Terms such as thesis or dissertation are sometimes used synonymously.

According to Good and Haat, "Preparing a scientific report is fundamentally different from argument-oriented essays or literary writings. There is a conscious presentation of affirmative and injunctive evidence, as well as the method by which the evidence is derived."

In the Indian academic context, dissertations presented for achievements of Ph.D., D.Litt. or D.Sc. level are often referred to as thesis and short dissertations submitted for the degrees of Master's level and M.Phil level are often termed as dissertation. But in this regard, no very strict and strong tradition can be followed. The reality is that the research report can be presented in the form of thesis

Note

(dissertation), dissertation (short dissertation), articles to be published in journals or papers read under seminars. In this context, there is a slight difference in the methods of presentation of research reports. In this unit, detailed information about report presentation is given.

6.2 Components of a Thesis Dissertation

A dissertation or short dissertation can be called a comprehensive report about the entire research. Under this, the background of the research, its problem-issue or question, the objectives of the study, the hypotheses formulated and evaluated in it, the sample in respect of which the research has been completed, the design of the research which has been followed, the research tool formed or selected for the compilation of the base materials. , which has been applied in research, used for the analysis of base materials, qualitative or quantitative rendering method, selection of base material and generalizations and conclusions drawn through it and suggestions regarding future research possibilities, a clear description of related research literature and references from which appropriately cited etc. is presented. In this way, there are three main components of dissertation, the following elements come under the initial part.

6.2.1 Preliminaries

1. Title Page
2. The Researcher's Declaration
3. The Certificate of the Research Supervisor
4. Acknowledgment
5. Table of Contents
6. List of Tables
7. List of Figures

Initial parts are often represented by Roman numerals. This is called the title page. It is given "Title of Study" at the top of the page, while in the middle the name of the university and faculty where the dissertation is being presented, the degree or course requiring presentation of the dissertation in whole or in part, and the session of study in which it is being deposited is marked. In the lower part

Note

of this page, the researcher's name, with educational qualification, on the right side and the name of the guide (supervisor) on the left, his educational qualification and his post and the name of the institution in which that supervisor is posted.

In the next pages of the dissertation, the certificate of the researcher and after this the certificate of the guide or supervisor of the research is attached. It is presented in the form of a certificate given by the guide and is a proof of the proper completion of the entire research and its authenticity. The guide marks his signature to do the memorandum of acceptance of the research work.

The third page of the introductory pages is known as the foreword. It can run for two or three pages in a row. Under this, a very brief description of the main issues of research and various chapters of the report is presented. At the end a memorandum of gratitude or a show of gratitude is also marked. Sometimes there is a tradition to express it separately. In this, gratitude is expressed to the individuals, authors or experts who have been genuinely helpful for the research. Apart from this, it is also considered necessary to thank those institutions and authorities who have encouraged or helped the researcher in a formal or informal manner. Normally the size of the statement should not exceed two or three pages.

The next page is of the Table of Contents, in which the chapters, their contents and page numbers are given for the whole report in a chronological order so that the reader of the dissertation can easily find the appropriate page. In this, the title of the chapter is marked with capital letters and its sub-sections with small letters. Each chapter and its sub-sections are represented by the correct page.

The next page in this sequence shows the table, pictures. It starts from a new page. Under this, the title of the table and the page of the dissertation report on which it is available is marked.

On completion of this tabular list, a list of diagrams or drawings presented in the report. This list is also given on the new page. The title of the diagram or drawing under it, its abbreviation, statement and the page of the report where it is displayed.

Note

6.2.2 Main Body of the Report

It consists of the following chapters :

1. Introduction
   (i) Role or Preface
   (ii) Problem Statement
   (iii) Limitation of Study
   (iv) Conceptual Definitions of Fundamentals
   (v) Abbreviation, if any
   (vi) Requirement Statement
2. Review of Related Literature
3. Methodology
4. Data Presenting and Analysis
5. Conclusions and Summary
6. Recommendations

The main content of the report is presented in several chapters immediately after the opening pages.

Logically, its main sub-sections are as follows :

1. **Introduction :** In the first chapter, the introduction of the study area and its importance in the course of education are highlighted. Accordingly, the problem is presented as its main component. Before the statement of the problem, it is appropriate to mention the theories related to it and the conclusions of some previous research, so that it can be clear that there are some important questions related to the problem, which are still unanswered and naturally to find answers to these questions. For this the problem presented has emerged. Hypotheses should be formulated in the same order and if the main problem is in question form, then only those sub-questions should be given the answers to which it is desired to be obtained. The significance of the problem and the purpose of the study should also be given in a concise and clear form. Limits, limitations should be mentioned in point form. This chapter is very important. Therefore, the expression of ideas should be in simple and flowing language, that is, in

Note

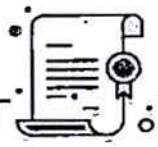
brief, the background of the research related issues, the concept of research and the limitations of research are mentioned.

2. **Survey or Review of Related Literature :** Review of literature related to chapter II. This part serves as a background for the present study. In this chapter, the researcher mentions the work done by other researchers, scholars and scientists updated on his research problem. The researcher studies journals, books and dissertations related to research because on the basis of this study, he is able to demarcate the problem properly. It also determines how the related work is different from other tasks. Through this, knowledge and guidelines of useful elements of problem solving would be obtained. This should also be written by the researcher in a systematic and logical manner and in the end, the related literature which has been used in this chapter should be referenced along with the page, that is, in short we can say that in this the researcher should refer to the work done by other scholars related to his research problem.

3. **Research Method :** Under this, the method used in research, the process of selection or manufacture of research equipment used for collecting the sample base material selected for it, and the research design which has been followed, are mainly discussed.

4. **Presentation and Analysis of Data :** This is an important chapter of the research report. In this chapter, the collected data is systematically arranged and analyzed and interpreted. After analyzing the data, the conclusions or results obtained from it are compared with other research results and it is seen that the result is consistent or opposite, then the reasons of these should be described. In short, it can be said that in this chapter, the systematic form, analysis and selection of base materials are presented in a systematic manner.

5. **Conclusion and Summary :** In this chapter, the researcher mentions the result of his work and in this he shows the solution of his problem which he had made the focal point in the introductory statement. In this chapter the researcher also



Note

tells about the acceptance or rejection of his hypothesis. If there is a conflict of his conclusions with any previous theory or opinion, as well as it makes it clear that what is the contribution of his research to the development of knowledge. This chapter is a summary of the entire research.

6. **Recommendations :** This is the last chapter of the research. In this chapter, the researcher makes recommendations based on his findings, which are useful for policy makers and readers. The recommendations should clearly reflect what the researcher wants to say. In this chapter, the researcher also indicates the related subject and in which areas work should be done.

6.2.3 Reference part

1. Foot notes
2. Bibliography
3. Appendix
4. Index

The detailed description of the above is as follows :

1. **Footnote :** In this footnotes are written at the bottom of the page or at the end of the chapter.
2. **Bibliography :** After the main content of the report is presented, further bibliography from the new page is given. In this list, such texts whose references have come directly or indirectly under the report, are mainly included.
3. **Appendix:** Under this, such relevant material used in research is presented, which increases the size of the dissertation by linking the report with the main subject. Some of the main materials to be included in the Appendix can be as follows. A copy of the questionnaire used in the research, randomly provided base material used in the research, special types of tables obtained by statistical analysis, etc.
4. **Index :** If there is a plan to publish the dissertation in book or monograph form, the researcher has to prepare an index

Note

immediately after the appendix. Two types of indexes are often formed. Subject matter index in which the posts related to the topic and author-index in which the authors referred on different pages are mentioned. It is presented at the end of the dissertation. It is also common practice to give a brief introduction of the researcher under the cover pages.

6.3 Features of Scientific Writing in Thesis Dissertation

First of all, we have to remember that there is enough difference in the style of scientific writing and literary writing.

"Here the object of the researcher is not to entertain or persuade the readers of the dissertation, or to simply express his opinion about a problem, make suggestions about its solution, and give reasoning on the basis of general observations not interested. On the contrary, he proposes hypotheses, cites base materials supported by facts and advertises whether they prove or disprove the hypothesis. Scientific reporting requires a clear, objective and logical presentation and analysis of evidence, not exaggerated arguments or exaggerations (translated) by emotions.

The characteristics of scientific writing under dissertation presentation are as follows :

1. **Language :** Application of common and familiar terminology in scientific and Emphasis is placed on clear expression.
2. **Organization :** The law chapters of the report and the sub-headings in them are kept in a logically related perspective. The whole content is presented under headings and sub-headings, keeping in mind the logical relationship of ideas and concepts from one chapter to another and within each chapter.
3. **Unity and Clarity :** Under scientific writing, there is an emphasis on unity and clarity. To ensure that the researchers presents ideas in a coherent and coherent form. The application of short but mutually logical sentences to express them is considered. Thus the article is constituted by such hoaxes. As pointed out by **Bon Dalen,** the researcher keeps on evaluating

Note

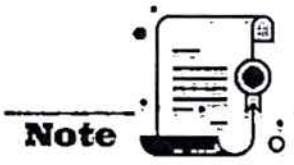
his chapters, paragraphs and sentences from this point of view again and again. From this it can be decided whether similar ideas can be presented together or not, whether one idea naturally adds to another idea, how to cut inconsistent and repetitive terminology and express the whole content how to increase the required fluency and comprehension in the language used to deliver, etc. In all, the researcher has to see that the words, sentences and paragraphs used under the dissertation are logically related to each other and there is a kind of ease in the communication of the whole report.

4. **Proportion and Emphasis :** It is necessary to organize the entire dissertation in such a way that the ratio of the chapters involved and the ideas described in them and the emphasis placed on them appears to be justified.

5. **Use of Conventions :** Under dissertation writing, some special conventions are used to place and mark the citations, authors and material taken from other researches or texts in the report. These can be classified into the following sections for convenience :

   (i) **Page and Chapter Format :** Pages are often given three types of headings : central headings, marginal headings which are called at the beginning of the paragraphs and paragraph headings which are called denoting small subsections. Each chapter starts on a new page. Even if the last page of the preceding chapter is of two or three lines, the subsequent chapter is started on a new page. Each chapter number is fixed and to be displayed in capital letters without punctuation.

   It is customary to number each page of the dissertation by a page number, but on many pages this number is not displayed. The page numbers are expressed differently in the two sequences. There is a practice of marking the opening pages by letters of Roman (i, ii, iii, iv etc.) or Hindi. The title page is denoted by (i) or the first letter but it is not marked or typed on that page.



(ii) **Use of Quotations :** The consistency of citations is very important in dissertations. Its size should not be so long that the reader of the report forgets whose views he is assessing. The following customs are generally prevalent in this regard :

(a) Direct citations are used only when the original words used by the author have been presented in such an effective and restrained manner that the dissertation author is not in a position to improve them easily or the researcher wants to comment on them or their tending to refute or of the ideas referred to in them interested in analysis.

(b) The exact nature of the words of the author or of quotations taken from any official record should be presented.

(c) If the tense given in the quotation does not match with its preface, a pronoun that is not specific is used and if the need is felt, the subject 'Sesame' can be used in the quoted subjects.

(d) When a quote is too lengthy, some part of the original article may be omitted.

(e) When quoted as a footnote, the double quotation mark is always used.

(iii) **Footnotes :** Footnotes are used immediately at the bottom of a page or at the end of a chapter or at the end of a dissertation or research paper. It is applied for the authenticity of a point, or statement so that they can freely take advantage of the appropriate sources. Particularly notable among the information provided under footnotes are the source of the information, usually the author's name, the title of the source, the actual page in the referenced source, publication date, publisher and publication location, year. Some conventional traditions for footnotes are as follows :

When the footnote is first introduced then the full name of the author, the year of publication, the name of the book, the publisher and the page number as displayed next when referring to it.

Generally a footnote should contain the following informations :

1. Name of the author, 2. Explanation of the source, 3. Version 4. Place of publication, 5. Name of the publisher, 6. Date of publication, 7. Page number.

There are many traditions of writing footnotes, which are described as follows :

(a) A general work should be written in footnotes as follows:

Krishan Kumar, Library Manual, New Delhi, Vikas, 1985, p. 77.

(b) Once the reference is given in the footnote, if it is coming again, there is no need to write the name of the author, publisher etc. again. For this the acceptance abbreviation ibid *i.e.,* 'same context' can be used. This will avoid duplication of documents. *For example,*

Ranganathan, S.R., Library manual, 2n ed., Bombay, Asia Publishing House, 1960, p. 190.

Ibid, p. 200

(iv) Abbreviation : Some Anglo and Latin word abbreviations which are used in footnote, bibliography etc. from the point of view of saving space, the following is a list of :

Abbreviations	Word	Abbreviations	Word
annon.	anonymous	loc. cit	the place cited
ante	before	mimeo	miemeographed
bk., bks.	book, books	m.s.	manuscript
chap.	Chapter	n.	number

Note

col., cols.	column, columns	n.d.	no date given
comn.	Commentary	n.n.	no name given
diss.	Dissertation	n.p.	no place given
ed., eds	editor, editors	NB.	(Note bene) taken notice
eng.	Englarged	o.p.	out of print
et.al	and others	op.cit	Previouslu cited
e.g.	(exempli gratia) for example	p., p.p.	page, pages
et alibi	and elsewhere	pt., pts.	part, parts
et. Seq	and the following	Passim.	here and there
fig., figs.	figure, figures	rev.	revised
fn.	Footnote	supra.	above
ibid.	same reference	tr.	Translation
idem.	the same	tr.	Translator
i.e.	that is	vide.	see, refer to
illus.	illustrated	viz.	namely
infra.	Below	vol.	volume
introd.	Introduction	vs.	against

(v) Tables and Figures : The information which is given by tables, they are presented through tables and those which are presented through pictorial media like graphs, pictures, charts, diagrams and maps etc. are called diagrams. It has to be remembered here that tables and diagrams cannot be considered an indispensable medium for presenting information in every dissertation. Under certain researches, only proper verbal descriptions may be expected to be communicated. It is considered necessary to follow the following conventions while writing tables and diagrams in dissertations :

(a) It is useful to display tables and diagrams closely under the main object of the thesis so that the required information can be understood accurately with the help of each other.

Note

(b) If the display of a table or diagram takes more than half a page if there is a possibility, they should be given some subject matter in between on different pages. It would be more expedient to submit without.

(c) Every table or diagram has to be presented under the heading. The title under which the information is displayed is given in a very brief form like a picture introductory description rather than sentence by sentence. It is self sufficient from the point of view of meaning.

(d) All the tables in a consecutive order and the diagrams in a different consecutive order to be marked with number signs. Tables and diagrams are sometimes accompanied by footnotes to explain a point in them or to define the abbreviations or symbols available under them.

(vi) Reference Memorandum: The research report includes those texts, articles, documents and manuscripts under the references which have been used by the researcher in completing his research and writing the dissertation. It is generally customary to display such references under 'bibliography'.

(a) The works from which quotations are taken in the dissertation may be included within the main subject matter or as footnotes.

(b) The source materials that have been used in research have to be presented in a comprehensive bibliography and books and research papers are also included in this.

(c) Specific bibliography which specifically mentions those sources which the researcher has used specially for his research.

(d) A concise attested bibliography regarded as such a list of referenced literature in which the contents of texts and compositions (which have been used) and a brief note about usability is infrequent.

Note

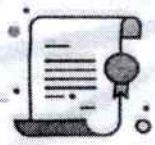
(vii) **Use of Appendix :** Such supporting evidences are included under the Appendix, which, if presented along with the main subject matter of the report, may hinder its legibility and make the whole subject comprehensible or complicated. Each appendix is separately marked with a special number. It is customary to mention it within the main subject or under the footnote.

6. **Size of the Report :** The size of the dissertations presented for the degrees of Ph.D., D.Lit. or D.Sc. in the Indian context is slightly larger (about 200-400 pages) and minor research at the level of M.A., M.Ed. and M.Phil. It is a common practice to keep the size of the arrangements relatively small (about 60 to 150 pages).

7. **Type-script of the Report :** Before typing the dissertation, read and edit the first motion of the pre-written draft very carefully.

   One should neither rely on typing nor relying on it for correction of inaccuracies. This action is required to prevent or prevent errors.

8. **Mechanical Aspect of the Report :** The dissertation is never submitted in unbounded pages. The typed copies are arranged in first, second, third and fourth copies and they are bound with proper get-up. It is common practice to have the cover page printed. It should not be an attempt to make it unnecessarily attractive. Submitting the entire dissertation with binding in a very clean manner. It is considered appropriate.



Exercises

Very Short Answer Questions

1. State the main components of a dissertation.
2. What is included in the preliminary part of the dissertation?
3. What is presented in the body of the report?
4. What is presented in context?

Short Answer Questions

1. Write the characteristics of scientific writing.
2. What is a footer note?
3. Explain the Appendix.

Long Answer Questions

1. What do you understand by research report? Explain its need.
2. Create a short research report on your selected topic.
3. Describe in detail the main components of the dissertation.
4. Write a short note on the footnote bibliography, appendix.