CHAPTER 1

INTRODUCTION

In recent days we hear talking about 'Statistics' from a common person to highly qualified person. It only shows how 'Statistics' has been intimately connected with wide range oı activities in daily life.

Statistics can be used either as plural or singular. When it is used as plural, it is a systematic presentation of facts and figures. It is in this context that majority of people use the word 'Statistics'. They only meant mere facts and figures. These figures may be with regard to production of foodgrains in different years, area under cereal crops in different years, per capita income in a particular state at different times, etc., and these are generally published in Trade Journals, Economics and Statistics Bülletins. Newspapers, etc. When statistics is used as singular, it is a science which deals with collection, classification, tabulation, analysis and interpretation of data.

Statistics as a science is of recent origin. The word 'Statistics' has been derived from a Latin word which means 'State' which in turn means 'politically organised people' i.e., government. Since governments used to collect the relevant data on births and deaths, defence personnel, financial status of the peoples, import and export, etc. Statistics was identified with Government. Recently, it pervades all branches of sciences, social sciences and even in Humanities like English literature. For example, in English literature the style of a particular poet or an author can be assessed with the help of statistical tools.

In the opinion of Fisher 'Statistics' has got three important functions to play (*i*) Study of statistical populations (*ii*) study of the variation within the statistical populations (*iii*) study of the methods of reduction of data.

P.C. Mahalanobis compares 'Statistician' with a 'Doctor' where Doctor prescribes medicine according to the disease of the patient whereas statistician suggests statistical technique according to the data in hand for proper analysis and interpretation.

Bowley defined statistics as 'the science of measurements of the social organism regarded as a whole in all its manifestations.' Another definition says that it is 'quantitative data affected to a marked extent by a multiplicity of causes.' Yet another definition says that it is a 'Science of counting' or 'Science of averages' and so on. But all these definitions are incomplete and are complementary to each other.

There are some of the limitations of 'Statistics' also when the data are not properly handled. People start disbelieving in statistics when the (1) data are not reliable (2) computing spurious relationships between variables (3) generalizing from a small sample to a population without taking care of error involved.

If one is ensured that data are reliable and is properly handled by a 'skilled statistician', the mistrust of statistics will disappear and in place of it precise and exact revelation of data will come up for reasonable conclusions.

CHAPTER 2

COLLECTION, CLASSIFICATION AND TABULATION OF DATA

2.1. Collection of Data

The data are of two kinds: (*i*) Primary data (*ii*) Secondary data.

Primary data are based on primary source of information and the secondary data are based on secondary source of information.

2.1.1. Primary data are collected by the following methods :

(*i*) By the investigator himself. (*ii*) By conducting a large scale survey with the help of field investigators. (*iii*) By sending questionnaires by post.

(*i*) The first method is limited in scope since the investigator himself cannot afford to bear the expenses of a large scale survey and also the time involved therein. Therefore, this method is of much use only in small pilot surveys like case studies. This method is being adopted by individual Investigators who submit dissertation for Masters and Doctoral degrees in rural sociology, Ag. Extension, Ag. Economics, Home Management, etc.

(*ii*) In this method the schedules which elicit comprehensive information will be framed by the Chief Investigator with the help of other experts based on objectives of the survey. The field investigators would be trained with the methodology and survey, mode of filling the schedules and the skill of conducting interviews with the respondents, etc. The field investigators will furnish the schedules by personal interview method and submit the schedules to the Chief Investigator for further statistical analysis. This method of collecting data requires more money and time since wide range of information covering large area is to be collected. But the findings based on the large scale survey will be more comprehensive and helpful for policy making decisions. The Decennial Census in India, National sample survey rounds conducted by Govt. of India, Cost of Cultivation schemes, PL 480 schemes, etc., are some of the examples of this method.

(*iii*) In the third method, the questionnaire containing different types of questions on a particular topic or topics systematically arranged in order which elicit answers of the type yes or no or multiple choice will be sent by post and will be obtained by post. This method is easy for collecting the data with minimum expenditure but the respondents must be educated enough so as to fill the questionnaires properly and send them back realizing the importance of a survey. The Council of Scientific and Industrial Research (CSIR) conducted a survey recently by adopting this method for knowing the status of scientific personnel in India.

2.1.2. The secondary data can be collected from secondary source of information like newspapers, journals and from third person where first hand knowledge is not available. Journals like Trade Statistics, Statistical Abstracts published by State Bureau of Economics and Statistics, Agricultural situation in India, import and export statistics and Daily Economic times are some of the main sources of information providing secondary data.

2.2. Classification of Data

The data can be classified into two ways : (*i*) classification according to attributes (Descriptive classification) (*ii*) classification according to measurements (Numerical classification).

2.2.1. Descriptive Classification: The classification of individuals (or subjects) according to qualitative characteristic (or characteristics) is known as descriptive classification.

(*a*) *Classification by Dichotomy:* The classification of individuals (or objects) according to one attribute is known as simple classification. Classification of fields according to irrigated and un-irrigated, population into employed and unemployed, students as hostellers and not-hostellers, etc., are some of the examples of simple classification.

(*b*) *Manifold Classification:* Classification of individuals (or objects) according to more than one attribute is known as manifold classification. For example, flowers can be classified according to colour and shape; students can be classified according to class, residence and sex, etc.

2 2.2. Numerical Classification: Classification of individuals (or objects) according to quantitative charactertistics such as height, weight, income, yield, age, etc., is called as numerical classification.

EXAMPLE: 227 students are classified according to weight as follows:

TABLE 2.1

Weight (lbs)	90-100	100-110	110-120-	120-130	130-140	140-150
No. of students	20	35	50	70	42	10

2.3. Tabulation of Data

Tabulation facilitates the presentation of large information into concise way under different titles and sub-titles so that the data in the table can further be subjected to statistical analysis. The following are the different types of tabulation:

2.3.1. Simple Tabulation: Tabulation of data according to one characteristic (or variable) is called as simple tabulation.

Tabulation of different high yielding varieties of wheat in a particular state, area under different types of soils are some of the examples.

2.3 2. Double Tabulation: Tabulation of data according to two attributes (or variables) is called double tabulation. For example, tabulation can be done according to crops under irrigated and unirrigated conditions.

2.3.3. Triple Tabulation: Tabulation of data according to three characteristics (or variables) is called triple tabulation.

For example, population tabulated according to sex, literacy and employment.

2.3.4. Manifold Tabulation: Tabulation of data according to more than three characteristics is called manifold tabulation.

EXAMPLE: Tabulated data of students in a college according to native place, class, residence and sex is given in Table 2.2.

TABLE 2.2. STUDENTS

Class	Rural				Urban			
	Male		Female		Male		Female	
	Hostellers	Day Scholars	Hostellers	Day Scholars	Hostellers	Day Scholars	Hostellers	Day Scholars
Intermediate								
Graduate								
Post-graduate								

2.3.5. The following are some of the precautions to be taken in tabulation of data.

(*a*) The title of the table should be short and precise as far as possible and should convey the general contents of the table.

(*b*) The sub-titles also should be given so that whenever a part of information is required it can be readily obtained from the marginal totals of the table.

(*c*) The various items in a table should follow in a logical sequence. For example, the names of the states can be put in an alphabetical order, the crops can be written according to importance on the basis of consumption pattern, the age of students in an ascending order, etc.

(*d*) Footnotes should be given at the end of a table whenever a word or figure has to be explained more elaborately.

(*e*) Space should be left after every five items in each column of the table. This will not only help in understanding of the items for comparison but also contributes for the neatness of the table.

EXERCISES

1. Draw up two independent blank tables, giving rows,

columns and totals in each case, summarising the details about the members of a number of families, distinguishing males from females, earners from dependants and adults from children.

2. At an examination of 600 candidates, boys outnumber girls by 16 per cent. Also those passing the examination exceed the number of those failing by 310. The number of successful boys choosing science subjects was 300 while among the girls offering arts subjects there were 25 failures. Altogether only 135 offered arts and 33 among them failed. Boys failing the examination numbered 18. Obtain all the class frequencies.

3. In an Agricultural University 1200 teachers are to be classified into 600 Agricultural, 340 Veterinary, 200 Home Science and 60 Agricultural Engineering Faculties. In each Faculty there are three cadres such as Professors, Associate Professors and Assistant Professors and in each Cadre there are three types of activity as teaching, Research and Extension. Draw the appropriate table by filling up the data.

4. Classify the population into Male and Female; Rural and Urban, Employed and unemployed, Private and Government and draw the table by filling up with hypothetical or original data.

CHAPTER 3

FREQUENCY DISTRIBUTION

3.1. Frequency Distribution

Frequency may be defined as the number of individuals (or objects) having the same measurement or enumeration count or lies in the same measurement group. Frequency distribution is the distribution of frequencies over different measurements (or measurement groups). The forming of frequency distribution is illustrated here.

EXAMPLE. Below are the heights (in inches) of 75 plants in a field of a paddy crop.

17,	8,	23,	24,	26,	13,	31,	16,	14,	35,	6,	11,
12,	11,	15,	21,	10,	4,	3,	19,	35,	36,	19,	40,
28,	17,	12,	2,	27,	31,	11,	21,	16,	34,	39,	1,
7,	12,	13,	10,	6,	21,	24,	22,	26,	28,	17,	6,
5,	15,	11,	16,	28,	4,	3,	19,	27,	35,	37,	14,
2,	9,	8,	16,	13,	22,	8,	26,	13,	12,	16,	14,
27,	31,	6.									

The difference between highest and lowest heights is 40—1 = 39. Supposing that 10 groups are to be formed, the class interval for each class would be 39/10=3.9. The groups (or classes) will be formed with a class interval of 4 starting from 1 continuing upto 40. The number of plants will be accounted in each class with the help of vertical line called 'tally mark'. After every fourth tally mark the fifth mark is indicated by crossing the earlier four marks. This procedure is shown in the following Table 3.1.

TABLE 3.1

Class	Tally marks	Frequency
1-4	卌 ||	7
5-8	卌 ||||	9
9-12	卌 卌 |	11
13-16	卌 卌 ||||	14
17-20	卌 |	6
21-24	卌 |||	8
25-28	卌 ||||	9
29-32	|||	3
33-36	卌	5
37-40	|||	3
		75

3.1.1. Inclusive Method of Grouping: The different groups formed in Table 3.1 belong to inclusive method of grouping since both upper and lower limits are included in each class. For example, in the first group, plants having heights 1″ and 4″ are included in that group itself. The width of each class is called class interval. The mid-value of the class interval is called class mark.

3.1.2. Exclusive Method of Grouping: In this method the upper limit of each group is excluded in that group and included in the next higher group. The inclusive method of grouping in Table 3.1 can be converted to exclusive method of grouping by modifying the classes 1-4, 5-8, 9-12, 13-16,......to 0.5-4.5, 4.5-8.5, 8.5-12.5, 12.5-16.5. However, the class interval in each group in exclusive method increased is 4. Here the upper limit, 4.5 is excluded in the first group and included in the next higher group 4.5-8.5. In other words, plants having heights between 0.5″ to 4.4″ are included in the group 0.5-4.5 and having heights from 4.5″ to 8.4″ are included in the next group 4.5 to 8.5 and so on.

3.1.3. Discrete Variable: A variable which can take only fixed number of values is known as discrete variable. In other words, there will be a definite gap between any two values. The number of children per family, the number of petals per flower, the number of tillers per plant, etc., are discrete variables. This variable is also called as 'discontinuous variable'.

3.1.4. Discrete Distribution: The distribution of frequencies of discrete variable is called discrete distribution. The frequency distribution of plants according to number of tillers is given in Table 3.2.

TABLE 3.2

No. of tillers	0	1	2	3	4	5	6	7
No. of plants	10	25	42	65	72	18	16	3

3.1.5. Continuous Variable: A variable which can assume any value between two fixed limits is known as continuous variable. The height of plant, the weight of an animal, the income of an individual, the yield per hectare of paddy crop, etc. are continuous variables.

3.1.6. Continuous Distribution: The distribution of frequencies according to continuous variable is called continuous distribution. For example, the distribution of students according to weights, is given in Table 3.3.

TABLE 3.3

Weight (lbs)	90-100	100-110	110-120	120-130	130-140	140-150
No. of students	6	15	42	18	12	5

3.2. Diagrammatic Representation

The representation of data with the help of a diagram is called diagrammatic representation.

(*i*) *Bar Diagram:* In this diagram, the height of each bar is directly proportional to the magnitude of the variable. The width of each bar and the space between bars should be same.

EXAMPLE: The yearwise data on area under irrigation in a particular state is represented by bar diagram in Fig. 3.1.

TABLE 3.4

Year	*Area under irrigation (million hectares)*
1970	15
1971	17
1972	18
1973	18
1974	20
1975	22

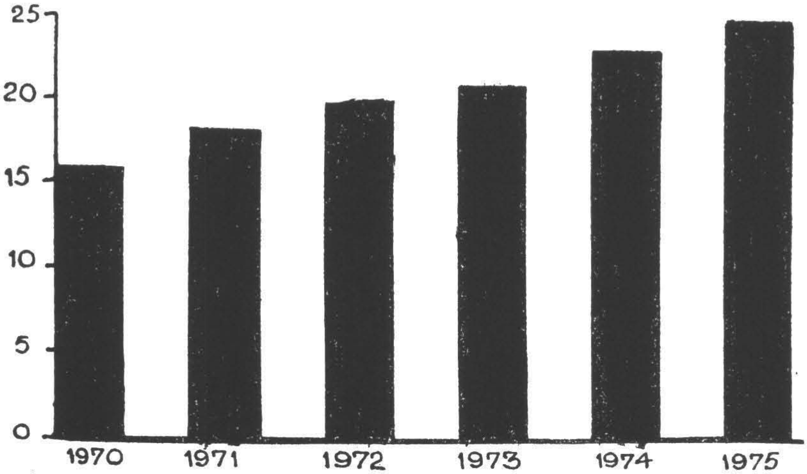

Fig. 3.1. Bar diagram.

(*ii*) *Component Bar Diagram:* In this case, the heights of the component parts of the bar are directly proportional to the magnitude of the ...

...ailable. Here also the
...n the bars should be
... of advantage if the

... Table 3.4 is further
...on and is presented

...on tons)	
...ll (...)	Total
	15
	17
	18
	18
	20
	22

...ng the data in Table

... This diagram ... of the v...
width of the bars and the space betwee...
same. This diagram would not be much ...
component parts are more than three.

EXAMPLE: The area under irrigation in...
sub-divided according to source of irrigati...
in Table 3.5.

TABLE 3.5

Year	Area under irrigation (in milli...		
	Canal (C)	Tank (T)	We... (W...
1970	7	5	3
1971	7	6	4
1972	8	6	4
1973	8	6	4
1974	8	6	6
1975	9	6	7

The component bar diagram representi...
3.5 are given in Fig. 3.2.

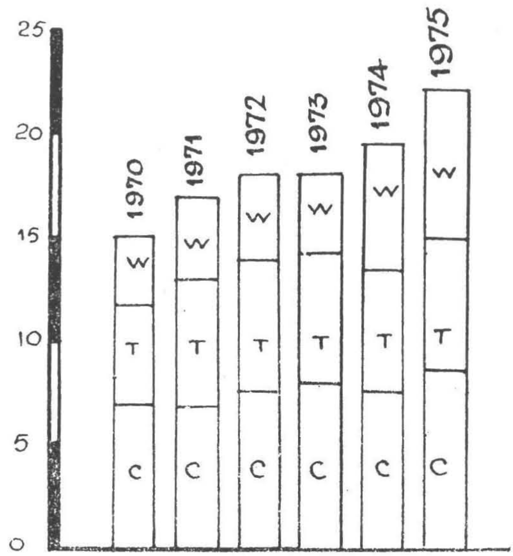

Fig. 3.2. Component bar diagram.

(*iii*) *Multiple Bar Diagram:* In this diagram, the height of each bar in a group of bars is directly proportional to the magnitude of individual item in a group of items. For example, yearwise cereal production, sex-wise literacy in different years, election yearwise, number of seats secured by different political parties in a Parliament (or State assembly) can be represented by Multiple bar diagram.

EXAMPLE: The following is the data on wages for different categories of agriculture labour in different years.

TABLE 3.6. LABOUR WAGES

Year	*Male* (*M*)	*Female* (*F*)	*Child* (*C*)
1950	0.75	0.50	0.30
1960	1.50	1.00	0.75
1970	2.50	2.00	1.50
1975	4.00	3.00	2.50

The multiple bar diagram representing the data in Table 3.7 is given in Fig. 3.3.

(*iv*) *Pie Diagram:* This is also known as Pie-chart. It is useful when the number of component parts of the variable is more than three. Here the areas of different sectors of a circle is

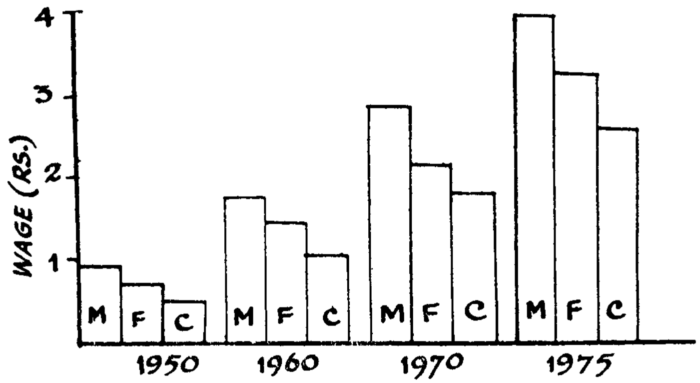

Fig. 3.3. Multiple bar diagram.

directly proportional to the magnitudes of the different component parts of the variable.

Let m_1 be the magnitude of the first component out of m, the total magnitude of the variable.

Then, $\theta_1 = 2\pi . \frac{m_1}{m} = 360 . \frac{m_1}{m}$

where θ_1 is the angle of a first sector. Similarly θ_2, θ_3...... can be obtained by multiplying 2π with $\frac{m_2}{m}$, $\frac{m_3}{m}$, etc.

After obtaining θ_1, θ_2,......the different sectors can be drawn on a circle each representing the individual component. The radius of the circle is proportional to the total magnitude of the variable.

EXAMPLE: Represent the expenditure of a salaried employee on different items by Pie-diagram. The details are given in Table 3.7.

TABLE 3.7

Items	*Expenditure* (*Rs.*)	*Sector angles* (θ_i)
Food	120	$360 \times \frac{120}{350} = 123.43$
House rent	70	$360 \times \frac{70}{350} = 72.00$
Clothing	50	$360 \times \frac{50}{350} = 51.43$
Education for children	35	$360 \times \frac{35}{350} = 36.00$

TABLE 3.7. (contd.)

Transport	25	$360 \times \frac{25}{350} = 25.71$
Miscellaneous	50	$360 \times \frac{50}{350} = 51.43$
	360	

The Pie-diagram representing the data in Table 3.7 is given in Fig. 3.4.

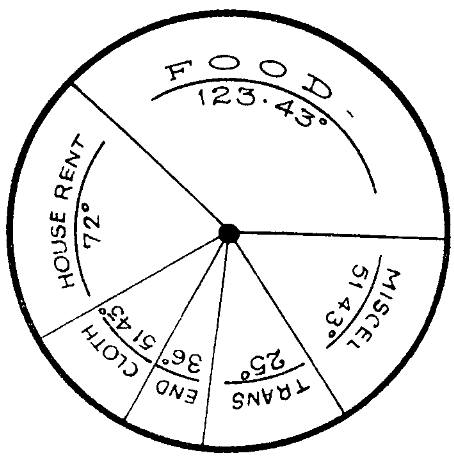

Fig. 3.4. Pie diagram.

It may be noted that the expenditure on food and house rent is accounted for a major share of the employee's salary.

Also the expenditure targets on different items in five year plans can be purposefully represented by pie diagram. If more than one employee is involved in the above example, as many circles may be drawn representing as many employees with the radius of each circle is proportional to the square root of the total salary of the corresponding employee.

(*v*) *Pictograms:* These are also called as pictorial charts. In this each variable is represented by the corresponding picture and the volume of a picture is directly proportional to the magnitude of the variable. For example, wheat production can be represented by the size of the wheat bag (or wheat ear or the number of wheat bags of the same size) according to particular scale, the size of the army by the size of the soldier (or soldiers of same size), the strength of navy by the size of battle ship (or the battle ships of same size), number of tractors by the size of tractor (or the tractors of same size) according to particular scale, etc.

Advantages: A diagram is always more appealing to eye than mere numerical data. It is easy for making comparisons and contrasts when more than one diagram is involved. It is easy to understand even for a layman.

Disadvantages: The main disadvantage of this representation is that it only gives rough idea of the variable but not the exact value. Also whenever the number of items are more it is difficult to depict on the diagrams since they require more space, time and unweildy for comparison.

3.3. Graphic Representation

Just as in the case of diagrammatic representation, here different methods of graphic representation are presented.

3.3.1. Histogram: It consists of rectangles erected with bases equal to class intervals of frequency distribution and heights of rectangles are proportional to the frequencies of the respective classes in such a way that the areas of rectangles are directly proportional to the corresponding frequencies.

EXAMPLE: Represent the following frequency distribution of farms according to area in a particular village by a histogram.

TABLE 3.8

Area (hectares)	0-2	2-4	4-6	6-8	8-10	10-12
No. of farms	40	48	25	18	12	7

From Fig. 3.5 one can infer that the maximum number of farms are lying in the group (2-4) and the minimum number in the group (10-12). The total area under the histogram is equal to the total frequency.

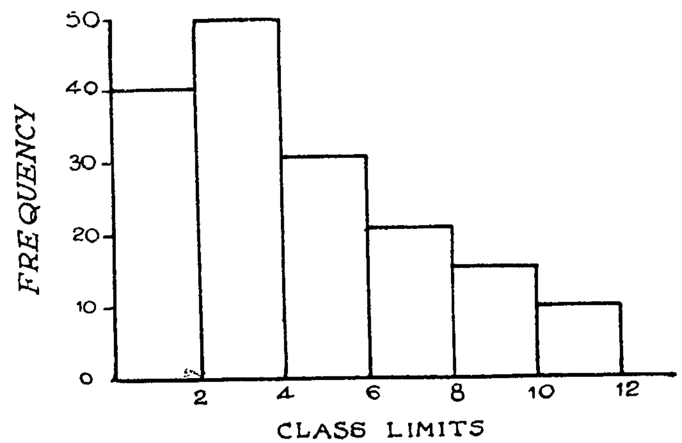

Fig. 3.5. Histogram.

3.3.2. Frequency Polygon : If the points are plotted with midvalues of the class intervals on the X-axis and the corresponding frequencies on the Y-axis, the figure obtained by joining these points with the help of a scale is known as frequency polygon.

EXAMPLE: The frequency polygon for the data given in Table 3.8 is as follows.

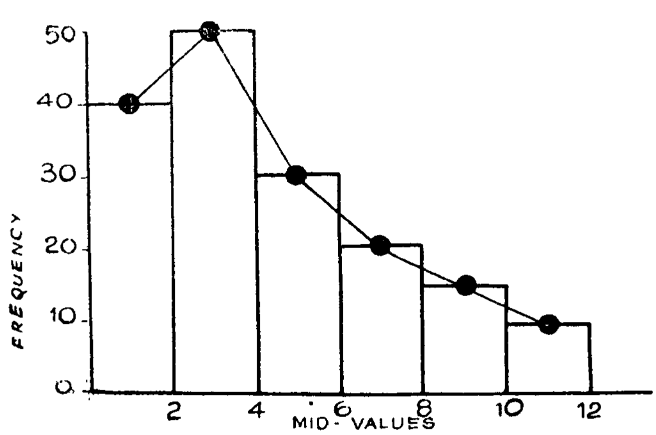

Fig. 3.6. Frequency polygon.

The frequency polygon in Fig. 3.6 is drawn with the assumption that the frequencies are concentrated at the mid-values of the corresponding classes. It may be noted that the area under histogram is equal to the area under frequency polygon.

3.3.3. Frequency Curve: If the points are plotted with midvalues of the class intervals on X-axis and the corresponding frequencies on Y-axis, the figure formed by joining these points with a smooth hand is known as frequency curve.

EXAMPLE: The frequency curve for the example given in Table 3.8 is given in Fig. 3.7.

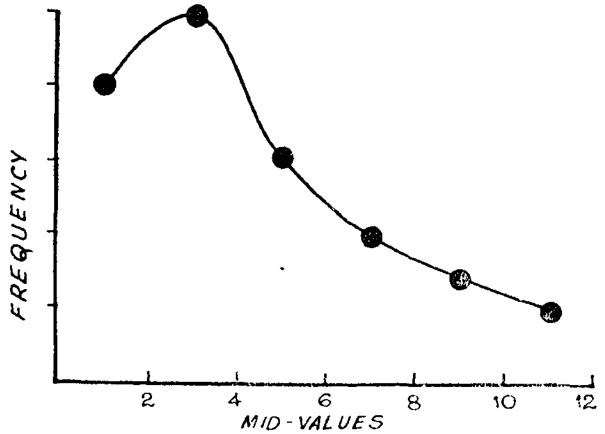

Fig. 3.7. Frequency curve.

3.3.4. Cumulative Frequency Curve (ogive): If the points are plotted with upper limits of classes on X-axis and the corresponding cumulative frequencies (less than) on Y-axis, the figure formed by joining these points with a smooth hand is known as cumulative frequency curve (less than). If the lower limits of classes are taken on X-axis and the corresponding cumulative frequencies (greater than) on Y-axis, the curve so obtained is called cumulative frequency curve (greater than).

EXAMPLE: Represent the distribution of rainfall on different days from July to September months in a particular locality and in a particular year by cumulative frequency curves.

TABLE 3.9

Rainfall (in cm)	*No. of days*	*Cum. fre. (less than)*	*Cum. fre. (greater than)*
0-3	6	6	92
3-6	9	15	86
6-9	10	25	77
9-12	25	50	67
12-15	19	69	42
15-18	15	84	23
18-21	8	92	8

The X-co-ordinate of the point of intersection of two cumulative frequency curves is the median value. The

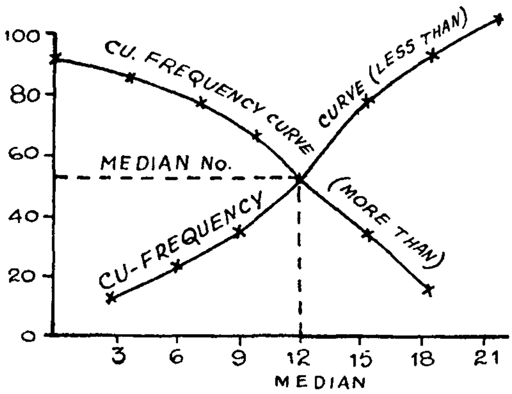

Fig. 3.8.

Y-co-ordinate will correspond to median no. i.e., $\frac{N+1}{2}$ where N is the total frequency. From the Fig. 3.8 first quartile and third quartile can be obtained from X-axis for the corresponding values (N+1)/4 and $\frac{(3N+1)}{4}$ respectively on Y-axis. The reader is advised to refer Sections 4.2 and 5.2 respectively for definitions of median and quartiles.

3.3.5. Lorenz Curve: It is the curve drawn between two variates which are expressed in percentage cumulative frequencies. This curve is useful to depict the income distribution of individuals where cumulative percentage of individuals are taken on the X-axis and the corresponding cumulative percentage of incomes are taken on the Y-axis. This is commonly used in graphic representation of the inequality aspect of the income distribution. This is due to Italian statisticians, Gini and Lorenz. This curve can also be used for the distribution of any non-negative variate, with a continuous type of distribution as for example, for the distribution of factories by capital size, (or number of employees), etc. The equality of the income distribution is depicted as a straight line drawn with 45° connecting the two diagonal points, and which is known as 'egalitarian line'. If the income distribution is not even then the egalitarian line will take a curve shape. This curve is called 'Lorenz curve'. If Lorenz curve is closer towards 'egalitarian line' there is less of inequality of income distribution. If the Lorenz curve is away from the 'egalitarian line' there is more of inequality of income distribution.

From Fig. 3.9, it can be inferred that the distribution of income in year y_2 has tended towards equality in comparison to the year y_1.

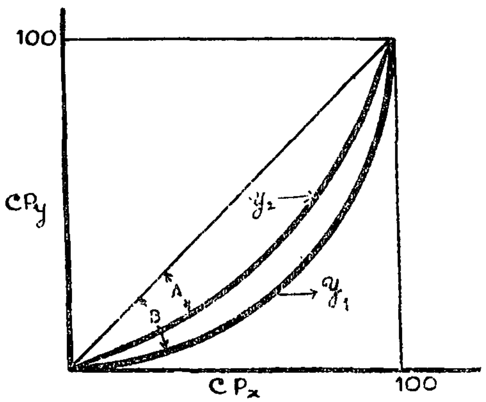

Fig. 3.9. Lorenz curve.

Since A<B where A is the area between y_2 Lorenz curve and egalitarian line and B is the area between y_1 and egalitarian line as shown in Fig. 3.9. The procedure for finding out the area A or B is given in the following sub-section of 'Fitting of Lorenz Curve'. If any curve coincides with the 'egalitarian line then the area would become zero and the Gini's concentration' ratio would be zero.

3.3.6. Fitting of Lorenz Curve: The approximate procedure of fitting 'Lorenz curve' as will as the method of finding out the area between 'Lorenz curve' and 'egalitarian line' is given here.

TABLE 3.10

Income class	*No. of persons*	*Mid value of income class*	*Prop. of persons* (P_x)	*Prop. of income* (P_y)	*Cum. prop. of persons* (CP_x)	*Cum. prop. of income* (CP_y)
Y_0-Y_1	f_1	y_1^1	P_1	q_1	$p_1 = p_1^1$	$q_1 = q_1^1$
Y_1-Y_2	f_2	y_2^1	p_2	q_2	$p_1 + p_2 = p_2^1$	$q_1 + q_2 = q_2^1$

TABLE 3.10 (Contd.)

1	2	3	4	5	6	7
Y_2-Y_3	f_3	y_3^1	p_3	q_3	$p_1+p_2+p_3=p_3^1$	
...	:	:	:	:	:	$q_1+q_2+q_3=q_3^1$
						:
Y_{k-1}-Y_k	f_k	y_k^1	p_k	q_k	p_k^1	q_k^1
	Σf_i	Σy_i	$\Sigma p_i=1$	$\Sigma q_i=1$		

Let $\triangle$ be the area of the trepezium between Lorenz curve and the X-axis in Fig. 3.10.

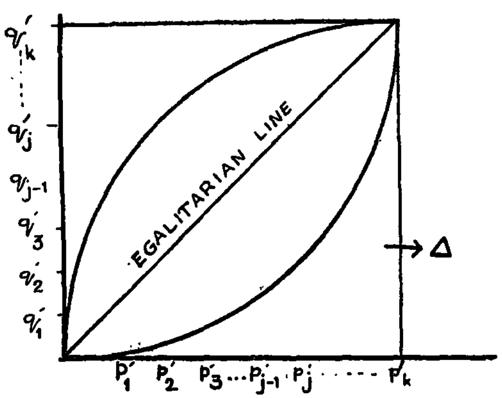

Fig. 3.10. Lorenz curve.

The area between the 'Lorenz curve' and 'egalitarian line' can be obtained by subtracting $\triangle$ from 0.5.

$$\text{Area of trepezium, } \triangle = \sum_{i=1}^{k} \frac{\left(q_j^1+q_{j-1}^1\right)\left(p_j^1-p_{j-1}^1\right)}{2}$$

Area between 'Lorenz curve' and 'egalitarian line' is

$$L=\tfrac{1}{2}-\triangle=\frac{1-2\triangle}{2}$$

It may be noted that the above method is an approximate one for finding out the area of trepezium.

EXAMPLE: The following is the distribution of income of different staff in an educational institution. Represent the data by Lorenz curve and also find the proportionate number of persons having income upto 20 per cent.

TABLE 3.11

Income group	*No. of staff members (X)*	*Mid-value of income group (Y)*	*Percentage prop. of Y (P_y)*	*Percentage prop. of X (P_x)*	*Cum. of prop. of Y C_{py}*	*Cum. of prop. of X C_{px}*
Less than 200	20	100	2.06	9.39	2.06	9.39
200-400	35	300	6.19	16.43	8.25	25.82
400-600	62	500	10.31	29.11	18.56	54.93
600-800	48	700	14.43	22.54	32.99	77.47
800-1000	25	900	18.56	11.73	51.55	89.20
1000-1200	16	1100	22.68	7.51	74.23	96.71
1200 & above	7	1250	25.77	3.29	100.00	100.00
	213	4850				

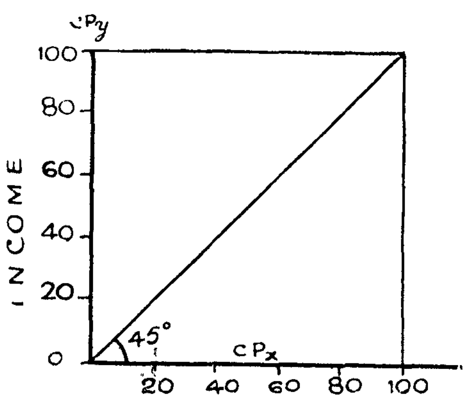

Fig. 3.11. Egalitarian line.

From Fig. 3.11, the proportion of persons having income upto 20 per cent is 60 per cent.

3.3.7. Remarks: The graphic representation generally depicts the trend when the number of observations is large. Also it provides intermediary values, though roughly.

EXERCISES

1. The following is the distribution of heights of plants of a particular crop.

Height (inches)	35-39	40-44	45-49	50-54	55-59	60-64
No. of plants	12	20	15	29	9	3

Draw the (*i*) Histogram (*ii*) Frequency Polygon (*iii*) Frequency curve and (*iv*) Ogive.

2. Draw the histogram of the following distribution of marriages classified according to the age of the bridegroom, and give your comments.

Class boundaries (age in years)	18-21	21-24	24-27	27-30	30-33
Frequencies (in thousands)	11	61	73	57	33
Class boundaries (age in years)	33-36	36-39	39-42		
Frequencies (in thousands)	21	14	9		

(*B.Sc. Madras, April, 1969*)

3. Draw the 'Ogive' for the following frequency distribution.

Age (in years)	10-15	15-20	20-25	25-30	30-35	35-40
No. of persons	42	60	150	70	35	20

4. The following are the data regarding the area under grape cultivation in different years.

Year	1960	1961	1962	1963	1964	1965
Area (100 acres)	20	22	27	30	32	34

Represent the above data by a suitable diagram.

5. Represent the following data by a bar diagram and comment on their relationship.

Country	*Birth-rate*	*Death rate*	*Infant mortality*
A	15.2	11.3	56
B	16.9	10.4	49
C	26.8	17.2	112
D	32.6	23.1	165

(*B.Sc. Madras, Sept., 1969*)

6. Represent the following data by sub-divided bar diagram drawn on the percentage basis.

Heads of expenditure	*State A*	*State B*
	(in lakhs of rupees)	
Agriculture	517	578
Irrigation	648	910
Industry	186	496
Transport	566	984
Miscellaneous	148	106

(*B.Sc. Madras, April, 1969*)

7. The data given below relates to the income of workers' families in an industrial area.

Income per week	*Number of families*	*Average income per family to the nearest rupee*
Less than Rs. 25	92	16
25-35	335	24
35-45	402	36
45-55	246	44
55-65	144	52
65-75	42	65
above Rs. 75	36	82

Draw a Lorenz curve to represent the data and determine therefrom, what percentage of the total income of the working classes is earned by the highly paid 25 per cent of the families.

(B.Sc. Madras, April, 1967)

8. The expenditure pattern of two cultivators on one hectare farm for different items of agricultural inputs and the corresponding sector angles are given in the following table.

Item	*Cultivator I*		*Cultivator II*	
	Expenditure (Rs.)	*Sector angle (degrees)*	*Expenditure (Rs.)*	*Sector angle (degrees)*
Land reclamation	200	54.14	100	22.78
Hybrid seeds	300	81.20	350	79.75
Fertilisers	600	162.41	800	182.28
Tractor rent	80	21.65	120	27.34
Electricity charges for pumping water	30	8.12	60	13.67
Labour charges	120	32.48	150	34.18
	1330	360.00	1580	360.00

CHAPTER 4

MEASURES OF LOCATION

It is always advisable to represent group of data by a single observation provided it does not loose any important information contained in the data and brings out every important information from it. This single value, which represents the group of values, is termed as a 'measure of central tendency' (or a measure of location or an 'average'). This should be a representative value or a typical member of the group. The different measures of location are 1. Arithmetic mean, 2. Median, 3. Mode, 4. Geometric Mean, and 5. Harmonic mean.

4.1. Arithmetic Mean

It is defined as the sum of the observations divided by its number.

Let $X_1, X_2, \ldots, X_n$ be n observations then the Arithmetic mean (A.M), $\overline{X}$ is defined as $\frac{X_1+X_2+\ldots+X_n}{n}$ which can be written as $\frac{1}{n}\ \Sigma X_i$, where 'Σ' is the summation which indicates the summing up of the observations from X_1 to X_n.

EXAMPLE: Compute the mean daily milk yield of a buffalo given the following milk yields (in kgs) for the consecutive 10 days.

15, 18, 16, 9, 13, 20, 16, 17, 21, 19

$$\overline{X} = \frac{15+18+\ldots+19}{10} = 16.4 \text{ kg.}$$

4.1.1. Linear Transformation Method: If the observation values are large, more in number and the deviation among themselves is small, the linear transformation method will save time in computation.

Let $d_i = X_i - A$ where A is called arbitrary mean and which is taken as round figure mid way between highest and lowest values.

$$\bar{X} = A + \bar{d}$$

$$= A + \frac{\Sigma d_i}{n}$$

For the above example, let A = 15

TABLE 4.1

Sl. No.	X_1	$d_i = (X_i - A)$	
1.	15	0	
2.	18	3	
3.	16	1	
4.	9	−6	
5.	13	−2	
6.	20	5	$\bar{X} = A + \frac{\Sigma d_i}{n}$
7.	16	1	$= 15 + \frac{14}{10} = 16.4$ kg.
8.	17	2	
9.	21	6	
10.	19	4	
		14	

4.1.2. Discrete Frequency Distribution: Let $f_1, f_2, \ldots, f_n$ be n frequencies corresponding to the variate values $X_1, X_2, \ldots, X_n$ respectively, then $\bar{X} = \frac{\Sigma f_i X_i}{\Sigma f_i}$

EXAMPLE: Compute the mean number of flowers per plant for the following data.

TABLE 4.2

No. of flowers (X_i)	*No. of plants* (f_i)	$f_i X_i$	
0	5	0	
1	10	10	
2	12	24	
3	16	48	$\bar{X} = \frac{161}{60} = 2.68$
4	8	32	
5	7	35	
6	2	12	
	60	161	

4.1.3 Grouped Frequency Distribution: (*a*) *Direct method.* Let $f_1, f_2, ..., f_n$ be n frequencies corresponding to the mid-values of the class intervals $X_1, X_2, ..., X_n$ then the A.M. is given by

$$\bar{X} = \frac{\Sigma f_i X_i}{\Sigma f_i} = \frac{\Sigma f_i X_i}{N}$$

EXAMPLE: Find the mean breadth of leaf given the following distribution.

TABLE 4.3

Breadth of leaf (in cms)	*No. of leaves* (f_i)	*Mid-value* (X_i)	f_iX_i	$d_i = \frac{X_i - A}{C}$	f_id_i
2-4	7	3	21	−2	−14
4-6	10	5	50	−1	−10
6-8	19	7	133	0	0
8-10	15	9	135	1	15
10-12	9	11	99	2	18
12-14	3	13	39	3	9
	63		477		18

$$\bar{X} = \frac{477}{63} = 7.57 \text{ cm.}$$

(*b*) *Linear transformation method*

$$\bar{X} = A + C\frac{\Sigma f_i d_i}{N} \text{ where } d_i = \frac{X_i - A}{C}, \text{ Here } A = 7$$

From Table 4.3, we have $\bar{X} = 7 + \frac{18 \times 2}{63} = 7.57$ C=Class interval.

Whenever the class interval is same it is always convenient to take $d_i = \frac{X_i - A}{C}$ where C is class interval to simplify the calculations. Consequently in the formula of $\bar{X}$ the second expression is multiplied by C.

The characteristics of a satisfactory average are listed here.

Characteristics of a satisfactory average: (*a*) It should have well defined formula, (*b*) It should be based upon all the observations, (*c*) It should be comprehensible, (*d*) It should be least affected by sampling fluctuations, (*e*) It should be easily computed, and (*f*) It should be capable of algebraic treatments.

Merits of A.M.: It possesses all the characteristics of satisfactory average which include algebraic properties such as (*i*) The algebraic sum of the deviations taken from A.M. is zero, *i.e.*, $\Sigma(X_i - \bar{X}) = 0$.

(*ii*) Let $\bar{X}_1$, be the mean of n_1 observations, $\bar{X}_2$ be the mean of the n_2 observations, ...,$\bar{X}_k$ be the mean of n_k observations then the mean $\bar{X}$ of $n=(n_1+n_2+\ldots+n_k)$ observations is given by

$$\bar{X} = \frac{n_1\bar{X}_1 + n_2\bar{X}_2 + \ldots + n_k\bar{X}_k}{n_1 + n_2 + \ldots + n_k}$$

Demerits of A.M.: (*a*) It may not be always identified with any one of the observations from which it is calculated, (*b*) It gives more weightage to extreme items whenever they are present, (*c*) It is difficult to calculate whenever the extreme classes in the continuous frequency distribution are not well defined.

4.2. Median

It is defined as that value of the variate below which half of the values lie and above which the remaining half lie when the variate values are arranged in ascending order of magnitude.

Case (*i*) *Variate values:*

(*a*) The number of observations is odd :

EXAMPLE: Find the median score of the following scores obtained by students in a particular one hour examination.

6, 9, 13, 4, 11, 8, 12, 9.5, 7

Arranging the scores in ascending order of magnitude, we have 4, 6, 7, 8, 9, 9.5, 11, 12, 13

Median number $= \frac{n+1}{2} = \frac{9+1}{2} = 5$, where n=number of observations.

The value of 5th observation is 9 which is the median value.

(*b*) If n is even

Let the scores be: 6, 9, 13, 4, 11, 8, 12, 9.5, 7, 8.5.

After arranging in ascending order of magnitude, we have 4, 6, 7, 8, 8.5, 9, 9.5, 11, 12, 13

$$\text{Median No.} = \frac{n+1}{2} = \frac{10+1}{2} = 5.5$$

Median value: 5th value+0.5 (6th value—5th value)
=8.5 + 0 5 (9—8.5) = 8.75

Case (*ii*) *Grouped frequency distribution:*

Let $f_1, f_2, \ldots, f_n$ be n frequencies corresponding to the mid-values of the classes $X_1, X_2, \ldots, X_n$ respectively then the Median is given by

$$M = l + \frac{\frac{N+1}{2} - m}{f} \times C$$, where l = lower limit of the median Class, C= Class interval of the median class, $\frac{N+1}{2}$ =median number, m=cum. fre. just preceding to the median class, f=frequency of the median class.

Here we assume that the groups are formed in ascending order of magnitude. Median class is that class in which the median number $\frac{N+1}{2}$ lies.

EXAMPLE: Obtain the median from the following distribution of weights of children in a particular locality.

TABLE 4.4

	Weights (Kg.)	*No. of children*	*Cum. fre.*
	0-4	3	3
	4-8	9	12
	8-12	18	30
Median class	12-16	20	50
	16-20	16	66
	20-24	7	73

$$M = l + \left(\frac{\frac{N+1}{2} - m}{f} \right) \times C$$

$\frac{N+1}{2} = \frac{73+1}{2} = 37$ Since 37 lies between the cumulative frequencies 30 and 50, the class (12-16) is the median class.

$l=12$, $C=4$, $m=30$, $f=20$.

$$M=12+\frac{37-30}{20}\times 4=13.4$$

Merits of Median: (*a*) It can be calculated even if the extreme classes are not well defined, (*b*) It can be easily located on frequency curve, (*c*) It is useful whenever the qualitative characters are under consideration, (*d*) It is having one important algebraic property *i.e.,* the sum of the absolute values of the deviations is least when the deviations are taken from the Median.

Demerits of Median: (*a*) It is not based on all the observations, (*b*) It is not widely used in practice, and (*c*) It is not well defined.

4.3. Mode

Mode is that value of the variate which occurs most frequently. Case (*i*) *Variate values.*

EXAMPLE: Find the model height (in inches) from the heights of 20 students.

60, 65, 64, 58, 69, 72, 64, 64, 65, 60, 61, 67, 64, 63, 67, 64, 68, 63, 64 and 66.

Here height 64″ is repeated more number of times and hence model height=64″

Case (*ii*) *Grouped distribution:*

Let $f_1, f_2, \ldots, f_n$ be n frequencies corresponding to the mid-values of the classes $X_1, X_2, \ldots, X_n$ respectively then the Mode (M′) is given by the formula

$$M'=l+\frac{f-f_p}{2f-f_p-f_s}\times C$$ where l=lower limit of the modal class, f=frequency of the modal class, f_p=frequency just preceding to modal class, f_s=frequency just succeeding to modal class and C=class interval of the modal class and modal class is that class in which maximum frequency exists.

EXAMPLE: Calculate the 'Mode' for the following distribution of wages in a certain factory.

TABLE 4.5

Daily wages	2-4	4-6	6-8	8-10	10-12	12-14	14-16	16-18
No. of employees	29	43	75	135	90	60	35	33

$$M' = l + \frac{f - f_p}{2f - f_p - f_s} \times C$$

$l = 8$, $f = 135$, $f_p = 75$, $f_s = 90$

$$M' = 8 + \frac{135 - 75}{270 - 75 - 90} \times 2 = 9.14$$

If the maximum frequency occurs more than once in the distribution, the method of grouping is adopted for locating the modal class and then the above formula will be used for finding the modal value.

Merits: (*a*) It can be easily located on frequency curve, (*b*) It can be calculated even when extreme classes are not well defined except the modal class, (*c*) It is used mostly in business. For example, Cloth merchant would like to keep a certain quality of cloth having maximum sales, shoe-maker would like to keep a shoe of size or sizes having maximum sales, and (*d*) It can easily be calculated except in the case where the maximum frequency occurs more than once.

Demerits: (*a*) It is not based upon all the observations, (*b*) It is not having algebraic properties, and (*c*) It is not stable, since different methods of forming class intervals would lead to different modal values.

The empirical relationship between mean, median and mode is Mean—Mode=3 (Mean —Median).

4.4. Geometric Mean

Let $X_1, X_2, \ldots, X_n$ be n observations then the geometric Mean is the n-th root of their product.

Case (*i*) *Variate values:*

The geometric mean of n observations, say, $X_1, X_2, \ldots, X_n$ is the n-th root of their product.

Hence G.M. $= (X_1, X_2 \ldots X_n)^{1/n}$

EXAMPLE: Compute the G.M. of the following observations: 6, 8, 11, 12, 21, 13.

$$\text{G.M.} = (6 \times 8 \times 11 \times 12 \times 21 \times 13)^{1/6}$$

$$\text{Log G.M.} = \tfrac{1}{6}(\log 6 + \log 8 + \ldots + \log 13)$$

$$= 1/6(0.7782 + 0.9031 + 1.0414 + 1.0792 + 1.3222 + 1.1139) = 1.0397$$

$$\text{G.M.} = \text{Anti log}\ (1.0397) = 10.96$$

Case (*ii*) *Grouped distribution*

Let $f_1, f_2, \ldots, f_n$ be n frequencies corresponding to the mid-values of the class intervals $X_1, X_2, \ldots, X_n$ then the Geometric mean is given by the formula:

$$G.M. = \left(X_1^{f_1}, X_2^{f_2} \ldots X_n^{f_n} \right)^{1/n}$$

$$\log G.M. = 1/N(f_1 \log X_1 + f_2 \log X_2 + \ldots + f_n \log X_n)$$

$$G.M. = \text{Anti log } 1/N(f_1 \log X_1 + f_2 \log X_2 + \ldots + f_n \log X_n)$$

EXAMPLE: Find the G.M. for the following distribution.

TABLE 4.6

Class	*Frequency* (f_i)	*Mid value* (X_i)	*Log* X_i	f_i *log* X_i
0-5	4	2.5	0.3979	1.5916
5-10	10	7.5	0.8751	8.7510
10-15	28	12.5	1.0969	30.7132
15-20	17	17.5	1.2430	21.1310
20-25	6	22.5	1.3522	8.1132
25-30	2	27.5	1.4393	2.8786
	67			73.1786

Log G.M. = 1/67 (73.1786) = 1.0922

G.M. = 12.37

Merits: (*a*) It has well defined formula, (*b*) It is based upon all the observations, (*c*) It is used in computing index numbers and also in time series analysis whenever ratios are under consideration. It is also used in finding out rate of change in population and computing compound interest, and (*d*) It possesses algebraic properties.

Demerits: (*a*) It is difficult to calculate, (*b*) It cannot be calculated whenever zero value is present in the observations, and (*c*) It may not be identified with any of the given observations.

4.5. Harmonic Mean

It is defined as the reciprocal of the arithmetic mean of the reciprocals.

Case (*i*) *Variate values*

Let $X_1, X_2, \ldots, X_n$ be n observations, then the Harmonic mean is given by the formula

$$\text{H.M.} = \frac{1}{\frac{1}{n}\left(\frac{1}{X_1}+\frac{1}{X_2}+\ldots+\frac{1}{X_n}\right)} = \frac{n}{\left(\frac{1}{X_1}+\frac{1}{X_2}+\ldots+\frac{1}{X_n}\right)}$$

EXAMPLE : The following are the quantities of onion (in kgs) sold per rupee in 6 markets. Find the average quantity per rupee.

1.5, 0.75, 0 5, 2.0, 2.5, 1.4

$$\text{H.M.} = \frac{6}{\left(\frac{1}{1.5}+\frac{1}{0.75}+\ldots+\frac{1}{1.4}\right)} = 1.07 \text{ kg.}$$

Case (*ii*) *Grouped distribution*

Let $f_1, f_2, \ldots, f_n$ be n frequencies corresponding to the mid-values of the classes $X_1, X_2, \ldots, X_n$ respectively, then the Harmonic mean is given by the formula:

$$\text{H.M.} = \frac{N}{\frac{f_1}{X_1}+\frac{f_2}{X_2}+\ldots+\frac{f_n}{X_n}}$$

EXAMPLE: Compute the H.M. for the following grouped distribution.

TABLE 4.7

Class	*Frequency* (f_i)	*Mid-value* (X_i)	$\frac{f_i}{X_i}$
3-6	6	4.5	1.33
7-10	9	8.5	1.06
11-14	14	12.5	1.12
15-18	20	16.5	1.21
19-22	10	20.5	0.49
23-26	2	24.5	0.08
	61		5.29

$$\text{H.M.} = \frac{61}{5.29} = 11.53$$

Merits: (*a*) It is well defined, (*b*) It is based upon all the observations, and (*c*) It is useful in the cases like finding out average rate of work per hour, average quantity of a commodity per rupee, average distance travelled per hour, etc.

Demerits: (*a*) It is not much used in practice except in few cases mentioned above, (*b*) It is difficult to calculate, and (*c*) It gives more weightage to smaller values.

EXERCISES

1. The following table gives the yield of wheat from 10 equal plots.

Plot No.	1	2	3	4	5	6	7	8	9	10.
Yield (in kg.)	60	40	50	45	60	55	65	50	65	55.

If the area of each plot is 242 square yards, find the average yield per acre ? (*U.P. Board, 1963*)

2. The following is the distribution of heights of 85 plants.

Height (cms)	30-32	33-35	36-38	39-41	42-44	45-47
No. of plants	8	13	20	29	10	5

Find the Mean, Mode and Median heights of the plant.

3. Find the median for the following table relating to the number of grains per wheat blade.

No. of grains	20-24	24-28	28-32	32-36	36-40	40-44	44-48
No. of wheat ears	6	10	25	35	14	5	8

Locate also the 'median' from the cumulative frequency curve.

4. Find the median and mode for the following table.

No. of days absent		*No. of students*
More than	40	10
,,	30	25
,,	25	47
,,	15	47
,,	10	49
,,	5	67
,,	0	85

5. Compute the Geometric average of relative prices of the following commodities for the year 1939. (Base year 1938—Price 100).

Commodity	Rice	Corn	Wheat	Oats	Barley	Potates	Sugar
Relative price	118	129	100	131	150	144	126
Weight	17	1385	561	408	100	194	142

Calculate also the weighted Geometric Mean using the weights.

6. If a city had a population of 2,50,000 in a given year and 3,00,000 five years later. What was the average annual per cent change ?

7. Rice is being sold at the following rates (kg. per 5 rupees) at 10 different markets.

1.00, 0.80, 0.90, 0.70, 0.60, 1.10, 0.90, 0.75, 0.65, 0.45
Compute the average quantity of rice per 5 rupees.

8. The following is the distribution of Fat (percentage) in 100 samples collected from different milk centres in villages

Fat (%)	1-3	3-5	5-7	7-9	9-11
Samples	40	26	30	2	2

Compute Mean, Median, Mode, G.M and H.M. of Fat content per sample.

9. The following is the distribution of body weights of 100 calves at the 1st lactation

Body weight (kg)	30-40	40-50	50-60	60-70	70-80
Calves	12	26	34	20	8

Find Mean, Median, Mode, G.M. and H.M. of body weight of calves.

10. Compute the Arithmetic Mean yield (bags) of paddy given in the following distribution.

Yield (bags)	less than 20	less than 25	less than 30
Farms	6	18	30
Yield (bags)	less than 35	less than 40	less than 45
Farms	34	16	14

11. The following is the distribution of Annual Income (Rs.) of families in a locality of a city.

Less than 20,000	20,000-40,000	40,000-80,000
64	45	32
80,000-1,60,000	1,60,000-3,20,000	3,20,000 and above
26	10	13

Find A.M. ?

12. The following is the distribution of grades (10-point scale) obtained by a student in a semester final examination in different subjects.

Subject	A	B	C	D	E	F	G
Credits	3	4	2	3	1	2	4
Grade	8.2	7.6	8.7	9.0	8.5	7.4	8.8

Find the Grade point average obtained by the student in that semester.

13. Compute the average rainfall in a rainy season in a city of a particular year.

Rain fall (cms)	lessthan 2	2-4	4-6	6-8
Days	10	14	20	26
Rain fall (cms)	8-10	10 and above		
Days	8	12		

CHAPTER 5

MEASURES OF DISPERSION

Measure of Dispersion: It is a measure which can give the wide spread or scattering of observations among themselves or from a central point. The different measures of dispersion are (*i*) Range, (*ii*) Quartile deviation, (*iii*) Mean deviation, and (*iv*) Standard deviation.

5.1. Range

It is defined as the difference between the highest and lowest values in a series of observations.

EXAMPLE: Find the 'Range' for the following weights of 15 goats.

30, 25, 14, 42, 18, 26, 21, 11, 35 32, 29, 23, 20, 19, 13.

Range: 49—11=31.

Range is not much used in practice since it depends upon two extreme values. Therefore presence of any extremely high and low values in the observations will affect the range considerably. However, this measure is easy to compute. This is useful when the data are of homogeneous nature. It is also used in the preparation of control charts and for the data based on daily temperatures, rainfall, etc.

5.2. Quartile Deviation: (Semi-inter quartile range)

It is given by the formula, $Q.D.=(Q_3—Q_1)/2$ where Q_1= First quartile, Q_3=Third quartile. First and third quartiles are also called as lower and upper quartiles respectively.

5.2.1. First Quartile: It is that value of the variate below which one-fourth of the values lie and above which the remaining three-fourth of the values lie when the values are arranged in ascending order of magnitude.

5.2.2. Third Quartile: It is that value of the variate below which three-fourth of the values lie and above which the remaining one-fourth of the values lie when the values are arranged in ascending order of magnitude.

Case (*i*) *Variate values*

EXAMPLE: Find the Q.D. for the following observations on number of mesta plants in 10 equi-sized plots.

13, 9, 16, 4, 8, 19, 7, 23, 21, 12.

Arranging the values in ascending order of magnitude, we have 4, 7, 8, 9, 12, 13, 16, 19, 21, 23.

$$Q.\ No. = \frac{n+1}{4} = \frac{10+1}{4} = 2.75$$

Q_1 No. = 2nd value + 0.75 (3rd value—2nd value)

$$= 7 + 0.75 \times 1 = 7.75$$

$$Q_3\ No. = \frac{3n+1}{4} = \frac{3 \times 10+1}{4} = 7.75$$

Q_3 No. = 7th value + 0.75 (8th value—7th value)

$$= 16 + 0.75\ (19-16) = 18.25$$

$$Q.D. = \frac{18.25-7.75}{2} = 5.25$$

Case (*ii*) *Continuous distribution*

Let $f_1, f_2, \ldots, f_n$ be n frequencies corresponding to the mid-values of the classes $X_1, X_2, \ldots, X_n$ respectively then the first quartile, Q_1 is given by

$$Q_1 = l_1 + \frac{\left(\frac{(N+1)}{4} - m_1\right)}{f_1} \times C_1$$ where l_1 = lower limit of the first quartile class, $N = \Sigma f_1$, $\frac{N+1}{4}$ = first quartile number, m_1 = cumulative frequency just preceding to the first quartile class, f_1 = frequency of the first quartile class, C_1 = Class interval of the first quartile class. First quartile class is that class in which the cumulative frequency $\frac{N+1}{4}$ exists.

Similarly the third quartile is obtained by the formula

$$Q_3 = l_3 + \frac{\left(\frac{(3N+1)}{4} - m_3\right)}{f_3} \times C_3$$

where the symbols indicate the same as in the case of first quartile with first quartile replaced by third quartile.

EXAMPLE: Compute the Q.D. for the following distribution of marks obtained in an examination by 80 students.

TABLE 5.1

Marks	*No. of students*	*Cum. fre.*
0-5	3	3
5-10	10	13
10-15	18	31
15-20	25	56
20-25	9	65
25-30	8	73
30-35	7	80

Q_1 No.$=\frac{N+1}{4}=\frac{80+1}{4}=20.25$, **(10—15) is the first quartile** class since 20.25 lies in that class

$$Q_1 = 10 + \frac{(20.25-13)}{18}\times 5 = 12.01$$

Q_3 No.$=\frac{3N+1}{4}=\frac{3\times 80+1}{4}=60.25$, **(20—25) is the third quartile class since 60.25 lies in that class.**

$$Q_3 = 20 + \frac{(60.25-56)}{9}\times 5 = 22.36$$

$$\text{Q.D.} = \frac{Q_3-Q_1}{2} = \frac{22.36-12.01}{2} = 5.18$$

Unlike Range, the presence of abnormal values do not affect the Q.D. since the end values on either side do not figure in the definition.

5.3. Mean Deviation

It is the mean of the absolute values of the deviations taken from some average.

5.3.1. Mean Deviation about Mean: Let $X_1, X_2, \ldots, X_n$ be n observations, the Mean deviation about mean is given by the formula

$$\text{M.D. about mean} = \frac{1}{n}\Sigma \mid X_i - \bar{X} \mid \text{ where } \bar{X} = \text{A.M.}$$

By linear transformation method, we have

$$\text{M.D. about mean} = \frac{1}{n}\Sigma \mid d_i - \bar{d} \mid \text{ where } d_i = (X_i - A) \text{ and}$$

$$\bar{d} = \Sigma \frac{d}{n}$$

EXAMPLE Find the M.D. about mean for the following data.
4, 9, 10, 14, 7, 8, 6, 14.

TABLE 5.2

									Total
X_i	4	9	10	14	7	8	6	14	72
$\mid X_i - \bar{X} \mid$	5	0	1	5	2	1	3	5	22

$\bar{X}=9$, M.D. about mean$=22/8=2.75$

Case (*ii*) *Continuous frequency distribution*

Let $f_1, f_2, \ldots, f_n$ be n frequencies corresponding to the mid-values of the classes $X_1, X_2, \ldots, X_n$ respectively then the M.D. about mean is given by the formula

$$\text{M.D. about mean} = \frac{1}{N}\Sigma f_i \mid X_i - \bar{X} \mid \text{ where } \bar{X} = \text{Mean}$$

EXAMPLE: Find the M.D. about mean for the following grouped distribution.

TABLE 5.3

Yield of milk per day (in kgs)	*No. of dairy animals* (f_i)	*Mid-value* X_i	$f_i X_i$	$X_i - \bar{X}$	$f_i \mid X_i - \bar{X} \mid$
0-2	6	1	6	5.66	33.96
2-4	10	3	30	3.66	36.60
4-6	14	5	70	1.66	23.24
6-8	18	7	126	0.34	6.12
8-10	11	9	99	2.34	25.74
10-12	7	11	77	4.34	30.38
12-14	5	13	65	6.34	31.70
	71		473		187.74

$\bar{X}=473/71=6.66$

M.D. about mean=187.74/71=2.64

M.D. can be computed by taking deviations from any other average like Median, Mode, etc. on the similar lines given above wherein Median, Mode, etc. will be substituted in the formula in place of Mean. M.D. is least when deviations are taken from the Median.

5.4. Standard Deviation

It is defined as the square root of the mean of the squares of the deviations taken from Arithmetic mean.

5.4.1. Variate Values: Let $X_1, X_2, ..., X_n$ be n observations then standard deviation (S.D.) is given by the formula

$$\sigma=\sqrt{\frac{1}{n}\Sigma(X_i-\bar{X})^2} \qquad \text{where } \bar{X}=\text{A.M.}$$

Simplifying the above formula, we have

$$\sigma=\sqrt{\frac{1}{n}\left[\Sigma X_i^2-\frac{(\Sigma X_i)^2}{n}\right]} \quad , \sigma^2=\text{variance}$$

By linear transformation method, we have

If $d_i=(X_i-A)$ where A =Arbitrary mean

$$\sigma=\sqrt{\frac{1}{n}\left[\Sigma d_i^2-\frac{(\Sigma d_i)^2}{n}\right]}$$

EXAMPLE: Compute the S.D. of the following data based on number of seeds germinated out of 20 in each of the ten petty dishes.

15, 13, 10, 17, 8, 12, 14, 11, 13, 15

TABLE 5.4

											Total
Xi	15	13	10	17	8	12	14	11	13	15	128
Xi^2	225	169	100	189	64	144	196	121	169	225	1702

$$\sigma=\sqrt{\frac{1}{10}\left[1702-\frac{(128)^2}{10}\right]} \qquad =2.52$$

$\sigma^2 = 6.35$

5.4.2. Continuous Frequency Distribution. Let $f_1, f_2, ..., f_n$ be n frequencies corresponding to the mid-values of the classes X_1

$X_2, \ldots, X_n$ respectively, then the standard deviation is given by the formula

$$\sigma=\sqrt{\frac{1}{N}\Sigma f_i\ (X_i-\bar{X})^2}$$

Simplifying, we have

$$\sigma=\sqrt{\frac{1}{N}\left[\Sigma f_i X_i^2-\frac{(\Sigma f_i X_i)^2}{N}\right]}$$

By linear transformation method, we have

$$\sigma=C\times\sqrt{\frac{1}{N}\left[\Sigma f_i d^2_i-\frac{(\Sigma f_i d_i)^2}{N}\right]}$$

Where $d_i = \frac{X_i-A}{C}$, A=Arbitrary mean and C=class interval.

EXAMPLE: Find the standard deviation and variance for the following distribution of lengths of wheat ears.

TABLE 5 5

Length of wheat ear	*No. of ears* (f_i)	*Mid-value* (X_i)	$f_i X_i$	$f_i X_i^2$	d_i	$f_i d_i$	$f_i d_i^2$
7-9	8	8	64	512	—2	—16	32
9-11	18	10	180	1800	—1	—18	18
11-13	25	12	300	3600	0	0	0
13-15	15	14	210	2940	1	15	15
15-17	6	16	96	1536	2	12	24
	72		850	10388		—7	89

(*i*) *Direct method*

$$\sigma=\sqrt{\frac{1}{72}\left[10388-\frac{(850)^2}{72}\right]}=2.22$$

Variance=4.91

$$\sigma=2\times\sqrt{\frac{1}{72}\left[89-\frac{(-7)^2}{72}\right]}=2.22$$

Variance=4.91

This is most commonly used measure of dispersion as a counterpart of mean in the case of 'measures of location'. This gives minimum value when the deviations are taken from the mean.

5.5 Coefficient of Variation

Sometimes it is necessary to express variation of a series of data relative to an average. For example, the variation of 2 to 3 quintals per acre in the production would be significant for a local variety of paddy but not so in the case high yielding variety. Hence, coefficient of variation (C.V.) which is the percentage ratio of S.D. to Mean is calculated for each of the variety. The one which is having less coefficient of variation (C.V.) is considered more consistent variety. Since C.V. is independent of units, it is useful for comparison of any two series with different units. The C.V. is also found useful to compare two players with respect to their consistency in scoring.

$$\text{C.V.} = \frac{\text{S.D.}}{\text{Mean}} \times 100$$

EXAMPLE : The scores of two candidates A and B in different one-hour examinations are given below. Examine who is the more consistent scorer.

TABLE 5.6

Candidate	*One-hour examination*					
	I	*II*	*III*	*IV*	*V*	*VI*
A	9.0	8.0	7.5	8.5	9.0	8.0
B	5.5	9.5	6.5	8.5	10.0	8.0

Arbitrary Mean = 8.0, $\Sigma d_i = 2.00$, $\Sigma d_i^2 = 2.50$

Candidate A : Mean = 8 + 2/6 = 8.33

$$\text{S.D.} = \sqrt{1/6\left[2.5 - \frac{(2)^2}{6}\right]} \qquad \text{C.V.} = \frac{0.55}{8.33} \times 100$$

$$= 0.55 \qquad\qquad = 6.60$$

Candidate B : Mean = 8 + 0/6 = 8.00, $\Sigma d_2 = 0$, $\Sigma d^2_2 = 15.00$

$$\text{S.D.} = \sqrt{1/6\left[15 - \frac{(0)^2}{6}\right]} \qquad \text{C.V.} = \frac{1.58}{8.00} \times 100$$

$$= 1.58 \qquad\qquad = 19.75$$

Threrefore, candidate A is more consistent.

5.6 Statistical Population

An aggregate of animate or inanimate objects is called statistical population. For example, large group of data on heights, weights, etc., is known as Statistical population.

5.7. Sample

To study particular character it is always not possible to study the whole lot or whole population as it requires more time and money. Therefore, we have to rely upon a part of the population for our study. This part or portion of a population is called sample. However, the sample should be as far as possible representative of the population with respect to character under consideration. This can be ensured by drawing the units from population at random so that each and every sample of equal size will be selected in the sample with equal probability.

Now the value of the mean based on sample of observations need not be equal to the population mean. The difference between the sample mean and the population mean is called 'sampling error'. Suppose if we take all possible samples of equal size, the means based on these samples follow a distribution known as 'sampling distribution'. The mean of all the means of samples of equal size is an estimate of the population mean, and the standard deviation of the means of these samples is known as 'standard error of mean'.

Since it is difficult, in general, to study all the possible samples, we have to depend on a single sample. The standard error of mean based on a single sample is given as

$$S.E.(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$ where σ=S.D. in the population, $\bar{X}$= Mean of a sample, n=size of the sample.

If σ is not known, it is estimated from a sample of observations as follows:

Case (*i*) If n is large sample (Say $>$ 30)

$S.E.(\bar{X}) = S/\sqrt{n}$ where $S = \sqrt{1/n \quad \Sigma(X_i - \bar{X})^2}$

Case (*ii*) If n is small sample (Say $<$ 30)

$$S.E.(\bar{X}) = \frac{s}{\sqrt{n}}$$ where s=unbased estimate of σ

$$s = \sqrt{\frac{1}{n-1}\Sigma(X_i - \bar{X})^2}$$

EXERCISES

1. The country's foodgrains output (in million tons) for 20

years are given as :

75, 74, 80, 81, 85, 86, 84, 81, 90, 87, 92, 94, 95, 93, 98, 96, 94, 99, 109, 110.

Obtain the values of Range, Mean deviation about Mean and Median and Coefficient of Variation.

2. The following table gives the yield of paddy in maunds per acre based on crop cutting experiments in a certain area. during 1940-41.

Yield (Maunds Acre)	*Freq.*	*Yield (Maunds/Acre)*	*Freq.*
0	4	24	128
3	4	27	73
6	32	30	50
9	81	33	13
12	135	36	12
15	198	39	5
18	210	42	1
21	144		

Calculate the Arithmetic mean, Standard deviation and quartile deviation of the distribution. *(I.A.S., 1949)*

3. Find the 'mean deviation about mode' for the following grouped distribution.

Class	3-7	7-11	11-15	15-19	19-23	23-27
Freq.	5	7	13	31	18	4

4. A distribution consists of three components with frequencies 200, 250 and 300 having means 25, 20 and 15 with standard deviations 3, 6 and 5 respectively. Find the 'mean' and 'standard deviation' of the combined distribution.

5. If any two series, where d_1 and d_2 represent the deviations from the same arbitrary mean, 15, the following results are given.

$n_1=12$ $\Sigma d_1=25$ $\Sigma d_1^2=650$

$n_2=20$ $\Sigma d_2=-20$ $\Sigma d_2^2=480$

Compute the coefficient of variation for both the series and determine which is more consistent series.

6. Below are the scores of two cricketeers in 10 innings. Find who is the more 'consistent scorer'.

A	204	68	150	30	70	95	60	76	24	19
B	99	190	130	94	80	89	69	85	65	40

7. Compute the coefficient of variation and standard error of mean given the following distribution of protein (percentage) content in 100 samples of red gram collected from different farms.

Protein (%)	less than 4	4-8	8-12	12-16	16-20
Samples	28	20	26	10	16

8. The following is the distribution of yields (quintals) per hectare in different farms in a Research Zone.

Yield (quintals)	10-16	16-22	22-28	28-34	34 and above
Farms	8	14	26	22	10

Compute coefficient of variation for the above distribution of yields.

9. Find the standard deviation of the following distribution of leaf areas (square mm) of sun flower crop in an experimental field.

Leaf area (sq.mm)	20-30	30-40	40-50	50-60
Leaves	8	14	28	16
Leaf area (sq.mm)	60-70	70-80	80 and above	
Leaves	10	20	10	

10. The following is the distribution of ear lengths (cm) of paddy crop of high yielding variety in an experimental field.

Ear length (cm)	less than 4	4-6	6-8	8-10
Ears	16	20	24	18
Ear length (cm)	10-12	12 and above		
Ears	15	7		

Obtain coeofficient of variation, quartile deviation and mean 'deviation' from mean for the above data.

11. The following is the distribution of heights of maize plants (cms) in an experimental field in a reasearch station.

Height (cms)	80-100	100-120	120-140	140-160
Plants	28	16	30	17
Height (cms)	160-180			
Plants	9			

Obtain 'standard deviation' and 'standard error of mean' for the aobve distribution of heights.

12. Find mean deviation from 'Median' and 'Mode' for the following distribution of rain fall (mm) in July month at an Agricultural Research station.

60,	76,	16,	18,	32,	18,	34,	46,	76,
76,	90,	92,	93,	84,	80,	54,	62,	70,
100,	85,	22,	23,	26,	48,	78,	76,	91,
56,	50,	66,	81					

13. Compute mean deviation from 'Mode' for the following distribution of yields (kg) of grapes in different grape gardens in an year.

Grape yields (kgs)	300-400	400-500	500-600
Gardens	8	16	19
Grape yields (kgs)	600-700	700-800	
Gardens	9	3	

14. The following are the minimum temperatures (celcius) recorded in a hill station of North India in the month of January.

6,	3,	2,	0,	−1,	−6,	3,	2,	10,	7
8,	9,	3,	4,	11,	5,	3,	0,	6,	3
−4,	−2,	1,	0,	3,	−1,	6,	2,	1,	3,
5.									

Compute 'Range' and 'Quartile' deviation'.

15. The following is the frequency distribution of number of Custard Apples per tree in a graden.

Custard Apples	150	172	175	184	189
Trees	6	10	15	8	12
Custard Apples	201	210			
Trees	13	5			

'Find Mean deviation' from 'Median' for the above distribution.

CHAPTER 6

MOMENTS, SKEWNESS AND KURTOSIS

6.1. Moments

Let $X_1, X_2, \ldots, X_n$ be n observations, then 'k'-th raw moment is defined by

$V_k = 1/n \Sigma (X_i - A)^k$ where A = Arbitrary mean

The 'k'-th central moment is given by

$\mu_k = 1/n\ \Sigma (X_i - \bar{X})^k$ where $\bar{X}$ = A.M.

In the case of a frequency distribution, the k-th raw and central moments are given by V_k and μ_k respectively, as

$$V_k = \frac{1}{N} \Sigma f(X_i - A)^k, \quad \mu_k = \frac{1}{N} \Sigma f(X_i - \bar{X})^k \quad \ldots (6.1)$$

where $N = \Sigma f_i$ = Total frequency.

The relation between k-th central moment and raw moments is given by

$$\mu_k = V_k - \binom{k}{1} V_{k-1} V_1 + \binom{k}{2} V_{k-2} V_1^2 - \binom{k}{3} V_{k-3} V_1^3 + \ldots + (-1)^k V_1^k \ldots (6.2)$$

where $V_{k-1}, V_{k-2}, \ldots V_1$, are (k—1)-th, (k—2)-th, ..., 1st raw moments respectively. $\binom{k}{1}$, $\binom{k}{2}$ etc., are the number of combinations taking 1 at a time, 2 at a time, etc., respectively out of k values.

If $k=1$, $\mu_1 = V_1 - V_1 = 0$

If $k=2$, $\mu_2 = V_2 - V_1^2$ = variance

If $k=3$, $\mu_3 = V_3 - 3V_2V_1 + 2V_1^3$

If $k=4$, $\mu_4 = V_4 - 4V_3V_1 + 6V_2V_1^2 - 3V_1^4$

The central moments are useful in measuring 'skewness' and 'kurtosis' of curves.

6.2. Skewness

Sometimes, even if the two measures like 'Mean' and 'standard deviation' are same for the distributions still the shape of the two curves may differ. For example, one curve may be symmetric and the other may be asymmetric. We shall define here 'symmetric' and 'asymmetric' curves for the Uni-modal frequency distribution.

6.2.1. Symmetric Curve: A symmetric curve is one where the shape of the curve on either side of the mean is identical. That is, in a frequency distribution, the frequencies on either side of a mean should be equal. In this curve, the mean, median and mode coincide at one point and the ordinate drawn from peak of the curve to mean on the X-axis would bifurcate the area under the curve into two equal halves. The skewness (or bending) in this case is zero. The symmetric curve is depicted in Fig. 6.1.

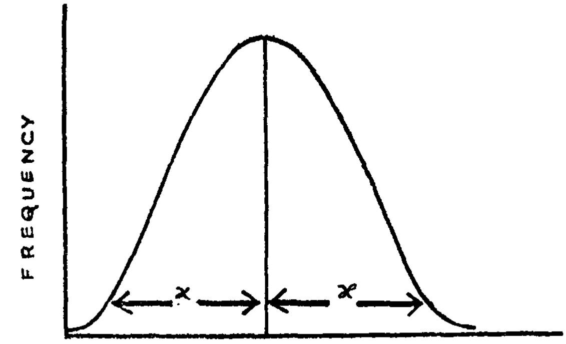

Fig. 6.1. Symmetric curve.

6 2.2. Asymmetric Curve: A curve which is not symmetric is known as 'Asymmetric' or 'skewed' curve. In this case, the peak of the curve may bend towards right or towards left with respect to mean.

The curve bending towards right from the mean and having a long tail on left is said to be negatively skewed. This curve is shown in Fig. 6.2.

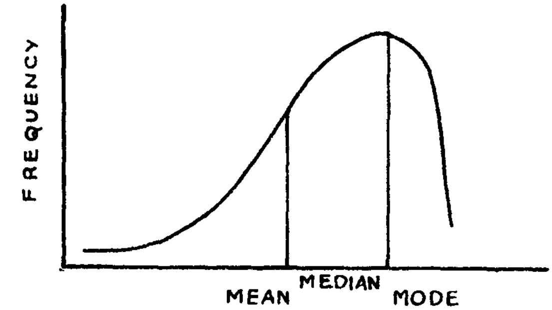

Fig 6.2. Negative skewness.

The curve bending towards left from the mean and having long tail towards right is said to be positively skewed. The curve is shown in Fig. 6.3.

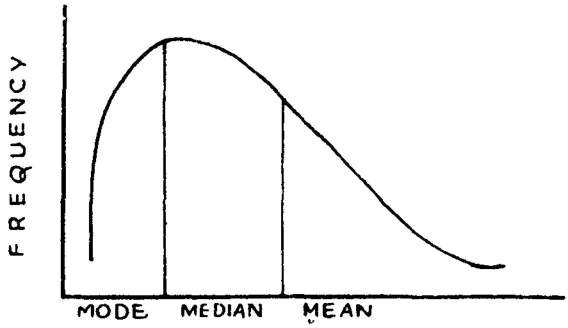

Fig. 6.3. Positive skewness.

The different measures of coefficient of skewness are given by

1. Pearson's Coefficient of skewness

$$=\frac{(\text{Mean}-\text{Mode})}{\text{S.D.}} \quad ...(6.3)$$

This measure is due to K. Pearson. In Fig. 6.2, Mode will be greater than Mean and hence (Mean—Mode) is negative. Therefore, the coefficient of skewness is negative. In Fig. 6.3, Mode will be less than Mean and hence (Mean—Mode) is positive. Hence, the coefficient of skewness is positive since the denominator in the formula is always positive. In Fig. 6.1, Mean is equal to Mode and hence (Mean—Mode) is zero. In the formula of coefficient of skewness, S.D. is used in order to make the coefficient independent of units so as to facilitate the comparison of two or more distributions. Median always lies between Mean and Mode in Figs. 6.2 and 6.3.

2. Quartile coefficient of skewness $=\frac{(Q_3-Q_2)-(Q_2-Q_1)}{(Q_3-Q_1)}$ ...(6.4)

where Q_1, Q_2 and Q_3 are the 1st, 2nd and 3rd quartiles respectively. This coefficient always lies between -1 and $+1$. In this case also the denominator is taken to make the coefficient independent of units.

3. Moment coefficient of skewness, $\sqrt{\beta_1}=\frac{\mu_3}{\mu_2^{3/2}}$ ...(6.5)

where μ_2 and μ_3 are 2nd and 3rd central moments respectively. In this formula μ_3 measures the excess of negative deviations over positive deviations and excess of positive deviations over negative deviations in Fig. 6.2 and Fig. 6.3, respectively. Here also the denominator is used to make the coefficient independent of units.

6.3. Kurtosis

The shape of the Vertex of the curve is known as Kurtosis. The measure of Kurtosis is known as coefficient of Kurtosis and is denoted by β_2.

6.3.1. Platykurtic: The peak or Vertex of the curve is more flat and the tails on both sides are long compared to normal curve (chapter 9). Here $\beta_2 < 3$.

6.3.2. Mesokurtic: The peak of the curve is normal and the tails on both sides are also normal. Here $\beta_2 = 3$.

6.3.3. Leptokurtic: The peak of the curve is narrow and sharp and the tails on both sides are short compared to normal curve. Here $\beta_2 > 3$.

The above three curves are depicted in Fig. 6.4. $(\beta_2 - 3)$ is taken as the departure from normality. This quantity would be negative for Platykurtic, zero for Mesokurtic and positive for Leptokurtic. The measure of Kurtosis is denoted by coefficient of Kurtosis and is given by the formula, $\beta_2 = \frac{\mu_4}{\mu_2^2}$...(6.6)

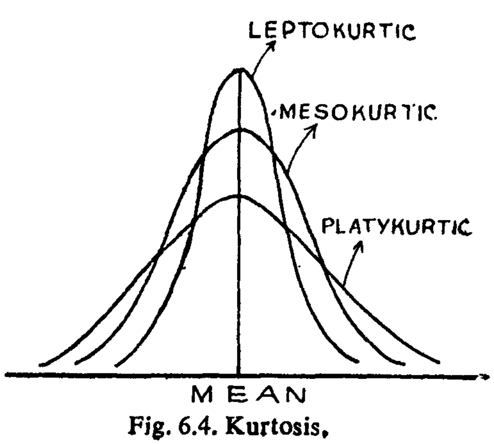

Fig. 6.4. Kurtosis.

Sometimes it is convenient to express the coefficients of skewness and kurtosis in terms of γ_1 and γ_2 respectively, where

$\gamma_1 = \beta_1$, $\gamma_2 = (\beta_2 - 3)$.

EXAMPLE: Compute the different coefficients of skewness and kurtosis for the following data on milk yield.

TABLE 6.1

Milk yield (kgs)	*No. of cows (f_i)*	*Mid-value X_i*	d_i	$f_i d_i$	$f_i d_i^2$	$f_i d_i^3$	$f_i d_i^4$
4-6	8	5	−3	−24	72	−216	648
6-8	10	7	−2	−20	40	−80	160
8-10	27	9	−1	−27	27	−27	27
10-12	38	11	0	0	0	0	0
12-14	25	13	1	25	25	25	25
14-16	20	15	2	40	80	160	320
16-18	7	17	3	21	63	189	567
	135		0	15	307	51	1747

$d_1 = \frac{X - A}{C}$ where A = 11, C = 2.

$V_1 = C/N\ \Sigma f_i\ d_1 = 2 \times 15/135 = 0.222$

$V_2 = C^2/N\ \Sigma f_i\ d_1^2 = 4 \times 307/135 = 9.096$

$V_3 = C^3/N\ \Sigma f_i\ d_1^3 = 8 \times 51/135 = 3.022$

$V_4 = C^4/N\ \Sigma f_i\ d_1^4 = 16 \times 1747/135 = 207.052$

$\mu_2 = V_2 - V_1^2 = 9.096 - (0.222)^2 = 9.047$

$\mu_3 = V_3 - 3\ V_2\ V_1 + 2V_1^3$

$= 3.022 - 3(9.096)\ (0.222) + 2\ (0.222)^3 = -3.014$

$\mu_4 = V_4 - 4\ V_3 V_1 + 6\ V_2 V_1^4$

$= 207.052 - 4\ (3.022)\ (0.222) + 6(9.096)\ (0.222)^2 - 3 \times (0.222)^4$

$= 207.051$

Moment coefficient of skewness, $\sqrt{\beta_1} = \frac{\mu_3}{\mu_2^{3/2}}$

$= -3.014/(9.047)^{3/2} = -0.1108$

Mean = 11.222, Mode = 10.917, S.D. = 3.008

Pearson's coefficient of skewness $= \frac{11.222 - 10.917}{3.008} = 0.1014$

$Q_1=9.185$, $Q_2=11.211$, $Q_3=13.480$

Quartile coefficient of skewness

$$=\frac{(13.480-11.211)-(11.211-9.185)}{(13.480-9.185)}=0.0566$$

Coefficient of Kurtosis, $\beta_2=\frac{\mu_4}{\mu_2^2}$

$$=207.051/(9.047)^2=2.530$$

The coefficients of skewness obtained by Pearson's coefficient of skewness and quartile coefficient of skewness are positive whereas the Moment coefficient of skewness gave negative value. The coefficient of Kurtosis indicates that the curve is platykurtic since (β_2-3) i.e., (2.529-3) is negative.

EXERCISES

1. In a frequency distribution, size 701, range 30-150 divided into 8 class intervals of equal width, the first three moments measured in terms of the scale units u from the mid point of the fourth class interval from the top are $\Sigma fu=-150$, $\Sigma fu^2=1532$, $\Sigma fu^3=-750$. Determine A.M. and the values of the first three moments from the mean in terms of the original unit. Calculate the standard deviation. *(B.Sc Madras, 1944)*

2. If the first three moments about an arbitrary mean 4 are 2, 15, and 84. Calculate the mean, variance and third moment about mean.

3. Define skewness and arrange median, mode and mean in ascending order of their magnitude for positively skewed curve.

4. Define r-th moment about mean and give the formula for 4-th moment about mean in terms of the moments about arbitrary mean.

5. For any two groups of data A and B the statistical constants are

	A	B
Median	19.64	24.46
Lower quartile	13.46	15.64
Upper quartile	25.94	37.76

Comment on the dispersion and skewness of A and B.

6. Define the various measures of dispersion and discuss their relative advantages.

Find the standard deviation and the Pearson coefficient of skewness for the following distribution.

Percentage ash content	3.0-3.9	4.0-4.9	5.0-5.9	6.0-6.9	7.0-7.9	8.0-8.9	9.0-9.9	10.0-10.9
Frequency	3	7	28	78	84	45	28	7

7. Find the 3rd and 4th central moments given the following raw moments

$V_4 = 40$, $V_3 = 10$, $V_2 = 5$ and $V_1 = 2$.

8. Obtain the 'quartile coefficient of skewness' given the following data on minimum temperatures [celcius] on 11 days in December month in a city.

4, 5, 6, 7, 10, 3, 1, 5, 12, 18, 16

9. Find the coefficient of Kurtosis given the following raw moments on rain fall data and also specify the type of kurtosis.

$V_1 = 2$, $V_2 = 6$, $V_3 = 8$ and $V_4 = 20$

10. Find the 'coefficient of skewness' given the following on yields of wheet in a region

Mean = 40, Mode = 50 Variance = 6

11. Obtain 3rd and 4th central moments for the following data on disease effected birds in 10 poultry farms

6, 13, 10, 2, 21, 6, 150, 96, 74, 65

12. Compute coefficients of 'skewness' and 'kurtosis' for the following data on rainfall (cm) in the month of August at a Regional Agricultural Research Station in Coastal Andhra Pradesh.

2, 0, 6, 8, 13, 7, 2, 2, 4, 10, 11, 2, 0, 5, 11, 0, 0, 8, 7, 9, 6, 5, 12, 3, 4, 5, 0, 6, 1, 2, 4

13. Find the coefficients of 'skewness' and Kurtosis for the following data on iron content (percentage) in samples of leafy vegetable sold in markets and also draw diagrams.

Iron content (%)	Samples
0 – 2	6
2 – 4	12
4 – 6	18
6 – 8	9
8 – 10	5