UNIT 3 SUMMARISATION OF UNIVARIATE DATA*

Structure

3.0 OBJECTIVES

After going through this unit, you will be able to:

- compute numerical quantities that measure the central tendency of a set of data such as, mean, median, mode, geometric mean and harmonic mean;
- explain the concept of dispersion;
- compute numerical quantities that measure the dispersion of a set of data;
- explain chebychev's inequality;
- compute the coefficient of variation; and
- find a measure for concentration of certain distribution of data.

* Adapted from IGNOU study material of EEC 13: Elementary Statistical Methods and Survey Techniques, Units 4, 5 and 6 written by R S Bharadwaj with modifications by K. Barik

3.1 INTRODUCTION

In the previous Unit we had discussed about condensation of raw data by grouping them into a few class intervals and presenting in the form of a table or diagram. Such tables or diagrams provide a rough idea of the distribution of observations. Often we need to compare between distributions. In such situations it is difficult to compare tables or diagrams simply by looking at them. It is much more convenient and useful for comparison if we could find out a single numerical value for describing the data.

Measures of Central Tendency (or Location) constitute one of the major statistics designed for this purpose. There are five main measures of central tendency. These are Arithmetic Mean, Geometric Mean, Median and Mode. You will learn about each one of these measures below.

3.2 MEASURES OF CENTRAL TENDENCY

In frequency distributions of observations discussed in Unit 2 we notice that the observations tend to cluster around a central value. This phenomenon of clustering around a central value in a frequency distribution is called '*Central Tendency'*. Thus, it is of interest to locate such a value around which clustering of observations takes place. There are several measures of central tendency (or location) of a frequency distribution. These measures produce numbers that summarise a frequency distribution in terms of one its properties, namely, central tendency.

3.2.1 Arithmetic Mean

The *average* or the *arithmetic mean*, or simply the *mean* when there is no ambiguity, is the most common measure of central tendency. It is defined as the sum total of all values in the sample divided by the number of observations. It is denoted by a bar above the symbol of the variable being averaged. Thus $\overline{X}$ stands for the mean of *X*-values in the sample. If in a sample a particular *X*-value, say X_i occurs with frequency $f_i (i = 1,2,...,n)$, its contribution to the total of X-values is $f_i X_i$. Thus, we can compute the mean of X-values by

$$\overline{X} = \frac{1}{N}(f_1X_1 + f_2X_2 + ... + f_nX_n) = \frac{\sum_{i=1}^{n} f_iX_i}{N}, \qquad \text{where } N = \sum_{i=1}^{n} f_i.$$

When observations are classified into class intervals, as for continuous variables, individual observations falling into a class intervals are not separately identifiable and be contribution of the individual observations from a class intervals to the total cannot be calculated. To avoid this difficulty, it is assumed that every observation falling into a class interval has a value equal to the *mid-point* into which these observations fall. Such a procedure will not give the exact mean had we computed it from raw data and may require what is called corrections for grouping.

Example 3.1: Compute the mean for discrete frequency distribution of Table 3.1.

Table 3.1

Frequency distribution of 100 households by size

Household Size (X_i)	**Frequency (f_i)**
1	3
2	16
3	25
4	33
5	12
6	7
7	2
8	2
Total	100

Let us compute the arithmetic mean of the data given in the above table.

$$\overline{X} = \frac{\sum_{i=1}^{n} f_i X_i}{N} = \frac{1\times3 + 2\times16 + 3\times25 + 4\times33 + 5\times12 + 6\times7 + 7\times2 + 8\times2}{100} = \frac{.374}{100} = 3.74$$

Thus, mean household size based on 100 households is 3.74.

Example 3.2: Compute the mean for grouped frequency distribution of Table 3.2.

Table 3.2

Frequency distribution of 100 households by average monthly household expenditure on food

Expenditure class (Rs.)	**Frequency**
262.5 – 286.5	1
286.5 – 310.5	14
310.5 – 334.5	16
334.5 – 358.5	28
358.5 – 382.5	26
382.5 – 406.5	15
Total	**100**

For computation of the mean we have to construct table as given below.

Class interval (Rs.) (1)	**Mid-point (X_i) (2)**	**Frequency (f_i) (2)**	**$f_i X_i$ (4)**
262.5 -286.5	274.5	1	274.5
286.5 – 310.5	298.5	14	4179.0
310.5 – 334.5	322.5	16	5160.0
334.5 – 358.5	346.5	28	9702.0
358.5 – 382.5	370.5	26	9633.0
382.5 – 406.5	394.5	15	5917.5
Total		**100**	**34866.0**

Thus, mean of monthly average household expenditure on food is

$$\overline{X} = \frac{34866}{100} = \text{Rs.}348.66.$$

We should note from the above example that to find column (3) we need to multiply the corresponding values of column (1) and (2), and often hand computations are long for each multiplication. These computations can be simplified, particularly when successive column (1) values are equidistant (but applicable otherwise also), by making the following simple transformation.

For $i = 1, 2, \ldots, n$

$$u_i = \frac{X_i}{h} \qquad \text{i.e., } X_i = A + hu_i \text{ and so } \overline{X} = A + h\overline{u}.$$

Often A is called the 'assumed mean' and $h\overline{u}$ as its correction to get $\overline{X}$. Choice of A and h are made so that computation of $\overline{u}$ becomes simple. Usually A is take as that X value for which the frequency is largest. For equidistant successive X-values is column (1), h may be taken as the difference between two successive X-values. For equal length class intervals, the difference between successive mid-points is the same as the length of each class interval.

We will explain this method by re-computing the mean of the monthly average household food expenditure data given in Table 3.2. We construct Table 3.3 by using A and h as explained below.

We define A = Mid-point of the class with largest frequency = 346.5 and

h = Common length of each class interval = 24.

Thus, $$u_i = \frac{X_i - 346.5}{24}$$

Table 3.3

Computation of Mean of Frequency Distribution of Table 3.2

Class interval (Rs.)	**Mid-point (X_i)**	$u_i = \frac{X_i - 346.5}{24}$	**frequency (f_i)**	**$f_i u_i$**
262.5 – 286.5	274.5	– 3	1	–3
286.5 – 310.5	298.5	–2	14	–28
310.5 – 334.5	322.5	–1	16	–16
334.5 – 358.5	346.5	0	28	0
358.5 – 382.5	370.5	1	26	26
382.5 – 406.5	394.5	2	15	30
Total			**100**	**9**

We find out that

$$\bar{u} = \frac{1}{N}\sum_{i=1}^{n} f_i u_i = \frac{1}{100} \times 9 = \frac{9}{100}$$

Thus, $X = A + h \times \bar{u} = 346.5 + 24 \times \frac{9}{100} = \text{Rs.}348.66$ as was computed earlier.

Properties of Arithmetic Mean

1) *The algebraic sum of deviations of a given set of observations is zero when taken from the arithmetic mean.*

   Let $X_1, X_2, ..., X_n$ be n observations with respective frequencies as $f_1, f_2, ..., f_n$. Mathematically, this property implies that $\sum_{i=1}^{n} f_i (X_i - \bar{X}) = 0$, where $X_I - \bar{X}$ is the deviation of i^{th} observation from mean. To prove the above property, we write

$$\sum_{i=1}^{n} f_i (X_i - \bar{X}) = \sum_{i=1}^{n} f_i X_i - \bar{X} \sum_{i=1}^{n} f_i = \sum_{i=1}^{n} f_i X_i - n.\bar{X} = 0.$$

   Hence, the result,

2) The sum of squares of deviations of a given set of observations is minimum when taken from the arithmetic mean.

   Mathematically, this property implies that for any arbitrarily chosen origin, A,

$$S = \sum_{i=1}^{n} f_i (X_i - A)^2 \text{ is minimum when } A = \bar{X}.$$

To prove this property, we note that the magnitude of *S* will depend upon the selected value of *A*. thus, we can say that *S* is a function of *A*.

We want to find that value of A for which S is minimum. Using calculus, this value is given by the equation $\frac{dS}{dA} = 0$ such that $\frac{d^2S}{dA^2} > 0$.

(Remember that the value of a function is minimum when first derivative is zero and second derivative is positive.)

Differentiating S with respect to A and equating to zero, we get

$$\frac{dS}{dA} = -2\sum_{i=1}^{n} f_i(X_i - A) = 0$$

This implies that

$$\sum_{i=1}^{n} f_i X_i - A\sum_{i=1}^{n} f_i = 0 \quad \text{or} \quad A = \frac{\sum_{i=1}^{n} f_i X_i}{\sum_{i=1}^{n} f_i} = \overline{X}.$$

Further, it can be shown that $\frac{d^2S}{dA^2} > 0$ when $A = \overline{X}$.

3.2.2 Median

Median of a distribution locates a central point which divides a distribution into two equal halves, i.e., it is the middle most value among a set of observations. Let us start with examples in a discrete case. Consider a data set having 5 distinct observations: 2, 4, 9, 12, 19 (arranged is ascending order). Here 9 is the middle most value since an equal number of observations are to its left and to its right. Thus, 9 is the median of the above observations. Consider another data set having 6 distinct observations: 3, 8, 15, 25, 35, 43. Here any point between 15 and 25 has the property that equal number of observations are to its left and to its right. Any point in the interval 15 to 25 may be used as a median. Conventionally we take the middle point of such an interval to define median uniquely. Thus 20 is the median of 3, 8, 15, 25, 35, 43.

When a data set has non-distinct observations – a situation more common in practice – difficulties' may arise. In such situations, it may not be always possible to locate the middle most value or the central point that divides the distribution into two equal halves, For example, in the case of the data set having 5 observations 2, 9, 9, 12, 19 the value 9 is repeated twice. Thus, a form definition of median is needed to overcome such difficulties.

A median of a distribution is a point or a central value such that ***at least*** *50% of the observations are less than or equal to it and* ***at least*** *50% of the observations are greater than or equal to it*. With this definition of median and the convention of taking the middle point of a class in which each point i a median, median of a distribution can always be specified uniquely. Thus, median of observations 2, 9, 9, 12, 19 is 9 because 3 of the 5 observations (60%) are less than or equal to 9 and 4 of the 5 observations (80%) are greater than or equal to 9.

Let us find out the median household size from the frequency distribution in Table 3.1. We notice that 77 (out of 100) households have family size of less than or equal to 4 and 56 households have family size of more than or equal to 4. Thus median family size in this case is 4.

Median for a grouped frequency distribution of a continuous variable is easier to understand if we look at the associated histogram with height of a rectangle equal to the frequency density, $\frac{f}{h}$, of the class. In such a histogram, the area of a rectangle gives the frequency of the corresponding class. The median, in this case, is a point in one of the classes such that the areas to its left and to its right are 50% each. First step is to locate the class, up to the right boundary of which the total areas is at least 50% (*called the median class*). Then the median is computed by adding, to the lower boundary value of this class, the length of a part of this class interval in proportion to the frequency needed to achieve 50%. A convenient method of finding out the median class is to compute the cumulative frequency (discussed in Unit 2, Section 2.3.3) and identifying the class interval in which the $\frac{N}{2}$th observation lies.

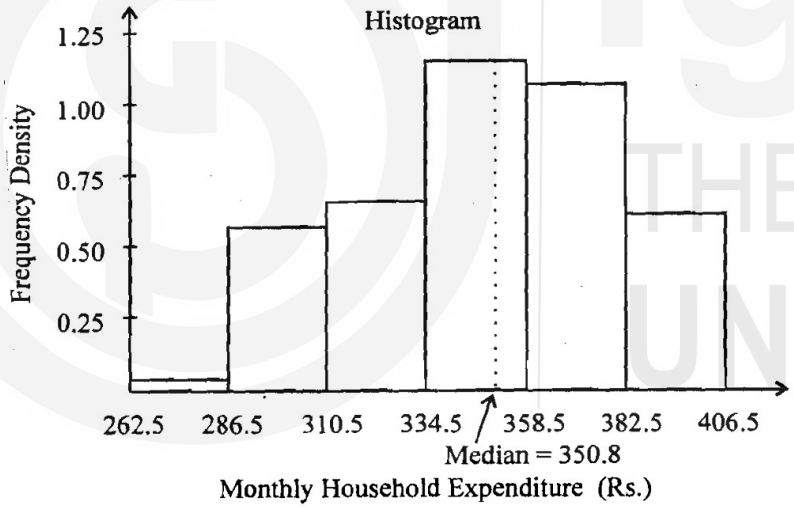

Fig. 3.1

Area up to the class boundary 334.5 is 131 and up to 358.5 is 59. Hence the median lies in the class 334.5 – 358.5. We now want to find a point in this classes so that the area from 334.5 to the point is (50 – 31) = 19, where are up to 334.5 is 31. Since the rectangle over the interval 334.5 – 358.5 has an area of 28, and is of length 24, to get an area of 19 we need $\frac{19}{28}$th part of 24. This works out to be $\frac{19}{28} \times 24 = 16.3$. Thus the median is 334.5 + 16.3 = 350.8. Note also that the area in the class 350.8 to 358.5 is 28 – 19 = 9 and to the right of 350.8 is 9 + 41 = 50, as it should be.

Based on the above procedure, we can write a formula for the computation of median.

$$M_d = l_m + \frac{\frac{N}{2} - C}{f_m} \times h, \quad \text{where}$$

l_m is the lower limit of the median class, i.e., the class in which median lies,

N is the total frequency,

C is the cumulative frequency of classes preceding the median class (not that C = 31 in the above example).

f_m is the frequency of median class, and

h is the width of median class.

3.2.3 Mode

As has been pointed out earlier, often observations tend to cluster around a central value. A simple measure of this phenomenon is called mode.

Mode or modal value of a discrete variable is defined as that value of the variable for which frequency is the maximum. Mode, however, is not the majority, i.e., it does not imply that most (50% or more) of the observations have the modal value.

From Table 3.1 we find that the mode or modal value of household size is 4 as this value occurs with largest frequency of 33 among 100 households.

There are, however, data sets when mode cannot be defined uniquely, i.e., the distribution has multiple mode. Raw data with 7 hypothetical observations with values 4, 3, 4, 1, 2, 5, 3 have two modes, 3 and 4. Distributions having two modes are called *bimodal distributions*, though the frequently encountered distributions have only one mode or are *unimodal*.

For observations on the continuous variable, like monthly household expenditure on food, no two observations are likely to have same value and so mode is not a meaningful measure of such raw data. However, central tendency comes out clearly when these raw data are grouped into various class intervals. For grouped data *modal class* is defined as the class having largest frequency. Since large class intervals are likely to include large number of observations and smaller class intervals are likely to have few observations, definition of modal class is meaningful only when class intervals have equal length.

For discrete data it is easier to find out the mode. But in the case of continuous data computation of the mode is done by the following formula:

$$M_0 = l_m + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times h, \quad \text{where}$$

l_m is the lower limit of the modal class, i.e., the class in which mode lies,

$\Delta_1 (= f_m - f_{m-1})$ is the difference of the frequencies of the modal class and its preceding class.

$\Delta_2 (= f_m - f_{m+1})$ is the difference of the frequencies of the modal class and its following class, and

h is the width of the modal class.

Let us look back to Table 3.2. Here modal class is 334.5 – 358.5 as it has the highest frequency, 28.

Thus, $l_m = 334.5,\ \Delta_1 = 28 - 16 = 12, \Delta_2 = 28 - 26 = 2$ and $h = 24$.

Hence $M_0 = 334.5 + \frac{12}{12+2} \times 24 = 355.07$

Mode is a useful measure of central tendency when a frequency distribution has a strong peak and it is particularly useless when a frequency distribution is almost flat.

Check Your Progress 1

1) The frequency distribution of a family size for 250 families in a ward of an industrial town is given below:

   Find the mean, median and mode.

Family Size	**Frequency**
1	4
2	22
3	25
4	45
5	52
6	41
7	36
8	15
9	7
10	3
Total	**250**

…….....

…….....

…….....

…….....

2) Compute the mean, median and mode for the following frequency distribution.

I.Q.	Frequency
160 – 169	2
150 – 159	3
140 – 149	7
130 – 139	19
120 – 129	37
110 – 119	79
100 – 109	69
90 – 99	65
80 – 89	17
70 – 79	5
60 – 69	3
50 – 59	2
40 – 49	1
Total	**309**

……

……

……

……

……

……

3.3 OTHER MEASURES OF CENTRAL TENDENCY

Besides the arithmetic mean, median and mode there are other averages which are relatively unimportant but may be appropriate in particular situations. These are Geometric Mean and Harmonic Mean.

Often we see that all the observations do not have equal importance. In such cases we need to give differential importance to different items. Here we use weighted means – arithmetic, geometric or harmonic – instead of simple means. This we will discuss in Section 3.3.2.

3.3.1 Geometric Mean and Harmonic Mean

Often we have to deal with that are time dependent, i.e., time series data which are unlike one-time data of Tables 3.1 and 3.2. For time dependent data, it is often of interest to find the pattern of change over time. Consider the following two data sets.

Set I: 1000 1100 1200 1300 1400 1500 1600

Set II: 1100 1210 1331 1464 1611 1772 1949

The first set looks like the basic salary (in Rs.) of an employee for 7 years with annual increment of Rs. 100 per year.

The second set looks more like his gross salary (in Rs.). Annual increase in the two sets is given below.

Set I: 100 100 100 100 100 100

Set II: 110 121 133 147 161 177

Arithmetic mean of the annual increase is 100 for Set I and 141.5 for Set II. On the basis of these average annual increases, if we find out the figures for the two sets, starting from the initial values, we would get the following:

Set I: 1000 1100 1200 1300 1400 1500 1600

Set II: 1100 1241.5 1383 1524.5 1666 1807.5 1949

We find that arithmetic mean has worked well for Set I. However, it has not worked well for Set II. It is because the progression of original numbers in the two sets is different. In set I, increment has been a fixed quantum whereas in Set II, figures have increased at a fixed rate. Fixed quantum of increase is called *arithmetic progression* and arithmetic mean is appropriate to describe the increase. Fixed rate of increase is called *geometric progression* and geometric mean is most appropriate to describe the increase.

For *n* numbers $X_1, X_2, ...X_n$ the geometric mean (GM) is defined as the nth root of the product of these n numbers, i.e.,

$$GM = (X_1, X_2, ..., X_n)^{\frac{1}{n}} = \left[\prod_{i=1}^{n} X_i\right]^{\frac{1}{n}}$$

Clearly, GM is not defined unless all the n numbers are positive. If any number is negative or zero, we cannot calculate GM. By taking logarithm of GM, we have

$$\log GM = \left(\frac{1}{n}\right)\left(\log X_1 + \log X_2 + ... + \log X_n\right) = \frac{1}{n}\sum_{i=1}^{n} \log X_i$$

which shows that now GM can be computed by using a log-table. Anti-logarithm of the arithmetic mean of log X values is GM. For the second data set, gross salary increased at the rate of 11% every year. In practice, however,

increase/decrease will not be at a fixed rate over the years; and it is meaningful to talk about average rate because fixed rate situation is rare.

In general, GM is more appropriate average for percentage (or proportionate) rates of change than arithmetic mean as in the case of rise in various price indices, cost of living indices, etc.

Finally, we discuss about another measure of location called the 'harmonic mean' (HM). This measure of central tendency comes naturally in many situations as in the following illustration. A stockist stocks Rs. 5000 worth of an item at the beginning of every month. Unit rate (in Rs.) of the item for five successive months had been 10.75, 11.80, 14.00, 11.45. and 12.00. The stockist wants to find average rate per unit of the item he has stocked for five months. Computation is presented below:

Month	**Amount Spent (Rs.)**	**Unit Rate (Rs.)**
1	5000	10.75
2	5000	11.80
3	5000	14.00
4	5000	11.45
5	5000	12.00
Total	**25000**	

$$\text{Average price (in Rs. of his entire stock} = \frac{\text{Total Money Spent}}{\text{Total Quantity Purchased}}$$

$$= \frac{5 \times 5000}{\frac{5000}{10.75} + \frac{5000}{11.80} + \frac{5000}{14.00} + \frac{5000}{11.45} + \frac{5000}{12.00}}$$

$$= \frac{5}{\frac{1}{10.75} + \frac{1}{11.80} + \frac{1}{14.00} + \frac{1}{11.45} + \frac{1}{12.00}}$$

$$= \frac{1}{\frac{1}{5}\left(\frac{1}{10.75} + \frac{1}{11.80} + \frac{1}{14.00} + \frac{1}{11.45} + \frac{1}{12.00}\right)} = 11.91$$

The last expression is 'the reciprocal of the arithmetic mean of the reciprocals' and is called harmonic mean (HM). For a set of n values $X_1, X_2, ..., X_n$, the HM is defined as

$$HM = \frac{n}{\frac{1}{X_1} + \frac{1}{X_2} + ... + \frac{1}{X_n}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{X_i}}$$

You should note that HM is not defined when any observation is zero.

If the stockiest, instead of stocking Rs. 5000 worth of items, stocks 3000 items at the beginning of every month at the given prices, the appropriate average would be arithmetic mean. To verify this, we can write

$$\text{Average Price} = \frac{\text{Total Money Spent}}{\text{Total Quantity Purchased}}$$

$$= \frac{3000 \times 10.75 + 3000 \times 11.80 + 3000 \times 14.00 + 3000 \times 11.45 + 3000 \times 12.00}{3000 \times 5}$$

$$= \frac{10.75 + 11.870 + 14.00 + 11.45 + 12.00}{5} = \text{AM of the given prices.}$$

3.3.2 Weighted Means

For many practical applications weighted means (arithmetic, geometric or harmonic) reflect phenomenon more clearly than unweighted or simple means that have been computed so far. For computation of, say, consumer price index, not all commodities are equally important. Increase in fuel cost may affect consumer price index more than an increase in agricultural prices. For stock market, stock of some key companies may be a trend setter. Weighted means are more appropriate in such situations. To find weighted mean, a weight w_i is attached to each X_i and the means are computed as if wi's are, symbolically, frequencies of the corresponding X_i's. The computational formulae are as given below:

$$\text{Weighted AM} = \frac{\sum_{i=1}^{n} w_i X_i}{\sum_{i=1}^{n} w_i}$$

$$\text{Weighted GM} = \left(\prod_{i=1}^{n} X_i^{w_i} \right)^{\frac{1}{\sum w_i}} \quad \text{and}$$

$$\text{Weighted HM} = \frac{\sum_{i=1}^{n} w_i}{\sum_{i=1}^{n} \frac{w_i}{X_i}}$$

Weighted mean is equal to unweighted mean when each w_i is the same or equal to unity.

3.3.3 Pooled Mean

Often we come across situations when the means have been computed for different sources or samples. In such situations we become interested to find an overall mean if it is meaningful. This is done by computing what is called a *pooled mean*. The procedure of computing a pooled mean is given below.

Let $m_1, m_2, ..., m_r$ be *r* arithmetic (or geometric or harmonic) means, computed on the basis of $n_1, n_2, ..., n_r$ observations respectively. Then

Pooled arithmetic mean $= \frac{1}{n}\sum_{i=1}^{r} m_i n_i$, where $n = \sum_{i=1}^{r} n_i$

Pooled geometric mean $= \left(\prod_{i=1}^{r} m_i^{n_i}\right)^{\frac{1}{n}}$ and

Pooled harmonic mean $= \frac{n}{\sum_{i=1}^{r} \frac{n_i}{m_i}}$

where $n = n_1 + n_2 + ... + n_r$

Note that the above expressions are similar to the expressions for weighted means.

3.3.4 Choosing a Measure of Central Tendency

It has already been discussed when a particular mean, AM or GM or HM, is more appropriate than the other two. However, when we have grouped data in which either of the end classes are open ended, i.e., of the type 'up to c_1' and / or c_{k-1} and above', mid-points of such classes cannot be computed. Consequently, no mean can be computed. There is, however, no problem in computing median or mode in such cases. On the other hand, a pooled median or mode cannot be computed, like the case for mean, unless all the sets of data are made available in their entirety. These problems are related to computational difficulties and not to appropriateness of a measure.

Since graphical representation of data is more appealing, median or mode are more useful in such a situation because their crude values can be obtained easily without having to go through any computations. Also, median and mode are simple concepts for communication and comparison between graphs. It has, however, been observed that median is less stable than arithmetic mean in repeated sampling and we need to be careful when comparing graphs.

For data that has a distribution close in shape to what is called the normal distribution, with one peak and going down symmetrically on either side, we may use of mean, median or mode. It is because, for a normal distribution, these measures have the same value.

You should note that choosing an appropriate measure of central tendency is not an end to data analysis, and much still remains. For example, by saying that household average monthly expenditure on food is Rs. 348.66, it does not say whether a large number of households have very low monthly average expenditure on food or a few households have a very good menu. Next set of analysis aims at answering such questions.

3.4 PERCENTILES

Concept of percentiles will be explained by using mainly Table 3.2 data on average monthly household expenditure. Percentiles are used in two directions, depending on the question to be answered. Direction of a question may be, what per cent of households have monthly average food expenditure upto Rs. 350.80? Or it may be, what is the maximum monthly average food expenditure of the lower 50% of the households? Note, from our earlier computation of median of Table 3.2 distribution, that the answer to one question is the figure in other, i.e., 50% of the households have Rs. 350.80 as maximum average monthly food expenditure. Depending on interest, percentage below a cut-off point may be called for: when a poverty line is decided, it is of interest to know the percentage below the poverty line. In the other direction, it may also be of interest to find the status of lower 10% or upper 5% of the population. These are answered by using what are called percentiles.

3.4.1 Percentile: Definition and Computation

For any given percentage v, the v^{th} percentile is P_v, a value of the variable being studied, so that at leasat v percent of the observations are less than or equal to P_v and at least (100 – v) percent of the observations are greater than or equal to P.

For example, for Table 3.1, distribution of household size, $P_v = 5$ for any from 78 to 79.

For grouped data, percentiles are more clearly understood when we look at the cumulative distribution function. Let F(X) be the proportion of observations less than or equal to X. Any given value X_0 is then the 100 $F(X_0)$th percentile. For Table 3.2, class boundaries, we have F(286.5) = 0.01, F(310.5) = 0.15, F(334.5) = 0.31, F(358.5) = 0.59 and F(382.5) = 0.85, and consequently Rs. 286.5 = P_{10}, Rs. 310.5 = P_{15}, Rs 334.5 = P_{31}, Rs. 358.5 = P_{59}, and Rs. 382.5 = P_{85}.

You should note that any amount less than Rs. 262.5 (lower boundary of first class interval) is zero-th percentile and any amount more than Rs. 406.5 (upper boundary of last class interval) is 100^{th} percentile.

3.4.2 Quartiles and Deciles

Depending on its use, some specific percentiles go by different names. Every 25^{th} percentile is called a quartile, and every 10^{th} percentile is called a decile. For example,

25^{th} percentile $= P_{25} = Q_1 =$ first quartile

50^{th} percentile $= P_{50} = Q_2 =$ second quartile

75^{th} percentile $= P_{75} = Q_3 =$ third quartile

10^{th} percentile $= P_{10} = d_1 =$ first decile

20^{th} percentile $= P_{20} = d_{21} =$ second decile, etc., and

$$P_{50} = Q_2 = d_5 = \text{median}$$

The formulae for Q_1 and Q_2 are similar to the formula for the median. These can be directly written as given below.

$$Q_1 = l_{Q_1} + \frac{\frac{N}{4} - C}{f_{Q_1}} \times h, \text{ and}$$

$$Q_{31} = l_{Q_3} + \frac{\frac{3N}{4} - C}{f_{Q_3}} \times h,$$

where *C* denotes the cumulative frequency of classes preceding the first (or third quartile class and *h* is the corresponding class width.

Using similar notations, it is possible to write the formula for any partition value. For example, the formula for 40^{th} percentile can be written as

$$P_{40} = l_{P_{40}} + \frac{\frac{40N}{100} - C}{f_{P_{40}}}$$

Percentiles also go by the name of fractiles when proportions, instead of percentage, are used. For example, P_{30} is 0.3 fractile.

Just as we do not get a complete picture of a distribution by looking at a measure of location, too many percentiles may be needed to describe the spread or dispersion of a distribution. It is felt that there should be some simple measures of dispersion. This is the topic of discussion of the next Section.

Check Your Progress 2

1) Given below are the prices is ratios for five commodities with the corresponding weights. Calculate the Weighted Arithmetic Mean and Geometric Mean.

Commodity	Price Ratio	weight
1	2.20	30
2	1.85	25
3	1.80	22
4	2.05	13
5	1.75	10

……

……

……

……

……

……

2) The earnings of five nationalised banks, in crores of rupees, is given below.

217.40 330.50 682.55 1263.59 2249.63

Find the Geometric Mean of the earnings.

……

……

……

……

……

……

3) The distribution of age of males at the time of marriage was as follows:

Age (years)	No. of Males
18 – 20	5
20 – 22	18
22 – 24	28
24 – 26	37
26 – 28	24
28 – 30	22

Find at the time of marriage (i) the average age, (ii) modal age, (iii) the median age, (iv) third quartile, (v) sixth decile, (vi) nineteenth percentile.

……

……

……

……

4) In a factory, a mechanic takes 15 days to fabricate a machine, the second mechanic takes 18 days, the third mechanic takes 30 days and the fourth mechanic takes 90 days. Find the average number of days taken the workers to fabricate the machine. Which average would you use, and why?

……

……

……

……

……

5) The amount of interest paid on each of the three different sums of money yielding 10%, 12% and 15% simple interest per annum are equal. What is the average yield percent on the total sum invested?

……

……

……

……

……

3.5 MEASURES OF DISPERSION

So far we have discussed various measures of central tendency, viz., arithmetic mean, median, mode geometric mean and harmonic mean. However, in many situations these measures do not represent the distribution of data. For example, look into the following three sets of data:

Set A: 2, 5, 17, 17, 44.

Set B: 17, 17, 17, 17, 17.

Set C: 13, 14, 17, 17, 24.

In all the sets the numerical value of the mean, median and mode are the same, that is, 17. Still all three sets are so different! While in Set B all the observations are equal, in Set A they are so dispersed. Definitely we need another measure which will account for such dispersion of data.

The word dispersion is used to denote the degree of heterogeneity in the data. It is an important characteristic indicating the extent to which observations vary amongst themselves. The dispersion of a given set of observations will be zero

when all of them are equal (as in Set B given above). The wider the discrepancy from one observation to another, the larger would be the dispersion. (Thus dispersion in Set A should be larger than that is Set C). A measure of dispersion is designed to state numerically the extent to which individual observations vary on the average. There are quite a few measures of dispersion. We discuss them below.

3.5.1 Range

Of all measures dispersions, range is the smallest. It is defined as *the difference between the largest and the smallest observations*. Thus for the data given at Set A the range is 44 – 2 = 42. Similarly, for Set B the range is 17 – 17 = 0 and for Set C it is 11. Now let us look into some grouped data. For Table 3.2 data (look back to the previous Unit), the range is Rs. 406.5 – Rs. 262.5 = Rs. 144. Notice that, for grouped data, largest and the smallest observations are not identifiable. Hence we take *the difference between two extreme boundaries of the classes*.

It is intuitive that, because of central tendency, if one selects a small sample, observations are more likely to be around its mode than away from it. Less likely or extreme values will be included in the sample when its size is large. This, in other words, implies that range will increase with increase in sample size. Also, it is known that in repeated sampling with same sample size, range varies considerably making it a less suitable measure for comparisons. However, range is a measure which is easy to understand and can be computed quickly.

3.5.2 Inter-quartile Range

Range as a measure of dispersion does not reflect a frequency distribution well, as it depends on the two extreme values. Even one very large or small observation, away from general pattern of other observations in the data set, makes the range very large. For example, in Set A, the range is found to be excessively large (44 – 2 = 42) because of the present of very large one observation, that is 44. To avoid such extreme observations, particularly when there is a strong central tendency, inter-quartile range is useful as a measure of dispersion. It is defined as

Inter-quartile Range $= Q_3 - Q_1 = P_{75} - P_{25}$.

Inter-quartile range is the range of the middle most 50% of the observations. If the observations are compact around median, i.e., a strong mode close to the median exists inter-quartile range will be smaller than half of the range. If the data are flat, having no central tendency, this measure will be large, and its value will be close to half of the range.

Let us look into the discrete data given in Table 3.1 of the previous Unit. Here, $P_{75} = 4$ and $P_{25} = 3$. Hence, the inter-quartile range of household size is 4 – 3 = 1. This shows that a strong central tendency exists in the distribution of household size range was observed to be 7 (since 8 – 1 = 7).

For Table 3.2 data, P_{25} of the average monthly expenditure on food was seen to be Rs. 325.50; P_{75} computed similarly works out to be Rs. 377.88 and inter-quartile range is Rs 377.88 – Rs. 325.50 = Rs. 52.38. Compared to Rs. 52.38, the range was observed to be Rs. 146.00 or 2.79 times larger. This shows not so strong central tendency for average monthly household expenditure on food.

3.5.3 Mean Deviation

While the range depends on the two extreme observations, inter-quartile range depends on the two extreme observations among the middle most 50 percent of the observations. Thus, one talks only about the percentage of observations between minimum, P_{25} and maximum, P_{75}. Thus both range and inter-quartile range do not depend upon all the observations in the sample. Hence, while computing range or inter-quartile range we do not say anything about the distribution of observations within the group.

Among many possibilities to quantify spread or dispersion of observations, one possibility is to used the deviation of observations from some central value.

Since mean is the most commonly used measure of central tendency, it is often taken as the central value with reference to which the deviations are computed. These deviations are then suitably combined to a get a measure of dispersion.

Mean deviation treats every single observation with equal weight, in the form of arithmetic mean of deviations based on each observation.

For observations $X_1, X_2, ..., X_n$, if we take deviations as simple difference, then for the i^{th} observations the deviation is $(X_i - \overline{X})$ where $\overline{X}$ is the mean. Mean of these deviations is

$$\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right) = \frac{1}{n}\sum_{i=1}^{n}1 = \overline{X} - \overline{X} = 0.$$

Since simple difference do not lead to any measure, absolute differences are used to define mean deviation.

$$\text{Mean Deviation} = \frac{1}{n}\sum_{i=1}^{n}\left|X_i - \overline{X}\right|, \text{ where}$$

the two vertical bars indicate that the sign of the difference with in two bars is to be taken as positive. For example, $|2 - 4| = 4$ (and not –2).

For frequency data, discrete or continuous type, the formula becomes

$$\text{Mean Deviation} = \frac{1}{N}\sum_{i=1}^{n} f_i\left|X_i - \overline{X}\right|,$$

where $N = \sum_{i=1}^{n} f_i$ and $X'_i s$ are distinct observations and f_i is the frequency of X_i in the discrete case and X_i is the mid-point of i^{th} class and f_i is its frequency for the continuous case. The need for such a measure is illustrated below.

Following summary values have been computed for two data sets.

	Data Set I	Data Set II
Number of observations	7	7
	7	7
P_{25}	12	12
Median $= P_{75}$	17	17
P_{75}	20	20
Range	10	10
inter-quartile range	12	12
Mean		

Thus, based on the above measures only, and not looking at the data sets I and II, it would appear that two persons separately may have worked out on the same data set. However, the two data sets may have been as given below.

Data Set I: 3 7 8 12 14 17 23

Data Set II: 2 7 11 12 13 17 22

One may construct much more different looking data sets having identical values for the above type of measures. This comparison indicates that more measures are needed and mean deviation is one such. This is not to imply that the above measures and mean deviation together completely describe a data set.

For data set I

Mean deviation =

$$\frac{1}{7}\left(|3-12|+|7-12|+|8-12|+|12-12|+|14-12|+|17-12|+|23-12|\right)$$

$$=\frac{9+5+4+0+2+5+11}{7}=\frac{36}{7}=51.4$$

For data set II

Mean deviation =

$$\frac{1}{7}\left(|2-12|+|7-12|+|11-12|+|12-12|+|13-12|+|17-12|+|22-12|\right)$$

$$=\frac{10+5+1+0+1+5+10}{7}=\frac{32}{7}4.57.$$

Thus, observations in data set I are more dispersed from mean than that of data set II.

Let us now compute mean deviation of household size and household average monthly food expenditure.

For household size data of Table 3.1, mean $= \overline{X} = 3.74.$ Mean deviation is now computed as

$$\text{Mean deviation} = \frac{1}{N}\sum_{I=1}^{N} f_i \left|X_i - \overline{X}\right|$$

$$= \frac{1}{100}(3|1-3.74| + 16|2-3.74| + ... + 2|8-3.74) = \frac{109.12}{100} = 1.0912.$$ For

Table 3.2 distribution on average household expenditure on food, mean $\overline{X} = \text{Rs.}348.66.$

The mean deviation=

$$= \frac{1}{100}\left(2|274.5-348.66| + ... + 15|394.5-348.66|\right) = \frac{2510.88}{100} = 25.11$$

So far we have considered mean deviation from mean. The mean deviation from median or from mode can also be defined in a similar way.

3.4.4 Variance and Standard Deviation

The most frequently used measures of dispersion are variance and standard deviation. Variance is so commonly used that it is also called dispersion.

Variance is a measure which suitably combines individual deviations from the mean, treating each observation with equal weight as in mean deviation. For variance, however, measure of individual deviation is taken as the *squared difference from the mean*. Since it is more manageable to use the squared difference rather than absolute difference, particularly while doing format mathematics, use of variance has become more popular. Conventionally variance for a population is denoted by σ^2 (pronounced *sigma-squared*) and variance for a sample is denoted by σ^2. Variance is defined as the mean of the squared deviations of observations form their mean. Variance from raw data is computed by

$$\text{Variance} = \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

For frequency data, discrete or continuous type, the formula becomes

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{n} f_i\left(X_i - \overline{X}\right)^2, \text{ where } N = \sum_{I=1}^{N} f_i$$

In the same scale of measurement, for example, observations with a variance of 2 are less dispersed then observations with variance more than 2. To talk about a distribution in terms of a measure of central tendency and a measure dispersion, it is a practical need to use both measures in the same unit. Mean and mean deviation are in the same unit. Since each deviation has been squared for Based on variance, an equally or more popular measure of dispersion in the same unit as that of observations is *standard deviation* , abbreviated as s.d. Standard deviation

as the positive *square root of variance*, i.e., s.d. = σ. As it is the positive square root of variance, it cannot be negative.

Let us compute the s.d. for household size data of Table 3.1

$$\sigma^2 = \frac{1}{100}\left[3(1-3.74)^2 + 16(2-3.74)^2 + ... + 2(8-3.74)^2\right] = \frac{199.24}{100} = 1.9924 \text{ and}$$

$\sigma = 1.4115$.

Similarly for Table 3.2 distribution of average monthly household expenditure on food, variance in Rs. Square is given by

$$\sigma^2 = \frac{1}{100}\left[2(274.50-348.66)^2 + ... + 15(394.5-348.66)^2\right] = \frac{95725.437}{100} = 957.25,$$

and s.d. is

$\sigma = \text{Rs.}3094.$

For computational convenience, the formula for variance is written in alternative form as

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} f_i(X_i - \overline{X})^2 = \frac{1}{n}\sum_{i=1}^{N} X_i^2 - \overline{X}^2$$

or

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} f_i(X_i - \overline{X})^2 = \frac{1}{n}\sum_{i=1}^{N} f_i X_i^2 - \overline{X}^2$$

as the case may be. Thus, variance is viewed as

variance = Mean of Squares – Square of the Mean

Using the above formulae, you may compute the variance for the data given in Tables 3.1 and 3.2 and verify the earlier results.

The computation of variance may be greatly simplified by changing X_i to $u_i = \frac{X_i - A}{h}$, as was done in the computation of mean.

Note that, since

$$u_i - \overline{u} = \frac{X_i - A}{h} - \frac{\overline{X} - A}{h} = \frac{X_i - \overline{X}}{h}, \text{ we can write}$$

$$X_i - \overline{X} = h(u_i - \overline{u})$$

Hence,

$$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left\{h\left(u_i - \overline{u}\right)\right\}^2 = h^2\sigma_u^2$$

where σ_x^2 is the variance of X_i and σ_u^2 is the variance of the *u* values.

Since the magnitude of *u* values is smaller, it is easier to compute variance of *u* values. Then the variance of X values can be easily computed by using the above formula.

Let us compute the variance by applying the above method for the data given in Table 3.2.

If we write $u_i = \frac{X_i - 346.5}{24}$, the u values are

–3, –2, –1, 0, 1, 2 and the respective frequencies are 1, 14, 16, 28, 26, 15.

The mean of u values = $\frac{-3\times1-2\times14-1\times16+0\times28+1\times26+2\times15}{100} = 0.09$

The mean of squares of u values =

$$\frac{9\times1+4\times14+1\times16+0\times28+1\times26+4\times15}{100} = 1.67$$

Thus $\sigma_u^2 = 1.67 - (0.09)^2 = 1.6619$ and

$$\sigma_X^2 = (24)^2.(1.6619) = 957.25.$$

Even though change from *X* to *u* is for computational ease, it brings up an important issue. Notice that $\sigma_2^2 = 1.6619$ but $\sigma_X^2 = 957.25,$ where *u* was obtained from X by a simple linear transformation, i.e., by change of origin and scale of X values. Typical such natural case are pounds and kilograms for weight, gallons and litres for liquid volume, etc. Since 1 kg. = 2.2046 lbs., s.d. of 5 kg. when measured in kilograms is same as 11.023 lbs. when measured in pounds; or since 1 litre = 0.22 gallon, s.d. of 5 litres when measured in litres is same as s.d of 1.1 gallons when measured in gallons. Thus, whereas variance and standard deviation are supposed to measure spread of observations, not much can be made out of these measures due to their dependence on the unit of measurement.

In this context, the single most useful result about the spread of observations based on mean and standard deviation, irrespective of unit of measurement, is due to Chebychev.

Check Your Progress 3

1) What is dispersion? What are the common measures of dispersion?

……

……

……

……

……

……

2) In a batch of 10 children the marks obtained by a dull boy are 25 marks below the average marks of other children. Show that the standard deviation of marks for all the children is at least 7.5. If this standard deviation is actually 12.0, find the standard deviation when the dull boy is left out.

……

……

……

……

……

……

3) The following data shows the daily profits (in Rs.) made by a shopkeeper on 15 successive days.

116, 87, 91, 81, 98, 102, 97, 100, 105, 101, 115, 98, 102, 98, 93

Determine the range, the mean deviation about mean and the standard deviation for the data.

……

……

……

……

……

……

4) Compute the arithmetic mean, standard deviation and the mean devaiton of the following data.

Scores	4 – 5	6 – 7	8 – 9	10 – 11	12 – 13	14 – 15	Total
f	4	10	20	15	8	3	60

……

……

……

……

……

5) The mean and the s.d. of a sample of 100 observations were calculated as 40 and 5.1 respectively by a student who by mistake took one observation as 50 instead of 40. Calculated the correct s.d.

……

……

……

……

3.6 RELATIONSHIP BETWEEN DISPERSION AND STANDARD DEVIATION

You have earlier learnt that when all the values in a set of data are located near their mean, then exhibit a small amount of dispersion or variation and those set of data in which some values are located far form their mean have a large amount of dispersion. A useful rule that illustrates the relationship between dispersion and standard deviation is given by Chebychev's theorem.

3.6.1 Chebychev's Theorem

For any set of observation and positive constant k (>1), the proportion of observations lying within k standard deviations of the mean is certain to be at least $1-\frac{1}{k^2}$.

Note that the theorem is not useful for any positive k less than or equal to 1, since $1-\frac{1}{k^2}$ is at the most equal to zero. For other values of k, the minimum proportion can be computed easily. For example, proportion of observations within 1.5 s.d. of the mean is certain to be at least $1-\frac{1}{1.5^2}=0.556$ or 55.6%. The following figure indicates spread of data based on Chebychev's theorem. For the household size data of Table 3.1, $\overline{X}=3.74$ and s = 1.4115. If we take k=2, we can say that at least $\left[\left(1-\frac{1}{2^2}\right)\times 100\right]=75\%$ of the households are certain to have their size between $3.74\pm 2\times 1.4115$, i.e., between 0.917 and 6.563.

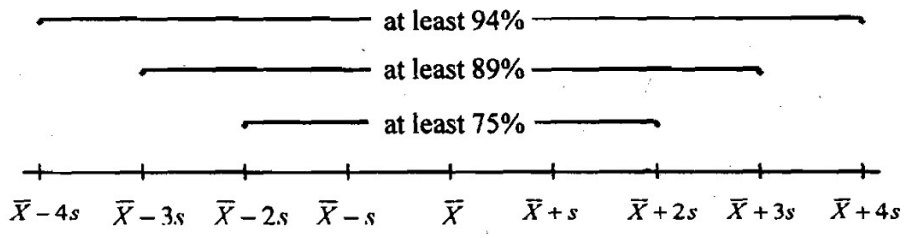

Fig. 3.2

For the Table 3.2 distribution of average monthly household expenditure on food $\overline{X} = \text{Rs.}348.66$ and s = Rs. 30.94, at least 55.6% (for k = 1.5) of households are certain to have monthly average food expenditure between Rs. 302.25 and Rs. 395.07. You can find the relevance of this theorem when we study normal distribution later in Unit 11.

3.6.2 Shape of Distribution

For methodological studies in many situations, a distribution is adequately described by measures of central tendency and dispersion. Yet other measures are also in use to describe distribution in practical situations, particularly for economic variables such as income, consumption, economic assets, etc., which are non-negative. Two such measures are *coefficient of variation and concentration ratio*. These measures will be viewed here essentially as measures of inequality in the distribution of economic variables.

3.6.3 Coefficient of Variation

Let us propose to economic status of households in two villages. The summary figures of monthly calories intake of households are given below for the two villages.

	Villages	
	A	**B**
Number of Households (n)	817	561
Mean calorie intake $(\overline{X})$	2417	2235
s.d. of calorie intake (σ)	418	232

The problem is to identify the village that has more inequality as far as calorie intake is concerned. Village A has higher mean calorie intake but has larger s.d. and larger number of households compared to village B. Village A may actually have more number of poorer households in than in village B. Therefore, in village A, inequality between households may be more than that in village B. One index which measures the quantum of such disparity is called the coefficient of variation, abbreviated as c.v. It is defined as percentage standard deviation per unit of mean, i.e.,

$$\text{c.v} = \frac{\sigma}{\overline{X}} \times 100$$

Since σ and $\overline{X}$ have the same unit of measurement, c.v. is unit free and is not affected by the choice of unit of measurement.

For village A, $\text{c.v.} = \frac{418}{2417} \times 100 = 17.29$ and for village B,

$$\text{c.v.} = \frac{232}{2235} \times 100 = 10.38.$$

Since the coefficient of variation in village A is greater than the coefficient of variation in village B, the inequalities are given in village A compared to village B.

To compare the extent of inequalities, we compute

$\frac{17.29-10.38}{10.38}\times 100 = 66.57$ which implies that compared to village B, 66.57% more inequality exists in village A.

3.6.4 Concentration Ratio

Above was a comparison of inequality between two villages, without quantifying the level of inequality within each village. If a distribution has a long right tail, it shows that a few have a large share. In other words, a majority of population has a very small share. Let us consider the distribution of income of a hypothetical economy.

Suppose there are three classes of people in the economy – the upper class, the middle class and the lower class. Let 10%, 30% and 60% be the share of population in these three classes respectively. Suppose the lower class receives only 20% of the national income, the middle class 30% and the upper class the rest, i.e., the remaining 50%. We can now present the data in a percentage cumulative frequency distribution form. Thus, the lowest 60% of the population receives only 20% of the income, the lowest 90% receive 50% (= 20 + 30) of the income and obviously, 100% of the population receive 100% of the income. If we take a graph paper where the percent cumulative total income is plotted on the vertical axis and we plot the point (0, 0), (60, 20), (90, 50) and (100, 100), then the curve joining these points is what we call the ***curve of concentration*** or ***Lorenz curve***. The straight line joining the points (0,0) and (100, 100) give the line of *equal distribution* or the *equitable line*. The equitable line is that one which shows that the proportion of share is exactly the same as the proportion of population who are supposed to share. The area between the line of equal distribution and the curve of concentration, called the *area of concentration* is an indicator of the degree of concentration; the larger the area the greater is the concentration.

Coefficient of Inequality

Let us take coordinates of the above points in per unit terms instead of percentage terms. Thus, the coordinates of the points, in the above example, can be written as (0,0), (0.60), (0.20), (0.90, 0.50) and (1.00, 1.00). The coefficient of inequality of income distribution is then defined as the ratio of the area of concentration to total area of the triangle. Since the area of the triangle is 0.5 (since $\frac{1}{2}\times 1\times 1 = 0.5),$ the coefficient of inequality is equal to twice the area of concentration when coordinates of various points are taken per unit rather than in percentage.

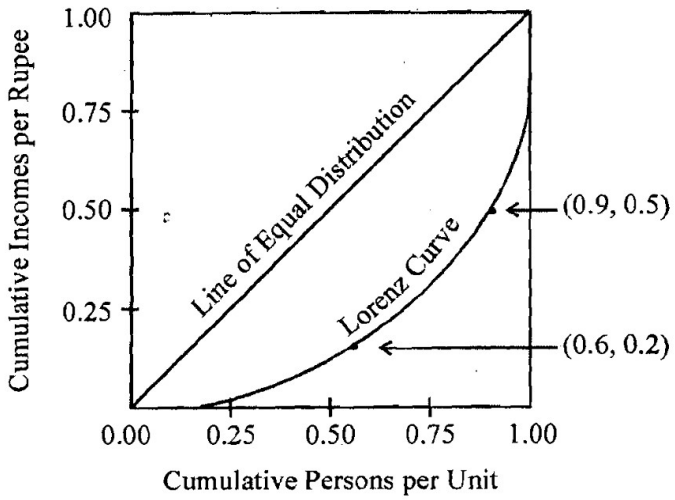

Fig. 3.3: Lorenz Curve

Check Your Progress 4

1) The following figures give the crude birth rate per 1000 people in Switzerland from 1968 to 1980.

   Crude birth rate (X): 17.1, 16.5, 15.8, 15.2, 14.3, 13.6, 12.9, 12.3, 11.7, 11.5, 11.3, 11.3, 11.6.

   Calculate the Variance, Standard Deviation and Coefficient of Variation.

   ……

   ……

   ……

   ……

   ……

2) The following table gives the distribution of age of lady teachers of a school as revealed by records.

Age Group (years)	No. of lady teachers
15 – 19	3
20 – 24	13
25 – 29	21
30 – 34	15
35 – 39	5
40 – 44	4
45 – 49	2

Calculate coefficient of variation, and (ii) number of teachers between the age 26 and 33 years.

……

……

……

……

……

3.7 LET US SUM UP

In this unit you have learned to compute various measures of central tendency. These measures of central tendency can be divided into two broad categories, namely mathematical averages and positional averages. Positional averages are mode, median, quartile, percentiles, etc., while arithmetic mean, geometric mean and harmonic mean are mathematical averages. Geometric Mean is most suitable for averaging ratio and proportional rates of growth while Arithmetic mean or Harmonic mean can be used to find average rates like price, speed, etc. depending upon the nature of the given condition.

You also learned about the measures of dispersion. The most important measures of dispersion you learned about in the unit are the variance, standard deviation and the concentration ratio. You have also learned to compute variance, standard deviation and coefficient of variation using both ungrouped and grouped data. The coefficient of variation is used to compare the dispersion of two distributions having either different means (even when their variables are measured in same units) or different units of measurement of their variables.

3.8 ANSWERS OR HINTS TO CHECK YOUR PROGRESS EXERCISES

Check Your Progress 1

1) 5.1, 5, 5

2) 108.48 ; 108.41 ; 111.42

Check Your Progress 2

1) Rs. 1.96 ; Rs. 1.95

2) Rs. 674.31 crore

3) i) 25.83 years (ii) 24.82 years (iii) 24.86 years (iv) 27.30 years (v) 25.59 years (vi) 28.79 years.

4) Arithmetic Mean, 38.25 days

5) Harmonic Mean, 12%.

Check Your Progress 3

1) Do it yourself.
2) 9.9
3) 35, 6.46, 8.85
4) 9.23, 2.49, 2.03
5) 5.0

Check Your Progress 4

1) 3.085, 2.021, 15.004%
2) 23.47%, 25 (rounded figure)

ignou
THE PEOPLE'S
UNIVERSITY