BLOCK IV:
MEASURES OF INEQUALITY

Unit 1 : Measures of Inequality, Standard Deviation and Variance, Coefficient of Variation

Unit 2 : Lorenz curve and Gini coefficient

Unit 3 : Pareto's Law of Income Distribution, deprivation index

Unit-1

Measures of Inequality, Standard Deviation and Variance, Coefficient of Variation

Unit Structure:

1.1 Introduction

1.2 Objectives

1.3 Certain Measures of Inequality

- 1.3.1 Range
- 1.3.2 Mean Deviation
- 1.3.3 Standard Deviation
- 1.3.4 Mathematical Properties of Standard Deviation
- 1.3.5 Standard Deviation of Logarithm
- 1.3.6 Coefficient of Variation
- 1.3.7 Illustrations and Examples
- 1.3.8 Uses of different measures of inequality

1.4 Summing Up

1.5 Key words

1.6 Model Questions

1.7 References and Suggested Readings

1.1 Introduction

In this unit, you will learn about the measures of inequality, concepts of variance and coefficient of variation. In addition, you will learn about the standard deviation of logarithm. It is considered to be the most useful measure of dispersion or standard deviation or root mean square deviation about the mean. Although inequality has long been topic of intense interest to sociologists, few have bothered to carefully specify what they mean by the term. Inequality can be viewed from different perspectives, all of which are related. Most common metric is *Income Inequality*, *Inequality of Wealth, and Inequality of Opportunity, lifetime Inequality etc.* Each of these inequality theories is connected to the others and offers unique yet complementary perspectives on the origins and effects of inequality. As a result, governments are better guided when developing particular policies to combat inequality.

1.2 Objectives

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

- Learn about the significance of range and mean deviation
- Learn about the significance of variance and coefficient of variance

1.3 Certain Measures of Inequality

1.3.1 Range

Let us consider the distribution of income of n persons. Let y_i denote the income of the i^{th} person, i=1,2...n .Then the range R is defined as the gap between the highest and the lowest income levels as a ratio of mean income. Thus the range R is given by

$$R = \frac{\text{Max}\, Y_i - \text{Min}\, Y_i}{\overline{Y}} \quad \text{Where} \quad \overline{Y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

If income is divided equally, then R=0. The main limitation of range as a measure of income inequality is that it ignores the distribution in between the extremes.

Advantages

It can be easily understood

It is easy to calculate and it is the simplest method of measuring dispersion

Disadvantages

1. It is too indefinite to be used as a practical measure of dispersion because it depends entirely upon extreme values.
2. It is not based on all observations
3. It is affected by sampling fluctuations

Uses

It is used in quality control

1.3.2 The relative mean deviation

Let y_i denote the income of the i^{th} person, i=1,2...n .The relative mean deviation M is given by

$$M = \frac{\sum_{i=1}^{n} |y_i - \overline{y}|}{n\overline{y}}$$

With perfect equality M=0 and with all income going to one person $M = \frac{2(n-1)}{n}$

The main problem with the relative mean deviation is that it is not at all sensitive to income transfers whatsoever unless they cross the dividing line of μ on the way. It is therefore, rather arbitrary.

Advantages

I is easy to understand and compute.

Mean deviation about an arbitrary point is least when the point is median.

Disadvantages

In mean deviation the signs of all deviations are taken as positive and therefore it is not suitable for further algebraic treatments .

It is rarely used in social .

It is often not useful for statistical inferences.

Uses

Mean deviation and its coefficients are used in studying economic problems such as distribution of income and wealth in a society.

1.3.3 The Standard Deviation

In the present measurement concept, both the precision concept and the concept of uncertainty are expressed in terms of standard deviation or times standard deviation. Among the various measures of dispersion the standard deviation is considered to be the most useful as the other measures lack adequacy and accuracy. The range is not satisfactory as its magnitude is determined by most extreme cases in the entire group. Mean deviation method is also an unsatisfactory measure as it ignores the algebraic signs of deviation. To some extent standard deviation is one such measure which helps to get rid of the negative sign without committing algebraic violence. The most widely accepted answer for a concise expression to understand the dispersion of data is to square the difference of each value from the group mean which will in fact give positive values. When these squared deviations are added up and then divided by the the number of values in the data, the result is the variance, which is always a positive number, but in different units than the mean. This inconvenience is removed by using the square root of the variance which is the population standard deviation or

S.D. In other words S.D is the square root of the averaged squared deviations from the mean . S.D is also sometimes referred to as "root – mean –square deviation."The measure has precise mathematical significance and eliminates the disregard of signs in the average deviation, by squaring the deviations. Squaring of deviations provides added weight to the extreme items and also the deviations are recorded from the arithmetic mean. The measure is defined as the root-mean square measure of dispersion, a quadratic mean of deviations about the arithmetic mean. Thus if x_i denotes the income of the i^{th} person, i=1,2, …,..n and $\bar{x}$ denotes their average or mean income then variance is denoted by V and is defined by

$$V = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$$

Check Your Progress

1. What purpose does a measure of inequality serve ?
2. Why is standard deviation considered to be a good measure?
3. What is the mean and variance of a standard normal variate?

Now the square root of the mean of the squares of the deviations of individual items from their arithmetic mean defined by standard deviation is given by

$$\sigma = \sqrt{V} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}} \qquad ...(1)$$

For grouped data (discrete variable)

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} f_i (x_i - \bar{x})^2}{\sum_{i=1}^{n} f_i}} \qquad ...(2)$$

And for grouped data (continuous variable)

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} f_i (M - \bar{x})^2}{\sum_{i=1}^{n} f_i}} \qquad ...(3)$$

Where M is the mid -value of the group.

If we take deviations from assumed mean then we get the following two alternative methods of calculating standard deviation:

a) Assumed Mean Method : In case of individual series we have the following formula:

$$\sigma = \sqrt{\frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2} \quad , d = x - A \qquad ...(4)$$

In case of frequency distribution

$$\sigma = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2} \qquad ...(5)$$

b) Step deviation method: If the deviations from assumed mean A have some common factors then these deviations are divided by the highest common factor h (say) and the step deviations denoted by 'd' are obtained. Thus $d' = d / h$ where h is the H.C.F of the deviations d . In this method we have the following formula:

For individual series ,

$$\sigma = \sqrt{\frac{\sum d'^2}{n} - \left(\frac{\sum d'}{n}\right)} \times h \qquad ...(6)$$

For frequency distribution

$$\sigma = \sqrt{\frac{\sum fd'^2}{N} - \left(\frac{\sum fd'}{N}\right)^2} \times h \qquad ...(7)$$

Where $d' = \frac{d}{h} = \frac{x - A}{h}$, A = Assumed mean

Since the normal distribution generally is stated in terms of standardized deviations, an innumerable number of standards exist for determining the pattern of dispersion compared to the theoretical case. Three often used standards are as follows:

$\bar{x} \pm 1\sigma$ $\bar{x} \pm 1\sigma$

includes 68.3% of the frequencies

$\bar{x} \pm 2\sigma$ $\bar{x} \pm 2\sigma$

includes 95.5% of the frequencies

$\bar{x} \pm 3\sigma$ $\bar{x} \pm 3\sigma$

includes 99.7% of the frequencies

In the absence of dispersion , the value of the standard deviation is zero . The size of the standard deviation varies directly in relation to the amount of dispersion. The greater the dispersion the larger its value, and the converse. Although the size of each observation affects the value of the standard deviation, extreme deviations, because of the squaring process, exert an undue weight upon the size of the standard deviation.

Some of the advantages and limitations of standard deviation are

Advantages:

1. The value of standard deviation is based on every observation in a set of data. It is the only measure of variation capable of algebraic treatment and less affected by fluctuations of sampling as compared to other measures of inequality.
2. It is possible to calculate the combined standard deviation of two or more sets of data.
3. S.D is useful in further statistical investigation. For example it plays a vital role in comparing skewness, correlation, and widely used in sampling theory.

Limitations

1. In comparison to other measures of variation, calculation of standard deviations is difficult.
2. The main demerit of variance is, that its unit is the square of the unit of measurement of variate values . For clarity , say, the variable X is measured in cms, the unit of variance is cm^2.Generally , this value is large and makes it difficult to decide about the magnitude of variation
3. While calculating standard deviation , more weight is given to extreme values and less to those near mean . This is because of the fact that while calculating S.D , the deviations from the mean are squared , therefore large deviations when squared are proportionately more than small deviations .

Uses

It is most widely used as a measure of dispersion

It is widely used in biological studies

It is used in fitting a normal curve to a frequency distribution

1.3.4 Mathematical Properties of Standard Deviation

1. Combined Standard deviation

Let a distribution having n_1 observations has mean $\overline{x}_1$ and standard deviation σ_1 and let another distribution having n_2 observations has mean $\overline{x}_2$ and standard deviation σ_2. Then the standard deviation σ of the distribution obtained by combining the two distributions having altogether (n_1+n_2) observations is given by:

$$\sigma = \sqrt{\frac{n_1\sigma_1^2 + n_2\sigma_2^2 + n_1 d_1^2 + n_2 d_2^2}{n_1 + n_2}}$$

Where $d_1 = x_1 - \overline{x}$, $d_2 = x_2 - \overline{x}$ and $\overline{x} = \frac{n_1 x_1 + n_2 x_2}{n_1 + n_2}$

Where $\overline{x}$ is the mean of the combined distribution.

In a similar way standard deviation of the combined distribution of two or more distributions can be obtained.

2. Standard deviation of the first n natural numbers is given by

$$\sigma = \sqrt{\frac{1(n^2 - 1)}{12}}$$

3. Standard deviation is independent of change of origin but not of scale.

1.3.5 The Standard Deviation of Logarithm

In contrast with taking of variance or standard deviation of actual values , if we take the variance and standard deviations of logarithms of the actual values then each of the variance and standard deviation gives greater importance to income transfers at the lower end. The standard deviation of logarithms is denoted by H and it is defined by

$$H = \sqrt{\frac{\sum_{i=1}^{n} (\log y_i - \log \overline{y})^2}{n}}$$

One advantage of the use of logarithm is that it eliminates the arbitrariness of the units and therefore of absolute values , since a change of units, which takes the form of a multiplication of the absolute values, comes out in the logarithmic form as an addition of a constant , and therefore goes out in the wash when pairwise differences are being taken. It is ,therefore ,no wonder that the standard deviation of the logarithm has frequently cropped up as a suggested measure of inequality."[Amartya Sen , "On Economic Inequality",pp.28-29]

> **Check Your Progress**
>
> 1. When is standard deviation of logarithm recommended?
> 2. What are the mathametical properties of standard deviation?
> 3. What are the user, advantages and limitations of Standard Deviation?

1.3.6 Coefficient of Variation

The standard deviation cannot be the sole basis for comparing two distributions. If we have a standard deviation of 10 and a mean of 5, the values vary by an amount twice as large as the mean itself. On the other hand, if we have a standard deviation of 10 and a mean of 5000, the variation relative to the mean is insignificant. Therefore we cannot know the dispersion of a set of data until we know the standard deviation, the mean and how the standard deviation compares with the mean. Also if two series differ in their units of measurements , their variability cannot be measurd by any measure discussd so far. What we need is a relative measure that will give us a fuel for the magnitude of the deviation relative to the magnitude of the mean. Hence in situations where either the two series have different units of measurements , or their means differ sufficiently in size , the coefficient of variation should be used as a measure of dispersion.

The coefficient of variation is one such relative measure of dispersion. It is a simple statistic using the mean and standard deviation. It relates the standard deviation and the mean by expressing the standard deviation as a percentage of the mean. It is a unitless measure of dispersion. The unit of measure then is "percent" rather than the same unit as the original data. The formula for the coefficient of variation is

$$\text{Coefficient of variation} = \frac{s\tan dard\ \ deviation}{mean} \times 100$$

$$= \frac{\sigma}{\bar{x}} \times 100$$

Check Your Progress
1. What is the utility of coefficient of variation?   2. Write a short note on coefficient of variation.

1.2.7 Illustration and Examples

Example 1: The following nine measurements are the heights in inches in a sample of nine soldiers. Compute the standard deviation .

Height(X) : 69 66 67 69 64 63 65 68 72

Solution : Here $\sum_{i=1}^{n} x_i$ = 69+66+….+72= 603

Mean, $\bar{x}$ =603/9= 67 inches

The standard deviation is given by

$$\sigma = \sqrt{V} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}} \qquad ...(1)$$

$$\sigma = \sqrt{\frac{64}{9}}$$

Example 2: Calculate the M.D from the A.M of the series 20,22,27,30,31,32,35,40,45,48.

Solution : We can tabulate the result as follows:

Serial No	1	2	3	4	5	6	7	8	9	10	Total
Marks (X)	20	22	27	30	31	32	35	40	45	48	330
$\lvert X - \bar{X} \rvert$	13	11	6	3	2	1	2	7	12	15	72

Here Mean $\bar{X} = \frac{\sum X}{n} = \frac{330}{10} = 33$

Therefore M.D from mean = $\bar{X} = \frac{\sum \lvert X - \bar{X} \rvert}{n} = \frac{72}{10} = 7.2$

Example 3 : Compute the standard deviation for the following data:

11,12,13,14,15,16,17,18,19,20,21

Solution : Here first we calculate the mean as

$$\overline{X} = \frac{\sum x}{n} = \frac{176}{11} = 16$$

And then calculate the standard deviation as follows:

x	$(x-\overline{x})$	$(x-\overline{x})^2$
11	-5	25
12	-4	16
13	-3	9
14	-2	4
15	-1	1
16	0	0
17	1	1
18	2	4
19	3	9
20	4	16
21	5	25

Thus by formula (1)

$$\sigma = \sqrt{\frac{110}{11}} = \sqrt{10} = 3.16$$

Example 4 : Find the standard deviation from the following record of number of scooter accidents in a street:

No. of accidents	1	2	4	5	6
No. of days	2	3	3	1	1

Solution : The solution is worked out using Assumed Mean Method

x	f	d=x-4	fd	fd^2
1	2	-3	-6	18
2	3	-2	-6	12
4	3	0	0	0
5	1	1	1	1
6	1	2	2	4
	N=10		$\sum fd = -9$	$\sum fd^2 = 35$

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

$$= \sqrt{\frac{35}{10} - \left(\frac{-9}{10}\right)^2}$$

$$= \sqrt{3.5 - 0.81}$$

$$= 1.64$$

Example 5: Calculate standard deviation from the following data:

Age (year)	20-30	30-40	40-50	50-60	60-70	70-80	80-90
No. of persons	3	61	132	153	140	51	2

Solution : Standard deviation is calculated from the following table :

Age	Mid value(x)	Frequency (f)	d =x-55	$d' = d/10$	fd'	fd'^2
20-30	25	3	-30	-3	9	27
30-40	35	61	-20	-2	-122	244
40-50	45	132	-10	-1	-132	132
50-60	55	153	0	0	0	0
60-70	65	140	10	1	140	140
70-80	75	51	20	2	102	204
80-90	85	2	30	3	6	18
		N=542			$\sum fd' = -15$	$\sum fd'^2 = 765$

$$\sigma = \sqrt{\frac{\sum fd'^2}{N} - \left(\frac{\sum fd'}{N}\right)^2} \times h$$

$$= \sqrt{\frac{765}{542} - \left(\frac{-15}{542}\right)^2} \times 10$$

$$= \sqrt{1.4114 - 0.0009} \times 10$$

$$= 11.876$$

i.e. S.D=11.876 years

Example 6: From the prices of shares X and Y below, state which is stable or more consistent in value:

X	55	54	52	53	56	58	52	50	51	49
Y	108	107	105	105	106	107	104	103	104	101

Solution : We have to calculate coefficient of variation for two series as follows:

X	$\lvert X-\overline{X}\rvert$	$\lvert X-\overline{X}\rvert^2$	Y	$\lvert Y-\overline{Y}\rvert$	$\lvert Y-\overline{Y}\rvert^2$
55	2	4	108	4	16
54	1	1	107	3	9
52	-1	1	105	1	1
53	0	0	105	1	1
56	3	9	106	2	4
58	5	25	107	3	9
52	-1	1	104	0	0
50	-3	9	103	-1	1
51	-2	4	104	0	0
49	-4	16	101	-3	9
Total		70			50

For X, Mean $(\overline{X}) = \frac{\sum_{1=1}^{n} X_i}{n} = \frac{530}{10} = 53$ and for Y, Mean

$\overline{Y}) = \frac{\sum_{1=1}^{n} Y_i}{n} = \frac{1050}{10} = 105$

S.D ($(\sigma) = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}}$

$= \sqrt{\frac{70}{10}} = \sqrt{7} = 2.65$

For the series X

$C.V = \frac{\sigma}{\overline{x}} \times 100$

$= \frac{2.65}{53} \times 100$

$= 5$

S.D $(\sigma) = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{n}}$

$= \sqrt{\frac{50}{10}} = \sqrt{5} = 2.23$

For the series Y

$C.V = \frac{\sigma}{\overline{x}} \times 100$

$= \frac{2.23}{105} \times 100$

$= 2.123$

Since C.V for Y <C.V for X, therefore the share Y is more consistent or stable in value.

Stop to Consider

Following is the statement of marks obtained by two students: A and B in 10 examination papers. Comment on whose marks are more consistent, A or B?

Marks scored by A	44	80	76	48	52	72	68	56	60	54
Marks scored by B	48	75	54	60	63	69	72	51	57	66

(The learners are advised to find out the solutions)

Example 7: The following datagive the number of passengers travelling by Jet Airways from Guwahati to Delhi in one week .

115 122 129 113 119 124 132 120 110 116

Calculate the mean and standard deviation and determine the percentage of class that lie between (i) μ±δ (ii) μ±2δ .What percentage of cases lie outside these limits?

Solution : The calculation for mean and standard deviation are shown in the following table

x	$(x-\bar{x})$	$(x-\bar{x})^2$
115	-5	25
122	2	4
129	9	81
113	-7	49
119	-1	1
124	4	16
132	12	144
120	0	0
110	-10	100
116	-4	16
1200	0	436

$$\mu = \frac{\sum x}{N} = \frac{1200}{10} = 120 \quad and \quad \sigma^2 = \frac{\sum (x-\bar{x})^2}{N} = \frac{436}{10} = 43.6$$

$$Therefore \quad \sigma = \sqrt{\sigma^2} = \sqrt{43.6} = 6.60$$

The percentage of cases that lie between a given limit are as follows :

Interval	Values within interval	Percentage of Population	Percentage falling outside
μ± σ =120±6.60 =113.4 and 126.6	113,115,116,119,120,122,124	70%	30%
μ± σ = 120±2(6.60) =106.80 and 133.20	110,113,115,116,119,120,122,124,129,132	100%	nil

Stop to Consider

Example 8: The superintendent of a Civil hospital wanted to see the number of days patients stay in the hospital after surgery and for this he chose randomly 200 patients . The data are given below:

Hospital stay(in days)	1-3	4-6	7-9	10-12	13-15	16-18	19-21	22-24
Number of patients	18	90	44	21	9	9	4	5

(a) Calculate the mean number of days patients stay in the hospital along with standard deviation of the same .

(b) How many patients are expected to stay between 0 and 17 days.

(The learners are advised to find out the solution)

Example 9: For a group of 50 male workers, the mean and the standard deviation of their monthly wages are Rs 6300 and Rs 900 respectively. For a group of 40 female workers, these are
Rs 5400 and Rs 600 respectively. Find the standard deviation of monthly wages for the combined group of workers.

Solution : Given that

$n_1=50 \quad ,\bar{x}_1=6300 \quad ,\sigma_1=900$

$n_2=40 \quad ,\bar{x}_2=5400 \quad ,\sigma_2=600$

The combined mean

$$\bar{x}=\frac{n_1x_1+n_2x_2}{n_1+n_2}=\frac{50\times 6300+40\times 5400}{50+40}=\frac{315000+216000}{90}=\frac{531000}{90}=\text{Rs}\,5900$$

The combined standard deviation

$$\sigma=\sqrt{\frac{n_1\sigma_1^2+n_2\sigma_2^2+n_1d_1^2+n_2d_2^2}{n_1+n_2}}$$

where $d_1=x_1-\bar{x}=\text{Rs.}(6300-5900)=\text{Rs}400$

$d_2=x_2-\bar{x}=\text{Rs.}(5400-5900)=\text{Rs}-500$

$$\sigma=\sqrt{\frac{50\times(900)^2+40\times(600)^2+50\times(400)^2+40\times(-500)^2}{50+40}}$$

$=\text{Rs}\,900$

Example 10: Mr. Goswami wants to invest Rs 10,000 in one of the two companies A or B. Average return in a year from company A is Rs 16,000 with a standard deviation of Rs. 125, while in company B, the average return in a year is Rs .20,000 with a standard deviation of Rs 200.

Which company will you recommend to Mr. Goswami for investment? Justify your answer .

Solution: Coefficient of variation for company $A = \frac{15}{16,000} \times 100 = 0.78\%$

Coefficient of variation for company $B = \frac{200}{20,000} \times 100 = 1\%$

Since the coefficient of variation for company A is less, company A is more consistent and Mr. Goswami should invest in company A.

1.3.8 Uses of different Measures of Inequality

Though range is a very crude measure of dispersion it is mployed for a number of purposes like quality control , fluctuations in share prices, variations in money rats and rates of exchange , weather forecast etc. For example, in industry it is used for quality control . In stock markt range is used to study the fluctuations in share prices . The meteorological department uses rang to determine the difference between maximum temperature and the minimum temperature. This is on account of the fact that the public is generally interested in knowing the limits within which the temperature fluctuates on a particular day .

The mean deviation has found favour particularly with conomists and business statisticians because of the fact that the other measure –standard deviation – gives more weightage to the extreme values . Mean deviation is extensively employed in studies on the distribution of personal wealth and in studies relating to forecasting business cycles . It is the standard deviation that is most extensively employed. It forms the basis of a number of statistical techniques and is used in many other techniques as well. The theory of sampling , regression and correlation , analysis of variance etc all use standard deviation extensively.

1.4 Summing Up

- Range is one of the most basic measures of variation. It is the difference between the smallest data item in the set and the largest.

- Mean deviation is the average of the absolute deviations taken from a central value, generally the mean or median .
- Standard deviation (σ) is defined as the square root of the variance. It measures the variability about the mean of a data set: the closer to the mean, the lower the standard deviation. Its symbol is **σ** (the Greek letter sigma)
- Standard Deviation is a statistical tool that is used widely by statisticians, economists, financial investors, mathematicians, and government officials. It allows these experts to see how variable a collection of data is.
- The standard deviation σ or its square, the variance, cannot be very useful in comparing two series where either the units are different or the mean values are different.
- The disadvantage of the standard deviation lies in the amount of work involved in its calculation and the large weight it attaches to extreme values because of the process of squares involved in its calculation.

1.5 Key Words

Range : It is the difference between the largest and the smallest observations in a set.
Mean deviation: It is the average of the absolute deviations taken from a central value, generally the mean or median
Variance: It is the average of the squares of the deviations taken from mean
Standard deviation: The positive square root of the variance is called standard deviation
Coefficient of variation: It is a relative measure of dispersion which relates the standard deviation and the mean by expressing the standard deviation as a percentage of the mean

1.6 Model Questions

A. Multiple Choice Questions

1. Which of the following is a method of measuring deviations from the average?

(a) Root mean square (b) Lorenz Curve

(c) Gini Coefficient (d) None of the above

2. Coefficient of variation is given by

(a) $\frac{\sigma}{\overline{x}}$ (b) $\frac{\overline{x}}{\sigma}$

(c) $\frac{\overline{x}}{\sigma} \times 100$ (d) $\frac{\sigma}{\overline{x}} \times 100$

3. Which of the following is a unit free number

(a) S.D (b) variance

(c) M.D (d) C.V

4. Root mean square deviation from mean is

(a) Standard deviation (b) Coefficient of variation

(c) both (d) none

B. Fill in the blanks

1. The square of standard deviation , namely σ^2 is termed as __________

2. If in a series , coefficient of variation is 64 and mean is 1, the standard deviation shall be __________

C. State whether true or false

1. The range is the easiest measure to measure inequality

2. variance and coefficient of variance are the same

D. Short –Answer Questions

1. Define mean deviation.

2. Define standard deviation

3. Write two advantages of standard deviation .

4. What is coefficient of variation?

5. Why is standard deviation preferred to mean deviation?

E. Long Answer Questions

1. Discuss the various measures of Inequality.
2. Discuss the variance and the coefficient of variation as measures of inequality.
3. Find out the standard deviation and variance from the following frequency distribution.

Marks:	0-4	4-8	8-12	12-16
No of students	4	8	2	1

   For a group of 50 male workers, the mean and standard deviation of their daily wages are Rs 72 and Rs 9 respectively. For another group of 40 female workers these are Rs 54 and Rs 6 respectively. Find the standard deviation for the combined group of 90 workers

4. From the data given below, state which series is more variable?

Variable	Series A	Series B
10-20	10	18
20-30	18	22
30-40	32	40
40-50	40	32
50-60	22	18
60-70	18	10

5. The following are the scores of two batsmen P and Q in a series of innings.

A	12	15	106	173	117	10	19	136	184	29
B	147	112	6	42	14	0	37	148	13	101

   Who is the better run-getter? Who is more consistent?

6. In two factories A and B , engaged in the same activity , the average weekly wages and standard deviation are as follows:
   (i) Which factory pays larger amount as weekly wages?
   (ii) Which factory shows greater variability in the distribution of wages?

(iii) What is the mean and standard deviation of all workers in these two factories taken together?

7. The first of two samples has 50 items with mean 54.4 and standard deviation 8. If the whole group has 150 items with mean 51.7 and S.D 7.6, find the mean and standard deviation of the second group.

8. Explain the procedure of calculating the mean deviation from grouped frequency distribution . How standard deviation is superior than mean deviation ?

9. Weights of the students in kgs . are recorded by a machine as under:

   50 55 57 49 54 61 64 59 58 56

   If the weighing machine shows weight less by 5 kg, find correct values of range , standard deviation and coefficient of variation without calculating the correct weights.

10. A group of 100 selected students average 163.8 cm in height with a coefficient of variation of 3.2%, what was the standard deviation of their height.

1.8 References and Suggested Readings

1. Hazarika, P.L. Essential Statistics for Economics and Business Statistics. New Delhi. Akansha Publishing House,2012
2. Sharma. J.K Business Statistics .Pearson Education , New Delhi,2007
3. Gupta, S.C Fundamental of Statistics , New Delhi : S. Chand and sons, 2005
4. Gupta, S.P .Statistical Methods .New Delhi. S. Chand and Sons, 2005
5. Hooda, R.P. Statistics for Business and Economics. New Delhi: Macmillan India Ltd 2002

******

Unit-2
Lorenz Curve and Gini Coefficient

Unit Structure:

2.1 Introduction
2.2 Objectives
2.3 Lorenz Curve
2.4 Gini Coefficient
2.5 Summing Up
2.6 Model Questions
2.9 References and Suggested Readings

2.1 Introduction

In this unit you will learn about the Lorenz Curve and Gini Coefficient. Lorenz Curve is a graphic method for studying dispersion or variation in income inequality. It was developed by Max. O. Lorenz in 1905 for representing inequality in the wealth distribution. The Gini Coefficient is a measure of statistical dispersion and it is the most popular method for operationalizing income inequality in the public health literature.

2.2 Objectives

After going through this unit, you will be able to

- understand the significance of Lorenz Curve,
- discuss the function of Gini Coefficient.

2.3 The Lorenz Curve

The Lorenz Curve is a graphic method of studying variation. It was developed by Max O Lorenz in 1905 for representing inequality of the wealth distribution i.e the variability of the distribution of income and wealth is quantified using Lorenz curves. Consequently, the Lorenz Curve is a measurement of how far a statistical series' actual distribution deviates from the line of equal distribution. The Lorenz Coefficient measures the magnitude of this deviation. There is greater inequality or unpredictability in the series if the Lorenz Curve is farther away from the line of equal distribution, and vice versa. However, it can be applied with equal advantage for comparing the distribution of profits amongst different groups of business and such other things.

Construction of Lorenz Curve

The steps involved in the construction of a Lorenz Curve are as follows:

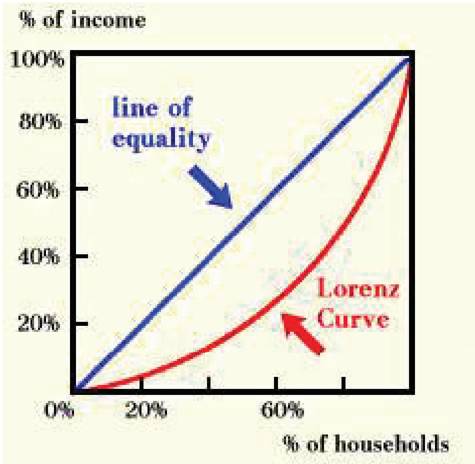
Step 1: The first step is to convert the given series into a cumulative frequency series. Then, the various items in the series are converted into percentages of the cumulative sum using the assumption that the cumulative sum of the items (or the mid-values of the class intervals) equals 100.

Step 2: The cumulative frequencies and items are plotted on a graph's X- and Y-axes, respectively, in the second step. For drawing Lorenz curve, the percentage of the population arranged from the poorest to the richest are represented on the horizontal axis (x-axis). The share of total income received by each percentage of population is represented along the (y-axis).This is also cumulated to 100. The zero percentage on the x-axis must be joined with the 100% along the y-axis. This is called the Line of Equal Distribution. This line makes an angle of 45^0 with the X-axis. Thus it is obvious that 0% of the population enjoys 0% of the income and 100% of the population enjoys 100% of income. If the wealth is equally distributed among the people then the Lorenz curve is a straight line or simply the diagonal. The further the Lorenz Curve is away from the line of equal distribution, the more unequal is the distribution of income.

Step 3: In the last step plot the actual data on the graph and obtain a curve joining the plotted points. This curve shows the actual distribution of the given statistical series.

The actual distribution curve is known as **Lorenz Curve.** If there is closeness in the Lorenz Curve to the Equal Distribution Line, it means that there is lesser variation in the distribution. However, if there is larger gap between the Lorenz Curve and the Equal Distribution Line, it means that there is greater variation in the distribution.

Besides, if two Lorenz Curves are drawn on the same graph paper, then the one which is further away from the equal distribution line shows greater variation.



Application of Lorenz Curve

A graphic measure of dispersion in a statistical series is known as Lorenz Curve. It provides the user with an immediate glimpse of the degree of variation in the given statistical distribution from its mean value; hence, is a simple measure. Prof. Lorenz first used this measure of dispersion for the measurement of economic inequality related to the distribution of income and wealth across different nations or different time periods for the same nation. Since then, the application of the Lorenz Curve has spread widely for the measurement of disparity of distribution related to various parameters of wages and profits.

The parameters in which the Lorenz Curve is now applied for the measure of dispersion are as follows:

- Distribution of Income
- Distribution of Wages
- Distribution of Wealth
- Distribution of Profits
- Distribution of Production
- Distribution of Population

Government agencies are especially interested in Lorenz curves, especially for net worth and income distributions within their country. Lorenz curves inform governments of how public policy is working (or not working). It may also be an indicator of how a government should establish its tax brackets based on gaps or ranges of income.

Merits of Lorenz Curve

The merits of the Lorenz Curve are as follows:

1. It is attractive and gives a rough idea of the extent of dispersion.
2. Lorenz Curve makes it easy to compare two or more series.

Demerits of Lorenz Curve

Lorenz curve suffer from a serious limitation:

1. With the help of the Lorenz Curve, one can only get a relative idea of the dispersion of the given distribution as compared with the line of equal distribution. Also, it does not provide the user with any numerical value of the variability for the given distribution. The inequality measured is not expressed in quantitative terms. It merely gives a picture of the extent to which an income distribution is pulled away from the line of equal distribution which ensures that everyone has the same income.
2. It is difficult to draw Lorenz Curve.

Example: Draw the Lorenz curve for the data relating to the profit of 50 manufacturing firms and show the extent of inequality present in the profits:

Profit (Rs'000)	10-20	20-30	30-40	40-50	50-60
No of firms	5	13	18	10	4

Solution : Computation of Lorenz Curve

Profit(Rs'000)	Mid Point (m)	Frequency (No. of firms ,f)	Cumulative frequency	% cumulative frequency(X)	Total value (mxf)	Cumulative total values	% cumulative total values (Y)
10-20	15	5	5	10	75	75	4.4
20-30	25	13	18	36	325	400	23.5
30-40	35	18	36	72	630	1,030	60.6
40-50	45	10	46	92	450	1,480	87.1
50-60	55	4	50	100	220	1,700	100.0

Now the Lorenz curve is drawn taking the percentage cumulative frequency (X) on X-axis and percentage cumulative total values (Y) on Y-axis . Look at figure carefully and study how it is drawn .

In recent studies, the Lorenz curve technique is used as a tool to inter distributional considerations in economic analysis. The concept of the Lorenz curve has been extended and generalized to study the relationships amogst distributions of different economic variables. The generalized Lorenz curve is called concentration curves and the Lorenz curve is only a special case to curves, viz, the concentration curve for i.

Stop to Consider

1. What is a Lorenz curve ?
2. Explain Lorenz curve with the help of a diagram .

2.4 Gini Coefficient

A measure that has been widely used to represent the extent of inequality is the Gini Coefficient developed by the Italian statistician and sociologist Corrada Gini and published in his paper 'Variability and Mutability' in the year 1912. It is also known as Gini index or Gini ratio.

Gini coefficient is based on the Lorenz curve and is defined as the ratio of the area between the diagonal and the Lorenz curve to the total area of the half square in which the curve lies. The Gini coefficient is usually defined mathematically based on the Lorenz curve, which plots the proportion of the total income of the population (y-axis) that is cumulatively earned by the bottom X% of the population as illustrated by the figure. The line at 45^0 represents perfect equality of incomes. The Gini coefficient can be considered as the ratio of the area that lies between the line of equality and Lorenz curve (say A) over the total area under the line of equality (say A and B)

i.e $G = \frac{A}{A+B}$

Theoretically Gini coefficient can lie between the two extreme values of 0 to 1. A Gini coefficient of zero expresses perfect equality where all values are the same i.e. where everyone has an exactly equal income. A Gini coefficient of '1' expresses maximal inequality among values where only one person has all the income. However a value greater than 1 may occur if some persons have negative income or wealth. The application areas of Gini coefficient are in the study of inequalities in discipline as diverse as sociology, economics, health science , ecology , chemistry, engineering and agriculture . In fact, Gini coefficient lies between 0.5 and 0.7 for countries having highly unequal income distributions and between 0.2 and 0.35 for countries having relatively equitable distributions. The World Bank is the main organization that provides the Gini index data. However, data is only available for 130 countries. Numerous other organizations provide statistics on income inequality and the ranking of countries using the World Bank's Gini index data.

Stop to Consider

- The Lorenz Curve – The Lorenz Curve is a graphic method of studying deviations from the average.
- Gini coefficient - Gini coefficient is based on the Lorenz curve and is defined as the ratio of the area between the diagonal and the Lorenz curve to the total area of the half square in which the curve lies.

Inspite of being popular with economists and statisticians, Gini Coefficient suffers from few drawbacks:

1) The Gini index is a relative measure that fails to capture absolute differences in income. It is possible for the Gini index of a country to rise due to increasing income inequality while the number of people living in absolute poverty is actually declining. This is because the Gini index violates the Pareto improvement principle, which says income inequality can increase with an increase in all incomes in a given society.
2) The measurement of the area between the diagonal line OP and the Lorenz Curve is at times difficult .The area between the diagonal and the Lorenz curve can be obtained by using integral calculus provided we know the functional form of the Lorenz curve.
3) Two countries could have different income distributions but the same Gini index. For example, in a country where 50% of the people have no income and the other 50% of the people have equal income, the Gini index is 0.5. In another scenario, where 75% of people with no income account for 25% of a country's total income, and the top 25% of people with an income account for 75% of the country's total income, the Gini index will also be 0.5. Consequently, as a basis for ranking the differences in income inequality between countries, the Gini index could be misleading.

4) The index does not capture social benefits or interventions that bridge inequality between rich and poor.

5) The Gini index also does not capture social benefits or other interventions aimed at bridging inequality between rich and poor. Subsidised housing, healthcare, education and social grants for the vulnerable are measures that subsidise household incomes, reducing income inequality to someextent.

6) Demographic changes or characteristics of the population are not reflected by the Gini index. Countries with high ratios of elderly people whose main sources of income are pensions, or countries with high student ratios are likely to have higher levels of income inequality as measured by the Gini index.

Check Your Progress

1. Explain the merits and demerits of Lorenz curve.
2. What is Gini co-efficient?
3. Discuss the draw backs of Gini Co-efficient.

2.5 Summing Up

- The Lorenz Curve is a graphic method of studying variation. It was developed by Max O Lorenz in 1905 for representing inequality of the wealth distribution.
- The closer the Lorenz curve is to the line of perfect equality, the less the inequality and smaller the Gini coefficient.
- If the wealth is equally distributed among the people then the Lorenz curve is a straight line or simply the diagonal. This line is called the line of equal distribution.
- Gini coefficient is based on the Lorenz curve and is defined as the ratio of the area between the diagonal and the Lorenz curve to the total area of the half square in which the curve lies.
- Gini coefficient can lie between the two extreme values of 0 to 1.
- A Gini coefficient of zero expresses perfect equality where all values are the same i.e. where everyone has an exactly equal income.

- A Gini coefficient of '1' expresses maximal inequality among values where only one person has all the income.
- The application areas of Gini coefficient are in the study of inequalities in discipline as diverse as sociology, economics, health science, ecology, chemistry, engineering and agriculture.
- The World Bank is the main organisation that provides the Gini index data.

2.6 Model Questions

A. Multiple Choice Questions

1. ___ is a measure of statistical dispersion

   (a) Standard deviation (b) Lorenz Curve

   (c) Gini Coefficient (d) None of the above

2. ___depicts inequality across a population

   (a) Standard deviation (b) Lorenz Curve

   (c) Gini Coefficient (d) None of the above

B. Fill in the blanks

1. The Lorenz curve was developed by Max O Lorenz in _____
2. Gini Coefficient of) indicates _______

C. State whether true or false

1. A Gini Coefficient of 1 measures perfect equality.
2. The Lorenz curve is a graphic method of studying variation.

D. Short –Answer Questions :

1. Why was Lorenz curve developed?
2. Give one limitation of Lorenz Curve.
3. Who developed Gini Coefficient?
4. What is Gini Index?
5. What purpose does a Gini Coefficient serve ?
6. Give one limitation of Gini Coefficient.

E. Long Answer Questions

1. Draw a Lorenz curve of the data given below:

Income:	100	200	400	500	800
No of persons:	80	70	50	30	20

2. Discuss the importance of Gini Coefficient.
3. Why is the Lorenz Curve Important?
4. How Does the Lorenz Curve Measure Inequality?

2.7 References and Suggested Readings

Hazarika, P.L. Essential Statistics for Economics and Business Statistics, New Delhi. Akansha Publishing House,2012

Sharma. J.K Business Statistics, Pearson Education , New Delhi,2007

Gupta, S.C Fundamental of Statistics , New Delhi : S. Chand and sons, 2005

Gupta , S.P .Statistical Methods, New Delhi, S. Chand and Sons, 2005

*****

Unit-3

Pareto's Law of Income Distribution, deprivation index

Unit Structure:

3.1 Introduction

3.2 Objectives

3.3 Pareto Law of Income Distribution

3.4 Deprivation Index

3.5 Summing Up

3.6 Model Questions

3.7 References and Suggested Readings

3.1 Introduction

In this unit you will learn about the Pareto's Law of Income Distribution and Deprivation Index. The Pareto Distribution is used in describing social , scientific and geophysical phenomena in society. Pareto created a mathematical formula in the early 20th century that described the inequalities in wealth distribution that existed in his native country of Italy. Deprivation index was associated with the construction of Human Development index (HDI) in Human Development Report, 1990 published by the United Nations Development Programme (UNDP, 1990).

3.2 Objectives

After going through this unit, you will be to:

- understand Pareto's Law of Income Distribution,
- discuss Deprivation Index.

3.3 Pareto Law of Income Distribution

In the 1890s, Vilfredo Pareto (1848-1923), who was an Italian Economist, Sociologist and Engineer studied income tax data from England, Ireland, several Italian cities, German states, and Peru. He wanted to study the problem of distribution of income among the citizens of a state. He plotted the number of people earning an income above a certain threshold against the respective threshold on double logarithmic paper and revealed a linear relationship. It is sometimes referred to as the Pareto Principle or the 80-20 Rule. Pareto felt that he had discovered a new type of "universal law" that was the result of underlying economic mechanisms. Since then, Pareto's discovery has been confirmed and generalized to the distribution of firm size (Axtell (2001)) and wealth, which also follow Pareto distributions, at least in the upper tail. Pareto observed that 80% of the country's wealth was concentrated in the hands of only 20% of the population. The theory is now applied in many disciplines such as incomes, productivity, populations, and other variables. The Pareto distribution serves to show that the level of inputs and outputs is not always equal.

Pareto's Law of income distribution can precisely be stated as follows:

"In all places and at all times , the distribution of income in a stable economy is given approximately by the empirical formula

$y=A(x-a)^{-\beta}$

where y is the number of people having income x or greater,'a' is the lowest income at which the curve begins and A and â are certain parameters"

The graph of the equation approaches the line x=a as x tends to a and it approaches the X-axis when x tends to infinity. Thus the Pareto curve is asymptotic to the lines x=a and y=0 and it looks like a hyperbola.

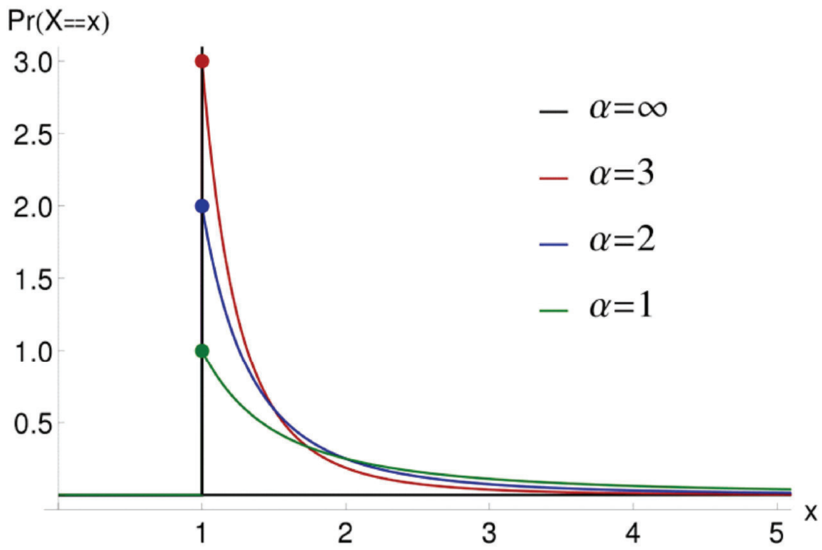
If we shift the origin to the point (a,0) then the Pareto curve takes the form

$\mathbf{Y=Ax^{-\beta}}$

Pareto observed that in many countries the value of â varied from 1.2 to 1.9 and consequently the value of gamma can be taken approximately as 1.5

Thus, gamma can be interpreted as the income elasticity of y i.e as the elasticity of the decrease in the number of persons when passing to a higher income class. Thus, the value of gamma provides a certain measure of the inequality of distribution of income. An increase in the value of gamma implies a corresponding increase in the difference between the incomes of various classes of people.

On a chart, the Pareto distribution is represented by a slowly declining tail, as shown below. Pareto chart is also called a **Pareto diagram** and **Pareto analysis**



The chart is defined by the variables α and x. It provides two main applications. One of the applications is to model the distribution of wealth among individuals in a country. The chart shows the extent to which a large portion of wealth in any country is owned by a small percentage of the people living in that country. The second application is to model the distribution of city populations, where a large percentage of the population is concentrated in the urban centers and a lower amount in the rural areas. The population in urban centers continues to increase while the rural population continues to decline as younger members of the population migrate to urban centers.

Applications of Pareto Chart

Pareto charts are the best chart to do the analysis of the bulk of data. In business industries, these charts are used very often. Let us see some of its more applications.

- For the analysis of the revenue growth of the organisation with respect to the time period.
- To choose for any specific data and work on it, in a broad set of data available.
- To explain to other people the set of data that one has.

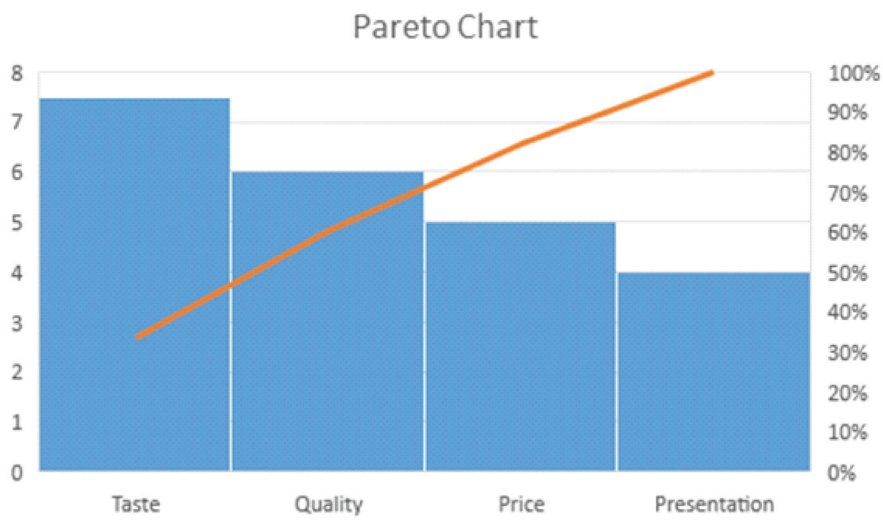
- For the analysis of population growth in a city or country or all over the world every year.
- To check the global problems and focus on resolving the major one.
- To check the major complaints coming from the public and resolve them on priority

Pareto Chart Example

Let us take an example, where we need to prepare a chart of feedback analysis for XYZ restaurant, as per the reviews and ratings received from the customers. Here the customers are given a checklist of four points based on which they have to rate the restaurant out of 10. The four points are:

1. Taste of the Food
2. Quality of the food
3. Price
4. Presentation

Now, let us draw the Pareto chart for the Feedback of XYZ restaurant as per the data received



Thus, Pareto chart considers the percentage of frequency (or measure) and cumulative percentage of measures to draw a line along with bars. Also, the cumulative percentage adds up to 100.

Limitations of Pareto's Law

While the 80-20 Pareto distribution rule applies to many disciplines, it does not necessarily mean that the input and output must be equal to 100%. For example, 20% of the company's customers could contribute 70% of the company's revenues. The ratio brings a total of 90%. It shows that the Pareto concept is merely an observation that suggests that the company should focus on certain inputs more than others.

Pareto's curve does not seem to fit better for lower incomes although it fits better for higher incomes. Pareto's law of income distribution is not effective for all types of economies. It is usually relevant to incomes in the capitalist countries and also the countries with feudal and early capitalist conditions . The definition is very rigid and stated in a very general form. Because of this Pareto's Law has been subjected to serious criticism by a number of scientific investigations. However, the distribution of income in every stable economy is found to follow approximately Pareto pattern.

3.4 Deprivation Index

Deprivation index was associated with the construction of Human Development index (HDI) in Human Development Report, 1990 published by the United Nations Development Programme (UNDP, 1990). Deprivation index is a composite index. One minus the deprivation index is the development index.

There are various components of deprivation. Usually the following components of deprivation in respect of a particular geographical region are taken into consideration.

(i) Life expectancy at birth(in years)

(ii) Adult literacy rate (in per cent)

(iii) Combined primary, secondary and tertiary enrolment ratio

(iv) Per capita Gross Domestic Product(GDP)

The deprivation indicator in respect of a particular above-mentioned component is obtained by using the following formula

Where maximum value and minimum value are specified for a particular geographical region and actual value denotes the value of the indicator at a particular point of time in the given region.

The average of the deprivation indicators in respect of all the four components of deprivation as mentioned above is taken. This is the composite deprivation index. One minus this composite deprivation index is the development index for the region.

Check Your Progress

1. What is a Pareto chart used for?
2. What is the 80/20 rule of Pareto charts?
3. Give example of a Pareto Chart.
4. What is development index?

Stop to Consider

Pareto Law of Income Distribution : Pareto created a mathematical formula in the early 20th century that described the inequalities in wealth distribution that existed in his native country of Italy

Pareto Curve: On a chart, the Pareto distribution is represented by a slowly declining tail called Pareto Curve

Deprivation Index: It is an index used in the construction of human development index

Development Index: One minus the deprivation index is the development index

3.5 Summing Up

Pareto's Law of Income Distribution forms the basis of the well-known, but often overlooked, 'eighty-twenty' rule that a small proportion of customers (or donors) are accountable for a very large share of sales turnover or income.

Pareto Law of Income Distribution is used to model the distribution of wealth among individuals in a country

Deprivation index is a composite index

One minus the deprivation index is the development index

3.6 Model Questions

1. How do you create a Pareto chart?
2. What are the applications of Pareto Chart?
3. What is the deprivation index ?
4. What are the components of the deprivation index?

3.7 References and Suggested Readings

Hazarika, P.L. Essential Statistics for Economics and Business Statistics. New Delhi. Akansha Publishing House,2012

Sharma. J.K Business Statistics .Pearson Education, New Delhi,2007

Gupta, S.C Fundamental of Statistics, New Delhi : S. Chand and sons,2005

Gupta, S.P .Statistical Methods. New Delhi . S. Chand and Sons,2005

Hooda, R.P. Statistics for Business and Economics. New Delhi: Macmillan India Ltd 2002

******