Department of Distance Education

Punjabi University, Patiala

Class : M.A. I (Economics) **Semester : 2**

Paper : III (Basic Quantitative Mathod) **Unit : II**

Medium : English

Lesson No.

Department website : www.pbidde.org

M.A. (ECONOMICS) PART-I (Semester-II) **PAPER-III (BASIC QUANTITATIVE METHODS)**

LESSON NO. 2.1 **AUTHOR : DR. VIPLA CHOPRA**

INTERPOLATION

Structure

2.1.1 Introduction :

In the course of statistical investigation, it may be found that the values for certain past years are available but the values for some other years are either not collected or destroyed due to fire, flood etc. to remove such difficulties, it becomes essential to estimate such values which are either missing or not collected. Interpolation, thus, refers to the insertion of an intermediate value in a series of items. In the present lesson our objective is to study the meaning, significance and different methods of interpolation.

Meaning : Interpolation refers to the estimation of a figure or value within the given limits of the available data under certain assumptions. It provides the missing link. For example, the census of population in India takes place every 10 years, i.e., we have census figures for 1951, 1961, 1971, 1981, 1991 and 2001. Now, if we require population figures for 1988 or 2000, what should we do? To talk of a census for 1988 or 2000 is impracticable. What is desirable is to obtain the required estimates by analysing the available data through interpolation. But it should be noted that interpolated values are only the best substitutes of the original figures. Sometimes, due to certain organizational and financial difficulties, collection of figures on the census basis is not possible and the sampling techniques are adopted. For filling the gaps, the method of interpolation is resorted to. The necessity of interpolating missing figures arises in the case of acturial work, commerce, economic and social studies. A few definition of interpolation are as follows :

Interpolation is "the art of reading between the lines of the table". - **Theile**

"Interpolation is the estimation of a most likely estimate in given conditions. The technique of estimating a past figure is termed as interpolation, while that of estimating a probable figure for the future is called extrapolation." - **Hirach**

2.1.2 Necessity of Interpolation :

There are certain cases under which interpolation is a necessity. These various cases are as follows :-

1. Interpolation has been found useful to calculate the values of median and mode when the data are available in the form of class intervals and class frequencies.
2. Interpolation is also useful in estimating the population figure for intercensul years. It also helps forecasting or projecting the future values on the basis of the past and present results.
3. Sometimes, the already collected data may not be adequate for the purpose of enquiry. To fill gaps in coverage, interpolation is a necessity.
4. Sometimes, gaps in statistical data may arise. These gaps in statistical data may arise due to various reasons. It may not be possible to collect the data either due to financial or organisational difficulties. In some cases, the data may have been lost due to flood, fire etc. Such gaps in statistical data can be filled by the process of interpolation.

2.1.3 Assumptions :

The following assumptions are made while making use of the techniques of interpolation.

1. No sudden or violent fluctuations :

While interpolating a figure, we always presume that there are no sudden ups and downs in the data. For example, if we are given the population figures for 1961, 1971, 1981, 1991 & 2001 and we are asked to interpolate the figure for 1999, this would be done on the assumption that throughout this period from 1961-2001 there has been no violent changes in population. That is, during the intervening period (1961-2001), there are no wars, floods, famines or any kind of abnormalities.

2. Regularity or Uniformity of Changes :

Another assumption made while interpolating values is that the rate of change of figures from one period to another is uniform. That is, the rate of increase or decrease should almost be the same. In the above example, our assumption would be that the rate of change of population from 1961-2001 has been uniform.

2.1.4 Methods of Interpolation :

There are two methods by which the figures may be interpolated. They are - (1) Graphic and (2) Algebraic.

The algebraic methods have assumed various forms :

1. Binominal Expansion Method
2. Newton's Method
3. Lagrange's Method

In the syllabus only three algebraic methods have been mentioned namely :

(a) Newton's formula for leading differences,
(b) Lagrange's formula, and
(c) Binomial expansion method

The present lesson gives in detail the above mentioned methods only.

2.1.4.1 Binomial Expansion Method :

This method of interpolation is simple to understand and requires very little calculations. This method is based on the binomial theorem of algebra. We know that the binomial expansion of $(q+p)^n$ as follows :

$$(q+p)^n = q^n + nq^{n-1}\, p + \frac{n(n-1)}{1.2}\, q^{n-2}\, p^2 + \frac{n(n-1)\,(n-2)}{1.2.3}\, q^{n-3}\,.\, p^3 + \text{--------}$$

where q stands for failure and p for success. This method is applicable only in those situations where the following two conditions are satisfied :

(i) the x-series or independent variable advances by equal intervals. The value of the independent variable are given at equidistant i.e. 1971, 1981, 1991 and 2001 etc.

(ii) the values to be interpolated corresponds to one of the equidistant values.

For example, if we are given the population figures for 1971, 1981 and 2001 and we are to estimate the population figure for 1995, then we apply binomial expansion method.

When this method is used we expand the binomial $(y-1)^n$ and equate it to zero.

$$(y-1)^n = y^n - ny^{n-1} + \frac{n(n-1)}{\lfloor 2} y^{n-2} \frac{n(n-1)(n-2)}{\lfloor 3} y^{n-3} + \ldots\ldots\ldots = 0$$

where n is the number of known values of y.

Steps : This method involves the following steps :

(i) The values of the independent variable (x) are denoted by x_0, x_1, x_2 etc., and the values of the dependent variable are denoted y_0, y_1, y_2 etc. Out of the various y values one or two values may be unknown or missing.

(ii) The known values of the dependent variable (y) are then counted. If the known values of y are n, then we take the n^{th} leading difference to be zero. Symbolically,

$$\Delta_0^n = 0$$

In case of 6 known values,

$$\Delta_0^6 = 0$$

Here Δ_0^n, Δ_0^6 are the n^{th} and 6^{th} leading differences.

(iii) Using Binomial theorem.

$$\Delta_0^n = (y-1)^n$$

$$\Delta_0^n = y^n - ny^{n-1} + \frac{n(n-1)}{1.2} y^{n-2} - \frac{n(n-1)(n-2)}{1.2.3} y^{n-3+} \ldots\ldots\ldots\ldots$$

where n = number of known values.
In case the known values are 5, then

$$\Delta_0^5 = (y-1)^5 = y^5 - 5y^{5-1} + \frac{5(5-1)}{1 \times 2} y^{5-2} - \frac{5(5-1)(5-2)}{1 \times 2 \times 3} y^{5-3}$$

$$+ \frac{5(5-1)(5-2)(5-3)}{1 \times 2 \times 3 \times 4} y^{5-4} - \frac{5(5-1)(5-2)(5-3)(5-4)}{1 \times 2 \times 3 \times 4 \times 5} y^{5-5}$$

$$= y_5 - 5y_4 + 10y_3 - 10y_2 + 5y_1 - y_0 = 0$$

Note : In this formula, the powers of the y's are used as subscripts of the y's i.e. y^6 will be taken as y_6, y^7 as y_7 and so on.

(iv) The following procedure may be adopted for simplifying the binomial expansion $(y-1)^n$:

(a) First, we write the different values of y's. If the known values are 6, first we write y_6 and then subsequently subtracting 1 from it until we get y_0. Thus,

$y_6, y_5, y_4, y_3, y_2, y_1, y_0$

(b) The plus and minus signs are to be placed alternatively starting from the first, which will be plus.

For example : $+ y_5 - y_4 + y_3 - y_2 + y_1 - y_0$

(c) The numerical co-efficients for y will be determined as follows : The numerical co-efficient of the first value of y's is always unity. The numerical co-efficient of other values of y's are obtained by using the formula :

$$\frac{\text{Coefficient of previous y} \times \text{subscript of the previous y}}{\text{Serial order of previous y}}$$

In the above example,

$$\text{I}.y_5 - 1\times\frac{5}{1}y_4 + \frac{5\times 4}{2}y_3 - \frac{10\times 3}{3}y_2 + ------$$

$$= y_5 - 5y_4 + 10y_3 - 10y_2 + 5y_1 - y_0 = 0$$

Some important Binomial Expansions

No. of known values	Formula	Binomial Expansion
2 or Δ_0^2	$(y-1)^2 = 0,$	$y_2 - 2y_1 + y_0 = 0$
3 or Δ_0^3	$(y-1)^3 = 0,$	$y_3 - 3y_2 + 3y_1 - y_0 = 0$
4 or Δ_0^4	$(y-1)^4 = 0,$	$y_4 - 4y_3 + 6y_2 - 4y_1 + y_0 = 0$
5 or Δ_0^5	$(y-1)^5 = 0,$	$y_5 - 5y_4 + 10y_3 - 10y_2 + 5y_1 - y_0 = 0$
6 or Δ_0^6	$(y-1)^6 = 0,$	$y_6 - 6y_5 + 15y_4 - 20y_3 + 15y_2 - 6y_1 + y_0 = 0$
7 or Δ_0^7	$(y-1)^7 = 0,$	$y_7-7y_6+21y_5-35y_4+35y_3-21y_2+7y_1-y_0=0$
8 or Δ_0^8	$(y-1)^8 = 0,$	$y_8-8y_7+28y_6-56y_5+70y_4-56y_3+28y_2-8y_1+y_0=0$

Example-1 : Compute the population for 1941 from the following table :

Year	Population (in millions)
1921	253

1931	287
1941	?
1951	315
1961	319

Solution : Since the known value are four, the fourth leading differences will be zero i.e. $(y-1)^4 = 0$ or $\Delta_0^4 = 0$

Year	Population (in millions)	
1921	253	y_0
1931	287	y_1
1941	?	y_2
1951	315	y_3
1961	319	y_4

Substituting these values, we have

$319 - 4\times315 + 6y_2 - 4\times287 + 253 = 0$

$6y_2 = 1836,\ y_2 = 306$

Hence the population of 1941 = 306 millions.

Example-2 : Estimate the missing term in the following table :

x :	1	2	3	4	5	6	7
y :	2	4	8	?	32	64	128

Solution : Since the known figures are six, the sixth leading differences will be zero.

$$(y-1)^6 \text{ or } \Delta_0^6 = 0$$

$$y_6 - 6y_5 + 15y_4 - 20y_3 + 15y_2 - 6y_1 + y_0 = 0$$

Substituting the values,

$$\Delta_0^6 = 128 - 6\times64 + 15\times32 - 20xy_3 + 15\times8 - 6\times4+2 = 0$$

$$= 128 - 384+480 - 20y_3 + 120 - 24+2 = 0$$

$$20y_3 = 128 - 384 + 480 + 120 - 24 + 2 = 322$$

$$y_3 = 16.1$$

Thus the most probable value of y corresponding to x=4, is 16.1

Two or More Missing Figures :

When two figures are missing in a series of items, there shall be two unknown quantities in the equation. Then a second equation is also obtained by ignoring the last given value in the series of items. With the help of these two equations, the two unknown quantities are obtained.

Steps :

(i) On the basis of the total numbers if known values and taking the leading difference of that order to be zero, the binomial expansion is written.

(ii) Then we increase the subscript of the y's in the binomial expansion by 1 so that at the end we get y_1 instead of y_0.

(iii) Then we substitute the values in the two equations. Solving the two equations simultaneously, we get the missing values.

Example-3 : Interpolate the missing figures in the following table :

Year	1951	1952	1953	1954	1955	1956	1957	1958
Sales in (000 Rs.)	10	-	18	25	-	35	37	45

Solution : As the 6 values are known and two values are missing, we shall assume that the sixth first leading difference and sixth second leading difference would be zero. Thus,

$$\Delta_0^6 = y_6 - 6y_5 + 15y_4 - 20y_3 + 15y_2 - 6y_1 + y_0 = 0 \quad \text{----- (i)}$$

$$\Delta_1^6 = y_7 - 6y_6 + 15y_5 - 20y_4 + 15y_3 - 6y_2 + y_1 = 0 \quad \text{----- (ii)}$$

We are given

Year	1951	1952	1953	1954	1955	1956	1957	1958
Sales in (000 Rs.)	10	-	18	25	-	35	37	45
	y_0	y_1	y_2	y_3	y_4	y_5	y_6	y_7

By substituting the value in (i) and (ii) we get,

$$37 - 6(35) + 15y_4 - 20(25) + 15(18) - 6y_1 + 10 = 0 \quad \text{---- (iii)}$$

$$45 - 6(37) + 15(35) - 20y_4 + 15(25) - 6(18) + y_1 = 0 \quad \text{---- (iv)}$$

From (iii) and (iv) we get

$$15y_4 - 6y_1 = 393$$

$$-20y_4 + y_1 = -615$$

Solving we get

$$120y_4 - 48y_1 = 3144$$

$$-120y_4 + 6y_1 = -3690$$

$$-42y_1 = -546$$

$$y_1 = 546/42 = 13$$

Put the value of y_1 in (i)

$$15y_4 - 6 \times 13 = 393$$

$$15y_4 = 471 \therefore y_4 = \frac{471}{15} = 31.4$$

Example-4 : Estimate the production of sugar for the year 1980 and 1990 from the following data :

Year	1965	1970	1975	1980	1985	1990	1995
Production (m. tonnes)	100	120	150	?	210	?	320

Solution : As five figures are known we shall assume that the fifth order difference will be zero. In the problem there are two unknown figures hence two equations will be required to determine them. They are :

$\Delta_0^5 = y_5 - 5y_4 + 10y_3 - 10y_2 + 5y_1 - y_0 = 0$ ----- (i)

$\Delta_1^5 = y_6 - 5y_5 + 10y_4 - 10y_3 + 5y_2 - y_1 = 0$ ----- (ii)

We are given

Year	1965	1970	1975	1980	1985	1990	1995
Production (m. tonnes)	100	120	150	?	210	?	320
	y_0	y_1	y_2	y_3	y_4	y_5	y_6

Put these values in (i) and (ii)

$y_5 - 5\times210 + 10y_3 - 10\times150 + 5\times120 - 100 = 0$ ---- (iii)

$320 - 5y_5 + 10\times210 - 10y_3 + 5\times150 - 120 = 0$ ---- (iv)

Solving (iii) and (iv) we get,

$y_5 + 10y_3 = 2050$

$-5y_5 - 10y_3 = 3050$

or $5y_5 + 50y_3 = 10250$

$-5y_5 - 10y_3 = -3050$

$50y_3 = 7200$

or $y_3 = 180$

Put the value of y_3 in eq. (i)

$y_5 + 10\times180 = 2050$

$y_5 = 250$

Hence the missing values corresponding to the periods 1980 and 1990 are 180 and 250 respectively.

2.1.4.2 Newton's Method for Leading Differences :

Newton's Method for Leading Differences is used when (i) the x-series or independent variable advances by equal intervals and (i) the value to be interpolated is not one of the value of the equidistant value (iii) the value to be interpolated lies in the beginning of the data.

This method is based on the principle that “just as the differences can be calculated from the y's, the y's may be built up from differences." As it is based on the differences between the various values of y, a difference table is required to be made in this method.

For example, observe the data :

Year	1951	1961	1971	1981	1991	2001
Population	y_0	y_1	y_2	y_3	y_4	y_5

If we are to estimate the population for 1957, then we apply Newton's formula. The formula for interpolation is :

$$y_x = y_0 + x\,\Delta_0^1 + \frac{x(x-1)}{L^2}\Delta_0^2 + \frac{x(x-1)(x-2)}{L^3}\Delta_0^3 + \frac{x(x-1)(x-2)(x-3)}{L^4}\Delta_0^4 + \text{-----}$$

where y_0 represents the value of y at origin.

y_x = Figure to be interpolated.

Δ 's = Differences.

The value of x in the formula is obtained as follow :

$$x = \frac{\text{The value to be interpolated - value at the origin}}{\text{Difference between two adjoining values}}$$

Steps :

(i) Denote the values of the independent variable (x) by x_0, x_1, x_2, ------ and the values of the dependent variable by y_0, y_1, ---------.
The values to be interpolated for given value of x is denoted by y_x.

(ii) Calculate $x = \frac{\text{The value to be interpolated - value at the origin}}{\text{Difference between two adjoining values}}$

(iii) Then, we compute $\Delta_0^1, \Delta_0^2, \Delta_0^3$ ---- etc.

$\Delta_0^1, \Delta_0^2, \Delta_0^3$ ------- are called 1st, 2nd and 3rd leading differences. The following is the table of differences.

TABLE SHOWING FINITE OR ADVANCING DIFFERENCES

x	**y**	**DIFFERENCES**							
		1st Difference Δ^1		**2nd Differences** Δ^2		**3rd Differences** Δ^3		**4th Differences** Δ^4	
x_0	y_0								
		$y_1 - y_0$	Δ_0^1						
x_1	y_1			$\Delta_1^1 - \Delta_0^1$	Δ_0^2				
		$y_2 - y_1$	Δ_1^1			$\Delta_1^2 - \Delta_0^2$	Δ_0^3		
x_2	y_2			$\Delta_2^1 - \Delta_1^1$	Δ_1^2		Δ_1^3	$\Delta_1^3 - \Delta_0^3$	Δ_0^4
		$y_3 - y_2$	Δ_2^1			$\Delta_2^2 - \Delta_1^2$			
x_3	y_3			$\Delta_3^1 - \Delta_2^1$	Δ_2^2				
		$y_4 - y_3$	Δ_3^1						
x_4	y_4								

In this table, the first difference in each column is called leading difference.

Example 5 : Find out expectation of life at the age 25 from the following data :

Age (in Years)	**Expectation of Life (Years)**
20	51
30	44
40	35
50	24

Solution :

TABLE-EXPECTATION OF LIFE

Age (Years) x		**Expectation of Life y**		**Difference Δ** Δ_1		Δ_2		Δ_3	
20	x_0	51	y_0						
				-7	Δ_0^1				
30	x_1	44	y_1			-2	Δ_0^2		
				-9	Δ_1^1			0	Δ_0^3
40	x_2	35	y_2			-2	Δ_1^2		
				-11	Δ_2^1				
50	x_3	24	y_3						

Newton's formula is :

$$y_x = y_0 + x\ \Delta_0^1 + \frac{x(x-1)}{1 \times 2}\Delta_0^2 + \frac{x(x-1)\,(x-2)}{1 \times 2 \times 3}\Delta_0^3 + ------$$

where y_x is the figure to be interpolated, Δ 's are the calculated algebraic differences and

$$x = \frac{\text{Year of Interpolation - Year of origin}}{\text{Distance between two x's}}$$

$$51 + .5 \times (-7) + \frac{.5(.5-1)}{1 \times 2}$$

$$\therefore x = \frac{25-20}{30-20} = .5$$

and $y_x = 51{+}.5 + \frac{.5(1-7)(.5-1)}{1 \times 2}(-2) + \frac{.5(.5-1)\,(.5-2)}{1.2.3}(0)$

= 51 - 3.5 + .25 = 47.74 years

2.1.4.3 Lagrange's Method :

This method has been developed by a famous French mathematician, Dr. Lagrange. It is a universal method. It can be applied in all cases - whether the x-series or independent variable advances by equal or unequal intervals. But in practice this method is used in those cases where the Newton's and Binomial Expansion methods are not applicable. A peculiar feature of this method is it's applicability for any data, whether our series advances by regular or irregular intervals or whether the value to be interpolated is in the beginning or in the end.

The formula is as follows :

$$f(x) = f(a)\frac{(x-b)(x-c)(x-d)}{(a-b)(a-c)(a-d)} + f(b)\frac{(x-a)(x-c)(x-d)}{(b-a)(b-c)(b-d)}$$

$$+f(c)\frac{(x-a)(x-b)(x-d)}{(c-a)(c-b)(c-d)} + f(d)\frac{(x-a)(x-b)(x-c)}{(d-a)(d-b)(d-c)}$$

where f(x) = figure to be interpolated.

a,b,c,d. etc. = given values of 'x' variable and f(a), f(b), f(c), f(d) etc. = Corresponding given values of 'y' variable.

The following example shall illustrate the Lagrange's method.

Example 6 : From the data given below find the score y if x is 5 by Langrange's Method.

x	:	3	1	2	4
y	:	3	4	1	2

Solution :

To find the score of y when x=5, by Lagrange's method

x	:	3(a),	1(b),	2(c),	4(d),	5(x)
y	:	3f(a),	4f(b),	1f(c),	2f(d),	f(x)

$$f(5) = \frac{3(5-1)(5-2)(5-4)}{(3-1)(3-2)(3-4)} + \frac{4(5-3)(5-2)(5-4)}{(1-3)(1-2)(1-4)}$$

$$+\frac{1(5-3)(5-1)(5-4)}{(2-3)(2-1)(2-4)} + \frac{2(5-3)(5-1)(5-2)}{(4-3)(4-1)(4-2)}$$

$$= \frac{3 \times 4 \times 3 \times 1}{2 \times 1 \times -1} + \frac{4 \times 2 \times 3 \times 1}{-2 \times -1 \times -3} + \frac{1 \times 2 \times 4 \times 1}{-1 \times 1 \times 2}$$

$$= +\frac{2 \times 2 \times 4 \times 3}{1 \times 3 \times 2}$$

f(5) = -18 – 4 - 4 + 8 = –10

Note : When x=5, y=-10. This appears to be an abnormal figure. This is so because the data does not fulfill the assumption of interpolation. Thus when all assumption are not fulfilled, the result through interpolation or extrapolation shall be untrustworthy.

Example 7 : The following table gives the normal weight of a baby during the first six months of life :

Age in months (x)	:	0	2	3	5	6
Weight in lbs (y)	:	5	7	8	10	12

Estimate the weight of baby at the age of 4 months.

Solution: Putting x = 4 and substituting the above values in Lagrange's formula, we get

$$y_4 = 5 \times \frac{(4-2)(4-3)(4-5)(4-6)}{(0-2)(0-3)(0-5)(0-6)}$$

$$7 \times \frac{(4-0)(4-3)(4-5)(4-6)}{(2-0)(2-3)(2-5)(2-6)} + 8 \times \frac{(4-0)(4-2)(4-5)(4-6)}{(3-0)(3-2)(3-5)(3-6)}$$

$$10 \times \frac{(4-0)(4-2)(4-3)(4-6)}{(5-0)(5-2)(5-3)(5-6)} + 12 \times \frac{(4-0)(4-2)(4-3)(4-5)}{(6-0)(6-2)(6-3)(6-5)}$$

$$= \frac{1}{9} - \frac{7}{3} + \frac{64}{9} + \frac{16}{3} - \frac{4}{3} = \frac{80}{9} = 8\frac{8}{9}$$

Estimated weight of baby at the age of 4 months is $8\frac{8}{9}$ lbs.

2.1.5 Merits and Demerits

The chief merits and demerits of the Lagrange method of interpolation may be outlined as under:

Merits:

1. This method of interpolation is applicable for all type of series.
2. This method involves a very simple procedure of interpolation, although its model appear to be very lengthy and clumsy.

Demerits :

This method proves to be very tidious and difficult, when the number of known values of the y variable happens to be large.

2.1.6 List of Questions

2.1.6.1 Short Questions

1. What is meant by Interpatations?
2. Explain the assumptions of Interpolations.
3. State Newton's Formula for Interpolation for equal intervals.

2.1.6.2 Long Questions

1. Interpolate the missing figure in the following table with the help of a suitable formula:

1961	1962	1963	1964	1965	1966	1967
1,331	1,728	2,197	?	3,375	4,096	4,913

2. Using Lagrange's Formula estimate from the following data the number of workers getting income not exceeding Rs. 26 per month:

Income not exceeding (Rs.)	15	25	30	35
No. of Workers	36	40	45	48

2.1.6 Suggested Readings

1. Croxton and Cowden : Applied General Statistics
2. Kapur, J.N. and Saxena, H.C : Mathematical Statistics
3. Chiang Alpha. C : Fundamental Methods of Mathematical Economics/

M.A. (ECONOMICS) PART-I (Semester-II) **PAPER-III (BASIC QUANTITATIVE METHODS)**

LESSON NO. 2.2 **AUTHOR : DR. RAJINDER SINGH**

INDEX NUMBERS

Structure

2.2.1 Introduction

An index number intends to measure the combined fluctuations in a group of series. We often come across statements like. "The general level of prices has registered an increase of (say) two percent". When we talk of general price level it is obvious that we are referring to the prices of all those commodities that are exchanged for money in a given country. Now, some of these prices may be going up while others may be coming down. Again, the rate of increase or decrease in different prices may be different. So to express the change of this nature by a single representative figure, as in the statement above, we have to make use of a statistical device which may enable us to do so. This device consists of constructing index numbers.

Thus we can define an index number as "a device of combining the variation in a group of related variables over a period of time, or at different places, with a view to obtain a figure that represents the net result of change in the constituent variable". These variables may be the price of commodities, the physical quantity of goods produced, marketed, consumed or such concepts as "intelligence", "beauty" or "efficiency".

Index numbers are used to measure change, in industrial output, fluctuations in the level of business or variations in the sizes of agricultural output, etc.

Many aspects of modern business are described by the use of index number. Both government and private agencies are making increasing efforts to the construction of index numbers as aids in management in the interpretation of changes in general economic life.

Generally, the index numbers are expressed in percentage with respect to a base and the index for base is taken as 100. Thus the statement, "The wholesale price index for 1971 with 1962 as the base is 125, there is a net increase in the prices of wholesale commodities to the extent of 25 percent.

2.2.2 Types of Index Numbers

An examination of the financial section of a newspaper will reveal many different index numbers which describe various aspect of business and economics. These index numbers may be classified as the (1) Price indices (2) Quantity indices and (3) Value indices.

Price Indices

The oldest and best known indices are those dealing with prices. Prices have been of great interest because these are of considerable importance in an economy. Well known examples of these type of indices are consumer price index, wholesale price index etc. The necessary data for price index numbers arise from the exchange of commodities (1) at different stages of production, raw material, semi-finished goods, and completely fabricated products, (2) at several level of distribution whole sale, retail and (3) for a variety of group of items-consumer goods, producer goods imports and exports, stock and bonds, durable and non-durable goods etc.

Quantity Indices

Quantity indices measure the physical volume of production construction, or employment. They are computed for (1) industry in general (2) specific industry and (3) specific operation or stages of production or distribution.

Because of the nature of the data quantity-index numbers are frequently less reliable than price index numbers. This is because the business records are designed to include chiefly those aspects of business which could be expressed in momentary units and consequently data in physical units for certain period of time are difficult to obtain. A well known example of this type of index is index of industrial production.

Value Indices

A value index is one that represents a comparison of the total value of production or sales at two period or at two different places, without regard to weather the observed difference is a result of differences in quantity or price or both thus an index number that compares the level of retail sales in two time periods is an example of value index.

2.2.3 Construction of Index Number

To illustrate the problem faced in the construction of index numbers, we shall explain the construction of the most commonly used index numbers viz., the price index number. However, the same procedure can be applied for the construction of other indices.

2.2.3.1 Simple Index Number

A simple index number is constructed from a single series of data which either extends over a period or simultaneously represents several different locations. In constructing such an index number, one particular period or place is selected as the base and the index for this base is taken as 100. The other items in the series are then expressed as percentage of this base. For example let P_0 be the price of a commodity in the base period and let P_1 be the price in the Ist or current period. Then the simple index in the current period is

$$P_{01} = \frac{P_1}{P_0} \times 100$$

The ratio $\frac{P_1}{P_0}$ is called the price relative.

Illustration 1

Let P_1 be the price of wheat in 1980 (average) and P_0 be price of wheat in 1960. For example if P_1 = 120 and P_0 = Rs.60, then the Price index of wheat for 1980 with respect to base (1960) is.

$$P_{01} = \frac{P_1}{P_0} \times 100 = \frac{120}{60} \times 100 = 200$$

This indicates that price of wheat in 1980 is 200 percent of the price of wheat in 1960.

2.2.3.2 Composite Index Numbers

Most of the index numbers in common use are composite. In this case we have more than one series and our interest is to find a common index number for all these series. Two basic method of constructing these index numbers are (1) Average of relatives index (2) Aggregative index.

Simple average of price relative index is usually found by taking arithmetic mean geometric mean and harmonic mean, and less frequently by taking median or mode of relatives. So using arithmetic mean, we get :

$$P_{01} = \frac{1}{K} \Sigma \left(\frac{P_1}{P_0} \right) \text{x } 100$$

Where P_0 is the price of commodity in the base period.

P_1 is the price of commodity in the current period.

K is the number of commodities.

P_{01} is index number for the current period, taking o as base period.

Illustration 2

Let us take the example of food grains price index. Let the prices of wheat, rice and maize increased in 1980 by Rs.120, Rs.200 and Rs.100 respectively compared to Rs. 60, Rs. 100 and Rs. 70 in 1960. Construct price index for 1980 compared to 1960 as base:

$$P_{o1} = \sum \left(\frac{P_1}{P_0} \times 100 \right) \quad \text{Where } k = 3$$

P_1	120	200	100
P_0	60	100	70

$$P_{01} = \frac{1}{3}\Sigma\left(\frac{120}{60} + \frac{200}{100} + \frac{100}{70}\right) \times 100$$

$$= 1.80 \text{ x } 100 = 180$$

Which means that the prices of these food grains in 1980 are 80 percent of the prices in 1960. This index is a composite or combined index for the three food grains.

Similarly, if we use geometric mean, the formula will be

$$In = \left(\frac{pn1}{P_{01}} \times \frac{pn2}{P_{02}} \times \frac{pnk}{P_{0k}}\right)^{\frac{Ink}{nk}}$$

Illustration 3

Taking data of illustration 2. If we use geometric mean :

$$In = \frac{1}{-}\left\{\frac{120}{60} + \frac{200}{100} + \frac{100}{70}\right\}^{\frac{1}{3}} \times 100$$

X X

$$= 178.77$$

We can get a single aggregative index by comparing the simple aggregate of actual prices for current period with that for the base period.

Symbolically :

$$In = \frac{\sum_{i=1}^{n} P_{ni}}{\sum_{i=1}^{n} P_{oi}} \times 100$$

Illustration 4

Let us consider four commodities of same group like oilseeds, pulses etc. Let the price of four oilseed crops, A, B, C and D in nth year be Rs.200, Rs.230, Rs.240 and Rs.170 and in 0th year (base year) Rs.70, Rs.100, Rs.120 and Rs.100 respectively. Then the Price index for nth year w.r.t. the base will be

$$In = \frac{\Sigma pni}{\Sigma poi} \times 100$$

$$= \frac{200 + 230 + 240 + 170}{(70 + 100 + 120 + 100)} \times 100$$

$$= \frac{840}{390} \times 100 = 215.4$$

We are to multiply each formula by 100 to express the index in percentage form. However, this factor is generally omitted from the formula and is introduced at the last stage.

2.2.4 Weighted Index Numbers

All the commodities included in the index number may not be of equal importance. For example, in constructing a wholesale price index for India, rice should have greater importance than tobacco. So we must give appropriate weights to the different commodities included in the index number. These weights should truly reflect the importance of each commodity.

If wi is the weight attached to the *i*th commodity, then the weighted arithmetic mean of price relative is given by :

$$P_{01} = \frac{\sum_{i=1}^{n} \left(\frac{P_1}{P_0}\right) wi}{\Sigma\ wi} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(i)}$$

and the weighted geometric mean by

$$P_{01} = \left[\Sigma \left(\frac{P_1}{P_0}\right) w_i\right] \frac{1}{\Sigma\ wi}$$

and the weighted harmonic mean by :

$$P_{01} = \frac{\Sigma \frac{\Sigma\ wi}{P_0}}{P_1} wi$$

Similarly if w_i is the weight attached to i

aggregative index is given by :

$$P_{01} = \frac{\Sigma P_1\ wi}{\Sigma P_0\ wi}$$

Illustration 5

To illustrate the above formula, let us consider the case of price index of food grains for, some year 'i' with respect to year 'o'. Let the food grains be wheat, rice, maize and barley. Let the prices of these food grains in nth year be Rs.120, 200, 100 and 70 and in year 'o' be Rs.80, 150, 70 and 40 respectively. Let the weights to be given to these crops according to their economic importance be 0.3, 0.4, 0.2 and 0.1 respectively. We are clearly considering rice to be most important followed by wheat. So in our case :

P_1	P_0	W
120	80	0.3
200	150	0.4
100	70	0.2
70	40	0.1

Then weighted index number

$$P_{01} = \frac{\sum P_1 w_i}{\sum P_0 w_i} \times 100$$

$$= \frac{120 \times 0.3 + 200 \times 0.4 + 100 \times 0.2 + 70 \times 0.1}{80 \times 0.3 + 150 \times 0.4 + 70 \times 0.2 + 40 \times 0.1} \times 100$$

$$= \frac{36 + 80 + 20 + 7}{24 + 60 + 14 + 4} \times 100$$

= 140.20

The Greek letter (Capital 'pi') is used as sign of multiplication.

Discussed below are some particular weighted index numbers of prices. If in (i) wi is taken q_0. The base period quantities, then we get :

$$P_{01} = \frac{\Sigma P_1\, q_0}{\Sigma P_0\, q_0} \times 100$$

Which is known as Laspeyre's formula. This formula is also the same as (i) with w_i equal to q_0.

Again taking q_1, the current period quantity, as wi in (i) we get

$$P_{01} = \frac{\Sigma P_1\, q_1}{\Sigma P_0\, q_1} \times 100$$

Which is known as Paashe's formula. This formula is also the same as (6) with w_i equal to q_1 the current period values.

Taking wi as $\frac{(q_1 + q_0)}{2}$ the average of current period and base period in

(i) we get :

$$P_{01} = \frac{\Sigma P_1 (q_1 + q_0)}{\Sigma P_0 (q_1 + q_0)}$$

Which is known as Marshall-Edgeworth formula.

Another commonly used index number is Fisher's ideal index number which is obtained by taking the geometric mean of **Paasche's** and **Laspeyr's** formula.

$$P_{01} = \sqrt{\frac{\Sigma P_1\, q_0}{\Sigma P_0\, q_0} \times \frac{\Sigma P_1\, q_1}{\Sigma P_0\, q_1}}$$

This is called ideal because it satisfies certain tests of consistency discussed in the next lesson.

Illustration 6

Construct price index numbers from the following data taking 1980 as base. Use

(i) **Laspeyre's Method** (ii) **Paasche's Method**

(iii) Bowley's Method (iv) Fisher's Method

Commodities	**1980**		**1981**	
	Price	**Value**	**Price**	**Value**
A	10	140	12	144
B	12	144	15	105
C	15	300	16	96
D	20	300	25	200
E	25	100	30	90

Sol. In the data price and value are given both for base year. Quantity can be calculated by dividing the value by corresponding price in each of base and current year. Value in 1980 is demoted by P_0q_0 and in 1981 by P_1q_1

Construction of Price Index Numbers

Commodity	1980			1981				
	Price (P_0)	Value (P_0q_0)	Quantity (q_0)	Price (P_1)	Value (P_1q_1)	Quantity q_1	(P_0q_1)	(P_1q_0)
A	10	140	14	12	144	12	120	168
B	12	144	12	15	105	7	84	180
C	15	300	20	16	96	6	90	320
D	20	300	15	25	200	8	160	375
E	25	100	4	30	90	3	75	120
		$\sum P_0q_0$ =984			$\sum P_1q_1$ =635		$\sum P_0q_1$ =529	$\sum P_1q_0$ =1163

Laspeyr's Index $P_{01} = \frac{\sum P_1q_0}{\sum P_0q_0} \times 100 = \frac{1163}{984} \times 100 = 118.19$

Paasche's Index $P_{01} = \frac{\sum P_1q_1}{\sum P_0q_1} \times 100$

$$= \frac{635}{529} \times 100 = 120.03$$

Bowley's Index $P_{01} = \frac{L+P}{2} = \frac{118.19+120.03}{2}$

$$= \frac{238.22}{2} = 119.11$$

Fisher's Ideal Index $P_{01} = \sqrt{L \times P}$

$$= \sqrt{(118.19)(120.03)} = \sqrt{14186.34}$$

$$= 119.10$$

Cost of Living Index Number

A cost of living index number measures the relative changes in the amount of money required to yield equivalent satisfaction under two different situations. The cost of living index always relates to a designated group of people e.g. the agricultural labourers, industrial workers etc.

This index is constructed by comparing the consumer prices for the two situations. This index covers food, clothing, fuel and lighting, house rent and miscellaneous group of items. For each of these groups a separate index number is calculated by weighted average of price relatives of the different items of the group, the weight being proportional to their consumption expenditure. For each item there will be number of price quotations covering different brands and marks. The price relative of an item is the simple average of the price relatives for different quotations of the item. The general index is then found by taking the weighted average of the group indices, the weights being proportional to the consumption expenditure on the different groups.

Illustration 7

Suppose it is required to determine the cost of living index number for the working class people of Bombay city for July 1956, with the year 1939 as base from the following group indices and weights.

Group	Weight W	Index of Prices I	I W
Food	53	469.2	24867.6
Fuel & lighting	8	301.2	2409.6
Clothing	9	407.7	3669.3
House Rent	14	106.3	1488.2
Miscellaneous	16	346.7	5547.2
	$\sum W=100$		$\sum IW = 37981.9$

The general cost of living index is : $\frac{\sum IW}{\sum W}$

P_{01} = (53 x 469.2 + 8 x 301.2 + 9 x 407.7 + 14 x 106.3 + 16 x 346.7) /100 = 379.8

This means that in July 1956 a worker in Bombay required 379.8 times more money as compared to that required in 1939 to have the same satisfaction in the two situations.

Example 8 : Construct the cost of living index for the following data by using (i) aggregate expenditure method (ii) Family budget method. Treat 1975 as base year

Com	Quantity in 1975	Price in 1975	Price in 1980
Rice	4	3.00	6.0
Wheat	5	1.50	2.0
Gram	3	1.80	3.6
Ghee	1	10.00	9.0
Sugar	5	3.50	5.0

Sol.

(a) Aggregate Expenditure Method :

Com.	q_0	P_0	P_1	P_1q_0	P_0q_0
Rice	4	3.00	6.0	24.00	12.00
Wheat	5	1.50	2.0	10.00	7.50
Gram	3	1.80	3.6	10.80	5.40
Ghee	1	10.00	9.0	9.00	10.00
Sugar	5	3.50	5.0	25.00	17.50
				78.80	52.40

Index No. $= \dfrac{\Sigma P_1 q_0}{\Sigma P_0 q_0} \times 100$ or $= \dfrac{78.80}{52.40} \times 100$ = 150.38

(b) Family Budget Method :

Com.	$\mathbf{P_o}$	$\mathbf{P_1}$	$\mathbf{q_o}$	$1 = \dfrac{P_1}{P_0} \times 100$	$\mathbf{V=P_0q_0}$	**I.V.**
Rice	3.0	6.00	4	200.00	12.00	240.00
Wheat	1.50	2.00	5	133.33	7.50	999.8
Gram	1.80	3.60	3	200.00	5.40	1080.00
Ghee	10.0	9.00	1	90.00	10.00	900.00
Sugar	3.50	5.00	5	142.86	17.50	2500.50
					52.40	7880.30

$$\text{Index No.} = \frac{\Sigma IV}{\Sigma V} \text{ or } \frac{7880.03}{52.40} = 150.38$$

Thus we find that result by both the methods is same.

WEIGHTED INDEX NUMBER

Problems of Selecting Suitable Weights

As discussed already, a weighted average of relatives is normally required in place of simple averages. The problem of how the weights should be determined has been discussed here. Any system of weights must have two requirements.

1. The numbers used as weights must measure the comparative importance of the different relatives to be averaged.
2. There must be numbers that can be properly added, since the formula for weighted average involves dividing by the sum of the weights.

The weights should be computed from the actual data that measure the relative importance of the different items; although if we do not have any data available, weights might be estimated. This is because the basic reason for using statistical methods is to replace opinion by objectively derived facts of computed weights should always be used if any kind of information can be obtained to measure the relative importance of the items included in the index.

Many different kinds of data may represent the basis for computation of weights. Two most commonly used weights are (i) quantity and (ii) value. A quantity weights means the quantity of a commodity produced, consumed or distributed. A value weight is the product of price per unit and quantity produced, consumed or distributed.

Quantity weights are to be used if we adopt the method of aggregate because the product of price and quantity will always be in rupees. On the other hand, if the method used is that of price relative, quantity can be used as weights for the product of price relatives and quantities expressed in different units would yield products expressed in different units. For example, Kilograms multiplied by price relatives would give kilograms. Tons multiplied by price relatives would give tons and so on. So the product of price relatives and quantities cannot be added and averaged. To overcome this difficulty the price relatives are weighted by value figures and the product is always in rupees only.

The next question is that of the time the value or quantity which is to be used as weights. These weights may be quantity or value for the base year or for the current year or it may be sum or average of the two. Sometimes the quantities or value of a typical year or the average to several years which are thought as typical may also be used as weights. The value weights may also be obtained by the product of price in one time period and quantity in any other time period.

2.2.6 Methods of Selecting a Base

There are two methods of selecting a base :

(i) Fixed Base Method : In this method one particular year or an average of a few years may be taken as base (=100) and prices in the subsequent years may be expressed as relatives of that in the base period.

(ii) Chain Base Method : In this method, there is no fixed base Prices in one year serve as base for the price relatives in the succeeding year. So prices in 1977 will serve as "base for the price relatives in 1978, 1978 as base for 1979 so on. But in this method, long period comparisons cannot be made."

The following steps are used in the construction of chain indices :

(i) Find Link relative by applying the formula :

$$\text{Link Relative} = \frac{\text{Current Year's Price}}{\text{Previous Year's Price}} \text{ x } 100$$

(ii) Find the average of Link Relative for each other :

$$\text{Average of Link Relative} = \frac{\text{Total of Link Relatives}}{\text{No. of Link Relatives}}$$

(iii) The average of link relatives are chained together to a common base. These resulting indices are called chain indices. The formula is :

$$\text{Chain index for current year} = \frac{\text{Average Link relatives of Current year xChain index of previous year}}{100}$$

Example 1 :

From the following data relating to average price of the groups of commodities given in rupees per unit, find chain base indices with 1980 as base.

Group	**1980**	**1981**	**1982**	**1983**	**1984**
A	3	4	5	6	6
B	5	8	10	12	15
C	8	10	15	18	20

Sol : Calculation of Chain Indices :
Using the three steps mentioned above we have

Group	1980	Link Relative Price	1981	Link Relative Price	1982	Link Relative Price	1983	Link Relative Price	1984	Link Relative Price
A	3	100	4	133.33	5	125	6	120	6	100
B	5	100	8	160.00	10	125	12	120	15	125
C	8	100	10	125.00	15	150	18	120	20	111.11
Total of Link Relatives		300		418.33		400		360		336.11
Average of Link Relatives		100		139.44		133.33		120		112.04
Chain Indices		100		$\frac{100 \times 139.44}{100}$ = 139.44		$\frac{139.44 \times 133.33}{100}$ = 185.32		$\frac{185.32 \times 120}{100}$ = 223.1		$\frac{223.1 \times 112.04}{100}$ = 249.05

Conversion of Fixed Base Index into Chain Base Index

To find chain base index, the fixed base index being given, we take the first index as 100 and then apply the following formula.

$$\text{Current year Chain base Index} = \frac{\text{Current year fixed base number}}{\text{Previous year fixed base number}} \times 100$$

Example 2 :

Following data relate to Fixed base index number. Convert these into chain base index numbers :

Year	1980	1981	1982	1983	1984
Index Number	100	112	120	150	165

Solution

Conversion of fixed base indices into chain Base indices :

Year	Fixed Base Index Numbers	Chain Base Index Numbers
1980	100	100
1981	112	$\frac{112}{100}\times 100 = 112$
1982	120	$\frac{120}{112}\times 100 = 107.14$
1983	150	$\frac{150}{120}\times 100 = 125$
1984	165	$\frac{165}{150}\times 100 = 110$

Conversion from Chain Base Indices to Fixed Base Indices

Formula for Changing Chain Base Indices to Fixed Base Indices :

$$\text{Current year's fixed index} = \frac{\text{Current year chain index X Previous year fixed index}}{100}$$

Conversion of chain base index number into fixed base index number

Year	Chain Base Index Number	Fixed Base Index Number
1980	140	100
1981	155	$\frac{155}{100} \times 100 = 155$
1982	170	$\frac{170}{100} \times 155 = 263.5$
1983	180	$\frac{180}{100} \times 263.5 = 474.3$

10.7 Tests for Consistency of Index Number

There are four mathematical tests of consistency that an index number is supposed to meet. They are (i) Time reversal test, (ii) Factor reversal test and (iii) Circular test, (iv) Unit test.

(i) Time Reversal Test

According to this, for any formula to be accurate it should be time consistent, i.e. we should get the picture of the change in the price level if the base period and the current period are interchanged. Thus if the price is doubled from 1938 price and 1938 price is 50 percent of the 1959 price. Thus if the index number for two year secured by the same method but with base reversed are reciprocals of each other, then that index is said to meet the time reversal test.

Symbolically

$P_{01} \text{ x } P_{10} = 1$

(ii) Factor Reversal Test

The value of all the commodities included in an index is the sum of values for the various commodities. Thus the ratio of the values for the two years give the value index.

$$P_{01} \times Q_{01} = \frac{\Sigma P_1 q_1}{\Sigma P_0 q_0}$$

According to this test, if the price and quantity factors in the price index formula (b) be interchanged so that a quantity index formula (a) is obtained, then the product of these two indices should give the value index.

Fisher's Ideal Formula discussed in the previous lesson, is the only price index which satisfies this test, for this formula :

$$P_{01} \times Q_{01} = \sqrt{\frac{\Sigma P_1 q_0}{\Sigma P_0 q_0} \times \frac{\Sigma P_1 q_1}{\Sigma P_0 q_1}} \times \sqrt{\frac{\Sigma q_1 P_0}{\Sigma q_0 P_0} \times \frac{\Sigma q_1 P_1}{\Sigma q_0 P_1}}$$

$$= \sqrt{\frac{\Sigma P_1 q_0}{\Sigma P_0 q_0} \times \frac{\Sigma q_1 P_1}{\Sigma q_1 P_0} \times \frac{\Sigma q_1 P_0}{\Sigma q_0 P_0} \times \frac{\Sigma P_1 q_1}{\Sigma P_1 q_0}}$$

$$= \sqrt{\frac{(\Sigma P_1 q_1)^2}{(\Sigma P_0 q_0)^2}} = \frac{\Sigma P_1 q_1}{\Sigma P_0 q_0}$$

Clearly $P_{01} \times Q_{01} = \dfrac{\Sigma P_1 \, q_1}{\Sigma P_0 \, q_0}$

(iii) Circular Test

In case of three given years a method is said to satisfy the circular test if :

$P_{12} \text{ x } P_{21} \text{ x } P_{31} = 1$

Where P_{12} is the index number of 2nd year with Ist year as base.

P_{21} is index number of Ist year with 2nd year as base.

and P_{31} is index number of Ist year with 3rd year as base.

It can be shown that Fisher's formula does not satisfy this test.

(iv) Unit Test

The unit test requires that the formula for constructing an index should be independent of the units in which, or for which prices and quantities are quoted. Except for the simple unweighted aggregative index all other formula satisfy this test.

Example 3 :

With the help of data given below how Fisher's Ideal Index satisfies the factor reversal test :

	1971-72		1979-80	
Commodity	**Price**	**Qty.**	**Price**	**Qty.**
Rice	56	71	50	26
Barley	32	107	30	83
Maize	41	62	28	48

	Base Year	1971-72	Current Year	1979-80				
Commodity	P_0	q_0	P_1	q_1	P_0q_0	P_0q_1	P_1q_0	P_1q_1
Rice	56	71	50	26	3976	1456	3550	1300
Barley	32	107	30	83	3424	2656	3210	2490
Maize	41	62	28	48	2542	1968	1736	1344
					9942	6080	8496	5134

Now for factor Reversal test to be satisfied :

$$P_{01} \times Q_{01} = \sqrt{\frac{\Sigma P_1 q_0}{\Sigma P_0 q_0} \times \frac{\Sigma P_1 q_1}{\Sigma P_0 q_1} \times \frac{\Sigma P_0 q_1}{\Sigma P_1 q_1} \times \frac{\Sigma P_o q_o}{\Sigma P_1 q_0}}$$

$$= \sqrt{\frac{8496}{9942} \times \frac{5134}{6080} \times \frac{6080}{9942} \times \frac{5134}{8496}}$$

$$= \sqrt{\frac{(5134)^2}{(9942)^2}}$$

$$= \frac{5134}{9942} = \frac{\Sigma P_1 q_1}{\Sigma P_0 q_0}$$

Thus Fisher's Ideal Index satisfies the Factor Reversal Test :

Example 4 :

From the following data show how Fisher's Index number satisfies the Time Reversal Test.

Commodity	Base Year Price	Base Year Qts.	Current Year Price	Current Year Qts.
A	20	4	24	5
B	15	5	24	3
C	30	2	12	5
D	50	1	40	2

Commodity	Base Yr. Price per unit (P_0)	Base Yr. qt. (q_0)	Current Yr. price per unit (P_1)	Current Yr. qt. (q_1)	P_0q_0	P_0q_1	P_1q_0	P_1q_1
A	20	4	24	5	80	100	96	120
B	15	5	24	3	75	45	120	72
C	30	2	12	5	60	150	24	60
D	50	1	40	2	50	100	40	80
					265	395	280	332

Now for Time Reversal test to be satisfied.

$$P_{01} \times P_{10} = \sqrt{\frac{\Sigma P_1 q_0}{\Sigma P_0 q_0} \times \frac{\Sigma P_1 q_1}{\Sigma P_0 q_1} \times \frac{\Sigma P_0 q_1}{\Sigma P_1 q_1} \times \frac{\Sigma P_0 q_0}{\Sigma P_1 q_0}}$$

$$\text{or } P_{01} \times P_{10} = \sqrt{\frac{280}{265} \times \frac{332}{395} \times \frac{265}{280} \times \frac{395}{332}} = 1$$

Thus Fisher's Ideal formula satisfies Time Reversal Test.

Example 5 :

Commodity	Base Year Price (Rs.)	Base Year Qty. (Kgs.)	Current Year Price (Rs.)	Current Year Qty. (Kgs.)
A	6	50	10	56
B	2	100	2	120
C	4	60	6	60
D	10	30	12	24
E	8	40	12	36

Show that Fisher's ideal formula satisfies factor and time reversal test.

Commodity	P_0	q_0	P_1	q_1	P_0q_0	P_0q_1	P_1q_0	P_1q_1
A	6	50	10	56	300	336	500	560
B	2	100	2	120	200	240	200	240
C	4	60	6	60	240	240	360	360
D	10	30	12	24	300	240	360	288
E	8	40	12	36	320	288	480	432
					1360	1344	1900	1880

$$P_{01} \times P_{10} = \sqrt{\frac{\Sigma P_1q_0}{\Sigma P_0q_0} \times \frac{\Sigma P_1q_1}{\Sigma P_0q_1} \times \frac{\Sigma P_0q_1}{\Sigma P_1q_1} \times \frac{\Sigma P_0q_0}{\Sigma P_1q_0}}$$

or $$P_{01} \times P_{10} = \sqrt{\frac{1900}{1360} \times \frac{1880}{1344} \times \frac{1344}{1880} \times \frac{1360}{1900}}$$

$$P_{01} \times P_{10} = 1$$

Hence Fisher's Formula satisfies the Time Reversal Test.

Now to satisfy Factor Reversal Test :

$$P_{01} \times q_{01} = \sqrt{\frac{\Sigma P_1q_0}{\Sigma P_0q_0} \times \frac{\Sigma P_1q_1}{\Sigma P_0q_1} \times \frac{\Sigma P_0q_1}{\Sigma P_0q_0} \times \frac{\Sigma P_1q_1}{\Sigma P_1q_0}}$$

$$= \sqrt{\frac{1900}{1360} \times \frac{1880}{1344} \times \frac{1344}{1360} \times \frac{1880}{1900}}$$

$$P_{01} \times q_{01} = \frac{1880}{1360} = \frac{\Sigma P_1q_1}{\Sigma P_0q_0}$$

Thus Fisher's Formula satisfies the Factor Reversal Test.

Example 6

With the help of following data show that Fisher's formula does not satisfy circular test.

	1981				1982						1983			
Comm.	P_1q_1	P_1q_1	P_1q_2	P_2q_1	P_2	q_2	P_2q_2	P_3q_1	P_1q_3	P_2q_3	P_3	q_3	P_3q_3	P_3q_2
A	2 5	10	2	20	4	1	4	25	4	8	5	2	10	5
B	3 4	12	6	20	5	2	10	24	3	5	6	1	6	12
C	4 3	12	20	18	6	5	30	24	16	24	8	4	32	46
		34	28	58			44	73	23	37			48	57

Using Fisher's ideal index number formula

$$P_{12} = \sqrt{\frac{\Sigma P_2q_1}{\Sigma P_1q_1} \times \frac{\Sigma P_2q_2}{\Sigma P_1q_2}}$$

$$= \sqrt{\frac{58}{34} \times \frac{44}{28}} = \sqrt{\frac{\mathbf{319}}{\mathbf{119}}}$$

Again $$P_{23} = \sqrt{\frac{\Sigma P_3q_2}{\Sigma P_2q_2} \times \frac{\Sigma P_3q_3}{\Sigma P_2q_3}}$$

$$= \sqrt{\frac{57}{44} \times \frac{48}{37}} = \sqrt{\frac{684}{407}}$$

and $$P_{31} = \sqrt{\frac{\Sigma P_1q_3}{\Sigma P_3q_3} \times \frac{\Sigma P_1q_1}{\Sigma P_3q_1}}$$

$$= \sqrt{\frac{23}{48} \times \frac{34}{73}} = \sqrt{\frac{391}{1752}}$$

P12 x P23 x P31

$$= \sqrt{\frac{319}{119}} \times \sqrt{\frac{684}{407}} \times \sqrt{\frac{391}{1752}} \quad =$$

So Fisher's formula does not satisfy circular test.

2.2.8 BOOKS FOR STUDY

1. Croxton and Cowden : Applied General Statistics
2. Elhance, D.N. : Fundamentals of Statistics
3. Mills, F.C. : Statistical Methods

2.2.9 List of Questions

2.2.9.1 Short Questions:

1. What is an index number?
2. Index numbers are economic barometers" Comment.
3. Distinguish between Laspeyre's and Paasche's index number formulae.
4. Why is Fisher's Index called Ideal?
5. What is circular test?

2.2.9.2 Long Questions

1. Calculate the Fisher's Index from the data given below base

	1968		1969	
Commodities	P	Q	P	Q
A	6.4	93	4.6	49
B	11.9	146	10.4	89
C	13.2	111	12.8	68
D	3.8	32	4.0	26

2. From the Fixed base index numbers given below construct Chain Base Index Number:

Year	1981	1982	1983	1984	1985	1986
Fixed Assests Index Number	94	98	102	95	98	100

3. What are index numbers? Examine the various problems involved in the constructions of index numbers?

LESSON NO.2.3 AUTHOR : DR. S.K. SRIVASTAVA

ANALYSIS OF TIME SERIES

Structure

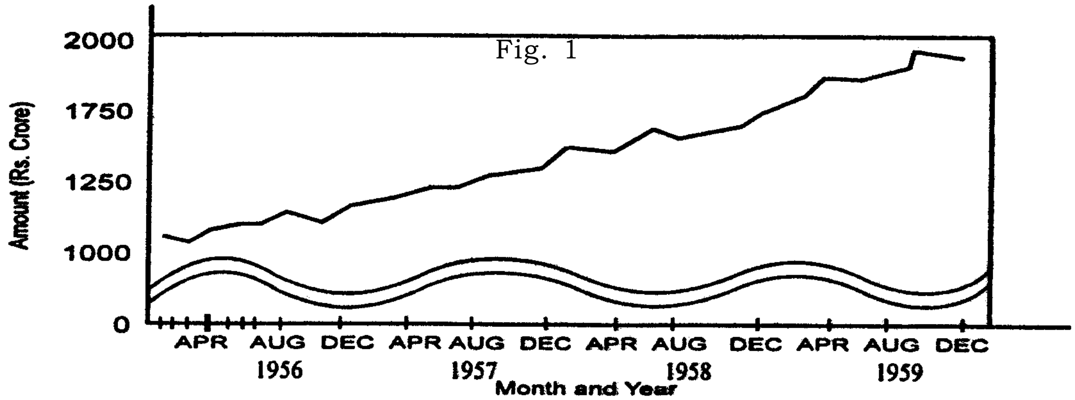
Introduction

Time Series analysis is mainly concerned with the development of technique for analysing the changes that occur in a series of data with the passage of time. Such an analysis differs from the ordinary statistical analysis because of the difference in the data and the object. In most statistical analysis the sample through which we study the population is assumed to have arisen from the stationary population. But in time series each observation is in general form because with the passage of time several changes might be taking place which have far reaching effect on the series under study.

An important task before economists and businessmen is to make estimate for future. For example, an economist is interested in estimating the population of a country in the coming year and also in following years so that proper planning can be carried out with regard to supply of food, jobs for the people etc. Similarly a business- man is interested in estimating his sales in the coming years so that he could plan and adjust his production accordingly and avoid the possibility of either unsold stocks or inadequate production to meet the demand. Such estimates pertaining to future will naturally be based on data of the past. In this connection one usually deals with statistical data which are recorded at successive intervals of time. Such data are called Time Series. Examples are price of a commodity at different points of time, daily sales in a departmental store, bank deposits or bank, clearings, weekly production of ice cream in a factory etc.

A time series is, thus, a series of values taken to be a variable at different points of time. The value assumed by a variable at 't' will be denoted by u(t) or Ut. Here the time sequence is of prime importance. Although the variable 't' will always be thought of as a time parameter, the theory has obvious application to variation in a space. For example, if we consider the variation in thickness of a cotton thread along its length I. The variable I may be interpreted in a manner analogous to it.

In a time series data, the values of the variable under consideration changes from time to time. These fluctuations are affected not by a single force but are due to the net effect of multiplicity of forces pulling it up and down and, if these forces were in equilibrium the series would remain constant. For example, the retail price of a particular commodity is influenced by a number of factors viz., the crop yield which further depends upon factors, such as transport facilities, government taxes, consumer demand etc.

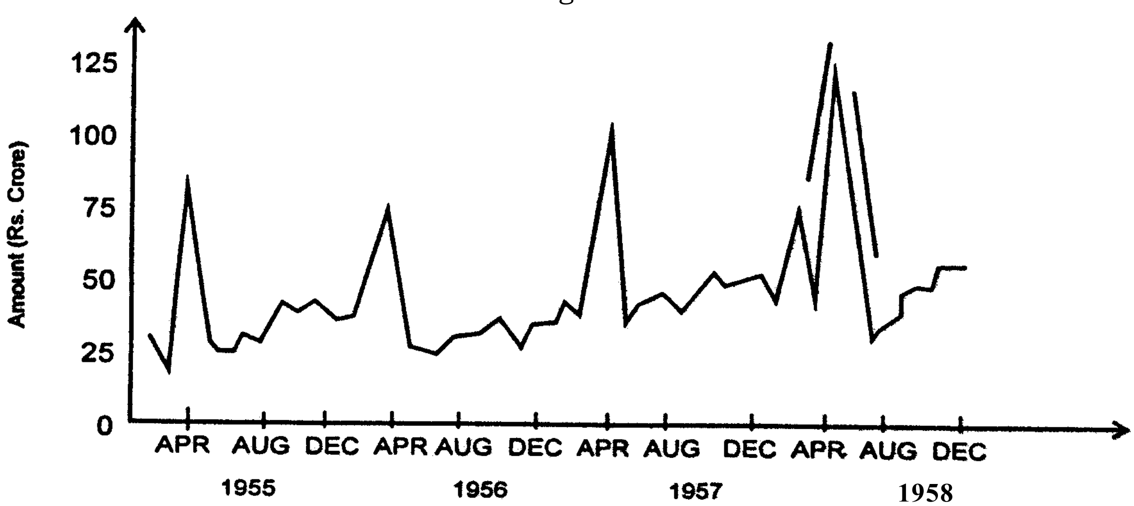


Deposit liabilities of schedule banks in India

Some examples of Time Series

We now give a few examples of time series. In examining any particular time series it's graph helps us to obtain a general picture of its behaviour. For illustration, we give four graphs of four different types of time series data in Fig. 1 to 4.

Fig. 2



Revenue expenditure and Defence Drawings

Fig.1 shows that graph of the data of monthly deposit liabilities of scheduled banks in India for the year 1956-59. Fig.2. depicts the graph of monthly revenue expenditure and defence drawing by Govt. of India for the year 1955-58. Fig.3 is the graph of an artificial time series data which gives the Welfer's sunspot numbers for the year 1851-1900. And fig.4 shows the graph of the annual crude birth rates (reported quarterly) of cattle in Great Britain for the years 1940-1945.

Such graphical representation of time series reveal the pattern of changes over time. A series which exhibits no change during the period under consideration will give a horizontal line. However all actual time series show continual changes over time giving us an overall impression of haphazard movement. A critical observation of the graphs of these series will, however, reveal that changes are not totally haphazard and a part of it, at least, can be accounted for. The part which can be accounted for is the systematic one and one remaining part, the unsystematic or irregular. The analysis of time series helps us in depicting the effect of various forces that result in the continual change in the variable.

Fig. 3

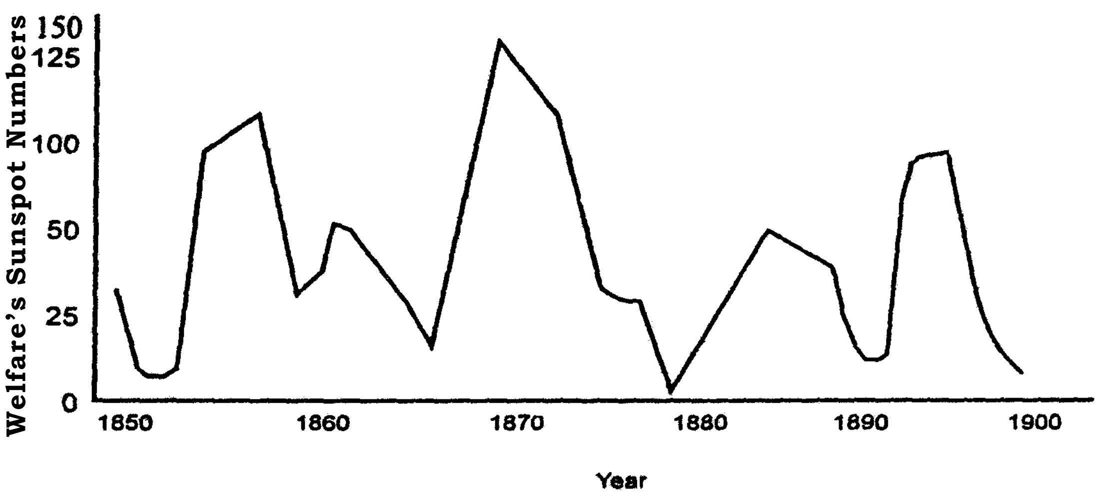
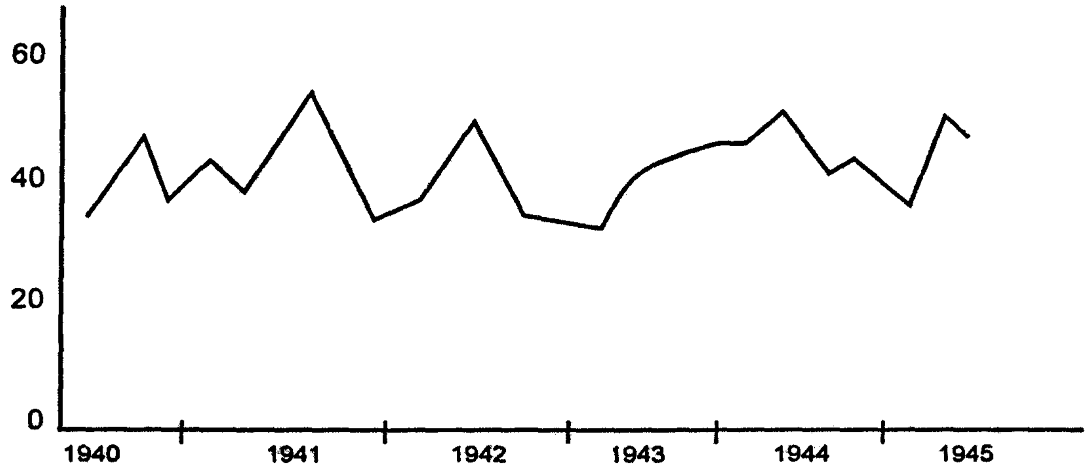


Fig. No.3

Fig. 4



Crude birth rates of cattle in Great Britain

2.3.2 Components of Time Series

The forces at work affecting a time series can be broadly classified into the following four categories of variation :

1. Secular Trend
2. Seasonal Variations
3. Cyclical Variations
4. Random or Irregular movements.

The value of the time series u_t at any time 't' regarded as the resultant of the impact of the above components of variations. In the following sequences we shall briefly explain them one by one.

1. Secular Trend

By secular trend or simply trend we mean the general tendency of the data to increase or decrease during long period of time. It need not have any particular shape, but the idea of trend implies a persistent movement in one direction or the other. For example, an upward tendency would be seen in data pertaining to population, agriculture, production, currency in circulation, prices etc., while a downward tendency would be noticed in data of birth and deaths, epidemics etc. as a result of advancement in medical sciences, better medical facilities, literacy etc.

These are all parts of trends, some series increase slowly and some increase fast, other decrease at varying rates and some after a period of decline reverse themselves and enter a period of growth. Figure 1 illustrates a series exhibiting an upward trend, other systematic components being almost absent.

Remarks

1. Trend is the result of those forces which are either constant or change very gradually over a long period of time. These would be, for example, due to change in the size of population or tastes and habits of people in a specified population or due to continuously improving technology in the manufacture of certain products etc.

2. It is not necessary that the rise or fall must continue in the same direction throughout the period of the series. One must observe the general tendency of the data. As long as one can say that the period as a whole was characterised by upward movement, or a downward movement, secular trend is present. For example, if we observe the prices over a period of 20 years and find that except for a year or two in between, the prices are continuously rising we would say that a trend is present.

3. When we say that secular trend refers to the general tendency of the data to increase or decrease over a 'long period of time' we would be interested to know as to what constitutes a 'long period time'. Does it mean several days, several years or several decades? The term 'long period of time' cannot be defined precisely. In some cases a period as small as a week may be fairly long while in some other cases a period as long as two years may not be long enough. To take an example, if we are studying the figure of production of agriculture commodity for 1978 and 1979 and if we find that the production has gone up in 1979, the increase cannot be called as secular trend because this is too short a period of time to conclude that the production is showing an increasing tendency. On the other hand, if we put a germicide into bacterial culture and count the number of organisms alive after every half minute for an hour, these 120 observations showing a general pattern would be secular trend. Hence the nature of data would determine whether a particular period should be called as long enough or not. One should be careful in this regard because an upward or downward trend over period may only be an upward or downward cyclical movement contained within a longer period.

2. Seasonal Variations

It would be observed that in many social, economic and business phenomena, apart from the growth factor in time series, there are forces at work which present the smooth flow of the series in a particular direction and trend to repeat themselves over a period of time. Such forces do not act continuously but operate more or less in a regular manner and are called periodic movements.

One rhythmic force which is inherent in most of the time series is what is called seasonal variations. By seasonal variations, we mean periodic and regular movements in time series where the period is no longer than one year. For example, sales and profits in a departmental store, prices and consumption of commodities, back clearing etc. all exhibit seasonal variation Fig.2 illustrates a series containing marked seasonal variation with period equal to one year, other components being negligible.

Nearly *every* type of business activity is susceptible to seasonable influence to a greater or less degree and, as such, these variations are regarded as normal phenomenon recurring periodically. Although the word seasonal seem to imply a connection with the season of the year the term is meant to include a kind of variation which is of periodic nature and whose repeated cycles are of duration not exceeding one year. Seasonal variations may be attributed to the following two causes :

(a) those resulting from natural forces, and

(b) those resulting from man made conventions.

In (a) we have those variables which "depends upon changes in the climate and weather conditions such as rainfall, heat, humidity etc. For example, production of agricultural commodities is influenced by the climate and seasons which, therefore, directly affect the income of the farmer, by affecting the price of the specified commodity, etc. Even production of other commodities such as eggs, woolen cloth, cold drinks etc. also vary seasonally. In (b) we have those variables which depends upon customs, traditions etc. which are man-made. For example, the sales and profits in departmental stores show a marked rise during festivals such as Holi, Diwali, Dussehra etc. or, in the first week of every month there are heavy withdrawals of money from the banks. To take another example, the sale of books and the note books increase during the first month of the opening of schools and colleges.

3. Cyclical Variations

The oscillatory movement in time series with period of oscillation more than one year are termed as cyclical variations. The terms 'cycle' refers to the recurrent variations in time series that usually are a period longer than a year and are regular neither in amplitude nor in length. Fig.3 illustrate a series exhibiting cyclical variations.

Most of the time series relating to economic and business show some kind of cyclical variation. Cyclical fluctuations are long-term movements that represent consistently recurring rises and decline in activity. The cyclical movements in a time series are generally attributed to the so called 'business cycle' which are composed of alternating periods of prosperity (or boom) and depression in an economy. Most of the economic and business series e.g. series relating to prices, production, wages etc. are affected by business cycles.

4. Irregular Movements

Irregular variations refer to such variations in a time series as do not repeat in any definite pattern. These variations in fact, include all types of variations other than those accounted for by the trend,' seasonal and cyclical movements. These fluctuations largely refer to purely random fluctuations, being the result of chance factors and wholly uncontrollable factor which are unpredicable. Irregular variations are also caused by isolated special occurrences as earth-quakes, wars etc.

2.3.3 Analysis of Time Series

The Immediate objective of the analysis of time series is to break down the series into the main components, which reflect the secular trend, seasonal movement, cyclical movements and random movement. In other words, the movement of the series is regarded as being compounded of these elements, and an attempt is made to isolate the magnitude of each of these elements separately, showing how the movements of the separate components together account for the movement of the series.

The method of analysis depends very largely on the hypothesis as to how the components of the time series combined. The simplest hypothesis is to assume that the separate influences have value 'which are additive and independent of each other. Such a hypothesis is called the additive model of the time series and can be expressed as:

$$Ut= Tt + St + Ct + Rt$$

Where Ut is time series value at time t, Tt is the trend value, St is the seasonal Variation, Ct is the cyclical variation and Rt is the random variation. Such a model of course, implies that Ct will have positive value according to whether we are in an above normal or below normal' phase of cycle. The total of positive and negative values for the cycle will be zero. Similar considerations apply to the St terms, St will be positive or negative according to the season of the year, and the total of the St for a year will be zero. Obviously the term St will not appear in series of annual data. Rt will have positive or negative values and in the long run. ΣR will be zero. Occasionally Rt may take an extreme value on account of some extraordinary occurrence such a movement should be regarded as an isolated occurrence,

An alternative hypothesis to the above one is :

$$Ut = T_t \times S_t \times C_t \times R_t$$

and is referred, to as the multiplicative model. In this case St, Ct, Rt instead of assuming positive and negative values are indices fluctuating above or below unity. With this model, the geometric mean of St in a year, Ct is a cycle and Rt in a long term period are unity. It may be pointed out the multiplicative decoposition of a time series is same the additive decomposition of the logarithmic values of the original series:

$$\log Ut = \log Tt + \log St + \log Ct + \log Rt$$

Although the additive assumption is undoubtedly true in some cases, the multiplicative assumption characterizes the majority of economic time series. Consequently the multiplicative model is not only considered the standard or traditional assumption for time series analysis but is more often employed in practice. For the reason we shall use only the multiplicative model in our subsequent discussion.

We shall now give methods of measurements of systematic components of time series and the analysis of the random component of time series:

2.3.4 Measurement of Trend

Trend can be studied and/or measured by the following methods:

(i) Graphic method or the method of the free drawing.

(ii) Method of Semi averages.

(iii) Method of moving averages.

(iv) Method of least squares.

Each of these methods will now be discussed.

2.3.4.1 (i) Free hand drawing of graphic method

This is the simplest method of the studying trend. In this method we first graph the series and join the points by lines. Then we draw a free-hand smooth curve which seem to fit the data best. This smooth curve enables us to form an idea about the general trend of the series. Smoothing the curve eliminates other components. viz. regular and irregular components.

This method does not involve any mathematical technique and is the simplest method of studying trend. The method can be used to describe all types of trend, linear and non-linear. The simplicity and flexibility are strong points of the method.

The main drawback of his method is that it is highly subjective because the trend curve will depend upon the personal judgement of the investigator and therefore, different persons may draw different curves from the same set of data. As such this method should be used by experienced statisticians and this limits the utility and popularity *of* the method.

2.3.4.2 (ii) Method of Semi-averages

In this method, the whole data are divided into two equal parts with respect to time. For example, if we are given data from 1961 to 1972 i.e. over a period of 12 years, the two equal parts will be the data from 1961 to 1966 and 1967 to 1972. In case of odd number of years the two parts are obtained by omitting the value corresponding to middle years e.g. for the data from 1961 to 1966 and 1968 to 1973, the value corresponding to middle years, 1967, being omitted.

After the data have been divided into two parts, the arithmetic mean of each part is obtained. We thus get two values. Each value is plotted at the mid-point of the interval converted by these respective part. The line obtained by joining these two points is the required trend line.

For even number of years which are multiple of 4 such as 8, 12 etc., the centering of average of each part will be done by taking the mid-value of the two middle years of each part we will illustrate the method by an example:

Example 1 :

Fit a trend line to the following data by the method of semi average: (Table-1)

Table-1

Year	Bank Clearings (Crore Rs.)	Year	Bank Clearings (Crore Rs.)
1956	53	1963	87
1957	79	1964	79
1958	75	1965	104
1959	66	1966	97
1960	69	1967	92
1961	94	1968	101
1962	195		

Since the number of years for which the series is given is 13, an odd number, we omit the middle year 1962, and divide the data in two parts 1956-1961, and 1963-1968.

The average of the first part, 1956-1961 $= \dfrac{437}{6}$

$= 72.83$

The average of the second part, 1963-1968 $= \dfrac{560}{6}$

$= 93.33$

These two means are to be plotted in the middle of the period of the two parts and then joined as shown in figure 5.

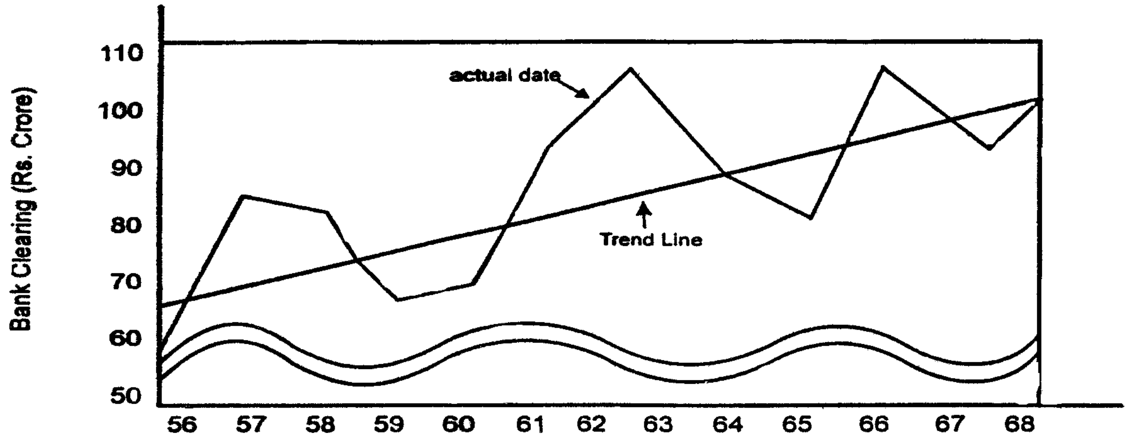


Fig. No. 5 *Trend by the method of Semi-averages*

Merits and Limitations :

As compared with graphic method, the merit of this method lies in its objectivity in the sense that everyone who applies this method is bound to get the same result. Also this method is simpler to understand, compared to the method of least squares and the methods of moving averages.

The limitations of the method is that it assumes a straight line relationship between the plotted points, which may or may not be true. Also, the limitations of arithmetic mean as an average stand in its way.

2.3.4.3 (ii) Method of Moving Averages :

A third method for obtaining the trend is an attempt to smooth out the bumps in the series by a process of averaging. This can be done by using what are called moving averages.

Moving averages of extent (or period) m in a series of successive arithmetic means of m at a time, starting with 1st, 2nd, 3rd terms etc. Thus the first average is the mean of the first m terms of the series, the second is the mean of the m terms from 2nd to (m + 1)th term, the third is the mean of m terms from 3rd to (m + 2)th term and so on.

If m is odd, say m = 2k + 1, moving average is placed against the middle term i.e., (k + 1)th term of the series. And if m is even, say m = 2k, the moving average is placed between the two middle terms, i.e., between kth and (k + 1)th term of the series. In the latter case the moving average does not coincide with an original time period. An attempt, therefore, must be made to synchronise moving averages with original data. This is done by centring the moving averages which is done by taking a moving average of two of these moving averages and putting these values in the centre. The graph obtained on plotting the moving average against time gives the trend.

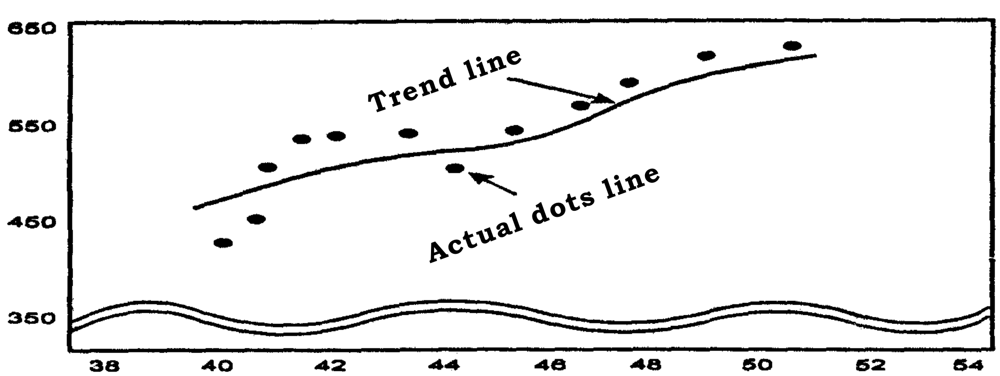
We illustrate the method by following example.

Example 2 :

The following table describes data relating to the annual production of tea in India during 1937-54. The 5 year moving average are also calculated.

Table 2

Year	Production (Million Ibs)	5-Year Moving Total	Trend Value (5 Yearly Moving Averages)
1937	396	--	--
1938	419	--	--
1939	417	2126	425.2
1940	430	2245	449.0
1941	464	2344	468.8
1942	515	2394	478.8
1943	518	2466	493.2
1944	467	2542	508.4
1945	502	2584	516.8
1946	540	2637	527.4
1947	557	2756	551.2
1948	571	2866	573.2
1949	586	2973	594.6
1950	612	3031	606.2
1951	647	3061	613.6
1952	615	3126	625.2
1953	608	--	--
1954	644	--	--



Trend fitted by the method of moving average for example

Fig. No.6

The data are plotted in Figure 6 and the smooth curve shows the trend obtained by moving average of extent 5.

If the series consists of random variations around a trend, a moving average will trend to reduce or smooth out these variations. The larger the number of terms in the moving average, the more smooth will be the resulting series; but the larger the number of terms the more information is lost at the beginning and the end of the series. A three year moving average is often used for this kind of work.

On the other hand, if the series consists of oscillatory movements about a trend, then the moving average completely eliminates the oscillatory movements provided (i) the extent of moving average is exactly equal to the period of oscillation and (ii) the trend is linear, Since different cycles vary in amplitude and, period, the appropriate period of moving average should be equal to the mean period of the cycles in the data.

Merits and Limitations :

The method of moving averages is simpler than the method of least square and is quite flexible in the sense that the addition for a few more figures to the data simply results in some more trend values and the previous calculations are not affected at all. Another merit of moving average method is that if follows the general movement of the data and that its shape is determined by the data rather than the statistician's choice of the mathematical function.

The main limitation of the method is that it does not provide trend values for all the terms. Though the values for the beginning of the period may not be important, the trend values for the end of the period are important. Another important, limitations of the method is that it cannot be used for forecasting or predicting future trend, which is the main objective of trend analysis and is accomplised by the method of least square.

Below we give another example by illustrating the computation of trend by the method of moving average of extent 4 (an even number).

Example 3

The following data relate to the production of tea in India. Assuming a 4 year cycle, we show below the calculation of trend by the method of moving averages.

Table 3

Calculation of Trend by the Moving Averages Method

Year	Production of Tea (Million Ibs)	4-Yearly Moving Total	4-Yearly Moving Averages	4-Yearly Moving average Centered
1961	465	--	--	--
1962	514	--	--	--
		← 1964	← 491.0	--
1963	518			← 495.7
		← 2001	← 500.25	
1964	467			← 503.6
		← 2027	← 506.75	
1965	502			← 511.6
		← 2066	← 516.50	
1966	540			← 529.5
		← 2170	← 542.50	
1967	557			← 553.0
		← 2254	← 563.50	
1968	571			← 572.5
		← 2326	← 581.50	
1969	536	--	--	--
1970	612	--	--	--

2.3.4.4 (iv) Method of least squares

A fourth method for obtaining the trend is to fit mathematical curve to the time series. The method of fitting the curve is by least squares and has certain advantages. Once the form of the curve has been decided, fitting of one curve is quite objective. Further more the form of mathematical curve can be interpreted in terms of the behaviour of the variable, so that a particular form of curve implies a certain type of curve. Thus, for example, a curve of the type.

$$Ut = a + bt$$

Where Ut is the value of the variable and t measures time. It means that the variable Ut is changing by a constant amount per unit of time. This result in a convenient and logical description of the trend. This is an advantage not shared by moving averages.

The method of least square consists in determining the unknown constants of the mathematical curve so that the sum of squares of the deviations between the given values of u and their estimates Ut, obtained by the mathematical curve i.e.

$\sum(u - u_t)^2$ is minimum

This method of curve fitting by the method of least squares is used quite often in trend analysis, particularly when one is interested in making projections for future times.

2.3.4.4.1 Linear Trend

The simplest form of trend is linearly expressed as :

$u_t = a + bt$

The constants a and b are determined so as to minimise the sum of squares.

$Z = (u - a - bt)^2$

Z will be minimum where its partial differential coefficients with respect to a and b are equal to zero.

$$\frac{\delta Z}{\delta a} = -2\,(u - a - bt) = 0, \text{ and}$$

$$\frac{\delta Z}{\delta b} = -2t\,(u - a - bt) = 0$$

This is equivalent to :

$aN + b\sum t = \sum u$ .... (1)

$a\sum t + b\sum t^2 = \sum tu$ .... (2)

Where N represents the total number of terms in the series. The equations (1) and (2) are called the normal equations whose solution gives the values of a and b.

We can measure the variable from any point of time as origin such as the first year. But the calculations are simplified when the mid point time is taken as origin because in that case the negative values in the first half of the series and positive values in the second half will balance somewhat.

We illustrate the method by the following example :

Example 4 :

Fit a linear trend to the annual figures of freight carried by an airlines agency.

Year	:	1955	1956	1957	1958	1959	1960	
Freight (Million ton miles)	:	9.4	10.5	11.2	12.1	15.4	24.6	
Year	:	1961	1962	1963	1964	1965	1966	1967
Freight (Million ton miles)	:	30.1	30.4	33.1	38.6	51.8	61.8	62.9

For fitting a straight line we need $\sum t$, $\sum t^2$, $\sum u$ and $\sum tu$. The calculation will be much simplified if we take the origin for t at year 1961. Thus we write t = 0 for 1961. t = 1 for 1962. t = 2 for 1963 etc. and t = -1 for 1960. t = -2 for 1959 etc., It will be convenient to do computation in a tabular form which is shown in the following table 4.

Table 4

Year	Freight u	t	t^2	tu	Trend Value ut
1955	9.4	-6	36	-56.4	1.77
1956	10.5	-5	25	-52.5	6.50
1957	11.2	-4	16	-44.8	11.23
1958	12.1	-3	9	-36.3	15.96
1959	15.4	-2	4	-30.8	20.96
1960	24.6	-1	1	-24.6	25.42
1961	30.1	0	0	0	30.15
1962	30.4	1	1	30.4	34.88
1963	33.1	2	4	66.2	39.61
1964	38.6	3	9	115.8	44.34
1965	51.8	4	16	207.2	49.07
1966	61.8	5	25	309.0	53.80
1967	62.9	6	36	377.4	58.83
	$\sum u = 391.9$	$\sum t = 0$	$\sum t^2 = 182$	$\sum tu = 860.6$	

Writing $\sum t = 0$ in the normal equation (1) and (2)

$$a = \frac{\Sigma u}{\Sigma n} = \frac{391.9}{13} = 30.15$$

$$b = \frac{\Sigma tu}{\Sigma t^2} = \frac{860.6}{182} = 4.73$$

And the trend line is

ut = 30.15 + 4.73t

where t is number of the year with origin at 1961.

The last column of table 4 is obtained by putting successive values of t in the fitted equation.

2.3.4.4.2 Exponential Trend :

The linear trend is one in which ut increases by a constant absolute amount per unit of time. Many other types of trends are available. A commodity used is one where 'u' increases by a constant percentage per unit of time. Such a trend is defined by :

$$u_t = ab^t$$

This type of trend is called an exponential trend.

An exponential trend is readily fitted by taking logarithms, it can be written as :

$$\log u_t = \log a + t \log b$$

Which is linear form in t and log u_t and so the constant log a and log b can be determined from equations (1) and (2) by replacing u_t by log u_t, a and b by log a and log b.

Thus $\log aN + \log b\sum t = \sum \log u$

and $\log a\sum t + \log b\sum t^2 = \sum t \log u$

Example 5

The fitting of an exponential curve is illustrated for the data of example 4. The calculations are illustrated in Table 5.

The values of the constants are

$$\log a = \frac{\Sigma \log u}{N} = \frac{18.0766}{13} = 1.3905$$

$$\log b = \frac{\Sigma t \log u}{\Sigma t^2} = \frac{13.7310}{182} = 0.0754$$

Taking antilogarithms, we have

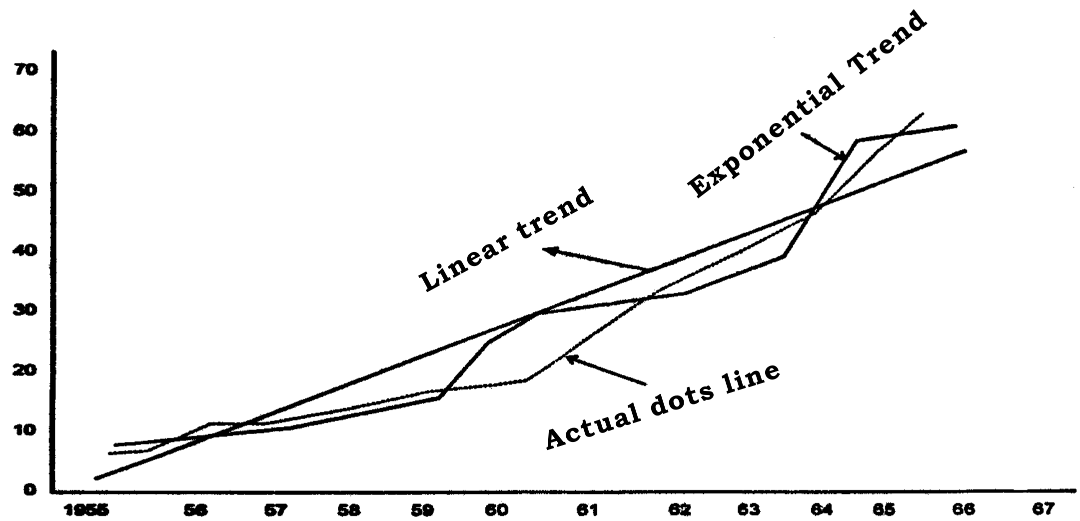
a = 24.85 and b = 1.19

Table 5

Year	Freight (u)	t	t^2	log u	t log u	u_t
1955	9.4	-6	36	0.9731	-5.8386	8.67
1956	10.5	-5	25	1.0212	-5.1060	10.31
1957	11.2	-4	16	1.0492	-4.1968	12.28
1958	12.1	-3	9	1.0827	-3.2484	5.60
1959	15.4	-2	4	1.1875	-2.3750	17.37
1960	24.6	-1	1	1.3909	-1.3909	20.65
1961	30.1	0	0	1.4786	0	24.65
1962	30.4	1	1	1.4829	1.4829	29.43
1963	33.1	2	4	1.5198	3.0396	34.77
1964	38.5	3	9	1.5866	4.7598	41.27
1965	51.8	4	16	1.7143	6.8572	49.21
1966	61.8	5	25	1.7910	8.9550	58.55
1967	62.9	6	36	1.7988	10.7922	69.65
Total			182	18.0766	13.7310	

and the exponential trend is :

$$u_t = 24.85\ (1.19)^t$$



Linear trend line and exponential trend

Fig.7

The last column of the table 5 gives the exponential trend values which is obtained by using successive values of t. For this we have first calculated log ut from the equation.

$$\log u_t = 1.8006 + 0.754t$$

and then take antilogarithms.

Figure 7 shows the graphs of the given series in example 4 and 5 alongwith the trend as fitted by the straight line and the exponential curve.

Figure 7 shows the linear trend along with exponential trend (shown by broken line) for data. It can be seen that in this case the exponential trend gives a better description of the under movement of the series than the linear one.

2.3.4.4.3 Other mathematical curves

The linear and exponential curves are simplest of trend types. Polynomials of the form $u_t = a + bt + ct^2 + dt^3 + \ldots\ldots\ldots\ldots$ (t + 1) terms may also be used and these can be fitted by the method of least squares. For example to fit the parabola.

$$u_t = a + bt + ct^2$$

we minimise,

$$Z = \sum(u - a - bt - ct^2)^2$$

The normal equations are :

$$aN + b\sum t + c\sum t^2 = \sum u$$

$$a\sum t + b\sum t^2 + c\sum t^3 = \sum tu$$

$$a\sum t^2 + b\sum t^3 + c\sum t^4 = \sum t^2 u$$

There are three equations in three unknowns a, b and c which could be solved.

More complicated forms also take some time, for which the fitting may be rather more difficult. For example a well known curve is the logistic curve whose equation is :

$$\frac{1}{y} = K + ab^x, \text{ where k, a and b are constants.}$$

The graph of equation is S shaped which starts growing slowly, then grows rapidly and finally reach a saturation point. The example where such a curve will fit well are the sales of a new product and the population of insects reproducing in a confined space etc.

The equation of another growth curve is :

$$y = Ka^{bx}$$

Which is called **Gompertz curve.**

The constants of these curves cannot be determined by the principles of least squares as explained above. Special techniques have been devised for fitting these curves to the given set of data. Due to mathematical complications we shall not discuss the methods here.

The choice of a particular type of equation that best describes the data is often difficult and needs considerable amount of judgement and experience.

Merits and Limitations :

The methods of least squares gives the best fit of the assumed mathematical equation to the data. In the sense that the sum of the squares of the deviations of observed values from their estimates is minimised. Also the method is not subjective.

The main limitation of the method is to justify the selection of the mathematical form of the curve to be fitted. Variables do change in a more or less systematic manner over time but this can usually be attributed to the operation of several explanatory variables on time.

ANALYSIS OF TIME SERIES-II

Measurement of Seasonal Variation :

Where data are expressed annually there is no seasonal variation. But monthly or quarterly data frequently exhibit strong seasonal movement and hence the need for developing a pattern of average seasonal variation. It may be desired to complete the seasonal pattern of different series, but more often we may want to know the extent to which we should discount the most recently available statistics for seasonal factors. For example, if we observe the sales of a book seller, we find that for the quarter July-September sales are maximum. If we know by how much the sales of this quarter are usually above or below the previous quarter for seasonal reasons, we shall be able to answer a very basic question, namely was this due to an underlying upward tendency or simply because this quarter is usually seasonally higher than the previous quarter.

In order to analyse seasonal variation it is necessary to assume that the seasonal pattern is super-imposed on a series of values and is independent of these in the sense that the same pattern is superimposed irrespective of the level of the series, e.g. the July-September quarter always contributes so much more or so much less of the series.

Seasonal patterns are exhibited by most of the business and economic phenomena and their study is necessitated by the following reasons :

(i) To isolate the seasonal variation, i.e. to determine the effect of seasons on the size of the variable, and

(ii) To eliminate them, i.e. to study as to what would be the value of the variable if there were no seasonal swings.

For the study of seasonal variation the data must be given for 'parts' of a year viz. monthly or quarterly, daily or weekly. The decision whether weekly, monthly or quarterly seasonal indices are to be calculated depends upon the nature of the problem and the type of the data available. The following methods are some of the simple and more popular methods of measuring seasonal variations :-

1. Method of Simple Averages
2. Ratio-to-Trend Method
3. Ratio to Moving Average Method
4. Link Relative Method

Method of Simple Averages :

This is the simplest of all the methods of measuring seasonal component of a time series and consists of the following steps :

(i) Arrange the data by years and months (or quarters data are given).

(ii) Compute the average $\overline{x}_i$ (i = 1, 2................12) for the month for all the years (i = 1, 2.................12) represents January, February ....December respectively).

(iii) Compute the average $\bar{x}$ of these monthly averages

$$\overline{x} = \frac{\sum x}{12}$$

(iv) Seasonal indices for different months are then obtained by expressing these monthly averages as percentage of x. Thus

Seasonal index for its month = $\frac{\overline{x}i}{x} x100$ (i = 1, 2................12)

Remarks :

1. If, instead of monthly averages we use monthly totals for all the years, we will get the same result.
2. Total of seasonal indices is 12 × 100 = 1200 for monthly data and 4 × 100 = 400 for quarterly data.

Merits and Demerits :

This method is based on the assumption that the data do not contain any trend any cyclical components and consist in eliminating irregular components by averaging the monthly (or quarterly) values over years. Since most of the economic time series have trends, these assumptions are not in general true and as such this method is seldom of any value. The method, however, is very simple.

Example 1

The data given below gives the average quarterly prices of commodity for four years. Assuming that trends is absent what are the seasonal indices for various quarters ?

Year	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
1967	40.3	44.8	46.0	48.0
1968	50.1	55.1	55.3	59.5
1969	47.2	50.1	52.1	55.2
1970	55.4	57.0	61.6	65.3

Solution : The difference, if any, in the averages of various quarters will be due to seasonal variations, because we have assumed that trend is absent. To calculate the seasonal indices, we first compute the total over years of the four quarters and then averages from each quarter.

	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
Total	193.0	207.0	215.0	228.0
Average	48.25	51.75	53.75	57.0
Seasonal Index	91.57	98.21	102.01	108.18

The total for the first quarter = 40.3 + 50.1 + 47.2 + 55.4
= 193.0

The average for the Ist quarter $= \frac{193.0}{4} = 48.25$

And similarly for other quarters.

The average of averages $= \frac{48.25 + 51.75 + 53.75 + 57.0}{4} = 52.69$

Seasonal Index for Ist quarter $= \frac{48.25}{52.69} \times 100 = 91.57$

Seasonal Index for 2nd quarter $= \frac{51.75}{52.69} \times 100 = 98.21$

Seasonal Index for 3rd quarter $= \frac{53.75}{52.69} \times 100 = 102.01$

Seasonal Index for 4th quarter $= \frac{57.0}{52.69} \times 100 = 108.18$

Example 2

The following data gives the monthly consumption of electric power in million of kw. hours for street lighting in a big city during 1971-75. The data are given in the first six columns of table 1. Find the monthly seasonal indices by the method of monthly averages.

Table 1

Consumption of monthly electricity power and computation of seasonal indices.

Consumption of Electric Power

Month	1971	1972	1973	1974	1975	Total of Cols. (2) to (6)	Averages	Seasonal index
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Jan.	318	342	367	392	420	1839	367.8	116.1
Feb.	281	309	328	349	373	1645	329.0	103.0
Mar.	278	299	320	342	370	1809	321.8	101.4
Apr.	250	268	287	311	334	1450	290.0	91.4
May	231	249	269	290	314	1353	270.6	85.5
June	216	236	251	273	96	1272	254.3	80.3
July	223	242	259	282	305	1211	262.2	82.8
Aug.	245	262	284	305	330	1426	285.2	90.0
Sep.	259	288	309	328	356	1550	310.0	98.0
Oct.	302	321	345	364	396	1728	345.6	109.1
Nov.	352	342	367	389	425	1845	369.2	116.6
Dec.	346	364	394	417	452	1974	394.8	124.7
					Total	19002	38004	1200
					Average of averages		316.7	

First six columns give the raw data. Column (7) is obtained by taking the total over years i.e. by adding columns (2) to (6). An average of each month is obtained by diving the column (7) by 5. The average of averages is then obtained by summing the twelve values is column (8) and dividing by 12. The seasonal index for each month is obtained by expressing the monthly average as percentage of the monthly averages.

Thus the seasonal index for January $= \frac{367.8}{316.7} \times 100 = 116.1$

and the seasonal index for February $= \frac{329.0}{316.7} \times 100 = 123.9$ and so on

Ratio to Trend Method

This method is an improvement over the simple averages method and is based on the assumption that seasonal variation for any given month is a constant fraction of trend. The ratio to trend method presumably isolates the seasonal factor in the following manner. Trend is eliminated when the ratios are computed.

$$= \frac{T \times S \times C \times R}{T} = S \times C \times R$$

Random elements are eliminated when the ratios are averaged. A careful selection of the period of years used in the computation may remove the cyclical effects. This method, therefore, is recommended for use either when cyclical variation is known to be absent or when it is not so even if present.

The measurement of seasonal variation by this method consists of the following steps.

(i) Obtain the trend values by the method of the least squares by fitting a mathematical curve.

(ii) Express the original data as percentage of the trend values, i.e. divide the figure corresponding to each month by the corresponding trend value and multiply these ratio by 100. The values so obtained are now free from trend.

(iii) The Cyclical and irregular components are then wiped out by averaging the percentages for different months (quarters) if the data monthly (quarterly), thus leaving us with indices of seasonal variations. Either arithmetic mean or median can be used for averaging. If there are a few very high or low values. Modified mean (which consists of calculating arithmetic mean by dropping out of these extreme values) may be used. In such situations it is often desirable to use median which is not affected by very high and low values.

(iv) Finally these indices, obtained in step, (iii) are adjusted to a total 1200 for monthly data and 400 for quarterly data by multiplying each index by a constant k given by

$$k = \frac{120}{\text{the sum of 12 indices}}$$

For monthly data and

$$k = \frac{400}{\text{the sum of 4 indices}}$$

for quarterly data. This gives the final seasonal index.

Merits and Demerits

Since this method attempts at ironing out the cyclical or irregular components by the process of averaging, the purpose will be accomplished only if the cyclical variation are known to be absent or they are not so pronounced even if present. On the other hand if the series exhibits pronounced cyclical swings the trend values obtained by the method of least squares will not follow the actual data as closely as a 12-month moving averages for monthly data. As such, the seasonal indices by ratio to trend method are liable to be more biased then those obtained by ratio to moving average method described earlier.

Compared with the method of simple averages, this method is certainly a more logical procedure for measuring seasonal variation. It has an advantage over the method of moving average too, for it has a ratio- to trend value of each month for which the data is at distance advantage especially when the period covered by the time series is very short. This method is not easy to compute and easy to understand.

Remarks :

The calculations are simplified to a great extent by fitting trend equation to yearly (total or averages) and then obtaining the monthly (or quarterly) trend values by a suitable modification of the trend equation, as illustrated in the following example.

Example 3

Calculate seasonal variation for the following data series in thousand rupees of a firm by ratio- to trend method.

Year	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
1973	30	40	36	34
1974	34	52	50	44
1975	40	58	54	48
1976	54	76	68	62
1977	40	92	86	82

Solution :

First of all we determine the trend values for the yearly averages (y) by fitting a linear trend by the method of least squares, as indicated in table 2.

Table 2

Calculation of trend by method of least squares

Year	Yearly total	Yearly average y	u = x - 1975	uy	u^2	Trend Value
1973	140	35	-2	-70	4	32
1974	180	45	-1	-45	1	44
1975	200	50	0	0	0	56
1976	260	65	1	65	1	68
1977	340	85	2	170	4	80
	Total	280	0	120	10	

To fit the line y = a + bu the normal equation are

$na + b\sum u = \sum y$

$a\sum u + b\sum u^2 = \sum uy$

and $a = \frac{\sum y}{n}$ and $b = \frac{\sum uy}{\sum u^2}$

In this case $a = \frac{280}{5} = 56$

and $b = \frac{120}{10} = 12$

And the trend line is

$y = 56 + 12\,u$

The trend value for each is obtained by putting u = -2, -1, 0, 1 and 2

For u = -2 the trend value 56 + 12x (-2) = 32 and so on.

The yearly increment in the trend value is equal to 12 (=b) and hence the quarterly increment is $\frac{12}{4} = 3$.

Next we determine the quarterly trend values as follow :

For 1973, the trend value for the middle quarter, i.e. half of second quarter and half of third quarter, is 32. Since the quarterly increment in the value is 3, the trend value for the second quarter is $32 - \frac{3}{2} = 32 - 1.5 = 30.5$ and that for the third quarter is 32 + 1.5 = 33.5. The trend value for the first quarter will be 3 less than the value of the second quarter i.e. equal to 30.5 - 3 = 27.5. The trend value for the fourth quarter will be 3 more than the value of the third quarter i.e., equal to 33.5 + 3 = 36.5.

Similarly we get the trend values for each quarter in other years. The values are given in table 3.

Table 3
Trend Values

Year	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
1973	27.5	30.5	33.5	36.5
1974	39.5	42.5	45.5	48.5
1975	51.5	54.5	57.5	60.5
1976	63.5	66.5	69.5	72.5
1977	75.5	78.5	81.5	84.5

The trend is then eliminated by expressing the given values as percentages of the corresponding trend values. Thus for the quarter 1973 the percentage is $\frac{30}{27.5} \times 100 = 109.1$ and for the second quarter of 1973 the percentage is $\frac{40}{30.5} \times 100 = 131.1$ and so on. The values are obtained in table 4. The arithmetic mean of S values for each quarter as then obtained. These are seasonal indices which are now adjusted by multiplying by the constant.

$$k = \frac{400}{92.78 + 118.26 + 102.92 + 87.32} = \frac{400}{401.28}$$

Table 4
(Calculation of Seasonal Indices)

Trend eliminated value (Percentage)

Year	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
1973	109.1	131.1	107.5	93.1
1974	86.1	122.4	109.9	90.7
1975	77.7	106.4	93.9	70.3
1976	85.0	114.3	97.8	85.5
1977	106.0	117.1	105.5	97.0
Total	463.9	591.3	514.6	436.0
Average (A.M.)	92.78	118.26	102.92	87.32
Seasonal Index (Adj)	92.9	117.7	102.5	86.9

Ratio to Moving average method :

As explained earlier, periodic fluctuations in a series are eliminated by taking a moving average of period equal to the period of the fluctuation. So from monthly data seasonal fluctuations can be removed by taking a 12 month moving average which must again be centred by taking a further - point moving average. The average value may, therefore, be supposed to give no estimates of and trend cyclical variations.

The ratios of original values to the moving averages etc. are therefore, expected to represent variation and random component.

$$\frac{T \times S \times C \times R}{T \times C} = S \times R$$

The different values for cash month are then averaged so that irregular fluctuations are removed. The adjusted seasonal indices are then obtained by multiplying by a constant as in the ratio-to-trend method. The different steps for the computation of seasonal indices by this method are as follows :

(i) Calculate the 12 month moving average of the given data and then centre them by further calculating a 2 point moving average.

(ii) Express the original data (except for 6 months in the beginning and 6 months at the end for which moving averages are not obtained) as percentage of the centred moving average values.

(iii) the preliminary seasonal indices are now obtained by averaging the percentage. If the variations in the set of values of a month is only due to irregular fluctuations, the values will vary only by a small amount and the arithmetic mean may be used. If however, there are extreme values, one should use the median or modified mean.

(iv) The sum of these indices = S (say) will not, in general, be 1200. An adjustment is done to make the sum of indices 1200 by multiplying each index by a constant

$$k = \frac{1200}{S}$$

These are the final seasonal indices.

Merits and Demerits

Of all the methods of measuring seasonal variation, the ratio to the moving average method is the most satisfactory, flexible and widely used. These indices do not fluctuate too much as the indices based on straight line trends. Mathematical methods of eliminating the effects of business cycles are not usually needed, for the 12 month moving average follows the cyclical course of the actual data quite closely.

The main drawback of this method is that it does not completely utilise the data. The 12-month moving average can not be obtained for the first 6 months and the last 6 months.

Example 4

Apply ratio to moving average method to ascertain seasonal indices from the data of number of persons visiting a place of interest as given in the first two columns of table 5.

Solution :

Table 5 and 6 give the necessary computation for the calculation for the seasonal indices.

Table-5

12-month moving averages and percentages

Years (1)	month (2)	No. of persons visiting a place of interest (3)	12-point moving totals (4)	12-point moving average	Centered 12-point moving averages (5)	Ratio to moving average $(3 \div 5) \times 100$ (6)
1971	Jan.	90				
	Feb.	85				
	Mar.	70				
	Apr.	60				
	May	55				
	June	45	898	74.83		
	July	30	908	75.67	75.3	39.8
	Aug.	40	912	76.00	75.8	52.8
	Sep.	70	916	76.33	76.2	9.9
	Oct.	120	918	76.50	76.4	157.1
	Nov.	115	918	76.50	76.5	150.3
1972	Dec.	118	920	76.66	76.7	130.4
	Jan.	100	920	76.66	76.6	154.0
	Feb.	89	923	76.91	76.8	115.9
	Mar.	74	918	76.50	76.7	96.5
	Apr.	62	925	77.16	76.8	80.7
	May	55	928	77.33	70.2	71.2
	June	47	930	77.50	77.9	60.7
	July	30	940	78.33	77.4	8.5
	Aug.	43	944	78.66	78.5	54.8

	Oct.	127	948	79.00	78.8	82.5
	Nov.	118	952	79.33	79.2	160.4
	Dec.	120	955	79.58	79.5	148.4
1973	Jan.	10	948	79.03	79.0	151.3
	Feb.	93	953	79.41	79.5	138.9
	Mar.	78	955	79.58	79.5	117.0
	Apr.	66	962	80.19	79.6	97.6
	May	58	965	80.41	80.3	82.2
	June	40	965	80.41	80.4	72.1
	July	35	969	80.75	80.6	49.6
	Aug.	45	---	---	---	---
	Sep.	72	---	---	---	---
	Oct.	130	---	---	---	---
	Nov.	118	---	---	---	---
	Dec.	124	---	---	---	---

Table-6

(Calculation of adjusted seasonal indices)

Month	Year 1971	1972	1973	Seasonal (Index Arith. Mean)	Adjusted Seasonal Index
Jan.	--	130.4	138.9	134.7	135.0
Feb.	--	115.9	117.0	116.5	116.7
Mar.	--	96.5	97.6	97.1	97.3
Apr.	--	80.7	82.2	81.5	81.6
May	--	71.2	72.1	71.7	71.8
June	--	60.7	49.6	55.2	55.3
July	39.8	38.5	--	39.2	39.3
Aug.	52.8	54.8	--	53.4	53.9
Sep.	91.9	82.5	--	87.2	87.3
Oct.	157.1	160.4	--	158.4	159.1
Nov.	150.3	148.4	--	149.4	149.7
Dec.	154.0	151.3	--	152.7	153.0
	Total			1197.7	1200.00

The correction faction (k) for obtaining adjusted seasonal indices

$$= \frac{1200}{1197.7} = 1.0019$$

Method of Link Relatives

This method, also known as Pearson Method, is based on averaging the link relatives which estimates approximately the ratio of successive seasonal indices. Link Relative is the value of one month (or quarter) expressed as a percentage of the proceeding month (or quarter), for monthly data.

The link relative for any month = $\frac{\text{Current Month's figure}}{\text{Previous month's figure}} \times 100$

When this method is adopted, the following steps are required for monthly data (i) Calculate the link relative of the monthly figure as follows

Link relative of Feb. = $\frac{\text{Feb}}{\text{Jan}} \times 100$

Link relative of May = $\frac{\text{May}}{\text{April}} \times 100$

etc.

(i) Calculate the average link relative for each month. Mean or median may be used for averaging.

C.R. for Feb. = $\frac{\text{L.R. of Feb.} \times \text{C.R. for Jan.}}{100}$, C.R. donate chain relative

C.R. for Mar. = $\frac{\text{L.R. of Mar.} \times \text{C.R. for Feb.}}{100}$

etc.

C.R. for Dec. = $\frac{\text{L.R. of Dec.} \times \text{C.R. for Nov.}}{100}$

Now by taking this Dec. value as base, a new C.R. for Jan. can be obtained as

$= \frac{\text{L.R. of Jan.} \times \text{C.R. for Dec.}}{100}$

Usually this value will be 100 due to trend and so we must adjust the chain relative trend.

(iv) The adjustment is done by substracting a correction factor from cash chain relative. Ltd. d 1/12 (Computed C.R. for Jan–100), than assuming linear trend the correction factor for Feb. Mar........, Dec. is d, 2d,.......11d respectively.

(v) Finally, adjust the correct chain relatives; total 1200 by expressing these as percentage of their arithmetic mean.

These values give the required seasonal indices.

Merits and Demerits

This method utilizes data more completely than moving average method.

There the loss is of only relative with the use of moving average method results in loss of six months at each end. The method is not simple and though the trend is eliminated by applying correction, the method is effective only if the growth is at a constant rate.

Example 5 :

Below are given the average quarterly prices of a commodity for given years. Calculates seasonal variation by method of link relatives.

Quarters

Year	I	II	III	IV
1966	30	26	22	31
1967	35	28	22	36
1968	31	29	28	32
1969	31	31	25	35
1970	34	36	26	35

Table 7

Calculation of Seasonal Indices by the Method of Link Relative

	Link Relative			
Year	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
1966	--	86.7	84.6	140.9
1967	112.9	80.8	78.6	163.6
1968	86.1	93.5	96.6	114.3
1969	96.6	100.0	80.7	140.0
1970	97.1	105.9	72.2	126.9
Arithmetic Average	$\frac{393.0}{4}$ = 98.24	$\frac{446.1}{5}$ = 89.22	$\frac{412.7}{5}$ = 82.52	$\frac{658.7}{5}$ = 137.14
Chain Relative	100	$\frac{89.22\times100}{100}$ = 89.22	$\frac{82.55\times89.22}{100}$ = 73.64	$\frac{137.14\times73.64}{100}$ = 100.68
Adjusted Chain Relatives	100	89.41	74.02	101.55
Seasonal Indices	106.45	95.68	79.21	108.76

The Second C.R. for the first quarter = $\frac{98.25 \div 100.98}{100}$ = 92.21

Correction = (99.21 - 100) = - 0.79

Average of adjusted chain relatives = $\frac{100.00+89.41+74.02+101.55}{24}$

= 93.49

Seasonal index for any quarter = $\frac{\text{Adjusted C.R.}}{93.49} \times 100$

(vi) To deseasonalize a series, for which we have an index of seasonal variation we divide the observed data by the index.

Measurement of Cyclical Variations :

Business cycles are perhaps the most important type of fluctuations in economic data. Certainly they have received a lot of attention but they are the most difficult-type of fluctuations to be measured. This is because successive cycles vary widely in timing, amplitude and pattern.

We shall now describe a crude measuring cyclical component of a time series. The method is called 'residual method' It consists, in removing for the given time series the three components i.e. trend, seasonal and irregular variation, in any order.

According to the multiplicative model, we have y = T × S × C × R. The trend value is first calculated by some method and indices are then obtained preferably by the method of moving averages.

y is then divided by T × S. The remainder will be C × R.

In the final stage, it is necessary to remove R from C × R by some process of smoothing. This is generally done by using moving averages of a suitable period. The resulting values give the cyclical component.

Since the method requires for its success a number of conditions to be satisfied, it is very seldom used.

A more sophisticated method of determining the cyclical component is the method of harmonic analysist which is too mathematical to be given here.

Random Component

No formula, however approximate, can be used to measure the random component directly at any point of the series. Usually the non-random component are determined and then a random residual which is left unaccounted for by these components is obtained. Even this becomes difficult when oscillations of irregular type appear in the series. However a method called the cariate-difference method enables us to estimate the variance of the random component. This enables us to estimate the extent of the randomness in the series. This method is too mathematical to be described here.

11.5 Books for Study

1. Elhance D.N	:	Fundamentals of Statistics.
2. Goon, A.M Gupta M.K. and Das Gupta B.	:	Fundamentals of Statictics (Vol. 1)
3. Gupta, S.P	:	Elementary Statistical Methods.
4. Mills, F.C	:	Statistical Methods

11.6 List of Questions

11.6.1 Short Questions

1. What is Time Series?
2. Distingush between seasonal variation and cyclical fluctions in a time series.
3. What is Secular Trend?
4. What is the different Components of a time series analysis?

11.6.2 Long Questions

1. What are the different components of an economic time series? How would you determine seasonal index?
2. The population of Tamil Nadu at the (Successive census) Years is given below

Year (x)	1911	1921	1931	1941	1951	1961	1971
Population(y) (in lakhs)	193	209	216	235	263	301	337

Fit a curve of the type y= ab

3. Fit a straight line trend by the methods of least squares of the following data

Year :	1	2	3	4	5	6	7	8	9	10
size of item	110	125	115	135	150	165	155	175	180	200

4. Fit y = ab to the following

x	:	1	2	3	4	5
Y	:	1.6	4.5	12.8	40.2	125

5. What are the different methods of present growth rates?