UNIT - V
CORRELATION, REGRESSION, INDEX NUMERS AND TIME SERIES ANALYSIS

Correlation coefficient:

The quantity r, called the linear correlation coefficient, measures the strength and the direction of a linear relationship between two variables. The linear correlation coefficient is sometimes referred to as the Pearson product moment correlation coefficient in honor of its developer Karl Pearson.

The correlation coefficient between two variables x and y is given by

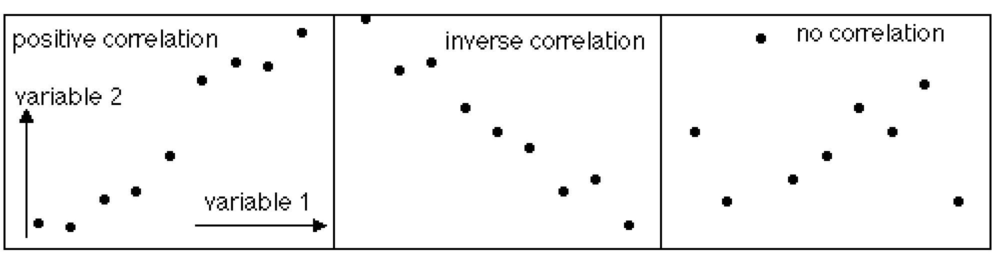
$$r = cov\frac{(x, y)}{\sqrt{var\ (x)\ var(y)}}$$

$$r = \frac{\sum x_i y_i \ - n\bar{x}\ \bar{y}}{\sqrt{(\sum x_i^2 \ - n\bar{x}^2)}\sqrt{(\sum y_i^2 \ - n\bar{y}^2)}}$$

where   is the number of pairs of data.

The value of r is such that $-1 < r < +1$. The $+$ and – signs are used for positive linear correlations and negative linear correlations, respectively.

A perfect correlation of ± 1 occurs only when the data points all lie exactly on a straight line. If $r = +1$, the slope of this line is positive. If $r = -1$, the slope of this line is negative.



Positive correlation:

If x and y have a strong positive linear correlation, r is close to $+1$. An r value of exactly $+1$ indicates a perfect positive fit. Positive values indicate a relationship between x and y variables such that as values for x increases, values for y also increase.

Negative correlation:

If x and y have a strong negative linear correlation, r is close to -1. An r value of exactly -1 indicates a perfect negative fit. Negative values indicate a relationship between x and y such that as values for x increase, values for y decrease.

No correlation:

If there is no linear correlation or a weak linear correlation, r is close to 0. A value near zero means that there is a random, nonlinear relationship between the two variables.

Note: r is a dimensionless quantity; that is, it does not depend on the units employed.

A correlation greater than 0.8 is generally described as *strong*, whereas a correlation less than 0.5 is generally described as *weak*. These values can vary based upon the "type" of data being examined. A study utilizing scientific data may require a stronger correlation than a study using social science data.

Coefficient of Determination r^2 or R^2 ***:***

The coefficient of determination, r^2, is useful because it gives the proportion of the variance (fluctuation) of one variable that is predictable from the other variable. It is a measure that allows us to determine how certain one can be in making predictions from a certain model/graph.

The coefficient of determination is the ratio of the explained variation to the total variation. The coefficient of determination is such that $0 < r^2 < 1$, and denotes the strength of the linear association between x and y. The coefficient of determination represents the percent of the data that is the closest to the line of best fit.

For example, if $r = 0.922$, then $r^2 = 0.850$, which means that 85% of the total variation in y can be explained by the linear relationship between x and y (as described by the regression equation). The other 15% of the total variation in y remains unexplained.

The coefficient of determination is a measure of how well the regression line represents the data. If the regression line passes exactly through every point on the scatter plot, it would be able to explain all of the variation. The further the line is away from the points, the less it is able to explain.

Example:

Emotion seems to play a pivotal role in determining popularity of a celebrity. In an exclusive survey made available to them by ad agency, Denstu-India, shows that the top 29 celebrity rankings are hugely impacted by the love/like quotient besides other parameters like performance. As per an article in Economic Times dt. 16^{th} October 2006, the following are the scores for some of the Indian celebrities for the years 2005 and 2006.

Celebrity	**Like Score 2005***	**Like Score 2006****
Rahul Dravid	59	53
Amitabh Bachchan	56	51
Sachin Tendulkar	43	50
Aishwarya Rai	56	50
Sania Mirza	21	49
Yuvaraj Singh	61	46
Sushmita Sen	56	46
Virendra Sehwag	64	46
Aamir Khan	57	45
Rani Mukherjee	57	45

Find the correlation coefficient between Like score 2005 and 2006.

Solution:

Here n=10

x	y	xy	x^2	y^2
59	53	3127	3481	2809
56	51	2856	3136	2601
43	50	2150	1849	2500
56	50	2800	3136	2500
21	49	1029	441	2401

61	46	2806	3721	2116
56	46	2576	3136	2116
64	46	2944	4096	2116
57	45	2565	3249	2025
57	45	2565	3249	2025
530	481	25418	29494	23209

$$\sum x_i = 530,\ \sum y_i = 481,\ \sum x_i^2 = 29494\ ,\ \sum y_i^2 = 23209,\ \sum x_i\, y_i = 25418$$

$$\boldsymbol{\bar{x}} = \frac{\sum x_i}{n} = \frac{530}{10} = \mathbf{53}$$

$$\boldsymbol{\bar{y}} = \frac{\sum y_i}{n} = \frac{481}{10} = \mathbf{48.1}$$

$$\rho = \frac{\sum x_i y_i - n\bar{x}\,\bar{y}}{\sqrt{(\sum x_i^2 - n\bar{x}^2)}\sqrt{(\sum y_i^2 - n\bar{y}^2)}}$$

$$\rho = \frac{25418 - 10 * 53 * 48.1}{\sqrt{(29494 - 10 * 53^2)\,(23209 - 10 * 48.1^2)}}$$

$$\rho = \frac{-75}{\sqrt{1404}\sqrt{72.9}}$$

$$\boldsymbol{\rho = -0.23}$$

Example:

The following data gives the closing prices of BSE Sensex, and the stock price of an individual company viz. ICICI bank during the 10 trading days during the period from 6th to 21st March 2006.

Date	SENSEX	ICICI Bank
6-3-2006	10735	613.20
7-3-2006	10725	600.65
8-3-2006	10509	590.55
9-3-2006	10574	601.75
10-3-2006	10765	612.90

13-3-2006	10804	603.10
16-3-2006	10878	607.50
17-3-2006	10860	605.25
20-3-2006	10941	605.40
21-3-2006	10905	597.80

Find the correlation coefficient between SENSEX and ICICI Bank.

Solution:

x	y	xy	x^2	y^2
10735	613.2	6582702	115240225	376014.2
10725	600.65	6441971	115025625	360780.4
10509	590.55	6206090	110439081	348749.3
10574	601.75	6362905	111809476	362103.1
10765	612.9	6597869	115885225	375646.4
10804	603.1	6515892	116726416	363729.6
10878	607.5	6608385	118330884	369056.3
10860	605.25	6573015	117939600	366327.6
10941	605.4	6623681	119705481	366509.2
10905	597.8	6519009	118919025	357364.8
107696	**6038.1**	**65031519**	**1160021038**	**3646281**

$$\sum x_i = 107696, \quad \sum y_i = \mathbf{6038.1}, \quad \sum x_i^2 = \mathbf{1160021038}, \quad \sum y_i^2 = \mathbf{3646281},$$

$$\sum x_i\, y_i = \mathbf{65031519}$$

$$\boldsymbol{\bar{x}} = \frac{\sum x_i}{n} = \frac{107696}{10} = \mathbf{10769.6}$$

$$\boldsymbol{\bar{y}} = \frac{\sum y_i}{n} = \frac{\mathbf{6038.1}}{10} = \mathbf{603.81}$$

$$\rho = \frac{\sum x_i y_i - n\bar{x}\,\bar{y}}{\sqrt{(\sum x_i^2 - n\bar{x}^2)}\sqrt{(\sum y_i^2 - n\bar{y}^2)}}$$

$$\rho = \frac{\mathbf{65031519} - 10 * 107696 * \mathbf{6038.1}}{\sqrt{(\mathbf{1160021038} - 10 * 107696^2)\,(\mathbf{3646281} - 10 * \mathbf{6038.1}^2)}}$$

$$\rho = \frac{3597.24}{\sqrt{178196.4}\sqrt{415.84}}$$

$$\boldsymbol{\rho = 0.42}$$

Regression analysis:

A statistical measure that attempts to determine the strength of the relationship between one dependent variable (usually denoted by Y) and a series of other changing variables (known as independent variables).

Types of Regression:

The two basic types of regression are linear regression and multiple regressions.

- Linear regression uses one independent variable to explain and/or predict the outcome of Y.
- Multiple regressions use two or more independent variables to predict the outcome.

The general form of each type of regression is:

Linear Regression: $Y = a + bX$

Multiple Regressions: $Y = a + b1X1 + b2\,X2 + b3X3 + \ldots + btXt$

Where Y= the variable that we are trying to predict

X= the variable that we are using to predict Y

a= the intercept

b= the slope

In multiple regressions the separate variables are differentiated by using subscripted numbers. Regression takes a group of random variables, thought to be predicting Y, and tries to find a mathematical relationship between them. This relationship is typically in the form of a straight line (linear regression) that best approximates all the individual data points. Regression is often used to determine how many specific factors such as the price of a commodity, interest rates, particular industries or sectors influence the price movement of an asset.

Finding the regression line using method of least squares:

The regression line y on x is given by $y = a + bx$

where

$$b = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}$$

$$a = \bar{y} - b\bar{x}$$

The regression line x on y is given by $x = c + dy$

Where

$$d = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum y_i^2 - n\bar{y}^2}$$

$$a = \bar{x} - d\bar{y}$$

$$\bar{x} = \frac{\sum x_i}{n} \quad and \quad \bar{y} = \frac{\sum y_i}{n}$$

and n is the number of pairs of data.

Standard error of estimator (Regression Line):

The square root of the residual variance is called the standard error of regression line. The residual variance for the regression line $y = a + bx$ is given by

$$\sigma_e^2 = \frac{\sum(y_i - \hat{y_i})^2}{n} \quad where \quad \hat{y_i} = a + bx_i$$

Example:

A tyre manufacturing company is interested in removing pollutants from the exhaust at the factory, and cost is a concern. The company has collected data from other companies concerning the amount of money spent on environmental measures and the resulting amount of dangerous pollutants released (as a percentage of total emissions)

Money spent (Rupees in lakhs)	8.4	10.2	16.5	21.7	9.4	8.3	11.5	18.4	16.7	19.3	28.4	4.7	12.3
Percentage of dangerous pollutants	35.9	31.8	24.7	25.2	36.8	35.8	33.4	25.4	31.4	27.4	15.8	31.5	28.9

a) Compute the regression equation.

b) Predict the percentage of dangerous pollutants released when Rs. 20,000 is spent on control measures.

c) Find the standard error of the estimate (regression line).

Solution:

Let x and y represents money spent and percentage of dangerous pollutants respectively.

Here $n = 13$

x	y	xy	x^2	$\bar{y} = a + bx$	$y - \bar{y}$	$(y - \bar{y})^2$
8.4	35.9	301.56	70.56	30.3472	5.5528	30.83359
10.2	31.8	324.36	104.04	30.1006	1.6994	2.88796
16.5	24.7	407.55	272.25	29.2375	-4.5375	20.58891
21.7	25.2	546.84	470.89	28.5251	-3.3251	11.05629
9.4	36.8	345.92	88.36	30.2102	6.5898	43.42546
8.3	35.8	297.14	68.89	30.3609	5.4391	29.58381
11.5	33.4	384.1	132.25	29.9225	3.4775	12.09301
18.4	25.4	467.36	338.56	28.9772	-3.5772	12.79636
16.7	31.4	524.38	278.89	29.2101	2.1899	4.795662
19.3	27.4	528.82	372.49	28.8539	-1.4539	2.113825
28.4	15.8	448.72	806.56	27.6072	-11.8072	139.41
4.7	31.5	148.05	22.09	30.8541	0.6459	0.417187
12.3	28.9	355.47	151.29	29.8129	-0.9129	0.833386
185.8	384	5080.27	3177.12			310.8354

a) $\bar{y} = a + bx$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{185.8}{13} = 14.29$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{384}{13} = 29.54$$

$$b = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}$$

$$= \frac{5080.27 - 13 * 14.29 * 29.54}{3177.12 - 13 * 14.29^2}$$

$$= \frac{-407.65}{2972.92} = -0.137$$

$$a = \bar{y} - b\bar{x} = 29.54 - (-0.137) \times 14.29 = 31.498$$

$$y = 31.498 - 0.137\, x$$

b) When Rs. 20,000 is spent on control then the percentage of dangerous pollutants released is

$$y = 31.498 - 0.137 \times 0.2$$

$$y = 31.471$$

c) Standard error of the estimate

$$\sigma_e^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n} \quad where \quad \hat{y}_i = a + bx_i$$

$$\sigma_e^2 = \frac{310.8354}{13} = 23.91$$

$$Standard\ error = \sqrt{\sigma_e^2} = \sqrt{23.91} = 4.89$$

Example:

A national level organization wishes to prepare a manpower plan based on the ever-growing sales offices in the country. Find the regression coefficient of Manpower on Sales Offices for the following data:

Year	Manpower	Sales Offices
2001	370	22
2002	386	25
2003	443	28
2004	499	31
2005	528	33
2006	616	38

Solution:

Let x and y represents sales offices and manpower respectively. Here $n = 6$,

x	y	xy	x^2
22	370	8140	484
25	386	9650	625
28	443	12404	784
31	499	15469	961
33	528	17424	1089
38	616	23408	1444
177	2842	86495	5387

$$\sum x_i = 177, \ \sum y_i = 2842, \ \sum x_i y_i = 86495, \ \sum x_i^2 = \mathbf{5387}$$

The regression line of Y on X is given by $Y = a + bX$

$$b = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{177}{6} = 29.5$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{2842}{6} = 473.67$$

$$b = \frac{86495 - 6 * 177 * 2842}{5387 - 6 * 29.5^2}$$

$$b = 16.04$$

$$a = \bar{y} - b\bar{x} = 473.67 - 16.04 \times 29.5 = 0.49$$

The regression line of manpower on sales offices is given by

$$y = 0.49 + 16.04\, x$$

Example:

The quantity of a raw material purchased by a company at the specified prices during the 12 months of 1992 is given

MONTH	PRICE/KG	QUANTITY (KG)
Jan	96	250

Feb	110	200
Mar	100	250
Aprl	90	280
May	86	300
June	92	300
July	112	220
Aug	112	220
Sep	108	200
Oct	116	210
Nov	86	300
Dec	92	250

a) Find the regression equation based on the above data

b) Can you estimate the appropriate quantity likely to be purchased if the price shoot upon Rs 124/kg?

c) Hence or otherwise obtain the coefficient of correlation between the price prevailing and the quantity demanded

Solution:

Let x and y represents price and quantity of a raw material purchased by the company respectively.

x	y	xy	x^2	y^2
96	250	24000	9216	62500
110	200	22000	12100	40000
100	250	25000	10000	62500
90	280	25200	8100	78400
86	300	25800	7396	90000
92	300	27600	8464	90000
112	220	24640	12544	48400
112	220	24640	12544	48400
108	200	21600	11664	40000

116	210	24360	13456	44100
86	300	25800	7396	90000
92	250	23000	8464	62500
1200	**2980**	**293640**	**121344**	**756800**

$\sum x_i = 1200, \ \sum y_i = 2920, \ \sum x_i y_i = 293640, \ \sum x_i^2 = \mathbf{121344}, \sum y_i^2 = \mathbf{756800}$

a) The regression line of Y on X is given by $Y = a + bX$

$$b = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{1200}{12} = 100$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{2920}{12} = 248.33$$

$$b = \frac{293640 - 12 * 100 * 248.33}{121344 - 12 * 100^2}$$

$$b = -3.24$$

$$a = \bar{y} - b\bar{x} = 248.33 — (3.24)\ 100$$

$$a = 572.33$$

The regression line of Price/kg on quantity in kg is given by

$$y = 572.33 - 3.24\ x$$

b) Given $x = 124/kg$ substituting this in (1) we get

$$y = 572.33 - 3.24\ (124)$$

$$y = 170.57$$

If the price is Rs $124/kg$, then the appropriate quantity likely to be purchased is approximately 171 kg.

c) To find Correlation coefficient:

$$\rho = \frac{\sum x_i y_i - n\bar{x}\,\bar{y}}{\sqrt{(\sum x_i^2 - n\bar{x}^2)}\sqrt{(\sum y_i^2 - n\bar{y}^2)}}$$

$$= \frac{293640 - 12 * 100 * 248.33}{\sqrt{(121344 - 12 * 100^2)}\sqrt{(756800 - 12 * 248.33^2)}}$$

$$\rho = -0.92$$

INDEX NUMBERS

An index number is an economic data figure reflecting price or quantity compared with a standard or base value.

Index numbers are names after the activity they measure. Their types are as under :

Price Index : Measure changes in price over a specified period of time. It is basically the ratio of the price of a certain number of commodities at the present year as against base year.

Quantity Index : As the name suggest, these indices pertain to measuring changes in volumes of commodities like goods produced or goods consumed, etc.

Value Index : These pertain to compare changes in the monetary value of imports, exports, production or consumption of commodities.

The base usually equals 100 and the index number is usually expressed as 100 times the ratio to the base value. For example, if a commodity costs twice as much in 1970 as it did in 1960, its index number would be 200 relative to 1960.

Index numbers are used especially to compare business activity, the cost of living, and employment. They enable economists to reduce unwieldy business data into easily understood terms.

In economics, index numbers generally are time series summarizing movements in a group of related variables. In some cases, however, index numbers may compare geographic areas at a point in time. An example is a country's purchasing power parity. The best-known index number is the consumer price index, which measures changes in retail prices paid by consumers. In addition, a cost-of-living index (COLI) is a price index number that measures relative cost of living over time. In contrast to a COLI based on the true but unknown utility function, a superlative index number is an index number that can be calculated. Thus, superlative index numbers are used to provide a fairly close approximation to the underlying cost-of-living index number in a wide range of circumstances.

Simple index numbers:

A simple index number Index numbers is a number that expresses the relative change in price, quantity, or value from one period to another.

Let p_0 be the base period price, and p_1 be the price at the selected or given period. Thus, the simple price index is given by:

$$P = \frac{p_1}{p_0}\,(100)$$

Example: The population of the Canadian province of British Columbia in 2004 was 4,196,400 and for Ontario it was 12,392,700. What is the population index of British Columbia compared to Ontario?

Solution:

$$Here\ p_1 = 4196400, p_0 = 12392700$$

$$Simple\ index\ P = \frac{p_1}{p_0}\,100\ = 419640012392700\ \times 100$$

$$P = 33.9$$

Weighted index:

$$Weighted\ Index\ Numbers\ = \times \frac{\sum index\ number * weight}{\sum weights}$$

There are two types of weighted indexes, they are

- Lapeyre's index
- Paasche' s index

Lapeyre's index:

$$Lapeyre's\ index\ = \frac{\sum p_1 q_0}{\sum p_0 q_0} * 100$$

$wher\ \ p_0\ \ is\ the\ price\ in\ the\ base\ period$

$p_1\ is\ the\ price\ in\ the\ current\ period$

$q_0\ \ is\ the\ quantity\ used\ in\ the\ base\ period$

Advantages: Requires quantity data from only the base period. This allows a more meaningful comparison over time. The changes in the index can be attributed to changes in the price.

Disadvantages: Does not reflect changes in buying patterns over time. Also, it may overweight goods whose prices increase.

Paasche's index:

$$Paasche's\ index = \frac{\sum p_1 q_1}{\sum p_0 q_1} * 100$$

$wher\ \ p_0\ \ is\ the\ price\ in\ the\ base\ period$

$p_1\ is\ the\ price\ in\ the\ current\ period$

$q_0\ \ is\ the\ quantity\ used\ in\ the$ current $period$

Advantages: Because it uses quantities from the current period, it reflects current buying habits.

Disadvantages: It requires quantity data for the current year. Because different quantities are used each year, it is impossible to attribute changes in the index to changes in price alone. It tends to overweight the goods whose prices have declined. It requires the prices to be recomputed each year.

Fisher' s index:

Laspeyres' index tends to overweight goods whose prices have increased. Paasche's index, on the other hand, tends to overweight goods whose prices have gone down. Fisher's ideal index was developed in an attempt to offset these shortcomings. It is the geometric mean of the Laspeyres and Paasche indexes.

$$Fisher's\ index = \sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} * \frac{\sum p_1 q_1}{\sum p_0 q_1}} * 100$$

$wher\ \ p_0\ \ is\ the\ price\ in\ the\ base\ period$

$p_1\ is\ the\ price\ in\ the\ current\ period$

$q_0\ \ is\ the\ quantity\ used\ in\ the$ base $period$

$q_1\ \ is\ the\ quantity\ used\ in\ the$ current $period$

Example:

The following table gives data about prices and consumption of four commodities A, B, C and D

Commodities	base year		current year	
	Price	quantity	Price	quantity
A	2	7	6	6
B	3	6	2	3
C	4	5	8	5
D	5	4	2	4

Find Laspeyre's index numbers, Paasche's index numbers and Fisher's index numbers.

Solution:

	p_0	q_0	p_1	q_1	p_0q_0	p_0q_1	p_1q_0	p_1q_1
A	2	7	6	6	14	12	42	36
B	3	6	2	3	18	9	12	6
C	4	5	8	5	20	20	40	40
B	5	4	2	4	20	20	8	8
Sum					72	61	102	90

$$Lapeyre\text{'}s\ index = \frac{\sum p_1q_0}{\sum p_0q_0} * 100 = \frac{102}{72} * 100 = 141.67$$

$$Paasche\text{'}s\ index = \frac{\sum p_1q_1}{\sum p_0q_1} * 100 = \frac{90}{61} * 100 = 147.54$$

$$Fisher\text{'}s\ index = \sqrt{\frac{\sum p_1q_0}{\sum p_0q_0} * \frac{\sum p_1q_1}{\sum p_0q_1}} * 100$$

$$= \sqrt{1.4167 * 1.4754} * 100 = 144.58$$

Unweighted indexes:

There are two methods to find unweighted indexes, they are

- Simple Average of the Price Indexes
- Simple Aggregate Index

Simple Average of the Price Indexes:

$$Simple\ Average\ of\ the\ Price\ Indexes = \frac{\sum p_i}{\boldsymbol{n}}$$

where p_i refer the simple index to each of the terms

Simple Aggregate Index:

$$Simple\ Aggregate\ Index = \frac{\sum p_1}{\sum p_0} * 100$$

where p_o is the base price in the base period.

p_1 is the price in the current period.

Example:

Find the Simple average of price index and simple aggregate index for the food price 1995=100 and 2005 (in rupees)

1995 price	**2005 price**
0.77	0.89
1.85	1.84
0.88	1.01
1.46	1.56
1.58	1.7
4.4	4.62

Solution:

$\mathbf{1995\ \ price\ p_0}$	$\mathbf{2005\ \ price\ p_1}$	$\mathbf{Simple\ Index = \frac{p_1}{p_0} * 100}$
0.77	0.89	115.6
1.85	1.84	99.5
0.88	1.01	114.8
1.46	1.56	106.8
1.58	1.7	107.6
4.4	4.62	105

$$Simple\ Average\ of\ the\ Price\ Indexes = \frac{\sum p_i}{\boldsymbol{n}} = \frac{649.3}{6} = 108.2$$

$$Simple\ Aggregate\ Index = \frac{\sum p_1}{\sum p_0} * 100 == \frac{11.62}{10.94} \times 100 = 106.2$$

Consumer price index:

Consumer price index (CPI) is a measure of change in the average retail prices of goods and services commonly purchased by a particular group of people in a particular area.

A consumer price index (CPI) measures changes through time in the price level of consumer goods and services purchased by households.

A measure that examines the weighted average of prices of a basket of consumer goods and services, such as transportation, food and medical care. The CPI is calculated by taking price changes for each item in the predetermined basket of goods and averaging them; the goods are weighted according to their importance. Changes in CPI are used to assess price changes associated with the cost of living.

Two basic types of data are needed to construct the CPI: price data and weighting data. The price data are collected for a sample of goods and services from a sample of sales outlets in a sample of locations for a sample of times. The weighting data are estimates of the shares of the different types of expenditure in the total expenditure covered by the index. These weights are usually based upon expenditure data obtained from expenditure surveys for a sample of households or upon estimates of the composition of consumption expenditure in the National Income and Product Accounts. Although some of the sampling of items for price collection is done using a sampling frame and probabilistic sampling methods, many items and outlets are chosen in a commonsense way (purposive sampling) that does not permit estimation of confidence intervals. Therefore, the sampling variance cannot be calculated. In any case, a single estimate is required in most of the purposes for which the index is used.

The index is usually computed monthly, or quarterly in some countries, as a weighted average of sub-indices for different components of consumer expenditure, such as food, housing, clothing, each of which is in turn a weighted average of sub-sub-indices. At the most detailed level, the elementary aggregate level, (for example, men's shirts sold in department stores in San Francisco), detailed weighting information is unavailable, so indices are computed using an unweighted arithmetic or geometric mean of the prices of the sampled product offers. (However, the growing use of scanner data is gradually making weighting information available even at the most detailed level.) These indices compare prices each month with prices in the price-reference month. The weights used to combine them into the higher-level aggregates, and then into the overall index, relate to the estimated expenditures during a preceding whole year of the consumers covered by the index on the products within its scope in the area covered. Thus the index is a fixed-weight index, but rarely a true Laspeyres index, since the weight-reference period of a year and the price-reference period, usually a more recent single month, do not coincide. It takes time to assemble and process the information used for weighting which, in addition to household expenditure surveys, may include trade and tax data.

Ideally, the weights would relate to the composition of expenditure during the time between the price-reference month and the current month. There is a large technical economics literature on index formulae which would approximate this and which can be shown to approximate what economic theorists call a true cost of living index. Such an index would show how consumer expenditure would have to move to compensate for price changes so as to allow consumers to maintain a constant standard of living. Approximations can only be computed retrospectively, whereas the index has to appear monthly and, preferably, quite soon. Nevertheless, in some countries, notably in the United States and Sweden, the philosophy of the index is that it is inspired by and approximates the notion of a true cost of living (constant utility) index, whereas in most of Europe it is regarded more pragmatically.

The coverage of the index may be limited. Consumers' expenditure abroad is usually excluded; visitors' expenditure within the country may be excluded in principle if not in practice; the rural population may or may not be included; certain groups such as the very rich or the very poor may be excluded. Saving and investment are always excluded, though the prices paid for financial services provided by financial intermediaries may be included along with insurance.

The index reference period, usually called the base year, often differs both from the weight-reference period and the price reference period. This is just a matter of rescaling the whole time-series to make the value for the index reference-period equal to 100. Annually revised weights are a desirable but expensive feature of an index, for the older the weights the greater is the divergence between the current expenditure pattern and that of the weight reference-period.

Calculating the CPI for a single item

$$\frac{CPI_2}{CPI_1} = \frac{Price_2}{Price_1}$$

where 1 is usually the comparison year and $CPI1$ is usually an index of 100.

Alternately, the CPI can be performed as

$$CPI = \frac{updated\ cost}{base\ period\ cost} * 100$$

The "updated cost" is the price of an item at a given year (say, the price of bread in 1982), divided by the initial year (the price of bread in 1970), and multiplied by one hundred.

Calculating the CPI for multiple items

Example: The prices of 95,000 items from 22,000 stores, and 35,000 rental units are added together and averaged. They are weighted this way: Housing: 41.4%, Food and Beverage: 17.4%, Transport: 17.0%, Medical Care: 6.9%, Other: 6.9%, Apparel: 6.0%, Entertainment: 4.4%. Taxes (43%) are not included in CPI computation.

$$CPI = \sum CPIn * weight$$

Weighting

Weights and sub-indices

Weights can be expressed as fractions or ratios summing to one, as percentages summing to 100 or as per mille numbers summing to 1000.

In the European Union's Harmonized Index of Consumer Prices, for example, each country computes some 80 prescribed sub-indices, their weighted average constituting the national Harmonized Index. The weights for these sub-indices will consist of the sum of the weights of a number of component lower level indices. The classification is according to use, developed in a national accounting context. This is not necessarily the kind of classification that is most appropriate for a Consumer Price Index. Grouping together of substitutes or of products whose prices tend to move in parallel might be more suitable.

For some of these lower level indexes detailed reweighing to make them be available, allowing computations where the individual price observations can all be weighted. This may be the case, for example, where all selling is in the hands of a single national organization which makes its data available to the index compilers. For lower level indexes, however, the weight will consist of the sum of the weights of a number of elementary aggregate indexes, each weight corresponding to its fraction of the total annual expenditure covered by the index. An 'elementary aggregate' is a lowest-level component of expenditure, one which has a weight but within which, weights of its sub-components are usually lacking. Thus, for example: Weighted averages of elementary aggregate indexes (e.g. for men’s shirts, raincoats, women’s dresses etc.) make up low level indexes (e.g. Outer garments),

Weighted averages of these in turn provide sub-indices at a higher, more aggregated level,(e.g. clothing) and weighted averages of the latter provide yet more aggregated sub-indices (e.g. Clothing and Footwear).

Some of the elementary aggregate indexes, and some of the sub-indexes can be defined simply in terms of the types of goods and/or services they cover, as in the case of such products as newspapers in some countries and postal services, which have nationally uniform prices. But where price movements do differ or might differ between regions or between outlet types, separate regional and/or outlet-type elementary aggregates are ideally required for each detailed category of goods and services, each with its own weight. An example might be an elementary aggregate for sliced bread sold in supermarkets in the Northern region.

Most elementary aggregate indexes are necessarily 'unweighted' averages for the sample of products within the sampled outlets. However in cases where it is possible to select the sample of outlets from which prices are collected so as to reflect the shares of sales to consumers of the different outlet types covered, self-weighted elementary aggregate indexes may be computed. Similarly, if the market shares of the different types of product represented by product types are known, even only approximately, the number of observed products to be priced for each of them can be made proportional to those shares.

Estimating weights

The outlet and regional dimensions noted above mean that the estimation of weights involves a lot more than just the breakdown of expenditure by types of goods and services, and the number of separately weighted indexes composing the overall index depends upon two factors:

1. The degree of detail to which available data permit breakdown of total consumption expenditure in the weight reference-period by type of expenditure, region and outlet type.
2. Whether there is reason to believe that price movements vary between these most detailed categories.

How the weights are calculated, and in how much detail, depends upon the availability of information and upon the scope of the index. In the UK the RPI does not relate to the whole of consumption, for the reference population is all private households with the exception of a) pensioner households that derive at least three-quarters of their total income from state pensions and benefits and b) "high income households" whose total household income lies within the top four per cent of all households. The result is that it is difficult to use data sources relating to total consumption by all population groups.

For products whose price movements can differ between regions and between different types of outlet:

- The ideal, rarely realizable in practice, would consist of estimates of expenditure for each detailed consumption category, for each type of outlet, for each region.
- At the opposite extreme, with no regional data on expenditure totals but only on population (e.g. 24% in the Northern region) and only national estimates for the shares of different outlet types for broad categories of consumption (e.g. 70% of food sold in supermarkets) the weight for sliced bread sold in supermarkets in the Northern region has to be estimated as the share of sliced bread in total consumption × 0.24 × 0.7.

The situation in most countries comes somewhere between these two extremes. The point is to make the best use of whatever data are available.

The nature of the data used for weighting

No firm rules can be suggested on this issue for the simple reason that the available statistical sources differ between countries. However, all countries conduct periodical Household Expenditure surveys and all produce breakdowns of Consumption Expenditure in their National Accounts. The expenditure classifications used there may however be different. In particular:

- Household Expenditure surveys do not cover the expenditures of foreign visitors, though these may be within the scope of a Consumer Price Index.
- National Accounts include imputed rents for owner-occupied dwellings which may not be within the scope of a Consumer Price Index.

Even with the necessary adjustments, the National Account estimates and Household Expenditure Surveys usually diverge.

The statistical sources required for regional and outlet-type breakdowns are usually weaker. Only a large-sample Household Expenditure survey can provide a regional breakdown. Regional population data are sometimes used for this purpose, but need adjustment to allow for regional differences in living standards and consumption patterns. Statistics of retail sales and market research reports can provide information for estimating outlet-type breakdowns, but the classifications they use rarely correspond to COICOP categories.

The increasingly widespread use of bar codes, scanners in shops has meant that detailed cash register printed receipts are provided by shops for an increasing share of retail purchases. This development makes possible improved Household Expenditure surveys, as Statistics Iceland has demonstrated. Survey respondents keeping a diary of their purchases need to record only the total of purchases when itemized receipts were given to them and keep these receipts in a special pocket in the diary. These receipts provide not only a detailed breakdown of purchases but also the name of the outlet. Thus response burden is markedly reduced, accuracy is increased, product description is more specific and point of purchase data are obtained, facilitating the estimation of outlet-type weights.

There are only two general principles for the estimation of weights: use all the available information and accept that rough estimates are better than no estimates.

Reweighing

Ideally, in computing an index, the weights would represent current annual expenditure patterns. In practice they necessarily reflect past expenditure patterns, using the most recent data available or, if they are not of high quality, some average of the data for more than one previous year. Some countries have used a three-year average in recognition of the fact that household survey estimates are of poor quality. In some cases some of the data sources used may not be available annually, in which case some of the weights for lower level aggregates within higher level aggregates are based on older data than the higher level weights.

Infrequent reweighing saves costs for the national statistical office but delays the introduction into the index of new types of expenditure. For example, subscriptions for Internet Service entered index compilation with a considerable time lag in some countries, and account could be taken of digital camera prices between re-weightings only by including some digital cameras in the same elementary aggregate as film cameras.

Producer price index:

The Producer Price Index (PPI) is a family of indexes that measure the average change over time in the prices received by domestic producers of goods and services. PPIs measure price change from the perspective of the seller. This contrasts with other measures, such as the Consumer Price Index (CPI). CPIs measure price change from the purchaser's perspective.

Sellers' and purchasers' prices can differ due to government subsidies, sales and excise taxes, and distribution costs.

The Producer Price Index (PPI) program measures the average change over time in the selling prices received by domestic producers for their output. The prices included in the PPI are from the first commercial transaction for many products and some services.

Sensex: (NIFTY, BSE, NSE)

The Sensex is an "index". i.e., an indicator. It gives you a general idea about whether most of the stocks have gone up or most of the stocks have gone down.

The Sensex (SENSITIVITY INDEX) is an indicator of all the major companies of the BSE (Bombay Stock Exchange). Sensex consists of the 30 largest and most actively traded stocks, representative of various sectors.

The Nifty (NATIONAL FIFTY) is an indicator of all the major companies of the NSE (National Stock Exchange). Nifty consists of 50 stocks ie. it has 50 listed companies.

If the Sensex goes up, it means that the prices of the stocks of most of the major companies on the BSE have gone up. If the Sensex goes down, this tells you that the stock price of most of the major stocks on the BSE have gone down.

Just like the Sensex represents the top stocks of the BSE, the Nifty represents the top stocks of the NSE.

The BSE is situated at Bombay and the NSE is situated at Delhi. These are the major stock exchanges in the country. There are other stock exchanges like the Calcutta Stock Exchange etc. but they are not as popular as the BSE and the NSE.

Most of the stock trading in the country is done though the BSE & the NSE. The 'BSE Sensex' or 'Bombay Stock Exchange' is value-weighted index composed of 30 stocks and was started in January 1, 1986. The Sensex is regarded as the pulse of the domestic stock markets in India. It consists of the 30 largest and most actively traded stocks, representative of various sectors, on the Bombay Stock Exchange. These companies account for around fifty per cent of the market capitalization of the BSE. The base value of the Sensex is *100* on April 1, 1979, and the base year of BSE-SENSEX is *1978-79*.

At regular intervals, the Bombay Stock Exchange (BSE) authorities review and modify its composition to be sure it reflects current market conditions. The index is calculated based on a free float capitalization method; a variation of the market cap method. Instead of using a company's outstanding shares it uses its float, or shares that are readily available for trading. The free-float method, therefore, does not include restricted stocks, such as those held by promoters, government and strategic investors.

Initially, the index was calculated based on the ‘full market capitalization’ method. However this was shifted to the free float method with effect from September 1, 2003. Globally, the free float market capitalization is regarded as the industry best practice.

As per free float capitalization methodology, the level of index at any point of time reflects the free float market value of 30 component stocks relative to a base period. The Market Capitalization of a company is determined by multiplying the price of its stock by the number of shares issued by the company. This Market capitalization is multiplied by a free float factor to determine the free float market capitalization. Free float factor is also referred as adjustment factor. Free float factor represent the percentage of shares that are readily available for trading.

The Calculation of Sensex involves dividing the free float market capitalization of 30 companies in the index by a number called Index divisor. The Divisor is the only link to original base period value of the Sensex. It keeps the index comparable over time and is the adjustment point for all Index adjustments arising out of corporate actions, replacement of scrips, etc.

The index has increased by over ten times from June 1990 to the present. Using information from April 1979 onwards, the long-run rate of return on the BSE Sensex works out to be 18.6% per annum, which translates to roughly 9% per annum after compensating for inflation.

Time series and Forecasting

Time series:

A sequence of numerical data points in successive order, usually occurring in uniform intervals. A time series is simply a sequence of numbers collected at regular intervals over a period of time.

Components of time series:

The four components of time series are:

1. Secular trend
2. Seasonal variation
3. Cyclical variation
4. Irregular variation

Secular trend:

A time series data may show upward trend or downward trend for a period of years and this may be due to factors like increase in population, change in technological progress, large scale shift in consumer’s demands, etc. For example, population increases over a period of time, price increases over a period of years, production of goods on the capital market of the country increases over a period of years. These are the examples of upward trend. The sales of a commodity may decrease over a period of time because of better products coming to the market. This is an example of declining trend or downward trend. The increase or decrease in the movements of a time series is called Secular trend.

Seasonal variation:

Seasonal variation is short-term fluctuation in a time series which occur periodically in a year. This continues to repeat year after year. The major factors that are responsible for the repetitive pattern of seasonal variations are weather conditions and customs of people. More woollen clothes are sold in winter than in the season of summer .Regardless of the trend we can observe that in each year more ice creams are sold in summer and very little in winter season. The sales in the departmental stores are more during festive seasons that in the normal days.

Cyclical variations:

Cyclical variations are recurrent upward or downward movements in a time series but the period of cycle is greater than a year. Also these variations are not regular as seasonal variation. There are different types of cycles of varying in length and size. The ups and downs in business activities are the effects of cyclical variation. A business cycle showing these oscillatory movements has to pass through four phases-prosperity, recession, depression and recovery. In a business, these four phases are completed by passing one to another in this order.

Irregular variation:

Irregular variations are fluctuations in time series that are short in duration, erratic in nature and follow no regularity in the occurrence pattern. These variations are also referred to as residual variations since by definition they represent what is left out in a time series after trend, cyclical and seasonal variations. Irregular fluctuations results due to the occurrence of unforeseen events like floods, earthquakes, wars, famines, etc.

Forecasting:

Forecasting involves the generation of a number, set of numbers, or scenario that corresponds to a future occurrence. It is absolutely essential to short-range and long-range planning. By definition, a forecast is based on past data, as opposed to a prediction, which is more subjective and based on instinct, gut feel, or guess. For example, the evening news gives the weather "forecast" not the weather "prediction." Regardless, the terms forecast and prediction is often used inter-changeably. For example, definitions of regression—a technique sometimes used in forecasting—generally state that its purpose is to explain or "predict."

Forecasting is based on a number of assumptions:

1. The past will repeat itself. In other words, what has happened in the past will happen again in the future.

2. As the forecast horizon shortens, forecast accuracy increases. For instance, a forecast for tomorrow will be more accurate than a forecast for next month; a forecast for next month will be more accurate than a forecast for next year; and a forecast for next year will be more accurate than a forecast for ten years in the future.
3. Forecasting in the aggregate is more accurate than forecasting individual items. This means that a company will be able to forecast total demand over its entire spectrum of products more accurately than it will be able to forecast individual stock-keeping units (SKUs). For example, General Motors can more accurately forecast the total number of cars needed for next year than the total number of white Chevrolet Impalas with a certain option package.
4. Forecasts are seldom accurate. Furthermore, forecasts are almost never totally accurate. While some are very close, few are "right on the money." Therefore, it is wise to offer a forecast "range." If one were to forecast a demand of 100,000 units for the next month, it is extremely unlikely that demand would equal 100,000 exactly. However, a forecast of 90,000 to 110,000 would provide a much larger target for planning.

A number of characteristics those are common to a good forecast:

- Accurate—some degree of accuracy should be determined and stated so that comparison can be made to alternative forecasts.
- Reliable—the forecast method should consistently provide a good forecast if the user is to establish some degree of confidence.
- Timely—a certain amount of time is needed to respond to the forecast so the forecasting horizon must allow for the time necessary to make changes.
- Easy to use and understand—users of the forecast must be confident and comfortable working with it.
- Cost-effective—the cost of making the forecast should not outweigh the benefits obtained from the forecast.

Forecasting techniques range from the simple to the extremely complex. These techniques are usually classified as being qualitative or quantitative.

QUALITATIVE TECHNIQUES

Qualitative forecasting techniques are generally more subjective than their quantitative counterparts. Qualitative techniques are more useful in the earlier stages of the product life cycle, when less past data exists for use in quantitative methods. Qualitative methods include the Delphi technique, Nominal Group Technique (NGT), sales force opinions, executive opinions, and market research.

QUANTITATIVE TECHNIQUES

Quantitative forecasting techniques are generally more objective than their qualitative counterparts. Quantitative forecasts can be time-series forecasts (i.e., a projection of the past into the future) or forecasts based on associative models (i.e., based on one or more explanatory variables). Time-series data may have underlying behaviors that need to be identified by the forecaster. In addition, the forecast may need to identify the causes of the behavior. Some of these behaviors may be patterns or simply random variations. Among the patterns are:

- Trends, which are long-term movements (up or down) in the data.
- Seasonality, which produces short-term variations that are usually related to the time of year, month, or even a particular day, as witnessed by retail sales at Christmas or the spikes in banking activity on the first of the month and on Fridays.
- Cycles, which wavelike variations are lasting more than a year that is usually tied to economic or political conditions.
- Irregular variations that do not reflect typical behavior, such as a period of extreme weather or a union strike.
- Random variations, which encompass all non-typical behaviors not accounted for by the other classifications.

The simple technique is the use of averaging. To make a forecast using averaging, one simply takes the average of some number of periods of past data by summing each period and dividing the result by the number of periods. This technique has been found to be very effective for short-range forecasting.

Variations of averaging include the moving average, the weighted average, and the weighted moving average. A moving average takes a predetermined number of periods, sums their actual demand, and divides by the number of periods to reach a forecast. For each subsequent period, the oldest period of data drops off and the latest period is added. Assuming a three-month moving average and using the data from Table given below,

Period	Actual Demand (000's)
January	**45**
February	**60**
March	**72**
April	**58**
May	**40**

one would simply add 45 (January), 60 (February), and 72 (March) and divide by three to arrive at a forecast for April:

$$45 + 60 + 72 = 177 \div 3 = 59$$

To arrive at a forecast for May, one would drop January's demand from the equation and add the demand from April. The following table presents an example of a three-month moving average forecast.

Three Month Moving Average Forecast

Period	Actual Demand (000's)	Forecast (000's)
January	45	
February	60	
March	72	
April	58	59
May	40	63
June		57

A weighted average applies a predetermined weight to each month of past data, sums the past data from each period, and divides by the total of the weights. If the forecaster adjusts the weights so that their sum is equal to 1, then the weights are multiplied by the actual demand of each applicable period. The results are then summed to achieve a weighted forecast. Generally, the more recent the data the higher the weight, and the older the data the smaller the weight. Using the demand example, a weighted average using weights of 0.4, 0 .3, 0.2, and 0.1 would yield the forecast for June as:

$$60(0.1) + 72(0.2) + 58(0.3) + 40(0.4) = 53.8$$

Forecasters may also use a combination of the weighted average and moving average forecasts. A weighted moving average forecast assigns weights to a predetermined number of periods of actual data and computes the forecast the same way as described above. As with all moving forecasts, as each new period is added, the data from the oldest period is discarded. The following table shows a three-month weighted moving average forecast utilizing the weights 0.5, 0.3, and 0.2.

Three–Month Weighted Moving Average Forecast

Period	Actual Demand (000's)	Forecast (000's)
January	45	
February	60	
March	72	
April	58	59
May	40	63
June		61

Exponential smoothing:

A more complex form of weighted moving average is exponential smoothing, so named because the weight falls off exponentially as the data ages. Exponential smoothing takes the previous period's forecast and adjusts it by a predetermined smoothing constant, ά (called alpha; the value for alpha is less than one) multiplied by the difference in the previous forecast and the demand that actually occurred during the previously forecasted period (called forecast error). Exponential smoothing is expressed formulaically as such: New forecast = previous forecast + alpha (actual demand − previous forecast)

$$F_{t+1} = F_t + \alpha(X_t - F_t)$$

Exponential smoothing requires the forecaster to begin the forecast in a past period and work forward to the period for which a current forecast is needed. A substantial amount of past data and a beginning or initial forecast are also necessary. The initial forecast can be an actual forecast from a previous period, the actual demand from a previous period, or it can be estimated by averaging all or part of the past data.

Trend analysis:

Associative or causal techniques involve the identification of variables that can be used to predict another variable of interest. For example, interest rates may be used to forecast the demand for home refinancing. Typically, this involves the use of linear regression, where the objective is to develop an equation that summarizes the effects of the predictor (independent) variables upon the forecasted (dependent) variable. If the predictor variable were plotted, the object would be to obtain an equation of a straight line that minimizes the sum of the squared deviations from the line (with deviation being the distance from each point to the line). The equation would appear as: $y = a + bx,$ where y is the predicted (dependent) variable, x is the predictor (independent) variable, b is the slope of the line, and a is equal to the height of the line at the y-intercept. Once the equation is determined, the user can insert current values for the predictor (independent) variable to arrive at a forecast (dependent variable).

If there is more than one predictor variable or if the relationship between predictor and forecast is not linear, simple linear regression will be inadequate. For situations with multiple predictors, multiple regressions should be employed, while non-linear relationships call for the use of curvilinear regression.

ECONOMETRIC FORECASTING

Econometric methods, such as autoregressive integrated moving-average model (ARIMA), use complex mathematical equations to show past relationships between demand and variables that influence the demand. An equation is derived and then tested and fine-tuned to ensure that it is as reliable a representation of the past relationship as possible. Once this is done, projected values of the influencing variables (income, prices, etc.) are inserted into the equation to make a forecast.

MAKING A FORECAST

The basic steps in the forecasting process:

- Determine the forecast's purpose. Factors such as how and when the forecast will be used, the degree of accuracy needed, and the level of detail desired determine the cost (time, money, employees) that can be dedicated to the forecast and the type of forecasting method to be utilized.

- Establish a time horizon. This occurs after one has determined the purpose of the forecast. Longer-term forecasts require longer time horizons and vice versa. Accuracy is again a consideration.
- Select a forecasting technique. The technique selected depends upon the purpose of the forecast, the time horizon desired, and the allowed cost.
- Gather and analyze data. The amount and type of data needed is governed by the forecast's purpose, the forecasting technique selected, and any cost considerations.
- Make the forecast.
- Monitor the forecast. Evaluate the performance of the forecast and modify, if necessary.

Example:

The following data given below relates to quarterly deposits of a commercial bank.

Deposits of a Commercial Bank

(Rs in Hundreds of Crores)

Year	March	June	September	December
2003	391	439	452	480
2004	509	562	572	622
2005	625	685	687	745
2006	743	808	805	867

Find i) the secular trend using method of least squares and also forecast the deposit in the year 2007.

ii) Find the seasonal indicates of various quarters using 4-Moving average method.

Solution:

t	y	t_y	t_2
1	391	391	1
2	439	878	4
3	452	1356	9

4	**480**	**1920**	**16**
5	**509**	**2545**	**25**
6	**562**	**3372**	**36**
7	**572**	**4004**	**49**
8	**622**	**4976**	**64**
9	**625**	**5625**	**81**
10	**685**	**6850**	**100**
11	**687**	**7557**	**121**
12	**745**	**8940**	**144**
13	**743**	**9659**	**169**
14	**808**	**11312**	**196**
15	**805**	**12075**	**225**
16	**867**	**13872**	**256**
136	**9992**	**95332**	**1496**

$$\sum t_i = 136, \quad \sum y_i = 9992, \quad \sum t_i y_i = 95332, \quad \sum t_i^2 = 1496\,, \quad n = 16$$

The regression line of Y on X is given by

$$y = a + bt$$

where

$$b = \frac{\sum t_i y_i - n\bar{t}\,\bar{y}}{\sum t_i^2 - \bar{t}^2}$$

$$\bar{t} = \frac{\sum ti}{n} = \frac{136}{16} = 8\,.5$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{9992}{16} = 624\,.5$$

$$b = \frac{95332 - 16 * 8.5 * 624.5}{1496 - 16 * (8.5)^2}$$

$$b = 30.59$$

$$a = \bar{y} - b\bar{x} = 624.5 - (30.59) \times 8.5$$

$$a = 364.49$$

The regression line of electric supply on demand for electric motors is given by

$$y = 364.49 + 30.59\, t$$

4-Moving average method:

Year		Deposits (1)	Moving average of 4 Quarters (2)	Centered Moving average of 4 Quarters (3)	Ratio to Moving average = (1)/(3)*100 (4)	Deseasonalisation = Original value/Seasonal index *100
2003	March	391				398.98
	June	439				429.13
	September	452	440.5			458.28
	December	480	470	455.25	99.29	475.62
2004	March	509	500.75	485.38	98.89	519.39
	June	562	530.75	515.75	98.69	549.36
	September	572	566.25	548.5	102.46	579.95
	December	622	595.25	580.75	98.49	616.33
2005	March	625	626	610.63	101.86	637.76
	June	685	654.75	640.38	97.6	669.6
	September	687	685.5	670.13	102.22	696.54
	December	745	715	700.25	98.11	738.21
2006	March	743	745.75	730.38	102	758.19
	June	808	775.25	760.5	97.7	789.83
	September	805	805.75	790.5	102.21	816.18
	December	867				859.1

Seasonal Index:

	March	June	September	December
2003			99.29	98.89
2004	98.69	102.46	98.49	101.86
2005	97.6	102.22	98.11	102
2006	97.7	102.21		
Average	**98**	**102.3**	**98.63**	**100.92**

Quarter	Seasonal Index
March	**98**
June	**102.3**
September	**98.63**
December	**100.92**

Example:

The data relates to the actual sales volume of a company during the period from January 2006 to December 2006.

Month	Jan	Feb	March	April	May	June	July	Aug	Sept	Oct	Nov	Dec
Sales	45	55	45	55	60	55	75	50	70	65	80	65

Forecast for the values of $\alpha = 0.2$ assuming the forecast for January 2006 to be 40 units by exponential smoothing method.

Solution:

Given $\alpha = 0.2$, and $F_t = 40$ $for\ January$ 2006

We know that

$$F_{t+1} = F_t + \alpha(X_t - F_t)$$

Month	$\boldsymbol{X_t}$	$\boldsymbol{F_t}$	$\boldsymbol{X_t - F_t}$	$\boldsymbol{\alpha\,(X_t - F_t)}$	$\boldsymbol{F_t + 1 = F_t + \alpha(X_t - F_t)}$
Jan	45	40	5	1	41
Feb	55	41	14	2.8	43.8
March	45	43.8	1.2	0.24	44.04
April	55	44.04	10.96	2.19	46.23

May	60	46.23	13.77	2.75	48.98
June	55	48.98	6.02	1.2	50.18
July	75	50.18	24.82	4.96	55.14
Aug	50	55.14	−5.14	−1.03	54.11
Sep	70	54.11	15.89	3.18	57.29
Oct	65	57.29	7.71	1.54	58.83
Nov	80	58.83	21.17	4.23	63.06
Dec	65	63.06	1.94	0.39	63.45

Therefore the forecast sales value for January 2007 is Rs. 63.45 Lakhs.