

# Correlation

Himadri Barman

Centre for Management Studies, Dibrugarh University

Correlation is the monotonic relationship between any two variables X and Y. Monotonic relationship here refers to any of the following: An increase in one leads to increase in the other or an increase in one leads to decrease in the other. There is no monotonic relationship if there is no increase/decrease because of increase/decrease in the other. Correlation may be numerically represented using the *Karl Pearson's correlation coefficient*. It is represented using the symbol  $r$ . The value of  $r$  will lie between -1 and +1. +1 refers to perfect positive correlation, -1 refers to perfect negative correlation and 0 refers to no correlation. Any positive value of  $r$  is positive correlation while negative values of  $r$  refer to negative correlation.

Without calculating the value also, we can have an idea of the relationship using a special type of graph called scatter diagrams. The following dataset is an example of perfect positive correlation.

X	1	2	3	4	5	6	7	8	9
Y	2	4	6	8	10	12	14	16	18

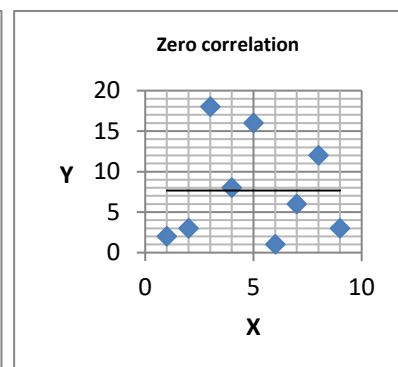
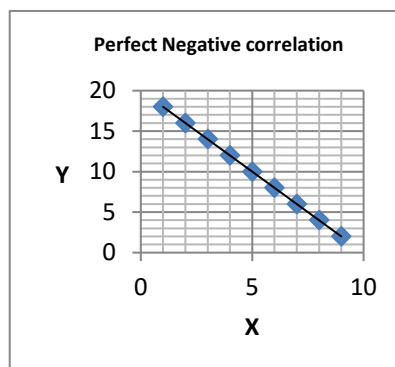
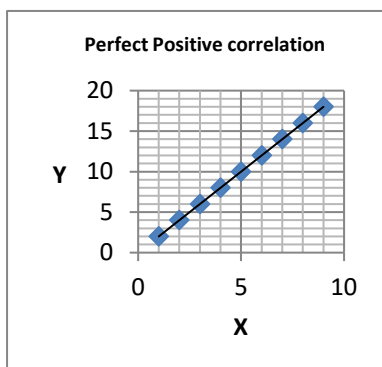
The dataset below shows perfect negative correlation.

X	1	2	3	4	5	6	7	8	9
Y	18	16	14	12	10	8	6	4	2

The dataset below is an example of zero correlation.

X	1	2	3	4	5	6	7	8	9
Y	2	3	18	8	16	1	6	12	3

The graphs below illustrate perfect positive correlation, perfect negative correlation and zero correlation for the above datasets.



Graphs generated using MS Excel

Karl Pearson's Correlation Coefficient is given by the formula:

$$r = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X) \cdot \text{var}(Y)}} = \frac{\frac{\sum(X - \bar{X})(Y - \bar{Y})}{n}}{\sqrt{\frac{\sum(X - \bar{X})^2}{n} \cdot \frac{\sum(Y - \bar{Y})^2}{n}}} = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum(X - \bar{X})^2 \sum(Y - \bar{Y})^2}}$$

$\bar{X}$  and  $\bar{Y}$  are the averages (means) of the values of  $X$  and  $Y$  respectively

The process for calculating the correlation coefficient is stated below using a tabular form:

$X$	$Y$	$(X - \bar{X})$	$(Y - \bar{Y})$	$(X - \bar{X})(X - \bar{X})$	$(Y - \bar{Y})(Y - \bar{Y})$	$(X - \bar{X})(Y - \bar{Y})$
3	4	-3	-1	9	1	3
8	6	2	1	4	1	2
6	7	0	2	0	4	0
8	5	2	0	4	0	0
5	3	-1	-2	1	4	2
$\bar{X} = 6$	$\bar{Y} = 5$			$\sum(X - \bar{X})^2 = 18$	$\sum(Y - \bar{Y})^2 = 10$	$\sum(X - \bar{X})(Y - \bar{Y}) = 7$

$$\bar{X} = \frac{3 + 8 + 6 + 8 + 5}{5} = 6; \bar{Y} = \frac{4 + 6 + 7 + 5 + 3}{5} = 5$$

$$\text{So, } r = \frac{7}{\sqrt{18 \cdot 10}} = \frac{7}{\sqrt{180}} = 0.521749$$

*Spearman's Rank correlation coefficient (also called Spearman's rho)* is used when the data is ordinal (rank) in nature. For example, consider the ranks obtained by nine students in CAM and QT as follows:

Rank in CAM	3	5	1	6	7	2	8	9	4
Rank in QT	5	3	2	6	8	1	7	9	4

Here, we will use the rank correlation coefficient. The formula is given by:

$$R = 1 - \frac{6 \sum D^2}{n(n^2 - 1)}$$

The process for calculating the rank correlation coefficient is stated below using a tabular form where  $D$  is the difference in rank and  $n$  is the number of values.

Rank in CAM ( $X$ )	Rank in QT ( $Y$ )	$D = X - Y$	$D^2$
3	5	-2	4
5	3	2	4
1	2	-1	1
6	6	0	0
7	8	-1	1
2	1	1	1
8	7	1	1
9	9	0	0
4	4	0	0
			$\sum D^2 = 12$

$$S_{o,R} = 1 - \frac{6.12}{9(81 - 1)} = 1 - \frac{72}{720} = 0.9$$

When equal ranks are associated with some entries, an adjustment to the original formula is made by adding  $\frac{1}{12}(m^3 - m)$  to the value of  $\sum D^2$ , where  $m$  stands for the number of items whose ranks are same. The revised formula is thus:

$$S_{o,R} = 1 - \frac{6\{\sum D^2 + 1/12(m_1^3 - m_1) + 1/12(m_2^3 - m_2) + \dots\}}{n(n^2 - 1)}$$

We take an example below where marks for the students are given.

Marks in CAM	56	64	64	70	50	34	60	65
Marks in QT	60	70	50	70	65	50	45	55

We will first find the rank and proceed further. The highest marks is given Rank 1. We see that in CAM, two students have the same marks 64 and they would occupy places 3 and 4. The average of 3 and 4 is 3.5 and so both are given a rank of 3.5. Similarly in QT, students getting 70 are occupying places 1 and 2 and 3, the average of which is 2. So, all get a rank of 2.

Marks in CAM	Rank in CAM (X)	Marks in QT	Rank in QT (Y)	D = X - Y	D <sup>2</sup>
56	6	60	5	1	1
64	3.5	70	2	1.5	2.25
64	3.5	50	7	-3.5	12.25
70	1	70	2	-1	1
50	7	65	4	3	9
34	8	70	2	6	36
60	5	45	8	-3	9
65	2	55	6	-4	16
					$\sum D^2 = 86.5$

Here  $m_1 = 2$  and  $m_2 = 3$

$$S_{o,R} = 1 - \frac{6\{86.5 + 1/12(2^3 - 2) + 1/12(3^3 - 3)\}}{8(64 - 1)} = 1 - \frac{534}{504} = 1 - 1.05952 = -0.05952$$

Karl Pearson's correlation coefficient can be found out in MS Excel using the function CORREL. There is no separate function to find the Spearman's rank correlation coefficient. To find the Spearman's correlation coefficient, use the CORREL function with the ranked data.

It is to be noted that for the above example, when using the CORREL function of Excel to find Spearman's correlation coefficient, Excel will give a value of -0.06136. This is because Excel uses a different formula to calculate the Spearman's correlation coefficient. It uses the

formula of Karl Pearson's correlation coefficient only by replacing the values of the variables with their corresponding ranks, i.e.,

$$\frac{\sum(R_X - \bar{R}_X)(R_Y - \bar{R}_Y)}{\sqrt{\sum(R_X - \bar{R}_X)^2 \sum(R_Y - \bar{R}_Y)^2}}$$

The Spearman's correlation is the non-parametric version of the Karl Pearson's correlation. While Karl Pearson's gives us a linear relationship (a stricter version of monotonic relationship where variables change in the same direction at a constant rate), Spearman's gives us a general monotonic relationship where variables tend to change together but not necessarily at a constant rate. Spearman's correlation coefficient should be used when the underlying assumptions of Karl Pearson's correlation coefficient are not met:

- If the data exhibits a non-linear relationship or is not normal.
- If at least one variable is ordinal.
- If there are significant outliers in the data set. Spearman's correlation coefficient is not sensitive to outliers because it performs calculations on the ranks, so the difference between actual values has no meaning.

# Regression

Himadri Barman

Centre for Management Studies, Dibrugarh University

---

Regression analysis is a technique of estimation where we estimate the value of the dependent variable if we know the value of the independent variable(s). The technique exploits the statistical relationship between the two variables. If there is only one independent variable, we call it linear regression. In linear regression, we have two lines of regression. If the independent variable is considered as  $X$  and the dependent variable is  $Y$ , we determine the values of  $Y$  from the values of  $X$ . For this, we construct the regression line of  $Y$  on  $X$ . Similarly, if we want to determine the values of  $X$  from the values of  $Y$ , we have the regression line of  $X$  on  $Y$ .

Regression line of  $Y$  on  $X$  is given by  $Y = a + bX$ . Here  $a$  and  $b$  are constants which we can find by solving the following two equations, where  $n$  is the number of data sets given.

$$\sum Y = na + b\sum X$$

$$\sum XY = a\sum X + b\sum X^2$$

Let us look at the following dataset:

X	6	2	10	4	8
Y	9	11	5	8	7

We prepare the following table

X	Y	$X^2$	$XY$
6	9	36	54
2	11	4	22
10	5	100	50
4	8	16	32
8	7	64	56
$\sum X = 30$	$\sum Y = 40$	$\sum X^2 = 220$	$\sum XY = 214$

The equations formed are:  $40 = 5a + 30b$  and  $214 = 30a + 220b$ .

On solving the two equations, we get  $a = 11.9, b = -0.65$ .

Therefore, the required regression line of  $Y$  on  $X$  is  $Y = 11.9 - 0.65X$ .

Similarly, the regression line of  $X$  on  $Y$  is given by  $X = c + dY$ . Here  $c$  and  $d$  are constants which we can find by solving the following two equations, where  $n$  is the number of data sets given.

$$\sum X = nc + d\sum Y$$

$$\sum XY = c\sum Y + d\sum Y^2$$

Taking the previous data set, we prepare the following table:

X	Y	Y <sup>2</sup>	XY
6	9	81	54
2	11	121	22
10	5	25	50
4	8	64	32
8	7	49	56
$\sum X = 30$	$\sum Y = 40$	$\sum X^2 = 340$	$\sum XY = 214$

The equations formed are:  $30 = 5a + 40b$  and  $214 = 40a + 340b$ .

On solving the two equations, we get  $a = 16.4, b = -1.3$ .

Therefore, the required regression line of X on Y is  $X = 16.4 - 1.3Y$ .

We can represent the lines of regression using regression coefficients. The regression line of Y on X is given by  $Y - \bar{Y} = b_{yx} (X - \bar{X})$ .

$b_{yx}$  is the regression coefficient of Y on X and  $b_{yx} = r \frac{\sigma_y}{\sigma_x}$ , where  $r$  is the correlation coefficient,  $\sigma_y$  is the standard deviation of Y and  $\sigma_x$  is the standard deviation of X. The formula to find  $b_{yx}$  is given by:

$$b_{yx} = \frac{\sum XY - \frac{\sum X \sum Y}{n}}{\sum X^2 - \frac{(\sum X)^2}{n}}$$

Similarly, the regression line of X on Y is given by  $X - \bar{X} = b_{xy} (Y - \bar{Y})$ .

$b_{xy}$  is the regression coefficient of X on Y and  $b_{xy} = r \frac{\sigma_x}{\sigma_y}$ , where  $r$  is the correlation coefficient,  $\sigma_y$  is the standard deviation of Y and  $\sigma_x$  is the standard deviation of X. The formula to find  $b_{xy}$  is given by:

$$b_{xy} = \frac{\sum XY - \frac{\sum X \sum Y}{n}}{\sum Y^2 - \frac{(\sum Y)^2}{n}}$$

The relationship between  $b_{yx}, b_{xy}$  and  $r$  is given by  $r = \sqrt{b_{yx} \cdot b_{xy}}$

$b_{yx}$  and  $b_{xy}$  must be of the same sign, i.e., either both should be positive or both should be negative.

Let us take the following dataset to work out a problem for regression coefficients.

X	1	2	3	4	5	6	7	8	9	10
Y	2	4	8	7	10	5	14	16	2	20

We prepare the following table:

X	Y	$X^2$	$Y^2$	XY
1	2	1	4	2
2	4	4	16	8
3	8	9	64	24
4	7	16	49	28
5	10	25	100	50
6	5	36	25	30
7	14	49	196	98
8	16	64	256	128
9	2	81	4	18
10	20	100	400	200
$\Sigma X = 55$	$\Sigma Y = 88$	$\Sigma X^2 = 385$	$\Sigma Y^2 = 1114$	$\Sigma XY = 586$

$$b_{yx} = \frac{586 - \frac{(55)(88)}{10}}{385 - \frac{(55)(55)}{10}} = \frac{102}{82.5} = 1.24$$

$$b_{xy} = \frac{586 - \frac{(55)(88)}{10}}{1114 - \frac{(88)(88)}{10}} = \frac{102}{339.6} = 0.30$$

The line of regression of Y on X is:

$$Y - 8.8 = 1.24 (X - 5.5). \left[ \bar{Y} = \Sigma Y/n = \frac{88}{10} = 8.8; \bar{X} = \Sigma X/n = \frac{55}{10} = 5.5 \right]$$

The line of regression of X on Y is:

$$X - 5.5 = 0.30 (Y - 8.8)$$

$$\text{And } r = \sqrt{(1.24)(0.30)} = 0.609$$

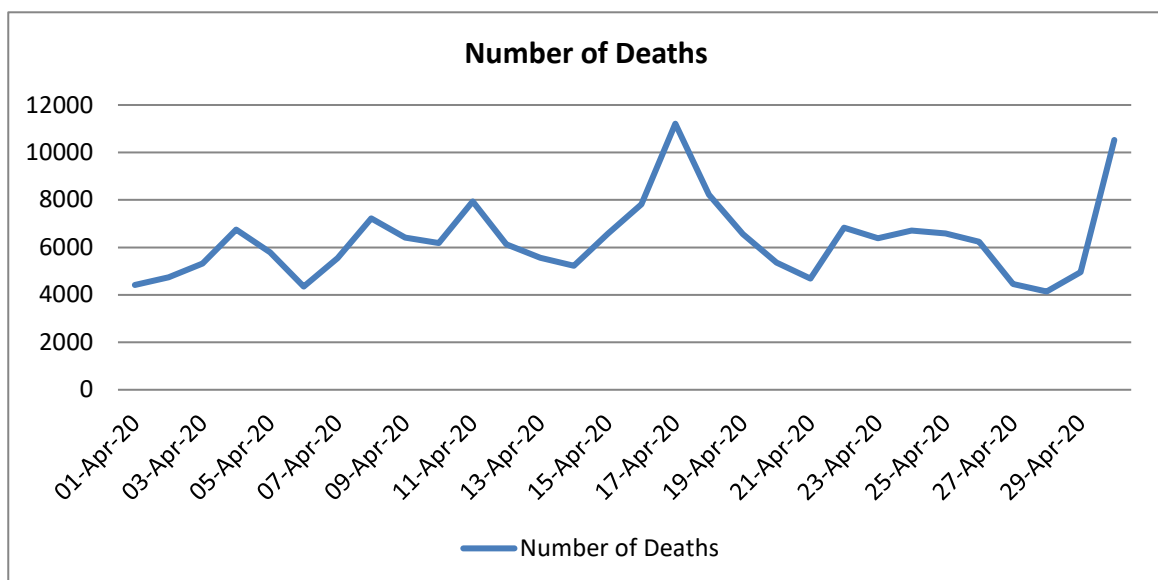
# Time Series

Himadri Barman

Centre for Management Studies, Dibrugarh University

Numerical data which are collected, observed or recorded at successive intervals or points of time is referred to as time series data. The number of deaths recorded each day of the Covid 19 pandemic is an example of time series data. Time sequence is of prime importance and the past data are analysed so as to understand the future better and plan accordingly. Time series data often require to be subjected to some preliminary adjustments before they may be analysed. Monthly infection data for the Covid 19 pandemic are not comparable because the number of days in each month is not the same. Hence, this data has to be brought to per-day basis before they can be put to further analysis.

Time series analysis provides knowledge about the fluctuations in economic and business phenomena. The causes of changes may also be found out. Time series analysis is a good tool for forecasting as past trends can be projected into future trends.



A graphical representation of a time series represents the changes over time as shown above. If there are no changes, we will get a horizontal line. Usually, time series data will show continual changes over time, giving us the impression of a haphazard movement overall. A critical study of the series will, however, reveal that the changes are not totally haphazard and a part of the variation can be accounted for. The part which can be accounted for is the systematic one and the remaining the unsystematic or irregular. The systematic part is attributed to secular trend, seasonal variation and cyclical variation. Thus there are four components of a time series, three of the systematic and the fourth the irregular variation.

Secular trend (or simply trend) is the smooth, regular, long-term movement of the series if observed long enough. The trend may be upward, downward or may remain more or less constant. The upward or downward trend may be seen over a longer period but frequent

changes are incompatible with the idea of a trend. By seasonal variation or fluctuation, we mean a periodic movement in a time series where the period is no longer than a year. Examples of seasonal variations are found in people visiting an ATM in a 24-hour period, sales in Big Bazaar over a one-year period, etc. By cyclical variation, we mean the oscillatory movement in a time series such that the period of oscillation is more than a year. One complete period is called a cycle. Cyclical variations are not necessarily periodic. Example is boom and depression in a business which follow one another in an irregular manner. Irregular variations are those which are either wholly unaccountable or caused by such unforeseen events as wars, floods, pandemics, etc. This category includes everything not accounted by the other three.

Two basic models are used for the analysis:

- Additive Model:  $O = T + S + C + I$
- Multiplicative Model:  $O = T \times S \times C \times I$

T, S, C, and I represent the Trend, Seasonal variation, Cyclical variation and Irregular variation respectively.

Following techniques / formulas can be used to calculate the variations:

- Trend – Graphic or Free hand method, Semi-average method, moving average method, method of least squares
- Seasonal variation – Method of monthly averages, ratio to trend method, ratio to moving average method, link relative method
- Seasonal variation =  $\frac{\sum(\text{original data} - \text{trend values})}{\text{number of terms (seasons)}}$
- Irregular variation = *short time oscillations – seasonal variations*

Let us calculate the four yearly moving average from the following production data (in '000 kgs):

Year	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Production	464	515	518	467	502	540	557	571	586	612

The process is simple. We find the average of four years at a time starting from 2006. So, the first average is found for the period 2006 – 2009. Subsequent periods are 2007 – 2010, 2008 – 2011, 2009 – 2012, and so on.

Year	Production	Moving Average
2006	464	---
2007	515	---
2008	518	---
2009	467	491.00
2010	502	500.50
2011	540	506.75
2012	557	516.50
2013	571	542.50
2014	586	563.50
2015	612	581.50

# ANOVA

Himadri Barman

Centre for Management Studies, Dibrugarh University

---

Consider the following situations

The mean cholesterol content of various food diets

The average telephone bill of households belonging to four different income groups

We are interested to compare the sample data sets drawn from various populations. In the first case, the various populations consist of the people going for various food diets. In the second case, we have four populations for four income groups. The technique of ANOVA (Analysis of Variance) becomes handy as it helps to compare the differences among the means of all the populations simultaneously. R A Fisher developed the theory concerning ANOVA. The basic principle underlying the technique is that the total variation in the dependent variable is broken into two parts – one which can be attributed to some specific causes and the other that may be attributed to chance. The one which is attributed to the specific causes is termed as the variation between samples and the one which is attributed to chance is termed as the variation within samples. Therefore, in ANOVA, the total variance may be decomposed into various components corresponding to the sources of the variation. For example, the sales of chairs could differ because of the various styles and the sizes of the stores selling them. Similarly, one could study the differences among the various types of drugs for curing a specific disease or differences in the yield of crops due to varieties of seed, fertilizers or soils.

In general, the ANOVA technique can investigate any number of factors which are supposed to influence the dependent variable of interest. It is also possible to investigate the differences in various categories within each of these factors. In ANOVA, the dependent variable in question is metric (interval or ratio scale), whereas the independent variables are categorical (nominal). If there is one independent variable (one factor) divided into various categories, we have one-way or one-factor ANOVA. In two-way or two-factor ANOVA, two factors each divided into the various categories are involved.

In ANOVA, it is assumed that each of the samples is drawn from a normal population and each of these populations has an equal variance. Another assumption that is made is that all the factors except the one being tested are controlled (kept constant). Basically, two estimates of the population variances are made. One estimate is based upon within the samples. The two estimates of variance can be compared for their equality using F-statistic.

### Problem (One-way ANOVA)

A consumer marketing group desired to examine whether supermarket chains operating in a city differed in their “out of stock” levels for advertised specials. The group identified the relevant response variable as the percentage of the items advertised not in stock. The following table provides the data collected from three supermarket chains in the city.

*Percentage of the items “Out of Stock” on Advertised Specials*

Chain 1	Chain 2	Chain 3
15	10	17
14	14	12
20	9	14
15	10	15
16	11	12

The marketing group would like to know whether there are significant differences among the three chains with regard to mean percentage out of stock on advertised specials. How would you analyse the situation?

### Solution

H0: The population means of percentage out of stocks position for all the three chains are equal

H1: The population means of percentage out of stocks for all the three chains are not equal

The basic calculations are given below:

- (a) Table showing Grand Mean and Treatment Means. Treatment means here denote the mean of each chain because the chains are the treatments.

	Chain 1	Chain 2	Chain 3
Treatment Means	16	10.8	14
Grand Mean	13.6		

- (b) Table showing Total Sum of Squares (TSS). Value of each cell is the square of (original value – grand mean)

Chain 1	Chain 2	Chain 3
1.96	12.96	11.56
0.16	0.16	2.56
40.96	21.16	0.16
1.96	12.96	1.96
5.76	6.76	2.56

TSS = 123.6

(c) Table showing Treatment Sum of Squares (TRSS) is given below. Each cell is the square of (Treatment Mean – Grand Mean)

5.76	7.84	0.16
------	------	------

**TRSS** = 68.8 (Sum of 5.76, 7.84 and 0.16 multiplied by 5 since five values are there for each chain)

The Error Sum of Squares (ESS) = TSS – TRSS = 123.6 – 68.8 = 54.8

### ANOVA Table

Source of Variation	SS	df	MS	F computed	F critical
<b>Treatment (Between Groups)</b>	68.8	2	34.40	7.53	3.89
<b>Error (Within Groups)</b>	54.8	12	4.57		
<b>Total</b>	123.6	14			

At 5% level of significance, H<sub>0</sub> is rejected.

### Explanation of the ANOVA Table

- The first column is labelled as “Sources of Variation”. This tells us the factors responsible for causing the variation. The treatment variation represents the variation caused by the chains with regard to the out-of-stock situation. The error variation represents the unexplained residual part. The word treatment is generic and as such may denote different methods, machines, different advertisement copy platforms, different strategies, different brands and the like.
- The second column is labelled as SS meaning sum of squares. The sum of squares is worked out corresponding to each source of variation.
- The third column is labelled as df denoting the degrees of freedom for each factor. There are 3 chains and so the chain factor has 2 degrees of freedom. The total number of observations is 15 and so the total df is 14. So, Error df is 12.
- The fourth column is labelled as MS meaning Mean Square. Each mean square is obtained by dividing the sum of squares by the corresponding df. The mean squares are not additive. So, total MS cannot be found.
- The fifth column is labelled as F computed, meaning the F statistic that we use for the interpretation of the result. This is the computed F. Statistically, F distribution is a ratio of two independent Chi-Square and so it has a pair of degrees of freedom, one for the numerator and the other for the denominator. In this case, computed F is the ratio of treatment mean square and error mean square. Both these mean squares follow a Chi-Square distribution. The *df* of F in this case is (2, 12).
- The sixth column is labelled as F critical meaning the critical F value for (2, 12) df at the given level of significance (5%).

# Index Numbers

Himadri Barman

Centre for Management Studies, Dibrugarh University

---

Index number in very simple terms is the ratio of two numbers. For example, Intelligence Quotient (IQ) is the ratio of the mental age to the chronological age. Index number is formally defined as a statistical measure designed to show changes in variables or a group of related variables with respect to time, geographical location or other characteristics.

Some important characteristics of index numbers are:

- It is the average of percentages
- It is a relative measurement
- It is a specialized average
- It measures changes not capable of direct measurement

The main purposes and uses of index numbers are:

- To study trends over a period of time
- To help in the formulation of policies
- To measure the purchasing power of money
- To deflate various values
- To act as economic barometers
- To measure seasonal variations and cyclical variations in a time series

The main limitations of index numbers are:

- They lack precision and mathematical accuracy as they are prepared on the basis of sample studies
- They present a general trend
- They are not very useful in studying the relative changes in economic variables
- They lack uniformity as they are prepared using a plethora of different methods

Some well-known indices are the CPI (Consumer Price Index), the WPI (Wholesale Price Index), the BSE Sensex (Sensitivity Index), etc.

Let us see the calculation of index using simple aggregative method for the following dataset:

Commodity	Prices in the year (₹/ unit)	
	2019 ( $P_0$ )	2020 ( $P_1$ )
A	15	20
B	25	30
C	60	70
D	50	60
	$\sum P_0 = 150$	$\sum P_1 = 180$

The Simple Price Index is given by  $P_{01} = \frac{\sum P_1}{\sum P_0} \times 100 = \frac{180}{150} \times 100 = 120$

Now, for the following data, we calculate the index by weighted aggregative method:

Commodity	Weight ( $W$ )	Base Year price ( $P_0$ )	Current Year price ( $P_1$ )
A	4	5	2
B	6	15	4
C	10	10	8

$$\text{The Weighted Price Index is given by } P_{01} = \frac{\sum P_1 W}{\sum P_0 W} \times 100 = \frac{8+24+80}{20+90+100} \times 100 = 53.33$$

Finally, we look at the formulas for Laspeyre's Index, Paasche's Index and Fisher's Index using the following data:

Commodity	2017		2020	
	Price ( $P_0$ )	Quantity ( $Q_0$ )	Price ( $P_1$ )	Quantity ( $Q_1$ )
A	6	50	10	56
B	2	100	2	120
C	4	60	6	60
D	10	30	12	24

$$\text{Laspeyres's Index : } L_{01} = \frac{\sum P_1 Q_0}{\sum P_0 Q_0} \times 100$$

$$\text{Paasche's Index : } P_{01} = \frac{\sum P_1 Q_1}{\sum P_0 Q_1} \times 100$$

$$\text{Fisher's Index : } F_{01} = \sqrt{\frac{\sum P_1 Q_0}{\sum P_0 Q_0} \times \frac{\sum P_1 Q_1}{\sum P_0 Q_1}} \times 100$$

We construct the table to calculate the indices:

Commodity	( $P_0$ )	( $Q_0$ )	( $P_1$ )	( $Q_1$ )	( $P_0 Q_0$ )	( $P_0 Q_1$ )	( $P_1 Q_0$ )	( $P_1 Q_1$ )
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
					$\sum P_0 Q_0$ = 1040	$\sum P_0 Q_1$ = 1056	$\sum P_1 Q_0$ = 1420	$\sum P_1 Q_1$ = 1448

$$\text{Laspeyres's Index : } L_{01} = \frac{1420}{1040} \times 100 = 136.54$$

$$\text{Paasche's Index : } P_{01} = \frac{1448}{1056} \times 100 = 137.12$$

$$\text{Fisher's Index : } F_{01} = \sqrt{\frac{1420}{1040} \times \frac{1448}{1056}} \times 100 = \sqrt{1.3654 \times 1.3712} \times 100 = 136.83$$

Fisher's Index satisfies the Time Reversal Test,

$$F_{01} \times F_{10} = 1 \text{ (the factor 100 is ignored), i.e., } \sqrt{\frac{\sum P_1 Q_0}{\sum P_0 Q_0} \times \frac{\sum P_1 Q_1}{\sum P_0 Q_1}} \times \sqrt{\frac{\sum P_0 Q_1}{\sum P_1 Q_1} \times \frac{\sum P_0 Q_0}{\sum P_1 Q_0}} = 1$$

Fisher's Index also satisfies the Factor Reversal Test,

$$\sqrt{\frac{\sum P_1 Q_0}{\sum P_0 Q_0} \times \frac{\sum P_1 Q_1}{\sum P_0 Q_1}} \times \sqrt{\frac{\sum Q_1 P_0}{\sum Q_0 P_0} \times \frac{\sum Q_1 P_1}{\sum Q_0 P_1}} = \frac{\sum P_1 Q_1}{\sum P_0 Q_0}$$

Fisher's Index is the geometric mean of Laspeyres's and Paasche's indices. It is free from bias. It is considered an ideal index because it combines the Laspeyres's Index and the Paasche's Index besides satisfying the Time Reversal and Factor Reversal tests.