Auditing Course Material

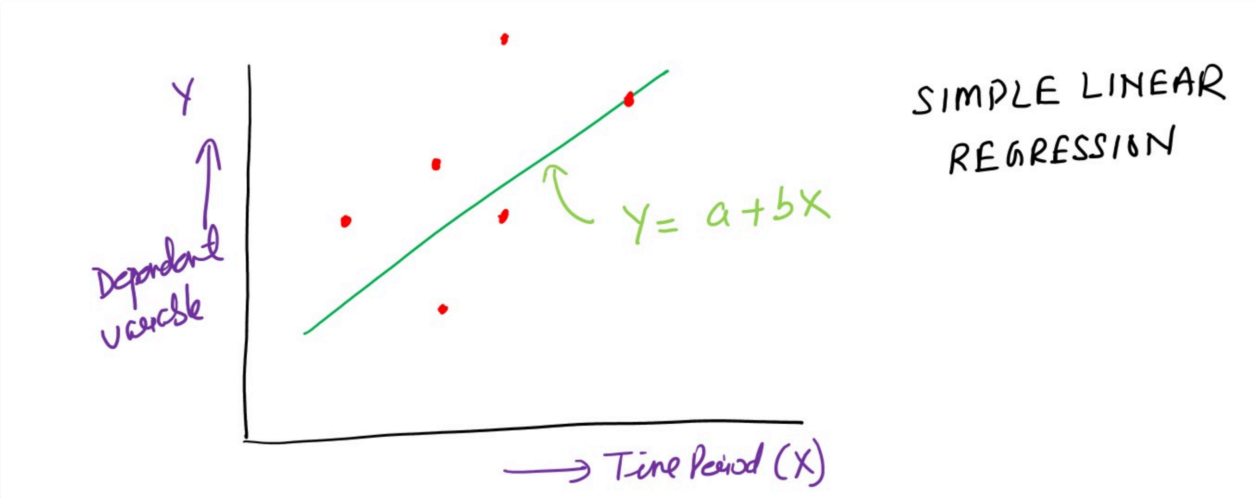
Part 49 of 61 (Chapters 4801-4900)

9. Simple Linear Regression

Let us now discuss the method of **trend projection**. This technique fits a trend line to a series of historical data points and then projects the slope of the line into the future for medium- to long-range forecasts. Several mathematical trend equations can be developed (for example, exponential and quadratic), but in this section, we will look at linear (straight-line) trends only.

To describe the linear association between quantitative variables, a statistical procedure called **regression** is used to construct a model. Regression is used to assess the contribution of one or more “explanatory” variables (called **independent variables**) to one “response” (or **dependent**) variable. It can also be used to predict the value of one variable based on the values of others. When there is only one independent variable and when the relationship can be expressed as a straight line, the procedure is called **simple linear regression**.

The objective in linear regression is to obtain an equation of a straight line that minimizes the sum of squared vertical deviations of data points from the line (i.e., the **least squares criterion**).



Regression equation of Y on X is given by:

Y= a + bX

where:
Y= Dependent variable
a = y-intercept
b = slope
X = Independent variable

a and b are calculated using the following equations:

b = $\frac{ \sum{xy-n\overline{xy}}}{ \sum{x^2-n\overline{x}^2}}$
a = $\overline{y}- b\overline{x}$

Illustration

With the help of the following data, project the trend of sales for the next 4 years:

Years	2002	2003	2004	2005	2006	2007
Sales (in lakhs)	100	110	115	120	135	140

Solution:

Year	**Time deviations from the middle of 2004 and 2005 (1 deviation = 6 months)**	**Sales (in lakh ₹)**	**Squares of time deviation**	**Product of time deviation and sales**

	X	Y	X^2	XY
2002	-5	100	25	-500
2003	-3	110	9	-330
2004	-1	115	1	-115
2005	+1	120	1	+120
2006	+3	135	9	+405
2007	+5	140	25	+700
N = 6	∑X = 0	∑Y = 720	$\sum X^2$ = 70	∑XY = 280

Regression equation of Y on X:

Y= a + bX

To find the values of a and b

b = $\frac{ \sum{xy-n\overline{xy}}}{ \sum{x^2-n\overline{x}^2}}$= $= \frac{ \sum{XY}}{ \sum{X^2}}= \frac{280}{70}$ = 4 $(\overline{x}=0$ since ∑X =0)

a = $\overline{y}- b\overline{x}$ = $\frac{ \sum{Y}}{n}= \frac{720}{6}$= 12 $(\overline{x}=0$ since ∑X = 0)

The Regression equation becomes:

Y = 120 + 4X

Sales forecast for the next years, i.e., 2008 to 2012:
Y_{2008} = 120 + 4(-7) = 120 + 28 = Rs 148 lakhs
Y_{2009} = 120 + 4(+9) = 120 + 36 = Rs 156 lakhs
Y_{2010} = 120 + 4(+11) = 120 + 44 = Rs 164 lakhs
Y_{2011} = 120 + 4(+13) = 120 + 52 = Rs 172 lakhs
Y_{2012} = 120 + 4(+15) = 120 +60 = Rs 180 lakhs

Standard error of Estimate

One indication of how accurate a prediction might be for a linear regression line is the amount of scatter of the data points around the line. If the data points tend to be relatively close to the line, predictions using the linear equation will tend to be more accurate than if the data points are widely scattered. The scatter can be summarized using the **standard error of estimate**. It can be computed by finding the vertical difference between each data point and the computed value of the regression equation for that value of x, squaring each difference, adding the squared differences, dividing by n - 2, and then finding the square root of that value.

S_e = $\sqrt{ \frac{ \sum{(y-y_c)^2}}{n-2}}$

Where,
S_e = Standard error of estimate
y = y value of each data point
n = Number of data points

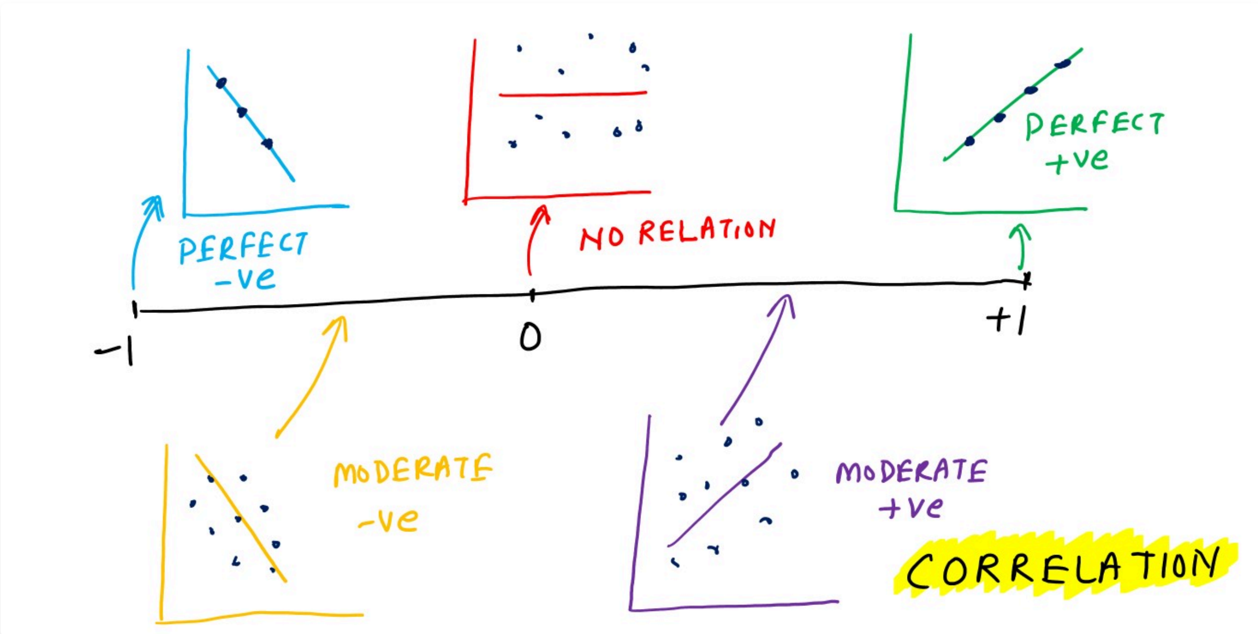
10. Correlation

The regression equation is one way of expressing the nature of the relationship between 2 variables. Regression lines are not "cause-and-effect" relationships. They merely describe the relationships among variables. The regression equation shows how one variable relates to the value and changes in another variable.

Another way to evaluate the relationship between two variables is to compute the coefficient of correlation. This measure expresses the degree or strength of the linear relationship (but note that correlation does not necessarily imply causality).

Correlation in a linear regression equation is a measure of the strength of the relationship between the independent and dependent variables.

Thus, the correlation describes the **strength of the relationship**. It is not concerned with 'cause' and 'effect'. A correlation is a relationship between two statistical variables measured from the same population.



Linear correlation is of 3 types:

1. **Positive Linear Correlation:** *High* values for one variable tend to correspond to *high* values for the second variable. Examples: Height vs. Weight for adults .
2. **Negative Linear Correlation:** *High* values for one variable tend to correspond to *low* values for the second variable. Examples: Age vs. Value of a car.
3. **No Linear Correlation:** No relationship between the variables or a non-linear relationship. Examples: Height vs. Number of Years of Education.

The correlation coefficient is represented by 'r' and the formula is:

$r = \frac{n \sum{xy}- \sum{x} \sum{y}}{ \sqrt{[n \sum{x^2-(\sum{x})^2}][n \sum{y^2} -(\sum{y})^2] } }$

The value of r varies between - 1.00 and + 1.00, with a value of +1.00 indicating a strong linear relationship between the variables. If r = 1.00, then an increase in the independent variable will result in a corresponding linear increase in the dependent variable. If r = –1.00, an increase in the dependent variable will result in a linear decrease in the dependent variable. A value of r near zero implies that there is little or no linear relationship between variables.

10. Correlation

Another measure of the strength of the relationship between the variables in a linear regression equation is the coefficient of determination. It is computed by squaring the value of r. It indicates the percentage of the variation in the dependent variable that is a result of the behavior of the independent variable. It is represented by r^2.

For example, if $r^2 = 0.897$, it means that 89.7% of the amount of variation in data can be attributed to the given factor (with the remaining 10.3% due to other unexplained factors). Thus, some amount of forecast error can be expected.

---

11. Multiple Regression

Multiple regression is an extension of linear regression. However, unlike in linear regression where the dependent variable is related to one independent variable, multiple regression develops a relationship between a dependent variable and multiple independent variables. The general formula for multiple regression is as follows:

$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + \ldots\ldots$

where:
Y = Dependent variable
a = y intercept
b_i = coefficients that represent the influence of the independent variables on the dependent variable.
X_i= Independent variables

For example, the dependent variable might be sales and the independent variables might be number of sales representatives, number of store locations, area population, and per capita income.

Multiple regression is a powerful tool for forecasting and should be used when multiple factors influence the variable that is being forecast. However, multiple regression does significantly increase data and computational requirements needed for forecasting. Fortunately, most standard statistical software programs have multiple regression capabilities.

---

12. Seasonal Adjustments

A seasonal pattern is a repetitive increase and decrease in demand. Many demand items exhibit seasonal behaviour.

Greeting card demand increases in conjunction with special days such as Valentine's Day and Mother's Day. The demand of gold increases on Dhanteras festival. Seasonal patterns can also occur on a monthly, weekly, or even daily basis. Some restaurants have higher demand in the evening than at lunch or on weekends as opposed to weekdays. Traffic—hence sales—at shopping malls picks up on Friday and Saturday.

A **seasonal index** indicates how a particular season (e.g., month or quarter) compares with an average season. When no trend is present, the index can be found by dividing the average value for a particular season by the average of all the data. Thus, an index of 1 means the season is average. For example, if the average sales in January were 120 and the average sales in all months were 200, the seasonal index for January would be less than 1 which means January is below average.

The resulting seasonal factors between 0 and 1.0 are, in effect, the portion of total annual demand assigned to each season. These seasonal factors are multiplied by the annual forecasted demand to yield adjusted forecasts for each season.

The next example illustrates how to compute seasonal indices from historical data and to use these in forecasting future values.

Illustration

Hero Honda sells motorcycles throughout the year. However, its peak season is during the fourth quarter of the year, from October to December. Hero Honda has experienced the demand for motorcycles for the past 3 years shown in the following table:

Year	**Demand (1000s) per Quarter**				
	1	**2**	**3**	**4**	**Total**
2008	12.6	8.6	6.3	17.5	**45.0**
2009	14.1	10.3	7.5	18.2	**50.1**
2010	15.3	10.6	8.1	19.6	**53.6**
Total	42.0	29.5	21.9	55.3	148.7

Solution:

Because we have 3 years of demand data, we can compute the seasonal factors by dividing total quarterly demand for the three years by total demand across all three years:

$S_1 = \frac{D_1}{\sum{D}} = \frac{42.0}{148.7} = 0.28$

$S_2 = \frac{D_2}{\sum{D}} = \frac{29.5}{148.7} = 0.20$

$S_3 = \frac{D_3}{\sum{D}} = \frac{21.9}{148.7} = 0.15$

$S_4 = \frac{D_4}{\sum{D}} = \frac{55.3}{148.7} = 0.37$

Next, we want to multiply the forecasted demand for the next year, 2011, by each of the seasonal factors to get the forecasted demand for each quarter. To accomplish this, we need a demand forecast for 2011. In this case, since the demand data in the table seem to exhibit a generally increasing trend, we compute a linear trend line for the three years of data in the table to get a rough forecast estimate:

$y = 40.97 + 4.30x$
$= 40.97 + 4.30\,(4)$
$= 58.17$

Thus, the forecast for 2011 is 58.17. Using this annual forecast of demand, we find that the seasonally adjusted forecasts for each quarter, SF_i, for 2011 are:

$SF_1 = (S_1)(F_5) = (0.28)(58.17) = 16.28$

$SF_2 = (S_2)(F_5) = (0.20)(58.17) = 11.63$

$SF_3 = (S_3)(F_5) = (0.15)(58.17) = 8.73$

$SF_4 = (S_4)(F_5) = (0.37)(58.17) = 21.53$

---

13. Accuracy of Forecast

The accuracy of a forecast refers to the closeness between the predicted value and the actual outcome. This discrepancy, termed **forecast error**, quantifies the difference between the forecasted and observed values.

Given the significant impact of demand forecasts across various business functions, precision in estimation becomes crucial. Evaluating the accuracy of forecasting techniques involves employing common measures to assess how closely the forecasts align with the actual outcomes.

Let us look at some of methods of forecasting error.

---

13. Accuracy of Forecast

The **Cumulative Error** formula is given by:

$= \sum_{t=1}^{n} (D_t-F_t)$

where: D_t is Actual Demand for period t,
F_t is Forecast Demand for period t,
n is Number of time period used.

Illustration

The actual and forecasted data of time periods 1 to 4 is given below.

CUMULATIVE ERROR

Time Period	Demand, D_t	Forecast, F_t	(D_t-F_t)
1	37	37	0
2	40	37	3
3	41	38	3
4	37	46	-9
			-3

Cumulative Error

$= \sum (D_t-F_t)$

$= -3$

---

13. Accuracy of Forecast

It is the mean of the deviations of the forecast demands from the actual demands. The Formula is given by:

MFE = $\frac{ \sum_{t=1}^n (D_t-F_t)}{n}$

It is also called **Average Error**.

Illustration

The actual and forescated data of time periods 1 to 4 is given below.

MEAN FORECAST ERROR

Time Period	Demand, D_t	Forecast, F_t	$(D_t - F_t)$
1	37	37	0
2	40	37	3
3	41	38	3
4	37	46	-9
			-3

Mean Forecast Error

$$= \frac{\Sigma (D_t - F_t)}{n} = \frac{-3}{4} = \boxed{0.75}$$

Also called AVERAGE ERROR

13. Accuracy of Forecast

It is the mean of absolute deviation of forecast demand from actual demand. It is also called **Mean Absolute Error**. The formula is given by:

MAD = $\frac{ \sum_{t=1}^n{\mid D_t-F_t\mid}}{n}$

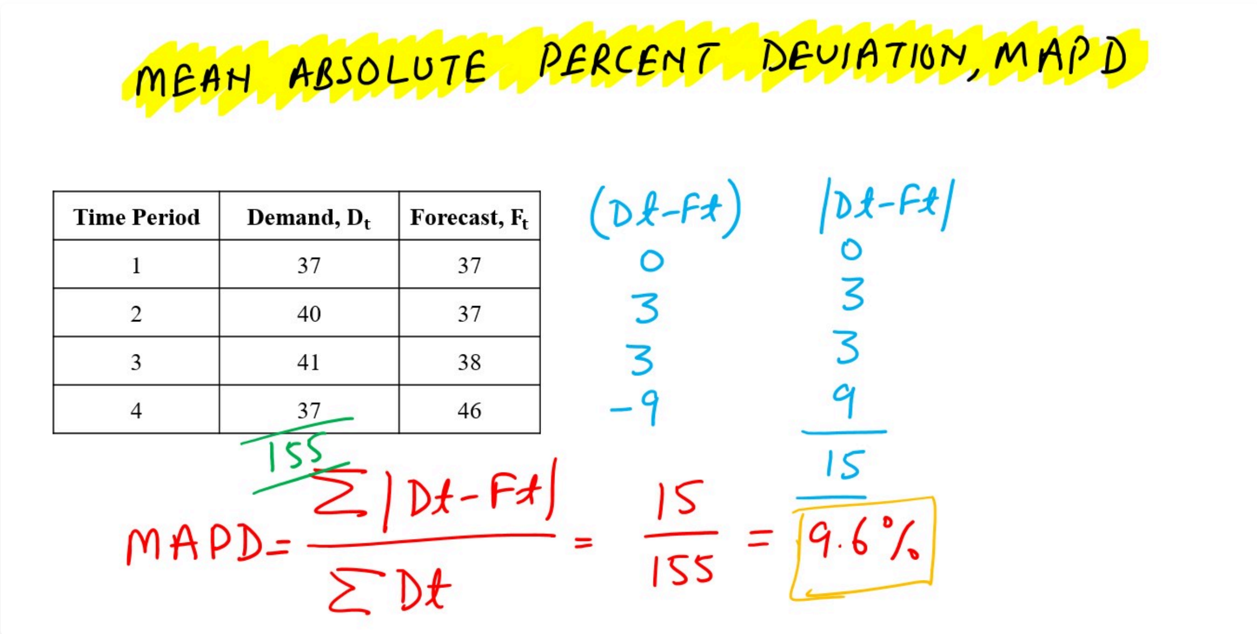
where: D_t is Actual Demand for period t,
F_t is Forecast Demand for period t,
n is Number of time period used.

The smaller the value of MAD, the more accurate the forecast.

Although if viewed alone, MAD is difficult to assess, as it gives absolute value. One benefit of MAD is to compare the accuracy of several different forecasting techniques.

Illustration

The actual and forescated data of time periods 1 to 4 is given below.



Time Period	Demand, D_t	Forecast, F_t
1	37	37
2	40	37
3	41	38
4	37	46

13. Accuracy of Forecast

It is the mean of the percent deviations of the forecast demands from the actual demands.

MAPE = $\frac{1}{n} \frac{ \sum_{t=1}^n{\mid D_t-F_t\mid}}{D_t}$ × 100

Illustration

The actual and forescated data of time periods 1 to 4 is given below.

MEAN ABSOLUTE PERCENT ERROR, MAPE

Time Period	Demand, D_t	Forecast, F_t	$\mid D_t-F_t \mid$
1	37	37	0
2	40	37	3
3	41	38	3
4	37	46	9
	155		15

$$MAPE = \frac{\frac{\sum \mid D_t-F_t \mid}{\sum D_t}}{n} = \frac{\frac{15}{155}}{4} = \boxed{2.4\%}$$

---

13. Accuracy of Forecast

Mean Square Error is the mean of the squares of the deviations of the forecast demands from the actual demands. The formula is given by:

MSE = $\frac{ \sum{_{t=1}^n} (D_t-F_t)^2}{n}$

Illustration

The actual and forescated data of time periods 1 to 4 is given below.

Time Period	Demand, D_t	Forecast, F_t	D_t-F_t	$(D_t-F_t)^2$
1	37	37	0	0
2	40	37	3	9
3	41	38	3	9
4	37	45	-8	64
				82

MEAN SQUARED ERROR, MSE $= \frac{\sum (D_t-F_t)^2}{n} = \frac{82}{4} = 20.5$

---

14. Tracking Signal

There are several ways to monitor forecast error over time to make sure that the forecast is performing correctly—that is, the forecast is in control. Forecasts can go “out of control” and start providing inaccurate forecasts for several reasons, including a change in trend, the unanticipated appearance of a cycle, or an irregular variation such as unseasonable weather, a promotional campaign, new competition, or a political event that distracts consumers.

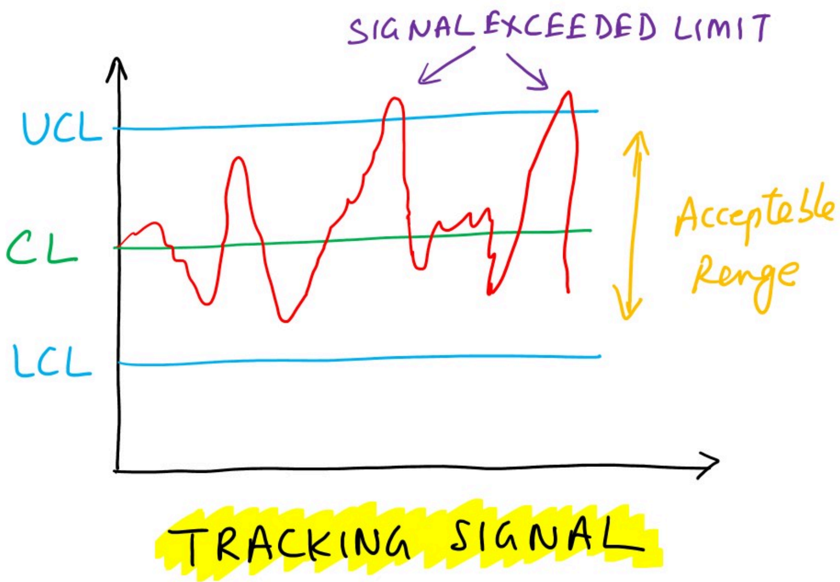
Thus, when there is a difference between forecast and actual values, one problem is to identify whether the difference is caused by random variation or is due to a bias in the forecast. **Forecast bias** is a persistent tendency for a forecast to be over or under the actual value of the data. We cannot do anything about random variation, but bias can be corrected.

One way to control for forecast bias is to use a tracking signal. A **tracking signal** is a tool used to monitor the quality of the forecast. It is computed as the ratio of the algebraic sum of the forecast errors divided by MAD.

Tracking signal = $\frac{Algebric \ sum \ of forecast \ errors}{MAD}= \frac{ \sum_{t=1}^n{\mid Actual - Forecast \mid}}{MAD}$

As the forecast errors are summed over time, they can indicate whether there is a bias in the forecast. To monitor forecast accuracy, the values of the tracking signal are compared against predetermined limits. These limits are usually based on judgment and experience and can range from ±3 to ±8. If errors fall outside these limits, the forecast should be reviewed.

Once tracking signals are calculated, they are compared with predetermined control limits. When a tracking signal exceeds an upper or lower limit, there is a problem with the forecasting method, and management may want to reevaluate the way it forecasts demand. The figure shows the graph of a tracking signal that is exceeding the range of acceptable variation. If the model being used is exponential smoothing, perhaps the smoothing constant needs to be readjusted.



---

15. Adaptive forecasting

Adaptive forecasting refers to computer monitoring of tracking signals and self-adjustment if a signal passes a preset limit. For example, when applied to exponential smoothing, the α and β coefficients are first selected on the basis of values that minimize error forecasts and then adjusted accordingly whenever the computer notes an errant tracking signal. This process is called **adaptive smoothing**.

---

16. Focus Forecasting

Some companies use forecasts based on a “best recent performance” basis. This approach, called focus forecasting and was developed by Bernard T. Smith, described in several of his books. It involves the use of several forecasting methods (e.g., moving average, weighted average, and exponential smoothing) all being applied to the last few months of historical data after any irregular variations have been removed. The method that has the highest accuracy is then used to make the forecast for the next month. This process is used for each product or service and is repeated monthly.

---

17. Diffusion Models

When new products or services are introduced, historical data are not generally available on which to base forecasts. Instead, predictions are based on rates of product adoption and usage spread from other established products, using mathematical diffusion models. These models take into account such factors as market potential, attention from mass media, and word of mouth. Although the details are beyond the scope of this text, it is important to point out that diffusion models are widely used in marketing and to assess the merits of investing in new technologies.

---

1. Introduction

National income refers to the total income earned by the citizens of a country in a given period of time, usually a year. It includes all income generated within a country's borders, whether earned by domestic or foreign entities, and is generally calculated by adding up all the incomes received by households, businesses, and government within the country, including wages, profits, interest, and rent. National income is a key measure of a country's economic performance and is often used to compare the economic well-being of different countries over time.

One of the earlier books on concept of national economy was written by Adam Smith, named *An Enquiry into the Nature and Cause of the Wealth of Nations*.

---

2. Types of Goods

In order to calculate National Income, it is important to distinguish between different types of goods that are produced within the economy. Goods can be classified into two categories - Final Goods and Intermediate Goods.

Final Goods

Final Goods are those goods that are meant for final use by consumers and will not pass through any more stages of production or transformations. They are the end products that are consumed directly by the consumer. For example, milk used by the consumer for self-consumption is a Final Good. Another example is a shirt that is ready to be sold to the consumer for final use.

There are two categories of Final Goods: Consumption Goods and Capital Goods.

1. **Consumption Goods** or Consumer Goods are goods that are purchased by their ultimate consumers for consumption. They are goods that are consumed when purchased by the end consumer. Examples of Consumer Goods include food, clothing, and services like recreation.
2. On the other hand, **Capital Goods** make the production of other commodities feasible, but they themselves do not get transformed in the production process. Capital Goods are also final goods, yet they are not final goods to be ultimately consumed. Examples of Capital Goods include tools, implements, and machines.

Consumer Durables are a unique category of Final Goods that are used for ultimate consumption, but have one characteristic in common with Capital Goods - they are also durable. That is, they are not extinguished by immediate or even short period consumption, and they have a relatively long life compared to articles such as food or even clothing. Examples of Consumer Durables include television sets, automobiles, and home computers.

Intermediate Goods

Intermediate Goods are goods that are used by other producers as material inputs. They are mostly used as raw material or inputs for the production of other commodities. Examples of Intermediate Goods include steel sheets used for making automobiles and copper used for making utensils. Intermediate Goods are not final goods since they are not meant for ultimate consumption, but rather for further production.

---

3. Stock and Flow Variables

When it comes to calculating National Income, it is important to understand the concept of Stock and Flow variables.

- **Stock variables** refer to any quantity that is measured at a particular point in time, such as the number of machines in a plant or the amount of money in a bank account on a specific date. These variables represent a snapshot of a particular moment in time and do not reflect any change that may occur over time.
- **Flow variables**, on the other hand, represent any quantity that is measured per unit of time. For example, income or expenditure over a period of one month or one year. Flow variables represent the change that occurs over time and reflect the rate at which a particular quantity is changing.

To understand the difference between Stock and Flow variables, let us consider an example. Suppose a company produces 100 units of a product in a given month. The production of 100 units is a flow variable as it represents the rate at which the company is producing the product over a specific period of time. However, the total number of units produced by the company over a year is a stock variable, as it represents the quantity produced at a particular point in time.

Another example of a stock variable is the number of cars owned by a family. The number of cars owned by a family is a quantity measured at a particular point in time and does not reflect any change that may occur over time. On the other hand, the amount spent on gasoline by the family each month is a flow variable as it represents the rate at which the family is spending money on gasoline over a specific period of time.

---

4. Economic Fluctuations

Economic fluctuations refer to the variations in the level of economic activity relative to the long-term growth trend of the economy. Business cycles or economic fluctuations involve various stages that vary in length and intensity. These cycles affect nearly all dimensions of economic activity, including production, employment, investment, and consumption, among others.

Depression

Depression is a severe and prolonged economic contraction characterized by a sharp reduction in an economy's total output accompanied by high unemployment lasting more than a year. The Great Depression of the 1930s is an example of a significant economic depression that affected many countries worldwide.

Recession

Recession is a decline in the economy's total output lasting at least two consecutive quarters, or six months. It is a less severe contraction than depression but still causes significant economic challenges, such as unemployment and reduced investment. An example of a recession is the Global Financial Crisis of 2008, which resulted in a sharp economic contraction and high unemployment rates worldwide.

Inflation

Inflation refers to an increase in the economy's average price level, leading to a decrease in the purchasing power of money. High inflation rates make it challenging to plan and invest in the future, as prices are continually changing. An example of high inflation is the hyperinflation that occurred in Zimbabwe in the late 2000s, which led to an almost complete collapse of the economy.

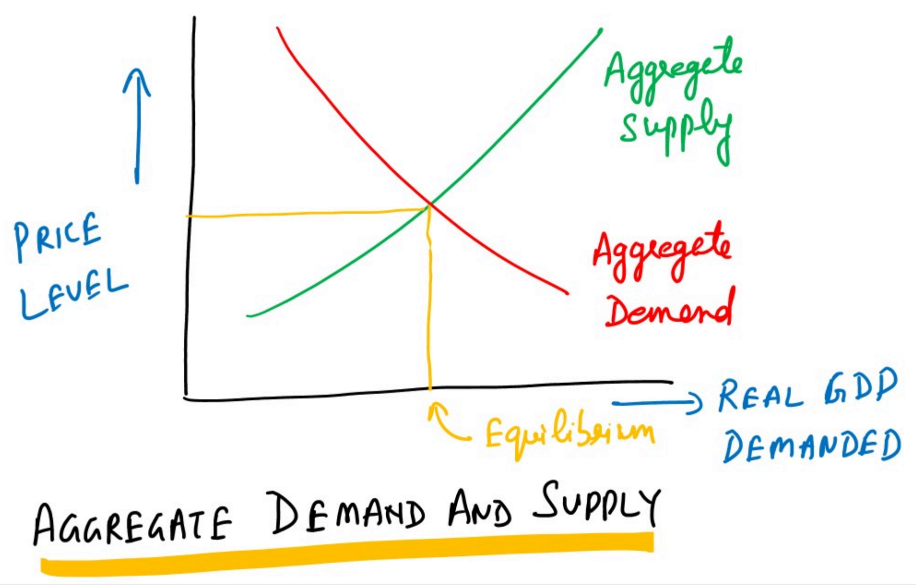
Economic Indicators

Economic indicators are variables that provide information on the state of the economy. They are crucial for policymakers, investors, and businesses in making decisions about investments, production, and consumption. There are three types of economic indicators: leading, coincident, and lagging indicators.

1. **Leading economic indicators** are variables that predict a recession or recovery. These indicators help policymakers, businesses, and investors prepare for changes in the economy before they occur. Examples of leading indicators include consumer confidence, stock market prices, business investment, and big-ticket purchases such as automobiles and homes.
2. **Coincident economic indicators** are variables that reflect peaks and troughs in economic activity as they occur. Examples of coincident indicators include employment, personal income, and industrial production. These indicators provide information on the current state of the economy and help policymakers and investors assess the severity of a recession or recovery.
3. **Lagging economic indicators** are variables that follow or trail changes in overall economic activity. Examples of lagging indicators include the interest rate and the average duration of unemployment. These indicators provide information on the consequences of changes in the economy, such as changes in the interest rate after a recession or changes in the duration of unemployment after a recovery.

5. Aggregate Demand and Supply

The **Aggregate Demand Curve** represents the inverse relationship between the price level and the quantity of real GDP demanded in the economy. As the price level increases, the quantity of real GDP demanded decreases, and vice versa, assuming other factors remain constant. This relationship is often represented in a graph where the vertical axis shows the price level, and the horizontal axis shows the quantity of real GDP demanded.



For instance, suppose the price level in an economy rises due to an increase in the cost of production or a decrease in the money supply. In that case, consumers may cut back on their spending, leading to a decrease in real GDP demanded, resulting in a movement up the aggregate demand curve.

The **Aggregate Supply Curve** represents the relationship between the economy's price level and the quantity of real GDP supplied in the economy. The curve shows how much firms will produce and sell at each price level, assuming other factors remain constant. The aggregate supply curve is upward-sloping because as prices increase, firms are willing to produce more output to increase their profits.

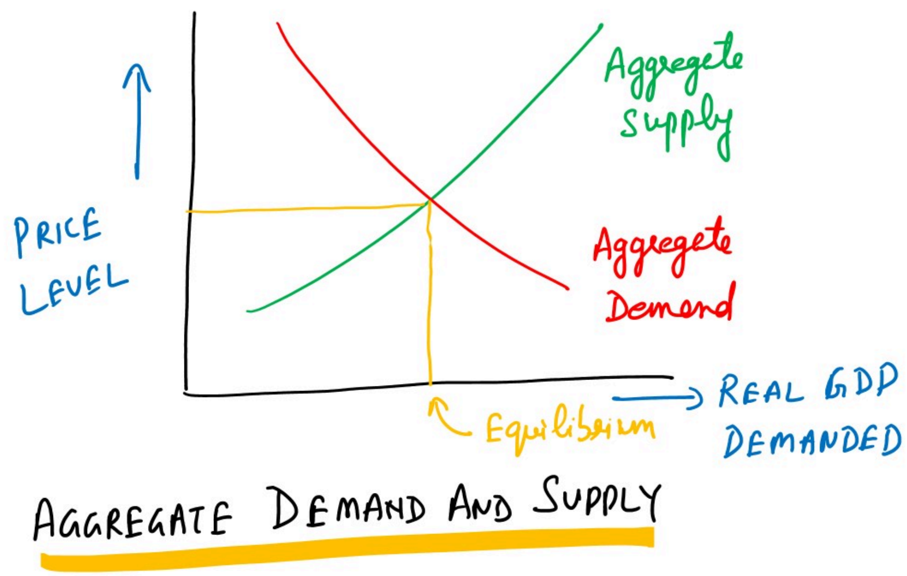
For instance, if a new technology is introduced in the economy, it may decrease the cost of production, and firms will be willing to supply more output at each price level. This results in a rightward shift in the aggregate supply curve, leading to a decrease in the price level and an increase in the quantity of real GDP supplied.

The **equilibrium** in an economy occurs at the point where the aggregate demand curve intersects with the aggregate supply curve. At this point, the quantity of real GDP demanded is equal to the quantity of real GDP supplied, and the economy is in a state of balance.

For example, suppose the economy is initially in equilibrium, but there is a decrease in the money supply. This will shift the aggregate demand curve to the left, resulting in a lower quantity of real GDP demanded. As a result, the economy moves to a new equilibrium point where the quantity of real GDP demanded is equal to the quantity of real GDP supplied, and the price level decreases.

The Great Depression of the 1930s can be explained by a shift to the left of the aggregate demand curve. The trigger for the shift of the aggregate demand curve in the Great Depression was the stock market crash of 1929. This event led to grim business expectations, which in turn led to a decrease in investment and consumer spending. As banks failed, the nation's money supply dropped, and world trade was severely restricted, aggregate demand declined further. A leftward shift of the aggregate demand curve implies a decrease in the quantity of aggregate output demanded at any given price level.

6. Demand side economics - Keynesian Economics



The central idea behind Keynesian economics, as explained in John Maynard Keynes' book *The General Theory of Employment, Interest, and Money*, is that aggregate demand is inherently unstable due to unpredictable "animal spirits" that guide investment decisions. Keynes argued that if businesses become pessimistic about the economy, they will invest less, reducing aggregate demand, output, and employment. Keynes believed that there were no natural market forces that could ensure that the economy would automatically recover, even if given sufficient time to adjust.

To combat this instability, Keynes proposed that the government should increase aggregate demand through **expansionary fiscal policy**. This could be achieved by increasing government spending or cutting taxes to stimulate consumption and investment. However, such policies could result in a budget deficit, as government outlays exceed government revenues.

Keynesian economics focuses on the **demand side of the economy**, arguing that changes in aggregate demand can promote full employment. The Keynesian approach suggests that government stimulus can shock the economy out of its depression. Once investment returns to normal levels, the government's intervention would no longer be necessary.

In terms of the Aggregate Demand and Aggregate Supply Curves, Keynes believed that shifts in the Aggregate Demand curve could be used to increase output and employment. By increasing government spending or cutting taxes, the Aggregate Demand curve would shift to the right, leading to an increase in real GDP and employment. This would compensate for the instability of private spending, particularly investment, and promote economic growth.

Stagflation

Stagflation is a term used to describe a situation where an economy experiences both stagnant growth or a contraction in output, and high inflation rates at the same time. This is a rare and challenging situation, as it goes against the standard economic theory that suggests a tradeoff between inflation and unemployment. Inflation is generally seen as a sign of strong demand, while unemployment signals weak demand. Stagflation, however, is a combination of high inflation and high unemployment, which is difficult for policymakers to tackle.

Stagflation occurs when the aggregate supply curve shifts to the left, as happened during the period of 1973-1980. The decrease in aggregate supply could be due to factors such as crop failures, disruptions in the supply of essential commodities, or production shortages. This leftward shift in the aggregate supply curve leads to a decrease in real GDP and an increase in prices. However, the aggregate demand curve can still be high due to factors such as government spending, increased investment, or rising wages, which may lead to higher inflation rates.

7. Supply side economics

To combat stagflation, attention has turned from aggregate demand to aggregate supply. By increasing aggregate supply, the price level can be lowered, while output and employment are increased. Supply-side economics advocates believe that this will expand real GDP and reduce the price level, resulting in economic growth.

Supply-side economics aims to shift the aggregate supply curve to the right through measures such as tax cuts or other incentives to increase production. This approach is based on the belief that by lowering tax rates, the government can increase after-tax wages, which would provide incentives to increase the supply of labour and other resources, leading to an increase in aggregate supply.

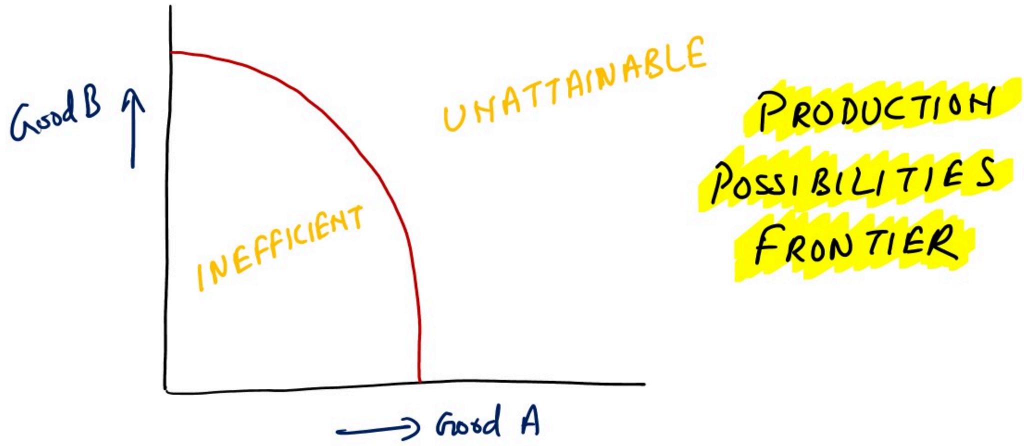
By increasing the supply of labour and other resources, aggregate supply can be increased, resulting in a shift of the aggregate supply curve to the right. As a result, the equilibrium point moves to the right, increasing both output and the price level. This increase in output and decrease in price level reduces the risk of stagflation.

However, the success of supply-side economics depends on a range of factors, such as how responsive workers and firms are to tax incentives, the level of government spending, and the overall state of the economy. Implementing tax cuts may not always have the desired effects and can result in increasing budget deficits.

8. Productivity and Growth

The production possibilities frontier (PPF) shows what the economy can produce if available resources are used efficiently.

Production Possibilities Frontier model



Production Possibilities Frontier (PPF) is a model used to represent the maximum amount of two goods that can be produced by an economy using all its available resources and technology efficiently during a given period. The PPF shows the trade-offs that must be made between the production of two goods due to scarcity of resources. The model makes simplifying assumptions that the resources are fixed in both quantity and quality, the available technology remains constant, and output is limited to just two broad classes of products: consumer goods and capital goods.

The PPF curve depicts the efficient combinations of two goods that can be produced by an economy using all its resources fully and efficiently. The curve is downward sloping, indicating that there is an inverse relationship between the production of the two goods. As we produce more of one good, we have to sacrifice the production of the other. Points inside the curve represent **inefficient combinations**, points on the curve represent efficient combinations, and points outside the curve are **unattainable combinations** given the available resources and technology.

The **law of increasing opportunity cost** states that as the economy produces more of one good, the opportunity cost of producing an additional unit of that good increases, and the shape of the PPF curve is bowed outwards. The increasing opportunity cost occurs because resources are not perfectly adaptable to the production of both types of goods. The more we produce of one good, the more we have to divert resources away from the production of the other good, and the resources that are left for the production of the other good are increasingly less suited to its production.

Let us consider an economy that produces two goods, consumer goods (such as pizzas and haircuts) and capital goods (such as pizza ovens and higher education). Suppose the economy produces 50 million units of consumer goods and no capital goods. The economy can use its resources to produce capital goods, but this will involve giving up some consumer goods. Suppose the economy diverts some of its resources to produce 10 million units of capital goods. The opportunity cost of producing 10 million units of capital goods is the production of two million units of consumer goods (50 million - 48 million).

It may be noted that the slope of the PPF shows the opportunity cost of an increment of capital. As the economy moves down the curve, the curve becomes steeper, reflecting the higher opportunity cost of capital goods in terms of forgone consumer goods.

8. Productivity and Growth

Economic growth is nothing but an outward shift of the production possibilities frontier. The growth can be caused by an increase in resources, such as a growth in the labour supply or in the capital stock.

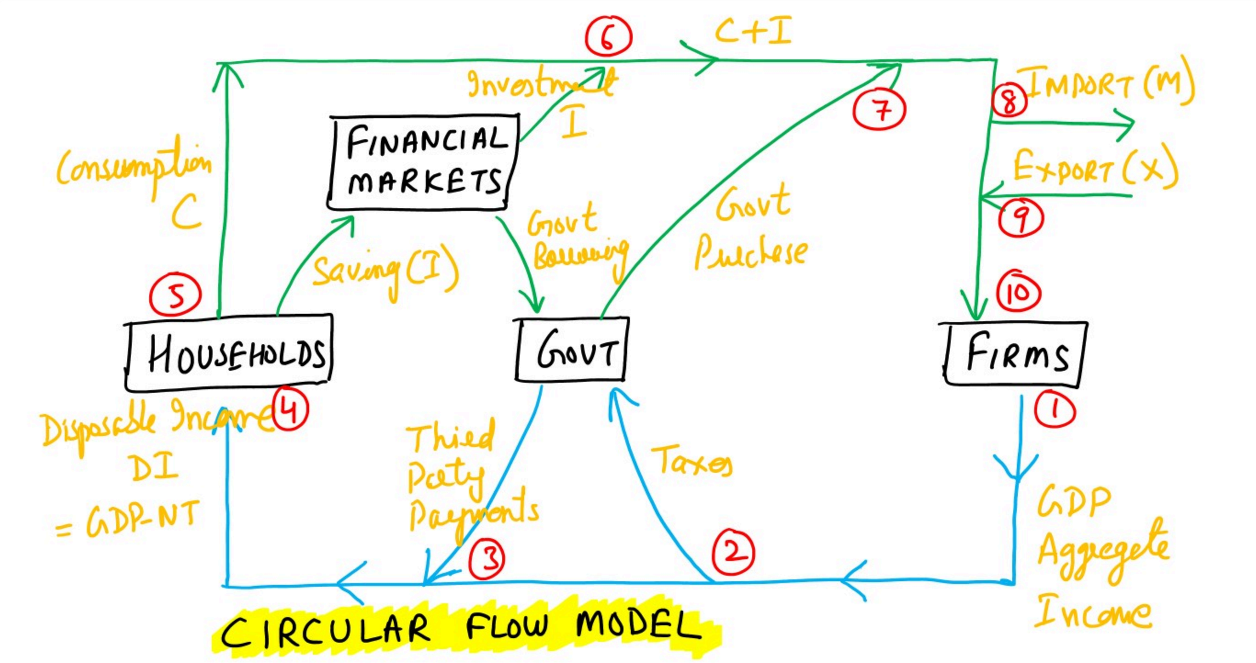
Labour supply can increase either because of population growth or because the existing population is willing to work more. The capital stock expands if the economy produces more capital this year.

Breakthroughs in technology also expand the frontier by making more efficient use of existing resources. Technological change often improves the quality of capital, but it can enhance the productivity of any resource.

Finally, any improvement in the rules of the game that nurtures production and exchange promotes growth and expands the PPF. For example, the economy can grow as a result of a patent laws that encourages more inventions or legal reforms that reduce transaction costs.

Thus, the economy grows because of a greater availability of resources, an improvement in the quality of resources, technological change that makes better use of resources, or improvements in the rules of the game that enhance production.

9. Circular flow model

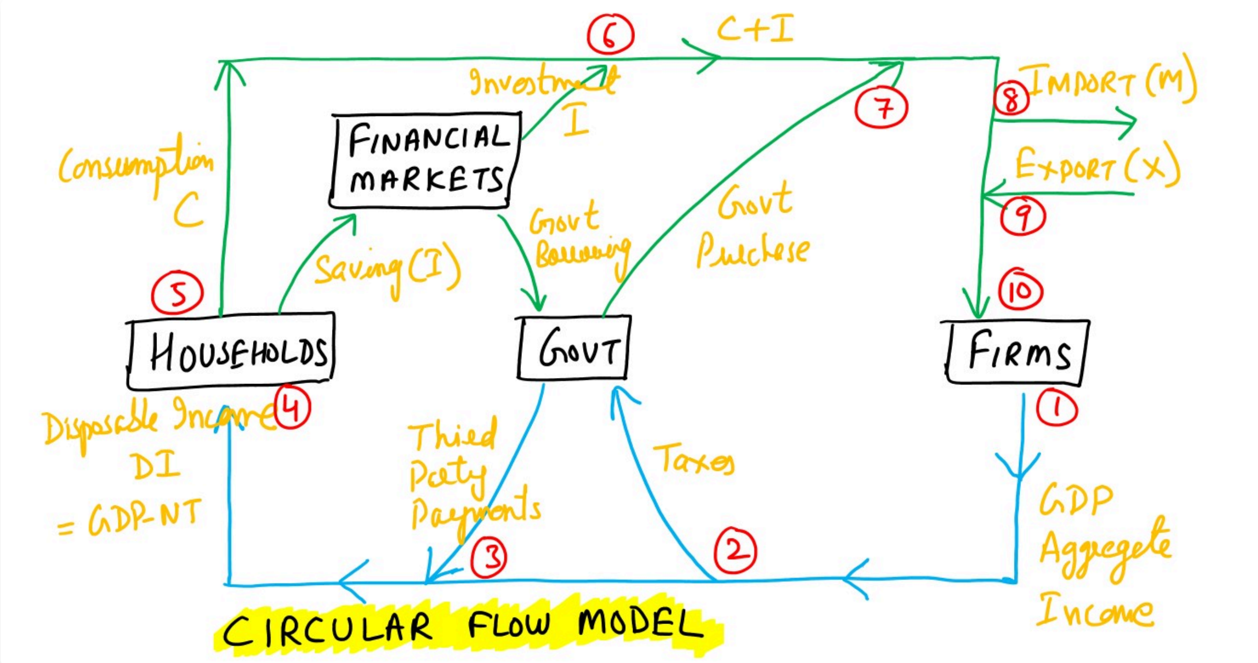


The Circular Flow of Income and Expenditure is a basic economic model that illustrates how money flows through an economy. It shows the flow of money and goods and services between households, firms, and the government.

The model is based on two key concepts: the circular flow of income and the circular flow of expenditure. The circular flow of income represents the flow of money through the economy from households to firms and back to households. The circular flow of expenditure represents the flow of goods and services through the economy from firms to households and back to firms.

The circular flow is a continuous process, but the logic of the model is clearest if we begin at juncture (1), where the firms make production decisions. The production must occur before output can be sold and income earned.

9. Circular flow model



Households supply their labour, capital, natural resources, and entrepreneurial ability to make products that sell to pay wages, interest, rent, and profit. Production of aggregate output, or GDP, gives rise to an equal amount of aggregate income. Thus, at juncture (1), aggregate output equals aggregate income.

But not all that income is available to spend. At juncture (2), governments collect taxes. Some of these taxes return as transfer payments to the income stream at juncture (3). By subtracting taxes and adding transfers, we transform aggregate income into **disposable income**, DI, which flows to households at juncture (4). Disposable income is take-home pay, which households can spend or save. **Transfer payments** are payments made without any productive service rendered in return, like old-age pensions, scholarships, and taxes.

The bottom half of this circular flow is the income half because it focuses on the income arising from production. **Aggregate income** is the total income from producing GDP, and disposable income is the income remaining after taxes are subtracted and transfers added. To simplify the discussion, we define **net taxes**, NT, as taxes minus transfer payments. So disposable income equals GDP minus net taxes.

Put another way, we can say that aggregate income equals disposable income plus net taxes:

GDP = Aggregate income = DI + NT

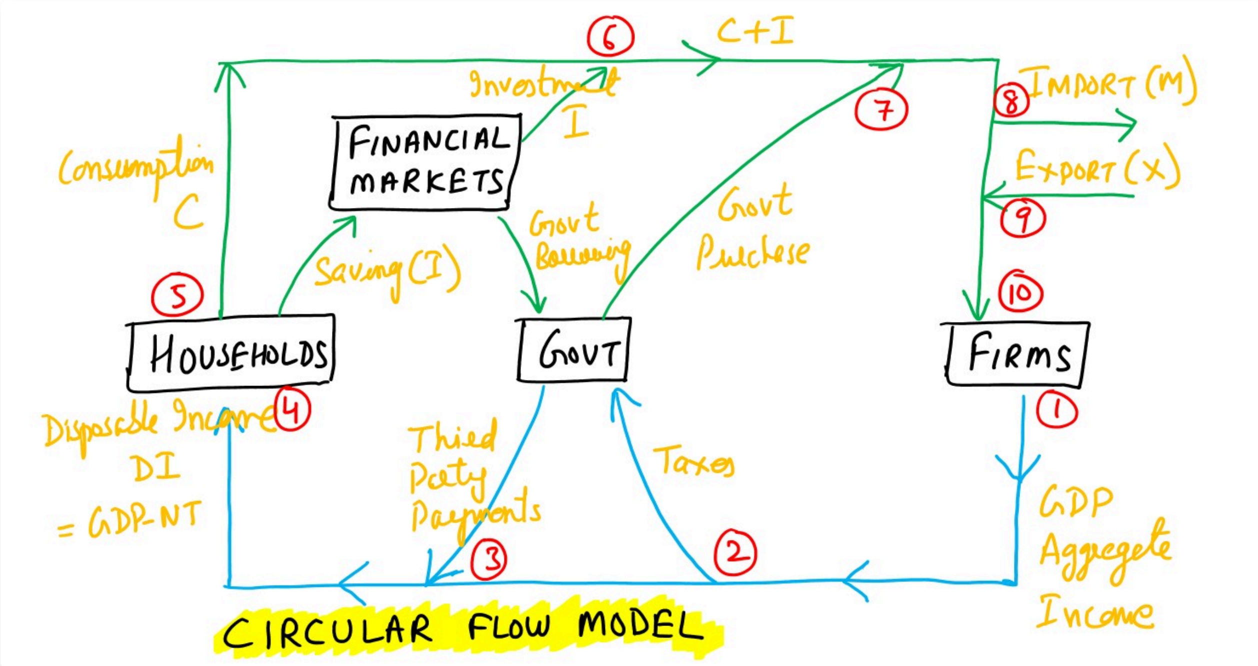
At juncture (4), firms have produced output and have paid resource suppliers; governments have collected taxes and made transfer payments. With the resulting disposable income in hand, households must now decide how much to spend and how much to save. Because firms have already produced the output and have paid resource suppliers, firms wait to see how much consumers want to spend. Any unsold production gets added to firm inventories.

9. Circular flow model

Disposable income splits at juncture (5). Part goes for consumption, C, and the rest is saved, S. Thus,

DI = C + S

Spending on consumption remains in the circular flow and is the biggest aggregate expenditure. Household saving flows to financial markets, which consist of banks and other financial institutions that link savers to borrowers.



For simplicity, the figure shows households as the only savers, though governments, firms, and the rest of the world could save as well. The primary borrowers are firms and governments, but households borrow too, particularly for new homes, and the rest of the world also borrows. In reality, financial markets should be connected to all four economic decision makers, but we have simplified the flows.

In our simplified model, firms pay resource suppliers an amount equal to the entire value of output. With nothing left for investment, firms must borrow to finance purchases of physical capital plus any increases in their inventories. Households also borrow to purchase new homes. Therefore, investment, I, consists of spending on new capital by firms, including inventory changes, plus spending on residential construction. Investment enters the circular flow at juncture (6), so aggregate spending at that point totals C + I.

Governments must also borrow whenever they incur deficits, that is, whenever their total outlays—transfer payments plus purchases of goods and services—exceed their revenues. Government purchases of goods and services, represented by G, enter the spending stream in the upper half of the circular flow at juncture (7). Remember that G excludes transfer payments, which already entered the stream as income at juncture (3).

Some spending by households, firms, and governments goes for imports. Because spending on imports flows to foreign producers, spending on imports, M, leaks from the circular flow at juncture (8). But the rest of the world buys Indian products, so foreign spending on Indian exports, X, enters the spending flow at juncture (9). Net exports, the impact of the rest of the world on aggregate expenditure, equal exports minus imports, X – M, which can be positive, negative, or zero.

The upper half of the circular flow, the expenditure half, tracks components of aggregate expenditure: consumption (C), investment (I), government purchases (G), and net exports (X –M). Aggregate expenditure flows into firms at juncture (10).

Aggregate expenditure equals the market value of aggregate output, or GDP. In other words,

C + I + G + (X – M) = Aggregate expenditure = GDP

9. Circular flow model

Thus we see that in the lower half of the circular flow, aggregate income equals disposable income plus net taxes. In the upper half, aggregate expenditure equals the total spending on country's output. The aggregate income arising from production equals the aggregate expenditure on that production. This is the first accounting identity. Thus, aggregate income (disposable income plus net taxes) equals aggregate expenditure (spending by each sector), or:

DI + NT = C + I + G + (X – M)

Because disposable income equals consumption plus saving, we can substitute C + S for DI in the above equation to yield:

C + S + NT = C + I + G + (X – M)

After subtracting C from both sides and adding M to both sides, the equation reduces to:

S + NT + M = I + G + X

Note that at various points around the circular flow, some of the flow leaks from the main stream. Saving, S, net taxes, NT, and imports, M, are **leakages** from the circular flow. **Injections** into the main stream also occur at various points around the circular flow. Investment, I, government purchases, G, and exports, X, are injections of spending into the circular flow.

As you can see from the preceding equation, leakages from the circular flow equal injections into that flow. This leakages-injections equation demonstrates a second accounting identity based on double-entry bookkeeping.

---

10. Macroeconomic Indicators

Gross Domestic Product (GDP) is a measure of the economic output of a country. It is calculated as the total value of all final goods and services produced within the borders of a country in a given year. GDP includes income earned by foreigners within the country and excludes income earned by nationals outside the country. GDP can be measured in two ways, Nominal GDP and Real GDP.

Nominal GDP is the value of all goods and services produced in a country during a given year, measured at current market prices. It does not adjust for inflation and can overestimate the economic output if prices are rising. For example, if a country produces 100 units of goods at a price of Rs 1 each in one year, its nominal GDP will be Rs 100.

However, if in the next year, the price of the same goods rises to Rs 2 each and the country produces the same 100 units, the nominal GDP will be Rs 200. This increase in nominal GDP does not necessarily indicate an increase in economic output.

Real GDP, on the other hand, adjusts for inflation and provides a more accurate measure of economic output. It measures the value of goods and services produced in a country using a constant set of prices. To calculate real GDP, a base year is chosen when the general price level is normal. Prices in the base year are set to 100 (or 1), and the prices of other years are compared to the base year prices to determine the inflation rate. The real GDP for a given year is then calculated by multiplying the nominal GDP by a ratio of the price level in the base year to the price level in the given year.

For example, suppose the base year is 2011-12 and the nominal GDP for 2022 is Rs 10 trillion with a GDP deflator of 150. Then the real GDP for 2022 will be Rs 6.67 trillion (10 trillion × 100/150).

GDP deflator is an index of the price level of all goods and services included in GDP. It is calculated as the ratio of nominal GDP to real GDP multiplied by 100. The GDP deflator provides a measure of inflation or deflation in the economy. If the GDP deflator is higher than 100, it indicates that prices have increased since the base year, while a GDP deflator less than 100 indicates that prices have decreased.

GDP Deflator = $\frac{Nominal \ GDP}{Real \ GDP}$× 100

In India, since January 2015, the base year has been changed to 2011-12 for calculation of GDP.

---

10. Macroeconomic Indicators

GDP or Gross Domestic Product measures the total value of goods and services produced within the borders of a country in a specific period. On the other hand, GNP or **Gross National Product** measures the total value of goods and services produced by a country's residents and businesses, regardless of their location, in a specific period. This means that GNP takes into account the income earned by Indian citizens and businesses outside the country's borders, as well as the income earned by foreign citizens and businesses within India's borders.

GNP = GDP + Net factor income from abroad

Net factor income from abroad = Factor income earned by the domestic factors of production employed in the rest of the world – Factor income earned by the factors of production of the rest of the world employed in the domestic economy.

---

10. Macroeconomic Indicators

Personal Income is the total income received by individuals from all sources, including wages, salaries, rent, interest, dividends, transfer payments, and capital gains, before the payment of direct taxes in one year. However, Personal Income is not equal to National Income because the former includes transfer payments, which are not included in National Income. **Transfer payments** are payments made without any productive service rendered in return, like old-age pensions, scholarships, and taxes.

On the other hand, **National Income** is the total income earned by all factors of production in a country during a year, and it is equal to the sum of wages, salaries, rent, interest, and profit earned in that year. National Income is a measure of the total production in a country and is used to calculate Gross Domestic Product (GDP).

To derive Personal Income from National Income, we need to adjust for income earned but not received and income received but not earned. Income earned but not received includes employer's share of social security taxes, taxes on production (like sales and property taxes) net of subsidies, corporate income taxes, and undistributed corporate profits. Income received but not earned includes transfer payments from the government, receipts from private pension plans, and interest paid by government and consumers.

Personal Income (PI) = National Income – Income earned but not received (Undistributed profits + Net interest payments made by households + Corporate tax) + Income received but not earned (Transfers to households from the government and firms.)

Disposable Income is the amount of income that households have available for spending or saving after the payment of direct taxes. It is calculated by subtracting personal tax payments (like income tax) and transfers to the government from Personal Income.

Personal Disposable Income (PDI) ≡ PI - Personal tax payments + Transfers to the government

National Disposable Income is the maximum amount of goods and services that the domestic economy has at its disposal. It is calculated by adding Net National Product at market prices, net factor income from abroad (i.e., income earned by domestic factors of production in foreign countries minus income earned by foreign factors of production in the domestic country), and other current transfers from the rest of the world (like gifts and aids).

National Disposable Income = Net National Product at market prices + Net factor income from abroad + Other current transfers from the rest of the world.

---

10. Macroeconomic Indicators

Private Income is income obtained by private individuals from any source, productive or otherwise, and the retained income of corporations.

Private Income = Factor income from net domestic product accruing to the private sector + National debt interest + Net factor income from abroad + Current transfers from government + Other net transfers from the rest of the world.

---

10. Macroeconomic Indicators

Income generated (or earned) by factors of production within the country from its own resources is called domestic income or domestic product.

Domestic income includes:

1. Wages and salaries,
2. Rents, including imputed house rents,
3. Interest,
4. Dividends,
5. Undistributed corporate profits, including surpluses of public undertakings,
6. Mixed incomes consisting of profits of unincorporated firms, self-employed persons, partnerships, etc.,
7. Direct taxes.

Since domestic income does not include income earned from abroad, it can also be shown as:
Domestic Income = National Income - Net income earned from abroad.

---

11. Factor Cost and Market Prices

Two methods are used to calculate national income - factor cost and market prices. Both factor cost and market prices play a crucial role in determining the national income of a country.

Factor Cost

Factor cost is the sum of all payments made to factors of production such as labour, capital, and land, for their contribution to the production process. It is the actual cost incurred to produce goods and services. Factor cost is calculated by adding up the cost of wages, interest, rent, and profit earned by the factors of production in the production process.

For example, if a farmer produces 100 kilograms of wheat at a cost of Rs. 4 per kilogram, the total factor cost of producing wheat will be Rs. 400 (100 × Rs. 4). This cost includes all the payments made to the factors of production involved in the production process, such as the farmer's labour, capital investment, and rent paid for the land.

Market Prices

Market prices refer to the prices at which goods and services are sold in the market. It includes indirect taxes and excludes subsidies. Indirect taxes are taxes that are paid by producers while subsidies are payments made to producers by the government to reduce the cost of production. Market prices help to calculate the total value of goods and services produced, including the taxes paid.

For example, if a company produces 100 chairs and sells them for Rs. 500 per chair, the market price of chairs produced will be Rs. 50,000 (100 × Rs. 500). If the company pays Rs. 2,000 in taxes and receives a subsidy of Rs. 1,000 from the government, then the net market price of chairs will be Rs. 49,000 (Rs. 50,000 - Rs. 2,000 + Rs. 1,000).

Difference between factor cost and market prices

The difference between factor cost and market prices lies in the inclusion or exclusion of indirect taxes and subsidies. While factor cost includes all the payments made to factors of production, market prices account for taxes paid and subsidies received by the producers.

To calculate the national income of a country, both factor cost and market prices are used. The national income at factor cost represents the total value of goods and services produced in a country, including the cost of production. The national income at market prices represents the total value of goods and services produced in a country, including taxes paid and subsidies received.

11. Factor Cost and Market Prices

1. We need to consider **Depreciation**, when we calculate "Net (N)" from "Gross (G)" or vice versa.

Gross = Net + Depreciation
Net = Gross – Depreciation

Example:

GDP_{MP} = NDP_{MP} + Depreciation
NDP_{MP} = GDP_{MP} + Depreciation

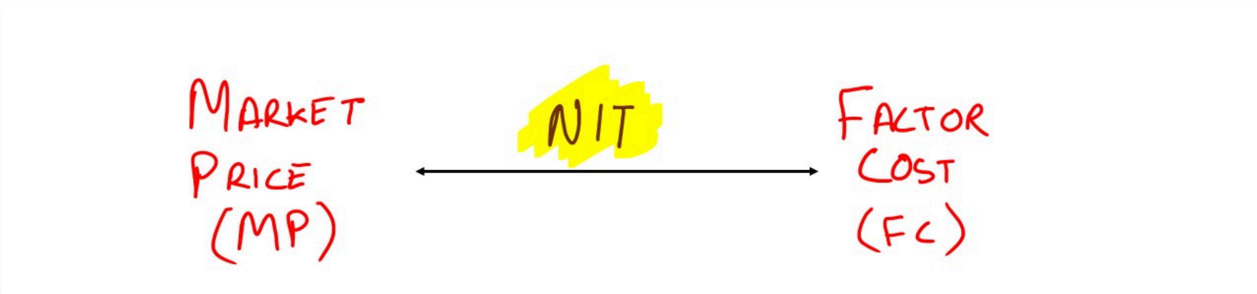
2. We need to consider **Net Factor Income from Abroad, NFIA**, when we calculate "National (N)" from "Domestic (D)" or vice versa.

National = Domestic + NFIA
Domestic = National - NFIA

Example:

GDP_{MP} = GNP_{MP} - NFIA
GNP_{MP} = GDP_{MP} + NFIA

3. We need to consider **Net Indirect Taxes, NIT**, when we calculate "Market Price (MP)" from "Factor Cost (FC)" or vice versa.



Market Price = Factor Cost + NIT
Factor Cost = Market Price - NIT

Example:

GDP_{MP} = GDP_{FC} + NIT
GDP_{FC} = GDP_{MP} – NIT

Factor Cost (FC) refers to the amount paid to factors of production for their contribution in the production process.
Factor Cost, FC = MP – NIT, where NIT is (Indirect Taxes – Subsidies)

Market Price (MP) refers to the price at which product is actually sold in the market.
Market Price, MP = FC + NIT, where NIT is (Indirect Taxes – Subsidies)

Net Indirect Tax (NIT) is the difference between Indirect Taxes (IT) and subsidies.

NIT = IT – Subsidies

---

12. Aggregates of National Income

Gross Domestic Product at Market Price (GDP_{MP})

GDP_{MP} is defined as the gross market value of the final goods and services produced within the domestic territory of a country during an accounting year by all production units.

a. 'Gross' in GDP_{MP} signifies that depreciation is included, i.e., no provision has been made for depreciation.
b. 'Domestic' in GDP_{MP} signifies that it includes all the final goods and services produced by all the production units located within the economic territory (irrespective of the fact whether produced by residents or non-residents).
c. 'Market Price' in GDP_{MP} signifies that indirect taxes are included and subsidies are excluded, i.e., it shows that Net Indirect Taxes (NIT) have been included.
d. 'Product' in GDP_{MP} signifies that only final goods and services have to be included and intermediate goods should not be included to avoid the double counting.

Gross Domestic Product at Factor Cost (GDP_{FC})

GDP_{FC} is defined as the gross factor value of the final goods and services produced within the domestic territory of a country during an accounting year by all production units excluding Net Indirect Tax.
$GDP_{FC} = GDP_{MP}$ – Net Indirect Taxes

Net Domestic Product at Market Price (NDP_{MP})

NDP_{MP} is defined as the net market value of all the final goods and services produced within the domestic territory of a country by its normal residents and non-residents during an accounting year.
$NDP_{MP} = GDP_{MP}$ – Depreciation

Net Domestic Product at Factor Cost (NDP_{FC})

NDP_{FC} refers to a total factor income earned by the factor of production within the domestic territory of a country during an accounting year.
$NDP_{FC} = GDP_{MP}$ – Depreciation – Net Indirect Taxes

NDP_{FC} is also known as Domestic Income or Domestic factor income.

Gross National Product at Market Price (GNP_{MP})

GNP_{MP} refers to market value of all the final goods and services produced by the normal residents of a country during an accounting year.
$GNP_{MP} = GDP_{MP}$ + Net factor income from abroad

It must be noted that GNP_{MP} can be less than GDP_{MP} when NFIA is negative. However, GNP_{MP} will be more than GDP_{MP} when NFIA is positive.

NFIA is Net Factor Income Abroad = Factor Income earned from abroad – Factor Income paid to abroad.

Gross National Product at Factor Cost (GNP_{FC}) or Gross National Income

GNP_{FC} refers to gross factor value of all the final goods and services produced by the normal residents of a country during an accounting year.
$GNP_{FC} = GNP_{MP}$ – Net Indirect Taxes

Net National Product at Market Price (NNP_{MP})

NNP_{MP} refers to net market value of all the final goods and services produced by the normal residents of a country during an accounting year.
$NNP_{MP} = GNP_{MP}$ – Depreciation

Net National Product at Factor Cost (NNP_{FC})

NNP_{FC} refers to net money value of all the final goods and services produced by the normal residents of a country during an accounting year.

$NNP_{FC} = GNP_{MP}$ – Depreciation – Net Indirect Taxes

It must be noted that NNP_{FC} is also known as **National Income**.

Net National Product at factor cost is also called as "National Income". Net National Product at factor cost is equal to sum total of value added at factor cost or net domestic product at factor cost and net factor income from abroad.

NNP at Factor Cost = NDP at Factor Cost + Net Factor Income from Abroad (NFIA)
where:
NDP at Factor Cost = NDP at Market Price - Indirect Taxes + Subsidies
where:
NDP at Market Price = GDP at Market Price - Depreciation

Thus we conclude:

National Income = NNP at Factor Cost = NNP at Market Price - Net Indirect Tax

We can also write

National Income = NNP at Factor Cost = NNP at Market Price – Indirect taxes + Subsidies

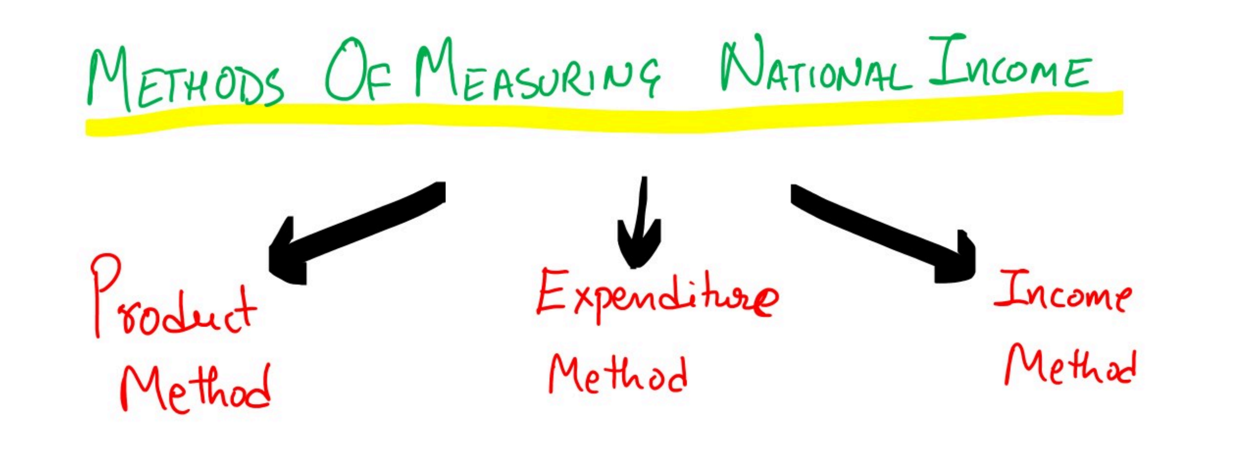
(Net Indirect Tax = Indirect taxes - Subsidies)

---

13. Measuring National Income

National income accounts are based on the fact that one person's spending is another person's income. Gross domestic product (GDP) includes only **final goods** and services, which are goods and services sold to the final or end user. The National income is almost equal to GDP. It requires some statistical adjustments (corporate income taxes, undistributed corporate profits etc).

In order to measure national income, there are 3 main methods used: the product method, expenditure method, and income method.



---

13. Measuring National Income

The product method is also known as the value added method. It involves calculating the aggregate annual value of goods and services produced in an economy during a given time period, typically a year. The value added is the difference at each stage of production between the selling price of a product and the cost of intermediate goods purchased from other firms.

The national economy is considered as an aggregate of productive units of different sectors, such as agriculture, mining, manufacturing, trade and commerce, services, etc. This method is used when the entire national economy is considered as an aggregate of producing units.

The value added of a firm is distributed among its 4 factors of production, namely, labour, capital, entrepreneurship, and land. Wages, interest, profits, and rents paid out by the firm must add up to the value added of the firm. The sum total of the gross value added of all the firms in the economy in a year is Gross Domestic Product (GDP). Value added is a flow variable.

GDP ≡ Sum total of gross value added of all the firms in the economy.

If we include the value of depreciation in value added, then the measure of value added that we obtain is called **Gross Value Added (GVA)**. If we deduct the value of depreciation from gross value added, we obtain **Net Value Added (NVA)**.

---

13. Measuring National Income

The expenditure method is used when the national economy is viewed as a collection of spending units, including individuals, households, firms, and government. This method measures national income at the final expenditure stages.

National income is calculated by adding the purchase of consumer goods and services by residents and households, government expenditure on goods and services, business enterprises' expenditure on capital goods and stocks, and net exports (exports minus imports).

National Income (GDP) = C + G + I + NX

where:
C is Consumption, which is purchase of consumer goods and services by residents and households. It is also called personal consumption expenditures.
G is Government expenditure on goods and services
I is Business enterprises' expenditure on capital goods and stocks. It is also called *gross private domestic investment*. A net increase to inventories also counts as investment.
NX is Net exports (exports - imports)

---

13. Measuring National Income

The income method is used when the national economy is considered as a combination of individuals and households owning different kinds of factors of production, which they use themselves or sell factor services to make their livelihood. The income method is used to measure national income as a combination of factor-owners and users.

The factors of production are land, labour, capital, and enterprise. Each factor of production, when put into use, provides income, which is a factor income. GDP by income method is calculated by summing up the factor incomes, including:

- compensation of employees (income earned from labour),
- rent (income earned from owning land),
- interest (income earned from owning capital),
- proprietors' income (income earned from organizing production), and
- corporate profits (income earned from organizing production).

GDP by income method = Sum of the Factor Incomes

There is another term used in factor payments. It is **operating surplus**. It is defined as the sum of rent and royalty, interest and profits (known as **RIP**).

The Central Statistics Office (CSO) of the Government of India reports GDP at factor cost and at market prices. In January 2015, the CSO replaced GDP at factor cost with the Gross Value Added (GVA) at basic prices (measure of supply side), and the GDP at market prices (measure of output). In May 2019, the Government merged Central Statistics Office (CSO) and National Sample Survey Office (NSSO) and created a new body called National Statistical Office (NSO).

1. Inflation

In economics, **inflation** refers to the upward movement in the overall price level of goods and services within an economy over a period of time. As the general price level rises, the purchasing power of a unit of currency decreases, leading to a decrease in the real value of money. Another way to conceptualize inflation is as an excess of money chasing a limited supply of goods and services.

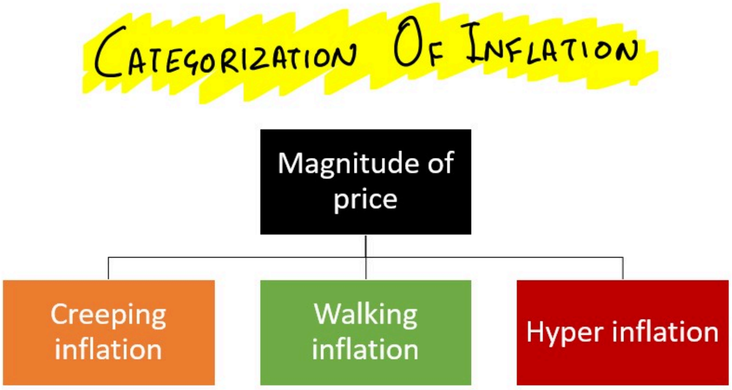
Conversely, **deflation** represents the opposite scenario, where there is a decline in the general price levels. Deflation occurs when the inflation rate falls below 0% or when there is negative inflation. Deflation has the effect of increasing the real value of money, allowing individuals to purchase more goods with the same amount of money over time. Deflation can be triggered by factors such as a reduction in the money supply or credit availability, as well as decreased spending, whether in the form of reduced government, personal, or investment spending. An unfortunate side effect of deflation is its potential to increase unemployment due to decreased demand in the economy.

Disinflation refers to a decrease in the rate of inflation. In such a situation, the rate of inflation decreases but is still positive. Disinflation is different from deflation, where there is an actual decrease in the general price level of goods and services.

2. Types of Inflation

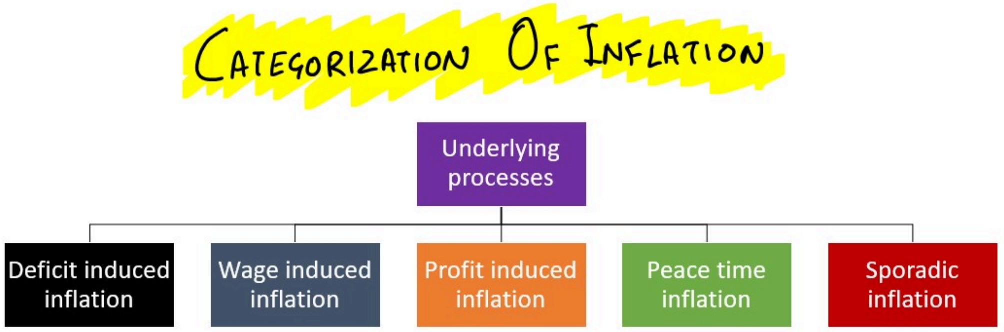
The inflation can be categorized into many types, based on its characteristics.

Categorization of Inflation Based on the Magnitude of Price Increases:



- **Creeping inflation** represents the mildest form of inflation, and some economists do not view it as detrimental to the economy. In fact, certain economists consider it a valuable tool for promoting economic development. They argue that it helps prevent the national economy from succumbing to stagnation.
- **Walking inflation** signifies a more noticeable increase in prices compared to creeping inflation. It serves as a warning sign of the potential emergence of running and jumping inflation, in which prices escalate at a faster rate.
- **Hyperinflation**, also known as galloping inflation, sees prices soaring continuously, with no apparent limit to how high they may climb. Hyperinflation is a clear indication of extreme dysfunction within a country's monetary system.

Classification of Inflation Based on the Underlying Processes:



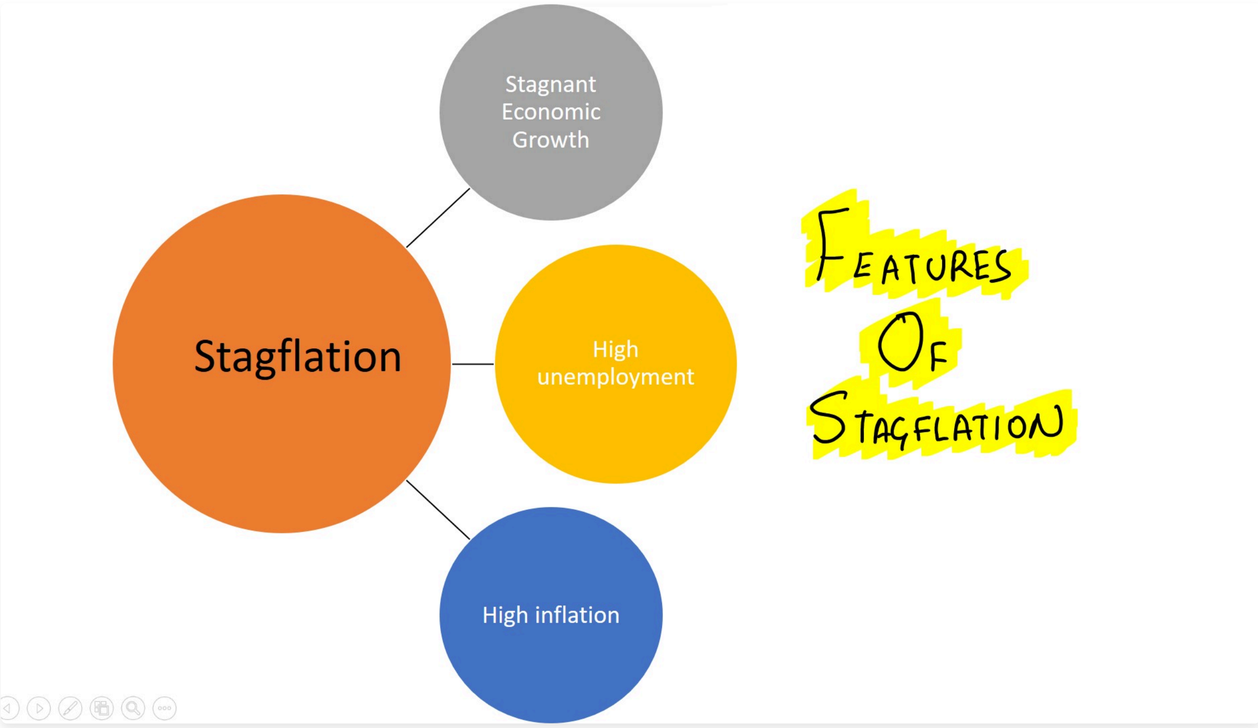
- **Deficit-induced inflation** arises from practices like deficit financing or excessive government spending exceeding its revenue receipts.
- **Wage-induced inflation** occurs due to a rise in money wages, contributing to overall price increases.
- **Profit-induced inflation** stems from increased profits among manufacturers, resulting in higher prices for goods and services.
- **Peace-time inflation** denotes price hikes during periods of peace, often due to increased government expenditures.
- **Sporadic inflation** is localized and specific to certain goods, often caused by abnormal shortages. For instance, food prices may surge due to poor monsoon conditions.

Inflation can also be categorized as open or suppressed. **Open inflation** refers to a situation where prices consistently and unabatedly rise. **Suppressed inflation** occurs when price increases are held in check through government policies, such as effective price controls and rationing of essential goods.

3. Stagflation

The term "stagflation" was coined by British politician Iain Macleod in 1965 to describe a situation where an economy experiences both inflation and stagnation simultaneously.

Stagflation is characterized by a combination of stagnant economic growth (stagnation), high unemployment rates, and high inflation. This term represents an unusual occurrence because inflation typically accompanies robust economic growth, while periods of economic slowdown tend to alleviate inflationary pressures.



Key features of stagflation include:

1. **Stagnant Economic Growth**: The economy experiences sluggish or stagnant growth rates, leading to reduced production, lower consumer demand, and limited job creation.

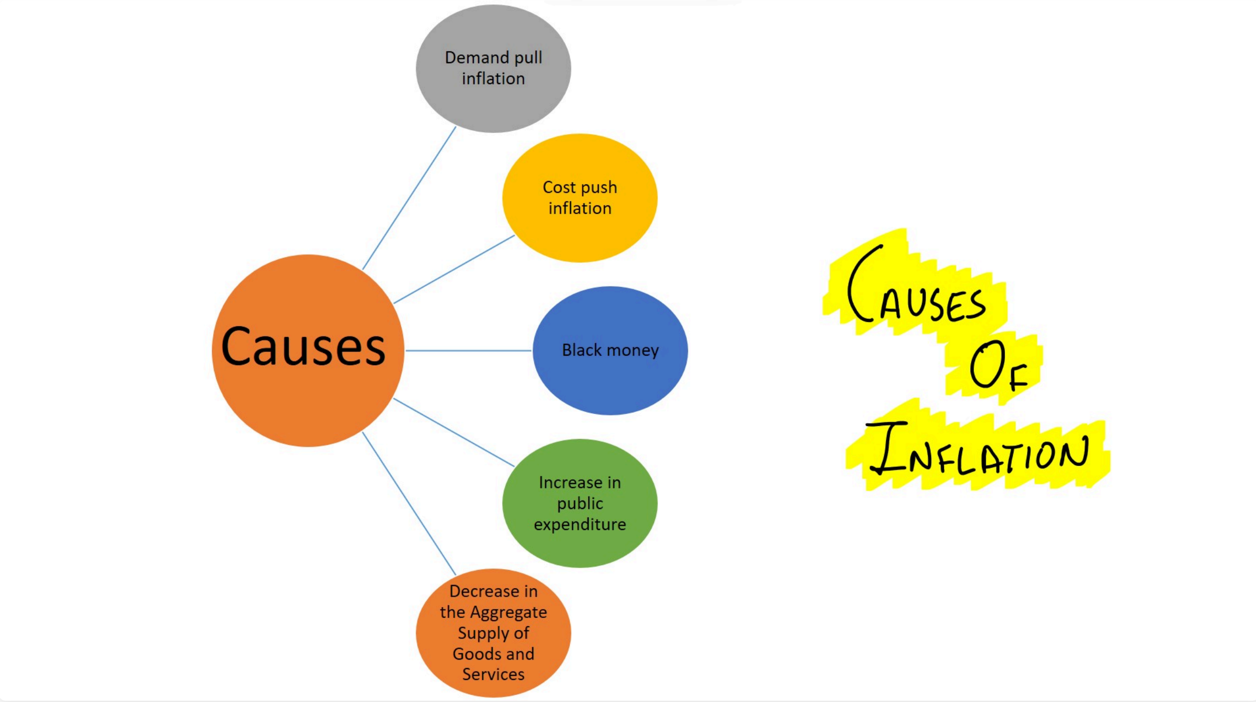
2. **High Unemployment**: Unemployment rates tend to be elevated as businesses restrain hiring or even lay off workers due to reduced economic activity.

3. **High Inflation**: Despite the slowdown in economic growth, prices of goods and services continue to rise at a rapid pace, leading to increased costs for consumers and businesses.

Stagflation poses challenges for policymakers since traditional measures to stimulate economic growth, such as reducing interest rates or increasing government spending, might exacerbate inflationary pressures. Managing stagflation requires a delicate balance between policies aimed at boosting economic activity while curbing inflation.

4. Causes of Inflation

The causes underlying inflation can be generally divided according to the source through which inflation originates.



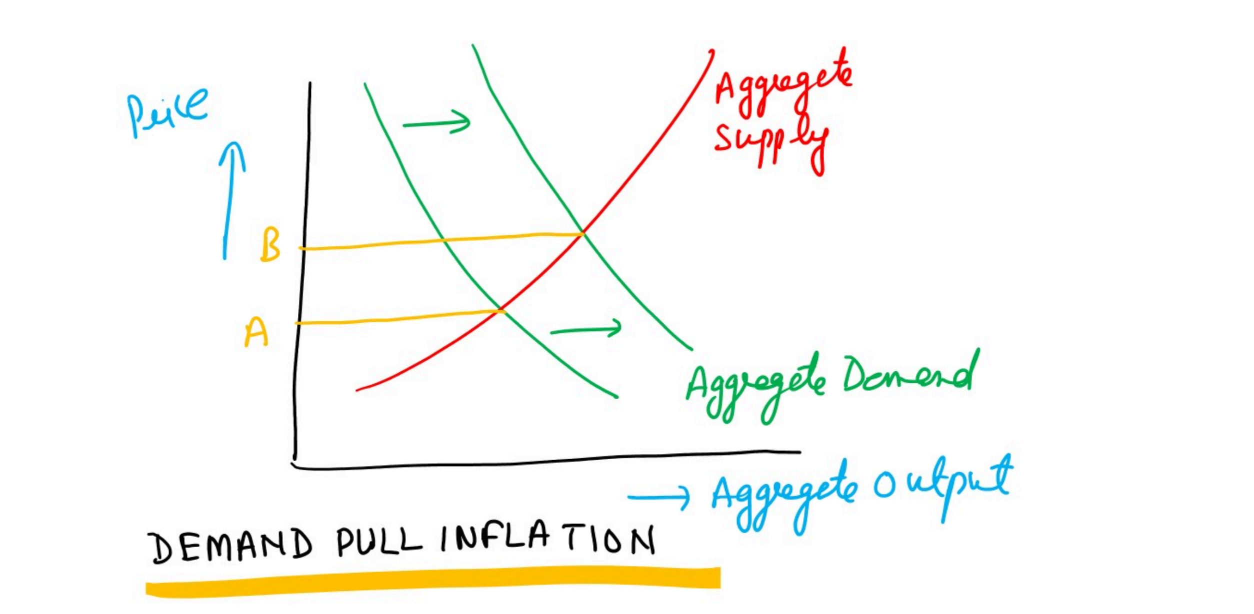
Depending upon initial process, we can identify 2 types of inflation:

a. Demand-pull inflation or, demand-side inflation; and
b. Cost-push inflation or, cost-side Inflation or Supply-side

Other than demand pull and cost push, there are other causes for inflation as well:

1. Black money
2. Increase in Public Expenditure
3. Decrease in the Aggregate Supply of Goods and Services

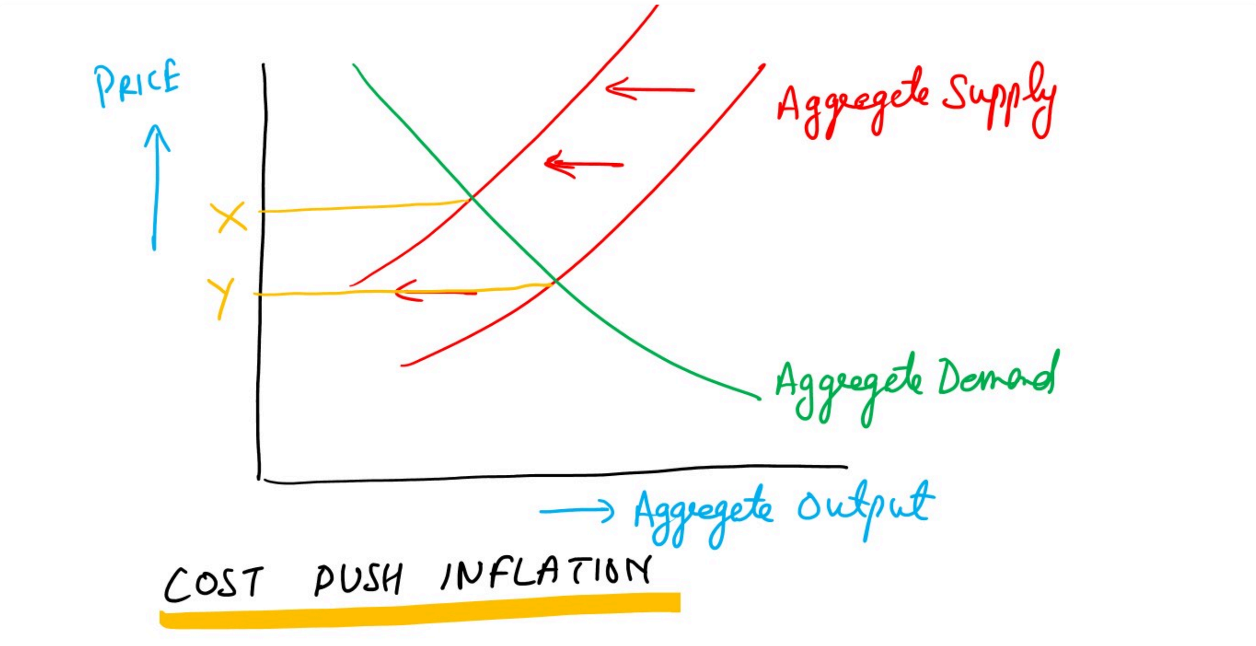
4. Causes of Inflation



Demand-pull inflation occurs when aggregate demand for goods and services in an economy exceeds the available supply, leading to an increase in the price level. The figure shows that an increase in aggregate demand raises the economy's price level from A to B. In such cases, a shift to the right of the aggregate demand curve pulls up the price level. Some of the factors contributing to demand-pull inflation are:

- **Increase in Money Supply:** When the central bank injects more money into the economy, it leads to an increase in demand for goods and services, which in turn pushes up prices.
- **Increase in Exports:** When the exports of a country increase, foreign exchange inflows increase, leading to an increase in the purchasing power of the people. This, in turn, results in an increase in demand, which leads to inflation.
- **Increase in Disposable Income:** When people have more disposable income, they tend to spend more, leading to an increase in demand and pushing up prices.
- **Deficit Financing:** When the government spends more money than it earns, it leads to an increase in the money supply, which can result in demand-pull inflation.
- **Foreign Exchange Reserves:** When a country's foreign exchange reserves increase, it leads to an increase in the money supply, leading to an increase in demand and prices.

4. Causes of Inflation



Cost-push inflation occurs when the cost of production of goods and services increases, leading to an increase in prices. It arises from reductions in aggregate supply, as shown in the figure, where a leftward shift of the aggregate supply curve raises the price level from Y to X.

Some of the factors contributing to cost-push inflation are:

- **Rise in Administered Prices:** When the government increases the prices of goods and services it controls, it leads to an increase in the cost of production, which can result in cost-push inflation.
- **Erratic Agriculture Growth:** If the agricultural sector, which is the main source of raw materials, does not grow at a consistent rate, it can result in an increase in the cost of production, leading to inflation.
- **Agricultural Price Policy:** When the government sets minimum support prices for agricultural products, it can lead to an increase in the cost of production, which can result in cost-push inflation.
- **Inadequate Industrial Growth:** When industrial growth is inadequate, it can lead to a shortage of goods and services, resulting in an increase in prices.

---

4. Causes of Inflation

In addition to demand-pull and cost-push inflation, there are other causes of inflation, such as:

- **Black Money:** When fake currency enters the economy, it leads to an increase in the money supply, which can result in inflation.
- **Increase in Public Expenditure:** When the government spends more money, it leads to an increase in the money supply, which can result in inflation.
- **Decrease in the Aggregate Supply of Goods and Services:** If there is a decrease in the supply of goods and services, it can result in inflation as demand exceeds supply.

---

5. Controlling Inflation

Here are some of the steps that governments may take when inflation is high:

i. **Monetary Policy Tightening:** Central banks may increase interest rates to make borrowing more expensive, thereby reducing consumer and business spending. This helps in curbing inflation by reducing overall demand in the economy.
ii. **Fiscal Policy Measures:** Governments can implement fiscal policies, such as reducing government spending or increasing taxes, to reduce the overall demand in the economy. This reduces pressure on prices by controlling excessive government expenditure.
iii. **Exchange Rate Policies:** Managing the exchange rate can be a tool to control inflation. Governments may allow their currency to appreciate (increase in value) to reduce the prices of imported goods, which can help lower inflation.
iv. **Supply-Side Reforms:** Governments can enact policies aimed at improving the supply side of the economy. This includes measures to increase productivity, reduce regulations, and promote competition. A more efficient supply side can help reduce cost-push inflation.
v. **Price Controls:** In extreme cases, governments may implement price controls on essential goods and services to prevent excessive price increases. However, this measure is often seen as a short-term solution and can lead to other economic problems if not managed carefully.
vi. **Public Awareness and Education:** Governments can also work on increasing public awareness and education about inflation and its effects. This can help in managing inflation expectations, which can play a significant role in controlling inflation.
vii. **Strengthening Financial Regulation:** Ensuring a well-regulated financial sector can help control inflation by preventing excessive lending and speculation that can fuel demand and asset bubbles.
viii. **Social Safety Nets:** To alleviate the impact of high inflation on vulnerable populations, governments may expand social safety net programs, such as food subsidies or direct cash transfers, to provide relief to those most affected by rising prices.

6. Monetary Policy

The Monetary Policy refers to the policy of the central bank with regard to the use of monetary instruments under its control to achieve the specified goals.

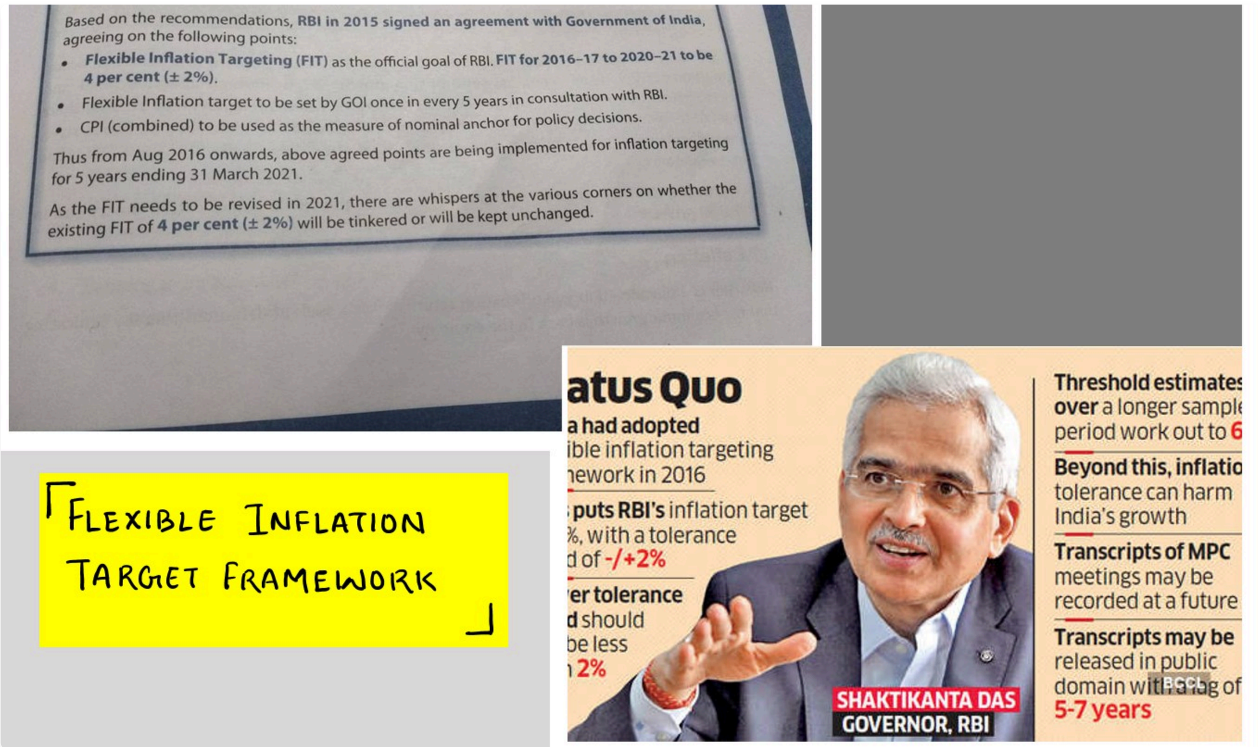
The Committee to Review the Working of the Monetary System (under chairmanship of **Dr. Sukhamoy Chakravarty**) recommended in 1985 a new monetary policy framework.

Before 2016, the objective of monetary policy was to maintain price stability and ensuring adequate flow of credit to the productive sectors of the economy. However, pursuant to the amendment to RBI Act 1934, in May 2016, the primary objective of monetary policy is to **maintain price stability while keeping in mind the objective of growth**.

Further, in 2016 amendment, it was added that, it is essential to have a modern **Monetary Policy Framework** to meet the challenges of an increasing complex economy.

---

6. Monetary Policy



Now there is a Flexible Inflation Targeting Framework (FITF) in India (after the 2016 amendment to the Reserve Bank of India (RBI) Act, 1934. The amended RBI Act provides for the inflation target to be set by the Government of India, in consultation with the RBI, once every 5 years.

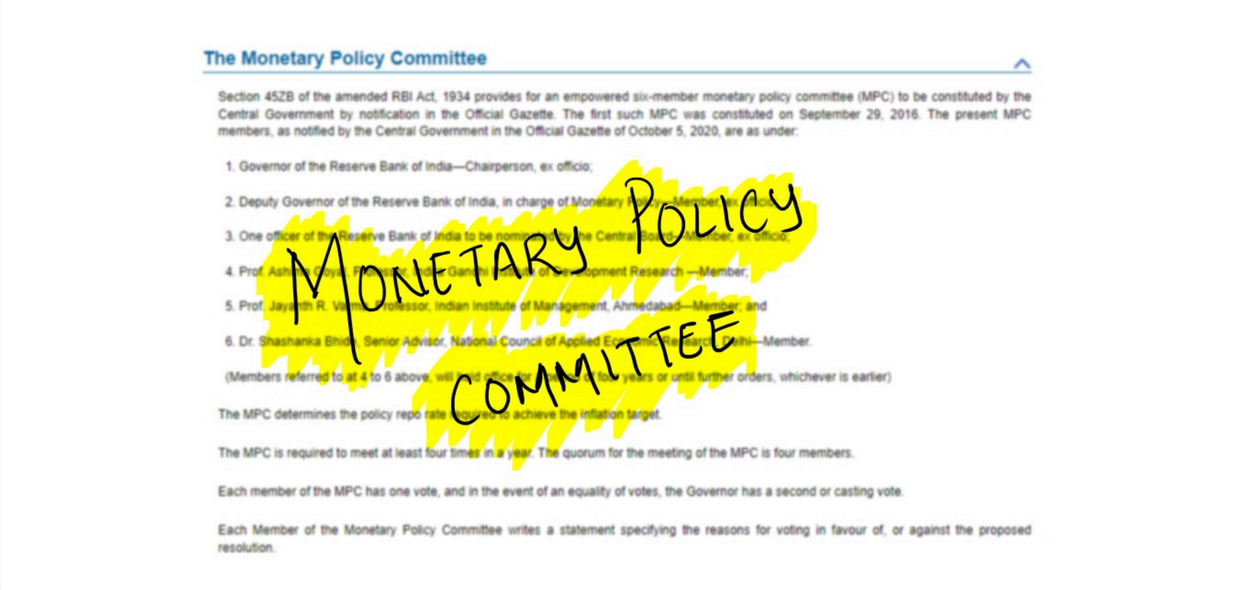
The Central Government has notified 4% Consumer Price Index (CPI) inflation as the target for the period from August 5, 2016, to March 31, 2021, with the upper tolerance limit of 6% and the lower tolerance limit of 2%.

Factors that constitute a failure to achieve the inflation target:

i. the average inflation is more than the upper tolerance level of the inflation target for any 3 consecutive quarters, OR
ii. the average inflation is less than the lower tolerance level for any 3 consecutive quarters.

While the Government of India sets the FITF, it is the RBI which operates the Monetary Policy Framework of the country. The amended RBI Act explicitly provides the legislative mandate to the Reserve Bank to operate the monetary policy framework of the country.

6. Monetary Policy



The Monetary Policy Committee

Section 45ZB of the amended RBI Act, 1934 provides for an empowered six-member monetary policy committee (MPC) to be constituted by the Central Government by notification in the Official Gazette. The first such MPC was constituted on September 29, 2016. The present MPC members, as notified by the Central Government in the Official Gazette of October 5, 2020, are as under:

1. Governor of the Reserve Bank of India—Chairperson, ex officio;
2. Deputy Governor of the Reserve Bank of India, in charge of Monetary Policy—Member, ex officio;
3. One officer of the Reserve Bank of India to be nominated by the Central Board—Member, ex officio;
4. Prof. Ashima Goyal, Professor, Indira Gandhi Institute of Development Research —Member;
5. Prof. Jayanth R. Varma, Professor, Indian Institute of Management, Ahmedabad—Member; and
6. Dr. Shashanka Bhide, Senior Advisor, National Council of Applied Economic Research, Delhi—Member.

(Members referred to at 4 to 6 above, will hold office for a period of four years or until further orders, whichever is earlier)

The MPC determines the policy repo rate required to achieve the inflation target.

The MPC is required to meet at least four times in a year. The quorum for the meeting of the MPC is four members.

Each member of the MPC has one vote, and in the event of an equality of votes, the Governor has a second or casting vote.

Each Member of the Monetary Policy Committee writes a statement specifying the reasons for voting in favour of, or against the proposed resolution.

Now in India, the policy interest rate required to achieve the inflation target is decided by the Monetary Policy Committee (MPC). The first such MPC was constituted on September 29, 2016. The present MPC members, as notified by the Central Government in the Official Gazette of October 5, 2020, are as under:

1. Governor of the Reserve Bank of India – Chairperson, ex officio;
2. Deputy Governor of the Reserve Bank of India, in charge of Monetary Policy – Member, ex officio;
3. One officer of the Reserve Bank of India to be nominated by the Central Board – Member, ex officio;
4. Prof. Ashima Goyal, Professor, Indira Gandhi Institute of Development Research —Member;
5. Prof. Jayanth R. Varma, Professor, Indian Institute of Management, Ahmedabad—Member; and
6. Dr. Shashanka Bhide, Senior Advisor, National Council of Applied Economic Research, Delhi— Member.

Members referred to at 4 to 6 above, will hold office for a period of 4 years or until further orders, whichever is earlier.

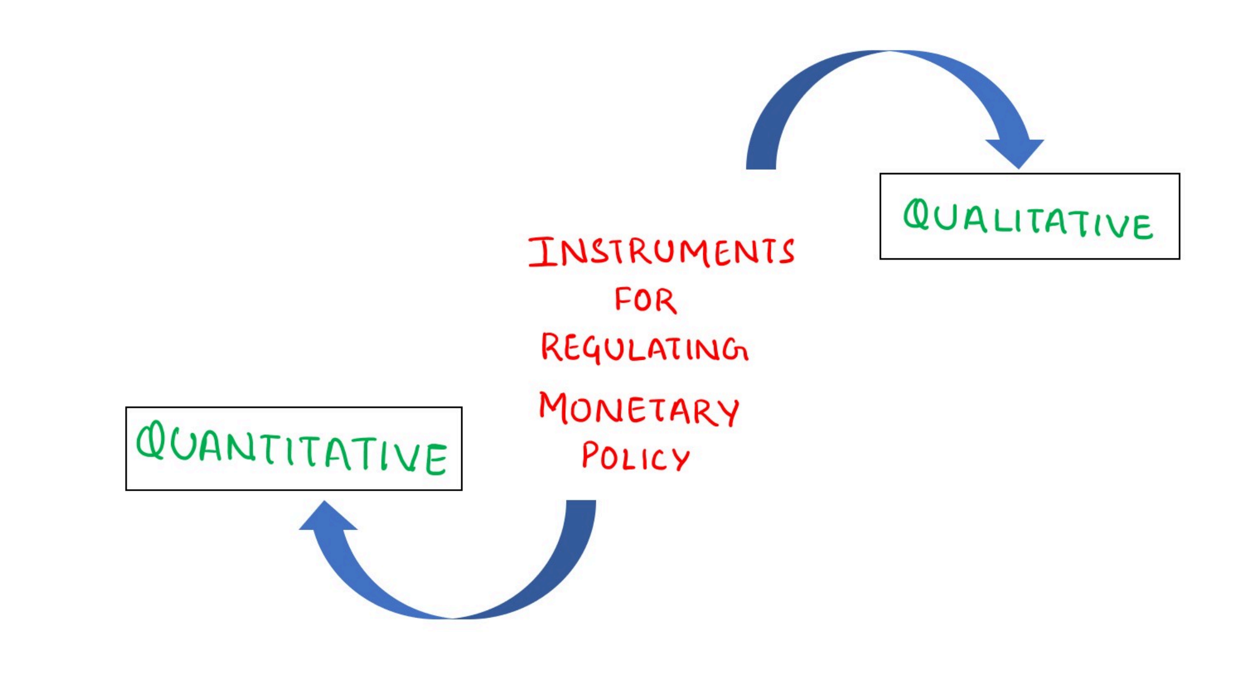
The MPC is required to meet at least 4 times in a year. The quorum for the meeting of the MPC is 4 members. Each member of the MPC has 1 vote, and in the event of an equality of votes, the Governor has a second or casting vote.

The resolution adopted by the MPC is published after the conclusion of every meeting of the MPC. Once in every 6 months, the Reserve Bank is required to publish a document called the **Monetary Policy Report** to explain:

i. the sources of inflation and
ii. the forecast of inflation for 6-18 months ahead.

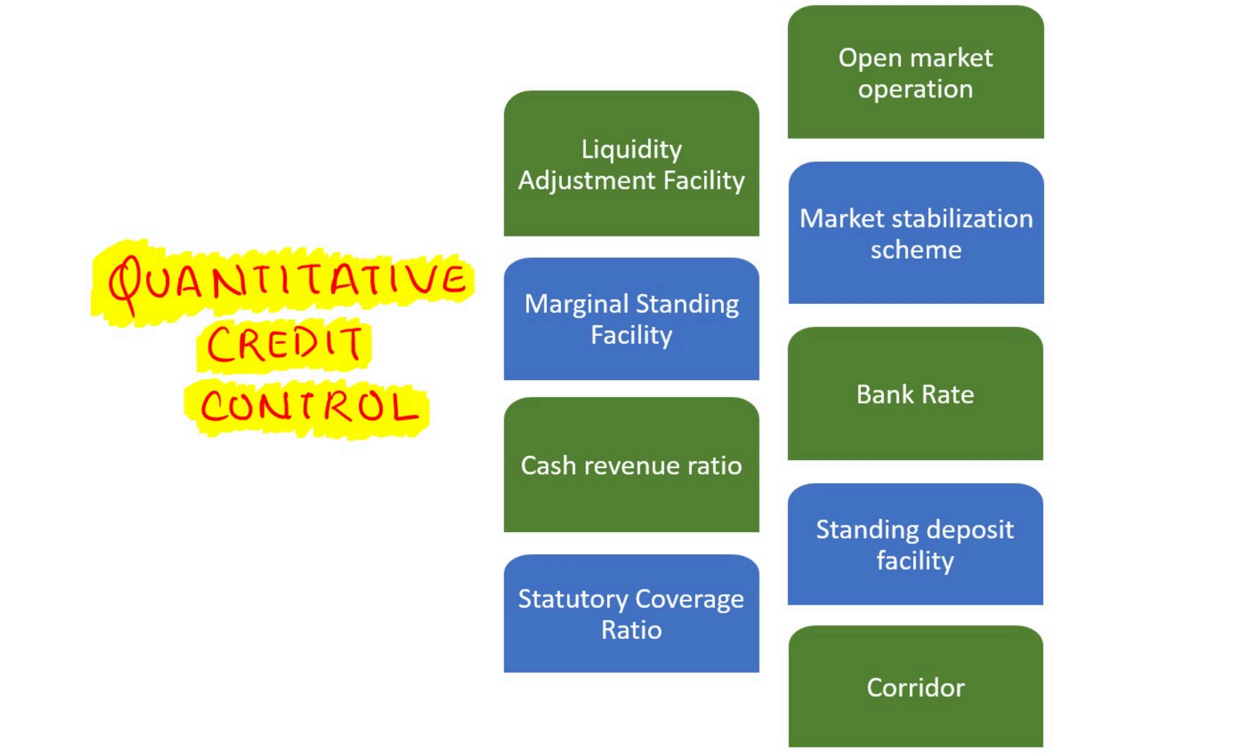
7. Regulating Monetary Policy

The Reserve Bank of India uses various instruments for regulating monetary policy. These instruments can be qualitative or quantitative in nature. Whereas, quantitative instruments influence the volume of Money and credit supply in the system, the qualitative instruments regulate credit supply in certain selective sectors (directions) of the economy.



7. Regulating Monetary Policy

The Quantitative Credit Control (**controls the volume**) measures are discussed below.



Liquidity Adjustment Facility (LAF): It is a tool used in monetary policy, which operates through repo and reverse repo auctions of the securities. The RBI conducts repos, when there is excess liquidity in the system. Under this, the RBI sells securities with an agreement to re-purchase it at a pre-determined date and rate (called **repo rate**). Reverse repo is just the opposite of repo, in which the RBI purchases securities with a committment to sell at a pre-determined date and rate (called *reverse repo rate*). This leads to injection of liquidity in the system.

Thus, the *Repo Rate* is the (fixed) interest rate at which the Reserve Bank provides overnight liquidity to banks against the collateral of government and other approved securities under the liquidity adjustment facility (LAF). The *Reverse Repo Rate* is the (fixed) interest rate at which the Reserve Bank absorbs liquidity, on an overnight basis, from banks against the collateral of eligible government securities under the LAF.

The LAF consists of overnight as well as term repo auctions.

Marginal Standing Facility (MSF): A facility under which scheduled commercial banks can borrow an additional amount of overnight money from the RBI by dipping into their Statutory Liquidity Ratio (SLR) portfolio up to a limit at a penal rate of interest. This provides a safety valve against unanticipated liquidity shocks to the banking system.

Cash Reserve Ratio (CRR): The average daily balance that a bank is required to maintain with the Reserve Bank as a share of such percent of its Net Demand and Time Liabilities (NDTL) that the Reserve Bank may notify from time to time in the Gazette of India.

Statutory Liquidity Ratio (SLR): The share of NDTL that a bank is required to maintain in safe and liquid assets, such as unencumbered government securities, cash and gold. Changes in SLR often influence the availability of resources in the banking system for lending to the private sector.

Open Market Operations (OMOs): These include both, outright purchase and sale of government securities by the RBI, for injection and absorption of durable liquidity, respectively.

Market Stabilisation Scheme (MSS): This instrument for monetary management was introduced in 2004. Surplus liquidity of a more enduring nature arising from large capital inflows is absorbed through the sale of short-dated government securities and treasury bills. The cash so mobilized, is held in a separate government account with the Reserve Bank.

Bank Rate: It is the rate at which the Reserve Bank is ready to buy or rediscount bills of exchange or other commercial papers. The Bank Rate is published under Section 49 of the Reserve Bank of India Act, 1934. This rate has been aligned to the MSF rate and, therefore, changes automatically as and when the MSF rate changes alongside policy repo rate changes.

Sterilization
1. Form of monetary action in which a central bank seeks to limit the effect of inflows and outflows of capital on the money supply 2. Most frequently involves the purchase or sale of financial assets by a central bank 3. Designed to offset the effect of foreign exchange intervention 4. Used to manipulate the value of one domestic currency relative to another 5. Initiated in the foreign exchange market

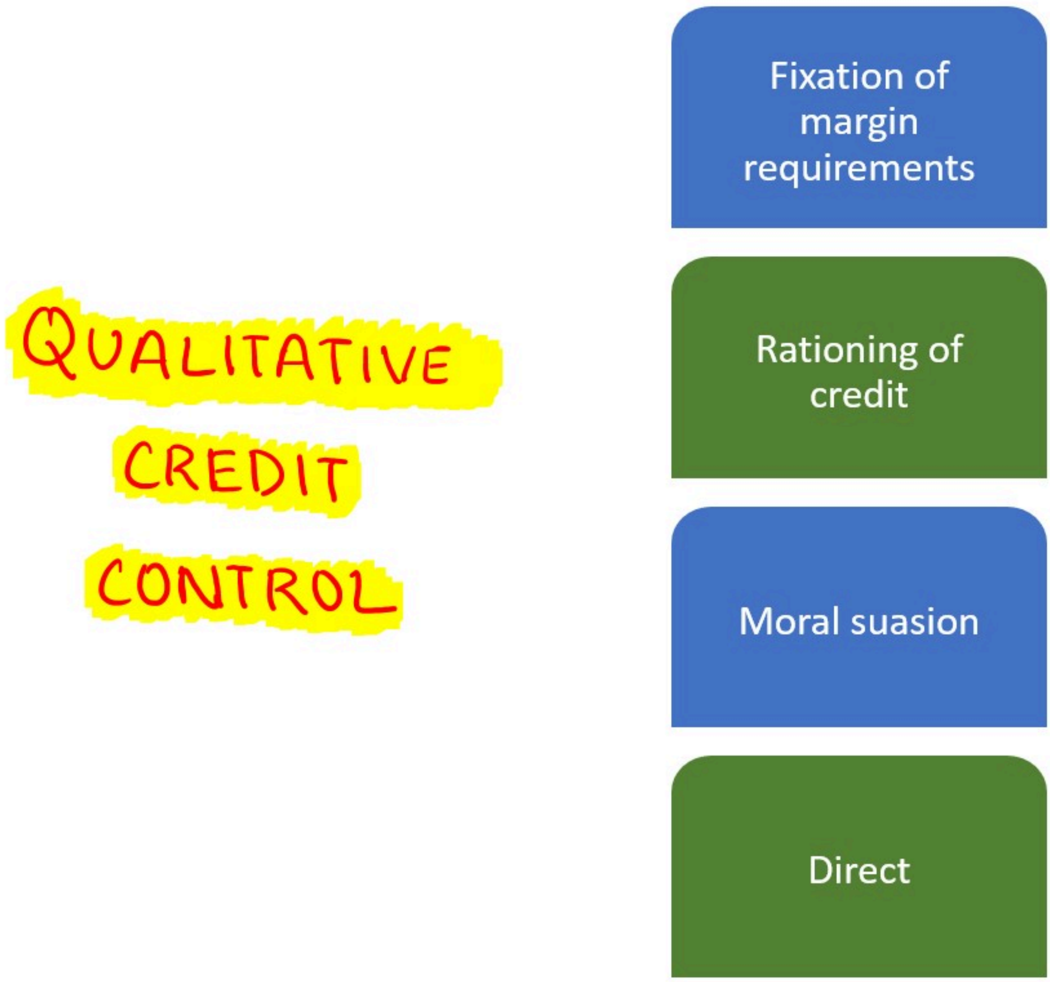
Quantitative easing
1. Monetary policy, adopted by the government to increase money supply in the economy to further increase lending by commercial banks and spending by consumers 2. Central bank infuses a pre-determined quantity of money into the economy by buying financial assets from commercial banks and private entities leading to an increase in banks' reserves

The **Standing Deposit Facility** will allow the RBI to absorb surplus funds from banks without collateral. This concept was first recommended by the Urjit Patel committee report in 2014. The Government approved it in February 2018.

The **Corridor** in monetary policy refers to the difference between the Reverse Repo rate and the high cost MSF rate. The corridor structure for the policy rate is a helping guide for the RBI to design its monetary policy operations.

7. Regulating Monetary Policy

The Qualitatitive Credit Control (**controls the distribution**) measures are discussed below.



Fixation of Margin Requirements: The term Margin denotes that percentage of the loan amount, which cannot be borrowed from bank. Hence this portion of finance is to be compulsorily brought by the borrower from own source. By using this method, during the period of inflation with a view to control credit, the RBI raises the margin and during deflation it lowers the margin to expand the credit.

Rationing of Credit: Rationing of credit is another method of selective credit control. It is made by regulating the purposes for which the loans are given among the various member banks. Finance is to be distributed to various sectors as per these requirements. Priority sector should be given preference in lending loans.

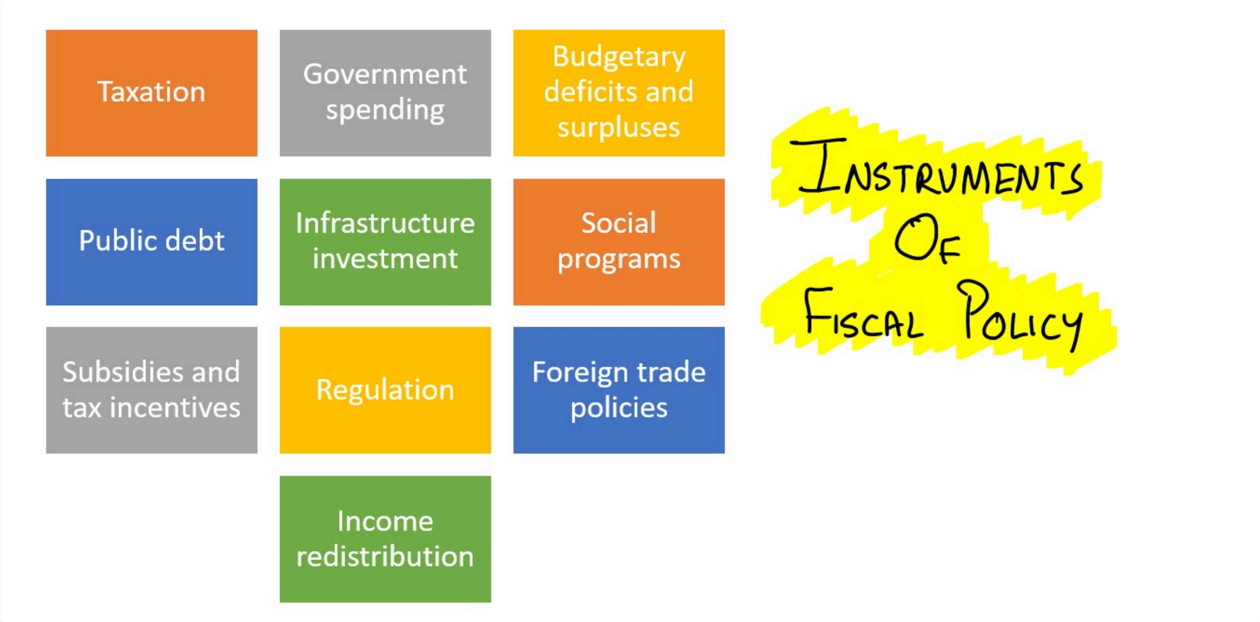
Moral Suasion: Under Moral Suasion, RBI issues periodical letters, advises, meetings etc. to bank to exercise control over credit in general or advances against particular commodities.

Direct Action: Under this method, if the commercial Banks do not follow the policy of the Central Bank, then the Central Bank has the only recourse to direct action (like punishment).

8. Fiscal Policy

Fiscal policy refers to the utilization of government revenue generation, primarily through taxes, and government expenditure, or spending, as a means to influence and manage the economy. It is frequently employed to stabilize the economy throughout the various phases of the business cycle.

Key instruments of fiscal policy include:



- **Taxation:** Governments can adjust tax rates and structures to influence consumer and business behavior. For example, reducing income taxes can stimulate consumer spending and investment, while increasing taxes can reduce overall demand.
- **Government Spending:** Governments can control their level of spending on various programs and projects. Increasing government expenditure can boost economic activity, while decreasing it can help reduce inflationary pressures.
- **Budgetary Deficits and Surpluses:** Governments can intentionally run budget deficits (spending more than they collect in revenue) or surpluses (collecting more than they spend) to influence economic conditions. Deficits can stimulate economic growth, while surpluses can help control inflation.
- **Public Debt:** Governments can manage their public debt levels to finance deficits or pay off debt when economic conditions allow. Effective debt management ensures that borrowing costs are manageable and sustainable.
- **Infrastructure Investment:** Investing in infrastructure projects, such as transportation and public facilities, can stimulate economic growth and job creation.

- **Social Programs:** Governments can use social programs like unemployment benefits, food assistance, and housing subsidies to provide a safety net during economic downturns.
- **Subsidies and Tax Incentives:** Governments may provide subsidies to specific industries or offer tax incentives to encourage particular economic activities, such as research and development or renewable energy production.
- **Regulation:** Regulatory policies can shape economic behavior and market dynamics. For example, financial regulations can influence lending practices and stability in the banking sector.
- **Foreign Trade Policies:** Governments can use trade policies, such as tariffs and trade agreements, to impact international trade flows and protect domestic industries.
- **Income Redistribution:** Tax and transfer policies can be used to redistribute income and wealth, reducing inequality and addressing social issues.

---

9. Measurement of Inflation

Index numbers are an important statistical tool used to measure changes in a group of related variables. They represent the general trend of diverging ratios and are expressed in terms of percentage. The purpose of index numbers is to provide a comparison between the current state of a variable and its state in a particular base period. The base period is the period with which the comparison is to be made, and it is assigned an index number of 100.

For instance, suppose you want to track the price changes of a particular good or service from 2010 to 2020. In this case, 2010 becomes the base year, and its value is assigned an index number of 100. If the price of the good or service is 120 in 2020, then the index number for 2020 would be 120/100 = 1.2. This means that the price has increased by 20% over the base year.

There are different types of index numbers used to measure changes in various variables.

i. **Price index numbers** are used to track the prices of certain goods and permit comparison between them. For example, the Consumer Price Index (CPI) measures the changes in the prices of goods and services consumed by households. Similarly, the Producer Price Index (PPI) measures the changes in the prices of goods produced by businesses.
ii. **Quantity index numbers**, on the other hand, measure the changes in the physical volume of production, construction or employment. These are used to track changes in output levels, employment levels or other physical variables.

Inflation is a major concern for most countries, and index numbers are used to track inflation rates. The type of inflation that reflects the rate of change in prices of all goods and services in an economy over a period of time is called **Headline Inflation.** This is often referred to as the 'All Commodities Inflation Rate'. It is a measure of the overall inflation rate in the economy.

Core inflation is another measure of inflation that excludes the prices of food and energy items. These items are excluded because their prices are much more volatile and can distort the inflation rate. Core inflation is calculated by excluding food and energy items from the headline inflation rate.

There are 2 main set of inflation indices for measuring price level changes in India – the Wholesale Price Index (WPI) and the Consumer Price Index (CPI). In addition, Gross Domestic Product (GDP) deflator and Private Final Consumption Expenditure (PFCE) deflator from the National Accounts Statistics (NAS under Ministry of Statistics & Programme Implementation) provide an implicit economy-wide inflation estimate.

---

9. Measurement of Inflation

In case of Consumer Price Index (CPI), there are 4 indices to measure it:

i. CPI for all India or CPI combined
ii. CPI for Agricultural Labourers (AL)
iii. CPI for Rural Labourers (RL)
iv. CPI for Industrial Workers (IW)

Two Ministries, Ministry of Statistics and Programme Implementation (MoSPI) and Ministry of Labour and Employment (MOLE) are engaged in the construction of different CPIs for different groups/sectors. CPI inflation is also called *Retail inflation* as the prices are quoted from retailers.

CPI by MoSPI

The Central Statistics Office CSO, (now renamed NSO) which comes under MoSPI, is constructing the Rural, Urban and the Combined CPIs (for both All India level and State wise). They are published from 2011 onwards. Of these, the **CPI Combined** is the most important of all the CPIs as it is relevant for all categories of people. Its base year is 2012 and methodology is "Geometric mean for elementary item index and Laspeyres Index Formula for higher level index".

Note: The Laspeyres Index is calculated by working out the cost of a group of commodities at current prices, dividing this by the cost of the same group of commodities at base period prices, and then multiplying by 100.

There are 6 groups of items in CPI:

1. **Food and Beverages** (45.86% weightage): Cereals and products, Meat and fish, Egg, Milk and products, Oils and fats, Fruits, Vegetables, Pulses and products, Sugar and confectionery, Spices, Non-alcoholic beverages, Prepared meals, snacks, sweets etc.
2. **Pan, tobacco and intoxicants** *(2.38% weightage)*
3. **Clothing and footwear** *(6.53% weightage)*
4. **Housing** *(10.07% weightage)*
5. **Fuel and light** *(6.84% weightage)*
6. **Miscellaneous** *(28.32% weightage)*: Household goods and services, Health, Transport and communication, Recreation and amusement, Education, Personal care and effects.

Consumer Food Price Index (CFPI) has weightage of 39.06% in Combined CPI. Out of 12 sub-groups contained in 'Food and Beverages' group, CFPI is based on 10 sub-groups, excluding 'Non-alcoholic beverages' and 'Prepared meals, snacks, sweets etc.'.

In April 2014, the RBI selected the All India CPI (of NSO) as the inflation index to target inflation under its new inflation targeting monetary policy framework.

CPIs by MoLE

The Labour Bureau, Ministry of Labour and Employment (MoLE) is preparing different indices for various categories of people. These are CPI for Rural Labourares (CPI-RL), CPI for Agricultural Labourares (CPI-AL) and CPI for Industrial Workers (CPI-IW). There was a CPI For Urban Non-Manual Employees (CPI-UME), but it was discontinued from April 2010. The methodology of measurement is "Weighted arithmetic mean according to Laspeyres Index Formula". Since these CPIs were for specific categories of workers, it lacked the quality of an all India index.

CPI is used in calculating purchasing power of money and real wage.

Purchasing power of money = $\frac{1}{Cost \ of \ living \ index}$

Real wage = $\frac{Money \ Wage}{Cost \ of \ living \ index}$ × 100

---

9. Measurement of Inflation

The Wholesale Price Index WPI is published, on monthly basis, by the Office of Economic Adviser, Ministry of Commerce and Industry. It is in use since 1942 and is being published from 1947 regularly. It has a long history for serving as the nationwide inflation indicator till the emergence of the combined CPI in 2011. An important feature of the WPI which separate it from the CPI is that prices are collected from **Wholesalers**. The methodology of measurement is "Weighted arithmetic mean according to Laspeyres Index Formula". The base year is 2011-12.

There are 3 groups of commodities in WPI:

1. **Primary Articles** (22.62% weightage): It is divided into 2 further categories:

i. Food Articles (15.26%): Cereals, Paddy, Wheat, Pulses, Vegetables, Potato, Onion, Fruits, Milk, Eggs, Meat & Fish
ii. Non-Food Articles (4.12%): Oil Seeds
iii. Minerals
iv. Crude Petroleum & Natural gas: Crude Petroleum

2. **Fuel and Power** (13.15% weightage): LPG, Petrol, HSD

3. **Manufactured Products** (64.23% weightage): Food Products (9.12%), Vegetable & Animal Oils and Fats, Beverages, Tobacco Products, Textiles, Wearing Apparel, Leather and Related Products, Wood and of Products of Wood and Cork, Paper and Paper Products, Chemicals and Chemical Products, Pharmaceuticals, Medicinal Chemical and Botanical Products, Rubber and Plastics Products, other Non-Metallic Mineral Products (Cement, Lime and Plaster), Basic Metals (Mild Steel - Semi Finished Steel), Fabricated Metal Products, except Machinery and Equipment

For a long time, India has been using WPI to measure the inflation. In 2014, the Reserve bank of India (RBI) adopted new CPI Consumer Price Index (CPI) as the key measure of inflation. The very first reason for replacing WPI from CPI is to include services too. Currently, services account for nearly 60% of GDP and a large number of these services are not traded with other countries. WPI didn't serve to the services such as healthcare, education, transportation, and tourism etc. whereas CPI does.

Consumer Price Index (CPI) used to calculate for rural and urban part of the country simultaneously so that specific measures can be taken for rural and urban benefits. CPI gives the general idea of retail inflation and prices driven by the consumer demand and supply. It is a better indicator of inflation for guiding monetary policy decisions & interest rate decisions than WPI Inflation.

---

9. Measurement of Inflation

India's first official Housing Price Index (named RESIDEX) is published by the National Housing Bank (NHB). It is published based on data from Central Registry of Securitisation Asset Reconstruction and Security Interest of India (CERSAI).

The base year is 2012-13 and it is updated every quarter.

The scope has been widened under NHB RESIDEX brand, to include housing price indices (HPI), land price indices (LPI) and building materials price indices (BMPI), and also housing rental index (HRI).

---

10. Inflation and Interest Rate

Inflation and interest rates have a close relationship. Inflation is the rate at which the general level of prices for goods and services is rising, and interest rates are the rates at which borrowers are charged to borrow money from lenders. Inflation affects interest rates, and interest rates affect inflation. Changes in one will often lead to changes in the other.

The nominal interest rate is the rate that appears on the loan agreement and is often discussed in the news media. The real interest rate, on the other hand, is the nominal rate minus the inflation rate. If inflation is high, then the real interest rate is lower than the nominal rate, and lenders lose purchasing power. For this reason, lenders and borrowers are more concerned about the real interest rate than the nominal rate.

The equation for the real interest rate shows how inflation affects interest rates. When inflation is high, the real interest rate is low, and when inflation is low, the real interest rate is high. For example, if inflation is 3% and the nominal interest rate is 5%, the real interest rate is only 2%. This means that the actual purchasing power of the lender's money is decreasing at a rate of 3%, while they are earning interest at a rate of 5%. The lender is, therefore, losing purchasing power, and this discourages lending.

Inflationary expectations play a significant role in determining interest rates. If lenders and borrowers expect inflation to increase, they will demand higher nominal interest rates to compensate for the expected decrease in purchasing power of the money lent or borrowed. Similarly, if inflation is expected to decrease, lenders and borrowers will accept lower nominal interest rates since they expect to gain in purchasing power.

Let's say that an investor lends Rs 100 for one year at a nominal interest rate of 6%. If the inflation rate during that year is 2%, then the real interest rate earned by the investor is 4% (6% - 2% = 4%). However, if the inflation rate rises to 4%, then the real interest rate earned by the investor falls to 2% (6% - 4% = 2%). Conversely, if the inflation rate falls to 1%, then the real interest rate earned by the investor rises to 5% (6% - 1% = 5%).

Central banks play a crucial role in regulating inflation and interest rates. When inflation is high, central banks may raise interest rates to discourage borrowing and spending, which in turn reduces demand and eventually leads to a decrease in inflation. On the other hand, when inflation is low, central banks may lower interest rates to encourage borrowing and spending, which increases demand and eventually leads to an increase in inflation.

---

10. Inflation and Interest Rate

The Fisher effect is an economic theory created by economist Irving Fisher that describes the relationship between inflation and both real and nominal interest rates.

$(1+i) = (1+r)(1+\pi)$

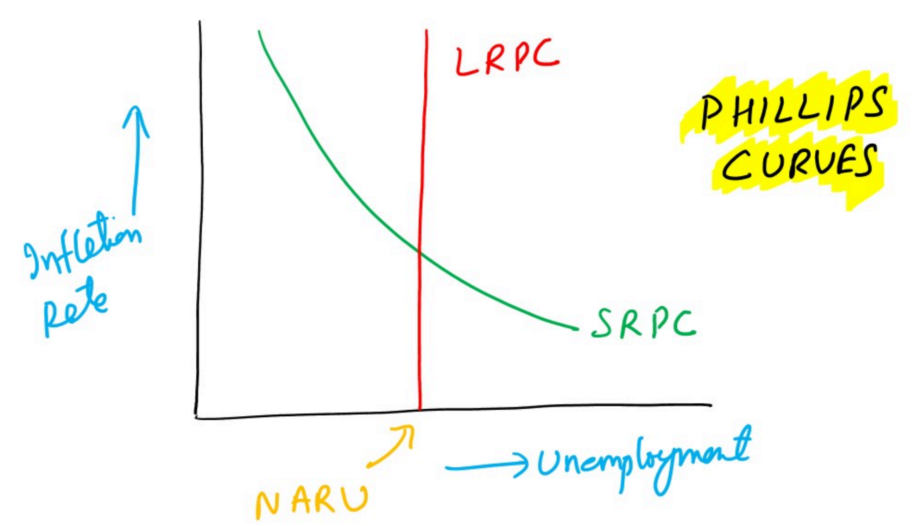
which is approximated to the equation, $r = i - \pi$

where:
r = Real interest rate
i = Nominal interest rate
π = Rate of inflation

---

11. Inflation and Unemployment

The relationship between inflation and unemployment is often referred to as the Phillips curve. The **Phillips curve** is a graphical representation of the inverse relationship between the rate of inflation and the rate of unemployment in an economy. According to the Phillips curve, as the rate of unemployment decreases, the rate of inflation increases, and vice versa.



When unemployment is high, there is a large pool of workers looking for jobs. This increases competition among job seekers, which puts downward pressure on wages. As a result, firms can hire workers at lower wages, which decreases their costs and increases their profits. This leads to lower prices, or lower inflation, in the economy.

On the other hand, when unemployment is low, there are fewer workers available for jobs. This creates upward pressure on wages, as firms must compete for workers. When firms must pay higher wages, their costs increase, which leads to higher prices, or higher inflation, in the economy.

The Phillips curve is not a fixed relationship, and it can shift over time due to changes in the economy. For example, an increase in productivity can lead to lower costs for firms, which can result in lower prices and lower inflation, even at low levels of unemployment.

In some cases, the Phillips curve can break down entirely, as was seen during the stagflation of the 1970s. During this period, the economy experienced both high inflation and high unemployment, which was contrary to the relationship predicted by the Phillips curve. This breakdown was due to a number of factors, including supply shocks that increased the cost of production and disrupted markets.

Overall, the Phillips curve highlights the trade-off between inflation and unemployment in the short run. Policymakers can use this relationship to guide their decisions on monetary and fiscal policy. For example, if the economy is experiencing high levels of unemployment, policymakers may choose to pursue expansionary policies, such as lower interest rates or increased government spending, in order to stimulate economic growth and decrease unemployment. However, these policies may also lead to higher inflation in the short run.

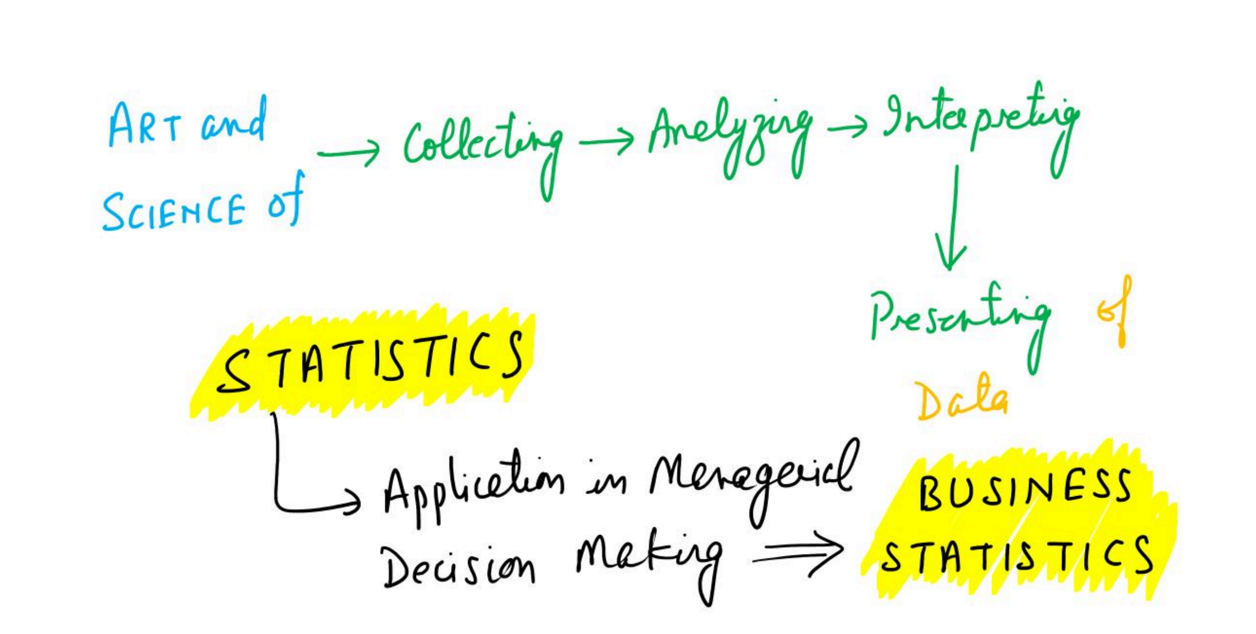
The Phillips curve describes the inverse relationship between unemployment and inflation. In the short run, when the expected price level changes, there is a corresponding change in the short-run aggregate supply curve (Short Run Phillips Curve SRPC). The intersection of this curve with different aggregate demand curves produces the short-run Phillips curve, which shows the relationship between unemployment and inflation. If inflation turns out as expected, unemployment equals the natural rate. If inflation exceeds expectations, unemployment falls below the natural rate, and if inflation is less than expected, unemployment exceeds the natural rate.

In the long run (Long Run Phillips Curve LRPC), wages and prices adjust to changes in aggregate demand, and the economy returns to its natural rate of unemployment (Natural Rate of Unemployment NARU). This adjustment results in a vertical long-run Phillips curve, indicating that in the long run, unemployment is independent of the rate of inflation. Thus, policy-makers cannot choose between unemployment and inflation in the long run, but only among alternative rates of inflation.

12. Inflationary Gap

An inflationary gap is the difference between the current level of real Gross Domestic Product (GDP) in an economy and the potential GDP that would be achieved if the economy were operating at full employment. It signifies a situation where the economy is producing more than its sustainable capacity, resulting in excess demand for goods and services. This excess demand can lead to rising prices and inflationary pressures. Policymakers often monitor and address inflationary gaps to maintain price stability and economic equilibrium by implementing measures like raising interest rates or reducing government spending.

1. Statistics

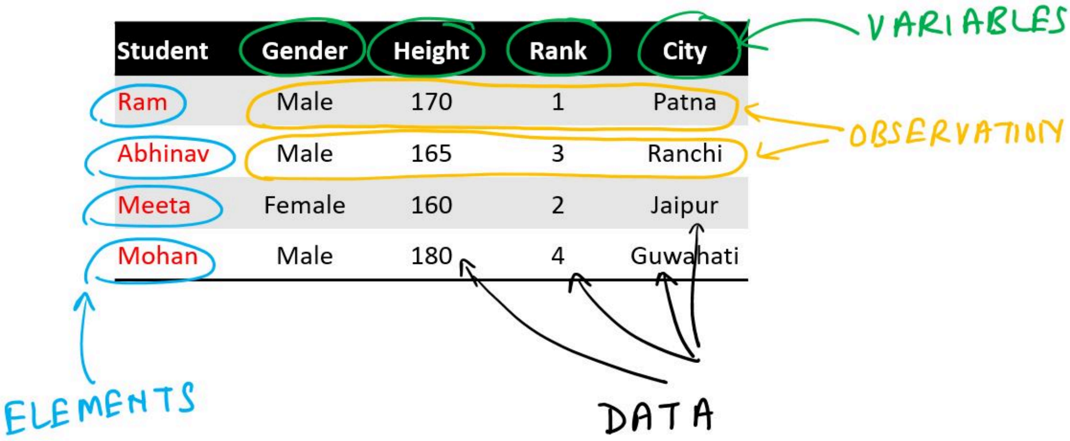


Statistics is both an art and a science encompassing the systematic gathering, organization, analysis, interpretation, and presentation of data. It involves employing mathematical techniques and methodologies to explore and understand patterns, trends, and relationships within datasets.

Managerial statistics, also known as business statistics, is the application of statistical methods and techniques to aid managerial decision-making in business and organizational settings. It involves using data analysis to solve business problems, make informed decisions, and improve operational efficiency and effectiveness.

2. Data

In statistics, data is the most important thing for any study. Data are the distinct factual pieces of the information. The data is considered as a plain fact. It is also called as the raw data, from which the statistics are extracted.



Data are the facts and figures collected, analyzed, and summarized for presentation and interpretation.

All the data collected in a particular study are referred to as the **data set** for the study.

The data set is comprised of Elements, Variables and Observations.

Elements are the entities, based on which the data is collected. Elements are very important feature of the data.

Variables are the characteristics of interest for the corresponding elements. A quantity whose value changes across the population and can be measured is called variable. For instance, consider a sample of employed individuals. The variables for this set of the population can be industry, location, gender, age, skills, job-type, etc. The value of the variables will differ with each employee.

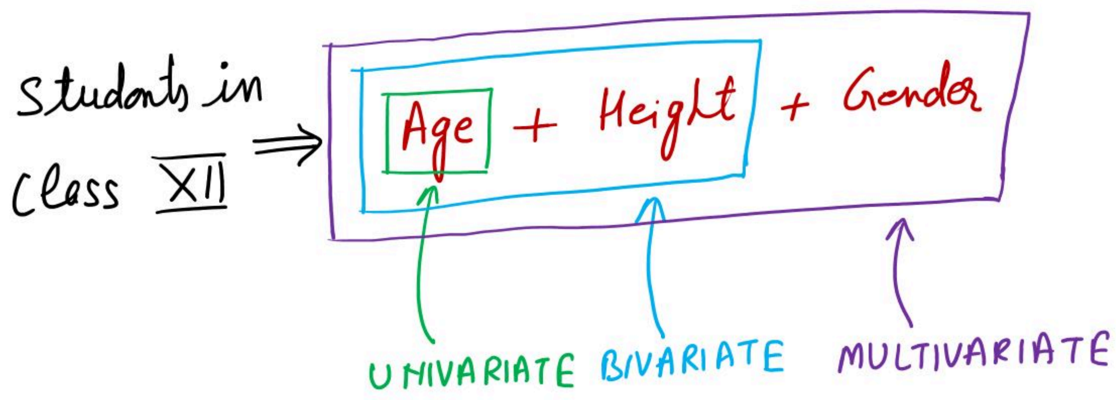
Observations are the set of measurements that are collected for a specific element.

We will understand this with the help of an example.

Earnings (in thousand Rs.)	Teacher	Doctor	Painter
Weekly	5	8	2
Monthly	150	100	80
Annually	1800	2000	800

In the above table:
Elements: Teacher, Doctor, Painter
Variables: Weekly, Monthly, and Annual earning
Observations: 5, 150, 1800, 8, 100, 2000, 2, 80, 800

3. Types of Variables



There are three types of variables. We will use example of data on houses to explain these 3 types.

1. Univariate Variables

Univariate data involves a single variable or characteristic, with one piece of information recorded for each item.

In the example of univariate data for housing prices, you might have a dataset consisting solely of the selling prices of houses in a particular area. Each entry records a single piece of information (price) for each house. Analyzing this univariate dataset would involve calculating summary statistics such as the average selling price, median price, range of prices, or standard deviation, providing insights into the typical pricing and the variability of housing prices in that area.

2. Bivariate Variables

Bivariate data involves exactly two variables recorded for each item, allowing the study of the relationship between these two variables.

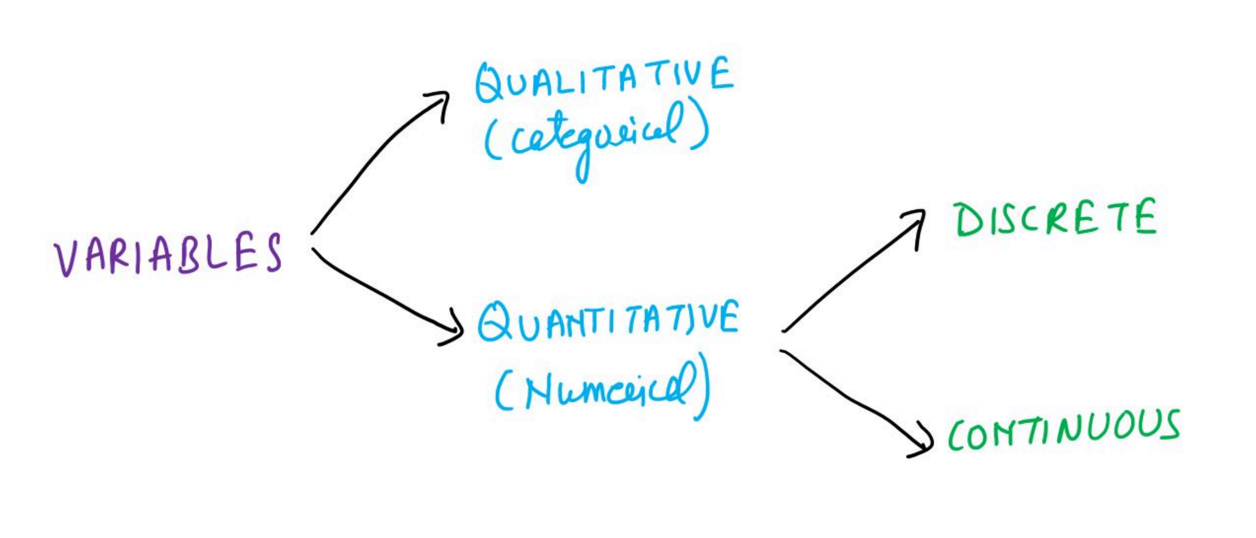
Moving to bivariate data for housing prices, you could expand the dataset to include two variables per entry. For instance, alongside the selling price, you could include the square footage of each house. By studying these two variables together, you can explore the relationship between house size and selling price. Using techniques like scatter plots, you can visualize how changes in square footage relate to changes in house prices, allowing you to understand if there's a correlation or pattern between the two.

3. Multivariate Variables

Multivariate data involves three or more variables recorded for each item, enabling the study of relationships among multiple variables.

Expanding further to a multivariate dataset for housing prices, you could include additional factors influencing house prices, such as location, number of bedrooms, and neighborhood crime rate. Each entry in the dataset would now consist of multiple variables—selling price, square footage, location, number of bedrooms, and crime rate. Analyzing these multiple variables together enables a comprehensive understanding of how various factors collectively influence housing prices. Employing advanced statistical methods like regression analysis, you can predict housing prices based on these multiple variables and uncover complex relationships among them, providing valuable insights for real estate decisions.

4. Categorical and Quantitative Variables



Data can be classified as either categorical or quantitative.

1. Qualitative Variables (Categorical Variables)

Qualitative variables represent characteristics or qualities and can't be measured numerically. They are categorical and often described by labels or categories.

Examples of Qualitative Variables are:

- Hair Color: Categorizing individuals based on hair color—blonde, brunette, red, black, etc.
- Vehicle Types: Classifying vehicles by type—sedan, SUV, truck, motorcycle, etc.
- Educational Levels: Grouping individuals by education level—high school graduate, bachelor's degree, master's degree, etc.

2. Quantitative Variables (Numerical Variables)

Quantitative variables, in contrast to qualitative variables, can be measured numerically. They can further be classified into discrete and continuous variables.

2.1 Discrete Variables: Discrete variables represent countable and finite values. They take on specific numerical values and often whole numbers.

Examples of Discrete Variables are:

- Number of Children in Families: Counting the number of children in different families—0, 1, 2, 3, etc. (whole numbers).
- Number of Books in a Library: Counting the quantity of books in different libraries—100, 500, 1000, etc. (whole numbers).
- Customer Complaints in a Week: Recording the count of complaints received by a company—0, 5, 10, 20, etc. (whole numbers).

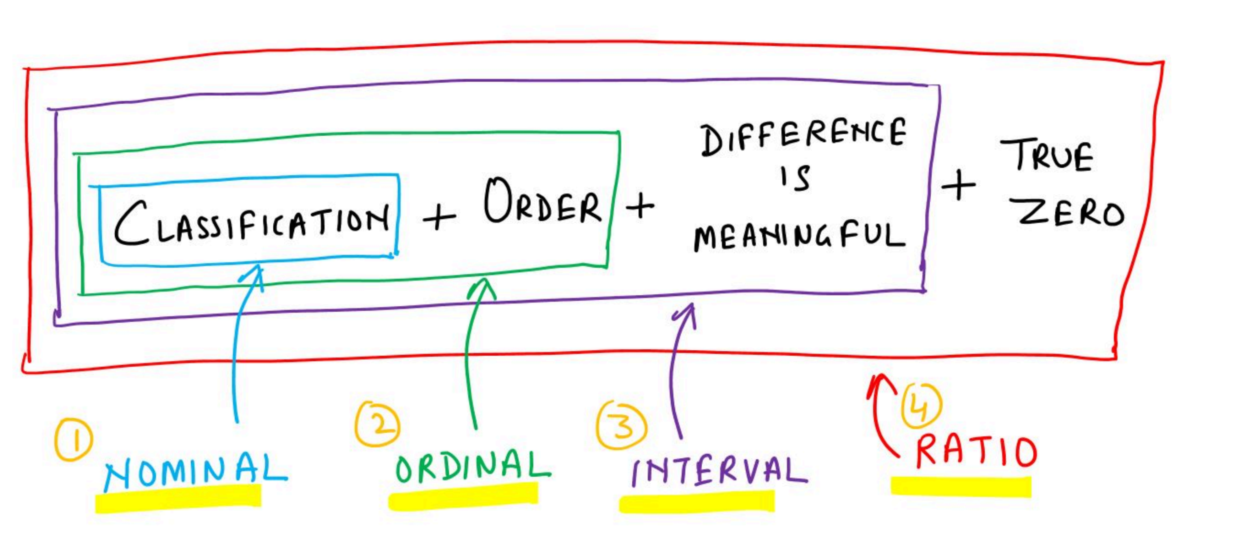
2.2 Continuous Variables: Continuous variables can take on any value within a given range. They are measured and can include fractional or decimal values.

Examples of Continuous Variables are:

- Weight of Individuals: Measuring the weight of people in kilograms or pounds—62.5 kg, 75.2 kg, 150 lbs, etc. (fractional or decimal values).
- Temperature: Measuring temperatures in degrees Celsius or Fahrenheit—23.5°C, 68.2°F, etc. (fractional or decimal values).
- Time Taken for a Process: Measuring time in hours, minutes, or seconds—3.25 hours, 45.6 minutes, etc. (fractional or decimal values).

5. Scales of Measurement

Nominal, Ordinal, Interval, and Ratio are defined as the four fundamental levels of measurement scales that are used to capture data in the form of surveys and questionnaires.



1. Nominal Scale

Nominal Scale, also called the categorical variable scale, is defined as a scale used for labeling variables into distinct classifications and doesn't involve a quantitative value or order. This scale is the simplest of the four variable measurement scales. Calculations done on these variables will be futile as there is no numerical value of the options.

Examples:

- Categorizing individuals by their political party preference— BJP, Congress, AAP, TMC etc.
- Classifying pets owned by households— dog, cat, fish, bird, etc.
- Grouping individuals based on marital status—single, married, divorced, widowed.
- Preference of smartphone- Apple – 1, Samsung – 2, OnePlus – 3.

2. Ordinal Scale

Ordinal Scale is defined as a variable measurement scale used to simply depicting the order of variables and not the difference between each of the variables. These scales are generally used to depict non-mathematical ideas such as frequency, satisfaction, happiness, a degree of pain, etc. It is quite straightforward to remember the implementation of this scale as 'Ordinal' sounds similar to 'Order', which is exactly the purpose of this scale.

In the Ordinal Scale, while there exists a clear order among categories, the distinctions between them lack precise numerical significance. Additionally, the presence of a zero point holds no meaningful value in this scale.

Examples:

- Classifying educational levels in order—High School Diploma, Associate's Degree, Bachelor's Degree, Master's Degree, Doctorate.
- Ranking movies by viewer preference—1 star, 2 stars, 3 stars, 4 stars, 5 stars.
- Categorizing pain levels—Mild, Moderate, Severe, Extreme.
- Ordering job positions within a company—Intern, Associate, Manager, Director, Vice President.

3. Interval Scale

The interval scale of measurement encompasses all the qualities of ordinal data, but in addition, the intervals between values represent a consistent and meaningful measurement. Interval data are always numeric (never categorical). In an interval scale, the differences between values hold significance, reflecting a fixed unit of measurement. However, the presence of a zero point doesn't indicate an absence but rather serves as a reference point without inherent meaning.

For example, three students with SAT scores of 620, 550, and 470 can be ranked or ordered in terms of best performance to poorest performance. In addition, the differences between the scores are meaningful. For instance, student 1 scored 620 – 550 = 70 points more than student 2, while student 2 scored 550 – 470 = 80 points more than student 3.

Examples:

- Temperature in Celsius or Fahrenheit
- Dates of Month
- Years on the Calendar
- IQ Scores
- Ph Level
- Credit Card Numbers
- Latitude and Longitude Coordinates

4. Ratio Scale

The scale of measurement for a variable is a ratio scale if the data have all the properties of interval data and the ratio of two values is meaningful. Variables such as distance, height, weight, and time use the ratio scale of measurement. This scale requires that a zero value be included to indicate that nothing exists for the variable at the zero point.

Examples:

- Temperature in Kelvin
- Height in Centimeters
- Weight in Kilograms
- Distance in Meters
- Time in Seconds
- Age in Years
- Income in Rupees

---

6. Cross-Sectional and Time Series Data

Data can also be grouped as cross sectional and time series data.

1. Cross-sectional Data

Cross-sectional data refer to observations collected from distinct individuals, entities, or groups at a specific point in time. This data collection method allows for comparisons among different subjects but only at a single moment, focusing on variation across various units within a population that share similar characteristics.

Examples of Cross-sectional Data:

- **Household Income Survey**: Gathering income data from various households in a country at a particular time to compare income levels among different socio-economic groups.
- **Educational Assessment**: Comparing test scores of students from different schools within a city on a specific date to analyze academic performance across diverse educational institutions.
- **Market Research on Consumer Preferences**: Collecting data on product preferences from customers of various age groups at a specific time to understand market trends among different demographics.

2. Time-series Data

Time-series data entails observations recorded at consistent intervals over time. This data collection method focuses on tracking changes, patterns, or trends within a single subject or variable, allowing for analysis of developments or variations over a specified time frame.

Examples of Time-series Data:

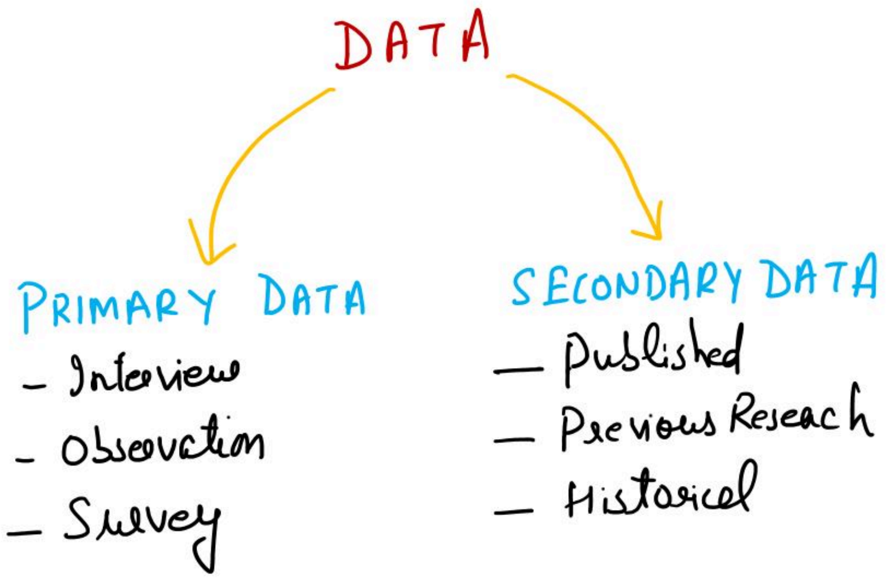
- **Stock Price Movements**: Monitoring the daily closing prices of a particular stock over several months to analyze its price fluctuations and trends.
- **Climate Data**: Recording daily temperature readings over a year to identify seasonal changes and patterns in temperature variations.
- **Economic Indicators**: Collecting monthly unemployment rates over several years to assess long-term trends and fluctuations in employment levels within an economy.

In summary, cross-sectional data involve observations from different units at one specific point in time, facilitating comparisons among distinct entities. Time-series data, on the other hand, comprises observations of a single entity or variable at regular time intervals, enabling analysis of trends, patterns, and changes over time.

---

7. Sources of Data

Data collection methods are used for collection of data for social research. While deciding about the method of data collection to be used for the study the research should be well acquainted with types of data: Primary and Secondary.



The **primary** data are those which are collected afresh and for the first time and thus happens to be original in character. The **secondary** data are those sort of data that researcher would be using for the study and accordingly he will have to select one or the other method of data collection.

Primary data is the data collected by the researcher themselves. Some of the sources for collecting primary data area:

- Interview
- Observation
- Action research
- Case studies
- Life histories
- Questionnaires
- Ethnographic research
- Longitudinal studies

Secondary sources are data that already exists and is sourced from some place, that has originally collected it. Secondary sources include:

- Previous research
- Official statistics
- Mass media products
- Diaries
- Letters
- Government reports
- Web information
- Historical data and information

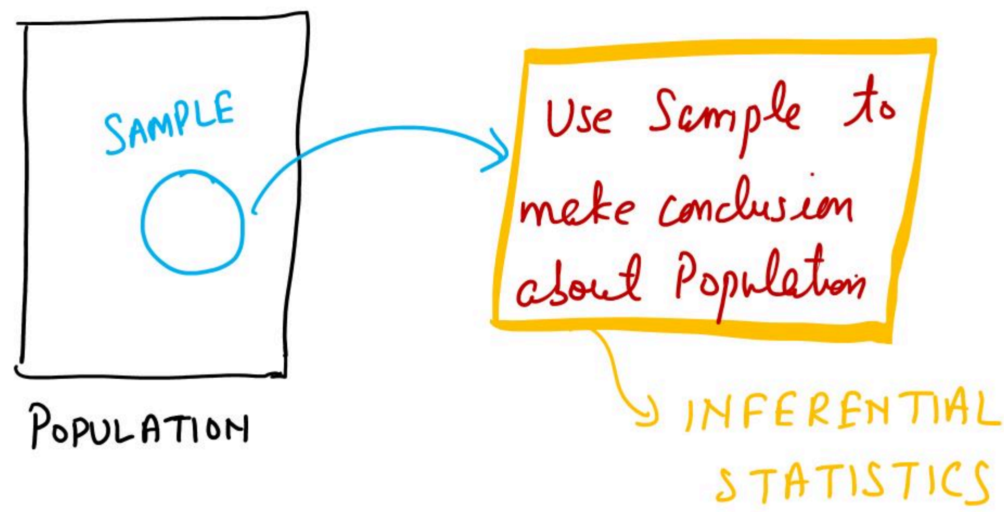
The combination of qualitative and quantitative and primary and secondary research is known as **triangulation** or **methodological pluralism**.

Tertiary sources of information are based on a collection of primary and secondary sources. Examples of tertiary sources include:

- textbooks (sometimes considered as secondary sources)

- dictionaries and encyclopedias
- manuals, guidebooks, directories, almanacs
- indexes and bibliographies

8. Inferential statistics



Inferential statistics involves using data from a sample to make conclusions or predictions about a larger population. It helps in generalizing findings beyond the observed sample to a broader population. This branch of statistics aids in drawing inferences, making predictions, and testing hypotheses about populations based on sample data.

Sample and Population

The **population** refers to the entire group that you want to draw conclusions about. It's the larger group or universe of interest, but it's often impractical or impossible to collect data from every individual within it.

A **sample** is a subset of the population that is selected for study. It's chosen to represent the population accurately and is used to make inferences or generalizations about the entire population.

Consider a scenario where you want to determine the average height of all students in a school (population). It's impractical to measure the height of every student, so you select a group of 100 students (sample) and measure their heights. By using inferential statistics, you can make estimations or draw conclusions about the average height of all students in the school based on the measured heights of the sample.

Descriptive vs. Inferential Statistics

Descriptive statistics summarize and describe data from a sample or population. For instance, calculating the mean, median, or standard deviation of the heights of the 100 students in the school sample represents descriptive statistics.

Inferential statistics, on the other hand, allow you to make predictions or inferences about a larger population based on data collected from a sample. For example, using the heights of the 100 students to estimate the average height of all students in the school is an application of inferential statistics.

Descriptive statistics would involve calculating the average height of the 100 measured students in the sample. Meanwhile, inferential statistics would take this information further to estimate the average height of all students in the school, providing a confident interval or hypothesis test result to infer if the sample's average represents the entire student body.

In essence, descriptive statistics help in summarizing and describing data, while inferential statistics aid in making predictions or drawing conclusions about larger populations based on sample data.

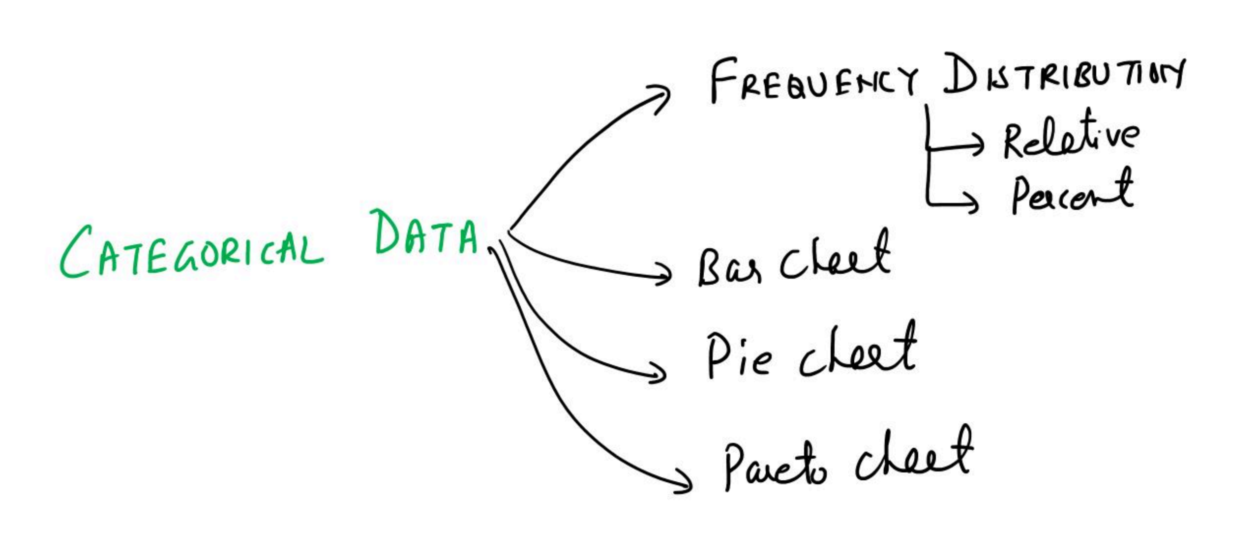
1. Introduction

The presentation of information is vital in uncovering patterns, trends, and insights. Tabular and graphical methods serve as indispensable tools for organizing and visualizing data in a coherent and interpretable manner.

Tabular methods involve structured arrangements of data into rows and columns, facilitating clear and systematic presentation, while graphical methods employ visual representations such as charts, graphs, or diagrams to illustrate patterns and relationships within datasets. Both approaches play pivotal roles in conveying complex information effectively, offering different perspectives for data interpretation and aiding in informed decision-making.

Whether through organized tables or visually appealing graphs, these methods serve as fundamental pillars in transforming raw data into comprehensible and actionable insights.

2. Summarizing Categorical Data



The categorical data can be summarized through any of the following methods:

1. Frequency Distribution
2. Relative Frequency Distribution and Percent Frequency Distribution
3. Bar Charts and Pie Charts

Let us discuss them, one by one.

2. Summarizing Categorical Data

A frequency distribution is a tabular summary of data showing the number (frequency) of items in each of several non-overlapping classes.

Frequency distribution displays how often different values or ranges of values occur in a dataset. It organizes raw data into groups or intervals, known as classes or bins, and then records the number of occurrences or frequencies within each class. This method helps in summarizing and understanding the distribution or pattern of data, especially when dealing with large datasets.

Let us use the following example to demonstrate the construction and interpretation of a frequency distribution for categorical data. Coke, Pepsi, and Sprite are three popular soft drinks.

Data From a sample of 30 Soft Drink Purchases is tabled below.

Coke	Sprite	Sprite	Pepsi	Coke	Sprite
Pepsi	Coke	Pepsi	Coke	Pepsi	Pepsi
Sprite	Pepsi	Coke	Sprite	Sprite	Coke
Pepsi	Coke	Pepsi	Coke	Pepsi	Pepsi
Coke	Pepsi	Sprite	Pepsi	Coke	Sprite

To develop a frequency distribution for these data, we count the number of times each soft drink appears in above table. Coke appears 10 times, Pepsi appears 12 times and Sprite appears 8 times.

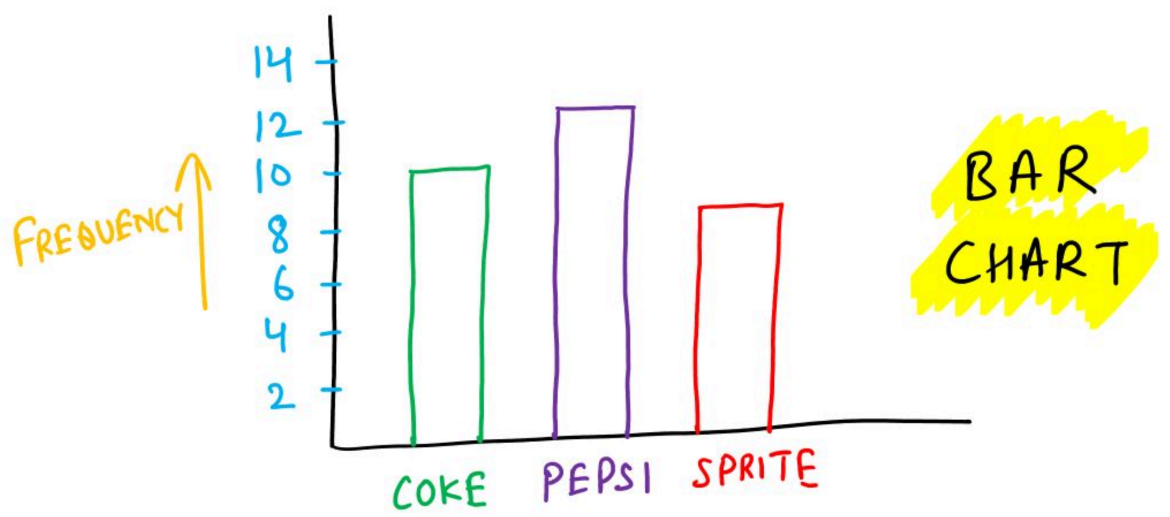
These counts are summarized in the frequency distribution as shown below:

Soft Drink	Frequency
Coke	10
Pepsi	12
Sprite	8
Total	30

Viewing the frequency distribution, we see that Pepsi is the leader, Coke is second and Sprite is third. The frequency distribution summarizes information about the popularity of the three soft drinks.

2. Summarizing Categorical Data

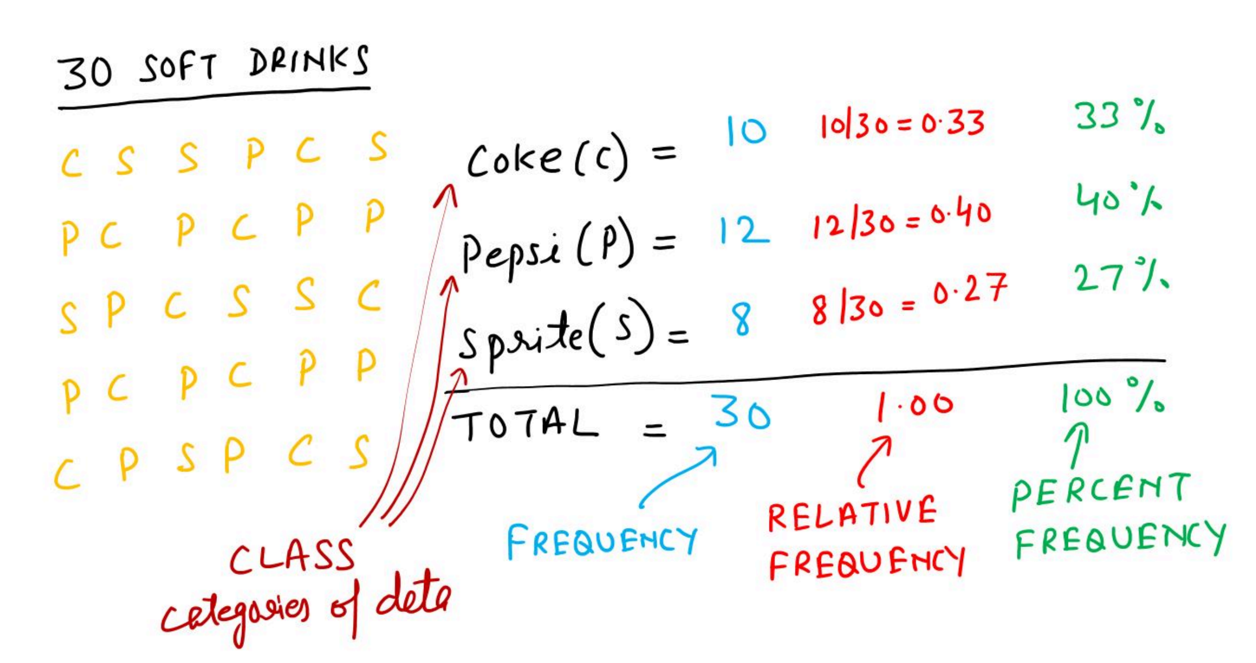
Another widely used qualitative data graphing technique is the **bar graph** or **bar chart**. A bar chart is a graphical device for depicting categorical data summarized in a frequency, relative frequency, or percent frequency distribution. On one axis of the graph (usually the horizontal axis), we specify the labels that are used for the classes (categories). A frequency, relative frequency, or percent frequency scale can be used for the other axis of the chart.



The bar graph is qualitative because the categories are non-numerical, and it may be either horizontal or vertical. A bar graph generally is constructed from the same type of data that is used to produce a pie chart. However, an advantage of using a bar graph over a pie chart for a given set of data is that for categories that are close in value, it is considered easier to see the difference in the bars of bar graph than discriminating between pie slices.

A side-by-side bar chart uses sets of bars to show the joint responses from two categorical variables.

2. Summarizing Categorical Data



The **relative frequency** of a class equals the fraction or proportion of items belonging to a class. For a data set with n observations, the relative frequency of each class can be determined as follows:

$$\text{Relative Frequency} = \frac{\text{Frequency of Class}}{n}$$

The **percent frequency** of a class is the relative frequency multiplied by 100.

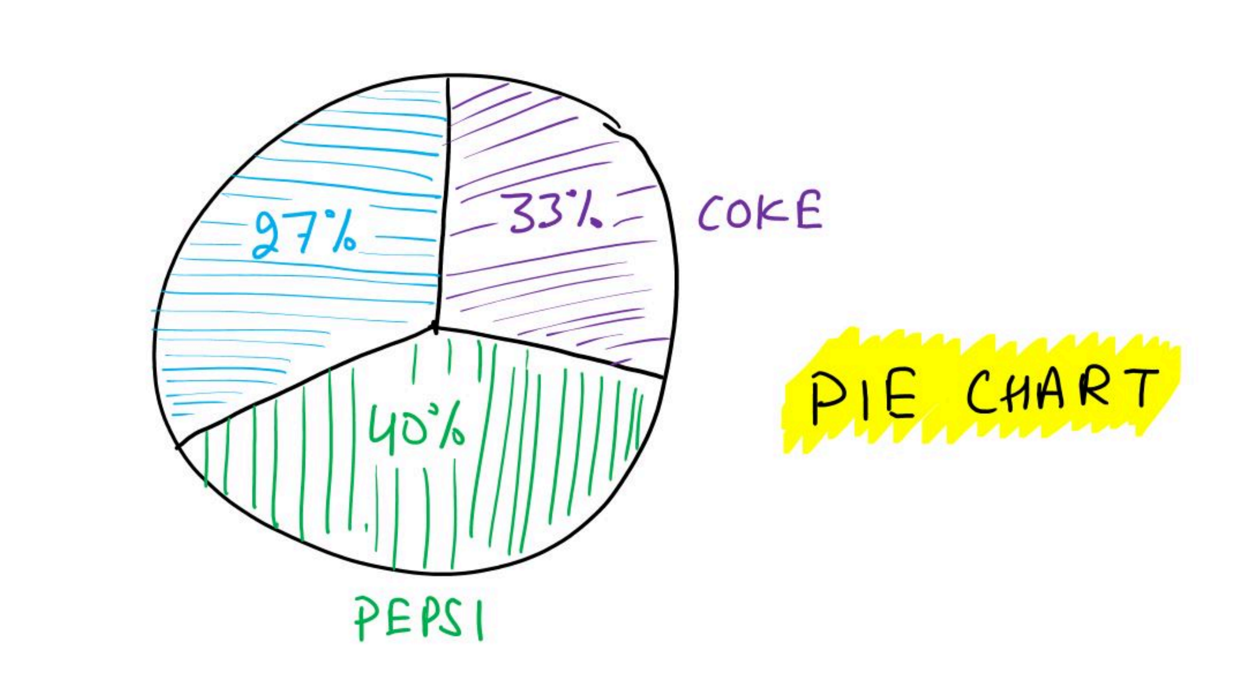
$$\text{Percent Frequency} = \frac{\text{Frequency of Class}}{n} \times 100$$

In the previous example, Relative frequency of coke = $\frac{10}{30}$ = 0.33

And Percent frequency of coke = $\frac{10}{30}$ $\times100$ = 33%

2. Summarizing Categorical Data

A **Pie chart** is a circular depiction of data where the area of the whole pie represents 100% of the data and slices of the pie represent a percentage breakdown of the sublevels. Pie charts show the relative magnitudes of the parts to the whole. They are widely used in business, particularly to depict such things as budget categories, market share, and time/resource allocations. Generally, it is more difficult for the viewer to interpret the relative size of angles in a pie chart than to judge the length of rectangles in a bar chart.



The pie chart provides another graphical device for presenting relative frequency and percent frequency distributions for categorical data. To construct a pie chart, we first draw a circle to represent all the data. Then we use the relative frequencies to subdivide the circle into sectors, or parts, that correspond to the relative frequency for each class.

---

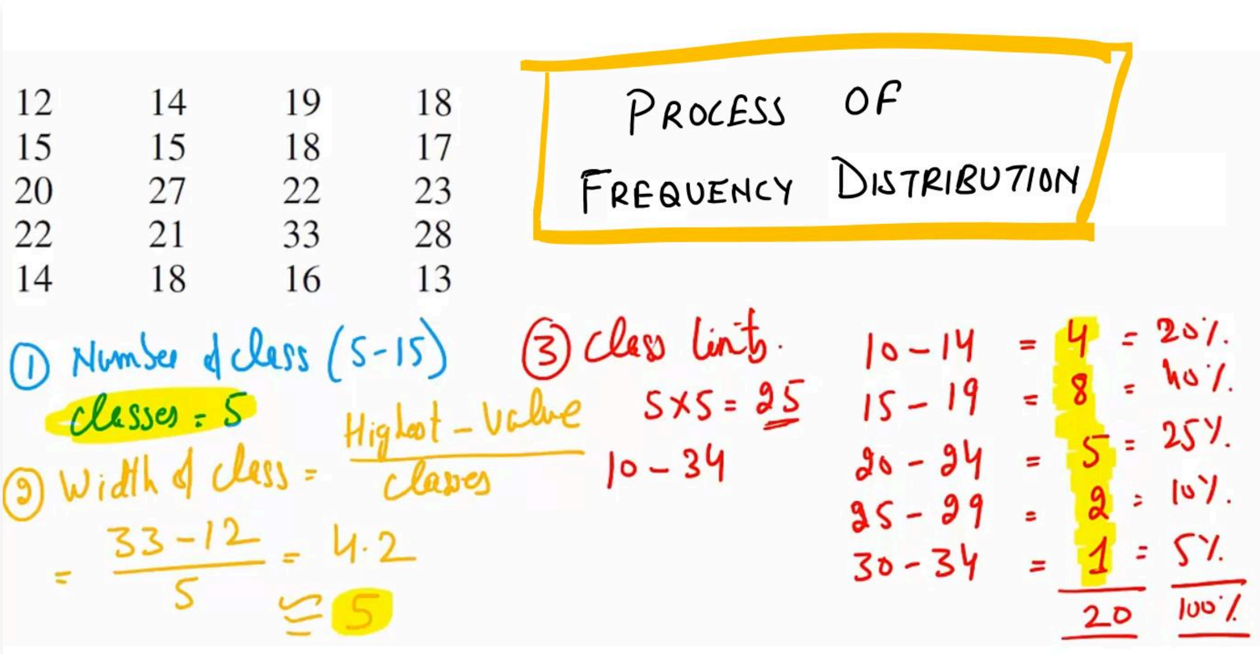
3. Summarizing Quantitative Data

The Quantitative Data can be summarized through any of the following methods:
1. Frequency Distribution
2. Dot Plot
3. Histogram and Frequency Polygon
4. Cumulative Frequency Distributions
5. Ogive
6. Stem-and leaf display
7. Crosstabulations
8. Scatter Diagram

Let us discuss them, one by one.

---

4. Frequency Distribution



As understood earlier, the frequency distribution is a tabular summary of data showing the number (frequency) of items in each of several non-overlapping classes. This definition holds good for quantitative as well as qualitative data.

It is important to understand concept of class here. Consider the quantitative data in table given below. These data show the time in days for 20 software projects undertaken by TCS.

Projects Time (In Days)			
Project 1	Project 2	Project 3	Project 4
12	14	19	18
15	15	18	17
29	27	22	23
22	21	33	28
14	18	16	13

Grouping of Data into Classes

Classes in a frequency distribution refer to the categories or intervals into which data is grouped or divided. They represent the ranges within which the raw data is organized for analysis and presentation. These classes are essential in summarizing and understanding the distribution of data, especially when dealing with large datasets.

Let us go through the steps to divide the data into classes.

Step 1 - Number of Classes

First, we decide the number of classes. For a small number of data items, as few as 5 or 6 classes may be used to summarize the data. For a larger number of data items, a larger number of classes is usually required. The goal is to use enough classes to show the variation in the data, but not so many classes that some contain only a few data items.

In our example, we take n = 5, because we decided to summarize the data with five classes.

Step 2 - Width of Class

The second step in constructing a frequency distribution for quantitative data is to choose a width for the classes. As a general guideline, we recommend that the width will be the same for each class. Thus, the choices of the number of classes and the width of classes are not independent decisions. A larger number of classes means a smaller class width, and vice versa. To determine an approximate class width, we begin by identifying the largest and smallest data values.

$$\text{Class Width} = \frac{\text{Largest Value} - \text{Smallest Value}}{\text{Number of Classes}}$$

In our example, class width of given data = $\frac{33-12}{5} = 4.2$

We decide to round up and use a class width of 5 days.

The width of class is also known as *Class Interval Width*.

Step 3- Class Limits

Class limits must be chosen so that each data item belongs to one and only one class. The lower class limit identifies the smallest possible data value assigned to the class. The upper class limit identifies the largest possible data value assigned to the class. In developing frequency distributions for qualitative data, we did not need to specify class limits because each data item naturally fell into a separate class. But with quantitative data class limits are necessary to determine where each data value belongs.

Using this TCS data, we selected 10 days as the lower class limit and 14 days as the upper class limit for the first class. This class is denoted as 10 –14. The smallest data value, 12, is included in the 10 –14 class. We then selected 15 days as the lower class limit and 19 days as the upper class limit of the next class. We continued defining the lower and upper class limits to obtain a total of five classes: 10–14, 15–19, 20–24, 25–29, and 30–34. The difference between the lower class limits of adjacent classes is the class width. Using the first two lower class limits of 10 and 15, we see that the class width is 15 – 10 = 5.

Let us make a new table to show class widths and class limits:

Time (days)	Frequency
10-14	4
15-19	8
20-24	5
25-29	2
30-34	1
Total	20

Mid Point of Class

In some applications, we want to know the midpoints of the classes in a frequency distribution for quantitative data. The class midpoint is the value halfway between the lower and upper class limits. The mid point for 10-14 class is 12, for 15-19 class is 17, for 20-24 class is 22 and so on. The mid point of a class is also called *Class Mark*.

Frequency Density

Frequency density is calculated as the ratio of the frequency of observations in a class to the class width. It allows for a standardized comparison of frequencies across different class intervals, especially when the intervals vary in width.

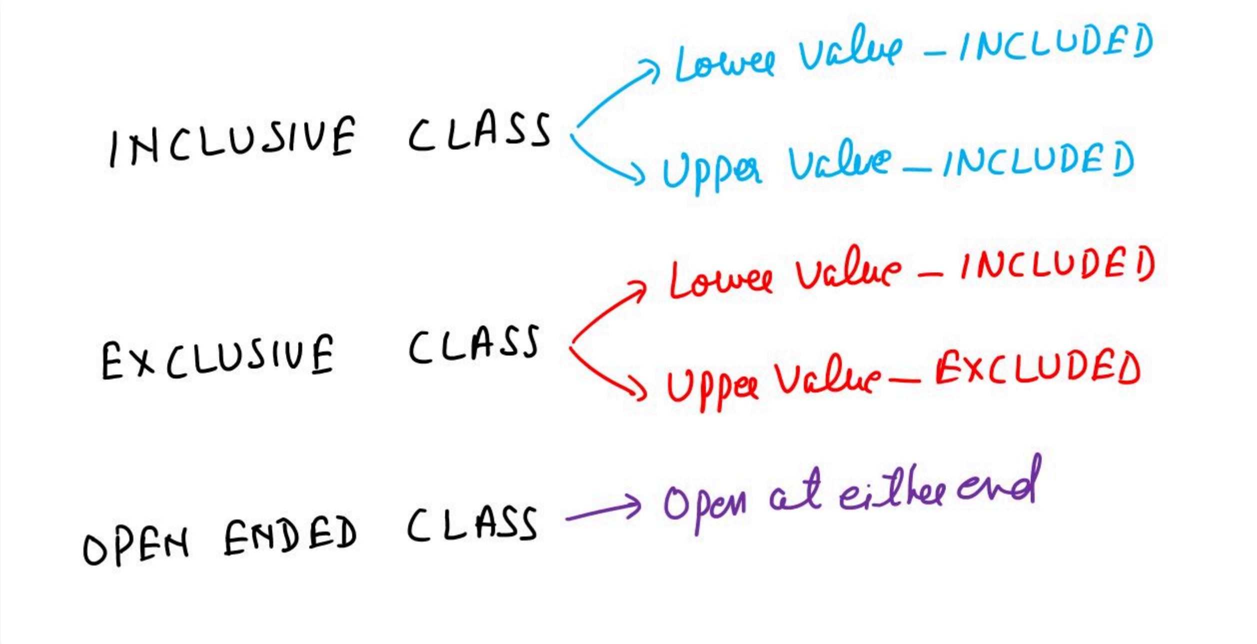
$$\text{Frequency Density} = \frac{\text{Class Frequency}}{\text{Width of Class}}$$

Relative Frequency and Percent Frequency

Similarly, we can calculate Relative Frequency and Percent Frequency Distributions for TCS data, which is shown below:

Time (days)	Frequency	Relative Frequency	Percent Frequency
10-14	4	0.20	20%
15-19	8	0.40	40%
20-24	5	0.25	25%
25-29	2	0.10	10%
30-34	1	0.05	5%
Total	20	1.00	100%

5. Types of Classes



Classes in frequency distributions can be categorized into three types based on their boundaries and how they encompass data values:

1. Inclusive Classes

Inclusive classes include both the lower and upper limits as part of the interval. The values at the endpoints are considered part of the class.

Consider a dataset of ages grouped into intervals. An inclusive class might be defined as "10-20," where both the ages 10 and 20 are included in the interval, meaning any value from 10 to 20 (inclusive) falls within this class.

2. Exclusive Classes

Exclusive classes include the lower limit but exclude the upper limit from the interval. The upper value is part of the next class.

Continuing with the age dataset, an exclusive class could be "10-20." Here, the lower age, 10, is included in the class, but the upper age, 20, is excluded, implying that values between 10 and up to, but not including 20, belong to this class. The value of 20 or 20.1 will be part of class "20-30".

3. Open-ended Classes

Open-ended classes have one limit undefined, typically used for the lowest or highest interval, where the value extends indefinitely.

In the age dataset, an open-ended class might be "60 and above." Here, 60 is the lower limit, indicating that the class includes all ages equal to or greater than 60, without specifying an upper limit.

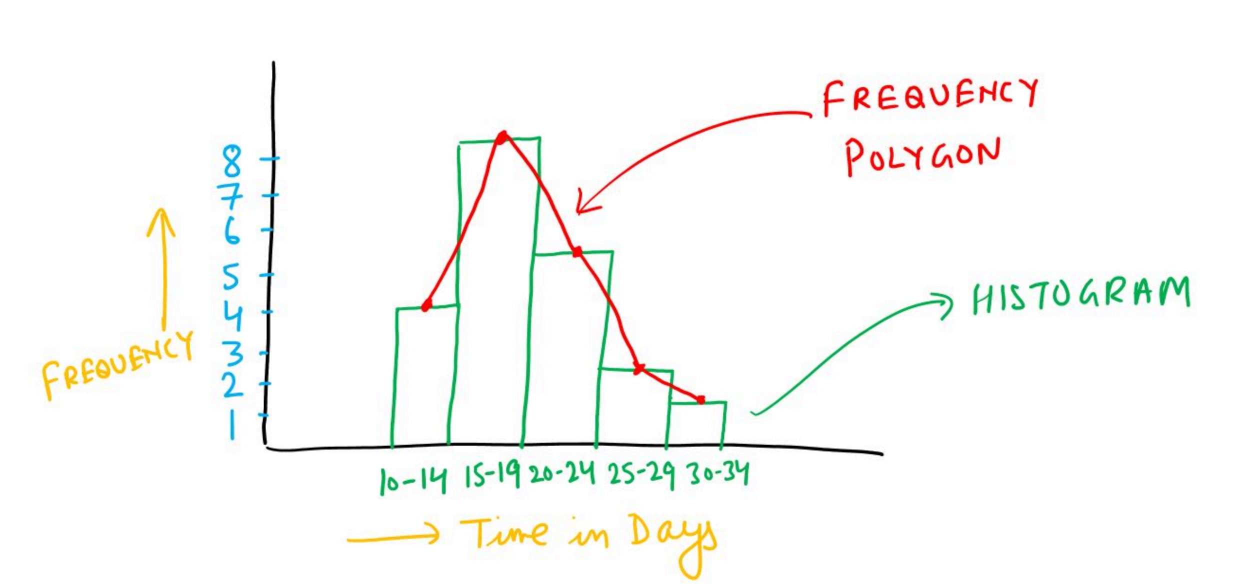
6. Histogram and Frequency Polygon

Till now, we have discussed the tabular representation of data. Let us now discuss Histogram, which is graphical representation of data.

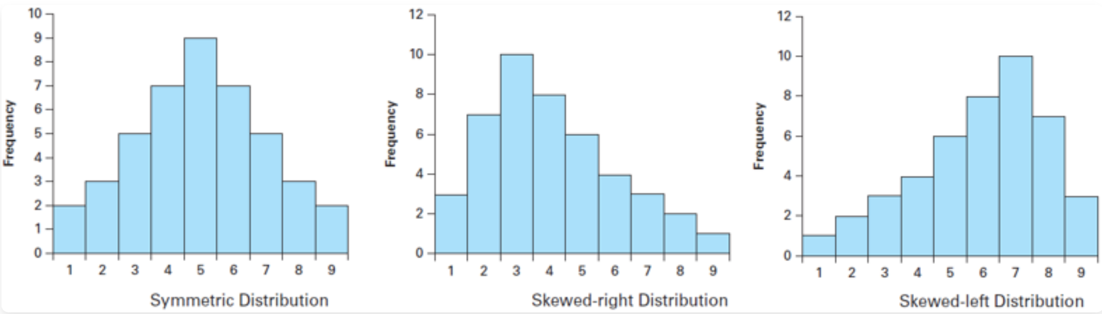
Histogram can be prepared for data previously summarized in either a frequency, relative frequency, or percent frequency distribution.

A histogram is constructed by placing the variable of interest on the horizontal axis and the frequency, relative frequency, or percent frequency on the vertical axis. The frequency, relative frequency, or percent frequency of each class is shown by drawing a rectangle whose base is determined by the class limits on the horizontal axis and whose height is the corresponding frequency, relative frequency, or percent frequency.

The histogram for TCS data is shown here.



We can describe graphically the shape of the distribution by a histogram. That is, we can visually determine whether data are evenly spread from its middle or center. Sometimes the center of the data divides a graph of the distribution into two “mirror images,” so that the portion on one side of the middle is nearly identical to the portion on the other side. Graphs that have this shape are symmetric; those without this shape are asymmetric or skewed.



The shape of a distribution is said to be **symmetric** if the observations are balanced, or approximately evenly distributed, about its center.

A distribution is skewed, or asymmetric, if the observations are not symmetrically distributed on either side of the center. A **skewed-right distribution** (sometimes called positively skewed) has a tail that extends farther to the right. A **skewed-left distribution** (sometimes called negatively skewed) has a tail that extends farther to the left.

Frequency Polygon

A **frequency polygon**, like the histogram, is a graphical display of class frequencies. However, instead of using bars or rectangles like a histogram, in a frequency polygon each class frequency is plotted as a dot at the class midpoint, and the dots are connected by a series of line segments. Construction of a frequency polygon begins by scaling class midpoints along the horizontal axis and the frequency scale along the vertical axis. A dot is plotted for the associated frequency value at each class midpoint. Connecting these midpoint dots completes the graph.

The frequency polygon is also called *Percent Polygon*.

---

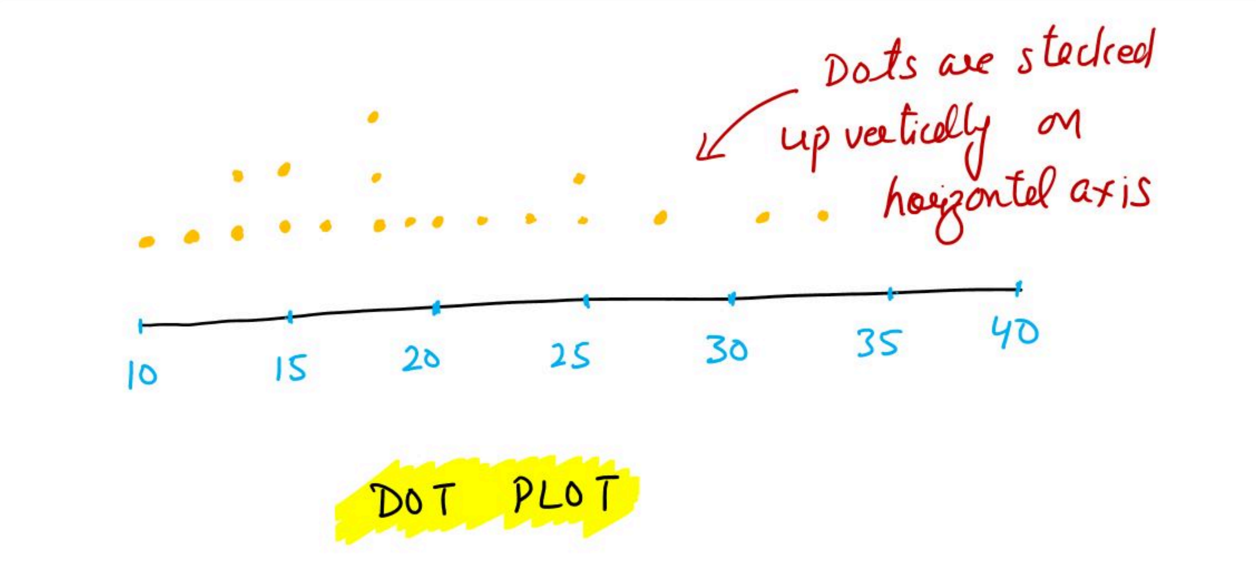
7. Dot Plot

One of the simplest graphical summaries of data is a **dot plot**.

A relatively simple statistical chart that is generally used to display continuous, quantitative data is the dot plot. In a dot plot, each data value is plotted along the horizontal axis and is represented on the chart by a dot. If multiple data points have the same values, the dots will stack up vertically. If there are a large number of close points, it may not be possible to display all of the data values along the horizontal axis.

Dot plots can be especially useful for observing the overall shape of the distribution of data points along with identifying data values or intervals for which there are groupings and gaps in the data.

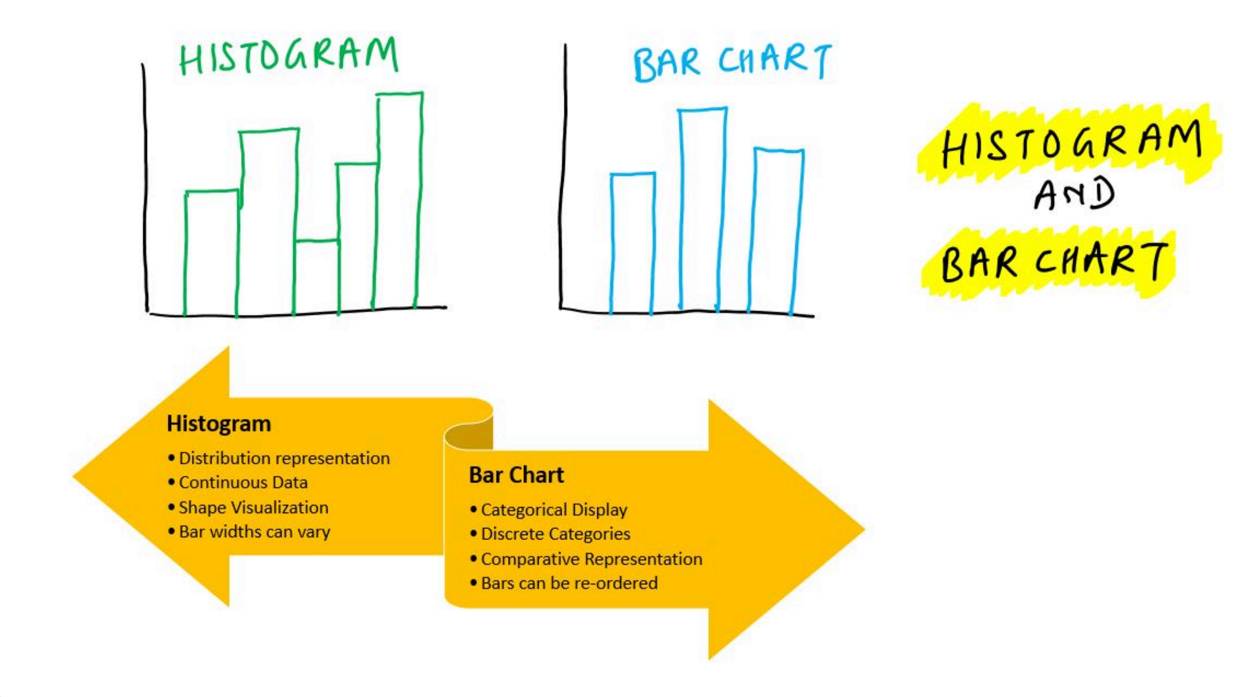
Dot plot for TCS Data in previous example is shown below.



---

8. Bar Chart and Histogram

Histograms and bar charts are both tools used to represent frequency distributions, but they differ in their suitability based on the type of data being depicted.



Bar Chart

Data Type: Primarily used for categorical (qualitative) data, showcasing distinct categories.
Appropriate Usage: Ideal for displaying non-continuous data and discrete categories. Bars can be rearranged based on preference.
Examples: Displaying revenue generated by different products or the percentage of people preferring specific coffee brands.

Histogram

Data Type: Suitable for numerical (quantitative) data, especially when dealing with continuous variables.
Appropriate Usage: Perfect for representing continuous data and providing insight into the shape of distribution (like skewness). Bars in a histogram represent ranges of values rather than distinct categories.
Examples: Representing the heights of students in a class, weights of products, scores of exam takers, or response times at a ticket counter.

Key Differences

Data Type: Bar charts are for categorical data, while histograms are for numerical data.
Data Representation: Histograms show continuous data distribution with bars representing ranges, whereas bar charts depict individual categories.
Bar Ordering: Bars in a bar chart can be rearranged, but a histogram's bars are fixed in their order, as they represent ranges of values.
Class Width: Histograms can have unequal class widths to represent different ranges, while bar charts for categorical data maintain uniform bar widths.

In summary, while both histograms and bar charts serve the purpose of depicting frequency distributions, the choice between them depends on the nature of the data, whether it's categorical or numerical, and whether the focus is on the distribution's shape and characteristics (suitable for histograms) or the discrete categories or proportions (suitable for bar charts).

9. Cumulative Frequency Distributions

The cumulative frequency distribution uses the number of classes, class widths, and class limits developed for the frequency distribution. However, rather than showing the frequency of each class, the cumulative frequency distribution shows the number of data items with values less than or equal to the upper class limit of each class.

To understand how the cumulative frequencies are determined, consider the class with the description “less than or equal to 24.” The cumulative frequency for this class is simply the sum of the frequencies for all classes with data values less than or equal to 24.

Cumulative frequency distribution for TCS Data is shown below:

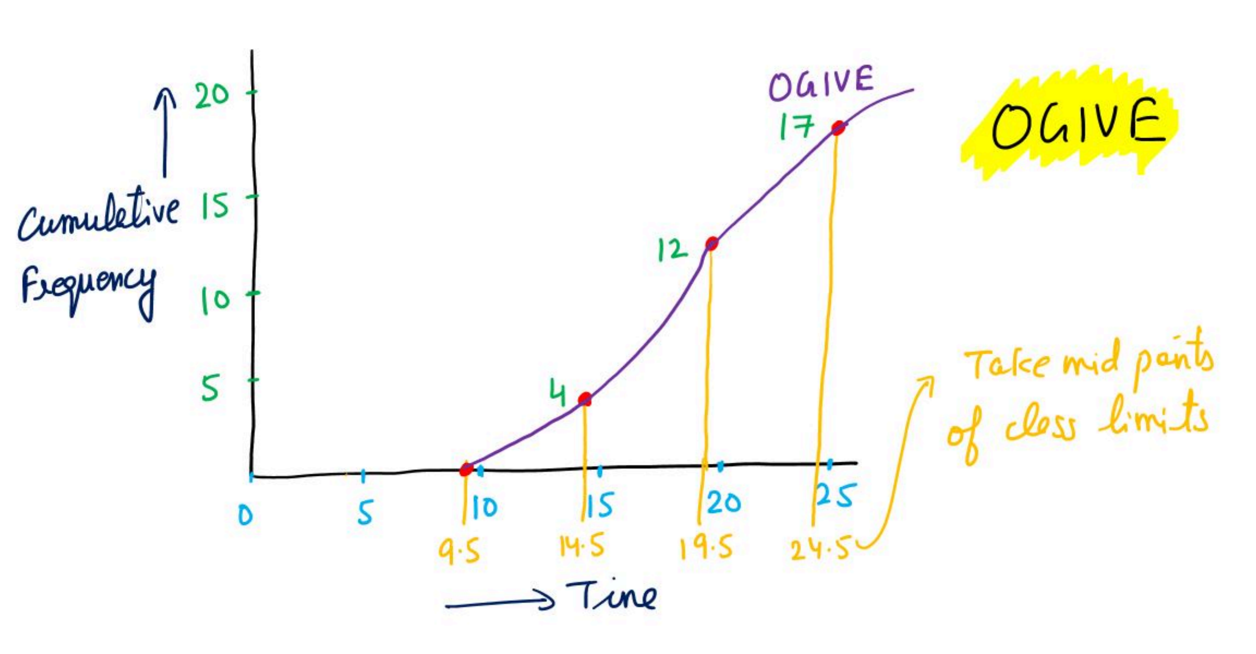
Time (days)	Cumulative Frequency	Cumulative Relative Frequency	Cumulative Percent Frequency
Less than or equal to 14	4	0.20	20%
Less than or equal to 19	12	0.60	60%
Less than or equal to 24	17	0.85	85%
Less than or equal to 29	19	0.90	95%
Less than or equal to 34	20	1.00	100%

From this Cumulative Frequency analysis of TCS data, we can conclude that the TCS finished 60% of projects in less than or equal to 19 days.

10. Ogive

A graph of a cumulative frequency distribution, called an ogive, shows data values on the horizontal axis and either the cumulative frequencies, the cumulative relative frequencies, or the cumulative percent frequencies on the vertical axis. The ogive is constructed by plotting a point corresponding to the cumulative frequency of each class.

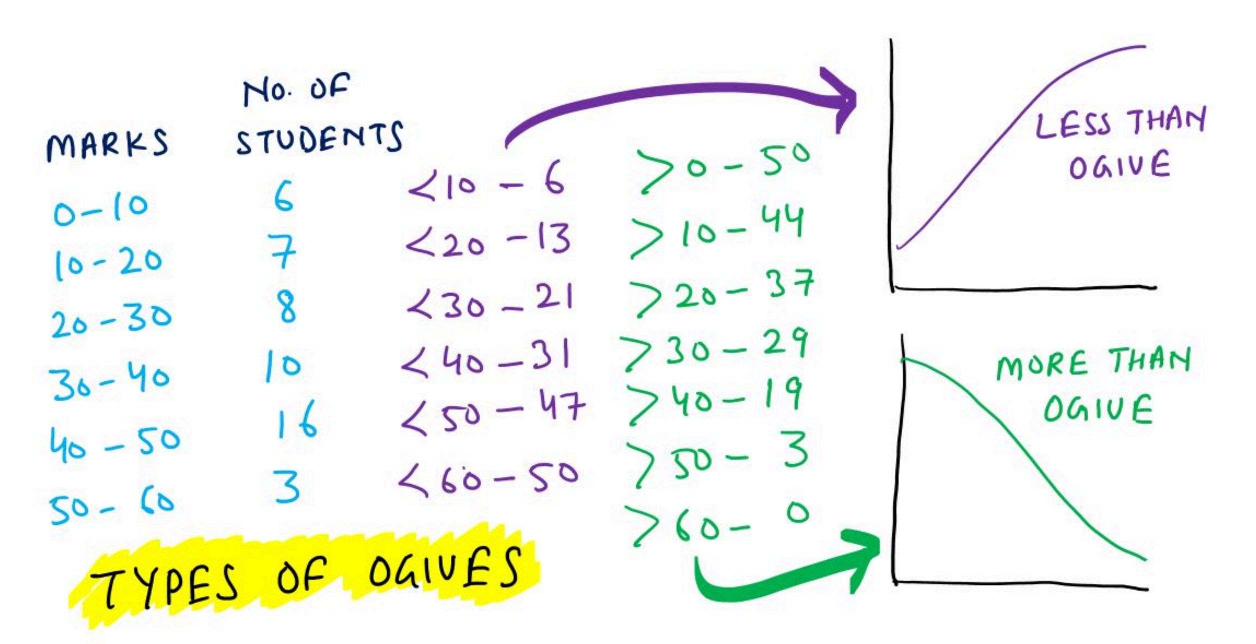
The Ogive is also known as Cumulative Frequency Polygon or Cumulative Percent Polygon.



Cumulative Frequency Ogive for the TCS data is given in the figure. Similarly, we can draw ogive for Cumulative Relative Frequency or Cumulative Percent Frequency.

10. Ogive

Ogivesshow the cumulative frequencies of data in a distribution. They come in two main types.



Less Than Ogive

This type of ogive plots cumulative frequencies **less than the upper boundary of each class interval**. It starts from the left-hand side of the graph and increases gradually as each data point's frequency is added cumulatively.

It helps visualize the number of data points that fall below a certain value or interval, providing insights into the distribution's lower end.

Suppose you want to understand how many students scored below a certain mark. A less than ogive would be useful here. For instance, if you're interested in knowing how many students scored below 80%, the less than ogive would help visualize the cumulative count of students who achieved less than that threshold.

Similarly, when analyzing income distribution in a population, a less than ogive could show how many individuals earn below a particular income level. For example, it could illustrate the cumulative number of people earning less than Rs 50,000 monthly.

More Than Ogive

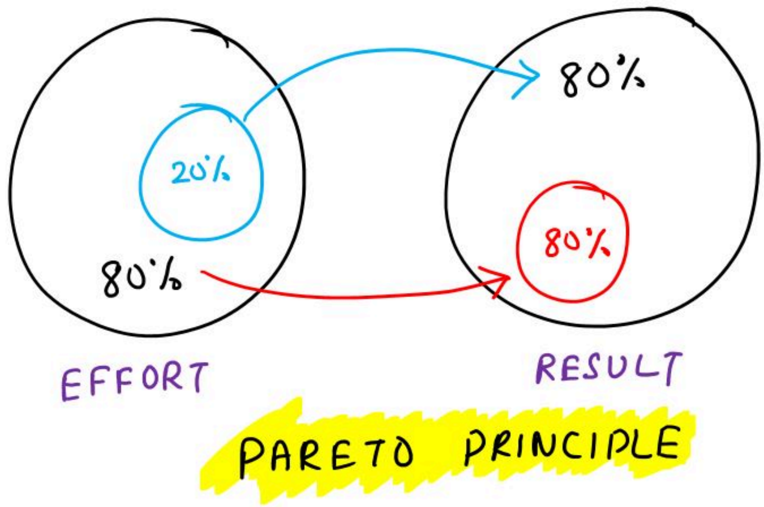
Contrary to the Less Than Ogive, this type represents cumulative frequencies **greater than the upper boundary of each class interval**. It begins from the right-hand side of the graph and ascends gradually as cumulative frequencies are added, showing the count of data points above certain values or intervals.

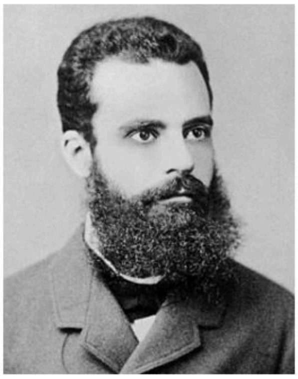
It offers a view of the data points exceeding specific values or intervals, revealing information about the distribution's upper end.

In a service industry, if you want to understand how many customers waited more than a certain duration, a more than ogive would be beneficial. For instance, it could depict the cumulative count of customers who waited more than 10 minutes in a queue.

Similarly, analyzing delivery times for packages, a more than ogive might display the cumulative count of deliveries that took longer than a specified time frame, like more than 3 days for delivery.

11. Pareto Chart




Vilfredo Pareto

A Pareto chart is used to identify and prioritize the most significant factors contributing to a problem or a situation.

It is based on the Pareto Principle, also known as the 80/20 rule, which suggests that roughly 80% of effects come from 20% of the causes. It combines both bar and line graphs to represent data and their cumulative impact.

Pareto charts were named after an Italian economist, Vilfredo Pareto, who observed more than 100 years ago that most of Italy's wealth was controlled by a few families who were the major drivers behind the Italian economy.

Pareto charts are versatile and can be applied across various fields to prioritize and address critical issues. Here are few examples:

1. Quality Control in Manufacturing
Use Pareto charts to identify the most common types of defects in a manufacturing process. This helps focus efforts on fixing the most frequent issues that contribute to a significant portion of product defects.

2. Customer Complaint Analysis
Analyze customer complaints in a service-oriented business. Determine the most frequent complaints to prioritize improvements that will have the most substantial impact on customer satisfaction.

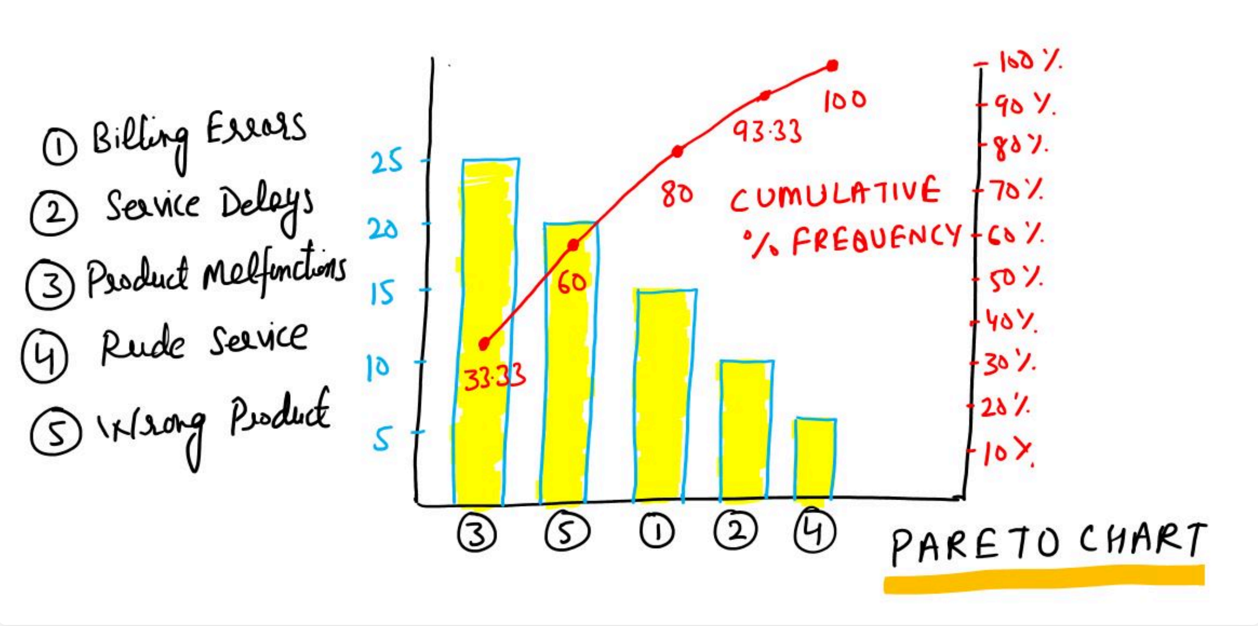
3. Inventory Management
Utilize Pareto analysis to identify the most commonly stocked items that contribute to the majority of sales. This allows for efficient inventory management by focusing on the vital few items instead of the less impactful ones.

4. Healthcare Services
Apply Pareto charts in healthcare to identify the most common medical errors or issues causing patient dissatisfaction. Prioritize addressing these issues to improve overall patient care and safety.

5. Project Management
In project management, use Pareto charts to identify the most common reasons for project delays or issues. This enables project managers to address key factors that significantly impact project timelines and success.

11. Pareto Chart

We will use example of analyzing customer complaints in a service center over a month.



1. Data Collection
Gather data on different categories or factors related to a specific issue or problem.

We will gather data on various types of complaints received:
Billing Errors
Service Delays
Product Malfunctions
Rude Service
Wrong Product Sent

2. Frequency Determination
Count the occurrences or frequency of each category.

Count the occurrences of each type of complaint:
Billing Errors: 15
Service Delays: 10
Product Malfunctions: 25
Rude Service: 5
Wrong Product Sent: 20

3. Ordering
Arrange categories from the most frequent to the least frequent.

Arrange the categories from most frequent to least frequent:
Product Malfunctions (25)
Wrong Product Sent (20)
Billing Errors (15)
Service Delays (10)
Rude Service (5)

4. Bar Chart
Construct a bar chart where the bars represent the frequency of each category in descending order. The bar chart is shown in the figure.

5. Line Graph
Plot a line graph showing the cumulative percentage total on the secondary axis.

Calculate the percentage for each category:
Product Malfunctions: 25 / Total Complaints (75) = 33.33%
Wrong Product Sent: 20 / Total Complaints (75) = 26.67%
Billing Errors: 15 / Total Complaints (75) = 20%
Service Delays: 10 / Total Complaints (75) = 13.33%
Rude Service: 5 / Total Complaints (75) = 6.67%

The line graph of the cumulative percentage is shown in the figure.

Conclusion

This Pareto chart visually demonstrates that addressing Product Malfunctions and Wrong Products Sent could potentially resolve over 60% of the complaints, making them the most critical areas for improvement.

---

12. Exploratory Data Analysis

The techniques of exploratory data analysis consist of simple arithmetic and easy-to-draw graphs that can be used to summarize data quickly.

Let us look at the method of Stem and leaf display.

---

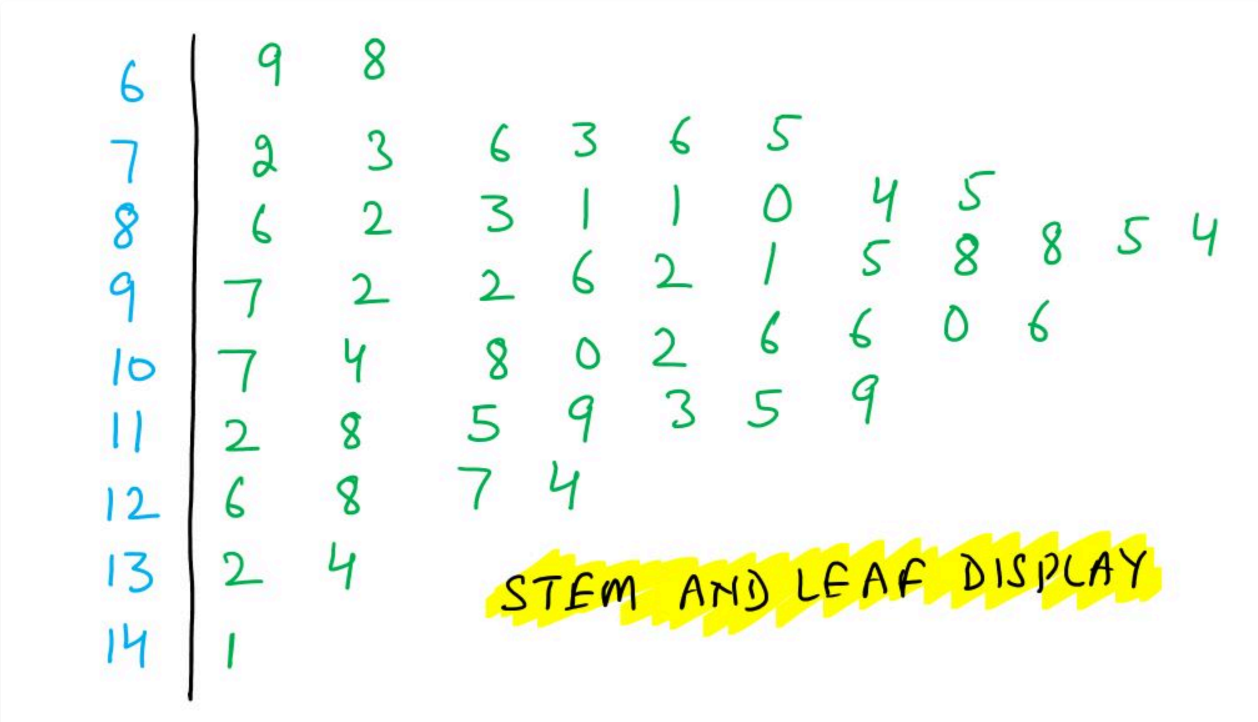
12. Exploratory Data Analysis

Stem-and leaf display—can be used to show both the rank order and shape of a data set simultaneously.

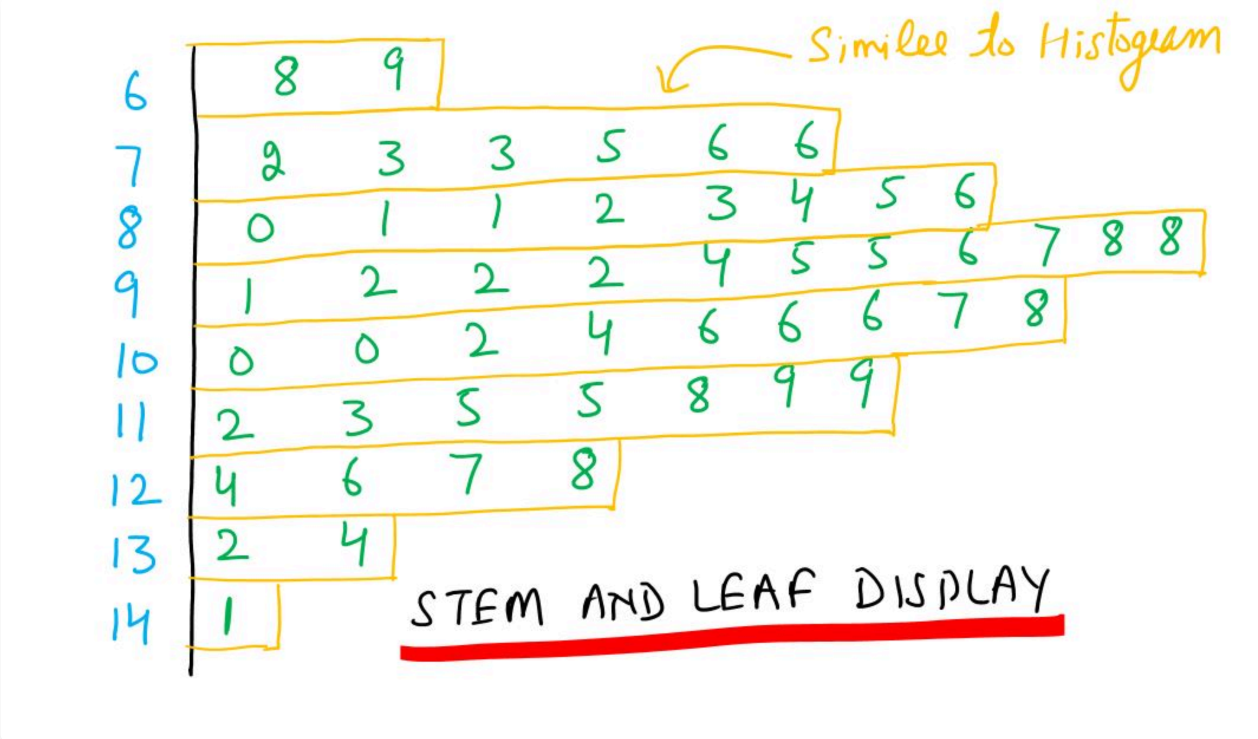
To understand the Exploratory Data Analysis lets us take data on number of questions answered correctly in an aptitude test. These data result from a 150-question aptitude test given to 50 job aspirants, recently interviewed for a position at Honda manufacturing plant. The data indicate the number of questions answered correctly.

112	72	69	97	107
73	92	76	86	73
126	128	118	127	124
82	104	132	134	83
92	108	96	100	92
115	76	91	102	81
95	141	81	80	106
84	119	113	98	75
68	98	115	106	95
100	85	94	106	119

To develop a stem-and-leaf display, we first arrange the leading digits of each data value to the left of a vertical line. To the right of the vertical line, we record the last digit for each data value. Continuing to place the last digit of each data value on the line corresponding to its leading digit(s) provides the following:



With this organization of the data, sorting the digits on each line into rank order is simple. Doing so provides the stem-and-leaf display shown here.



The numbers to the left of the vertical line (6, 7, 8, 9, 10, 11, 12, 13, and 14) form the **stem**, and each digit to the right of the vertical line is a **leaf**. It is shown in the left figure.

To focus on the shape indicated by the stem-and-leaf display, let us use a rectangle to contain the leaves of each stem. Doing so, we obtain figure, shown on the right.

Rotating this page counterclockwise onto its side provides a picture of the data that is similar to a histogram with classes of 60–69, 70–79, 80–89, and so on.

Although the stem-and-leaf display may appear to offer the same information as a histogram, it has two primary advantages.
1. The stem-and-leaf display is easier to construct by hand.
2. Within a class interval, the stem-and-leaf display provides more information than the histogram because the stem-and-leaf shows the actual data.

13. Crosstabulations

Crosstabulation is used to summarize data in a way that reveals the relationship between two variables. A **crosstabulation** is a tabular summary of data for two variables. It is also called Contingency Table.

Let us illustrate the use of a crosstabulation by considering the following application based on data from Zomatos's review of 300 restaurants in Mumbai. The quality rating and the meal price data were collected for a sample of 300 restaurants.

Quality rating is a categorical variable with rating categories of good, very good, and excellent. Meal price is a quantitative variable that ranges from Rs 10 to Rs 49.

A crosstabulation of the data for meal price is shown in the table.

	Meal Price			
Quality Rating	**Rs. 10-19**	**Rs. 20-29**	**Rs. 30-39**	**Rs. 40-49**
Good	42	40	2	0
Very Good	34	64	46	6
Excellent	2	14	28	22
Total	78	118	76	28

Above example of cross tabulation shows frequency distribution. Similarly, we can also show Relative Frequency or Percent Frequency.

13. Crosstabulations

The data in two or more crosstabulations are often combined or aggregated to produce a summary crosstabulation showing how two variables are related. In such cases, we must be careful in drawing a conclusion because a conclusion based upon aggregate data can be reversed if we look at the unaggregated data. The reversal of conclusions based on aggregated and unaggregated data is called **Simpson's paradox.**

To provide an illustration of Simpson's paradox we consider an example involving the review of two restaurants in Patna. Lakshmi restaurant and Khana Khazana both serve Chinese food and Mughlai food.

For each restaurant a crosstabulation was developed based upon two variables: Review (Good or Poor) and Type of Food (Chinese and Mughlai). Suppose that the two crosstabulations were then combined by aggregating the type of review data.

The resulting aggregated crosstabulation contains two variables: Review (Good or Poor) and Restaurant (Lakshmi or Khana Khazana).

This crosstabulation shows the number of reviews in which the food was Good or Poor for both restaurants.

The following crosstabulation shows these results along with the column percentages in parentheses next to each value:

Review	Lakshmi	Khana Khazana	Total
Good	129 (86%)	110 (88%)	239
Poor	21 (14%)	15 (12%)	36
Total	150 (100%)	125 (100%)	275

A review of the column percentages shows that 86% of the reviews were Good for Lakshmi restaurant, while 88% of the reviews were good for restaurant Khana Khazana. From this aggregated crosstabulation, we conclude that Khana Khazana has good reviews.

The following unaggregated crosstabulations show the specialties of Lakshmi and Khana Khazana restaurants.

Restaurant Laxmi Review

Review	Chinese	Mughlai	Total
Good	29 (91%)	100 (85%)	129
Poor	3 (9%)	18 (15%)	21
Total	32 (100%)	118 (100%)	150

Restaurant Khana Khazana Review

Review	Chinese	Mughlai	Total
Good	90 (90%)	20 (80%)	110
Poor	10 (10%)	5 (20%)	15
Total	100 (100%)	25 (100%)	125

From the crosstabulation and column percentages for restaurant Lakshmi, we see that the review for Chinese were good in 91% cases and for reviews for Mughlai were good in 85% cases. While for Restaurant Khana Khazana, we see that the 90% reviews were good for Chinese, and 80% reviews were good for Mughlai. Thus, when we unaggregate the data, we see that restaurant Lakshi has better reviews. This result contradicts the conclusion we reached with the aggregated data crosstabulation that showed restaurant Khana Khazana has better reviews. This reversal of conclusions based on aggregated and unaggregated data illustrates **Simpson's paradox**.

14. Scatter Diagram and Trendline

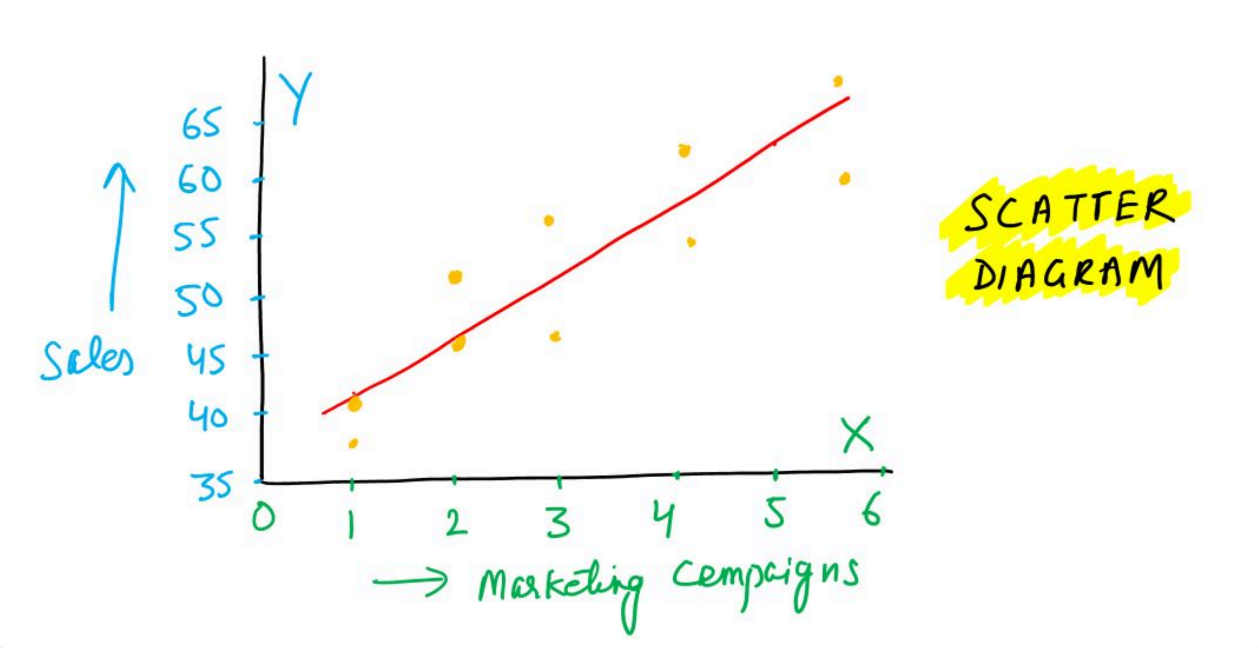
A **scatter diagram** is a graphical presentation of the relationship between two quantitative variables, and a trendline is a line that provides an approximation of the relationship.

As an illustration, consider the relationship between marketing campaigns and sales. During past 10 months, the company launched many marketing campaigns to promote sales. The managers want to investigate whether a relationship exists between the number of marketing campaigns and sales.

The data on the number of commercials shown and sales in different months is shown in the table.

Months	Number of Commercial (X)	Sales (in Lakhs) (Y)
1	2	50
2	5	57
3	1	41
4	3	54
5	4	54
6	1	38
7	5	63
8	3	48
9	4	59
10	2	46

Scatter diagram and trendline for above data is shown below:



The scatter diagram in the figure indicates a positive relationship between the number of marketing campaigns and sales. Higher sales are associated with a higher number of marketing campaigns. The relationship is not perfect in that all points are not on a straight line. However, the general pattern of the points and the trendline suggest that the overall relationship is positive.

14. Scatter Diagram and Trendline

Scatter diagrams, also known as scatter plots, are graphs that display the relationship between two variables. They can exhibit various patterns, indicating different types of relationships between the variables being studied.

Here are three types of scatter diagrams and the relationships they suggest:

1. Positive Relationship
In a positive relationship, as one variable increases, the other variable also tends to increase. The points on the graph tend to form an upward-sloping pattern from left to right.

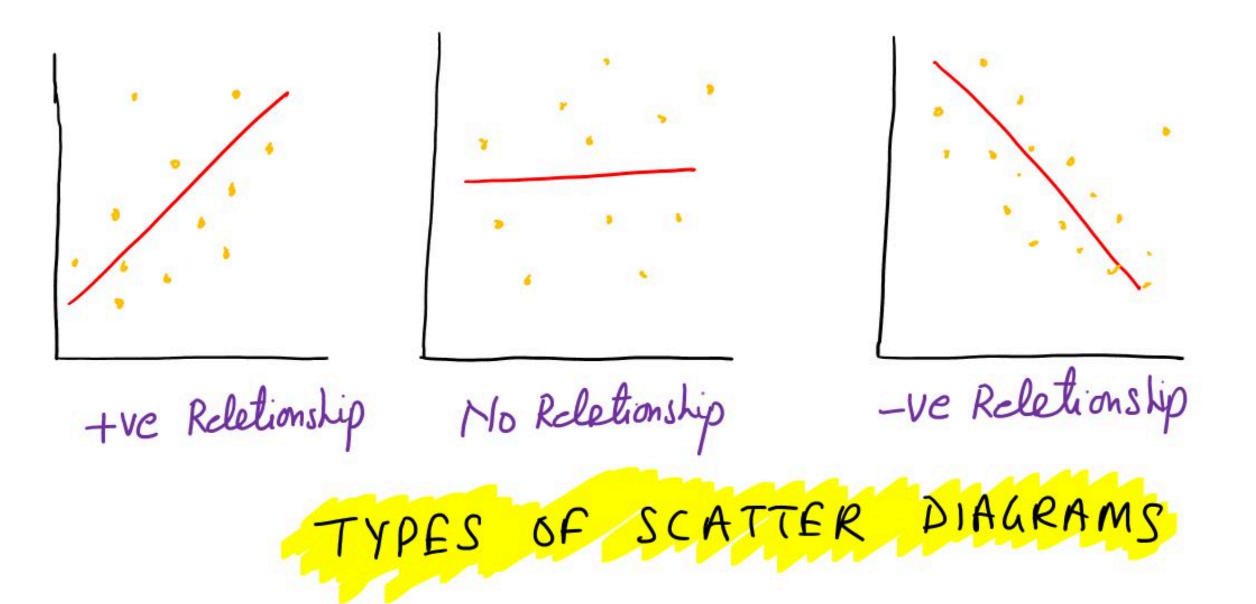
For example, as the number of hours studied increases, the exam scores also tend to increase. Each increase in one variable is associated with an increase in the other.

2. No Apparent Relationship
When there's no discernible pattern or trend between the variables, they show no apparent relationship. Points on the graph are scattered randomly with no observable trend, forming a shapeless cloud or cluster.

For instance, when comparing shoe size and IQ scores, there might not be a significant correlation or connection between the two variables.

3. Negative Relationship
In a negative relationship, as one variable increases, the other variable tends to decrease. The points on the graph exhibit a downward-sloping pattern from left to right.

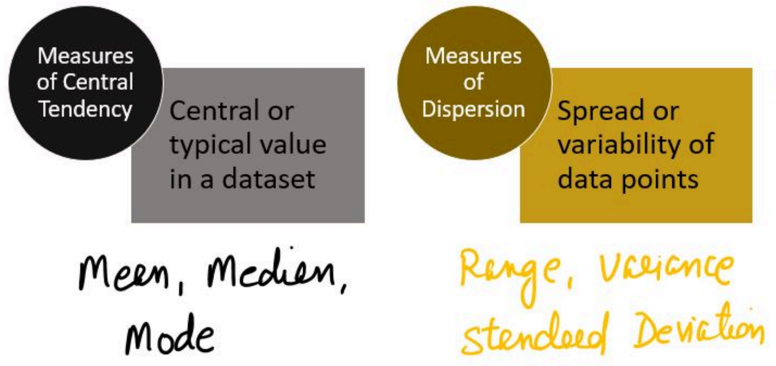
For example, as the amount of rainfall decreases, the instances of flooding tend to increase. Each increase in one variable corresponds to a decrease in the other.



The left most diagram depicts a positive relationship. The middle diagram shows no apparent relationship between the variables. The right most diagram depicts a negative relationship where y tends to decrease as x increases.

1. Introduction

Descriptive statistics is the term given to the analysis of data that helps describe, show or summarize data in a meaningful way.

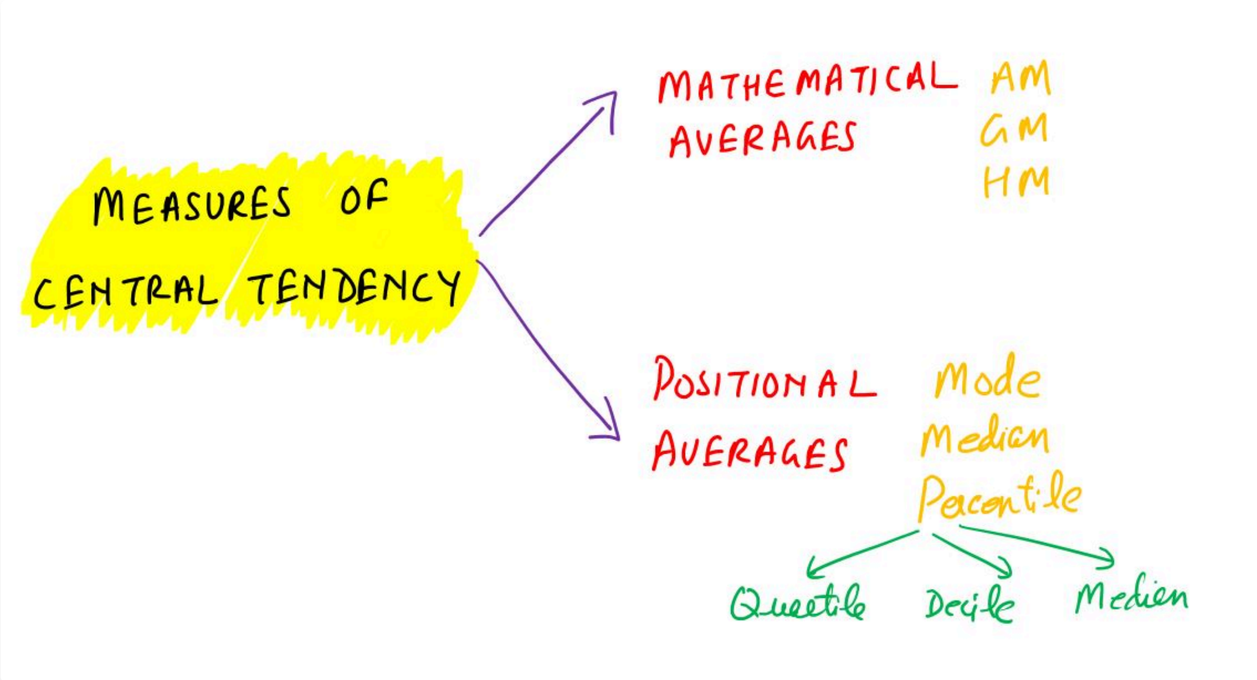


Typically, there are following types of statistic that are used to describe data:

- Measures of Central Tendency
- Measures of Dispersion
- Measures of Skewness

---

2. Measures of Central Tendency



Measures of central tendency are statistical benchmarks that succinctly convey the overall importance of data, serving as a single indicator pinpointing the central position within a distribution. In essence, they represent the focal point or location of the dataset's concentration.

The main objectives of Measure of Central Tendency are:
1) To condense data in a single value.
2) To facilitate comparisons between data.

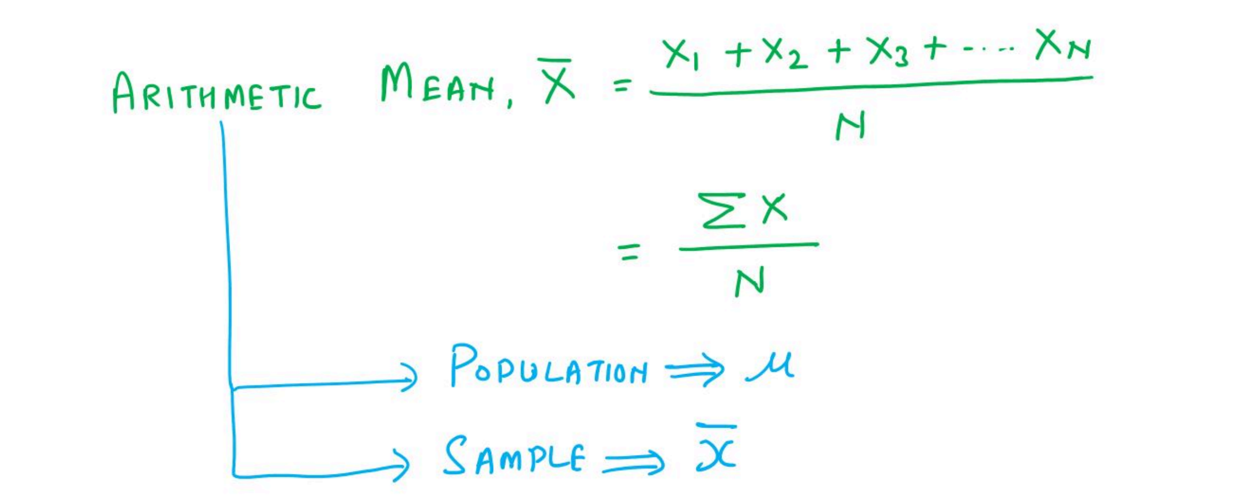
The following are the measures of average or central tendency that are in common use:
(i) Arithmetic Average or Arithmetic Mean or Simple Mean
(ii) Median
(iii) Mode
(iv) Geometric mean and Harmonic mean
(v) Quartiles, Deciles and Percentile

Arithmetic mean, Geometric mean and Harmonic means are called **Mathematical averages** (capable of algebraic treatment) while Mode, Median, Quartiles, Deciles and Percentile are called **Positional averages**.

Let us discuss them one by one.

3. Arithmetic Mean

Arithmetic Mean (AM) is the most commonly used measure of central tendency. It is defined as the sum of the values of all observations divided by the number of observations and is usually denoted by X. In general, if there are N observations as $X_{1,}$ $X_{2,}$ $X_{3,}$ ...., X_N



The Arithmetic Mean and Number of observations for a **Population** is represented by μ and N, respectively. The Arithmetic Mean and Number of observations for a **Sample** is represented by $\bar{x}$ and n, respectively.

There are three methods to find the arithmetic mean, which we will understand by using Income data of 10 families:

Families	**Income**
A	850
B	700
C	100
D	750
E	5000
F	80
G	420
H	2500
I	400
J	360
Total	11160

Let us understand these 3 methods one by one.

3. Arithmetic Mean

In direct method, the AM is calculated by direct formula, given below:

DIRECT METHOD

$$\overline{X} = \frac{X_1 + X_2 + X_3 - - X_N}{N} = \frac{\sum X}{N}$$

Illustration

Family	Income (X)
A	850
B	700
C	100
D	750
E	5000
F	80
G	420
H	2500
I	400
J	360

11160

DIRECT METHOD

$$\sum X = 11160$$

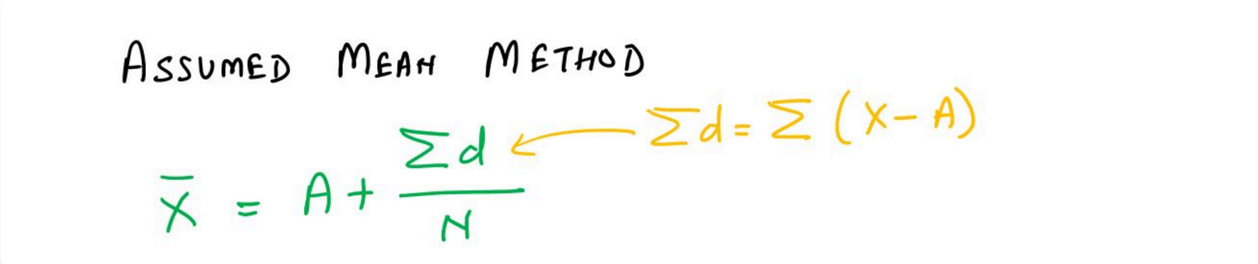
$$\overline{X} = \frac{\sum X}{n} = \frac{11160}{10} = \boxed{1116}$$

3. Arithmetic Mean

If the number of observations in the data is more, it is difficult to compute arithmetic mean by direct method. The computation can be made easier by using assumed mean method.

Here you assume a particular figure in the data as the arithmetic mean, on the basis of logic/experience. Then you may take deviations of the said assumed mean from each of the observation. You can, then, take the summation of these deviations and divide it by the number of observations in the data.

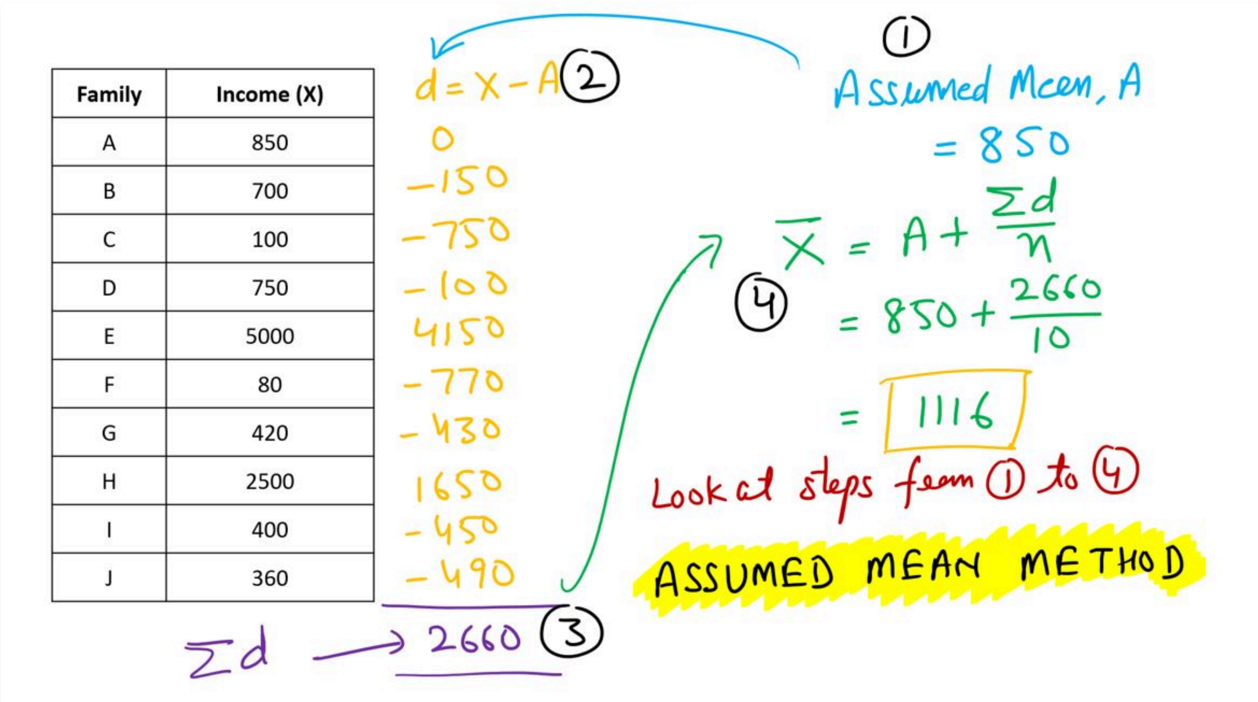
The actual arithmetic mean is estimated by taking the sum of the assumed mean and the ratio of sum of deviations to number of observations.
Let, A = assumed mean
X = individual observations
N = total numbers of observations
d = deviation of assumed mean from individual observation, i.e. d = X – A



Then sum of all deviations is taken as Σd = Σ(X-A)

Then we find $\frac{ \sum{d} }{N}$ and add A to it to get $\bar{X}$

Illustration



Family	Income (X)
A	850
B	700
C	100
D	750
E	5000
F	80
G	420
H	2500
I	400
J	360

3. Arithmetic Mean

The calculations can be further simplified by dividing all the deviations taken from assumed mean by the common factor 'c'. The objective is to avoid large numerical figures, i.e., if d = X – A is very large, then find d`

d`= \(\frac{d}{c} = \frac{X-A}{c} \)

The formula for AM becomes:

STEP DEVIATION METHOD

$$\overline{X} = A + \frac{\Sigma d'}{N} \times C$$

$$d' = \frac{d}{C} = \frac{X-A}{C}$$

Illustration

Family	Income (X)	② $d = X - A$	③ $d' = d/10$
A	850	0	0
B	700	−150	−15
C	100	−750	−75
D	750	−100	−10
E	5000	4150	415
F	80	−770	−77
G	420	−430	−43
H	2500	1650	165
I	400	−450	−45
J	360	−490	−49

④ $\Sigma d' = 266$

① Assumed Mean, A = 850

Assume C = 10

⑤ $\overline{X} = A + \frac{\Sigma d'}{n} \times C$

$= 850 + \frac{266}{10} \times 10$

$= \boxed{1116}$

Look at steps ① → ⑤

STEP DEVIATION METHOD