Regression Analysis & Times Series

Linear regression

Linear regression is a simple statistical tool used to model the dependence of a variable (say, costs) on one (or more) variables (say, volume). This functional relationship may then be formally stated as an equation, with associated statistical values that describe how well this equation fits the data.

The equation of a straight line is:

$$y = a + bx$$

where y = dependent variable
a = intercept (on y-axis)
b = gradient
x = independent variable

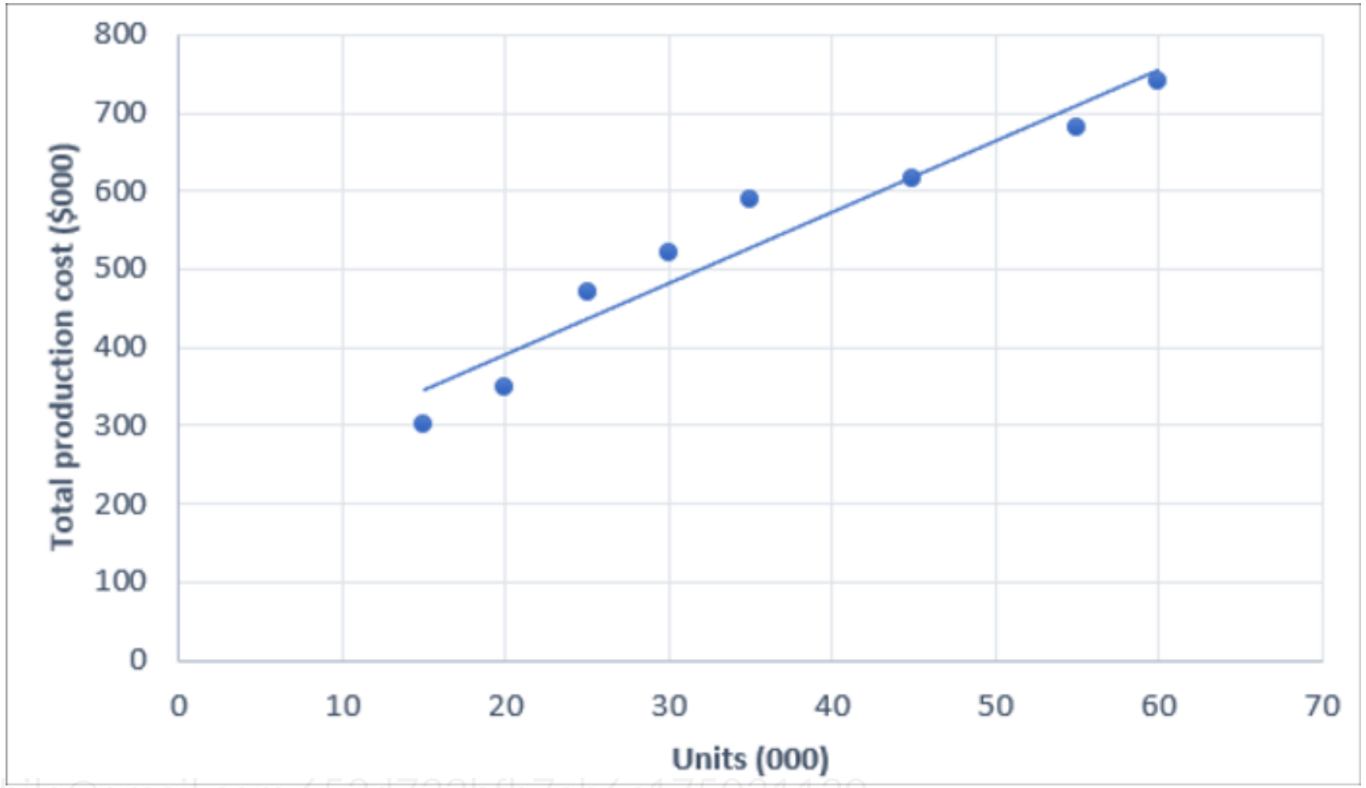
and

$$b = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2}$$

where n = number of pairs of data

and

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



Regression Analysis

Advantages

1.Regression analysis help in making prediction and forecasting for business in near and long term.

2. It supports business decisions by providing necessary information related to dependent target and predictors.

3. Business used regression analysis technique for identifying new opportunities in market.

Dis Advantages

Limitations of simple linear regression

(1) Assumes a linear relationship between the variables.

(2) Only measures the relationship between two variables. In reality, the dependent variable is affected by many independent variables.

(3) Only interpolated forecasts tend to be reliable. The equation should not be used for extrapolation.

(4) Regression assumes that the historical behaviour of the data continues into the foreseeable future.

Correlation

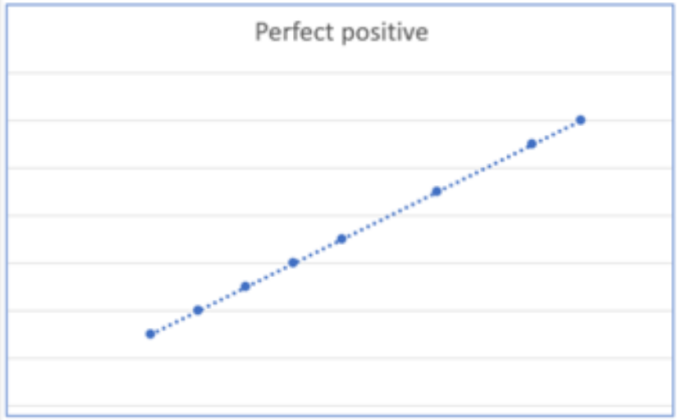
Regression analysis attempts to find the relationship between a number of variables. Correlation is concerned with establishing how strong the relationship is.

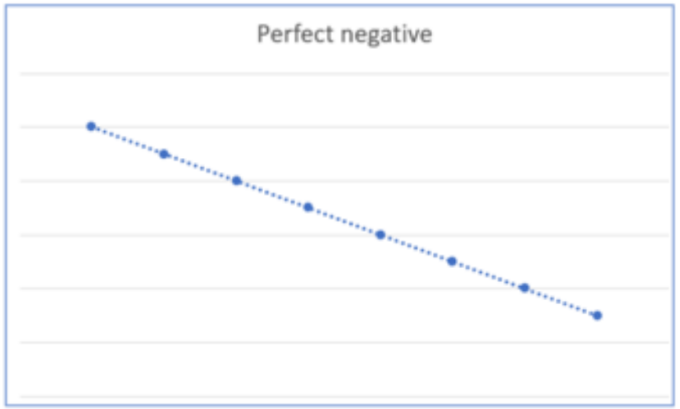
Positive and negative correlation

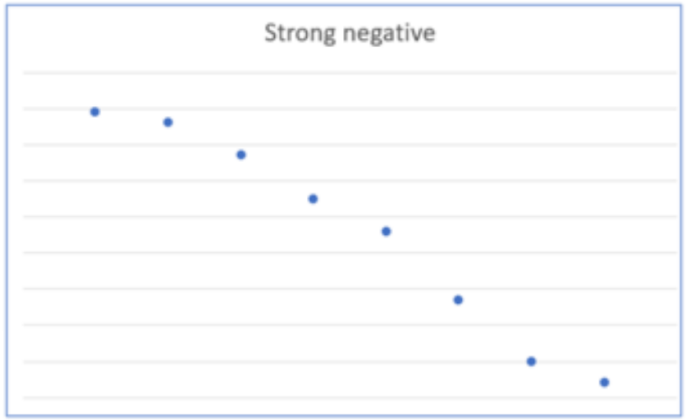
Correlation can be positive or negative.

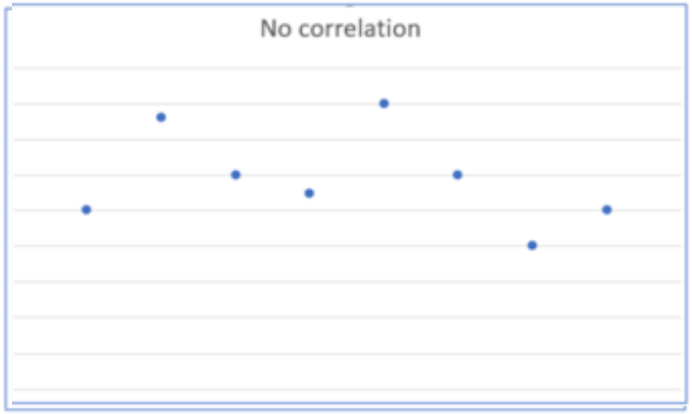
Positive correlation means that high values of one variable are associated with high values of the other and that low values of one are associated with low values of the other.

Negative correlation means that low values of one variable are associated with high values of the other and vice versa.









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

r must always be between –1 and +1.

The coefficient of determination

coefficient of determination which is calculated as r^2.

Limitations of correlation

(1) Correlation could be misleading used on sample data. Because an apparent correlation in a sample is not necessarily present in the population from which the sample came from. It could be only due to chance coincidence – a random sampling error. This is why a correlation must be accompanied by a significance test to assess its reliability.

(2) Correlation and causality ae not the same thing. Although a correlation demonstrates that a relationship exists between two variables, it does not automatically imply that one causes the other. Just because X and Y are correlated in some way does not mean that X causes a change in Y, or vice versa.

Times Series

Time series forecasting methods are based on analysis of historical data. They make the assumption that past patterns in data, such as seasonality, can be used to forecast future data points. This means that its future predictions are more curved than linear.

Characteristic time series components

The **basic trend** (long-term)

Seasonal variations (short-term)

Cyclical variations (medium-term)

Random variations (short-term)

Addictive Model

Multiplicative Model

Times Series

Advantages and disadvantages of time series analysis

Advantages	Disadvantages
• Identifies seasonal variations	• Complicated
• Can be non-linear	• 'Seasons' may change
• Accurate	• Based on historical data
	• Less useful on the long term

There are a number of problems with using time series analysis in order to estimate or forecast future results.

- The main problem is the inherent weakness of extrapolation. In order to estimate the trend for the future, the trend line is extended on the graph and the figures read off. However, although the time series has moved in that particular manner in the past, it does not necessarily mean that it will continue to do so in the future.
- The seasonal adjustments used to find the forecast for the future are also based upon historic figures that may well already be out of date. There is no guarantee that the seasonal variations will remain the same in the future. If the time series has a large residual or random variation element, then this will make any forecasts even less reliable.