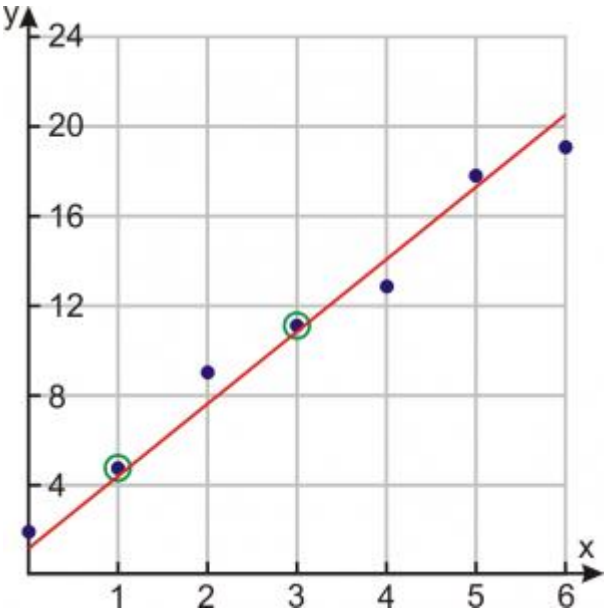
BSCS-507
DCS-UOK
Supervisor: Dr. Shaista Rais

Least Squares Regression:

The concept of least square regression is a mathematical process of regression analysis of a set of data that shows their relationship with each other. It shows the line of best fit, which is a graphical representation of all the data points that fit best in a straight line.

The ordinary least squares regression is a visual representation which shows the relation between an independent variable that is known and a dependent variable which is unknown. It is extremely popular and widely used by analysts, mathematicians, and even traders and investors to identify price and performance trends and also spot opportunities for investment.

Let us consider two variables, x and y. These are plotted on a graph with values of x on the x-axis and y on the y-axis. The dependent variables are all plotted on the y-axis and independent variables are on the x-axis. The dots represent these values in the below graph. A straight line is drawn through the dots – referred to as the line of best fit.



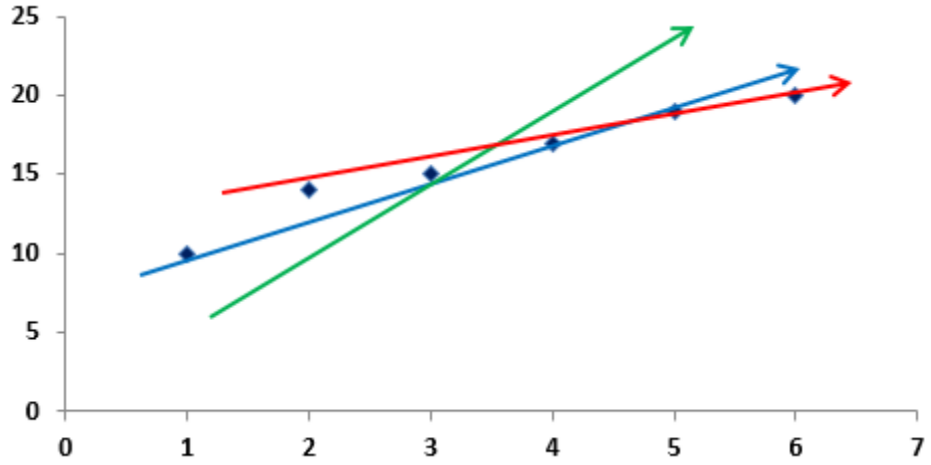
The objective of least squares regression is to ensure that the line drawn through the set of values provided establishes the closest relationship between the values.

Through this method, the behavior of the dependent variables can be analysed and predicted which is useful in the financial market. This linear relationship helps in reducing the vertical distance between the straight regression line and the data points. The least square explains the least value of summation of the squares of each error, which is also known as variance.

Line of Best Fit in the Least Square Regression

The line of best fit
is a straight line drawn through a scatter of data points that best represents the relationship between them.

Let us consider the following graph wherein a data set plot along the x and y-axis. These data points represent using the blue dots. Three lines are drawn through these points – a green, a red, and a blue line. The green line passes through a single point, and the red line passes through three data points. However, the blue line passes through four data points, and the distance between the residual points and the blue line is minimal compared to the other two lines.



In the above graph, the blue line represents the line of best fit as it lies closest to all the values and the distance between the points outside the line to the line is minimal (the distance between the residuals to the line of best fit – also referred to as the sums of squares of residuals). However, in the other two lines, the orange and the green, the distance between the residuals and the lines is greater than the blue line.

The least-squares method provides the closest relationship between the dependent and independent variables by minimizing the distance between the residuals, and the line of best fit, i.e., the sum of squares of residuals is minimal under this approach. Hence, the term "least squares."

BSCS-507
DCS-UOK
Supervisor: Dr. Shaista Rais

Scatter plots are the graphs that present the relationship between two variables in a data-set. It represents data points on a two-dimensional plane. The independent variable or attribute is plotted on the X-axis, while the dependent variable is plotted on the Y-axis. These plots are often called scatter graphs or **scatter diagrams**.