

Simple Linear Regression

Introduction

Regression analysis is a statistical procedure for assessing the connections among variables. It incorporates numerous methods for demonstrating and investigating several variables, when the emphasis is on the relationship between a dependent variable and one or more independent variables. Usually, more than one independent variable impacts the dependent variable. Simple linear regression is one of the methods used in regression analysis where there is only one independent variable. In simple linear regression we predict one variable from the second variable. In this process when prediction of unknown variable are plotted with the help of known variable a straight line is formed.

Real World Applications

Regression is used largely in business purposes. We predict an unknown variable with the data we have from a known variable. In its most rudimentary form, regression analysis is the estimation of the ratio between two variables. Say you want to estimate the growth in meat sales (MS Growth), based on economic growth (GDP Growth). If past data indicates that the growth in meat sales is around one and a half times the growth in the economy, the regression would look as follows: $MS\ Growth = (GDP\ Growth) * 1.5$. The relationship between many variables also involves a constant.

The most well-known utilization of regression in business is to anticipate occasions that have yet to happen. Demand analysis, for instance, predicts what number of units shoppers will buy. Numerous other key parameters other than demand are needy variables in regression models, in any case. Anticipating the quantity of customers who will go before a specific announcement or the quantity of viewers who will watch the Super Bowl may offer administration some assistance with assessing what to pay for a promotion. Insurance agencies vigorously depend on regression examination to appraise what number of approach holders will be included in mishaps or be casualties of robberies, for instance.

Another key utilization of regression models is the advancement of business procedures. A industry administrator may, for instance, fabricate a model to comprehend the relationship between broiler temperature and the time frame of realistic usability of the treats prepared in those stoves. An organization working a call focus may wish to know the relationship between hold up times of guests and number of protests. A principal driver of improved efficiency in business and quick monetary progression around the world amid the twentieth century was the successive utilization of statistical tools in assembling and administration commercial ventures. Today, supervisors considers regression a vital apparatus.

Theory

Here, in this article we are dealing with simple linear regression. In simple linear regression we have only one independent variable with respect to the dependent variable that we deal with. The variable which we are predicting is known as criterion variable and the variable on which we base our predictions is known as predictor variable. In simple linear regression the connection between these two variables is a straight line on which we base our future predictions.

Suppose we have the criterion variable as Y , which we want to predict and let X be the predictor variable on which the predictions are based. Now we name the new variable Y' which we are predicting with the variables we have. As discussed earlier, this relation between Y' and X should be a straight line. Therefore,

$$Y' = bX + A$$

Here, b is the slope of the equation and

$$b = r \frac{S_Y}{S_X}$$

And the intercept A on Y

$$A = M_Y - bM_X$$

M_X – Mean of X

M_Y – Mean of Y

S_X – Standard deviation of X

S_Y – Standard deviation of Y

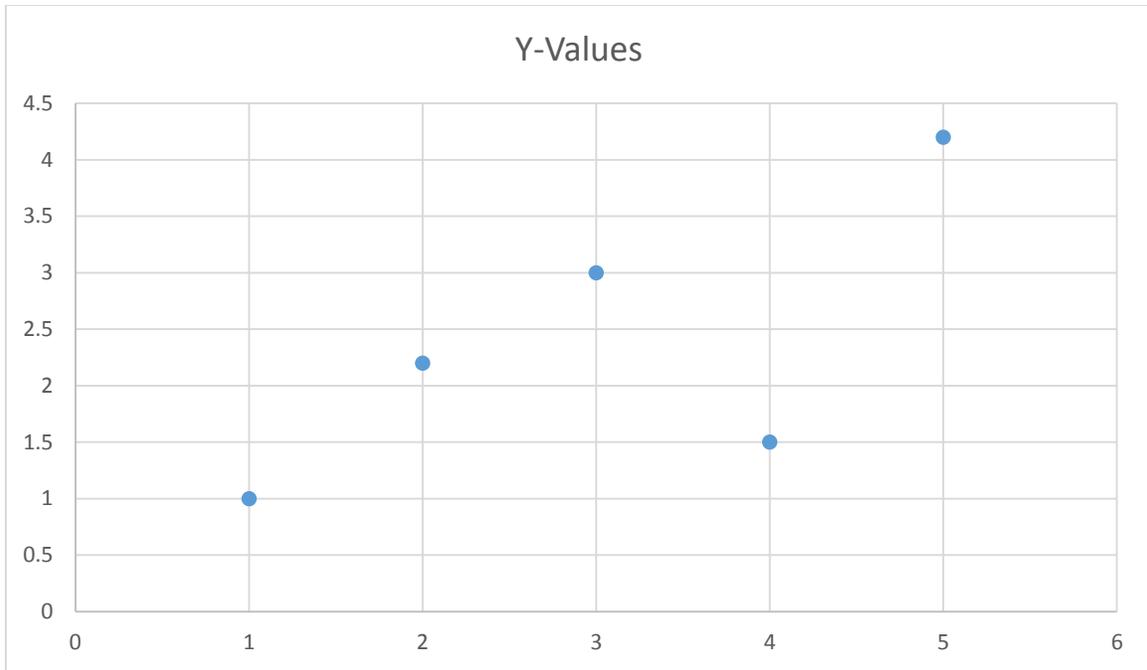
r = correlation between X and Y

Numerical Example

We take a simple example of two variables and plot a regression line with the above theory.

X	Y
1.0	1.0
2.0	2.2
3.0	3
4.0	1.5
5.0	4.2

We can plot these values to see the pattern



Now, with all the values we can proceed to evaluate our regression line.

$$M_X = 3$$

$$M_Y = 2.38$$

$$S_X = 1.58$$

$$S_Y = 1.3$$

$$r = 0.7$$

$$b = 0.58$$

$$A = 0.64$$

From these values we have the regression line as $Y' = 0.58X + 0.64$

