

Linear Regression Analysis

Explorable.com 123.9K reads

Linear regression analysis is a powerful technique used for predicting the unknown value of a variable from the known value of another variable.

More precisely, if X and Y are two related variables, then linear regression analysis helps us to predict the value of Y for a given value of X or vice verse.

For example age of a human being and maturity are related variables. Then linear regression analyses can predict level of maturity given age of a human being.



The banner features the Explorable logo at the top center. Below it, three quiz cards are displayed in a row. Each card has a square image at the top and a title below it. The first card shows a pair of red roller skates on a wooden deck, titled 'Quiz: Psychology 101 Part 2'. The second card shows a fan of colorful pens, also titled 'Quiz: Psychology 101 Part 2'. The third card shows a Ferris wheel at sunset, titled 'Quiz: Flags in Europe'. To the right of the cards is a link that says 'See all quizzes =>'.

EXPLORABLE
Quiz Time!

 Quiz: Psychology 101 Part 2

 Quiz: Psychology 101 Part 2

 Quiz: Flags in Europe

[See all quizzes =>](#)

Dependent and Independent Variables

By linear regression, we mean models with just one independent and one dependent variable. The variable whose value is to be predicted is known as the dependent variable [1] and the one whose known value is used for prediction is known as the independent variable [2].

Two Lines of Regression

There are two lines of regression- that of Y on X and X on Y. The line of regression of Y on X is given by $Y = a + bX$ where a and b are unknown constants known as intercept and slope of the equation. This is used to predict the unknown value of variable Y when value of variable X

is known.

$$Y = a + bX$$

On the other hand, the line of regression of X on Y is given by $X = c + dY$ which is used to predict the unknown value of variable X using the known value of variable Y. Often, only one of these lines make sense.

Exactly which of these will be appropriate for the analysis in hand will depend on labeling of dependent and independent variable in the problem to be analyzed.

Choice of Line of Regression

For example, consider two variables crop yield (Y) and rainfall (X). Here construction of regression line of Y on X would make sense and would be able to demonstrate the dependence of crop yield on rainfall. We would then be able to estimate crop yield given rainfall.

Careless use of linear regression analysis could mean construction of regression line of X on Y which would demonstrate the laughable scenario that rainfall is dependent on crop yield; this would suggest that if you grow really big crops you will be guaranteed a heavy rainfall.

Regression Coefficient

The coefficient of X in the line of regression of Y on X is called the regression coefficient of Y on X. It represents change in the value of dependent variable (Y) ^[1] corresponding to unit change in the value of independent variable (X) ^[2].

For instance if the regression coefficient of Y on X is 0.53 units, it would indicate that Y will increase by 0.53 if X increased by 1 unit. A similar interpretation can be given for the regression coefficient of X on Y.

Once a line of regression has been constructed, one can check how good it is (in terms of predictive ability) by examining the coefficient of determination (R²). R² always lies between 0 and 1. All software provides it whenever regression procedure is run.

R² - coefficient of determination

The closer R² is to 1, the better is the model and its prediction. A related question is whether the independent variable significantly ^[3] influences the dependent variable. Statistically, it is equivalent to testing the null hypothesis ^[4] that the regression coefficient is zero. This can be done using t-test.

Assumption of Linearity

Linear regression does not test whether data is linear. It finds the slope and the intercept

assuming that the relationship between the independent and dependent variable can be best explained by a straight line.

One can construct the scatter plot to confirm this assumption. If the scatter plot reveals non linear relationship [5], often a suitable transformation can be used to attain linearity.

Source URL: <https://explorable.com/linear-regression-analysis>

Links

[1] <https://explorable.com/dependent-variable>

[2] <https://explorable.com/independent-variable>

[3] <https://explorable.com/statistically-significant-results>

[4] <https://explorable.com/null-hypothesis>

[5] <https://explorable.com/non-linear-relationship>