Linear Regression

Maths › Linear Regression

Download PDF   NCERT Solutions >   Popular Textbook Solutions >   CBSE   >

Introduction to Linear Regression

Is this page helpful?   ⋮

Linear regression is used to predict the relationship between two variables by applying a linear equation to observed data. There are two types of variable, one variable is called an independent variable, and the other is a dependent variable. Linear regression is commonly used for predictive analysis. The main idea of regression is to examine two things. First, does a set of predictor variables do a good job in predicting an outcome (dependent) variable? The second thing is which variables are significant predictors of the outcome variable? In this article, we will discuss the concept of the Linear Regression Equation, formula and Properties of Linear Regression.

Examples of Linear Regression

The weight of the person is linearly related to their height. So, this shows a linear relationship between the height and weight of the person. According to this, as we increase the height, the weight of the person will also increase. It is not necessary that one variable is dependent on others, or one causes the other, but there is some critical relationship between the two variables. In such cases, we use a scatter plot to simplify the strength of the relationship between the variables. If there is no relation or linking between the variables then the scatter plot does not indicate any increasing or decreasing pattern. In such cases, the linear regression design is not benefici

The measure of the relationship between two variables is shown by the correlation coefficient. The range of the coefficient lies between -1 to +1. This coefficient shows the strength of the association of the observed data between two variables.

Linear Regression Equation is given below:

Y=a+bX

where X is the independent variable and it is plotted along the x-axis

Y is the dependent variable and it is plotted along the y-axis

Here, the slope of the line is b, and a is the intercept (the value of y when x = 0).

Linear Regression Formula

As we know, linear regression shows the linear relationship between two variables. The equation of linear regression is similar to that of the slope formula. We have learned this formula before in earlier classes such as a linear equation in two variables. Linear Regression Formula is given by the equation

Y= a + bX

We will find the value of a and b by using the below formula

$$a= \frac{\left(\sum_Y\right)\left(\sum_{X^2}\right) - \left(\sum_X\right)\left(\sum_{XY}\right)}{n\left(\sum_{x^2}\right) - \left(\sum_x\right)^2}$$

$$b= \frac{n\left(\sum_{XY}\right) - \left(\sum_X\right)\left(\sum_Y\right)}{n\left(\sum_{x^2}\right) - \left(\sum_x\right)^2}$$

Simple Linear Regression

Simple linear regression is the most straight forward case having a single scalar predictor variable x and a single scalar response variable y. The equation for this regression is given as y=a+bx

The expansion to multiple and vector-valued predictor variables is known as multiple linear regression. It is also known as multivariable linear regression. The equation for this regression is given as Y = a+bX. Almost all real-world regression patterns include multiple predictors. The basic explanations of linear regression are often explained in terms of multiple regression. Note that, in these cases, the dependent variable

The most popular method to fit a regression line in the XY plot is found by using least-squares. This process is used to determine the best-fitting line for the given data by reducing the sum of the squares of the vertical deviations from each data point to the line. If a point rests on the fitted line accurately, then the value of its perpendicular deviation is 0. It is 0 because the variations are first squared, then added, so their positive and negative values will not be cancelled. Linear regression determines the straight line, known as the least-squares regression line or LSRL. Suppose Y is a dependent variable and X is an independent variable, then the population regression line is given by the equation;

$Y = B_0 + B_1 X$

Where

B_0 is a constant

B_1 is the regression coefficient

When a random sample of observations is given, then the regression line is expressed as;

$\hat{y} = b_0 + b_1 x$

where b_0 is a constant

b_1 is the regression coefficient,

x is the independent variable,

$\hat{y}$ is known as the predicted value of the dependent variable.

Properties of Linear Regression

For the regression line where the regression parameters b_0 and b_1 are defined, the following properties are applicable:

- The regression line reduces the sum of squared differences between observed values and predicted values.

- The regression line passes through the mean of X and Y variable values.

- The regression constant b_0 is equal to the y-intercept of the linear regression.

- The regression coefficient b_1 is the slope of the regression line. Its value is equal to the average change in the dependent variable (Y) for a unit chang~ the independent variable (X)

$Y = B_0 + B_1 X$

Where

B_0 is a constant

B_1 is the regression coefficient

Given below is the formula to find the value of the regression coefficient.

$B_1 = b_1 = \Sigma[(x_i - x)(y_i - y)] / \Sigma[(x_i - x)^2]$

Where x_i and y_i are the observed data sets.

And x and y are the mean value.

Importance of Regression Line

A regression line is used to describe the behaviour of a set of data, a logical approach that helps us study and analyze the relationship between two different continuous variables. Which is then enacted in machine learning models, mathematical analysis, statistics field, forecasting sectors, and other such quantitative applications. Looking at the financial sector, where financial analysts use linear regression to predict stock prices and commodity prices and perform various stock valuations for different securities. Several well-renowned companies make use of linear regressions for the purpose of predicting sales, inventories, etc.

Key Ideas of Linear Regression

- Correlation explains the interrelation between variables within the data.

- Variance is the degree of the spread of the data.

- Standard deviation is the dispersion of mean from a data set by studying the variance's square root.

- Residual (error term) is the actual value found within the dataset minus the expected value that is predicted in linear regression.

Important Properties of Regression Line

- Regression coefficient values remain the same because the shifting of origin takes place because of the change of scale. The property says that if the variables x and y are changed to u and v respectively u= (x-a)/p v=(y-c) /q, Here p and q are the constants.Byz =q/p*bvu Bxy=p/q*buv.

property, the intersection of the two regression lines is (x`, y`), which is the solution of the equations for both the variables x and y.

- You will understand that the correlation coefficient between the two variables x and y is the geometric mean of both the coefficients. Also, the sign over the values of correlation coefficients will be the common sign of both the coefficients. So, if according to the property regression coefficients are byx= (b) and bxy= (b') then the correlation coefficient is r=+-sqrt (byx + bxy) which is why in some cases, both the values of coefficients are negative value and r is also negative. If both the values of coefficients are positive then r is going to be positive.

- The regression constant (a_0) is equal to the y-intercept of the regression line and also a_0 and a_1 are the regression parameters.

Regression Line Formula:

A linear regression line equation is written as-

Y = a + bX

where X is plotted on the x-axis and Y is plotted on the y-axis. X is an independent variable and Y is the dependent variable. Here, b is the slope of the line and a is the intercept, i.e. value of y when x=0.

Multiple Regression Line Formula: $y = a + b_1x_1 + b_2x_2 + b_3x_3 + ... + b_tx_t + u$

Assumptions made in Linear Regression

- The dependent/target variable is continuous.

- There isn't any relationship between the independent variables.

- There should be a linear relationship between the dependent and explanatory variables.

- Residuals should follow a normal distribution.

- Residuals should have constant variance.

- Residuals should be independently distributed/no autocorrelation.

Solved Examples

1. Find a linear regression equation for the following two sets of data:

| y | 3 | 7 | 5 | 10 |

Sol: To find the linear regression equation we need to find the value of Σx, Σy, Σx

2

2

and Σxy

Construct the table and find the value

x	y	x²	xy
2	3	4	6
4	7	16	28
6	5	36	30
8	10	64	80
Σx = 20	Σy = 25	Σx² = 120	Σxy = 144

The formula of the linear equation is y=a+bx. Using the formula we will find the value of a and b

$$a= \frac{\left(\sum_Y\right)\left(\sum_{X^2}\right) - \left(\sum_X\right)\left(\sum_{XY}\right)}{n\left(\sum_{x^2}\right) - \left(\sum_x\right)^2}$$

Now put the values in the equation

$$a = \frac{25 \times 120 - 20 \times 144}{4 \times 120 - 400}$$

120

b= $\dfrac{n\left(\sum_{XY}\right) - \left(\sum_{X}\right)\left(\sum_{Y}\right)}{n\left(\sum_{x^2}\right) - \left(\sum_{x}\right)^2}$

Put the values in the equation

$$b = \frac{4 \times 144 - 20 \times 25}{4 \times 120 - 400}$$

b=$\dfrac{76}{80}$

b=0.95

Hence we got the value of a = 1.5 and b = 0.95

The linear equation is given by

Y = a + bx

Now put the value of a and b in the equation

Hence equation of linear regression is y = 1.5 + 0.95x

Is this page helpful?

⋮

Prepare better for CBSE Class 10
Try ~~Vedantu PRO~~ **for free**

▶ LIVE classes with top teachers

Aa In-class doubt-solving

Practice tests and quizzes

Start 7-day free trial

Book your Free Demo session

Get a flavour of LIVE classes here at Vedantu

Introducing **Ai LIVE**

Full year course starts from

₹20,000 for 2 years

Select and buy

| |
| Enter your name |

| |
| Enter mobile number (India only) |

| |
| Book a free demo |

Vedantu Improvement Promise

We promise improvement in marks or get your fees back. T&C Apply*

Latest Vedantu courses for you

Grade 11 Science | ALLBOARDS | JEE | English   ⌄

▶

JEE 2-Year (2022-24)

📅 Academic year 2022-24   🔤 ENGLISH   📖 Unlimited access till final school exam

✅ JEE Full course   ✅ 1000+ chapter specific short courses

⚛ Physics   ⚗ Chemistry   📏 Maths

₹20,000 for 2 years
EMI starts from ₹1,666.67 per month

Select and buy

Recently Updated Pages

Introducing Ai LIVE
   Full year course starts from
₹20,000 for 2 years

Select and buy

○ ○

Trending topics

○ ○ ○

FAQs on Linear Regression

1. Why are regression lines considered to be important?　　⌃

Regression lines are used in the financial sector by companies where various financial analysts implement linear regressions to predict stock prices, commodity prices and to perform valuations for many different securities.

2. How do you define Slope?　　⌄

3. How will you explain the difference between Linear regression and multiple regression?　　⌄

Introducing **Ai LIVE**

Full year course starts from

₹20,000 for 2 years

Select and buy

6. What are the Types of Linear Regression? ⌄

7. What are the Differences Between Linear and Logistic Regression? ⌄

8. How Does a Linear Regression Work? ⌄

Comment

Enter your comment here

Post Comment
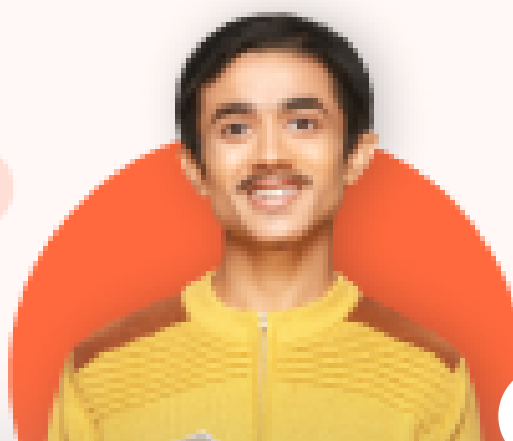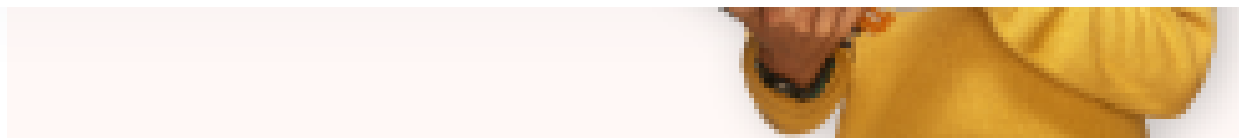
NCERT Book Solutions

NCERT

NCERT Solutions

NCERT Solutions for Class 12 Maths

NCERT Solutions for Class 12 Physics

NCERT Solutions for Class 12 Chemistry

NCERT Solutions for Class 12 Biology

NCERT Solutions for Class 11 Maths

NCERT Solutions for Class 11 Physics

NCERT Solutions for Class 11 Chemistry

NCERT Solutions for Class 11 Biology

NCERT Solutions for Class 10 Maths

NCERT Solutions for Class 10 Science

NCERT Solutions for Class 9 Maths

NCERT Solutions for Class 9 Science

Reference Book Solutions

HC Verma Solutions

RD Sharma Solutions

RD Sharma Class 10 Solutions

RD Sharma Class 9 Solutions

RS Aggarwal Solutions

RS Aggarwal Class 10 Solutions

ICSE Class 10 solutions

Lakmir Singh Solutions

Text Book Solutions

Important Formulas

Introducing **Ai LIVE**

Full year course starts from

₹20,000 for 2 years

Select and buy

Toll Free:
1800-120-456-456

Mobile:
+91 988-660-2456

Introducing Ai LIVE

Full year course starts from

₹20,000 for 2 years

Select and buy