

How can I make predictions using linear regression?

Authored by
stats writer

May 4, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I make predictions using linear regression?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=142563>

Linear regression is a statistical method used for making predictions by establishing a relationship between a dependent variable and one or more independent variables. This method calculates the linear relationship between the variables and uses it to predict the value of the dependent variable based on the values of the independent variables. To make predictions using linear regression, one needs to first collect a set of data points and then use a regression model to analyze the data and determine the best fitting line. This line can then be used to predict the values of the dependent variable for new data points. The accuracy of these predictions can be improved by choosing appropriate variables, evaluating the model's performance and adjusting it accordingly. Linear regression is widely used in many fields, including finance, economics, and social sciences, for making reliable predictions based on historical data.

Make Predictions with Linear Regression

Linear regression is a method we can use to quantify the relationship between one or more predictor variables and a .

One of the most common reasons for fitting a regression model is to use the model to predict the values of new observations.

We use the following steps to make predictions with a regression model:

Step 1: Collect the data. Step 2: Fit a regression model to the data. Step 3: Verify that the model fits the data well. Step 4: Use the fitted regression equation to predict the values of new observations.

The following examples show how to use regression models to make predictions.

Example 1: Make Predictions with a Simple Linear Regression Model

Suppose a doctor collects data for height (in inches) and weight (in pounds) on 50 patients.

She then fits a simple linear regression model using "weight" as the predictor variable and "height" as the response variable.

The fitted regression equation is as follows:

$$\text{Height} = 32.7830 + 0.2001 * (\text{weight})$$

After checking that of the linear regression model are met, the doctor concludes that the model fits the data well.

He can then use the model to predict the height of new patients based on their weight.

For example, suppose a new patient weighs 170 pounds. Using the model, we would predict that this patient would have a height of 66.8 inches:

$$\text{Height} = 32.7830 + 0.2001*(170) = 66.8 \text{ inches}$$

Example 2: Make Predictions with a Multiple Linear Regression Model

Suppose an economist collects data for total years of schooling, weekly hours worked, and yearly income on 30 individuals.

He then fits a multiple linear regression model using "total years of schooling" and "weekly hours worked" as the predictor variable and "yearly income" as the response variable.

$$\text{Income} = 1,342.29 + 3,324.33*(\text{years of schooling}) + 765.88*(\text{weekly hours worked})$$

After checking that all of the linear regression model assumptions are met, the economist concludes that the model fits the data well.

He can then use the model to predict the yearly income of a new individual based on their total years of schooling and weekly hours worked.

For example, suppose a new individual has 16 years of total schooling and works an average of 40 hours per

week. Using the model, we would predict that this individual would have a yearly income of \$85,166.77:

$$\text{Income} = 1,342.29 + 3,324.33*(16) + 765.88*(45) = \$85,166.77$$

On Using Confidence Intervals

When using a regression model to make predictions on new observations, the value predicted by the regression model is known as a .

Although the point estimate represents our best guess for the value of the new observation, it's unlikely to *exactly* match the value of the new observation.

So, to capture this uncertainty we can create a confidence interval - a range of values that is likely to contain a population parameter with a certain level of confidence.

For example, instead of predicting that a new individual will be 66.8 inches tall, we may create the following confidence interval:

95% Confidence Interval =

We would interpret this interval to mean that we're 95% confident that the true height of this individual is between 64.8 inches and 68.8 inches.

Cautions on Making Predictions

Keep in mind the following when using a regression model to make predictions:

1. Only use the model to make predictions within the range of data used to estimate the regression model.

For example, suppose we fit a regression model using the predictor variable "weight" and the weight of individuals in the sample we used to estimate the model ranged between 120 pounds and 180 pounds.

It would be invalid to use the model to estimate the height of an individual who weighted 200 pounds because this falls outside of the range of the predictor variable that we used to estimate the model.

It's possible that the relationship between weight and height is different outside of the range of 120 to 180 pounds, so we shouldn't use the model to estimate the height of an individual who weighs 200 pounds.

2. Only use the model to make predictions for the population you sampled.

For example, suppose the population that an economist draws a sample from all lives in a particular city.

We should only use the fitted regression model to predict the yearly income of individuals in this city since the entire sample that was used to fit the model lived in this city.

ARABPSYCHOLOGY.COM