How can R-Squared be calculated by hand?

Authored by

stats writer

May 2, 2024

R-Squared is a statistical measure that indicates the level of variation in a dependent variable that is explained by an independent variable(s). It is commonly used to evaluate the performance of a regression model. The calculation of R-Squared involves several steps that can be performed manually by hand. First, the data points for the dependent and independent variables are plotted on a scatter plot. Then, the best-fit line is drawn through the data points. Next, the distance between each data point and the best-fit line is calculated. This distance is squared and added together to find the sum of squared errors (SSE). The total sum of squares (SST) is also calculated by finding the squared distance between each data point and the mean of the dependent variable. Finally, R-Squared is obtained by dividing the difference between SST and SSE by SST. This result is then subtracted from 1 to get the final R-Squared value. This process can be repeated for different regression models to compare their R-Squared values and determine the best fitting model.

Calculate R-Squared by Hand

In statistics, R-squared (R2) measures the proportion of the variance in the that can be explained by the predictor variable in a regression model.

We use the following formula to calculate R-squared:

R2 = 2

The following step-by-step example shows how to calculate R-squared by hand for a given regression model.

Step 1: Create a Dataset

First, let's create a dataset:

x	y
3	22
5	24
5	28
7	20
9	28
12	31
14	37
17	33

Step 2: Calculate Necessary Metrics

Next, let's calculate each metric that we need to use in the R2 formula:

	x	x^2	y	y^2	xy
	3	9	22	484	66
	5	25	24	576	120
	5	25	28	784	140
	7	49	20	400	140
	9	81	28	784	252
	12	144	31	961	372
	14	196	37	1369	518
	17	289	33	1089	561
Σ	**72**	**818**	**223**	**6447**	**2169**

Step 3: Calculate R-Squared

Lastly, we'll plug in each metric into the formula for R2:

R2 = 2R2 = 2R2 = 0.6686

Note: The *n* in the formula represents the number of observations in the dataset and turns out to be n = 8 observations in this example.

Assuming *x* is the predictor variable and *y* is the response variable in this regression model, the R-squared for the model is 0.6686.

This tells us that 66.86% of the variation in the variable *y* can be explained by variable *x*.