

What is the interpretation of Root Mean Square Error (RMSE)?

Authored by
stats writer

April 29, 2024

RECOMMENDED CITATION

stats writer (2024). *What is the interpretation of Root Mean Square Error (RMSE)?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=141216>

Root Mean Square Error (RMSE) is a commonly used metric for evaluating the accuracy of a prediction model. It measures the average difference between the predicted values and the actual values, taking into account both the magnitude and direction of the errors. In simpler terms, RMSE indicates how far, on average, the predicted values are from the actual values. A lower RMSE value indicates a better fit of the model to the data. In other words, it is a measure of the model's performance and can be used to compare different models or to track the improvement of a model over time. RMSE is often used in fields such as statistics, machine learning, and data analysis to assess the effectiveness of a model in predicting outcomes.

Interpret Root Mean Square Error (RMSE)

is a technique we can use to understand the relationship between one or more predictor variables and a .

One way to assess how well a regression model fits a dataset is to calculate the root mean square error, which is a metric that tells us the average distance between the predicted values from the model and the actual values in the dataset.

The lower the RMSE, the better a given model is able to "fit" a dataset.

The formula to find the root mean square error, often abbreviated RMSE, is as follows:

$$\text{RMSE} = \sqrt{\sum(P_i - O_i)^2 / n}$$

where:

Σ is a fancy symbol that means "sum"
 P_i is the predicted value for the i th observation in the dataset
 O_i is the observed value for the i th observation in the dataset
 n is the sample size

The following example shows how to interpret RMSE for a given regression model.

Example: How to Interpret RMSE for a Regression Model

Suppose we would like to build a regression model that uses "hours studied" to predictor "exam score" of students on a particular college entrance exam.

We collect the following data for 15 students:

Hours Studied	Exam Score
1	68
1	78
1	75
2	83
2	80
2	78
2	89
2	93
3	90
3	91
4	94
5	88
5	84
5	90
6	94

We then use statistical software (like Excel, SPSS, R, Python) etc. to find the following fitted regression model:

$$\text{Exam Score} = 75.95 + 3.08 * (\text{Hours Studied})$$

We can then use this equation to predict the exam score of each student, based on how many hours they studied:

Hours Studied	Exam Score	Predicted Score
1	68	79.03
1	78	79.03
1	75	79.03
2	83	82.11
2	80	82.11
2	78	82.11
2	89	82.11
2	93	82.11
3	90	85.19
3	91	85.19
4	94	88.27
5	88	91.35
5	84	91.35
5	90	91.35
6	94	94.43

We can then calculate the squared difference between each predicted exam score and the actual exam score. Then we can take the square root of the mean of these differences:

Hours Studied	Exam Score	Predicted Score	Squared Difference
1	68	79.03	121.661
1	78	79.03	1.061
1	75	79.03	16.241
2	83	82.11	0.792
2	80	82.11	4.452
2	78	82.11	16.892
2	89	82.11	47.472
2	93	82.11	118.592
3	90	85.19	23.136
3	91	85.19	33.756
4	94	88.27	32.833
5	88	91.35	11.223
5	84	91.35	54.023
5	90	91.35	1.823
6	94	94.43	0.185
		RMSE	5.681

The RMSE for this regression model turns out to be 5.681.

Recall that the of a regression model are the differences between the observed data values and the predicted values from the model.

$$\text{Residual} = (P_i - O_i)$$

where

P_i is the predicted value for the i th observation in the dataset
 O_i is the observed value for the i th observation

in the dataset

And recall that the RMSE of a regression model is calculated as:

$$\text{RMSE} = \sqrt{\sum(P_i - O_i)^2 / n}$$

This means that the RMSE represents the square root of the variance of the residuals.

This is a useful value to know because it gives us an idea of the average distance between the observed data values and the predicted data values.

This is in contrast to the R^2 of the model, which tells us the proportion of the variance in the response variable that can be explained by the predictor variable(s) in the model.

Comparing RMSE Values from Different Models

The RMSE is particularly useful for comparing the fit of different regression models.

For example, suppose we want to build a regression model to predict the exam score of students and we want to find the best possible model among several

potential models.

Suppose we fit three different regression models and find their corresponding RMSE values:

RMSE of Model 1: 14.5 RMSE of Model 2: 16.7 RMSE of Model 3: 9.8

Model 3 has the lowest RMSE, which tells us that it's able to fit the dataset the best out of the three potential models.

ARABPSYCHOLOGY.COM