

How do you calculate RMSE (Root Mean Square Error) using SAS?

Authored by
stats writer

June 24, 2024

RECOMMENDED CITATION

stats writer (2024). *How do you calculate RMSE (Root Mean Square Error) using SAS?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=149869>

Root Mean Square Error (RMSE) is a commonly used metric in statistics to measure the difference between the actual values and the predicted values in a given dataset. In SAS (Statistical Analysis System), the RMSE can be calculated by first finding the squared differences between the actual and predicted values, then taking the mean of these squared differences, and finally taking the square root of the mean value. This is done using the PROC MEANS and PROC SQRT functions in SAS, which allow for efficient and accurate computation of the RMSE. By using SAS, researchers and analysts can easily and reliably calculate the RMSE to evaluate the performance of their predictive models and make data-driven decisions.

Calculate RMSE in SAS

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 symbol that represents "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 step-by-step example shows how to calculate the RMSE for a simple linear regression model in SAS.

Step 1: Create the Data

For this example, we'll create a dataset that contains the total hours studied and final exam score for 15 students.

We'll fit a simple linear regression model using *hours* as the predictor variable and *score* as the response variable.

The following code shows how to create this dataset in SAS:

```
/*create dataset*/  
data exam_data;  
input hours score;  
datalines;  
1 64  
2 66
```

4 76

5 73

5 74

6 81

6 83

7 82

8 80

10 88

11 84

11 82

12 91

12 93

14 89

;

run;

/*view dataset*/

proc printdata=exam_data;

Obs	hours	score
1	1	64
2	2	66
3	4	76
4	5	73
5	5	74
6	6	81
7	6	83
8	7	82
9	8	80
10	10	88
11	11	84
12	11	82
13	12	91
14	12	93
15	14	89

Step 2: Fit the Simple Linear Regression Model

Next, we'll use `proc reg` to fit the simple linear regression model:

```
/*fit simple linear regression model*/  
proc regdata=exam_data;  
model score = hours;  
run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: score

Number of Observations Read	15
Number of Observations Used	15

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	847.26698	847.26698	63.91	<.0001
Error	13	172.33302	13.25639		
Corrected Total	14	1019.60000			

Root MSE	3.64093	R-Square	0.8310
Dependent Mean	80.40000	Adj R-Sq	0.8180
Coeff Var	4.52852		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	65.33395	2.10599	31.02	<.0001
hours	1	1.98237	0.24796	7.99	<.0001

Step 3: Extract RMSE from Regression Model

If you only want to view the RMSE of this model and none of the other output results, you can use the following code:

```
/*fit simple linear regression model*/  
proc regdata=exam_data outest=outest noprint;  
model score = hours / rmse;  
run;
```

```
quit;
```

```
/*print RMSE of model*/
```

```
proc printdata=outest;
```

```
var _RMSE_;
```

```
run;
```

Obs	_RMSE_
1	3.64093

Notice that only the RMSE value of 3.64093 is shown in the output.

Note: The argument noprint in proc reg tells SAS not to print the entire output of regression results as it did in the previous step.

The following tutorials explain how to perform other common tasks in SAS: