

# How to Easily Calculate RMSE Using PROC GPOWER in SAS

Authored by  
**stats writer**

November 19, 2025

## RECOMMENDED CITATION

stats writer (2025). *How to Easily Calculate RMSE Using PROC GPOWER in SAS*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=97867>

The calculation of the Root Mean Square Error (RMSE) is a fundamental task in assessing the performance of predictive analytics models. While various SAS procedures can output RMSE, the most straightforward approach for standard regression model fitting involves using **PROC REG**. This powerful procedure enables users to specify the model structure and, crucially, utilize the RMSE option within the **MODEL** statement to generate the required metric directly in the output or save it to a dataset. This guide focuses on employing **PROC REG** for a clean and efficient RMSE extraction.

## Understanding Root Mean Square Error (RMSE)

One of the most essential metrics used to evaluate how effectively a regression model captures the underlying relationship in a dataset is the **Root Mean Square Error**. The RMSE quantifies the average magnitude of the errors--the differences between predicted values generated by the model and the actual observed values in the dataset. Conceptually, it represents the standard deviation of the residuals (prediction errors), providing a measure of how spread out these residuals are.

A primary advantage of using RMSE is that it is expressed in the same units as the response variable, making it highly interpretable in practical contexts. Furthermore, because the errors are squared before averaging, the RMSE naturally gives greater weight to larger errors, which is often desirable when seeking a model that minimizes significant deviations.

Crucially, the magnitude of the RMSE provides a direct indication of model quality: **the lower the RMSE value, the better the given model is able to "fit" the dataset**, meaning its predictions are, on average, closer to the actual observations. When comparing competing models, the model with the lowest RMSE is typically preferred, assuming all other statistical assumptions are met.

## The Mathematical Formula for RMSE

To fully appreciate the metric, it is helpful to review the mathematical structure underlying the **Root Mean Square Error**, often abbreviated as RMSE. The formula involves summing the squared differences between observed and predicted values, dividing by the sample size, and finally taking the square root to return the units to the original scale of the dependent variable.

The formal mathematical expression for calculating RMSE is provided below:

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

Understanding the notation used in this formula is key to comprehending the calculation process:

$\Sigma$  is the standard symbol for **summation**, indicating that the operation must be performed across

all data points in the sample.

$P_i$  represents the **predicted value** generated by the regression model for the  $i$ th observation in the dataset.

$O_i$  represents the **observed value** (the actual measurement) for the  $i$ th observation in the dataset.

$n$  is the **sample size**, representing the total number of observations available.

The subsequent sections provide a detailed, step-by-step tutorial demonstrating how to practically implement this calculation for a simple linear regression model using the SAS environment.

## Step 1: Preparing the Data in SAS

Before calculating any model metrics, we must first ensure that the dataset is properly structured and loaded into the SAS environment. For this illustrative example, we will construct a hypothetical dataset containing two variables: the total *hours studied* (the predictor variable) and the resulting *final exam score* (the response variable) for a sample of 15 students.

We aim to fit a simple linear regression model to determine if study time is a significant predictor of exam performance. The data creation process utilizes the **DATA** step in SAS, followed by the use of **DATALINES** to input the raw data observations directly.

The following code block demonstrates the standard procedure for creating and subsequently viewing this dataset in SAS. The use of **PROC PRINT** confirms the successful creation and structure of the `exam\_data` table.

```
/*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 print data=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

The resulting output image confirms that the input data contains 15 observations, with clear values for both the predictor (hours) and the response (score).

## Step 2: Fitting the Simple Linear Regression Model using PROC REG

Once the data is ready, the next necessary step is to apply the appropriate statistical procedure to fit the **simple linear regression model**. In SAS, the **PROC REG** procedure is the standard tool for least squares fitting of linear regression models. This procedure not only generates parameter estimates but also provides a comprehensive set of diagnostic statistics, including the RMSE implicitly within its standard output.

To execute the model fitting, we invoke **PROC REG** and specify the `exam\_data` dataset. The **MODEL** statement defines the relationship, where the variable `score` is designated as the dependent variable and `hours` as the independent variable. SAS automatically performs the

necessary calculations to estimate the coefficients and calculate the goodness-of-fit statistics.

Executing the following code will produce the standard, extensive output that includes the Analysis of Variance (ANOVA) table, parameter estimates, and various summary statistics.

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

**The REG Procedure**  
**Model: MODEL1**  
**Dependent Variable: score**

<b>Number of Observations Read</b>	15
<b>Number of Observations Used</b>	15

  

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

  

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

  

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

As illustrated in the summary output above, the RMSE value is present under the "Root MSE" label in the summary statistics section. This value serves as the initial calculation of the required metric for this specific regression model.

### Step 3: Explicitly Extracting RMSE for Reporting

While Step 2 displays the RMSE within the standard output, often in large statistical projects, analysts need to extract specific metrics programmatically for further analysis or reporting, rather than manually parsing large output tables. SAS provides functionality within **PROC REG** to suppress the voluminous standard output and save only the desired statistics to a new dataset.

To achieve this clean extraction, we utilize two key options: **NOPRINT** and **OUTEST**. The **NOPRINT** option suppresses the standard listing output, focusing the procedure on data creation. The **OUTEST=outest** option creates an output dataset (`outest`) containing parameter estimates and associated goodness-of-fit statistics. Crucially, the **RMSE** option must be added to the **MODEL** statement to ensure the Root MSE is included in the output dataset variable named **\_RMSE\_**.

The subsequent **PROC PRINT** statement then selectively displays only the **\_RMSE\_** variable from the newly created `outest` dataset, resulting in a minimal, focused output containing only the required metric.

```
/*fit simple linear regression model*/  
proc reg data=exam_data outest=outest noprint;  
model score = hours / rmse;  
run;  
quit;  
  
/*print RMSE of model*/  
proc print data=outest;  
var _RMSE_;  
run;
```

Obs	_RMSE_
1	3.64093

The final output clearly shows only the extracted RMSE value of **3.64093**, confirming successful implementation of the extraction method.

**A critical note on efficiency:** The argument **noprint** in **proc reg** is vital when processing large datasets or fitting numerous models iteratively. It instructs SAS not to generate the lengthy, detailed output, thereby conserving computational resources and streamlining the workflow to

produce only the targeted metric.

## Interpreting the Resulting RMSE

An RMSE of 3.64093, derived from our example, means that on average, the predicted exam score deviates from the actual exam score by approximately 3.64 points. Since the response variable (score) is measured on a scale of 0 to 100, this value provides a meaningful measure of predictive accuracy.

The interpretation of whether 3.64 is "good" or "bad" depends heavily on the context and the variability of the original data. If the scores ranged widely (e.g., from 30 to 95), an error of 3.64 might be considered excellent. However, if all scores clustered tightly between 75 and 85, an error of 3.64 would indicate poor predictive performance relative to the data's inherent low variability.

When comparing this **simple linear regression model** against other potential models--such as a polynomial regression model or a model using multiple predictors--the model yielding the lowest RMSE would be considered the most accurate in terms of average prediction error. This metric is thus essential for objective model selection.

## Conclusion

Calculating the Root Mean Square Error is a critical step in validating any statistical regression model. SAS offers robust procedures, particularly **PROC REG**, that simplify this calculation significantly. By understanding how to fit the model and use the **OUTEST** and **RMSE** options, analysts can efficiently extract this metric for comprehensive model evaluation and reporting.

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