

# What is Residual Variance? (Definition & Example)

Authored by  
**stats writer**

April 28, 2024

## RECOMMENDED CITATION

stats writer (2024). *What is Residual Variance? (Definition & Example)*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=140547>

Residual variance is a statistical measure that represents the amount of variation in a data set that is not accounted for by the regression model. It is calculated by taking the sum of the squared differences between the actual values and the predicted values from the regression line. In other words, it measures the difference between the observed data and the values predicted by the regression equation. A lower residual variance indicates a closer fit between the data and the regression model, while a higher residual variance suggests that the model is not accurately predicting the data. For example, if a regression model is used to predict the sales of a product, the residual variance would measure the variation in actual sales that is not explained by the model.

## **What is Residual Variance? (Definition & Example)**

**Residual variance (sometimes called "unexplained variance") refers to the variance in a model that cannot be explained by the variables in the model.**

**The higher the residual variance of a model, the less the model is able to explain the variation in the data.**

**Residual variance appears in the output of two different statistical models:**

- 1. ANOVA: Used to compare the means of three or more independent groups.**
- 2. Regression: Used to quantify the relationship between one or more predictor variables and a .**

**The following examples show how to interpret residual variance in each of these methods.**

## Residual Variance in ANOVA Models

Whenever we fit an ANOVA ("analysis of variance") model, we end up with an ANOVA table that looks like the following:

ANOVA

<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Between Groups	192.2	2	96.1	2.357532	0.113848	3.354131
Within Groups	1100.6	27	40.76296			
Total	1292.8	29				

The value for the residual variance of the ANOVA model can be found in the SS ("sum of squares") column for the Within Groups variation.

This value is also referred to as "sum of squared errors" and is calculated using the following formula:

$$\sum(X_{ij} - X_j)^2$$

where:

$\Sigma$ : a greek symbol that means "sum"  
 $X_{ij}$ : the  $i$ th observation in group  $j$   
 $X_j$ : the mean of group  $j$

In the ANOVA model above we see that the residual

variance is 1,100.6.

To determine if this residual variance is "high" we can calculate the mean sum of squared for within groups and mean sum of squared for between groups and find the ratio between the two, which results in the overall F-value in the ANOVA table.

$$F = MS_{\text{between}} / MS_{\text{within}} = 96.1 / 40.76296 = 2.357$$

This means we don't have sufficient evidence to say that the mean difference between the groups we're comparing is significantly different.

This tells us that the residual variance in the ANOVA model is high relative to the variation that the model actually can explain.

#### Residual Variance in Regression Models

In a regression model, the residual variance is defined as the sum of squared differences between predicted data points and observed data points.

It is calculated as:

$$\sum (y_i - \hat{y}_i)^2$$

where:

$\Sigma$ : a greek symbol that means "sum"  
 $\hat{y}_i$ : The predicted data points  
 $y_i$ : The observed data points

When we fit a regression model, we typically end up with output that looks like the following:

<b>Regression Statistics</b>					
Multiple R	0.98294208				
R Square	0.96617513				
Adjusted R	0.95651089				
Standard E	0.91826226				
Observatio	10				

  

<b>ANOVA</b>					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	2	168.5976	84.29878	99.97417	7.11748E-06
Residual	7	5.9024	0.843206		
Total	9	174.5			

  

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	17.1158537	1.171711716	14.60756	1.68E-06
x1	1.01829268	0.348522419	2.921742	0.022285
x2	0.39634146	0.325643032	1.217104	0.263001

The value for the residual variance of the ANOVA model can be found in the SS ("sum of squares") column for the Residual variation.

The ratio of the residual variation relative to the total

variation in the model tells us the percentage of variation in the response variable that can't be explained by the predictor variables in the model.

For example, in the table above we would calculate this percentage as:

**Unexplained variation = SS Residual / SS Total**  
**Unexplained variation = 5.9024 / 174.5**  
**Unexplained variation = .0338**

We can also calculate this value using the following formula:

**Unexplained variation = 1 - R<sup>2</sup>**  
**Unexplained variation = 1 - 0.96617**  
**Unexplained variation = .0338**

The R-squared value for the model tells us the percentage of variation in the response variable that can be explained by the predictor variable.

Thus, the lower the unexplained variation, the better a model is able to use the predictor variables to explain the variation in the response variable.