

# What is Explained Variance? (Definition & Example)

Authored by  
**stats writer**

June 28, 2024

## RECOMMENDED CITATION

stats writer (2024). *What is Explained Variance? (Definition & Example)*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=155822>

Explained variance refers to the proportion of variation in a dataset that can be accounted for by a particular variable or set of variables. It is a statistical measure that helps to understand how much of the variability in a dataset is explained by a specific factor or factors. In other words, it measures the extent to which a model or set of predictors can explain the observed outcomes.

For example, if a study is looking at the factors that influence students' grades, the explained variance would measure how much of the variation in students' grades can be explained by factors such as attendance, studying habits, and socioeconomic status. A high explained variance would indicate that these variables have a significant impact on students' grades, while a low explained variance would suggest that other factors may play a more significant role.

In summary, explained variance is a useful tool in understanding the relationship between variables and their impact on outcomes. It allows researchers to determine the strength of a model or set of predictors in explaining the variability in a dataset.

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

**Explained variance (sometimes called "explained variation") refers to the variance in the response variable in a model that can be explained by the predictor variable(s) in the model.**

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

**Explained 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 response variable.

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

**Note:** The opposite of explained variance is known as .

### Explained 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>
Explained Variance	Between Groups	192.2	2	96.1	2.357532	0.113848	3.354131
Unexplained Variance	Within Groups	1100.6	27	40.76296			
	Total	1292.8	29				

The explained variance can be found in the SS ("sum of squares") column for the .

In the ANOVA model above we see that the explained variance is 192.2.

To determine if this explained 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.

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

The F-value in the ANOVA table above is 2.357 and the corresponding p-value is 0.113848.

Since this p-value is not less than  $\alpha = .05$ , we do not have sufficient evidence to reject .

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

### Explained Variance in Regression Models

In a regression model, the explained variance is summarized by R-squared, often written  $R^2$ .

This value represents the proportion of the variance in the response variable that can be explained by the predictor variable(s) in the model.

The value for R-squared can range from 0 to where:

A value of 0 indicates that the response variable cannot be explained by the predictor variable(s) at all. A value of 1 indicates that the response variable can be perfectly explained without error by the predictor variable(s).

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

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

<i>ANOVA</i>					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
<b>Explained Variance</b>	Regression	168.5976	84.29878	99.97417	7.11748E-06
<b>Unexplained Variance</b>	Residual	5.9024	0.843206		
	Total	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

We can see that the explained variance is 168.5976 and

**the total variance is 174.5.**

**Using these values, we can calculate the R-squared value for this regression model as:**

**R-squared:  $SS \text{ Regression} / SS \text{ Total}$   
R-squared:  $168.5976 / 174.5$   
R-squared: 0.966**

**Since the R-squared value for this model is close to 1, it tells us that the explained variance in the model is extremely high.**

**In other words, the model is able to do a good job of using the predictor variables to explain the variation in the response variable.**

**Related:**