

# How to calculate Regression Sum of Squares (SSR)

Authored by  
**stats writer**

December 22, 2025

## RECOMMENDED CITATION

stats writer (2025). *How to calculate Regression Sum of Squares (SSR)*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=108312>

The Regression Sum of Squares (SSR) is a foundational metric in statistical regression analysis, providing crucial insight into the performance and explanatory power of a regression model. It quantifies the variation in the response variable that is successfully explained by the predictive relationship established in the model. Essentially, SSR measures how well the established regression line--often determined using the Least Squares Method--captures the underlying trends within the dataset. A higher SSR, relative to the total variation, generally indicates a better fit, suggesting that the predictor variables are effectively modeling the changes observed in the dependent variable. Understanding SSR is indispensable for model evaluation, particularly when comparing the efficacy of different models or assessing the overall contribution of the predictors.

Calculating the SSR involves comparing the variance captured by the model against the overall variance of the data. This metric forms the numerator of the R-squared statistic, which is perhaps the most widely recognized measure of goodness of fit in linear regression. By focusing solely on the deviation between the fitted values and the observed mean of the response variable, the SSR isolates the variability attributable to the relationship defined by the regression equation. This isolation allows statisticians and data analysts to determine precisely how much of the scatter in the data points is structured and predictable, as opposed to random or unexplained. It is a powerful tool for validating the core assumption that the independent variables genuinely influence the dependent variable in a measurable way.

Furthermore, the utility of the SSR extends beyond mere descriptive statistics; it is integral to performing F-tests for overall model significance. In this context, the ratio of the Mean Square Regression (MSR, derived from SSR) to the Mean Square Error (MSE) determines whether the model's explanatory power is statistically significant compared to a model with no predictors. Consequently, grasping the methodology behind calculating and interpreting the Regression Sum of Squares is fundamental to anyone performing rigorous quantitative analysis, ensuring that derived models are robust, reliable, and scientifically sound. This detailed guide will walk through the exact calculation process, contextualize its meaning, and provide tools for practical application.

## Understanding Regression Analysis Fundamentals

Regression analysis serves as the bedrock of predictive modeling in statistics, aiming to model the relationship between a dependent variable (the response) and one or more independent variables (the predictors). The primary goal is to create a line or plane of best fit that minimizes the distances, or residuals, between the observed data points and the predicted values. This line represents the systematic relationship inherent in the data. The success of this modeling effort is contingent upon how much of the total variation in the response variable can be systematically accounted for by the variations in the predictor variables. Metrics like SSR are essential because they provide a quantitative measure of this success.

In simple linear regression, the model attempts to fit a straight line, defined by the equation  $\hat{y} = a + bx$ , where  $\hat{y}$  represents the predicted value,  $x$  is the predictor variable,  $a$  is the intercept, and  $b$  is the slope coefficient. The slope coefficient,  $b$ , is critical, as it dictates the change in the response variable for a one-unit change in the predictor variable. The efficiency of the calculation of these coefficients relies heavily on minimizing the squared errors, which naturally leads to the derivation of the Sums of Squares metrics. Understanding the geometry of the regression line and its relationship to the observed data points is crucial before delving into the arithmetic of the SSR calculation, as the SSR fundamentally measures the distance between the predicted value ( $\hat{y}$ ) and the overall mean of the response variable ( $\bar{y}$ ).

Statistical theory mandates that every data point exhibits some degree of natural variation. This variation, or scatter, must be dissected into components that are explained by the model and those that remain unexplained. The explained component is precisely what the Regression Sum of Squares seeks to isolate. When a regression line perfectly models the data--a rare occurrence in real-world scenarios--the SSR would equal the total variation, signifying a perfect fit. Conversely, if the regression line explains none of the variation, the SSR would be zero. By providing a clear metric for the explained variance, SSR enables researchers to quantify the confidence they can place in their model's ability to forecast or explain phenomena within the scope of the analyzed data.

## Defining the Components of Variation

To accurately calculate the Regression Sum of Squares, one must first understand how the total observed variation in the dependent variable ( $Y$ ) is partitioned into its constituent parts. Total variation is measured by the Total Sum of Squares (SST). The SST measures the overall variation of the observed response values around their grand mean. Mathematically, it is the sum of the squared differences between each observed data point ( $y_i$ ) and the mean of all observed data points ( $\bar{y}$ ). This represents the benchmark of variability that the model attempts to explain.

The total variation (SST) is rigorously decomposed into two distinct components: the explained variation and the unexplained variation. The explained variation is captured by the SSR, which is the focus of this discussion. The SSR measures the variation between the value predicted by the regression line ( $\hat{y}_i$ ) and the grand mean of the response variable ( $\bar{y}$ ). When the regression line is far from the mean line, it indicates that the model is doing a good job of predicting values that are significantly different from the average, resulting in a large SSR. This large explained variation is what signifies a successful model fit.

The second component is the unexplained variation, known as the Residual Sum of Squares (SSE), sometimes referred to as the Error Sum of Squares. SSE measures the variation that the model fails to capture. It is calculated by summing the squared differences between the actual

observed data point ( $y_i$ ) and the predicted value from the regression line ( $\hat{y}_i$ ). These differences are the residuals, representing the vertical distance between the data point and the best-fit line. The fundamental relationship connecting these three metrics is arguably the most important identity in linear regression: **SST = SSR + SSE**. This identity confirms that the total variability of the data is perfectly accounted for by the combination of the variability explained by the model (SSR) and the variability left unexplained (SSE).

## The Formula for Regression Sum of Squares (SSR)

The calculation of the Regression Sum of Squares is derived directly from the components of variation explained above. The formula quantifies the degree to which the fitted regression line deviates from the null hypothesis model, which is simply a horizontal line at the mean of the response variable. The mathematical expression for SSR in its most general form is defined by the following summation, where  $\hat{y}_i$  denotes the predicted value for the  $i$ -th observation, and  $\bar{y}$  represents the mean of the response variable:

$$SSR = \sum (\hat{y}_i - \bar{y})^2$$

To execute this calculation manually for a set of  $N$  observations, one must first determine the mean of the response variable  $Y$ . Next, the regression line must be fitted to the data, typically using the Least Squares Method, to obtain the predicted values ( $\hat{y}_i$ ) for every corresponding predictor value ( $x_i$ ). For each data point, the difference between its predicted value ( $\hat{y}_i$ ) and the overall mean ( $\bar{y}$ ) is calculated. This difference is then squared to eliminate negative values and weight larger deviations more heavily. Finally, all these squared differences are summed across all observations to yield the total Regression Sum of Squares. This meticulous process ensures that the contribution of the predictor variables is accurately isolated.

It is important to note that while the core definition involves the predicted values, the SSR can also be calculated indirectly using the other sums of squares, provided they are already known. Since **SST = SSR + SSE**, if both the Total Sum of Squares (SST) and the Residual Sum of Squares (SSE) have been calculated, the SSR can be determined by subtraction: **SSR = SST - SSE**. This alternative method is often simpler in practice, especially when the SSE is already computed as part of the minimization process during model fitting. Regardless of the method used, the resulting SSR value represents a crucial measurement in evaluating the effectiveness of the statistical model constructed during the regression analysis.

```
@import url('https://fonts.googleapis.com/css?family=Droid+Serif|Raleway');
```

```
h1 {
text-align: center;
```

```
font-size: 50px;
margin-bottom: 0px;
font-family: 'Raleway', serif;
}

p {
color: black;
margin-bottom: 15px;
margin-top: 15px;
font-family: 'Raleway', sans-serif;
}

#words {
padding-left: 30px;
color: black;
font-family: Raleway;
max-width: 550px;
margin: 25px auto;
line-height: 1.75;
}

#words_summary {
padding-left: 70px;
color: black;
font-family: Raleway;
max-width: 550px;
margin: 25px auto;
line-height: 1.75;
}

#words_text {
color: black;
font-family: Raleway;
max-width: 550px;
margin: 25px auto;
line-height: 1.75;
}

#words_text_area {
display:inline-block;
color: black;
```

```
font-family: Raleway;  
max-width: 550px;  
margin: 25px auto;  
line-height: 1.75;  
padding-left: 100px;  
}
```

```
#calcTitle {  
text-align: center;  
font-size: 20px;  
margin-bottom: 0px;  
font-family: 'Raleway', serif;  
}
```

```
#hr_top {  
width: 30%;  
margin-bottom: 0px;  
border: none;  
height: 2px;  
color: black;  
background-color: black;  
}
```

```
#hr_bottom {  
width: 30%;  
margin-top: 15px;  
border: none;  
height: 2px;  
color: black;  
background-color: black;  
}
```

```
#words_table label, #words_table input {  
display: inline-block;  
vertical-align: baseline;  
width: 350px;  
}
```

```
#buttonCalc {  
border: 1px solid;  
border-radius: 10px;
```

```
margin-top: 20px;

cursor: pointer;
outline: none;
background-color: white;
color: black;
font-family: 'Work Sans', sans-serif;
border: 1px solid grey;
/* Green */
}

#buttonCalc:hover {
background-color: #f6f6f6;
border: 1px solid black;
}

#words_table {
color: black;
font-family: Raleway;
max-width: 350px;
margin: 25px auto;
line-height: 1.75;
}

#summary_table {
color: black;
font-family: Raleway;
max-width: 550px;
margin: 25px auto;
line-height: 1.75;
padding-left: 20px;
}

.label_radio {
text-align: center;
}

td, tr, th {
border: 1px solid black;
}

table {
```

```
border-collapse: collapse;
}
td, th {
min-width: 50px;
height: 21px;
}
.label_radio {
text-align: center;
}

#text_area_input {
padding-left: 35%;
float: left;
}

svg:not(:root) {
overflow: visible;
}
```

Utilize this interactive tool to efficiently calculate the Regression Sum of Squares (SSR) for a given dataset. This calculation is derived from the relationship between your independent (**predictor**) and dependent (**response**) variables, essential for evaluating model fit.

To begin the computation, please input your lists of corresponding data values--one for the predictor variable (X) and one for the response variable (Y)--into the respective text boxes below. Ensure the lists are comma-separated and of equal length before activating the calculation:

**Predictor Variable Values (X):**

6, 7, 7, 8, 12, 14, 15, 16, 16, 19

**Response Variable Values (Y):**

14, 15, 15, 17, 18, 18, 16, 14, 11, 8

Calculate SSR

**Calculated Regression Sum of Squares (SSR): 19.6122**

```
function calc() {

//get input data
```

```
var x = document.getElementById('x').value.split(',').map(Number);
var y = document.getElementById('y').value.split(',').map(Number);

//check that both lists are equal length
if (x.length - y.length == 0) {
document.getElementById('error_msg').innerHTML = "";

var xbar = math.mean(x);
var ybar = math.mean(y);

let xbar2_hold = 0
for (let i = 0; i < x.length; i++) {
xbar2_hold += Math.pow(x, 2);
}

var xbar2 = xbar2_hold / x.length;

let sxx = 0
for (let i = 0; i < x.length; i++) {
sxx += Math.pow(x - xbar, 2);
}

let syy = 0
for (let i = 0; i < y.length; i++) {
syy += Math.pow(y - ybar, 2);
}

let sxy = 0
for (let i = 0; i < x.length; i++) {
sxy += (x - xbar)*(y-ybar);
}

let sxx2 = 0
for (let i = 0; i < x.length; i++) {
sxx2 += (x - xbar)*(Math.pow(x, 2)-xbar2);
}

let sx2x2 = 0
for (let i = 0; i < x.length; i++) {
sx2x2 += Math.pow((Math.pow(x, 2)-xbar2), 2);
}
```

```

let sx2y = 0
for (let i = 0; i < x.length; i++) {
  sx2y += (Math.pow(x, 2)-xbar2)*(y-ybar);
}

var b = ((sxy*sx2x2)-(sx2y*sxx2)) / ((sxx*sx2x2)-Math.pow(sxx2, 2));
var c = ((sx2y*sxx)-(sxy*sxx2)) / ((sxx*sx2x2)-Math.pow(sxx2, 2));
var a = ybar - (b*xbar) - (c*xbar2);

var sst = syy;
var ssr = (sxy/sxx)*sxy;
var sse = sst-ssr;

document.getElementById('ssr').innerHTML = ssr.toFixed(4);
}

//output error message if boths lists are not equal
else {
//document.getElementById('out').innerHTML = "";
document.getElementById('error_msg').innerHTML = 'The two lists must be of equal length.';
}

} //end calc function

```

## Interpreting the Results of SSR

The magnitude of the Regression Sum of Squares provides immediate and critical insight into the quality of the regression model. A large SSR signifies that the predicted values (??) are substantially different from the overall mean (y). This large difference means the regression line has successfully captured a significant portion of the total variability inherent in the response data. In essence, the independent variable(s) are strong predictors of the dependent variable, and the model holds considerable explanatory power. Conversely, an SSR value close to zero indicates that the predicted values are very similar to the mean, suggesting the regression line is nearly horizontal and adds little explanatory benefit beyond simply using the average of the response variable as the prediction.

However, the SSR value itself is context-dependent and should not be interpreted in isolation. For instance, a high SSR in a dataset with inherently large total variation (high SST) might not be as impressive as a moderate SSR in a dataset with minimal total variation. To standardize this evaluation, the SSR is often used in conjunction with the Total Sum of Squares (SST) to calculate the Coefficient of Determination, or **R-squared (R<sup>2</sup>)**. The R<sup>2</sup> value is calculated as the ratio of SSR

to SST ( $R^2 = SSR / SST$ ), and it represents the proportion of the variance in the dependent variable that is predictable from the independent variables. This normalization allows for meaningful comparisons of model fit across diverse datasets.

Ultimately, a strong SSR indicates a tight fit between the model and the data, making the model more reliable for prediction and inference. If the SSR is low relative to the Residual Sum of Squares (SSE), it suggests that the majority of the variation remains unexplained. This scenario often prompts model refinement, such as adding more relevant predictor variables, transforming existing variables, or considering non-linear regression techniques. Interpreting SSR is thus an iterative process; it informs the analyst whether the current model formulation adequately addresses the underlying statistical relationships in the sample data.

### Relationship to SST and SSE: The Variance Decomposition

The foundational principle of variance decomposition in linear regression dictates that the total variation in the dependent variable must be perfectly split between the variation explained by the model (SSR) and the variation remaining as error (SSE). This identity,  **$SST = SSR + SSE$** , is more than just a convenient arithmetic relationship; it defines the statistical structure upon which the entire regression framework is built. Understanding this breakdown is key to appreciating how well the regression analysis performs relative to a baseline of no prediction.

The Total Sum of Squares (SST) represents the benchmark of variability. If we had no predictive variables, the best estimate for any new observation would simply be the grand mean ( $\bar{y}$ ), and the total error would be equal to the SST. The goal of fitting a regression model is to reduce this error. The SSR quantifies the reduction in error achieved by using the predicted values (??) instead of just the mean. The remaining error, the SSE, represents the irreducible residual error that even the best linear model cannot account for, potentially due to measurement noise, omitted variables, or inherent randomness.

When SSR is large and SSE is small, the model has successfully absorbed most of the total variation, signifying a strong fit. This relationship is formalized in the R-squared value,  $R^2 = 1 - (SSE / SST)$ . As the SSR increases, the SSE decreases proportionally, driving the R-squared value closer to 1 (or 100%), indicating that the model explains almost all the variability in the data. Conversely, if SSR is small compared to SSE, the  $R^2$  approaches 0, suggesting that the model is no better at predicting the outcome than simply guessing the average value. This constant interplay between the three sums of squares is vital for conducting formal hypothesis tests, such as the F-test, which compares the relative contribution of the explained variation (SSR) against the unexplained variation (SSE).

## Practical Applications and Model Comparison

The primary application of the Regression Sum of Squares lies in its role in model assessment and comparison. Because SSR directly relates to the explained variance, it provides a quantitative basis for determining the practical utility of a predictive model. In multiple regression, where analysts might test several combinations of predictor variables, comparing the SSR values--or more commonly, the resultant R-squared values--allows for an objective choice of the most efficacious model. A model that achieves a substantially higher SSR while maintaining a comparable level of complexity is generally preferred, as it demonstrates superior ability to capture the systematic patterns in the data.

Beyond simple R-squared calculation, the SSR is indispensable for computing the F-statistic in the Analysis of Variance (ANOVA) table associated with regression. The F-statistic tests the null hypothesis that all regression coefficients are zero, implying that the model has no explanatory power. The F-statistic is typically calculated as the ratio of the Mean Square Regression ( $MSR = SSR / df_R$ ) to the Mean Square Error ( $MSE = SSE / df_E$ ). A large F-statistic, driven by a large SSR relative to the SSE, provides statistical evidence to reject the null hypothesis, confirming that the overall regression model is statistically significant and useful for prediction.

Moreover, the concept of SSR is extended to sequential (or hierarchical) regression models, where researchers investigate the marginal contribution of adding new predictor variables. By comparing the SSR of a smaller model (Model A) to the SSR of a larger model (Model B, which includes all variables from A plus a new variable), one can perform an incremental F-test. The difference in the SSR values ( $SSR_B - SSR_A$ ) quantifies the unique contribution of the newly added variable to the explained variance. This technique is critical in fields like econometrics and psychology for determining the independent importance of various factors influencing an outcome.

## Conclusion: The Significance of SSR in Data Analysis

The Regression Sum of Squares (SSR) stands as a pillar of statistical evaluation in regression modeling, offering a precise metric for the goodness of fit achieved by a model. By isolating the variation in the dependent variable that is directly attributable to the predictor variables, SSR translates the success of the fitted regression line into a tangible, quantifiable value. This measurement is crucial for model validation, providing the necessary input for constructing the R-squared value and performing the rigorous F-tests required to prove the model's overall statistical significance. Without the SSR, assessing the explanatory power of a model would be subjective and incomplete.

Whether utilized directly in model optimization, indirectly through the calculation of R-squared, or as a key component in ANOVA tables, the SSR guides analysts toward more accurate and insightful conclusions. A strong understanding of its calculation--as the sum of squared deviations

between predicted values and the response mean--empowers data scientists to not only fit models but to critically evaluate their performance against established statistical benchmarks. As demonstrated by the interactive calculator above, this concept is highly practical and forms the basis for automated statistical software and complex machine learning algorithms alike.

In summation, mastering the calculation and interpretation of SSR is essential for moving beyond merely finding a line of best fit to truly understanding the causal or predictive relationships within a dataset. It ensures that any claims regarding the efficacy of a regression analysis are founded on robust, objective measures of explained variance, thereby enhancing the credibility and utility of the resulting predictive model across all areas of quantitative research.

ARABPSYCHOLOGY.COM