

# How can I use R to calculate the Sum of Squares Total (SST), Sum of Squares Regression (SSR), and Sum of Squares Error (SSE)?

Authored by  
**stats writer**

April 26, 2024

## RECOMMENDED CITATION

stats writer (2024). *How can I use R to calculate the Sum of Squares Total (SST), Sum of Squares Regression (SSR), and Sum of Squares Error (SSE)?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=139567>

R is a statistical programming language that can be used to calculate various statistical measures, including the Sum of Squares Total (SST), Sum of Squares Regression (SSR), and Sum of Squares Error (SSE). SST represents the total variation in a dataset, SSR represents the variation explained by a regression model, and SSE represents the variation that cannot be explained by the model. To calculate these measures using R, one can use the built-in functions such as "sum", "lm", and "anova". These functions take in the necessary input data and return the corresponding values for SST, SSR, and SSE. By understanding and utilizing these functions, one can effectively analyze and interpret the relationship between variables in a dataset.

## Calculate SST, SSR, and SSE in R

**We often use three different values to measure how well a actually fits a dataset:**

**1. Sum of Squares Total (SST) - The sum of squared differences between individual data points ( $y_i$ ) and the mean of the response variable ( $y$ ).**

$$SST = \sum(y_i - y)^2$$

**2. Sum of Squares Regression (SSR) - The sum of squared differences between predicted data points ( $\hat{y}_i$ ) and the mean of the response variable( $y$ ).**

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

**3. Sum of Squares Error (SSE) - The sum of squared differences between predicted data points ( $\hat{y}_i$ ) and observed data points ( $y_i$ ).**

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

The following step-by-step example shows how to calculate each of these metrics for a given regression model in R.

Step 1: Create the Data

First, let's create a dataset that contains the number of hours studied and exam score received for 20 different students at a certain college:

```
#create data frame
```

```
df <- data.frame(hours=c(1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 4, 4, 4, 5, 5, 6, 7, 7, 8),  
score=c(68, 76, 74, 80, 76, 78, 81, 84, 86, 83, 88, 85, 89, 94, 93, 94, 96, 89, 92, 97))
```

```
#view first six rows of data frame
```

```
head(df)
```

```
hours score
```

```
1 1 68
```

```
2 1 76
```

```
3 1 74
```

```
4 2 80
```

**5 2 76**

**6 2 78**

## Step 2: Fit a Regression Model

Next, we'll use the `lm()` function to fit a simple linear regression model using score as the and hours as the predictor variable:

```
#fit regression model  
model <- lm(score ~ hours, data = df)
```

```
#view model summary  
summary(model)
```

**Call:**

```
lm(formula = score ~ hours, data = df)
```

**Residuals:**

**Min 1Q Median 3Q Max**

**-8.6970 -2.5156 -0.0737 3.1100 7.5495**

**Coefficients:**

**Estimate Std. Error t value Pr(>|t|)**

**(Intercept) 73.4459 1.9147 38.360 < 2e-16 \*\*\***

**hours 3.2512 0.4603 7.063 1.38e-06 \*\*\***

---

**Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1**

**Residual standard error: 4.289 on 18 degrees of freedom**

**Multiple R-squared: 0.7348, Adjusted R-squared: 0.7201**

**F-statistic: 49.88 on 1 and 18 DF, p-value: 1.378e-06**

Step 3: Calculate SST, SSR, and SSE

**We can use the following syntax to calculate SST, SSR, and SSE:**

```
#find sse
```

```
sse <- sum((fitted(model) - df$score)^2)
```

```
sse
```

```
331.0749
```

```
#find ssr
```

```
ssr <- sum((fitted(model) - mean(df$score))^2)
```

```
ssr
```

```
917.4751
```

```
#find sst
```

```
sst <- ssr + sse
```

```
sst
```

```
1248.55
```

**Sum of Squares Total (SST): 1248.55**  
**Sum of Squares Regression (SSR): 917.4751**  
**Sum of Squares Error (SSE): 331.0749**

**We can verify that  $SST = SSR + SSE$ :**

```
SST = SSR + SSE  
1248.55 = 917.4751 + 331.0749
```

**We can also manually calculate the of the regression model:**

```
R-squared = SSR / SST  
R-squared = 917.4751 / 1248.55  
R-squared = 0.7348
```

**This tells us that 73.48% of the variation in exam scores can be explained by the number of hours studied.**

**You can use the following calculators to automatically calculate SST, SSR, and SSE for any simple linear regression line:**