

Calculate SST, SSR, and SSE in Excel

Authored by
Mohammed loot

November 5, 2025

RECOMMENDED CITATION

Mohammed loot (2025). *Calculate SST, SSR, and SSE in Excel*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=10836>

When undertaking the rigorous task of evaluating a statistical [regression model](#), analysts rely heavily on three core measures that meticulously quantify the agreement between the predicted outcomes and the observed data points. These metrics are essential because they systematically partition the overall variability inherent within the dataset, thereby offering critical, quantifiable insight into the effectiveness and explanatory power of the model used. Understanding the origin and calculation of these values is fundamental to mastering model assessment.

Defining the Three Sums of Squares

These three essential values--SST, SSR, and SSE--represent the core components utilized to assess the overall goodness of fit in any linear regression scenario. Each metric describes a distinct source of variance, allowing statisticians to determine how much variation is explained by the independent variables versus how much remains residual error.

1. Sum of Squares Total (SST) - This metric encapsulates the total variation present in the dependent, or response, variable (Y). It serves as the initial benchmark of variability before the application of the model. Mathematically, it is calculated as the sum of squared differences between each individual data point (y_i) and the overall mean of the response variable (\bar{y}). It represents the total amount of variation available to be explained.

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

2. Sum of Squares Regression (SSR) - Also frequently referred to as the explained sum of squares, the [SSR](#) measures the portion of the total variability that is successfully accounted for and explained by the [regression model](#). This value is derived by calculating the sum of squared differences between the predicted data points (\hat{y}_i) and the mean of the response variable (\bar{y}). A higher SSR relative to SST indicates a model with strong predictive capabilities.

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

3. Sum of Squares Error (SSE) - This metric, commonly known as the residual sum of squares, quantifies the unexplained variation, which is essentially the error or noise remaining in the model after fitting. The [SSE](#) represents the variability that the [regression model](#) was unable to capture or account for. It is computed as the sum of squared differences between the observed data points (y_i) and the corresponding predicted data points (\hat{y}_i). Minimizing SSE is the primary goal of the Ordinary Least Squares (OLS) method.

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

A thorough understanding of these components is critical, as they serve as the mathematical bedrock for calculating the [coefficient of determination \(R-squared\)](#), which is arguably the most widely used measure of linear model fit. The subsequent detailed, step-by-step example will

demonstrate exactly how to derive and calculate each of these metrics for a standard linear regression scenario using the robust analytical capabilities of **Microsoft Excel**.

The Fundamental Identity of Sums of Squares

A foundational principle in statistical modeling, particularly in the context of [Analysis of Variance \(ANOVA\)](#), is the inherent relationship shared among these three sums of squares. This identity states that the total variability observed in the data ([SST](#)) must always be precisely equal to the sum of the variability explained by the fitted model ([SSR](#)) and the remaining unexplained variability ([SSE](#)). This relationship provides a crucial mathematical check, confirming that the regression analysis has successfully and completely partitioned the data's overall variance into two distinct, additive components.

The identity is expressed succinctly as: **SST = SSR + SSE**. This straightforward equation forms the algebraic foundation not only for understanding variance decomposition but also for efficiently calculating key model fit statistics, most notably the [R-squared](#) value. If the results generated by Excel adhere to this fundamental identity, it provides immediate verification of the internal consistency and accuracy of the regression analysis output.

For practical calculation, we will leverage Excel's powerful, built-in **Data Analysis ToolPak**. This essential add-in automates the otherwise complex matrix algebra involved in linear regression and conveniently compiles these three critical values--SST, SSR, and SSE--directly within the resulting [ANOVA](#) summary table.

Step 1: Preparing the Dataset in Excel

To properly execute our demonstration, the first necessary action is to organize a dataset that is suitable for performing a standard linear regression analysis. Our specific goal is to examine the relationship between the number of hours a student dedicates to studying (which serves as the independent variable, X) and the corresponding score they achieve on a subsequent examination (the dependent variable, Y). The sample data we will use is based on observations collected from 20 distinct students.

Accurate data input is the single most crucial preparatory step. It is essential to ensure that your independent variables (X) and dependent variables (Y) are placed in adjacent columns in the spreadsheet. This arrangement is specifically required to facilitate seamless processing and range selection when utilizing the **Data Analysis ToolPak** in the following steps.

	A	B	C	D	E	F	G
1	Hours	Score					
2	1	68					
3	1	76					
4	1	74					
5	2	80					
6	2	76					
7	2	78					
8	2	81					
9	2	84					
10	3	86					
11	3	83					
12	3	88					
13	4	85					
14	4	89					
15	4	94					
16	5	93					
17	5	94					
18	6	96					
19	7	89					
20	7	92					
21	8	97					
22							
23							
24							
25							
26							
27							
28							

Step 2: Executing the Linear Regression Analysis

With the data properly structured, the next phase involves utilizing Excel's statistical computation features to run the formal [linear regression](#) analysis. This automated procedure is essential because it generates the underlying predicted values and residuals that are mathematically required for the derivation of the sums of squares metrics.

To begin, navigate to the main ribbon interface in Excel and click on the **Data** tab. Within the Data ribbon, locate and click the **Data Analysis** button, which is typically situated on the far right side of the options. A critical reminder: if this option is not immediately visible, you must first install and activate the **Data Analysis ToolPak**, which functions as a required add-in for all advanced statistical operations within Excel.



A new dialog box will pop up after clicking **Data Analysis**. Scroll through the list of available statistical tests, select **Regression** from the menu, and then click OK to proceed to the detailed regression setup window. The next step involves carefully defining the input and output parameters for the calculation.

In the subsequent Regression dialog box, specify the following required ranges:

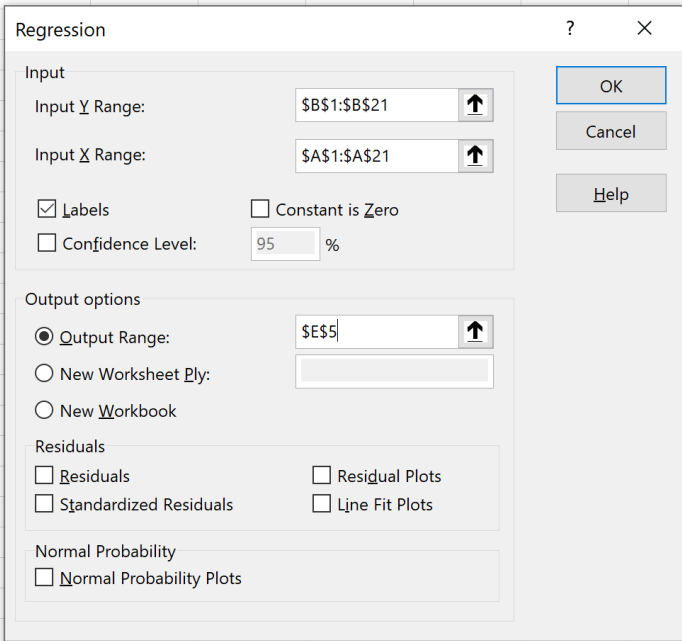
Input Y Range: Select the entire column range containing the exam scores (the dependent variable, Y).

Input X Range: Select the entire column range containing the hours studied (the independent variable, X).

Crucially, ensure the **Labels** checkbox is marked if you included the descriptive column headers in your initial range selections.

Designate an **Output Range** (e.g., cell D1) where you wish the resulting statistical tables to be displayed, ensuring it is a blank area.

	A	B	C	D	E	F	G	H	I
1	Hours	Score							
2	1	68							
3	1	76							
4	1	74							
5	2	80							
6	2	76							
7	2	78							
8	2	81							
9	2	84							
10	3	86							
11	3	83							
12	3	88							
13	4	85							
14	4	89							
15	4	94							
16	5	93							
17	5	94							
18	6	96							
19	7	89							
20	7	92							
21	8	97							
22									
23									
24									
25									
26									
27									



The image shows the 'Regression' dialog box in Microsoft Excel. The dialog box is overlaid on the spreadsheet data. It has a title bar with a question mark and a close button. The 'Input' section contains 'Input Y Range' set to '\$B\$1:\$B\$21' and 'Input X Range' set to '\$A\$1:\$A\$21'. There are checkboxes for 'Labels' (checked), 'Constant is Zero' (unchecked), and 'Confidence Level' (set to 95%). The 'Output options' section has 'Output Range' set to '\$E\$5' (selected with a radio button), and options for 'New Worksheet Ply' and 'New Workbook' are unselected. The 'Residuals' section has checkboxes for 'Residuals' (unchecked), 'Standardized Residuals' (unchecked), 'Residual Plots' (unchecked), and 'Line Fit Plots' (unchecked). The 'Normal Probability' section has a checkbox for 'Normal Probability Plots' (unchecked). Buttons for 'OK', 'Cancel', and 'Help' are on the right side.

Upon confirming these specific settings and clicking OK, Excel will instantly generate a comprehensive output sheet. This output details the complete results of the [linear regression](#) analysis, presenting several tables that are foundational for identifying our target variance metrics.

E	F	G	H	I	J
SUMMARY OUTPUT					
<i>Regression Statistics</i>					
Multiple R	0.857224				
R Square	0.734832				
Adjusted R Square	0.720101				
Standard Error	4.288712				
Observations	20				
ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	917.4751	917.4751	49.88162	1.37823E-06
Residual	18	331.0749	18.39305		
Total	19	1248.55			
<i>Coefficients and Standard Error</i>					
	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>		
Intercept	73.44585	1.914653	38.35988	1.03E-18	69.42331605
Hours	3.251152	0.460328	7.062692	1.38E-06	2.284039699

Step 3: Extracting SST, SSR, and SSE from the ANOVA Table

The most efficient and direct way to identify the three sums of squares metrics is to look within the **ANOVA** (Analysis of Variance) table, which is a key component of the regression output. This table is specifically designed to organize and present the decomposition of total variance within the model.

Focus your attention specifically on the column labeled **SS** (Sum of Squares). The descriptive rows in this section correspond directly and unambiguously to the required statistical metrics:

The row labeled **Regression** provides the crucial **Sum of Squares Regression (SSR)**, representing the explained variance.

The row labeled **Residual** (sometimes listed as Error) yields the **Sum of Squares Error (SSE)**, quantifying the unexplained variance.

The final row, labeled **Total**, presents the **Sum of Squares Total (SST)**, which is the baseline variability.

The image below clearly highlights the calculated values derived from our sample student data, showing how the total variability is broken down into its two constituent parts: explained and unexplained variance.

E	F	G	H	I	J
SUMMARY OUTPUT					
<i>Regression Statistics</i>					
Multiple R	0.857224				
R Square	0.734832				
Adjusted R Square	0.720101				
Standard Error	4.288712				
Observations	20				
<i>ANOVA</i>					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	917.4751	917.4751	49.88162	1.37823E-06
Residual	18	331.0749	18.39305		
Total	19	1248.55			
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>
Intercept	73.44585	1.914653	38.35988	1.03E-18	69.42331605
Hours	3.251152	0.460328	7.062692	1.38E-06	2.284039699

By referencing the SS column of the [ANOVA](#) table, we accurately extract the following numerical metrics for our specific dataset:

Sum of Squares Total (SST): 1248.55

Sum of Squares Regression (SSR): 917.4751

Sum of Squares Error (SSE): 331.0749

Step 4: Verification and Calculation of R-squared

Before moving to the final interpretation, it is standard practice--and a vital quality assurance step--to rigorously verify the fundamental algebraic identity: **SST = SSR + SSE**. Successfully satisfying this equation provides strong assurance regarding the internal consistency and accuracy of the results provided by Excel.

$$SST = SSR + SSE$$

$$1248.55 = 917.4751 + 331.0749$$

$$1248.55 \approx 1248.55$$

Since the identity holds true, we can now confidently proceed to calculate the [coefficient of determination](#), or [R-squared](#). This resulting metric is often the most accessible and widely cited measure of model performance, as it indicates the exact proportion of the variance in the

dependent variable (Y) that is successfully predictable by the independent variable(s) (X).

The formula for R-squared is derived directly from the relationship between the explained and total sums of squares:

$$R\text{-squared} = SSR / SST$$

$$R\text{-squared} = 917.4751 / 1248.55$$

$$R\text{-squared} = 0.7348$$

This final calculation reveals a value of **0.7348**, meaning that **73.48%** of the total variation observed in the students' exam scores can be effectively explained and predicted solely by the number of hours they spent studying. This high percentage strongly suggests a robust and well-fitting [linear regression model](#). Conversely, the remaining 26.52% of the variation is attributed to unexplained factors or random noise, as captured by the SSE metric.

Conclusion and Next Steps

Mastering the calculation and interpretation of SST, SSR, and SSE is indispensable for anyone performing serious statistical analysis. These metrics not only allow for the derivation of the powerful R-squared statistic but also offer a granular view into where the variance in your data originates--whether it is explained by the model or remains residual error. Utilizing the **Data Analysis ToolPak** in Excel provides a streamlined and reliable method for obtaining these values efficiently.

For further reading and exploring advanced applications of these core statistical concepts, consult authoritative resources focused on econometric modeling and inferential statistics.