

How do you calculate SST, SSR, and SSE in Excel?

Authored by
stats writer

April 26, 2024

RECOMMENDED CITATION

stats writer (2024). *How do you calculate SST, SSR, and SSE in Excel?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=139559>

The process of calculating SST, SSR, and SSE in Excel involves utilizing the built-in statistical functions and formulas provided by the software. SST (Total Sum of Squares) represents the total variation in a set of data, SSR (Regression Sum of Squares) measures the amount of variation explained by the regression model, and SSE (Error Sum of Squares) reflects the unexplained or residual variation. To calculate these measures, one must first input the data into an Excel spreadsheet and then use the appropriate functions, such as SUM, SUMSQ, and COUNT, to calculate the necessary values. Once these values are obtained, the formulas for SST, SSR, and SSE can be applied to calculate the respective sums. This process allows for a comprehensive analysis of the variation in a set of data and is a useful tool for statistical analysis and modeling.

Calculate SST, SSR, and SSE in Excel

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).

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

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).

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

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

observed data points (y_i).

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

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

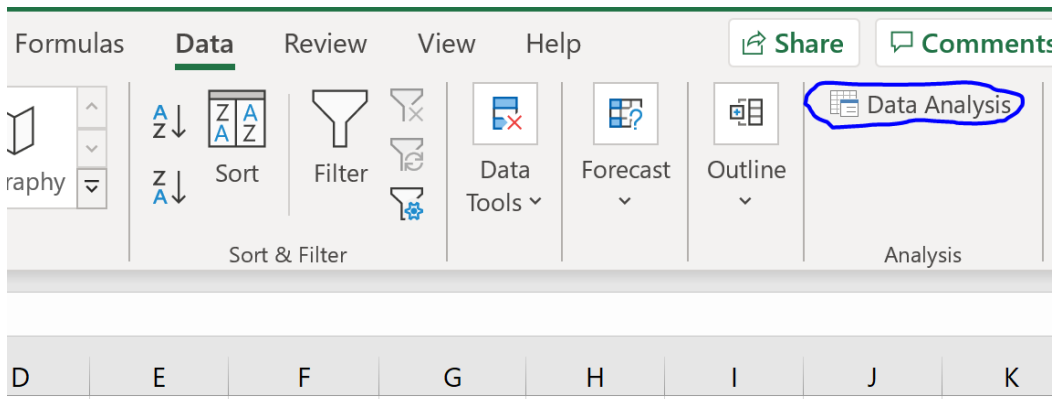
Step 1: Create the Data

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

| | 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: Fit a Regression Model

Along the top ribbon in Excel, click the Data tab and click on Data Analysis. If you don't see this option, then you need to first .



Once you click on Data Analysis, a new window will pop up. Select Regression and click OK.

In the new window that appears, fill in the following information:

| | 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 | | | | | | | | | |

Regression

Input

Input Y Range:

Input X Range:

Labels Constant is Zero

Confidence Level: %

Output options

Output Range:

New Worksheet Ply:

New Workbook

Residuals

Residuals Residual Plots

Standardized Residuals Line Fit Plots

Normal Probability

Normal Probability Plots

OK

Cancel

Help

| 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</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 |

Step 3: Analyze the Output

The three sum of squares metrics - SST, SSR, and SSE - can be seen in the SS column of the ANOVA table:

| 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 |

The metrics turn out to be:

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 $SST = SSR + SSE$:

$$SST = SSR + SSE \quad 1248.55 = 917.4751 + 331.0749$$

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.

ARABPSYCHOLOGY.COM