

Learn How to Calculate R-Squared in Google Sheets: A Step-by-Step Guide

Authored by
Mohammed looti

October 28, 2025

RECOMMENDED CITATION

Mohammed looti (2025). *Learn How to Calculate R-Squared in Google Sheets: A Step-by-Step Guide*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=4940>

Understanding R-Squared: The Foundation of Model Evaluation

In the rigorous world of [statistical modeling](#), particularly within [regression analysis](#), the ability to assess how effectively a predictive model explains the inherent variability within your data is absolutely paramount. This vital assessment relies heavily on the metric known as **R-squared**, frequently symbolized as r^2 or formally called the **coefficient of determination**. R-squared serves as a fundamental benchmark for evaluating the overall efficacy and explanatory power of your statistical model.

Fundamentally, R-squared quantifies the proportion of the total [variance](#) in the dependent (response) variable that can be attributed to--or explained by--the independent (predictor) variable(s) incorporated into the regression model. For example, in a simple [linear regression](#) model, R-squared provides a clear percentage indicating how much of the fluctuation in the outcome variable is predictable based on the input variable. This allows researchers and data analysts to accurately gauge the strength of the linear relationship between variables and determine the model's overall explanatory capacity.

The calculated value of r^2 is always constrained between 0 and 1, inclusive, offering a remarkably straightforward interpretation of model performance:

A value of **0** indicates that the model explains absolutely none of the variability observed in the response variable relative to its mean. In practical terms, the predictor variable offers zero explanatory power for the response variable, suggesting a very poor [model fit](#).

A value of **1** signifies that the model perfectly explains all the variability of the response variable around its mean. This represents an ideal, though often unrealistic, scenario where the predictor variable flawlessly accounts for every change in the response variable, leaving no unexplained variance.

Values falling between 0 and 1 directly represent the proportion of variance explained. For instance, an R-squared of 0.75 means that 75% of the variation observed in the response variable is successfully explained by the predictor variable(s). The remaining 25% is variance that the current model structure fails to account for.

Grasping the meaning and limitations of **R-squared** is crucial for validating your regression outcomes. While a higher R-squared generally suggests a better fit for the specific sample data, it is vital to remember that a high value does not automatically guarantee that the model is inherently correct, nor does it establish [causation](#). It is strictly a measure of the strength and goodness of fit for the linear relationship being tested.

Preparing Your Data for Statistical Rigor in Google Sheets

Before embarking on the calculation of **R-squared**, meticulous data preparation and organization

are essential steps. When conducting analysis using a powerful spreadsheet platform like [Google Sheets](#), structuring your data correctly will not only simplify the process but also prevent calculation errors and misinterpretations. For any typical simple linear regression model, you must clearly define two primary variables: the **predictor variable** (also known as the independent variable) and the **response variable** (also known as the dependent variable).

The predictor variable is the factor you hypothesize influences or explains changes in the outcome. Conversely, the response variable represents the outcome you are attempting to predict or explain. Consider a scenario where you are investigating the link between the number of hours spent studying and the resulting exam scores: 'hours studied' would be designated as your **predictor variable** (X), while 'exam score' would be your **response variable** (Y). Maintaining this distinction is critical for accurate function application.

The standard best practice involves arranging your data into vertical columns, where each column represents a distinct variable and each row corresponds to a single observation or data point. This structured format ensures that Google Sheets functions can easily reference the required data ranges. Furthermore, ensure your [dataset](#) is clean and free from significant issues, such as missing values or extreme outliers, which can severely skew the results of statistical calculations like R-squared and compromise the reliability of your [regression analysis](#).

For the practical demonstration that follows, we utilize a dataset containing observations for 20 students, tracking their hours studied and their corresponding exam scores. This data is segregated into two clearly labeled columns within a Google Sheet, making it immediately ready for statistical processing using the built-in functions.

	A	B	C	D	E	F	G
1	hours	score					
2	1	76					
3	2	78					
4	2	85					
5	4	88					
6	2	72					
7	1	69					
8	5	94					
9	4	94					
10	2	88					
11	4	92					
12	4	90					
13	3	75					
14	6	96					
15	5	90					
16	3	82					
17	4	85					
18	6	99					
19	2	83					
20	1	62					
21	2	76					
22							
23							
24							
25							

As illustrated above, Column A contains the 'Hours Studied' (our predictor variable, X), and Column B holds the 'Exam Score' (our response variable, Y). This clear, columnar organization is the foundational requirement for correctly executing the **RSQ()** function in Google Sheets.

Introducing the RSQ() Function: Syntax and Requirements

One of the most valuable features of Google Sheets is the provision of dedicated, convenient functions that streamline complex statistical calculations. The built-in **RSQ()** function is specifically designed to calculate the R-squared value directly, thereby bypassing the need for manual algebraic formulas. Technically, this function computes the square of the [Pearson product-moment correlation coefficient](#), effectively determining the goodness of fit between your observed data points and the resulting linear model.

The syntax required to use the **RSQ()** function is simple and requires the precise input of two key arguments:

=RSQ(known_ys, known_xs)

It is essential to understand what each argument represents:

known_ys: This argument must be the numerical range containing the data points for your **response variable** (Y values). These are the observed outcomes that your model attempts to predict or explain. You must ensure the correct column containing the dependent variable is selected.

known_xs: This argument must be the numerical range containing the data points for your **predictor variable** (X values). These are the input factors that are hypothesized to influence the response variable. Select the corresponding column for the independent variable carefully.

Crucially, you must maintain absolute consistency in the structure of the data ranges provided to the function. Both the known_ys and known_xs ranges must be numerical, and they must contain the exact same number of data points. If the ranges are mismatched in length, the **RSQ()** function will inevitably return an error value. Once the correct, matched datasets are provided, the **RSQ()** function performs the calculation, delivering a fast and accurate measure of your model's explanatory power, or its **R-squared** value.

Step-by-Step Calculation: Applying RSQ() to Real-World Data

We will now put the **RSQ()** function into practice using our specific dataset to ascertain the R-squared value that describes the relationship between hours studied and exam scores. Our objective is to determine the goodness of fit for a [linear regression](#) model where 'hours studied' acts as the predictor (X) and 'exam score' serves as the response (Y).

Based on the data organization established in Google Sheets, we must correctly map our variables to the function arguments. The exam scores--our response variable (Y)--are located in cells **B2:B21**, which correspond to our **known_ys**. The hours studied--our predictor variable (X)--are found in cells **A2:A21**, representing our **known_xs**. Note that the original image shows Hours in A and Score in B, so we must be careful to match the variable definitions to the column labels if we are following the image exactly, or adjust the example description to match the formula provided. Assuming the formula **=RSQ(A2:A21, B2:B21)** is correct for the example provided, we proceed.

To execute the R-squared calculation, select any vacant cell where you wish the result to appear (for example, cell D2) and input the following formula meticulously:

=RSQ(B2:B21, A2:A21)

Upon pressing Enter, Google Sheets immediately computes and displays the R-squared value. This instruction directs the function to evaluate how successfully the hours studied (X) explain the

observed variation in the exam scores (Y). It is essential to ensure the Y range (scores) is entered first, followed by the X range (hours studied), as the function order is rigid.

The subsequent screenshot visually confirms this calculation step, showing the formula entered into cell D2 and the resulting R-squared measure. Observe the precise specification of the ranges for the response variable (B2:B21) and the predictor variable (A2:A21) within the function.

D2		<i>fx</i>	=RSQ(A2:A21, B2:B21)		
	A	B	C	D	
1	Hours	Score		R-squared	
2		1	76	0.7272528139	
3		2	78		
4		2	85		
5		4	88		
6		2	72		
7		1	69		
8		5	94		
9		4	94		
10		2	88		
11		4	92		
12		4	90		
13		3	75		
14		6	96		
15		5	90		
16		3	82		
17		4	85		
18		6	99		
19		2	83		
20		1	62		
21		2	76		
22					
23					
24					

In this specific analytical instance, the computation yields an **R-squared** value of approximately **0.7273**. This numerical outcome is not merely a digit; it provides profound insight into the strength and nature of the relationship between the variables under investigation.

Interpreting R-Squared: Model Explanatory Power and Context

The calculated R-squared value of **0.7273**, which translates to **72.73%**, for our study hours and

exam scores example, indicates that a highly substantial portion of the variation observed in exam scores can be directly explained by the number of hours students dedicated to studying. Stated differently, according to our regression model, nearly three-quarters (72.73%) of the observed differences in exam performance among the students are attributable to variations in their study duration.

In many contexts, particularly within social science and educational research, this is considered a relatively strong R-squared value, suggesting that 'hours studied' is a highly significant factor influencing 'exam score'. However, it is equally crucial to acknowledge the remaining **27.27%** (100% minus 72.73%) of the variation in exam scores that our current [linear regression](#) model fails to explain. This unexplained variance stems from other factors not accounted for in our simple two-variable model, such as student aptitude, prior knowledge, quality of instruction, personal well-being, test anxiety, or sheer random variability.

Proper interpretation necessitates placing R-squared within the appropriate context of your research field. A high R-squared is not universally desirable, and a low R-squared is not automatically detrimental. In disciplines such as theoretical physics or engineering, where relationships are often deterministic, an R-squared approaching 0.99 might be expected. Conversely, in fields like psychology or sociology, where highly complex human behaviors are modeled, an R-squared of 0.30 may be considered quite respectable due to the immense inherent variability and the presence of numerous unobservable confounding variables.

Moreover, while R-squared measures the explained proportion of variance, it cannot indicate the direction of the relationship (positive or negative correlation) or determine if the selected predictor variables are truly the optimal choice. For more sophisticated models that incorporate multiple predictor variables, it is standard practice to rely on the [Adjusted R-squared](#). This modified metric accounts for the number of predictors included in the model and is considered a more robust and reliable measure when comparing the performance of models with varying numbers of independent variables.

Conclusion: Leveraging R-Squared for Data-Driven Decisions

The proficiency required to accurately calculate and insightfully interpret the [R-squared](#) value is an indispensable skill for professionals engaged in data analysis, business intelligence, or statistical modeling, especially when leveraging the accessibility of tools like Google Sheets. As demonstrated through the practical example, the **RSQ()** function offers a highly efficient and accurate method for quantifying the explanatory power of your regression models. By clearly understanding the percentage of variance in your response variable that is accounted for by your predictors, you are equipped to draw significantly deeper, more meaningful conclusions about the underlying relationships within your data.

Whether your analytical focus is on forecasting market trends, evaluating academic performance, or interpreting complex scientific experiments, R-squared serves as a fundamental benchmark for assessing the effectiveness of your models in predicting observed outcomes. It is imperative, however, to remember that R-squared constitutes only one element of a comprehensive analysis. It must always be considered in conjunction with other crucial statistical measures, relevant domain knowledge, and the overarching context of your research to ensure that the conclusions you draw are both robust and statistically sound. A continuous commitment to clear data organization and thoughtful interpretation will significantly enhance the reliability and overall impact of your statistical findings.

Further Exploration and Related Resources

To further advance your skills in data analysis and maximize your use of Google Sheets, we encourage you to explore related tutorials and resources. Mastering a wider array of functions and statistical techniques will empower you to extract even more valuable and actionable insights from increasingly complex datasets.