

Calculating Least Squares Regression: A Step-by-Step Guide Using Google Sheets

Authored by
Mohammed looti

November 11, 2025

RECOMMENDED CITATION

Mohammed looti (2025). *Calculating Least Squares Regression: A Step-by-Step Guide Using Google Sheets*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=16808>

The [method of least squares](#) stands as a cornerstone technique in statistics, providing a systematic approach to finding the optimal linear relationship within a dataset. Its primary goal is to derive the line of best fit--often referred to as the [regression line](#)--by minimizing the cumulative sum of the squared vertical distances between the observed data points and the line itself. This minimization process, which focuses on the residuals (errors), ensures that the resulting model is the most statistically sound representation of the linear trend, making it invaluable across scientific, economic, and engineering disciplines.

While dedicated statistical packages are often used for complex modeling, the power of modern spreadsheet software allows for efficient implementation of this sophisticated calculation. Specifically, [Google Sheets](#) offers a robust, built-in function that simplifies the entire process. This function, known as **LINEST()** (short for LINEar STatistics), is meticulously designed to calculate the critical parameters required to define the best-fit straight line based on the least squares principle.

The following video provides a brief explanation of this method:

This comprehensive tutorial will guide you through a practical, step-by-step example, detailing precisely how to leverage the **LINEST()** function within Google Sheets to derive, interpret, and visualize the key parameters of your best-fit regression line.

Understanding Linear Regression and the Least Squares Principle

Before executing any calculations in a spreadsheet, it is essential to solidify the conceptual foundation of what the [method of least squares](#) is intended to accomplish. Simple linear regression aims to define a predictable linear relationship between two variables: an independent predictor variable (X) and a dependent response variable (Y). This relationship is mathematically represented by the classic linear equation: $Y = mX + b$, where 'm' is the slope (determining the steepness and direction) and 'b' is the Y-intercept (the point where the line crosses the Y-axis).

The core difficulty in real-world data analysis is determining the unique values for 'm' and 'b' that genuinely capture the underlying trend across all collected data points, which are invariably scattered due to measurement error or natural variance. The least squares approach resolves this ambiguity by utilizing a mathematical optimization technique. It systematically searches for the line parameters ('m' and 'b') that result in the smallest possible sum of the squared differences (errors) between the actual observed Y values and the Y values predicted by the resulting linear equation.

The process of squaring the errors is critical; it ensures that large deviations (outliers) are penalized significantly more than small ones, thereby preventing positive and negative residuals from canceling each other out and ultimately yielding a statistically optimal and highly robust linear fit. In the context of [Google Sheets](#), the **LINEST()** function handles all this complex matrix algebra

behind the scenes, providing the necessary slope and intercept in its basic form, and offering advanced statistical metrics like the R-squared value and standard errors when used in its extended form.

Step 1: Preparing and Entering the Data in Google Sheets

The successful execution of the least squares calculation hinges entirely on the correct preparation and organization of your raw data within the spreadsheet environment. The **LINEST()** function has specific requirements regarding the arrangement of the known variables. Your dataset must be structured into two adjacent columns: one dedicated to the known dependent variables (Y values, which you are trying to predict) and one for the known independent variables (X values, the predictors).

For our practical example, we will structure a representative dataset. Following conventional spreadsheet practice, we will place the X values (the predictor variable) in Column A and the corresponding Y values (the response variable) in Column B. This specific order is mandatory because the input arguments for the **LINEST()** function explicitly require the range containing the Y values first, immediately followed by the range containing the X values.

The illustration below displays the organized dataset we will use for this simple linear regression analysis. Note the clear separation and labeling of the X and Y variables:

	A ▼	B	C	D
1	x	y		
2	2	12		
3	3	13		
4	3	15		
5	4	19		
6	6	18		
7	7	19		
8	8	20		
9	10	24		
10	12	19		
11	13	26		
12	13	28		
13	14	31		
14	19	29		
15	22	34		
16	24	39		
17				
18				

It is vital to ensure that your data is meticulously clean, complete, and perfectly aligned, spanning consistently from row 2 to row 16. Any misalignments, missing entries, or errors in the data entry process will inevitably propagate through the calculation, resulting in inaccurate regression [coefficients](#) and a fundamentally flawed resulting regression model.

Step 2: Utilizing the LINEST() Function for Regression Calculation

The core computational step in this process is the application of the [LINEST\(\)](#) function, which serves as the dedicated tool for executing the [method of least squares](#) calculation directly within Google Sheets. The fundamental syntax required to retrieve the two defining characteristics of the line--the slope (m) and the intercept (b)--is straightforward: `=LINEST(known_ys, known_xs)`.

A key characteristic of the **LINEST()** function that users must understand is its designation as an array function. This means it is designed not to return a single value, but rather multiple results that spill over into adjacent cells simultaneously. For a simple linear regression ($Y = mX + b$), the function requires two horizontal cells to display its primary output: the slope (m) is returned into the cell where the formula is initially entered, and the y-intercept (b) is automatically placed in the cell immediately to the right.

To fit a [regression line](#) to our specific dataset, where the Y values are contained in the range

B2:B16 and the X values are in the range A2:A16, we must enter the following formula into a single, unoccupied starting cell, such as **D1**:

=LINEST(B2:B16, A2:A16)

Once this formula is confirmed, [Google Sheets](#) instantaneously calculates the parameters and populates the results. Cell D1 will display the slope coefficient (1.07949), and the neighboring cell E1 will automatically be filled with the y-intercept coefficient (11.55211). The subsequent screenshot clearly illustrates this output, showing the two primary parameters of the fitted line occupying cells D1 and E1:

D1	fx =LINEST(B2:B16, A2:A16)				
	A	B	C	D	E
1	x	y		1.079489696	11.55210991
2	2	12			
3	3	13			
4	3	15			
5	4	19			
6	6	18			
7	7	19			
8	8	20			
9	10	24			
10	12	19			
11	13	26			
12	13	28			
13	14	31			
14	19	29			
15	22	34			
16	24	39			
17					
18					
19					

These numerical results represent the calculated regression [coefficients](#) that define the precise mathematical relationship derived from the observed data, rigorously determined by applying the least squares criterion.

Step 3: Interpreting the Regression Coefficients and Making Predictions

The output generated by the [LINEST\(\)](#) function provides all the essential numerical components

required to formally state the equation of the fitted [regression line](#). Based on the values we derived in Step 2 (Slope 'm' ≈ 1.07949 and Y-intercept 'b' ≈ 11.55211), we can now construct the specific prediction equation tailored for our dataset:

Fitted Regression Line Equation: $Y = 1.07949 * X + 11.55211$

This equation is the primary output of the analysis and allows for meaningful interpretation of the relationship between our variables in a practical context:

The Y-intercept (11.55211) represents the estimated baseline value of the dependent variable Y at the point where the independent variable X is equal to zero.

The Slope (1.07949) quantifies the rate of change: it indicates that for every one-unit increase in the independent variable X, the estimated value of the dependent variable Y increases by approximately 1.07949 units.

The most immediate and powerful practical application of this fitted equation is forecasting. We can use the statistical model to reliably estimate or predict the value of the dependent variable (Y) for any new, unobserved value of the independent variable (X). For instance, if the goal is to predict the expected Y value when X is equal to 10, we simply substitute $X = 10$ into our precise fitted equation, derived using the [method of least squares](#):

$$y = 11.55211 + 1.07949(10)$$

$$y = 11.55211 + 10.7949$$

$$y = \mathbf{22.34701}$$

Consequently, based on the robust statistical model derived from our data, we estimate that the value of y is approximately 22.347 when x is 10. This robust predictive capability underscores the immense utility and necessity of regression analysis in data-driven decision-making.

Step 4: Visualizing the Results with a Scatter Plot and Trendline

While the calculated regression [coefficients](#) provide the exact mathematical definition of the line of best fit, visualizing the results is an absolutely vital step. Visualization confirms the appropriateness of the linear assumption and serves as a powerful tool for communicating the findings to a broader audience. We can easily plot the original observed data points and overlay the fitted regression line directly within Google Sheets to complete our analysis.

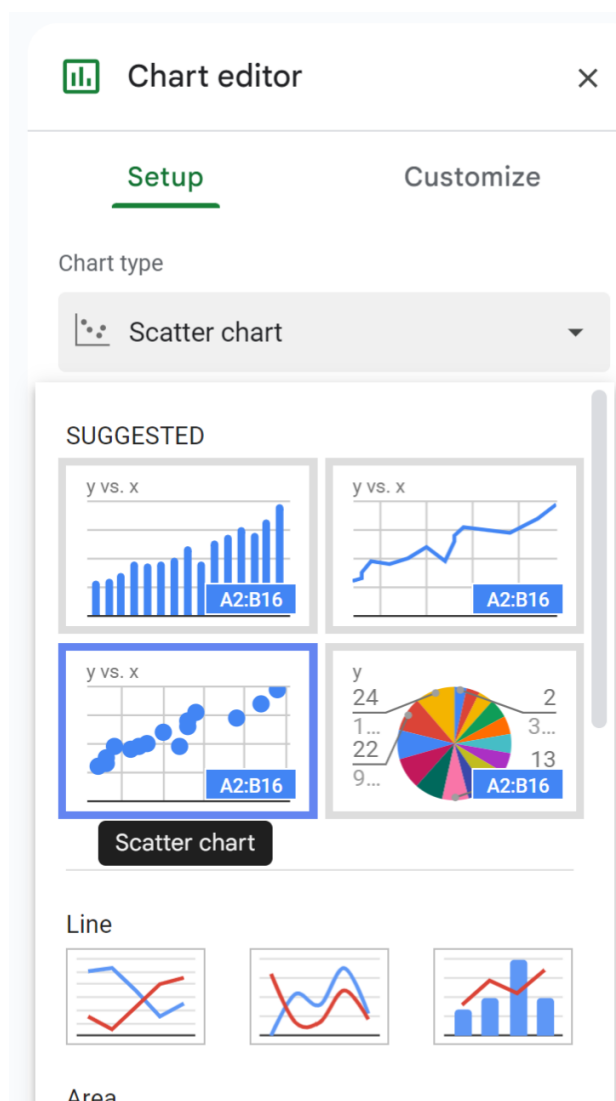
Use the following sequence of steps to efficiently generate the required visualization:

Select the entire data range that contains both your predictor and response variables, specifically cells **A2:B16** (the X and Y columns).

Navigate to the top menu ribbon and click the **Insert** tab, then select the **Chart** option from the dropdown menu.

In the **Chart editor** panel that subsequently appears on the right side of the screen, ensure that the designated chart type is a [Scatter chart](#). This plot type is essential for bivariate data as it clearly displays the relationship and any underlying linear trend.

The initial chart setup, displaying only the raw data points without the fitted line, should resemble the image provided below:



To integrate the statistically derived least squares line (which visually corresponds precisely to the coefficients calculated by the [LINEST\(\)](#) function), you must proceed to customize the chart settings:

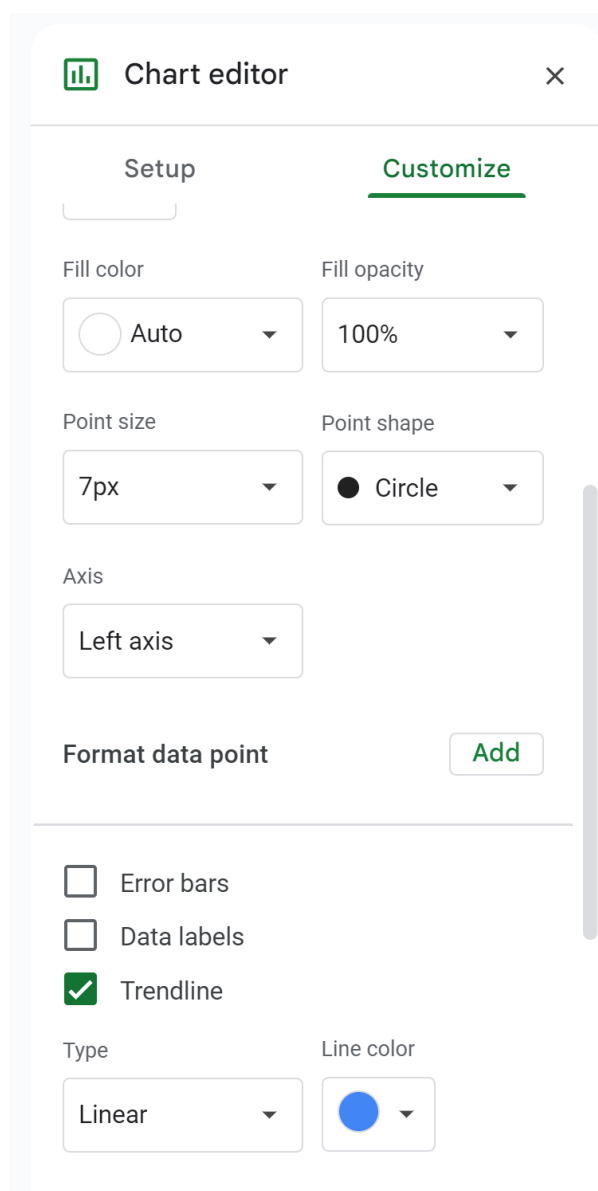
Click the **Customize** tab located within the **Chart editor** panel.

Expand the **Series** section of the customization options, which controls the visual properties of the

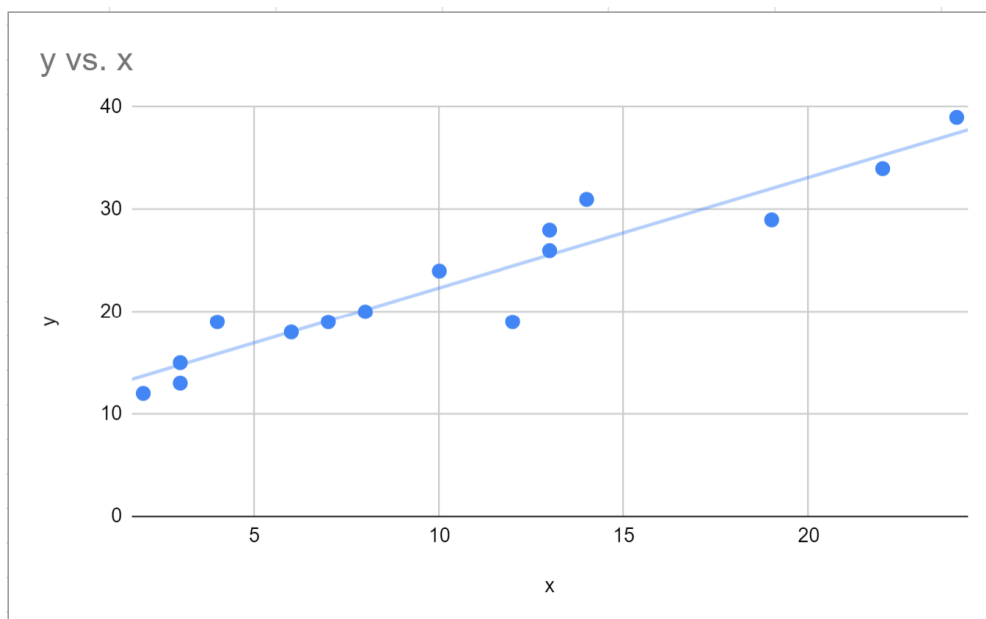
data points.

Scroll down through the Series settings until you locate the specialized option labeled **Trendline**. Check this box to automatically overlay the calculated linear fit onto your scatter plot.

This critical customization step ensures that the statistical model is visually superimposed onto the observational data, providing immediate confirmation of how well the linear assumption holds true for the underlying data structure:



The final result shows the fitted [regression line](#) prominently displayed on the scatter plot, offering a clear and immediate visual summary of the discovered linear relationship:



Each individual point on the plot represents an observed (x, y) pair from the original dataset. The straight line traversing the plot is the uniquely determined line that minimizes the squared distance to all these points, thus definitively representing the fitted regression line as determined by the [method of least squares](#).

Additional Resources for Statistical Analysis in Google Sheets

Mastering the **LINEST()** function and the robust visualization tools available in Google Sheets provides a powerful and accessible foundation for conducting meaningful statistical analysis without the need for specialized, expensive desktop software. The immediate ability to accurately calculate and visually represent the line of best fit using the method of least squares is essential for preliminary data exploration and the creation of accurate linear models.

For users keen on extending their capabilities beyond basic linear regression, Google Sheets offers a wealth of other statistical functions. The following resources offer guidance on executing other common statistical tasks, allowing you to deepen your understanding of data modeling and rigorous analysis within the spreadsheet environment:

Calculating Standard Deviation in Google Sheets

Performing T-Tests within Spreadsheet Environments

Advanced Data Filtering Techniques

By effectively leveraging these powerful built-in functionalities, users can efficiently manage the entire data analysis pipeline--from initial data entry and the calculation of precise regression [coefficients](#), to performing complex linear modeling and interpretation--all within a single, highly

accessible cloud environment.