

# How can I perform linear regression in Google Sheets?

Authored by  
**stats writer**

April 20, 2024

## RECOMMENDED CITATION

stats writer (2024). *How can I perform linear regression in Google Sheets?*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=137382>

Linear regression is a statistical method used to analyze the relationship between two continuous variables. In Google Sheets, this can be done by using the built-in function "LINEST". First, organize your data in two columns, with the independent variable in one column and the dependent variable in the other. Then, select a blank cell and enter the formula "`=LINEST(y-range, x-range, true, true)`", where "y-range" is the range of cells containing the dependent variable and "x-range" is the range of cells containing the independent variable. This will generate the regression coefficients, including the slope and intercept of the regression line. Additionally, the function will also provide other useful statistics such as the coefficient of determination (R-squared) and the standard error. By using the LINEST function, you can easily perform linear regression in Google Sheets and gain insights into the relationship between your variables.

## Perform Linear Regression in Google Sheets

**Linear regression is a method that can be used to quantify the relationship between one or more explanatory variables and a response variable.**

**We use simple linear regression when there is only one explanatory variable and multiple linear regression when there are two or more explanatory variables.**

**It's possible to perform both types of regressions using the LINEST() function in Google Sheets, which uses the following syntax:**

**LINEST(known\_data\_y, known\_data\_x, calculate\_b, verbose)**

**where:**

**known\_data\_y**: Array of response values  
**known\_data\_x**: Array of explanatory values  
**calculate\_b**: Indicates whether or not to calculate the y-intercept. This is TRUE by default and we leave it this way for linear regression.  
**verbose**: Indicates whether or not to provide additional regression statistics beyond just the slope and intercept. This is FALSE by default, but we will specify this to be TRUE in our examples.

The following examples show how to use this function in practice.

#### Simple Linear Regression in Google Sheets

Suppose we are interested in understanding the relationship between *hours studied* and *exam score*. studies for an exam and the exam score they receive.

To explore this relationship, we can perform simple linear regression using *hours studied* as an explanatory variable and *exam score* as a response variable.

The following screenshot shows how to perform simple linear regression using a dataset of 20 students with the following formula used in cell D2:

## =LINEST(B2:B21,A2:A21,TRUE,TRUE)

	A	B	C	D	E
1	<b>hours</b>	<b>score</b>			
2	1	76		5.250257467	67.16168898
3	2	78		0.7578474311	2.663270396
4	2	85		0.7272528139	5.280516451
5	4	88		47.99518133	18
6	2	72		1338.290628	501.9093718
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					

The following screenshot provide annotations for the output:

	A	B	C	D	E
1	<b>hours</b>	<b>score</b>		<i>slope</i>	<i>intercept</i>
2	1	76	<i>coefficient for slope, coefficient for intercept</i>	5.250257467	67.16168898
3	2	78	<i>Standard error for slope, standard error for intercept</i>	0.7578474311	2.663270396
4	2	85	<i>coefficient of determination, standard error</i>	0.7272528139	5.280516451
5	4	88	<i>F stat w/ degrees of freedom</i>	47.99518133	18
6	2	72	<i>Regression SS, Residual SS</i>	1338.290628	501.9093718
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					

**Here is how to interpret the most relevant numbers in the output:**

**R Square: 0.72725.** This is known as the coefficient of determination. It is the proportion of the variance in the response variable that can be explained by the explanatory variable. In this example, roughly 72.73% of the variation in the exam scores can be explained by the number of hours studied.

**Coefficients:** The coefficients give us the numbers

**necessary to write the estimated regression equation. In this example the estimated regression equation is:**

$$\text{Exam score} = 67.16 + 5.2503 * (\text{hours})$$

**We interpret the coefficient for hours to mean that for each additional hour studied, the exam score is expected to increase by 5.2503, on average. We interpret the coefficient for the intercept to mean that the expected exam score for a student who studies zero hours is 67.16.**

**We can use this estimated regression equation to calculate the expected exam score for a student, based on the number of hours they study. For example, a student who studies for three hours is expected to receive an exam score of 82.91:**

$$\text{Exam score} = 67.16 + 5.2503 * (3) = 82.91$$

**Multiple Linear Regression in Google Sheets**

**Suppose we want to know if the number of hours spent studying and the number of prep exams taken affects the score that a student receives on a certain college entrance exam.**

To explore this relationship, we can perform multiple linear regression using *hours studied* and *prep exams taken* as explanatory variables and *exam score* as a response variable.

The following screenshot shows how to perform multiple linear regression using a dataset of 20 students with the following formula used in cell E2:

**=LINEST(C2:C21,A2:B21,TRUE,TRUE)**

	A	B	C	D	E	F	G
1	<b>hours</b>	<b>prep_exams</b>	<b>score</b>				
2	1	1	76		-0.6016868046	5.555748295	67.67352554
3	2	3	78		0.9143850312	0.8991916992	2.815802228
4	2	3	85		0.734027217	5.365703261	#N/A
5	4	5	88		23.45815717	17	#N/A
6	2	2	72		1350.756885	489.4431152	#N/A
7	1	2	69				
8	5	1	94				
9	4	1	94				
10	2	0	88				
11	4	3	92				
12	4	4	90				
13	3	3	75				
14	6	2	96				
15	5	4	90				
16	3	4	82				
17	4	4	85				
18	6	5	99				
19	2	1	83				
20	1	0	62				
21	2	1	76				
22							
23							
24							
25							

**Here is how to interpret the most relevant numbers in the output:**

**R Square: 0.734.** This is known as the coefficient of determination. It is the proportion of the variance in the response variable that can be explained by the explanatory variables. In this example, 73.4% of the variation in the exam scores can be explained by the number of hours studied and the number of prep exams taken.

**Standard error: 5.3657.** This is the average distance that the observed values fall from the regression line. In this example, the observed values fall an average of 5.3657 units from the regression line.

**Estimated regression equation:** We can use the coefficients from the output of the model to create the following estimated regression equation:

$$\text{Exam score} = 67.67 + 5.56 * (\text{hours}) - 0.60 * (\text{prep exams})$$

**We can use this estimated regression equation to calculate the expected exam score for a student, based on the number of hours they study and the number of**

prep exams they take. For example, a student who studies for three hours and takes one prep exam is expected to receive a score of 83.75:

$$\text{Exam score} = 67.67 + 5.56*(3) - 0.60*(1) = 83.75$$

The following tutorials explain how to perform other common tasks in Google Sheets:

**[How to Create a Residual Plot in Google Sheets](#)**