

How can I use the LINEST function in Excel to calculate the linear regression coefficients for a set of data points?

Authored by
stats writer

June 30, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I use the LINEST function in Excel to calculate the linear regression coefficients for a set of data points?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=161484>

The LINEST function in Excel is a useful tool for calculating the linear regression coefficients for a set of data points. This function uses the least squares method to find the best fit line for the data and calculates the slope and intercept values. To use the LINEST function, simply select the data points and enter the function in a cell. This will return an array of values, including the slope and intercept, which can be used to create a linear regression equation for the data. This function is particularly helpful for analyzing trends and making predictions based on the data. Overall, the LINEST function is a powerful tool for performing linear regression analysis in Excel.

This article describes the formula syntax and usage of the **LINEST** function in Microsoft Excel.

Description

The **LINEST** function calculates the statistics for a line by using the "least squares" method to calculate a straight line that best fits your data, and then returns an array that describes the line. You can also combine **LINEST** with other functions to calculate the statistics for other types of models that are linear in the unknown parameters, including polynomial, logarithmic, exponential, and power series. Because this function returns an array of values, it must be entered as an array formula. Instructions follow the examples in this article.

The equation for the line is:

$$y = mx + b$$

-or-

$$y = m_1x_1 + m_2x_2 + \dots + b$$

if there are multiple ranges of x-values, where the dependent y-values are a function of the independent x-values. The m-values are coefficients corresponding to each x-value, and b is a constant value. Note that y, x, and m can be vectors. The array that the **LINEST** function returns is {m_n,m_{n-1},...,m₁,b}. **LINEST** can also return additional regression statistics.

Syntax

LINEST(known_y's, , ,)

The LINEST function syntax has the following arguments:

Syntax

known_y's Required. The set of y-values that you already know in the relationship $y = mx + b$.

If the range of **known_y's** is in a single column, each column of **known_x's** is interpreted as a separate variable.

If the range of **known_y's** is contained in a single row, each row of **known_x's** is interpreted as a separate variable.

known_x's Optional. A set of x-values that you may already know in the relationship $y = mx + b$.

The range of **known_x's** can include one or more sets of variables. If only one variable is used, **known_y's** and **known_x's** can be ranges of any shape, as long as they have equal dimensions. If more than one variable is used, **known_y's** must be a vector (that is, a range with a height of one row or a width of one column).

If **known_x's** is omitted, it is assumed to be the array {1,2,3,...} that is the same size as **known_y's**.

const Optional. A logical value specifying whether to force the constant b to equal 0.

If **const** is TRUE or omitted, b is calculated normally.

If **const** is FALSE, b is set equal to 0 and the m-values are adjusted to fit $y = mx$.

stats Optional. A logical value specifying whether to return additional regression statistics.

If **stats** is TRUE, **LINEST** returns the additional regression statistics; as a result, the returned array is {**mn,mn-1,....,m1,b;sen,sen-1,....,se1,seb;r2,sey;F,df;ssreg,ssresid**}.

If **stats** is FALSE or omitted, **LINEST** returns only the m-coefficients and the constant b.

The additional regression statistics are as follows.

Statistic	Description
se1,se2,....,sen	The standard error values for the coefficients m1,m2,....,mn.
seb	The standard error value for the constant b (seb = #N/A when const is FALSE).
r2	The coefficient of determination. Compares estimated and actual y-values, and ranges in value from 0 to 1. If it is 1, there is a perfect correlation in the sample -- there is no difference between the estimated y-value and the actual y-value. At the other extreme, if the coefficient of determination is 0, the regression equation is not helpful in predicting a y-value. For information about how r2 is calculated, see "Remarks," later in this topic.
sey	The standard error for the y estimate.
F	The F statistic, or the F-observed value. Use the F statistic to determine whether the observed relationship between the dependent and independent variables occurs by chance.

Statistic	Description
df	The degrees of freedom. Use the degrees of freedom to help you find F-critical values in a statistical table. Compare the values you find in the table to the F statistic returned by LINEST to determine a confidence level for the model. For information about how df is calculated, see "Remarks," later in this topic. Example 4 shows use of F and df.
ssreg	The regression sum of squares.
ssresid	The residual sum of squares. For information about how ssreg and ssresid are calculated, see "Remarks," later in this topic.

The following illustration shows the order in which the additional regression statistics are returned.

	A	B	C	D	E	F
1	m_n	m_{n-1}	...	m_2	m_1	b
2	se_n	se_{n-1}	...	se_2	se_1	se_b
3	r^2	se_y				
4	F	df				
5	ssreg	ssresid				

Examples

Example 1 - Slope and Y-Intercept

Copy the example data in the following table, and paste it in cell A1 of a new Excel worksheet. For formulas to show results, select them, press F2, and then press Enter. If you need to, you can adjust the column widths to see all the data.

Known y	Known x
1	0
9	4
5	2
7	3
Result (slope)	Result (y-intercept)
2	1
Formula (array formula in cells A7:B7)	
=LINEST(A2:A5,B2:B5,,FALSE)	

Example 2 - Simple Linear Regression

Copy the example data in the following table, and paste it in cell A1 of a new Excel worksheet. For formulas to show results, select them, press F2, and then press Enter. If you need to, you can adjust the column widths to see all the data.

Month	Sales
1	\$3,100
2	\$4,500
3	\$4,400
4	\$5,400
5	\$7,500
6	\$8,100
Formula	Result
=SUM(LINEST(B1:B6, A1:A6)*{9,1})	\$11,000
Calculates the estimate of the sales in the ninth month, based on sales in months 1 through 6.	

Example 3 - Multiple Linear Regression

Copy the example data in the following table, and paste it in cell A1 of a new Excel worksheet. For formulas to show results, select them, press F2, and then press Enter. If you need to, you can adjust the column widths to see all the data.

Floor space (x1)	Offices (x2)	Entrances (x3)	Age (x4)	Assessed value (y)
2310	2	2	20	\$142,000
2333	2	2	12	\$144,000
2356	3	1.5	33	\$151,000
2379	3	2	43	\$150,000
2402	2	3	53	\$139,000
2425	4	2	23	\$169,000
2448	2	1.5	99	\$126,000
2471	2	2	34	\$142,900
2494	3	3	23	\$163,000

Floor space (x1)	Offices (x2)	Entrances (x3)	Age (x4)	Assessed value (y)
2517	4	4	55	\$169,000
2540	2	3	22	\$149,000
-234.2371645				
13.26801148				
0.996747993				
459.7536742				
1732393319				
Formula (dynamic array formula entered in A19)				
=LINEST(E2:E12,A2:D12,TRUE,TRUE)				

Example 5 - Calculating the t-Statistics

Another hypothesis test will determine whether each slope coefficient is useful in estimating the assessed value of an office building in [Example 3](#). For example, to test the age coefficient for statistical significance, divide -234.24 (age slope coefficient) by 13.268 (the estimated standard error of age coefficients in cell A15). The following is the t-observed value:

$$t = m_4 \div se_4 = -234.24 \div 13.268 = -17.7$$

If the absolute value of t is sufficiently high, it can be concluded that the slope coefficient is useful in estimating the assessed value of an office building in [Example 3](#). The following table shows the absolute values of the 4 t-observed values.

If you consult a table in a statistics manual, you will find that t-critical, two tailed, with 6 degrees of freedom and Alpha = 0.05 is 2.447. This critical value can also be found by using the **TINV** function in Excel. **TINV**(0.05,6) = 2.447. Because the absolute value of t (17.7) is greater than 2.447, age is an important variable when estimating the assessed value of an office building. Each of the other independent variables can be tested for statistical significance in a similar manner. The following are the t-observed values for each of the independent variables.

Variable	t-observed value
Floor space	5.1
Number of offices	31.3
Number of entrances	4.8

Variable	t-observed value
Age	17.7

These values all have an absolute value greater than 2.447; therefore, all the variables used in the regression equation are useful in predicting the assessed value of office buildings in this area.

ARABPSYCHOLOGY.COM