

How to Easily Perform Cubic Regression Analysis in Excel

Authored by
stats writer

December 6, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Easily Perform Cubic Regression Analysis in Excel*.
PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=106116>

Welcome to this comprehensive guide on performing

Cubic regression in

Microsoft Excel.

Cubic regression is a sophisticated

regression technique

that becomes necessary when the relationship between your

predictor variable

(independent variable, X) and the

response variable

(dependent variable, Y) exhibits a distinct

non-linear relationship. Unlike simple linear regression, which models data using a straight line, cubic regression uses a third-degree polynomial curve (a curve with up to two turning points) to capture complex data patterns more accurately. This method is particularly useful in fields like engineering, economics, and biology where relationships often follow complex trajectories that cannot be adequately described by quadratic or simple linear models. By fitting a cubic model, we aim to find the coefficients that minimize the sum of squared errors, thereby providing the best possible fit for the observed data points. Understanding when and how to apply this technique is critical for generating robust predictive models.

While specialized statistical software is often used for advanced regression analysis,

Excel provides powerful built-in functionalities that allow users to perform complex calculations, including

cubic regression, using array formulas like

LINEST. This approach is highly effective for moderate datasets and offers immediate results within the familiar spreadsheet environment. This detailed, step-by-step example will guide you through the process of setting up the data, executing the necessary formulas, interpreting the resulting coefficients, and finally, visualizing the fitted cubic regression model directly onto a scatterplot in Excel. We will focus on ensuring the mathematical integrity of the model while providing a clear understanding of each procedural step required to achieve a clean and valid cubic fit, demonstrating how Excel handles the complexity of polynomial expansion.

Understanding Cubic Regression and its Mathematical Foundation

Before diving into the mechanics of Excel, it is essential to solidify the mathematical definition of a cubic regression model. A cubic model extends the standard linear model by introducing terms for the predictor variable squared (X^2) and the predictor variable cubed (X^3). The general form of the estimated cubic regression equation is written as: $\hat{y} = \beta_0 + \beta_1x + \beta_2x^2 + \beta_3x^3$, where $\hat{y}$ represents the predicted value of the

response variable, x is the value of the

predictor variable, and β_0 , β_1 , β_2 , and β_3 are the regression coefficients (the intercept and slopes)

that must be calculated from the observed data. Finding these coefficients using the principle of least squares is the primary goal of the regression process. The inherent complexity of this non-linear relationship allows the fitted line to change direction up to two times, providing modeling flexibility far beyond simple linear or quadratic models, which are often insufficient to capture real-world phenomena exhibiting saturation, decline, and subsequent recovery.

The decision to use a cubic model should ideally be driven by two factors: theoretical justification or visual inspection of the data. If previous research or subject matter expertise suggests a third-order polynomial relationship is likely--for instance, in dose-response curves in biology or efficiency curves in engineering--or if a preliminary scatterplot of the raw data clearly shows an S-shaped or similar curve that is not captured by lower-order polynomials, then cubic regression is the appropriate regression technique. It is crucial, however, to avoid the pitfall of **overfitting**, where adding higher-order terms merely complexifies the model without adding real predictive power, often resulting in a model that fits the current sample data perfectly but performs poorly on new, unseen data. Therefore, careful model selection, often utilizing statistical measures like adjusted R-squared and p-values (which LINEST can also provide if requested), is always recommended when choosing a higher-degree polynomial model.

Prerequisites: Preparing Data for the LINEST Function

Before executing the calculation, a crucial step involves organizing and preparing the dataset within the Excel environment. For the LINEST function to correctly compute the coefficients for a cubic model, your independent variable (X) must be implicitly expanded into the required polynomial components: X^1 , X^2 , and X^3 . When performing standard multiple linear regression, you would need separate columns for each predictor. However, Excel's array functionality within the LINEST function simplifies this by allowing us to use a single data range for X and instruct the function to calculate the powers internally.

To ensure success, verify that your data is clean, without missing values, and that the dependent and independent variables are aligned correctly. Although the specific column order is not mathematically mandatory for the LINEST calculation itself, maintaining a standard structure--such as placing the X values (the predictor variable) adjacent to the Y values (the response variable)--significantly improves readability and minimizes errors during range selection for the array formula. This preparatory organization is the foundation upon which the computational analysis relies.

Step 1: Setting Up the Sample Data

To demonstrate this procedure effectively, we will start by creating a sample dataset. This fictitious data includes a range of values for the independent variable (X) and corresponding values for the dependent variable (Y) which are engineered to exhibit a clear cubic trend. We will use a reasonably sized sample of 12 data points to ensure sufficient variation for the third-degree polynomial to be properly fitted. Ensure that your X values (Column A) and Y values (Column B) are correctly entered into your spreadsheet, typically starting from the second row (A2 and B2, reserving row 1 for headers).

This organization is crucial because the array formula in the next step relies on precise range references. If the data is misaligned or spread across non-contiguous cells, the formula will fail to correctly pair the X and Y observations. We recommend double-checking the entered numerical values against the visual aid to guarantee accuracy before proceeding to the calculation phase.

First, let's create the exemplary dataset in Excel:

	A	B	C	D	E	F
1	x	y				
2	6	14				
3	9	28				
4	12	50				
5	16	64				
6	22	67				
7	28	57				
8	33	55				
9	40	57				
10	47	68				
11	51	74				
12	55	88				
13	60	110				
14						
15						
16						
17						
18						
19						
20						
21						
22						

Step 2: Implementing the Cubic Regression Model using LINEST

The core of performing

cubic regression in Excel lies in the robust

LINEST function. Since a cubic model requires calculating four coefficients (β_0 , β_1 , β_2 , and β_3), the output of the

LINEST function must occupy a range of 1 row by 4 columns. Before entering the formula, it is **essential** to select a contiguous block of four empty cells horizontally where the results will be displayed, as

LINEST is an array function designed to return multiple statistical values simultaneously. Failure to select the entire output range will result in only the first coefficient being displayed.

To fit the cubic model, we utilize an array construction within the

LINEST function. The general syntax requires specifying the known Y values and the known X values. For cubic regression, the known_x's argument must be adapted to include the first, second, and third powers of the X data range. We achieve this by using the exponentiation operator (^) combined with an array constant {1, 2, 3}. This tells Excel to internally generate three distinct sets of predictor variables (X^1 , X^2 , and X^3) from the single input column, treating the problem as a multiple regression against these three generated powers.

Using the specific data ranges from our example (Y in B2:B13 and X in A2:A13), the formula must be typed into the selected four cells. If you are using an older version of Excel, you must confirm this formula using **Ctrl + Shift + Enter** to execute it as an array formula, surrounding the formula with curly braces {}. Newer versions of Excel (Microsoft 365) handle dynamic array output automatically. The powerful and concise formula structure is as follows:

=LINEST(B2:B13, A2:A13^{1,2,3})

Executing this array formula correctly will populate the four chosen cells with the calculated coefficients. The coefficients are returned in descending polynomial order: the leftmost cell will contain the coefficient for X^3 (β_3), followed by the coefficient for X^2 (β_2), the coefficient for X^1 (β_1), and finally, the constant intercept (β_0). This specific ordering is a standard output convention for the

LINEST function when dealing with polynomial array inputs.

The following screenshot illustrates the result of applying this powerful array formula to our particular dataset, yielding the four required coefficients:

D1								
=LINEST(B2:B13, A2:A13^{1,2,3})								
	A	B	C	D	E	F	G	H
1	x	y		0.003302	-0.32142	9.832054	-32.0118	
2	6	14						
3	9	28						
4	12	50						
5	16	64						
6	22	67						
7	28	57						
8	33	55						
9	40	57						
10	47	68						
11	51	74						
12	55	88						
13	60	110						
14								
15								
16								
17								
18								
19								
20								
21								
22								
23								

Understanding and Interpreting the Regression Output

Once the

LINEST function successfully executes, the four resulting values represent the optimal coefficients for the

cubic regression model based on the principle of least squares estimation. Let's assume the four output values, reading from left to right in the output array, are: **0.0033**, **-0.3214**, **9.832**, and **-32.0118**. These numbers directly correspond to the coefficients β_3 , β_2 , β_1 , and β_0 , respectively.

Using these calculated coefficients, we can formally write the estimated cubic regression model for predicting the

response variable (?) based on the

predictor variable (x). This equation is the mathematical representation of the fitted curve that best describes the

non-linear relationship observed in the data. The derived equation is:

$$\hat{y} = -32.0118 + 9.832x - 0.3214x^2 + 0.0033x^3$$

Interpreting these coefficients requires care, as unlike simple linear regression, the interpretation of

individual slopes (β_1 , β_2 , β_3) does not represent a constant marginal change. Instead, they collectively define the complex curvature of the third-degree polynomial. The constant (β_0 , -32.0118) represents the predicted value of Y when X is equal to zero, though in many practical applications, an X value of zero may fall outside the meaningful range of the data, meaning interpretation of the intercept should be cautious. The sign and magnitude of the higher-order terms (β_2 and β_3) are key determinants in shaping the inflection points of the fitted curve.

Step 3: Creating the Initial Scatterplot Visualization

While the coefficients derived from LINEST provide the precise mathematical definition of the model, visualizing the fit is crucial for validation and intuitive understanding of the regression technique. A

scatterplot allows us to see how well the cubic curve tracks the original data points, confirming the appropriateness of the cubic structure and identifying potential outliers or areas where the fit might be weaker.

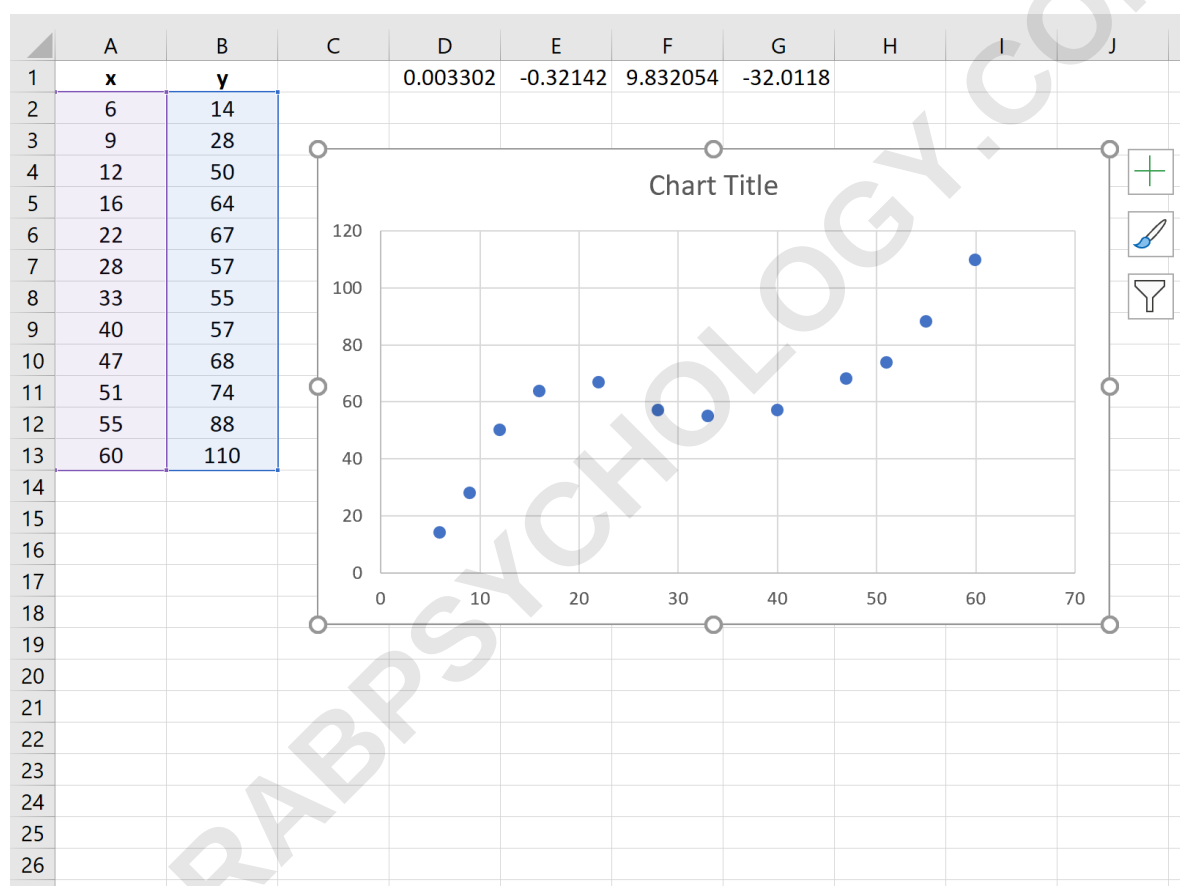
The first step in visualization is to generate the base chart. Begin by highlighting the entire range of your data, encompassing both the X and Y columns (A2:B13 in our example). This ensures that Excel correctly identifies the paired data points needed for the plot, assigning the leftmost column (A) as the independent variable (X-axis) by default in a scatter chart.

Highlight the data range as shown here:

	A	B	C	D	E	F	G
1	x	y		0.003302	-0.32142	9.832054	-32.0118
2	6	14					
3	9	28					
4	12	50					
5	16	64					
6	22	67					
7	28	57					
8	33	55					
9	40	57					
10	47	68					
11	51	74					
12	55	88					
13	60	110					
14							
15							
16							
17							
18							
19							
20							
21							
22							

Next, navigate to the **Insert** tab located in the top ribbon of Excel. Within the **Charts** group, locate the **Insert Scatter (X, Y) or Bubble Chart** option. Click this button and select the first option, which displays only markers. Executing this step will generate the initial scatterplot, showing the distribution of your raw data points. At this stage, you should be able to visually confirm the suspected non-linear relationship that necessitated the use of cubic regression over simpler models.

This action produces the basic chart structure:

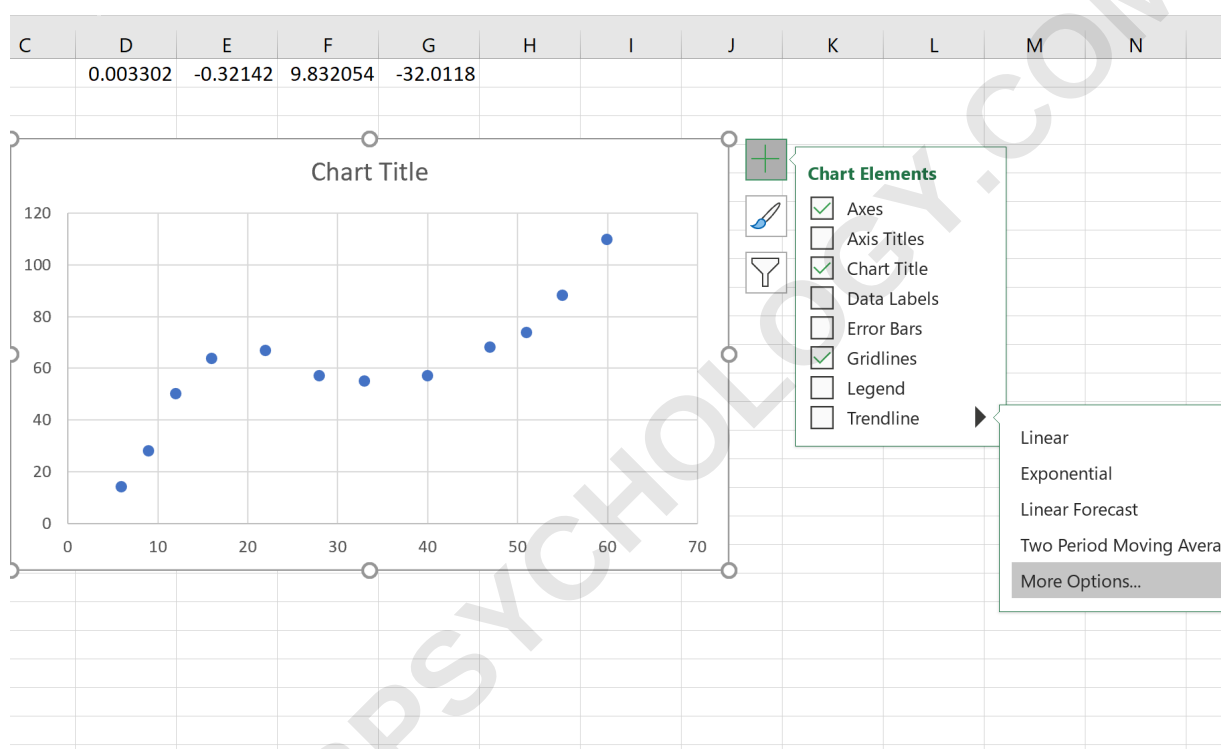


Step 4: Adding and Configuring the Polynomial Trendline

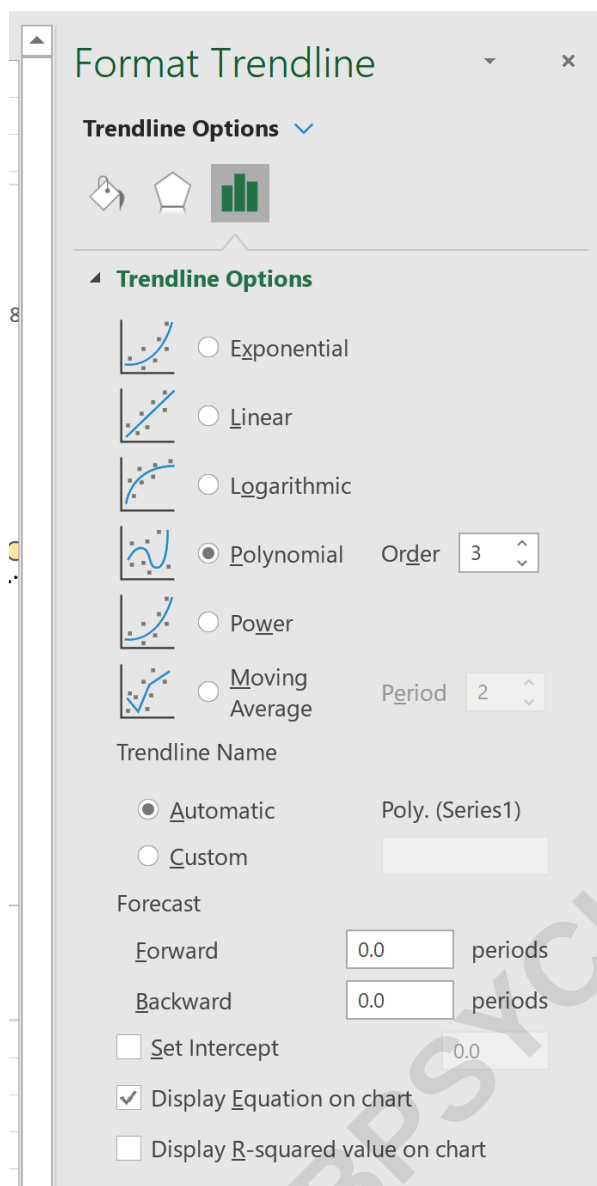
To overlay the visual representation of the cubic model onto the raw data, we must add a Polynomial trendline of the third order. With the chart selected, click the green plus sign (Chart Elements) that appears in the top right corner of the chart area. Hover over the **Trendline** option in the menu that appears, and then click the arrow to the right of it. This action reveals advanced trendline options that are necessary for specifying a polynomial degree higher than the default linear setting. In the subsequent dropdown menu, select **More Options...** to open the detailed trendline formatting pane on the right side of your screen.

This step is vital for ensuring the visualization accurately reflects the mathematical model calculated using

LINEST. The menu path allows us to bypass the simpler trendline types (like Linear or Exponential) and specifically request a higher-order fit. Once the Trendline Options pane is open, select the radio button labeled **Polynomial**. Crucially, immediately after selecting Polynomial, you must adjust the **Order** setting. Since we are performing cubic regression, set the Order dropdown value to **3**. This tells Excel's charting tool to calculate and draw a curve based on the X , X^2 , and X^3 terms, matching our mathematical model derived from the LINEST array formula.



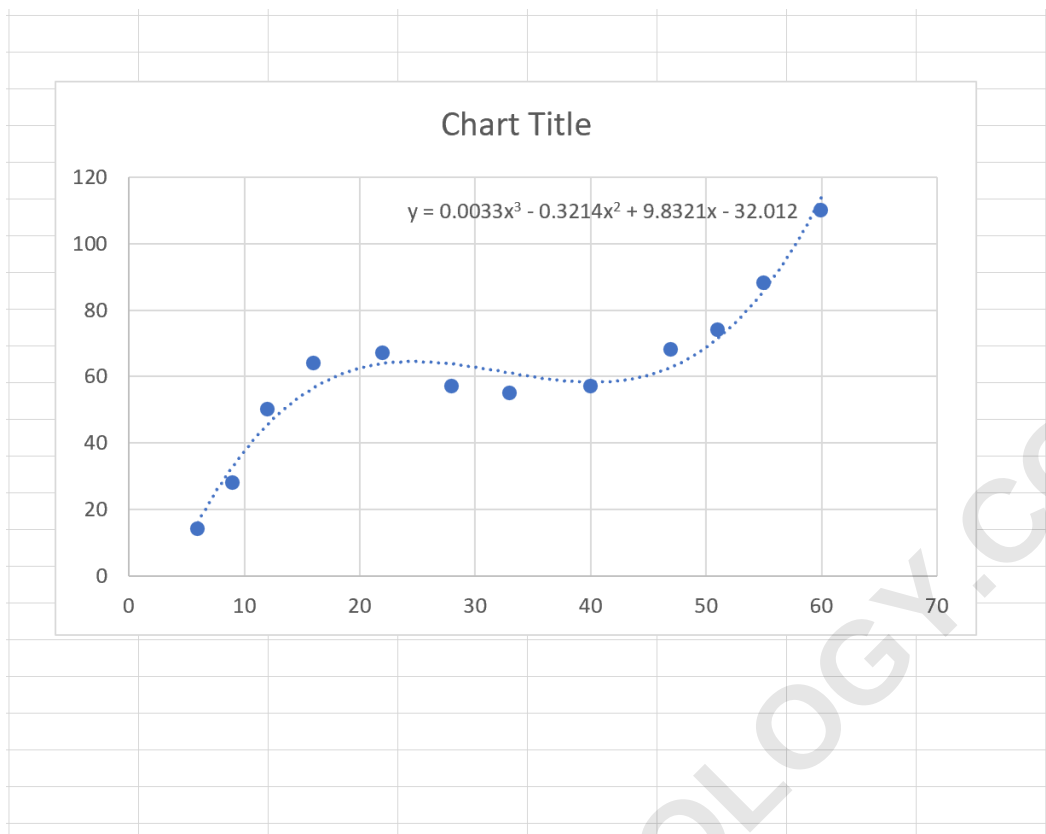
To visually verify that the chart-generated equation matches the one derived from the LINEST function, check the box labeled **Display Equation on chart**. It is also highly recommended to check the box **Display R-squared value on chart**, as the R-squared value provides a concise measure of how well the cubic model fits the data, indicating the proportion of variance in the response variable that is predictable from the predictor variable. A high R-squared value (close to 1) suggests a strong fit for this Polynomial trendline.



Analyzing the Final Results and Model Validation

Upon completing the configuration in Step 4, the polynomial curve will be drawn through your data points, and the corresponding equation and R-squared value will be displayed directly on the chart. This final visual output confirms the successful execution of the cubic regression analysis.

The visual output will present the fitted curve and the equation as follows:



A crucial validation step is comparing the equation generated by the chart visualization tool to the coefficients calculated earlier using the LINEST function. You will notice that the numerical coefficients in the chart equation (e.g., $y = 0.0033x^3 - 0.3214x^2 + 9.832x - 32.011$) precisely match the values returned by the array formula (0.0033, -0.3214, 9.832, -32.0118). This concordance confirms the mathematical consistency between the direct regression calculation method and the graphical trendline method, providing strong assurance that your regression technique has been applied correctly. The visual fit of the curve to the data points further reinforces the conclusion that the cubic model is appropriate for describing this specific non-linear relationship.

Conclusion: Leveraging Cubic Regression for Complex Modeling

Mastering

cubic regression in Excel equips you with a powerful, accessible tool for analyzing datasets where the underlying relationship is highly non-linear. By employing the array functionality of the LINEST function, analysts can bypass the need for external statistical software for basic polynomial fitting, allowing for rapid deployment of predictive models. This methodology results in a precise mathematical model ($y = \beta_0 + \beta_1x + \beta_2x^2 + \beta_3x^3$). Furthermore, the ability to overlay a

Polynomial trendline of order 3 onto a scatterplot provides an invaluable visual check, ensuring the curve accurately maps the trajectory of the observed data points and confirms that the chosen model adequately captures the complexity of the relationship between the predictor variable and the response variable. This dual approach--calculation via LINEST and validation via visualization--ensures both mathematical accuracy and interpretive clarity in your regression modeling efforts, making complex analysis accessible directly within your spreadsheet environment.

ARABPSYCHOLOGY.COM