

# Cubic Regression Analysis in Excel: A Step-by-Step Tutorial

Authored by  
**Mohammed loot**

November 5, 2025

## RECOMMENDED CITATION

Mohammed loot (2025). *Cubic Regression Analysis in Excel: A Step-by-Step Tutorial*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=10305>

When analyzing complex datasets, simple linear models often fall short in describing the true relationship between variables. [Cubic regression](#), a specialized form of polynomial modeling, is the essential statistical tool employed when the relationship between a [predictor variable](#) and a response variable demonstrates a significant [non-linear](#) curvature. This technique is distinguished by its use of a third-degree polynomial ( $x^3$ ) to accurately capture curves, peaks, and inflection points present within the observed data structure.

The ability to fit a precise cubic curve is indispensable across disciplines such as financial engineering, biological modeling, and empirical economics, where underlying phenomena rarely adhere to perfectly straight paths. By using a cubic model, researchers can achieve a far more accurate and representative statistical model that reflects the dynamic complexities of the process being studied. This comprehensive tutorial provides a rigorous, step-by-step methodology for executing and validating a cubic regression analysis directly within the familiar environment of Microsoft Excel, utilizing its robust, built-in statistical functionalities.

## Preparing the Environment and Understanding LINEST Requirements

Before initiating the calculation process, it is fundamental to understand the core engine we will be leveraging: the [LINEST](#) function. LINEST is Excel's powerful [array formula](#) designed to compute the statistics necessary for fitting a line (or curve) to data using the "least squares" method. While its application for standard linear regression is straightforward, deploying it for advanced polynomial modeling, such as cubic regression, necessitates specific structural modifications to the formula input.

A crucial prerequisite for successful cubic modeling is the proper handling of the independent variable inputs. Since a cubic model requires the predictor variable ( $X$ ) to be present in its first, second, and third powers ( $x$ ,  $x^2$ ,  $x^3$ ), we must explicitly instruct the function to generate and consider these three degrees. This is accomplished by defining the 'known\_x' argument as an array constant, which compels Excel to treat the single predictor column as three distinct columns of calculated inputs. This manipulation is vital for achieving the three necessary [coefficients](#) plus the intercept.

Furthermore, users must remember the unique execution requirement of LINEST. Since it functions as an [array formula](#), it requires specific entry confirmation. Traditionally, this means selecting the entire desired output range first and then confirming the formula input using the combination of **Ctrl + Shift + Enter**. In modern Excel versions that support dynamic arrays, pressing Enter may suffice, but understanding the array nature is key. Failure to execute this step correctly will result in an incomplete output, typically displaying only the first coefficient and rendering the model unusable.

## Step 1: Organizing and Preparing Raw Data in Excel

The first operational step requires meticulously organizing the dataset into two adjacent columns: one dedicated to the **predictor variable** (X, the independent variable) and the other for the response variable (Y, the dependent variable). For the purpose of this demonstration, we utilize a simulated dataset intentionally crafted to exhibit a pronounced curvilinear pattern, making it an exemplary candidate for a comprehensive cubic analysis. It is imperative at this stage to ensure your data is meticulously cleaned and verified for accuracy before proceeding to the calculation phase.

In our running example, we designate Column A for the X values and Column B for the corresponding Y values. This structured, contiguous layout is essential, as the **LINEST** function relies on accurately defined ranges for precise calculation. The data arrangement must clearly separate the known Y values from the known X values to facilitate the array processing in the subsequent step.

The initial data structure, demonstrating the relationship between the X and Y variables, is visualized below. This relationship will serve as the foundation upon which our cubic equation will be modeled:

	A	B	C	D	E	F
1	<b>x</b>	<b>y</b>				
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						

It is highly recommended practice to generate a preliminary scatterplot at this juncture. A visual review of the data distribution allows for quick confirmation that a [non-linear](#) model, specifically a cubic function, is indeed the most appropriate and statistically sound choice for capturing the underlying trend.

## Step 2: Executing the Cubic Regression Calculation Using LINEST

The pivotal step in deriving the cubic regression model involves the sophisticated array application of the [LINEST](#) function. Unlike a simple linear fit, the cubic model requires the function to simultaneously account for three degrees of the independent variable ( $X$ ,  $X^2$ ,  $X^3$ ). We achieve this necessary complexity by defining the 'known\_x' argument as an array constant, which dynamically calculates the required powers for every data point.

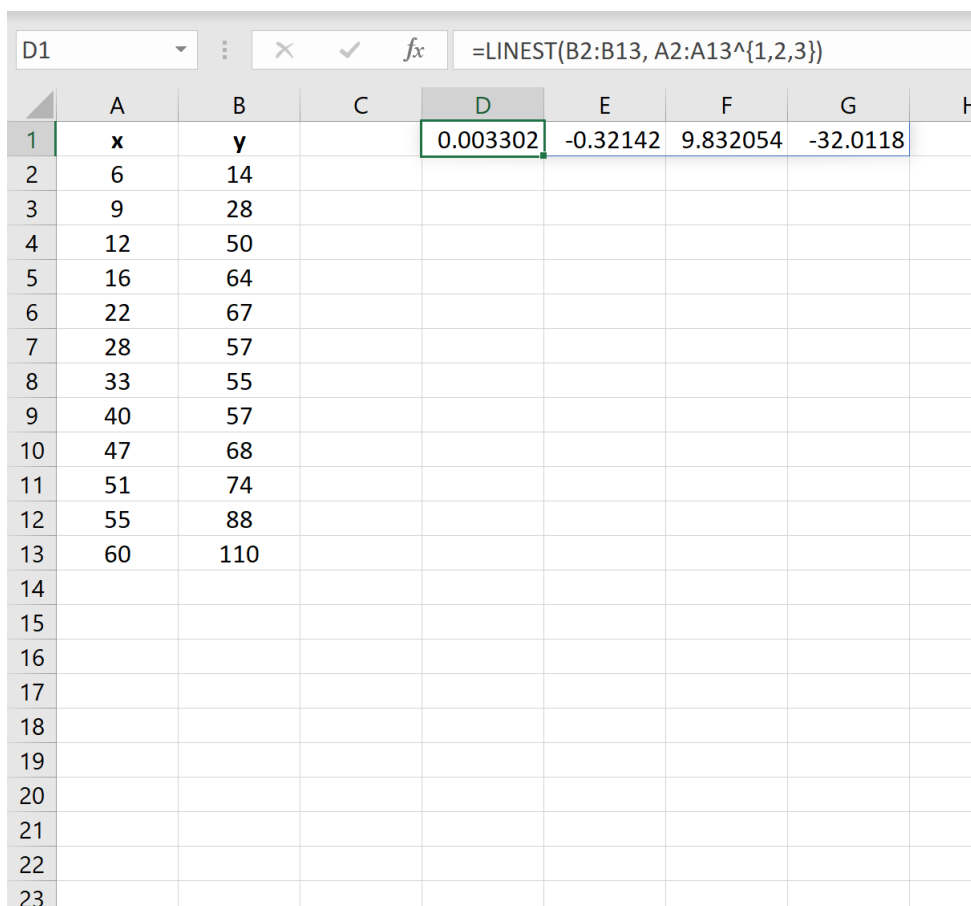
To execute the cubic regression, you must first select a contiguous range of cells--specifically, a 1-row by 4-column block. This size is necessary because a cubic model yields four critical output values: three [coefficients](#) (for  $x^3$ ,  $x^2$ , and  $x^1$ ) and one Y-intercept. In this selected output range, enter the following array formula precisely as written. Note the critical inclusion of `{1, 2, 3}`, which is the mechanism that instructs Excel to generate the first, second, and third power columns

from the input data range A2:A13:

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

In this structure, B2:B13 represents the dependent variable (known Y values), while the manipulated range A2:A13^{1,2,3} represents the known X values raised to the required polynomial powers. After typing the formula, it must be confirmed as an [array formula](#) (using **Ctrl + Shift + Enter** in older versions) to ensure the calculation populates all four cells in the selected output range with the complete statistical results.

The successful application of this command will provide the necessary [coefficients](#) that define your model, displayed horizontally in the output range. This calculated output forms the mathematical basis of your estimated cubic curve, as illustrated in the figure below:



The screenshot shows an Excel spreadsheet with the following data and formula:

	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								

The formula bar shows: `=LINEST(B2:B13, A2:A13^{1,2,3})`

## Interpreting the Derived Coefficients and Model Equation

The horizontal output row generated by the LINEST function contains the estimated [coefficients](#) (or parameters) for the cubic equation. It is vital to understand the order in which Excel presents

these values. When the simplest form of LINEST is used, the coefficients are consistently ordered from the highest power of X down to the Y-intercept. Therefore, for our cubic model, the four values correspond sequentially to the coefficient of  $x^3$ , the coefficient of  $x^2$ , the coefficient of  $x^1$ , and finally, the intercept ( $b_0$ ).

By mapping these calculated numerical values to the standard cubic equation form ( $y = b_0 + b_1x + b_2x^2 + b_3x^3$ ), we can formally define our estimated [regression](#) model. This precisely formulated model is the tool needed to predict the value of the response variable ( $y$ ) for any given value of the [predictor variable](#) ( $x$ ), provided the value remains within the reasonable scope of the observed data range (interpolation).

Based on the numerical output derived from our example dataset in Step 2, the estimated cubic regression model is explicitly formulated as follows:

$$y = -32.0118 + 9.832x - 0.3214x^2 + 0.0033x^3$$

This equation represents the unique curve calculated via the least squares methodology--the curve that minimizes the collective sum of squared errors between the model's predicted values and the actual observed values in our dataset. Thus, it provides the optimal fit for the observed [non-linear](#) trend.

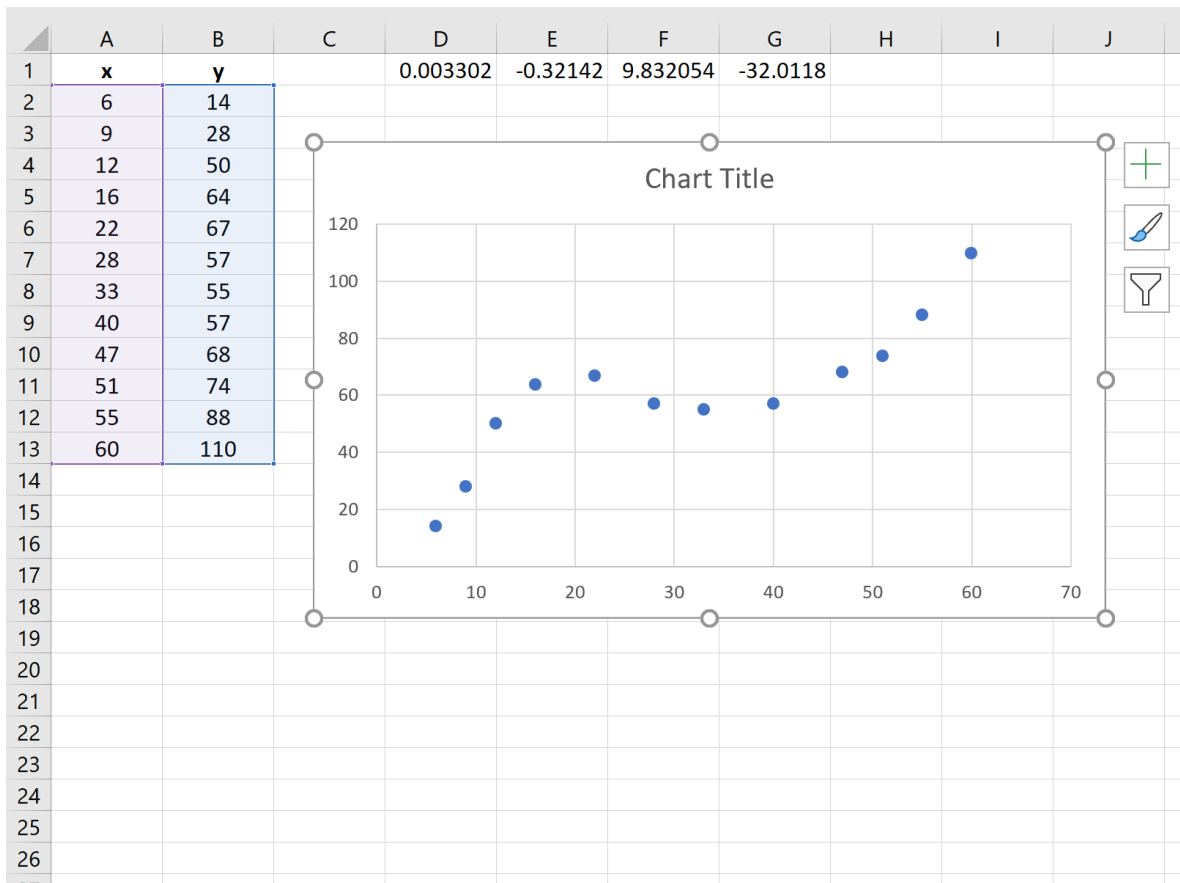
### Step 3: Visual Confirmation and Model Validation via Scatterplot

While calculating the [coefficients](#) using LINEST is essential for mathematical prediction, the visualization step is equally critical for confirming the model's appropriateness and understanding the characteristic shape of the [cubic regression](#) curve. We will now generate a scatterplot and overlay the calculated third-order trendline using Excel's native charting capabilities.

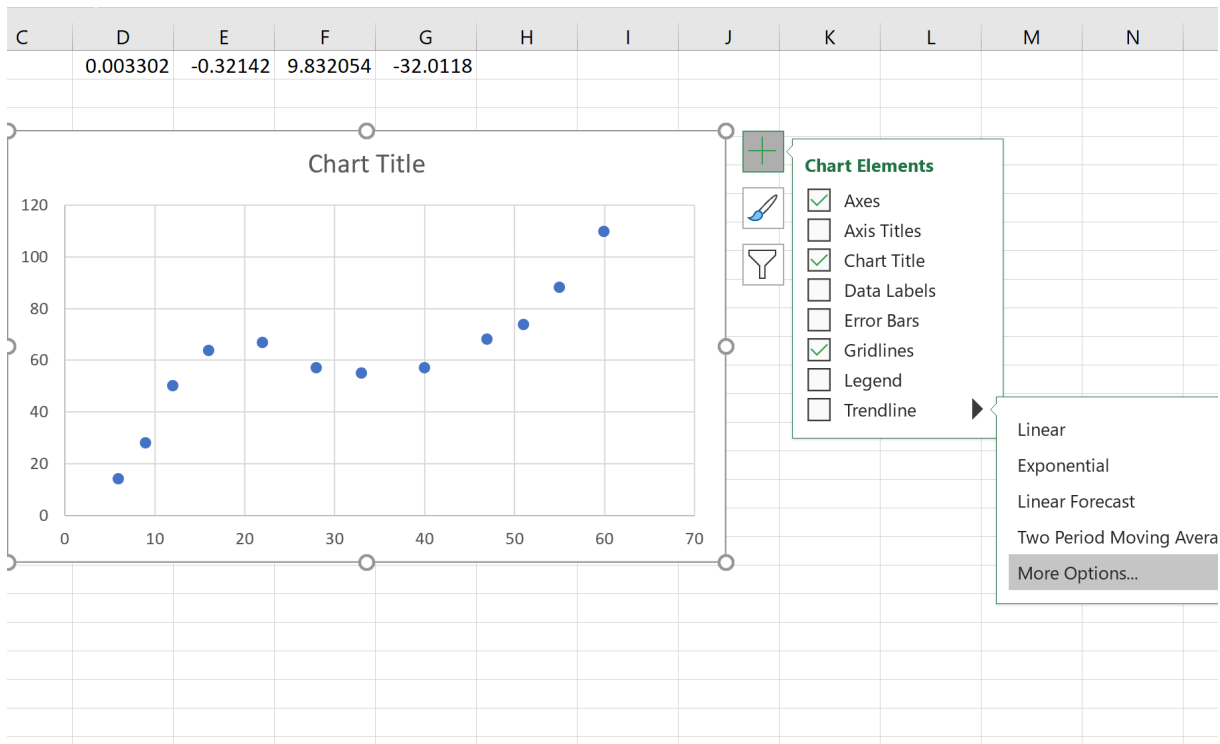
To begin the visualization process, highlight the entire range of your raw data, ensuring both the X and Y columns are selected. This selection informs Excel of the data pairing necessary for accurate chart generation and variable plotting.

	A	B	C	D	E	F	G
1	<b>x</b>	<b>y</b>		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							

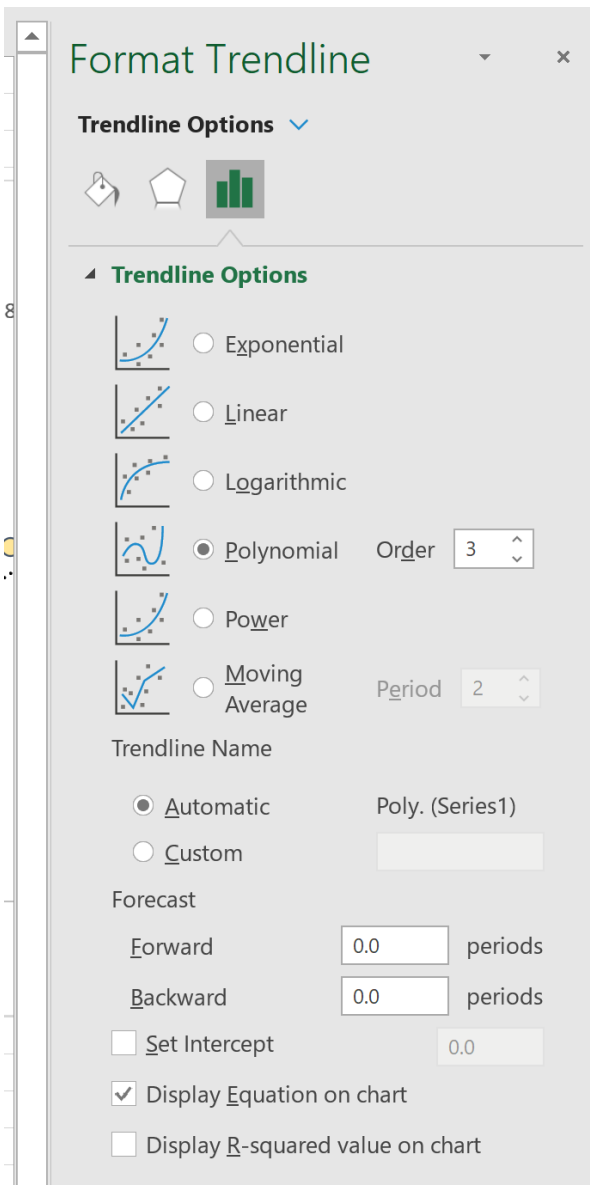
Navigate to the **Insert** tab on the main ribbon. Within the **Charts** group, select the **Insert Scatter (X, Y)** option and choose the basic scatterplot type (displaying points only). This action immediately renders the observed relationship between the predictor and response variables, illustrating the initial non-linear pattern:



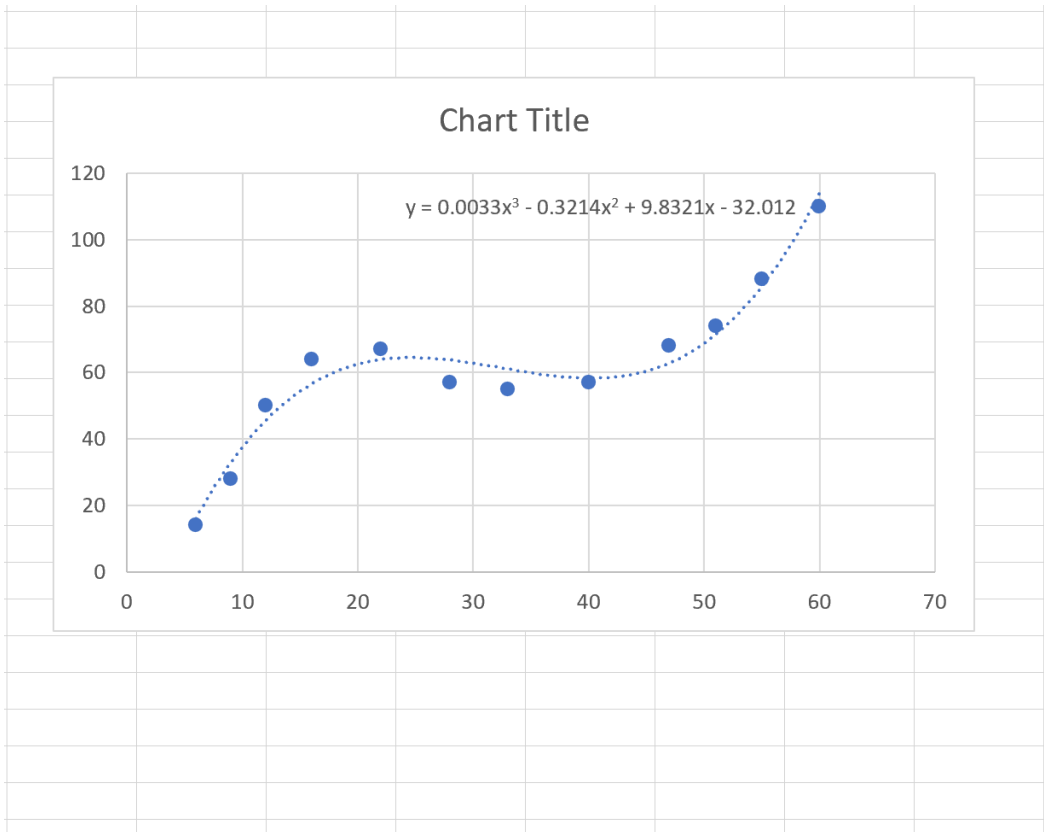
To overlay the cubic trendline, click directly on the generated chart. The green plus sign (Chart Elements) will appear near the top-right corner. Click this sign, hover over the **Trendline** option, and click the arrow to the right. From the subsequent dropdown menu, select **More Options** to open the detailed Format Trendline pane, which provides advanced configuration settings.



In the Format Trendline pane, override the default Linear setting by selecting the **Polynomial** option under Trendline Options. Crucially, adjust the **Order** setting to **3**, which explicitly specifies that we are fitting a cubic curve to the data points. For immediate verification, also check the box labeled "Display Equation on chart." This action overlays the mathematical representation of the curve directly onto the graphical output.



The finalized chart provides powerful visual evidence, confirming the excellent fit of the cubic model to the raw data, as the curved line passes closely through the observed points. The displayed equation on the chart is a vital cross-validation element:



By comparing the chart's equation with the manual calculations, we confirm that the equation automatically calculated and displayed by the charting feature precisely matches the [coefficients](#) derived manually using the **LINEST()** array function. This alignment validates the accuracy and reliability of both the calculated statistical parameters and the resulting graphical representation.

## Conclusion: Mastering Advanced Non-Linear Modeling in Excel

Successfully fitting a [cubic regression](#) model in Excel provides a highly effective and accessible method for analyzing datasets that exhibit complex, curvilinear relationships. By skillfully combining the [LINEST](#) array formula with the power constant `{1, 2, 3}`, users can efficiently and accurately derive the necessary [coefficients](#) that mathematically define the best-fit cubic curve.

The subsequent visualization phase--involving the creation of a scatterplot and the addition of a third-order polynomial trendline--is indispensable. This step not only confirms the model's visual accuracy but also serves as a critical cross-check for the derived equation. Mastering these techniques ensures that Excel remains a powerful and readily available platform for performing advanced statistical analysis, even when navigating the complexities of [non-linear](#) data structures.

## **Additional Resources for Deepening Regression Expertise**

To further enhance your command of polynomial regression and related statistical methodologies, we recommend reviewing authoritative resources. These include official documentation detailing Excel's statistical functions or exploring academic introductions that provide theoretical depth on modeling complex non-linear data structures.