Learning Quadratic Regression Analysis Using Microsoft Excel

Authored by

Mohammed looti

November 9, 2025

Understanding Regression and the Need for Non-Linear Models

Regression analysis stands as a cornerstone statistical technique used to meticulously model and analyze the relationship between a dependent variable (the response) and one or more independent variables (the predictors). The primary objective of any **regression** model is to estimate the conditional expectation of the response variable given specific values of the independent variables, thereby enabling robust prediction and profound explanation of observed phenomena. The most frequently encountered model is linear regression, which inherently assumes a simple, constant, straight-line relationship between the predictor and response variables. In this linear framework, any change in the predictor variable is expected to yield a proportionate and constant change in the response variable.

Although **linear regression** is remarkably powerful and widely applicable across numerous disciplines, its utility is confined strictly to scenarios where the underlying relationship between variables is genuinely linear. For instance, a linear model might effectively describe how increasing the dosage of a medication relates to a reduction in symptoms, assuming the effect remains consistent across all dosage levels. However, complexity often defines real-world data; many relationships exhibit patterns that feature distinct curves, defined peaks, or noticeable troughs. When a preliminary data visualization, such as a scatterplot, clearly reveals a significant curvature, forcing the data into a linear model inevitably leads to a poor statistical fit, resulting in biased predictions and fundamentally inaccurate interpretations of the underlying variable relationship.

When the relationship deviates significantly from a predictable straight line, statisticians must advance their modeling approach by adopting non-linear techniques. One of the most accessible and highly interpretable forms of non-linear modeling is polynomial regression, of which **quadratic regression** is a specific and essential type. This model is perfectly suited for describing scenarios where the response variable increases up to an optimal point and subsequently begins to decrease, or vice-versa, creating a recognizable symmetrical, parabolic curve. Recognizing the critical need to transition from a simple linear approach to a more sophisticated polynomial methodology, such as **quadratic regression**, is paramount for achieving accuracy in statistical modeling and ensuring the robustness of forecasts.

Why Choose Quadratic Regression? Identifying the Curvilinear Pattern

The defining mathematical characteristic of a **quadratic relationship** is its parabolic geometry--it typically manifests as a "U" shape or an inverted "U" shape when rendered graphically. This inherent curvature mathematically dictates that the effect exerted by the predictor variable on the response variable is not constant; rather, it changes continuously as the predictor variable itself increases. In stark contrast to the constant slope characteristic of linear models, the slope in a quadratic model is dynamic, reflecting an impact that is either diminishing or accelerating depending on the value of the predictor. This complex structure is algebraically achieved by incorporating the predictor variable squared (X^2) as an additional, distinct term within the model equation, thereby granting the regression line the necessary flexibility to bend and closely follow the true distribution of the data points.

Consider the illustrative example of the relationship between hours worked per week and an individual's reported happiness level. Initially, as a person increases their work commitment from a low baseline (e.g., 10 hours) up to a moderate level (e.g., 30 hours), they often experience greater financial security, increased fulfillment, and a stronger sense of accomplishment, all contributing to elevated happiness. Yet, upon crossing a certain critical threshold--perhaps 45 or 50 hours per week--the marginal benefits dissipate, and additional work begins to introduce severe stress, chronic fatigue, loss of vital leisure time, and eventual burnout. At this crucial inflection point, further increases in hours worked paradoxically lead to a measurable decline in reported happiness. This classic peak-and-decline dynamic cannot be accurately represented by a linear model; it mandates the use of a **quadratic regression** model to precisely identify and capture this turning point.

In scenarios such as these, where empirical evidence or theoretical reasoning suggests that the relationship between two variables possesses an optimal point--either a maximum or a minimum--the **quadratic regression** methodology provides the essential analytical framework required to model this complex behavioral pattern. By systematically incorporating the squared term alongside the original predictor, the model is fully equipped to identify this inflection point with statistical rigor, offering significantly greater explanatory power and predictive accuracy than a simple straight line could ever provide. Our subsequent steps will rigorously demonstrate the exact methodology required to implement this powerful statistical technique effectively within the familiar computational environment of **Microsoft Excel**.

Setting Up the Data and Visual Confirmation in Excel

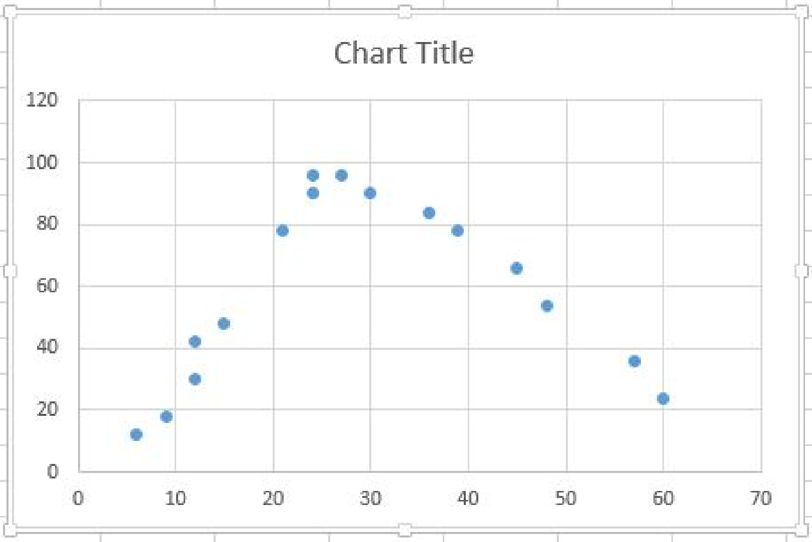
To clearly illustrate the practical steps involved in executing **quadratic regression**, we will utilize a small, representative dataset. This dataset comprises observations detailing the number of hours worked per week and the corresponding reported happiness level (measured on a standardized scale of 0 to 100) for 16 distinct individuals. This initial arrangement and accurate input of data serves as the fundamental foundation upon which all subsequent statistical analysis will be built.

The first and perhaps most vital step in any sound statistical analysis is the visual inspection of the data. Before committing to the specific structure of a regression model, we must visually verify the data's pattern using a scatterplot to confirm the empirical presence of a non-linear, parabolic relationship. If the plotted data points visibly trace a U-shape or an inverted U-shape, this observation provides compelling initial evidence favoring the use of **quadratic regression** over its simpler linear counterpart.

To generate this essential initial visualization in Excel, follow this straightforward procedure: First, highlight the entire data range containing both the predictor variable (Hours Worked, Column A) and the response variable (Happiness Level, Column B). For our example, this corresponds to cells **A2:B17**. Next, navigate to the **INSERT** tab located on the top ribbon. Within the *Charts* group, select the *Scatter* plot icon. This action immediately generates a scatterplot, which serves as a powerful visual diagnostic tool, instantly revealing the empirical relationship between the two variables:

	A	B	C
1	Hours worked	Reported Happiness	
2	6	12	
3	9	18	
4	12	30	
5	12	42	
6	15	48	
7	21	78	
8	24	90	
9	24	96	
10	27	96	
11	30	90	
12	36	84	
13	39	78	
14	45	66	
15	48	54	
16	57	36	
17	60	24	
18			

	A	B
1	Hours worked	Reported Happiness
2	6	12
3	9	18
4	12	30
5	12	42
6	15	48
7	21	78
8	24	90
9	24	96
10	27	96
11	30	90
12	36	84
13	39	78
14	45	66
15	48	54
16	57	36
17	60	24



Careful observation of the resulting scatterplot emphatically confirms that the relationship connecting hours worked and reported happiness is decidedly *not* linear. The data points clearly delineate an inverted "U" shape, providing unequivocal support for the decision that **quadratic regression** represents the most statistically appropriate modeling technique for analyzing this specific dataset.

Preparing the Quadratic Predictor Variable

The core algebraic representation of a quadratic regression equation is defined as: **yhat = b0 + b1x1 + b2x1**^{2}. To successfully solve and estimate the coefficients of this equation utilizing Excel's standard linear regression tool, we are required to treat the squared predictor term ($x?^2$) as an entirely separate independent variable within the spreadsheet setup. Consequently, the next necessary preparatory step involves systematically modifying our existing dataset to explicitly include a new column that meticulously calculates and represents the square of the original predictor variable (Hours Worked).

We must first logically organize the columns to meet the structural requirements of the Excel Regression Analysis tool. Since the tool mandates that all predictor variables (the X variables) must be placed in adjacent columns, we will temporarily relocate the Happiness data from Column B to Column C. Highlight the existing values in column B and move them to column C. This action effectively clears column B, dedicating it for the insertion of our new calculated squared term:

	A	B	C	D
1	Hours worked		Reported Happiness	
2	6		12	
3	9		18	
4	12		30	
5	12		42	
6	15		48	
7	21		78	
8	24		90	
9	24		96	
10	27		96	
11	30		90	
12	36		84	
13	39		78	
14	45		66	
15	48		54	
16	57		36	
17	60		24	
18				

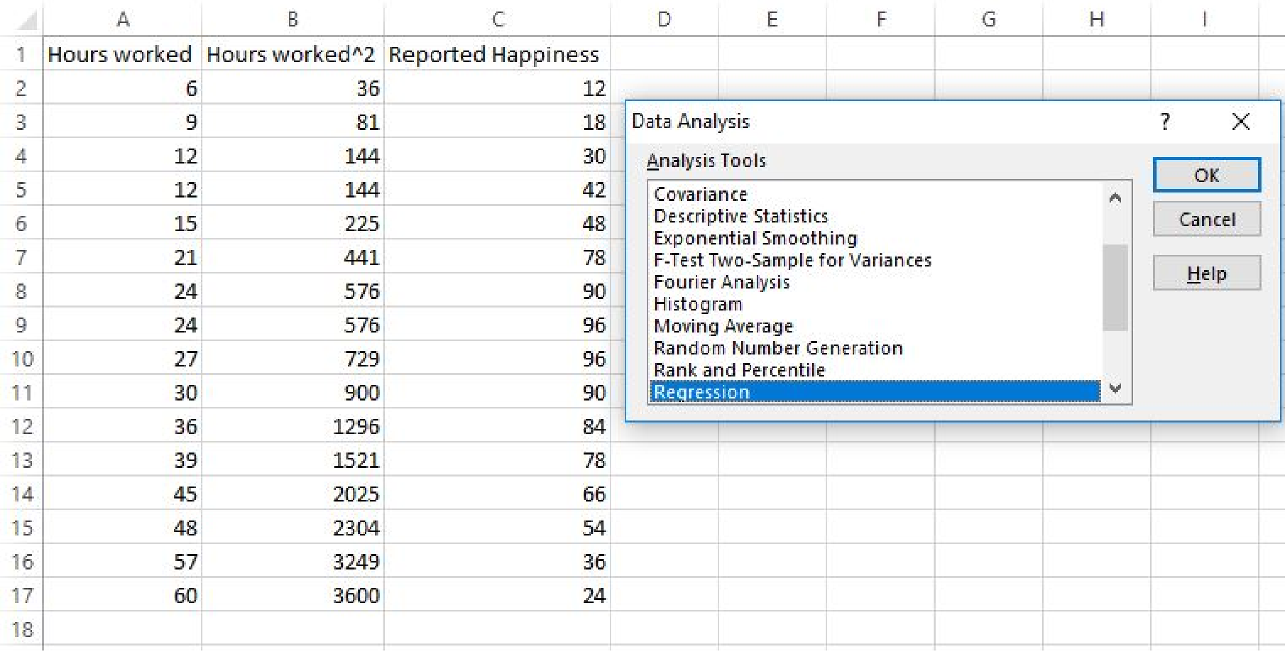
Next, we proceed to calculate the squared values of the Hours Worked variable. In cell B2, enter the simple formula: **=A2^2**. This calculation instructs Excel to take the value housed in cell A2 (the hours worked for the first observation) and square it, producing the result **36**. After the formula is entered, click and hold the bottom right corner of cell B2 (known as the fill handle) and drag the formula downwards, ensuring it populates the remainder of Column B. This essential step guarantees that every single observation in the dataset now has a corresponding squared value, thereby completing the necessary preparation of our independent variables for the final regression model.

	A	B	C	D
1	Hours worked	Hours worked^2	Reported Happiness	
2	6	36	12	
3	9	81	18	
4	12	144	30	
5	12	144	42	
6	15	225	48	
7	21	441	78	
8	24	576	90	
9	24	576	96	
10	27	729	96	
11	30	900	90	
12	36	1296	84	
13	39	1521	78	
14	45	2025	66	
15	48	2304	54	
16	57	3249	36	
17	60	3600	24	
18				
19				

Executing the Regression Analysis Using the Data Analysis ToolPak

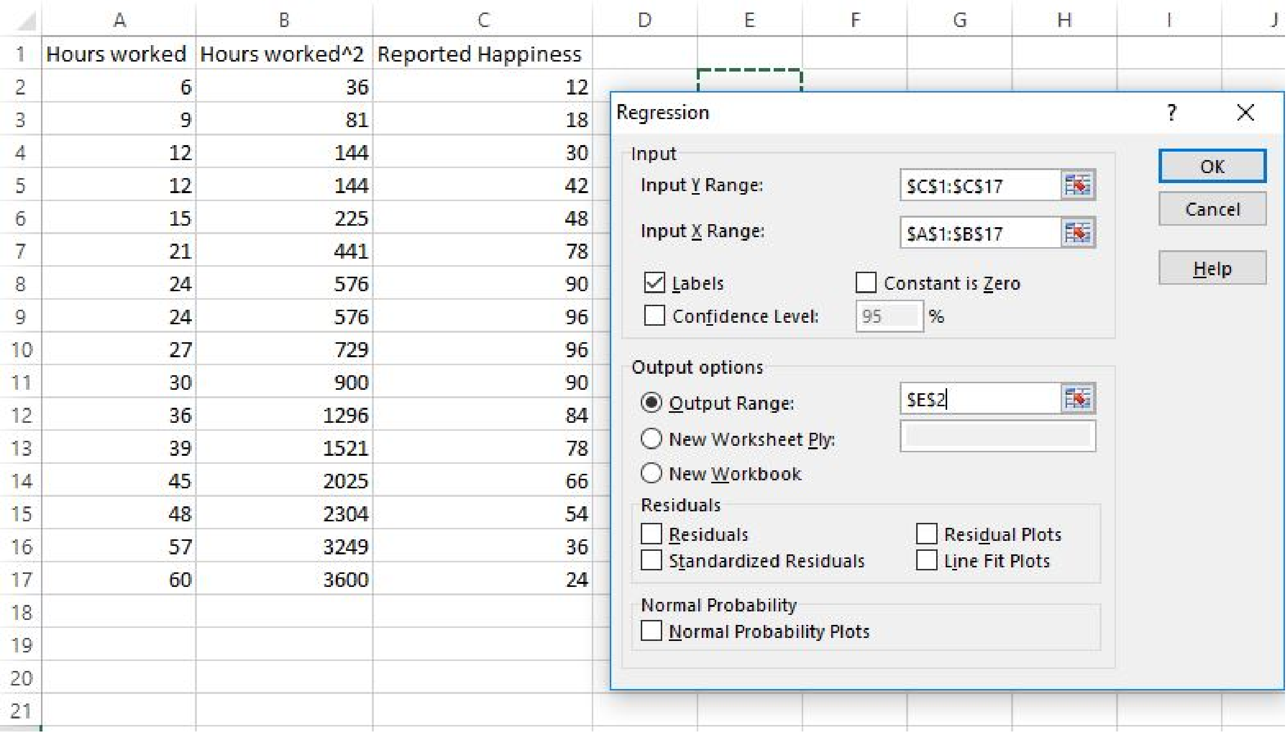
With the data now correctly and meticulously structured--featuring the response variable (Happiness Level) in one column and the two required predictor variables (Hours Worked and Hours Worked Squared) situated in adjacent columns--we are fully prepared to execute the **quadratic regression** using Excel's robust, built-in statistical capabilities. This advanced process fundamentally relies upon the Data Analysis ToolPak, an indispensable add-in required for advanced statistical computation within the Excel environment. It is crucial to verify that the *Data Analysis* option is visible in your top ribbon; if it is absent, you must first enable the ToolPak via the Excel Options menu before proceeding.

To initiate the formal analysis, click on the **DATA** tab located along the top ribbon, and subsequently select the *Data Analysis* option, which is typically found positioned on the far right. A comprehensive dialog box will immediately appear, providing a list of available statistical tools. From this extensive list, scroll down and explicitly select the **Regression** tool, and then click the *OK* button to proceed.



	A	B	C
1	Hours worked	Hours worked^2	Reported Happiness
2	6	36	12
3	9	81	18
4	12	144	30
5	12	144	42
6	15	225	48
7	21	441	78
8	24	576	90
9	24	576	96
10	27	729	96
11	30	900	90
12	36	1296	84
13	39	1521	78
14	45	2025	66
15	48	2304	54
16	57	3249	36
17	60	3600	24

In the subsequent *Regression* input dialog box, it is essential to define the data ranges for both the response variable and the predictor variables with precision. The **Input Y Range** must accurately encompass the response variable (Happiness Level), corresponding to cells **C1:C17** (including the column header). Crucially, the **Input X Range** must contain *both* predictor variables necessary for the quadratic model: the original Hours Worked and the newly calculated Hours Worked Squared. This range corresponds precisely to cells **A1:B17**. Ensure that the *Labels* checkbox is selected, acknowledging the inclusion of the header rows in the defined ranges. Finally, specify an appropriate *Output Range* (for example, cell E1) to display the comprehensive results clearly on the current worksheet, and then click *OK* to successfully run the **quadratic regression** model.



	A	B	C
1	Hours worked	Hours worked^2	Reported Happiness
2	6	36	12
3	9	81	18
4	12	144	30
5	12	144	42
6	15	225	48
7	21	441	78
8	24	576	90
9	24	576	96
10	27	729	96
11	30	900	90
12	36	1296	84
13	39	1521	78
14	45	2025	66
15	48	2304	54
16	57	3249	36
17	60	3600	24

Interpreting the Key Statistical Outputs

Following the successful execution of the model, Excel automatically generates a detailed output table that includes various crucial statistical metrics necessary for evaluating the quality, fit, and statistical significance of the resultant **quadratic regression** model. A thorough understanding of these metrics is absolutely vital for drawing meaningful and statistically defensible conclusions regarding the modeled relationship between hours worked and reported happiness.

E	F	G	H	I	J	K	L	M
SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.953526							
R Square	0.909211							
Adjusted R Square	0.895244							
Standard Error	9.519413							
Observations	16							
ANOVA								
	df	*SS*	*MS*	*F*	*gnificance F*			
Regression	2	11797.7	5898.85	65.09491	1.69E-07			
Residual	13	1178.05	90.61923					
Total	15	12975.75						
	Coefficients	*andard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	-30.2529	8.766065	-3.45113	0.004299	-49.1908	-11.3149	-49.1908	-11.3149
Hours worked	7.173061	0.631631	11.35642	4.04E-08	5.808506	8.537616	5.808506	8.537616
Hours worked^2	-0.10699	0.009461	-11.3078	4.25E-08	-0.12743	-0.08655	-0.12743	-0.08655

The output commences with the Regression Statistics block, which features the critical metric known as **R Square**. Also formally known as the coefficient of determination, this value quantifies the proportion of the total variance observed in the response variable that can be statistically explained or accounted for by the combined predictor variables (X and X^2). In the context of this specific example, the R-Square value is exceptionally high at **0.9092**. This high value directly indicates that 90.92% of the variation observed in reported happiness levels is successfully accounted for by the number of hours worked and its derived squared term. Such a high coefficient strongly suggests an excellent statistical fit of the quadratic model to the observed data points.

Another critical measure provided is the **Standard Error** of the regression. This value serves to quantify the average distance that the actual observed data points deviate or fall from the estimated regression line. It operates as a direct and immediate measure of the model's overall precision and reliability. In our case, the standard error is calculated as **9.519 units**. Interpreted practically, this means that, on average, the model's predicted happiness level deviates from the actual observed happiness level by approximately 9.519 points on the scale. A primary goal in all regression modeling is to minimize the standard error, as a lower value signifies a tighter, more precise clustering of the actual observations around the calculated prediction line.

Finally, the ANOVA table furnishes the essential **F Statistic**, which is employed to test the overall statistical significance of the entire regression model. The F statistic is derived by dividing the mean square of the regression by the mean square of the residuals (Regression MS / Residual MS). In essence, this calculation formally tests the null hypothesis that all regression coefficients (excluding the intercept) are statistically zero, which would imply that the model, as a whole, possesses no predictive power whatsoever. Our calculated F statistic is **65.09**, and the corresponding p-value is extremely small (far less than 0.0001). Since this p-value is significantly below the conventional statistical significance threshold of 0.05, we definitively reject the null hypothesis, concluding that the overall **quadratic regression** model is statistically significant and highly valuable for predictive purposes.

Formulating and Applying the Quadratic Regression Equation

The most immediately practical and actionable component of the entire Excel output resides within the coefficients table. This table provides the necessary constants required to mathematically construct the estimated regression equation. These coefficients represent the intercept (b?) and the specific weights, or slopes, applied to the predictor variables (b? for X and b? for X^2). The standard, general form of the estimated quadratic equation remains:

yhat = b0 + b1x1 + b2x12

By carefully extracting the numerical values from the coefficients column within the Excel output, we can substitute them precisely into the general formula. The specific coefficients derived from our analysis are: Intercept (b?) = -30.252, X Variable 1 (Hours worked, b?) = 7.173, and X Variable 2 (Hours worked squared, b?) = -0.106.

Therefore, the specific estimated **quadratic regression** equation formulated for the purpose of predicting reported happiness based on work hours is:

reported happiness level = -30.252 + 7.173(Hours worked) - 0.106(Hours worked)2

This newly derived equation now functions as a powerful, empirically verified tool capable of predicting the expected happiness level for any number of hours worked within the range of our collected data. For instance, to calculate the expected happiness level of an individual hypothesized to work 30 hours per week, we substitute the value 30 into the equation: reported happiness level = -30.252 + 7.173(30) - 0.106$(30)^2$. The resulting calculated expected happiness level is precisely **88.649**. This calculated prediction is founded upon the strong, statistically significant relationship established and validated by the **quadratic regression** model, offering a far more accurate and justifiable estimate than any simpler linear model could possibly provide for this curvilinear data.

Additional Resources