

Learning Quadratic Regression Analysis in Stata: A Step-by-Step Tutorial

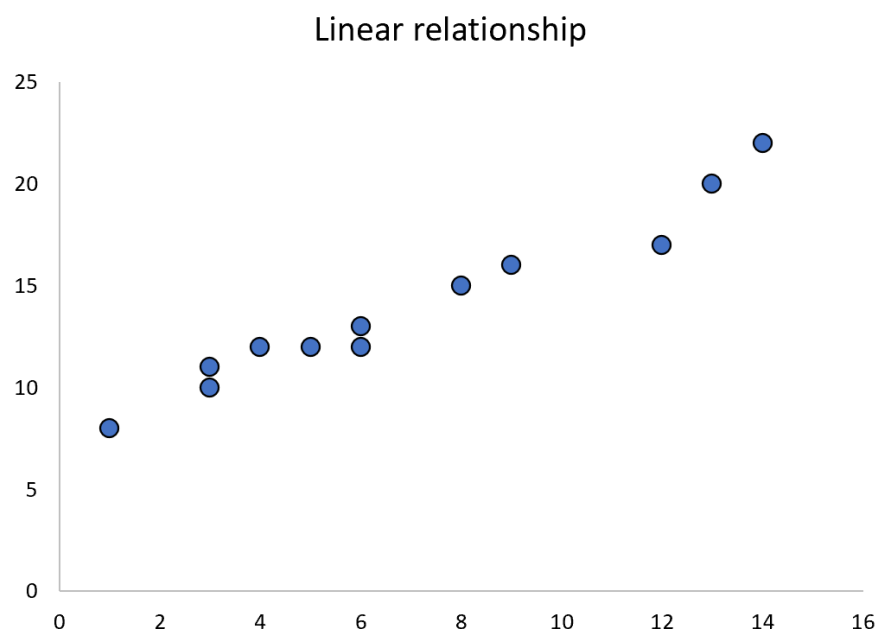
Authored by
Mohammed looti

November 8, 2025

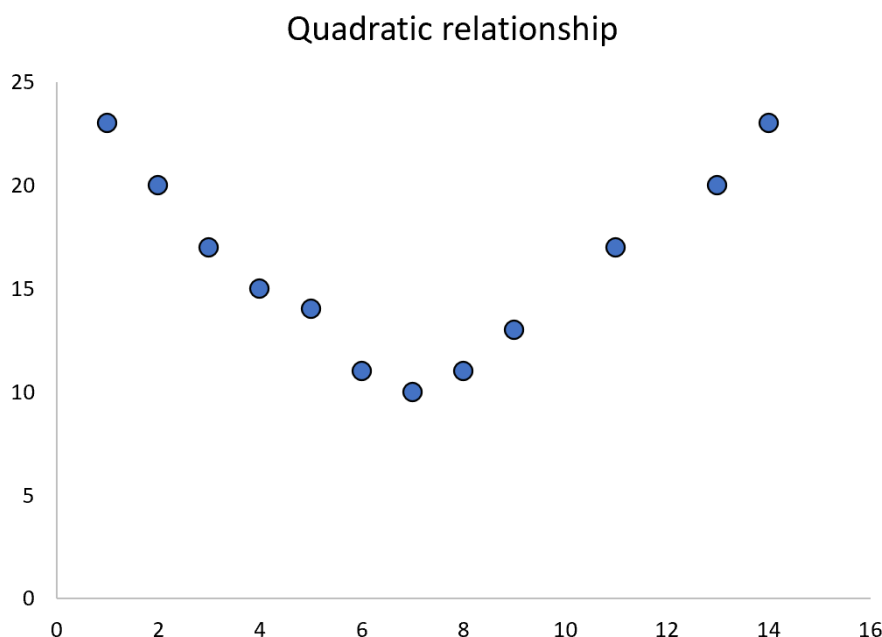
RECOMMENDED CITATION

Mohammed looti (2025). *Learning Quadratic Regression Analysis in Stata: A Step-by-Step Tutorial*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=13711>

In statistical modeling, determining the correct functional form is paramount. When analyzing the relationship between two variables, the simplest approach is often to assume a [linear relationship](#). This method, known as **linear regression**, is powerful for quantifying straight-line associations.



However, many real-world phenomena do not follow a simple linear path. When the relationship between two variables exhibits a curve--specifically a parabolic shape--we must transition from linear methods to [quadratic regression](#). This technique allows us to accurately model data where the effect of the predictor variable eventually diminishes or reverses after reaching a peak or trough.



This comprehensive tutorial serves as a step-by-step guide to performing and interpreting a **quadratic regression** analysis within the statistical software [Stata](#).

Understanding the Need for Quadratic Regression

Quadratic regression is a specialized form of polynomial regression that incorporates a squared term of the independent variable, enabling the model to capture curvilinear relationships effectively. This modeling approach is essential in fields like psychology, economics, and ecology where observed effects often peak or bottom out. For example, increasing input (such as fertilizer, advertising spend, or hours worked) might lead to positive results up to an optimal inflection point, after which additional input yields diminishing returns or even detrimental outcomes.

To illustrate this methodology, we will use a practical example exploring the connection between the number of hours an individual works per week and their self-reported happiness level. This case study demonstrates how to identify, model, and interpret a significant quadratic association using Stata's powerful command interface.

Setting Up the Data in Stata (The Happiness & Hours Dataset)

Our primary objective is to understand how weekly work hours influence happiness. We have collected observational data from 16 distinct participants, recording their total hours worked per week and their corresponding happiness level, measured on a standardized scale ranging from 0 (lowest happiness) to 100 (highest happiness). This dataset forms the foundation for performing the subsequent steps of the analysis in Stata.

	hours	happiness	
1	6	12	
2	9	18	
3	12	30	
4	12	42	
5	15	48	
6	21	78	
7	24	90	
8	24	96	
9	27	96	
10	30	90	
11	36	84	
12	39	78	
13	45	66	
14	48	54	
15	57	36	
16	60	24	

To replicate this analysis precisely, you must manually input this specific data into Stata. This procedure is accomplished by navigating through the top menu interface: **Data > Data Editor > Data Editor (Edit)**. Once the data is successfully entered and verified, we can proceed with the diagnostic and modeling steps required for fitting the quadratic equation.

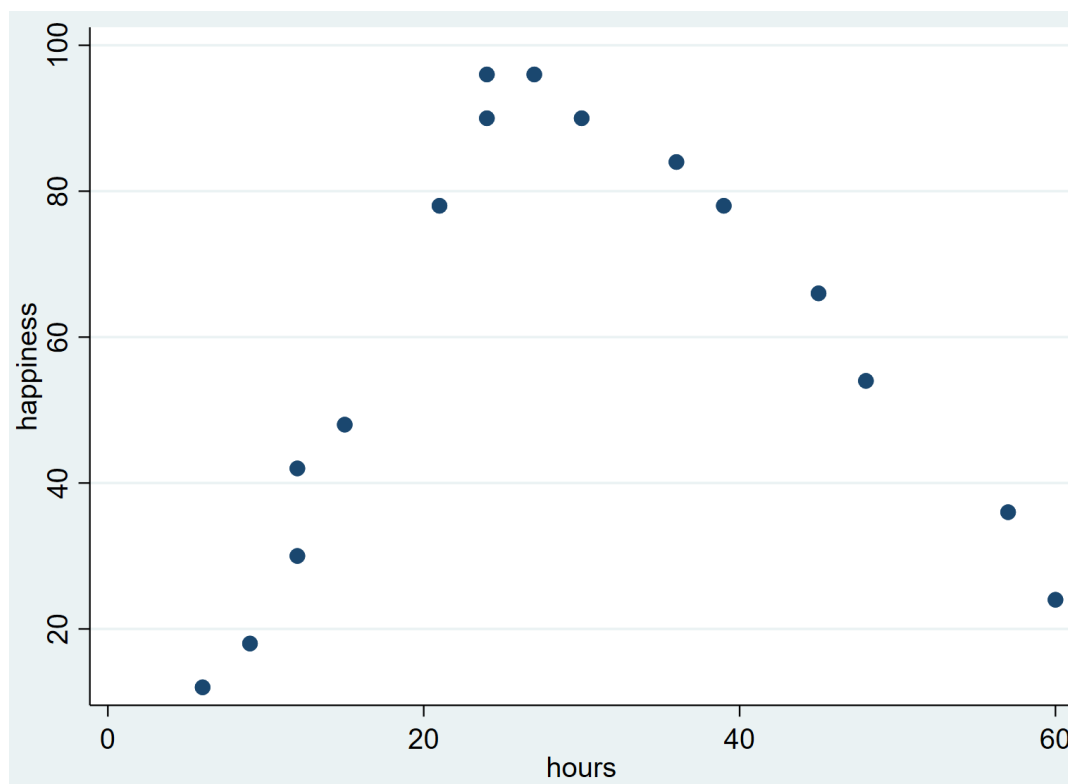
Step 1: Visualizing the Curvilinear Relationship

Before implementing any complex model, it is a crucial diagnostic step to visually confirm that the relationship between the [explanatory variable](#) (hours worked) and the [response variable](#) (happiness) is indeed curvilinear. A visual inspection using a [scatterplot](#) provides the necessary preliminary evidence for selecting a quadratic model over a linear one.

In the Stata Command box, execute the following command to generate the initial scatterplot:

scatter happiness hours

This command produces a graphical representation of the data distribution:



The resulting visualization clearly shows a distinct inverted "U" shape. Happiness tends to increase rapidly as weekly work hours rise from zero, reaching a perceptible maximum around 30 hours, after which happiness levels begin to drop significantly as hours increase further. This unmistakable pattern strongly supports the decision to use **quadratic regression** to accurately quantify this non-linear association, confirming that a simpler linear model would be inappropriate and misleading.

Step 2: Preparing the Model (Creating the Squared Predictor)

A fundamental requirement of quadratic regression is the inclusion of the squared term of the predictor variable. While a linear model relates Y to X , the quadratic model relates the response Y to both the linear predictor X and its squared counterpart, X^2 . Therefore, the next essential technical step in Stata is to create a new variable that holds the square of our existing predictor variable, *hours*.

To generate this new variable, which we will name *hours2*, input the following command into the Stata Command box:

```
gen hours2 = hours*hours
```

To verify the successful creation and calculation of this new variable, you can inspect the data

editor by navigating to **Data > Data Editor > Data Editor (Browse)**. The dataset should now include the *hours2* column, where each entry is precisely the square of the corresponding *hours* value, ensuring the model is ready for execution.

	hours	happiness	hours2	
1	6	12	36	
2	9	18	81	
3	12	30	144	
4	12	42	144	
5	15	48	225	
6	21	78	441	
7	24	90	576	
8	24	96	576	
9	27	96	729	
10	30	90	900	
11	36	84	1296	
12	39	78	1521	
13	45	66	2025	
14	48	54	2304	
15	57	36	3249	
16	60	24	3600	

Step 3: Executing and Interpreting the Quadratic Regression Model

With the squared predictor variable (*hours2*) successfully generated, we are now prepared to execute the quadratic regression. In this model specification, *happiness* serves as the response variable, while both the linear term *hours* and the quadratic term *hours2* act as the explanatory variables. To fit the model and generate the output, use the standard **regress** command, ensuring both predictors are included:

regress happiness hours hours2

Stata will produce a detailed output table summarizing the fit statistics and the calculated coefficients for the quadratic model:

```
. regress happiness hours hours2
```

Source	SS	df	MS	Number of obs	=	16
Model	11797.7001	2	5898.85003	F(2, 13)	=	65.09
Residual	1178.04994	13	90.6192262	Prob > F	=	0.0000
				R-squared	=	0.9092
				Adj R-squared	=	0.8952
Total	12975.75	15	865.05	Root MSE	=	9.5194

happiness	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
hours	7.173061	.6316306	11.36	0.000	5.808506	8.537616
hours2	-.1069887	.0094615	-11.31	0.000	-.127429	-.0865484
_cons	-30.25287	8.766065	-3.45	0.004	-49.1908	-11.31493

The interpretation of this output focuses on several key statistical indicators required for drawing sound conclusions:

Prob > F (Overall Model Significance): The reported value is 0.000. Since this [p-value](#) is substantially less than the conventional significance level of 0.05, it indicates that the overall model is statistically significant. This confirms that the combined predictors (*hours* and *hours*²) have a relationship with the response variable (*happiness*) that is statistically reliable.

R-squared (Coefficient of Determination): The value is 0.9092, which suggests a remarkably strong model fit. The [R-squared](#) statistic quantifies the proportion of the variance in the response variable that can be explained by the predictors. In this example, 90.92% of the total variation observed in happiness levels is accounted for by the quadratic relationship with weekly hours worked.

Step 4: Applying and Reporting the Results

The culmination of the analysis is the construction of the specific [regression equation](#) using the coefficient values reported in the Stata output table. This mathematical formula allows us to predict the happiness level for any individual given their number of hours worked per week. The equation follows the general quadratic form, incorporating the constant and the coefficients for both the linear and squared terms:

$$\text{predicted happiness} = -30.25287 + 7.173061(\text{hours}) - 0.1069887(\text{hours}^2)$$

We can utilize this equation to generate specific predictions. For instance, calculating the predicted happiness for an individual working 60 hours per week demonstrates the diminishing returns of excessive work:

predicted happiness = $-30.25287 + 7.173061(60) - 0.1069887(60^2) = \mathbf{14.97}$.

Conversely, for an individual working 30 hours per week--which aligns with the peak of the observed curve--the predicted happiness is considerably higher:

predicted happiness = $-30.25287 + 7.173061(30) - 0.1069887(30^2) = \mathbf{88.65}$.

Finally, the results must be reported in a formal, concise manner, typically including the overall F-statistic, degrees of freedom, and p-value to fully quantify the relationship. An example of a formal summary of the quadratic regression analysis is provided below:

A quadratic regression analysis was performed to quantify the relationship between the number of hours worked by an individual and their corresponding happiness level (measured from 0 to 100). A sample of 16 individuals was used in the analysis.

Results showed that there was a statistically significant curvilinear relationship between the explanatory variables *hours* and *hours*² and the response variable *happiness* ($F(2, 13) = 65.09$, $p < 0.0001$).

Combined, these two explanatory variables accounted for 90.92% of the explained variability in happiness ($R^2 = 0.9092$).

The derived regression equation for predicting happiness based on weekly hours worked is:
predicted happiness = $\mathbf{-30.25287 + 7.173061(hours) - 0.1069887(hours^2)}$.