

How to Run a Quadratic Regression in Stata: A Step-by-Step Guide

Authored by
stats writer

December 28, 2025

RECOMMENDED CITATION

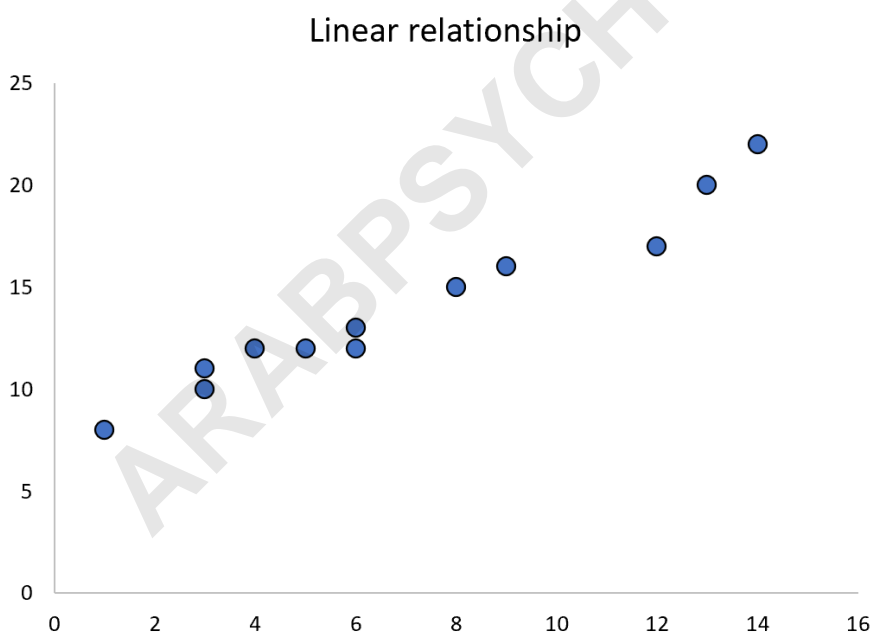
stats writer (2025). *How to Run a Quadratic Regression in Stata: A Step-by-Step Guide*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=109591>

Performing **Quadratic Regression** in **Stata** is a powerful statistical technique used when the relationship between your measured variables follows a curved path, rather than a simple straight line. This method involves executing a standard **regression** analysis but crucially includes both the original **independent variables** and their mathematically squared terms as distinct predictors.

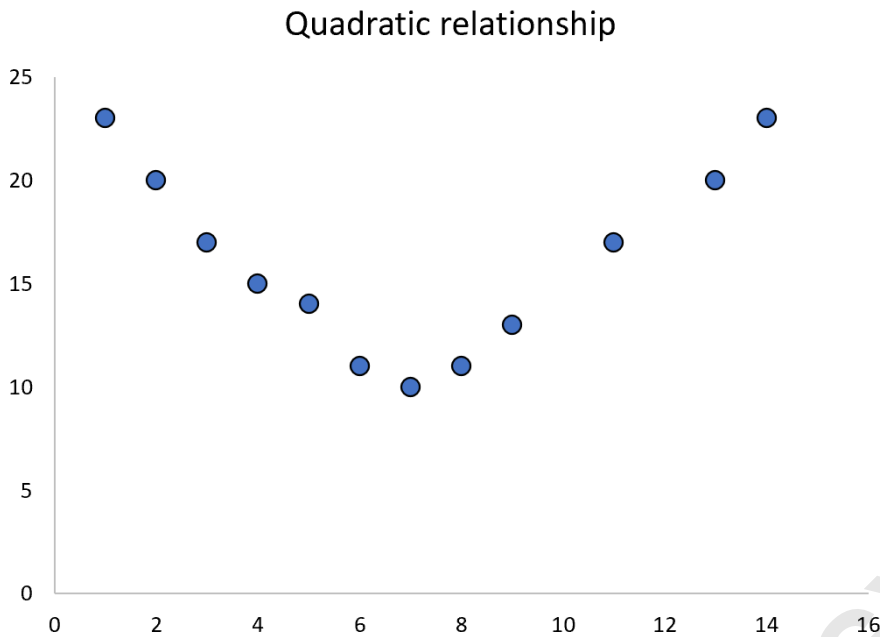
The resulting **coefficients** provide sophisticated insights: the coefficient for the base independent variable captures the linear rate of change, while the coefficient for the squared term quantifies the non-linear curvature, or the quadratic effect. In Stata, this advanced modeling is achieved simply by utilizing the standard `regress` command, provided you have first generated the necessary squared predictor variable.

Understanding Linear vs. Quadratic Relationships

Before implementing a quadratic model, it is crucial to understand the visual distinction between linear and curvilinear data patterns. When two variables exhibit a **linear relationship**, their association can be accurately modeled using simple linear regression, where the rate of change remains constant across the entire range of the independent variable. This relationship is represented graphically by a straight line:



However, many observational phenomena, particularly in social sciences, display complex associations characterized by diminishing returns or non-linear growth. When a relationship follows a distinctive curved path--specifically a parabolic or "U"-shaped curve--it strongly suggests a quadratic model is required. In these circumstances, standard linear models are insufficient, and you must utilize **quadratic regression** to quantify the complexity of the association accurately.



The following tutorial provides a practical, step-by-step walkthrough detailing how to execute and interpret a quadratic regression analysis specifically within the Stata statistical software environment.

Case Study: Modeling Hours Worked and Happiness

To demonstrate the practical application of this technique, we will use a classic example focusing on the relationship between labor input and psychological well-being. Our objective is to determine the relationship between the number of hours an individual works per week and their self-reported level of happiness. We hypothesize that happiness increases up to an optimal workload but then decreases significantly due to strain--a classic inverted "U" shape, indicating a quadratic effect.

We have compiled a sample dataset from 16 individuals, recording their weekly hours worked and their corresponding happiness scores (measured on a validated scale from 0 to 100):

	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	

This example can be easily replicated. To input the data shown above, open Stata and navigate to **Data > Data Editor > Data Editor (Edit)** along the top menu bar to manually create and populate the variables *hours* and *happiness*.

Follow the subsequent steps to execute the quadratic regression analysis in Stata.

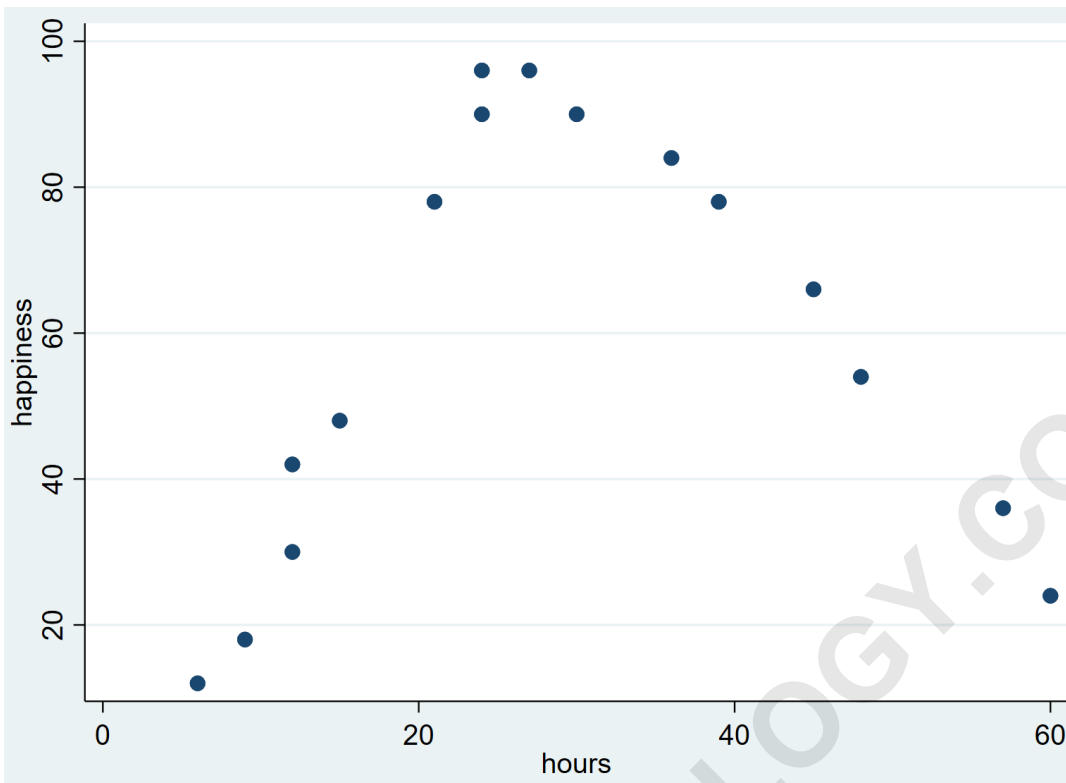
Step 1: Visual Confirmation via Scatterplot

The initial step in fitting any complex model is data exploration and visual validation. We must confirm that the pattern between our **explanatory variable** (*hours*) and the **response variable** (*happiness*) is indeed quadratic. If the relationship appeared linear, a simpler model would suffice.

To create a preliminary visual assessment, generate a **scatterplot** by executing the following command in the Stata Command window:

```
scatter happiness hours
```

This command produces the visualization below, allowing for an immediate inspection of the underlying trend:



Interpreting the Curvilinear Pattern

The resulting scatterplot clearly supports the hypothesis of a curvilinear relationship. We observe that happiness levels demonstrate a positive correlation with hours worked, rising steadily as hours increase from zero up to an apparent peak.

Critically, an inflection point is observed around 30 hours per week. Beyond this point, the trend sharply reverses: happiness begins to decline significantly as work hours continue to increase past this optimal point. This pronounced inverted "U" shape confirms the presence of a **quadratic relationship**, thereby validating the need for quadratic regression modeling.

Step 2: Generating the Squared Predictor Variable

To incorporate the quadratic effect into the regression model, we must explicitly create a new variable that represents the square of our predictor variable, `hours`. We use Stata's powerful data manipulation command, `gen` (generate), for this purpose.

Enter the following syntax into the Command box to create the variable `hours2`:

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

You can confirm the successful creation of this variable by inspecting your dataset. Go to **Data >**

Data Editor > Data Editor (Browse). As shown below, the new variable `hours2` is now included, containing the squared value of the corresponding observation in the `hours` column, making it ready for use in the regression:

	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: Running the Quadratic Regression Model

Now that the squared predictor variable is available in the dataset, we can proceed to fit the quadratic model. The `regress` command is executed by listing the **response variable** first, followed by all **explanatory variables**. For this model, we include `happiness` as the dependent variable, and both `hours` and `hours2` as the predictors.

Execute the following command to perform the quadratic regression analysis:

```
regress happiness hours hours2
```

Stata processes the data and generates a comprehensive output table detailing the ANOVA, model fit statistics, and the calculated coefficients for each term:

```
. 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

Interpreting the Stata Output Statistics

The output table contains several critical statistics necessary for evaluating the quality and validity of the fitted quadratic model. We focus specifically on the overall model significance and the coefficient of determination.

Overall Significance (Prob > F): This value represents the **p-value** for the F-test, which assesses whether the predictors (`hours` and `hours2`) collectively have a significant relationship with `happiness`. The obtained value of 0.0000 is significantly below the standard 0.05 alpha level, leading us to conclude that the overall quadratic regression model is highly **statistically significant**.

Model Fit (R-squared): The R-squared value is 0.9092. This statistic measures the proportion of the total variability in the response variable (`happiness`) that is explained by the model's predictors. In this case, 90.92% of the variation in individual happiness scores is successfully accounted for by the quadratic function of hours worked, indicating an exceptionally strong model fit.

Step 4: Formulating and Applying the Regression Equation

We construct the specific quadratic prediction model using the coefficients reported in the output table. The general algebraic form is $Y = b_0 + b_1X + b_2X^2$. By substituting the intercept and the coefficients for `hours` and `hours2`, we derive the exact predictive model:

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

This formula is now a powerful tool for forecasting happiness given any specific number of hours worked per week. We can utilize it to contrast the predicted outcomes for low, moderate, and

extreme workloads:

Prediction for Optimal Workload (30 hours per week): Substituting 30 into the equation, which corresponds to the observed peak of the curve:

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

Prediction for Extreme Workload (60 hours per week): Substituting 60, where we expect significant burnout and decline in happiness:

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

These predictions confirm the quadratic nature of the relationship, showing that doubling the workload from 30 to 60 hours per week results in a massive drop in predicted happiness due to the increasing negative quadratic effect.

Step 5: Reporting the Results in Academic Format

The final and most professional step is to summarize the findings of the quadratic regression concisely and accurately. The report must include the statistical test results, model fit indices, and the final predictive equation. Below is an example of standard reporting language:

A quadratic regression 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** relationship between the explanatory variables *hours* and *hours2* and the response variable *happiness* ($F(2, 13) = 65.09, p < 0.0001$).

Combined, these two explanatory variables accounted for 90.92% of explained variability in happiness.

The regression equation was found to be:

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