

Learning Quadratic Regression in R: A Step-by-Step Guide

Authored by
Mohammed loot

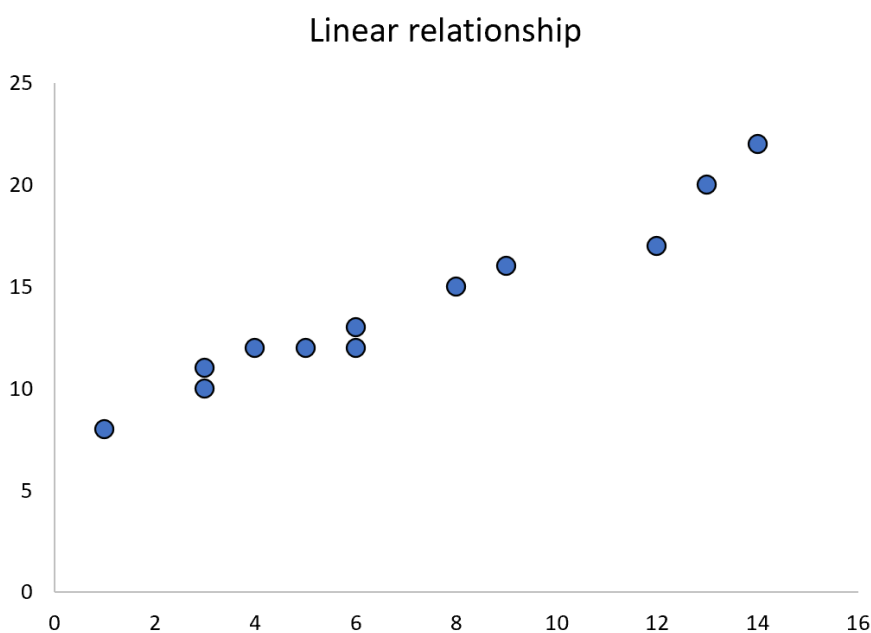
November 8, 2025

RECOMMENDED CITATION

Mohammed loot (2025). *Learning Quadratic Regression in R: A Step-by-Step Guide*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=13033>

When analyzing relationships between variables in statistics, we frequently begin by assuming a linear correlation. If this assumption holds true, tools like [simple linear regression](#) provide a powerful framework for quantifying the relationship and making predictions. A linear relationship implies that a change in the independent variable results in a constant, proportional change in the dependent variable.

However, not all real-world phenomena follow a straight line. Many relationships exhibit curvature, where the effect of the independent variable diminishes or reverses after a certain point. It is critical to visualize the data first to determine if a simple linear model is appropriate, or if a more complex, non-linear approach is necessary to accurately capture the underlying structure of the data.



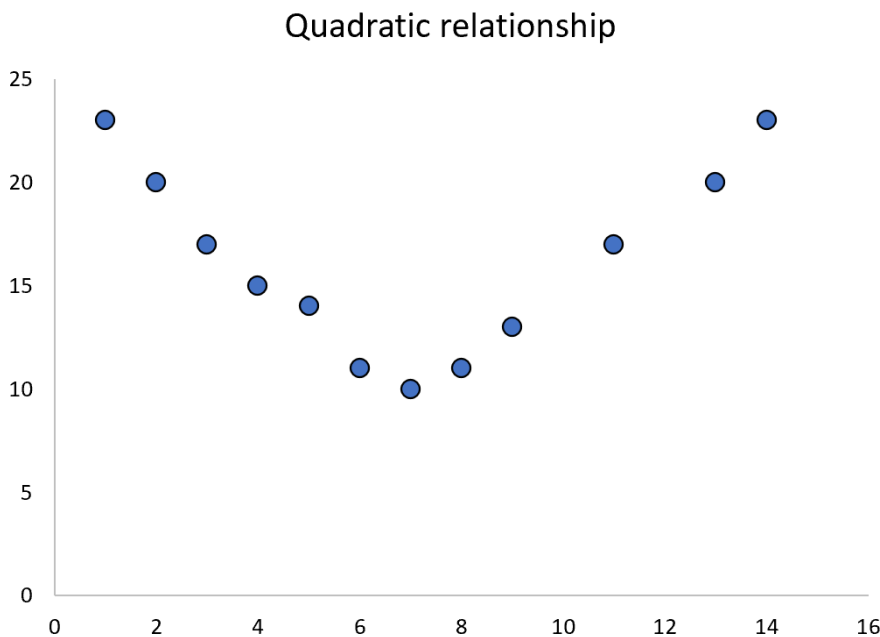
The Need for Quadratic Regression

In cases where the data clearly displays a curved pattern, often resembling a parabola (U-shape or inverted U-shape), relying on linear regression will lead to poor model fit and inaccurate conclusions. This is where [quadratic regression](#) becomes essential.

[Quadratic regression](#), a form of polynomial regression, models the relationship between the independent variable (X) and the dependent variable (Y) using a second-degree polynomial. This technique allows us to capture the crucial curvature that simple linear models cannot, significantly improving the explanatory power of our analysis. The mathematical form of a quadratic regression model is typically represented as:
$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \epsilon$$

This tutorial will guide you through the process of performing a [quadratic regression](#) analysis within the statistical environment, [R](#). We will demonstrate how to structure the model to include a squared

term and how to interpret the resulting statistical output, contrasting its effectiveness with the results from a standard linear model.



Case Study: Hours Worked vs. Happiness

To illustrate the application of [quadratic regression](#), let us consider a hypothetical study investigating the relationship between the number of hours an individual works per week and their reported level of happiness. It is often hypothesized that happiness initially increases as hours worked increase (due to productivity or financial stability), but eventually decreases as working hours become excessive, leading to burnout. This type of relationship suggests an inverted U-shaped, or quadratic, curve.

We have collected data from 11 individuals, recording the number of hours they work weekly and their self-reported happiness level, measured on a scale from 0 to 100. This dataset provides the foundation for our statistical modeling exercise in [R](#).

Hours	Happiness
6	14
9	28
12	50
14	70
30	89
35	94
40	90
47	75
51	59
55	44
60	27

Step 1: Inputting and Structuring the Data in R

The initial step in any statistical analysis involves accurately inputting the data into the chosen statistical environment. Using [R](#), we will create a data frame named `data` to store our two primary variables: `hours` (the independent variable) and `happiness` (the dependent variable). Ensuring the data is correctly structured is paramount before proceeding with visualization or modeling.

The following [R](#) commands define the data frame and then display the structured data, confirming that all observations have been correctly loaded into the environment.

#create data

```
data <- data.frame(hours=c(6, 9, 12, 14, 30, 35, 40, 47, 51, 55, 60),  
happiness=c(14, 28, 50, 70, 89, 94, 90, 75, 59, 44, 27))
```

#view data

```
data
```

```
hours happiness
```

```
1 6 14
```

```
2 9 28
```

```
3 12 50
```

```
4 14 70
```

```
5 30 89
```

```
6 35 94
```

```
7 40 90
```

```
8 47 75
```

```
9 51 59
```

```
10 55 44
```

11 60 27

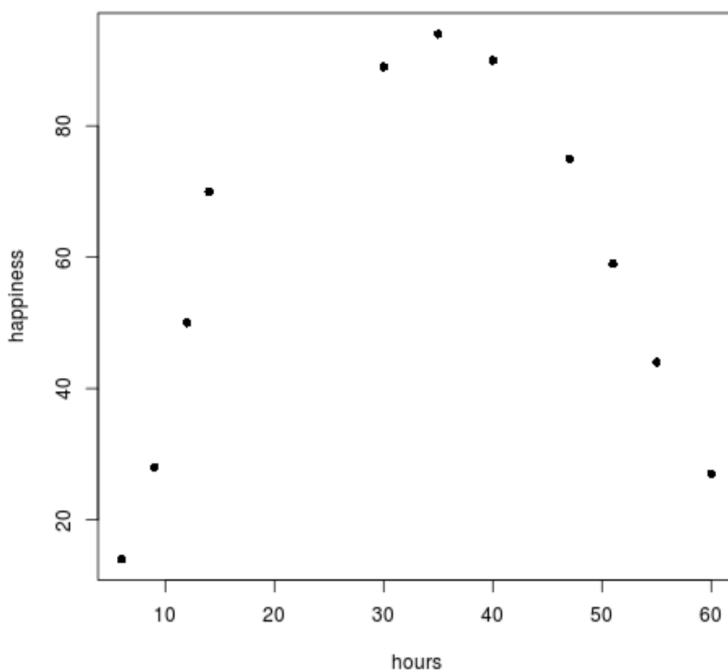
Step 2: Visualizing the Relationship and Testing Linearity

Before fitting any model, data visualization is a crucial diagnostic step. We must create a [scatterplot](#) to visually inspect the form of the relationship between hours worked and happiness. The visual evidence often provides the strongest initial indicator of whether a linear, quadratic, or other non-linear model is required.

The [scatterplot](#) generated below clearly demonstrates a curvilinear pattern; specifically, the data points rise steeply, reach a peak around 35-40 hours, and then begin to decline. This inverted U-shape confirms our hypothesis that a simple linear model will likely be insufficient for capturing the variance in happiness.

#create scatterplot

```
plot(data$hours, data$happiness, pch=16)
```



To quantify the inadequacy of the linear assumption, we fit a simple linear regression model using R's built-in `lm()` function. As expected from the visualization, the summary statistics reveal that the linear model explains very little of the variance in happiness.

#fit linear model

```
linearModel <- lm(happiness ~ hours, data=data)
```

```
#view model summary
summary(linearModel)
```

Call:

```
lm(formula = happiness ~ hours)
```

Residuals:

```
Min 1Q Median 3Q Max
```

```
-39.34 -21.99 -2.03 23.50 35.11
```

Coefficients:

```
Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept) 48.4531 17.3288 2.796 0.0208 *
```

```
hours 0.2981 0.4599 0.648 0.5331
```

```
---
```

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 28.72 on 9 degrees of freedom

Multiple R-squared: 0.0446, Adjusted R-squared: -0.06156

F-statistic: 0.4201 on 1 and 9 DF, p-value: 0.5331

The output confirms the visual assessment: the total variance in happiness explained by this simple linear model is merely **4.46%**, as indicated by the Multiple [R-squared](#) value. This low value necessitates moving forward with a non-linear approach.

Step 3: Fitting the Quadratic Regression Model

To perform [quadratic regression](#) in R, we must first create a new variable that represents the independent variable squared (X^2). This variable will be included alongside the original linear term (X) in our regression formula. This technique transforms the problem into a multiple linear regression where the predictors are X and X^2 , allowing us to use the same `lm()` function.

We define the new column `hours2` by squaring the values in the `hours` column. We then fit the `quadraticModel` using the formula `happiness ~ hours + hours2`. The inclusion of the squared term is the defining characteristic of the quadratic model.

```
#create a new variable for hours2
```

```
data$hours2 <- data$hours^2
```

```
#fit quadratic regression model
```

```
quadraticModel <- lm(happiness ~ hours + hours2, data=data)
```

```
#view model summary
summary(quadraticModel)

Call:
lm(formula = happiness ~ hours + hours2, data = data)

Residuals:
Min 1Q Median 3Q Max
-6.2484 -3.7429 -0.1812 1.1464 13.6678

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) -18.25364 6.18507 -2.951 0.0184 *
hours 6.74436 0.48551 13.891 6.98e-07 ***
hours2 -0.10120 0.00746 -13.565 8.38e-07 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.218 on 8 degrees of freedom
Multiple R-squared: 0.9602, Adjusted R-squared: 0.9502
F-statistic: 96.49 on 2 and 8 DF, p-value: 2.51e-06
```

The resulting model summary shows a dramatic improvement in model fit. The total variance in happiness explained by the model skyrocketed to **96.02%**, a massive leap from the 4.46% achieved by the simple linear model. Furthermore, the coefficients for both `hours` and `hours2` are statistically significant ($P < 0.001$), confirming that the quadratic term is essential for explaining the variance in happiness.

We can visually confirm the superior fit of the quadratic model by plotting the regression curve onto the original [scatterplot](#). We must first generate a sequence of predicted values across the range of hours worked using the `predict()` function.

```
#create sequence of hour values
```

```
hourValues <- seq(0, 60, 0.1)
```

```
#create list of predicted happiness levels using quadratic model
```

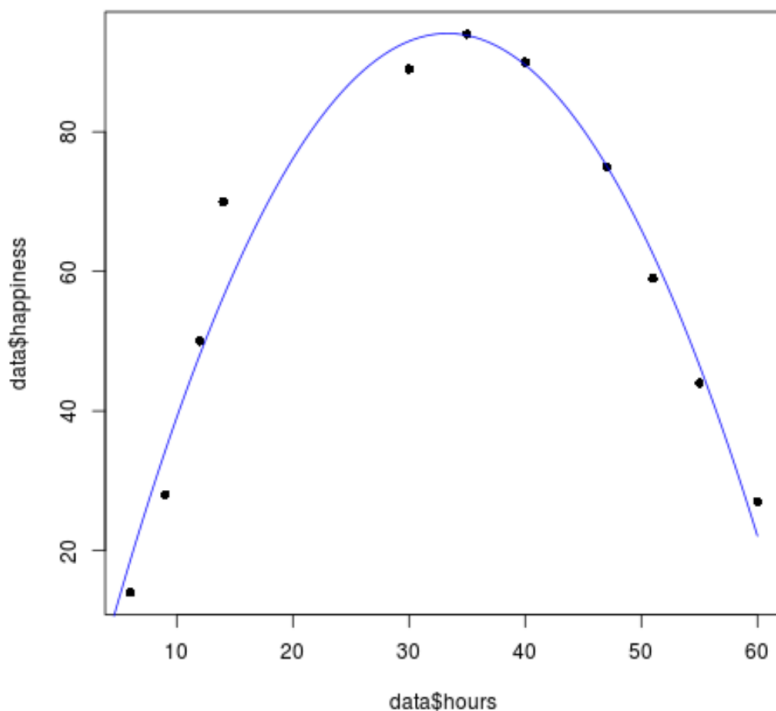
```
happinessPredict <- predict(quadraticModel,list(hours=hourValues, hours2=hourValues^2))
```

```
#create scatterplot of original data values
```

```
plot(data$hours, data$happiness, pch=16)
```

```
#add predicted lines based on quadratic regression model
```

```
lines(hourValues, happinessPredict, col='blue')
```



The blue curve demonstrates that the [quadratic regression](#) line follows the actual data points very closely, providing strong visual confirmation that this model form is appropriate for our variables.

Step 4: Interpreting and Applying the Model

The final step involves interpreting the coefficients derived from the [quadratic regression](#) output and using the resulting equation for prediction. The coefficients define the exact shape and position of the fitted parabolic curve.

The key coefficients from our model summary were:

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-18.25364	6.18507	-2.951	0.0184 *
hours	6.74436	0.48551	13.891	6.98e-07 ***
hours2	-0.10120	0.00746	-13.565	8.38e-07 ***

Based on these estimates, we can construct the precise fitted regression equation for predicting happiness (Y) based on hours worked (X):

$$\text{Happiness} = -0.1012(\text{hours})^2 + 6.7444(\text{hours}) - 18.2536$$

The interpretation of the coefficients in a [quadratic regression](#) is more complex than in a linear model because the effect of hours (X) on happiness (Y) is not constant; it depends on the current value of hours. Since the coefficient for the squared term (hours^2) is negative (-0.1012), the parabola opens downwards, confirming the inverted U-shaped relationship where happiness peaks and then declines.

This equation can now be utilized to predict the happiness level for any number of hours worked within the observed range. For example, consider an individual working 60 hours per week:

Prediction for 60 hours: Happiness = $-0.1012(60)^2 + 6.7444(60) - 18.2536 = 22.09$. This suggests a significantly lower happiness score due to overwork.

Conversely, let us examine an individual working 30 hours per week, a point closer to the peak of the curve:

Prediction for 30 hours: Happiness = $-0.1012(30)^2 + 6.7444(30) - 18.2536 = 92.99$. This predicted score is substantially higher, demonstrating the efficacy of the [quadratic regression](#) model in capturing the non-linear dynamics between these two variables.