When should you use polynomial regression?

Authored by

stats writer

May 11, 2024

RECOMMENDED CITATION

stats writer (2024). *When should you use polynomial regression?*. PSYCHOLOGICAL SCALES. Retrieved from https://scales.arabpsychology.com/?p=143650

Polynomial regression is a statistical method used to model relationships between variables when the data points do not follow a straight line. It is commonly used in situations where the relationship between the variables is believed to be nonlinear. This regression technique is useful when the data has multiple peaks or curves, and a traditional linear regression model would not accurately capture the underlying pattern. Polynomial regression can also be used when there is a strong correlation between the independent and dependent variables, but the relationship is not strictly linear. It is particularly effective in predicting outcomes in fields such as economics, engineering, and social sciences. Overall, polynomial regression is a powerful tool for analyzing and predicting complex relationships between variables and should be used when the data suggests a nonlinear relationship.

When Should You Use Polynomial Regression?

Polynomial regression is a technique we can use to fit a regression model when the relationship between the predictor variable(s) and the response variable is nonlinear.

A polynomial regression model takes the following form:

$$Y = \beta 0 + \beta 1X + \beta 2X2 + ... + \beta hXh + \varepsilon$$

In practice, there are three easy ways to determine if you should use polynomial regression compared to a simpler model like .

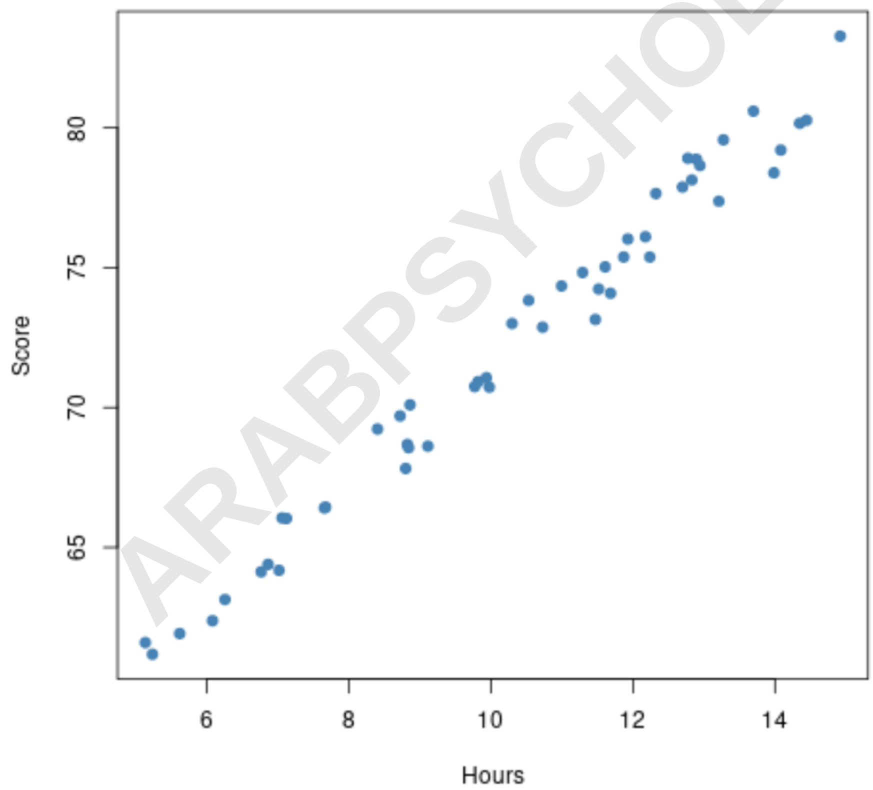
1. Create a Scatterplot of the Predictor Variable and Response Variable

The easiest way to determine if you should use

polynomial regression is to create a simple scatterplot of the predictor variable and the response variable.

For example, suppose we'd like to use the predictor variable "hours studied" to predict the score that a student will receive on a final exam.

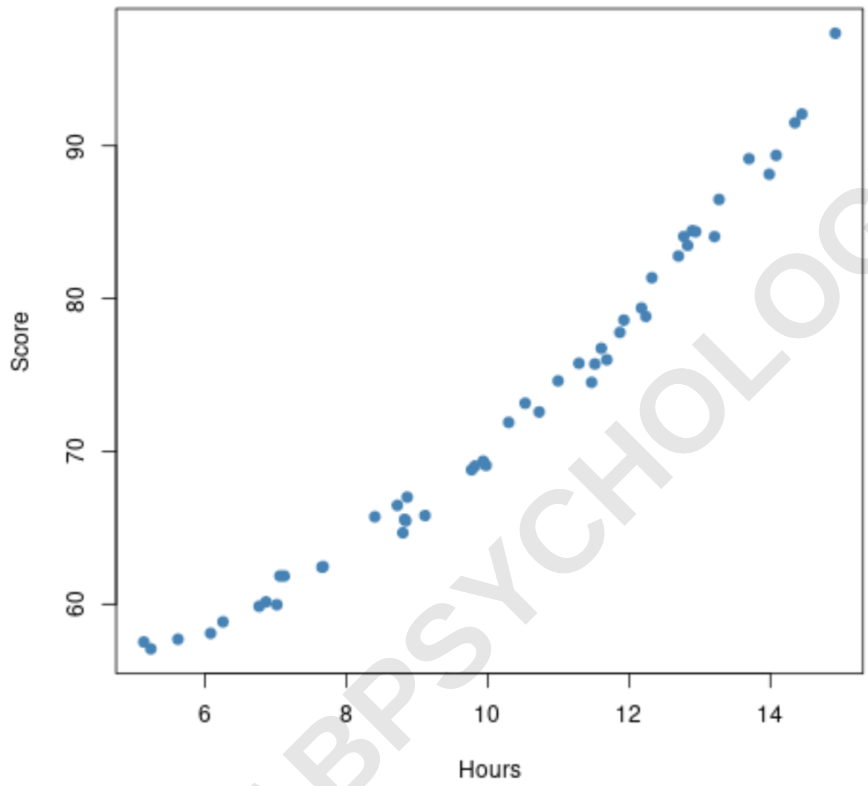
Before fitting a regression model, we can first create a scatterplot of hours studied vs. exam score. Suppose our scatterplot looks like the following:



The relationship between hours studied and exam score

looks linear, so it would make sense to fit a simple linear regression model to this dataset.

However, suppose the scatterplot actually looked like the following:



This relationship looks a bit more nonlinear, so this tells us that it may be a good idea to fit a polynomial regression model instead.

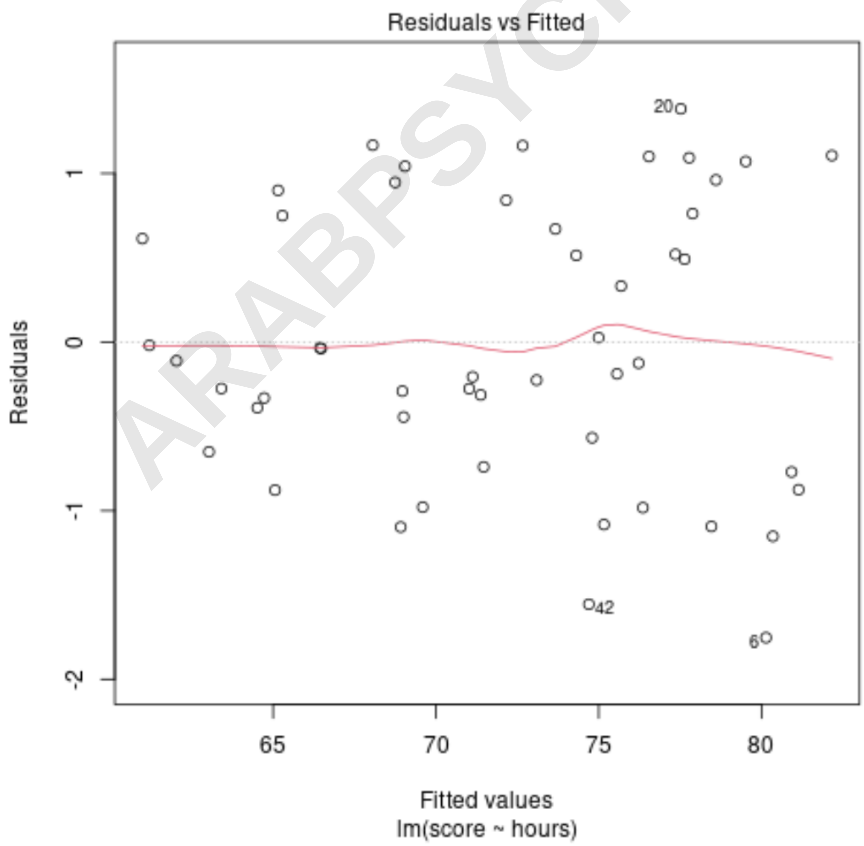
2. Create a Fitted Values vs. Residual Plot

Another way to determine if you should use polynomial

regression is to fit a linear regression model to the dataset and then created a fitted values vs. residuals plot for the model.

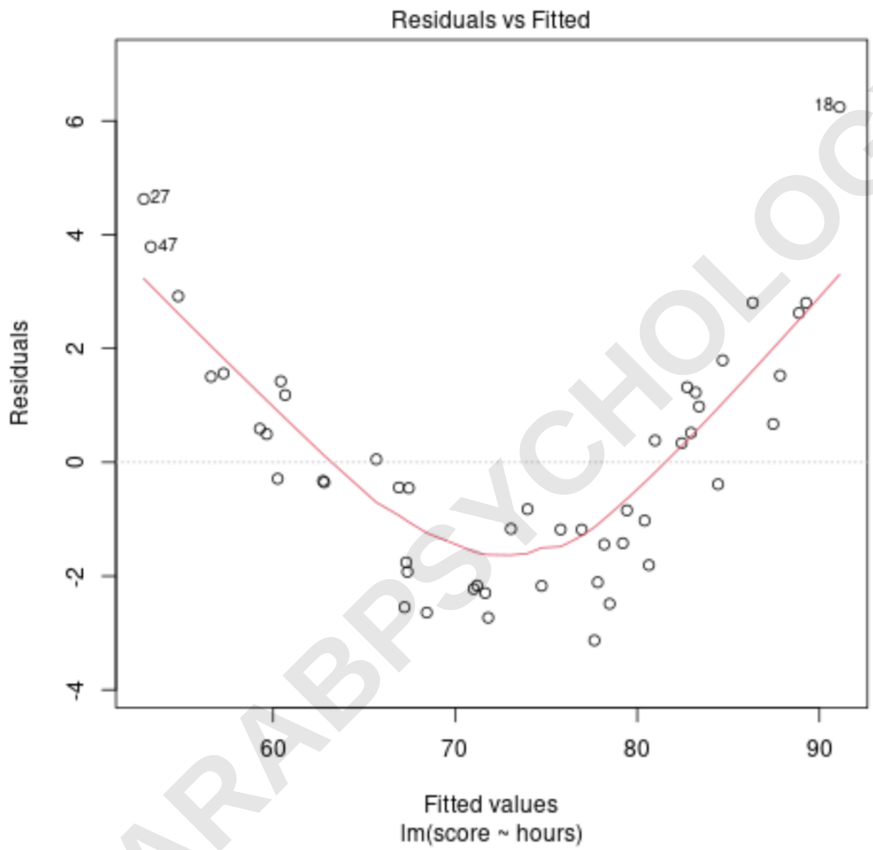
If there is a clear nonlinear pattern in the residuals, then this is an indication that polynomial regression could offer a better fit to the data.

For example, suppose we fit a linear regression model using hours studied as a predictor variable and exam score as a response variable, then create the following fitted values vs. residuals plot:



The residuals are randomly scattered around zero with no clear pattern, which indicates that a linear model provides an appropriate fit to the data.

However, suppose our fitted values vs. residuals plot actually looked like the following:



From the plot we can see that there is a clear nonlinear pattern in the residuals - the residuals exhibit a "U" shape.

This tells us that a linear model is not appropriate for

this particular data and it could be a good idea to instead fit a polynomial regression model.

3. Calculate the Adjusted R-Squared Value of the Model

Another way to determine if you should use polynomial regression is to fit both a linear regression model and a polynomial regression model and calculate the adjusted R-squared values for both models.

The adjusted R-squared represents the proportion of the variance in the response variable that can be explained by the predictor variables in the model, *adjusted* for the number of predictor variables in the model.

The model with the higher adjusted R-squared represents the model that is better able to use the predictor variable(s) to explain the variation in the response variable.

The following tutorials explain how to perform polynomial regression using different statistical software: