How can I perform polynomial regression in SAS?

Authored by

stats writer

June 26, 2024

Polynomial regression is a statistical technique used to model the relationship between a dependent variable and one or more independent variables by fitting a polynomial curve to the data. In SAS, this can be performed by using the PROC REG procedure, which allows users to specify the degree of the polynomial and the independent variables to be included in the model. The procedure also provides various diagnostic measures to evaluate the goodness of fit and identify outliers. By using SAS for polynomial regression, users can effectively analyze and interpret complex data sets and make informed decisions based on the results.

Perform Polynomial Regression in SAS

The most common type of regression analysis is simple linear regression, which is used when a predictor variable and a have a linear relationship.

However, sometimes the relationship between a predictor variable and a response variable is nonlinear.

In these cases it makes sense to use polynomial regression, which can account for the nonlinear relationship between the variables.

The following example shows how to perform polynomial regression in SAS.

Example: Polynomial Regression in SAS

Suppose we have the following dataset in SAS:

/*create dataset*/

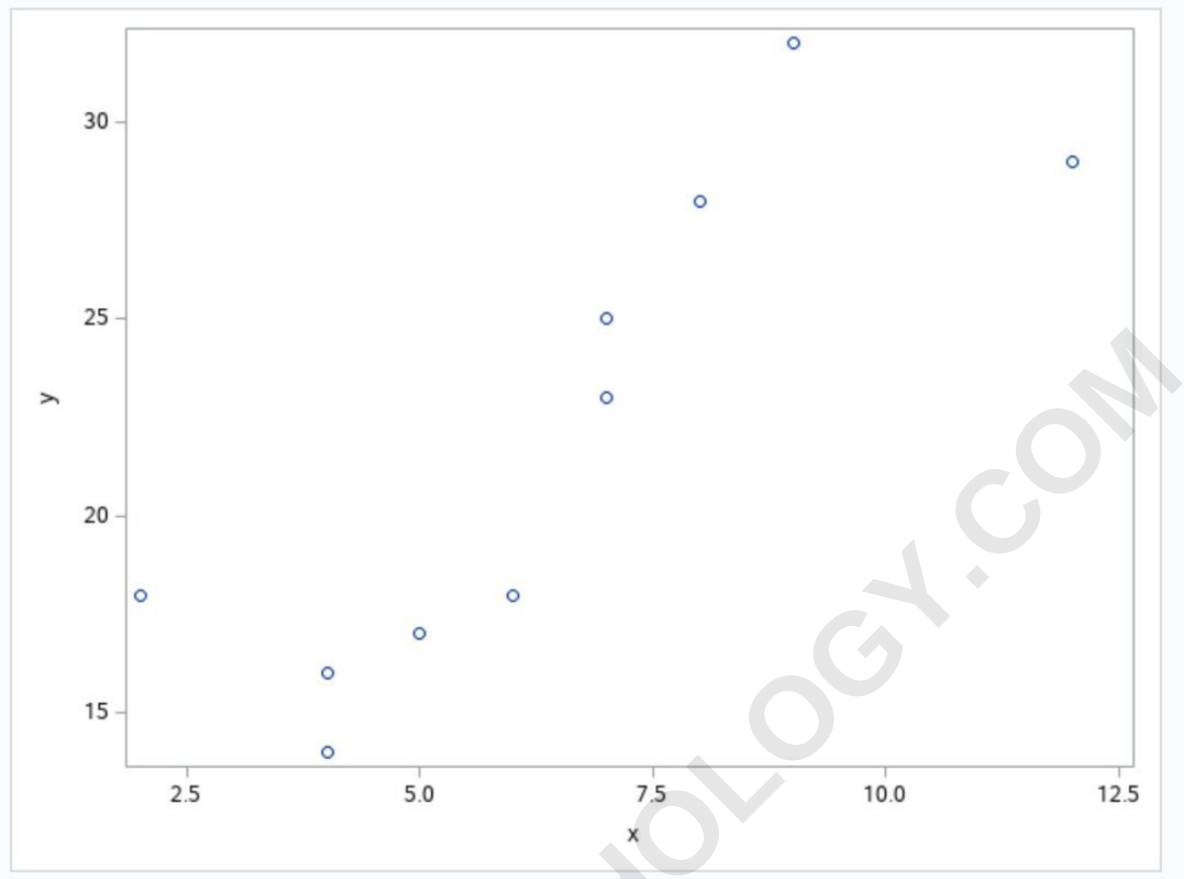
```
data my_data;
input x y;
datalines;
2 18
4 14
4 16
5 17
6 18
7 23
7 25
8 28
9 32
12 29
;
run;

/*view dataset*/
proc printdata=my_data;
```

Obs	x	y
1	2	18
2	4	14
3	4	16
4	5	17
5	6	18
6	7	23
7	7	25
8	8	28
9	9	32
10	12	29

Now suppose we create a scatter plot to visualize the relationship between the variables x and y in the dataset:

```
/*create scatter plot of x vs. y*/
proc sgplotdata=my_data;
scatter x=x y=y;
run;
```



From the plot we can see that the relationship between x and y appears to be cubic.

Thus, we can define two new predictor variables in our dataset (x2 and x3) and then use proc reg to fit a polynomial regression model using these predictor variables:

```
/*create dataset with new predictor variables*/
data my_data;
input x y;
```

```
x2 = x**2;
x3 = x**3;
datalines;
2 18
4 14
4 16
5 17
6 18
7 23
7 25
8 28
9 32
12 29
;
run;

/*fit polynomial regression model*/
proc regdata=my_data;
model y = x x2 x3;
run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: y

Number of Observations Read	10
Number of Observations Used	10

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	343.46329	114.48776	80.47	<.0001
Error	6	8.53671	1.42278		
Corrected Total	9	352.00000			

Root MSE	1.19281	R-Square	0.9757
Dependent Mean	22.00000	Adj R-Sq	0.9636
Coeff Var	5.42184		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	37.21341	3.97355	9.37	<.0001
x	1	-14.23806	2.12779	-6.69	0.0005
x2	1	2.64819	0.33621	7.88	0.0002
x3	1	-0.12648	0.01582	-7.99	0.0002

From the Parameter Estimates table we can find the coefficient estimates and write our fitted polynomial regression equation as:

y = 37.213 - 14.238x + 2.648x2 - 0.126x3

For example if x has a value of 4 then y is expected to have a value of 14.565:

y = 37.213 - 14.238(4) + 2.648(4)2 - 0.126(4)3 = 14.565

We can also see the polynomial regression model has an adjusted R-squared value of 0.9636, which is extremely close to one and tells us that the model does an excellent job of fitting the dataset.

Related:

The following tutorials explain how to perform other common tasks in SAS: