

Chapter 375

Polynomial Model Fit – Y vs One X

Introduction

This program fits polynomials and the ratio of two polynomials of up to fifth order. Examples of these types of models are:

$$Y = A0 + A1X + A2X^2$$

and

$$Y = \frac{A0 + A1X + A2X^2 + A3X^3 + A4X^4 + A5X^5}{1 + B1X + B2X^2 + B3X^3 + B4X^4 + B5X^5}$$

Both types of polynomial models are used to approximate functions. Polynomial models are related to the Taylor series expansion of a function. The ratio of two polynomials is related to rational function approximation (see Pade approximation).

These models approximate many different curves. Since these are approximating curves and usually have no physical interpretation, care must be taken outside the range of the data. You must study the resulting model graphically to determine that the model behaves properly between data points.

Problems with Ratio of Polynomials Models

As stated above, polynomials are used to approximate a function in a specific range close to a fixed point (such as zero). The approximation is only accurate within a narrow range. Outside this range, the polynomial approximation is less accurate.

For example, consider the polynomial ratio model

$$Y = \frac{10 + 11X + X^2}{4 - 5X + X^2}$$

Note that these two polynomials can be factored as follows

$$Y = \frac{(X + 1)(X + 10)}{(X - 1)(X - 4)}$$

Suppose the range of X is from 0 to 10. We note that when X is equal to 1 or 4, a division by zero occurs and the value of Y goes toward infinity, so the model may not be useful. However, if the range of the data was 5 to 10, the roots of the denominator polynomial are missed, and no division by zero occurs.

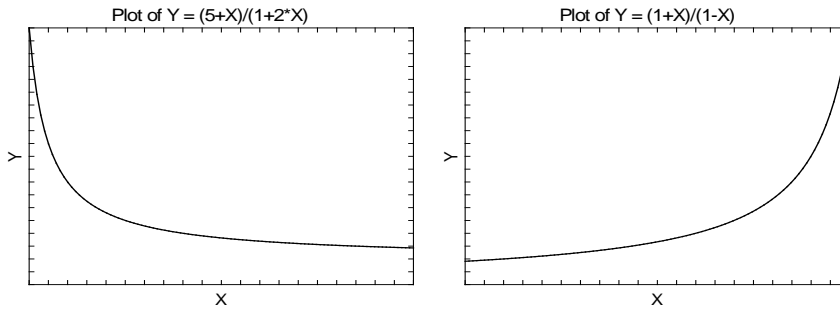
As this example points out, when the roots of the denominator polynomial are within the range of the data, serious errors in the approximation will often be seen.

Examples of Ratio of Polynomials Models

Here are some examples of ratio of polynomials models

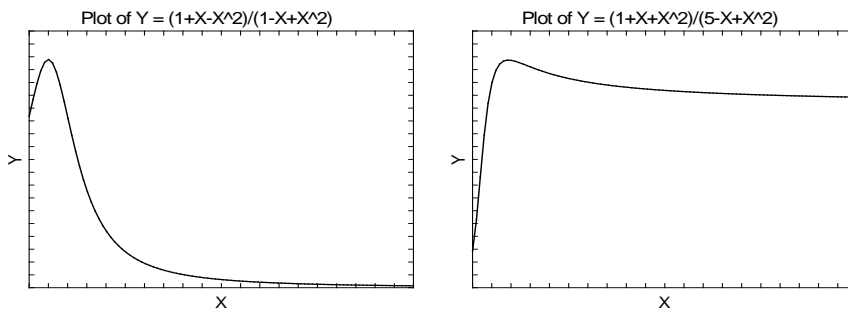
$$Y = (A + BX) / (1 + CX)$$

The ratio of first-order polynomials model is a slight extension of the Michaelis-Menten model.



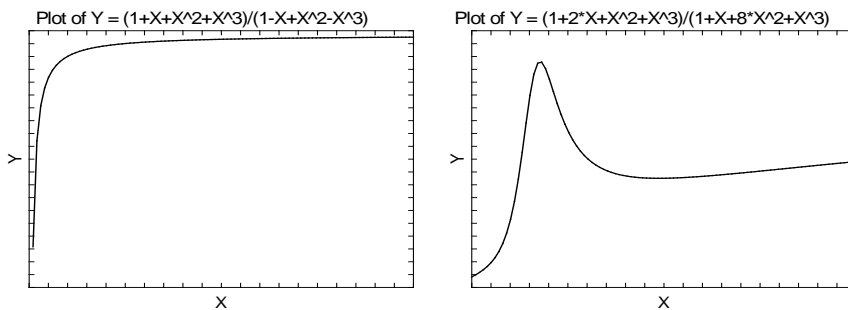
$$Y = (A + BX + CX^2) / (1 + DX + EX^2)$$

The ratio of second-order polynomials model may be used to approximate many complicated models.



$$Y = (A + BX + CX^2 + DX^3) / (1 + EX + FX^2 + GX^3)$$

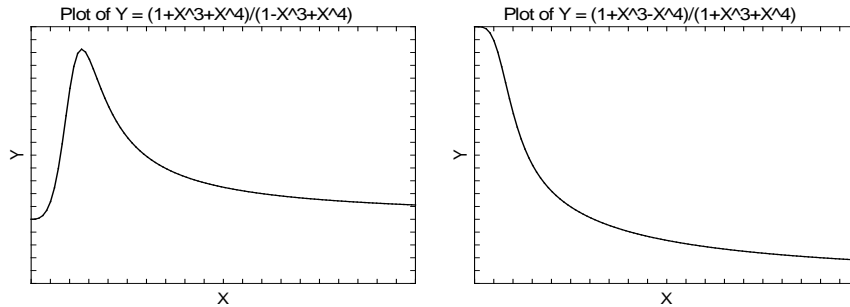
The ratio of third-order polynomials model may be used to approximate many complicated models. However, care must be used when estimating such high-degree models because of the common occurrence of $Y = 0$ in the denominator polynomial.



Polynomial Model Fit – Y vs One X

$$Y = (A + BX + CX^2 + DX^3 + EX^4) / (1 + FX + GX^2 + HX^3 + IX^4)$$

The ratio of fourth-order polynomials model may be used to approximate many complicated models. However, care must be used when estimating such high-degree models because of the common occurrence of $Y = 0$ in the denominator polynomial.



Starting Values

Starting values are determined by the program. You do not have to supply starting values.

Assumptions and Limitations

Usually, nonlinear regression is used to estimate the parameters in a nonlinear model without performing hypothesis tests. In this case, the usual assumption about the normality of the residuals is not needed. Instead, the main assumption needed is that the data may be well represented by the model.

Data Structure

The data are entered in two variables: one dependent variable and one independent variable.

Missing Values

Rows with missing values in the variables being analyzed are ignored in the calculations. When only the value of the dependent variable is missing, predicted values are generated.

Example 1 – Fitting a Polynomial Model

This section presents an example of how to fit a polynomial model. In this example, we will fit a cubic polynomial to the variables Y and X of the FnReg1 database.

Setup

To run this example, complete the following steps:

1 Open the FnReg1 example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **FnReg1** and click **OK**.

2 Specify the Polynomials Curve Fit – One X Variable procedure options

- Find and open the **Ratio of Polynomials Fit – One Variable** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Y VariableY

X VariableX

Model to Fit.....**3: Y = A0 + A1X + A2X² + A3X³ (Cubic)**

3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Model Estimation

Model Estimation

Rows Used: 100 of 100

Polynomial Order: 3

Polynomial Model: $Y = A0 + A1X + A2X^2 + A3X^3$

Y: Y

X: X

Pseudo-R²: 0.951

Coefficient			Standard Error	95% Confidence Interval Limits for the Coefficient	
Name	Term	Estimate		Lower	Upper
A0	Intercept	10.31521	0.1484675	10.02051	10.60992
A1	X	1.194476	0.1164404	0.9633433	1.425608
A2	X ²	-0.3086269	0.02517666	-0.3586021	-0.2586516
A3	X ³	0.01718805	0.001579796	0.01405218	0.02032392

Estimated Model (Double Precision)

Y =

$10.3152119487976 + (1.19447568568676)(X) - (0.308626865292296)(X)^2 + (0.0171880511495343)(X)^3$

Rows Used

The number of rows from the dataset that were used in the analysis followed by the number of rows that exist on the dataset. This allows you to determine how many observations were skipped.

Polynomial Order

The order of the polynomial. This is the value of the highest exponent in any term in the model.

Polynomial Model

The polynomial model that was estimated. Use this to double check that the model estimated was what you wanted.

Pseudo-R²

There is no direct R² defined for nonlinear regression. This is a pseudo R² constructed to approximate the usual R² value used in multiple regression. We use the following generalization of the usual R² formula:

$$R^2 = (ModelSS - MeanSS)/(TotalSS - MeanSS)$$

where *MeanSS* is the sum of squares due to the mean, *ModelSS* is the sum of squares due to the model, and *TotalSS* is the total (uncorrected) sum of squares of Y (the dependent variable).

Polynomial Model Fit – Y vs One X

This version of R^2 tells you how well the model performs after removing the influence of the mean of Y. Since many nonlinear models do not explicitly include a parameter for the mean of Y, this R^2 may be negative (in which case we set it to zero) or difficult to interpret. However, if you think of it as a direct extension of the R^2 that you use in multiple regression, it will serve well for comparative purposes.

Note that for models that do not have a denominator polynomial, the pseudo- R^2 = regular R^2 .

Coefficient Name

The name of the parameter whose results are shown on this line.

Coefficient Term

The name of the term (function of X) that is displayed on this line.

Coefficient Estimate

The estimated value of this parameter.

Standard Error

An estimate of the standard error of the parameter based on asymptotic (large sample) results.

95% Lower and Upper Confidence Interval Limits for the Coefficient

The lower and upper values of a 95% confidence limit for this parameter. This is a large sample (at least 25 observations for each parameter) confidence limit.

Estimated Model (Double Precision)

The model that was estimated with the coefficients replaced with their estimated values (in double-precision). This expression may be copied and pasted as a variable transformation in the spreadsheet. This will allow you to predict for additional values of X.

Analysis of Variance Table

Analysis of Variance

Source	DF	Sum of Squares	Mean Square
Mean	1	10559.48	10559.48
Model	4	10689.7	2672.424
Model (Adjusted)	3	130.2166	43.40553
Error	96	6.671717	0.06949705
Total (Adjusted)	99	136.8883	
Total	100	10696.37	

Source

The labels of the various sources of variation.

Polynomial Model Fit – Y vs One X

DF

The degrees of freedom.

Sum of Squares

The sum of squares associated with this term. Note that these sums of squares are based on Y, the dependent variable. Individual terms are defined as follows:

Mean	The sum of squares associated with the mean of Y. This may or may not be a part of the model. It is presented since it is the amount used to adjust the other sums of squares.
Model	The sum of squares associated with the model.
Model (Adjusted)	The model sum of squares minus the mean sum of squares.
Error	The sum of the squared residuals. This is often called the sum of squares error or just "SSE."
Total	The sum of the squared Y values.
Total (Adjusted)	The sum of the squared Y values minus the mean sum of squares.

Mean Square

The sum of squares divided by the degrees of freedom. The Mean Square for Error is an estimate of the underlying variation in the data.

Parameter Correlation Matrix

Parameter Correlation Matrix

	A0	A1	A2	A3
A0	1.000	-0.924	0.838	-0.768
A1	-0.924	1.000	-0.976	0.935
A2	0.838	-0.976	1.000	-0.988
A3	-0.768	0.935	-0.988	1.000

This report displays the asymptotic correlations of the parameter estimates. When these correlations are high (absolute value greater than 0.95), the precision of the parameter estimates is suspect.

Predicted Values and Residuals

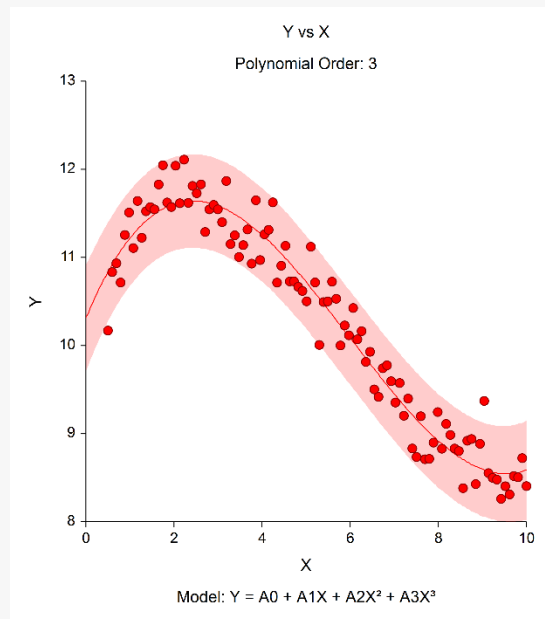
Predicted Values and Residuals

Row	X	Y	Predicted Value	95% Prediction Interval Limits for the Predicted Value		Residual
				Lower	Upper	
1	0.50000	10.16989	10.83744	10.27663	11.39826	-0.66755
2	0.59596	10.83415	10.92109	10.36539	11.47680	-0.08694
3	0.69192	10.93412	10.99963	10.44837	11.55089	-0.06551
4	0.78788	10.71519	11.07314	10.52572	11.62056	-0.35795
5	0.88384	11.25534	11.14171	10.59757	11.68586	0.11363
6	0.97980	11.50980	11.20544	10.66408	11.74681	0.30436
7	1.07576	11.10627	11.26442	10.72538	11.80345	-0.15814
8	1.17172	11.64065	11.31873	10.78163	11.85583	0.32192
9	1.26768	11.22304	11.36847	10.83295	11.90399	-0.14543
10	1.36364	11.52402	11.41373	10.87949	11.94798	0.11029

The section shows the values of the residuals and predicted values. If you have observations in which the independent variable is given, but the dependent (Y) variable was left blank, a predicted value and prediction limits will be generated and displayed in this report.

Function Plot(s)

Function Plot(s)

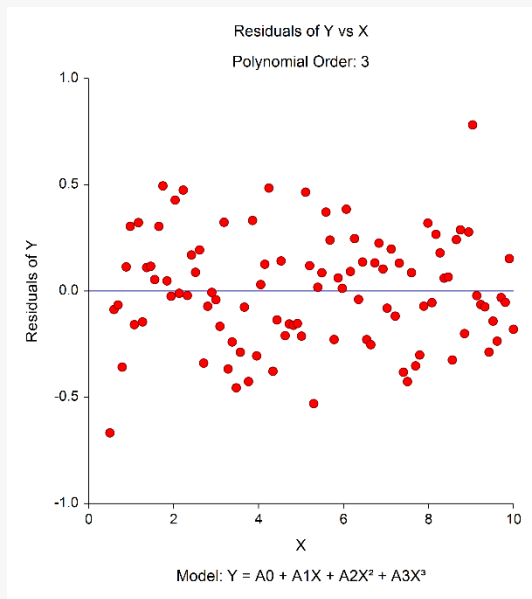


This plot displays the data along with the estimated function and prediction limits. It is useful in deciding if the fit is adequate and if the prediction limits are appropriate.

In poorly fit models, we have found that it is often necessary to disable the prediction limits so that the data will show up. In these cases, the prediction limits may be so wide that the scale of the plot does not allow the data values to be separated.

Residual Plot(s)

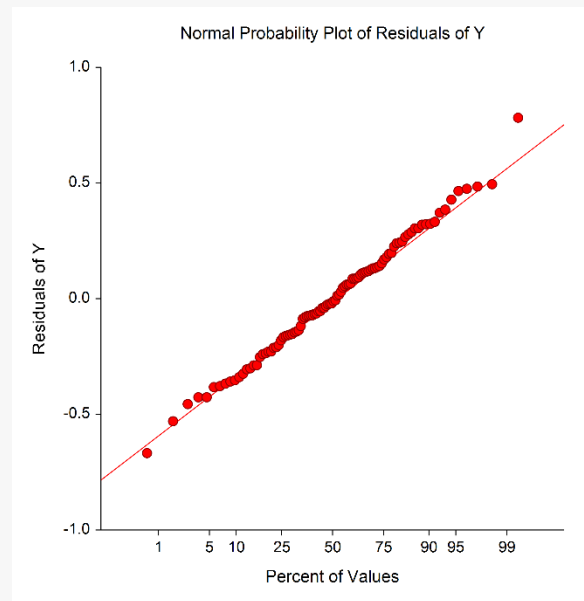
Residual Plot(s)



This is a scatter plot of the residuals versus the independent variable, X. The preferred pattern is a rectangular shape or point cloud. Any nonrandom pattern may require a redefining of the model.

Probability Plot(s)

Probability Plot(s)



If the residuals are normally distributed, the data points of the normal probability plot will fall along a straight line. Major deviations from this ideal picture reflect departures from normality. Stragglers at either end of the normal probability plot indicate outliers, curvature at both ends of the plot indicates long or short distributional tails, convex or concave curvature indicates a lack of symmetry, and gaps or plateaus or segmentation in the normal probability plot may require a closer examination of the data or model. We do not recommend that you use this diagnostic with small sample sizes.

Predicting for New Values

You can use your model to predict Y for new values of X. Here's how. Add new rows to the bottom of your database containing the values of the independent variable that you want to create predictions for. Leave the dependent variable blank. When the program analyzes your data, it will skip these rows during the estimation phase, but it will generate predicted values for all rows, regardless of whether the Y variable is missing or not.

Example 2 – Fitting a Ratio of Polynomials Model

This section presents an example of how to fit a polynomial ratio model. In this example, we will fit a 3 / 4 polynomial ratio model to the variables Y and X of the FnReg3 dataset.

Setup

To run this example, complete the following steps:

1 Open the FnReg3 example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **FnReg3** and click **OK**.

2 Specify the Polynomial Model Fit – One X procedure options

- Find and open the **Polynomial Model Fit – One X** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Y Variable **Y**
 X Variable **X**
 Model to Fit..... **3 / 4: Y = (A0 + A1X + A2X² + A3X³) / (1 + B1X + B2X² + B3X³ + B4X⁴)**

3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Model Estimation

Model Estimation

Rows Used: 51 of 51

Iterations: 270

Polynomial Order: 3 / 4

Polynomial Model: $Y = (A0 + A1X + A2X^2 + A3X^3) / (1 + B1X + B2X^2 + B3X^3 + B4X^4)$

Y: Y

X: X

Pseudo-R²: 0.990

Coefficient			95% Confidence Interval Limits for the Coefficient		
Name	Term	Estimate	Standard Error	Lower	Upper
A0	Intercept	11.7843	1.012324	9.74275	13.82584
A1	X	-0.2794355	0.09171839	-0.4644032	-0.09446771
A2	X ²	0.005810308	0.003520239	-0.001288929	0.01290955
A3	X ³	-4.172031E-05	2.865621E-05	-9.951107E-05	1.607044E-05
B1	X	-0.0783304	0.002023107	-0.08241038	-0.07425041
B2	X ²	0.002391857	4.693836E-05	0.002297197	0.002486517
B3	X ³	-2.86472E-05	4.337468E-06	-3.739453E-05	-1.989986E-05
B4	X ⁴	1.171313E-07	1.026248E-06	-1.952496E-06	2.186758E-06

Estimated Model (Double Precision)

Y =

$$(11.7842966508335 - (0.279435473777121)*(X) + (0.00581030839632261)*(X)^2 - (4.17203122468724E-05)*(X)^3) / (1 - (0.0783303966070252)*(X) + (0.00239185672828884)*(X)^2 - (2.86471961423406E-05)*(X)^3 + (1.17131250064605E-07)*(X)^4)$$

Predicted Values and Residuals

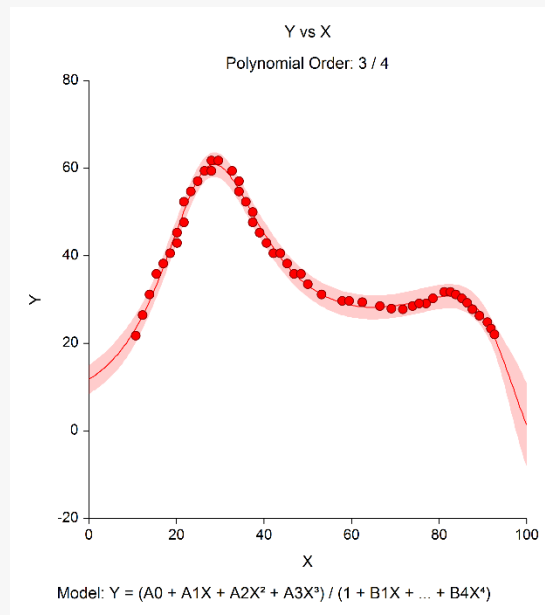
Predicted Values and Residuals

Row	X	Y	Predicted Value	95% Prediction Interval Limits for the Predicted Value		Residual
				Lower	Upper	
1	10.69182	21.76471	23.38152	20.21170	26.55134	-1.61681
2	12.26415	26.47059	26.23679	23.28620	29.18738	0.23380
3	13.83648	31.17647	29.48607	26.68323	32.28891	1.69040
4	15.40880	35.88235	33.13580	30.36270	35.90890	2.74656
5	16.98113	38.23529	37.15681	34.36014	39.95348	1.07848
6	18.55346	40.58823	41.46602	38.66883	44.26321	-0.87778
7	20.12579	42.94118	45.90926	43.14496	48.67357	-2.96809
8	20.12579	45.29412	45.90926	43.14496	48.67357	-0.61515
9	21.69811	47.64706	50.25371	47.51837	52.98905	-2.60665
10	23.27044	54.70588	54.20129	51.46606	56.93652	0.50459

The section shows the values of the residuals and predicted values. If you have observations in which the independent variable is given, but the dependent (Y) variable was left blank, a predicted value and prediction limits will be generated and displayed in this report.

Function Plot(s)

Function Plot(s)

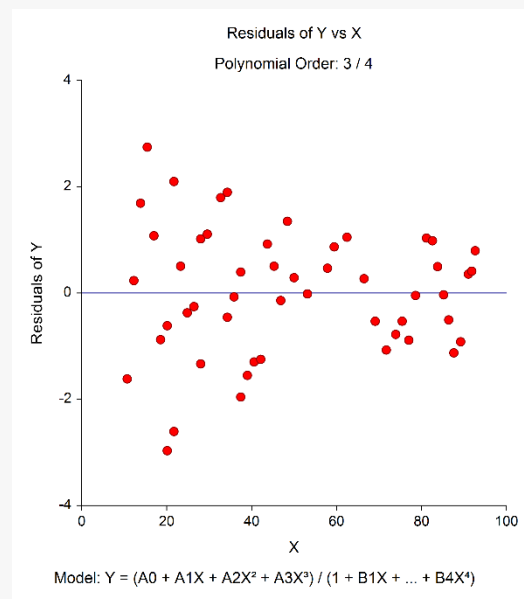


This plot displays the data along with the estimated function and prediction limits. It is useful in deciding if the fit is adequate and if the prediction limits are appropriate.

In poorly fit models, we have found that it is often necessary to disable the prediction limits so that the data will show up. In these cases, the prediction limits may be so wide that the scale of the plot does not allow the data values to be separated.

Residual Plot(s)

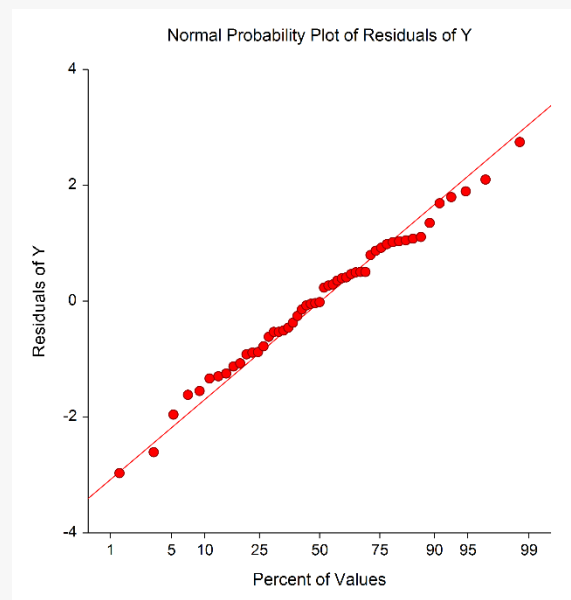
Residual Plot(s)



This is a scatter plot of the residuals versus the independent variable, X. The preferred pattern is a rectangular shape or point cloud. Any nonrandom pattern may require a redefining of the model.

Probability Plot(s)

Probability Plot(s)



If the residuals are normally distributed, the data points of the normal probability plot will fall along a straight line.