

PolynomialFit Help

This program fits datasets to polynomials. Calling the independent variable x and the dependent variable y , then the program fits to:

$$y = \sum_{i=0}^9 a(i) x^i = a(0) + a(1)x + a(2)x^2 + \dots$$

The program determines the values and errors of the fit coefficients $a(i)$. Controls allow you to choose which powers of the polynomial you wish to include in the fit.

The dataset is in a file that either you or some other program has created.

Overview

When the program starts you will be prompted to choose the file that contains the dataset to which you wish to fit.

The variables defined in the dataset will then be displayed in the Fit Variables part of the screen.

You choose the terms of the polynomial by clicking on the buttons labeled with Choose the Powers for the Fit, located near the bottom of the window. Usually these buttons are the only things that you will need to change to fit your data to a polynomial.

As you choose powers, the polynomial you are fitting is shown below the buttons that choose the powers.

When the program starts, Powers 0 and 1 are turned on, and all others are turned off. This is a fit to a straight line.

You may change the powers at any time and the results of the new fit will be displayed. The most important of these results are:

- Fit Coefficients: Value \pm Error. This is a table of the values of the fit coefficients and their estimated errors.
- Residuals plot. The residuals are the values of the dependent (y) variables as given by the fit minus the values in the dataset. For a good fit the residuals should be small and randomly distributed around zero. If this is not the case either there is something wrong with the data or you are trying to fit the data to an incorrect model.

- Fit Result Plot. This shows the data and the results of the fit.

Examples

Powers	Description
0 1	<p>This fits the data to a straight line, the default. The fitter returns the intercept with the dependent (y) variable axis, $a(0)$, and the slope of the straight line, $a(1)$. Often one writes the equation of a straight line as:</p> $y = mx + b$ <p>Then, the notation of the fitter is:</p> $a(1) = m$ $a(0) = b$
0 1 2	<p>Adds a term proportional to the square of the independent (x) variable to the straight line fit of the previous example. The fitter returns the intercept, $a(0)$, the slope, $a(1)$, and the coefficient of the second-order term $a(2)$.</p>
1	<p>Fit the data to a straight line with the intercept forced to zero. The fitter returns the slope $a(1)$.</p>

Evaluating the Fit

As already mentioned, if the data are good and the model being fit to is appropriate then the residuals should be small and randomly distributed around zero.

Here are a couple of tips for determining the best fit to the data.

1. If the value of a parameter is zero within errors, remove that power from the fit.
2. Even if the value of a parameter is not zero within errors, if removing that power from the fit does not make any appreciable difference in the Fit Result Plot you may wish to consider leaving the power out of the fit.

The program also displays some numeric ways of evaluating the fit.

The Dataset Does Not Contain Explicit Errors

If there are no explicit errors in the data, then the Sum of the Squares of the residuals is shown. Although smaller values indicate a better fit, defining what is "small" is not simple.

The Dataset Contains Explicit Errors

If there are explicit errors in one or both coordinates of the dataset then the Chi-Squared is shown. This is related to the sum of the squares of the residuals, except that the data are weighted with the explicit errors.

To interpret the chi-squared one needs to also look at the Degrees of Freedom of the fit, whose value is also shown. This number is the number of datapoints in the data minus the number of parameters to which you are fitting.

If the Chi-Squared is much greater than the Degrees of Freedom then the fit is poor.

If the Chi-Squared is much less than the Degrees of Freedom, then the fit is "too good to be true" and is suspect. A common cause of this behavior is when the explicit errors are larger than is appropriate.

The perfect result is when the Chi-Squared is approximately equal to the Degrees of Freedom.

Options

There are switches that allow you to change the default behavior of the fitter. You will seldom if ever need to use these.

Adjust Significant Figures

By default the fitter adjusts the values of the parameter values based on their errors. This makes the values much easier to read. Clicking on the Adjust Sig Figs switch allows you to turn this adjustment off; clicking on it again turns the adjustment back on.

These adjusted values are what is used in displaying the Fit Result in the graph. Occasionally rounding errors can cause the resulting graph to look silly. Sometimes turning off the adjustment fixes this.

Reweight

If the data do not have explicit errors, by default the data are weighted by assigning an error in the dependent (y) variable equal to the square root of (sum of the squares / degrees of freedom). This does not effect the values of

the fit coefficients, but does usually make the estimates of their errors more realistic. Clicking on the Reweight switch allows you to turn off this reweighting and have the fitter assume that there are no errors in the dependent variable; clicking on the switch again turns reweighting back on. For most experimental data, reweighting is the right thing to do.

The Reweight switch is only visible when the data do not have explicit errors.

This document was written by David M. Harrison, Dept. of Physics, Univ. of Toronto in June 2007. Last modified July 18, 2007.