REGRESSION

Regression and interpolation differ in that interpolation attempts to connect all data points while regression attempts to define a single function, in a least-squares sense, through all data points. New data points can be predicted using the resulting function. Mathcad supports the following regression types:

- Linear regression
- Generalized regression
- Polynomial regression
- Specialized regression

> ***NOTE:***
> All regression methods are discussed in depth in the tutorial and QuickSheet sections of Mathcad.

POLYNOMIAL REGRESSION

The information below will be used to help illustrate the following discussion on polynomial regression.

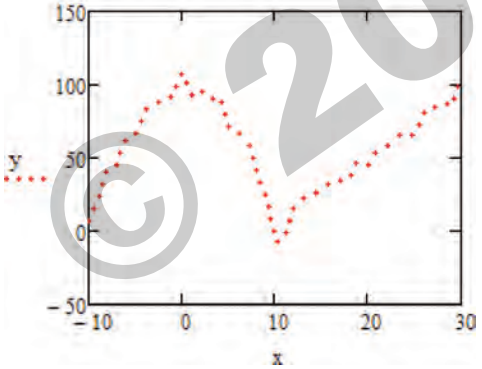
D =

	0	1
0	-10	7.04
1	-9	19.78
2	-8	43.39
3	-7	45.55
4	-6	61.42
5	-5	65.7
6	-4	...

$x := D^{\langle 0 \rangle}$

$y := D^{\langle 1 \rangle}$

rows(D) = 41



Polynomial regression fits a polynomial that minimizes the vertical distance of all data points to a polynomial curve of degree n. Polynomial regression is a two-step process that involves the regress and interp functions.

The **regress** function is initially used, and includes the order of the regressed polynomial. The regress function has the form regress(vx, vy, n), where:

- *vx* is a vector of real numbers in increasing order.
- *vy* is a vector of real numbers.
- *n* is the order of the regressed polynomial.

Here, we have calculated the regression coefficients for a third and fourth order polynomial using the regress function.

$$P3 := regress(x, y, 3) \qquad P4 := regress(x, y, 4)$$

$$P3 = \begin{pmatrix} 3 \\ 3 \\ 3 \\ 3 \\ 79.885 \\ -0.338 \\ -0.453 \\ 0.017 \end{pmatrix} \qquad P4 = \begin{pmatrix} 3 \\ 3 \\ 4 \\ 90.971 \\ -3.584 \\ -0.698 \\ 0.058 \\ -1.019 \times 10^{-3} \end{pmatrix}$$

The first three elements in the vectors are values used internally by the interp function. They are not the regression coefficients. Translating the coefficient values from P3 into a regression curve would give:

$$R3(t) := 0.017t^3 - 0.453t^2 - 0.338t + 79.885$$

The coefficients must manually be built into a regression curve if a function is required.

The **interp** function has the form interp(vs, vx, vy, x), where:

- *vs* is a vector of spline coefficients returned by spline function, such as lspline
- *vx* is a vector of real numbers in increasing order.
- *vy* is a vector of real numbers
- *x* is the value of the independent variable at which you want an interpolated value or range of values.

The interp function can be used if a numerical evaluation is required.

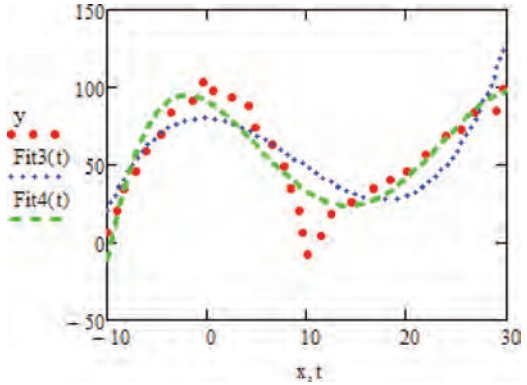
$$Fit3(t) := interp(P3, x, y, t)$$

$$Fit4(t) := interp(P4, x, y, t)$$

Next, the domain is defined by range variables. This range variable will be defined as t in increments of 0.01, using the minimum and maximum values of *x*:

$$t := min(x), min(x) + 0.01 .. max(x)$$

The original data, third order polynomial, and fourth order polynomial are plotted on same graph:



STATISTICS FUNCTIONS

To further analyze your data or to determine the suitability of a regression result, you may wish to apply other statistics functions for data analysis. There are many data analysis functions including:

- variance and standard deviation functions.
- finding the arithmetic, geometric, or harmonic means.
- frequency distributions.