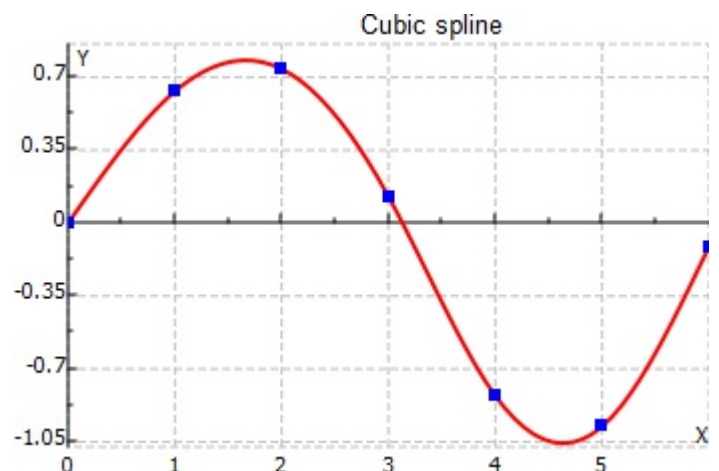
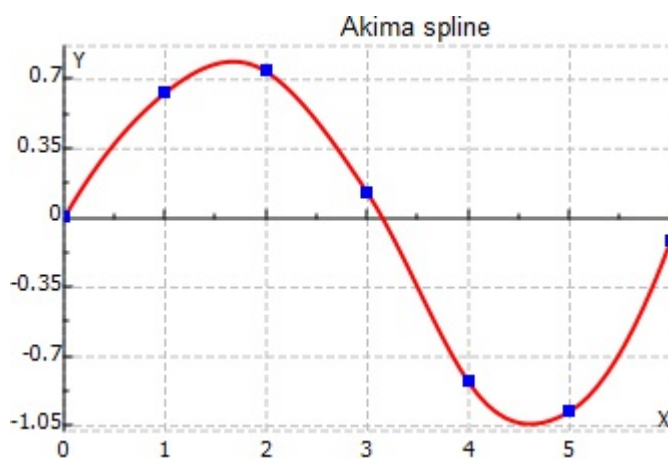
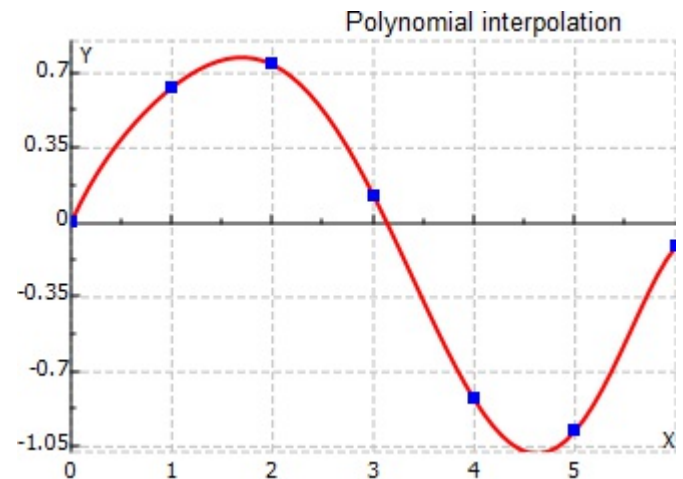
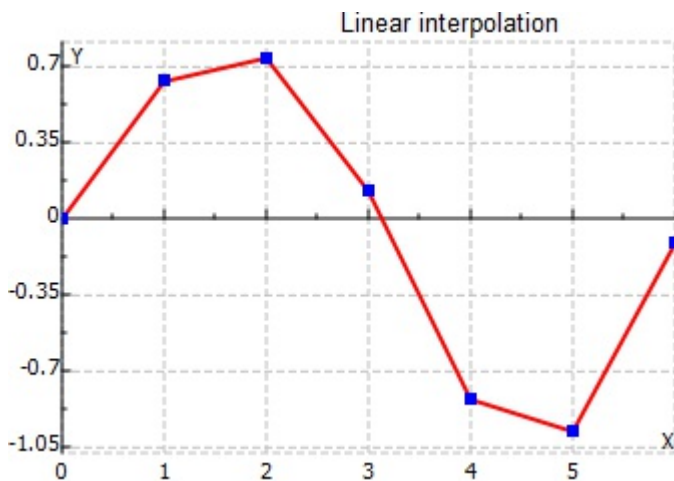


Curve fitting

Curve fitting is the process of constructing a mathematical function which has the best fit to a series of data points. Curve fitting refers to either **a) interpolation**, where an exact fit to the data is required, or **b) regression**, where an unique function with a random error is constructed that approximately fits the data.

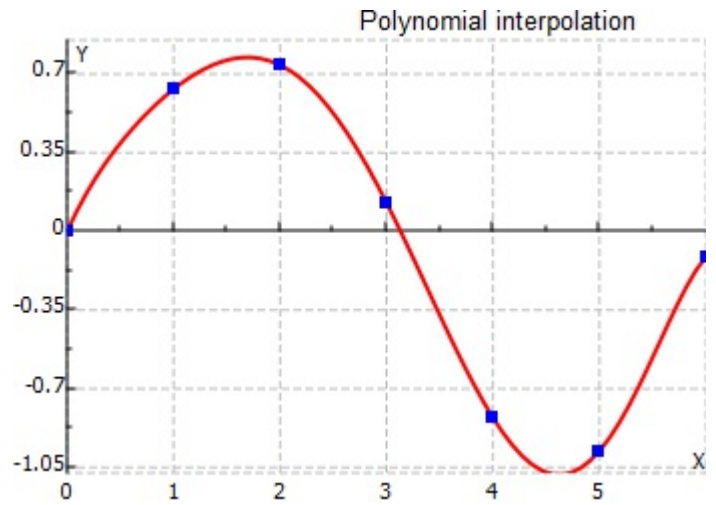
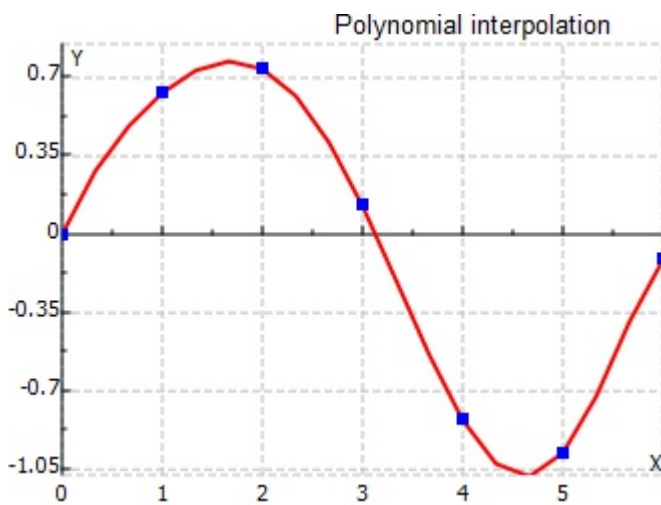
Interpolation is the method of estimation of a value within a sequence of known values. In MatDeck you have the following interpolation methods : Linear interpolation, Polynomial interpolation, Ration interpolation, Cubic spline, Akima spline, Hermite spline, Cubic B spline and Bezier interpolation.

$$a := \begin{bmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 0 & 0.6325 & 0.7402 & 0.1274 & -0.8293 & -0.9768 & -0.1133 \end{bmatrix}$$

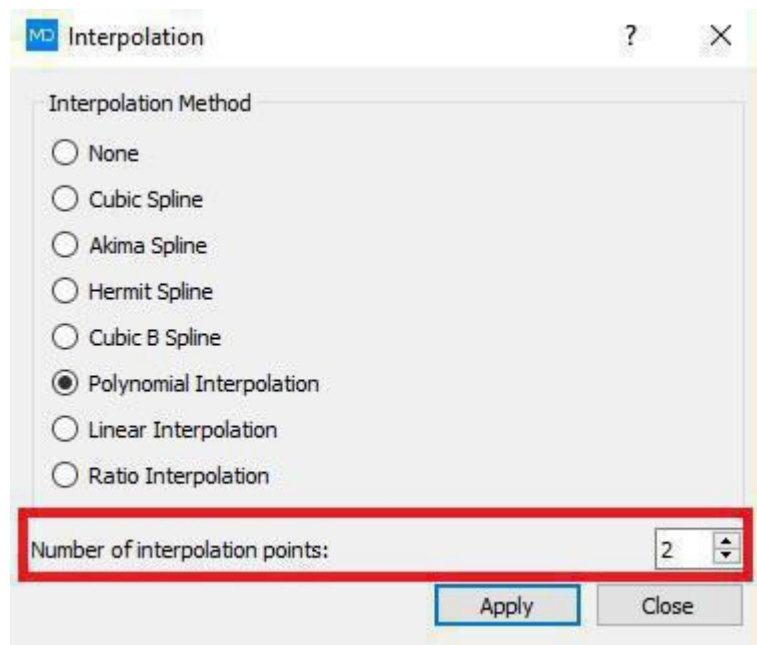


The above example demonstrates the differences between interpolation methods. Linear interpolation uses a linear function for each source data interval $[x_k, x_{k+1}]$, while spline interpolation uses low-degree polynomials in each of the intervals.

The number of interpolation points is a parameter which tells the interpolation function how many points to create between every two source data point intervals $[x_k, x_{k+1}]$.



Interpolation points parameters can be set from the Graph tab, Interpolation icon, as shown on the picture bellow. Interpolation window refers to the selected graph, and you can change interpolation method or number of interpolation points and after pressing the Apply button, change can be seen on the graph.



To create interpolation equations for each source data interval $[x_k, x_{k+1}]$ use the functions: `linearinterpmat`, `polyinterpeq`, `cubicsplinemat`, `akimasplinemat`, `hermitsplinemat`.

Linear
interpolation equations

$$\text{linearinterpmat}(\mathbf{a}) = \begin{bmatrix} [0,1] & [1,2] & [2,3] & [3,4] \\ 0.633x & 0.108x+0.525 & -0.613x+1.966 & -0.957x+2.997 \end{bmatrix}$$

Hermit
spline equations

$$\mathbf{a1} := \text{mat transpose}(\text{hermitsplinesmat}(\mathbf{a}, 1))$$

$$\mathbf{a1} = \begin{bmatrix} [0,1] & -0.262x^3 + 0.262x^2 + 0.633x \\ [1,2] & -0.098x^3 + 0.129x^2 + 0.406x + 0.196 \\ [2,3] & 0.188x^3 - 1.678x^2 + 4.201x - 2.455 \\ [3,4] & 0.577x^3 - 5.937x^2 + 19.273x - 19.822 \\ [4,5] & 0.101x^3 - 0.907x^2 + 1.861x - 0.219 \\ [5,6] & -0.506x^3 + 8.594x^2 - 47.665x + 85.696 \end{bmatrix}$$

To create a interpolation equation for the specific source data interval $[x_k, x_{k+1}]$ use the functions: linearinterpeq, cubicsplineeq, akimasplineeq, hermitsplineeq, polyinterpeq, bezierinterpeq.

$$\text{linearinterpeq}(\mathbf{a}, 2) = -0.613x + 1.966 \quad \text{OR} \quad \text{linearinterpeq}(\mathbf{a}, [2, 3]) = -0.613x + 1.966$$

$$\text{cubicsplineeq}(\mathbf{a}, 0) = 0.731x - 0.098x^3 \quad \text{OR} \quad \text{cubicsplineeq}(\mathbf{a}, [0, 1]) = 0.731x - 0.098x^3$$

$$\text{akimasplineeq}(\mathbf{a}, 4) = 24.185 - 14.225x + 2.610x^2 - 0.154x^3$$

OR

$$\text{akimasplineeq}(\mathbf{a}, [4, 5]) = 24.185 - 14.225x + 2.610x^2 - 0.154x^3$$

$$\text{hermitsplineeq}(\mathbf{a}, 1, 1) = -0.098x^3 + 0.129x^2 + 0.406x + 0.196$$

OR

$$\text{hermitsplineeq}(\mathbf{a}, 1, [1, 2]) = -0.098x^3 + 0.129x^2 + 0.406x + 0.196$$

On the other hand, **regression** methods are statistical processes for estimating the relationships among variables. They are used to define relations among the inner data values, but also a prediction

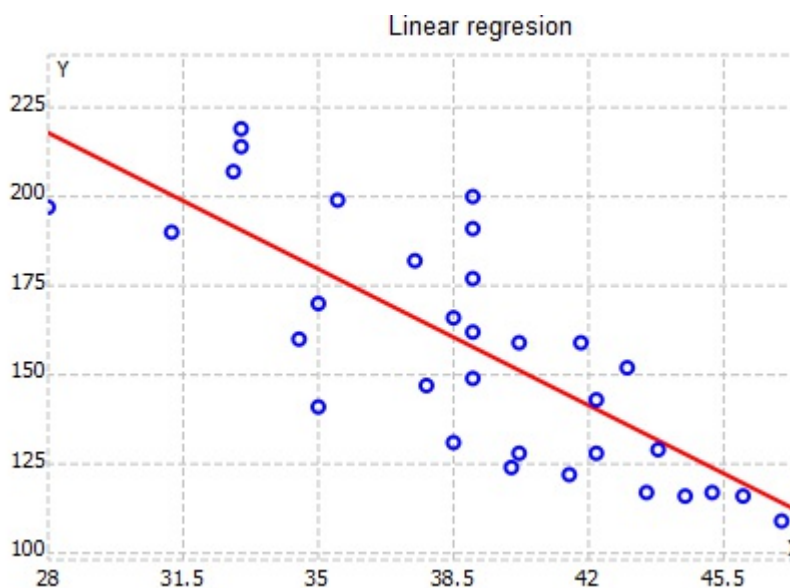
of values beyond the range of the observed data.

When regression is used on graph, always turn on Data points for a better illustration of the problem you are solving.

Regression methods that you have at your disposal are:

- Linear regression
- Exponential regression
- Logarithmic regression
- Power regression
- Polynomial regression

`b:=excel read("R-square.xlsx" , "Sheet1" , "B2:C50" , false)`



To determine how well the chosen model fits the data, you should use the functions regressiontable and regression. It is a sum of several statistics, displayed in a table: R-squared, Adjusted R-squared, Standard Error of the Regression and the regression equation. These best-of-fit statistics helps us in choosing the right regression method. You can also use them separately by using the functions: rsquared (R-squared), rsquaredadjusted (Adjusted R-squared), esterror (Standard Error of the Regression).

`regression(b , "linear") =`

	Value
Reg. formula	-5.461 x+ 370.825
RMSE	20.01
R-sq	0.608
R-sq(adj)	0.6

To evaluate the dependence of the starting parameters and display the dependence (regression)

formula, use the following functions:

- linfit - Linear regression
- expfit - Exponential regression
- logfit - Logarithmic regression
- polyfit - Polynomial regression
- powfit - Power regression

R-squared is a statistical measure of how close the data is to the fitted regression line, the higher it's value is, the better. R-squared value is always between 0 and 100%.

The main limitations of R-squared is that it does not indicate if we have chosen an adequate regression model, adding a predictor to a model always increases a R-squared's value. That means that the model with more terms has higher R-squared's values, this property can easily mislead us to make the wrong conclusion about the model itself.

For the above example, best-of-fit statistics values are

$$\text{rsquare}(b, \text{"linear"}) = 0.608$$

R-squared value

Adjusted R-squared is a modified version of R-squared that is adjusted to the number of model predictors.

If the new predictor improves the model, the adjusted R-squared will increase. It will decrease if a predictor improves the model less than can be expected by chance.

Adjusted R-squared values are always equal or lower than an R-squared value, and it can also be a negative. For example, if we use polynomial regression on the same data source, we can see that the Adjusted R-squared value increases when the polynomial degree is higher. This way you can find the peaks for an Adjusted R-squared value and when it's values start to decline, you have found the optimal number of predictors to include in your model.

$$\text{rsquare}\left(b, \begin{bmatrix} \text{"polynomial"} \\ 2 \end{bmatrix}\right) = 0.609$$

$$\text{rsquareadjusted}\left(b, \begin{bmatrix} \text{"polynomial"} \\ 2 \end{bmatrix}\right) = 0.601$$

$$\text{rsquare}\left(b, \begin{bmatrix} \text{"polynomial"} \\ 3 \end{bmatrix}\right) = 0.618$$

$$\text{rsquareadjusted}\left(b, \begin{bmatrix} \text{"polynomial"} \\ 3 \end{bmatrix}\right) = 0.610$$

Standard Error of the Regression (RMSE) is the average value that the source data are located from the regression line. Smaller values of error tells us that the data is closer to the regression line. Use these statistics to assess the precision of the model. About 95% of the source data should be

located within the area of $2 \times \text{RMSE}$ from the regression line.

```
bx := col2vec(b , 0)
```

```
by := col2vec(b , 1)
```

Split source matrix into two vectors

```
y2 := by + 2 * esterror(b , "linear")
```

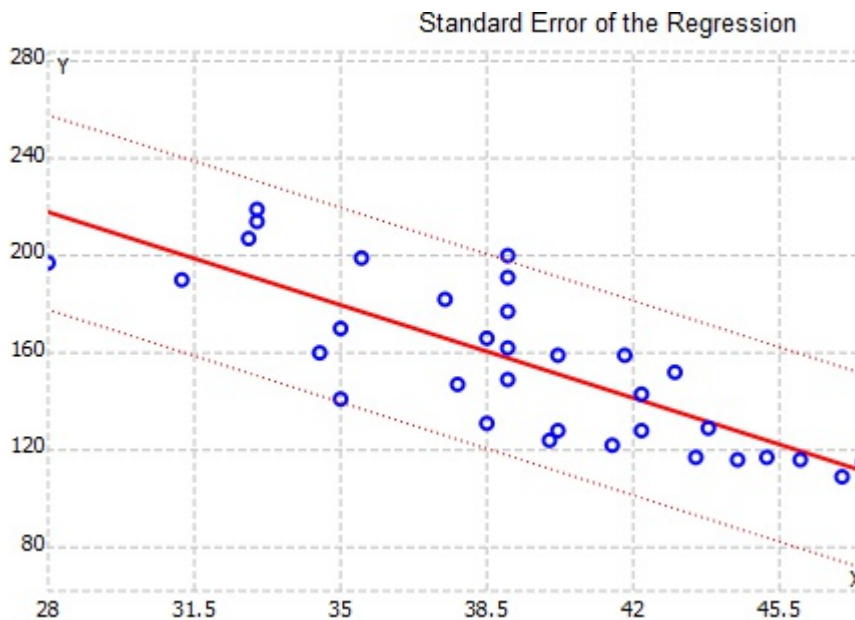
```
y3 := by - 2 * esterror(b , "linear")
```

Translate regression line for 2S value in each direction

```
c := join mat cols(bx , y2)
```

```
d := join mat cols(bx , y3)
```

Create two matrices to display area of expectancy for points



To make a decision about what is the best fitting model for this data set, we will have to use function **regression** and compare given results for different fitting techniques. The results are:

```
regression(b , "linear") =
```

	Value
Reg. formula	-5.461 x + 370.825
RMSE	20.01
R-sq	0.608
R-sq(adj)	0.6

regression(**b** , "exponential") =

	Value
Reg. formula	$602.759 \cdot e^{-0.035 x}$
RMSE	20.095
R-sq	0.605
R-sq(adj)	0.596

regression(**b** , "logarithmic") =

	Value
Reg. formula	$-208.824 \ln(x) + 921.341$
RMSE	20.033
R-sq	0.607
R-sq(adj)	0.599

regression(**b** , "power") =

	Value
Reg. formula	$19828.935 x^{-1.32766}$
RMSE	20.452
R-sq	0.591
R-sq(adj)	0.582

regression(**b** , $\left[\begin{array}{c} \text{"polynomial"} \\ 2 \end{array} \right]$) =

	Value
Reg. formula	$0.041 x^2 - 8.654 x + 431.916$
RMSE	19.984
R-sq	0.609
R-sq(adj)	0.601

regression(**b** , $\left[\begin{array}{c} \text{"polynomial"} \\ 3 \end{array} \right]$) =

	Value
Reg. formula	$0.022 x^3 - 2.483 x^2 + 87.146 x - 764.915$
RMSE	19.746
R-sq	0.618
R-sq(adj)	0.610

As we can see, while the degree of polynomial fit is growing, the R-squared and Adjusted R-squared values also grow and the RMSE decreases. We will have to calculate a few more polynomial fits with higher degree to determine the border from which the statistics values start falling.

This is the longer way to determine the best fit. The easiest way is to use a function called **regressiontable**, where you can compare all the fits simultaneously. The results are placed in a table from where you can easily compare all the fitting methods and conclude which is the best.

regressiontable(b) =

Method	R-square	Adj. R-sq	RMSE	SSE	Coeff Num.
"lin"	0.608	0.6	20.01	19618.850	2
"exp"	0.605	0.596	20.095	19787.302	2
"log"	0.607	0.599	20.033	19664.546	2
"pow"	0.591	0.582	20.452	20496.691	2
"poly2"	0.609	0.601	19.984	19567.921	3
"poly3"	0.618	0.610	19.746	19104.919	4
"poly4"	0.633	0.626	19.352	18350.989	5
"poly5"	0.634	0.626	19.34	18327.103	6
"poly6"	0.637	0.63	19.247	18152.65	7
"poly7"	0.373	0.36	25.306	31379.281	8
"poly8"	0.639	0.631	19.214	18089.898	9
"poly9"	0.641	0.633	19.162	17992.215	10

From this table we can see that the best fitting method is the polynomial with ninth degree, because the R-squared values are at their highest and the error values are at their smallest. The same values for tenth degree polynomial fitting start to fall and errors start to become higher.

