After creating dataset you can analyze it in different ways. First, you can calculate statistics. Open Statistics dialog, Common tabsheet, click Calculate.

**Min, Max:** minimal and maximal field values after normalization. The numbers in parentheses with # symbol mean the numbers of the record where the maximum and minimum values were reached.

The same with **Mean.** The number just below the Mean word means the number of object closest to the average point.

**Stand.dev.:** standard deviations after normalization. The numbers in parentheses indicate how many objects are in the interval \((m-\sigma,m+\sigma)\), where \(m\) – average, \(\sigma\) – standard deviation. The number just below the Stand.dev. word means overall standard deviation.

**Real min, Real max, Real mean** – Mean, Max, Mean values before normalization.

You can print out all numbers in Excel.
Euclidean Distances matrix (after normalization). Choose Distances tabsheet, click Calculate. (Not very useful if you have thousands of objects. ;))
Correlations matrix. Red color indicate values greater than 0.8, blue – greater then 0.6, green – greater than 0.4.
Principal components analysis. Coordinates of principal vectors are shown. Red color indicate the maximum value. Green – several big values.

First line - eigen values of the corresponding vector.
Second line – overall contribution in summary dispersion. From this you can determine **effective linear dimension** of the dataset.

In this example **Iris** dataset is 2D dimensional with accuracy 0.05 and 3D dimensional with accuracy 0.01.

**Tip**: print the table in Excel to plot different diagrams.
Histogram for every color of points in the dataset. Choose the field in the left listbox.
If you want to see the combined histogram, click **Unified**.
More sophisticated individual-object analysis. Choose an object in the left listbox. You can change the Mark of the object in the Field combobox. On the right diagram you can see distribution of other objects by their distances from the selected one.

In the middle you can evaluate the probabilities of field values for the object. In Iris dataset you have 150 records. For the 119th object (selected on the picture) you have probability of N2 field = \( \frac{4}{150} \), and probability of N2 in the same class = \( \frac{1}{150} \). It means that there are only 4 objects (including 119th) in the dataset with value of N2 in the same interval as the 119th object has. And only 119th object has this value of N2 in it’s class (green).
On Clustering tabsheet select method of clustering, number of clusters (not for all methods) and click Analyze. You will see the results on the Map panel.

Here you see the colors of the clusters, number of objects in every cluster and interclass deviation value (compactness).

Cancel will cancel the clustering.

Remember colors will assign for every point it’s current color (result of clustering).

Numbers in table will put information about cluster’s numbers in the datatable.

Distances in table will put information about distances from the centroids in the datatable.
You can see records colored accordingly to the cluster number.
If you have clicked the **Numbers in table** button, you will see the numbers of clusters.

**Tip:** right-click on the table and print the table in Excel worksheet or CSV file for the further analysis.
Hierarchical clustering, constructing minimal spanning tree. There are two modes for the method.

In **Hierarchical**, specify **number of clusters** method you have to specify the number of clusters.

In **Hierarchical** method you need to specify **Sensitivity** parameter on the **Sensitivity** trackbar. Sensitivity is the maximal length of edge to be cut.
You can change sizes of the points accordingly to some criteria.
For example, Distance from the closest node of the constructed net or Value of a field.
Choose criteria, specify minimal and maximal size of the point and click Analyze.
**Linear discriminative analysis.** Select the color of the class to be separated from others, mark the coordinates you want to use in decision function and click **Analyze.** You will see the result. Other classes are indicated by black color. Big points indicate error level of classification. In addition you see the coefficient’s values. For example, in this situation the decision function is

\[ f = -0.69 + 0.0032N1 - 0.14N2 + 0.42N3 + 0.57N4 \]

(you should use normalized values of N1, N2, N3, N4).
**Linear regression analysis.** Select the field to be calculated, mark the coordinates you want to use in regression function and click **Analyze.** Big points indicate error level of calculation. For example, in this situation the function is

\[ N1 = 5.84 + 0.28N2 + 1.24N3 - 0.42N4 \]

(you should use normalized values of N2, N3, N4, but the resulting value is not normalized).

Try to change value of **Quality** to test accuracy level.
You can visualize values of two linear functions: from Linear Discriminative Analysis and from Linear Regression Analysis using map coloring.

For example, on the picture you can see 0-1 valued linear decision function.
Select points by clicking on the **Select** button and dragging the mouse over the map. You can change properties of displaying of the selected points in the **Selection** dialog.
VIDAEXPERT: DATA ANALYSIS

Click this to switch between **Map** and **Table** panels.

Here you can see the selected points as records in the datatable.
You can assign for some points labels with their short description. In **Annotation** dialog choose which points do you need to label, content of the labels and click **Apply**.