

Non linear multiple regression with XLSTAT

[demoNLIN2.xls](#)

Dataset to run a non linear multiple regression

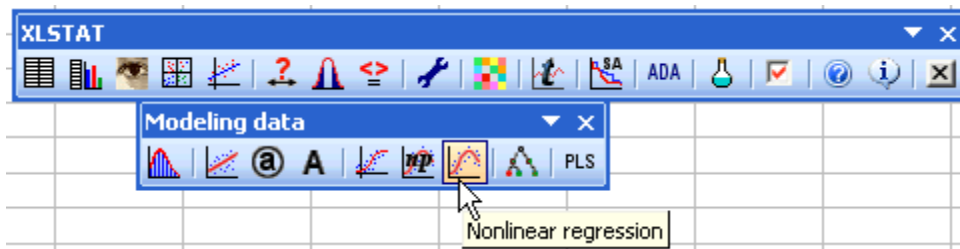
An Excel sheet with both the data and the results can be downloaded by clicking [here](#). Our purpose is to study the effect of the concentration of two components, C1 and C2, on the viscosity of a yogurt. The model that we want to fit writes:

$$F(C1, C2) = pr5 / (1 + \text{Exp}(-pr1 - pr2 * C1 - pr3 * C2 - pr4 * C1 * C2))$$

pr1, ..., pr5 are the parameters of the model. This logistic-like model allows you to take into account both the concentrations of the components and the interaction between them.

Setting up a non linear multiple regression

After opening XLSTAT, select the **XLSTAT / Modeling data / Nonlinear regression** command, or click on the corresponding button of the **Modeling Data** toolbar (see below).

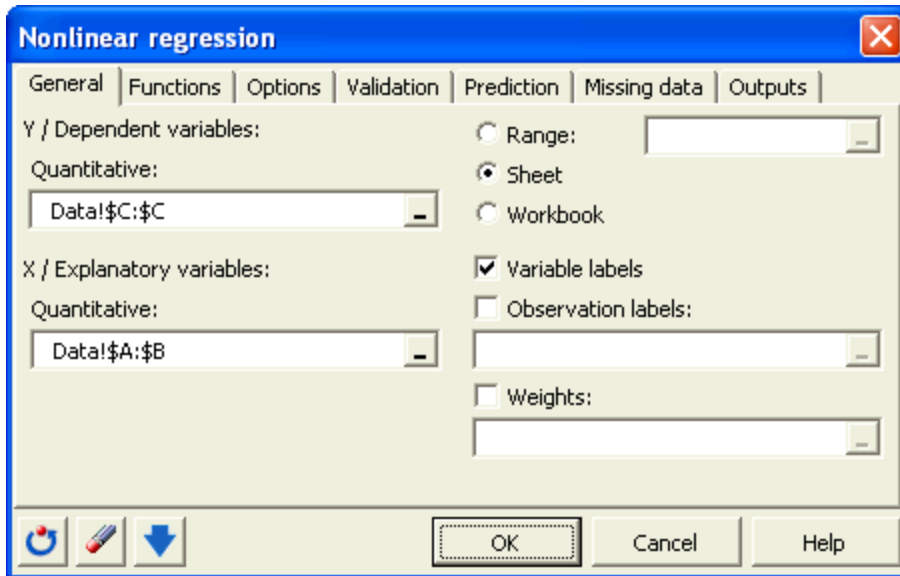


Once you've clicked on the button, the nonlinear regression dialog box appears. Select the data on the Excel sheet.

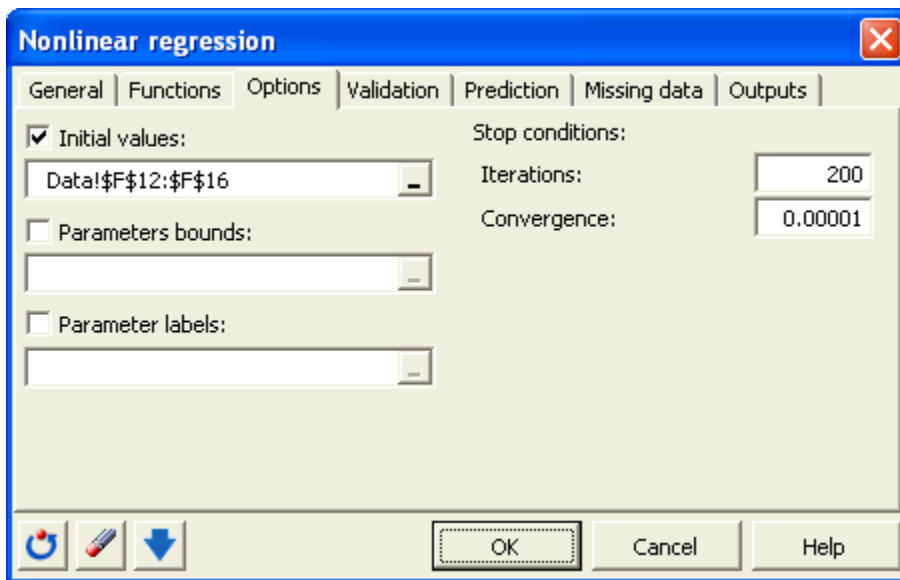
The **Dependent variable** (or response variable) is in our case the "Viscosity".

The **quantitative explanatory variables** are the concentration of the two components "C1" and "C2".

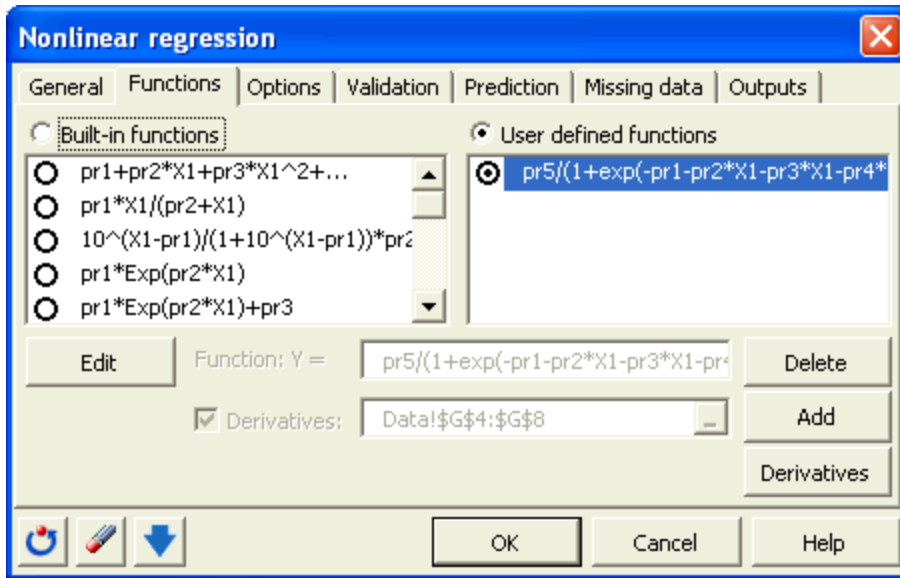
As we selected the column headers, we left the option **Variable labels** option activated. We left the **Residuals** option activated as well, because we want to analyze the predictions and the residuals.



In the **Options** tab we selected the values of the initial values of the five parameters.



In the **Functions** tab, the various functions are displayed. As the function we want to use is not listed in the **Preprogrammed functions** (you can notice the univariate version of the function in the list), we needed to enter the model: we first clicked on **Add**, then entered the function, then checked **Derivatives**, then selected them on the Excel sheet. In order to add this function to the user functions library, we clicked on **Save**. The function is then automatically added and selected.



The computations begin once you have clicked on the **OK** button. The results will then be displayed.

Interpreting the results of a non linear multiple regression

The first table gives the basic statistics of the selected variables.

The second table (see below) displays the goodness of fit coefficients, including the R^2 (coefficient of determination), and the SSE (sum of square of errors), the later being the criterion used for the model optimization. The R^2 corresponds to the % of the variability of the dependant variable (the viscosity) that is explained by the two explanatory variables (the components). The closer to 1 the R^2 is, the better the fit.

Goodness of fit statistics:	
Observatio	20,000
DF	15,000
R^2	0,923
SSE	0,004
MSE	0,000
RMSE	0,017
Iterations	200,000

In our case, 99% of the variability is explained by the two variables and their interaction, which is an excellent result that confirms that the selected model is appropriate.

The next table shows the results for the model parameters. As we can see, the ratios (parameter)/(std deviation) are larger for pr5 and pr4. As the same ratio is the largest for pr5 we

deduce that the interaction between the two components has a greater effect on the viscosity than the concentrations themselves.

Model parameters:		
Parameter	Value	Standard error
pr1	0,759	3,683
pr2	2,605	6,333
pr3	0,316	1,297
pr4	0,143	5,637
pr5	5,555	7,047

The following chart allows you to visualize the quality of the fit by comparing the predicted values to the observed values.

