

# Nonlinear regression with XLSTAT

[demoNLIN.xls](#)

## Dataset for nonlinear regression

An Excel sheet with both the data and the results can be downloaded by clicking [here](#). The data have come from [Ratkowsky D.A. (1983). Nonlinear Regression Modeling. New York, Marcel Dekker].

## Goal of this nonlinear regression

Using nonlinear regression, our goal is to see how the dry weight of onions vary with growth time. If it seems intuitive that a 3 parameters logistic regression would be suitable, the model suggested by Ratkowsky is more complex.

2 parameters Logistic model

$$f(x) = \frac{1}{1 + \exp(-p_1 - p_2 x)}$$

3 parameters Logistic model

$$f(x) = \frac{p_3}{1 + \exp(-p_1 - p_2 x)}$$

4 parameters Ratkowsky's model

$$f(x) = \frac{p_3}{1 + \exp(-p_1 - p_2 x)^{1/p_4}}$$

In order to show the capabilities of XLSTAT, and although the 3 parameters logistic model is directly available in XLSTAT, we show in this example how the user can create and add himself new functions to the available functions.

First, as the XLSTAT nonlinear regression does not include the Ratkowsky's model among the preprogrammed functions, we need to write the derivatives of the model for each of the 4 parameters. In the table below, we show the 4 derivatives and their transcription using the XLSTAT conventions: Excel syntax, with "pri" for parameter i and "Xj" for variable j. Note: if a derivative starts with a "-", you need to add a quote ' at the beginning to avoid that Excel detects an error.

Mathematical writing

$$\frac{\partial f}{\partial p_1} = \frac{p_3}{p_4} \frac{\text{Exp}(-p_1 - p_2 x)}{(1 + \text{Exp}(-p_1 - p_2 x))^{1+1/p_4}}$$

XLSTAT syntax

$(pr3/pr4)*exp(-pr1-pr2*X1)/(1+exp(-pr1-pr2*X1))^{(1+1/pr4)}$

Mathematical writing

$$\frac{\partial f}{\partial p_2} = \frac{p_3 x}{p_4} \frac{Exp(-p_1-p_2 x)}{(1+Exp(-p_1-p_2 x))^{1+1/p_4}}$$

XLSTAT syntax

$(pr3*X1/pr4)*exp(-pr1-pr2*X1)/(1+exp(-pr1-pr2*X1))^{(1+1/pr4)}$

Mathematical writing

$$\frac{\partial f}{\partial p_3} = \frac{1}{(1+Exp(-p_1-p_2 x))^{1+1/p_4}}$$

XLSTAT syntax

$1/(1+exp(-pr1-pr2*X1))^{(1/pr4)}$

Mathematical writing

$$\frac{\partial f}{\partial p_4} = \frac{p_3}{p_4^2} \frac{Ln(1+Exp(-p_1-p_2 x))}{(1+Exp(-p_1-p_2 x))^{1+1/p_4}}$$

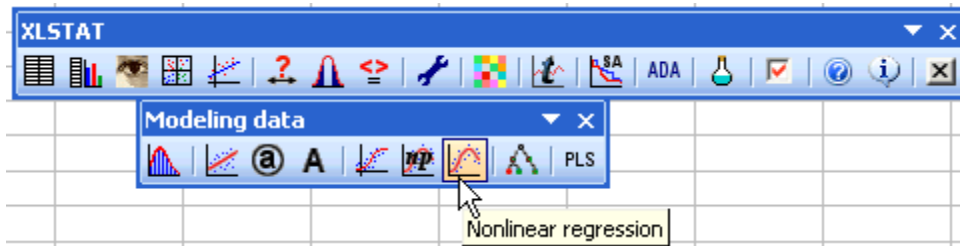
XLSTAT syntax

$(pr1/pr4^2)*ln(1+exp(-pr1-pr2*X1))/(1+exp(-pr1-pr2*X1))^{(1/pr4)}$

Last, when a function has a complex structure, it is advised that you specify a starting point. In our case, it seems that [0, 0, 725, 1] would be fine. 725 corresponds to the maximum of the dependent variable (Dry weight). The starting point and the derivatives must be vertically entered in Excel. Then, the analysis can start.

## Setting up a nonlinear 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).

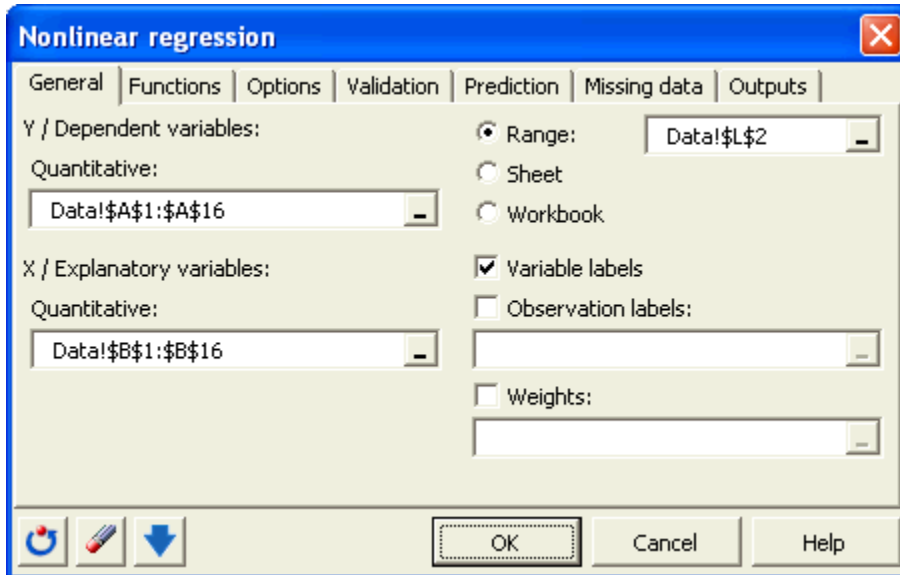


Once you've clicked on the button, the nonlinear regression dialog box appears. Select the data on the Excel sheet.

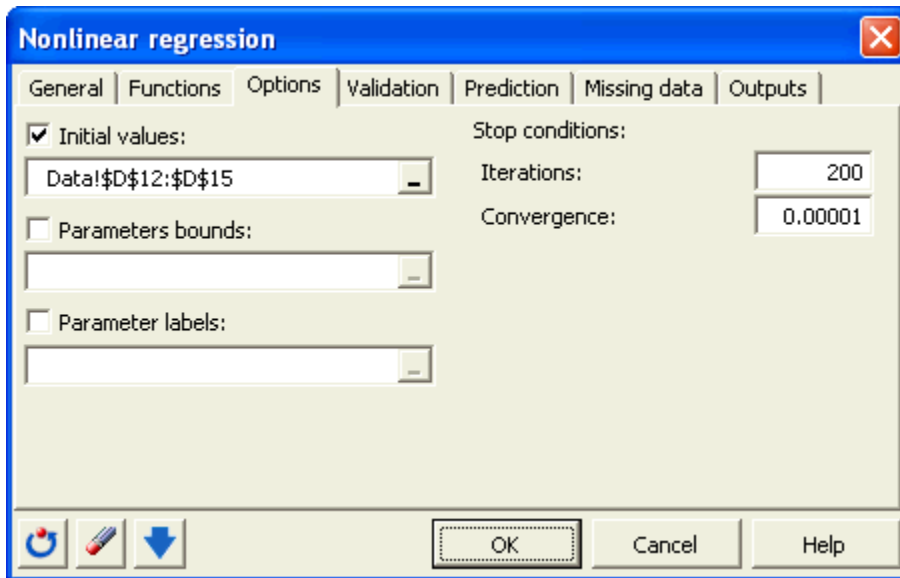
The **Dependent variable** (or variable to model, or response variable) is in our case the "Dry weight".

The **quantitative explanatory variable** is the "Time". We want to explain the variability of the "Dry Weight" by the "Time".

As we selected the column titles, we left the option **Variable labels** activated.

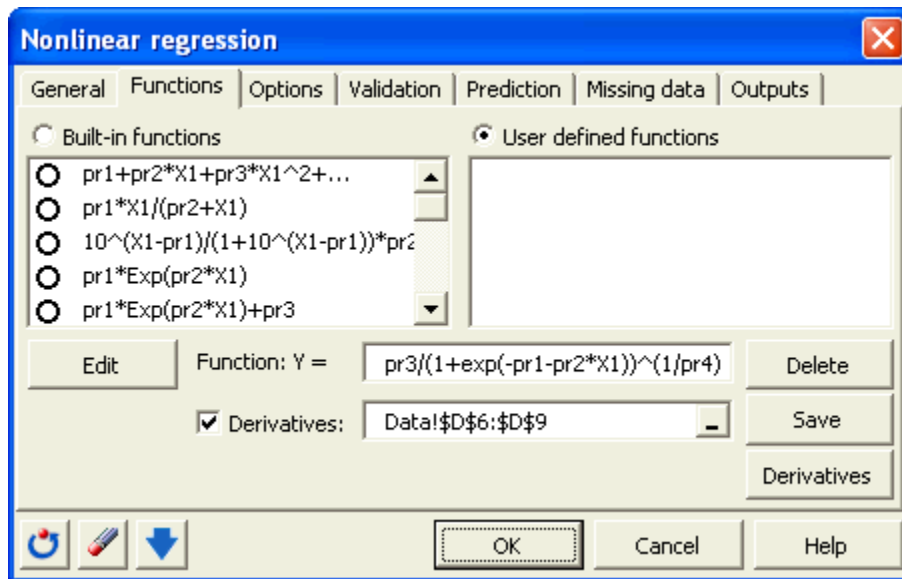


In the **Options** tab, we select the data that correspond to the starting point, that corresponds to the initial value of each parameter. *Note: a header must not be selected here.*



In the **Functions** tab, as the Ratkowsky's model is not listed in the **Preprogrammed functions** (on the opposite you can notice the 3 parameters logistic regression is available), we needed to enter the Ratkowsky's model: to add a user defined function, activate the corresponding option, then click on **Add**. Enter the function in the **Function: Y =** box.

Then we selected the derivatives on the Excel sheet (there is one derivative per parameter). 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 nonlinear regression

The first table gives basic statistics for the dependent and explanatory 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 dry weight) that is explained by the explanatory variable (the time). The closer to 1 the  $R^2$  is, the better the fit.

Goodness of fit statistics:	
Observatio	15,000
DF	11,000
$R^2$	0,992
SSE	8786,405
MSE	798,764
RMSE	28,262
Iterations	10,000

In our case, 99% of the variability of the Dry weight is explained by the Time, which is an excellent result.

The next table shows the detailed results for the model parameters. As we can see, the parameter pr3, which initial value was 725, is equal to 699.64 after the fitting. The standard deviation gives some idea of the reliability of result obtained for each parameter. The value of the parameter pr4 being close 1, one can imagine that a 3 parameters logistic regression would give an almost as good result, a hypothesis that you can easily test as this function is available in the preprogrammed functions list.

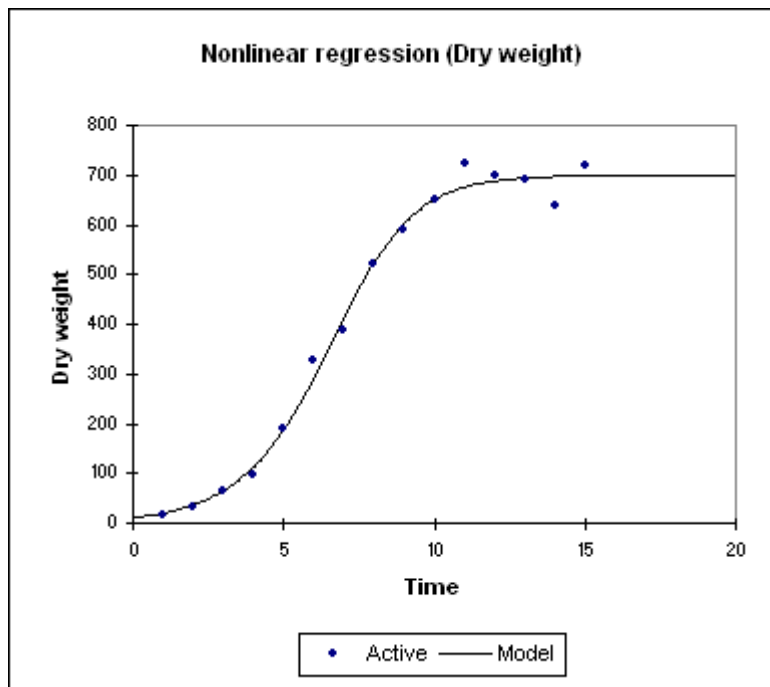
Model parameters:		
Parameter	Value	Standard error
pr1	-5,277	1,916
pr2	0,760	0,180
pr3	699,641	14,996
pr4	1,279	0,633

The equation of the model is displayed and can easily be reused in Excel.

Equation of the model:									
Dry weight = 699,641439748242/(1+exp(+5,27713504767677-0,759630407462012*Time))^(1/1,27925112599094)									

The next table (see the Excel sheet) shows the analysis of residuals. One can notice that the model is not well fitted for the observations 11 and 14.

The first chart (see below) allows to visualize the data and the fitted curve. The other charts allow to visually analyze the residuals, and are useful when the number of data is big.



## **Conclusion for this nonlinear regression**

As a conclusion, in the context of this analysis and of the selected model, the growth time allows to explain very well the dry weight of an onion.