

Running a multinomial logit model with XLSTAT

[DemoLOGmult.xls](#)

Multinomial logit model

The multinomial logit model is a generalization of the logit model when the response variable has more than two categories. This method is very useful when one wants to understand, or to predict, the effect of a series of variables on an unordered qualitative response variable (a variable which can take more than two values). Multinomial logit model can be helpful to model the effect of some descriptive variables on the choice of a brand in a market with more than two brands. All results are given relatively to a reference category (for example, the brand that is best established).

With XLSTAT you can run the multinomial logit on raw data. The dialog box for the multinomial logit model is the same as the one used for the logistic regression.

The methodology of multinomial logit model aims at modeling the probability associated to each category depending on the values of the explanatory variables, which can be categorical or numerical variables.

Dataset for running a multinomial logit model

The example treated here is a marketing case where we want to detect if customers are likely to choose one of three brands depending on their age and sex. An Excel sheet with both the data and the results can be downloaded by clicking [here](#).

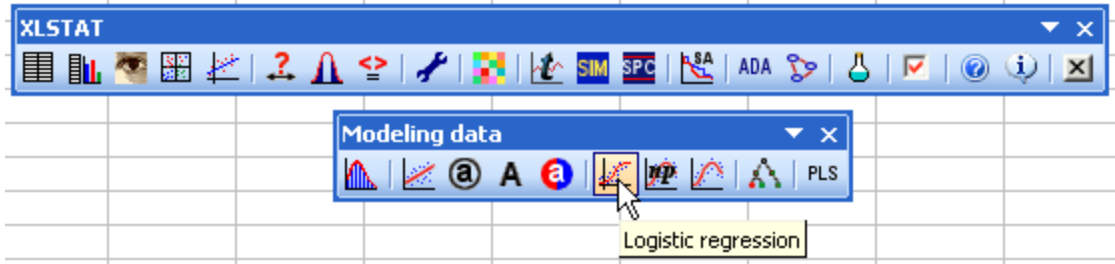
The data consist of 750 observations. The reference category is brand 1.

Goal of this multinomial logit model

Our goal is to understand if customers are more likely to choose brand 2 or 3 than brand 1 depending on their age and sex.

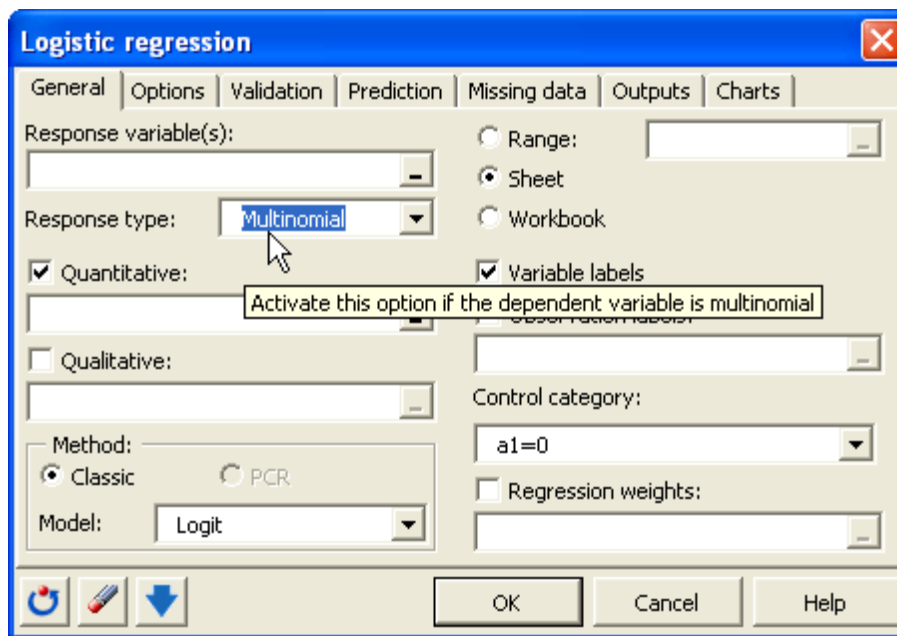
Setting up a multinomial logit model

To activate the Multinomial Logit Model dialog box, start XLSTAT, then select the **XLSTAT / Modeling data / Logistic regression** for binary response data command, or click on the **logistic regression** button of the **Modeling Data** toolbar (see below).



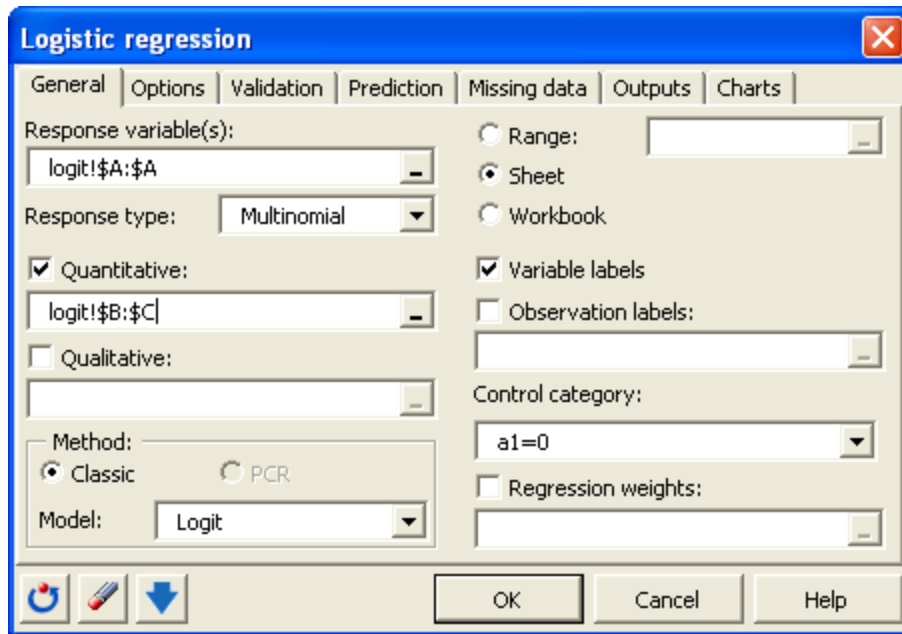
When you click on the button, the **Logistic regression** dialog box appears.

To activate the multinomial logit model, change the response type and choose **multinomial**. A new box appears where you can choose the control or reference category (in our case we choose $a_1=0$, meaning we set to 0 the relative effect for the first category).



Select the data on the Excel sheet. The **Response** corresponds to the column where the variable to be explained is stored. In this particular case we have two **quantitative explanatory variables**.

As we selected the column titles of all variables, we have selected the option **Variable labels included**.



Many options are available in the dialog box.

Note: The default options correspond to the basic choice one would make; please consult the XLSTAT help for more details.

The computations begin once you have clicked on the **OK** button.

Interpret the results of a multinomial logit model

The following table gives several indicators on the model quality (or goodness of fit). These results are equivalent to the R² and to the analysis of variance table in linear regression and ANOVA. The most important value to look at is the probability of the Chi-square test on the log ratio. This is equivalent to the Fisher's F test: we try to evaluate if the variables bring significant information by comparing the model as it is defined, with a simpler model with only one constant. In this case, as the probability is lower than 0.0001, we can conclude that significant information is brought by the variables.

Goodness of fit statistics (Variable BRAND):		
Statistic	Independent	Full
Observations	735	735
Sum of weights	735,000	735,000
DF	734	729
-2 Log(Likelihood)	1591,792	1405,941
R ² (McFadden)	0,000	0,117
R ² (Cox and Snell)	0,000	0,223
R ² (Nagelkerke)	0,000	0,252
AIC	1595,792	1417,941
SBC	1604,991	1445,541
Iterations	0	6

Test of the null hypothesis H0: Y=0,418 (Variable BRAND)			
Statistic	DF	Chi-square	Pr > Chi ²
-2 Log(Likelihood)	4	185,850	< 0,0001
Score	4	163,954	< 0,0001
Wald	4	129,797	< 0,0001

The next table gives details on the model. This table is helpful in understanding the effect of the various variables on the categories of the response variable. It is quite different from the logistic regression table. Parameters are obtained for each variable and for each category of the response variable (except the reference category). Odds ratios are also available for a better understanding of the results.

Model parameters (Variable BRAND):										
Category	Source	Value	Standard error	Wald Chi-Square	Pr > Chi ²	Lower bound	Upper bound	Odds ratio	Lower bound	Upper bound (95%)
2,000	Intercept	-11,775	1,775	44,024	< 0,0001	-15,253	-8,296			
	FEMALE	0,524	0,194	7,272	0,007	0,143	0,905	1,688	1,154	2,471
	AGE	0,368	0,055	44,813	< 0,0001	0,260	0,476	1,445	1,297	1,610
3,000	Intercept	-22,721	2,058	121,890	< 0,0001	-26,755	-18,688			
	FEMALE	0,466	0,226	4,247	0,039	0,023	0,909	1,594	1,023	2,482
	AGE	0,686	0,063	119,954	< 0,0001	0,563	0,809	1,986	1,756	2,245

Parameters interpretation is not immediate. The model equation for modality 2 is:

$$\text{Log}(P(\text{Response}=2)/P(\text{Response}=1)) = -11.775 + 0.524 * \text{FEMALE} + 0.368 * \text{AGE}$$

For example, we can say that for one unit change in the variable AGE, the log of the ratio of the two probabilities, $P(\text{Response}=2)/P(\text{Response}=1)$, will be increased by 0.368. Therefore, we can say that, in general, the older a person is, the more he will prefer brand 2. The ratio of the probability of choosing one outcome category over the probability of choosing the reference category is often referred as odds ratios (and it is also sometimes referred as relative risks). So another way of interpreting the regression results is in terms of odds ratios. We can say that for

one unit change in the variable AGE, we expect the relative risk of choosing brand 2 over 1 to increase by 1.445.

By looking at the probability of the Chi-squares, we see that the variable that most influences the response variable for both category 2 and 3 is the age of the customer. The intercepts are significant. The marketing experts should focus on older people if they want to increase the market share of the brand 1.