

The conditional logit model with XLSTAT-Conjoint

[demoCondLogit.xls](#)

The conditional logit model

The conditional logit model is a statistical method similar to logistic regression.

The conditional logit model is a method mostly used in its evolved form as part of the conjoint analysis, it is nevertheless useful when analyzing a certain type of data.

It is McFadden (1973) who introduced this model. Instead of having one line per individual, there will be as many lines as alternatives. Thus, it is no longer the characteristics of individuals that are modeled but those alternatives.

If one seeks to study travel modes, we will have four travel modes (car / train / air / bus), each travel mode has its own characteristics (price, speed) but an individual can choose only one of the four modes.

As part of a conditional logit model, we have for N individuals, $N*4$ rows with 4 rows associated with the four choices. The binary response variable will indicate the choice of the individual (1) and 0 if the individual did not choose this option.

A column associated with the name of the individuals (with 4 lines per individual for our example) has to be selected in XLSTAT. The explanatory variables will also have $N * 4$ rows.

Dataset for the conditional logit model

The example discussed below is a classic case in which one seeks to compare the travel modes proposed to go on vacation. It comes from Greene, W.H. (2003). *Econometric Analysis*, 5th edition. Upper Saddle River, NJ: Prentice Hall.

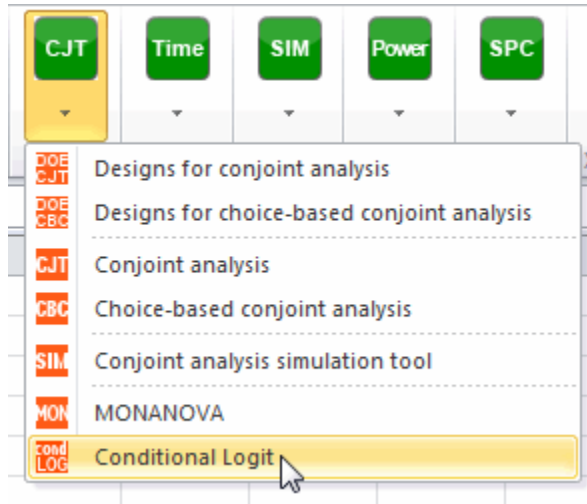
An Excel file containing both the data and the results can be downloaded by clicking [here](#).

The data correspond to a sample of 210 individuals, each one having 4 possibilities (air, car, bus and train). We asked each of them the travel mode they would choose to go on vacation.

The data set has 840 rows. The first column identifies the individual, the second is the binary variable modeling the travel mode. Then there are two quantitative variables, respectively, the overall cost and the waiting time during the trip associated with each travel mode for each individual. Finally, the categorical variable associated with the transportation is in the last column (air, train, bus or car).

Set a conditional logit model

To activate the dialog box, start XLSTAT, then select **XLSTAT / XLSTAT-Conjoint / Conditional Logit** , or click the corresponding button on the **XLSTAT-Conjoint** toolbar (see below).



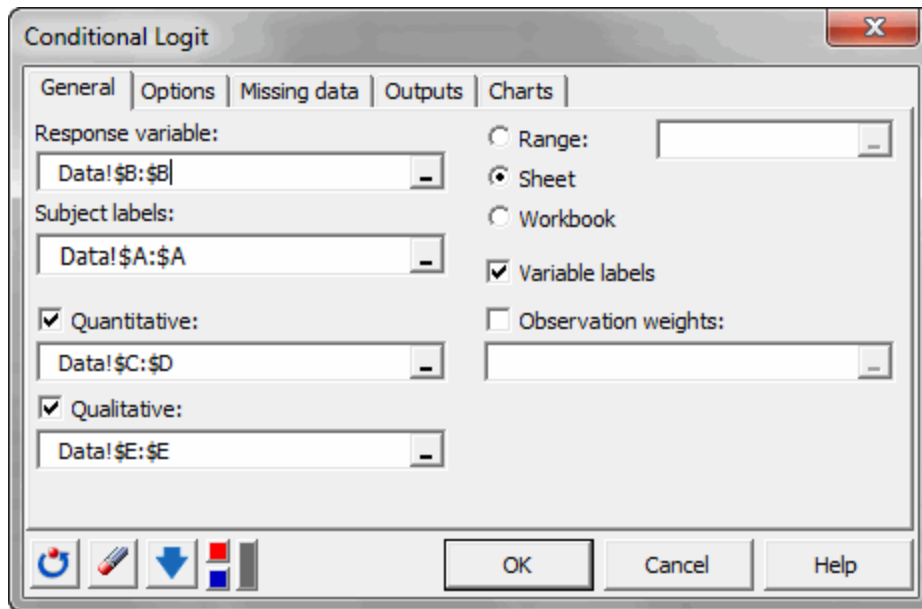
Once you have clicked the button, the dialog box appears.

Select the data on the Excel sheet.

The response variable corresponds to the binary variable.

The subject labels correspond to the numbers associated with the individuals (you can also have names of individuals instead).

In our case there are three predictors, one qualitative - the travel mode - and two quantitative - global cost and waiting time. As we selected the labels of the variables, we must select the variables labels option.



Once you click the OK button, the calculations are performed and the results displayed.

Interpret the results of a conditional logit model

The following table gives several indicators of the quality of the model (or goodness of fit). These results are similar to R^2 and to the analysis of variance table of linear regression and ANOVA. The most important value is the Chi-square associated with the log ratio (LR). This is the equivalent of the F test of Fisher's linear model: an attempt to assess whether the variables provide a significant amount of information to explain the variability of the binary variable. In our case, as the probability is less than 0.0001, we can conclude that the variables provide a significant amount of information.

Statistic	Independent	Full
Observations	840	840
Sum of weigh	840,000	840,000
DF	839	835
-2 Log(Likelih	582,244	399,953
R^2 (McFadden	0,000	0,313
R^2 (Cox and S	0,000	0,195
R^2 (Nagelkerk	0,000	0,390
AIC	584,244	409,953
SBC	587,591	426,689
Iterations	0	5

These goodness of fit statistics show that our model is significantly better than the model without any predictors. The following table confirms these initial impressions:

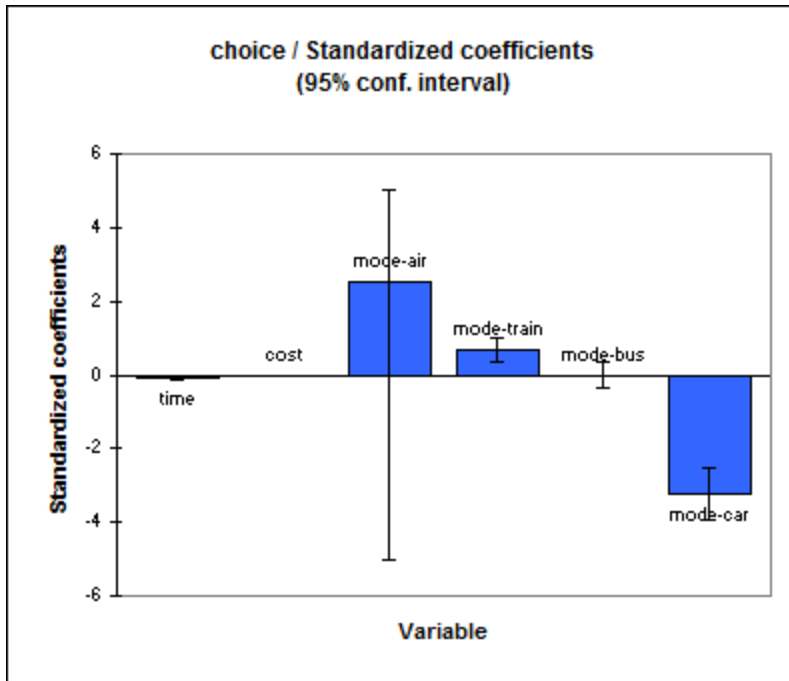
Test of the null hypothesis H0: Y=0 (Variable choice)			
Statistic	DF	Chi-square	Pr > Chi ²
-2 Log(Likelihood)	5	182,290	< 0,0001
Score	5	172,487	< 0,0001
Wald	5	103,592	< 0,0001

The p-values are all very small and the impact of the three variables is significant in the type III analysis table.

Type III analysis (Variable choice)			
Source	DF	Chi-square (Wald)	Pr > Wald
time	1	86,569	< 0,0001
cost	1	12,969	0,000
mode	3	86,465	< 0,0001

Finally, the coefficients of the model show that the air is preferred and that the waiting time has a significant negative effect on the choice of travel mode.

Standardized coefficients (Variable choice):						
Source	Value	Standard error	Wald Chi-Square	Pr > Chi ²	Lower bound	Upper bound (95%)
time	-0,097	0,010	86,569	< 0,0001	-0,118	-0,077
cost	-0,016	0,004	12,969	0,000	-0,024	-0,007
mode-air	2,549	0,323	62,226	< 0,0001	-4,996	4,996
mode-train	0,695	0,167	17,311	< 0,0001	0,368	1,023
mode-bus	-0,017	0,180	0,009	0,926	-0,369	0,336
mode-car	-3,228	0,364	78,547	< 0,0001	-3,941	-2,514



The analysis of residuals may also be useful and provide other information about individuals' choices.