

# Conduct and Interpret a Multinomial Logistic Regression

## *What is Multinomial Logistic Regression?*

Multinomial Logistic Regression is the linear regression analysis to conduct when the dependent variable is nominal with more than two levels. Thus it is an extension of logistic regression, which analyzes dichotomous (binary) dependents. Since the SPSS output of the analysis is somewhat different to the logistic regression's output, multinomial regression is sometimes used instead.

Like all linear regressions, the multinomial regression is a predictive analysis. Multinomial regression is used to describe data and to explain the relationship between one dependent nominal variable and one or more continuous-level(interval or ratio scale) independent variables.

Standard linear regression requires the dependent variable to be of continuous-level(interval or ratio) scale. Logistic regression jumps the gap by assuming that the dependent variable is a stochastic event. And the dependent variable describes the outcome of this stochastic event with a density function (a function of cumulated probabilities ranging from 0 to 1). Statisticians then argue one event happens if the probability is less than 0.5 and the opposite event happens when probability is greater than 0.5.

How can we apply the logistic regression principle to a multinomial variable (e.g. 1/2/3)?

Example:

*We analyze our class of pupils that we observed for a whole term. At the end of the term we gave each pupil a computer game as a gift for their effort. Each participant was free to choose between three games – an action, a puzzle or a sports game. The researchers want to know how the initial baseline (doing well in math, reading, and writing) affects the choice of the game. Note that the choice of the game is a nominal dependent variable with more than two levels. Therefore multinomial regression is the best analytic approach to the question.*

How do we get from logistic regression to multinomial regression? Multinomial regression is a multi-equation model, similar to multiple linear regression. For a nominal dependent variable with  $k$  categories the multinomial regression model estimates  $k-1$  logit equations. Although SPSS does compare all combinations of  $k$  groups it only displays one of the comparisons. This is typically either the first or the last category. The multinomial regression procedure in SPSS allows selecting freely one group to compare the others with.

What are logits? The basic idea behind logits is to use a logarithmic function to restrict the probability values to  $(0,1)$ . Technically this is the log odds (the logarithmic of the odds of  $y = 1$ ). Sometimes a probit model is used instead of a logit model for multinomial regression. The following graph shows the difference for a logit and a probit model for different values  $(-4,4)$ . Both

models are commonly used as the link function in ordinal regression. However, most multinomial regression models are based on the logit function. The difference between both functions is typically only seen in small samples because probit assumes normal distribution of the probability of the event, when logit assumes the log distribution.

At the center of the multinomial regression analysis is the task estimating the k-1 log odds of each category. In our k=3 computer game example with the last category as reference multinomial regression estimates k-1 multiple linear regression function defined as

$$\text{logit}(y=1) = \log\left(\frac{p(y=1)}{1-(p=1)}\right) = \beta_0 + \beta_1 \cdot x_{i2} + \beta_2 \cdot x_{i2} + \dots + \beta_p \cdot x_{in} \text{ for } i = 1 \dots n.$$

$$\text{logit}(y=2) = \log\left(\frac{p(y=2)}{1-(p=2)}\right) = \beta_0 + \beta_1 \cdot x_{i2} + \beta_2 \cdot x_{i2} + \dots + \beta_p \cdot x_{in} \text{ for } i = 1 \dots n.$$

Multinomial regression is similar to the Multivariate Discriminant Analysis. Discriminant analysis uses the regression line to split a sample in two groups along the levels of the dependent variable. In the case of three or more categories of the dependent variable multiple discriminant equations are fitted through the scatter cloud. In contrast multinomial regression analysis uses the concept of probabilities and k-1 log odds equations that assume a cut-off probability 0.5 for a category to happen. The practical difference is in the assumptions of both tests. If the data is multivariate normal, homoscedasticity is present in variance and covariance and the independent variables are linearly related, then we should use discriminant analysis because it is more statistically powerful and efficient. Discriminant analysis is also more accurate in predictive classification of the dependent variable than multinomial regression.

### ***The Multinomial Logistic Regression in SPSS***

For multinomial logistic regression, we consider the following research question:

*We conducted a research study with 107 students. The students were measured on a standardized reading, writing, and math test at the start of our study. At the end of the study, we offered every pupil a computer game as a thank you gift. They were free to choose one of three games – a sports game, a puzzle and an action game. How does the pupils' ability to read, write, or calculate influence their game choice?*

First we need to check that all cells in our model are populated. Although the multinomial regression is robust against multivariate normality and therefore better suited for smaller samples than a probit model, we still need to check. We find that some of the cells are empty. We must therefore collapse some of the factor levels. The easiest way to check is to create the contingency table (*Analyze/Descriptive Statistics/Crosstabs...*).

But even if we collapse the factor levels of our multinomial regression model down to two levels (performance good vs. not good) we observe empty cells. We proceed with the analysis regardless, noting and reporting this limitation of our analysis.

Multinomial Regression is found in SPSS under *Analyze/Regression/Multinomial Logistic...*

This opens the dialog box to specify the model. Here we need to enter the dependent variable *Gift* and define the reference category. In our example it will be the last category since we want to use the sports game as a baseline. Then we enter the three collapsed factors into the multinomial regression model. The factors are performance (good vs. not good) on the math, reading, and writing test.

In the menu *Model...* we need to specify the model for the multinomial regression. The huge advantage over ordinal regression analysis is the ability to conduct a stepwise multinomial regression for all main and interaction effects.

If we want to include additional measures about the multinomial regression model to the output we can do so in the dialog box *Statistics...*

***Output, syntax, and interpretation can be found in our downloadable manual: [Statistical Analysis: A Manual on Dissertation Statistics in SPSS](#) (included in our member resources). [Click here to download.](#)***