

Multinomial Logistic Regression Models

Multinomial logistic regression models estimate the association between a set of predictors and a multicategory nominal (unordered) outcome. Examples of such an outcome might include “yes,” “no,” and “don’t know”; “Apple iPhone,” “Android,” and “Samsung Galaxy”; or “walk,” “bike,” “car,” “public transit.” The most common form of the model is a logistic model that is a generalization of the binary outcome of standard logistic regression involving comparisons of each category of the outcome to a referent category. There are J total categories of the outcome, indexed by the subscript j , and the number of comparisons is then $J - 1$. The equation for the model is written in terms of the logit of the outcome, which is a comparison of a particular category to the referent category, both denoted π_j here.

$$\ln\left(\frac{\pi_j}{\pi_j}\right) = \alpha_j + \beta_j X$$

The natural log of the ratio of the two proportions is the same as the logit in standard logistic regression, where $\ln(\pi_j/\pi_j)$ replaces $\ln[\pi/(1-\pi)]$, and is sometimes referred to as the *generalized logit*. The binary logistic model is therefore a special case of the multinomial model. In generalized linear modeling terms, the link function is the generalized logit and the random component is the multinomial distribution. The model differs from the standard logistic model in that the comparisons are all estimated simultaneously within the same model. The j subscript on both the intercept, α_j , and slope, β_j , indicate that there is an intercept and a slope for the comparison of each category to the referent category. Note that in the ordinal logistic model, there is only one slope coefficient for each predictor. Odds ratios for each coefficient (for predicting the difference of one category response from the referent) are computed as usual, with $OR = e^{\beta_j}$, and represent the odds increase (or decrease) for category j compared with the referent category for each unit increase in X .

The predicted probabilities can be computed from the model parameters for a specific value of X . For the standard logistic regression, we used the logistic transformation to find the probability according to the logistic cumulative distribution function (cdf; see the “Logistic Regression” handout). For a simple logistic regression with one predictor, we used

$$\pi = \frac{1}{1 + e^{\alpha + \beta X}}$$

Entering in a specific value of X and the model estimates of α and β and using the exponential function, the estimate of the expected probability can be computed for the specific value of X . (For additional predictors, the values for X and β for those variables are added to the exponent in the denominator.) Predicted probabilities (y-axis) are then often plotted with a separate line for each comparison as a function of the X variable values (x-axis).

The cdf transformation for the multinomial distribution must add the exponent functions of the intercepts and the coefficients for each of the comparisons to the referent category.¹ For a single predictor, the predicted probability can be computed by generalizing the above equation for standard logistic, using the following equation with as many additional $J - 1$ terms in the denominator for every comparison to the referent category:

$$\pi_j = \frac{1}{1 + e^{\alpha_j + \beta_j X} + e^{\alpha_{j-1} + \beta_{j-1} X} \dots e^{\alpha_{j-1} + \beta_{j-1} X}} = \frac{1}{1 + \sum e^{\alpha_j + \beta_j X}}$$

¹ The referent category in the logistic or the multinomial logistic have e^0 , which is 1. This is why 1 appears in the numerator and denominator.

The result is the estimated proportion for the referent category relative to the total of the proportions of all categories combined (1.0), given a specific value of X and the intercept and slope coefficient(s).

Maximum likelihood is the most common estimation used for multinomial logistic regression. And, as with logistic regression, model fit tests, such as the likelihood ratio test with degrees of freedom equal to $J - 1$,² are used to determine whether together all of the comparisons to the referent are significant.

The multinomial logistic models assume that there is *independence of irrelevant alternatives* (IIA). The assumption is that if an additional category was to be added to the outcome, the proportions for the original categories would be equally affected by adding the new category (e.g., adding a third party candidate would equally impact votes for the two major party candidates).³ As this example suggests, the IIA assumption is not particularly realistic in many situations, even though it is needed for truly unbiased estimates of the observed and predicted proportions. Although tests have been suggested to investigate violation of the assumption, they do not appear to perform well (e.g., Cheng & Long, 2005; Fry & Harris, 1996).

Multinomial Probit Models

Multinomial probit models analogous to the binary probit model are also possible, and have been considered as one potential solution that would be free of the IIA assumption. For the multinomial probit model, the probit link is used with multivariate normal distribution random component. The use of the multivariate normal distribution instead of the logistic distribution allows correlations among possible alternatives. The multinomial probit model, however, has other difficulties that make it a less than optimal alternative. For the multinomial probit model to have a mathematically identified solution, the predictors must be included that are associated with each specific alternative. Restrictions on the correlations among errors are also necessary to obtain estimates, and there may not be a theoretical basis for making particular restrictions. Multinomial probit is not available in the current version of SPSS but can be estimated in Limdep (maximum likelihood) or MNP (Bayesian) package in R, PROC MDC in SAS (maximum likelihood).

Discrete Choice Models

The term *discrete choice model* is very generally applied to binary and multiple-category (ordinal or nominal) outcomes. No particular estimation approach or statistical model is implied by the use of the term discrete choice, but most often the reference to discrete choice is to multinomial models (most typically multinomial logistic). The terminology is applied in consumer choice studies in marketing (McFadden, 1974), transportation research, and economics among other fields. The goal is simply to predict one decision over another using a set of explanatory factors. Model results are often described in terms of the economic theory of utility (systematic function of the predictors and random error) of one choice over another, representing the amount of gain to an individual by selecting one choice over another.

References and Further Reading

- Ben-Akiva, M., & Lerman, S. (1985). *Discrete choice models*. London: MIT Press.
- Cheng, S., & Long, J.S. (2007). Testing for IIA in the Multinomial Logit Model. *Sociological Methods & Research*, 35, 583-600.
- Fry, T. R. L. F., & Harris, M.N. (1996). A Monte Carlo study of tests for the independence of irrelevant alternatives property. *Transportation Research Part B: Methodological* 30, 19-30.
- Long, J.S. (2012). Regression models for nominal and ordinal outcomes. In Best and Wolf (Eds), *Regression Models*. Thousand Oaks, CA: Sage Publications
- McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics*. New York: Academic Press.

² SAS prints the score and Wald test for the model as well.

³ The widely used example is the assumption that the original transportation choices of car and red bus would be equally affected if the choice was between a car, red bus, or blue bus.