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 $\ln[\pi/(1\text{-}\pi)]$, 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:

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

[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.

Multinomial Logistic Example
SPSS

Multinomial logistic models can be estimated in SPSS using the `nomreg` procedure and in R using the `mlogit` package or the `nnet` package and the `multinom` function.

```
GET FILE='c:\jason\spsswin\cdaclass\multinomial.sav'.

nomreg work (base = 1) with age srh married
/print = paramter summary cps mfi.
```

[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.

Case Processing Summary

		N	Marginal Percentage
work work status	1.00 full	29	5.7%
	2.00 part	38	7.4%
	3.00 unemployed	19	3.7%
	4.00 retired	426	83.2%
Valid		512	100.0%
Missing		54	
Total		566	
Subpopulation		172[a]	

a. The dependent variable has only one value observed in 117 (68.0%) subpopulations.

Model Fitting Information

Model	Model Fitting Criteria: -2 Log Likelihood	Likelihood Ratio Tests: Chi-Square	df	Sig.
Intercept Only	434.296			
Final	382.533	51.762	9	.000

Pseudo R-Square

Cox and Snell	.096
Nagelkerke	.134
McFadden	.080

Parameter Estimates

work work status[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B): Lower Bound	Upper Bound
2.00 part	Intercept	-3.311	4.091	.655	1	.418			
	age A3a-age of R	.078	.056	1.929	1	.165	1.081	.968	1.207
	srh B1-your health at present	-.685	.264	6.742	1	.009	.504	.301	.845
	married marital status	.026	.543	.002	1	.962	1.026	.354	2.974
3.00 unemployed	Intercept	-2.291	4.752	.232	1	.630			
	age A3a-age of R	.058	.065	.777	1	.378	1.059	.932	1.204
	srh B1-your health at present	-.655	.305	4.597	1	.032	.520	.286	.945
	married marital status	-.563	.625	.812	1	.367	.569	.167	1.938
4.00 retired	Intercept	-7.577	3.438	4.857	1	.028			
	age A3a-age of R	.175	.047	13.549	1	.000	1.191	1.085	1.307
	srh B1-your health at present	-.771	.222	12.100	1	.001	.463	.300	.714
	married marital status	-.180	.436	.171	1	.679	.835	.355	1.962

a. The reference category is: 1.00 full.

R

```
> #use lessR routine for listwise deletion
> library(lessR)
> mydata <-Subset(work!='NA' & age!='NA' & srh!='NA' & married!='NA')

> #make sure dv is a factor
> d$work <- factor(d$work)
>
> library(nnet)
> d$work <- relevel(d$work, ref = 1)
> model <- multinom(work ~ age + srh + married, data = d)
# weights:  20 (12 variable)
initial  value 709.782713
iter  10 value 325.755760
iter  20 value 297.417808
iter  30 value 297.133699
iter  40 value 297.125872
final  value 297.125651
converged
```

```
> summary(model)
Call:
multinom(formula = work ~ age + srh + married, data = d)

Coefficients:
  (Intercept)        age        srh     married
2   -3.311567 0.07788846 -0.6846427  0.02574789
3   -2.291723 0.05759751 -0.6545843 -0.56346491
4   -7.577511 0.17462753 -0.7710237 -0.18024777

Std. Errors:
  (Intercept)        age       srh   married
2    4.090534 0.05608322 0.2636701 0.5429907
3    4.752491 0.06532525 0.3052976 0.6251271
4    3.438072 0.04744140 0.2216583 0.4359652

Residual Deviance: 594.2513
AIC: 618.2513
>
> #obtain odds ratios
> exp(cbind(OR=coef(model), confint(model)))
   (Intercept)      age       srh   married
2 0.0364589851 1.081002 0.5042704 1.0260822 0.00001202087
3 0.1010921469 1.059289 0.5196580 0.5692333 0.96847498115
4 0.0005118338 1.190803 0.4625393 0.8350633 0.30076485647
```

SAS

```
proc logistic data=one ;
model work (ref=first) = age srh married / link=glogit;
run;
```

The LOGISTIC Procedure

Model Information

Data Set	WORK.ONE	
Response Variable	work	work status
Number of Response Levels	4	
Model	generalized logit	
Optimization Technique	Newton-Raphson	

Number of Observations Read	566
Number of Observations Used	512

Response Profile

Ordered Value	work	Total Frequency
1	full	29
2	part	38
3	retired	426
4	unemployed	19

Logits modeled use work='full' as the reference category.
NOTE: 54 observations were deleted due to missing values for the response or explanatory variables.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	652.014	618.251
SC	664.729	669.111
-2 Log L	646.014	594.251

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	51.7624	9	<.0001
Score	45.5293	9	<.0001
Wald	40.2975	9	<.0001

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
age	3	25.1953	<.0001
srh	3	12.2197	0.0067
married	3	1.2012	0.7527

Analysis of Maximum Likelihood Estimates

Parameter	work	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	part	1	-3.3093	4.0902	0.6546	0.4185
Intercept	retired	1	-7.5753	3.4377	4.8558	0.0276
Intercept	unemployed	1	-2.2894	4.7522	0.2321	0.6300
age	part	1	0.0779	0.0561	1.9274	0.1650
age	retired	1	0.1746	0.0474	13.5471	0.0002
age	unemployed	1	0.0576	0.0653	0.7766	0.3782
srh	part	1	-0.6846	0.2637	6.7422	0.0094
srh	retired	1	-0.7710	0.2216	12.0998	0.0005
srh	unemployed	1	-0.6546	0.3053	4.5970	0.0320
married	part	1	0.0257	0.5430	0.0022	0.9622
married	retired	1	-0.1803	0.4359	0.1710	0.6792
married	unemployed	1	-0.5635	0.6251	0.8125	0.3674

Odds Ratio Estimates

Effect	work	Point Estimate	95% Wald Confidence Limits	
age	part	1.081	0.968	1.207
age	retired	1.191	1.085	1.307
age	unemployed	1.059	0.932	1.204
srh	part	0.504	0.301	0.845
srh	retired	0.463	0.300	0.714
srh	unemployed	0.520	0.286	0.945
married	part	1.026	0.354	2.974
married	retired	0.835	0.355	1.962
married	unemployed	0.569	0.167	1.938

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.