

Multinomial Logistic Regression using SAS and PROC CATMOD¹

Multinomial Logistic Regression can be used with a categorical dependent variable that has more than two categories. Multinomial Logistic Regression can be done with SAS using PROC CATMOD. For this example, the dependent variable marcat is marital status. The syntax for proc catmod is as follows.

```
PROC CATMOD ;
weight adjwt ;          * This procedure DOES NOT normalize weights. If using a
                        weight, be sure to adjust so the mean of weights is 1. See:
                        http://www.albany.edu/csda/adjsas.pdf
DIRECT black age anychild ; * To make results more easily interpretable, include all
                             independent variables in the DIRECT statement (dummy coded
                             and continuous variables) ;
MODEL marcat = black age anychild / ML NOITER NODESIGN NOPROFILE NOGLS;
* ML asks for Maximum Likelihood estimates. The options that begin with "NO" ask SAS
  to refrain from showing you some output that it would show you by default.
```

The dependent variable marcat is marital status. This example uses 1990 IPUMS data. All black and white women aged 25 to 45 were included in the sample. The independent variables are:

- 1) Black Black women are coded 1, and white women are coded 0.
- 2) Age Woman's age
- 3) Anychild Coded 1 if the woman has an "own" child living in her household with her.

The weighted means of all of the variables are:

Variable	N	Mean	Std Dev	Minimum	Maximum
marcat	399307	3.2822040	1.1245330	1.0000000	4.0000000
black	399307	0.1293563	0.3355943	0	1.0000000
AGE	399307	34.5148307	5.8836194	25.0000000	45.0000000
anychild	399307	0.6661534	0.4715863	0	1.0000000

The weighted frequencies for the dependent variable are:

	marcat	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1 never married		69252.05	17.34	69252.05	17.34
2 widowed		4277.163	1.07	73529.21	18.41
3 div, sep		70310.48	17.61	143839.7	36.02
4 Married, spouse present		255467.3	63.98	399307	100.00

By default, PROC CATMOD uses the highest numbered category as the comparison group. It is possible to make the comparison group the lowest category by sorting the data in descending order, and then specifying "order = data" on the PROC CATMOD statement. For example:

```
proc sort; by descending marcat ;
PROC CATMOD order = data ;
```

However, you should remember that sorting can be time, CPU and disk intensive with large data sets. It may be better to create the dependent variable to conform with SAS defaults.

¹Prepared by Patty Glynn, Deenesh Sohoni, and Laura Leith, University of Washington, 3/14/02 C:\all\help\helpnew\catmod.wpd, modified 12/5/03

The CATMOD Procedure
Data Summary

Response	marcat	Response Levels	4
Weight Variable	ADJWT	Populations	84
Data Set	XX	Total Frequency	399307
Frequency Missing	0	Observations	399307

Maximum Likelihood Analysis
Maximum likelihood computations converged.
Maximum Likelihood Analysis of Variance

Source	DF	Chi-Square	Pr > ChiSq
Intercept	3	20481.74	<.0001
black	3	25635.56	<.0001
AGE	3	15149.88	<.0001
anychild	3	63473.99	<.0001
Likelihood Ratio	240	6242.95	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	Function Number	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	2.9108	0.0302	9287.64	<.0001
	2	-7.4275	0.1140	4242.79	<.0001
	3	-1.4108	0.0275	2639.26	<.0001
black	1	2.1762	0.0144	22717.52	<.0001
	2	1.5502	0.0371	1749.19	<.0001
	3	1.3374	0.0125	11531.87	<.0001
AGE	1	-0.0931	0.000912	10428.48	<.0001
	2	0.1024	0.00290	1249.08	<.0001
	3	0.0205	0.000756	732.42	<.0001
anychild	1	-2.9085	0.0118	61029.54	<.0001
	2	-0.8530	0.0336	643.88	<.0001
	3	-1.0820	0.00945	13096.45	<.0001

There are three lines of output for each independent variable. Examine the column labeled as "Function Number".

- 1 = lowest category compared to highest (never married / married spouse present)
- 2 = 2nd lowest category compared to highest (widowed / married spouse present)
- 3 = 3rd lowest category compared to highest (sep, divorced / married spouse present)

An example of interpreting results: Women who have any of their own children living with them are less likely to be never-married (-2.9085) , widowed (-0.8530) , or divorced or separated (-1.0820), when controlling for race and age.

The odds ratio - or exponent of the regression coefficient is not presented in the output. It can be calculated with a spreadsheet - for example, using Excel, =exp(-2.9085) = 0.054558

An example of presenting results for multinomial logistic regression follows.

Results of Multinomial Logistic Regression, Marital Status of Black and White Women Age 25-45.

	Never Married	Widowed	Divorced/Separated
Black	2.18*** (.01)	1.55*** (.04)	1.34*** (.01)
Age	-0.09*** (.00)	0.10*** (.00)	0.02*** (.00)
Own Child in home	-2.91*** (.01)	-0.85*** (.03)	-1.08*** (.01)
Intercept	2.91*** (.03)	-7.43*** (.11)	-1.41*** (.03)
N	69,252	4,277	70,310

Total N = 399,307

Notes: Reference category for the equation is Married with Spouse Present.

Standard errors in parentheses.

* p # .05 ** p # .01 *** p # .001 (two-tailed tests).

The program used for these results follows.

```
* mlogit2.sas ;
title1 'mlogit2, use 19902 IPUMS, white and black women age 25 to to 45' ;
options compress = yes linesize = 72 ;
libname t2 'C:\all\help\helpnew\mlogit' ;

proc format ;
  value marcat
    4 = 'Married, spouse present'    3 = 'div, sep'
    2 = 'widowed'                    1 = 'never married' ;

data xx ; set t2.mlogit;
proc means ;
weight adjwt ;
var marcat black age anychild ;
run ;
proc freq ;
tables marcat ;
weight adjwt ;
format marcat marcat. ; run ;
PROC CATMOD ;
weight adjwt ;
DIRECT black age anychild ;
MODEL marcat = black age anychild / ML NOITER NODESIGN NOPROFILE NOGLS;
format marcat marcat. ;
run ;
```