

## Multinomial Logistic Regression using STATA and MLOGIT<sup>1</sup>

Multinomial Logistic Regression can be used with a categorical dependent variable that has more than two categories. Maximum-likelihood multinomial (polytomous) logistic regression can be done with STATA using `mlogit`. For this example, the dependent variable `marcat` is marital status. This example uses 1990 IPUMS data, and includes black and white women 25 to 45. 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:

```
. sum marcat black age anychild [weight= adjwt]
(analytic weights assumed)
```

Variable	Obs	Weight	Mean	Std. Dev.	Min	Max
marcat	399307	399306.998	3.282204	1.124533	1	4
black	399307	399306.998	.1293563	.3355943	0	1
age	399307	399306.998	34.51483	5.883619	25	45
anychild	399307	399306.998	.6661534	.4715863	0	1

The weighted frequencies for the dependent variable are:

```
. tab marcat [iweight= adjwt]
```

marcat	Freq.	Percent	Cum.
1Never Mar	69252.0493	17.34	17.34
2Widow	4277.16333	1.07	18.41
3Div/sep	70310.481	17.61	36.02
4Married	255467.304	63.98	100.00
Total	399306.998	100.00	

Remember that STATA is case sensitive - for variable names as well as commands. The STATA command to ask for multinomial logistic regression is:

```
mlogit marcat black age anychild [pweight= adjwt], basecategory(4)
```

The option "pweight" is described in STATA documentation: "pweights, or sampling weights, are weights that denote the inverse of the probability that the observation is included due to the sampling design." STATA normalizes weights in this procedure so it is not necessary to adjust their mean. By default, Stata will use the most frequent category for the comparison group. The "basecategory" option allows you to specify the category to be used for comparison.

The results follow:

---

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

Multinomial regression

Number of obs = 399307  
Wald chi2(9) = 69988.28  
Prob > chi2 = 0.0000  
Pseudo R2 = 0.1708

Log likelihood = -312559.9

marcat	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
<b>1Never Mar</b>						
black	2.176214	.0182961	118.94	0.000	2.140355	2.212074
age	-.0931311	.0010538	-88.37	0.000	-.0951966	-.0910656
anychild	-2.908525	.0128889	-225.66	0.000	-2.933787	-2.883264
_cons	2.910785	.0349539	83.27	0.000	2.842276	2.979293
<b>2Widow</b>						
black	1.550249	.0444795	34.85	0.000	1.463071	1.637428
age	.1023641	.0033663	30.41	0.000	.0957663	.108962
anychild	-.8529765	.0389209	-21.92	0.000	-.9292601	-.7766929
_cons	-7.427532	.1327063	-55.97	0.000	-7.687632	-7.167433
<b>3Div/sep</b>						
black	1.33742	.0147771	90.51	0.000	1.308457	1.366383
age	.0204502	.0008426	24.27	0.000	.0187988	.0221017
anychild	-1.081962	.0106939	-101.18	0.000	-1.102921	-1.061002
_cons	-1.410772	.0306294	-46.06	0.000	-1.470804	-1.350739

(Outcome marcat==4Married is the comparison group)

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 = 2<sup>nd</sup> lowest category compared to highest ( widowed / married spouse present )
- 3 = 3<sup>rd</sup> 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.

If you want odds ratios reported rather than coefficients, add the option rrr as follows:

```
mlogit marcat black age anychild [pweight= adjwt ], rrr basecategory(4)
```

You can change the comparison group by adding the option "base (value)" For example:

```
mlogit marcat black age anychild [pweight= adjwt ], rrr base (1)
```

The commands used for these results follow.

```
log using "C:\all\help\helpnew\mlogit\mlogit_stata.log"  
set memory 1000m  
use "C:\all\help\helpnew\mlogit\mlogit.dta" , clear  
set more off  
label define marcat 1 "1Never Mar" 2 "2Widow" 3 "3Div/sep" 4 "4Married"  
label values marcat marcat  
sum marcat black age anychild [weight= adjwt]  
tab marcat [iweight= adjwt]  
mlogit marcat black age anychild [pweight= adjwt], basecategory(4)  
log close
```

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).