

## Post-Estimation Commands for MLogit

Richard Williams, University of Notre Dame, <https://www3.nd.edu/~rwilliam/>  
Last revised September 27, 2024

These notes borrow heavily (sometimes verbatim) from Long & Freese, 2014 [Regression Models for Categorical Dependent Variables Using Stata, 3rd Edition](#).

Many/most of the Stata & `spost13` post-estimation commands work pretty much the same way for `mlogit` as they do for `logit` and/or `ologit`. We'll therefore concentrate primarily on the commands that are somewhat unique.

**Making comparisons across categories.** By default, `mlogit` sets the base category to the outcome with the most observations. You can change this with the `basecategory` option. `mlogit` reports coefficients for the effect of each independent variable on each category relative to the base category. Hence, you can easily see whether, say, `yr89` significantly affects the likelihood of your being in the SD versus the SA category; but you can't easily tell whether `yr89` significantly affects the likelihood of your being in, say, SD versus D, when neither is the base. You could just keep rerunning models with different base categories; but `listcoef` makes things easier by presenting estimates for all combinations of outcome categories.

```
. use https://www3.nd.edu/~rwilliam/statafiles/ordwarm2.dta
(77 & 89 General Social Survey)
. mlogit warm i.yr89 i.male i.white age ed prst, b(4) nolog
```

```
Multinomial logistic regression              Number of obs   =       2293
                                             LR chi2(18)     =       349.54
                                             Prob > chi2     =       0.0000
Log likelihood = -2820.9982                 Pseudo R2      =       0.0583
```

	warm	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
SD	yr89					
	1989	-1.160197	.1810497	-6.41	0.000	-1.515048 - .8053457
	male					
	Men	1.226454	.167691	7.31	0.000	.8977855 1.555122
	white					
	White	.834226	.2641771	3.16	0.002	.3164484 1.352004
	age	.0316763	.0052183	6.07	0.000	.0214487 .041904
	ed	-.1435798	.0337793	-4.25	0.000	-.209786 -.0773736
	prst	-.0041656	.0070026	-0.59	0.552	-.0178904 .0095592
	_cons	-.7221679	.4928708	-1.47	0.143	-1.688177 .2438411
D	yr89					
	1989	-.4255712	.1318065	-3.23	0.001	-.6839071 -.1672352
	male					
	Men	1.326716	.137554	9.65	0.000	1.057115 1.596317
	white					
	White	.4126344	.1872718	2.20	0.028	.0455885 .7796804
	age	.0292275	.0042574	6.87	0.000	.0208832 .0375718
	ed	-.0513285	.0283399	-1.81	0.070	-.1068737 .0042167
	prst	-.0130318	.0055446	-2.35	0.019	-.023899 -.0021645
	_cons	-.3088357	.3938354	-0.78	0.433	-1.080739 .4630676

```

A
   |
   |   yr89
1989 |   -.0625534   .1228908   -0.51   0.611   -.3034149   .1783082
   |
   |   male
Men   |   .8666833   .1310965   6.61   0.000   .6097389   1.123628
   |
   |   white
White |   .3002409   .1710551   1.76   0.079   -.0350211   .6355028
age   |   .0066719   .0041053   1.63   0.104   -.0013744   .0147181
ed    |   -.0330137   .0274376   -1.20   0.229   -.0867904   .020763
prst  |   -.0017323   .0052199   -0.33   0.740   -.0119631   .0084985
_cons |   .3932277   .3740361   1.05   0.293   -.3398697   1.126325
-----
SA   |   (base outcome)
-----

```

```
. listcoef 1.yr89, help
```

```
mlogit (N=2293): Factor change in the odds of warm
```

```
Variable: 1.yr89 (sd=0.490)
```

```
-----
                |           b           z           P>|z|           e^b           e^bStdX
-----+-----
SD      vs D      |      -0.7346      -4.434      0.000           0.480           0.698
SD      vs A      |      -1.0976      -6.705      0.000           0.334           0.584
SD      vs SA     |      -1.1602      -6.408      0.000           0.313           0.567
D       vs SD     |       0.7346       4.434      0.000           2.085           1.433
D       vs A      |      -0.3630      -3.395      0.001           0.696           0.837
D       vs SA     |      -0.4256      -3.229      0.001           0.653           0.812
A       vs SD     |       1.0976       6.705      0.000           2.997           1.712
A       vs D      |       0.3630       3.395      0.001           1.438           1.195
A       vs SA     |      -0.0626      -0.509      0.611           0.939           0.970
SA     vs SD     |       1.1602       6.408      0.000           3.191           1.765
SA     vs D      |       0.4256       3.229      0.001           1.530           1.232
SA     vs A      |       0.0626       0.509      0.611           1.065           1.031
-----

```

```

b = raw coefficient
z = z-score for test of b=0
P>|z| = p-value for z-test
e^b = exp(b) = factor change in odds for unit increase in X
e^bStdX = exp(b*SD of X) = change in odds for SD increase in X

```

Based on the above, we see that yr89 has little effect on strongly agreeing versus agreeing. In every other contrast though, the difference is significant.

It is possible to get overwhelmed with output, at least if you do this for all variables. The `pvalue` option can limit the output to differences which are significant. Also, the `positive` option only shows the positive differences (if you flip the comparison the coefficient will go negative.)

```
. listcoef , help pvalue(.01) positive
```

```
mlogit (N=2293): Factor change in the odds of warm (P<0.01)
```

```
Variable: 1.yr89 (sd=0.490)
```

		b	z	P> z	e^b	e^bStdX
D	vs SD	0.7346	4.434	0.000	2.085	1.433
A	vs SD	1.0976	6.705	0.000	2.997	1.712
A	vs D	0.3630	3.395	0.001	1.438	1.195
SA	vs SD	1.1602	6.408	0.000	3.191	1.765
SA	vs D	0.4256	3.229	0.001	1.530	1.232

```
Variable: 1.male (sd=0.499)
```

		b	z	P> z	e^b	e^bStdX
SD	vs SA	1.2265	7.314	0.000	3.409	1.844
D	vs A	0.4600	4.403	0.000	1.584	1.258
D	vs SA	1.3267	9.645	0.000	3.769	1.938
A	vs SA	0.8667	6.611	0.000	2.379	1.541

```
Variable: 1.white (sd=0.329)
```

		b	z	P> z	e^b	e^bStdX
SD	vs SA	0.8342	3.158	0.002	2.303	1.316

```
Variable: age (sd=16.779)
```

		b	z	P> z	e^b	e^bStdX
SD	vs A	0.0250	5.578	0.000	1.025	1.521
SD	vs SA	0.0317	6.070	0.000	1.032	1.701
D	vs A	0.0226	6.789	0.000	1.023	1.460
D	vs SA	0.0292	6.865	0.000	1.030	1.633

```
Variable: ed (sd=3.161)
```

		b	z	P> z	e^b	e^bStdX
D	vs SD	0.0923	3.374	0.001	1.097	1.339
A	vs SD	0.1106	3.945	0.000	1.117	1.418
SA	vs SD	0.1436	4.251	0.000	1.154	1.574

```
Variable: prst (sd=14.492)
```

```
    b = raw coefficient
```

```
    z = z-score for test of b=0
```

```
    P>|z| = p-value for z-test
```

```
    e^b = exp(b) = factor change in odds for unit increase in X
```

```
    e^bStdX = exp(b*SD of X) = change in odds for SD increase in X
```

Using the .01 level of significance (which may be wise given the many comparisons that are being done) we see that white only clearly distinguished between those who strongly agree and those who strongly disagree. prst does not have any significant effects.

## Using mlogtest for tests of the Multinomial Logistic Model.

The `mlogtest` command provides a convenient means for testing various hypotheses of interest. Incidentally, keep in mind that `mlogit` can also estimate a logistic regression model; ergo you might sometimes want to use `mlogit` instead of `logit` so you can take advantage of the `mlogtest` command.

**Tests of independent variables.** `mlogtest` can provide likelihood-ratio tests for each variable in the model. To do this yourself, you would have to estimate a series of models, store the results, and then use the `lrtest` command. `mlogtest` can automate this process.

```
. mlogtest, lr
```

```
LR tests for independent variables (N=2293)
```

```
Ho: All coefficients associated with given variable(s) are 0
```

	chi2	df	P>chi2
1.yr89	58.853	3	0.000
1.male	106.199	3	0.000
1.white	11.152	3	0.011
age	83.119	3	0.000
ed	21.087	3	0.000
prst	8.412	3	0.038

From the above, we can see that each variable's effects are significant at the .05 level.

If you happen to have a very large data set or a very complicated model, LR tests can take a long time. It may be sufficient to simply use Wald tests in such cases. Remember, a Wald test only requires the estimation of the constrained model. In Stata, we could just do this with a series of `test` commands. Again, `mlogtest`, using the `wald` parameter, can automate the process and also present results more succinctly:

```
. mlogtest, wald
```

```
Wald tests for independent variables (N=2293)
```

```
Ho: All coefficients associated with given variable(s) are 0
```

	chi2	df	P>chi2
1.yr89	53.812	3	0.000
1.male	97.773	3	0.000
1.white	10.783	3	0.013
age	79.925	3	0.000
ed	20.903	3	0.000
prst	8.369	3	0.039

We see that both tests lead to very similar conclusions in this case. That is fairly common; it seems they are most likely to differ in borderline cases.

You can also use `mlogtest` to test sets of variables, e.g.

```
. mlogtest, lr set(1.white prst \ 1.white ed \ 1.yr89 1.male )
```

LR tests for independent variables (N=2293)

Ho: All coefficients associated with given variable(s) are 0

	chi2	df	P>chi2
1.yr89	58.853	3	0.000
1.male	106.199	3	0.000
1.white	11.152	3	0.011
age	83.119	3	0.000
ed	21.087	3	0.000
prst	8.412	3	0.038
set_1	19.282	6	0.004
set_2	30.334	6	0.000
set_3	167.621	6	0.000

set\_1 contains: 1.white prst

set\_2 contains: 1.white ed

set\_3 contains: 1.yr89 1.male

**Tests for combining dependent categories.** If none of the IVs significantly affects the odds of outcome m versus outcome n, we say that m and n are indistinguishable with respect to the variables in the model. If two outcomes are indistinguishable with respect to the variables in the model, you can obtain more efficient estimates by combining them. I often use this command to see if I can combine categories, even if, say, I am using a command like ologit. Again, you can use both Stata or spost13 commands, and you can do LR or Wald tests.

```
. mlogtest, lrcomb
```

LR tests for combining alternatives (N=2293)

Ho: All coefficients except intercepts associated with a given pair of alternatives are 0 (i.e., alternatives can be collapsed)

	chi2	df	P>chi2
SD & D	43.864	6	0.000
SD & A	153.130	6	0.000
SD & SA	215.033	6	0.000
D & A	98.857	6	0.000
D & SA	191.730	6	0.000
A & SA	54.469	6	0.000

Based on the above, we see that no categories should be combined. Doing the same thing with Wald tests,

```
. mlogtest, combine
```

```
Wald tests for combining alternatives (N=2293)
```

```
Ho: All coefficients except intercepts associated with a given pair  
of alternatives are 0 (i.e., alternatives can be combined)
```

	chi2	df	P>chi2
SD & D	41.018	6	0.000
SD & A	135.960	6	0.000
SD & SA	183.910	6	0.000
D & A	93.183	6	0.000
D & SA	167.439	6	0.000
A & SA	51.441	6	0.000

**Independence of Irrelevant Alternatives (IIA) Tests.** The Stata 12 Reference Manual (P. 710) explains the IIA assumption this way:

A stringent assumption of multinomial and conditional logit models is that outcome categories for the model have the property of independence of irrelevant alternatives (IIA). Stated simply, this assumption requires that the inclusion or exclusion of categories does not affect the relative risks associated with the regressors in the remaining categories. One classic example of a situation in which this assumption would be violated involves the choice of transportation mode; see McFadden (1974). For simplicity, postulate a transportation model with the four possible outcomes: rides a train to work, takes a bus to work, drives the Ford to work, and drives the Chevrolet to work. Clearly, “drives the Ford” is a closer substitute to “drives the Chevrolet” than it is to “rides a train” (at least for most people). This means that excluding “drives the Ford” from the model could be expected to affect the relative risks of the remaining options and that the model would not obey the IIA assumption.

The 3<sup>rd</sup> edition of Long & Freese (section 8.4, pp. 407-411) explains the assumption further, and also explains ways of testing it. Long & Freese include tests for IIA in their programs but do NOT encourage their use. They note that these tests often provide conflicting results (e.g. some tests reject the null while others do not) and that various simulation studies have shown that these tests are not useful for assessing violations of the IIA assumption. They further argue that the multinomial logit model works best when the alternatives are dissimilar and not just substitutes for one another (e.g. if your choices were take your car to work, take a blue bus, or take a red bus, the two bus alternatives would be very similar and the IIA assumption would likely be violated, whether the tests showed it or not).

Paul Allison has also raised concerns about the IIA tests; see his blog entry at <http://www.statisticalhorizons.com/iaa>.

But, if some reviewer says you need to test the IIA assumption, here is how you can do it with `mlogtest`.

```
. mlogtest, iia
```

Hausman tests of IIA assumption (N=2293)

Ho: Odds(Outcome-J vs Outcome-K) are independent of other alternatives

	chi2	df	P>chi2
SD	-0.177	14	.
D	-10.884	14	.
A	-3.009	13	.
SA	-1.606	14	.

Note: A significant test is evidence against Ho.

Note: If  $\chi^2 < 0$ , the estimated model does not meet asymptotic assumptions.

suest-based Hausman tests of IIA assumption (N=2293)

Ho: Odds(Outcome-J vs Outcome-K) are independent of other alternatives

	chi2	df	P>chi2
SD	18.651	14	0.179
D	20.289	14	0.121
A	23.480	14	0.053
SA	11.381	14	0.656

Note: A significant test is evidence against Ho.

Small-Hsiao tests of IIA assumption (N=2293)

Ho: Odds(Outcome-J vs Outcome-K) are independent of other alternatives

	lnL(full)	lnL(omit)	chi2	df	P>chi2
SD	-1025.061	-1018.448	13.226	14	0.509
D	-718.007	-711.796	12.422	14	0.572
A	-678.789	-673.072	11.433	14	0.652
SA	-936.474	-928.840	15.268	14	0.360

Note: A significant test is evidence against Ho.

In this example the tests say IIA has not been violated. Long & Freese give examples of where different tests reach different conclusions with the same set of data.

Measures of Fit. The `fitstat` command can be used the same as before, e.g.

```
. quietly mlogit warm i.yr89 i.male i.white age ed prst, b(4) nolog
. quietly fitstat, save
. * Now drop prst, white & ed, the three least significant vars
. quietly mlogit warm i.yr89 i.male age , b(4) nolog
. fitstat, dif
```

	Current	Saved	Difference
-----			
Log-likelihood			
Model	-2848.592	-2820.998	-27.594
Intercept-only	-2995.770	-2995.770	0.000
-----			
Chi-square			
D (df=2281/2272/9)	5697.184	5641.996	55.188
LR (df=9/18/-9)	294.357	349.544	-55.188
p-value	0.000	0.000	0.000
-----			
R2			
McFadden	0.049	0.058	-0.009
McFadden (adjusted)	0.045	0.051	-0.006
Cox-Snell/ML	0.120	0.141	-0.021
Cragg-Uhler/Nagelkerke	0.130	0.153	-0.023
Count	0.412	0.424	-0.013
Count (adjusted)	0.061	0.081	-0.020
-----			
IC			
AIC	5721.184	5683.996	37.188
AIC divided by N	2.495	2.479	0.016
BIC (df=12/21/-9)	5790.035	5804.486	-14.451

Note: Likelihood-ratio test assumes current model nested in saved model.

Difference of 14.451 in BIC provides very strong support for current model.

Incidentally, note that the chi-square and AIC tests favor the full model; however, the BIC test prefers the model that drops the least significant variables, `prst`, `white` & `ed`. As we have seen before, the BIC test tends to lead to more parsimonious models, especially when the sample size is large.

**Outliers.** The `leastlikely` command can be used to identify the cases where the observed value was farthest from the predicted value. You might want to check such cases for coding errors or think if there are ways to modify the model so these cases are not so discrepant.

```
. quietly mlogit warm i.yr89 i.male i.white age ed prst, b(4) nolog
. leastlikely warm yr89 male white age ed prst
```

Outcome: 1 (SD)

	Prob	warm	yr89	male	white	age	ed	prst
112.	.0389264	SD	1989	Women	White	46	16	57
167.	.0355258	SD	1989	Women	White	37	15	61
212.	.0423206	SD	1989	Women	White	50	16	62
271.	.0352297	SD	1989	Women	White	20	12	31
286.	.0407416	SD	1989	Women	NotWhite	54	12	34

Outcome: 2 (D)

	Prob	warm	yr89	male	white	age	ed	prst
414.	.1286143	D	1989	Women	White	19	12	50
563.	.1175782	D	1989	Women	NotWhite	41	18	69
675.	.1322747	D	1989	Women	White	25	16	50
803.	.107113	D	1989	Women	NotWhite	30	16	60
1001.	.1288399	D	1989	Women	White	32	18	62

Outcome: 3 (A)

	Prob	warm	yr89	male	white	age	ed	prst
1305.	.1621244	A	1977	Men	White	79	8	41
1344.	.1575535	A	1977	Men	White	72	7	22
1404.	.1625481	A	1977	Men	White	74	8	26
1449.	.1398363	A	1977	Men	White	71	4	23
1729.	.1303623	A	1977	Men	White	81	5	36

Outcome: 4 (SA)

	Prob	warm	yr89	male	white	age	ed	prst
1963.	.0313339	SA	1977	Men	White	64	6	26
2093.	.0372785	SA	1977	Men	White	48	4	17
2107.	.034017	SA	1977	Men	White	69	8	33
2119.	.0345335	SA	1977	Men	White	58	4	41
2138.	.0316978	SA	1977	Men	White	57	3	37

**Aids to Interpretation.** These are much the same as we talked about before. Standardized coefficients, however, are a noteworthy exception:

```
. listcoef, std
option std not allowed after mlogit
```

This is because the  $y^*$  rationale does not hold in a multinomial logit model, i.e. there is no underlying latent variable. (As we saw earlier, however, the `listcoef` command will still do X-standardization.)

Other commands, however, behave identically or almost identically to what we have seen before. For example, we can use the `predict` command to come up with predicted probabilities:

```
. quietly mlogit warm i.yr89 i.male i.white age ed prst, b(4) nolog
. predict SDlogit Dlogit Alogit SAllogit
(option pr assumed; predicted probabilities)
```

```
. list warm yr89 male white age ed prst SDlogit Dlogit Alogit SAllogit in 1/10, clean
```

	warm	yr89	male	white	age	ed	prst	SDlogit	Dlogit	Alogit	SAllogit
1.	SD	1977	Women	White	33	10	31	.14696	.2569168	.375222	.2209013
2.	SD	1977	Men	White	74	16	50	.1931719	.4962518	.2510405	.0595358
3.	SD	1989	Men	White	36	12	41	.074012	.3257731	.4686748	.1315401
4.	SD	1977	Women	White	73	9	36	.277139	.383207	.2358743	.1037797
5.	SD	1977	Women	White	59	11	62	.2066857	.2824558	.3317693	.1790893
6.	SD	1989	Men	White	33	4	17	.1461631	.383301	.3885765	.0819594
7.	SD	1977	Women	White	43	7	40	.2276894	.2719278	.3321202	.1682626
8.	SD	1977	Women	White	48	12	48	.1571982	.2740046	.358632	.2101651
9.	SD	1977	Men	White	27	17	69	.0970773	.259278	.477971	.1656736
10.	SD	1977	Men	White	46	12	50	.1997817	.3800453	.3360028	.0841702

The extremes (use `findit extremes`) command helps you to see who is most likely and least likely to be predicted to strongly disagree:

```
. extremes SDlogit warm yr89 male white age ed prst
```

```
+-----+
| obs:   SDlogit  warm  yr89  male    white  age  ed  prst |
+-----+
| 1214.  .0078837   A   1989  Women  NotWhite  27  20  68 |
| 2048.  .011555   SA  1989  Women  NotWhite  26  17  52 |
| 2241.  .0127511   SA  1989  Women  NotWhite  21  15  61 |
| 1855.  .0131329   A   1989  Women  NotWhite  25  16  36 |
| 803.   .0142298   D   1989  Women  NotWhite  30  16  60 |
+-----+

+-----+
| 612.   .4276913   D   1977  Men    White   80  5   45 |
| 171.   .4289597   SD  1977  Men    White   67  3   32 |
| 282.   .4378463   SD  1977  Men    White   68  3   37 |
| 87.    .4426529   SD  1977  Men    White   83  5   51 |
| 863.   .479314   D   1977  Men    White   54  0   40 |
+-----+
```

Based on the results, we see that fairly young white women in 1989 with high levels of education and occupational prestige were predicted to be the least likely to strongly disagree. Conversely, nonwhite elderly males in 1977 with low levels of education and generally low levels of occupational prestige had almost a 50% predicted probability of strongly disagreeing.

Other comments. See Long and Freese for detailed explanations of how different commands are working, e.g. they often show you how the same things could be done in Stata without their commands (albeit in a much more tedious process). They also offer detailed advice on graphing results.