Multinomial Logit Models - Overview

This is adapted heavily from Menard's <u>Applied Logistic Regression analysis</u>; also, Borooah's <u>Logit and Probit: Ordered and Multinomial Models</u>; Also, Hamilton's <u>Statistics with Stata, Updated for Version 7</u>.

When categories are unordered, Multinomial Logistic regression is one often-used strategy. Mlogit models are a straightforward extension of logistic models.

Suppose a DV has M categories. One value (typically the first or last) of the DV is designated as the reference category. The probability of membership in other categories is compared to the probability of membership in the reference category.

For a DV with M categories, this requires the calculation of M-1 equations, one for each category relative to the reference category, to describe the relationship between the DV and the IVs.

Hence, if the first category is the reference (which is the default for Stata's `mlogit`; SPSS's NOMREG by default uses the last category for the reference but you can change it with the Base=First parameter); then, for m = 2, ..., M,

$$\ln\frac{P(Yi = m)}{P(Yi = 1)} = \alpha_m + \sum_{k=1}^{K} \beta_{mk} X_{ik} = Z_{mi}$$

Hence, for each case, there will be M-1 predicted log odds, one for each category relative to the reference category. (Note that when m = 1 you get $\ln(1) = 0 = Z_{11}$, and $\exp(0) = 1$.)

When there are more than 2 groups, computing probabilities is a little more complicated than it was in logistic regression. For m = 2, ..., M,

$$P(Y_i = m) = \frac{\exp(Z_{mi})}{1 + \sum_{h=2}^{M} \exp(Z_{hi})}$$

For the reference category,

$$P(Y_i = 1) = \frac{1}{1 + \sum_{h=2}^{M} \exp(Z_{hi})}$$

In other words, you take each of the M-1 log odds you computed and exponentiate it. Once you have done that the calculation of the probabilities is straightforward.

Note that, when M = 2, the mlogit and logistic regression models (and for that matter the ordered logit model) become one and the same.

We'll redo our Challenger example, this time using SPSS's NOMREG routine.

```
NOMREG
  distress  (base = first) WITH temp date
  /CRITERIA = CIN(95) DELTA(0) MXITER(100) MXSTEP(5) CHKSEP(20) LCONVERGE(0)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /MODEL
  /INTERCEPT = INCLUDE
  /PRINT = PARAMETER SUMMARY LRT
  /Save = ESTPROB (MLog) .
```

Nominal Regression

Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	49.911			
Final	35.767	14.143	4	.007

Pseudo R-Square

Cox and Snell	.459
Nagelkerke	.519
McFadden	.283

Likelihood Ratio Tests

Effect	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	40.714	4.946	2	.084
TEMP	42.739	6.972	2	.031
DATE	47.243	11.475	2	.003

The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.

Parameter Estimates

DISTRESS thermal distress incidents[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B) Lower Bound	Upper Bound
2 1 or 2	Intercept	-8.4059	10.471	.644	1	.422			
	TEMP	-.10541	.134	.616	1	.433	.900	.692	1.171
	DATE	.001769	.001	1.502	1	.220	1.002	.999	1.005
3 3 plus	Intercept	-40.433	25.179	2.579	1	.108			
	TEMP	-.29647	.157	3.573	1	.059	.743	.547	1.011
	DATE	.006775	.003	3.987	1	.046	1.007	1.000	1.014

a. The reference category is: 1 none.

For group 2 (one or two distress incidents), the coefficients tell us that lower temperatures and higher dates increase the likelihood that you will have one or two distress incidents as opposed to none. We see the same thing in group 3, but the effects are even larger.

From the above, we see that

$$Z_{2i} = -8.4059 - .10541*Temp + .001769*Date$$

$$Z_{3i} = -40.433 - .29647*Temp + .006775*Date.$$

Hence, for flight 13, where Temp = 78 and Date = 9044, we get

$$Z_2 = -8.4059 - .10541*78 + .001769*9044 = -.629$$

$$Z_3 = -40.433 - .29647*78 + .006775*9044 = -2.2846$$

In each case, the negative numbers tell us flight 13 was more likely to fall in the reference category. From these numbers, we can compute that, for Flight 13,

$$P(Y_i = 1) = \frac{1}{1 + \sum_{h=2}^{M} \exp(Z_{hi})} = \frac{1}{1 + \exp(-.629) + \exp(-2.2846)} = .6116$$

$$P(Y_i = 2) = \frac{\exp(Z_{1i})}{1 + \sum_{h=2}^{M} \exp(Z_{hi})} = \frac{\exp(-.629)}{1 + \exp(-.629) + \exp(-2.2846)} = .326$$

$$P(Y_i = 3) = \frac{\exp(Z_{2i})}{1 + \sum_{h=2}^{M} \exp(Z_{hi})} = \frac{\exp(-2.2846)}{1 + \exp(-.629) + \exp(-2.2846)} = .0623$$

These numbers are similar to what we got with the ordinal regression. If we do similar calculations for Challenger, we get P(Y = 1) = .0005367, P(Y = 2) = .0000593, P(Y = 3) = .9999404.

So, in this case, both the multinomial and ordinal regression approaches produce virtually identical results, but the ordinal regression model is somewhat simpler and requires the estimation of fewer parameters. Note too that in the Ordered Logit model the effects of both Date and Time were statistically significant, but this was not true for all the groups in the Mlogit analysis; this probably reflects the greater efficiency of the Ordered Logit approach. Particularly in a model with more X variables and/or categories of Y, the ordinal regression approach would be simpler and hence preferable, provided its assumptions are met.

In short, the models get more complicated when you have more than 2 categories, and you get a lot more parameter estimates, but the logic is a straightforward extension of logistic regression.

Because we included the parameter /Save = ESTPROB (MLog), we can also get the estimated probabilities for each case of falling into each of the three groups (again with the exception of the case we really want, case 25).

```
Formats mlog1_1 mlog2_1 mlog3_1 (f8.4).
List flight temp date distress mlog1_1 mlog2_1 mlog3_1 .
```

List

FLIGHT	TEMP	DATE	DISTRESS	MLOG1_1	MLOG2_1	MLOG3_1
1	66	7772	1	.8340	.1654	.0005
2	70	7986	2	.8398	.1595	.0007
3	69	8116	1	.7884	.2094	.0022
4	80	8213	.	.	.	.
5	68	8350	1	.6829	.3049	.0123
6	67	8494	2	.5868	.3756	.0376
7	72	8569	1	.6870	.2964	.0166
8	73	8642	1	.6797	.3003	.0200
9	70	8732	1	.5427	.3856	.0717
10	57	8799	2	.0716	.2256	.7028
11	63	8862	3	.1849	.3458	.4693
12	70	9008	3	.3317	.3841	.2842
13	**78**	**9044**	**1**	**.6124**	**.3251**	**.0625**
14	67	9078	1	.1627	.2924	.5449
15	53	9155	3	.0027	.0245	.9728
16	67	9233	3	.0773	.1827	.7400
17	75	9250	3	.3277	.3437	.3286
18	70	9299	3	.1100	.2131	.6769
19	81	9341	2	.5081	.3325	.1594
20	76	9370	2	.2599	.3033	.4368
21	79	9407	1	.3577	.3248	.3174
22	75	9434	3	.1684	.2444	.5872
23	76	9461	2	.1824	.2499	.5677
24	58	9508	3	.0011	.0110	.9879
25	**31**	**9524**	**.**	**.**	**.**	**.**

Number of cases read: 25   Number of cases listed: 25

Stata example. In Stata, we use the `mlogit` command. In Stata, the most frequent category is the default reference group:

```
. mlogit  distress date temp

Iteration 0:   log likelihood = -24.955257
Iteration 1:   log likelihood = -19.232647
Iteration 2:   log likelihood = -18.163998
Iteration 3:   log likelihood = -17.912395
Iteration 4:   log likelihood = -17.884218
Iteration 5:   log likelihood = -17.883654
Iteration 6:   log likelihood = -17.883653
```

```
Multinomial logistic regression                    Number of obs   =        23
                                                   LR chi2(4)      =     14.14
                                                   Prob > chi2     =    0.0069
Log likelihood = -17.883653                        Pseudo R2       =    0.2834


------------------------------------------------------------------------------
    distress |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
1 or 2       |
        date |   .0017686   .0014431     1.23   0.220    -.0010599     .004597
        temp |  -.1054113   .1343361    -0.78   0.433    -.3687052    .1578826
       _cons |  -8.405851   10.47099    -0.80   0.422    -28.92862    12.11692
-------------+----------------------------------------------------------------
3 plus       |
        date |   .0067752   .0033931     2.00   0.046     .0001248    .0134256
        temp |  -.2964675   .1568354    -1.89   0.059    -.6038594    .0109243
       _cons |  -40.43276   25.17892    -1.61   0.108    -89.78254    8.917024
------------------------------------------------------------------------------
(Outcome distress==none is the comparison group)
```

To have Stata compute the Z values and the predicted probabilities of being in each group:

`. predict z2, xb outcome(2)`

`. predict z3, xb outcome(3)`

`. * You could predict z1 – but it would be 0 for every case!`

`. predict  mnone monetwo mthreeplus, p`

`. list  flight temp date distress z2 z3 mnone monetwo mthreeplus`

```
     +--------------------------------------------------------------------------------+
     | flight   temp   date   distress         z2         z3      mnone   monetwo  mthree~s |
     |--------------------------------------------------------------------------------|
  1. |  STS-1     66   7772       none    -1.6178   -7.342882  .8340411  .1654192  .0005398 |
  2. |  STS-2     70   7986     1 or 2  -1.660975   -7.078863  .8397741  .1595182  .0007077 |
  3. |  STS-3     69   8116       none  -1.325651   -5.901621  .7884166   .209427  .0021563 |
  4. |  STS-4     80   8213          .  -2.313626   -8.505571  .9098317  .0899842  .0001841 |
  5. |  STS-5     68   8350       none  -.8063986   -4.019761  .6828641  .3048736  .0122624 |
     |--------------------------------------------------------------------------------|
  6. |  STS-6     67   8494     1 or 2  -.4463157   -2.747666  .5868342  .3755631  .0376027 |
  7. |  STS-7     72   8569       none  -.8407306   -3.721865  .6870095  .2963726  .0166179 |
  8. |  STS-8     73   8642       none  -.8170375   -3.523744  .6797047  .3002516  .0200437 |
  9. |  STS-9     70   8732       none  -.3416339   -2.024575  .5426942   .385643  .0716627 |
 10. | STS_41-B   57   8799     1 or 2   1.147206     2.28344  .0716345  .2256043  .7027612 |
     |--------------------------------------------------------------------------------|
 11. | STS_41-C   63   8862     3 plus   .6261569    .9314718   .184889   .345818   .469293 |
 12. | STS_41-D   70   9008     3 plus   .1464868    -.154624  .3317303   .384064  .2842057 |
 13. | STS_41-G   78   9044       none  -.6331355   -2.282458  .6123857  .3251306  .0624836 |
 14. | STS_51-A   67   9078       none   .5865193    1.209041  .1626547  .2924077  .5449376 |
 15. | STS_51-C   53   9155     3 plus   2.198456    5.881276  .0027153  .0244682  .9728165 |
     |--------------------------------------------------------------------------------|
 16. | STS_51-D   67   9233     3 plus   .8606451    2.259195  .0772794  .1827414  .7399792 |
 17. | STS_51-B   75   9250     3 plus   .0474203    .0026329    .32774  .3436559  .3286041 |
 18. | STS_51-G   70   9299     3 plus   .6611357    1.816955    .11001  .2130884  .6769016 |
 19. | STS_51-F   81   9341     1 or 2   -.424109   -1.159631  .5081418  .3325039  .1593543 |
 20. | STS_51-I   76   9370     1 or 2   .1542354    .5191875   .259914  .3032586  .4368274 |
     |--------------------------------------------------------------------------------|
 21. | STS_51-J   79   9407       none   -.096562   -.1195333  .3577449  .3248158  .3174394 |
 22. | STS_61-A   75   9434     3 plus   .3728341    1.249267  .1683607  .2444334  .5872059 |
 23. | STS_61-B   76   9461     1 or 2   .3151737    1.135729  .1823506   .249911  .5677384 |
 24. | STS_61-C   58   9508     3 plus   2.295699    6.790579  .0011107  .0110305  .9878589 |
 25. | STS_51-L   31   9524          .     5.1701    14.90361   3.37e-07  .0000593  .9999404 |
     +--------------------------------------------------------------------------------+
```

Closing Comments. A few other things you may want to consider:

- You may want to combine some categories of the DV, partly to make the analysis simpler, and partly because the number of cases in some categories may be very small. Remember, the more categories you have, the more parameters you will estimate, and the more difficult it may be to get significant results. It is simplest, of course, to only have two categories, but you'll have to decide whether or not that is justified for your particular problem.

- Make sure you understand what the reference category is, since different programs do it differently. You may need to recode the variable if there is no other way of changing the reference category.

- If the DV is ordinal, other techniques may be appropriate and more parsimonious.