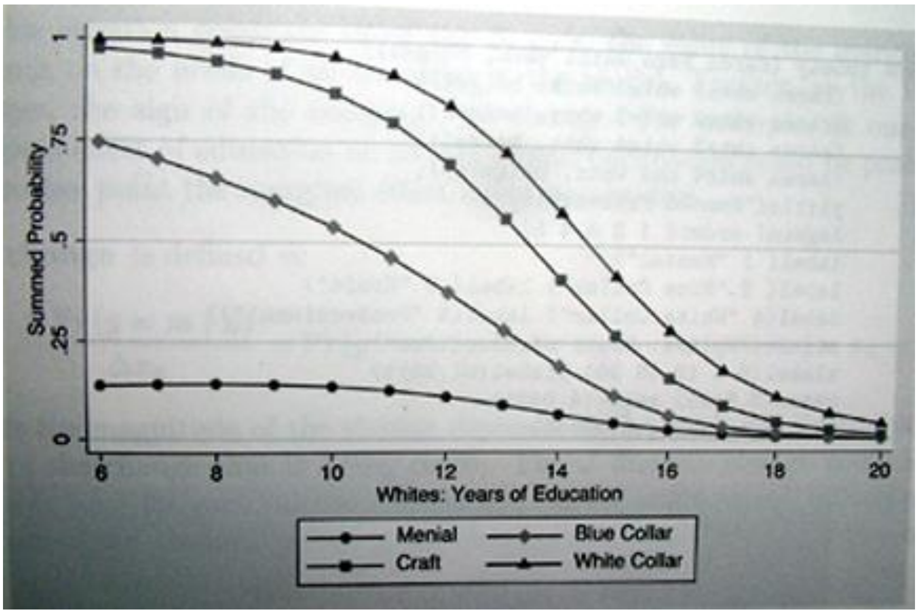
Sociology 7704: Regression Models for Categorical Data
Instructor: Natasha Sarkisian

Multinomial logit

We use multinomial logit models when we have multiple categories but cannot order them (or we can, but the parallel regression assumption does not hold). Here the order of categories is unimportant. Multinomial logit model is equivalent to simultaneous estimation of multiple logits where each of the categories is compared to one selected so-called base category. But if we would estimate them separately, we would lose information, as each logit would be estimated on a different sample (selected category plus base category, with all other categories omitted from analyses). To avoid that, we use multinomial logit.

Multinomial logit does not assume parallel slopes – so if we estimate it for ordinal level variable and then plot cumulative probabilities, we would see something like this (note the variation in slope!):



Let's estimate a multinomial logit model for the same variable we used above:

```
. mlogit natarmsy age sex childs educ born
Iteration 0:   log likelihood = -1410.9409
Iteration 1:   log likelihood = -1388.2174
Iteration 2:   log likelihood = -1387.8455
Iteration 3:   log likelihood = -1387.8455
Multinomial logistic regression                   Number of obs   =       1337
                                                  LR chi2(10)     =      46.19
                                                  Prob > chi2     =     0.0000
Log likelihood = -1387.8455                       Pseudo R2       =     0.0164
------------------------------------------------------------------------------
    natarmsy |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
too_little   |
         age |     .00548   .0039204     1.40   0.162    -.0022039    .0131639
         sex |  -.1919797   .1251455    -1.53   0.125    -.4372605     .053301
      childs |  -.0194531   .0411446    -0.47   0.636     -.100095    .0611887
        educ |  -.0102552   .0210369    -0.49   0.626    -.0514869    .0309764
        born |  -.8933254   .2685341    -3.33   0.001    -1.419643   -.3670082
       _cons |   .9484192   .4877278     1.94   0.052    -.0075097    1.904348
-------------+----------------------------------------------------------------
about_right  |  (base outcome)
```

```
-------------+----------------------------------------------------------------
too_much     |
         age |  -.0135326   .0049789    -2.72   0.007     -.023291   -.0037742
         sex |   .0420268   .1485803     0.28   0.777    -.2491853    .3332389
      childs |  -.0128663   .0519464    -0.25   0.804    -.1146793    .0889468
        educ |   .0475599   .0257811     1.84   0.065    -.0029701      .09809
        born |   .1980988   .2326137     0.85   0.394    -.2578157    .6540132
       _cons |  -1.054006   .5377872    -1.96   0.050     -2.10805    .0000374
------------------------------------------------------------------------------
```

Model Interpretation

1. Coefficients and Odds Ratios

Note that we now have two sets of coefficients to interpret. So here, we can see that variable born differentiates between categories "too little" and "about right" while variable age differentiates between "too much" and "about right."

Also note that it automatically omitted the category "about right" -- it usually omits the category with the largest number of observations unless you specify otherwise. Here's how we change that:

```
. mlogit natarmsy age sex childs educ born, b(1)
Iteration 0:   log likelihood = -1410.9409
Iteration 1:   log likelihood = -1388.2174
Iteration 2:   log likelihood = -1387.8455
Iteration 3:   log likelihood = -1387.8455
Multinomial logistic regression                   Number of obs   =       1337
                                                  LR chi2(10)     =      46.19
                                                  Prob > chi2     =     0.0000
Log likelihood = -1387.8455                       Pseudo R2       =     0.0164
------------------------------------------------------------------------------
    natarmsy |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
too_little   |  (base outcome)
-------------+----------------------------------------------------------------
about_right  |
         age |    -.00548   .0039204    -1.40   0.162    -.0131639    .0022039
         sex |   .1919797   .1251455     1.53   0.125     -.053301    .4372605
      childs |   .0194531   .0411446     0.47   0.636    -.0611887     .100095
        educ |   .0102552   .0210369     0.49   0.626    -.0309764    .0514869
        born |   .8933254   .2685341     3.33   0.001     .3670082    1.419643
       _cons |  -.9484192   .4877278    -1.94   0.052    -1.904348    .0075097
-------------+----------------------------------------------------------------
too_much     |
         age |  -.0190126   .0051423    -3.70   0.000    -.0290914   -.0089338
         sex |   .2340066   .1550509     1.51   0.131    -.0698876    .5379007
      childs |   .0065869   .0537937     0.12   0.903    -.0988468    .1120205
        educ |   .0578152   .0270313     2.14   0.032     .0048347    .1107956
        born |   1.091424   .2962107     3.68   0.000     .5108619    1.671987
       _cons |  -2.002425   .5858736    -3.42   0.001    -3.150716   -.8541341
------------------------------------------------------------------------------
```

This allows us to see that variables age, educ and born differentiate between categories too much and too little. Variables sex and childs appear not to be able to differentiate between any categories.

Interpretation of results is again very similar. Since we cannot interpret sizes of regular coefficients, let's examine odds ratios. To obtain odds ratios in multinomial logit models, we use option rrr rather than or.

```
. mlogit natarmsy age sex childs educ born, rrr
Iteration 0:   log likelihood = -1410.9409
Iteration 1:   log likelihood = -1388.2174
Iteration 2:   log likelihood = -1387.8455
Iteration 3:   log likelihood = -1387.8455
Multinomial logistic regression                   Number of obs   =       1337
```

```
                                                  LR chi2(10)     =      46.19
                                                  Prob > chi2     =     0.0000
Log likelihood = -1387.8455                       Pseudo R2       =     0.0164
------------------------------------------------------------------------------
    natarmsy |        RRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
too_little   |
         age |   1.005495    .003942     1.40   0.162     .9977985    1.013251
         sex |   .8253236   .1032856    -1.53   0.125     .6458032    1.054747
      childs |   .9807349   .0403519    -0.47   0.636     .9047515      1.0631
        educ |   .9897972   .0208223    -0.49   0.626     .9498161    1.031461
        born |   .4092924    .109909    -3.33   0.001     .2418004     .692804
       _cons |   2.581625    1.25913     1.94   0.052     .9925184    6.715028
-------------+----------------------------------------------------------------
about_right  |  (base outcome)
-------------+----------------------------------------------------------------
too_much     |
         age |   .9865586   .0049119    -2.72   0.007     .9769782    .9962329
         sex |   1.042922   .1549578     0.28   0.777     .7794356    1.395481
      childs |   .9872161   .0512823    -0.25   0.804      .891652    1.093022
        educ |   1.048709   .0270369     1.84   0.065     .9970343    1.103062
        born |   1.219083   .2835753     0.85   0.394     .7727376    1.923244
       _cons |   .3485387   .1874396    -1.96   0.050     .1214747    1.000037
------------------------------------------------------------------------------
 (Outcome natarmsy==about right is the comparison group)
```

Here we can, for example, say that being foreign born decreases one's odds of saying that the U.S. spends too little versus that the U.S. spends "about right" on national defense by approximately 60%.

We can also use listcoef which generates odds ratios for all possible models group comparisons -- one table per variable:

```
. listcoef
mlogit (N=1337): Factor change in the odds of natarmsy
Variable: age (sd=17.396)
--------------------------------------------------------------------------------
                              |        b         z     P>|z|       e^b   e^bStdX
------------------------------+-------------------------------------------------
too little   vs about right   |     0.0055    1.398    0.162     1.005     1.100
too little   vs too much      |     0.0190    3.697    0.000     1.019     1.392
about right  vs too little    |    -0.0055   -1.398    0.162     0.995     0.909
about right  vs too much      |     0.0135    2.718    0.007     1.014     1.265
too much     vs too little    |    -0.0190   -3.697    0.000     0.981     0.718
too much     vs about right   |    -0.0135   -2.718    0.007     0.987     0.790
--------------------------------------------------------------------------------
Variable: sex (sd=0.498)
--------------------------------------------------------------------------------
                              |        b         z     P>|z|       e^b   e^bStdX
------------------------------+-------------------------------------------------
too little   vs about right   |    -0.1920   -1.534    0.125     0.825     0.909
too little   vs too much      |    -0.2340   -1.509    0.131     0.791     0.890
about right  vs too little    |     0.1920    1.534    0.125     1.212     1.100
about right  vs too much      |    -0.0420   -0.283    0.777     0.959     0.979
too much     vs too little    |     0.2340    1.509    0.131     1.264     1.124
too much     vs about right   |     0.0420    0.283    0.777     1.043     1.021
--------------------------------------------------------------------------------
Variable: childs (sd=1.698)
--------------------------------------------------------------------------------
                              |        b         z     P>|z|       e^b   e^bStdX
------------------------------+-------------------------------------------------
too little   vs about right   |    -0.0195   -0.473    0.636     0.981     0.968
too little   vs too much      |    -0.0066   -0.122    0.903     0.993     0.989
about right  vs too little    |     0.0195    0.473    0.636     1.020     1.034
about right  vs too much      |     0.0129    0.248    0.804     1.013     1.022
too much     vs too little    |     0.0066    0.122    0.903     1.007     1.011
too much     vs about right   |    -0.0129   -0.248    0.804     0.987     0.978
--------------------------------------------------------------------------------
Variable: educ (sd=3.042)
--------------------------------------------------------------------------------
```

```
                             |          b         z     P>|z|       e^b   e^bStdX
-----------------------------+---------------------------------------------------
too little   vs about right  |    -0.0103    -0.487     0.626     0.990     0.969
too little   vs too much     |    -0.0578    -2.139     0.032     0.944     0.839
about right  vs too little   |     0.0103     0.487     0.626     1.010     1.032
about right  vs too much     |    -0.0476    -1.845     0.065     0.954     0.865
too much     vs too little   |     0.0578     2.139     0.032     1.060     1.192
too much     vs about right  |     0.0476     1.845     0.065     1.049     1.156
---------------------------------------------------------------------------------
Variable: born (sd=0.276)
---------------------------------------------------------------------------------
                             |          b         z     P>|z|       e^b   e^bStdX
-----------------------------+---------------------------------------------------
too little   vs about right  |    -0.8933    -3.327     0.001     0.409     0.781
too little   vs too much     |    -1.0914    -3.685     0.000     0.336     0.740
about right  vs too little   |     0.8933     3.327     0.001     2.443     1.280
about right  vs too much     |    -0.1981    -0.852     0.394     0.820     0.947
too much     vs too little   |     1.0914     3.685     0.000     2.979     1.352
too much     vs about right  |     0.1981     0.852     0.394     1.219     1.056
---------------------------------------------------------------------------------
```

We can also use all the same options with listcoef that we used with binary logit, and some additional options that help restrict which comparisons are shown: positive, negative, adjacent, gt (greater than), lt (less than). For example:

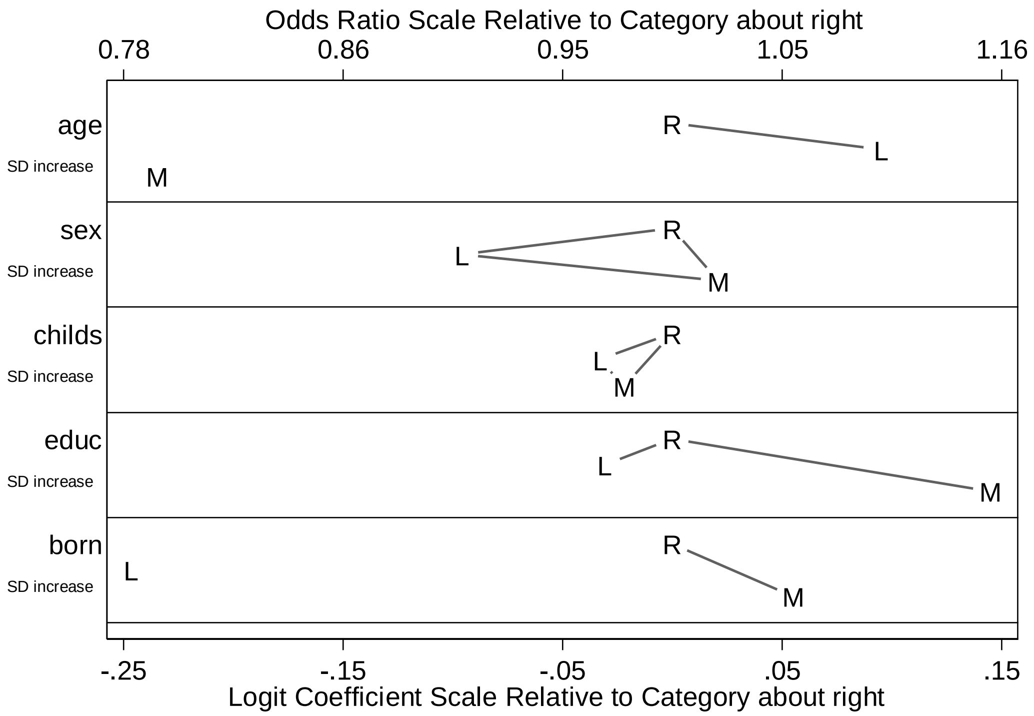
```
. listcoef, positive
mlogit (N=1337): Factor change in the odds of natarmsy
Variable: age (sd=17.396)
---------------------------------------------------------------------------------
                             |          b         z     P>|z|       e^b   e^bStdX
-----------------------------+---------------------------------------------------
too little   vs about right  |     0.0055     1.398     0.162     1.005     1.100
too little   vs too much     |     0.0190     3.697     0.000     1.019     1.392
about right  vs too much     |     0.0135     2.718     0.007     1.014     1.265
---------------------------------------------------------------------------------
Variable: sex (sd=0.498)
---------------------------------------------------------------------------------
                             |          b         z     P>|z|       e^b   e^bStdX
-----------------------------+---------------------------------------------------
about right  vs too little   |     0.1920     1.534     0.125     1.212     1.100
too much     vs too little   |     0.2340     1.509     0.131     1.264     1.124
too much     vs about right  |     0.0420     0.283     0.777     1.043     1.021
---------------------------------------------------------------------------------
Variable: childs (sd=1.698)
---------------------------------------------------------------------------------
                             |          b         z     P>|z|       e^b   e^bStdX
-----------------------------+---------------------------------------------------
about right  vs too little   |     0.0195     0.473     0.636     1.020     1.034
about right  vs too much     |     0.0129     0.248     0.804     1.013     1.022
too much     vs too little   |     0.0066     0.122     0.903     1.007     1.011
---------------------------------------------------------------------------------
Variable: educ (sd=3.042)
---------------------------------------------------------------------------------
                             |          b         z     P>|z|       e^b   e^bStdX
-----------------------------+---------------------------------------------------
about right  vs too little   |     0.0103     0.487     0.626     1.010     1.032
too much     vs too little   |     0.0578     2.139     0.032     1.060     1.192
too much     vs about right  |     0.0476     1.845     0.065     1.049     1.156
---------------------------------------------------------------------------------
Variable: born (sd=0.276)
---------------------------------------------------------------------------------
                             |          b         z     P>|z|       e^b   e^bStdX
-----------------------------+---------------------------------------------------
about right  vs too little   |     0.8933     3.327     0.001     2.443     1.280
too much     vs too little   |     1.0914     3.685     0.000     2.979     1.352
too much     vs about right  |     0.1981     0.852     0.394     1.219     1.056
---------------------------------------------------------------------------------
```

We can also filter by p-value:

```
. listcoef, pvalue(.05)
mlogit (N=1337): Factor change in the odds of natarmsy (P<0.05)
Variable: age (sd=17.396)
--------------------------------------------------------------------------------
                             |          b         z     P>|z|       e^b   e^bStdX
-----------------------------+--------------------------------------------------
too little   vs too much     |     0.0190     3.697     0.000     1.019     1.392
about right  vs too much     |     0.0135     2.718     0.007     1.014     1.265
too much     vs too little   |    -0.0190    -3.697     0.000     0.981     0.718
too much     vs about right  |    -0.0135    -2.718     0.007     0.987     0.790
--------------------------------------------------------------------------------

Variable: sex (sd=0.498)

Variable: childs (sd=1.698)

Variable: educ (sd=3.042)
--------------------------------------------------------------------------------
                             |          b         z     P>|z|       e^b   e^bStdX
-----------------------------+--------------------------------------------------
too little   vs too much     |    -0.0578    -2.139     0.032     0.944     0.839
too much     vs too little   |     0.0578     2.139     0.032     1.060     1.192
--------------------------------------------------------------------------------

Variable: born (sd=0.276)
--------------------------------------------------------------------------------
                             |          b         z     P>|z|       e^b   e^bStdX
-----------------------------+--------------------------------------------------
too little   vs about right  |    -0.8933    -3.327     0.001     0.409     0.781
too little   vs too much     |    -1.0914    -3.685     0.000     0.336     0.740
about right  vs too little   |     0.8933     3.327     0.001     2.443     1.280
too much     vs too little   |     1.0914     3.685     0.000     2.979     1.352
--------------------------------------------------------------------------------
```

Mlogitplot command can assist you in interpreting all these sets of odds ratios further:

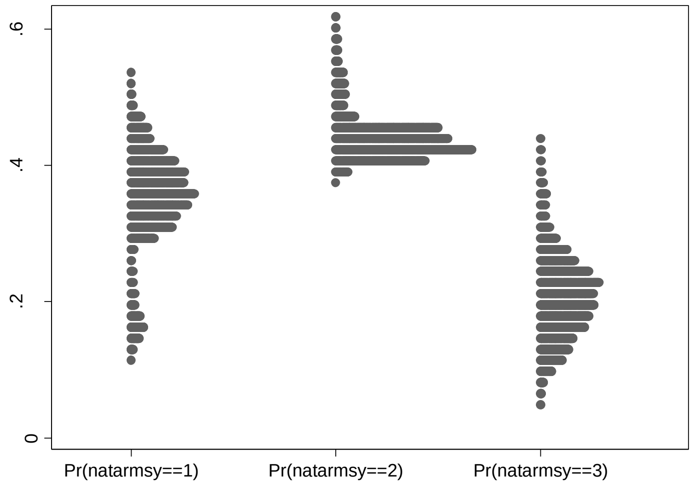
```
. mlogitplot, symbols(L R M) sig(.05)
```



2. Predicted probabilities and changes in predicted probabilities.

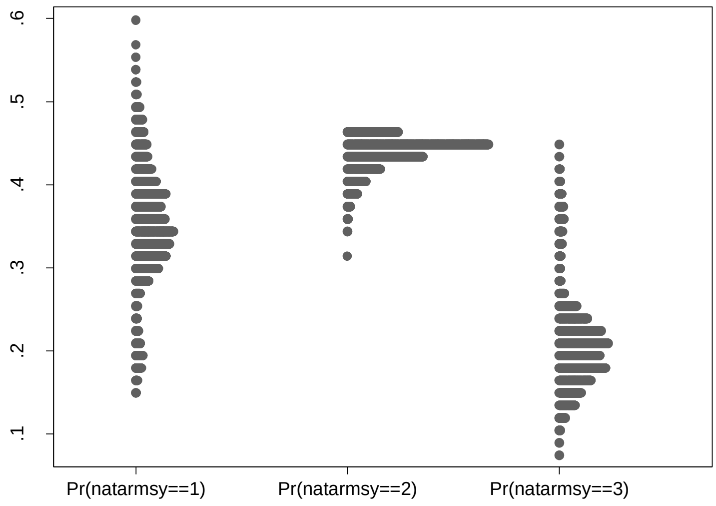
We can also examine predicted probabilities or changes in predicted probabilities. That is, we can use prvalue, prtab and prgen, and prchange just like we did for ordered logit.

```
. predict pm1 pm2 pm3
(option p assumed; predicted probabilities)
(26 missing values generated)

. dotplot pm1 pm2 pm3
```



If we compare this to the dotplot for ologit (obtained earlier), we will see some differences in the middle category; this is common. Overall, however, if the differences are substantial and affect other categories as well, mlogit may be more appropriate than ologit.

From ologit:

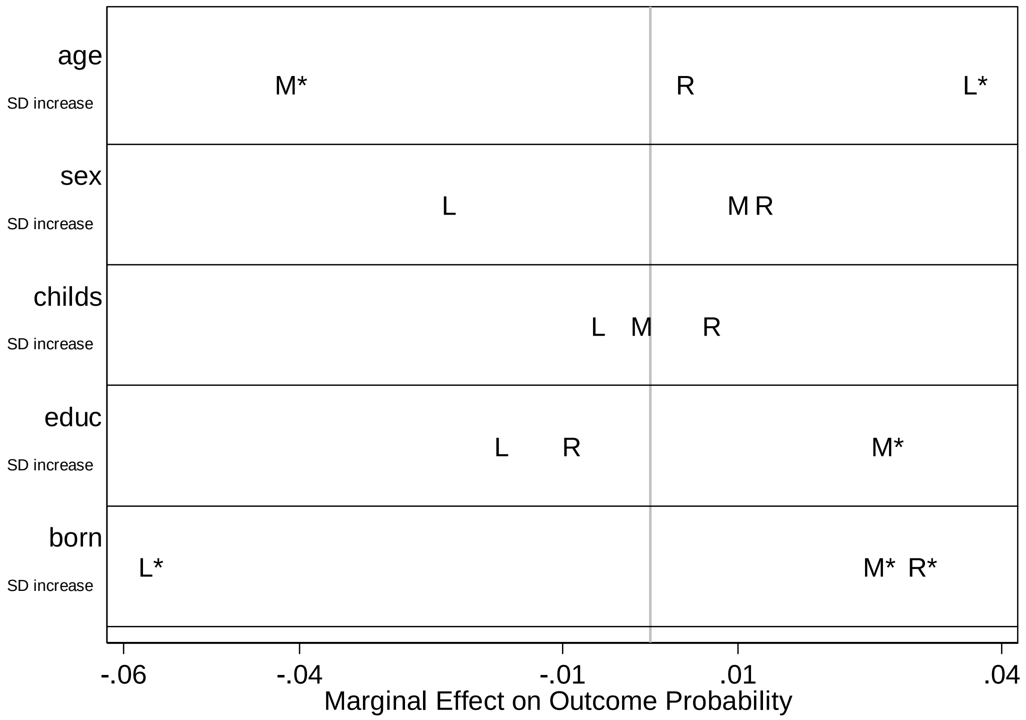



```
. mtable, atmeans

Expression: Pr(natarmsy), predict(outcome())

too_little  about_right  too_much
---------------------------------
     0.352        0.446     0.202

Specified values of covariates
```

```
           |      age       sex    childs      educ      born
 ----------+--------------------------------------------------
   Current |     46.4      1.55      1.85      13.4      1.08
```

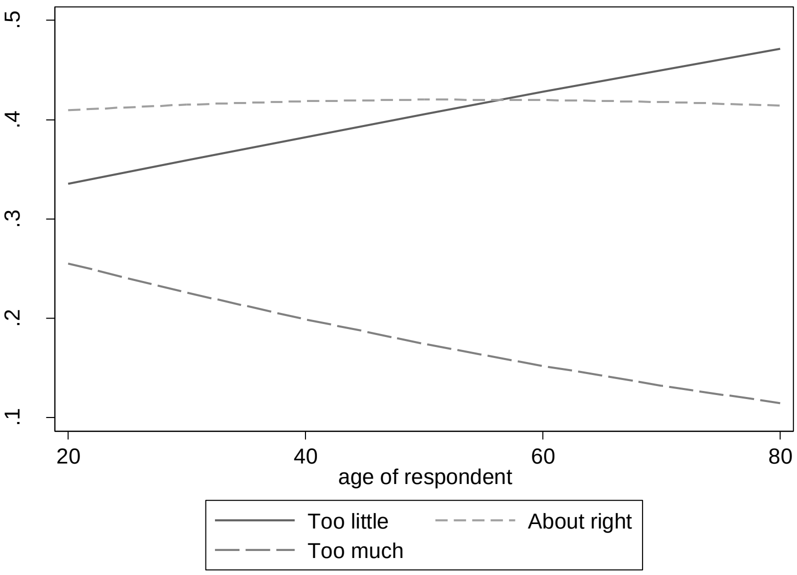
```
. mchange
mlogit: Changes in Pr(y) | Number of obs = 1337
Expression: Pr(natarmsy), predict(outcome())
             | too lit~e  about r~t   too much
-------------+--------------------------------
age          |
          +1 |     0.002      0.000     -0.003
     p-value |     0.008      0.665      0.001
         +SD |     0.037      0.004     -0.041
     p-value |     0.011      0.798      0.000
    Marginal |     0.002      0.000     -0.003
     p-value |     0.008      0.657      0.001
sex          |
          +1 |    -0.045      0.024      0.020
     p-value |     0.067      0.377      0.396
         +SD |    -0.023      0.013      0.010
     p-value |     0.072      0.360      0.380
    Marginal |    -0.046      0.026      0.020
     p-value |     0.077      0.344      0.363
childs       |
          +1 |    -0.003      0.004     -0.001
     p-value |     0.688      0.649      0.927
         +SD |    -0.006      0.007     -0.001
     p-value |     0.687      0.649      0.926
    Marginal |    -0.003      0.004     -0.001
     p-value |     0.689      0.648      0.928
educ         |
          +1 |    -0.006     -0.003      0.008
     p-value |     0.197      0.538      0.033
         +SD |    -0.017     -0.009      0.027
     p-value |     0.186      0.512      0.038
    Marginal |    -0.006     -0.003      0.008
     p-value |     0.203      0.551      0.031
born         |
          +1 |    -0.178      0.087      0.091
     p-value |     0.000      0.078      0.042
         +SD |    -0.057      0.031      0.026
     p-value |     0.000      0.028      0.015
    Marginal |    -0.214      0.120      0.094
     p-value |     0.000      0.020      0.008

Average predictions

             | too lit~e  about r~t   too much
-------------+--------------------------------
  Pr(y|base) |     0.355      0.438      0.207

. mchange, amount(sd) brief

mlogit: Changes in Pr(y) | Number of obs = 1337

Expression: Pr(natarmsy), predict(outcome())

             | too lit~e  about r~t   too much
-------------+--------------------------------
age          |
         +SD |     0.037      0.004     -0.041
     p-value |     0.011      0.798      0.000
sex          |
         +SD |    -0.023      0.013      0.010
     p-value |     0.072      0.360      0.380
childs       |
         +SD |    -0.006      0.007     -0.001
     p-value |     0.687      0.649      0.926
educ         |
         +SD |    -0.017     -0.009      0.027
```

```
     p-value |     0.186      0.512      0.038
born         |
         +SD |    -0.057      0.031      0.026
     p-value |     0.000      0.028      0.015

. mchangeplot, symbols(L R M) sig(.05)
```



We can also use marginsplot and mgen commands to create graphs of probabilities, for example:

```
. mgen, at(age=(20(10)80) sex=1 born=1) atmeans noatlegend  stub(mn_)
Predictions from: margins, at(age=(20(10)80) sex=1 born=1) atmeans noatlegend
predict(outcome())
Variable   Obs Unique      Mean       Min       Max  Label
----------------------------------------------------------------------------------------
mn_pr1       7      7  .4044002   .335254  .4711151  pr(y=too little) from margins
mn_ll1       7      7  .3519058  .2777555  .3981721  95% lower limit
mn_ul1       7      7  .4568945  .3927526  .5440581  95% upper limit
mn_age       7      7        50        20        80  age of respondent
mn_Cpr1      7      7  .4044002   .335254  .4711151  pr(y<=too little)
mn_pr2       7      7  .4165292   .409603  .4202045  pr(y=about right) from margins
mn_ll2       7      7  .3642952  .3443215   .379344  95% lower limit
mn_ul2       7      7  .4687631  .4606194  .4842867  95% upper limit
mn_Cpr2      7      7  .8209293  .7448571  .8854191  pr(y<=about right)
mn_pr3       7      7  .1790707  .1145808  .2551429  pr(y=too much) from margins
mn_ll3       7      7  .1398252  .0754612  .1966254  95% lower limit
mn_ul3       7      7  .2183162  .1537005  .3136605  95% upper limit
mn_Cpr3      7      2         1  .9999999         1  pr(y<=too much)
----------------------------------------------------------------------------------------
Specified values of covariates

      sex     childs       educ       born
-------------------------------------------
        1   1.854899   13.35228          1

. lab var mn_pr1 "Too little"
. lab var mn_pr2 "About right"
. lab var mn_pr3 "Too much"

. graph twoway (line mn_pr1 mn_pr2 mn_pr3 mn_age, sort lpattern(solid dash longdash)
ytitle("Predicted probability"))
```



Measures of Fit and Hypotheses Testing:

We can obtain fit statistics using fitstat like we did for binary and ordered logit.

To test hypotheses, you can use either tests based on likelihood ratos or Wald tests; the results are typically the same. Here, I demonstrate only the likelihood ratio-based options; see help mlogtest for Wald test options if desired. Compared to ordered logit, for multinomial logit hypotheses tests become more complicated. Here, if we want to drop a variable from the model, we want to test that it is not significant across all outcome categories (regardless of which one we omit). For that we use mlogtest command:

```
. mlogtest, lr
**** Likelihood-ratio tests for independent variables
 Ho: All coefficients associated with given variable(s) are 0.
   natarmsy |       chi2   df   P>chi2
-------------+-------------------------
        age |     14.266    2    0.001
        sex |      3.186    2    0.203
     childs |      0.231    2    0.891
       educ |      4.935    2    0.085
       born |     17.322    2    0.000
---------------------------------------
```

We conclude that variables sex, childs, and educ are not statistically significant across equations and could potentially be dropped (although we saw that educ was significant on .05 level in one of the models, when we join the results across categories it appears to be not significant). We can do the same with Wald test; the results look very similar but Wald test takes less computational resources (if the dataset is large and the model is very complex, for example) and Wald test can be used with robust SE (and LR test cannot).

```
. mlogtest, wald

Wald tests for independent variables (N=1337)

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

                 |      chi2    df   P>chi2
-----------------+-------------------------
             age |    13.702     2    0.001
             sex |     3.185     2    0.203
          childs |     0.231     2    0.891
            educ |     4.849     2    0.089
```

```
         born |    14.956     2    0.001
```

We can also test jointly whether these three variables are statistically significant as a set – i.e.. we can check if it makes sense to drop all three variables, sex, childs, and educ:

```
. mlogtest, lr set(sex childs educ)
**** Likelihood-ratio tests for independent variables
 Ho: All coefficients associated with given variable(s) are 0.
    natarmsy |       chi2   df   P>chi2
-------------+-------------------------
         age |     14.266    2    0.001
         sex |      3.186    2    0.203
      childs |      0.231    2    0.891
        educ |      4.935    2    0.085
        born |     17.322    2    0.000
-------------+-------------------------
      set_1: |      8.812    6    0.184
         sex |
      childs |
        educ |
---------------------------------------
```

The test indicates that we can drop all three (we interpret the probability for set_1).
Another test that we might want to do is to test whether it makes sense to combine some categories of our dependent variable – e.g. whether it makes sense to combine "too little" and "about right." We can combine them if all of our independent variables jointly do not differentiate between the two categories – nothing predicts that they are different.

```
. mlogtest, lrcomb
**** LR tests for combining outcome categories
 Ho: All coefficients except intercepts associated with given pair
     of outcomes are 0 (i.e., categories can be collapsed).
Categories tested |      chi2   df   P>chi2
------------------+------------------------
about_ri-too_much |    16.204    5    0.006
about_ri-too_litt |    16.993    5    0.005
too_much-too_litt |    41.557    5    0.000
-------------------------------------------
```

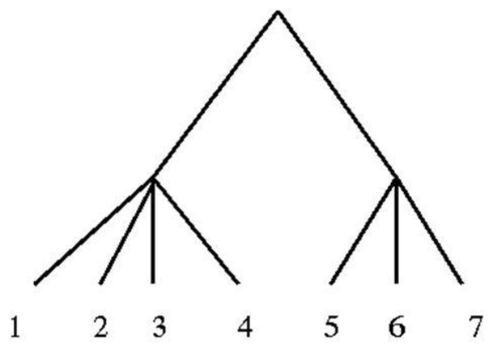
LR test and Wald test produce similar results - for all combinations of categories, we reject the hypotheses that our variables do not differentiate between categories. So we cannot combine any.

Diagnostics

1. Independence of Irrelevant Alternatives (IIA) assumption

One important assumption of multinomial logit is the assumption of Independence of Irrelevant Alternatives (IIA). That is, multinomial logit models assume that odds for each specific pair of outcomes do not depend on other outcomes available (deleting outcomes should not affect the odds among the remaining outcomes). Unfortunately, we do not have a good applied test for this assumption. The results of existing tests -- Hausman test and Small-Hsiao test – are inconsistent, and simulations show problematic conclusions – see pp. 407-410 in Long and Freese for discussion of this. Therefore, the main advice is that we should be sure that from a theoretical standpoint, the alternatives "can plausibly be assumed to be distinct and weighted independently in the eyes of each decision maker" (McFadden 1974, cited in Long and Freese). That is, we should not have a scenario where some of the alternatives are closer substitutes for each other than other alternatives.

If IIA indeed assumption does not hold, one alternative that allows partial relaxation of that assumption is a nested model, i.e. a model in which some categories are considered to share a nest together. IIA holds within a nest but not across nests.



The commands in Stata that you'd want to look into are nlogit and nlogitrum, but the data would have to be restructured with each alternative being a separate observation (separate line in the dataset) – see "Specification(s) of Nested Logit Models" by Florian Heiss: http://www.mea.mpisoc.mpg.de/uploads/user_mea_discussionpapers/dp16.pdf

2. Multicollinearity.
As was the case for binary and ordered logit, we can test for multicollinearity by running OLS model instead of multinomial logit and using vif.
3. Linearity and Additivity.
As usual, you should start the process by examining the univariate distributions and the bivariate relationships. Like in ordered logit, in order to examine bivariate relationships as well as to conduct many diagnostics, we should create the dichotomies corresponding to each equation:

```
. gen natarmsy1=(natarmsy==1) if (natarmsy==1 | natarmsy==3)
(2008 missing values generated)
. gen natarmsy2=(natarmsy==2) if (natarmsy==2 | natarmsy==3)
(1894 missing values generated)
```

For each of these dichotomous variables, we can then obtain lowess plots, just like we did for ordered logit. We can then use these dichotomies to run binary logits and conduct various multivariate diagnostics.

```
. logit natarmsy1 age sex childs educ born
Logistic regression                               Number of obs   =        751
                                                  LR chi2(5)      =      42.34
                                                  Prob > chi2     =     0.0000
Log likelihood = -473.24011                       Pseudo R2       =     0.0428
------------------------------------------------------------------------------
   natarmsy1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |    .020441   .0052802     3.87   0.000      .010092      .03079
         sex |   -.257952    .157136    -1.64   0.101    -.5659329     .050029
      childs |  -.0009124   .0532109    -0.02   0.986    -.1052039    .1033791
        educ |  -.0584523   .0282196    -2.07   0.038    -.1137618   -.0031428
        born |  -1.038649   .3007153    -3.45   0.001     -1.62804   -.4492576
       _cons |    1.91543   .5894602     3.25   0.001     .7601091     3.07075
------------------------------------------------------------------------------

. logit natarmsy2 age sex childs educ born
Logistic regression                               Number of obs   =        863
                                                  LR chi2(5)      =      15.22
                                                  Prob > chi2     =     0.0095
Log likelihood = -534.01018                       Pseudo R2       =     0.0140
------------------------------------------------------------------------------
   natarmsy2 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0128336   .0049079     2.61   0.009     .0032143    .0224529
         sex |  -.0536544   .1496431    -0.36   0.720    -.3469494    .2396406
      childs |   .0114876   .0522925     0.22   0.826    -.0910039    .1139791
        educ |  -.0426433   .0247853    -1.72   0.085    -.0912217     .005935
```

```
        born |  -.2192112    .232668    -0.94   0.346    -.675232    .2368097
       _cons |   1.062732   .5271903     2.02   0.044    .0294579    2.096006
------------------------------------------------------------------------------
```

Note that in order for this approach to work, each binary model should look similar to the corresponding equation of the multinomial model. That will typically be the case if the IIA assumption holds. But let's compare:

```
. mlogit natarmsy age sex childs educ born, b(3)
Multinomial logistic regression                   Number of obs   =       1337
                                                  LR chi2(10)     =      46.19
                                                  Prob > chi2     =     0.0000
Log likelihood = -1387.8455                       Pseudo R2       =     0.0164
------------------------------------------------------------------------------
    natarmsy |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
too little   |
         age |   .0190126   .0051423     3.70   0.000     .0089338    .0290914
         sex |  -.2340065   .1550509    -1.51   0.131    -.5379007    .0698876
      childs |  -.0065869   .0537937    -0.12   0.903    -.1120205    .0988468
        educ |  -.0578152   .0270313    -2.14   0.032    -.1107956   -.0048347
        born |  -1.091425   .2962101    -3.68   0.000    -1.671986   -.5108634
       _cons |   2.002426   .5858732     3.42   0.001     .8541352    3.150716
-------------+----------------------------------------------------------------
about right  |
         age |   .0135326   .0049789     2.72   0.007     .0037742     .023291
         sex |  -.0420268   .1485803    -0.28   0.777    -.3332389    .2491853
      childs |   .0128663   .0519464     0.25   0.804    -.0889467    .1146793
        educ |  -.0475599   .0257811    -1.84   0.065      -.09809    .0029701
        born |  -.1980986   .2326138    -0.85   0.394    -.6540133    .2578161
       _cons |   1.054006   .5377872     1.96   0.050    -.0000375     2.10805
------------------------------------------------------------------------------
(natarmsy==too much is the base outcome)
```

Looks similar. For each of these binary models, you can do the full range of linearity diagnostics that are appropriate for binary models – i.e., run Box-Tidwell test, etc. Like with ordered logit, you should be aware of the possibility that you might find different patterns for different binary models; in that case, you'll have to figure out how to reconcile them in mlogit.

You can also use fitint for these binary models (fitint does not work with mlogit), although keep in mind the warnings regarding interpreting interactions mentioned in the discussion of binary logit.

<u>4. Outliers and Influential Observations</u>

In order to do unusual data diagnostics for multinomial logit, we should also rely on separate binary models we've used in previous steps. All the same methods we discussed for binary logit apply here as well, and like in ordered logit, the fact that you'll have to do a separate search for unusual data for each binary model may complicate things if they suggest that different observations are influential. Make sure that you test the potential effects of these influential observations on your mlogit model (rather than just on individual binary logits).

<u>5. Error term distribution</u>

Like we did for binary and ordered logit, we can obtain robust standard errors for the multinomial logit model in order to check whether our assumptions about error distribution hold (compare with the model on pp.1-2):

```
. mlogit natarmsy age sex childs educ born, robust
Multinomial logistic regression                   Number of obs   =       1337
                                                  Wald chi2(10)   =      40.85
                                                  Prob > chi2     =     0.0000
Log pseudolikelihood = -1387.8455                 Pseudo R2       =     0.0164
------------------------------------------------------------------------------
             |               Robust
    natarmsy |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
too little   |
         age |     .00548   .0039155     1.40   0.162    -.0021943    .0131543
```

```
         sex |  -.1919798   .1254863    -1.53   0.126    -.4379285    .0539689
      childs |  -.0194531   .0405578    -0.48   0.631    -.0989449    .0600386
        educ |  -.0102552    .019935    -0.51   0.607     -.049327    .0288166
        born |  -.8933259   .2701132    -3.31   0.001    -1.422738   -.3639138
       _cons |   .9484196   .4706752     2.02   0.044     .0259132    1.870926
-------------+----------------------------------------------------------------
too much     |
         age |  -.0135326   .0050701    -2.67   0.008    -.0234697   -.0035955
         sex |   .0420268   .1482007     0.28   0.777    -.2484413    .3324949
      childs |  -.0128663   .0534559    -0.24   0.810     -.117638    .0919054
        educ |   .0475599   .0278666     1.71   0.088    -.0070576    .1021775
        born |   .1980986   .2302914     0.86   0.390    -.2532642    .6494614
       _cons |  -1.054006   .5745375    -1.83   0.067    -2.180079    .0720669
------------------------------------------------------------------------------
(natarmsy==about right is the base outcome)
```

<u>The problem of perfect prediction in logit, ologit and mlogit</u>

Sometimes when running analyses for categorical outcomes, we run into the problem of perfect prediction (perfect separation). For example:

```
. mlogit natarmsy age sex childs i.educ born
Iteration 0:   log likelihood = -1410.9409
Iteration 1:   log likelihood = -1367.5166
Iteration 2:   log likelihood = -1365.8514
Iteration 3:   log likelihood = -1365.6452
Iteration 4:   log likelihood =  -1365.603
Iteration 5:   log likelihood = -1365.5934
Iteration 6:   log likelihood = -1365.5918
Iteration 7:   log likelihood = -1365.5916
Iteration 8:   log likelihood = -1365.5916
Iteration 9:   log likelihood = -1365.5916
Multinomial logistic regression                   Number of obs   =       1337
                                                  LR chi2(48)     =      90.70
                                                  Prob > chi2     =     0.0002
Log likelihood = -1365.5916                       Pseudo R2       =     0.0321
------------------------------------------------------------------------------
    natarmsy |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
too_little   |
         age |   .0077433   .0040551     1.91   0.056    -.0002046    .0156912
         sex |  -.2088383   .1271909    -1.64   0.101    -.4581279    .0404513
      childs |  -.0220421   .0424435    -0.52   0.604    -.1052298    .0611457
             |
        educ |
          1  |  -14.02326   2287.734    -0.01   0.995      -4497.9    4469.853
          2  |   .7975166   1.408267     0.57   0.571    -1.962636    3.557669
          3  |  -14.72475   1617.191    -0.01   0.993     -3184.36    3154.911
          4  |   .6330178   1.880399     0.34   0.736    -3.052496    4.318532
          5  |  -.0348836   1.698759    -0.02   0.984    -3.364391    3.294624
          6  |   1.462163   1.461175     1.00   0.317    -1.401688    4.326014
          7  |   1.367193   1.742221     0.78   0.433    -2.047498    4.781884
          8  |  -.2593536   1.321068    -0.20   0.844    -2.848599    2.329892
          9  |   .8447427    1.29865     0.65   0.515    -1.700564    3.390049
         10  |    .571317   1.284897     0.44   0.657    -1.947035    3.089669
         11  |   .6201585   1.265171     0.49   0.624    -1.859531    3.099848
         12  |   .7967541   1.241752     0.64   0.521    -1.637035    3.230543
         13  |   1.138548   1.252149     0.91   0.363    -1.315618    3.592715
         14  |   .7783036   1.249805     0.62   0.533    -1.671269    3.227876
         15  |    .403707   1.268138     0.32   0.750    -2.081797    2.889211
         16  |   .6326915   1.251138     0.51   0.613    -1.819494    3.084877
         17  |   .6176581   1.294039     0.48   0.633    -1.918613    3.153929
         18  |   .4673819   1.272086     0.37   0.713    -2.025861    2.960624
         19  |   .2741944   1.382557     0.20   0.843    -2.435568    2.983957
         20  |   .2140612   1.321342     0.16   0.871    -2.375722    2.803844
             |
        born |  -.8631172    .275354    -3.13   0.002    -1.402801   -.3234333
       _cons |  -.0048823    1.30334    -0.00   0.997    -2.559381    2.549616
-------------+----------------------------------------------------------------
```

```
about_right  |  (base outcome)
-------------+----------------------------------------------------------------
too_much     |
         age |  -.0150876   .0051592    -2.92   0.003    -.0251994   -.0049758
         sex |   .0871751   .1507846     0.58   0.563    -.2083572    .3827074
      childs |  -.0174627   .0532681    -0.33   0.743    -.1218663    .0869409
             |
        educ |
          1  |  -15.44767   2992.642    -0.01   0.996    -5880.919    5850.023
          2  |  -.6565282   1.499769    -0.44   0.662     -3.59602    2.282964
          3  |  -15.41758   2115.643    -0.01   0.994    -4162.001    4131.166
          4  |   -14.1123   1632.554    -0.01   0.993     -3213.86    3185.635
          5  |  -14.76051   1192.335    -0.01   0.990    -2351.693    2322.172
          6  |  -.1012508   1.542967    -0.07   0.948    -3.125411    2.922909
          7  |     .47356   1.888627     0.25   0.802    -3.228081    4.175201
          8  |  -.6447085   1.327683    -0.49   0.627     -3.24692    1.957503
          9  |  -.6039934   1.336655    -0.45   0.651    -3.223788    2.015802
         10  |  -.8738507   1.320653    -0.66   0.508    -3.462283    1.714581
         11  |  -.4533993    1.27835    -0.35   0.723     -2.95892    2.052121
         12  |  -.5542129   1.251803    -0.44   0.658    -3.007701    1.899275
         13  |  -.8929498   1.274891    -0.70   0.484     -3.39169     1.60579
         14  |  -.7702706   1.264435    -0.61   0.542    -3.248517    1.707976
         15  |  -1.019888   1.291675    -0.79   0.430    -3.551524    1.511748
         16  |  -.4348901   1.262842    -0.34   0.731    -2.910014    2.040234
         17  |  -1.006427   1.338302    -0.75   0.452     -3.62945    1.616597
         18  |  -.0167748   1.277241    -0.01   0.990    -2.520121    2.486571
         19  |   .5239221   1.329945     0.39   0.694    -2.082722    3.130567
         20  |  -.3176245   1.316061    -0.24   0.809    -2.897056    2.261807
             |
        born |   .1878618   .2412132     0.78   0.436    -.2849074     .660631
       _cons |   .1783677   1.317699     0.14   0.892    -2.404275    2.761011
------------------------------------------------------------------------------
Note: 3 observations completely determined.  Standard errors questionable.
```

```
. tab educ natarmsy if e(sample)

   highest |
   year of |
    school |  national defense -- version y
 completed | too littl  about rig   too much |     Total
-----------+---------------------------------+----------
         0 |         1          2          1 |         4
         1 |         0          1          0 |         1
         2 |         4          5          2 |        11
         3 |         0          2          0 |         2
         4 |         1          1          0 |         2
         5 |         1          3          0 |         4
         6 |         4          3          2 |         9
         7 |         2          1          1 |         4
         8 |         6         17          6 |        29
         9 |        12         13          6 |        31
        10 |        14         20          7 |        41
        11 |        25         34         19 |        78
        12 |       147        161         75 |       383
        13 |        62         52         19 |       133
        14 |        71         84         35 |       190
        15 |        22         38         12 |        72
        16 |        58         76         42 |       176
        17 |        13         19          6 |        38
        18 |        20         31         24 |        75
        19 |         4          8         11 |        23
        20 |         7         15          9 |        31
-----------+---------------------------------+----------
     Total |       474        586        277 |     1,337
```

Same for logit:

```
. gen natarmsy_much=(natarmsy>2) if natarmsy<.
(1417 missing values generated)

. logit natarmsy_much age sex childs i.educ born

note: 1.educ != 0 predicts failure perfectly
      1.educ dropped and 1 obs not used

note: 3.educ != 0 predicts failure perfectly
      3.educ dropped and 2 obs not used

note: 4.educ != 0 predicts failure perfectly
      4.educ dropped and 2 obs not used

note: 5.educ != 0 predicts failure perfectly
      5.educ dropped and 4 obs not used

Iteration 0:   log likelihood = -680.03556
Iteration 1:   log likelihood =  -656.1523
Iteration 2:   log likelihood = -655.26998
Iteration 3:   log likelihood = -655.26951
Iteration 4:   log likelihood = -655.26951

Logistic regression                               Number of obs   =       1328
                                                  LR chi2(20)     =      49.53
                                                  Prob > chi2     =     0.0003
Log likelihood = -655.26951                       Pseudo R2       =     0.0364

------------------------------------------------------------------------------
natarmsy_much |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
          age |  -.0184596   .0048344    -3.82   0.000    -.0279348   -.0089843
          sex |    .177159   .1404164     1.26   0.207     -.098052    .4523701
       childs |  -.0082026   .0499406    -0.16   0.870    -.1060844    .0896792
              |
         educ |
           1  |          0  (empty)
           2  |  -.9725465   1.414436    -0.69   0.492     -3.74479    1.799697
           3  |          0  (empty)
           4  |          0  (empty)
           5  |          0  (empty)
           6  |  -.7142174   1.427659    -0.50   0.617    -3.512377    2.083942
           7  |   -.206547   1.654014    -0.12   0.901    -3.448355    3.035261
           8  |  -.5872592   1.258309    -0.47   0.641      -3.0535    1.878982
           9  |  -.9528357   1.259104    -0.76   0.449    -3.420635    1.514963
          10  |  -1.102306   1.248176    -0.88   0.377    -3.548687    1.344074
          11  |  -.7045497   1.206182    -0.58   0.559    -3.068623    1.659524
          12  |  -.8804889    1.18186    -0.75   0.456    -3.196891    1.435913
          13  |  -1.383427   1.202971    -1.15   0.250    -3.741207    .9743542
          14  |    -1.0862   1.193678    -0.91   0.363    -3.425766    1.253367
          15  |   -1.18731   1.221016    -0.97   0.331    -3.580458    1.205838
          16  |  -.6890343   1.191933    -0.58   0.563    -3.025181    1.647112
          17  |  -1.252424   1.265548    -0.99   0.322    -3.732853    1.228005
          18  |  -.2018643   1.204461    -0.17   0.867    -2.562565    2.158836
          19  |   .4046231   1.249601     0.32   0.746    -2.044549    2.853795
          20  |  -.4204136   1.242649    -0.34   0.735    -2.855961    2.015133
              |
         born |   .4849982   .2296187     2.11   0.035     .0349537    .9350427
        _cons |  -.4493042   1.243108    -0.36   0.718     -2.88575    1.987142
------------------------------------------------------------------------------
```

The default solution in logit vs. mlogit is different – logit drops out the problematic cases and estimates the model without them; mlogit estimates the model with them but reports that SE are problematic. I usually try to avoid presenting either solution if possible and try to group the dummy variables (this is most common when we use groups of dummies with some small categories). For example here:

```
. gen educ5=educ
```

```
(12 missing values generated)

. replace educ5=5 if educ<5
(30 real changes made)

. logit natarmsy_much age sex childs i.educ5 born

Iteration 0:   log likelihood = -682.13296
Iteration 1:   log likelihood = -657.74178
Iteration 2:   log likelihood = -656.81282
Iteration 3:   log likelihood = -656.81221
Iteration 4:   log likelihood = -656.81221

Logistic regression                               Number of obs   =       1337
                                                  LR chi2(19)     =      50.64
                                                  Prob > chi2     =     0.0001
Log likelihood = -656.81221                       Pseudo R2       =     0.0371

------------------------------------------------------------------------------
natarmsy_much |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          age |  -.0186419   .0048357    -3.86   0.000    -.0281198    -.009164
          sex |    .170222   .1402375     1.21   0.225    -.1046385    .4450824
       childs |  -.0068073   .0496539    -0.14   0.891    -.1041272    .0905127
              |
        educ5 |
          6   |   .5019065   1.033303     0.49   0.627     -1.52333    2.527143
          7   |   1.005343   1.326445     0.76   0.448    -1.594441    3.605128
          8   |   .6242693   .7822843     0.80   0.425    -.9089798    2.157518
          9   |   .2575394   .7806997     0.33   0.741    -1.272604    1.787683
         10   |   .1097225   .7581913     0.14   0.885    -1.376305     1.59575
         11   |   .5066539   .6876422     0.74   0.461       -.8411    1.854408
         12   |   .3311681     .64536     0.51   0.608    -.9337143    1.596051
         13   |  -.1716817   .6811657    -0.25   0.801    -1.506742    1.163379
         14   |   .1253993   .6628517     0.19   0.850    -1.173766    1.424565
         15   |   .0254604   .7100298     0.04   0.971    -1.366172    1.417093
         16   |   .5231261   .6594135     0.79   0.428    -.7693006    1.815553
         17   |  -.0368228    .778926    -0.05   0.962     -1.56349    1.489844
         18   |   1.012178   .6810217     1.49   0.137    -.3225998    2.346956
         19   |   1.618002    .759363     2.13   0.033     .1296779    3.106326
         20   |   .7934305   .7467434     1.06   0.288    -.6701597    2.257021
              |
         born |   .4729687   .2289636     2.07   0.039     .0242082    .9217292
        _cons |  -1.631795   .7728145    -2.11   0.035    -3.146483   -.1171062
------------------------------------------------------------------------------
```

And if combining dummies is not possible (e.g. this happens for a single dummy), I would opt for leaving out the problematic variable rather than leaving out cases.