The Multinomial Logit Model for Nominal Response Data

Lin Chen

Department of Public Health Sciences
The University of Chicago

Nominal Categorical Data

- A nominal response variable takes on discrete categories, but there is not an intrinsic ordering of the categories.
- Examples:
  1. Blood type: A, B, AB, or O.
  2. Hair color: blonde, brown, black, red, etc.
  3. Tumor cell pathologic feature categories
  4. College majors: maths, computer sciences, English literature, etc.

Nominal Response: notation

- Let $C_1, C_2, ..., C_k,\ k \geq 2$ denote the k categories for the response (no intrinsic order).
- Let Y_i be the response variable for the i^{th} individual, with Y_i taking the value j if the response is in category $C_j,\ j = 1, 2, ..., k$.
- $p_{ij} = P(Y_i = j) = P[$ *individual* i *responds in category* $C_j]$
- Does it make sense to model the cumulative probability $\gamma_{ij} = P(Y_i \leq j)$ as in the ordinal response case?

  Not really, in ordinal logistic regression the cumulative probability γ_{ij} is order-dependent. When modeling nominal categorical responses, there is no ordering of the categories. We cannot collapse any categories.

Standard Logit Model $\longrightarrow$ Multinomial Logit Model

- Standard logit model: $\log(\frac{p}{1-p}) = \boldsymbol{x}^T\boldsymbol{\beta}$, where p is the probability of response and $1-p$ is the probability of non-response.
- Multinomial logit model:
  - Pick a baseline/reference category (which plays the same role as category “non-response" in standard logit model), let’s say C_1
  - Model can be written as:

$$\log(\frac{p_j}{p_1}) = \alpha_j + \boldsymbol{x}^T\boldsymbol{\beta}_j, \quad j = 2, 3, ..., k \tag{1}$$

- Under model (1) and given the fact that $\sum_{j=1}^{j=k} p_j = 1$,

$$p_1 = \frac{1}{1 + \exp(\alpha_2 + \boldsymbol{x}^T\boldsymbol{\beta}_2) + \exp(\alpha_3 + \boldsymbol{x}^T\boldsymbol{\beta}_3) + \cdots + \exp(\alpha_k + \boldsymbol{x}^T\boldsymbol{\beta}_k)} \tag{2}$$

$$p_j = \frac{\exp(\alpha_j + \boldsymbol{x}^T\boldsymbol{\beta}_j)}{1 + \exp(\alpha_2 + \boldsymbol{x}^T\boldsymbol{\beta}_2) + \exp(\alpha_3 + \boldsymbol{x}^T\boldsymbol{\beta}_3) + \cdots + \exp(\alpha_k + \boldsymbol{x}^T\boldsymbol{\beta}_k)}, j = 2, 3, ..., k \tag{3}$$

Proportional Odds (i.e. Ordinal Logit) Model vs. Multinomial Logit Model

As in other logit models, $\boldsymbol{x}$ is a covariate vector without intercept and is of dimension p. Contrasting the two model extensions of the standard logit model:

- Proportional odds (i.e. Ordinal logit) model: $\log(\frac{\gamma_j}{1-\gamma_j}) = d_j - \boldsymbol{x}^T \boldsymbol{\beta}, \ \ j = 1, 2, ..., k-1$
- Multinomial logit model: $\log(\frac{p_j}{p_1}) = \alpha_j + \boldsymbol{x}^T \boldsymbol{\beta}_j, \ \ j = 2, 3, ..., k$

Proportional Odds Model vs. Multinomial Logit Model

- Proportional odds model: $\log(\frac{\gamma_j}{1-\gamma_j}) = d_j - x^T\beta,\ \ j = 1, 2, ..., k-1$ cumulative
- Multinomial logit model: $\log(\frac{p_j}{p_1}) = \alpha_j + x^T\beta_j,\ \ j = 2, 3, ..., k$ Individual
- Comparisons:
  1. Proportional odds model predicts cumulative probability (except the last category), whereas multinomial logit model predicts the probability for each category (see Slide 4).
  2. Proportional odds model has constant slope β: the effect of x, is the same for all $k-1$ ways to **collapse response into binary outcomes**. Multinomial logit model has different slope β_j depending on the response category. ologit has constant slope
  3. Proportional odds model, has $(k-1)$ intercepts plus p slopes, for a total of $k-1+p$ parameters to be estimated. Multinomial logit model, has $(k-1)$ intercepts plus $(k-1)\times p$ slopes, for a total of $(k-1)+p\times(k-1)$ parameters to be estimated.

Example: *High School Program Choice*

Students entering high school make a program choice among general, vocational, and academic program studies. Their choice might be related to their writing score and their socio-economic status (SES). These data describe 200 high school students.

```
. list prog ses write in 1/15, clean

           prog      ses   write
  1.   vocation      low      44
  2.   vocation   middle      41
  3.   academic      low      65
  4.   academic      low      50
  5.   academic      low      40
  6.   academic      low      41
  7.   academic   middle      54
  8.   academic      low      44
  9.   vocation   middle      49
 10.    general   middle      54
 11.   academic   middle      46
 12.   vocation   middle      44
 13.   vocation   middle      46
 14.   academic     high      41
 15.   vocation     high      39
. . .
```

High School Program Choice - looking at the data

Examine the variables using `codebook`. In this data, both `prog` and `ses` are numeric with labels attached, and can be analyzed directly.

```
. codebook prog
--------------------------------------------------------------------------------
prog                                                          type of program
--------------------------------------------------------------------------------
                type:   numeric (float)
               label:   sel

               range:   [1,3]                        units:  1
       unique values:   3                        missing .:  0/200

          tabulation:   Freq.   Numeric  Label
                           45         1  general
                          105         2  academic
                           50         3  vocation

. codebook ses
--------------------------------------------------------------------------------
ses                                                              (unlabeled)
--------------------------------------------------------------------------------
                type:   numeric (float)
               label:   sl

               range:   [1,3]                        units:  1
       unique values:   3                        missing .:  0/200

          tabulation:   Freq.   Numeric  Label
                           47         1  low
                           95         2  middle
                           58         3  high
```

HS Program Choice: exploratory analyses

```
. tab prog ses, row col chi2
|-------------------|
|     frequency     |
|  row percentage   |
| column percentage |
+-------------------+
   type of |                  ses
   program |       low     middle       high |     Total
-----------+---------------------------------+----------
   general |        16         20          9 |        45
           |     35.56      44.44      20.00 |    100.00
           |     34.04      21.05      15.52 |     22.50
-----------+---------------------------------+----------
  academic |        19         44         42 |       105
           |     18.10      41.90      40.00 |    100.00
           |     40.43      46.32      72.41 |     52.50
-----------+---------------------------------+----------
  vocation |        12         31          7 |        50
           |     24.00      62.00      14.00 |    100.00
           |     25.53      32.63      12.07 |     25.00
-----------+---------------------------------+----------
     Total |        47         95         58 |       200
           |     23.50      47.50      29.00 |    100.00
           |    100.00     100.00     100.00 |    100.00

          Pearson chi2(4) =  16.6044   Pr = 0.002
```

Test of association suggests strong relationship between SES and program choice

HS Program Choice: some exploratory analyses

```
. oneway write prog, means

             | Summary of
             |   writing
     type of |    score
     program |        Mean
-------------+------------
     general |   51.333333
    academic |   56.257143
    vocation |       46.76
-------------+------------
       Total |      52.775

                        Analysis of Variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------
Between groups      3175.69786      2   1587.84893     21.27     0.0000
 Within groups      14703.1771    197    74.635417
------------------------------------------------------------------------
    Total            17878.875    199    89.843593

Bartlett's test for equal variances:  chi2(2) =   2.6184  Prob>chi2 = 0.270
```

Difference in mean writing score by program choice. Writing score may affect how students choose different programs.

HS Program Choice: the multinomial logit model

Now consider program choice as a nominal categorical variable with three categories, general, vocational or academic (setting reference using the option `base(2)`). We will fit a multinomial logit model (using `mlogit`) comparing each category to reference. **This model produces odds ratios from sub-tables of the 3x3 table of program type choice by SES.**

multinomial

"academic" as baseline

```
. mlogit prog i.ses , base(2) nolog
Multinomial logistic regression                 Number of obs     =        200
                                                LR chi2(4)        =      16.78
                                                Prob > chi2       =     0.0021
Log likelihood = -195.70519                     Pseudo R2         =     0.0411
------------------------------------------------------------------------------
        prog |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
general      |
         ses |
     middle  |  -.6166071   .4334269    -1.42   0.155    -1.466108     .232894
       high  |  -1.368595   .5000522    -2.74   0.006    -2.348679   -.3885105
             |
       _cons |  -.1718503   .3393104    -0.51   0.613    -.8368865     .493186
-------------+----------------------------------------------------------------
academic     |  (base outcome)
-------------+----------------------------------------------------------------
vocation     |
         ses |
     middle  |   .1093299   .4369785     0.25   0.802    -.7471323    .9657921
       high  |  -1.332227   .5501196    -2.42   0.015    -2.410442   -.2540125
             |
       _cons |  -.4595323   .3687342    -1.25   0.213    -1.182238    .2631734
------------------------------------------------------------------------------
```

SES only.

HS Program Choice: the multinomial logit model

What are these parameters? Create a 2x2 table with outcome the program choice (general vs. academic) and the 'exposure' variable SES, with middle SES exposed and low SES unexposed:

	Exposure SES		
Program	Exposed middle	Unexposed low	
case+ general	20	16	36
case− academic	44	19	63

Note that the OR from this table is $(20 \times 19)/(44 \times 16) = 0.5398$. Taking the log yields the parameter from the above model $(-.6166)$. Individuals of middle SES are less likely than low SES individuals to choose the general program over the academic program. √

You may expand this table to a 2x3 table and consider different SES as multiple exposures. This will estimate the model comparing general versus academic programs.

HS Program Choice: the multinomial logit model

Another subtable. Create a 2x2 table with outcome the program choice (vocational vs. academic) and the 'exposure' variable SES, with middle SES exposed and low SES unexposed:

	SES		
Program	middle	low	
vocational	31	12	43
academic	44	19	63

OR from this table is $\overset{\sim 1}{1.1155}$. Taking the log yields the parameter from the above model (0.1093). Individuals of middle SES are not any more or less likely to choose the vocational program over the academic program. ~ no difference

Similarly, you may expand this table to consider all SESs to estimate the parameters comparing vocational and academic programs.

Model with writing score added:

```
. mlogit prog i.ses write, base(2) nolog
Multinomial logistic regression                  Number of obs    =       200
                                                 LR chi2(6)       =     48.23
                                                 Prob > chi2      =    0.0000
Log likelihood = -179.98173                      Pseudo R2        =    0.1182
------------------------------------------------------------------------------
        prog |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
general      |
         ses |
      middle |   -.533291   .4437321    -1.20   0.229     -1.40299    .336408
        high |  -1.162832   .5142195    -2.26   0.024    -2.170684   -.1549804
             |
       write |  -.0579284   .0214109    -2.71   0.007    -.0998931   -.0159637
       _cons |   2.852186   1.166439     2.45   0.014     .5660075    5.138365
-------------+----------------------------------------------------------------
academic     |  (base outcome)
-------------+----------------------------------------------------------------
vocation     |
         ses |
      middle |   .2913931   .4763737     0.61   0.541    -.6422822    1.225068
        high |  -.9826703   .5955669    -1.65   0.099     -2.14996    .1846195
             |
       write |  -.1136026   .0222199    -5.11   0.000    -.1571528   -.0700524
       _cons |     5.2182   1.163549     4.48   0.000     2.937686    7.498714
------------------------------------------------------------------------------
```

SES and writing

HS Program Choice: The fitted multinomial logit model

The model produces equations as follows:

$$\log\left(\frac{\hat{p}_{\mathit{general}}}{\hat{p}_{\mathit{academic}}}\right) = 2.852 - 0.533 \times \mathit{middle} - 1.16 \times \mathit{high} - 0.0579 \times \mathit{write}$$

v.s.
"compared to" baseline

$$\log\left(\frac{\hat{p}_{\mathit{vocational}}}{\hat{p}_{\mathit{academic}}}\right) = 5.218 + 0.291 \times \mathit{middle} - 0.983 \times \mathit{high} - 0.114 \times \mathit{write}$$

low

Note that only none or 1 of the covariates *middle* and *high* take on value 1 in any given prediction. The categories are mutually exclusive.

can't be both middle and high

~~middle=1~~
~~high=1~~

HS Program Choice: Model prediction

- What is the predicted relative chance of a student with writing score of 60 from high SES choosing vocational program over $\frac{\text{vocational}}{\text{academic}}$ academic program?

  Here is the fitted model:

$$\log\left(\frac{\hat{p}_{\mathit{vocational}}}{\hat{p}_{\mathit{academic}}}\right) = 5.218 + \cancel{0.291 \times \mathit{middle}} - 0.983 \times \mathit{high} - 0.114 \times \mathit{write}$$

  Based on the fitted model,

$$\log\left(\frac{\hat{p}_{\mathit{vocational}}}{\hat{p}_{\mathit{academic}}}\right) = 5.218 - 0.983 \times \mathit{1} - 0.114 \times \mathit{60}$$

  The relative chance is
  $\frac{\hat{p}_{\mathit{vocational}}}{\hat{p}_{\mathit{academic}}} = \exp(5.218 - 0.983 - 0.114 \times 60) = 0.074$ ✓

HS Program Choice: Model prediction

- What is the predicted absolute chance of a student with writing score of 60 from high SES choosing vocational program? (check Slide 4)

  Based on the fitted model:

  $\log\left(\frac{vocational}{academic}\right) - \log\left(\frac{general}{academic}\right) = \log\left(\frac{vocational}{academic} \div \frac{general}{academic}\right)$

  $= \log\left(\frac{vocational}{general}\right)$

$$\log\left(\frac{\hat{p}_{general}}{\hat{p}_{academic}}\right) = 2.852 - 1.16 \times 1 - 0.0579 \times 60$$

$$\log\left(\frac{\hat{p}_{vocational}}{\hat{p}_{academic}}\right) = 5.218 - 0.983 \times 1 - 0.114 \times 60$$

  The absolute chance is

  $\hat{p}_{vocational} = \frac{vocational}{1 + (total)}$

  vocational + general

$$\begin{aligned}
\hat{p}_{vocational} &= \frac{\exp\left(\hat{\alpha}_{vocational} + \boldsymbol{x}^T \hat{\boldsymbol{\beta}}_{vocational}\right)}{1 + \exp\left(\hat{\alpha}_{general} + \boldsymbol{x}^T \hat{\boldsymbol{\beta}}_{general}\right) + \exp\left(\hat{\alpha}_{vocational} + \boldsymbol{x}^T \hat{\boldsymbol{\beta}}_{vocational}\right)} \\
&= \frac{\exp(5.218 - 0.983 - 0.114 \times 60)}{1 + \exp(2.852 - 1.16 - 0.0579 \times 60) + \exp(5.218 - 0.983 - 0.114 \times 60)} \\
&= 0.06
\end{aligned}$$

HS Program Choice: the multinomial logit model

Note that a different set of ORs can be produced by changing the reference category (here to vocational=3)

```
. mlogit prog i.ses write , base(3) nolog

Multinomial logistic regression                 Number of obs     =        200
                                                LR chi2(6)        =      48.23
                                                Prob > chi2       =     0.0000
Log likelihood = -179.98173                     Pseudo R2         =     0.1182

------------------------------------------------------------------------------
        prog |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
general      |
         ses |
     middle  |  -.8246841   .4901229    -1.68   0.092    -1.785307    .1359392
       high  |  -.1801617    .648455    -0.28   0.781     -1.45111    1.090787
             |
       write |   .0556742   .0233313     2.39   0.017     .0099456    .1014028
       _cons |  -2.366014   1.174248    -2.01   0.044    -4.667498   -.0645293
-------------+----------------------------------------------------------------
academic     |
         ses |
     middle  |  -.2913931   .4763737    -0.61   0.541    -1.225068    .6422822
       high  |   .9826703   .5955669     1.65   0.099    -.1846195     2.14996
             |
       write |   .1136026   .0222199     5.11   0.000     .0700524    .1571528
       _cons |    -5.2182   1.163549    -4.48   0.000    -7.498714   -2.937686
-------------+----------------------------------------------------------------
vocation     |  (base outcome)
------------------------------------------------------------------------------
```

HS Program Choice: changing the reference category in the model

- In the above model, the first set of estimates changed - this is a different outcome - logit for general vs. vocational program choice - i.e., a different 2x2 subtable than earlier
- The second set of estimates is identical except for sign - why? - This is the logit for academic versus vocational - before we estimated the logit of vocational versus academic, i.e., the reciprocal
- General model fit summaries (log likelihood, etc) are identical to the earlier model with a different baseline outcome category

Summary

Multinomial Logit Model

- A straightforward extension of binary logit model (i.e. logit model or logistic regression)
- The multinomial logit model for outcomes with K categories involves $(K-1)$ sets of comparisons of each outcome category to the reference outcome category for calculating sets of parameters (log relative probability or change in log relative prob.). The simultaneous MLE estimation algorithm in the multinomial model can be statistically more efficient than estimating each pair of categories separately.
- These models have history in social science, analysis of multi-way contingency tables (log-linear models for table frequencies)