Ordinal and Multinomial Logistic Regression

A New Example Using Data From *Multilevel Thinking*

Andy Grogan-Kaylor

2023-10-16

Table of contents

1 Background

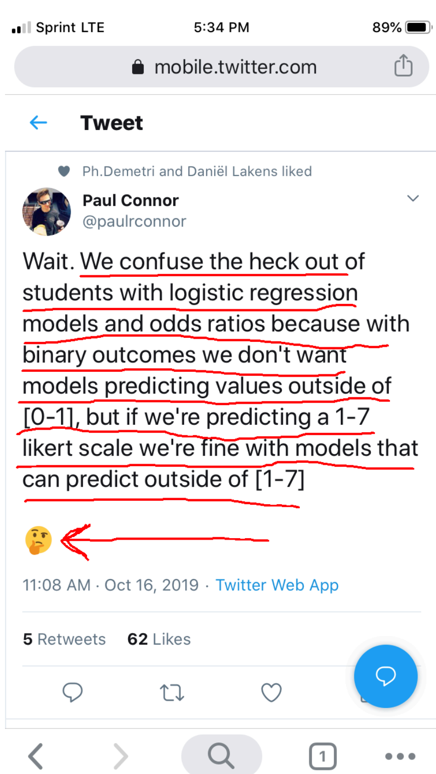

Figure 1: A Tweet

2 The Data

Data are simulated data on parent behaviors and child outcomes from *Multilevel Thinking*.

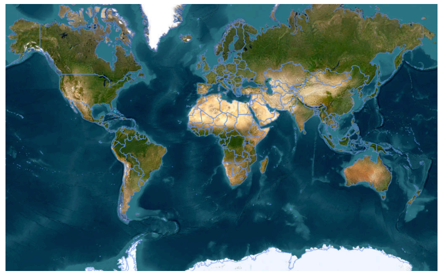

Figure 2: Simulated Data on Countries of the World

```
use "https://github.com/agrogan1/multilevel-thinking/raw/main/simulate-and-analyze-multilevel-data

describe
```

```
note:
      https://github.com/agrogan1/multilevel-thinking/raw/main/simulate-and-a
      > nalyze-multilevel-data/simulated_multilevel_data.dta redirected to
      https://raw.githubusercontent.com/agrogan1/multilevel-thinking/main/sim
      > ulate-and-analyze-multilevel-data/simulated_multilevel_data.dta


Contains data from https://github.com/agrogan1/multilevel-thinking/raw/main/sim
> ulate-and-analyze-multilevel-data/simulated_multilevel_data.dta
 Observations:         3,000
    Variables:             8                  21 Apr 2023 12:38
-------------------------------------------------------------------------------
Variable      Storage   Display    Value
    name         type    format    label      Variable label
-------------------------------------------------------------------------------
country         float   %9.0g                 country id
HDI             float   %9.0g                 Human Development Index
family          float   %9.0g                 family id
id              str7    %9s                   unique country family id
group           float   %9.0g                 arbitrary group variable
physical_puni~t float   %9.0g                 physical punishment in past week
warmth          float   %9.0g                 parental warmth in past week
outcome         float   %9.0g                 beneficial outcome
-------------------------------------------------------------------------------
Sorted by: country  family
```

3 Setup

We need to create a categorical outcome variable for demonstration purposes.

```
* create an outcome_group variable

egen outcome_group = cut(outcome), group(3) // divide outcome into groups

label define outcome_group 0 "low" 1 "medium" 2 "high" // define value labels

label values outcome_group outcome_group // attach value labels

tabulate outcome_group
```

```
Running C:\Users\agrogan\Desktop\GitHub\newstuff\categorical\ordinal-multinomia
> l-logistic-regression-2\profile.do .



outcome_gro |
         up |      Freq.     Percent        Cum.
------------+-----------------------------------
        low |      1,000       33.33       33.33
     medium |      1,000       33.33       66.67
       high |      1,000       33.33      100.00
------------+-----------------------------------
      Total |      3,000      100.00
```

4 Ordinal Logistic Regression

$$\ln \left(\frac{p(y \leq k)}{p(y > k)} \right) = \beta_0 + \beta_1 x_1 + ...$$

Because the data are clustered by countries, we will use the , `cluster(country)` option in each model. The `brant` command can be installed by typing `findit brant`, and installing the Long & Freese `spost` utilities.

```
ologit outcome_group physical_punishment warmth HDI i.group, or cluster(country) // ordinal logit

brant // brant test

margins, at(warmth = (1(1)7)) // margins at different values of warmth

marginsplot, title("Predicted Probabilities From Ordinal Logit") ///
plot(_outcome, labels("low" "medium" "high")) // graph w/ manual labels

graph export myologit.png, replace
```

```
Running C:\Users\agrogan\Desktop\GitHub\newstuff\categorical\ordinal-multinomia
> l-logistic-regression-2\profile.do .

Iteration 0:   Log pseudolikelihood = -3295.8369
```

```
Iteration 1:  Log pseudolikelihood = -3157.4676
Iteration 2:  Log pseudolikelihood = -3157.0335
Iteration 3:  Log pseudolikelihood = -3157.0333

Ordered logistic regression                             Number of obs =  3,000
                                                        Wald chi2(4)  = 242.78
                                                        Prob > chi2   = 0.0000
Log pseudolikelihood = -3157.0333                       Pseudo R2     = 0.0421

                               (Std. err. adjusted for 30 clusters in country)
------------------------------------------------------------------------------
             |               Robust
outcome_gr~p | Odds ratio   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
physical_p~t |   .7962002   .0197074    -9.21   0.000     .7584964    .8357781
      warmth |   1.282995    .026044    12.28   0.000     1.232951    1.335069
         HDI |    1.00389   .0058436     0.67   0.505     .9925017    1.015409
     2.group |   1.322192   .0754851     4.89   0.000     1.182221    1.478735
-------------+----------------------------------------------------------------
       /cut1 |    -.04647   .4096606                       -.84939    .7564499
       /cut2 |   1.446814    .426558                       .610776    2.282853
------------------------------------------------------------------------------
Note: Estimates are transformed only in the first equation to odds ratios.


Brant test of parallel regression assumption

                      |       chi2     p>chi2      df
 ---------------------+------------------------------
                  All |       1.98      0.739       4
 ---------------------+------------------------------
  physical_punishment |       1.45      0.229       1
               warmth |       0.20      0.656       1
                  HDI |       0.05      0.818       1
              2.group |       0.18      0.672       1

A significant test statistic provides evidence that the parallel
regression assumption has been violated.


Predictive margins                                       Number of obs = 3,000
Model VCE: Robust

1._predict: Pr(outcome_group==0), predict(pr outcome(0))
2._predict: Pr(outcome_group==1), predict(pr outcome(1))
3._predict: Pr(outcome_group==2), predict(pr outcome(2))

1._at: warmth = 1
2._at: warmth = 2
3._at: warmth = 3
4._at: warmth = 4
5._at: warmth = 5
6._at: warmth = 6
```

```
7._at: warmth = 7

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
_predict#_at |
        1 1  |   .4715116   .0239632    19.68   0.000     .4245446    .5184785
        1 2  |    .411902   .0219914    18.73   0.000     .3687996    .4550044
        1 3  |   .3547047   .0204707    17.33   0.000     .3145829    .3948265
        1 4  |   .3012864   .0194346    15.50   0.000     .2631953    .3393776
        1 5  |   .2526558   .0187163    13.50   0.000     .2159724    .2893391
        1 6  |   .2094156   .0180743    11.59   0.000     .1739907    .2448405
        1 7  |   .1717793   .0173168     9.92   0.000      .137839    .2057196
        2 1  |   .3210415   .0100789    31.85   0.000     .3012872    .3407958
        2 2  |   .3376888   .0091914    36.74   0.000     .3196739    .3557037
        2 3  |   .3465153   .0092644    37.40   0.000     .3283575    .3646731
        2 4  |   .3467361    .010075    34.42   0.000     .3269895    .3664827
        2 5  |   .3383307   .0114619    29.52   0.000     .3158658    .3607955
        2 6  |   .3220464   .0133672    24.09   0.000     .2958472    .3482456
        2 7  |   .2992734   .0156422    19.13   0.000     .2686153    .3299314
        3 1  |    .207447   .0183764    11.29   0.000     .1714298    .2434641
        3 2  |   .2504092   .0196723    12.73   0.000     .2118522    .2889661
        3 3  |     .29878    .021223    14.08   0.000     .2571838    .3403763
        3 4  |   .3519775   .0231631    15.20   0.000     .3065787    .3973762
        3 5  |   .4090136   .0255026    16.04   0.000     .3590294    .4589977
        3 6  |    .468538   .0280772    16.69   0.000     .4135078    .5235682
        3 7  |   .5289473   .0305829    17.30   0.000      .469006    .5888886
------------------------------------------------------------------------------

Variables that uniquely identify margins: warmth

file myologit.png saved as PNG format
```

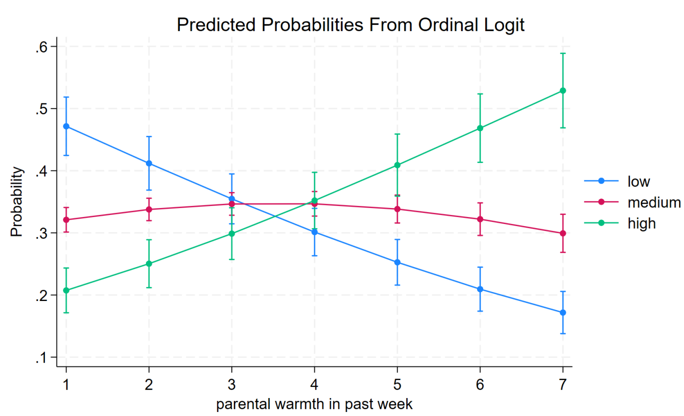

Figure 3: `marginsplot` from `ologit`

5 Multinomial Logistic Regression

$$\ln\left(\frac{P(y=y_2)}{P(y=y_1)}\right) = \ln\left(\frac{P(y=\text{something else})}{P(y=\text{something})}\right)$$

$$= \beta_0 + \beta_1 x_1 + ...$$

$$\ln\left(\frac{P(y=y_3)}{P(y=y_1)}\right) = \ln\left(\frac{P(y=\text{something else altogether})}{P(y=\text{something})}\right)$$

$$= \beta_0 + \beta_1 x_1 + ...$$

Because the *Brant* test was insignificant, the results below are likely to look similar. Imagine, however, if the *Brant* test were statistically significant, suggesting that we should estimate separate regression coefficients for each value of the outcome. Imagine, in addition, if we were estimating an outcome that were truly multinomial in nature, such as *post-secondary* education pursued: *none*, *vocational*, *university*. For heuristic purposes, we will relabel the outcome accordingly.

```
label define outcome_group2 0 "none" 1 "vocational" 2 "university" // define value labels

label values outcome_group outcome_group2 // attach NEW value labels

tabulate outcome_group

mlogit outcome_group physical_punishment warmth HDI i.group, rr cluster(country)

margins, at(warmth = (1(1)7)) // margins at different values of warmth

marginsplot, title("Predicted Probabilities From Multinomial Logit") ///
plot(_outcome, labels("none" "vocational" "university")) // graph w/ manual labels

graph export mymlogit.png, replace
```

```
Running C:\Users\agrogan\Desktop\GitHub\newstuff\categorical\ordinal-multinomia
> l-logistic-regression-2\profile.do .



outcome_gro |
         up |      Freq.     Percent        Cum.
------------+-----------------------------------
       none |      1,000       33.33       33.33
 vocational |      1,000       33.33       66.67
 university |      1,000       33.33      100.00
------------+-----------------------------------
      Total |      3,000      100.00


Iteration 0:   Log pseudolikelihood = -3295.8369
Iteration 1:   Log pseudolikelihood = -3159.3121
```

```
Iteration 2:  Log pseudolikelihood = -3157.2541
Iteration 3:  Log pseudolikelihood = -3157.2532
Iteration 4:  Log pseudolikelihood = -3157.2532

Multinomial logistic regression                         Number of obs =  3,000
                                                        Wald chi2(8)  = 216.92
                                                        Prob > chi2   = 0.0000
Log pseudolikelihood = -3157.2532                       Pseudo R2     = 0.0420

                               (Std. err. adjusted for 30 clusters in country)
------------------------------------------------------------------------------
             |               Robust
outcome_gr~p |        RRR   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
none         |  (base outcome)
-------------+----------------------------------------------------------------
vocational   |
physical_p~t |   .8284144   .0268834    -5.80   0.000     .7773647    .8828166
      warmth |   1.172042   .0323704     5.75   0.000     1.110284    1.237235
         HDI |   1.003045   .0039244     0.78   0.437     .9953822    1.010766
     2.group |   1.244621   .1034633     2.63   0.008     1.057495     1.46486
       _cons |   .7248303   .2045156    -1.14   0.254     .4169312     1.26011
-------------+----------------------------------------------------------------
university   |
physical_p~t |    .733425   .0260105    -8.74   0.000     .6841767    .7862183
      warmth |   1.402776   .0404291    11.74   0.000     1.325733    1.484296
         HDI |   1.005061   .0080327     0.63   0.528     .9894402    1.020929
     2.group |   1.454744   .1119325     4.87   0.000     1.251102    1.691534
       _cons |   .3950266    .227379    -1.61   0.107     .1278413    1.220623
------------------------------------------------------------------------------
Note: _cons estimates baseline relative risk for each outcome.


Predictive margins                                       Number of obs = 3,000
Model VCE: Robust

1._predict: Pr(outcome_group==none), predict(pr outcome(0))
2._predict: Pr(outcome_group==vocational), predict(pr outcome(1))
3._predict: Pr(outcome_group==university), predict(pr outcome(2))

1._at: warmth = 1
2._at: warmth = 2
3._at: warmth = 3
4._at: warmth = 4
5._at: warmth = 5
6._at: warmth = 6
7._at: warmth = 7

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
_predict#_at |
```

```
        1 1  |   .4655491   .0256453    18.15   0.000     .4152852     .515813
        1 2  |   .4108856   .0225268    18.24   0.000     .3667338    .4550374
        1 3  |   .3566849    .020455    17.44   0.000     .3165938    .3967761
        1 4  |   .3043247   .0194768    15.62   0.000     .2661507    .3424986
        1 5  |   .2551027   .0192162    13.28   0.000     .2174397    .2927657
        1 6  |    .210102   .0191257    10.99   0.000     .1726162    .2475877
        1 7  |    .170087   .0187808     9.06   0.000     .1332774    .2068966
        2 1  |   .3312655   .0149681    22.13   0.000     .3019286    .3606025
        2 2  |   .3403628    .010943    31.10   0.000      .318915    .3618106
        2 3  |   .3438888   .0090929    37.82   0.000     .3260671    .3617104
        2 4  |   .3414688    .010569    32.31   0.000     .3207539    .3621838
        2 5  |   .3331582    .014179    23.50   0.000     .3053679    .3609485
        2 6  |   .3194468   .0184628    17.30   0.000     .2832603    .3556333
        2 7  |    .301194   .0227261    13.25   0.000     .2566517    .3457363
        3 1  |   .2031854   .0183179    11.09   0.000     .1672829    .2390879
        3 2  |   .2487516   .0194812    12.77   0.000     .2105691    .2869341
        3 3  |   .2994263   .0210267    14.24   0.000     .2582148    .3406379
        3 4  |   .3542065   .0231943    15.27   0.000     .3087464    .3996666
        3 5  |   .4117391   .0260214    15.82   0.000     .3607381    .4627401
        3 6  |   .4704512   .0292975    16.06   0.000     .4130291    .5278733
        3 7  |    .528719   .0326555    16.19   0.000     .4647153    .5927227
------------------------------------------------------------------------------


Variables that uniquely identify margins: warmth

file mymlogit.png saved as PNG format
```

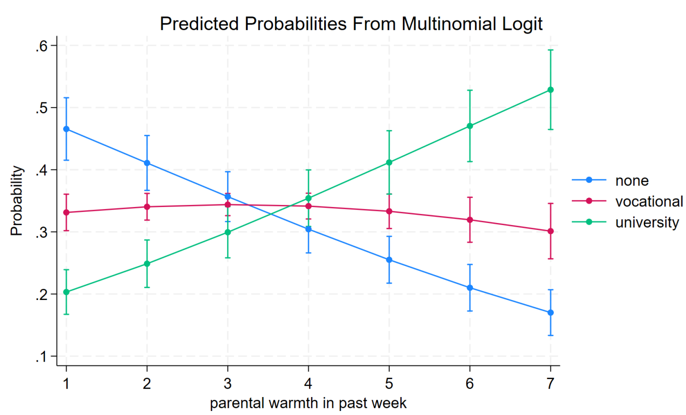

Figure 4: `marginsplot` from `mlogit`