Cutpoints in Ordered Logistic Regression

Andy Grogan-Kaylor

7 Oct 2020

Background

Cutpoints in ordered logistic regression are not terrifically substantively informative, but they do contain statistical information.

This handout draws heavily on the Stata documentation for `ologit`.

Ordered Logistic Regression

Setup

```
. clear all // clear the workspace
```

Get The Data

```
. use http://www.stata-press.com/data/r15/fullauto // use auto data set from Stata documenta
> tion
(Automobile Models)
```

Codebook

```
. codebook rep77 foreign // codebook
```

rep77			Repair Record 1977

```
              type:  numeric (int)
             label:  repair

             range:  [1,5]                        units:  1
     unique values:  5                    missing .:  8/74

       tabulation:  Freq.   Numeric  Label
                       3         1  Poor
                      11         2  Fair
                      27         3  Average
                      20         4  Good
                       5         5  Excellent
                       8         .
```

foreign			Foreign

```
              type:  numeric (int)
             label:  foreign
```

```
                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  0/74

            tabulation:  Freq.   Numeric  Label
                            52         0  Domestic
                            22         1  Foreign
```

Run The Model

```
. ologit rep77 foreign // estimate ordered logistic regression

Iteration 0:   log likelihood = -89.895098
Iteration 1:   log likelihood = -85.951765
Iteration 2:   log likelihood = -85.908227
Iteration 3:   log likelihood = -85.908161
Iteration 4:   log likelihood = -85.908161

Ordered logistic regression                     Number of obs   =         66
                                                LR chi2(1)      =       7.97
                                                Prob > chi2     =     0.0047
Log likelihood = -85.908161                     Pseudo R2       =     0.0444

      rep77 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]

    foreign |   1.455878   .5308951     2.74   0.006     .4153425    2.496413

      /cut1 |  -2.765562   .5988208                     -3.939229   -1.591895
      /cut2 |  -.9963603   .3217706                     -1.627019   -.3657016
      /cut3 |   .9426153   .3136398                      .3278925    1.557338
      /cut4 |   3.123351   .5423257                      2.060412     4.18629

. predict yhat* // predicted probabilities for different levels of dv
(option pr assumed; predicted probabilities)

. tabstat yhat1 yhat2 yhat3 yhat4 yhat5, by(foreign) // table of predicted probabilities

Summary statistics: mean
  by categories of: foreign (Foreign)
 foreign |     yhat1      yhat2      yhat3      yhat4      yhat5

Domestic |  .0592137   .2104439     .44997   .2382181   .0421543
 Foreign |  .0144652   .0648099    .295154   .4668096   .1587614

   Total |  .0459101   .1671473   .4039436   .3061777   .0768213
```

Calculations

We can use the cutpoints as another way of calculating these probabilities, with the logistic formula $1/(1+e^{u_j})$

For example, the Stata documentation notes that

> "For a foreign car, the probability of a poor record is the probability that $1.46 + u_j <= -2.77$, or equivalently, $u_j <= -4.23$. Making this calculation requires familiarity with the logistic distribution: the probability is $1/(1 + e^{4.23}) = 0.014$. On the other hand, for domestic cars, the probability of a poor record is the probability $u_j <= -2.77$, which is 0.059 $[1/(1 + e^{2.77})]$."