WWS 509 Generalized Linear Models: Precept 8
Multinomial and Sequential Logit Models

Kristin E. Bietsch

Office of Population Research, Princeton University

November 2012

Introducing the Data

This data comes from Germán's website. We are looking at how people rate their level of satisfaction in their housing. The options are high, medium, and low. Since there are more than 2 options, we can use multinomial logistic regression to predict responses. Controls we will consider include contact with neighbors (high or low), type of housing (tower, apartment, atrium, or terrace), and influence (on a scale of 0-2).

Multinomial Logit Model

- There are J-1 equations. One of the outcomes will serve as the baseline.
- Like the logit model, the log-odds is a linear function of the predictor.
- It makes no difference which category we pick as the reference cell, because we can convert from one formulation to another
- $log(\frac{\pi_{i1}}{\pi_{i2}}) = log(\frac{\pi_{i1}}{\pi_{i3}}) - log(\frac{\pi_{i2}}{\pi_{i3}})$

Sequential Logit Model

- Also known as the hierarchical logit model
- We are now looking at nested comparisions. For example, I have to be ok with my housing not to rate my satisfaction level as "low," and then I choose between medium or high.
  - Can you think of other data which might benefit from this model?

Interpretation

General

1. Why did I change contact and influence with "-1"?
2. Which housing type is the reference category? Where besides the logit output can you find this?

Multinomial Logit

1. Looking at the model only including housing type, what seems to be the best housing? The worst?
2. Does this first model fit the data?
3. Is influence, net of housing, predictive of satisfaction?
4. What assumption am I making my modeling influence as continuous and not categorical?
5. What is the odds ratio for a person with an influence level of 2 of having a satisfaction level of medium compared to low in this model? Of high compared to low?
6. What about in the model that also controls for neighbor contact?
7. What about contact with neighbors?
8. Does the model with neighbors fit the data? How do you know?

Sequential Logit

1. Do you see anything interesting here that we did not see in the multinomial model?
2. Which model do you prefer for this data: multinomial or sequential?

Appendices

Stata Output

```
(Housing Conditions in Copenhagen)

. quietly tab housing, gen(house_type)

. rename house_type1 tower
```

```
. rename house_type2 apartment

. rename house_type3 atrium

. rename house_type4 terrace

. local house apartment atrium terrace

. replace influence=influence-1
(72 real changes made)

. quietly tab influence, gen(influ_type)

. rename influ_type1 influ_low

. rename influ_type2 influ_med

. rename influ_type3 influ_high

. local influ influ_med influ_high

. gen con_high=contact-1

. *** Multinomial ***
. *Saturated
. quietly mlogit  satisfaction `house' `influ' con_high[fw=n]

. estimates store sat

. scalar ll_sat = e(ll)

. *Linear
. mlogit  satisfaction `house' [fw=n], baseoutcome(1)

Iteration 0:   log likelihood = -1824.4388
Iteration 1:   log likelihood = -1794.3345
Iteration 2:   log likelihood = -1794.1045
Iteration 3:   log likelihood = -1794.1044

Multinomial logistic regression                   Number of obs   =       1681
                                                  LR chi2(6)      =      60.67
                                                  Prob > chi2     =     0.0000
Log likelihood = -1794.1044                       Pseudo R2       =     0.0166

------------------------------------------------------------------------------
satisfaction |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
```

```
-------------+----------------------------------------------------------------
Low          |  (base outcome)
-------------+----------------------------------------------------------------
Medium       |
   apartment |  -.3646241   .1700011    -2.14   0.032    -.6978201   -.0314281
      atrium |   .1905641    .219739     0.87   0.386    -.2401164    .6212446
     terrace |  -.6062843   .2025693    -2.99   0.003    -1.003313   -.2092559
       _cons |   .0200007   .1414284     0.14   0.888    -.2571939    .2971953
-------------+----------------------------------------------------------------
High         |
   apartment |  -.5948893   .1486684    -4.00   0.000    -.8862741   -.3035045
      atrium |  -.2977324   .2028366    -1.47   0.142    -.6952848      .09982
     terrace |  -1.345052   .1921081    -7.00   0.000    -1.721577   -.9685266
       _cons |   .7031975   .1228862     5.72   0.000     .4623451      .94405
------------------------------------------------------------------------------

. lrtest . sat

Likelihood-ratio test                                 LR chi2(6)  =    118.13
(Assumption: . nested in sat)                         Prob > chi2 =    0.0000

. mlogit  satisfaction ‘house’  influence[fw=n], baseoutcome(1)

Iteration 0:   log likelihood = -1824.4388
Iteration 1:   log likelihood = -1745.0131
Iteration 2:   log likelihood = -1744.0996
Iteration 3:   log likelihood = -1744.0991
Iteration 4:   log likelihood = -1744.0991

Multinomial logistic regression                   Number of obs   =       1681
                                                  LR chi2(8)      =     160.68
                                                  Prob > chi2     =     0.0000
Log likelihood = -1744.0991                       Pseudo R2       =     0.0440

------------------------------------------------------------------------------
satisfaction |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Low          |  (base outcome)
-------------+----------------------------------------------------------------
Medium       |
   apartment |  -.3878463     .17075    -2.27   0.023    -.7225103   -.0531824
      atrium |    .193475   .2205748     0.88   0.380    -.2388437    .6257936
     terrace |  -.5894198   .2033776    -2.90   0.004    -.9880327    -.190807
   influence |   .3207971   .0871399     3.68   0.000      .150006    .4915883
       _cons |  -.2062289   .1547364    -1.33   0.183    -.5095066    .0970489
-------------+----------------------------------------------------------------
```

```
High         |
   apartment |  -.6609661   .1531665    -4.32   0.000     -.961167   -.3607652
      atrium |  -.3039443   .2084978    -1.46   0.145    -.7125925    .1047038
     terrace |  -1.306371   .1972983    -6.62   0.000    -1.693068    -.919673
   influence |   .7736902   .0807801     9.58   0.000      .615364    .9320164
       _cons |   .0654819   .1405259     0.47   0.641    -.2099437    .3409076
------------------------------------------------------------------------------

. lrtest . sat

Likelihood-ratio test                                 LR chi2(4)  =     18.11
(Assumption: . nested in sat)                         Prob > chi2 =    0.0012

. mlogit  satisfaction `house' influence con_high[fw=n], baseoutcome(1)

Iteration 0:   log likelihood = -1824.4388
Iteration 1:   log likelihood =  -1737.219
Iteration 2:   log likelihood = -1736.1067
Iteration 3:   log likelihood = -1736.1059
Iteration 4:   log likelihood = -1736.1059

Multinomial logistic regression                   Number of obs   =       1681
                                                  LR chi2(10)     =     176.67
                                                  Prob > chi2     =     0.0000
Log likelihood = -1736.1059                       Pseudo R2       =     0.0484

------------------------------------------------------------------------------
satisfaction |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Low          |  (base outcome)
-------------+----------------------------------------------------------------
Medium       |
   apartment |  -.4413551   .1723525    -2.56   0.010    -.7791598   -.1035504
      atrium |   .1216597   .2227338     0.55   0.585    -.3148905    .5582099
     terrace |  -.6713139   .2061066    -3.26   0.001    -1.075275   -.2673525
   influence |   .3487541   .0879542     3.97   0.000     .1763671    .5211411
    con_high |   .3635188   .1323484     2.75   0.006     .1041207    .6229169
       _cons |  -.3823228     .16806    -2.27   0.023    -.7117143   -.0529313
-------------+----------------------------------------------------------------
High         |
   apartment |   -.732001   .1551883    -4.72   0.000    -1.036164   -.4278376
      atrium |  -.4014077   .2112742    -1.90   0.057    -.8154975     .012682
     terrace |  -1.409386   .2000943    -7.04   0.000    -1.801564   -1.017208
   influence |   .8103557   .0818756     9.90   0.000     .6498824    .9708289
    con_high |   .4794532   .1240436     3.87   0.000     .2363322    .7225742
       _cons |  -.1732837   .1540446    -1.12   0.261    -.4752055    .1286381
```

```
------------------------------------------------------------------------------

. lrtest . sat

Likelihood-ratio test                                 LR chi2(2)  =      2.13
(Assumption: . nested in sat)                         Prob > chi2 =    0.3451

. *** Sequential Model ***
. gen ok = satisfaction > 1

. logit ok ‘house’ influence con_high[fw=n]

Iteration 0:   log likelihood = -1074.5419
Iteration 1:   log likelihood = -1010.5894
Iteration 2:   log likelihood = -1009.8002
Iteration 3:   log likelihood = -1009.7996
Iteration 4:   log likelihood = -1009.7996

Logistic regression                               Number of obs   =       1681
                                                  LR chi2(5)      =     129.48
                                                  Prob > chi2     =     0.0000
Log likelihood = -1009.7996                       Pseudo R2       =     0.0603

------------------------------------------------------------------------------
          ok |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   apartment |  -.6154755   .1427886    -4.31   0.000    -.8953361   -.3356149
      atrium |  -.1789057   .1921456    -0.93   0.352    -.5555043    .1976928
     terrace |  -1.080569   .1729336    -6.25   0.000    -1.419512    -.741625
   influence |   .6177193   .0734991     8.40   0.000     .4736638    .7617748
    con_high |   .4297313   .1105171     3.89   0.000     .2131217    .6463409
       _cons |   .4386837   .1394585     3.15   0.002       .16535    .7120174
------------------------------------------------------------------------------

. scalar ll_ok = e(ll)

. gen high = satisfaction==3

. logit high ‘house’ influence con_high[fw=n] if ok

Iteration 0:   log likelihood = -749.89688
Iteration 1:   log likelihood = -725.90151
Iteration 2:   log likelihood = -725.84153
Iteration 3:   log likelihood = -725.84153

Logistic regression                               Number of obs   =       1114
```

```
                                                  LR chi2(5)      =      48.11
                                                  Prob > chi2     =     0.0000
Log likelihood = -725.84153                       Pseudo R2       =     0.0321

------------------------------------------------------------------------------
        high |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   apartment |  -.3252394    .1571975    -2.07   0.039    -.6333408    -.017138
      atrium |   -.526402    .1997125    -2.64   0.008    -.9178313   -.1349727
     terrace |  -.7737943     .210587    -3.67   0.000    -1.186537   -.3610513
   influence |   .4712369    .0832203     5.66   0.000     .3081281    .6343457
    con_high |   .0878219    .1294341     0.68   0.497    -.1658643    .3415081
       _cons |   .2374671    .1574473     1.51   0.131    -.0711239     .546058
------------------------------------------------------------------------------

. scalar ll_high = e(ll)

. * Overall log-likelihood
. di ll_ok + ll_high
-1735.6411
```