WWS 509 Generalized Linear Models: Precept 4 Section 3.5

Kristin E. Bietsch

Office of Population Research, Princeton University

October 2012

1 Introducing the Data

Today we will be looking at the same data as last week, women giving birth after HIV testing. For today, I have created group data, based on age groups (15-19, 20-24, 25-29, 30-34, 35-39, 40-44) and HIV status.

2 Deviance

Deviance is the goodness of fit likelihood ratio chi-square statistic

- With a degree of freedom determined from the number of parameters added to the model
- To see if a model fits, you want to look at the deviance and check the χ^2 distribution.

Regardless if the group or individual data, the estimated effects, the standard errors, and likelihood ratio test based on differences between deviances will be the same (but the deviances will change).

3 Categorical Age and Binary HIV status

There are 5 potential models from this data, what are they?

Model	Notation	$\text{logit}(\pi_{ij})$	Deviance	D.F.

Using the above table:

1. How would I test the gross effect of age?
    (a) Test if the model including only age fits the data.
2. How would I test the gross effect of HIV?
    (a) Test if the model including only HIV fits the data.
3. Which variable, HIV or age, explains more (only looking at the single variable models)?
    (a) Can you test this?

Now look at the additive model in the table:

1. What can this model tell you?
    (a) What is the effect of adding age to the model containing HIV status?
    (b) What is the effect of adding HIV to the model containing age?
2. Does this model fit the data?

Now finally, look at the saturated model in the table:

1. How much deviance is in this model?
2. Is this model useful for interpretation?

4 Linear Age and Binary HIV status

Now age as a categorical variable uses up many of our degrees of freedom, so lets see if we can model things a bit differently. Refer to the output in Appendix

B. Fill in the deviance table for all the models in Appendix B (and don't forget to include the null model from Appendix A).

Model	Notation	logit(π_{ij})	Deviance	D.F.

1. What do you think about modeling age linearly?
    - (a) Is it an improvement over the null?
    - (b) Does this model fit the data?

I added HIV into the model:

1. Is this model an improvement over the previous model?
    - (a) Does this model fit the data?

Lets now try an interaction term?

1. Is the interaction term significant (look at the output)?
    - (a) Does this model offer an improvement over the additive model?

OK, linear effect of age and childbearing, not a good idea. But this doesn't mean we have to return to our categorical variables. I add a quadratic term for age.

1. How did I create this term in Stata?
2. What do you think about this model?
    - (a) Does this model offer an improvement over the model of age as a linear predictor?

For the last regression, I model age, age squared, and HIV status.

1. Does this model offer an improvement over the model of age and age squared?
2. Does this model fit the data?

Appendices

Appendix A: Grouped Factor Data Stata Output

```
. gen chi_dum=.
(3307 missing values generated)

. replace chi_dum=1 if child_born!=0
(1114 real changes made)

. replace chi_dum=0 if child_born==0
(2193 real changes made)

. tab age_range, gen(age_group)

  age_range |      Freq.     Percent        Cum.
------------+-----------------------------------
       17.5 |        817       24.71       24.71
       22.5 |        582       17.60       42.30
       27.5 |        569       17.21       59.51
       32.5 |        494       14.94       74.45
       37.5 |        396       11.97       86.42
       42.5 |        449       13.58      100.00
------------+-----------------------------------
      Total |      3,307      100.00

. gen n=1

. collapse (sum) chi_dum n, by (age_range hiv5)

. glm chi_dum, family(binomial n)

Iteration 0:   log likelihood = -216.53224
Iteration 1:   log likelihood = -212.84368
Iteration 2:   log likelihood =  -212.8429
Iteration 3:   log likelihood =  -212.8429

Generalized linear models                          No. of obs      =        12
Optimization     : ML                              Residual df     =        11
                                                   Scale parameter =         1
Deviance         =  366.1140405                    (1/df) Deviance =  33.28309
Pearson          =  349.6392631                    (1/df) Pearson  =  31.78539

Variance function: V(u) = u*(1-u/n)                [Binomial]
Link function    : g(u) = ln(u/(n-u))              [Logit]
```

```
                                                  AIC             =  35.64048
Log likelihood   = -212.8429045                   BIC             =  338.7801

------------------------------------------------------------------------------
             |                 OIM
     chi_dum |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |  -.6773133   .0367922   -18.41   0.000    -.7494246    -.605202
------------------------------------------------------------------------------

. glm chi_dum hiv5, family(binomial n)

Iteration 0:   log likelihood = -209.82664
Iteration 1:   log likelihood = -206.18839
Iteration 2:   log likelihood = -206.18768
Iteration 3:   log likelihood = -206.18768

Generalized linear models                         No. of obs      =        12
Optimization     : ML                             Residual df     =        10
                                                  Scale parameter =         1
Deviance         =  352.8035962                   (1/df) Deviance =  35.28036
Pearson          =  334.9688544                   (1/df) Pearson  =  33.49689

Variance function: V(u) = u*(1-u/n)               [Binomial]
Link function    : g(u) = ln(u/(n-u))             [Logit]

                                                  AIC             =  34.69795
Log likelihood   = -206.1876824                   BIC             =  327.9545

------------------------------------------------------------------------------
             |                 OIM
     chi_dum |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        hiv5 |  -.5553665   .1584862    -3.50   0.000    -.8659937   -.2447392
       _cons |   -.641307   .0379678   -16.89   0.000    -.7157225   -.5668916
------------------------------------------------------------------------------

. tab age_range, gen(age_group)

  age_range |      Freq.     Percent        Cum.
------------+-----------------------------------
       17.5 |          2       16.67       16.67
       22.5 |          2       16.67       33.33
       27.5 |          2       16.67       50.00
       32.5 |          2       16.67       66.67
       37.5 |          2       16.67       83.33
       42.5 |          2       16.67      100.00
------------+-----------------------------------
      Total |         12      100.00

. drop age_group1

. glm chi_dum age_group*, family(binomial n)

Iteration 0:   log likelihood = -49.074751
Iteration 1:   log likelihood =   -49.0383
Iteration 2:   log likelihood =   -49.0383

Generalized linear models                          No. of obs      =        12
Optimization     : ML                              Residual df     =         6
                                                   Scale parameter =         1
Deviance         =  38.50483154                    (1/df) Deviance =  6.417472
Pearson          =  36.28561409                    (1/df) Pearson  =  6.047602

Variance function: V(u) = u*(1-u/n)                [Binomial]
Link function    : g(u) = ln(u/(n-u))              [Logit]

                                                   AIC             =   9.17305
Log likelihood   = -49.03830004                    BIC             =  23.59539

------------------------------------------------------------------------------
             |                 OIM
     chi_dum |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  age_group2 |   1.174999   .1260839     9.32   0.000     .9278791    1.422119
  age_group3 |   1.625005   .1255021    12.95   0.000     1.379025    1.870984
  age_group4 |   1.528567   .1297104    11.78   0.000      1.27434    1.782795
  age_group5 |   1.109389   .1393674     7.96   0.000     .8362339    1.382544
  age_group6 |  -.1650128   .1627317    -1.01   0.311    -.4839611    .1539355
       _cons |  -1.593367   .0933763   -17.06   0.000    -1.776382   -1.410353
------------------------------------------------------------------------------

. glm chi_dum age_group* hiv5, family(binomial n)

Iteration 0:   log likelihood = -31.604792
Iteration 1:   log likelihood = -31.585329
Iteration 2:   log likelihood = -31.585328

Generalized linear models                          No. of obs      =        12
Optimization     : ML                              Residual df     =         5
                                                   Scale parameter =         1
```

```
Deviance         =   3.59888798                    (1/df) Deviance =  .7197776
Pearson          =  4.468746092                    (1/df) Pearson  =  .8937492

Variance function: V(u) = u*(1-u/n)                [Binomial]
Link function    : g(u) = ln(u/(n-u))              [Logit]

                                                   AIC             =  6.430888
Log likelihood   = -31.58532826                    BIC             = -8.825645

------------------------------------------------------------------------------
             |                 OIM
     chi_dum |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  age_group2 |   1.210116   .1264414     9.57   0.000     .9622954    1.457937
  age_group3 |   1.705656   .1267314    13.46   0.000     1.457268    1.954045
  age_group4 |   1.624498   .1312985    12.37   0.000     1.367158    1.881838
  age_group5 |   1.201172   .1407213     8.54   0.000     .9253632     1.47698
  age_group6 |  -.1244814   .1630386    -0.76   0.445    -.4440312    .1950684
        hiv5 |  -.9173204   .1638805    -5.60   0.000     -1.23852   -.5961205
       _cons |  -1.585984   .0934129   -16.98   0.000     -1.76907   -1.402898
------------------------------------------------------------------------------

. gen age_hiv2=age_group2*hiv5

. gen age_hiv3=age_group3*hiv5

. gen age_hiv4=age_group4*hiv5

. gen age_hiv5=age_group5*hiv5

. gen age_hiv6=age_group6*hiv5

. glm chi_dum age_group* hiv5 age_hiv*, family(binomial n)

Iteration 0:   log likelihood = -29.786201
Iteration 1:   log likelihood = -29.785884
Iteration 2:   log likelihood = -29.785884

Generalized linear models                          No. of obs      =        12
Optimization     : ML                              Residual df     =         0
                                                   Scale parameter =         1
Deviance         =  1.69961e-14                    (1/df) Deviance =         .
Pearson          =  1.01332e-16                    (1/df) Pearson  =         .

Variance function: V(u) = u*(1-u/n)                [Binomial]
Link function    : g(u) = ln(u/(n-u))              [Logit]
```

```
                                                  AIC             =  6.964314
Log likelihood   = -29.78588427                   BIC             =  1.70e-14

------------------------------------------------------------------------------
             |                 OIM
     chi_dum |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  age_group2 |   1.234313   .1278387     9.66   0.000      .983754    1.484873
  age_group3 |   1.730612   .1291826    13.40   0.000     1.477418    1.983805
  age_group4 |   1.638989   .1343145    12.20   0.000     1.375737     1.90224
  age_group5 |   1.182523   .1444718     8.19   0.000     .8993634    1.465682
  age_group6 |  -.1099815   .1652238    -0.67   0.506    -.4338142    .2138511
        hiv5 |   .3448405   .8072783     0.43   0.669    -1.237396    1.927077
    age_hiv2 |  -1.630209     .94745    -1.72   0.085    -3.487177    .2267589
    age_hiv3 |  -1.418832    .863965    -1.64   0.101    -3.112172    .2745082
    age_hiv4 |  -1.290682   .8622822    -1.50   0.134    -2.980724    .3993601
    age_hiv5 |  -.9483295   .8766014    -1.08   0.279    -2.666437    .7697778
    age_hiv6 |  -1.345306   1.097037    -1.23   0.220    -3.495458    .8048466
       _cons |  -1.597603   .0940268   -16.99   0.000    -1.781893   -1.413314
------------------------------------------------------------------------------
```

Appendix B: Grouped Data with Linear Age Stata Output

```
. glm chi_dum age_range, family(binomial n)

Iteration 0:   log likelihood = -214.93171
Iteration 1:   log likelihood = -210.95632
Iteration 2:   log likelihood = -210.95541
Iteration 3:   log likelihood = -210.95541

Generalized linear models                         No. of obs      =        12
Optimization     : ML                             Residual df     =        10
                                                  Scale parameter =         1
Deviance         =  362.3390561                   (1/df) Deviance =  36.23391
Pearson          =  345.5239143                   (1/df) Pearson  =  34.55239

Variance function: V(u) = u*(1-u/n)               [Binomial]
Link function    : g(u) = ln(u/(n-u))             [Logit]
```

```
                                                  AIC             =  35.49257
Log likelihood   = -210.9554123                   BIC             =    337.49

------------------------------------------------------------------------------
             |                 OIM
     chi_dum |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   age_range |   .0082429   .0042403     1.94   0.052     -.000068    .0165538
       _cons |  -.9100142    .125642    -7.24   0.000    -1.156268   -.6637604
------------------------------------------------------------------------------

. glm chi_dum age_range hiv5, family(binomial n)

Iteration 0:   log likelihood = -207.24585
Iteration 1:   log likelihood = -203.37119
Iteration 2:   log likelihood = -203.37042
Iteration 3:   log likelihood = -203.37042

Generalized linear models                         No. of obs      =        12
Optimization     : ML                             Residual df     =         9
                                                  Scale parameter =         1
Deviance         =  347.1690663                   (1/df) Deviance =  38.57434
Pearson          =  328.9774173                   (1/df) Pearson  =  36.55305

Variance function: V(u) = u*(1-u/n)               [Binomial]
Link function    : g(u) = ln(u/(n-u))             [Logit]

                                                  AIC             =  34.39507
Log likelihood   = -203.3704174                   BIC             =  324.8049

------------------------------------------------------------------------------
             |                 OIM
     chi_dum |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   age_range |   .0101268   .0042639     2.38   0.018     .0017698    .0184838
        hiv5 |  -.5955005   .1594579    -3.73   0.000    -.9080323   -.2829687
       _cons |  -.9245035   .1256216    -7.36   0.000    -1.170717   -.6782897
------------------------------------------------------------------------------

. gen range_hiv=age_range*hiv5

. glm chi_dum age_range hiv5 range_hiv, family(binomial n)

Iteration 0:   log likelihood = -206.39388
Iteration 1:   log likelihood = -202.52284
Iteration 2:   log likelihood =   -202.522
Iteration 3:   log likelihood =   -202.522

Generalized linear models                          No. of obs      =        12
Optimization     : ML                              Residual df     =         8
                                                   Scale parameter =         1
Deviance         =   345.472226                    (1/df) Deviance =  43.18403
Pearson          =  327.9595101                    (1/df) Pearson  =  40.99494

Variance function: V(u) = u*(1-u/n)                [Binomial]
Link function    : g(u) = ln(u/(n-u))              [Logit]

                                                   AIC             =  34.42033
Log likelihood   = -202.5219973                    BIC             =   325.593

------------------------------------------------------------------------------
             |                 OIM
     chi_dum |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   age_range |   .0111866   .0043401     2.58   0.010     .0026801     .019693
        hiv5 |    .355974   .7404394     0.48   0.631    -1.095261    1.807209
   range_hiv |  -.0301394   .0231563    -1.30   0.193     -.075525    .0152462
       _cons |  -.9542751   .1277653    -7.47   0.000     -1.20469   -.7038597
------------------------------------------------------------------------------

. gen age2=(age_range-27.5)^2

. glm chi_dum age_range age2, family(binomial n)

Iteration 0:   log likelihood = -50.210447
Iteration 1:   log likelihood = -50.174929
Iteration 2:   log likelihood = -50.174929

Generalized linear models                          No. of obs      =        12
Optimization     : ML                              Residual df     =         9
                                                   Scale parameter =         1
Deviance         =  40.77808903                    (1/df) Deviance =  4.530899
Pearson          =  38.68819948                    (1/df) Pearson  =  4.298689

Variance function: V(u) = u*(1-u/n)                [Binomial]
Link function    : g(u) = ln(u/(n-u))              [Logit]

                                                   AIC             =  8.862488
Log likelihood   = -50.17492879                    BIC             =  18.41393

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

```
             |                 OIM
     chi_dum |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   age_range |   .0492848   .0054826     8.99   0.000     .0385391    .0600305
        age2 |  -.0111378   .0006621   -16.82   0.000    -.0124355   -.0098401
       _cons |  -1.310733   .1487056    -8.81   0.000     -1.60219   -1.019275
------------------------------------------------------------------------------

. glm chi_dum age_range age2 hiv5, family(binomial n)

Iteration 0:   log likelihood = -32.684431
Iteration 1:   log likelihood = -32.661696
Iteration 2:   log likelihood = -32.661695

Generalized linear models                          No. of obs      =        12
Optimization     : ML                              Residual df     =         8
                                                   Scale parameter =         1
Deviance         =  5.751621084                    (1/df) Deviance =  .7189526
Pearson          =  6.656666375                    (1/df) Pearson  =  .8320833

Variance function: V(u) = u*(1-u/n)                [Binomial]
Link function    : g(u) = ln(u/(n-u))              [Logit]

                                                   AIC             =  6.110282
Log likelihood   = -32.66169482                    BIC             = -14.12763

------------------------------------------------------------------------------
             |                 OIM
     chi_dum |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   age_range |   .0538285   .0055535     9.69   0.000     .0429439    .0647132
        age2 |  -.0115719   .0006689   -17.30   0.000     -.012883   -.0102608
        hiv5 |  -.9203951   .1641225    -5.61   0.000    -1.242069    -.598721
       _cons |  -1.347511    .149278    -9.03   0.000     -1.64009   -1.054931
------------------------------------------------------------------------------
```