WWS 509 Generalized Linear Models: Precept 7
Poisson Models with Extra Variation

Kristin E. Bietsch

Office of Population Research, Princeton University

November 2012

Introducing the Data

This data looks at number of volunteer experiences in the past year. Control variables include race, gender, education, and income. Do not cite!

Extra Poisson Variation

1. **Over dispersion-** variance larger than the ______________
2. Poisson estimates are consistent when the variance is ______________ (not just equal) to the mean
3. Regular Poisson standard errors tend to be ______________ in the presence of over-dispersion
4. Correction can be done by estimating the ϕ using ______________ divided by its degrees of freedom $\hat{\phi} = \frac{\chi_p^2}{n-p}$
5. We are assuming that ______________ is actually extra Poisson variation

Negative Binomial Regression

1. Starts with a standard Poisson model and adds a ______________ θ_i to represent unobserved heterogeneity
2. Negative binomial distribution is best known as the distribution of the number of failures before k successes of ______________ trials

Zero Inflated Poisson

1. Sometimes in count data, you have many more zeros than the model ________________
2. Zero Inflated Poisson is divided into two analyses: finding the ________________ and those that follow the Poisson distribution.
3. You need to tell Stata which ________________ to use to predict the count and the "always zeros"
   (a) You do not have to use the same variables if you think they aren't predictive of those who never do something and of those who might do something and how often they will do it.

Interpreting Results

1. Why did I rename two of the variables?
2. The mean number of volunteer experiences in the past year is 0.36, and they variance is 0.86. What does this tell us?
3. Looking at the Poisson regression, does this model fit the data?
   (a) How do you know?
   (b) What does the line "di invchi2tail" mean?
   (c) In this model, how much larger is the variance than the mean?
   (d) How would you adjust your standard errors to account for extra variation using the above number?
4. What am I doing in the second regression?
   (a) Has the significance levels of any variables changed?
   (b) What is different in this model from the previous model?
   (c) What is the same?
   (d) What are some assumptions about error in this model?
5. Now looking at the negative binomial regression, what does alpha tell us?
   (a) Compare the coefficients for the variables in the 3 models. What do you think?
6. Now look at the last model, which kind of model am I using?
   (a) Are these coefficients similar to the first three models?
   (b) What do the bottom set of coefficients represent? The top?

(c) Looking at the bottom set of numbers, are any significant in determining who will be in the "always zero" class?

(d) Now looking at the top set of numbers, what can you tell me about predicting the number of volunteer experiences?

Appendices

Stata Output

```
. rename gender female

. rename race nonwhite

. drop if educate==.
(9 observations deleted)

. drop if income==.
(951 observations deleted)

. summarize volteer

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
     volteer |      1944    .3605967     .928261          0          9

. di r(Var)
.8616684

. ***************************************************************

. glm volteer female nonwhite educate income, family(poisson)

Iteration 0:   log likelihood = -1787.7576
Iteration 1:   log likelihood = -1686.0444
Iteration 2:   log likelihood = -1685.4303
Iteration 3:   log likelihood = -1685.4299
Iteration 4:   log likelihood = -1685.4299

Generalized linear models                          No. of obs      =      1944
Optimization     : ML                              Residual df     =      1939
                                                   Scale parameter =         1
Deviance         =  2465.513855                    (1/df) Deviance =  1.271539
```

```
Pearson          =  4349.349483                    (1/df) Pearson  =  2.243089

Variance function: V(u) = u                        [Poisson]
Link function    : g(u) = ln(u)                    [Log]

                                                   AIC             =  1.739125
Log likelihood   = -1685.429933                    BIC             = -12217.57

------------------------------------------------------------------------------
             |                 OIM
     volteer |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   .2613218    .077849     3.36   0.001     .1087407     .413903
    nonwhite |  -.2803773   .1083786    -2.59   0.010    -.4927954   -.0679592
     educate |   .1028018   .0144317     7.12   0.000     .0745161    .1310875
      income |   .0568278   .0156635     3.63   0.000     .0261278    .0875278
       _cons |  -3.158302   .2447916   -12.90   0.000    -3.638084   -2.678519
------------------------------------------------------------------------------

. di invchi2tail(1939,0.05)
2042.5553

. **************************************************************

. glm volteer female nonwhite educate income, family(poisson) scale(x2)

Iteration 0:   log likelihood = -1787.7576
Iteration 1:   log likelihood = -1686.0444
Iteration 2:   log likelihood = -1685.4303
Iteration 3:   log likelihood = -1685.4299
Iteration 4:   log likelihood = -1685.4299

Generalized linear models                          No. of obs      =      1944
Optimization     : ML                              Residual df     =      1939
                                                   Scale parameter =         1
Deviance         =  2465.513855                    (1/df) Deviance =  1.271539
Pearson          =  4349.349483                    (1/df) Pearson  =  2.243089

Variance function: V(u) = u                        [Poisson]
Link function    : g(u) = ln(u)                    [Log]

                                                   AIC             =  1.739125
Log likelihood   = -1685.429933                    BIC             = -12217.57

------------------------------------------------------------------------------
             |                 OIM
```

```
     volteer |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   .2613218   .1165939     2.24   0.025     .0328019    .4898418
    nonwhite |  -.2803773    .162318    -1.73   0.084    -.5985148    .0377601
     educate |   .1028018   .0216143     4.76   0.000     .0604385    .1451652
      income |   .0568278   .0234592     2.42   0.015     .0108486     .102807
       _cons |  -3.158302   .3666231    -8.61   0.000     -3.87687   -2.439734
------------------------------------------------------------------------------
(Standard errors scaled using square root of Pearson X2-based dispersion.)

. ***************************************************************

. nbreg volteer female nonwhite educate income

Fitting Poisson model:

Iteration 0:   log likelihood = -1685.4301
Iteration 1:   log likelihood = -1685.4299

Fitting constant-only model:

Iteration 0:   log likelihood = -1529.4738
Iteration 1:   log likelihood =  -1478.209
Iteration 2:   log likelihood =  -1441.165
Iteration 3:   log likelihood = -1441.1354
Iteration 4:   log likelihood = -1441.1354

Fitting full model:

Iteration 0:   log likelihood = -1421.3272
Iteration 1:   log likelihood = -1419.7876
Iteration 2:   log likelihood = -1419.7818
Iteration 3:   log likelihood = -1419.7818

Negative binomial regression                      Number of obs   =       1944
                                                  LR chi2(4)      =      42.71
Dispersion     = mean                             Prob > chi2     =     0.0000
Log likelihood = -1419.7818                       Pseudo R2       =     0.0148

------------------------------------------------------------------------------
     volteer |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   .2844083   .1240645     2.29   0.022     .0412463    .5275702
    nonwhite |  -.3110729   .1622372    -1.92   0.055     -.629052    .0069063
     educate |   .1119953   .0242689     4.61   0.000      .064429    .1595615
      income |    .051931   .0224104     2.32   0.020     .0080075    .0958545
```

```
       _cons |  -3.247383   .3768865    -8.62   0.000    -3.986067   -2.508699
-------------+----------------------------------------------------------------
    /lnalpha |   1.362715   .0979507                      1.170735    1.554695
-------------+----------------------------------------------------------------
       alpha |   3.906786   .3826725                      3.224362    4.733642
------------------------------------------------------------------------------
Likelihood-ratio test of alpha=0:  chibar2(01) =  531.30 Prob>=chibar2 = 0.000

. ***********************************************************

. zip volteer female nonwhite educate income, inflate(female nonwhite educate income)

Fitting constant-only model:

Iteration 0:   log likelihood = -1688.6627
Iteration 1:   log likelihood = -1476.8738
Iteration 2:   log likelihood = -1444.8905
Iteration 3:   log likelihood = -1441.6992
Iteration 4:   log likelihood = -1441.6908
Iteration 5:   log likelihood = -1441.6908

Fitting full model:

Iteration 0:   log likelihood = -1441.6908
Iteration 1:   log likelihood = -1433.2457
Iteration 2:   log likelihood = -1433.0106
Iteration 3:   log likelihood = -1433.0101
Iteration 4:   log likelihood = -1433.0101

Zero-inflated Poisson regression                  Number of obs   =       1944
                                                  Nonzero obs     =        376
                                                  Zero obs        =       1568

Inflation model = logit                           LR chi2(4)      =      17.36
Log likelihood  =  -1433.01                       Prob > chi2     =     0.0016

------------------------------------------------------------------------------
     volteer |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
volteer      |
      female |   .1400743   .1068452     1.31   0.190    -.0693386    .3494871
    nonwhite |  -.1001711   .1532696    -0.65   0.513     -.400574    .2002317
     educate |   .0595475   .0217629     2.74   0.006      .016893     .102202
      income |   .0413874   .0227047     1.82   0.068     -.003113    .0858878
       _cons |  -1.033588   .3601695    -2.87   0.004    -1.739508   -.3276692
-------------+----------------------------------------------------------------
```

```
inflate      |
      female |  -.1849913   .1477324    -1.25   0.210    -.4745414    .1045589
    nonwhite |   .2341655   .2002978     1.17   0.242     -.158411     .626742
     educate |  -.0650641   .0286693    -2.27   0.023    -.1212549   -.0088734
      income |  -.0164052   .0295685    -0.55   0.579    -.0743584     .041548
       _cons |   2.152383   .4718441     4.56   0.000     1.227585     3.07718
------------------------------------------------------------------------------
```