WWS 509 Generalized Linear Models: Precept 6
Introduction to Poisson Models

Kristin E. Bietsch

Office of Population Research, Princeton University

November 2012

1 Introducing the Data

This data is avaliable from Germán's website. It is looking at lung cancer deaths in the population. Variables include these deaths, the population in each group, age groups (40-44 to 80 plus), and smoking status (don't smoke, cigarettes only, pipe/cigar only, pipe/cigar and cigarettes). I use non-smokers age 40-44 as the reference group for the addative model. I have run the necessary models for you, they can be found in Appendix A.

2 A Discussion of Poisson

2.1 Some basics of Poisson data

1. With a poisson model, we can model _____ data.
2. An important feature of Poisson data is that the mean is equal to the _____.
   (a) Because of this, the normal assumtion of _____ is not appropriate for Poisson data.
3. Another useful property is that the _____ of independent Poisson random variables is also Poisson.
   (a) A consequence of this is that we can analyze _____ and _____ data with equivalent results.
      i. They have the same likelihood function.

2.2 A few more things about Poisson

1. With count data, what kind of numbers will we never have? _____
   (a) Thus, how should we deal withl this? Hint: think back to logits _____
   (b) How would we write this? _____
   (c) Ok, not if we just wanted the left side in terms of μ? _____
2. In the final model, an exponentiated regression coefficient e^{β_j} represents a _____ effect of the j^{th} predictor on the mean
   (a) Increasing x_j by 1 unit multiplies the mean by a factor of _____
   (b) An advantage of the log-link model: with count data, the effects of predictors are often multiplicative rather than _____
      i. _____ effects for small models
      ii. _____ effects for large models
3. Finally, we can use deviance to test the goodness of fit for the models. We can also use Pearson's chi-squared statistic. And advantage of using deviance is that we can compare _____ models.

3 Deviance Tables

Table 1: Goodness of Fit of Models

Model	Goodness of Fit	Degrees of Freedom

Table 2: Deviance of Models

Model	Deviance	Degrees of Freedom

4 Interpretation

In the null model:

1. What are we modeling? _____
2. What is the offset? _____
3. What does the constant represent?_____
4. What am I doing when I write the line of code "quietly sum dr [fw=pop]"?_____

Refer to the additive model:

1. What does the constant represent? _____
   (a) What kind of transformation do you need to do to make this interpretable? _____
   (b) Do that and interpret. _____
2. How much higher is the probablity of dying for a 76 year old compared to a 42 year old of the same smoking status? _____
3. How much higher is the probablity of dying for a cigar and cigarette smoker compared to a non-smoker of the same age? _____
4. Compare the probability of dying for a 42 year old non-smoker to a 75 year old pipe only smoker. _____

Appendices

Appendix A: Stata Output

```
. use http://data.princeton.edu/wws509/datasets/smoking
(Smoking and Lung Cancer)

. ********************************************************************
. quietly tab smoke, gen(sstatus)
. quietly tab age, gen(age_g)
. gen os=log(pop*100000)

. rename sstatus2 cigar_pipe
. rename sstatus3 cig_cigar_pipe
. rename sstatus4 cigarette
. rename age_g2 a45_49
. rename age_g3 a50_54
. rename age_g4 a55_59
. rename age_g5 a60_64
. rename age_g6 a65_69
. rename age_g7 a70_74
. rename age_g8 a75_79
. rename age_g9 a80plus

. local status cigar_pipe cig_cigar_pipe cigarette
. local age_group  a45_49 a50_54 a55_59 a60_64 a65_69 a70_74 a75_79 a80plus

. ************************************************************
. * NULL MODEL
. poisson dead, offset(os)

Iteration 0:   log likelihood = -2148.0027
Iteration 1:   log likelihood = -2148.0027
Poisson regression                                Number of obs   =         36
                                                  LR chi2(0)      =      -0.00
                                                  Prob > chi2     =          .
Log likelihood = -2148.0027                       Pseudo R2       =    -0.0000

------------------------------------------------------------------------------
        dead |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |  -13.32889   .0104656 -1273.59   0.000     -13.3494   -13.30838
          os |   (offset)
------------------------------------------------------------------------------
```

```
. estat gof

         Goodness-of-fit chi2  =  4055.982
         Prob > chi2(35)       =    0.0000

. gen dr= dead/(pop*100000)

. quietly sum dr [fw=pop]

. di r(mean)
1.627e-06

. di exp(_b[_cons])
1.627e-06

. ************************************************************
. * ONE FACTOR MODELS
. poisson dead 'status', offset(os)

Iteration 0:   log likelihood = -2075.3683
Iteration 1:   log likelihood = -2075.3629
Iteration 2:   log likelihood = -2075.3629
Poisson regression                                Number of obs   =         36
                                                  LR chi2(3)      =     145.28
                                                  Prob > chi2     =     0.0000
Log likelihood = -2075.3629                       Pseudo R2       =     0.0338

------------------------------------------------------------------------------
        dead |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  cigar_pipe |   .3667831   .0466918     7.86   0.000     .2752688    .4582974
cig_cigar_~e |   -.063346    .038233    -1.66   0.098    -.1382812    .0115893
   cigarette |   .0545969   .0392158     1.39   0.164    -.0222647    .1314586
       _cons |  -13.35289   .0349215  -382.37   0.000    -13.42134   -13.28445
          os |   (offset)
------------------------------------------------------------------------------

. estat gof

         Goodness-of-fit chi2  =  3910.702
         Prob > chi2(32)       =    0.0000

. di exp(_b[cigar_pipe]), exp(_b[cig_cigar_pipe]), exp(_b[cigarette])
1.4430848 .93861868 1.0561148
```

```
. poisson dead 'age_group', offset(os)

Iteration 0:   log likelihood = -239.63576
Iteration 1:   log likelihood = -215.94852
Iteration 2:   log likelihood = -215.87266
Iteration 3:   log likelihood = -215.87266
Poisson regression                                Number of obs   =         36
                                                  LR chi2(8)      =    3864.26
                                                  Prob > chi2     =     0.0000
Log likelihood = -215.87266                       Pseudo R2       =     0.8995

------------------------------------------------------------------------------
        dead |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      a45_49 |   .5560324   .0799878     6.95   0.000     .3992592    .7128055
      a50_54 |   .9881489   .0768149    12.86   0.000     .8375945    1.138703
      a55_59 |   1.371451   .0652555    21.02   0.000     1.243552    1.499349
      a60_64 |   1.628995   .0625358    26.05   0.000     1.506427    1.751563
      a65_69 |   1.957145   .0626921    31.22   0.000     1.834271    2.080019
      a70_74 |   2.205774   .0641042    34.41   0.000     2.080132    2.331416
      a75_79 |   2.457785   .0671346    36.61   0.000     2.326203    2.589366
     a80plus |   2.687489   .0708023    37.96   0.000     2.548719    2.826259
       _cons |  -14.90865   .0584206  -255.19   0.000    -15.02315   -14.79414
          os |   (offset)
------------------------------------------------------------------------------

. estat gof

         Goodness-of-fit chi2  =  191.7219
         Prob > chi2(27)       =    0.0000

. ************************************************************
. * TWO FACTOR MODEL
. poisson dead 'status' 'age_group', offset(os)

Iteration 0:   log likelihood = -159.13827
Iteration 1:   log likelihood = -130.85567
Iteration 2:   log likelihood = -130.75481
Iteration 3:   log likelihood = -130.75481
Poisson regression                                Number of obs   =         36
                                                  LR chi2(11)     =    4034.50
                                                  Prob > chi2     =     0.0000
Log likelihood = -130.75481                       Pseudo R2       =     0.9391

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

```
-------------+----------------------------------------------------------------
  cigar_pipe |   .0478066   .0469926     1.02   0.309    -.0442972    .1399103
cig_cigar_~e |   .2179548   .0386942     5.63   0.000     .1421156    .2937941
   cigarette |   .4169593   .0399121    10.45   0.000      .338733    .4951855
      a45_49 |   .5538766   .0799886     6.92   0.000     .3971018    .7106513
      a50_54 |   .9803865   .0768183    12.76   0.000     .8298253    1.130948
      a55_59 |   1.379458   .0652606    21.14   0.000     1.251549    1.507366
      a60_64 |   1.654228   .0625688    26.44   0.000     1.531596    1.776861
      a65_69 |   1.998171   .0627875    31.82   0.000      1.87511    2.121232
      a70_74 |   2.271405   .0643537    35.30   0.000     2.145274    2.397536
      a75_79 |   2.558575   .0677844    37.75   0.000      2.42572     2.69143
     a80plus |   2.846924   .0724225    39.31   0.000     2.704979     2.98887
       _cons |  -15.19295   .0682382  -222.65   0.000    -15.32669    -15.0592
          os |   (offset)
-------------------------------------------------------------------------------

. estat gof

         Goodness-of-fit chi2  =  21.48617
         Prob > chi2(24)       =    0.6099

. poisson, irr
Poisson regression                                Number of obs   =         36
                                                  LR chi2(11)     =    4034.50
                                                  Prob > chi2     =     0.0000
Log likelihood = -130.75481                       Pseudo R2       =     0.9391

-------------------------------------------------------------------------------
        dead |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+-----------------------------------------------------------------
  cigar_pipe |   1.048968   .0492937     1.02   0.309     .9566696    1.150171
cig_cigar_~e |   1.243531   .0481174     5.63   0.000      1.15271    1.341508
   cigarette |   1.517341   .0605602    10.45   0.000     1.403169    1.640803
      a45_49 |   1.739985   .1391789     6.92   0.000     1.487507    2.035316
      a50_54 |   2.665486   .2047582    12.76   0.000     2.292918    3.098591
      a55_59 |   3.972746   .2592638    21.14   0.000     3.495755    4.514824
      a60_64 |   5.229043   .3271752    26.44   0.000     4.625552    5.911272
      a65_69 |   7.375554   .4630923    31.82   0.000     6.521536     8.34141
      a70_74 |   9.693009   .6237814    35.30   0.000     8.544382    10.99605
      a75_79 |   12.91739   .8755977    37.75   0.000     11.31037    14.75275
     a80plus |   17.23469    1.24818    39.31   0.000       14.954    19.86323
          os |   (offset)
-------------------------------------------------------------------------------
```