WWS 509 Generalized Linear Models: Precept 10
Survival Analysis Using Poisson

Kristin E. Bietsch

Office of Population Research, Princeton University

November 2012

Introducing the Data

This week we are looking at the graduation rate of PhD students from Princeton, Columbia, and Berkeley. This data comes from Espenshade and Rodríguez 1997 and can be found on the website. We will be using Poisson models to look at piece-wise survival analysis. For more about survival analysis, Germán teaches a mini course every-other year (Survival Website). Also, Cox proportional hazards are taught in POP502/ECON572, which I hear has a great preceptor!

Survival Data

Survival data is a special type of data. Think about if you wanted to study mortality, but you don't want to wait until everyone has died to do it. Or maybe you want to study divorce, but not everyone is going to get divorced and some people will get divorced, but not yet.

- If an event has not occured at the time of data collection it is called ____________.
- Different observations can be at risk for different amounts of time, therefore, we need a variable measuring ____________.
  - If using the offset option in Stata, make sure to take the ____________.
- If we are using grouped data, we will need information on ____________ and ____________ of all the people in each group.
- If we are using individual data, we will need information on ____________ and ____________ for each observation.
- What are some interesting ways you can think to use survival analysis?

Interpretation

General

1. Looking at the local I created, what year did I choose as my reference group? Why do you think I chose this?
2. Why did I change the residence term into temporary?
3. Why did I take the log of exposure?

Null Model

1. What does the constant in this model represent?
2. Transform the constant into something easy to interpret.
3. How does this relate to the observed rate?

University Model

1. Interpret the coefficients for Berekely and Columbia. Remember the reference group is Princeton.
2. Does university matter in graduation rate?

Residence Model

1. Interpret the coefficent on "temporary."
2. Why do you think results are this way?

Time Model

1. What year has the highest graduation rate?
2. What is the graduation rate in that year?

University + Time Model

1. Controlling for time, what is difference between Berkeley, Columbia, and Princeton?
2. I bet you are happy you are at Princeton.

Appendices

Stata Output

```
. use http://data.princeton.edu/wws509/datasets/phd.dta
(Time to Ph.D. at Berkeley, Columbia and Princeton)

. tab year, gen(year_)
. local time year_1 year_2 year_3 year_4 year_6 year_7 year_8 year_9
year_10 year_11 year_12 year_13  year_14

. gen berkeley= university==1
. gen columbia= university==2
. gen temporary= residence-1
. gen logexp=log(exposure)

. *null model
. poisson events, offset(logexp)

Iteration 0:   log likelihood = -2712.4406
Iteration 1:   log likelihood = -2712.4406

Poisson regression                                Number of obs   =         73
                                                  LR chi2(0)      =       0.00
                                                  Prob > chi2     =          .
Log likelihood = -2712.4406                       Pseudo R2       =     0.0000

------------------------------------------------------------------------------
      events |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |    -3.1098   .0132465  -234.76   0.000    -3.135763   -3.083838
      logexp |   (offset)
------------------------------------------------------------------------------

. estat gof

         Goodness-of-fit chi2  =  5044.534
         Prob > chi2(72)       =    0.0000

. estimates store null

. quietly summarize events
. scalar nevents = r(sum)
. quietly summarize exposure
. di "Observed Rate = " nevents/r(sum)
```

```
Observed Rate = .04460987


. * Lets look if I should have stayed at Berkeley!
. poisson events berkeley columbia, offset(logexp)

Iteration 0:   log likelihood =  -2230.795
Iteration 1:   log likelihood = -2230.7381
Iteration 2:   log likelihood = -2230.7381

Poisson regression                                Number of obs   =         73
                                                  LR chi2(2)      =     963.41
                                                  Prob > chi2     =     0.0000
Log likelihood = -2230.7381                       Pseudo R2       =     0.1776

------------------------------------------------------------------------------
      events |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    berkeley |  -.8780076   .0333036   -26.36   0.000    -.9432814   -.8127337
    columbia |  -1.416278   .0462468   -30.62   0.000     -1.50692   -1.325635
       _cons |  -2.257475   .0290129   -77.81   0.000    -2.314339   -2.200611
      logexp |   (offset)
------------------------------------------------------------------------------

. estat gof

         Goodness-of-fit chi2  =  4081.128
         Prob > chi2(70)       =    0.0000

. lrtest null .

Likelihood-ratio test                                  LR chi2(2)  =    963.41
(Assumption: null nested in .)                         Prob > chi2 =    0.0000

. poisson events berkeley columbia temporary, offset(logexp)

Iteration 0:   log likelihood = -2132.2795
Iteration 1:   log likelihood = -2132.2677
Iteration 2:   log likelihood = -2132.2677

Poisson regression                                Number of obs   =         73
                                                  LR chi2(3)      =    1160.35
                                                  Prob > chi2     =     0.0000
Log likelihood = -2132.2677                       Pseudo R2       =     0.2139

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

```
      events |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    berkeley |  -.8829095   .0333058   -26.51   0.000    -.9481877   -.8176312
    columbia |   -1.41956   .0462475   -30.69   0.000    -1.510203   -1.328916
   temporary |   .4752226   .0323063    14.71   0.000     .4119033    .5385418
       _cons |  -2.338184   .0297592   -78.57   0.000    -2.396511   -2.279857
      logexp |   (offset)
------------------------------------------------------------------------------

. estat gof

         Goodness-of-fit chi2  =  3884.188
         Prob > chi2(69)       =    0.0000

. poisson events `time', offset(logexp)

Iteration 0:   log likelihood = -987.36059
Iteration 1:   log likelihood = -977.73285
Iteration 2:   log likelihood = -977.69997
Iteration 3:   log likelihood = -977.69997

Poisson regression                                Number of obs   =         73
                                                  LR chi2(13)     =    3469.48
                                                  Prob > chi2     =     0.0000
Log likelihood = -977.69997                       Pseudo R2       =     0.6395

------------------------------------------------------------------------------
      events |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      year_1 |  -2.100309   .0858483   -24.47   0.000    -2.268568   -1.932049
      year_2 |  -.7857379   .0540307   -14.54   0.000    -.8916361   -.6798398
      year_3 |  -.0695101   .0457912    -1.52   0.129    -.1592592     .020239
      year_4 |   .0544905   .0459175     1.19   0.235    -.0355061    .1444872
      year_6 |  -.1503525   .0520957    -2.89   0.004    -.2524581   -.0482469
      year_7 |  -.3434541   .0576071    -5.96   0.000    -.4563619   -.2305463
      year_8 |  -.6143716   .0664536    -9.25   0.000    -.7446182    -.484125
      year_9 |  -.8471369   .0762228   -11.11   0.000    -.9965307    -.697743
     year_10 |  -1.163863   .0922588   -12.62   0.000    -1.344687   -.9830387
     year_11 |  -1.645646   .1211291   -13.59   0.000    -1.883055   -1.408238
     year_12 |  -1.781405   .1356471   -13.13   0.000    -2.047269   -1.515542
     year_13 |  -2.548429   .1482973   -17.18   0.000    -2.839086   -2.257771
     year_14 |  -3.405446    .154551   -22.03   0.000    -3.708361   -3.102532
       _cons |  -2.439876   .0340404   -71.68   0.000    -2.506594   -2.373158
      logexp |   (offset)
------------------------------------------------------------------------------
```

```
. estat gof

         Goodness-of-fit chi2  =  1575.052
         Prob > chi2(59)       =    0.0000

. poisson events `time' berkeley columbia, offset(logexp)

Iteration 0:   log likelihood = -490.53092
Iteration 1:   log likelihood = -482.75404
Iteration 2:   log likelihood = -482.73038
Iteration 3:   log likelihood = -482.73038

Poisson regression                                Number of obs   =         73
                                                  LR chi2(15)     =    4459.42
                                                  Prob > chi2     =     0.0000
Log likelihood = -482.73038                       Pseudo R2       =     0.8220

------------------------------------------------------------------------------
      events |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      year_1 |  -2.160665   .0859071   -25.15   0.000    -2.329039    -1.99229
      year_2 |  -.8438316   .0541078   -15.60   0.000     -.949881   -.7377822
      year_3 |  -.1124038   .0458399    -2.45   0.014    -.2022484   -.0225591
      year_4 |   .0360547    .045927     0.79   0.432    -.0539605    .1260699
      year_6 |  -.1405086   .0520978    -2.70   0.007    -.2426185   -.0383988
      year_7 |   -.331341   .0576116    -5.75   0.000    -.4442577   -.2184243
      year_8 |  -.5965924   .0664687    -8.98   0.000    -.7268687   -.4663161
      year_9 |  -.8401931   .0762385   -11.02   0.000    -.9896178   -.6907684
     year_10 |  -1.178694   .0922862   -12.77   0.000    -1.359572   -.9978165
     year_11 |  -1.673799   .1211607   -13.81   0.000    -1.911269   -1.436328
     year_12 |  -1.828525   .1356796   -13.48   0.000    -2.094452   -1.562598
     year_13 |  -2.571328   .1483305   -17.34   0.000    -2.862051   -2.280606
     year_14 |  -3.398059   .1546342   -21.97   0.000    -3.701137   -3.094982
    berkeley |  -.7693476   .0335248   -22.95   0.000    -.8350549   -.7036402
    columbia |  -1.442275   .0463438   -31.12   0.000    -1.533108   -1.351443
       _cons |  -1.636069   .0431878   -37.88   0.000    -1.720716   -1.551422
      logexp |   (offset)
------------------------------------------------------------------------------

. estat gof

         Goodness-of-fit chi2  =  585.1131
         Prob > chi2(57)       =    0.0000

. * Estimating "survival" probabilities
```

```
. predict p_events
(option n assumed; predicted number of events)

. gen hazard  = events/exposure
. sort university year
. by university: gen cumhaz = sum(hazard)
. gen survival = exp( -cumhaz)
. tab year university, sum(survival) mean

                              Means of survival

   Year of |
  graduate |
    school |          University
    (1-14) |  Berkeley   Columbia  Princeton |     Total
-----------+---------------------------------+----------
         1 | .98140088  .98833132  .93995562 | .96989594
         2 | .90257922  .95536572  .72673407 | .86155967
         3 | .78443277  .88305119  .46713404 | .71153933
         4 | .61886293  .79426798  .30070929 | .57128007
         5 | .48875487  .72256139  .20441666 | .47191097
         6 | .40128219  .66628823  .16244519 |  .4100052
         7 | .34557866  .62381384  .14546854 | .37162034
         8 |  .3063942  .58608893  .13615279 | .38422381
         9 | .27941421  .56743196  .12993373 | .36472521
        10 | .26272123  .55089468  .12652953 | .30071666
        11 |  .2533576   .5423317  .12485366 | .29347514
        12 | .24600293   .5387401  .12064083 | .25440552
        13 | .24237683  .53783846  .11711153 | .28492591
        14 | .24041579  .53761292  .11686262 | .28382678
-----------+---------------------------------+----------
     Total | .45382674  .70790517  .31302056 | .49144411
```