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 **censored**.
- Different observations can be at risk for different amounts of time, therefore, we need a variable measuring **exosure**.
  - If using the offset option in Stata, make sure to take the **log**.
- If we are using grouped data, we will need information on **number of events** and **total exposure** of all the people in each group.
- If we are using individual data, we will need information on **if the event occured** and **their exposure** 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?
   - **I chose the 5th year, because this is a common year to finish (I hope it is the year I finish!).**
2. Why did I change the residence term into temporary?
   - **First, to make the constant interpretable.**
   - **Second, residence doesn't tell me anything, it is like using "gender." Now we have a dummy variable.**
3. Why did I take the log of exposure?
   - **We need to have exposure as a log for the offset option in Poisson.**

Null Model

1. What does the constant in this model represent?
   - **It is the log of the graduation rate.**
2. Transform the constant into something easy to interpret.
   - **di "fitted rate=" exp(_b[_cons]) fitted rate=.04460987**
   - **The overall graduation rate is 44.6 Ph.Ds per thousand grad-student years of exposure**
3. How does this relate to the observed rate?
   - **They are the same.**

University Model

1. Interpret the coefficients for Berekely and Columbia. Remember the reference group is Princeton.
   - **di exp(_b[berkeley]), exp(_b[columbia]) .41561017 .24261547**
   - **The graduation rate is almost 60% lower at Berkeley and 75% lower at Columbia than Princeton**
2. Does university matter in graduation rate?
   - **Yes, looking at the likelihood ratio test comparing this model to the null, there is a chi-square of 963 on 2 degrees of freedom. This is statistically significant at the 1% level.**

Residence Model

1. Interpret the coefficent on "temporary."
   - **exp(_b[temporary])= 1.6083722**
   - **Temporary residents have a 61% higher graduation rate than perminant residents.**
2. Why do you think results are this way?
   - **Perhaps people only become temporary status when they are about to finish.**

Time Model

1. What year has the highest graduation rate?
   - **Year 4**
2. What is the graduation rate in that year?
   - **exp(_b[_cons]+_b[year_4])=.09205352**
   - **92 graduations per 1000 person years.**

University + Time Model

1. Controlling for time, what is difference between Berkeley, Columbia, and Princeton?
   - **di exp(_b[berkeley]),exp(_b[columbia]) .46331525 .23638929**
   - **The graduation rate at Berkeley, controlling for year, is 54% less than at Princeton.**
   - **The graduation rate at Columbia, controlling for year, is 76% less than at Princeton.**
2. I bet you are happy you are at Princeton.