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 Do not cite!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 **mean**
2. Poisson estimates are consistent when the variance is **proportional** (not just equal) to the mean
3. Regular Poisson standard errors tend to be **smaller than they should be** in the presence of over-dispersion
4. Correction can be done by estimating the ϕ using **Pearsons chi-squared statistic** divided by its degrees of freedom $\hat{\phi} = \frac{\chi_p^2}{n-p}$
5. We are assuming that **lack of fit** is actually extra Poisson variation

Negative Binomial Regression

1. Starts with a standard Poisson model and adds a **random effect** 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 **Bernoulli** trials

Zero Inflated Poisson

1. Sometimes in count data, you have many more zeros than the model **predicts**
2. Zero Inflated Poisson is divided into two analyses: finding the **always zeros** and those that follow the Poisson distribution.
3. You need to tell Stata which **variables** 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? **This is a really big problem when people represent their results for dummy variables. The options for this question are male or female, not gender or no gender. Now I know when I interpret which category is the reference.**
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? **The variance is much greater than the mean, so we may be dealing with extra Poisson variation**
3. Looking at the Poisson regression, does this model fit the data? **No**
   (a) How do you know? **The high deviance and Pearson. How do I know for sure, look at the next question.**
   (b) What does the line "di invchi2tail" mean? **I am finding the 5% cut off for a chi-square with the number of degrees of freedom.**
   (c) In this model, how much larger is the variance than the mean? **124% Wow! (1/df) Pearson = 2.243089**
   (d) How would you adjust your standard errors to account for extra variation using the above number? **Take the square root = 1.49769**
4. What am I doing in the second regression? **I am accounting for extra variation. The scale x2 calculates the standard error adjustment for you.**
   (a) Has the significance levels of any variables changed? **Race is no longer at the same level of statistical significance.**

(b) What is different in this model from the previous model? **The standard errors are larger.**

(c) What is the same? **The coefficients are the same as the Poisson model.**

(d) What are some assumptions about error in this model? **We are assuming that the error in the original model was all because of extra variation. You have to be really careful here, because you are assuming that your model is awesome.**

5. Now looking at the negative binomial regression, what does alpha tell us? **It is the variance of the multiplicative random effect. It is highly significant (non-zero).**

   (a) Compare the coefficients for the variables in the 3 models. What do you think? **They are very similar. (Poisson and extra variance are the same)**

6. Now look at the last model, which kind of model am I using? **Zero Inflated Poisson**

   (a) Are these coefficients similar to the first three models? **No, now there are two sets of coefficients, one for determining if you are in the "always zero" group and one for determining number of volunteer experiences in the other group.**

   (b) What do the bottom set of coefficients represent? The top? **The bottom are the predictors for the "always zero" volunteer experiences group. The top is for those people who are not always zero, predicting the number of volunteer experiences.**

   (c) Looking at the bottom set of numbers, are any significant in determining who will be in the "always zero" class? **The only thing that looks predictive is education. It predicts that for each additional year of education a person is 6.3% lower odds of being in the "always zero" category.**

   (d) Now looking at the top set of numbers, what can you tell me about predicting the number of volunteer experiences? **Education is significant here as well, while income is significant at the 10% level.**