Chapter 4 Part 7 Case study

05/03/2021

Negative Binomial Modeling

- Another approach to dealing with overdispersion is to model the response using a negative binomial instead of a Poisson distribution.
- An advantage of this approach is that it introduces another parameter in addition to λ, which gives the model more flexibility and, as opposed to the quasi-Poisson model, the negative binomial model assumes an *explicit likelihood* model.
- You may recall that negative binomial random variables take on non-negative integer values, which is consistent with modeling counts.

Negative Binomial Modeling

- This model consists of selecting a λ for each institution and then generating a count using a Poisson random variable with the selected λ.

- With this approach, the counts will be more dispersed than would be expected for observations based on a single Poisson variable with rate λ.

- Mathematically, you can think of the negative binomial model as a Poisson model where λ is also random, following a gamma distribution.

- The overdispersion in this case is given by $\frac{\mu^2}{r}$, which approaches 0 as r increases (so smaller values of r indicate greater overdispersion).

- Lets conduct negative binomial poisson regression in R

Summarizing Results

- These results differ from the quasi-Poisson model.
- Several effects are now statistically significant at the .05 level
- The effect of type of institution for the Central region ($Z = 2.64, p = .008$)
- The difference between Northeast and Central regions for colleges ($Z = 2.76, p = .006$)
- The difference between Northeast and Central regions in type effect ($Z = -2.01, p = .044$)
- The difference between West and Central regions in type effect ($Z = 2.71, p = .007$).
- In this case, compared to the quasi-Poisson model, negative binomial coefficient estimates are generally in the same direction and similar in size, but negative binomial standard errors are somewhat smaller.

Summarizing The Case Study

- We explored the possibility of differences in the violent crime rate between colleges and universities, controlling for region.
- Our initial efforts seemed to suggest that there are indeed differences between colleges and universities, and the pattern of those differences depends upon the region.
- However, this model exhibited significant lack-of-fit which remained after the removal of an extreme observation. In the absence of additional covariates, we accounted for the lack-of-fit by using a quasilikelihood approach and a negative binomial regression, which provided slightly different conclusions. We may want to look for additional covariates and/or more data.