Stat 363: Poisson Regression (continued)

Example: Do animals bite more during a full moon? An article in the *British Medical Journal* by Bhattacharjee et al. (2000) investigated whether animals bite more often during full moons. To address this question, the researchers conducted a retrospective observational analysis of 1621 consecutive patients who presented at an English hospital ER between 1997 and 1999 with an animal bite. The data is found in **bites.csv**, and relevant R code can be found under **FullMoon.Rmd**. Variables include:

* lunar.cycle = period of lunar cycle (1-10), where 10 = full moon
* n.days = number of days in that period
* n.bites = number of patients presenting with an animal bite during that period

1. Is there initial evidence of more bites during full moon phases?

2. Why can’t we just perform a t-test comparing full moon periods to non-full moon periods using period-level data (n=10)?

3. What options can you think of for handling the fact that Period 5 is based on only 2 lunar days? What are the modeling implications of including number of days in a period as an *offset* term?

4. Interpret model parameters from fit1.

5. Is there any evidence of lack of fit in fit1? What factors may lead to lack of fit?

6. Does full moon have an effect above and beyond the linear cycle trend? Interpret the coefficient and a 95% confidence interval for the fullmoon term in fit3.

7. What is fit4 doing? Does it offer an improvement over fit3?

8. What evidence is there that fit5 (which uses **bitesbyday.csv**) is the analysis performed by the authors of this paper?

9. Based on this analysis, does it make you more wary of animal bites during full moons?

Now we will focus on addressing issues of overdispersion in fit5.

10. Is there evidence of lack of fit in fit5? Cite evidence both from a goodness of fit test and from comparing means and variances by period.

11. If overdispersion goes uncorrected, what are implications for p-values and CIs for model coefficients?

Overdispersion parameter adjustment. One solution is to simply add a second parameter to inflate variances, so that . This is called a “quasi-Poisson” or, in general, a “quasi-likelihood” approach, because our data no longer follows a true Poisson distribution.

Negative binomial modeling. Another solution is to model response using a negative binomial distribution, so that . This distribution comes about if , but the s themselves are randomly chosen according to a gamma distribution: . In this case, and , so that is the amount of overdispersion.

12. Compare the following estimates, tests, and intervals under usual Poisson regression, quasi-Poisson regression, and negative binomial regression:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Poisson | quasi-Poisson | Negative Binomial |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
| Wald-type test stat |  |  |  |
| Wald-type p-value |  |  |  |
| LRT-type test stat |  |  |  |
| LRT-type p-value |  |  |  |
| CI - profile for |  |  |  |
| CI – Wald-type for |  |  |  |