Homework 5

Noah Johnson

March 11, 2018

BYSH 4.15.1 Conceptual Exercise 5

Why is the log of means, log($\bar{Y}$), not $\bar{Y}$, plotted against X when assessing the assumptions for Poisson regression?

Because Poisson regression assumes that $log(\lambda_X)$ is linear in X. To test this assumption empirically, we estimate $log(\lambda_X)$ at each X by finding $log(\bar{Y})$. Then we plot these estimates against X looking for a linear relationship.

BYSH 4.15.1 Conceptual Exercise 8 (Fish)

A state wildlife biologist collected data from 250 park visitors as they left at the end of their stay. Each were asked to report the number of fish they caught during their one week stay. On average visitors caught 21.5 fish per week.

(a) Define the response.

We should model the response Y as a Poisson random variable, representing the number of fish a park visitor catches during one week.

(b) What are the possible values for the response?

As with all Poisson variables, the set of possible outcomes is $\{0, 1, 2, \dots\}$.

(c) What does λ represent?

$\lambda = 21.5$ represents $E[Y]$, the average number of fish a park visitor catches per week.

(d) Would a zero-inflated model be considered here? If so, what would be a "true zero?"

We should consider a zero-inflated model if during the process of exploratory data analysis it was determined that there were more zero counts of fish caught than we would expect to see if responses were coming from a Poisson distribution with $\lambda = 21.5$.

"True zeroes" would be those park visitors who will always report catching zero fish, because they never fish. Perhaps they are vegans, or have a degree in marine biology and know that fish are friends, not food. A zero-inflated Poisson model could estimate α, the proportion of "true zeroes".

BYSH 4.15.2 Guided Exercise 1 (Elephant Mating)

How does age affect male elephant mating patterns? An article by Poole (1989) investigated whether mating success in male elephants increases with age and whether there is a peak age for mating success. To address this question, the research team followed 41 elephants for one year and recorded both their ages and their number of matings. The data is found in elephant.csv, and relevant R code can be found under elephantMating.R.

The variables are:
MATINGS: the number of matings in a given year
AGE: the age of the elephant in years.

```r
elephant <- read.csv(paste("https://github.com/broadenyourstatisticalhorizons/bysh_book/",
    "raw/master/data/elephant.csv", sep = ""))
```

(f) Are the number of matings significantly related to age? Test with

i. a Wald test

ii. a drop in deviance test

(g) Add a quadratic term in AGE to determine whether there is a maximum age for the number of matings for elephants. Is a quadratic model preferred to a linear model? To investigate this question, use

i. a Wald test

ii. a drop in deviance test

(h) What can we say about the goodness of fit of the model with age as the sole predictor? Compare the residual deviance for the linear model to a χ^2 distribution with the residual model degrees of freedom.

BYSH 4.15.2 Guided Exercise 3 (Bullfrogs)

***Big Bullfrog on the Pond*: Use the field observation bullfrog data bullfrogs.csv to determine whether there is convincing evidence that the number of mates is related to the size of the bullfrog.**

```r
bullfrogs <- read.csv(paste("https://github.com/broadenyourstatisticalhorizons/bysh_book/",
    "raw/master/data/bullfrogs.csv", sep = ""))
```

(a) Graph number of mates by size and comment.

(b) Graph log(number of mates) by size and comment.

(c) Write out the likelihood for the Poisson regression model.

(d) Fit the Poisson regression model and interpret the coefficient. Provide a measure of uncertainty.

(e) Conduct a goodness-of-fit test.

(f) Look at the residuals and comment.