

Statistical Methods for Discrete Response, Time Series, and Panel Data (W271): Lab 1

K.C. Tobin, Weixing Sun, Winston Lin

July 2, 2017

The purpose of this lab is to practice using Poisson Regression models, conduct residual diagnostics, and examine goodness of fit.

- Do question 23, 24, 25 in C Chapter 5 (page 351 and 352) of Bilder and Loughin's *Analysis of Categorical Data with R*.
- Even though it is not included in these questions, conduct a comprehensive Exploratory Data Analysis (EDA) analysis, which includes both graphical and tabular analysis, as taught in this course.

5.23. Agresti (2007) provides data on the social behavior of horseshoe crabs. These data are contained in the HorseshoeCrabs.csv file available on our website. Each observation corresponds to one female crab. The response variable is Sat, the number of “satellite” males in her vicinity. Physical measurements of the female—Color (4-level ordinal), Spine (3-level ordinal), Width (cm), and Weight (kg)—are explanatory variables.

- a. Fit a Poisson regression model with a log link using all four explanatory variables in a linear form. Test their significance and summarize results.
 - b. Compute deviance/df and interpret its value.
 - c. Examine residual diagnostics and identify any potential problems with the model.
 - d. Carry out the GOF test described on page 296 using the PostFitGOFTest() function available in the PostFitGOFTest.R program of the same name from our website. Use the default number of groups. (M/5 when M = 100)
 - e. State the hypotheses, test statistic and p-value, and interpret the results.
- ii. Plot the Pearson residuals for the groups against the interval centers (available in the pear.res and centers components, respectively, of the list returned by the function). Use this plot and the residual plots from part (c) to explain the results.

5.24. Continuing Exercise 23, conduct an influence analysis. Interpret the results.

5.25. Continuing Exercise 23, notice that there is one crab with a weight that is substantially different from the rest. This can be seen, for example, in a histogram of the weights. Remove this crab from the data and repeat the steps from Exercise 23. Has this fixed any problems with the model? Are there any other problems with the model, and what could be done to solve these problems?