Statistics 550 Spring 2019 Exam 3 Take-Home Questions

1) The linked baseball data set has variables for doubles, triples, home runs, bases on balls, sacrifice hits, and stolen bases for several major league baseball teams. Fit a generalized linear model to predict the proportion ($x_i =$ triples, $n_i =$ at bats) of team triples, using the covariates mentioned above assuming a binomial distribution. Check residual plots (against predicted values, against covariates, against hat values) and Cooks distance values. Write a few sentences reporting the fitted model, and commenting about the model assumptions.

2) Exercise 15.9 in the text, just for the Binomial and Poisson distributions.

3) For the generalized linear model, we studied how the maximum likelihood estimators can be computed using an iterated weighted least squares (IWLS) procedure. The IWLS procedure uses a working response variable:

$$Z_i^{(l-1)} = \eta_i^{(l-1)} + (Y_i - \mu_i^{(l-1)})g'(\mu_i^{(l-1)})$$

with weights of $W_i^{(l-1)} = 1/\{[g'(\mu_i^{(l-1)})]^2 a_i v(\mu_i^{(l-1)})\}$.We examined these terms in lecture for Poisson-distributed count data with a log link function. Substitute the appropriate terms into the expressions for $Z_i^{(l-1)}$ and $W_i^{(l-1)}$ to show what these expressions are for a model with a binomially distributed Y_i (x_i successes in n_i tries) variable having an logit link function.

4) For the linked set of data, fit the following model:

$$Y_i = \beta_1(1 - \exp(-\beta_2 X_i)) + \varepsilon_i.$$

Report the fitted model, get predicted and residual plots, and comment on the fit of the model and the significance of the parameter estimates.