Stat 616 Generalized Linear Models, Homework 8 **Spring 2019**

Due: Thursday, April 4, 2019

1. Use your previous course material or online resource to review the Chi-Square test for independence of 2 (or more) categorical variables. The test is also known as the "Chi-Square test for contingency table."

From handout *R08_PoissonLoglinear_2019.R*.

2. Answer the 5 questions at the end of the handout. These questions apply to other software users as well.

From Dobson & Barnett, An Introduction to Generalized Linear Models, p. 183.

3. Exercise 9.2
   - In part (b), consider the main effects and the two-way interactions. Use variable selection or test the significance of the interactions to drop the non-significant interactions from the model. Note that the interaction between 2 categorical explanatory variables often involves more than one slope parameter.

4. Exercise 9.3 (a, b).

5. Exercise 9.5
   - Organize the dataset in a correct format for the software and the function.
   - In part (a), you will separate the dataset into 3 subsets according the type of housing. Then you will run the log-linear model for each of the sub-dataset.
   - In part (b), to analyze three variables simultaneously, start with a model with all 2-way and 3-way interactions. (Note that, in this case, a model with all 2-way and 3-way interaction terms is a saturated model.) Then use variable selection (AIC or other criteria) to choose the predictors for the log-linear model.

6. Extra Credit: Exercise 9.3 (c).

Before you start, read the manuscript *prop_odds.pdf* (http://www.stat.uchicago.edu/~pmcc/reports/prop_odds.pdf, by Prof. Peter McCullagh, Department of Statistics, University of Chicago). In particular, note that:

(1) Equation 2 in the manuscript is the proportional odds model, which uses the same parameterization as in R function polr( ) and in STATA.

(2) Figure 1 on p.3 of the manuscript illustrates how the explanatory variable and the linear component are related to the distribution of the latent variable (Z).

(3) Another way to connect the explanatory variable and the linear component to the latent variable (Z) is to assume that for different x-values, the distribution of the latent variable remains the same, but the boundaries of the Z-intervals depend on the predictor and the linear component.

This is the end of HW 8.