Homework # 6, Stat 526

1. The data is based on a sample described by Madsen(1976) of 1681 residuals of twelve areas in Copenhagen. The variable are type of housing (H), degree of contact with other residents (C), feeling of influence on apartment management (I), and satisfaction (Response) with housing conditions.

   (a) Fit a proportation odds model with all the main and interaction effects. Test the goodness of the fit. You need to relevel the Response such that "Low" less than "Medium", and "Medium" less than "High".

   (b) Fit a proportation odds model with all the main effects only. Test the goodness of the fit. Test the null hypothesis that at least one of the four interaction effects are significant.

   (c) Test each of the interaction effects sequentially by HIC, IC, HI, and HC by excluding one after another from the model in (a).

2. The following data (taken from Rosner (2000), Fundamentals of Biostatistics, 5th Edition, p. 39) shows data collected on patients discharged from a hospital. Using GLMs, determine which factors predict duration of stay. The variables are: Id number; Duration of hospital stay; Age; Sex (1=M, 2=F); Temp1-first temperature following admission; WBC1-first WBC (103) following admission; Antib-received antibiotic (1=Y, 2=N); Bact-received bacterial culture (1=Y, 2=N); Serv-service (1=medical, 2=surgical). The interesting model is based on only three predictors, Age, Temp1, and WBC1 to the response Duration. Thus, you only need to consider those four variables-all continuous. No interaction effects will be considered.

   (a) Fit a linear model and study the residual plot. Does it tell us the variances are equal. Use the Box-Cox procedure to find a good transformation on the response.

   (b) Fit a transformed linear model and Gamma GLM. Study the significance of the three variables. If I want to predict the response mean, which model is better for us to use. Why. You may use the plots to explain your answer.

   (c) If there is another person with Age=50, Temp1=98.6, and WBC1=10. Predict the mean of the duration and its 95% confidence interval.