STAT 442
October 2008 Assignment 3

Due: At the beginning of the lecture, Thursday, Oct. 23.

1. The data below is referred to in question 8 in section 5.7 of Simonoff (p. 185). The table below describes the number of homicides in the years 1990-2000.

Year	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Homicides	14.00	24.00	21.00	15.00	5.00	13.00	11.00	9.00	12.00	17.00	11.00

(a) Refer to the web-site for population data for the 15 counties that comprise Vermont. Use this data set to compute to compute yearly homicide rates over the period in question. Plot the rates versus year. What patterns do you observe in the plot?

(b) Fit a log-linear trend in rates using a Poisson regression model. Interpret the fitted model in terms of yearly percent change in the homicide rate. Examine the fit of the model using diagnostic plots. For each plot state clearly which assumptions can be examined through the plot, whether the plot indicates potential inadequacy of fit, and if potential inadequacies are apparent, the nature of the inadequacy identified.

(c) Examine the possibility of overdisperion by estimating an over-disperion scale factor. Assess it's statistical significance using a relevant score test. Re-fit the model following a quasi-likelihood approach, and re-interpret your findings in light of the new fit.

(d) Based on (b) and (c), would you conclude that there is evidence for an overall decreasing trend in the homicide rate?

2. Refer to the "vineyard" data from Simonoff (a link to the Simonoff's data sets is provided on the course web-page). The data describes the yearly production of *lugs* (i.e. baskets) of grapes for each of 52 rows of grapes in an Ohio vineyard over the period of 1983 to 1991.

(a) Begin by plotting examining the relationship of production to year and row number. Plot the data in three ways:

(i) Scatter plots of the dependent variable versus the explanatory variables.

(ii) Repeat the above, after transformation the depending variable by square roots

(iii) Boxplots of the transformed dependent variable versus the explanatory variables.

Are there any particular advantages to any of the three approaches above. Based on the plots, describe the patterns of relationship.

(b) Fit a Poisson regression model to the data with linear terms for year and row number. Perform a graphical examination of the goodness of the fit using plots of regression diagnostics. Describe any apparent deviations from assumptions of the Poisson model.

(c) The vineyard owners were particularly interested in the potential for a fertility variations over field in which the rows were planted. Construct a model (this may require an interative process of fitting and re-checking model fit) to best predict the pattern of variation in yields over rows in future years. For simplicity, enter year as a factor in all models.

(d) Augment the model in (c) to examine whether there is evidence for an interaction between year and row number. To further examine the nature of the interaction, plot the fitted values from the augmented model against row number, with separate lines by year (for example, by using *matplot(1:52,matrix(fitted(pfit),nrow=52),type="b")*, where *pfit* is the object returned from fitting the augmented model using *glm()*). Based on the plot, describe the most prominent departures from log-additivity. Are there any areas of the field which one would predict to provide consistenly lower than average yields ?