Statistics/Forestry/Horticulture 572 **Discussion 10**

Course website	http://www.stat.wisc.edu/courses/st572-larget/
TA	Xu He(Mario)
Email	hexu@stat.wisc.edu
Website of TA	www.stat.wisc.edu/∼hexu
Office	1275F, MSC
Office hour	Th 11:00-12:00, 4:00-6:00
Phone	(608)334-9792

1 Problem 1

In an experiment, the number of Ceriodaphnia organisms are counted in a controlled environment in which reproduction is occurring among the organisms. The experimenter places into the containers a varying concentration of a particular component of jet fuel that impairs reproduction. Hence, it is expected that as the concentration of jet fuel grows, the mean number of counts should decrease. The experiment is done on two different kinds of strains.

number: number of organisms
conc: concentration of jet fuel in grams per liter
strain: the strain type of the organism

(a) Use Poisson regression to fit a model predicting the number of organisms using conc and strain as inputs.

(b) How many observations are smaller than or equal to 5 in the real data? Use fake data simulation to simulate 1000 data sets from the fitted Poisson model. For each data set, compute the count of observations which are smaller than or equal to 5. Compare the real counts with the distribution of fake counts. Interpret this comparison.

(c) Do (a) and (b) again, using quasi-Poisson models.

2 Problem 2

A recent research topic in the treatment of HIV is to determine whether the genotype of a patient's HIV virus can be used to decide on what type of treatment a patient should receive if the patient is failing his/her current therapy. For this purpose, a scoring system called Genotypic Sensitivity Score (GSS) has been developed.

GSS: GSS score
VL: patient viral load

(a) Fit the data with VL be the response. Report the coefficients and standard errors.

(b) Compute 5 fake data sets. Fit a model for each data set and plot the residual. Is the pattern similar to the pattern of real data? What does it imply?

(c) Do a log transformation on VL, and then fit the model again. Compare residuals from fake data sets with real residuals.

(d) Use sim() to generate 1000 sets of plausible regression model coefficients of (c). Create a 95% confidence interval for prediction when the GSS is 10.