Stats 539 Homework 7: Due Wednesday May 3 by 5:00pm

1. Section 9.1.1 on p. 287-288 illustrates an example of the effects of ignoring correlation within subjects. Derive and verify the two expressions for var(b) and var(w) in equation (9.1) on p. 288.

2. Agresti Exercise 9.25 (p. 328).

3. Agresti Exercise 9.39 (p. 332).
   Link to data set: http://www.stat.ufl.edu/∼aa/glm/data/Rats.dat

4. The Skin Cancer Prevention Study was a randomized, double-blind, placebo-controlled clinical trial of beta carotene to prevent non-melanoma skin cancer in high-risk subjects (Greenberg et al., 1989, 1990; also see Stuckel, 1993). A total of 1805 subjects were randomized to either placebo or 50 mg of beta carotene per day for five years. Subjects were examined once a year and biopsied if a cancer was suspected to determine the number of new skin cancers occurring since the last exam. The outcome variable is a count of the number of new skin cancers per year. The outcome was evaluated on 1683 subjects comprising a total of 7081 measurements. The main objective of the analyses is to compare the effects of beta carotene on skin cancer rates. Variables are defined as follows:

ID	Identification number for subject
Center	Center where treated
Age	Age of subject in years
Skin	Skin type (1 = burns; 0 = otherwise)
Gender	1 = male; 0 = female
Exposure	Count of the number of previous skin cancers
Y	Count of the number of new skin cancers per year
Trt	1 = beta carotene; 0 = placebo
Year	Year of follow-up

   A link to the data set is below:

   http://www.math.montana.edu/shancock/courses/stat539/data/skin.txt

   (We are going to ignore the correlation that may be present with the Center variable; do not use the Center variable in any of the following analyses.)

   (a) Consider a Poisson generalized linear mixed model for the subject-specific log rate of skin cancers with randomly varying intercepts. Fit a model with linear trends for the log rate over time and allow the slopes to depend on the treatment group. Do not put any other covariates (besides Trt and Year) into the model. Report the fitted model equation.

   (b) Write an interpretation for each of the estimated fixed effects, β_j.

(c) What is the estimate of the standard deviation of the randomly varying intercepts? Give an interpretation of this value in context of the problem.

(d) What conclusions do you draw about the effect of beta carotene on the log rate of skin cancers using the model in part (a)? Provide results that support your conclusions.

(e) Obtain the predicted (empirical BLUP) random effect for each subject. (Only print the first few BLUPs.)

   i. Calculate the sample variance of the predictions. How does it compare to the estimate of the variance of the random intercepts obtained from your fitted model? Why might they differ?

   ii. Plot the predictions against age and the count of the number of previous skin cancers. What do you conclude from these plots?

(f) Fit a marginal model with the same link and fixed effects as your model from part (a), assuming an exchangeable correlation structure. Report the fitted model equation and write an interpretation for each of the estimated fixed effects, β_j.

(g) For the main objective of the analysis, is a GLMM or marginal model more appropriate? Explain.

(h) Using the method you chose in part (g), repeat the analysis adjusting for skin type, age, and the count of the number of previous skin cancers. What conclusions do you draw about the effect of beta carotene on the adjusted log rate of skin cancers?