**DATA 557**

**Homework Assignment 8**

Submissions are due by **5:00pm, Wednesday March 6.**

This homework is an individual (not group) assignment. You may work together to help each other solve problems, but you should do all the work, create your own solutions, and hand in your own work without copying others’ work.

It is not necessary to include R code with your solutions for this assignment.

Question 1.

1. Use Poisson regression to estimate the effect of dose on post-treatment cell count in the “cells” data from HW 7. Treat dose as a quantitative variable, and adjust for sex, age and pre-treatment cell count. Report the coefficient estimate table.
2. State the interpretation of the exponentiated coefficient estimate for dose.
3. Is there evidence of an effect of dose?
4. Are robust SEs necessary?
5. Assess the assumptions of the Poisson regression model.
6. Compare results with results from the linear regression model for HW 7, Q1.

Question 2.

1. Use Poisson regression with dose as a categorical (i.e., factor) variable. Report the coefficient estimate table.
2. Interpret the estimated effects of dose in this model using exponentiated coefficient estimates.

Question 3.

1. For this question you are to use the “Teeth” data to estimate the association between having at least one tooth extracted (i.e., EXTR > 0) and disease severity (PDALL), with adjustment for age and gender. Estimate this association using logistic regression. Report the coefficient estimate table.
2. Provide an interpretation of the association using an exponentiated coefficient estimate from the logistic regression model.
3. Estimate the association using Poisson regression. Report the coefficient estimate table.
4. Provide an interpretation of the association using an exponentiated coefficient estimate from the Poisson regression model.
5. Compare results from the two models.

Question 4. **This is an optional question - if you answer it correctly you will earn bonus points!**

An experiment is to be conducted to estimate the linear association between an explanatory variable X and a response variable Y. The association will be estimated using the least squares estimate of the regression coefficient for X in a simple linear regression model with X as explanatory variable and Y as response variable. The assumptions of independence of observations, linearity of the relationship between X and the mean of Y, and constant variance all are met. You have a fixed total sample size of n=50 observations. You may choose any values of the X variable for the n observations subject to the constraint that all values of X must be no greater than 40 and no smaller than 10. Prove that the standard error of the least-squares estimate is minimized by setting 25 of the X values equal to 10 and 25 values equal to 40.