Spring 2016 Exam

You may use SAS help at `sas.support.com` as well as **R** online help. You may also review your PROC IMPORT statements for tab-delimited text files with headers. The water quality data set used in this exam (Ecoli.xls) is available in Blackboard and on the course website as an Excel workbook.

We will be working with the water quality data in several ways. Import the Excel spreadsheet into your SAS work directory. You will find that you need to revise the original data set as you work through the exercises; all such changes can simply be included in the same DATA step as you proceed.

1. For this problem, we will study a logistic regression using the E. Coli data set.

   The pollution standard for E Coli is 126 colonies per milliliter. Create a binary variable that is set to 0 for E Coli readings below the standard (strictly less than 126) and set to 1 for E Coli readings that are above the standard.

   (a) Create a Month variable from the variable `Collection_Date` using the MONTH() function. Month will be numeric, but we will treat it as a categorical variable here. Generate a freqency table in PROC FREQ tabulating Month versus the binary variable and comment on any interesting trends.

   (b) Construct a binary regression model for the binary response variable as a function of the categorical variable Month. Is the model significant? What does the intercept term estimate? Construct a 95% confidence interval for the intercept, exponentiate the endpoints and interpret.

   (c) Use the table of odds ratio estimates to determine the odds ratio of Month 2 (February) vs Month 1 (January). What other information would you need to construct a confidence interval for this odds ratio?

   (d) Fit Month as a continuous variable. Interpret a confidence interval for the transformed slope parameter using the table of Odds Ratio Estimates. Use the tables of Model Fit Statistics to compare the fit for Month as a categorical variable vs the fit for Month as a continuous variable.

2. For the following questions, we will study a two-way table.

   (a) Create a three-level water quality variable for E Coli with levels 0 (< 50), 1 ($50 - 125^{-}$), and 2 (≥ 126) representing Good, Average and High readings for E Coli. Construct a two-way table of this variable against Month.

   (b) Is it reasonable to consider Month an ordinal variable? Why or why not? Report results of a test of ordinal association.

   (c) Are Month and water quality independent? Using studentized Pearson residuals, and thresholds for these values, prepare a table of signs indicating departures from independence (use + signs for standardized residuals greater than 3 and signs for standardized residuals less than 3. Report any patterns in your results.