## Midterm Exam

1. I generated data from the GSS (General Social Survey) website cross-tabulating Importance of Free Time and Relaxation (1=Unimportant up to 7=Very Important), Highest Degree (1=Less than High School, 2=High School, Junior/Community College or above) and Gender (F=Female, M=Male). The data set is stored in sheet Q1 of Excel workbook Midterm_2019.xlsx. It is available on the website.

   (a) Test whether there is a strong ordinal relationship between Highest Degree and Free Time and Relaxation.

   (b) Test for a strong linear-by-linear relationship using $M^2$ for both default scores and midranks.

   (c) Fit an independence model for Highest Degree and Free Time and Relaxation and compute standardized residuals. Identify any cells with residuals greater in absolute value than 2.0 and discuss patterns you see in the data set.

2. Using the same data set, dichotomize Free Time and Relaxation by assigning categories 1-5 to 0 and categories 6-7 to 1.

   (a) Using this binary variable as a response, fit a saturated logit model with the explanatory variables Highest Degree, Gender and their two-way interaction.

   (b) Interpret interaction, main effect and intercept terms.

3. Again using the same data set, trichotomize Free Time and Relaxation by assigning category 1-3 to 1, 4-5 to 2 and categories 6-7 to 3. Now consider a 3 by 3 cross-tabulation of this new 3-level variable with Highest Degree.

   (a) Conduct the usual large sample tests of independence. What are the results?

   (b) Use Monte Carlo simulation in SAS to approximate an exact chi-squared test of independence. Compare results to (a). Why may an exact test be infeasible here?

4. Four different labs are sent groundwater samples that include readings of an analyte, fluoride, either above or below a regulatory limit (1=Above limit, 0=Below limit). Results from the labs' tests are dichotomized by whether or not their readings correctly detected whether the analyte concentration was either above or below the regulatory limit (1=Correct, 0=Incorrect). The data can be found in Sheet Q2 of Excel workbook Midterm_2019.xlsx.

   (a) Compute the marginal odds ratio of the odds of correctly detecting fluoride above the regulatory limit and interpret.

   (b) Now compute the four sets of conditional odds ratios for each lab. Compare results to (a) and discuss.

5. (a) Using the same data set as above, treat the Lab Result as the response. Conduct a Cochran-Mantel-Haenszel test in PROC FREQ and report your results.

   (b) Now fit nested models in PROC LOGISTIC including lab, an additive model for lab and sample, and an interaction model for lab and sample and compare results to your results in (a).