BIOS663 Homework 4
Due Monday, April 8 in class.

1. The following questions are on the data and model described in Q3 of HW3:

   - Examine the tolerances and variance inflation factors from this model. Do you think any collinearity is present based on the tolerance and VIF? Why or why not?
   - Conduct an eigenanalysis of the scaled SSCP and correlation matrices, presenting a table formatted like Table 8.6.2 in Muller and Fetterman.
     (a) Does there appear to be any collinearity between the intercept and the covariates? Why or why not? If so, list the variables?
     (b) Does there appear to be any collinearity among the covariates? Why or why not? If so, list the variables?

2. Find the Box-Cox transformation of the simulated data (BoxCox.dat) and compare the residual plots of the raw and transformed data.

3. Investigators are interested in the effect of dermal nicotine exposure in a population of Latino tobacco workers in North Carolina. (Nicotine can be absorbed from tobacco leaves through the skin and can cause nicotine poisoning, which is characterized by nausea, vomiting, headache, and dizziness.) Data were collected on tobacco work tasks and risk factors for exposure to nicotine during a summer tobacco work season. Nicotine exposure was measured by levels of cotinine, a nicotine metabolite, contained in saliva. Other covariates of interest include age, body mass index, education, work conditions (working in wet conditions is believed to increase nicotine absorption), type of tobacco work ("priming" refers to picking or harvesting the tobacco and is expected to result in highest nicotine exposures, "barning" refers to putting the harvested tobacco into a barn for curing, "topping" refers to breaking the flower off the top of the plant, and "other" refers to farm work that does not involve tobacco contact, such as driving a truck), and smoking (smokers would also have nicotine exposure through cigarettes, and it is not known whether exposure to tobacco leaves would increase cotinine levels to a similar extent in both smokers and non-smokers).

   The variables are available in the file tobacco.dat and listed in the following order.

   - COTININE: salivary cotinine concentration (in ng/mL)
   - AGE: age (in years)
   - BMI: body mass index (in kg/m^2)
   - EDUC: years of education
   - WET: takes value 1 if work conditions on day of measurement were wet and takes value 0 otherwise

- TASK: takes value 1 for priming, 2 for barning, 3 for topping, and 4 for other work not involving tobacco contact
- LNNSMOKE: natural logarithm of (1 + number of cigarettes smoked per day)

To report a test, provide H_0, the test statistic, the degrees of freedom, the p-value, the decision (accept/reject H_0), and an interpretation of the result in terms of the subject matter.

- One-Way ANOVA: For these questions, use the log of salivary cotinine as the response and task as the only predictor.
  - Report a test of whether all cell means are equal.
  - If your overall test of the task effect was significant, examine all pairwise comparisons using the Scheffe correction. Summarize your findings in a table including columns for the estimated mean difference, degrees of freedom, F statistic, p-value, and a Scheffe confidence interval for the mean difference. Explain your findings in language the investigator can understand.
  - Provide a table of parameter estimates and standard errors using (a) cell mean coding and (b) reference cell coding, and give the interpretations of parameters in both coding schemes. In addition, provide the $\mathbf{C}$ and $\boldsymbol{\theta}_0$ matrices used to test the hypothesis that average cotinine levels for workers involved in priming are greater than the average continine levels for all other workers.
- Two-Way ANOVA: For these questions, use the log of salivary cotinine as the response and task and wet as predictors.
  - Fit the two-way ANOVA model with full interaction, and interpret all parameter estimates in your model, clearly stating which coding scheme you used. Discuss the validity of the HILE Gauss assumptions for this model.
  - *Based on this model*, create a table of the estimated mean log cotinine levels, associated standard errors, and how each estimated mean is obtained from the model parameters (e.g., $\widehat{\beta}_0 + \widehat{\beta}_1$) for each task-wet combination.
- The Full Model in Every Cell: For these questions, use the log of salivary cotinine as the response and task, and lnnsmoke as predictors.
  - Fit the full model in every cell. Provide and interpret estimates of all parameters for this model.
  - Report an appropriate test of whether task is related to cotinine levels. If this test is significant, report step-down tests to determine exactly where differences lie. For all tests reported, be sure to state H_0 clearly and give explicit justification for which tests were used and why.