BIO 226 Homework Assignment 2
Due Thursday February 26, 2015

Purpose:

To provide an introduction to the use of PROC MIXED for analyzing data from a single-group repeated measures design.

Instructions:

1. For each question requiring data analysis, support your conclusions by including the relevant SAS output in your answer.
2. Include your SAS program as an Appendix to your solutions (PLEASE don't include the output from your program!).
3. Homework should be turned by class on the due date.

Joint Modeling of Mean and Covariance in the Dental Growth Study

In a study of dental growth, measurements of the distance (mm) from the center of the pituitary gland to the pteryomaxillary fissure were obtained on 11 girls and 16 boys at ages 8, 10, 12 and 14 years (Potthoff and Roy, 1964).

The data are in the file "dental.txt" on the course web page. Each row of the data set contains the following six variables: subject ID, gender ("F" or "M"), and the measurements at ages 8, 10, 12 and 14 years, respectively.

For all analyses in this homework, subset the dataset to contain only the measurements for the girls.

PART A: Descriptive Analyses

1. Plot the observed trajectories of distance for the 11 girls (all on one plot). Briefly, describe the important patterns that are apparent in the plot.

2. Obtain descriptive statistics for the distances in girls including means and standard deviations at each measurement occasion. Plot the means against time and comment on the shape of their trajectory. Also, comment on how the standard deviations are changing with time.

3. Obtain the variance-covariance and correlation matrices for the repeated measurements over time. Briefly, describe the important patterns in the correlations.

4. Identify one unexpected feature of this correlation structure generated in question 3. Provide a possible reason for this unexpected feature.

PART B: Repeated Measures Models

5.   Use PROC MIXED to fit a repeated measures model in which there is separate parameter for the mean distance at each of the four ages [HINT: Use the NOINT option on the MODEL statement]. Use an unstructured variance-covariance structure. For estimating parameters, fit the model using (i) the METHOD=ML option and (ii) the METHOD=REML option on the PROC MIXED statement.

   a. Does either method of estimation give the same estimates of the variances and covariances as in your answer to question 3? If so, which one or ones?

   b. Compare the variances and covariances for the two methods of estimation (i.e. ML versus REML). If they are not the same, comment on why they are different and state which method should be preferred and why?

   c. Compare the estimated betas in the regression model to the sample means calculated in question 2. Provide a reason for any relationship you notice between these two sets of estimates.

6.   Using METHOD=ML and an unstructured variance-covariance structure, fit a model which assumes that the mean distance changes linearly with time.

   a. Why can this model be considered to be nested within the model fitted in question 5?

   b. By undertaking a likelihood ratio test, evaluate the goodness of fit of this model compared with the model fitted (with METHOD=ML) in question 5. Which model do you prefer and why?

   c. Why should you not use the REML log likelihood to compare models in this question?