

HOMEWORK 1
BIOSTATISTICS 755
DUE JANUARY 30TH, 2025

This question will use data on Orthodontic Measurements on Children. The data are from a study of dental growth measurements of the distance (mm) from the center of the pituitary gland to the pteryomaxillary fissure. Measurements were obtained on 11 girls and 16 boys at ages 8, 10, 12, and 14. Scientific goals of the study are to:

- determine whether distances over time are larger for boys than for girls, and
- whether the rate of change of distance over time is similar for boys and girls.

The data is available on the website in the file `dental.txt`, which contains the data and information about the study and variables.

Please use this dataset to answer the following questions. Note that `dental.txt` is given in the *wide format*. This format is good for questions 1, 4 and 5. For question 2 the *long format* will be needed. Please see the example from class on how to go from the wide to the long format.

1. Read the data into SAS (or R) and calculate sample means, standard deviations and variances of the distance measurements at each occasion.
2. Construct three time plots for data. For these plots:
 - (a) Make all the colors differ by subject.
 - (b) Make a panel plot which separate panels for boys and girls (make all lines black).
 - (c) Take the plot in (b) and add a blue mean line to each panel.
3. Using the plot in 2c:
 - (a) discuss the differences in the means for boys and girls, and
 - (b) discuss the pattern in variation over time.
4. Calculate the 4 x 4 covariance and correlation matrices for the four repeated measures. Does correlation appear to vary by the time difference between measurements?

5. Calculate the 4 x 4 covariance and correlation matrices for the four repeated measures separately for boys and girls. Comment on the differences in the variance and correlation between boys and girls.
6. Recall the 4 correlations “truths” from the second set of slides. Describe on what these truths mean in the context of this example. In your description, try to be more specific than restating each “truth” in context. For example, don’t use the word ‘correlation’ in your description, think about what the “truth” tells you about the data if taken as fact.
7. State three potential sources of variability in the data in context of the problem. Using the time plot you constructed in 2(c) give your hypothesized ordering of which source of variability is the largest.