

Biost 540: Homework 1

General Instructions:

- You can work in groups of **at most three students**. If you choose to work in groups, elect one member as the leading member who will be responsible for uploading the homework assignment solutions. Please note that all members in the group are expected to equally contribute to the assignment.
- Be sure to show work for all problems. R code should not appear in the main body of the homework; however, the code should appear at the end of the assignment as an Appendix. It should be possible for someone to use the code to reproduce any figures or numeric results.

Part A: EDA for Augmentation dataset

This dataset is from a clinical trial for augmentation treatment for depression (Sanacora et al., 2004). There were 100 patients total: half were randomly assigned to the augmentation group (fluoxetine + yohimbine) and half were randomly assigned to the control group (fluoxetine + placebo). Participants in the study were followed for an additional 6 weeks and their Hamilton depression scale ratings (HDRS) and Clinical Global Impressions Scale for severity (CGI). We will focus on the participants' HDRS.

```
library(reshape2)
aug <- read.csv("Data/augmentation.csv")
aug <- aug[, c("id", "Treatment_Group", "HD_t0", "HD_t1", "HD_t2",
              "HD_t3", "HD_t4", "HD_t5", "HD_t6")]
aug_long <- melt(aug, id=c("id", "Treatment_Group"))
aug_long$week <- as.numeric(gsub("HD_t", "", aug_long$variable))
```

1. Summarize the distribution of HDRS at baseline and weeks 1 – 6 in the two groups. Plot the mean HDRS over time for the two groups. What do you notice?
2. Plot individual series of longitudinal observations for all subjects. Comment on what you observe, specifically in regards to differences between the two groups.
3. Characterize the correlation among the HDRS overall and separately for each group. Comment on what you notice. Is the variability in responses different at different time points?
4. Provide summaries of the missing data. Here are a few things to consider: When does it occur? How much is missing in each group? For subjects missing data, what does the baseline HDRS look like (as compared to subjects with complete observations)?

Part B: Dental Growth

Scientific question of interest: is there a difference in dental growth rates between male and female children?

```
library(nlme)
data(Orthodont)
```

1. Summarize the distances at baseline (age 8) and at the end of the study (age 14) for males and females.
2. Compute a slope over time for each child and summarize these slopes.
3. Create a boxplot of slopes by sex. Based on this figure does there appear to be an association between sex and rate of dental growth?
4. Run a t-test or fit a simple linear regression model to address the scientific question of interest. What do you conclude? Be sure to include *scientific* and *statistical* conclusions (i.e. include point estimate(s), 95% confidence interval(s), p-value(s), and the corresponding interpretations).
5. Conduct a pre/post analysis using the 8 year and 14 year observations. What are the results from the Post, Change, and ANCOVA models? Include interpretations of the slope parameter. Which would be

most useful in answering the question of whether there is a difference in growth rates between males and females?