Midterm

Biost 540

General Instructions:

- You can work in groups of at most three students.
- When working in groups, elect one member as the leading member who will be responsible for uploading the midterm project. Please note that all members in the group are expected to equally contribute to the assignment!

Background of the data:

CD4 count refers to the numbers of T-lymphocyte cells in the body. These cells are affected by the HIV virus. A normal CD4 count is approximately 800 to 1000; a CD4 count below 200 is one of the diagnostic criteria for AIDS established by the Centers for Disease Control and Prevention (CDC).

The AIDS Clinical Trial Group (ACTG) study 193A was a randomized, double-blind study of AIDS patients with advanced immune suppression (CD4 counts of less than or equal to 50 cells/mm3). Patients were randomized to one of four daily regimens containing 600 mg of zidovudine:

1. Zidovudine alternating monthly with 400 mg didanosine
2. Zidovudine plus 2.25 mg of zalcitabine
3. Zidovudine plus 400 mg of didanosine
4. Zidovudine plus 400 mg of didanosine plus 400 mg of nevirapine

Measurements of CD4 counts were scheduled to be collected at baseline and at 8 week intervals during follow-up. However, the CD4 count data are unbalanced due to mistimed measurements and missing data that resulted from skipped visits and dropout. The number of measurements of CD4 counts during the first 48 weeks of follow-up varied from 1 to 9, with a median of 4.

Variables:

1. Subject ID
2. Treatment: coded

   1 = zidovudine alternating monthly with 400 mg didanosine;

   2 = zidovudine plus 2.25mg of zalcitabine,

   3 = zidovudine plus 400mg of didanosine, and

   4 = zidovudine plus 400mg of didanosine plus 400mg of nevirapine

3. Age: in years
4. Gender: coded 1=Males, 0=Females
5. Week: time since baseline in weeks
6. log(CD4 count + 1)

Scientific Goal:

To determine the relative clinical efficacy of zidovudine (AZT) plus didanosine (ddI), AZT plus zalcitabine (ddC), AZT alternating monthly with ddI, and AZT/ddI plus nevirapine in HIV-infected patients with advanced disease

Specific Questions:

1. Using graphical methods and/or numerical summary statistics describe the relationship between log-CD4 counts over time for the four treatment groups; examine whether there appears to be a difference comparing patients by gender or by baseline age; and investigate if individuals with missing data have different characteristics from those who have complete observations.
2. Prior authors indicated time-dependent growth of log-CD4 with a change-point at approximately week 16. Fit appropriate models that use all available information and allow you to answer the following questions:
    - Characterize treatment effectiveness over time. Is there a difference in treatment regimens in terms of changes in log-CD4 counts over time? What can you say about the relative effectiveness of the treatment regimens?
    - How does treatment effectiveness vary over time controlling for gender and baseline age?

Format:

In order to answer the scientific questions organize your short report as follows. Format: maximum of 2.5 pages, font size 12, line spacing 1.5, with the following sections:

1. **Introduction**: provide a brief background for the problem and the questions you will address
2. **Methods**: describe the statistical methods that you utilized (for both descriptive and inferential analyses). Justify any key decisions that you made.
3. **Results**: present the results of your analyses appropriate for a general scientific audience. Provide careful interpretation for any key regression coefficients that address the question of interest.
4. **Tables and Figures**: Include these at the end of your report. Provide appropriate captions and be sure to refer to them in the main text. Please note that unedited output from any statistical software is not acceptable. It is probably most useful to first consider the tables and figures you will present. For example, you may organize your Tables and Figures as follows:

- Table 1: Descriptive statistics for the patient characteristics by treatment group. The purpose of such a table is to allow the reader to assess the comparability of treatment groups with respect to other predictors of response such as age, sex, etc., while at the same time giving them an idea of the types of patients in the study.
- Table 2: Descriptive statistics for outcomes by treatment group. While we are ultimately interested in making inference about some summary measure (along with its precision as measured by a CI or a SE), we need to recognize that excessively high or low outcomes may indicate possible toxic treatments for individual patients (so ranges of the data and/or SD are also of interest). Hence, this table might focus more on the data itself, rather than the inference. (The inference is further described below.)
- Figure 1: A graphical display of outcomes. This could either be primarily by treatment group with superimposed smooths, or it could be primarily inferential (by showing point estimates with standard error bars or confidence intervals).
- Table 3: Inferential statistics presenting results by treatment group. This table would typically include point estimates, confidence intervals, and P values.

5. **Code**: provide a copy of the code that you utilized for the analysis. This code should have adequate documentation so that others could replicate the analysis.

Note: pages for tables, figures and code are not counted towards your maximum of 2.5 pages for the extended abstract.