STAT 6289, Assignment #1

Due Wednesday, October 16, 2019

Overall Objectives:

We would like to do a step-by-step analysis of the BMACS data. The data are described in Chapter 1.2.2 (Example 1) of my book "Nonparametric Models for Longitudinal Data: With Implementation in R" by Wu and Tian (2018, Chapman & Hall/CRC, Boca Raton). In short, this dataset is from the Baltimore site of the Multi-center AIDS Cohort Study (BMACS), which included 283 homosexual men who were infected by the human immunodeficiency virus (HIV) between 1984 and 1991. The covariates of this simple illustration dataset are age of HIV infection (age), smoking status coded as "always smokers (Smoke = 1)" and never smokers, and CD4 cell percent before HIV infection. The outcome we are interested in is the post infection CD4 percent. These variables were repeatedly observed between one to 14 times.

Part of the analysis for this BMAC data has been presented in Chapter 2.4.1 and a few other chapters of my book. We would like to perform some further detailed analyses of this dataset using different modeling schemes.

Specific Analysis for Model Construction:

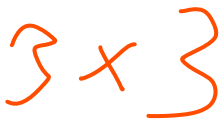
In the following, you need to present the mathematical expressions of your models and assumptions, and provide justifications and possible scientific interpretations of your modeling approach. For analysis and justifications, you need to show some preliminary analysis (such as plots and summary statistucs) similar to the ones given in slides LME1 and LME3 (both available on Blackboard), present your R code and output, and tabulate your results and write down the explanation of your findings.

1. First, we consider a simple linear mixed-effects model of post-infection CD4 as a function of time since HIV-infection. Using the two-stage modeling approach, you may consider the random effects as functions of (a) smoking status only, (b) pre-CD4 level only, (c) both smoking status and pre-CD4. Give your final chosen model under the following three scenarios for the random effects:

   1.1 only random intercept;

   1.2 only random slope;

   1.3 both random intercept and random slope.



2. To allow some flexibility, we next consider the linear mixed-effect model of post-CD4 as a polynomial function of time since HIV-infection. Using the two-stage modeling approach, you now have three possible random effect terms, which may be considered as functions of (a) smoking status only, (b) pre-CD4 level only, (c) both smoking status and pre-CD4. Give your final chosen model under the following four scenarios for the random effects:

   1.1 only random intercept;

1.2 only random slope;

1.3 both random intercept and random slope;

1.4 random intercept, random slope, and random coefficient for the second polynomial term.

3. Compare the fitness of the above models, and justify your final chosen linear mixed-effects model for the BMACS data. This can be done by the "anova" function in R.

4. Present the final estimates and inferences of your model, and discuss the conclusions of your analysis.