STAT 6225, Assignment #3

Due Monday, November 25, 2019

Overall Objectives:

As a follow-up analysis of the BMACS data, we would like to extend the linear mixed-effects models of Assignment #1 to nonparametric models for the post-HIV CD4 percent. For the purpose of illustration, we would like to simplify the analysis by only considering smoking status (always smokers, never smokers). The other two covariates, age at HIV infection and CD4 cell percent before HIV infection, are possibly important, but will not be considered due to the shortage of time.

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 present your estimates as well as their 95% CIs (this can be computed using the resampling bootstrap method). Similar to the two earlier assignments, the results here can be computed using the standard R mixed-effects model package, such as lme4. Please present your R code and output, and tabulate your results and write down the explanation of your findings.

1. First, without covariates, we consider some simple piece-wise linear mixed-effects models of post-infection CD4 as a function of time since HIV-infection. Suppose that we pick two interior knots at $t = 2$, 4 years after HIV infection. Give your final chosen model, and estimates and 95% CI's for the fixed effects under the following three scenarios:

   1.1 local constant fit, i.e. using the basis functions $B_1(t) = 1_{[t<2]}$, $B_2(t) = 1_{[2 \le t<4]}$ and $B_3(t) = 1_{[4 \le t]}$, with random-effects for each of the basis functions;

   1.2 local linear fit without assuming continuity at the knots, i.e. three separate straight lines at $t < 2$, $2 \le t < 4$ and $4 \le t$, with random-effects for the intercepts only;

   1.3 local linear fit with continuity at the knots, i.e., using $B_1(t) = 1$, $B_2(t) = t$, $B_3(t) = (t-2)_+$ and $B_4(t) = (t-4)_+$, with random-effects for each of the basis functions.

   1.4 What are your interpretations and clinical conclusions from the results obtained in #1.1, #1.2 and #1.3?

2. Extend the analysis in #1 by including smoking status as the covariate. In this case, you will consider the varying-coefficient mixed-effects models of the form

$$Y_i(t) = \beta_0(t) + b_{0i}(t) + \beta_1(t)\, X_i(t) + b_{1i}(t)\, X_i(t) + \epsilon(t),$$

   where $X_i(t)$ is the smoking status, $b_{0i}(t)$ and $b_{1i}(t)$ are subject-specific deviation curves and $\epsilon(t)$ is the independent measurement error. Give your final chosen model, and estimates and 95% CI's for the fixed effects under the following three scenarios:

2.1 local constant fit using the basis functions in #1.1 for $\beta_0(t)$, $b_{0i}(t)$, $\beta_1(t)$ and $b_{1i}(t)$;

2.2 local linear fit without assuming continuity at the knots for $\beta_0(t)$, $b_{0i}(t)$, $\beta_1(t)$ and $b_{1i}(t)$, i.e., using the basis functions as in #1.2;

2.3 local linear fit with continuity at the knots for $\beta_0(t)$, $b_{0i}(t)$, $\beta_1(t)$ and $b_{1i}(t)$, i.e., using the basis functions as in #1.3.

2.4 What are your interpretations and clinical conclusions about the effects of smoking status obtained in #2.1, #2.2 and #2.3?