Homework #5

BSTA 519 – Fall 2021

Applied Longitudinal Data Analysis

Due – Monday 11/15/2021 by midnight

Linear mixed effects models, Problem 1

Chapter 8, problem 8.1.

The data is provided in Exercise.txt. If you open up the dataset using WordPad, it will show the correct format. This is the dataset we used to illustrate the covariance pattern models.

In addition to the questions specified in the textbook, answering the following:

Perform basic model diagnostic for the random intercept and random slope model, and in particular,

a. Create scatter plots of all possible pairs of the responses at any two time points. What assumption do you check by using such a plot? Do you see any departure from this assumption?

b. Create a Q-Q plot for marginal residuals at each time point. How do plots speak to the normality of the data?

c. Create the following two plots
   i. Marginal residuals vs. predicted marginal mean
   ii. Conditional residuals vs. predicted conditional mean

   and comment on the patterns shown in these plots, whether there are any outlying observations, and any concerning patterns.

d. Obtain the transformed marginal residuals and transformed predicted marginal mean, and do the following:
   i. Create a Q-Q plot for the transformed marginal residuals. How does the plot speak to the normality of data? Any evidence to show that the data are skewed?
   ii. Create a plot of transformed marginal residuals vs. transformed predicted means.
   iii. Create a plot of absolute values of transformed marginal residual vs. transformed predicted means, and adding a lowess curve.

   Comment on the patterns shown in plots from ii) and iii), whether there are any outlying observations, and any concerning patterns.

e. Check whether there are any outlying individuals using transformed marginal residuals and Mahalanobis distance (d_i). Focus on the most extreme values of d_i.

Homework #5

BSTA 519 – Fall 2021

Applied Longitudinal Data Analysis

Due – Monday 11/15/2021 by midnight

Linear mixed effects models, Problem 2

The Childhood Asthma Management Program study

The Childhood Asthma Management Program (CAMP) was a clinical trial carried out in children with asthma. The trial was designed to determine the long-term effects of 3 treatments (budesonide, nedocromil, or placebo) on pulmonary function for up to 10 years period. The design of CAMP was a multicenter, masked, placebo-controlled, randomized trial. A total of 695 children aged 5-12 years were enrolled between December of 1993 and September of 1995.

For this analysis, you will only look at the data from the treatment group of nedocromil and the outcome variable is, FFratio, the ratio of FEV1 to forced vital capacity (FVC) where FEV1 is the Forced Expiratory Volume at 1 second. The data for this analysis is provided in the file ***CAMP HWK5.xlsx***.

Dataset Variables

id	Randomized participant ID
age_rz	Age in years at Randomization
GENDER	m=male, f=female
ETHNIC	w=white, b=black, h=hispanic, o=other
FFratio	FEV1/FVC ratio %
Fdays	Days since randomization

1) Create an appropriate time plot to show the trend of how FFratio changes over time and describe the trend.

2) Fit a random intercept model to evaluate both the cross-sectional association between age at randomization and FFratio, and the longitudinal association between age and FFratio.

    i. Write out the model you fit to evaluate both associations.

    ii. Is there a significant cross-sectional association? Is there a significant longitudinal association? Interpret your results.

iii. Obtain the estimated covariance and correlation matrix. What is the estimated intra-class correlation coefficient in this case? Is there a strong correlation among the repeated measures within the same subject?

iv. Based on this model, how do the subject-specific response curves differ among the individuals?

v. Plot the marginal mean response over time based on the fitted model.

vi. Fit a marginal model using the same mean response with an exchangeable correlation matrix, and the REML method. How do the results from the marginal model compare with the results above?

3) Fit a random intercept and random slope model with the above specification for the mean response, with an unstructured covariance matrix for the random effects.
   i. Write out the model you fit to evaluate both associations.

   ii. How does the resulted variance-covariance matrix differ from the one in 1)? Provide the estimated covariance and correlation matrix.

   iii. Does this model provide a better fit to the data, compared to the model in 1)?

   iv. Based on this model, how do the subject-specific response curves differ from each other? Could you provide the range of the slopes for the middle 95% subjects?

4) Fit a random intercept and random slope model to evaluate whether the longitudinal association between age and FFratio differs by gender after adjusting for ethnicity and age at randomization, and interpret your results.

5) Perform basic model diagnostic for the random intercept and random slope model in 4). In particular, do the following:
   a. Create the following two plots
      i. Marginal residuals vs. predicted marginal mean
      ii. Conditional residuals vs. predicted conditional mean

   and comment on the patterns shown in these plots, whether there are any outlying observations, and any concerning patterns.

   b. Obtain the transformed marginal residuals and transformed predicted marginal mean, and

i. Create a Q-Q plot for the transformed marginal residuals. How does the plot speak to the normality of data? Any evidence to show that the data are skewed?
ii. Create a plot of transformed marginal residuals vs. transformed predicted means.
iii. Create a plot of absolute values of transformed marginal residual vs. transformed predicted means, and adding a lowess curve.

Comment on the patterns shown in plots from ii) and iii), whether there are any outlying observations, and any concerning patterns.

c. Check whether there are any outlying individuals using transformed marginal residuals and Mahalanobis distance (d_i). Focus on the most extreme values of d_i.