Biostatistics 140.655, 2017-18
Lab 4

Topics:

- Generating longitudinal data from mixed model framework
- Interpretation of fixed effect parameters within mixed models
- Quantifying heterogeneity across subjects based on the random effect variance

Learning Objectives:

Students who successfully complete this lab will be able to:

- Write out the structure of data generated from a mixed model.
- Describe the steps required to generate data from a mixed model.
- Implement linear and logistic mixed effects regression models.
- Interpret the fixed effects parameters within linear and logistic mixed effects models
- Interpret the variance components generated within a mixed effects model.

Associated Quiz:

- While we will review and discuss parts of this exercise, there is a short quiz (Quiz 4) on Courseplus which will assess your basic knowledge of the course materials thus far with focus on ideas from this lab session.
- Quiz 4 is available on Courseplus Wednesday March 7th, please complete the quiz by 5pm on Friday March 9th.
- Please do not discuss the solution for the quiz with your peers until Saturday March 10th.

Scientific Background:

Recall the exercise therapy trial we explored in Lecture 7. Participants were randomized to receive increasing number of repetitions (TRT = 0) or increasing amount of weight (TRT = 1). Measures of strength were taken at baseline (day 0) and on days 2, 4, 6, 8, 10 and 12. The original trial had 37 participants.

Suppose you are planning a larger, more definitive trial to show that increasing the amount of weight is superior to increasing the number of repetitions. To understand the properties of our hypothesis test for a treatment effect (i.e. power or type II error), we will conduct a simulation study using results from the original trial of 37 participants.

I fit a linear mixed model to the data from the exercise therapy trial as follows:

$$Y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i}) \times time_{ij} + \beta_2 \times trt_i \times time_{ij} + \varepsilon_{ij}$$

$$\begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} {\sigma_0}^2 & \sigma_{01} \\ \sigma_{01} & {\sigma_1}^2 \end{bmatrix}\right), \varepsilon_{ij} \sim N(0, \sigma^2), Corr\left(b_{0i}, \varepsilon_{ij}\right) = 0, Corr\left(b_{1i}, \varepsilon_{ij}\right) = 0$$

I fit the model above to the exercise therapy trial and obtained the estimates below which we will assume are the true values for the parameters in our simulation study

$$\beta_0 = 81, \beta_1 = 0.11, \beta_2 = 0.06, {\sigma_0}^2 = 9.7, \sigma_{01} = -0.01, {\sigma_1}^2 = 0.03, \sigma^2 = 0.65$$

NOTE: $Corr(b_{0i}, b_{1i}) = \frac{-0.01}{\sqrt{9.7}\times\sqrt{0.03}} = -0.02$

Lab Exercise:

1. Stata and R code has been provided to you to simulate a hypothetical trial of 250 participants per treatment group. Review the code and order the steps below to come up with a road map for simulating data from a mixed model.

_____ Sample hypothetical participants from the trial; specifically, sample values of the random effects with one set of random effects representing a hypothetical individual

_____ Obtain "guesstimates" for the parameters within the hypothesized mixed model

_____ Hypothesize a mixed model that describes how you think the response is generated; this model should include relevant parameters for testing a hypothesis (i.e. do participants generate strength more quickly on TRT = 1 compared to TRT = 0)

_____ Calculate the expected mean response for each hypothetical individual at each follow-up

_____ Obtain an observed response for each hypothetical individual by sampling a random residual at each time point. This residual represents natural biological variation in the response.

2. Fit the mixed model to the simulated data and summarize the results.

   a. The mixed model acknowledges that there may be heterogeneity in the mean strength at baseline across participants. Give an interval containing roughly 95% of the values for the mean strength at baseline.

   b. The mixed model acknowledges that there may be heterogeneity in rate of change in strength across the follow-up period. Give an interval containing roughly 95% of the expected weekly changes in average strength for participants receiving TRT = 0. Give a similar interval for participants receiving TRT =1.

   c. Test the hypothesis that the expected weekly change in strength is the same in the two treatment groups.

3. Consider a new outcome; a binary indicator for improved strength comparing each follow-up to baseline.

$$biny_{ij} = 1 \; if \; Y_{ij} > Y_{i1}, 0 \; otherwise$$

The logistic mixed model is fit for $time_{ij} > 0$.

$$\log\left[\frac{P(biny_{ij} = 1 | b_{0i}, b_{1i}, trt_i, \; time_{ij})}{P(biny_{ij} = 0 | b_{0i}, b_{1i}, trt_i, \; time_{ij})}\right] = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i}) \times time_{ij} + \beta_2 \times trt_i \times time_{ij}$$

$$\begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} {\sigma_0}^2 & \sigma_{01} \\ \sigma_{01} & {\sigma_1}^2 \end{bmatrix}\right)$$

Fit the model above using 5, 7, and 14 integration points. Fill in the table below:

Parameter	Integration points: 5	Integration points: 7	Integration points: 14
β_0			
β_1			
β_2			
${\sigma_0}^2$			
${\sigma_1}^2$			
σ_{01}			

Do you think your estimates have "converged"? i.e. do you think you should continue to evaluate the model fit for larger number of integration points?

4. Based on the results of the mixed effects logistic regression model, summarize the results.

    a. Interpret the main effect of time, i.e. $\exp(\beta_1)$.

    b. Estimate and interpret $\exp(\beta_1 + \beta_2)$.

    c. This model acknowledges that participants will vary in how their odds of improved strength relative to baseline will change over time. Provide an interval that contains roughly 95% of the weekly odds of improved strength among participants receiving TRT = 0. Provide a similar interval for participants receiving TRT = 1.

