Biostatistics 140.655, 2017-18
Lab 4 Solution

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 9th, please complete the quiz by 5pm Friday March 11th.

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:

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

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

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.

__3__ Sample hypothetical participants from the trial; specifically, sample values of the random effects with one set of random effects representing a hypothetical individual

__2__ Obtain "guesstimates" for the parameters within the hypothesized mixed model

__1__ 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)

__4__ Calculate the expected mean response for each hypothetical individual at each follow-up

__5__ 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.

```
. mixed y time i.trt#c.time || id: time, cov(uns)

Mixed-effects ML regression                     Number of obs      =      3500
Group variable: id                              Number of groups   =       500

                                                Obs per group: min =         7
                                                               avg =       7.0
                                                               max =         7


                                                Wald chi2(2)       =    265.78
Log likelihood = -5865.8688                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   .1087612   .0122563     8.87   0.000     .0847393    .1327831
             |
  trt#c.time |
          1  |   .0591408   .0173104     3.42   0.001      .025213    .0930685
             |
       _cons |   81.25553   .1404492   578.54   0.000     80.98026    81.53081
------------------------------------------------------------------------------
```



```
------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
id: Unstructured             |
                   var(time) |   .0317507   .0023872      .0274003    .0367918
                  var(_cons) |   9.556108   .6238506      8.408376     10.8605
             cov(time,_cons) |  -.0085468   .0273626     -.0621766    .0450829
-----------------------------+------------------------------------------------
               var(Residual) |   .6609638   .0186949      .6253196    .6986397
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(3) =  6826.10   Prob > chi2 = 0.0000
```

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.

   The estimate of the population mean strength at baseline is 81.3 and the estimated random intercept variance is 9.6.

   Therefore, roughly 95% of the participants will have mean strength measurements at baseline ranging from 81.3 ± 2 * sqrt(9.6) → 75.1 to 87.5.

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.

   The estimated weekly increase in strength is 0.11 units on average across all participants receiving TRT = 0. The estimated random slope variance is 0.03. Therefore, roughly 95% of the participants receiving TRT = 0 will have expected weekly changes in strength ranging from 0.11 ± 2 * sqrt(0.03) → -0.24 to 0.46.

   The estimated weekly increase in strength is 0.11 + 0.06 = 0.17 units on average across all participants receiving TRT = 1. Therefore, roughly 95% of the participants receiving TRT = 1 will have expected weekly changes in strength ranging from -0.18 to 0.52.

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

   We look at the treatment x time interaction term to determine if the expected weekly change in strength is the same in the two treatment groups; p = 0.001. So, there is statistical evidence that the expected weekly change is different across the two treatment groups; however, the clinical relevance of an additional 0.06 unit increase in strength per week is questionable.

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	-0.082	-0.099	-0.099
β_1	0.253	0.265	0.266
β_2	0.100	0.101	0.102
${\sigma_0}^2$	1.017	1.823	1.819
${\sigma_1}^2$	0.101	0.127	0.131
σ_{01}	0.184	0.069	0.068

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?

There are fairly large differences comparing the estimates based on 5 integration points relative to those produced using 7 integration points. Moving from 7 to 14 integration points, we see changes but those changes are smaller. I would think it reasonable to stop at 7 integration points; you could also increase to 28 integration points and see how similar those results are to those obtained using 14.

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

   I assumed that the results based on 14 integration points were stable.

```
. meqrlogit biny time i.trt#c.time if time > 0 || id: time, cov(uns) intp(14)


Mixed-effects logistic regression               Number of obs      =      3000
Group variable: id                              Number of groups   =       500

                                                Obs per group: min =         6
                                                               avg =       6.0
                                                               max =         6

Integration points =  14                        Wald chi2(2)       =     80.15
Log likelihood =  -1386.407                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
        biny |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   .2655752   .0417484     6.36   0.000     .1837499    .3474006
             |
  trt#c.time |
          1  |   .1015917   .0451137     2.25   0.024     .0131704     .190013
             |
       _cons |  -.0995508   .1587037    -0.63   0.530    -.4106043    .2115027
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
id: Unstructured             |
                   var(time) |   .1306467   .0322961      .0804785    .2120886
                  var(_cons) |   1.819872   .8122917      .7587714    4.364865
             cov(time,_cons) |   .0683561   .0828858      -.094097    .2308092
------------------------------------------------------------------------------
LR test vs. logistic regression:     chi2(3) =   804.61   Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.
```

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

      For the typical or average participant receiving TRT = 0, the odds of improving strength relative to baseline strength increase by 30 percent per week of follow-up (exp(0.265) = 1.30).

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

      For the typical or average participant receiving TRT =1, the odds of improving strength relative to baseline strength increase by 44 percent per week of follow-up (exp(0.265 + 0.102) = 1.44).

   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.

      The random slope variance is estimated to be 0.13.

For 95% of the participants receiving TRT = 0, the relative odds of improved strength per week is expected to fall within 0.63 to 2.68 (exp(0.265 ± 2*sqrt(0.13))), therefore, for some participants the odds of improvements in strength relative to baseline are decreasing over time (weekly odds ratio of 0.63) whereas for other participants the odds of improvement are increasing over time (weekly odds ratio of 2.68).

For 95% of the participants receiving TRT =1, the relative odds of improved strength per week is expected to fall within 0.70 to 2.97.

