MS COMPREHENSIVE EXAM
Stat 505-506 Take Home
August 21–22, 2014, 100 Points

Timing: This exam is due 24 hours after you pick it up from Monique.

The Data: `http://www.math.montana.edu/~jimrc/MScomp/reading.csv`

This pdf file is posted as
`http://www.math.montana.edu/~jimrc/MScomp/TakeHomeAug2014.pdf`

Background Information:
A reading assessment containing 57 questions was given to each of 539 students in 15 schools. Schools were selected at random from all schools in a small state, and students were selected at random from within each selected school. Each of the questions was designed to assess one of nine target skill areas, and the scores you see are the averages for a given student for each of the nine target areas. The number of questions combined into a target skill area score varied from one target to the next is contained in the variable `itemCt` and shown here in this table. It is the same for each student.

Target	1	2	3	4	5	6	7	8	9	Total
Number of items	4	6	8	5	9	6	8	6	5	57

Researchers wonder:

- Are the target areas of different mean difficulty?
  If so, can we rank them or identify the most and least difficult? They think of them as grouped in this way:
  Targets 1 and 2 deal with "retrieving information", 3 through 7 are "structuring" questions, and 8 and 9 ask students to "evaluate". Do these categorizations help explain some of any potential differences?

- How do scores vary across schools and individuals?

Your task:
Write up a complete report for the researchers describing your findings, how you obtained them, and justifying decisions you made along the way. Include your computing code as an appendix and any plots near where you refer to them in your report. Do not clutter the report with unneeded computer output, but do include important components of your analysis if needed for the explanation. Assume that the researchers who will read your report are familiar with mixed effects models and have taken a full year course in data analysis from "The Sleuth". Be sure to include the following components:

- Assume that only these nine target areas are of interest.

  – Write out the model (Laird–Ware or Gelman notation) and explain each piece including all assumed distributions. (10 pts)
  – What information does this analysis provide? Provide a complete analysis which addresses the research questions. Include exploratory plots to give the researchers a thorough understanding of their data. Include diagnostics to address any assumptions you are making. (40 pts)

- Estimate the correlations between two scores from the same student and between two scores from the same school (different students). (10 pts)

- Explain the scope of inference given what you have been told about the data collection. (12 pts)

- The nine target areas could be considered to represent a large population of potential targets rather than (as above) viewing them as the only targets of interest. In this view, they would ignore the subclassification of 'retrieving', 'structuring', or 'evaluating'. How does the 'representative sample' concept change the model? Write out and explain the new model, but do not fit it. Explain, compare, and contrast how this model (versus the model above) might be used to answer the researchers' questions. Again, don't answer the questions, just explain what types of answers are possible under this alternative way of viewing the target effects. (18 pts)

Formatting, grammar, clarity: 10 pts

Email your report in pdf format to
`jimrc@math.montana.edu`
`jobo@math.montana.edu`.