36-463: Hierarchical Linear Models
Fall 2014
HW06 – Due Thu Oct 13, 2016

Announcements

- Please turn this assignment in on Blackboard as usual.
- Reading in Gelman & Hill (G&H):
  - G&H, Chapters 11–13. Please read and try out the ideas in these chapters.
    - We will return to ideas in these chapters (as well as Ch's 14-15) throughout the remainder of the semester.
    - In Ch 13, skip material on inverse-Wishart distribution (pp 286–287) for now—premature for us!
- This assigment and the associated data sets are in the "hw06" area of the class website.
- For all the exercises below (both pages!), you may find that you will need the "foreign", "arm", "lme4" and "ggplot2" libraries in R.

Exercises

1. Consider the following multilevel model for data y_i, $i = 1, \ldots, n$, arranged into J groups, $j = 1, \ldots, J$, where each group j has n_j observations:

$$\left.\begin{array}{rcl} y_i &=& \alpha_{j[i]} + \epsilon_i,\ \epsilon_i \overset{iid}{\sim} N(0, \sigma^2) \\ \alpha_j &=& \beta_0 + \eta_j,\ \eta_j \overset{iid}{\sim} N(0, \tau^2) \end{array}\right\} . \qquad (*)$$

Prove the following assertions:

(a) If $i \neq i'$ and $j[i] \neq j[i']$, then $\mathrm{Corr}\,(y_i, y_{i'}) = 0$.
(b) If $i \neq i'$ but $j[i] = j[i']$, then $\mathrm{Corr}\,(y_i, y_{i'}) = \frac{\tau^2}{\tau^2+\sigma^2}$.
(c) Let $\overline{y}_{j.} = \frac{1}{n_j} \sum_{i:j[i]=j} y_i$, the average of all observations in group j. Then $\mathrm{Var}\,(\overline{y}_{j.}) = \tau^2 + \sigma^2/n_i$
(d) Suppose we exactly replicate the experiment generating new data y_i^* following the model

$$\left.\begin{array}{rcl} y_i^* &=& \alpha_{j[i]} + \epsilon_i^*,\ \epsilon_i^* \overset{iid}{\sim} N(0, \sigma^2) \\ \alpha_j &=& \beta_0 + \eta_j,\ \eta_j \overset{iid}{\sim} N(0, \tau^2) \end{array}\right\} , \qquad (**)$$

so that the group level α's and η's (and β_0) are the same between $(*)$ and $(**)$ [the conditions we are measuring didn't change] but the new set of ϵ^*'s are independent of η's and ϵ's [we re-measured, and so we have new measurement error on each observation]. Form the group averages $\overline{y}_{j.}^*$, analogous to $\overline{y}_{j.}$. Then

$$\mathrm{Corr}\,(\overline{y}_{j.}, \overline{y}_{j.}^*) = \frac{\tau^2}{\tau^2 + \sigma^2/n_j} .$$

(This is another interpretation of the reliability coefficient $\frac{\tau^2}{\tau^2+\sigma^2/n_j}$.)

In all four parts, carefully state any assumptions that you need.

2. G&H Chapter 12, #9.

3. G&H Chapter 12, #6. When each part asks you to "write the model", you should write it in good, careful statistical notation with greek letters, appropriate subscripts, and so forth. I think you will find it easiest to do, using the "multilevel model" notation, rather than the "variance components" or "hierarchical Bayes" notation. Of course you will fit the models using lmer() in R.

4. G&H Chapter 13, #1. Again, write the models in good, careful statistical notation, and fit them using lmer() in R.

5. In some settings, like the London Schools data or the Radon measurement data, the clustering comes from natural social or administrating groupings of observations—houses in the same county, students at the same school, etc. As we discussed at the beginning of the semester, mixed effects models also get used for longitudinal data in which each individual has a different trajectory over time; now the "clusters" are measurements at different times on the same individual. The models that use random effects for the slopes and intercepts (or other coefficients) that vary from one one individual to the next are called "growth curve models". This exercise addresses a grow curve model.

   The data come from the folder CD4 in the hw06 area of the website. Most of the variable names are self-explanatory. You can find out more by googling "cd4 arv", for example.

   - G&H Chapter 12, #2, parts (a), (b) and (c) [I have no idea what (d) is asking for!].

     In general the idea of partial pooling is that it "shrinks" coefficients toward an overall mean level (mean intercept, mean slope, etc.), without forcing all the coefficients to be equal across groups or clusters (across children in this case). In part (c) you should compare and contrast the amount of partial pooling or shrinkage between parts (a) and (b).