3/23/2010 36-402/608 ADA-II H. Seltman

Handout #16: Mixed Models, part 1

1. Overview
   (a) The mixed (fixed plus random effects) linear model and the closely related hierarchical linear model are among of the most widely applicable and used statistical models.
   (b) They generalize and subsume (fixed) linear models, repeated measures models, and arima models.
   (c) The underlying concept can be thought about in two equivalent ways, including
      i. Repeated measurements on the same object or from the same group are correlated, and that correlation can be directly incorporated into the Normal likelihood (called a general linear model, not to be confused with a generalized linear model).
      ii. Repeated measurements on the same object share unmeasured covariates that would affect the outcome if they were measured. We can model these "upper level" latent variables as random effects: each "upper level" object (within which repeated measurements are made) is a random draw from a population of such objects and the latent variables induce a Normal distribution of group effects.
   (d) In either case, compared to the mean based on just the fixed (based on measured variables) effects, the repeated outcomes (or corresponding errors) for any one "upper level" object are correlated, but compared to the group mean, they may be uncorrelated (or may still have correlation that needs to be modeled).
   (e) Explanatory variables are "fixed" if their values are of primary interest, and "random" if the quantities being estimated are only a random value out of some distribution, in which case the s.d. (or variance) of the distribution is of interest (and the mean is modeled as zero). The latter uses only 1, instead of k-1, df.
   (f) Linear mixed models are the starting point for generalized linear mixed models which allow non-Normal outcomes.
   (g) Linear mixed models are the starting point for non-linear mixed models which means models other than of the form $\mathbf{X}\boldsymbol{\beta}$.
   (h) Mixed (hierarchical) models can have more than two levels of hierarchy (nesting).
   (i) Mixed models are likelihood based, so MAR missingness is appropriately handled.

2. Formal model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}$$
$$\mathbf{b} \sim N(\mathbf{0}, \mathbf{G})$$
$$\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{R})$$

   $\mathbf{y}$ is all outcomes, stacked, usually with multiple measurements from the same upper level object in adjacent rows. $\boldsymbol{\beta}$ is the usual vector of unknown, population parameters (of the fixed effects).

   $\mathbf{X}$ is all fixed covariates (including appropriately coded categorical variables), one row per measurement.

   $\mathbf{Z}$ is the matrix of random covariates, one row per measurement. Dropping $\mathbf{Z}$ gives a standard (fixed effects) linear model. The simplest, most common form of $\mathbf{Z}$ is one column per upper level group, coded as a group indicator variable. Many other forms are possible. $\mathbf{b}$ is the vector of random effects, one per column of $\mathbf{Z}$.

   $\mathbf{G}$ is the variance-covariance matrix for the $\mathbf{b}$s, and $\mathbf{R}$ is the variance-covariance matrix for the $\boldsymbol{\epsilon}$s. Often $\mathbf{R}$ is block diagonal with one block for each upper level group.

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$$
$$\text{Var}(\mathbf{y}) = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}$$

3. (Overly) simple example: Four class rooms are randomized to four levels of annoying background noise (10, 4, 6, and 8 units respectively). The number of students per classroom is 5, 4, 5, and 4. Gender is recorded for each student. The outcome is score on a science test.

4. Breakout and Discussion