3/23/2010 36-402/608 ADA-II H. Seltman
Breakout #16 Comments

(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.

What fixed effects model makes sense here? (Consider randomizing light to each student individually and accounting for gender.) What parameters would you estimate? In addition to the intercept we need fixed effects for the slope of the sound effect and for the gender effect. Using obvious notation, we would estimate $\beta_0,\ \beta_S,\ \beta_M$, where M represents male with a female baseline.

What are possible unmeasured classroom variables that we can subsume into a per-classroom random effect (intercept)? The teacher effect is most obvious. Other classroom environmental effects such as what is outside the windows and cheerfulness of the decorations may have an effect. Additionally the gender and age mix of the class, presence/absence of a bully, etc. may affect scores.

Where are the unmeasured student-level variables in the model? Variables such as age, ability to concentrate, reading ability, parental encouragements, diet, IQ, previous science teachers, etc. are all represented by the residual error, σ^2.

What is the appropriate within-classroom correlation structure? The structure of equicorrelation of students within a classroom and uncorrelated errors between classrooms seems most appropriate.

Write out $\mathbf{y}$, $\mathbf{X}$, $\boldsymbol{\beta}$, $\mathbf{Z}$, $\mathbf{b}$, $\boldsymbol{\epsilon}$, $\mathbf{G}$, and $\mathbf{R}$ using Greek letters for parameters, Roman letters for variables, and numbers for constants (make up the ones I didn't give you), and using ellipses to avoid excess tedium. If you can't figure one out, at least try to figure out its dimension.

$\mathbf{y}$ is all outcomes:

$$\begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{15} \\ y_{21} \\ \vdots \\ y_{43} \\ y_{44} \end{bmatrix}$$

$\mathbf{X}$ has an intercept, the noise level, and a gender indicator for each student:

$$\begin{bmatrix} 1 & 10 & 0 \\ 1 & 10 & 1 \\ \vdots & & \\ 1 & 10 & 1 \\ 1 & 4 & 0 \\ \vdots\,1 & 8 & 0 \\ 1 & 8 & 1 \end{bmatrix}$$

$\boldsymbol{\beta}$ is the fixed effects: $\boldsymbol{\beta}' = [\beta_0,\ \beta_S,\ \beta_M]$.

$\mathbf{Z}$ is 4 indicator variables for classroom:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The $\mathbf{b}$s are 4 random variables: $\mathbf{b}' = [b_1 b_2 b_3 b_4]$

$\mathbf{y}$ is all errors (beyond the random intercept):

$$\begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{15} \\ \epsilon_{21} \\ \vdots \\ \epsilon_{43} \\ \epsilon_{44} \end{bmatrix}$$

$\mathbf{G}$ is the variance-covariance matrix of the 4 random effects:

$$\begin{bmatrix} \tau^2 & 0 & 0 & 0 \\ 0 & \tau^2 & 0 & 0 \\ 0 & 0 & \tau^2 & 0 \\ 0 & 0 & 0 & \tau^2 \end{bmatrix}$$

$\mathbf{R}$ is the variance covariance matrix of the errors, so it is an 5+4+5+4=18 by 18 diagonal matrix with σ^2 in all off diagonal positions. This results in the following form for $\mathbf{ZGZ}' + \mathbf{R}$:

$$\begin{bmatrix} \mathbf{G} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{H} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{H} \end{bmatrix}$$

where $\mathbf{0}$ is a matrix of zeros of whichever size fits, $\mathbf{H}$ is

$$\begin{bmatrix} \tau^2 + \sigma^2 & \tau^2 & \tau^2 & \tau^2 \\ \tau^2 & \tau^2 + \sigma^2 & \tau^2 & \tau^2 \\ \tau^2 & \tau^2 & \tau^2 + \sigma^2 & \tau^2 \\ \tau^2 & \tau^2 & \tau^2 & \tau^2 + \sigma^2 \end{bmatrix}$$

and $\mathbf{G}$ is like $\mathbf{H}$, but 5 by 5.

You should see that any two students in the same class have the same correlation: $\frac{\tau^2}{\tau^2+\sigma^2}$, any two students in different classes are uncorrelated, and the variance of any test score around that predicted by noise and gender is $\tau 62 + \sigma^2$ which is the classroom random intercept effect plus the residual error effect.