Multi-Level Modeling

Nathaniel MacNell

EPID 799B

Fall 2016

Overview

- What is multi-level modeling? Why use it?
- What are the types of multi-level models?
- How do I put them into R?
- Other considerations

A Note on Terminology

- Lots of different classification systems
- Some example terms for what we are talking about:

  Multilevel models
  Heirarchical models
  Fixed, mixed, random effects models
  Nested models
  GEE

Why Multi-Level?

Useful when you have **clustered** data:

- GLM assumes independent observations

Benefits:

- Can be used to control for *unmeasured* confounders or remove nuisance parameters
- Can improve precision of estimates
- Can estimate effects at multiple scales

Multi-level Data

Data for multi-level modeling includes:

- Individual covariates
- Group-level covariates
- Group identifiers

Generalized Estimating Equations

Handles clustering in the outcome **covariance matrix**. (No confounding control; more popular in biostatistics than as an epidemiology method)

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

$$\begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix}$$

Random vs. Fixed Effects

Imagine that each cluster (person) has their own individual intercept (instead of an overall intercept). We could model this intercept in two ways:

1. Consider it like another "error" term, but at the level of the cluster (each observation has two intercepts that are added together). **[RANDOM]**
2. Consider it like another parameter of the model to estimate (make a categorical variable ID variable). **[FIXED]**

Mixing random and fixed effects

Coefficients can have fixed and random effects too! This bring us to the idea of mixed models:

1. Add an “error” term to individual-level beta coefficient (allowing the effect to differ between groups around some mean effect). **[RANDOM]**
2. Estimate beta coefficients separately for each group (i.e. add an interaction term between individual-level beta and group ID). **[FIXED]**