Ch 1 - Background

What is multilevel regression modeling?

A regression whose parameters (the regression coefficients) have a probability model. The model of these parameters have parameters of their own, known as hyperparameters.

Multilevel models have two key parts: 1. Varying coefficients, and 2. A model for those varying coefficients

While classical regression can accommodate for varying coefficients via indicator variables, multilevel models are capable of modeling variation between groups.

Models for regression coefficients

Consider an educational study using data on students from different schools. Students' grades y on a standardized test are predicted based on a pre-test score x.

Two levels of regressions can be fit for this data. The first is a student-level regression that predicts student grades, one model for each school. The second is a school-level regression that uses the pre-test coefficients of the schools as the target, and school traits as predictors.

For this data setup, you may have a *varying intercept model*, in which only intercepts vary between schools:

$$y_i = \alpha_{j[i]} + \beta x_i + \epsilon_i \qquad \text{for students i} = 1, ..., \text{n}$$

$$\alpha_j = a + bu_j + \eta_j \qquad \text{for schools j} = 1, ..., \text{J}$$

You can also have a *varying intecept, varying-slope model*:

$$y_i = \alpha_{j[i]} + \beta x_i + \epsilon_i \qquad \text{for students i} = 1, ..., \text{n}$$

$$\alpha_j = a_0 + b_0 u_j + \eta_{j1} \qquad \text{for schools j} = 1, ..., \text{J}$$

$$\beta_j = a_1 + b_1 u_j + \eta_{j2} \qquad \text{for schools j} = 1, ..., \text{J}$$

Labels

Multilevel models are also called hierarchical models because data may structured such that one group (students) is a subgroup of the other (schools). There is also a hierarchy in the model, with the within-school parameters at the bottom and school hyperparameters at the top.

There exist multilevel data structures that aren't technically nested. In these cases, the label multilevel model is more appropriate.

On random effects

Multilevel models are also referred to as random-effects or mixed-effects models.

Here, the regression coefficients being modeled are the random effects. Fixed effects correspond to parameters that don't vary by group, or to parameters that vary but aren't modeled.

A mixed-effects model includes both fixed and random effects.

In this book, the authors choose "multilevel models" as preferred terminology over "mixed-effects models". They give two reasons: - Fixed effects is a special case of random effects, where variance is set to 0. Thus technically all regression parameters are random, and multilevel can be used as an umbrella term. - Fixed, random, and mixed effects can be confusing and misleading terminology

Motivations for multilevel modeling

- Learning about treatment effects that vary by group
- Using all data to perform inferences for groups with small sample size
- Prediction
- Analysis of structured data
- More efficient inference of regression parameters
- Including predictors at two different levels