Chapter 11 Multilevel Models

Generalize multiple regression allowing intercept or slope or or slope & intercept to vary by group. $j[i]$ describes which of J groups contains the ith point.

$$y_i = \beta_{0j[i]} + \beta_1 x_i + \epsilon_i; \quad i = 1, \ldots, N; \; j = 1, \ldots, J$$

$$y_i = \beta_0 + \beta_{1j[i]} x_i + \epsilon_i; \quad i = 1, \ldots, N; \; j = 1, \ldots, J$$

$$y_i = \beta_{0j[i]} + \beta_{1j[i]} x_i + \epsilon_i; \quad i = 1, \ldots, N; \; j = 1, \ldots, J$$

Picture these models.

Clustered Data

We'd like to estimate the effect of city policies on enforcing child support payments.
Take into account individual variables like mom's race, dad's age, informal support (response?), city. (1367 rows of data)
and
City-wide variables: enforcement, benefit level. 20 rows of data.

- Individual-level regression uses individual responses (on the left), pretends that we're all in one city except for different policies. But cities differ in many ways.
- City-level analysis uses aggregate response and aggregate predictors like proportion hispanic.
- Two-steps: Run logistic regression on individuals with an "intercept" for each city, then treat the fitted city intercepts as data for a second city level regression.

Multilevel Models

Fit both of the two-step models at once.

$$\Pr(y_i = 1) = \text{logit}^{-1}(\mathbf{X}_i\boldsymbol{\beta} + \alpha_{j[i]}), \text{ for } i = 1, \ldots, n$$

with $\alpha_{j[i]} \sim N(\mathbf{U}_j\boldsymbol{\gamma}, \sigma_\alpha^2)$ for $j = 1, \ldots, 20$.
$\mathbf{X}_i$ is a row of the individual data matrix, U_j is a row of the city-level data matrix.
Note: this is observational, policies cannot be "set" as in an experiment. Need to measure all important "pre-treatment" variables and adjust for them.

§11.3 Repeated Measurements

Often we repeat the same measurements over time. For example 2000 adolescents were surveyed about smoking every 6 months.
$\mathbf{X}$ then contains variables which can change over time (age, date)
$\mathbf{U}$ contains info which does not change (gender, do parents smoke?) $u_i = 1$ if either parent smokes, $u_2 = 1$ for females, 0 for males.

$$\begin{aligned}\Pr(y_{jt} = 1) &= \text{logit}^{-1}(\beta_0 + \beta_1 u_{1j} + \beta_2 u_{2j} + \\ &\quad \beta_3(1 - u_{2j})t + \beta_4 u_{2j} t + \alpha_j)\end{aligned}$$

Wide Data Format

Each person is on a single line. Line length may vary, as each wave of the study adds two more columns.

ID	sex	msmoke	dsmoke	age1	smoke1	age2	smoke2	⋯
1	f	Y	Y	15:0	N	15:6	N	⋯
2	f	N	N	14:7	N	15:1	N	⋯
3	m	Y	N	15:1	N	15:7	Y	⋯
4	f	N	N	15:3	N	15:9	N	⋯
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋱

Long Data Format

Each observation is on its own line. Individuals will appear in multiple lines.

ID	sex	msmoke	dsmoke	age	smoke	wave
1	f	Y	Y	15:0	N	1
2	f	N	N	14:7	N	1
3	m	Y	N	15:1	N	1
4	f	N	N	15:3	N	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
1	f	Y	Y	15:6	N	2
2	f	N	N	15:1	N	2
3	m	Y	N	15:7	Y	2
4	f	N	N	15:9	N	2

Converting Wide to Long in R

- `stack` and `unstack` (not for factors)

```
longDFage <-  stack(wideDF, select=c("age1","age2","age3") )
longDFsmoke <-  stack(wideDF,  select=c("smoke1","smoke2","smoke3")
```

- `reshape` function

```
longDF <-  reshape(wideDF, varying=list(age=c("age1","age2","age3")
                                smoke=c("smoke1","smoke2","smoke3"),
                   direction="long")
```

- `reshape` package (Hadley Wickham) lets you `melt` and `recast` data frames.

They provide these data in Long format with `parsmk` (one or both parents smoke)

Time Series Cross-Sectional Data

Example: Surveys of opinion on death penalty taken over 23 years and in 34 states.

Allow us to compare states and look for time trends. Two clustering variables: states and times (and interactions). Years are crossed (not nested) with states.

Other crossed variables: Job category and state used to predict earnings.

$$y_i = \mathbf{X}_i\boldsymbol{\beta} + \alpha_{j[i]} + \gamma_{k[i]} + \epsilon_i$$

$$\alpha_j \sim N(\mathbf{U}_j a, \sigma_\alpha^2), \quad \gamma_k \sim N(\mathbf{V}_k g, \sigma_\gamma^2)$$

§11.4 Factors and Indicators

- "Fixed" effects: a J level factor is coded as a baseline (in the intercept) and $J-1$ indicators (adjustments to intercept)
- "Random" effects or Multilevel models: all coefficients have a common distribution (usually gaussian about a common mean, with common variance).
  Note: With two levels of random effects, one group of coefficients will need to be centered at 0 so that we don't have two "intercept" columns in the data matrix.

More Fixed vs Random

Read footnote p 245 – five definitions of "random" effects.

When to use fixed vs random?

- Fixed if group-level coefficients are of interest, random if we care about the underlying population mean and variance of effects.
- Fixed if all groups are represented, random if we have a sample of groups.
- G&H advice: always use multilevel models (random effects) and separate out groups of coefficients for further modeling $(\alpha_1, \ldots, \alpha_J)$

§11.5 Cost/Benefits

Classical Regression:

- Prediction for continuous or binomial (or Poisson) responses.
- Transformed response fixes nonconstant variance, nonlinear model.
- Categorical predictors as indicator variables.
- Interaction effects can be addressed.
- Causal inference for randomized (or ignorable) treatments.

Multilevel Models:

- Account for extra variation. City–child support example: classical regression could use city-level predictors, but city-to-city variation goes into overall error.
- Model distribution of individual-level coefficients.
- Blend estimates together as in radon in counties.

Complexity

More levels $\Rightarrow$ more complexity, but hey, that's the way life is.

Have to add more assumptions: random variables in each level must be independent of those in other levels. Equal spread, independence, proper model, normality.

Worth it?
Multilevel models fit more complex data in ways not suited to classical regression models. In limits ($\sigma_\alpha^2 \to \infty$ or $\sigma_\alpha^2 \to 0$), multilevel model and classical do agree.