Chapter 12: Multilevel linear models with varying-intercepts

Partial pooling with no predictors

Multilevel regression can be thought of as a method of comprimising between complete pooling (estimating average without differentiating groups) and no pooling (estimating average for each group).

Complete-pooling and no-pooling regression estimates

Complete pooling: Estimate average across all groups, ignoring crucial variation among groups.

No pooling: Estimate average for each group, overstating the variation.

Partial-pooling estimates from a multilevel model

Multilevel estimates represent a compromise between complete- and partial-pooling.

Multilevel estimate for a given group j can be approximated as weighted average of mean of observations in group $(\bar{y}_j)$ and mean over all groups $(\bar{y}_{all})$.

$$\hat{\alpha}_j^{multilevel} = \frac{\frac{n_j}{\sigma_y^2}\bar{y}_j + \frac{1}{\sigma_\alpha^2}\bar{y}_{all}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}}$$

Weighted average reflects relative information available about the individual group and the average of all the groups.

- If a group has a smaller sample size, more weight is given to the overall average
- If a group has a larger sample size, more weight is given to the group average

To apply the formula, you need estimates of variation within and between groups.

Partial pooling with predictors

In this scenario, no-pooling can refer to fitting a separate regression model within each group, or classical regression model that includes group indicators.

Complete-pooling regression

Recall, this will not take into account group-level variation; it pools all observations into a single group

$$y_i = \alpha + \beta x_i + \epsilon_i$$

No-pooling regression

Includes an indicator for each group (different intercept for each group, same slope)

$$y_i = \alpha_{j[i]} + \beta x_i + \epsilon_i$$

Note on R formulas: The `-` operator removes the specified terms, so that `(a+b+c)^2 - a:b` is identical to `a + b + c + b:c + a:c`. It can also used to remove the intercept term: when fitting a linear model `y ~ x - 1` specifies a line through the origin.

Problems with no-pooling and complete-pooling analyses

Complete-pooling ignores group-level variation, which is undesirable if you are analyzing those groups. No-pooling overstates variance between groups. Moreover, groups with small sample size result in un-trustworthy models.

Multilevel analysis

A simple multilevel model for data with one predictor can be written as

$$y_i \sim N(\alpha_{j[i]} + \beta x_i, \sigma_y^2), \qquad \text{for i} = 1, \ldots, \text{n}$$

where

$$\alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2), \qquad \text{for j} = 1, \ldots, \text{n}$$

The distribution of α_j has the effect of its estimate toward the mean μ_α, like a soft constraint. When $\sigma_\alpha \to \infty$ the soft constraint does nothing, leading to no-pooling. When $\sigma_\alpha \to 0$ you hae a hard constraint on α_j, resulintg in complete-pooling.

Groups with small sample sizes will have stronger pooling (closer to overall average model), whereas groups with larger sample sizes will have weaker pooling (closer to group-level model).

Average regression line and individual- and group-level variances

To examine multilevel models, it's more common to plot group-level models and varying parameters, as opposed to examining numerical estimates. It can be helpful to look at numerical summaries of hyperparameters though.

For example, if you have $\hat{\mu}_\alpha = 1.46$, $\hat{\beta} = -0.69$ then the regression line for all groups is $y = 1.46 - 0.69x$. If $\hat{\sigma}_y = 0.76$ and $\hat{\sigma}_\alpha = 0.33$ then you have error standard deviations of 0.76 at the individual level and 0.33 at the group level. Note that $\hat{\sigma}_y$ is within-group variance for all groups.

One way to interpret variation between counties σ_α is to compare them to the variance ratio $\frac{\sigma_\alpha^2}{\sigma_y^2} = \frac{0.33^2}{0.76^2} = 0.19$.

This tells you that the variance between groups is same as the variance of the average of 5 measurements in a group. Practically, this means that a group with a sample size > 5 has more information (lower variance estimate) in the within-group model (no-pooling) than using the group-level model (complete-pooling).

In summary:

- The multilevel regression line in a group is closer to complete-pooling when sample size N < 5
- The multilevel regression line in a group is closer to no-pooling when sample size N > 5

Intraclass correlation

The relative values of individual- and group-level variances are known as **intraclass correlation**,

$$\frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_y^2}$$

When this value is 0, $\sigma_\alpha^2 = 0$ and the grouping conveys no information. When the value is 1, $\sigma_y^2 = 0$ and all members of a group are identical.

A way to understand ICC is that is the proportion of model variance that is explained by group differences.

Partial pooling (shrinkage) of group coefficients α_j

Multilevel modeling partially pools group-level parameters α_j toward the mean level, μ_α. There is more pooling when the group-level standard deviation σ_α is small, and more smoothing for groups with fewer observations.

The multilevel-modeling estimate of α_j can be expressed as the weighted average of the no-pooling estimate for its group $(\bar{y}_j - \beta \bar{x}_j)$ and the mean μ_α.

Isolating α_j in the no-pooling estimate of $\bar{y}$ in $\bar{y} = \alpha_j + \beta \bar{x}_j$ gives $\alpha_j = \bar{y} - \beta \bar{x}_j$. The multilevel-modeling estimate of α_j can then be represented as a weighted average of the no-pooling estimate for the group, and the complete-pooled mean:

$$\alpha_j \approx \frac{\frac{n_j}{\sigma_y^2}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}} (\bar{y} - \beta \bar{x}_j) + \frac{\frac{1}{\sigma_\alpha^2}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}} \mu_\alpha$$

Classical regression as special case

$\sigma_\alpha \to 0$ yields complete-pooling and $\sigma_\alpha \to \infty$ yields no-pooling model.

Fitting multilevel models in R

Via `lmer` – mixed effects refers to random effects (coefficients that vary by group) and fixed effects (coefficients that do not vary).

Commands

Varying-intercept model with no predictors

```
mod <- lmer(y ~ 1 + (1 | county))
```

Varying-intercept model with individual-level predictor

```
mod <- lmer(y ~ x + (1 | county))
```

Model summary

To get a summary, run `display(mod)`. This gives you intercept and variable inferences, standard deviation at both group- and individual-level, and tells you how many data points and groups were fitted.

Getting coefficients

`coef(mod)` gives the estimated regression line within each group.

You can also look at the fixed effects (average for groups) and random effects (group-level errors) individually via `fixef(mod)` and `ranef(mod)`.

Fixed effects refers to coefficients that don't vary by group, for group-level coefficients, or for group averages.

More complicated models

`lmer()` works well when the sample size and number of groups is moderate to large. When number of groups is small or model becomes complicated, it can be better to use Bayesian inference.

Five ways to express a multilevel model

1. As an extension of classical regression, where coefficients vary across groups

In other words, your classical regression model

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \epsilon_i$$

becomes

$$y_i = \beta_{j[i]} + \beta_{j[i]} X_{i1} + \beta_{j[i]2} X_{i2} + \ldots + \epsilon_i$$

The group-level predictor here is X_{i2} and is encoded in the first level of the equation.

This is multilevel because a multivariate distribution is assigned to β within each group.

In the case of a varying-intercept model, the model would be

$$y_i = \alpha_{j[i]} + X_i \beta + \epsilon_i \qquad \text{1st-level}$$

and

$$\alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2) \qquad \text{2nd-level}$$

where X and β don't include the intercept.

2. As a combination of separate local regressions

Within each county j, fit a local regression

$$y_i \sim N(\alpha_j + \beta x_i, \sigma_y^2) \qquad \text{1st-level}$$

where the group-level predictor enters in at the second level of the model as μ_j

$$\alpha_j \sim N(\gamma_0 + \gamma_1 \mu_j, \sigma_\alpha^2) \qquad \text{2nd-level}$$

where μ_j is a group-level predictor.

Note that the local regression coefficients are the same in all J models, but the different intercepts are connected through the second-level model.

3. As a large regression model

You can combine local and group-level predictors into a single matrix X as well:

$$y_i \sim N(X_i\beta, \sigma_y^2)$$

with group indicators

$$\beta_j \sim N(0, \sigma_\alpha^2) \qquad\qquad \text{for j} = 3, \ldots, \text{J+2}$$

Varying the group indicator β_j encodes variance between group coefficients.

Here, β_0 is the intercept, β_1 is the individual-level predictor, β_2 is the group-level predictor.

4. As a regression with multiple error terms

You can re-express the model in (3), treating group indicator coefficients as error terms instead of regression coefficients.

This is commonly referred to as the mixed-effects model.

$$y_i \sim N(X_i\beta + \eta_{j[i]}, \sigma_y^2) \qquad\qquad \text{for i} = 1, \ldots, \text{n}$$

$$\eta_j \sim N(0, \sigma_\alpha^2)$$

where $j[i]$ represents the county that contains house i, and X = [constant term, individual predictor, group predictor]

5. As a large regression with correlated errors

$$y_i = X_i\beta + \epsilon_i^{all}, \epsilon^{all} \sim N(0, \Sigma)$$

When is multilevel modeling most effective?

Multilevel modeling are most effective when the model estimates veer closer to complete pooling. In this setting, estimates can vary by group, while still being estimated precisely.

When the multilevel esimate is close to complete pooling, it still allows for variation between groups. It also handles groups of both small and large sizes well.

Statistical significance

Statistical significance is not an appropriate criterion for including group indicators in a multilevel model. The purpose of multilevel models is not to see whether effects are significantly different between groups; rather, the goal is to estimate group effects for each group, while appropriately accounting for uncertainty.

How many groups and observations are needed to fit a multilevel model?

In general, you want to use multilevel modeling when you have data with many groups. When the number of groups is small, it's difficult to estimate between-group variation so multilevel modeling brings little to the table.

There isn't a great rule of thumb in terms of sample size, but you can have as little as one per group. The only issues is that you'll get imprecise estimations.