Multilevel Models

Patrick Lam

Outline

Outline

Multilevel Data

Suppose we have multilevel data (perhaps cross-country survey data) in which we have

- n individuals indexed by i
- J groups indexed by j where each individual i belongs to one group j

We may have more than two levels (individuals within countries within regions) or non-nested levels (individuals within certain age groups but also within certain levels of education).

Levels can be interpreted in different ways (in TSCS data, we have countries over time where each observation is an individual and observations are grouped by country or time).

We may also have both individual and group-level covariates.

Running Example

As a running example, we will use a dataset on levels of radon in houses in different counties.

```
> radon <- read.csv("radon.csv")
```

- Dependent variable (`log.radon`): the log of the measured radon level in a house
- Independent variable (`floor`): dummy variable for whether the measurement was taken on the first floor (1) or the basement (0)
- The houses are grouped by `county` (85 counties total).
- Group-level covariate (`u.full`): a measurement of soil uranium at the county level

So our units are houses and the houses are grouped by county.

How do we deal with the multilevel nature of the data?

1. We can assume the multilevel nature of the data is irrelevant by pooling all the observations together as if the groupings did not exist (complete pooling model).

2. We can assume the multilevel nature of the data gives us information by allowing the coefficients to be completely different for each group and estimating them individually for each group (no pooling model).

3. We can allow the coefficients for each group to be different but drawn from a common distribution (multilevel model).

Options 2 and 3 allow us to model unobserved heterogeneity not captured by our covariates (for example, there may be something special about group A that is not captured by our covariates)

Outline

The Complete Pooling Model

Suppose we have the following linear model:

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

The **complete pooling model** assumes that the groupings provide no information by constraining the coefficients to be the same for all groups (we basically pool all the observations together into a single group).

This is just the normal regression model where α and β do not vary by j.

```
> complete.pool <- lm(log.radon ~ floor, data = radon)
```

The complete pooling approach may be too restrictive in that it ignores any variation in the coefficients between groups (for example, β might vary over time or across countries.)

Outline

The No Pooling Model

In the **no pooling model**, one or more of the coefficients are allowed to vary by group.

For example, we can let the intercept differ by group:

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

where the notation $\alpha_{j[i]}$ denotes the α corresponding to the group j that individual i belongs to.

This model is commonly known as the **fixed effects model**.

The varying intercepts attempt to model unobserved heterogeneity between groups.

For example, in political science papers, you usually see models that use things like "country fixed effects".

The assumption is that there might be something different about being in a specific country that cannot be captured by the covariates.

Therefore, each country is giving its own intercept (fixed effect) to allow for a different baseline level of y (when all covariates equal 0).

The "fixed effects" terminology can be very confusing since it can also refer to slightly different quantities (as we will see later).

We can also let the intercept and slope(s) differ by group:

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

This model is less common.

It assumes that the relationship between x and y is different for each group.

One should almost never have a varying slope without a varying intercept unless there is a good theoretical reason to do so.

Dummy Variables Regression

A common way to estimate the fixed effects model is to include $J - 1$ dummy indicators for the J groups into the regression model (with one group left out as the baseline category)

This is sometimes called *dummy variables* regression.

Suppose our individuals belong to one of three groups. Let g_j be a dummy variable that takes on a value of 1 if i is in group j and 0 if not.

The varying intercept (fixed effects) model is

$$y_i = \alpha + \beta_1 x_i + \beta_2 g_2 + \beta_3 g_3 + \epsilon_i$$

where

$$\begin{aligned} \alpha_1 &= \alpha \\ \alpha_2 &= \alpha + \beta_2 \\ \alpha_3 &= \alpha + \beta_3 \end{aligned}$$

β_2 and β_3 can be interpreted as the effect on y of being in group 2 and group 3 relative to group 1 (the baseline category).

In our example, we can add "county fixed effects" by including county dummies in the regression.

```
> no.pool <- lm(log.radon ~ floor + as.factor(county), data = radon)
```

Outline

The Hierarchical Model

The **hierarchical model** is a compromise between the complete pooling and no pooling models in that it partially pools the coefficients by allowing them to differ according to a common distribution (usually the normal distribution).

Hierarchical models are also known as partially pooled models, multi-level models, mixed effects models, random effects models, or random coefficient models.

The varying intercept model can be written as

$$\begin{aligned} y_i &= \alpha_{j[i]} + \beta x_i + \epsilon_i \\ \alpha_j &\sim N(\alpha, \sigma_\alpha^2) \end{aligned}$$

or equivalently as

$$\begin{aligned} y_i &= \alpha_{j[i]} + \beta x_i + \epsilon_i \\ \alpha_j &= \alpha + \eta_j \\ \eta_j &\sim N(0, \sigma_\alpha^2) \end{aligned}$$

The estimate of the variance term σ_α^2 gives us a measure of how different the coefficients are between groups.

- The complete pooling model assumes $\sigma_\alpha^2 = 0$
- The no pooling model assumes $\sigma_\alpha^2 = \infty$

We can also incorporate group-level covariates into the model.

Suppose we think that some group-level covariate u has an effect on the average y within groups. We can think of u as affecting the intercepts α_j.

$$\begin{aligned} y_i &= \alpha_{j[i]} + \beta x_i + \epsilon_i \\ \alpha_j &\sim N(\gamma_0 + \gamma_1 u_j, \sigma_\alpha^2) \end{aligned}$$

or equivalently

$$\begin{aligned} y_i &= \alpha_{j[i]} + \beta x_i + \epsilon_i \\ \alpha_j &= \gamma_0 + \gamma_1 u_j + \eta_j \\ \eta_j &\sim N(0, \sigma_\alpha^2) \end{aligned}$$

We can interpret γ_1 as the effect of u on the average y within a group when all the individual-level covariates are 0.

The varying intercept and slope(s) model can be written as

$$\begin{aligned} y_i &= \alpha_{j[i]} + \beta_{j[i]} x_i + \epsilon_i \\ \begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} &\sim N\left(\begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix} \right) \end{aligned}$$

or with group-level covariates

$$\begin{aligned} y_i &= \alpha_{j[i]} + \beta_{j[i]} x_i + \epsilon_i \\ \begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} &\sim N\left(\begin{pmatrix} \gamma_0^\alpha + \gamma_1^\alpha u_j \\ \gamma_0^\beta + \gamma_1^\beta u_j \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix} \right) \end{aligned}$$

We can interpret γ_1^β as the effect of u on the relationship between x and y within a group.

Hierarchical models are often referred to as **random effects models** because of the randomness of the distribution of coefficients.

However, they are also commonly called **mixed effects** models because they consist of two types of effects:

- *fixed effects*, which do not vary across groups
- *random effects*, which do vary across groups

The terminology is confusing because the fixed effects referred to here are NOT the same as the fixed effects in the no pooling model, which refer to the varying intercepts.

In the simple varying intercepts hierarchical model, we can identify the fixed effects and random effects components.

$$\begin{aligned} y_i &= \alpha_{j[i]} + \beta x_i + \epsilon_i \\ \alpha_j &= \gamma_0 + \gamma_1 u_j + \eta_j \end{aligned}$$

By contrast, in the no pooling model, the "fixed effects" referred to α_j, which was allowed to vary.

The confusion in terminology have led some (Gelman, among others) to call for the end to the use of the terms "fixed effects" and "random effects".

Nevertheless, people still use them.

In substantive papers, the mention of "fixed effects" almost always refers to the no pooling model with intercepts varying by group.

Outline

Fixed Effects versus Random Effects

Unless there is a theoretical expectation that a complete pooling model is appropriate (when there is nothing important about the groupings), one should always take advantage of the information in the hierarchical nature of the data.

However, which model is better for dealing with hierarchical data: the fixed effects (no pooling) model or the random effects (partial pooling/hierarchical/mixed effects) model?

Many people have suggested reasons for using one over the other, such as

- use fixed effects if the group-level coefficients are of interest, and use random effects if the interest lies in the underlying population
- use fixed effects when the groups in the data represent all possible groups, and use random effects when the population includes groups not in the data
- use fixed effects when the number of groups is small (less than five) since there may not be enough to accurately estimate the variation in the coefficients

Gelman suggests always using the random effects model because it is less restrictive (the variance of the coefficients is estimated rather than assumed to be infinite).

Outline

Outline

The Complete Pooling Model

The complete pooling model is the simple linear regression model, which ignores the groupings:

$$y_i = \alpha + \beta \cdot floor + \epsilon_i$$

```
> complete.pool <- lm(log.radon ~ floor, data = radon)
> complete.pool

Call:
lm(formula = log.radon ~ floor, data = radon)

Coefficients:
(Intercept)        floor
      1.327       -0.613
```

Outline

The No Pooling Model

The no pooling (fixed effects) model adds dummy indicator variables for the groupings ("county fixed effects").

We have 86 coefficients corresponding to the intercept (baseline county), slope, and $m - 1$ (84) county dummies.

```
> length(coef(no.pool))
```

```
[1] 86
```

Alternatively, we could have estimated this without an intercept term but with m county dummies.

```
> no.pool.1 <- lm(log.radon ~ floor - 1 + as.factor(county), data = radon)
> length(coef(no.pool.1))
```

```
[1] 86
```

Outline

The Hierarchical Model

Remember that in our hierarchical/mixed effects model, we had two components: fixed effects and random effects.

Our varying intercept model is

$$\begin{aligned} y_i &= \alpha_{j[i]} + \beta x_i + \epsilon_i \\ \alpha_j &= \gamma_0 + \gamma_1 u_j + \eta_j \end{aligned}$$

We can substitute the second equation into the first to get

$$y_i = (\gamma_0 + \eta_{j[i]}) + \beta x_i + \gamma_1 u_{j[i]} + \epsilon_i$$