Representing models mathematically

Luke Miratrix

September 17, 2019

Abstract

This document runs through two complicated example multilevel models to illustrate different mathematical notation for the same thing, as well as how to move from one representation to another.

1 The Scenario

We have a collection of schools that we have randomized into treatment and control conditions. The treatment condition is a novel reading program and the control condition is business as usual. We hope that the treatment accomplishes two things: raising reading level overall, and reducing the gap in reading level between "at-risk" kids and not at risk kids (we assume we have a at-risk status as a dummy variable, measured for all kids and treatment and control prior to treatment).

2 The Two-Level Random Intercept Model

We will use the "double-indexing" that is the most common notation for multilevel models (not the Gelman and Hill bracket ($j[i]$) notation). Treatment is at the *school level*: let Z_j be an indicator of whether school j was treated (so a 0/1 variable). Then, for student i in school j we have

$$Y_{ij} = \alpha_j + \beta_1 R_{ij} + \beta_2 X_{ij} + \epsilon_{ij}$$
$$\alpha_j = \gamma_0 + \gamma_1 Z_j + \gamma_2 S_j + u_j$$

with Y_{ij} being the reading level of the student, R_{ij} being a dummy variable of student's "at risk" status, X_{ij} being an important student demographic variable (e.g., prior reading level), and S_j being a school-level covariate (such as a school quality measure).

This is the two-level model. Level 1 is the first equation with the distribution on the residuals of $\epsilon_{ij} \sim N(0, \sigma^2)$. Level 2 is the second equation with a distribution of random effects of

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

The σ_α^2 is the variance of the random intercept.

Call the β_{0j} the random intercept and u_j the random effect. The u_j is the residual of the level 2 model,. In R, we would say `coef()` for the intercept (including the mean γ_0) and `ranef()` for the random effect. In math, `coef()` gives $\gamma_0 + u_j$ and `ranef()` gives only u_j. Neither include the γ_1 or γ_2; these will be separate columns you get from `coef()`.

Remarks. We have *completely pooled* the coefficient for R_{ij} and X_{ij}: we are assuming all the schools have the same relationship between the outcome and these covariates.

The intercept α_j is the expected (predicted average) outcome of a not-at-risk student with $X_{ij} = 0$. Different schools have different means. In particular, treatment schools have a mean of γ_1 more than control; this is the treatment impact.

2.1 The Reduced Form Model

If we plug in our 2nd level into the first we get the following:

$$\begin{aligned}
Y_{ij} &= \beta_{0j} + \beta_1 R_{ij} + \beta_2 X_{ij} + \epsilon_{ij} \\
&= (\gamma_0 + \gamma_1 Z_j + \gamma_2 S_j + u_j) + \beta_1 R_{ij} + \beta_2 X_{ij} + \epsilon_{ij} \\
&= \gamma_0 + \gamma_1 Z_j + \gamma_2 S_j + \beta_1 R_{ij} + \beta_2 X_{ij} + (u_j + \epsilon_{ij})
\end{aligned}$$

The $u_{0j} + \epsilon_{ij}$ is our total random error. It is how much our prediction of a new, unknown student, would differ from their actual score if we didn't know the school's random effect. The rest of the model is the mean model or structural portion of the model.

This is also called the *reduced form*; it is what econometricians work with. They will write the entire residual as ε_{ij}, however:

$$Y_{ij} = \gamma_0 + \gamma_1 Z_j + \gamma_2 S_j + \beta_1 R_{ij} + \beta_2 X_{ij} + \varepsilon_{ij}$$

Remarks. The reduced form helps us see our treatment effect more clearly. It is a shift in outcome of γ_1 for treated students.

The γ_0 is the overall mean reading level for students with $X_{ij} = 0$ for not-at-risk students ($R_{ij} = 0$) in control schools with $S_j = 0$.

Note how we subscript school-level covariates with only a j vs. individual-level covariates get an ij. If you want, you can index everything by ij; the fact that S_{ij} will then be the same for all students i in school j is hidden in the data. But it does make it look very much like OLS with a weird error term:

$$Y_{ij} = \gamma_0 + \gamma_1 Z_{ij} + \gamma_2 S_{ij} + \beta_1 R_{ij} + \beta_2 X_{ij} + (u_j + \epsilon_{ij})$$

Finally, you might call all the different pieces by different letters to indicate whether you care about them or not. E.g.,

$$Y_{ij} = \mu + \tau Z_{ij} + \beta_1 S_{ij} + \beta_2 R_{ij} + \beta_3 X_{ij} + (u_{0j} + \epsilon_{ij}).$$

Here τ is our treatment effects of interest. The β's are just adjustments to be ignored. The μ is the grand mean (for those not treated, with $S_{ij} = 0$ and $R_{ij} = 0$ and $X_{ij} = 0$). People often use μ for mean and τ for treatment.

2.2 Fitting it in lmer

We fit it as:

```
lmer( Y ~ R + Z + X + S + (1|id), data=dat )
```

3 The Two-Level Random Slopes Model

Now let's get very complex to really unpack notational stuff. We are going to let treatment not only impact the average outcome in schools, but also allow treatment to differentially impact students who are "at risk". I.e., we are going to have two treatment impacts, one for not at risk, and one for at risk. This is an interaction of risk status and treatment.

Furthermore, we are going to let different schools have different gaps between at risk and not at risk, but allowing a random effect for the at risk coefficient.

Using our same variables as above, we have, for student i in school j

$$Y_{ij} = \beta_{0j} + \beta_{1j}R_{ij} + \beta_2 X_{ij} + \epsilon_{ij}$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + \gamma_{02}S_j + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + u_{1j}.$$

This is the two-level model. Level 1 is the first equation with the distribution on the residuals of $\epsilon_{ij} \sim N(0, \sigma^2)$. Level 2 are the second and third equations, and the distribution of random effects of

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{10} \\ \tau_{10} & \tau_{11} \end{pmatrix} \right]$$

The τ_{00} is the variance of the random intercept. τ_{11} is the variance of the random slope. τ_{10} is the covariance (not correlation) of the random effects. To get the *correlation* of random effects we have $\rho = \tau_{10}/\sqrt{\tau_{00}}\sqrt{\tau_{11}}$. (Note that $\tau_{10} = \tau_{01}$, meaning the covariance of A and B is the same as covariance of B and A, so we just write one of them.)

Call the β_{0j} the random intercept and β_{1j} a random coefficient. We might call them both random coefficients. Call the u_{0j}, u_{1j}, which are the residuals of the level 2 models, the random effects. In R, we would say `coef()` for the coefficients (including the means) and `ranef()` for the random effects.

Remarks. We have *completely pooled* the coefficient for X_{ij}: we are assuming all the schools have the same relationship between the outcome and X_{ij}. This is why we have no level 2 equation for β_2 and we do not index β_2 as β_{2j}.

The intercept β_{0j} is the expected (predicted average) outcome of a not-at-risk student with $X_{ij} = 0$. Different schools have different means.

The achievement gap of at-risk and not-at-risk students for control schools is measured by γ_{10}. For treatment schools it is $\gamma_{10} + \gamma_{11}$.

The γ_{01} is the average treatment effect for not-at-risk students. Then $\gamma_{01} + \gamma_{11}$ is the average treatment effect for the at-risk students. If we find $\gamma_{11} \neq 0$ then the average effects differ for our two types of students, and the change in the achievement gap induced by treatment is measured by γ_{11}.

In this model, the school-level covariate explains overall differences in reading between schools, but does not relate to the size of treatment impact in a school, or relate to the at-risk vs. not-at-risk achievement gap.

The level 2 covariate matrix. Sometimes people like to write the correlation matrix using other parameterizations. E.g., we might see

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \rho \sigma_\alpha \sigma_R \\ & \sigma_R^2 \end{pmatrix} \right]$$

to indicate the cross-school variation in the intercept (α) and the risk gap (R). Now we specifically have written our correlation of random effects as ρ.

3.1 The Reduced Form Model

If we plug in our 2nd level into the first we have to plug in both equations. If we do we get... a mess:

$$\begin{aligned}
Y_{ij} &= \beta_{0j} + \beta_{1j}R_{ij} + \beta_2 X_{ij} + \epsilon_{ij} \\
&= (\gamma_{00} + \gamma_{01}Z_j + \gamma_{02}S_j + u_{0j}) + (\gamma_{10} + \gamma_{11}Z_j + u_{1j})R_{ij} + \beta_2 X_{ij} + \epsilon_{ij} \\
&= \gamma_{00} + \gamma_{01}Z_j + \gamma_{02}S_j + u_{0j} + \gamma_{10}R_{ij} + \gamma_{11}Z_j R_{ij} + u_{1j}R_{ij} + \beta_2 X_{ij} + \epsilon_{ij} \\
&= \gamma_{00} + \gamma_{01}Z_j + \gamma_{02}S_j + \gamma_{10}R_{ij} + \gamma_{11}Z_j R_{ij} + \beta_2 X_{ij} + u_{0j} + u_{1j}R_{ij} + \epsilon_{ij} \\
&= \gamma_{00} + (\gamma_{01} + \gamma_{11}R_{ij})Z_j + \gamma_{02}S_j + \gamma_{10}R_{ij} + \beta_2 X_{ij} + (u_{0j} + u_{1j}R_{ij} + \epsilon_{ij})
\end{aligned}$$

The $u_{0j} + u_{1j}R_{ij} + \epsilon_{ij}$ is our total random error. It is how much our prediction of a new, unknown student, would differ from their actual score if we didn't know the school's random effect. The rest of the model is the mean model or structural portion of the model.

This is our *reduced form*; it is what econometricians work with. They will write the entire residual as ε_{ij}, however:

$$Y_{ij} = \gamma_{00} + (\gamma_{01} + \gamma_{11}R_{ij})Z_j + \gamma_{02}S_j + \gamma_{10}R_{ij} + \beta_2 X_{ij} + \varepsilon_{ij}$$

Remarks. The reduced form helps us see our treatment effects and treatment variation across groups more clearly. We can put both terms involving the treatment indicator in parenthesis (final line above) to show how treatment is different by γ_{11} for the at-risk students.

We also see that the difference in treatment effects between at-risk and not at-risk is an *interaction* between student risk and treatment assignment of the school (note the $Z_j R_{ij}$ term).

The γ_{00} is the overall mean reading level for students with $X_{ij} = 0$ for not-at-risk students in control schools. The γ_{10} is the average difference between at-risk and not-at-risk students in control schools, across all schools.

We can rearrange our equations above to get

$$Y_{ij} = (\gamma_{00} + u_{0j}) + (\gamma_{01} + \gamma_{11}R_{ij} + u_{1j})Z_j + \gamma_{02}S_j + \gamma_{10}R_{ij} + \beta_2 X_{ij} + \epsilon_{ij}.$$

This shows the random intercept and random slope all bundled up.

As with the intercept model, you might call all the different pieces by different letters to indicate whether you care about them or not. E.g.,

$$Y_{ij} = \mu + \tau Z_{ij} + \beta_1 S_{ij} + \beta_2 R_{ij} + \beta_3 X_{ij} + \tau_R Z_{ij}R_{ij} + (u_{0j} + u_{1j}R_{ij} + \epsilon_{ij}).$$

Here τ and τ_R are our treatment effects of interest. The β's are just adjustments to be ignored. The μ is the grand mean. This model is the same as above, we are just changing names around.

3.2 Fitting it in lmer

We fit it as:

```
lmer( Y ~ R * Z + X + S + (R|id), data=dat )
```

Two other ways of saying the same thing:

```
lmer( Y ~ 1 + R * Z + X + S + (1 + R|id), data=dat )
```

and

```
lmer( Y ~ 1 + Z + S + R + Z:R + X + (1 + R|id), data=dat )
```

Remarks. See how the reduced form and `lmer()` align, especially if we write out what R automatically does with `R * Z` (R will expand `R*Z` into `R + Z + R:Z` automatically).

3.3 Another form for the two-level model

The above is the "double-subscript" way of writing a model. By contrast, Gelman and Hill index with a nifty "bracket notation." First, let $j[i]$ indicate the school student i is attending. Then we have:

$$Y_i = \beta_{0j[i]} + \beta_{1j[i]}R_i + \beta_2 X_i + \epsilon_i$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + \gamma_{02}S_j + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + u_{1j},$$

This is basically identical to the above, but if you are not familiar with the bracketing then things can get messy.

The advantage of this is we can then imagine each student gets their own unique id, i, and then we can query where that student is via $j[i]$. This can be useful when looking at crossed effects models, where units have different random effects for different things (e.g., for a test we might have observation k corresponding to a single answer for a test question, with $i[k]$ being the student who answered it and $q[k]$ being the question item).