Applied Multilevel Regression Modeling

Day 2: Random intercepts

Constantin Manuel Bosancianu

WZB Berlin Social Science Center
Institutions and Political Inequality
bosancianu@icloud.com

July 30, 2019

Yesterday

Clustered data poses serious problems for the standard regression estimation strategy (OLS).

Most important ones refer to heteroskedastic errors and biased standard errors.

Though some solutions to these problems exist (LSDV or Huber–White estimator for SEs), they are *post hoc* and atheoretical.

A good framework of analysis would treat heterogeneity of effects as something to be *modeled*, rather than a *nuisance*.

Today

MLMs: compromise between *no pooling* and *complete pooling* approach.

Notation for MLMs: not the Gelman–Hill convention (!).[1]

Assessing degree of clustering: ICC.

Simplest form of MLM: random intercept specifications.

[1] Will distribute at the end of class a small comparison document for the different MLM notation styles encountered.

Between *no* and *complete* pooling

Working example

ISSP *Citizenship II* module, data collected between 2014 and 2016: 35 countries and 48120 valid observations on political efficacy, education, and urban residence.

Outcome—political efficacy:

- ✓ "People like me. . . no say in what the government does"
- ✓ "Don't think government cares about much what people like me think"
- ✓ "Have a pretty good understanding of the important political issues"
- ✓ "Most people are better informed about politics [. . .] than I am"

Each item measured on 5-point Likert scale; final index is an average of the four (or fewer) items.

Complete pooling

Predictors:

- ✓ education: number of years of full-time education (0–35)
- ✓ urban residence: dichotomous (1=big city, or suburbs of big city)

Complete pooling: fit model on the entire sample, without taking into account group membership.

Implication: information from the entire sample is used in estimating parameters.

Complete pooling

	DV: Political efficacy
(Intercept)	2.315 (0.011)***
Education (years)	0.047 (0.001)***
Urban residence	0.083 (0.007)***
R^2	0.066
Adj. R^2	0.066
Num. obs.	48120
RMSE	0.749

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

No pooling

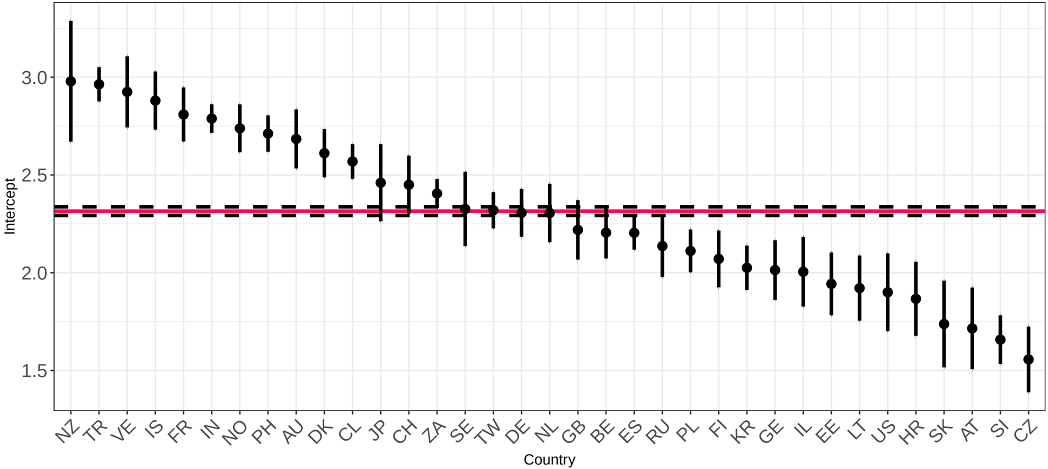
No pooling: fit same model separately for each group.[2]

Implication: the estimate for group *j* is obtained using information from just that group.

What kinds of questions can we answer with this approach?

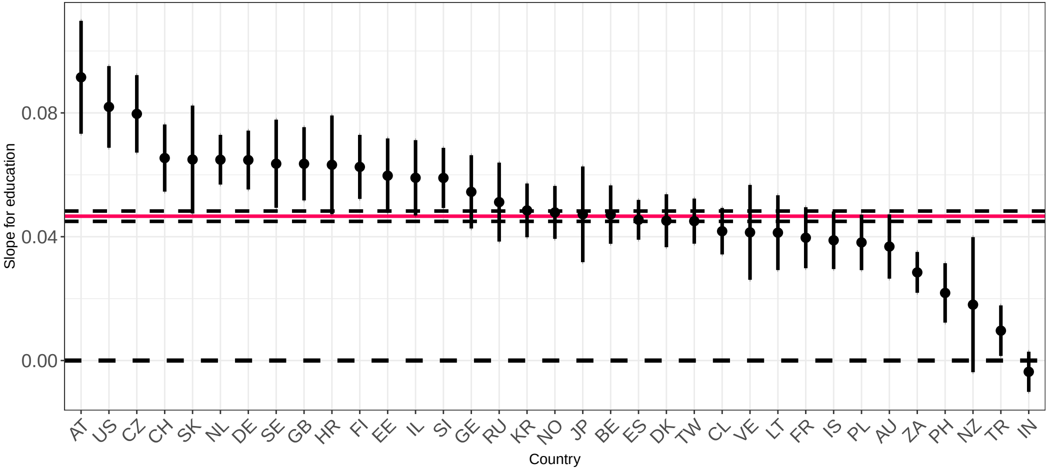
[2] A variant of this uses country dummies in the model.

Intercepts



Red line designates intercept from a *no pooling* model.

Slopes for education



Why the difference in CIs between, say, New Zealand and South Africa?

No vs. *Complete*

Complete pooling eliminates variation in estimates between groups, but minimizes uncertainty.

No pooling maximizes variation in estimates between groups, but results in maximum uncertainty.

Depending on group size, *no pooling* might also make groups seem more different from each other than they really are (Gelman & Hill, 2007).

Multilevel estimator

$$\hat{\alpha}_j^{MLM} \approx \frac{\frac{n_j}{\sigma_y^2}\bar{y}_j + \frac{1}{\sigma_\alpha^2}\bar{y}_{all}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}} \quad (1)$$

- $\bar{y}_j$: unpooled estimate
- $\bar{y}_{all}$: completely pooled estimate
- n_j: sample size in group j
- σ_y^2: within-group variance in outcome
- σ_α^2: variance in average level of outcome between groups

Multilevel estimator

$$\hat{\alpha}_j^{MLM} \approx \frac{\frac{n_j}{\sigma_y^2}\bar{y}_j + \frac{1}{\sigma_\alpha^2}\bar{y}_{all}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}} \tag{2}$$

Smaller $n_j \Rightarrow$ MLM estimate for group j is pulled closer to $\bar{y}_{all}$.

Greater homogeneity of groups $\Rightarrow$ greater differences in means between groups $\Rightarrow \sigma_y^2$ is lower and σ_α^2 is higher $\Rightarrow$ MLM estimate is pulled closer to $\bar{y}_j$.

MLM notation

From OLS to MLM (1)

$$EFF_i = \beta_0 + \beta_1 * EDU_i + \beta_2 * URB_i + \epsilon_i,\ with\, \epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2) \tag{3}$$

MLM extension is not that different.

$$EFF_{ij} = \beta_{0j} + \beta_{1j} * EDU_{ij} + \beta_{2j} * URB_{ij} + \epsilon_{ij} \tag{4}$$

i indexes level-1 units, while j indexes level-2 groups.

The equation denotes that the intercepts and slopes for each of the j groups are getting their own statistical specification.

From OLS to MLM (2)

For now, we only want a meaningful specification for the intercept.

$$\beta_{0j} = \gamma_{00} + \gamma_{01} * CORR_j + \upsilon_{0j} \tag{5}$$

Interpretation: the level of political efficacy in a country (controlling for level-1 factors) is associated with the level of (perceived) corruption in the country.

From OLS to MLM (3)

Though we could write a similar model for either, or both, of the level-1 slopes, let's keep them fixed for now.

$$\begin{cases} \beta_{1j} = \gamma_{10} \\ \beta_{2j} = \gamma_{20} \end{cases} \tag{6}$$

The implication: the effect of education on the level of political efficacy, and of urban residence on efficacy, *is identical* for each of the J groups.

From OLS to MLM (4)

Taken together, we have 4 equations, at 2 levels.

$$\begin{cases} EFF_{ij} = \beta_{0j} + \beta_{1j} * EDU_{ij} + \beta_{2j} * URB_{ij} + \epsilon_{ij} \\ \beta_{0j} = \gamma_{00} + \gamma_{01} * CORR_j + \upsilon_{0j} \\ \beta_{1j} = \gamma_{10} \\ \beta_{2j} = \gamma_{20} \end{cases} \tag{7}$$

Subscripts vary depending on whether the variable is measured at level-1 (`URB` or `EDU`) or at level-2 (`CORR`).

MLM specification—extended form

Extended form obtained by plugging in last 3 equations into the first one.

$$
\begin{aligned}
EFF_{ij} &= \gamma_{00} + \gamma_{01} * CORR_j + \upsilon_{0j} + \gamma_{10} * EDU_{ij} + \gamma_{20} * URB_{ij} + \epsilon_{ij} = \\
&= \gamma_{00} + \gamma_{10} * EDU_{ij} + \gamma_{20} * URB_{ij} + \gamma_{01} * CORR_j + \upsilon_{0j} + \epsilon_{ij}
\end{aligned}
\tag{8}
$$

Fixed effects: γ_{00}, γ_{10}, γ_{20}, and γ_{01}.

Random effects: υ_{0j}, and ϵ_{ij}.

The models which incorporate both began to be known as *mixed-effects models*.[3]

[3] The distinction between *fixed* and *random* comes from the experimental design literature. Gelman and Hill (2007) reject the name, which they consider confusing.

Fixed vs. random

Fixed-effects are interpreted in exactly the same way as coefficients in regression.

Random-effects, though, despite their name, are not interpreted as effects. They are, rather, variances of residuals.

The latter are useful to report—to the extent that they gradually become smaller, they indicate improvements in model fit.[4]

[4] A popular measure of model fit for multilevel models, called R^2, is based on these random-effects (Snijders & Bosker, 1999).

First model

	DV: Political efficacy
(Intercept) (γ_{00})	2.054^{***}
	(0.161)
Education (no. of years) (γ_{10})	0.044^{***}
	(0.001)
Urban residence (γ_{20})	0.053^{***}
	(0.007)
Perception of corruption (γ_{01})	0.005^{*}
	(0.002)
Num. obs.	48120
Num. groups: country	35
Var: country (Intercept) (υ_{0j})	0.069
Var: Residual (ϵ_{ij})	0.494

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

ICC

Are MLMs needed *always*?

Truthful answer is ***no***.

Go back to the formula for the estimator: The more heterogeneous the groups, the estimate is pulled toward $\bar{y}_{all}$, making a complete pooling model more attractive.

In extreme situations (groups are completely homogenenous inside, or are completely identical to each other), there's no need to invest in a MLM.

How can we tell when an MLM will be useful?

Clustering

We need a measure of how much clustering there is in a two-stage sample. This ties into the issue of *design effect* (Snijders & Bosker, 1999).

$$SE = \frac{SD}{\sqrt{n}} \tag{9}$$

Design effect: by how much do we have to adjust n to correct for the lack of independence among observations?

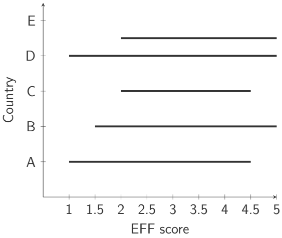
$$DE = 1 + (n - 1)\rho \tag{10}$$

Formula above is valid for equal group sizes. It shows, though, that the more homogeneous the groups (ρ is higher), the design effect is larger $\Rightarrow$ the effective sample size is smaller!

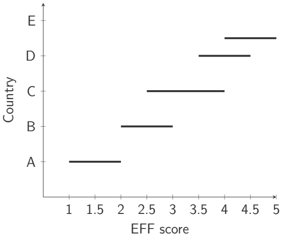
Clustering

The ICC (intraclass correlation coefficient): $\rho = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_y^2}$. In this formula, $\sigma_{total}^2 = \sigma_\alpha^2 + \sigma_y^2$ (between-group variance and within group variance).

Denotes the share of total variance that is between groups (can also be expressed as a correlation coefficient of individuals *within* the same group).



(a) Variation mainly within



(b) Variation mainly between

Calculating ICC

Derived based on a *null* model (with no predictors):

$$\begin{cases} EFF_{ij} = \beta_{0j} + e_{ij} \\ \beta_{0j} = \gamma_{00} + \upsilon_{0j} \end{cases} \tag{11}$$

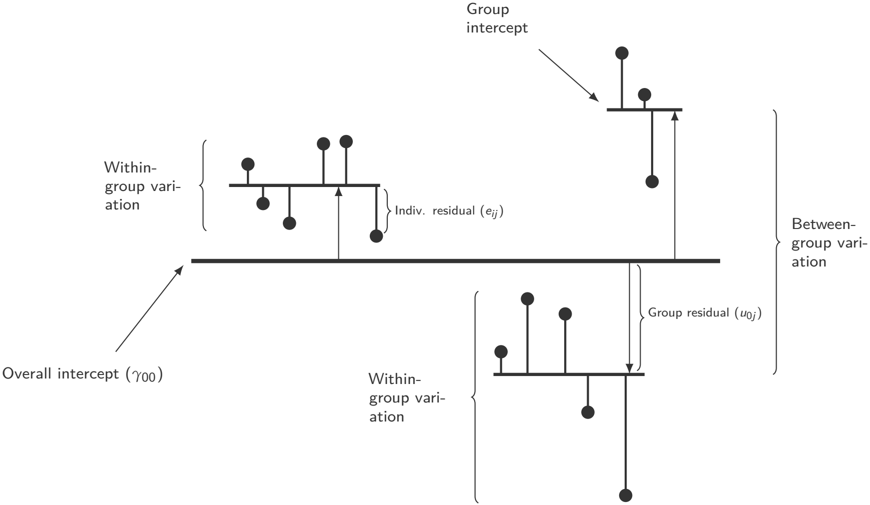
$$EFF_{ij} = \gamma_{00} + \upsilon_{0j} + e_{ij} \tag{12}$$

Directly from the model output in R you can compute:

$$ICC = \frac{\sigma^2_{\upsilon_{0j}}}{\sigma^2_{\upsilon_{0j}} + \sigma\epsilon_{ij}{}^2} \tag{13}$$

This is the way Luke (2004) introduces the ICC.

Visualizing the 3 elements



Adapted from Merlo et al. (2005).

Rules of thumb?

Not clear whether there are any, apart from those dictated by common sense.

With a model with very low ICC (e.g., below 0.05), it will be hard to find any group-level predictor that is statistically significant.

ICC for our example

	DV: Political efficacy
(Intercept) (γ_{00})	2.94374^{***}
	(0.04775)
Var: country (Intercept) ($\sigma^2_{\upsilon_{0j}}$)	0.08167
Var: Residual ($\sigma^2_{\epsilon_{ij}}$)	0.53158
Num. obs.	50954
Num. groups: countries	36

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

ICC for our example: $\frac{0.08167}{0.08167+0.53158} \approx 13.3\%$ of variance in outcome is at level-2 (between countries).

Thank you for the kind attention!

References

Gelman, A., & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.

Luke, D. A. (2004). *Multilevel Modeling*. Thousand Oaks, CA: Sage Publications.

Merlo, J., Chaix, B., Yang, M., Lynch, J., & Råstam, L. (2005). A Brief Conceptual Tutorial of Multilevel Analysis in Social Epidemiology: Linking the Statistical Concept of Clustering to the Idea of Contextual Phenomenon. *Journal of Epidemiology and Community Health*, *59*(6), 443-449.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage.