

Working with Categorical Predictors in Multilevel Models: A Brief Tutorial

Haley E. Yaremych, MS,¹ Kristopher J. Preacher, PhD,¹ & Donald Hedeker, PhD²

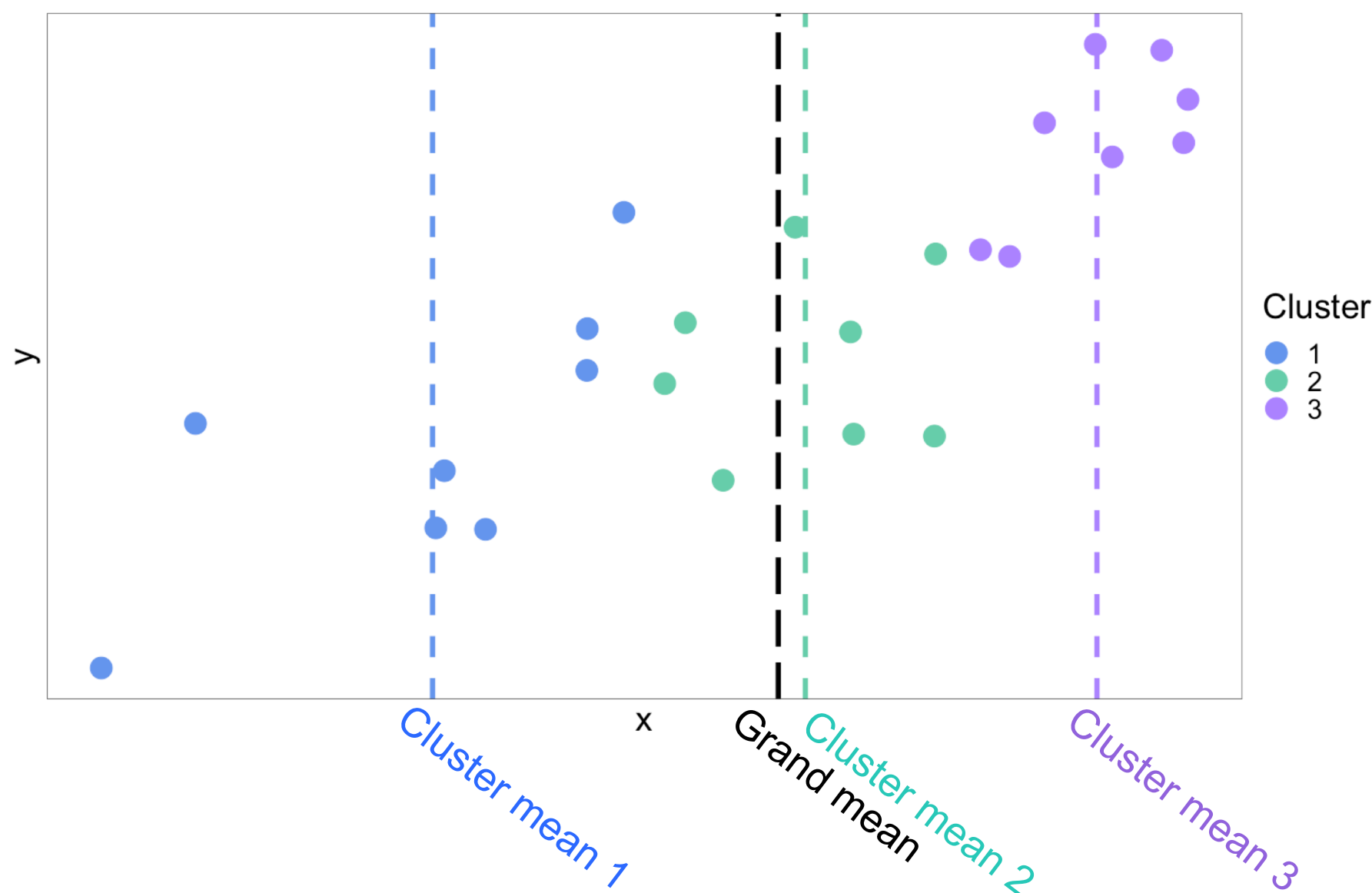
¹Department of Psychology & Human Development, Vanderbilt University ²Department of Public Health Sciences, University of Chicago



Background

- Within- and between-cluster effects of a predictor can be studied through use of multilevel models. Such effects can be uncovered by appropriately centering the level-1 predictors.¹
- Centering can be conducted in multiple ways when working with multilevel data.
- The mean around which predictors are centered influences the estimates given by the multilevel model and how those estimates must be interpreted.²

Fig 1: Example of different means for a level-1 continuous predictor from multilevel data.



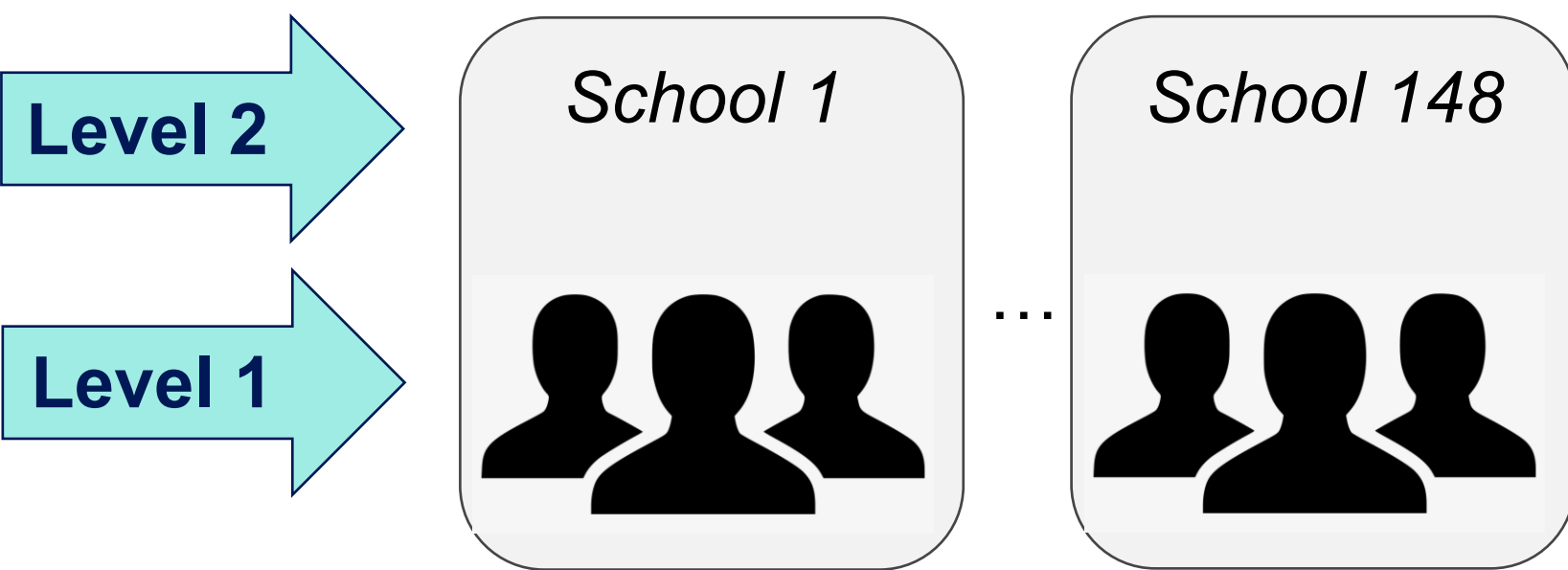
Current Aims

Demonstrate how coding variables representing categorical predictors should be centered in multilevel models.

Explain how the effects of categorical predictors should be interpreted.

Empirical Example

- $N = 3,435$ children nested within 148 schools³
- Level-1 outcome: academic achievement at age 16
- Level-1 predictor: four-group indicator of parental education
 - 1) neither parent highly educated
 - 2) mother educated, father not
 - 3) father educated, mother not
 - 4) both parents highly educated



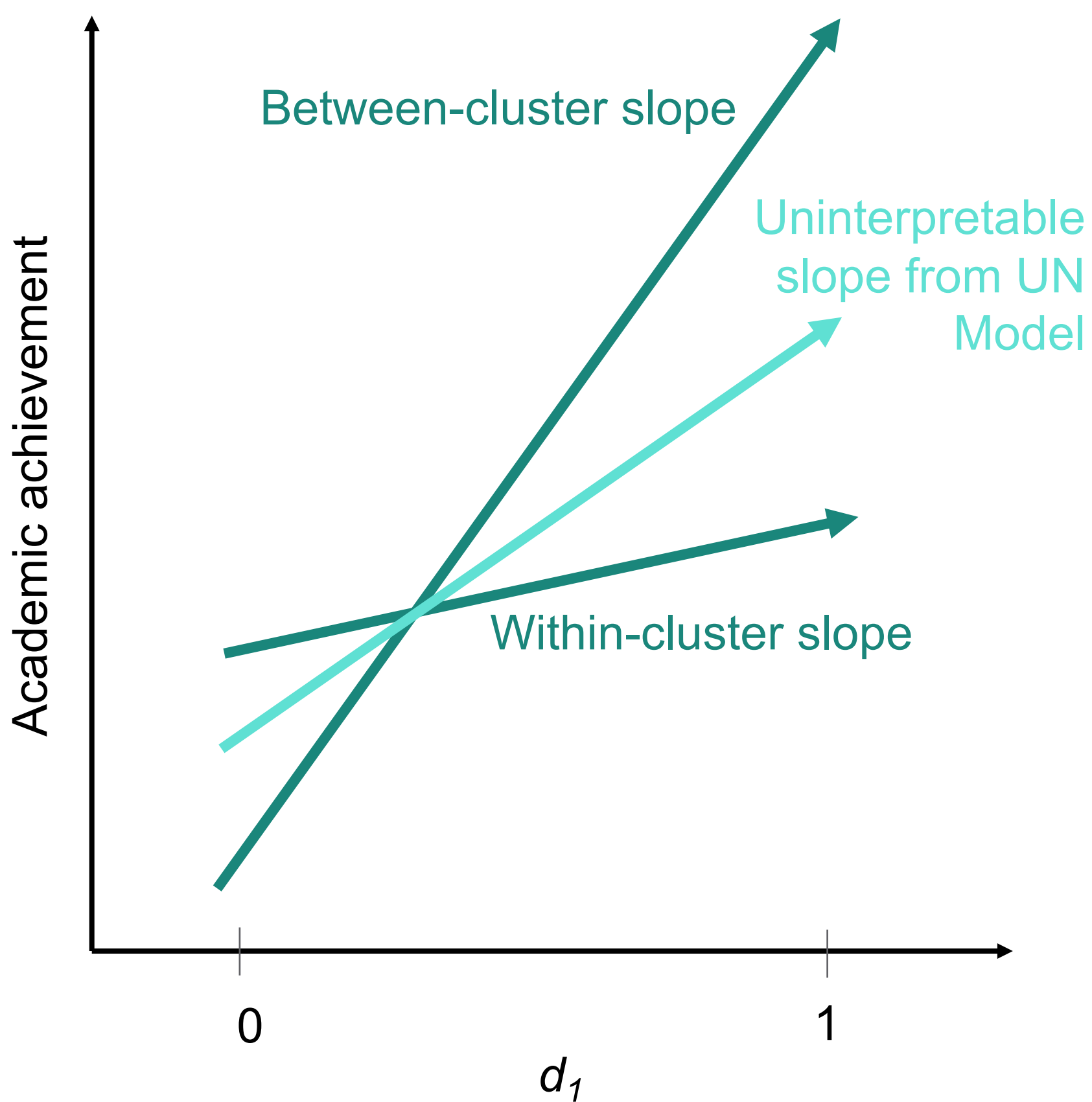
Choosing a Coding Scheme

Dummy codes				Unweighted effect codes				Contrast codes			
Interpret group mean differences with respect to the reference group				Interpret group mean differences with respect to the unweighted mean of all groups				Interpret group mean differences depending on how each code is uniquely constructed			
Group	Codes created in empirical example			Group	Codes created in empirical example			Group	Codes created in empirical example		
	d_1	d_2	d_3		e_1	e_2	e_3		c_1	c_2	c_3
1 (reference)	0	0	0	1	-1	-1	-1	1	-3/4	0	0
2	1	0	0	2	1	0	0	2	1/4	-1/3	-1/2
3	0	1	0	3	0	1	0	3	1/4	-1/3	1/2
4	0	0	1	4	0	0	1	4	1/4	2/3	0

Centering & Model Specification Options

The UN Model	The CWC(M) Model	The UN(M) Model
Include... • Original level 1 coding variables in their raw form, uncentered	Include... • Level 1 coding variables that have been centered around their corresponding cluster mean • Cluster means of coding variables as level-2 predictors themselves	Include... • Level 1 coding variables in their raw form, uncentered • Cluster means of coding variables as level-2 predictors themselves

Fig 2: Demonstration of within- and between-cluster slopes from the CWC(M) Model compared with the uninterpretable slope obtained from the UN Model fit to empirical example data.



Interpretation

The UN Model			
Slope estimates are uninterpretable! In the vast majority of situations the UN Model is not recommended.			
The CWC(M) Model		The UN(M) Model	
Level 1 slopes → group mean differences within clusters, on average.		Level 1 slopes → group mean differences within clusters, on average.	
Level 2 slopes → mean differences upon moving from a cluster composed entirely of one group to another.		Level 2 slopes → mean differences upon moving from a cluster composed entirely of one group to another, <i>holding the predictor constant at level 1.</i>	
Level 1 slope of d_1: <i>within schools, on average</i>		Level 2 slope of d_1: <i>between schools</i>	

References

1. Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, 12(2), 121–138. <https://doi.org/10.1037/1082-989X.12.2.121>

2. Yaremych, H.E., Preacher K.J., & Hedeker, D. (2021). Centering categorical predictors in multilevel models: Best practices and interpretation. *Psychological Methods*. <https://doi.org/10.1037/met0000434>

3. Paterson, L. (1991). Socio-economic status and educational attainment: A multi-dimensional and multi-level study. *Evaluation & Research in Education*, 5(3), 97–121.