

## Stat 414 - Day 19

### Adding Level 2 Variables (4.5)

---

#### Last Time: Adding Level 1 variables

- Adding Level 1 variables should explain Level 1 variability (reduce  $\hat{\sigma}$ )
    - Can calculate the percentage of Level 1 variability explained
    - Can calculate  $R^2$  values as the amount of total variation explained by the fixed effects
  - Can also change Level 2 variability (and ICC)
    - Explains Level 2 variability if the association at the group level is ‘consistent’
    - Can increase the Level 2 variability if the association is not consistent
    - Can use `performance::r2(model, by_group = TRUE)`
  - May refer to  $\sigma^2$  as the “conditional” variance and  $\tau^2 + \sigma^2$  as the “marginal” variance (across the groups).
- 

#### Example 1: Radon cont. Recall the radon dataset.

- (a) Now we want to add the soil uranium level to the model. Is this a level 1 or level 2 variable? How can you tell?
- (b) Show how to add this variable to the model equations. How many terms will it add to the model?
- Level 1:
- Level 2:
- (c) What do we have to do differently in the `lmer` command to tell R about this variable?
- (d) Interpret the coefficient of uranium in this model. Do you consider it statistically significant?
- (e) How did the variance components change? Do these changes make sense? Explain.

**Example 2:** Recall the Netherlands study on language test performance for Grade 8 students nested within schools. Another important type of Level 2 variable is aggregating a Level 1 variable, e.g., average school IQ. These are sometimes referred to as “contextual effects.”

- (a) Is the school mean verbal IQ related to (average) performance score?

(b) Adding `sch_iqv`, what are the Level 1 and Level 2 model equations?

Level 1:

Level 2:

(c) Interpret the slope of the `sch_iqv` variable in model 3 in context.

(d) Based on the output you have, is this new variable a significant addition to the model? How are you deciding?

(e) How much Level 1 variability did we explain? How much Level 2?

(f) Which do we understand more, why some schools have higher language scores or why some students have higher language scores? Explain your reasoning.

(g) What if we change to the “deviation” variable, `verbal_IQ - sch_iqv`? (aka Group Mean Centering)

```
devIQ = neth$IQ_verb - neth$sch_iqv
```

Interpret the slope coefficients for this model.

(h) Does the deviation variable explain variation at Level 1 and/or Level 2?

*Notes:*

- “Group mean centering” (as opposed to grand mean centering) creates a “within” group variable.
- Some recommend calculating group means before observations with missing values are deleted (or better yet, use imputation)
- When using  $x$  and  $\bar{x}$  (rather than deviation and  $\bar{x}$ ), the coefficient of  $\bar{x}$  is the difference between the within and between group effect. The significance of the group mean variable is akin to the Hausman specification test in econometrics: Is the difference between the “within group” and “between group” effects statistically significant?