

Stat 414 - Day 20

Random Slopes (5.1)

Last Week: Adding Level 1 and Level 2 variables to random intercept models

- Level 1 variables may explain association at Level 1 and/or Level 2 (if the distribution of the variable differs across the Level 2 units, i.e., variation at both levels)
 - Can also increase unexplained variation at Level 2
 - Some can only explain Level 1 variation (e.g., percentile, z-score in country)
- Level 2 variables only explain association at Level 2 (e.g., country uranium)
- Aggregating a Level 1 variable to Level 2 gives a nice “contextual” variable
 - Does coefficient of group mean variable represent the additional contribution or the group level effect?
 - $\hat{\beta}_1 x_{ij} + \hat{\beta}_2 \bar{x}_j$ ($\hat{\beta}_2$ is the additional contribution at Level 2)
 - If not significant, the Level 1 and Level 2 associations are “the same”
 - $\hat{\beta}_1 (x_{ij} - \bar{x}_j) + \hat{\beta}_2 \bar{x}_j$ ($\hat{\beta}_2$ is the effect at Level 2)
 - Nicely separates the Level 1 and Level 2 associations

Example 1: Data were collected from nine beaches along the Dutch coast. Five readings were taken for each beach. The response variable was species richness (number of different species), and available variables are NAP (the height of the sampling station relative to the mean tidal level), and Exposure (a composite measure of wave action, length of the surf zone, slope, grain size, and the depth of the anaerobic layer).

- (a) What are the Level 1 units in this study? How many are there? Any Level 1 variables?
- (b) What are the Level 2 units in this study? How many are there? Any Level 2 variables?
- (c) Does Richness vary with NAP? In the expected direction?
- (d) Suppose we treated the beaches as a “fixed effect,” write out the model equation:

Let’s consider instead the the “random intercepts” or “variance components” model.

$Y_{ij} = \beta_0 + u_j + \epsilon_{ij}$ for the i^{th} site on the j^{th} beach

- (e) How will \hat{u}_1 compare to $\hat{\beta}_1$?

Fit the random intercepts (null) model:

```
model0 = lmer(Richness ~ 1 + (1 | rikzdata$Beach), data = rikzdata)
```

- (f) According to this model, how much of the variation in Richness is due to the different beaches?

Fit the multilevel model that also allows Richness to vary with NAP, after adjusting for beach:
`model1 = lmer(Richness ~ NAP + (1 | Beach), data = rikzdata)`

(g) Does this model appear to be valid? Do any of the patterns in the residuals vs. fitted plots appear related to any particular beach? What does the pattern suggest doing with the model?

To allow the slopes to vary across the Level 2 units in the model equation, we add a j index to the slope too. Level 1 equation:

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + \epsilon_{ij}$$

where $\beta_{0j} = \beta_{00} + u_{0j}$ and $u_{0j} \sim N(0, \tau_0^2)$ and ...

(h) Write out the second level 2 equation

(i) Now create the composite equation.

(j) Give the expression for beach j 's intercept. Give the expression for beach j 's slope.

(k) What assumptions are we making about the distribution of the random slopes?

Fit the random slopes (or "random coefficients") model, allowing the slopes to vary across the beaches:

```
model2 = lmer(Richness ~ NAP + (1 + NAP | Beach), data = rikzdata, REML = FALSE)
# you get a warning (not an error) and can ignore it
```

(l) How many parameters does this add to the model? (What if beach was a fixed effect?)
 What do these new parameter estimate(s) tell you?

(m) Based on this model, what is the equation for Beach 1? What is the equation for Beach 9?

(n) Are the differences in the slopes statistically significant? (State hypotheses, df, test statistic, p-value, conclusion in context.)

Notes:

- In a random slopes model, be careful with the interpretation of the intercept variance and the intercept-by-slope covariance, they assume $x = 0$. Another reason why its always good practice to center your explanatory variables so the intercept is meaningful.
- You can use Likelihood-ratio tests to assess whether the random slopes model is “worth pursuing statistically.” As you can see, randomly slopes can improve the complexity of the model pretty quickly, so you should have empirical, or better yet theoretical, reasons for doing so.