Supplemental Material

CBE—Life Sciences Education

Theobald *et al.*

APPENDICES

Appendix 1: The mathematical explanation of random effects

There are two common kinds of random effects that result in three common types of multilevel models:

1) Random intercepts
2) Random slopes
3) Random slopes and intercepts

The mechanics of how these random effects influence the estimate are apparent from the mathematical equations that include these effects. Let's take a study design where student outcomes are measured but students (i) are clustered in sections (j). The research question asks if some predictor (x) has an impact on student performance (y)

A standard linear regression model that does not account for the sections is written as:

$$y_i = \alpha + \beta x_i + \epsilon_i$$

Where y_i is the outcome, α is the intercept, β is the effect of the predictor x, or the slope on x, and ϵ_i is the error.

A random intercepts model allows each section (j) to have a unique intercept:

$$y_{ij} = \alpha_j + \beta x_{ij} + \epsilon_{ij}$$

This unique intercept is apparent in the α_j term in that each of the j sections has a unique value. The non-independence of the errors within section is apparent in the error term ϵ_{ij} in that the j term indicates accounting for the correlation between the measurement errors in each section.

A random slopes model allows each section (j) to have a unique slope:

$$y_{ij} = \alpha + \beta_j x_{ij} + \epsilon_{ij}$$

This unique slope is apparent in the β_j term in that each of the j sections has a unique value.

A random intercepts and random slopes model allows each section (j) to have a unique intercept and a unique slope:

$$y_{ij} = \alpha_j + \beta_j x_{ij} + \epsilon_{ij}$$

The unique intercept and slope are apparent in the α_j and β_j terms respectively in that each of the j sections has a unique value.

Appendix 2: Using the t-statistic for hypothesis testing

Calculating the t-statistic relies on an accurate measure of the degrees of freedom in a model, and this is not straightforward in multilevel models (D. M. Bates, 2010). Thus, the t-statistic only follows a t-distribution in certain multilevel modeling cases (e.g., linear multilevel model, with nested random effects and balanced data).

A more appropriate method for hypothesis testing (instead of relying on t-values or p-values) is with model selection – specifically using model selection to determine what fixed effects are retained in the best fitting model (Burnham & Anderson, 2002). When model selection is employed to select the fixed effect structure (not just the random effect structure as is demonstrated here), it is considered best practice not to interpret significance based on p-values but rather to interpret significance based on what parameters have been retained in the final model (Burnham & Anderson, 2002); each model that is fit is considered a separate hypothesis with the null hypothesis being that additional parameters do not help explain variance/improve model fit.

Appendix 3: Annotated R Code

Appendix 4: Data