Additional Levels of Nesting

- Thus far we have encountered only two-level models
    - Repeated measures were nested in subjects
        - **Level 1 model**: Within-subjects (repeated measures)
        - **Level 2 model**: Between-subjects
        - Key assumption is that repeated measures are correlated, but the subjects are not
- Some research designs include additional levels of nesting
    - Repeated measures are nested in subjects which are nested in classrooms (or families)
    - In these situations subjects' score are expected to be correlated because of shared experiences of teacher/living experiences, etc.
    - This correlation should be modeled

- **Key assumption:** Correlation among the lower levels, but not for the highest level
  - For 3-level model:
    - Correlation among repeated measures
    - Correlation among student scores within classrooms
    - No correlation between scores across classrooms
- Random sampling assumption at highest level easiest way to guarantee the independence between classrooms

Three-Level Model

- Response variable now has three subscripts Y_{hij}
  - Order of subscripts signifies nesting
    - Outer subscript depicts nesting within the inner subscripts
    - Time points (*j*) nested in students (*i*) nested in classrooms (*h*)
- Suppose students are nested in classrooms A, B, ..., G
  - Information about classrooms is contained in the text file `classroom.txt`

```
## Read in the classroom data
> classroom <- read.table(file = "Classroom.txt", header = TRUE)
> head(classroom)

  subid classr
1     1      A
2     2      A
3     3      A
4     4      B
5     5      B
6     6      B

## Merge the Minneapolis and classroom data
> MPLS.3 <- merge(mpls.l, classroom, by = c("subid"))
> head(MPLS.3)

  subid X risk gen eth ell sped  att grade read grade5 dadv ethW classr
1     1 1  HHM   F Afr   0    N 0.94     5  172      0 DADV   NW      A
2     1 2  HHM   F Afr   0    N 0.94     6  185      1 DADV   NW      A
3     1 3  HHM   F Afr   0    N 0.94     7  179      2 DADV   NW      A
4     1 4  HHM   F Afr   0    N 0.94     8  194      3 DADV   NW      A
5     2 5  HHM   F Afr   0    N 0.91     5  200      0 DADV   NW      A
6     2 6  HHM   F Afr   0    N 0.91     6  210      1 DADV   NW      A
```

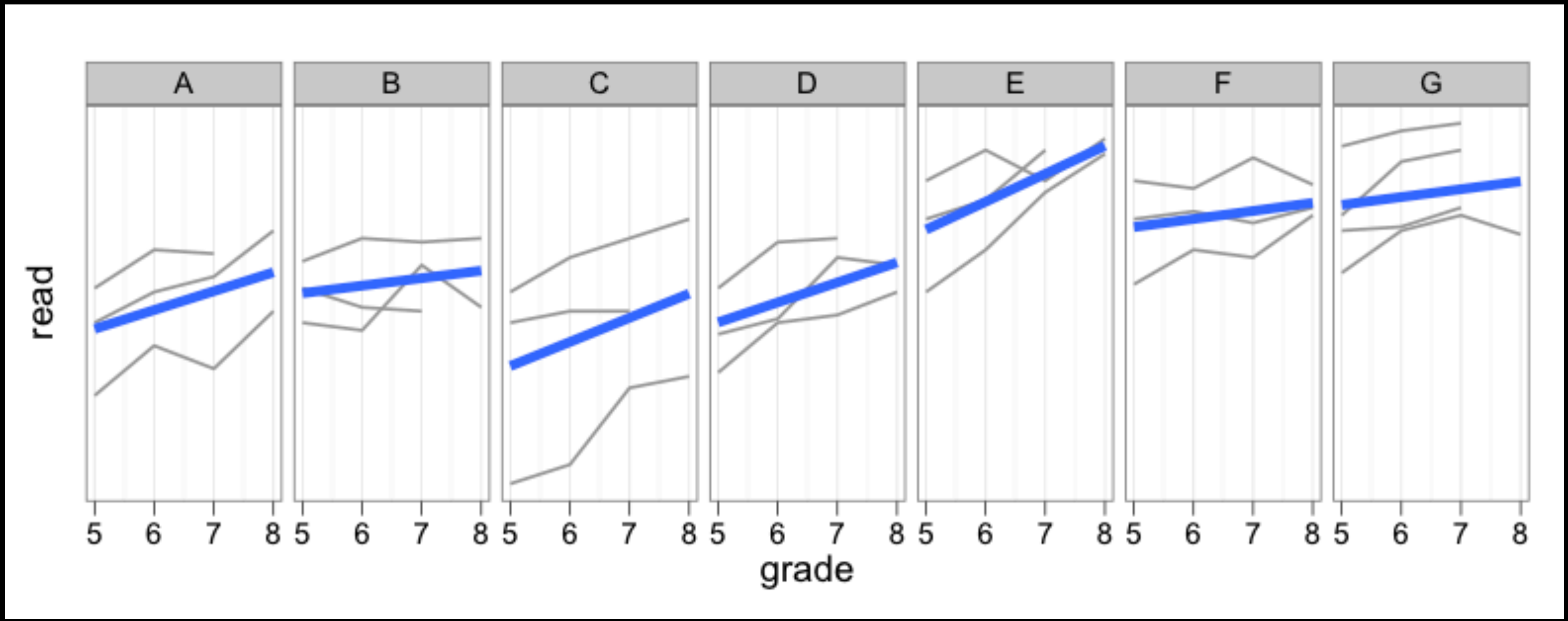
	classr	subid	grade
1	A	1	5
2	A	1	6
3	A	1	7
4	A	1	8
5	A	2	5
6	A	2	6
7	A	2	7
8	A	3	5
9	A	3	6
10	A	3	7
11	A	3	8
12	B	4	5
13	B	4	6
14	B	4	7
15	B	5	5
16	B	5	6
17	B	5	7
18	B	5	8
19	B	6	5
20	B	6	6
21	B	6	7
22	B	6	8
23	C	7	5
24	C	7	6

In the data frame–if it is sorted on the highest level, then the next highest, etc.–the lower orders of nesting (e.g., grade) change faster than higher orders (e.g., student and classroom)

- Plots used to get a sense of the
  - Variability of reading scores among classrooms, and
  - Variability of subjects reading scores within classrooms
  - Individual observed reading curves are graphed over grade
    - Faceted by classroom
    - Linear regression lines plotted in each panel using `lm()`

```
>  ggplot(data = MPLS.3, aes(x = grade, y = read, group = subid)) +
     geom_line(colour = "grey60") +
     facet_grid(. ~ classr) +
     opts(aspect.ratio = 2) +
     stat_smooth(aes(group = 1), method = "lm", se = F, lwd = 1.5) +
     scale_x_continuous(breaks = 5:8) +
     scale_y_continuous(breaks = 0:1) +
     theme_bw()
```

Individual curves (thin) and regression curves (thick) for reading by classroom



- The vertical scatter within each panel illustrates the variability of subjects nested within classrooms
- The horizontal scatter represents variability among classrooms
- Appears to be non-trivial variability among classrooms, at least with respect to the intercepts of the regression curves

- Similar to the two-level case, random effects are used to account for the dependency due to the nested structure of the data
  - Random effects at multiple levels must be considered
  - In the three-level model, random effects are included at the second and third levels
    - Level 1 is for the longitudinal data
    - Level 2 is for subjects
    - Level 3 is for classrooms

Level 1

$$y_{hij} = \beta^*_{0hi} + \beta^*_{1hi}(\texttt{grade}_{hij}) + \epsilon_{hij}$$

Level 2

$$\beta^*_{0hi} = \beta^\dagger_{0h} + b_{0hi}$$
$$\beta^*_{1hi} = \beta^\dagger_{1h} + b_{1hi}$$

Level 3

$$\beta^\dagger_{0h} = \beta_0 + b_{0h}$$
$$\beta^\dagger_{1h} = \beta_1 + b_{1h}$$

- Level 1 is a longitudinal model for the repeated measures nested within subjects, nested within classrooms
  - β_{0hi}^* is the random intercept of the *i*th subject, nested in the *h*th classroom
  - β_{1hi}^* is the random slope of the *i*th subject, nested in the *h*th classroom
  - The random error term, ε_{hij}, is assumed to be independent, normally distributed, with mean of 0 and constant variance

$$y_{hij} = \beta_{0hi}^* + \beta_{1hi}^*(\texttt{grade}_{hij}) + \epsilon_{hij}$$

- Level 2 is a model for subjects nested within classrooms
  - The time subscript (*j*) does not appear at Level 2
  - $\beta_{0h}^{\dagger}$ is the average intercept of the subjects
  - $\beta_{1h}^{\dagger}$ is the average slope of the subjects
  - The Level 2 random effects, b_{0hi} and b_{1hi}, represent the deviation of each subject from the subject average intercept and slope, respectively
  - The Level 2 random effects are assumed to be normally distributed with mean of 0, and variance-covariance matrix, **G**

$$\beta^*_{0hi} = \beta^{\dagger}_{0h} + b_{0hi}$$
$$\beta^*_{1hi} = \beta^{\dagger}_{1h} + b_{1hi}$$

- Level 3 is the classroom model
  - Neither the time subscript (*j*) nor the student subscript (*i*) appear at Level 2
  - β_0 being the average intercept of classrooms
  - β_1 being the average slope of classrooms
  - Level 3 random effects, b_{0h} and b_{1h}, represent the deviation of each classroom from the average classroom intercept and slope, respectively
  - Level 3 random effects are assumed to be normally distributed, with mean of 0, and variance-covariance matrix, **H**
  - The inter-level random effects are assumed to be completely uncorrelated, and uncorrelated with the Level 1 random error

$$\beta^{\dagger}_{0h} = \beta_0 + b_{0h}$$
$$\beta^{\dagger}_{1h} = \beta_1 + b_{1h}$$

Substituting higher level terms into the lower level equations yields the LMER model

$$y_{hij} = (\beta_0 + b_{0h} + b_{0hi}) + (\beta_1 + b_{1h} + b_{1hi})(\texttt{grade}_{ij}) + \epsilon_{hij}$$

- LMER model is perhaps easier to deal with than the multilevel model because there are fewer terms
  - Easier to see that β_0 is the mean intercept among subjects and classrooms, and there are two deviation terms, one for subjects (b_{0hi}), and one for classrooms (b_{0h})
  - Likewise, β_1 is the mean slope among subjects and classrooms, with individual deviations for subjects (b_{1hi}) and for classrooms (b_{1h})

Assumptions of the model can be written in the following manner,

$$b_{khi} \sim \mathcal{N}(\mathbf{0}, \mathbf{G}) \perp b_{kh}) \sim \mathcal{N}(\mathbf{0}, \mathbf{H}) \perp \epsilon_{hij} \sim \mathcal{N}(\mathbf{0}, \sigma_e^2 \mathbf{H}_{ji})$$

- Goals of the analysis are usually the same as with a two-level model
  - Draw inferences about β_1, and possibly β_0
  - The random effects are means to an ends to these goals, being included only to account for dependency due to nesting

- There is a relatively large number of variance components in the three-level model
  - There are seven in this model
    - Three unique elements in **G**
    - Three in **H**
    - The random error variance, σ_ε^2
- It may not always be the case that all the random effects are required

Because of the number of parameters, serious inferences cannot be made based on the example data. Nevertheless, the model is estimated to illustrate how a serious analysis might be performed

- The `lmer()` syntax for a three-level model requires a term for the additional level of random effects

- To appreciate the syntax, note that the random effects portion of the model can be written as

$$\underbrace{\left[b_{0hi}(1) + b_{1hi}(\mathtt{grade}_{hij})\right]}_{\text{Subjects}} + \underbrace{\left[b_{0h}(1) + b_{1h}(\mathtt{grade}_{hij})\right]}_{\text{Classrooms}}$$

- Each random intercept term appears as a single effect

- Each random slope term is multiplied by the time predictor (`grade` is used here, but `grade5` can also be used)

- The subject random effects must be pinned to the subject identification number
- The classroom random effects must be pinned to the classroom identification letter
- The random effects syntax is

```
(1 + grade5 | subid) + (1 + grade5 | classr)
```

- The 1 is optional, but included here for clarity

In the syntax below, the model is estimated, and the results are printed. It is again cautioned that **the data set is not large enough to make serious inferences** for this model.

```
## Create grade5 predictor
> MPLS.3$grade5 <- MPLS.3$grade - 5

## Fit three-level model
> thrlvl.1 <- lmer(read ~ 1 + grade5 + (1 + grade5 | subid) +
    (1 + grade5 | classr), data = MPLS.3, REML = FALSE)
```

```
## Print results
> print(thrlvl.1, cor = FALSE)

Linear mixed model fit by maximum likelihood
Formula: read ~ grade5 + (1 + grade5 | subid) + (1 + grade5 | classr)
   Data: MPLS.3
   AIC   BIC logLik deviance REMLdev
 580.4 601.8 -281.2    562.4   555.7
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 subid    (Intercept) 232.1735 15.2372
          grade5        4.2698  2.0663  -0.748
 classr   (Intercept) 145.2937 12.0538
          grade5        2.2691  1.5063  -0.178
 Residual              18.5997  4.3127
Number of obs: 80, groups: subid, 22; classr, 7

Fixed effects:
            Estimate Std. Error t value
(Intercept) 201.5319     5.6570   35.63
grade5        4.8454     0.8627    5.62
```

- The fixed effects table is identical in form to that of the two-level model
  - The estimated mean reading score at grade 5 is 201.53
  - The estimated mean increase per grade is 4.85
- The variance components table shows the variances for the subject intercepts and slopes, *and* for the classroom intercepts and slopes
  - The variability due to subjects is greater than that due to classrooms
  - The correlation between the subject random effects is also greater in absolute value
- The output indicates there are 80 total rows of the data frame used in the analysis, 22 subjects, and 7 classrooms

- The objects of inference in this case are the subjects (subjects are measured repeatedly, not classrooms)
- Inferences regarding subjects should take into account the nesting within classrooms, as we expect the intra-classroom correlation to be higher than the inter-classroom correlation
  - Influence of classrooms is reflected in the estimated $\mathbf{H}$ matrix
    - The estimated variance components are considered when computing the deviance and when computing the SEs of the fixed effects
    - When a two-level model is fit ignoring classrooms, the SEs are smaller than in this model
    - The adjustment for the nesting within classrooms results in wider CIs and larger *p*-values for the fixed effects

Static Predictors in Three-Level Models

- Static predictors may be used in three-level models
  - A complication is that static predictors can be included at different levels depending on the type of between-unit contrast they represent
  - For the example, there might be static predictors that represent *between-subject* differences or *between-classroom* differences
    - The former vary among subjects with no regard for classrooms, or at least not explicitly so.
    - Classroom- level predictors vary among classrooms, and they only vary among subjects because of the nesting within classrooms

Classroom-Level Static Predictors

- An example of a classroom-level static predictor might be the presence or absence of a student teacher
  - This would be a dichotomous classroom-level predictor
  - In this scenario, the student teachers are assigned to classrooms, and not subjects
    - The variability in subjects is due entirely to the variability in classrooms
- Such static predictors are included in the Level 3 model

Subject-Level Static Predictors

- An example of a subject-level static predictor is dadv
  - Values of this categorical predictor are assigned to subjects, not classrooms
  - For this reason, dadv is included in the Level 2 model
  - The level at which the static predictor is included is denoted by its subscript
    - The model with dadv uses the Level 2 subscript, *hi*

$$
\begin{aligned}
y_{hij} = {} & (\beta_0 + b_{0h} + b_{0hi}) + (\beta_1 + b_{1h} + b_{1hi})(\texttt{grade}_{hij}) \\
& + \beta_2(\texttt{dadv}_{hi}) + \beta_3(\texttt{grade}_{hij} \cdot \texttt{dadv}_{hi}) + \epsilon_{hij}
\end{aligned}
$$

- In `lmer()`, static predictors are specified as previously discussed, with nothing in the syntax to distinguish their level
- The level must be inferred from the nature of the static predictor, or by examination of the data frame

```
## Estimate model with subject-level static predictor (dadv)
> thrlvl.2 <- lmer(read ~ 1 + grade5 * dadv +
    (1 + grade5 | subid) + (1 + grade5 | classr),
    data = MPLS.3, REML = FALSE)

## Print model results
> print(thrlvl.2, cor = FALSE)
```

```
Linear mixed model fit by maximum likelihood
Formula: read ~ 1 + grade5 * dadv + (1 + grade5 | subid) + (1 +
grade5 |      classr)
   Data: MPLS.3
   AIC   BIC logLik deviance REMLdev
 570.2 596.4 -274.1    548.2   534.9
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 subid    (Intercept) 179.8901 13.4123
          grade5        3.8154  1.9533  -0.721
 classr   (Intercept)   4.1809  2.0447
          grade5        2.6355  1.6234  -1.000
 Residual              18.6114  4.3141
Number of obs: 80, groups: subid, 22; classr, 7

Fixed effects:
                Estimate Std. Error t value
(Intercept)     217.4972     4.5550   47.75
grade5            4.4400     1.3309    3.34
dadvDADV        -28.1467     6.1571   -4.57
grade5:dadvDADV   0.7009     1.7746    0.39
```

- The random effects table in the output suggests some problems
  - Perfect correlation among the Level 3 random effects for classroom indicates that not all elements of $\mathbf{H}$ need be estimated
- It is also odd that the variance of the classrooms random intercepts (4.18) has decreased so much as compared to the model without the static predictor (145.29)
  - This is due to the way the classroom variable was generated and it should not be taken as regular
- The fixed effects table suggests there might be a population intercept difference among the risk groups, but not a slope difference
  - The estimated fixed effect for `dadv` is −28.15, indicating the disadvantaged group curve has an intercept that is below that of the advantaged group

- The object of inference is again subjects, as a risk group status is assigned to a subject and not to a classroom
  - Though between-subject differences are the primary focus, the three-level model is important here because it accounts for the dependency associated with nesting within classrooms
  - As a final word, there are many varieties of nesting and crossing that can be handled by LMER. A number of these cases are discussed in Pinheiro and Bates (2000) and Bates (2011)

Bates, D. M. (2011). *lme4: Mixed-effects modeling with R.* New York: Springer.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer.