Introduction to Mixed-Effects Models

Andrew Zieffler

Educational Psychology

University of Minnesota

Driven to Discover℠

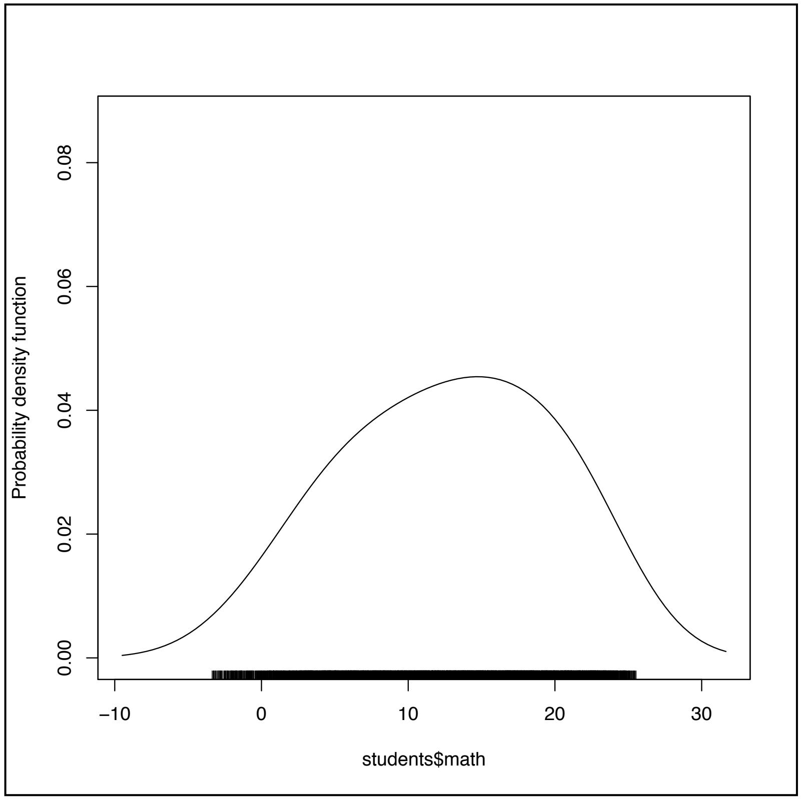
SES and Math Achievement

Research Questions

1. How much do U.S. high schools vary in their mean math achievement?
2. Does a high level of SES in a school predict high math achievement?
3. Is the connection between student SES and math achievement similar across schools? Or does the relationship show substantial variation?

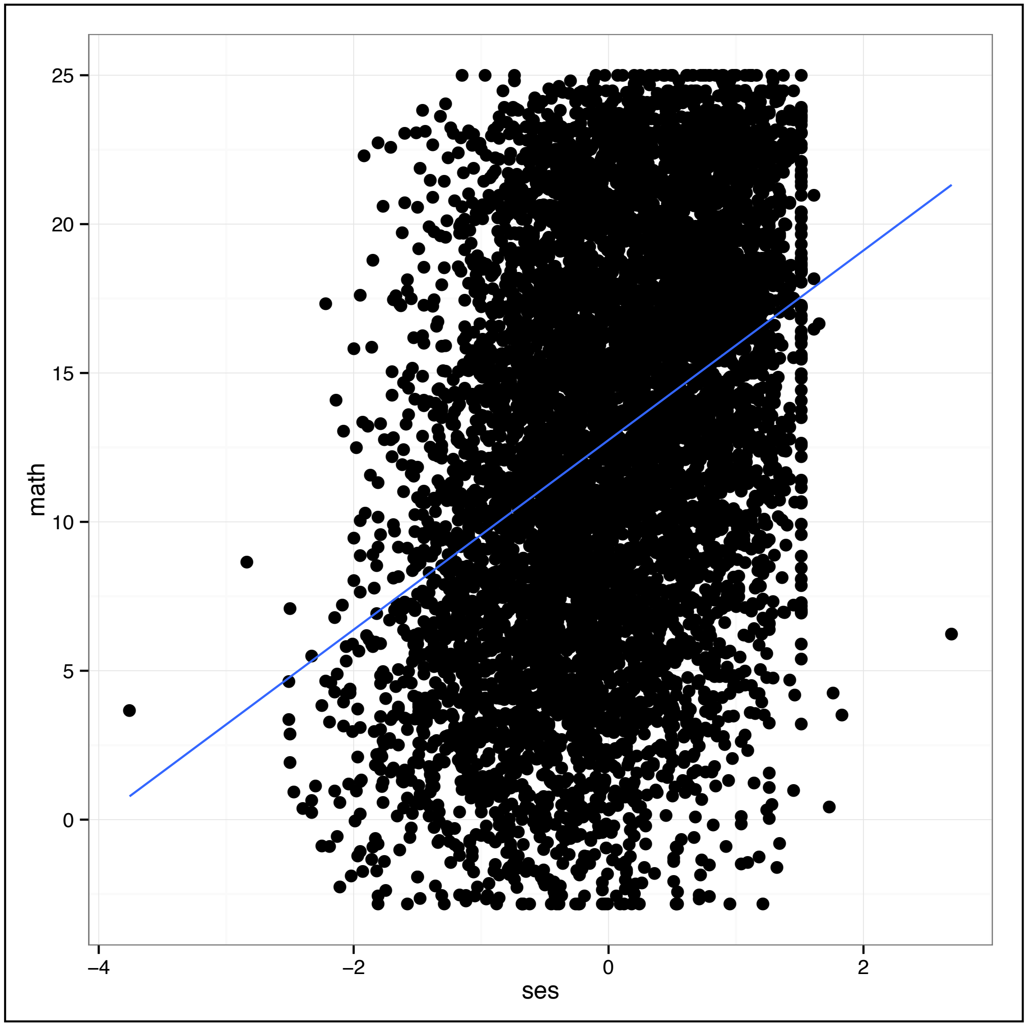
Raudenbush, S. W., & Bryk. A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, CA: Sage.

Examine Outcome



n	7185.00
mean	12.75
sd	6.88
median	13.13
trimmed	12.91
mad	8.12
min	-2.83
max	24.99
range	27.82
skew	-0.18
kurtosis	-0.92
se	0.08

Examine Relationship



	ses	math
ses	1.0000000	0.3607556
math	0.3607556	1.0000000

Positive relationship between SES andmath scores...students with higher SES tend to have higher math scores than students with lower SES

Fit LM at the Case-Level

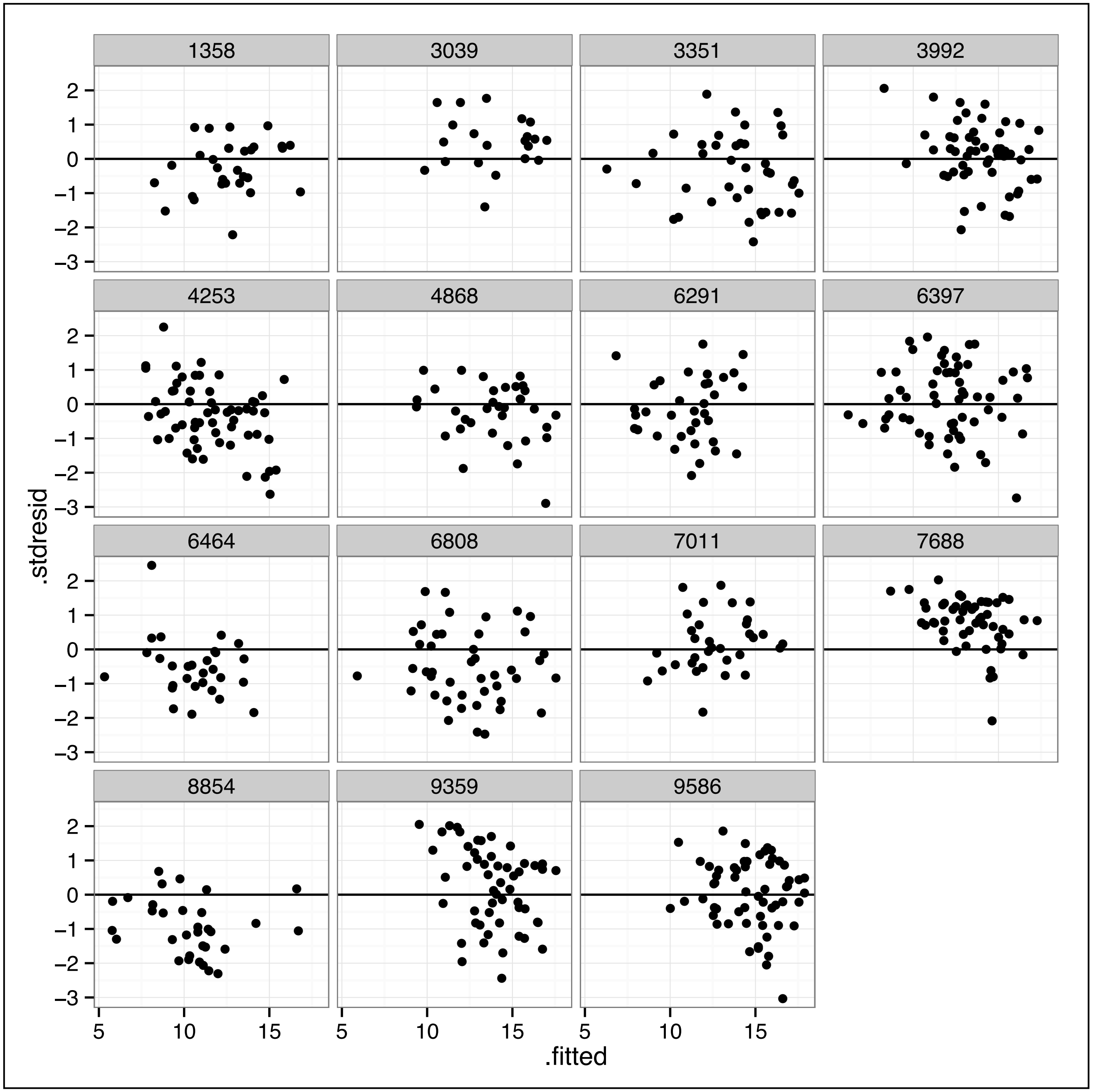
```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 12.74740    0.07569  168.42   <2e-16 ***
ses          3.18387    0.09712   32.78   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.416 on 7183 degrees of freedom
Multiple R-squared:  0.1301,   Adjusted R-squared:   0.13
F-statistic:  1075 on 1 and 7183 DF,  p-value: < 2.2e-16
```

Differences in SES seem to explain differences in mathematic scores, $F(1, 7183) = 1075$, $p < .001$. These differences explain about 13% of the variation in math scores.

$$\hat{\text{math}} = 12.74 + 3.19(\text{ses})$$

The average mathematics score for students with SES = 0 (average SES) is 12.74.

The average difference in mathematics scores between students who are one standard deviation different in their SES is 3.19.



Although the model seems to fit the overall data well, **within a school**, the residuals are either *mostly positive* or *mostly negative*

This is a violation of the independence assumption.

Why is this a Problem?

```
Analysis of Variance Table

Response: math
            Df Sum Sq Mean Sq F value    Pr(>F)    
ses          1  44233   44233  1074.7 < 2.2e-16 ***
Residuals 7183 295644      41                      
---
```

7185 independent observations = 7184 *df*

Model 1 *df* | **Residual** 7183 *df*

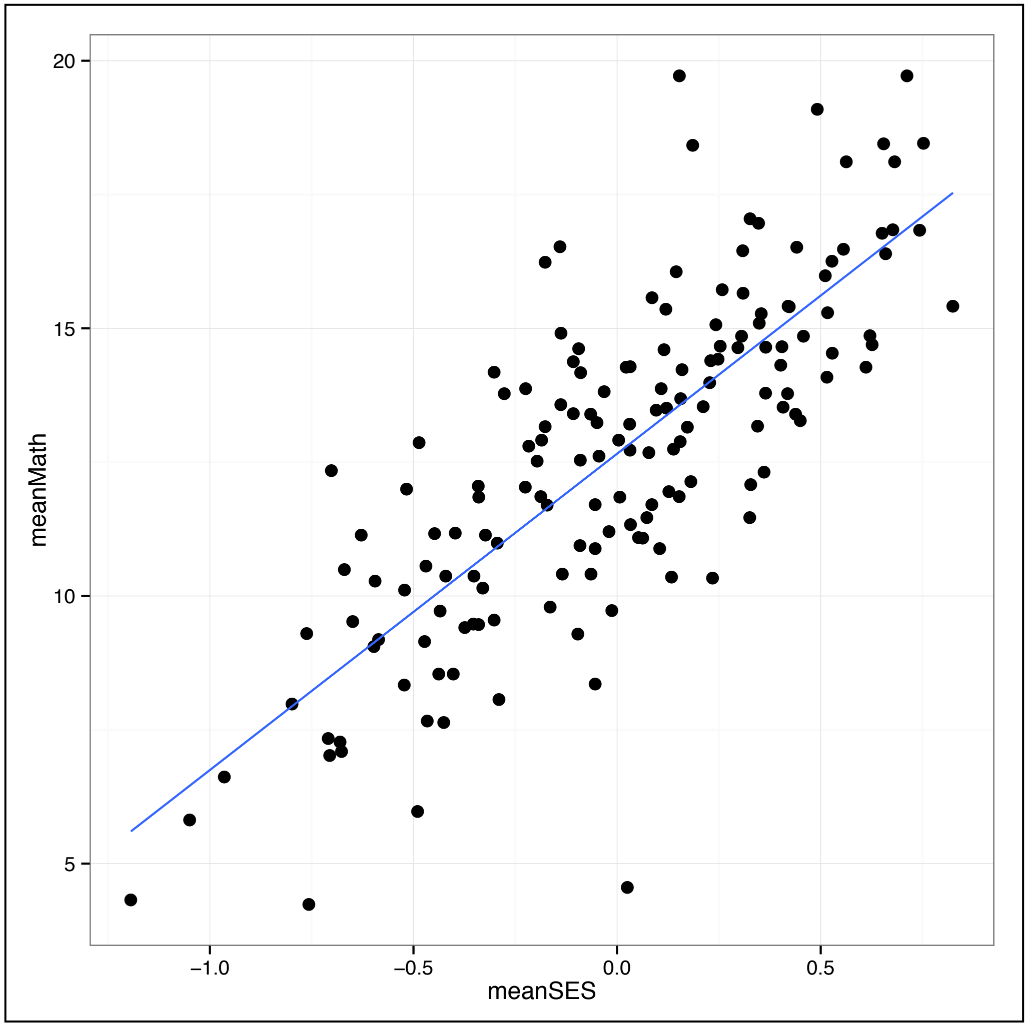
$$F = \frac{\mathrm{MS}_{\mathrm{Model}}}{\mathrm{MS}_{\mathrm{Residual}}}$$

$$F = \frac{\left(\frac{\mathrm{SS}_{\mathrm{Model}}}{df_{\mathrm{Model}}}\right)}{\left(\frac{\mathrm{SS}_{\mathrm{Residual}}}{\boxed{df_{\mathrm{Residual}}}}\right)}$$

Violation of the independence assumption indicates that it is wrong to think that the data constitutes 64 independent observations….which means that we have fewer *df* for the residual than we think...

Solution 1: Analysis on the Group Means

Rather than fitting the model to the 1785 cases, we can fit it to the 160 school-level observations.



Fit LM at the Group-Level

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.6573     0.1533   82.57   <2e-16 ***
meanSES       5.9093     0.3714   15.91   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.939 on 158 degrees of freedom
Multiple R-squared:  0.6157,   Adjusted R-squared:  0.6132
F-statistic: 253.1 on 1 and 158 DF,  p-value: < 2.2e-16
```

Differences in school-level SES seem to explain about 62% of the variation in differences in the school mean math scores, $F(1, 158) = 253.1$, $p < .001$.

Note that R^2 is higher in this analysis since we it only considers the *between-school variation* and **not** the *within-school variation*.

The *residual df*, which now reflects the more reasonable assumption that schools can be treated independently, is much lower than in the previous analysis (158 vs. 1783). Consequently, the *p*-value is higher, albeit still statistically reliable.

The estimated coefficients are not that different from the initial analysis.

Two Problems with the Group-Level Analysis

- We do not account for the variation in math scores within a particular school
- We do not account for variation in the number of students sampled within each school. In the group-level analysis, each school is weighted equally, even though the averages are based on more houses in some cities than others.

Solution: Fit a Multi-Level Model

Level-1: $$\text{math}_{ij} = \beta_0 + \beta_1(\text{ses}_{ij}) + \epsilon_{ij}$$

Level-2: $$\beta_0 = \mu_0 + \eta_{0j}$$
$$\beta_1 = \mu_1 + \eta_{1j}$$

The **level-1 model** specifies how the student-level predictors relate to the outcome (math scores)

The **level-2 model** specifies how each coefficient in the level-1 model is predicted by school-level predictors

This is called a multi-level model.

Level-1 Model

$$\text{math}_{ij} = \beta_0 + \beta_1(\text{ses}_{ij}) + \epsilon_{ij}$$

- Level-1 model specifies the **student-specific regression** lines
  - ✓ β_0 is the intercept for the *i*th student
  - ✓ β_1 is the slope for the *i*th student
  - ✓ ε_{ij} are the random errors around the *i*th student's regression line

- Only source of variation in Level-1 model is student-to-student variation
- Any unaccounted for student-to-student variation is absorbed into the residual, ε_{ij}

Level-2 Model

$$\beta_0 = \mu_0 + \eta_{0j}$$
$$\beta_1 = \mu_1 + \eta_{1j}$$

- Level-2 models are school-level regression models for the student-level regression coefficients
  - ✓ Outcome variables are the Level-1 regression parameters (intercepts and slopes)
  - ✓ Number of Level-2 equations determined by number of regression parameters in the Level-1 model
  - ✓ Covariates that account for school-to-school variation exclusively appear in Level-2 model

- Error terms in Level-2 model are called **random-effects**
  - ✓ Indifferent to the number of fixed effects at Level-2
  - ✓ Can only be as many random-effects as there are Level-2 equations
  - ✓ Not every Level-2 equation needs to have random-effect

Mixed-Effects Model

Level-1: $$\text{math}_{ij} = \beta_0 + \beta_1(\text{ses}_{ij}) + \epsilon_{ij}$$

Level-2: $$\beta_0 = \mu_0 + \eta_{0j}$$
$$\beta_1 = \mu_1 + \eta_{1j}$$

Substitute the Level-2 equations into the Level-1 equation to get the mixed-effects model

$$\text{math}_{ij} = (\mu_0 + \eta_{0j}) + (\mu_1 + \eta_{1j})\,(\text{ses}_{ij}) + \epsilon_{ij}$$

$$\text{math}_{ij} = \mu_0 + \eta_{0j} + (\mu_1)(\text{ses}_{ij}) + (\eta_{1j})(\text{ses}_{ij}) + \epsilon_{ij}$$

$$\text{math}_{ij} = \mu_0 + (\mu_1)(\text{ses}_{ij}) + \Big[\eta_{0j} + (\eta_{1j})(\text{ses}_{ij}) + \epsilon_{ij}\Big]$$

The model essentially helps us to partition the variation in the residual...that which is due to between-school (school-to-school) variation (the random effects) and that which is within-school (student-to-student) error.

$$\text{math}_{ij} = \mu_0 + (\mu_1)(\text{ses}_{ij}) + \boxed{\epsilon_{ij}}$$

$$\text{math}_{ij} = \mu_0 + (\mu_1)(\text{ses}_{ij}) + \left[\boxed{\eta_{0j} + (\eta_{1j})(\text{ses}_{ij}) + \epsilon_{ij}}\right]$$

The mixed-effects model is used to specify the R syntax for fitting the model

$$\text{math}_{ij} = \boxed{\mu_0 + (\mu_1)(\text{ses}_{ij})} + \left[\boxed{\eta_{0j} + (\eta_{1j})(\text{ses}_{ij})} + \epsilon_{ij}\right]$$

```
> library(lme4)
> lmer.1 = lmer(math ~ 1 + ses + (1 + ses | school), data = students)
```

Level-2 ID

The `lmer()` function uses syntax similar to the `lm()` function.

- The fixed-effects use the exact same syntax
- The random-effects appear in parentheses and reference the level-2 ID variable

```
> summary(lmer.1)

Fixed effects:
            Estimate Std. Error t value
(Intercept)  12.6650     0.1898   66.71
ses           2.3938     0.1181   20.27
```

Note the estimates, SE, and *t*-value are close to those from the analysis on the grouped data.

```
# Grouped analysis

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.6573     0.1533   82.57   <2e-16 ***
meanSES       5.9093     0.3714   15.91   <2e-16 ***
---
```

```
> summary(lmer.1)

Fixed effects:
            Estimate Std. Error t value
(Intercept)  12.6650     0.1898   66.71
ses           2.3938     0.1181   20.27
```

The fixed-effects are interpreted exactly the same as we interpret the coefficients in an LM.

- The predicted mathematics score for a student at the average SES (ses = 0) is 12.67.
- The average difference in mathematics scores between students who are one standard deviation different in their SES is 2.39.

There are no *p*-values given for results from a mixed-effects analysis in R. That is because it is theoretically unclear what value for the *df* should be used for the *t*-test.

The output associated with the random-effects provide variance estimates for the associated random-effects in the model and also for the residuals.

```
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 school   (Intercept)  4.8286  2.1974
          ses          0.4129  0.6426   -0.11
 Residual             36.8302  6.0688
Number of obs: 7185, groups: school, 160
```

Estimated within-school variance: Quantifies the amount of variation **within** schools

Estimated between-school variance: Quantifies the amount of variation **between** schools.

$$\beta_{0j} \sim \mathcal{N}\left(\mu_0, \sigma^2_{\beta_{0j}}\right) \qquad \hat{\sigma}^2_{\beta_{0j}} = 4.8286$$

$$\beta_{1j} \sim \mathcal{N}\left(\mu_1, \sigma^2_{\beta_{1j}}\right) \qquad \hat{\sigma}^2_{\beta_{1j}} = 0.4129$$

The between-school variance estimates tell us how much the school-level intercept and slopes vary.

By partitioning the error variance, we get a more appropriate **residual variance** (MS) on which to test the effect of SES.

$$\hat{\sigma}_\epsilon^2 = 36.8302$$

```
> anova(lmer.1)

Analysis of Variance Table
    Df Sum Sq Mean Sq F value
ses  1  15126   15126   410.7
```

$$F = \frac{15126}{36.8302} = 410.6956$$

Again, no *p*-values given for the same reason...we have no idea what the appropriate *df* are.

We could test this if we knew which *df* to use in the denominator….1783? 158? somewhere in between?

```
# Best case scenario
> 1 - pf(410.7, df1 = 1, df2 = 1783)

[1] 0


# Worst case scenario
> 1 - pf(410.7, df1 = 1, df2 = 158)

[1] 0
```

In this example it doesn't matter, but in practice when the sample size and the number of groups is smaller this can make a world of difference.

Testing Fixed Effects in Mixed-Effects Models

$$t = \frac{\hat{\beta}_j}{\mathrm{SE}_{\hat{\beta}_j}}$$

Fixed effects in multilevel regression are typically tested by creating a ratio of the slope/intercept estimate to the estimate of the standard error. This kind of ratio, the Wald ratio, usually distributed as a z or t, is used in many statistical tests.

In the HLM software, the *df* associated with the t-distribution is computed as

$$df = N - q - 1$$

where N is the number of level-2 units (groups), q is the number of predictors in the model.

In practice, this formula does not seem to be used precisely, and if you re-run the analyses several times, you will notice somewhat different degrees of freedom listed in the output under—**approximate df**. In addition, test of the effects of cross-level interactions often use degrees of freedom based on the number of level-1 units (i.e., total number of individuals in the sample).

Testing Fixed Effects in Mixed-Effects Models

In SPSS, the same ratio is called z—a Wald test. SPSS uses a Satterthwaite adjustment to the *df* based on the number of level-1 cases. Satterthwaite degrees of freedom are a way of proportionally adjusting the *df* to provide a more accurate *p*-value estimate from the family of distributions.

SAS has many options for *df* adjustment. The Kenward–Rogers adjustment makes further adjustments based on the fact the the values used in the variance–covariance matrix are estimates and not known quantities.

Note. For samples in which the number of cases at both level-1 and level-2 are large, all methods give almost equivalent results. For small samples, it is recommended that the Kenward–Rogers method be used (Verbeke & Molenberghs, 2000).

Obtaining *p*-values for Fixed-Effects in R

```
# Best case scenario
> library(lmerTest)

# Re-fit the model
> lmer.1 = lmer(math ~ 1 + ses + (1 + ses | school), data = students)

> summary(lmer.1)

Fixed effects:
            Estimate Std. Error       df t value Pr(>|t|)
(Intercept)  12.6650     0.1898 145.5100   66.71   <2e-16 ***
ses           2.3938     0.1181 157.3900   20.27   <2e-16 ***
```