Sample Lecture Notes

Multilevel Models for Hierarchical Data

Daniel J. Bauer & Patrick J. Curran

CenterStat.org

Copyright © 2024 Curran-Bauer Analytics

All rights reserved

No part of these notes may be reproduced, translated, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without written permission by the copyright holders.

Chapter 2
Basic Multilevel Models

2.1 The Random-Effects ANOVA Model 2-3

2.2 Incorporating Lower-Level Predictors 2-15

2.3 Incorporating Upper-Level Predictors 2-25

2.4 Model Estimation 2-35

2.5 Frequently Asked Questions 2-44

2.1 The Random-Effects ANOVA Model

Objectives

- Present the random-effects ANOVA model
  - Differentiate fixed v. random effects
  - Explicate the variance components of the model
- Calculate and interpret the intraclass correlation coefficient (ICC)
- Begin to establish notation for multilevel models
  - We follow notation of Raudenbush & Bryk (2002)

CenterStat

Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods, Second Edition*. Thousand Oaks, CA: Sage Publications, Inc.

The Random-Effects ANOVA Model

- The simplest multilevel model
  - No predictors, only a random intercept
- Random intercept captures mean/level differences in the dependent variable across groups/clusters
- Decomposes the variance of the outcome into within- and between-group parts

Level 1 (within-group differences): $y_{ij} = \beta_{0j} + r_{ij}$

Level 2 (between-group differences): $\beta_{0j} = \gamma_{00} + u_{0j}$

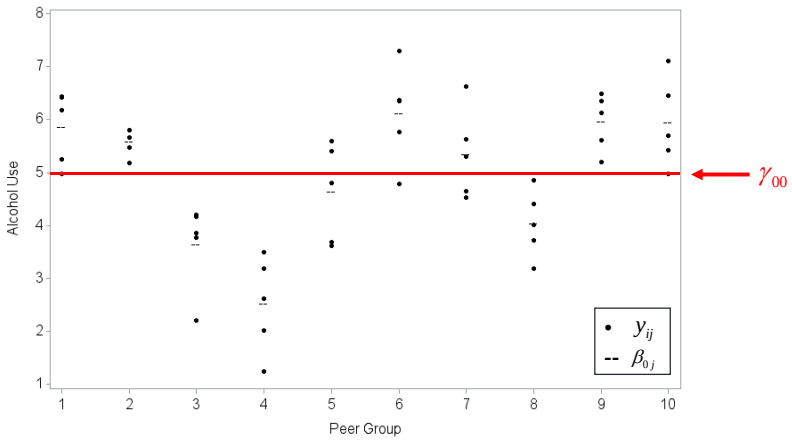
Reduced-Form (combined) Equation: $y_{ij} = \gamma_{00} + u_{0j} + r_{ij}$

CenterStat

The reduced-form equation (a.k.a. combined or "mixed-model" equation) is obtained by substituting the Level 2 equation into the Level 1 equation. The subscript *i* typically indicates individual, whereas *j* typically indicates group/cluster.

Visualization of Random-Effects ANOVA

- Two sources of variability around the overall average γ_{00}

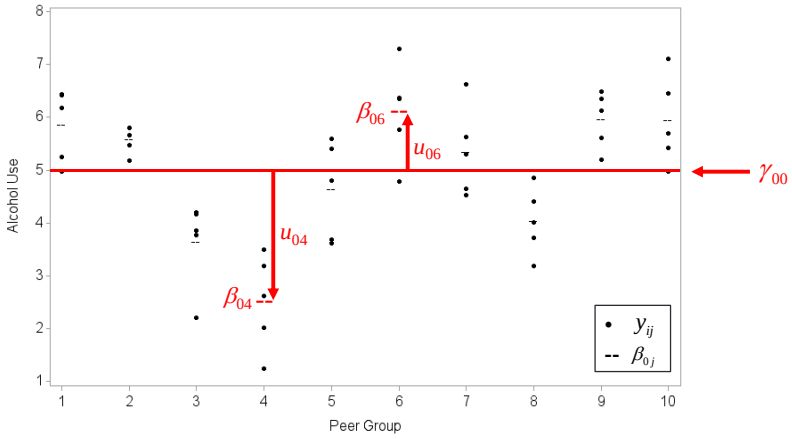


CenterStat

Similar plots were presented in Chapter 1, but now we see how the variability is represented within the random-effects ANOVA model. A minor difference from Chapter 1 is that we are visualizing the data here in terms of true population parameter values, rather than with respect to sample estimates.

Visualization of Random-Effects ANOVA

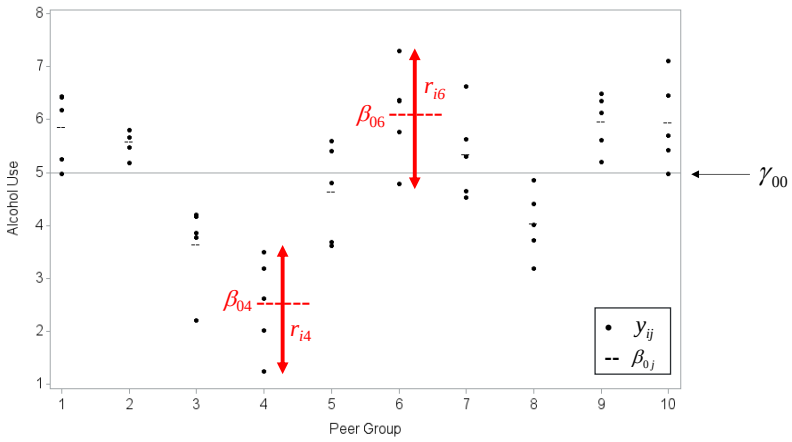
- Group mean relative to overall average: $\beta_{0j} = \gamma_{00} + u_{0j}$



CenterStat

Visualization of Random-Effects ANOVA

- Score on outcome relative to group mean: $y_{ij} = \beta_{0j} + r_{ij}$



CenterStat

Fixed Effects

- The reduced-form equation contains all terms within the model.

$$y_{ij} = \gamma_{00} + u_{0j} + r_{ij}$$

- *Fixed effects* are constant for all individuals in the population and hence carry no i or j subscript.
- The Greek symbol gamma is used for these effects (e.g., γ_{00}).

CenterStat

The numerical subscripts on gamma will become clearer later, once we have more effects in our model. Here, the first 0 on γ_{00} reflects that this is a term in the Level 2 equation for β_{0j} and the second 0 on γ_{00} reflects that it is the intercept within this Level 2 equation (a 0 being the numerical value used to designate an intercept). When we add predictors to the Level 2 equation for β_{0j}, we will see we will use symbols of γ_{01}, γ_{02}, etc., for the fixed-effects of the 1st, 2nd, and subsequent predictors.

Random Effects

- The reduced-form equation contains all terms within the model.

$$y_{ij} = \gamma_{00} + u_{0j} + r_{ij}$$

- Random effects are not constant and carry subscripts to indicate the units over which they vary.
  - u is used to indicate a group-level random effect.
  - r is used to indicate an individual-level random effect. We will, however, typically refer to r as a residual and reserve the term random effect for u.
- We do not estimate values for the random effect for each unit. We estimate the parameters of their distributions.

CenterStat

Confusingly, a random effect at Level 2 *is* a group-level residual and a residual at Level 1 *is* an individual-level random effect, but we tend to reserve the term *random effect* for the u's at Level 2 and the term *residuals* for the r's at Level 1.

Variance Components

- The reduced-form equation is $y_{ij} = \gamma_{00} + u_{0j} + r_{ij}$
- We interpret γ_{00} as the overall average, u_{0j} as a group difference and r_{ij} as an individual difference.
- The variance of r_{ij} is designated $V(r_{ij}) = \sigma^2$ (within-groups variance)
- The variance of u_{0j} is designated $V(u_{0j}) = \tau_{00}$ (between-groups variance)
- σ^2 and τ_{00} referred to as variance components or (co)variance parameters

CenterStat

In terms of notation, the double subscript on τ_{00} comes from the variance being a covariance of u_0 with itself, so u_0 with u_0 to get τ_{00}. This notation scheme will make more sense when we enter additional random effects to the model in Chapter 4 and need to track additional variances and covariances among the random effects.

The Variance Components of the Model

- The reduced-form equation is $y_{ij} = \gamma_{00} + u_{0j} + r_{ij}$
- We assume the residuals/random effects are normally distributed and uncorrelated across levels:

$$r_{ij} \sim N\left(0, \sigma^2\right) \qquad u_{0j} \sim N\left(0, \tau_{00}\right)$$

- This implies that:

$$E\left(y_{ij}\right) = \gamma_{00}$$

$$VAR\left(y_{ij}\right) = \tau_{00} + \sigma^2$$

$$COV\left(y_{ij}, y_{i'j}\right) = \tau_{00} \text{ for } i \neq i'$$

CenterStat

Given the presence of the random effects, residuals are typically assumed to be independent and identically distributed both within and across clusters. Similarly, random effects are typically assumed to be independent and identically distributed across clusters. These default assumptions can be modified if necessary.

The covariance expression here is a bit tricky to think about. We are accustomed to taking covariances (or their standardized values, correlations) between *variables* measured on the same people. Here, however, we are obtaining a covariance between *people* measured on the same variable.

We can see how the model captures within-groups dependence (covariance within groups) by considering two scenarios.

First, suppose we sample two people from <u>different groups</u>. We can designate them person i in group j and person i' in group j'. These two observations share nothing in common and are therefore independent. We can see how our model implies the independence of these two observations with some simple covariance algebra.

$$\begin{aligned} COV\left(y_{ij}, y_{i'j'}\right) &= COV\left(\gamma_{00} + u_{0j} + r_{ij}, \gamma_{00} + u_{0j'} + r_{i'j'}\right) \\ &= COV\left(u_{0j} + r_{ij}, u_{0j'} + r_{i'j'}\right) \\ &= COV\left(u_{0j}, u_{0j'}\right) + COV\left(u_{0j}, r_{i'j'}\right) + COV\left(r_{ij}, u_{0j'}\right) + COV\left(r_{ij}, r_{i'j'}\right) \\ &= 0 \end{aligned}$$

fixed effect drops out because it is a constant

Equals zero since assume r's independent over individuals

Equal zero since assume u and r to be uncorrelated

Equals zero since assume u's independent over groups

Next, suppose we sample two people from the same group. We can designate them person *i* in group *j* and person *i'* in group *j*. These two observations share the same group and therefore their observations on *y* will be dependent. We can again use covariance algebra to see how our model captures this dependence. Note that, relative to the prior equations, both observations now share the subscript *j*.

$$COV\left(y_{ij}, y_{i'j}\right) = COV\left(\gamma_{00} + u_{0j} + r_{ij}, \gamma_{00} + u_{0j} + r_{i'j}\right)$$

fixed effect drops out because it is a constant

$$= COV\left(u_{0j} + r_{ij}, u_{0j} + r_{i'j}\right)$$

$$= COV\left(u_{0j}, u_{0j}\right) + COV\left(u_{0j}, r_{i'j}\right) + COV\left(r_{ij}, u_{0j}\right) + COV\left(r_{ij}, r_{i'j}\right)$$

Equals zero since assume *r*'s independent over individuals

Equal zero since assume *u* and *r* to be uncorrelated

$$= COV\left(u_{0j}, u_{0j}\right)$$

Not zero since same group

$$= VAR\left(u_{0j}\right)$$

Covariance of something with itself is its variance

$$= \tau_{00}$$

Thus, the variance of the random intercept term, τ_{00}, which we interpret in this model to represent the variance of between-group mean differences in *y*, also represents the covariance between the *y* values of any two people sharing the same group. We thus see how between-group differences imply within-group dependence and how can capture both with the variance component τ_{00}.

The Intraclass Correlation

- The intraclass correlation (ICC) is obtained by standardizing the within-group covariance between observations into a within-group correlation:

$$\text{ICC} = \frac{\tau_{00}}{\sqrt{\tau_{00}+\sigma^2}\sqrt{\tau_{00}+\sigma^2}} = \frac{\tau_{00}}{\tau_{00}+\sigma^2}$$

- Since the denominator is equal to the total variance, the ICC can also be interpreted as measuring the proportion of variance due to between-group mean differences.
- Measures the degree of dependence in the data, or strength of the nesting (design) effect.

CenterStat

This is just like how a correlation between variables represents a standardized covariance. The only difference is that the correlation is now between people within a group on the same variable rather than between variables.

The Intraclass Correlation

$$\text{ICC} = \frac{\tau_{00}}{\tau_{00}+\sigma^2}$$

- If ICC = 0 then there are no between-group differences to produce dependence. The nesting of the data is irrelevant.
  - This is the assumption of GLM
  - Multilevel model reduces to GLM when ICC = 0
- If ICC = 1 then all differences are between-groups differences and all individuals within a group have identical scores.
- ICCs of .1 to .3 are common for individuals nested within groups.

CenterStat

Two Equivalent Interpretations of the ICC

- There are two equivalent ways to interpret the ICC
- Say the ICC was equal to .18
  1. Of the total variability observed in the outcome measure, approximately 18% can be attributed to between-group variation
  2. On average, observations that share the same group have outcomes that are correlated approximately .18
- Both of these are correct and either can be used in practice

CenterStat

Example: Siblings within Families

- Earlier we suggested that the birth order analysis might be compromised by dependence.
- 3313 kids nested within 2207 families:

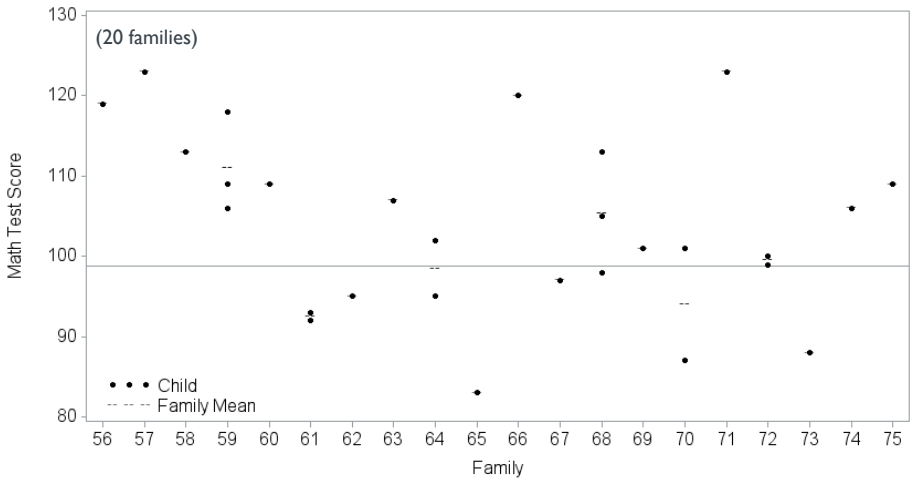
Frequency Count by Family

Obs in Data	*Frequency*	*Percent*	*Cumulative Frequency*	*Cumulative Percent*
1	1304	59.08	1304	59.08
2	719	32.58	2023	91.66
3	165	7.48	2188	99.14
4	19	0.86	2207	100.00

- Multiple children per family produces dependence; also provides an opportunity to separate within- v. between-family effects

CenterStat

Within- and Between-Family Math Variation

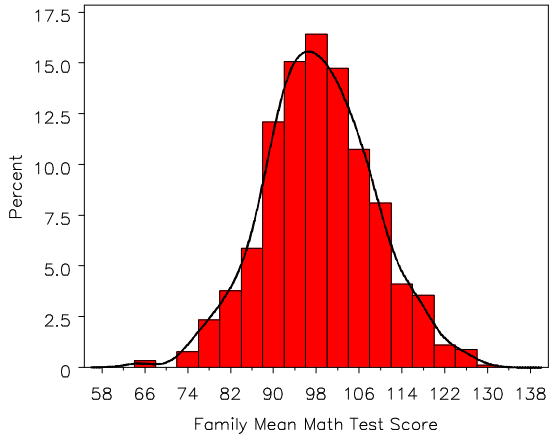


CenterStat

This index plot is again similar to the one described in Chapter 1, where family ID is arrayed on the x-axis for a subset of families (here, families 56-75) to show the distribution of math scores within and across families.

Between-Family Math Variation

- Histogram of family mean math scores for families with 2+ kids in data.



CenterStat

Note that there is a great deal of variation in the average math scores observed across families. We can more formally assess this variation via the Random-Effects ANOVA model.

Intra-Family Correlation

- To assess degree of dependence, and relative importance of within- and between-family differences, we shall fit a RE-ANOVA.

Level 1 (within-family differences):

$$Math_{ij} = \beta_{0j} + r_{ij} \qquad r_{ij} \sim N\left(0, \sigma^2\right)$$

Level 2 (between-family differences):

$$\beta_{0j} = \gamma_{00} + u_{0j} \qquad u_{0j} \sim N\left(0, \tau_{00}\right)$$

Reduced-Form (combined) Equation:

$$Math_{ij} = \gamma_{00} + u_{0j} + r_{ij} \qquad \text{ICC} = \frac{\tau_{00}}{\tau_{00} + \sigma^2}$$

CenterStat

RE-ANOVA Results

Covariance Parameter Estimates

	Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z	Lower	Upper
τ_{00}	Intercept	mom_id	52.12	4.40	11.85	<.0001	44.47	61.94
σ^2	Residual		102.30	4.03	25.36	<.0001	94.84	110.69

Solution for Fixed Effects

	Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
γ_{00}	Intercept	98.91	0.24	2206	414.50	<.0001

- Average Math is 98.91.
- About 1/3 of variance in Math can be attributed to between-family differences

$$\text{ICC} = \frac{52.12}{52.12 + 102.30} = .34$$

CenterStat

The ICC shows a high level of dependence and a relatively large contribution of between-family factors to Math.

The *p*-values shown here for the *z*-tests are one-tailed probabilities. A one-tailed *p*-value is used here because variances cannot be negative. Some software programs (e.g., SAS) report one-tailed *p*-values for variances and two-tailed *p*-values for covariances, whereas other software programs (e.g., SPSS) report two-tailed *p*-values for all variance and covariance estimates. Stata reports only confidence limits (although a *p*-value could be computed easily based on the reported estimate and standard error). In R, you mostly also rely on confidence limits.

"Lower" and "Upper" in the table above refer to 95% confidence limits. Importantly, the z-tests and confidence limits shown here are based on alternative assumptions about the sampling distributions of the variance parameters. In accordance with the asymptotic properties of maximum likelihood estimates (see Section 2.4), the *z*-tests assume a normal distribution (these are also known as Wald tests). Given the impossibility of negative variances, however, the sampling distributions of variances tend to be skewed in finite samples. The asymmetric confidence limits shown here are computed by SAS based on assuming a skewed sampling distribution (obtained via a Satterthwaite approximation). Similar in spirit, Stata computes z-type confidence limits for the log of the standard deviation, then transforms these values to obtain asymmetric limits for the variance. In R, you can request likelihood profile confidence limits. Although confidence limits obtained by any of these methods will generally be more accurate, they are not useful for null hypothesis tests because they will never include zero. (If the lower bound is close to zero, however, that is suggestive that the variance component is superfluous).

SPSS computes confidence limits assuming a normal sampling distribution, which has the virtue of permitting null hypothesis tests that agree with the reported *p*-values for the z-test, but have the downside of being inaccurate except with extremely large samples.

Later in this chapter we discuss the use of likelihood ratio tests to evaluate the significance of the variance at Level 2.

Summary of Analysis

- The random effects ANOVA indicates that there is a sizeable ICC for math scores within families.
  - Siblings within a family tend to have somewhat similar math test scores.
- Our previous GLM analysis results are thus formally incorrect
  - GLM assumes the ICC is zero.

CenterStat

Summary

- A fixed effect is constant for all individuals whereas a random effect has a distribution of values over units
- Random effects ANOVA model has
  - one fixed effect, γ_{00}, to capture the overall average
  - a residual, r_{ij}, to capture within-group differences
  - a random effect, u_{0j}, to capture between-group differences
  - within- and between-group variance components σ^2 and τ_{00}
- Key result of random-effects ANOVA is ICC
  - Correlation of outcome values within groups
  - Proportion variance arising from between-group differences

CenterStat

2.2 Incorporating Lower-Level Predictors

Objectives

- Include and evaluate predictors at Level 1
  - Interpret fixed effects
- Consider variance explained by predictors
  - For now, consider models with only random intercepts
  - Sometimes called *random intercept regression model*
  - Will consider random slopes for Level 1 predictors in Chapter 4

CenterStat

Including Lower-Level Predictors

- We will extend random-effects ANOVA model by including continuous and/or categorical predictors at Level 1
  - Categorical predictors included via coding variables
- Random intercept now captures group differences in the outcome that persist after controlling for predictors

CenterStat

A Single Continuous Predictor at Level 1

Level 1:

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

Level 2:

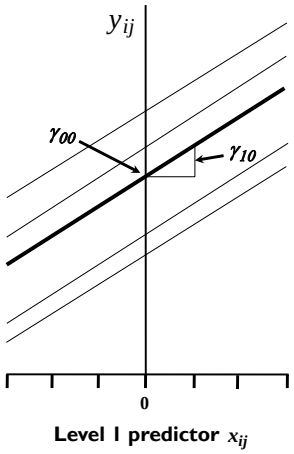
$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

Reduced-Form:

$$y_{ij} = \left(\gamma_{00} + u_{0j}\right) + \gamma_{10} x_{ij} + r_{ij}$$

$$= \left(\gamma_{00} + \gamma_{10} x_{ij}\right) + u_{0j} + r_{ij}$$



CenterStat

Note that a random intercept produces cluster-specific regression lines that are parallel, differing only in level. In Chapter 4, we will discuss models that do not assume the cluster-specific regression lines to be parallel (random slope models).

Explained Variance

- Unlike GLM, no model R^2
- Why? We no longer have one pool of variance; variance is partitioned into within- and between-group components
- We can, however, determine variance explained at each level of the model

CenterStat

Explained Variance

- Random-effects ANOVA

  Level 1: $y_{ij} = \beta_{0j} + r_{ij}$ $\quad r_{ij} \sim N\left(0, \sigma^2\right)$

  Level 2: $\beta_{0j} = \gamma_{00} + u_{0j}$ $\quad u_{0j} \sim N\left(0, \tau_{00}\right)$

  Reduced-Form Equation:

  $$y_{ij} = \gamma_{00} + u_{0j} + r_{ij}$$

- Here, σ^2 and τ_{00} are the unconditional (total) within- and between-groups variances for y

CenterStat

Explained Variance

- Extending the model to include a lower-level predictor

  Level 1: $y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + r_{ij}$ $\quad r_{ij} \sim N\left(0, \sigma^2\right)$

  Level 2: $\beta_{0j} = \gamma_{00} + u_{0j}$ $\quad u_{0j} \sim N\left(0, \tau_{00}\right)$

  $\beta_{1j} = \gamma_{10}$

  Reduced-Form Equation:

  $$y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + u_{0j} + r_{ij}$$

- Here, σ^2 and τ_{00} are the conditional (residual) within- and between-groups variances for y, after controlling for x

CenterStat

Explained Variance

- From Level 1 and 2 equations, it is not obvious that a Level 1 predictor can explain both Level 1 and Level 2 variance in y

Level 1: $y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + r_{ij}$

Level 2: $\beta_{0j} = \gamma_{00} + u_{0j}$

$\beta_{1j} = \gamma_{10}$

- In the Reduced-Form Equation, however, x is simply a predictor.

Reduced-Form Equation: $y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + u_{0j} + r_{ij}$

CenterStat

How can L1 predictor explain L2 variance?

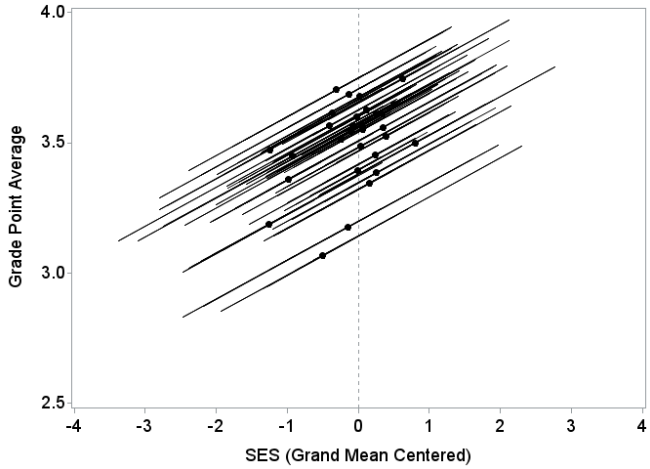
Reduced-Form Equation: $y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + u_{0j} + r_{ij}$

- Like the outcome variable, Level 1 predictors are also composed of two potential sources of variability, mean differences on x between groups and individual differences on x within groups.
  - Within-groups variability in x explains within-groups variability in y.
  - Between-groups variability in x explains between-groups variability in y.
- Example: SES as a predictor of GPA for students nested within schools
  - Within- and between-schools variability in SES predicts within- and between-schools variability in GPA

CenterStat

Within- & Between-Groups Variance in L1 Predictor

$$GPA_{ij} = \gamma_{00} + \gamma_{10} SES_{ij} + u_{0j} + r_{ij}$$



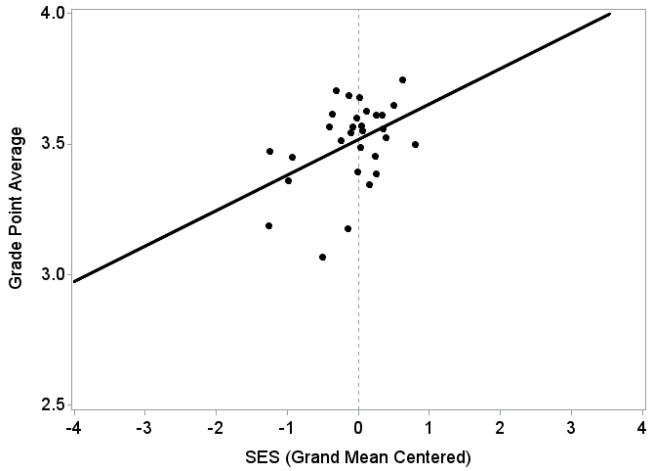
Within-school variance in SES predicts *within-school* variance in GPA

CenterStat

The regression lines show how within-groups variance in x explains within-groups variance in y. But note also that there is between-groups variance in both variables, depicted by the cluster means (dots).

Within- & Between-Groups Variance in L1 Predictor

$$GPA_{ij} = \gamma_{00} + \gamma_{10} SES_{ij} + u_{0j} + r_{ij}$$



Between-schools variance in SES predicts *between-schools* variance in GPA

CenterStat

Here we see the regression of the cluster means for y on the cluster means for x, demonstrating how between-groups variance in x can explain between-groups variance in y.

Computing Explained Variance

- Proportion of variance explained at Levels 1 and 2:

$$1-\frac{\sigma_C^2}{\sigma_U^2} \qquad 1-\frac{\tau_{00C}}{\tau_{00U}}$$

- U and C designate unconditional (total) and conditional (residual) variances at each level.
  - Estimates for parameters subscripted by U from random effects ANOVA.
  - Estimates for parameters subscripted by C from model including Level-1 predictor(s).
- Although useful, these are not true R^2 values (for instance, they can sometimes be negative).
- Helpful to keep in mind that these are "parts of parts" of variances
  - You can explain a very large proportion of a very small amount of variance

CenterStat

To make these computations, you must use the same sample when fitting the two models. Ensuring you have the same sample for each model gets a bit tricky if you have missing data: If there is missing data on x for cases with non-missing y then, by default, these observations will be included in the RE-ANOVA model but not the model including x. To compute variance explained, one would need to select only cases with observed, non-missing x values when fitting the RE-ANOVA.

Although these pseudo-R^2s are the most commonly reported measures of variance explained, other authors have advanced other measures. Rights & Sterba (2019) provide a general account of different ways to express explained variance in multilevel models, including a number of other measures not described here:

Rights, J.D., & Sterba, S.K. (2019). Quantifying explained variance in multilevel models: An integrative framework for defining R-squared measures. *Psychological Methods*, *24,* 309-338. https://doi.org/10.1037/met0000184

Example: Birth Order Effects on Math Scores

Level 1 (within-family differences):

$$Math_{ij} = \beta_{0j} + \beta_{1j}Second_{ij} + \beta_{2j}Third_{ij} + \beta_{3j}Fourth_{ij} + \beta_{4j}Old_{ij} + \beta_{5j}Second_{ij} \times Old_{ij} + \beta_{6j}Third_{ij} \times Old_{ij} + \beta_{7j}Fourth_{ij} \times Old_{ij} + r_{ij}$$

$$r_{ij} \sim N\left(0, \sigma^2\right)$$

Level 2 (between-family differences):

$$\beta_{0j} = \gamma_{00} + u_{0j} \qquad u_{0j} \sim N\left(0, \tau_{00}\right)$$

$$\beta_{pj} = \gamma_{p0} \text{ for } p > 0$$

CenterStat

Example: Birth Order Effects on Math Scores

Reduced-Form Equation:

$$Math_{ij} = \gamma_{00} + \quad \text{(intercept)}$$

$$\gamma_{10}Second_{ij} + \gamma_{20}Third_{ij} + \gamma_{30}Fourth_{ij} + \quad \text{(effect of birth order for younger cohort)}$$

$$\gamma_{40}Old_{ij} + \quad \text{(effect of cohort for first born)}$$

$$\gamma_{50}Second_{ij} \times Old_{ij} + \gamma_{60}Third_{ij} \times Old_{ij} + \gamma_{70}Fourth_{ij} \times Old_{ij} + \quad \text{(birth order x cohort)}$$

$$u_{0j} + \quad \text{(random family effect)}$$

$$r_{ij} \quad \text{(person residual)}$$

$$r_{ij} \sim N\left(0, \sigma^2\right)$$

$$u_{0j} \sim N\left(0, \tau_{00}\right)$$

(variance components)

CenterStat

Note that if $\tau_{00} = 0$ then the clustering of siblings within families is irrelevant and this model simplifies to the GLM that we fit in Chapter 1.

Results

Covariance Parameter Estimates

	Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z	Lower	Upper
τ_{00}	Intercept	mom_id	49.28	4.30	11.46	<.0001	41.83	58.92
σ^2	Residual		101.61	4.01	25.37	<.0001	94.20	109.94

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F	
brthordr	3	1098	11.53	<.0001	Main effects that pool over levels of other factor
cohort	1	460	30.02	<.0001	
cohort*brthordr	3	1098	0.34	0.7961	Global test of interaction between factors

CenterStat

For categorical predictors, Type III tests provide multi-degree of freedom tests of main effects and interactions (e.g., the 3 degree of freedom test of *brthordr* tests the global null hypothesis that there are no differences across birth orders).

Results

Solution for Fixed Effects

	Effect	brthordr	Estimate	Standard Error	DF	t Value	Pr > |t|	
γ_{00}	Intercept		101.50	0.47	2206	214.02	<.0001	
γ_{10}	brthordr	2	-1.53	0.67	1098	-2.29	0.0220	Still significant birth order effects
γ_{20}	brthordr	3	-2.74	0.80	1098	-3.40	0.0007	
γ_{30}	brthordr	4	-3.49	1.23	1098	-2.85	0.0045	
γ_{40}	cohort		-3.27	0.61	1098	-5.38	<.0001	
γ_{50}	cohort*brthordr	2	0.34	0.93	1098	0.37	0.7120	still no sig. difference in birth order effects for older cohort
γ_{60}	cohort*brthordr	3	-0.99	1.28	1098	-0.78	0.4378	
γ_{70}	cohort*brthordr	4	-0.08	2.09	1098	-0.04	0.9692	

Trends seen in fixed effects are fairly similar to the GLM results, but there is an important hidden result in the variance component estimates…

CenterStat

There are some minor differences in the fixed effect estimates relative to the GLM. The standard errors are also a bit smaller here than in the GLM. For Level 1 predictors, standard errors can sometimes be biased to be too large in GLM analyses, although more often they are too small.

Explained Variance

- Comparing the variance component estimates from the RE-ANOVA model to the model including birth order and cohort.
- Level 1 variance explained:

$$1-\frac{\hat{\sigma}_C^2}{\hat{\sigma}_U^2}=1-\frac{101.61}{102.30}=.007$$

- Level 2 variance explained:

$$1-\frac{\hat{\tau}_{00C}}{\hat{\tau}_{00U}}=1-\frac{49.28}{52.12}=.054$$

CenterStat

Summary of Results

- Less than 1% of within-family variance in math scores explained by birth order.
- Yet birth order explains 5.4% of between-family differences in math scores.
- Results are thus inconsistent with theories that posit birth order effects represent within-family dynamics.
  - this is reflected in much greater variability explained between-families (5.4%) compared to that explained within families (.7%)
- We can now bring in one or more Level 2 predictors to explicitly model factors associated with between-family differences.

CenterStat

Summary

- Extended the model to include predictors at Level 1
  - Sometimes called "random intercepts regression model"
- Fixed intercept and slopes interpreted much like in regular regression
- Level 1 predictors can explain variance at both levels
  - Within-group differences in x predict within-group differences in y
  - Between-group differences in x predict between-group differences in y
- Variance components now capture *residual* within- and between-group differences in outcome
  - Use to compute explained variance at each level of the model

CenterStat

2.3 Incorporating Upper-Level Predictors

Objectives

- Consider the incorporation of upper-level predictors in a multilevel model
  - Sometimes called *Means as Outcomes Model*
    - if only predictors are at Level 2
  - Sometimes called *Intercepts as Outcomes Model*
    - if predictors are included at both levels

CenterStat

Random Effects ANOVA Model

- Recall the random effects ANOVA model

Level 1: $y_{ij} = \beta_{0j} + r_{ij}$

Level 2: $\beta_{0j} = \gamma_{00} + u_{0j}$

Reduced-Form: $y_{ij} = \gamma_{00} + u_{0j} + r_{ij}$

- The coefficient β_{0j} represents the group mean of y for cluster j

CenterStat

Including Upper-Level Predictors

- We may wish to try to explain variation in β_{0j} across clusters using cluster-level predictors
- Constitutes prediction of group mean of y, hence sometimes referred to as *means as outcomes* model
- Example: Predicting behavior problems among children within a family as a function of mother's depression

CenterStat

Including Upper-Level Predictors

- We designate the Level 2 predictor (maternal depression) as w and we use it as a predictor of β_{0j} (family mean for child behavior problems)

Level 1: $y_{ij} = \beta_{0j} + r_{ij}$

Level 2: $\beta_{0j} = \gamma_{00} + \gamma_{01} w_j + u_{0j}$

Reduced-Form: $y_{ij} = \gamma_{00} + \gamma_{01} w_j + u_{0j} + r_{ij}$

- Note that u_{0j} now captures residual variation in the group means (variation between families not accounted for by maternal depression)

CenterStat

Variance Explained

- Can compute variance explained at Level 2 by w as

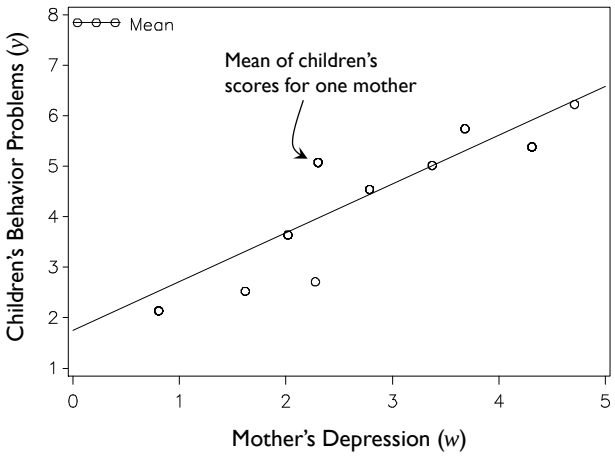
$$1-\frac{\tau_{00C}}{\tau_{00U}}$$

- where
  - τ_{00C} is the conditional (residual) variance obtained from the model with upper-level predictor(s)
  - τ_{00U} is the total variance at Level 2 obtained from the random effects ANOVA model
- Level 2 predictors can only explain Level 2 variance
  - No within-groups variance in w so can't explain within-groups variance in y

CenterStat

Means as Outcomes

Conceptually similar to regressing $\bar{y}_j$ on w.

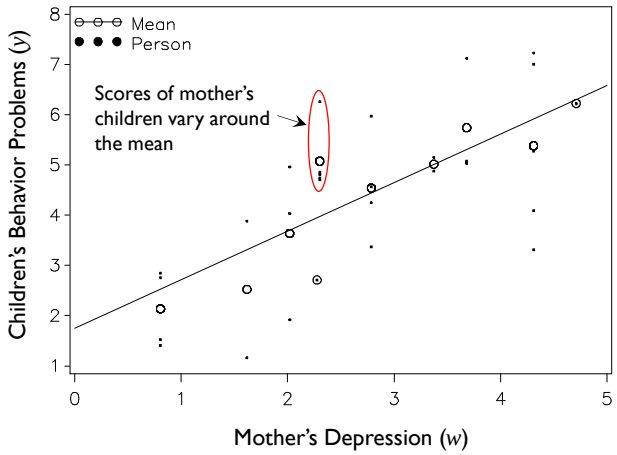


CenterStat

Literally computing and regressing cluster means for the outcome on cluster-level predictors is sometimes referred to as an aggregated analysis and is not recommended because it usually is conducted in way that ignores the standard errors of the cluster means which often differ between clusters due to different numbers of observations within clusters (among other possible reasons).

Means as Outcomes

But takes account of within-cluster variability and differential precision.



CenterStat

Differential precision occurs in part due to differences in the number of observations included per cluster. Multilevel models take account of this variation.

Including Predictors at Both Levels

- We may want to include predictors at both levels of the model.
  - For example, predicting child behavior problems by both the impulsivity of the child at Level 1 and maternal depression at Level 2.
- Level 2 predictors then explain between-groups variance that remains after controlling for Level 1 predictors.
  - After accounting for any family-level differences in impulsivity, what is the unique effect of maternal depression?
- Likewise, any effects of Level 1 predictors on between-group differences control for Level 2 predictors.
  - After controlling for maternal depression, does impulsivity still explain some of the between-family variance in child behavior problems?

CenterStat

Model with Just a Lower-Level Predictor

- Consider a model with just a single lower-level predictor

Level 1: $y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + r_{ij}$

Level 2: $\beta_{0j} = \gamma_{00} + u_{0j}$

$\beta_{1j} = \gamma_{10}$

Reduced-Form: $y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + u_{0j} + r_{ij}$

- We may wish to add Level 2 predictors to
  - Help explain between-cluster differences currently absorbed by u_{0j}
  - Control for between-cluster differences due to Level 2 predictors

CenterStat

Model With Predictor(s) at Each Level

- When we add an upper-level predictor (w) we again include it in the equation for β_{0j}

Level 1: $y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + r_{ij}$

Level 2: $\beta_{0j} = \gamma_{00} + \gamma_{01}w_j + u_{0j}$

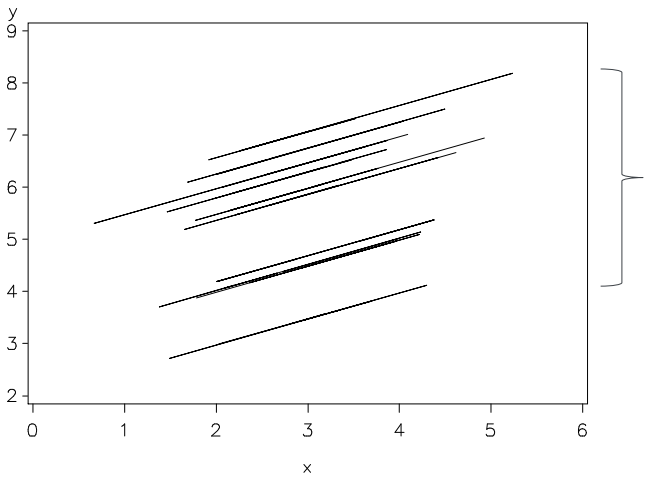
$\beta_{1j} = \gamma_{10}$

Reduced-Form: $y_{ij} = \gamma_{00} + \gamma_{01}w_j + \gamma_{10}x_{ij} + u_{0j} + r_{ij}$

- Now β_{0j} an intercept of regression equation (not a mean) so sometimes called an *intercepts as outcomes* model
- Can compute variance explained by w beyond that explained by x by considering proportional reduction in τ_{00} estimate.

CenterStat

Intercepts as Outcomes



Can level differences in within-group regression lines be explained by *w*?

CenterStat

Example: Birth Order

- In our earlier analysis we noted that birth order explains more between-family variance than within-family variance – why?
- Wichman, Rodgers & MacCallum (2006) offer an explanation:
  - Later born children, by definition, come from larger families
  - Larger family size is associated with lower maternal education, economic disadvantage, younger maternal age at first birth, etc.
  - Progressively higher proportion of later born children with disadvantaged mothers results in appearance of math score differences due to birth order.
- To evaluate this hypothesis, we will include the mother's age at first birth as an additional, upper-level predictor.

CenterStat

Example: Birth Order

- The model we fit previously, with just lower-level predictors, was:
  - Level 1:

$$Math_{ij} = \beta_{0j} + \beta_{1j}Second_{ij} + \beta_{2j}Third_{ij} + \beta_{3j}Fourth_{ij} + \beta_{4j}Old_{ij} + \beta_{5j}Second_{ij} \times Old_{ij} + \beta_{6j}Third_{ij} \times Old_{ij} + \beta_{7j}Fourth_{ij} \times Old_{ij} + r_{ij}$$

$$r_{ij} \sim N\left(0, \sigma^2\right)$$

  - Level 2:

$$\beta_{0j} = \gamma_{00} + u_{0j} \qquad u_{0j} \sim N\left(0, \tau_{00}\right)$$
$$\beta_{pj} = \gamma_{p0} \text{ for } p > 0$$

CenterStat

Example: Birth Order

- Now we add the confounder, age at first birth, to the intercept equation:
  - Level 1:

$$Math_{ij} = \beta_{0j} + \beta_{1j}Second_{ij} + \beta_{2j}Third_{ij} + \beta_{3j}Fourth_{ij} + \beta_{4j}Old_{ij} + \beta_{5j}Second_{ij} \times Old_{ij} + \beta_{6j}Third_{ij} \times Old_{ij} + \beta_{7j}Fourth_{ij} \times Old_{ij} + r_{ij}$$

$$r_{ij} \sim N\left(0, \sigma^2\right)$$

  - Level 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Birthage_j + u_{0j} \qquad u_{0j} \sim N\left(0, \tau_{00}\right)$$
$$\beta_{pj} = \gamma_{p0} \text{ for } p > 0$$

CenterStat

Example: Birth Order

Reduced-Form Equation for prior model:

$$Math_{ij} = \gamma_{00} + \gamma_{10}Second_{ij} + \gamma_{20}Third_{ij} + \gamma_{30}Fourth_{ij} + \gamma_{40}Old_{ij} + \gamma_{50}Second_{ij} \times Old_{ij} + \gamma_{60}Third_{ij} \times Old_{ij} + \gamma_{70}Fourth_{ij} \times Old_{ij} + u_{0j} + r_{ij}$$

Reduced-Form Equation for expanded model:

$$Math_{ij} = \gamma_{00} + \gamma_{01}BirthAge_{j} + \gamma_{10}Second_{ij} + \gamma_{20}Third_{ij} + \gamma_{30}Fourth_{ij} + \gamma_{40}Old_{ij} + \gamma_{50}Second_{ij} \times Old_{ij} + \gamma_{60}Third_{ij} \times Old_{ij} + \gamma_{70}Fourth_{ij} \times Old_{ij} + u_{0j} + r_{ij}$$

CenterStat

Results

Covariance Parameter Estimates

	Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z	Alpha	Lower	Upper
τ_{00}	Intercept	mom_id	43.33	4.14	10.48	<.0001	0.05	36.25	52.73
σ^2	Residual		101.65	3.99	25.46	<.0001	0.05	94.26	109.96

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
brthordr	3	1098	1.35	0.2557
cohort	1	460	7.30	0.0071
cohort*brthordr	3	1098	1.84	0.1385
brthage	1	2205	113.07	<.0001

Main effects and interaction controlling for *brthage*

Birth order main effect *ns* after controlling for mother's age at first birth

CenterStat

Results

Solution for Fixed Effects

	Effect	brthordr	Estimate	Standard Error	DF	t Value	Pr > |t|
γ_{00}	*Intercept*		99.11	0.52	2205	191.51	<.0001
γ_{10}	*brthordr*	2	0.02	0.67	1098	0.03	0.9797
γ_{20}	*brthordr*	3	0.17	0.84	1098	0.20	0.8403
γ_{30}	*brthordr*	4	0.12	1.26	1098	0.09	0.9261
γ_{40}	*cohort*		-0.28	0.66	1098	-0.42	0.6767
γ_{50}	*cohort*brthordr*	2	-0.59	0.92	1098	-0.64	0.5208
γ_{60}	*cohort*brthordr*	3	-2.80	1.28	1098	-2.19	0.0287
γ_{70}	*cohort*brthordr*	4	-2.47	2.08	1098	-1.19	0.2359
γ_{01}	*brthage*		0.73	0.07	2205	10.63	<.0001

Now no evidence of birth order effects for younger cohort

And still no interaction

Expected family math score increases by .7 points for each year delay childbirth

Birth order comparisons non-significant, controlling for mother's age at first birth

CenterStat

Although the coefficient γ_{60} is statistically significant, the overall test of the cohort × birth order interaction was not significant, indicating that this should probably not be interpreted. Moreover, the effect is unexpected theoretically, suggesting that the math differential between first and third born in the older cohort is significantly greater than in the younger cohort, despite no significant differences being observed for first versus second born or first versus fourth born children across cohorts.

Another possible point of confusion is that the Type 3 F-test for cohort reported on the prior slide and the t-test of the cohort coefficient on this slide have different p-values. This is due to the presence of the birth order by cohort interaction in the model. The Type 3 F-test of cohort pools across levels of birth order to provide a test of the "main effect" of cohort. In contrast, the t-test of the cohort coefficient is the effect of cohort when none of the interaction terms come into play, i.e., at the reference category for birth order, or first-born children. That is, the Type 3 F-test is for the main effect of cohort pooling over birth orders and the t-test is for the effect of cohort within first-born children. The difference in p-values is to be expected given that we are testing somewhat different things with the two statistics. In the absence of any interaction, the p-values would be identical (e.g., as they are here for birth age).

Summary of Results

- Previous model indicated that birth order effects explain between-family rather than within-family math score differences.
- Later born children tend to have mothers who began child bearing at a younger age.
- Controlling for maternal age at first birth, birth order differences in math scores become non-significant.
- Mothers who delay childbirth tend to have children with slightly higher math scores.

CenterStat

Summary

- Extended model to include predictors at Level 2
  - When only upper-level predictors, sometimes called *means as outcomes* model
  - Predicting between-group variation in outcome
- When including predictors at both levels
  - Level 2 predictors explain between-group variation that persists after controlling for Level 1 predictors
    - Between-mother differences in infant birth weights associated with beta interferon after controlling for gestational age
  - Also controlling for differences in L2 predictors when assessing L1 predictors
    - Birth order differences in children's math scores after controlling for mother's age at childbearing

CenterStat

2.4 Model Estimation

Objectives

- Describe full information maximum likelihood (ML) and restricted maximum likelihood (REML) estimation
- Compare and contrast ML and REML
- Describe likelihood ratio tests (LRTs)

CenterStat

Maximum Likelihood

- A likelihood can be defined that maps the parameters of a given model to the observed sample data.
- ML maximizes this likelihood to calculate a set of parameter estimates that are most consistent with the observed data.
- Desirable properties of ML include:
  - asymptotically unbiased, consistent, and efficient
  - asymptotic sampling distributions are normal
  - functions of ML parameters also ML estimates
  - availability of likelihood ratio tests for comparing nested models

CenterStat

Maximum Likelihood

- Potential limitation: ML estimates of (co)variance parameters conditioned on "known" population fixed effects, yet fixed effects estimated from sample data
- This results in bias in the variance component estimates
- To see this, let's look at a simple case, the estimation of a population variance in single-level data

CenterStat

Maximum Likelihood

- An unbiased estimate of the population variance can be obtained from the formula

$$\hat{\sigma}_y^2 = \left(\frac{1}{N}\right)\sum_{i=1}^{N}(y_i - \mu)^2$$

- The problem is that we usually do not know μ, so we substitute in the sample estimate

$$\hat{\sigma}_y^2 = \left(\frac{1}{N}\right)\sum_{i=1}^{N}(y_i - \bar{y})^2$$

- But this is known to produce a biased estimate of the population variance, particularly at small N

CenterStat

The ML estimate divides by the total sample size because it is assumed that there is no sampling variability in the estimation of the sample mean (i.e., sample mean is taken to be the population mean).

Restricted Maximum Likelihood (REML)

- REML account for our use of the sample estimate of the population mean to obtain an unbiased estimate.
- REML divides by $N-1$ to reflect loss of one degree-of-freedom due to the estimation of the sample mean

$$\hat{\sigma}_y^2 = \left(\frac{1}{N-1}\right)\sum_{i=1}^{N}\left(y_i - \bar{y}\right)^2$$

- This is precisely the difference between ML and REML.
  - ML treats fixed effects (like the mean) as if they were known when really estimated
  - REML adjusts for the fact that the fixed effects were estimated

CenterStat

The REML estimate uses the same numerator as the ML estimate, but it divides by one fewer than the sample size to serve as a correction for the fact that we are estimating the mean from the sample data, using up one piece of information (degree of freedom).

ML versus REML

- With larger J (Level 2 sample size), ML and REML will provide identical or nearly identical results for all effects
- With fewer groups, ML and REML will provide identical or nearly identical results for fixed effects but will provide different results for variance parameters
  - Usually, Level 1 residual variance is comparable
  - Variance estimates tend to diverge most at Level 2
  - ML tends to under-estimate Level 2 variances compared to REML
  - Differences between are larger with more fixed effects and lower J

CenterStat

Choosing between REML and ML

- Because ML is a full-information estimator, it will be more efficient (i.e., produce smaller standard errors)
- Because REML treats fixed effects as sample-based estimates, REML provides less biased estimates of variance components and standard errors when fewer Level-2 groups
- ML and REML converge as number of Level-2 groups becomes larger
- With fewer level 2 groups and/or many fixed effects, REML tends to be optimal
- However, there are also some important differences between REML & ML in testing nested models, and we explore this next

CenterStat

Here *N* refers to number of Level 2 units, as these are the independent sampling units. In general, with multilevel models, all *asymptotic* (i.e., large sample) properties pertain when the number of units at the highest level is large.

Testing Nested Models

- Define deviance to be -2 times the log of the likelihood:

$$D = -2LL$$

- A given model that is fitted to a sample of data results in a value of D. In isolation, D is not interpretable
- But if two models are nested, the difference between each model D is (usually) distributed as a chi-square
- Two models are nested if the parameters in one model are a direct subset of those of a second model
  - e.g., if model can be defined by imposing restrictions on the other model

CenterStat

Note that this use of the term "deviance" is somewhat loose. Deviance is properly defined relative to the -2LL of a saturated model, but for most multilevel models there is no identifiable saturated model.

Testing Nested Models

- Consider a null model deviance denoted D_0 with the number of estimated parameters q_0.
- Consider a less restricted model deviance denoted D_1 with the number of estimated parameters q_1.
- The difference between D_0 and D_1 is distributed as a large-sample chi-square on $df = q_1 - q_0$:

$$(D_0 - D_1) \sim \chi^2 (df = q_1 - q_0)$$

CenterStat

The difference between nested-model deviances is *asymptotically* distributed as a χ^2, that is, a sufficient sample size is needed to obtain this distribution. There are no universal rules that dictate what size "sufficient" might be.

Likelihood Ratio Test

$$(D_0 - D_1) \sim \chi^2 (df = q_1 - q_0)$$

- The difference between deviances is called a likelihood ratio test, or LRT.
- A significant LRT reflects that the inclusion of the additional parameters improves model fit.
- In contrast, a non-significant LRT reflects that the inclusion of the additional parameters does not improve model fit
  - That is, the more restricted (and thus more parsimonious) model is supported.

CenterStat

Testing Nested Models

- With ML, LRTs can be conducted for models that are nested with respect to either the fixed or random effects
- With REML, LRTs can only be conducted for models that are nested with respect to the random effects
- In general, REML is the recommended estimator unless LRTs are used to test nested fixed effects.
- But multi-parameter Wald tests of fixed effects can be validly conducted with either ML or REML because these are not based on LRTs.
- Thus, REML is the typical estimator of first choice.

CenterStat

Caveat: Testing Variance Components

- A problem is that test distribution of the LRT gets "wonky" if one of the tested parameters is a variance
  - e.g., testing whether a random intercept is needed to account for dependence within clusters with null model assuming $\tau_{00} = 0$
- The problem is that this is really a one-tailed hypothesis test because variances should only ever be positive
  - And less restricted model estimated with the constraint $\hat{\tau}_{00} \geq 0$
- Because the null hypothesis places τ_{00} on the boundary of permissible values, the LRT no longer has a simple chi-square distribution

CenterStat

Here we use the highly technical description "wonky" to convey the formal statistical property of weird, odd, or funky.

Caveat: Testing Variance Components

- The correct test distribution for the LRT when one or more parameters is restricted to a boundary is a weighted sum (or mixture) of chi-squares with different degrees of freedom
  - Sometimes called a "chi-bar-square" distribution
- ***Exactly what this looks like depends on the constraints within the null model relative to the less restricted model***
  - This can get a bit complicated, but we will consider common scenarios throughout the class
- Note that the standard LRT is always more conservative (higher *p*-value) than the correct LRT
  - So if significant by standard LRT then will be significant by correct LRT as well

CenterStat

Stram and Lee (1994, 1995) derived the chi-square-bar distribution for several common scenarios under the assumption of independent and homoscedastic residuals (see also Verbeke & Molenberghs, 2000). Although derived for ML, Morrell (1998) showed that the same results apply with REML estimation. In complex scenarios, the weights for the mixture may not be known exactly and may require approximation via simulation (Baey, Cournède & Kuhn, 2019).

Given the complications of conducting likelihood ratio tests involving random effect variance parameters, a variety of other inference procedures have been suggested. For instance, Du and Wang (2020) discuss the potential advantages of permutation tests relative to likelihood ratio tests. Presently, however, the likelihood ratio test remains the most commonly used approach for conducting inferences regarding the inclusion of random effects.

Baey, C., Cournède, P.-H. & Kuhn E. (2019). Asymptotic distribution of likelihood ratio test statistics for variance components in nonlinear mixed effects models. *Computational Statistics and Data Analysis, 135*, 107-122.

Du, H. & Wang, L. (2020). Testing variance components in linear mixed modeling using permutation. *Multivariate Behavioral Research, 55*, 120-136.

Morrell, C.H. (1998), Likelihood ratio testing of variance components in the linear mixed-effects model using restricted maximum likelihood. *Biometrics, 54*, 1560-1568.

Stram, D. & Lee, J. (1994). Variance components testing in the longitudinal mixed effects model.. *Biometrics, 50*, 1171–1177.

Stram, D. & Lee, J. (1995). Corrections to variance components testing in the longitudinal mixed effects model. *Biometrics, 51*, 1196.

Verbeke, G. & Molenberghs, G. (2000). *Linear mixed models for longitudinal data*: New York: Springer.

Caveat: Testing Variance Components

- One simple case is where the null model assumes $\tau_{00} = 0$ versus the alternative $\tau_{00} > 0$
  - Null and alternative models are equivalent otherwise
- Here the standard LRT would use the test distribution

$$(D_0 - D_1) \sim \chi^2(1)$$

- But the correct LRT uses

$$(D_0 - D_1) \sim .5\chi^2(0) + .5\chi^2(1)$$

  - where a $\chi^2(0)$ is just equal to zero
- In this case, the weights imply a *p*-value for the correct LRT that is ½ the *p*-value of the standard LRT

CenterStat

The case of testing a model with a random intercept only (e.g., random-effects ANOVA) versus a model without any random effects is one instance in which the correct LRT is quite easy to calculate. You simply divide the *p*-value of the standard LRT by 2. Note that this holds specifically under the common assumption that the Level-1 residuals are independent and homoscedastic.

Using the correct likelihood ratio test to make inferences about random effects is preferable to the Wald z or confidence limits reported by many software programs.

Summary

- Multilevel models usually fit by either ML or REML
- REML produces less biased variance component estimates whereas ML is provides more efficient estimates
  - Usually prioritize lower bias with REML but with many clusters these differences disappear
- Both estimators permit likelihood ratio tests, but ML more flexible
  - REML only allows LRTs when models nested in (co)variance parameters but have identical fixed effects
  - ML allows LRTs when models are nested with respect to (co)variance parameters and/or fixed effects
  - When testing variance against null of zero, LRT distribution gets wonky

CenterStat

2.5 Frequently Asked Questions

Objectives

- Questions addressed here:
  - How is random effects ANOVA related to the classical ANOVA model I learned about back in Intro Stats?
  - Can I obtain a significance test for the ICC?
  - Should I recompute and interpret the ICC after adding predictors?
  - You reported raw coefficients for all of your fixed effects – what about standardized coefficients?
  - Do I really have to use this chi-bar-square thing?

CenterStat

Fixed vs. Random Effects ANOVA Model?

- *The ANOVA I learned about in Intro Stats separates between-group and within-group variance to test group-mean differences. How is this different here?*
- There is much conceptual overlap between classical ANOVA and the MLM, particularly in the separation of within-groups and between-groups variability
- However, key to the MLM is that observed groups are considered a random sample drawn from a population of groups whereas in the classical ANOVA model groups are treated as fixed and known
- The goal of the MLM is thus to estimate the parameters that define the distributions believed to govern within- and between-group differences
  - and this allows inferences to be drawn back to the population of groups

CenterStat

Can you Test Significance of the ICC?

- Although it is easy to compute the numerical value of the ICC, (between-groups variance divided by total variance) a *p*-value or confidence interval is often not reported with the estimate
- Although *p*-values and CIs can be computed for the ICC, these are often of less use in practice
  1. The ICC can serve as a standardized effect size estimate and we need not hook our wagon to the null hypothesis testing infrastructure
  2. Rarely if ever make decisions about model building based on the statistical significance of the ICC
  3. We already have tests of significance of the variance components to go into computing the ICC and these are often of greater interest
- As such, we nearly always treat the ICC as a descriptive statistic

CenterStat

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1,* 30-46.

Is the ICC Interpretable with Predictors?

- The ICC is the ratio of between-groups variance to total variance
  - e.g., an ICC=.22 indicated that of the total variability observed in the outcome, approximately 22% can be attributed to between-group differences
- These variance estimates are obtained from the random effects ANOVA model with no predictors, thus capture <u>all</u> the variance in the outcome
  - however, level-1 predictors can explain both level-1 and level-2 variance, and level-2 predictors can explain level-2 variance
- Can again compute ratio, but this is now a <u>residual</u> ICC or ratio of <u>unexplained</u> variances
  - captures dependence for unexplained part of y that is represented by u <u>and</u> r
- Could be meaningful, but more common to report ICC without predictors

CenterStat

Can I Compute Standardized Coefficients?

- In the traditional regression model it is routine to estimate and report standardized estimates of the regression coefficients
  - a raw coefficient implies a one-unit change in x is associated with a one-unit change in y
  - a standardized coefficient implies that a one standard deviation change in x is associated with a one standard deviation change in y
  - This helps get a sense of the magnitude of effect that does not rely on the original metrics of the predictor and outcome variables
- However, it is much less common to see the calculation and reporting of standardized regression coefficients in the MLM, but why?

CenterStat

Can I Compute Standardized Coefficients?

- Consider a generic fixed effect γ
- Adapting the formula from traditional regression, we can obtain a standardized estimate by multiplying by the ratio of standard deviations:

$$\frac{SD_{predictor}}{SD_{outcome}}\hat{\gamma}$$

- Yet in MLM we have a decomposition of total variance into within-group and between-group components
  - Which SD should be used?

CenterStat

Can I Compute Standardized Coefficients?

- Often recommended to use "grand SD" pooling over all observations
  - implicit in commonly recommended strategy to standardize (z-score) variables before the analysis so that "raw" coefficients will be in SD units
- Yet Wang et al. (2019) note that this approach can be quite problematic
  - pooling over within- and between-group differences not always appropriate
  - better standardization options exist, though a bit more complicated
- Given these issues, we generally prefer the following strategy:
  - presentation and interpretation of raw regression coefficients
  - visual depictions of the magnitude of effects
  - quantification through reduction in unexplained variance at the appropriate level(s) of the model (i.e., pseudo-R^2 or change in pseudo-R^2)

CenterStat

Wang, L., Zhang, Q., Maxwell, S. E., & Bergeman, C. S. (2019). On standardizing within-person effects: Potential problems of global standardization. *Multivariate Behavioral Research, 54*, 382-403.

Do I Always Have to Use the Chi-Bar Mixture?

- It has recently become more widely expected that a proper chi-bar-square distribution be used for LRTs testing variance components
- However, this does become a bit tedious and model-dependent, and may not be always necessary
- The *p*-value for the traditional (and technically incorrect) LRT is always conservative, so using the correct chi-bar-square test distribution will only reduce the traditional *p*-value
- As such, if you have a significant *p*-value using the traditional method, it will always remain significant when conducting the correct LRT
- So the chi-bar-square LRT remains best, but there may be situations with large samples and high power where this is unnecessary

CenterStat

Summary

- Classic ANOVA and MLM both separate within-group and between-group variance, but only the MLM treats groups as random
- It is possible to compute a *p*-value for an ICC, but these are rarely seen in practice and the ICC is treated as an estimate of effect size
- Traditional ICC is computed in absence of predictors; can be computed with predictors but then interpreted with respect to residual variance
- Possible to obtain standardized estimates of regression coefficients, but complex and often best to instead examine explained variance
- Use of the chi-bar-square distribution for LRTs of random effects is best practice but not strictly necessary if standard LRT is already significant

CenterStat

Chapter Summary

- The random-effects ANOVA is the simplest multilevel model
  - Provides estimates of within- and between-groups variances and ICC
- Can include predictors solely at Level 1
  - Random intercepts regression model
- Can include predictors solely at Level 2
  - Means as outcomes model
- Can have predictors at both levels
  - Intercepts as outcomes model
- Estimates are typically obtained using ML or REML

CenterStat