

Chapter 2 Basic Multilevel Models

2.1	The Random-Effects ANOVA Model.....	2-3
2.2	Incorporating Lower-Level Predictors.....	2-13
2.3	Incorporating Upper-Level Predictors.....	2-23

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 Raudenbush & Bryk (2002), for the most part.

 CenterStat 2.3

For our field, the Raudenbush & Bryk (2002) notation is the most widely used notation for these models.

The Random-Effects ANOVA Model

- ▶ The simplest multilevel model.
 - ▶ No predictors, only a random intercept.
- ▶ A 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 2.4

Fixed and Random Effects

- ▶ The reduced-form equation is

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

- ▶ *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 2.5

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

Fixed and Random Effects

- ▶ The reduced-form equation is

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

- ▶ 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 2.6

The Variance Components of the Model

- ▶ 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 $V(r_{ij}) = \sigma^2$
- ▶ The variance of u_{0j} is designated $V(u_{0j}) = \tau_{00}$
- ▶ σ^2 and τ_{00} referred to as variance components or (co)variance parameters

 CenterStat 2.7

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:

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

- ▶ This implies that: $E(y_{ij}) = \gamma_{00}$

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

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

 CenterStat 2.8

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 assumptions can be modified.

An Expanded Case

- ▶ Consider subsample of 2 clusters of data with 3 persons each
- ▶ For instance, these might be siblings within families: Michael, Lindsay & Buster Bluth; Chris, Meg & Stewie Griffin

$$y_{ij} = \gamma_{00} + u_{0j} + r_{ij} \quad \Rightarrow \quad \begin{pmatrix} y_{11} \\ y_{21} \\ y_{31} \\ y_{42} \\ y_{52} \\ y_{62} \end{pmatrix} = \begin{pmatrix} \gamma_{00} \\ \gamma_{00} \\ \gamma_{00} \\ \gamma_{00} \\ \gamma_{00} \\ \gamma_{00} \end{pmatrix} + \begin{pmatrix} u_{01} \\ u_{01} \\ u_{01} \\ u_{02} \\ u_{02} \\ u_{02} \end{pmatrix} + \begin{pmatrix} r_{11} \\ r_{21} \\ r_{31} \\ r_{42} \\ r_{52} \\ r_{62} \end{pmatrix}$$

} Bluth Family
} Griffin Family

CenterStat 2.9

One would never really do a multilevel analysis with only two clusters (e.g., families), but the vectors get too long if this is expanded out to be more general.

An Expanded Case

- ▶ If r_{ij} independent of u_{0j} , $r_{ij} \sim N(0, \sigma^2)$ and $u_{0j} \sim N(0, \tau_{00})$, then

$$\begin{pmatrix} y_{11} \\ y_{21} \\ y_{31} \\ y_{42} \\ y_{52} \\ y_{62} \end{pmatrix} \sim N \left(\begin{pmatrix} \gamma_{00} \\ \gamma_{00} \\ \gamma_{00} \\ \gamma_{00} \\ \gamma_{00} \\ \gamma_{00} \end{pmatrix}, \begin{pmatrix} \tau_{00} + \sigma^2 & & & & & \\ & \tau_{00} + \sigma^2 & & & & \\ & & \tau_{00} + \sigma^2 & & & \\ & & & 0 & & \\ & & & & \tau_{00} + \sigma^2 & \\ & & & & & \tau_{00} + \sigma^2 \\ & & & & & & \tau_{00} + \sigma^2 \end{pmatrix} \right)$$

- ▶ Independent blocks
 - ▶ Characteristic of clustered data (each cluster defines a block)
- ▶ Variance of observations is $\tau_{00} + \sigma^2$
- ▶ Covariance between two observations within a block is τ_{00}

CenterStat 2.10

The covariance matrix of the observations has a block diagonal structure because clusters (e.g., families) are assumed to be independent (that is, the random effects/residuals are independent across clusters).

An Expanded Case

- ▶ What if we omitted u_{0j} from the model (implying $\tau_{00} = 0$)?

$$\begin{pmatrix} y_{11} \\ y_{21} \\ y_{31} \\ y_{42} \\ y_{52} \\ y_{62} \end{pmatrix} \sim N \left[\begin{pmatrix} \gamma_{00} \\ \gamma_{00} \\ \gamma_{00} \\ \gamma_{00} \\ \gamma_{00} \\ \gamma_{00} \end{pmatrix}, \begin{pmatrix} \sigma^2 & & & & & \\ 0 & \sigma^2 & & & & \\ 0 & 0 & \sigma^2 & & & \\ 0 & 0 & 0 & \sigma^2 & & \\ 0 & 0 & 0 & 0 & \sigma^2 & \\ 0 & 0 & 0 & 0 & 0 & \sigma^2 \end{pmatrix} \right]$$

- ▶ Variance of observations now pooled in σ^2 .
 - ▶ No partition of within- and between-family variance.
- ▶ Zero covariance, all observations implied to be independent.
 - ▶ Siblings within a family no more similar to one another than anyone else.

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.

The ICC equation draws on the general formula for computing a correlation, namely dividing the covariance by the product of the standard deviations. What is a little different here is that we are standardizing the covariance between two *observations* rather than the covariance between two *variables*.

The Intraclass Correlation

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

- ▶ 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 individuals within a group have identical scores.
- ▶ ICCs of .1 to .3 are common for individuals within groups.

 CenterStat 2.13

As we will see, higher ICCs are typical of other types of data, such as longitudinal data.

Example: Siblings within Families

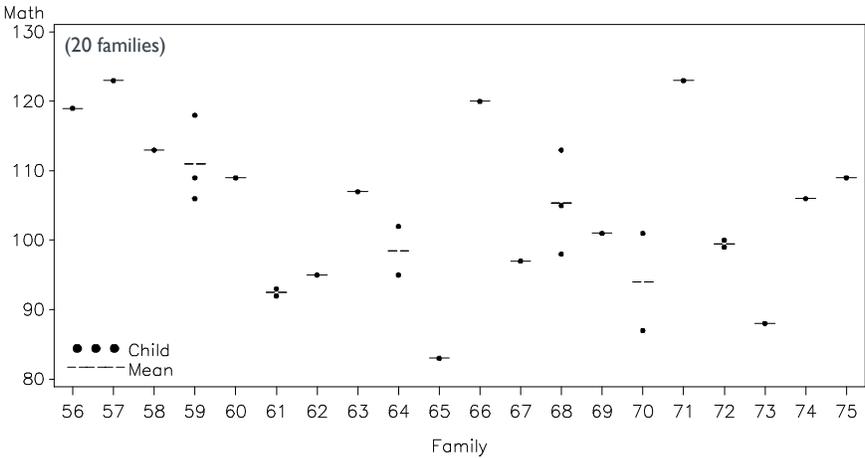
- ▶ Earlier we suggested that the birth order analysis might be compromised by dependence.
- ▶ 3313 kids nested within 2207 families:

<i>Obs in</i> <i>Data</i>	<i>Frequency Count by Family</i>		<i>Cumulative</i> <i>Frequency</i>	<i>Cumulative</i> <i>Percent</i>
	<i>Frequency</i>	<i>Percent</i>		
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 2.14

Within- and Between-Family Math Variation

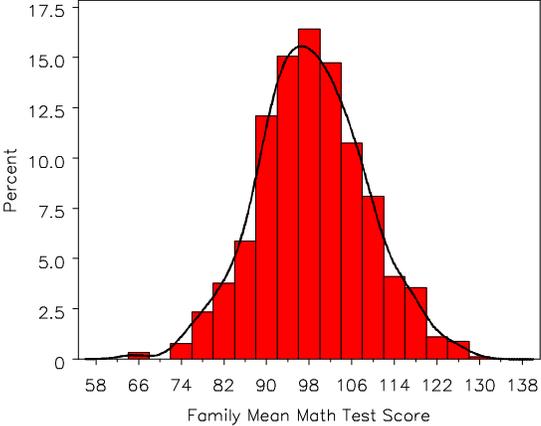


CenterStat 2.15

This index plot is similar to the one described in Chapter 1, where family ID is arrayed on the x-axis for a randomly chosen 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 2.16

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

Intrafamily 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):

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

Level 2 (between-family differences):

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

Reduced-Form (combined) Equation:

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

RE-ANOVA Results

Covariance Parameter Estimates								
			Standard					
	Cov Parm	Subject	Estimate	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						
		Standard				
	Effect	Estimate	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

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

CenterStat 2.18

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

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).

“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, the z -tests (or Wald tests) assume a normal distribution. 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. Although the confidence limits are generally 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.

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 2.19

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 2.20

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 a “random intercept regression model”
 - ▶ Will consider random slopes for Level 1 predictors in Chapter 4

 CenterStat 2.22

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 2.23

A Single Predictor at Level 1

Level 1:

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

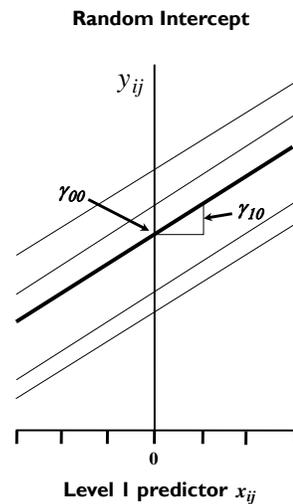
Level 2:

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

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

Reduced-Form:

$$\begin{aligned} y_{ij} &= (\gamma_{00} + u_{0j}) + \gamma_{10}x_{ij} + r_{ij} \\ &= (\gamma_{00} + \gamma_{10}x_{ij}) + u_{0j} + r_{ij} \end{aligned}$$



 2.24

Note that a random intercept produces cluster-specific regression lines that are parallel, differing only in level. We will later 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.

 2.25

Explained Variance

▶ Random-effects ANOVA

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

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

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 .

Explained Variance

▶ Extending the model to include a lower-level predictor

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

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

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

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 .

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

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

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

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

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

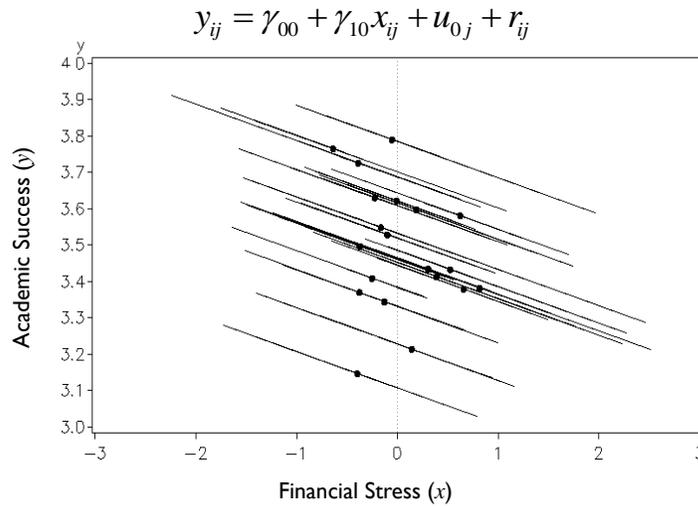
$$\text{Reduced-Form Equation: } y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + u_{0j} + r_{ij}$$

How can L1 predictor explain L2 variance?

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

- ▶ Like the outcome variable, Level 1 predictors are also composed of two 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: Financial stress as predictor of academic success among university students

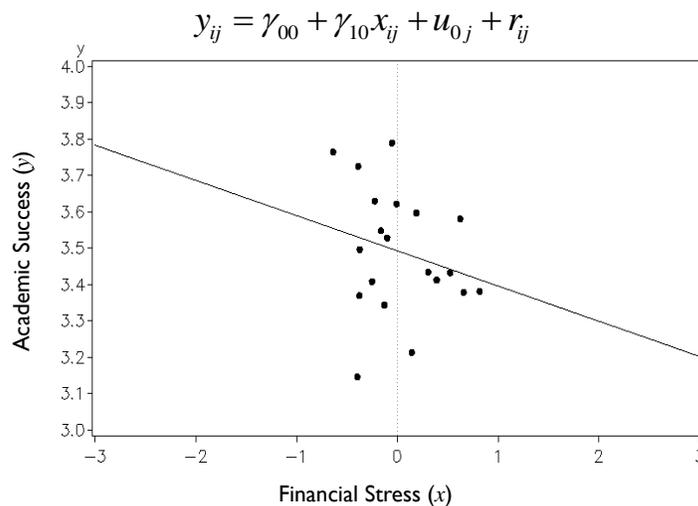
Within- & Between-Groups Variance in L1 Predictor



CenterStat 2.30

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



CenterStat 2.31

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 are obtained from random effects ANOVA.
 - ▶ Estimates for parameters subscripted by C are obtained from model with Level-I predictor(s).
- ▶ Although useful, these are not true R^2 values (for instance, they can sometimes be negative).

 CenterStat 2.32

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^2 s are the most commonly reported measures of variance explained, other authors have advanced other measures. Rights & Sterba (in press) 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. (In press). Quantifying explained variance in multilevel models: An integrative framework for defining R-squared measures. *Psychological Methods*.

Example: Birth Order Effects on Math Scores

Level 1 (within-family differences):

$$\begin{aligned} \text{Math}_{ij} = & \beta_{0j} + \beta_{1j}\text{Second}_{ij} + \beta_{2j}\text{Third}_{ij} + \beta_{3j}\text{Fourth}_{ij} + \beta_{4j}\text{Old}_{ij} + \\ & \beta_{5j}\text{Second}_{ij} \times \text{Old}_{ij} + \beta_{6j}\text{Third}_{ij} \times \text{Old}_{ij} + \beta_{7j}\text{Fourth}_{ij} \times \text{Old}_{ij} + r_{ij} \\ r_{ij} \sim & N(0, \sigma^2) \end{aligned}$$

Level 2 (between-family differences):

$$\begin{aligned} \beta_{0j} = & \gamma_{00} + u_{0j} & u_{0j} \sim N(0, \tau_{00}) \\ \beta_{pj} = & \gamma_{p0} \text{ for } p > 0 \end{aligned}$$

 2.33

Example: Birth Order Effects on Math Scores

Reduced-Form Equation:

$$\begin{aligned} \text{Math}_{ij} = & \gamma_{00} + & & \text{(intercept)} \\ & \gamma_{10}\text{Second}_{ij} + \gamma_{20}\text{Third}_{ij} + \gamma_{30}\text{Fourth}_{ij} + & & \text{(main effect of birth order)} \\ & \gamma_{40}\text{Old}_{ij} + & & \text{(main effect of cohort)} \\ & \gamma_{50}\text{Second}_{ij} \times \text{Old}_{ij} + \gamma_{60}\text{Third}_{ij} \times \text{Old}_{ij} + & & \text{(birth order x cohort)} \\ & \gamma_{70}\text{Fourth}_{ij} \times \text{Old}_{ij} + & & \\ & u_{0j} + & & \text{(random family effect)} \\ & r_{ij} & & \text{(person residual)} \\ r_{ij} \sim & N(0, \sigma^2) \\ u_{0j} \sim & N(0, \tau_{00}) \end{aligned}$$

 2.34

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								
			Standard	Z				
	Cov Parm	Subject	Estimate	Error	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
cohort	1	460	30.02	<.0001
cohort*brthordr	3	1098	0.34	0.7961

 CenterStat 2.35

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 simultaneously tests whether $\gamma_{10} = 0$, $\gamma_{20} = 0$, $\gamma_{30} = 0$, or the global null hypothesis that there are no differences across birth orders).

Results

Solution for Fixed Effects						
Effect	brthodr	Estimate	Standard Error	DF	t Value	Pr > t
γ_{00} Intercept		101.50	0.47	2206	214.02	<.0001
γ_{10} brthodr	2	-1.53	0.67	1098	-2.29	0.0220
γ_{20} brthodr	3	-2.74	0.80	1098	-3.40	0.0007
γ_{30} brthodr	4	-3.49	1.23	1098	-2.85	0.0045
γ_{40} cohort		-3.27	0.61	1098	-5.38	<.0001
γ_{50} cohort*brthodr	2	0.34	0.93	1098	0.37	0.7120
γ_{60} cohort*brthodr	3	-0.99	1.28	1098	-0.78	0.4378
γ_{70} cohort*brthodr	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 2.36

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 2.37

Summary of Results

- ▶ Less than 1% of within-family variance in math scores explained by birth order.
- ▶ Birth order explains 5.4% of between-family differences in math scores.
- ▶ Results are inconsistent with theories that posit birth order effects represent within-family dynamics.
- ▶ We can now bring in one or more Level 2 predictors to explicitly model factors associated with between-family differences.

 CenterStat 2.38

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 2.39

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 2.41

Random Effects ANOVA Model

- ▶ Recall the random effects ANOVA model

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

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

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

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

 CenterStat 2.42

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

 2.43

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)

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

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

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

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

 2.44

Variance Explained

- ▶ Can compute variance explained at Level 2 by w as

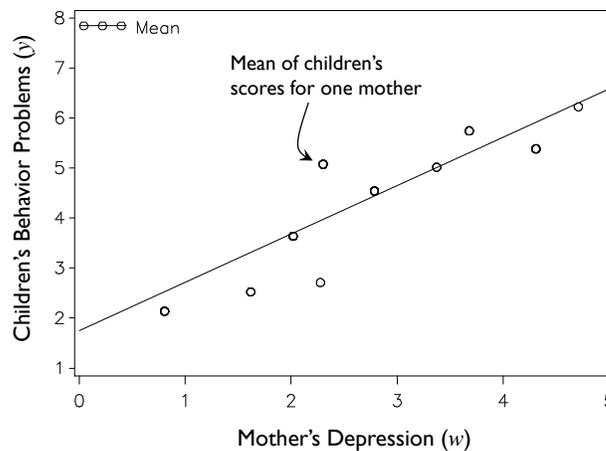
$$1 - \frac{\tau_{00C}}{\tau_{00U}}$$

- ▶ 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 .

Means as Outcomes

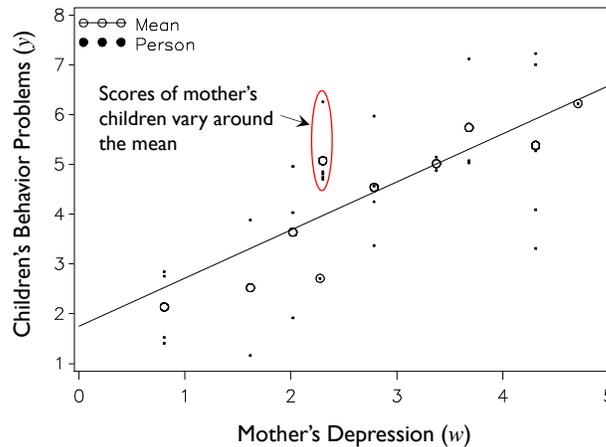
Conceptually similar to regressing \bar{y}_j on w .



Literally computing and regressing cluster means for the outcome on cluster-level predictors is sometimes referred to as an aggregated analysis.

Means as Outcomes

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



CenterStat 2.47

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 2.48

Model with Just a Lower-Level Predictor

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

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

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

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

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

- ▶ 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.

 2.49

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}

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

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

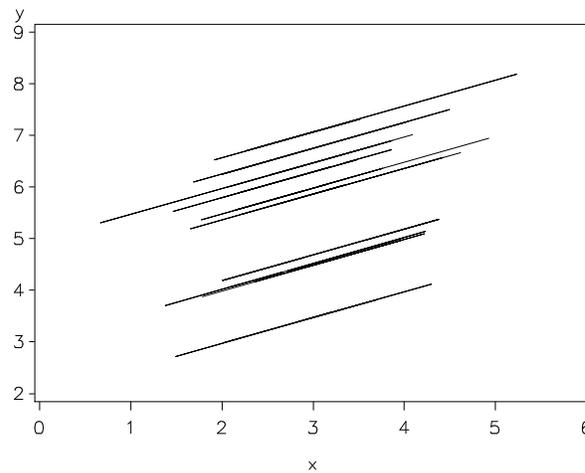
$$\beta_{1j} = \gamma_{10}$$

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

- ▶ 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.

 2.50

Intercepts as Outcomes



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

CenterStat 2.51

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 addition, upper-level predictor.

CenterStat 2.52

Example: Birth Order

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

- ▶ Level 1:

$$\begin{aligned} \text{Math}_{ij} = & \beta_{0j} + \beta_{1j}\text{Second}_{ij} + \beta_{2j}\text{Third}_{ij} + \beta_{3j}\text{Fourth}_{ij} + \beta_{4j}\text{Old}_{ij} + \\ & \beta_{5j}\text{Second}_{ij} \times \text{Old}_{ij} + \beta_{6j}\text{Third}_{ij} \times \text{Old}_{ij} + \beta_{7j}\text{Fourth}_{ij} \times \text{Old}_{ij} + r_{ij} \\ r_{ij} \sim & N(0, \sigma^2) \end{aligned}$$

- ▶ Level 2:

$$\begin{aligned} \beta_{0j} = & \gamma_{00} + u_{0j} & u_{0j} \sim N(0, \tau_{00}) \\ \beta_{pj} = & \gamma_{p0} \text{ for } p > 0 \end{aligned}$$

Example: Birth Order

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

- ▶ Level 1:

$$\begin{aligned} \text{Math}_{ij} = & \beta_{0j} + \beta_{1j}\text{Second}_{ij} + \beta_{2j}\text{Third}_{ij} + \beta_{3j}\text{Fourth}_{ij} + \beta_{4j}\text{Old}_{ij} + \\ & \beta_{5j}\text{Second}_{ij} \times \text{Old}_{ij} + \beta_{6j}\text{Third}_{ij} \times \text{Old}_{ij} + \beta_{7j}\text{Fourth}_{ij} \times \text{Old}_{ij} + r_{ij} \\ r_{ij} \sim & N(0, \sigma^2) \end{aligned}$$

- ▶ Level 2:

$$\begin{aligned} \beta_{0j} = & \gamma_{00} + \gamma_{01}\text{Birthage}_j + u_{0j} & u_{0j} \sim N(0, \tau_{00}) \\ \beta_{pj} = & \gamma_{p0} \text{ for } p > 0 \end{aligned}$$

Example: Birth Order

Reduced-Form Equation for prior model:

$$\begin{aligned} \text{Math}_{ij} = & \gamma_{00} + \gamma_{10}\text{Second}_{ij} + \gamma_{20}\text{Third}_{ij} + \gamma_{30}\text{Fourth}_{ij} + \\ & \gamma_{40}\text{Old}_{ij} + \gamma_{50}\text{Second}_{ij} \times \text{Old}_{ij} + \gamma_{60}\text{Third}_{ij} \times \text{Old}_{ij} + \gamma_{70}\text{Fourth}_{ij} \times \text{Old}_{ij} + \\ & u_{0j} + r_{ij} \end{aligned}$$

Reduced-Form Equation for expanded model:

$$\begin{aligned} \text{Math}_{ij} = & \gamma_{00} + \gamma_{01}\text{BirthAge}_{ij} + \gamma_{10}\text{Second}_{ij} + \gamma_{20}\text{Third}_{ij} + \gamma_{30}\text{Fourth}_{ij} + \\ & \gamma_{40}\text{Old}_{ij} + \gamma_{50}\text{Second}_{ij} \times \text{Old}_{ij} + \gamma_{60}\text{Third}_{ij} \times \text{Old}_{ij} + \gamma_{70}\text{Fourth}_{ij} \times \text{Old}_{ij} + \\ & u_{0j} + r_{ij} \end{aligned}$$

CenterStat 2.55

Results

Covariance Parameter Estimates								
			Standard	Z				
<i>Cov Parm</i>	<i>Subject</i>	<i>Estimate</i>	<i>Error</i>	<i>Value</i>	<i>Pr > Z</i>	<i>Alpha</i>	<i>Lower</i>	<i>Upper</i>
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				
<i>Effect</i>	<i>Num DF</i>	<i>Den DF</i>	<i>F Value</i>	<i>Pr > F</i>
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

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

CenterStat 2.56

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

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 2.57

Although the coefficient γ_{60} is statistically significant, the overall test of the cohort \times birth order interaction was not significant, indicating that this should probably not be interpreted. Moreover, the effect is unexpected theoretically, suggesting that the math IQ 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 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 test of cohort pools across levels of birth order to provide a true 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 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 2.58

Example: Beta Interferon

- ▶ Data collected on the pregnancies of mothers with Multiple Sclerosis (MS) and healthy controls, courtesy of Dr. Gideon Koren who runs the Motherisk Program at the University of Toronto, Hospital for Sick Children.
- ▶ 37 mothers participated in the study, some more than once, for a total of 46 pregnancies.
- ▶ Some mothers with MS discontinued use of Beta Interferon medication prior to conception while others did not.
- ▶ We want to see if Beta Interferon causes a difference in the birth weights of infants above and beyond MS.

 CenterStat 2.59

Original Analysis

- ▶ The authors originally did not account for the fact that some mothers had more than one pregnancy represented in the data.
- ▶ Number of mothers with 1, 2, 3 pregnancies:

COUNT	Frequency Count			
	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	31	83.78	31	83.78
2	3	8.11	34	91.89
3	3	8.11	37	100.00

- ▶ The authors believed this to be ignorable and used GLM.
- ▶ What happened?

Reviewer Comments

- ▶ Reviewer 1:

“The main problem with the data analysis is that it fails to account for the clustering of gestations within mother. The analysis should be based on a linear mixed model...”

- ▶ Reviewer 2:

“There appears to be an effect due to clustering of infants at the level of the mother, and I suspect given the size of this study the main results of this paper are no longer significant once this clustering effect is accounted for... Statistical tools for handling correlated responses are readily available today.”

Reanalysis

- ▶ We will first fit a Random-Effects ANOVA to determine the ICC and obtain baseline estimates of variance at the two levels of the data (mother and child).
- ▶ We will then add the upper-level predictor Group (Beta-Interferon, Discontinued, Healthy Control)
- ▶ We will then observe whether group differences persist after controlling for maternal weight gain and gestational age (Level 1 covariates).
- ▶ Given small sample sizes, we will not conduct tests on the variance components of the model and we will use the Kenward-Roger's method of testing fixed effects.

 CenterStat 2.62

Note: The Kenward-Roger's method for testing fixed effects involves two adjustments that aim to provide better protection against Type I errors in small samples. First, degrees of freedom are determined using a Satterthwaite procedure. Second, the standard errors of the fixed effects are adjusted upwards to guard against small sample bias and sampling error in the variance components

Random-Effects ANOVA

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

Level 1:

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

Level 2:

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

Reduced-Form Equation:

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

 CenterStat 2.63

Random-Effects ANOVA

Covariance Parameter Estimates			
Cov Parm	Subject	Estimate	
τ_{00}	Intercept	MID	0.95
σ^2	Residual		0.28

$$ICC = \frac{.95}{.95 + .28} = .77$$

Solution for Fixed Effects								
Effect	Estimate	Error	DF	t Value	Pr > t	Lower	Upper	
γ_{00}	Intercept	7.75	0.18	35.4	43.01	<.0001	7.38	8.11

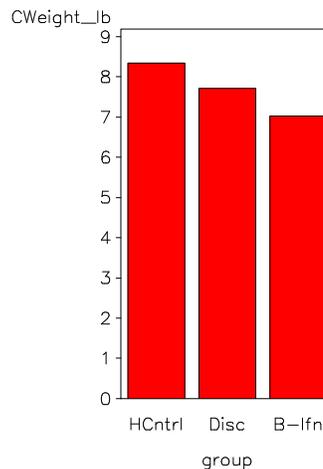
- ▶ Pooling over groups, average birth weight is 7.7 lbs.
- ▶ ICC indicates a very high level of dependence (similar birth weights for children born to same mother)

CenterStat 2.64

No standard errors or z-tests are presented for the variance parameters because the number of mothers contributing to this partitioning of variance is too small to permit useful inferences.

Examining Group Differences

- ▶ Group mean differences:



CenterStat 2.65

There is a difference in birth weights across groups in the sample, but we would like to determine whether this is sufficient evidence for a similar difference in the population.

Evaluating Upper-Level Predictor

- ▶ Assess to what extent the large variance attributable to mothers is due to the differences between the three groups.

Level 1:

$$CWeight_lb_{ij} = \beta_{0j} + r_{ij} \quad r_{ij} \sim N(0, \sigma^2)$$

Level 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}B_Ifn_j + \gamma_{02}Disc_j + u_{0j} \quad u_{0j} \sim N(0, \tau_{00})$$

Reduced-Form Equation:

$$CWeight_lb_{ij} = \gamma_{00} + \gamma_{01}B_Ifn_j + \gamma_{02}Disc_j + u_{0j} + r_{ij}$$

- ▶ A “means as outcomes” model.

Evaluating Upper-Level Predictor

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
τ_{00} Intercept	MID	0.71
σ^2 Residual		0.28

$$1 - \frac{\hat{\tau}_{00C}}{\hat{\tau}_{00U}} = 1 - \frac{.71}{.95} = .25$$

- ▶ 25% of variance between mothers explained by group.
- ▶ Mean differences between the three groups are significant.

Type 3 Tests of Fixed Effects				
Effect	Num		F Value	Pr > F
	DF	Den DF		
group	2	32.8	5.52	0.0086

Estimates of Fixed Effects

Solution for Fixed Effects									
			Standard						
	Effect	group	Estimate	Error	DF	t Value	Pr > t	Lower	Upper
γ_{00}	Intercept		8.30	0.24	35.1	34.59	<.0001	7.81	8.79
γ_{01}	group	B-lfn	-1.27	0.39	34.8	-3.22	0.0027	-2.07	-0.47
γ_{02}	group	Disc	-0.76	0.39	31.6	-1.96	0.0588	-1.54	0.030

- ▶ Expected birth weight for healthy control = 8.30 lbs
- ▶ For mothers with MS taking beta-interferon, expect infant birth weight to be 1.27 lbs less (15% lower) than healthy control.
- ▶ For mothers with MS who discontinued beta-interferon, expect infant birth weight to be .76 lbs less (9% lower) than healthy control (non-significant).

 CenterStat 2.68

Controlling for Lower-Level Covariates

- ▶ Determine whether group differences persist after controlling for maternal weight gain and gestational age.

Level 1:

$$CWeight_lb_{ij} = \beta_{0j} + \beta_{1j}GestAge_{ij} + \beta_{2j}MGain_lb_{ij} + r_{ij}$$

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

- ▶ Now an “intercepts as outcomes” model
- ▶ Gestational age and maternal weight gain centered at median values for healthy controls so that β_{0j} represents expected birth weight for child of mother j assuming typical gestation.

 CenterStat 2.69

We used median centering here because one mother in the healthy control group gained an unusually high amount of weight during pregnancy. As a measure of central tendency, the median is robust to extreme observations, whereas the mean is not. We discuss centering in greater detail in Chapter 3.

Controlling for Lower-Level Covariates

Level 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}B\text{-}I\text{-}fn_j + \gamma_{02}Disc_j + u_{0j} \quad u_{0j} \sim N(0, \tau_{00})$$

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

$$\beta_{2j} = \gamma_{20}$$

Reduced-Form Equation:

$$CWeight_lb_{ij} = \gamma_{00} + \gamma_{01}B\text{-}I\text{-}fn_j + \gamma_{02}Disc_j + \\ \gamma_{10}GestAge_{ij} + \gamma_{20}MGain_lb_{ij} + \\ u_{0j} + r_{ij}$$

 2.70

Controlling for Lower-Level Covariates

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
τ_{00} Intercept	MID	0.33
σ^2 Residual		0.33

Type 3 Tests of Fixed Effects				
Effect	Num Den		F Value	Pr > F
	DF	DF		
GestAge	1	30.7	15.41	0.0005
MGain_lb	1	31.4	1.77	0.1932
group	2	23.4	4.57	0.0211

Group effect still significant

 2.71

Note that these variance component estimates suggest a negative amount of variance explained at Level 1 by the two covariates. In part, this likely reflects instability in these estimates due to the small sample. Additionally, however, direct comparisons of the variance estimates between models are not appropriate in this case because two observations with missing maternal weight gain were excluded here but included in the prior models

Estimates of Fixed Effects

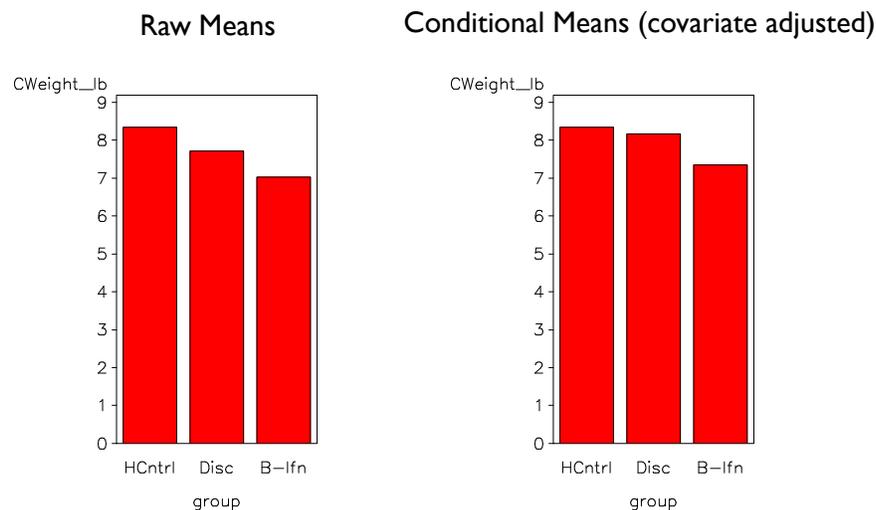
Solution for Fixed Effects										
		Standard								
Effect	group	Estimate	Error	DF	t Value	Pr > t	Lower	Upper		
γ_{00} Intercept		8.34	0.20	31.3	41.82	<.0001	7.93	8.74		
γ_{10} GestAge		0.27	0.07	30.7	3.93	0.0005	0.13	0.42		
γ_{20} MGain_lb		0.01	0.01	31.4	1.33	0.1932	-0.00	0.02		
γ_{01} group	B-Ifn	-0.99	0.34	30.6	-2.91	0.0066	-1.68	-0.30		
γ_{02} group	Disc	-0.17	0.35	20.4	-0.48	0.6369	-0.91	0.57		

- ▶ Differences from control smaller for both MS groups, particularly for discontinued group.
- ▶ Suggests lower mean birth weight observed for MS discontinue group is due to earlier delivery as opposed to abnormal fetal development.

CenterStat 2.72

In our software demonstration, we will show how these estimates can also be used to generate a contrast between the beta-interferon and discontinuation group (completing all pairwise comparisons).

Group Differences



CenterStat 2.73

Note that the difference between healthy controls and MS-discontinued is reduced when accounting for the covariates, whereas the difference between MS-discontinued and MS-beta interferon has increased.

Summary of Results

- ▶ “Correct” analysis reveals similar pattern of results as original analysis, but with greater richness.
 - ▶ High level of dependence within mothers, even after conditioning on predictors.
 - ▶ Both MS groups have infants that are underweight, but less so for mothers who discontinue beta-interferon.
 - ▶ For MS mothers who discontinue, lower birth weight largely accounted for by earlier gestational age and less maternal weight gain.
 - ▶ For MS mothers who take beta-interferon during pregnancy, birth weight remains lower than would be expected based on gestational age and maternal weight gain alone.

 CenterStat 2.74

Summary of Results

- ▶ Reviewers wrong that accounting for dependence would render all effects *ns*. Still, this was important to verify.
 - ▶ Partial exception: maternal weight gain was “marginally significant” in GLM, but clearly *ns* here.
 - ▶ In this analysis, this variable was only a covariate and not of key interest, but in other cases differences between GLM and multilevel analysis can be more impactful.

 CenterStat 2.75

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

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

