

# Chapter 1

## General Introduction

1.1	Nested Data Structures.....	1-3
1.2	Why Conventional Statistics are Inappropriate for Nested Data .....	1-8
1.3	Methods for Analyzing Nested Data .....	1-22



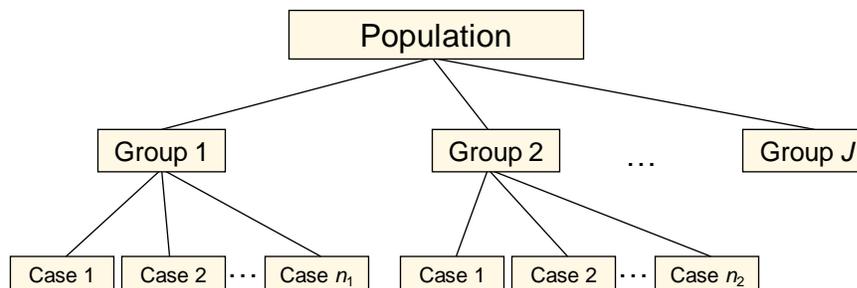
## 1.1 Nested Data Structures

### Objectives

- ▶ Discuss types of nested data that are common in psychological, social, and health research.
  - ▶ Hierarchically nested data (clustered observations)
  - ▶ Longitudinally nested data (repeated measures)
- ▶ Understand why nested data structures usually imply correlated observations, or dependence.

### Hierarchical Data Structures

- ▶ Hierarchical data structures are those in which multiple observations are taken from each of many clusters.
- ▶ A common hierarchical data structure is when individuals are sampled from naturally occurring groups.



The schematic shows the population to be composed of  $J$  groups, within which multiple individuals (cases) are sampled.

## Common Hierarchical Data Structures

---

- ▶ Students nested within schools
  - ▶ Friends nested within peer groups
  - ▶ Siblings nested within families
  - ▶ Spouses nested within dyads
  - ▶ Patients nested within treatment providers
- 

## Dependence in Hierarchical Data

---

- ▶ **Observations taken from the same unit tend to be correlated:**
    - ▶ Students attending the same school may have more similar academic outcomes than students attending different schools.
    - ▶ Internalizing behavior may be correlated across siblings within a family.
    - ▶ Adolescents within a given peer group may consume similar levels of alcohol.
    - ▶ Some interviewers may have better rapport with participants than others, eliciting fuller disclosure of sensitive and/or illegal behaviors
    - ▶ Treatment providers may differ in effectiveness, leading patients with different providers to show different outcomes.
- 

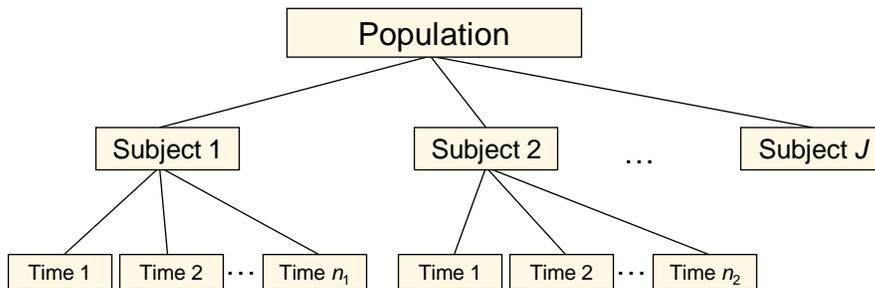
In some cases, hierarchical data arises as an explicit feature of the sampling design (e.g., cluster sampling). In other cases, however, the investigator may not have initially intended to collect clustered data (e.g., when participants are nested within interviewer or treatment provider).

## Dependence in Hierarchical Data

- ▶ Dependence (correlated observations) arises from:
  - ▶ Observations within units sharing common experiences/influences (e.g., quality of parenting affects all children in the family).
  - ▶ Unit formation affected by common characteristics (e.g., selection of school districts by parents may lead to demographic similarity of students within schools).
  
- ▶ Behavioral and social scientists know a great deal about the *selection* and *socialization* processes that produce correlated observations, but are often unsure how to analyze such data.

## Longitudinal Data Structures

- ▶ Longitudinal data structures arise when the same units are sampled repeatedly over time.
- ▶ A common longitudinal data structure is one in which repeated observations are made on the same individual over time.



The schematic shows the population to be composed of  $J$  individuals, for whom repeated measures are taken over time. The time intervals need not be equal, nor need they be the same across individuals, for multilevel models to be applied to the data.

## Common Longitudinal Data Structures

---

- ▶ Long-term longitudinal studies of development with annual or biennial assessments.
  - ▶ Short-term intensive longitudinal data, as gathered using experience sampling techniques, or in microgenetic studies.
  - ▶ Randomized clinical trials, with pretest(s), posttest(s) and possibly follow-up(s) assessments.
  - ▶ Within-subjects designs in which individuals sequentially participate in multiple conditions of an experiment.
- 

## Dependence in Longitudinal Data

---

- ▶ Repeated measures collected on the same individual tend to be correlated:
    - ▶ Some adolescents engage in consistently higher levels of alcohol use than others.
    - ▶ Blood pressure monitored over the course of the day differs between individuals with good versus poor cardiac health.
    - ▶ Toddlers differ in their rate of vocabulary acquisition (“early talkers” and “late talkers”).
    - ▶ Some patients with sickle cell disease consistently experience more pain than others.
- 

Differences in level and/or change *across* individuals imply correlated observations *within* individuals.

## Dependence in Longitudinal Data

---

- ▶ **Sources of dependence:**
    - ▶ Stability of personality or behavior over time (e.g., with trait-like characteristics) creates a pattern of constant correlation.
    - ▶ Individual differences in change over time creates more complex correlation patterns (such as decreasing correlation with increasing time lag).
  - ▶ There is greater recognition of dependence in this context, but continued use of statistical models that are often not optimal (e.g., repeated measures ANOVA or MANOVA).
- 

We will discuss several advantages of multilevel models relative to repeated measures ANOVA/MANOVA. There are both interpretational and statistical advantages, the latter including better treatment of missing data and data obtained using different assessment schedules across participants.

## Nested Data Structures

---

- ▶ We will focus on modeling both hierarchical and longitudinal data structures.
  - ▶ Both involve dependent observations so can be analyzed using similar modeling strategies.
    - ▶ Models for longitudinal data are often somewhat more complex, given the temporal ordering of observations.
  - ▶ Other types of data structures that involve dependence, but which will not be our principal focus:
    - ▶ Three-level data
    - ▶ Cross-classified data
    - ▶ Multiple membership data
    - ▶ Partially nested data
    - ▶ Spatial data
- 

We discuss extensions to some of these other data structures at the end of the course.

## 1.2 Why Conventional Statistics are Inappropriate for Nested Data

### Objectives

- ▶ Clarify why standard statistical techniques are inappropriate for nested data structures.
- ▶ Review the standard single-level general linear model.

Despite increasing recognition of problems associated with fitting standard single-level models to nested data, this practice continues to be widespread.

### The Challenge of Analyzing Nested Data

- ▶ The General Linear Model (GLM) assumes independence of observations after conditioning on predictors.

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} + r_i \quad r_i \stackrel{iid}{\sim} N(0, \sigma^2)$$

- ▶ Residuals assumed to be *independent* and *identically distributed*, following a *Normal* distribution with constant variance  $\sigma^2$ .
- ▶ Implies knowing one participant's score is not informative about any other participant's score.
- ▶ **With nested data, this assumption is almost always violated.**
  - ▶ Knowing the score of one participant in a group gives you some information about the scores of the others in that group.

## Within- and Between-Groups Variability

- ▶ Nested data are correlated because there is both within- and between-groups variability in our dependent variable.
- ▶ Suppose we collect alcohol use self-ratings on 5 individuals within each of 10 peer groups.

<i>group</i>	<i>person1</i>	<i>person2</i>	<i>person3</i>	<i>person4</i>	<i>person5</i>
1	6.18	6.43	5.25	6.42	4.98
2	5.66	5.80	5.19	5.47	5.80
3	3.78	4.16	2.20	3.86	4.20
4	1.24	2.63	3.50	3.18	2.01
5	5.60	3.63	3.68	4.80	5.41
6	7.29	5.76	6.36	4.78	6.35
7	6.62	4.52	5.62	5.31	4.66
8	4.01	3.72	4.41	3.19	4.86
9	6.12	5.21	6.49	6.36	5.62
10	5.70	7.11	4.98	5.42	6.45

Note that individuals within groups are ordered arbitrarily in these tables. That is, there is nothing to distinguish “person1” from “person2” other than that “person1” happened to be listed first in the data set.

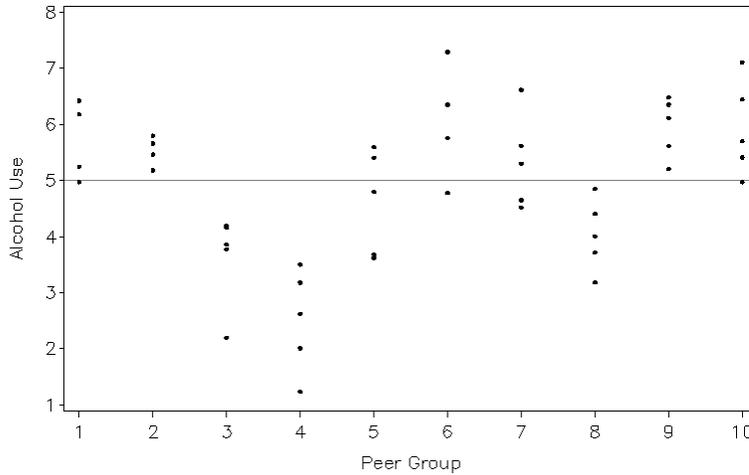
## Within- and Between-Groups Variability

- ▶ Grand mean is 4.96. Group means capture between-groups variability in alcohol use around the grand mean.
- ▶ Some groups use little alcohol, others a lot. Individuals within groups tend to have similar scores.

<i>group</i>	<i>person1</i>	<i>person2</i>	<i>person3</i>	<i>person4</i>	<i>person5</i>	<i>mean</i>
1	6.18	6.43	5.25	6.42	4.98	5.85
2	5.66	5.80	5.19	5.47	5.80	5.58
3	3.78	4.16	2.20	3.86	4.20	3.64
4	1.24	2.63	3.50	3.18	2.01	2.51
5	5.60	3.63	3.68	4.80	5.41	4.62
6	7.29	5.76	6.36	4.78	6.35	6.11
7	6.62	4.52	5.62	5.31	4.66	5.35
8	4.01	3.72	4.41	3.19	4.86	4.04
9	6.12	5.21	6.49	6.36	5.62	5.96
10	5.70	7.11	4.98	5.42	6.45	5.93

## Index Plot of Alcohol Use by Peer Group

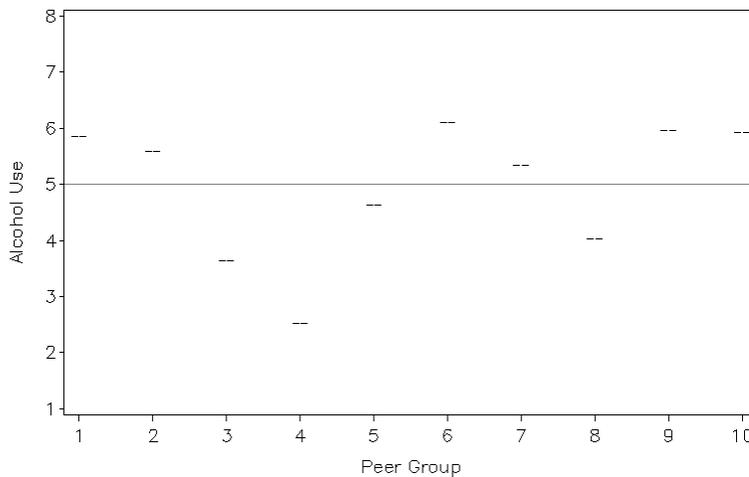
- ▶ Levels of individual use appear to vary by group.



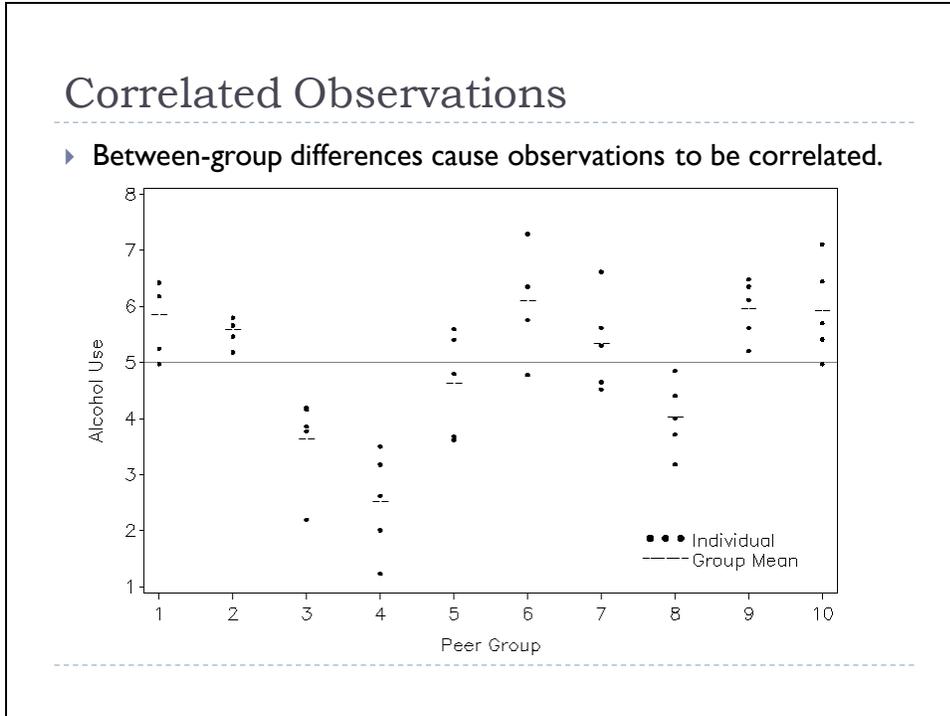
In an index plot, the x-axis indicates levels of an ID variable, here designating peer groups 1 through 10. The vertical bands of dots convey the distribution of alcohol use scores within each peer group, showing variability across individuals and across groups.

## Variability Between Groups

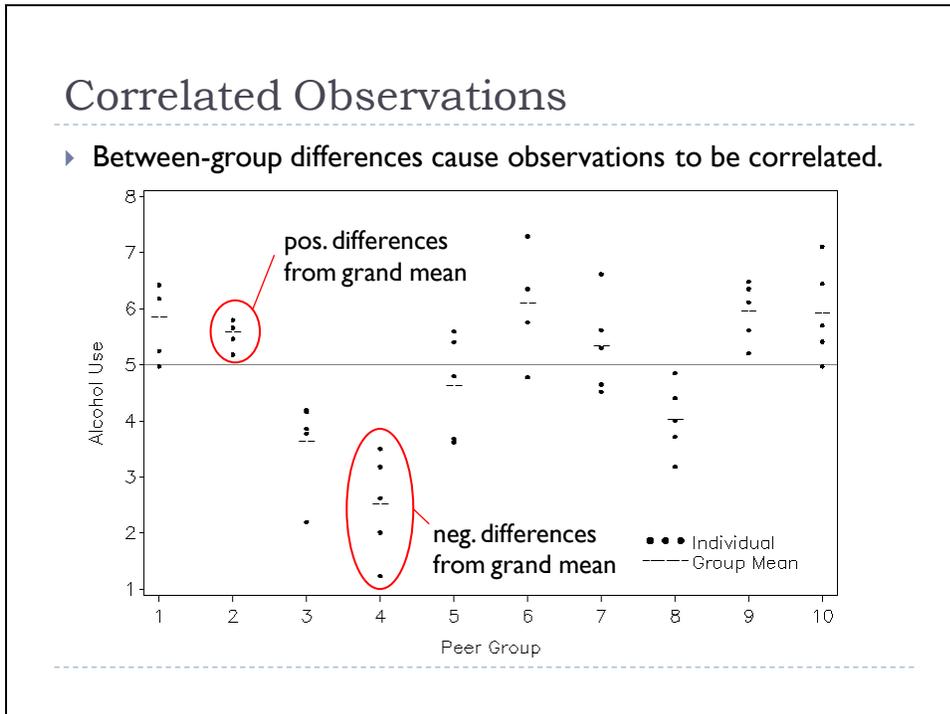
- ▶ Differences between groups captured by group means.



The between-group variability is summarized by plotting the group means.



Note the explicit depiction of both within- and between-groups variability in the plot.



Note that when the group mean is high, observations within the group tend to be above the grand mean (positive residuals), whereas when the group mean is low, observations within the group tend to be below the grand mean (negative residuals). The residuals are thus positively correlated within groups.

## Summarizing the Problem

---

- ▶ **GLM assumes independent observations**
    - ▶ Only source of variability (other than explained variability) is individual-level variability.
  - ▶ **For nested data structures there are usually two sources of variability**
    - ▶ Between-group differences
    - ▶ Individual differences within groups.
  - ▶ **What are the consequences of pooling between- and within-groups variability in a GLM?**
- 

## Example: Birth Order Effects on IQ

---

**The New York Times**  
nytimes.com

May 28, 1985

### **Spacing of Siblings Strongly Linked to Success in Life**

By DANIEL GOLEMAN

SCIENTIFIC interest in the effects of birth order on later development, an interest that has fallen in and out of vogue several times in the last few decades, seems to be coming back again with a rush.

Early findings showed that firstboms and only children had a higher proportion of successes in later life and that they tended to have higher I.Q.'s than later-born children have. Firstboms and only children were also shown to be overrepresented in such groups as those listed in "Who's Who," astronauts and students in graduate and professional schools. Critics have maintained that the birth-order effect has been exaggerated and is more ephemeral than its supporters suggest.

---

The popular appeal of birth order research is longstanding. Much of this research has been based on sampling a single individual per family (e.g., analyzing data from a single birth cohort).

## Example: Birth Order Effects on IQ

The New York Times  
nytimes.com

June 21, 2007

### Study Says Eldest Children Have Higher I.Q.s

By BENEDICT CAREY

The eldest children in families tend to develop higher I.Q.'s than their siblings, researchers are reporting today, in a large study that could settle more than a half-century of scientific debate about the relationship between I.Q. and birth order.

The average difference in I.Q. was slight — three points higher in the eldest child than in the closest sibling — but significant, the researchers said. And they said the results made it clear that it was due to family dynamics, not to biological factors like prenatal environment.

Three points on an I.Q. test may not sound like much. But experts say it can be a tipping point for some people — the difference between a high B average and a low A, for instance. That, in turn, can have a cumulative effect that could mean the difference between a admission to an elite private liberal-arts college and a less exclusive public one.

## Example: Birth Order Effects on IQ

- ▶ Theories of birth order effects all invoke within-family processes.
  - ▶ e.g., progressive division of parental attention and resources as family size grows.
- ▶ In data sets with one sibling per family, birth order effects may be contaminated by between-family differences.
  - ▶ e.g., birth order confounded with family disadvantage if later born children tend to be from larger, less advantaged families.
- ▶ To parse within- and between-group differences, we must collect and correctly analyze data consisting of multiple individuals per group (e.g., multiple siblings per family).

Since birth order effects are thought to reflect within-family processes, research on birth order effects should include multiple siblings per family so that these within-family processes can be examined directly. Otherwise, birth order differences may be confounded with between-family differences.

## Example: Birth Order Effects on IQ

- ▶ We will re-evaluate birth order effects using age-normed math IQ scores.
- ▶ 3313 kids from 2207 families taken from the NLSY child sample:
  - ▶ Young Cohort: Children who were 7-8 y.o. in 1986-1988
  - ▶ Old Cohort: Children who were 13-14 y.o. in 1986-1988.
- ▶ Siblings from the same family are in the sample:

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

The NLSY child sample consists of the children born to women who were part of the original NLSY sample. Some theories suggest that birth order effects vary with age, hence the inclusion of the two cohorts. Given the prospective design, it is also important to recognize that children within the older cohort were born to mothers who began childbearing at a younger age.

## Example: Birth Order Effects on IQ

- ▶ Given nested data, how to proceed?
- ▶ **Option 1: Take only one case per family.**
  - ▶ loss of information.
  - ▶ no ability to parse within-family from between-family differences.
- ▶ **Option 2: Ignore nesting of siblings within families.**
  - ▶ observations not independent, violates GLM assumption.
  - ▶ within- and between-family differences are pooled together.
- ▶ **Option 3: Use a statistical model that accounts for nested data.**
  - ▶ allows dependent observations.
  - ▶ potential to distinguish within- and between-family differences.
- ▶ Before proceeding to Option 3, let us consider Option 2.

## Example: Birth Order Effects on IQ

- ▶ Some initial summaries of the data:

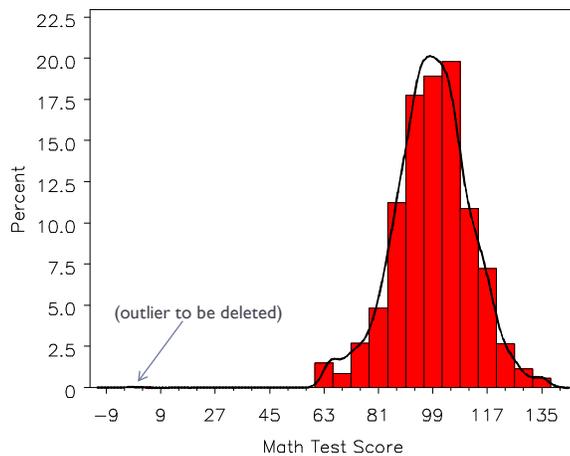
*Table of Birth Order by Cohort*

Birth Order	Cohort		
	Young	Old	Total
1	650	976	1626
2	559	507	1066
3	312	158	470
4	102	49	151
<i>Total</i>	1623	1690	3313

As one would expect, sample sizes diminish progressively for later birth orders.

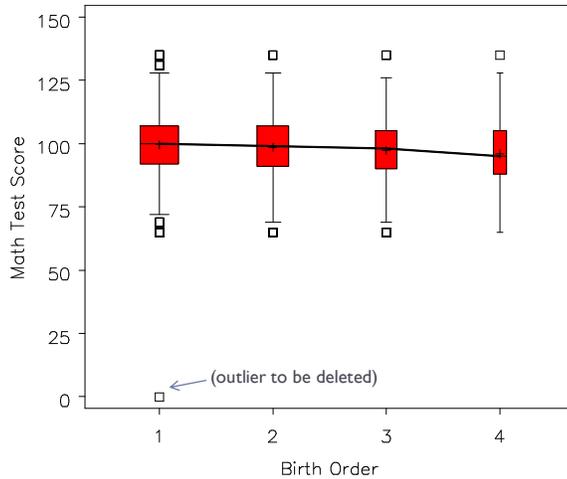
## Example: Birth Order Effects on IQ

- ▶ Some initial summaries of the data:



## Example: Birth Order Effects on IQ

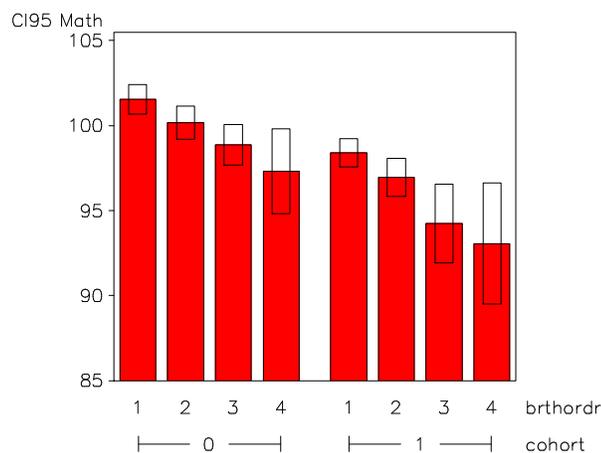
► Some initial summaries of the data:



The box widths in this side-by-side boxplot are proportional to sample size. Note that the outlier apparent here and in the preceding histogram was not flagged in the original analyses, illustrating why it is important to graph the data prior to conducting an analysis.

## Example: Birth Order Effects on IQ

► Some initial summaries of the data:



Confidence intervals are wider for later birth orders due to smaller sample sizes.

## Example: A Traditional Analysis

- ▶ We might fit a general linear model to the data (here of the form of an ANOVA, or regression with binary predictors).

$$\begin{aligned}
 \text{Math}_i &= \beta_0 + && \text{(intercept)} \\
 &\beta_1 \text{Second}_i + \beta_2 \text{Third}_i + \beta_3 \text{Fourth}_i + && \text{(main effect of birth order)} \\
 &\beta_4 \text{Old}_i + && \text{(main effect of cohort)} \\
 &\beta_5 \text{Second}_i \times \text{Old}_i + \beta_6 \text{Third}_i \times \text{Old}_i + \beta_7 \text{Fourth}_i \times \text{Old}_i + && \text{(birth order x cohort)} \\
 &r_i && \text{(person residual)} \\
 \\
 &r_i \overset{iid}{\sim} N(0, \sigma^2)
 \end{aligned}$$

Recall that any  $G$ -level categorical predictor can be summarized in a regression model by  $G-1$  coding variables. A variety of coding options can be entertained, e.g., dummy variables, effect code variables, or contrast variables. In the present case, Second, Third, and Fourth are binary dummy variables scored one if the child was second-, third-, or fourth-born, respectively, and zero otherwise. The absence of a dummy variable for First indicates that first-borns are the reference category. Similarly, Old is a binary dummy variable indicating whether the child is in the older cohort (scored one) or younger cohort (scored zero). The younger cohort is thus the reference category.

## Example: A Traditional Analysis

$$Math_i = \beta_0 + \beta_1 Second_i + \beta_2 Third_i + \beta_3 Fourth_i + \beta_4 Old_i + \beta_5 Second_i \times Old_i + \beta_6 Third_i \times Old_i + \beta_7 Fourth_i \times Old_i + r_i$$

- ▶ Expected value for math test:

Cohort	Birth Order			
	First	Second	Third	Fourth
Young	$\beta_0$	$\beta_0 + \beta_1$	$\beta_0 + \beta_2$	$\beta_0 + \beta_3$
Old	$\beta_0 + \beta_4$	$\beta_0 + \beta_1 + \beta_4 + \beta_5$	$\beta_0 + \beta_2 + \beta_4 + \beta_6$	$\beta_0 + \beta_3 + \beta_4 + \beta_7$

The “expected value” is the mean within the population as a whole (or one can think of it as the mean of an infinitely large sample). It may be conditional on predictors (e.g., the conditional mean for math IQ given birth order and cohort). The expected value of variable  $Y$  is often designated  $E(Y)$ . A conditional expected value may be designated  $E(Y|X=x)$ , indicating the expected value of  $Y$  when predictor  $X$  takes on the value  $x$ , or more simply as  $E(Y|X)$ .

## Example: A Traditional Analysis

$$Math_i = \beta_0 + \beta_1 Second_i + \beta_2 Third_i + \beta_3 Fourth_i + \beta_4 Old_i + \beta_5 Second_i \times Old_i + \beta_6 Third_i \times Old_i + \beta_7 Fourth_i \times Old_i + r_i$$

- ▶ Expected value for math test:

Cohort	Birth Order			
	First	Second	Third	Fourth
Young	$\beta_0$	$\beta_0 + \beta_1$	$\beta_0 + \beta_2$	$\beta_0 + \beta_3$
Old	$\beta_0 + \beta_4$	$\beta_0 + \beta_1 + \beta_4 + \beta_5$	$\beta_0 + \beta_2 + \beta_4 + \beta_6$	$\beta_0 + \beta_3 + \beta_4 + \beta_7$

- ▶ Intercept is expected value for first born in young cohort:
  - ▶ A point of contrast for effects of interest.

### Example: A Traditional Analysis

$$Math_i = \beta_0 + \beta_1 Second_i + \beta_2 Third_i + \beta_3 Fourth_i + \beta_4 Old_i + \beta_5 Second_i \times Old_i + \beta_6 Third_i \times Old_i + \beta_7 Fourth_i \times Old_i + r_i$$

- ▶ Expected value for math test:

		Birth Order			
Cohort	First	Second	Third	Fourth	
Young	$\beta_0$	$\beta_0 + \beta_1$	$\beta_0 + \beta_2$	$\beta_0 + \beta_3$	
Old	$\beta_0 + \beta_4$	$\beta_0 + \beta_1 + \beta_4 + \beta_5$	$\beta_0 + \beta_2 + \beta_4 + \beta_6$	$\beta_0 + \beta_3 + \beta_4 + \beta_7$	

- ▶ Main effect of Birth Order:
  - ▶ Are expected scores of first, second, third & fourth born kids different?

### Example: A Traditional Analysis

$$Math_i = \beta_0 + \beta_1 Second_i + \beta_2 Third_i + \beta_3 Fourth_i + \beta_4 Old_i + \beta_5 Second_i \times Old_i + \beta_6 Third_i \times Old_i + \beta_7 Fourth_i \times Old_i + r_i$$

- ▶ Expected value for math test:

		Birth Order			
Cohort	First	Second	Third	Fourth	
Young	$\beta_0$	$\beta_0 + \beta_1$	$\beta_0 + \beta_2$	$\beta_0 + \beta_3$	
Old	$\beta_0 + \beta_4$	$\beta_0 + \beta_1 + \beta_4 + \beta_5$	$\beta_0 + \beta_2 + \beta_4 + \beta_6$	$\beta_0 + \beta_3 + \beta_4 + \beta_7$	

- ▶ Main effect of Cohort:
  - ▶ Does the older cohort score better or worse than the younger cohort?

## Example: A Traditional Analysis

$$Math_i = \beta_0 + \beta_1 Second_i + \beta_2 Third_i + \beta_3 Fourth_i + \beta_4 Old_i + \beta_5 Second_i \times Old_i + \beta_6 Third_i \times Old_i + \beta_7 Fourth_i \times Old_i + r_i$$

- ▶ Expected value for math test:

Cohort	Birth Order			
	First	Second	Third	Fourth
Young	$\beta_0$	$\beta_0 + \beta_1$	$\beta_0 + \beta_2$	$\beta_0 + \beta_3$
Old	$\beta_0 + \beta_4$	$\beta_0 + \beta_1 + \beta_4 + \beta_5$	$\beta_0 + \beta_2 + \beta_4 + \beta_6$	$\beta_0 + \beta_3 + \beta_4 + \beta_7$

- ▶ Birth Order x Cohort interaction:
  - ▶ Is the birth order effect different in the older cohort?

## Example: A Traditional Analysis

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	12950.77	1850.11	12.20	<.0001
Error	3304	500967.82	151.62		
Total	3311	513918.59			

R-Square	Coefficient Var	Root MSE	math Mean
0.025	12.46	12.31	98.83

Source	DF	Type III SS	Mean Square	F Value	Pr > F
brthdr	3	2562.27	854.09	5.63	0.0008
cohort	1	5039.26	5039.26	33.24	<.0001
cohort*brthdr	3	216.83	72.28	0.48	0.6985

## Example: A Traditional Analysis

Parameter		Estimate	Standard		Pr >  t
			Error	t Value	
Intercept		101.54	0.48	210.24	<.0001
brthordr	2	-1.36	0.71	-1.92	0.0548
brthordr	3	-2.67	0.85	-3.15	0.0017
brthordr	4	-4.24	1.31	-3.23	0.0012
cohort		-3.14	0.62	-5.04	<.0001
cohort*brthordr	2	-0.08	0.98	-0.08	0.9366
cohort*brthordr	3	-1.49	1.35	-1.10	0.2718
cohort*brthordr	4	-1.10	2.23	-0.49	0.6217

## Problems with the Traditional Approach

- ▶ When assuming independence of nested data there is the potential for making incorrect inferences about misleading effect estimates.
  - ▶ MSE, standard errors, degrees of freedom, F-tests, t-tests all computed under assumption of independence – all thus incorrect.
  - ▶ Model presumes more information than is present:
    - ▶  $N-p-1$  degrees of freedom computed assuming conditionally independent observations (*iid* residuals)
    - ▶ With nested data we know observations are partially redundant (correlated residuals).
    - ▶ No parsing of within- and between-group variability in dependent variable, so effect estimates can represent within-group processes, between-group processes, or both.

In the birth order analysis, there were 3312 cases included in the analysis (excluding the outlier). The *effective sample size* is, however, really lower since siblings from the same family do not contribute independent information to the analysis. The degrees of freedom for the GLM are thus incorrect.

## 1.3 Methods for Analyzing Nested Data

### Objectives

---

- ▶ Review three approaches to modeling nested data structures:
    - ▶ adjust results of standard analysis (ignoring clustering).
    - ▶ the fixed-effects approach.
    - ▶ the random-effects approach (multilevel modeling).
- 

In particular, there is a great deal of contention about whether the fixed-effects approach or random-effects approach is more appropriate.

### Adjusting the Results

---

- ▶ One approach for dealing with dependence is to retain the general linear model but correct the standard errors and test statistics.
- ▶ For instance, suppose we have data on individuals nested within clusters, but we fit a standard regression model

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 w_i + r_i$$

where  $x$  is an individual-level predictor and  $w$  is a cluster-level predictor.

- ▶ e.g., predicting an individual's attachment to school ( $y$ ) as a function of engagement in extracurricular activities ( $x$ ) and number of teachers within the school who have graduate degrees ( $w$ ).
-

## Adjusting the Results

- ▶ To show what happens in this case, we've generated 1000 samples of artificial data with a modest level of dependence.
  - ▶ Each sample has 100 clusters of individuals
  - ▶ Clusters have from 1 to 20 individual members
  - ▶ Data generated on the outcome,  $y$ , an individual-level predictor,  $x$ , and a cluster-level predictor,  $w$ .
- ▶ We fit a typical regression model to the data, ignoring clustering.
- ▶ For comparison, both “naïve” standard errors and Huber-White corrected standard errors are computed.
- ▶ We can compare the mean standard error to the true standard error (standard deviation of estimates across samples) to determine whether our standard errors are “honest”.

Simulation studies are useful for examining the quality of model estimates, standard errors and inference tests under specific, known conditions.

## Adjusting the Results

- ▶ “Naïve” SEs for  $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$  are too small because the MSE fails to differentiate within- and between-cluster variance in  $y$ .
- ▶ SEs computed using a modified Huber-White (HW) sandwich estimator for cluster-correlated data corrects the problem.

Parameter	Value	SD	Naïve SEs		HW SEs	
			SE	SE/SD	SE	SE/SD
$\beta_0$	1.00	0.280	0.179	0.64	0.264	0.94
$\beta_1$	0.50	0.028	0.024	0.85	0.028	0.99
$\beta_2$	1.00	0.097	0.054	0.56	0.091	0.94

Note. SD=true standard deviation of sampling distribution; SE=mean estimated standard error

If the estimated standard errors accurately reflect the sampling variability in the data the ratio of SE/SD should be approximately 1. “Honest” standard error estimates will average out to equal the true SD.

## Advantages and Disadvantages

---

- ▶ This approach is useful if you are interested in overall effects (pooling individuals over groups) and you want “honest” SEs and test statistics.
  - ▶ Cannot be used to evaluate within- versus between-group effects
  - ▶ Cannot be used to evaluate whether effects vary across groups.
- 

Correcting standard errors can be useful, but the estimates of the model still pool over individuals from all groups and hence do not provide information on within-group versus between-group differences.

## The Fixed-Effects Approach

- ▶ Another approach is to account for dependence by adding the grouping variable as a fixed factor in the model.
- ▶ For example, suppose we want to evaluate the effect of peer alcohol use on academic failure and we have collected data on individuals from 20 schools.
- ▶ A possible model for this data would be:

$$\text{Grades}_i = \beta_1 \text{School}_1_i + \beta_2 \text{School}_2_i + \cdots + \beta_{20} \text{School}_{20}_i + \beta_{21} \text{PeerUse}_i + r_i$$

- ▶ The dummy variables effectively give each school its own intercept, accounting for differences across schools in levels of achievement. This removes the source of dependence.

The overall intercept of the model has been omitted here, permitting dummy variables for all 20 levels of the School variable to be included in the model. In this specification of the model there is no reference group, and the coefficients  $\beta_1$  through  $\beta_{20}$  can be interpreted as school-specific intercepts. The school-specific intercepts serve to absorb any mean differences in alcohol use across schools that remain after controlling for peer use.

## Advantages of the Fixed-Effects Approach

---

- ▶ It is the most appropriate method if the groups are truly fixed and not representative of a population of groups.
    - ▶ e.g., treatment groups, ethnic groups, religions, etc.
  - ▶ Useful when there are a small number of level-2 units.
    - ▶ e.g., data collected from only 3 schools.
  - ▶ Adjusts model for all unmeasured level-2 covariates.
    - ▶ e.g., estimating and removing main effects of group controls for all between-group differences.
    - ▶ Estimates of individual-level predictors unambiguously reflect within-group effects, as all between-group variance has been removed.
  - ▶ Allows for estimation within standard GLM.
- 

For fixed groups, one does not seek to make inferences to any groups outside of those included in the sample.

## Disadvantages of the Fixed-Effects Approach

---

- ▶ Difficult to evaluate effects of group-level variables (can be done with contrasts, but becomes cumbersome).
  - ▶ Many nuisance parameters to estimate (not parsimonious).
    - ▶ Problem with small groups (e.g., dyadic data)
    - ▶ Problem with “singletons” (e.g., “only children” in family data)
  - ▶ Grouping variable assumed to be a fixed predictor, so inferences must be restricted to these particular groups (not a broader population of groups).
-

## Random Effects Approach

---

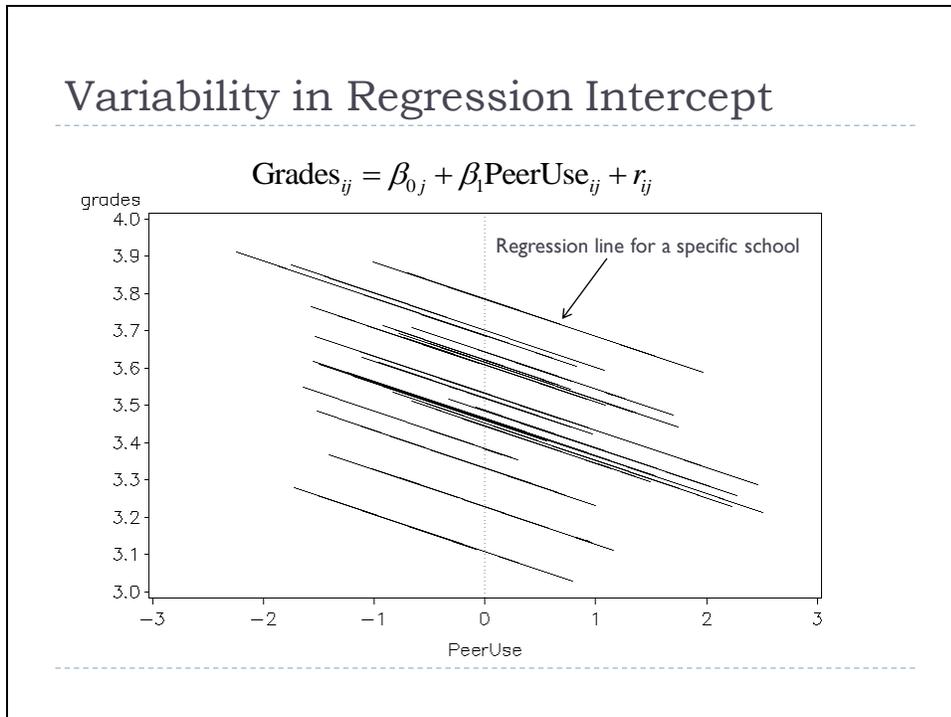
- ▶ Group differences are assumed to follow a particular distribution in the population. We sample from this distribution when we sample groups.
- ▶ Our concern is with the parameters of the population distribution, not with differences among the sampled groups.
- ▶ For the previous example, we write the regression model as

$$\text{Grades}_{ij} = \beta_{0j} + \beta_1 \text{PeerUse}_{ij} + r_{ij}$$

where we will assume a particular distribution for  $\beta_{0j}$  just as we customarily do for  $r_{ij}$ .

- ▶ Because  $\beta_{0j}$  is randomly sampled from a distribution when we randomly sample groups, we refer to it as a random effect.
- 

The notation we have used here differs slightly from what will be introduced in later chapters in order to facilitate comparison to the fixed-effect approach.



Like the fixed effects approach, allowing each school to have a different intercept absorbs level differences in alcohol use across schools (holding peer use constant). The difference is that these regression lines are viewed as a sample of all possible regression lines that might be observed from the population of schools and it is the population of schools that is that target for inference. In the fixed effect approach, inference is restricted to the 20 regression lines obtained for the schools in the sample.

### Advantages of Random-Effects Approach

---

- ▶ Parsimonious.
  - ▶ Permits inferences to the population of groups.
  - ▶ Conforms to the sampling design, with the random selection of groups followed by the random selection of individuals within groups.
  - ▶ Enables us to examine the effects of individual-level and group-level influences simultaneously.
- 

### Disadvantages of the Random-Effects Approach

---

- ▶ Requires a sufficient number of groups to enable inferences to population.
  - ▶ Requires that we assume a specific distribution for group effects (e.g., group differences are normally distributed).
  - ▶ Omitted group-level variables can be problematic (e.g., when correlated with individual-level variables that are included in the model).
  - ▶ Requires more complex methods of estimation (i.e., maximum likelihood).
- 

Advocates of the fixed-effects approach sometimes argue that effects for individual-level predictors are confounded in the random-effects approach. This assertion reflects a misunderstanding of how effects can be decomposed in the random-effects approach. We show this effect decomposition in Chapter 3.

## Why We Prefer the Random-Effects Approach

---

- ▶ Do not want our inferences to be limited to groups in the sample by using a model that fails to conform to sampling design.
  - ▶ Can check assumptions for the distributions of the random effects.
  - ▶ Can mitigate omitted variable bias problem through careful model specification and centering.
  - ▶ Simplifies to GLM if no evidence of dependence due to nesting.
  
  - ▶ *On balance, we feel that the advantages of the random-effects approach are substantial and the disadvantages can be minimized.*
- 

## Chapter Summary

---

- ▶ Two common types of nested data are hierarchical and longitudinal.
  - ▶ Traditional GLM/ANOVA/regression should not be applied to such data, given violation of independence assumption.
  - ▶ Two better approaches to analyzing nested data are the fixed-effects approach and the random-effects approach. Each has advantages and disadvantages.
-