

*Paul Roback and Julie Legler*

---

# ***GLMs and Multilevel Models***

***Broadening Your Statistical Horizons with Applications  
using R***



---

# *Contents*

---

<b>Preface</b>	<b>1</b>
<b>1 Two Level Longitudinal Data</b>	<b>3</b>
1.1 Learning objectives . . . . .	3
1.2 Case study: Charter schools . . . . .	4
1.3 Initial Exploratory Analyses . . . . .	5
1.3.1 Data organization . . . . .	5
1.3.2 Missing data . . . . .	6
1.3.3 Exploratory analyses for general multilevel models . .	8
1.3.4 Exploratory analyses for longitudinal data . . . . .	9
1.4 Preliminary two-stage modeling . . . . .	14
1.4.1 Linear trends within schools . . . . .	14
1.4.2 Effects of level two covariates on linear time trends . .	16
1.4.3 Error structure within schools . . . . .	19
1.5 Initial models . . . . .	20
1.5.1 Unconditional means model . . . . .	21
1.5.2 Unconditional growth model . . . . .	22
1.5.3 Modeling other trends over time . . . . .	24
1.6 Building to a final model . . . . .	27
1.6.1 Uncontrolled effects of school type . . . . .	27
1.6.2 Add percent free and reduced lunch as a covariate . .	30
1.6.3 A potential final model with three Level Two covariates	32
1.7 Covariance structure among observations . . . . .	35
1.7.1 Standard covariance structure . . . . .	35
1.7.2 Alternative covariance structures . . . . .	38

1.7.3	Covariance structure in non-longitudinal multilevel models . . . . .	39
1.7.4	Final thoughts regarding covariance structures . . . .	40
1.7.5	Details of covariance structures (Optional) . . . . .	40
1.8	Notes on Using R (Optional) . . . . .	41
1.9	Exercises . . . . .	43
1.9.1	Conceptual Exercises . . . . .	43
1.9.2	Guided Exercise . . . . .	46
1.9.3	Open-ended Exercises . . . . .	47

---

## *Preface*

---



**GLMs and Multilevel Models: Broadening Your Statistical Horizons (BYSH) with Applications using R** is intended to be accessible to undergraduate students who have successfully completed a regression course through, for example, a textbook like *Stat2* [Cannon et al., 2019]. We started teaching this course at St. Olaf in 2003 so students would be able to deal with the non-normal, correlated world we live in. It has been offered at St. Olaf every year since. Even though there is no mathematical prerequisite, we still introduce fairly sophisticated topics such as likelihood theory, zero-inflated Poisson, and parametric bootstrapping in an intuitive and applied manner. We believe strongly in case studies featuring real data and real research questions; thus, most of the data in the textbook and available at our GitHub repo<sup>1</sup> arises from collaborative research conducted by the authors and their students, or from student projects. Our goal is that, after working through this material, students will not necessarily be expert in these methods and associated theory, but that they will develop an expanded toolkit and a greater appreciation for the wider world of data and statistical modeling.

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

**Acknowledgements.** We would like to thank students of Stat 316 at St. Olaf College since 2010 for their patience as this book has taken shape with their feedback. We would especially like to thank these St. Olaf students for their summer research efforts which significantly improved aspects of this book: Ce-

---

<sup>1</sup><https://github.com/proback/BYSH>

cilia Noecker, Anna Johanson, Nicole Bettes, Kiegan Rice, Anna Wall, Jack Wolf, Josh Pelayo, Spencer Eanes, and Emily Patterson. Early editions of this book also benefitted greatly from feedback from instructors who used these materials in their classes, including Matt Beckman, Laura Boehm Vock, Beth Chance, Laura Chihara, Mine Dogucu, and Katie Ziegler-Graham. Finally, we have appreciated the support of two NSF grants (#DMS-1045015 and #DMS-0354308) and of our colleagues in Mathematics, Statistics, and Computer Science at St. Olaf.

# 1

---

## *Two Level Longitudinal Data*

---

### 1.1 Learning objectives

After finishing this chapter, you should be able to:

- Recognize longitudinal data as a special case of multilevel data, with time at Level One.
- Consider patterns of missingness and implications of that missing data on multilevel analyses.
- Apply exploratory data analysis techniques specific to longitudinal data.
- Build and understand a taxonomy of models for longitudinal data.
- Interpret model parameters in multilevel models with time at Level One.
- Compare models, both nested and not, with appropriate statistical tests and summary statistics.
- Consider different ways of modeling the variance-covariance structure in longitudinal data.

```
# Packages required for Chapter 9
library(GGally)
library(data.table)
library(Hmisc)
library(mice)
library(lattice)
library(nlme)
library(reshape2)
library(MASS)
library(mnormt)
library(lme4)
library(gridExtra)
library(knitr)
library(kableExtra)
library(tidyverse)
```

---

## 1.2 Case study: Charter schools

Charter schools were first introduced in the state of Minnesota in 1991 [U.S. Department of Education]. Since then, charter schools have begun appearing all over the United States. While publicly funded, a unique feature of charter schools is their independence from many of the regulations that are present in the public school systems of their respective city or state. Thus, charters will often extend the school days or year and tend to offer non-traditional techniques and styles of instruction and learning.

One example of this unique schedule structure is the KIPP (Knowledge is Power Program) Stand Academy in Minneapolis, MN. KIPP stresses longer days and better partnerships with parents, and they claim that 80% of their students go to college from a population where 87% qualify for free and reduced lunch and 95% are African-American or Latino [KIPP, 2017]. However, the larger question is whether or not charter schools are out-performing non-charter public schools in general. Because of the relative youthfulness of charter schools, data has just begun to be collected to evaluate the performance of charter versus non-charter schools and some of the factors that influence a school's performance. Along these lines, we will examine data collected by the Minnesota Department of Education for all Minnesota schools during the years 2008-2010.

Comparisons of student performance in charter schools versus public schools have produced conflicting results, potentially as a result of the strong differences in the structure and population of the student bodies that represent the two types of schools. A study by the Policy and Program Studies Service of five states found that charter schools are less likely to meet state performance standards than conventional public schools [Finnigan et al., 2004]. However, Witte et al. [2007] performed a statistical analysis comparing Wisconsin charter and non-charter schools and found that average achievement test scores were significantly higher in charter schools compared to non-charter schools, after controlling for demographic variables such as the percentage of white students. In addition, a study of California students who took the Stanford 9 exam from 1998 through 2002 found that charter schools, on average, were performing at the same level as conventional public schools [Buddin and Zimmer, 2005]. Although school performance is difficult to quantify with a single measure, for illustration purposes in this chapter we will focus on that aspect of school performance measured by the math portion of the Minnesota Comprehensive Assessment (MCA-II) data for 6th grade students enrolled in 618 different Minnesota schools during the years 2008, 2009, and 2010 [Minnesota Department of Education]. Similar comparisons could obviously be conducted for other grade levels or modes of assessment.



As described in [GreenIII et al. \[2003\]](#), it is very challenging to compare charter and public non-charter schools, as charter schools are often designed to target or attract specific populations of students. Without accounting for differences in student populations, comparisons lose meaning. With the assistance of multiple school-specific predictors, we will attempt to model sixth grade math MCA-II scores of Minnesota schools, focusing on the differences between charter and public non-charter school performances. In the process, we hope to answer the following research questions:

- Which factors most influence a school's performance in MCA testing?
- How do the average math MCA-II scores for 6th graders enrolled in charter schools differ from scores for students who attend non-charter public schools? Do these differences persist after accounting for differences in student populations?
- Are there differences in yearly improvement between charter and non-charter public schools?

---

## 1.3 Initial Exploratory Analyses

### 1.3.1 Data organization

Key variables in `chart_wide_condense.csv` which we will examine to address the research questions above are:

- `schoolid` = includes district type, district number, and school number
- `schoolName` = name of school
- `urban` = is the school in an urban (1) or rural (0) location?
- `charter` = is the school a charter school (1) or a non-charter public school (0)?
- `schPctnonw` = proportion of non-white students in a school (based on 2010 figures)
- `schPctsped` = proportion of special education students in a school (based on 2010 figures)
- `schPctfree` = proportion of students who receive free or reduced lunches in a school (based on 2010 figures). This serves as a measure of poverty among school families.
- `MathAvgScore.0` = average MCA-II math score for all sixth grade students in a school in 2008
- `MathAvgScore.1` = average MCA-II math score for all sixth grade students in a school in 2009

**TABLE 1.1:** The first six observations in the wide data set for the Charter Schools case study.

schoolid	schoolName	urban	charter	schPetnonw	schPetsped	schPetfree	MathAvgScore.0	MathAvgScore.1	MathAvgScore.2
Dtype 1 Dnum 1 Snum 2	RIPPLESIDE ELEMENTARY	0	0	0.0000	0.1176	0.3627	652.8	656.6	652.6
Dtype 1 Dnum 100 Snum 1	WRENSHALL ELEMENTARY	0	0	0.0303	0.1515	0.4242	646.9	645.3	651.9
Dtype 1 Dnum 108 Snum 30	CENTRAL MIDDLE	0	0	0.0769	0.1231	0.2615	654.7	658.5	659.7
Dtype 1 Dnum 11 Snum 121	SANDBURG MIDDLE	1	0	0.0977	0.0827	0.2481	656.4	656.8	659.9
Dtype 1 Dnum 11 Snum 193	OAK VIEW MIDDLE	1	0	0.0538	0.0954	0.1418	657.7	658.2	659.8
Dtype 1 Dnum 11 Snum 195	ROOSEVELT MIDDLE	1	0	0.1234	0.0886	0.2405	655.9	659.1	660.3

- **MathAvgScore.2** = average MCA-II math score for all sixth grade students in a school in 2010

This data is stored in WIDE format, with one row per school, as illustrated in Table ??.

For most statistical analyses, it will be advantageous to convert WIDE format to LONG format, with one row per year per school. To make this conversion, we will have to create a time variable, which under the LONG format is very flexible—each school can have a different number of and differently-spaced time points, and they can even have predictors which vary over time. Details for making this conversion in R can be found in Section 1.8, and the form of the LONG data in this study is exhibited in the next section.

### 1.3.2 Missing data

In this case, before we convert our data to LONG form, we should first address problems with missing data. Missing data is a common phenomenon in longitudinal studies. For instance, it could arise if a new school was started during the observation period, a school was shut down during the observation period, or no results were reported in a given year. Dealing with missing data in a statistical analysis is not trivial, but fortunately many multilevel packages (including the lme4 package in R) are adept at handling missing data.

First, we must understand the extent and nature of missing data in our study. Table 1.2, where 1 indicates presence of a variable and 0 indicates a missing value for a particular variable, is a helpful starting point. Among our 618 schools, 540 had complete data (all covariates and math scores for all three years), 25 were missing a math score for 2008, 35 were missing math scores in both 2008 and 2009, etc.

A frequency table of missing data patterns. The number of schools with a particular missing data pattern are listed in the left column; the remaining columns of 0's and 1's describe the missing data pattern, with 0 indicating a missing value. Some covariates that are present for every school are not listed. The bottom row gives the number of schools with missing values for specific variables; the last entry indicates that 121 total observations were missing.

**TABLE 1.2:** A frequency table of missing data patterns. The number of schools with a particular missing data pattern are listed in the left column; the remaining columns of 0's and 1's describe the missing data pattern, with 0 indicating a missing value. Some covariates that are present for every school are not listed. The bottom row gives the number of schools with missing values for specific variables; the last entry indicates that 121 total observations were missing.

	charter	MathAvgScore.2	MathAvgScore.1	MathAvgScore.0	
540	1	1	1	1	0
25	1	1	1	0	1
4	1	1	0	1	1
35	1	1	0	0	2
6	1	0	1	1	1
1	1	0	1	0	2
7	1	0	0	1	2
	0	14	46	61	121

Statisticians have devised different strategies for handling missing data; a few common approaches are described briefly here:

- Include only schools with complete data. This is the cleanest approach analytically; however, ignoring data from 12.6% of the study's schools (since 78 of the 618 schools had incomplete data) means that a large amount of potentially useful data is being thrown away. In addition, this approach creates potential issues with informative missingness. Informative missingness occurs when a school's lack of scores is not a random phenomenon but provides information about the effectiveness of the school type (e.g., a school closes because of low test scores).
- Last observation carried forward. Each school's last math score is analyzed as a univariate response, whether the last measurement was taken in 2008, 2009, or 2010. With this approach, data from all schools can be used, and analyses can be conducted with traditional methods assuming independent responses. This approach is sometimes used in clinical trials because it tends to be conservative, setting a higher bar for showing that a new therapy is significantly better than a traditional therapy. Of course, we must assume that a school's 2008 score is representative of their 2010 score. In addition, information about trajectories over time is thrown away.
- Imputation of missing observations. Many methods have been developed for sensibly "filling in" missing observations, using imputation models which base imputed data on subjects with similar covariate profiles and on typical observed time trends. Once an imputed data set is created (or several imputed data sets), analyses can proceed with complete data

**TABLE 1.3:** The first six observations in the long data set for the Charter Schools case study; these lines correspond to the first two observations from the wide data set illustrated in Table ??.

schoolName	charter	schPctsped	schPctfree	year08	MathAvgScore
RIPPLESIDE ELEMENTARY	0	0.1176	0.3627	0	652.8
RIPPLESIDE ELEMENTARY	0	0.1176	0.3627	1	656.6
RIPPLESIDE ELEMENTARY	0	0.1176	0.3627	2	652.6
WRENSHALL ELEMENTARY	0	0.1515	0.4242	0	646.9
WRENSHALL ELEMENTARY	0	0.1515	0.4242	1	645.3
WRENSHALL ELEMENTARY	0	0.1515	0.4242	2	651.9

methods that are easier to apply. Risks with the imputation approach include misrepresenting missing observations and overstating precision in final results.

- Apply multilevel methods, which use available data to estimate patterns over time by school and then combine those school estimates in a way that recognizes that time trends for schools with complete data are more precise than time trends for schools with fewer measurements. [Laird \[1988\]](#) demonstrates that multilevel models are valid under the fairly unrestrictive condition that the probability of missingness cannot depend on any unobserved predictors or the response. This is the approach we will follow in the remainder of the text.

Now, we are ready to create our LONG data set. Fortunately, many packages (including R) have built-in functions for easing this conversion, and the functions are improving constantly. The resulting LONG data set is shown in Table 1.3, where `year08` measures the number of years since 2008:

### 1.3.3 Exploratory analyses for general multilevel models

Notice the **longitudinal** structure of our data—we have up to three measurements of test scores at different time points for each of our 618 schools. With this structure, we can address questions at two levels:

- Within school—changes over time
- Between schools—effects of school-specific covariates (charter or non-charter, urban or rural, percent free and reduced lunch, percent special education, and percent non-white) on 2008 math scores and rate of change between 2008 and 2010.

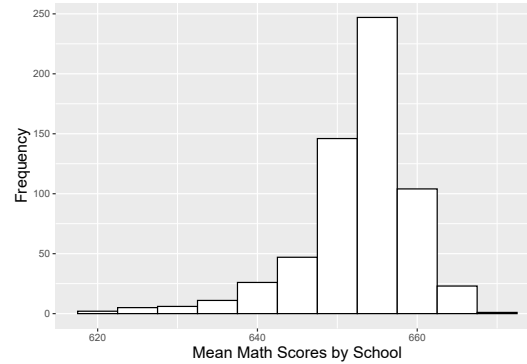
As with any statistical analysis, it is vitally important to begin with graphical and numerical summaries of important variables and relationships between

variables. We'll begin with initial exploratory analyses that we introduced in the previous chapter, noting that we have no Level One covariates other than time at this point (potential covariates at this level may have included measures of the number of students tested or funds available per student). We will, however, consider the Level Two variables of charter or non-charter, urban or rural, percent free and reduced lunch, percent special education, and percent non-white. Although covariates such as percent free and reduced lunch may vary slightly from year to year within a school, the larger and more important differences tend to occur between schools, so we used percent free and reduced lunch for a school in 2010 as a Level Two variable.

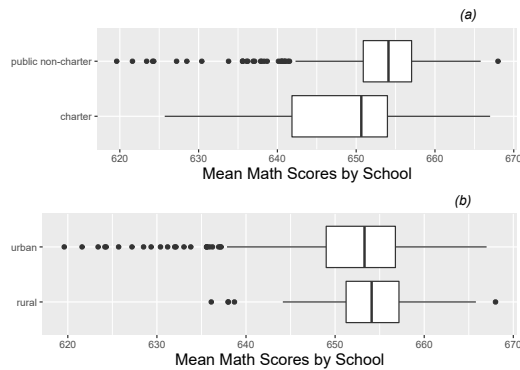
As in Chapter ??, we can conduct initial investigations of relationships between Level Two covariates and test scores in two ways. First, we can use all 1733 observations to investigate relationships of Level Two covariates with test scores. Although these plots will contain dependent points, since each school is represented by up to three years of test score data, general patterns exhibited in these plots tend to be real. Second, we can calculate mean scores across all years for each of the 618 schools. While we lose some information with this approach, we can more easily consider each plotted point to be independent. Typically, both types of exploratory plots illustrate similar relationships, and in this case, both approaches are so similar that we will only show plots using the second approach, with one observation per school.

Figure 1.1 shows the distribution of MCA math test scores as somewhat left skewed. MCA test scores for sixth graders are scaled to fall between 600 and 700, where scores above 650 for individual students indicate “meeting standards”. Thus, schools with averages below 650 will often have increased incentive to improve their scores the following year. When we refer to the “math score” for a particular school in a particular year, we will assume that score represents the average for all sixth graders at that school. In Figure 1.2, we see that test scores are generally higher for both schools in rural areas and for public non-charter schools. Note that in this data set there are 237 schools in rural areas and 381 schools in urban areas, as well as 545 public non-charter schools and 73 charter schools. In addition, we can see in Figure 1.3 that schools tend to have lower math scores if they have higher percentages of students with free and reduced lunch, with special education needs, or who are non-white.

```
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
```



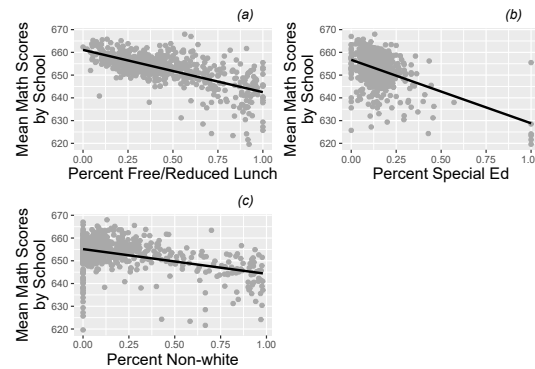
**FIGURE 1.1:** Histogram of mean sixth grade MCA math test scores over the years 2008-2010 for 618 Minnesota schools.



**FIGURE 1.2:** Boxplots of categorical Level Two covariates vs. average MCA math scores. Plot (a) shows charter vs. public non-charter schools, while plot (b) shows urban vs. rural schools.

### 1.3.4 Exploratory analyses for longitudinal data

In addition to the initial exploratory analyses above, longitudinal data—multilevel data with time at Level One—calls for further plots and summaries that describe time trends within and across individuals. For example, we can examine trends over time within individual schools. Figure 1.4 provides a **lattice plot** illustrating trends over time for the first 24 schools in the data set. We note differences among schools in starting point (test scores in 2008), slope (change in test scores over the three year period), and form of the relationship. These differences among schools are nicely illustrated in so-called **spaghetti plots** such as Figure 1.5, which overlays the individual schools' time trends (for the math test scores) from Figure 1.4 on a single set of axes. In order



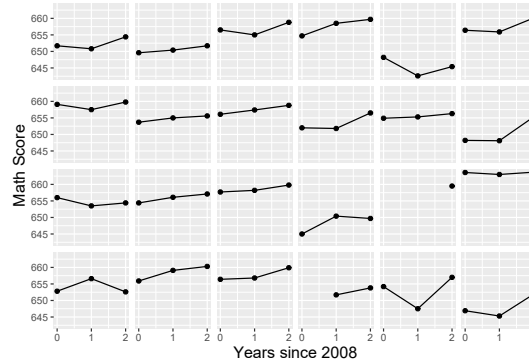
**FIGURE 1.3:** Scatterplots of average MCA math scores by (a) percent free and reduced lunch, (b) percent special education, and (c) percent non-white in a school.

to illustrate the overall time trend without making global assumptions about the form of the relationship, we overlaid in bold a nonparametric fitted curve through a **loess smoother**. LOESS comes from “locally weighted scatterplot smoother”, in which a low-degree polynomial is fit to each data point using weighted regression techniques, where nearby points receive greater weight. LOESS is a computationally intensive method which performs especially well with larger sets of data, although ideally there would be a greater diversity of x-values than the three time points we have. In this case, the loess smoother follows very closely to a linear trend, indicating that assuming a linear increase in test scores over the three year period is probably a reasonable simplifying assumption. To further examine the hypothesis that linearity would provide a reasonable approximation to the form of the individual time trends in most cases, Figure 1.6 shows a lattice plot containing linear fits through ordinary least squares rather than connected time points as in Figure 1.4.

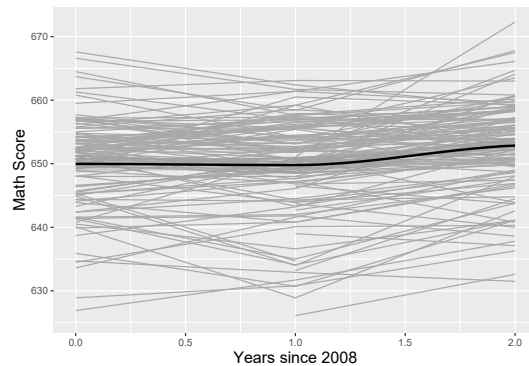
```
## `geom_smooth()` using formula 'y ~ x'
```

```
## `geom_smooth()` using formula 'y ~ x'
```

Just as we explored the relationship between our response (average math scores) and important covariates in Section 1.3.3, we can now examine the relationships between time trends by school and important covariates. For instance, Figure 1.7 shows that charter schools had math scores that were lower on average than public non-charter schools and more variable. This type of plot is sometimes called a **trellis graph**, since it displays a grid of smaller charts with consistent scales, where each smaller chart represents a



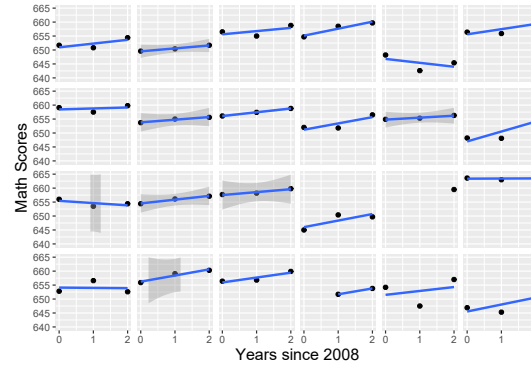
**FIGURE 1.4:** Lattice plot by school of math scores over time for the first 24 schools in the data set.



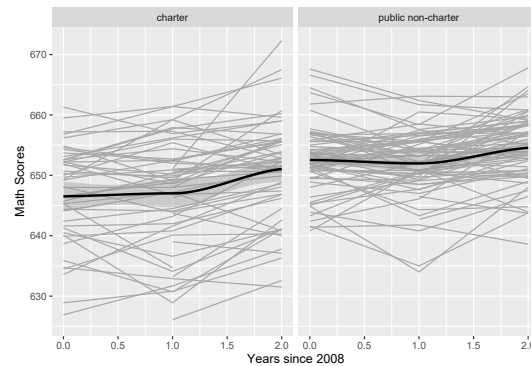
**FIGURE 1.5:** Spaghetti plot of math scores over time by school, for all the charter schools and a random sample of public non-charter schools, with overall fit using loess (bold).

condition—an item in a category. Trends over time by school type are denoted by bold loess curves. Public non-charter schools have higher scores across all years; both school types show little growth between 2008 and 2009, but greater growth between 2009 and 2010, especially charter schools. Exploratory analyses like this can be repeated for other covariates, such as percent free and reduced lunch in Figure 1.8. The trellis plot automatically divides schools into four groups based on quartiles of their percent free and reduced lunch, and we see that schools with lower percentages of free and reduced lunch students tend to have higher math scores and less variability. Across all levels of free and reduced lunch, we see greater gains between 2009 and 2010 than between 2008 and 2009.





**FIGURE 1.6:** Lattice plot by school of math scores over time with linear fit for the first 24 schools in the data set.

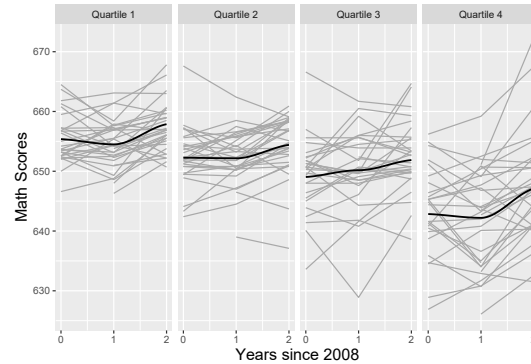


**FIGURE 1.7:** Spaghetti plots showing time trends for each school by school type, for a random sample of charter schools (left) and public non-charter schools (right), with overall fits using loess (bold).

## 1.4 Preliminary two-stage modeling

### 1.4.1 Linear trends within schools

Even though we know that every school's math test scores were not strictly linearly increasing or decreasing over the observation period, a linear model for individual time trends is often a simple but reasonable way to model data. One advantage of using a linear model within school is that each school's data points can be summarized with two summary statistics—an intercept and a slope (obviously, this is an even bigger advantage when there are more

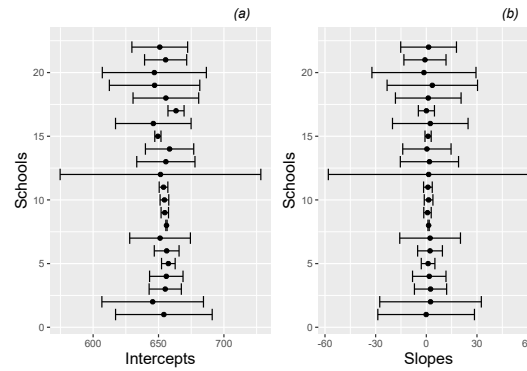


**FIGURE 1.8:** Spaghetti plots showing time trends for each school by quartiles of percent free and reduced lunch, with loess fits.

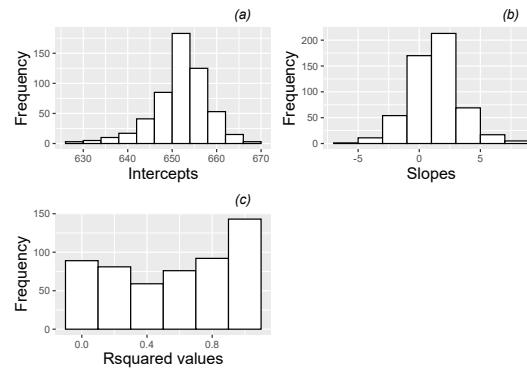
observations over time per school). For instance, we see in Figure 1.6 that sixth graders from the school depicted in the top right slot slowly increased math scores over the three year observation period, while students from the school depicted in the fourth column of the top row generally experienced decreasing math scores over the same period. As a whole, the linear model fits individual trends pretty well, and many schools appear to have slowly increasing math scores over time, as researchers in this study may have hypothesized.

Another advantage of assuming a linear trend at Level One (within schools) is that we can examine summary statistics across schools. Both the intercept and slope are meaningful for each school: the *intercept* conveys the school's math score in 2008, while the *slope* conveys the school's average yearly increase or decrease in math scores over the three year period. Figure 1.9 shows that point estimates and uncertainty surrounding individual estimates of intercepts and slopes vary considerably. In addition, we can generate summary statistics and histograms for the 618 intercepts and slopes produced by fitting linear regression models at Level One, in addition to R-square values which describe the strength of fit of the linear model for each school (Figure 1.10). For our 618 schools, the mean math score for 2008 was 651.4 (SD=7.28), and the mean yearly rate of change in math scores over the three year period was 1.30 (SD=2.51). We can further examine the relationship between schools' intercepts and slopes. Figure 1.11 shows a general decreasing trend, suggesting that schools with lower 2008 test scores tend to have greater growth in scores between 2008 and 2010 (potentially because those schools have more room for improvement); this trend is supported with a correlation coefficient of -0.32 between fitted intercepts and slopes. Note that, with only 3 or fewer observations for each school, extreme or intractable values for the slope and R-square are possible. For example, slopes cannot be estimated for those schools with just a single test score, R-square values cannot be calculated for those

schools with no variability in test scores between 2008 and 2010, and R-square values must be 1 for those schools with only two test scores.



**FIGURE 1.9:** Point estimates and 95% confidence intervals for (a) intercepts and (b) slopes by school, for the first 24 schools in the data set.

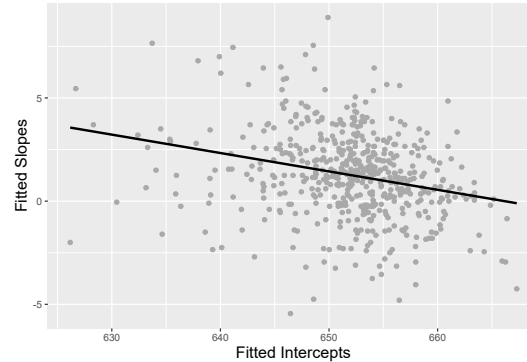


**FIGURE 1.10:** Histograms for (a) intercepts, (b) slopes, and (c) R-square values from fitted regression lines by school.

```
## `geom_smooth()` using formula 'y ~ x'
```

### 1.4.2 Effects of level two covariates on linear time trends

Summarizing trends over time within schools is typically only a start, however. Most of the primary research questions from this study involve comparisons among schools, such as: (a) are there significant differences between charter schools and public non-charter schools, and (b) do any differences between

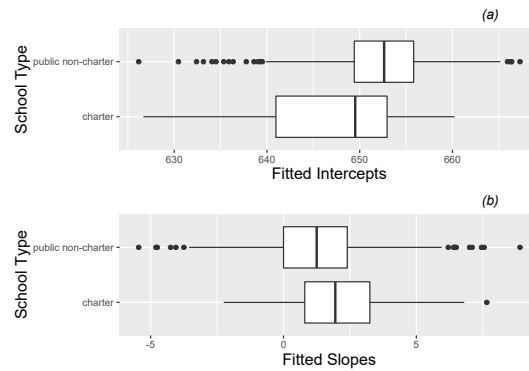


**FIGURE 1.11:** Scatterplot showing the relationship between intercepts and slopes from fitted regression lines by school.

charter schools and public schools change with percent free and reduced lunch, percent special education, or location? These are Level Two questions, and we can begin to explore these questions by graphically examining the effects of school-level variables on schools' linear time trends. By school-level variables, we are referring to those covariates that differ by school but are not dependent on time. For example school type (charter or public non-charter), urban or rural location, percent non-white, percent special education, and percent free and reduced lunch are all variables which differ by school but which don't change over time, at least as they were assessed in this study. Variables which would be time-dependent include quantities such as per pupil funding and reading scores.

Figure 1.12 shows differences in the average time trends by school type, using estimated intercepts and slopes to support observations from the spaghetti plots in Figure 1.7. Based on intercepts, charter schools have lower math scores, on average, in 2008 than public non-charter schools. Based on slopes, however, charter schools tend to improve their math scores at a slightly faster rate than public schools, especially at the seventy-fifth percentile and above. By the end of the three year observation period we would nevertheless expect charter schools to have lower average math scores than public schools. For another exploratory perspective on school type comparisons, we can examine differences between school types with respect to math scores in 2008 and math scores in 2010. As expected, boxplots by school type (Figure 1.13) show clearly lower math scores for charter schools in 2008, but differences are slightly less dramatic in 2010.

Any initial exploratory analyses should also investigate effects of potential confounding variables such as school demographics and location. If we discover, for instance, that those schools with higher levels of poverty (measured by the percentage of students receiving free and reduced lunch) display lower



**FIGURE 1.12:** Boxplots of (a) intercepts and (b) slopes by school type (charter vs. public non-charter).



**FIGURE 1.13:** Boxplots of (a) 2008 and (b) 2010 math scores by school type (charter vs. public non-charter).

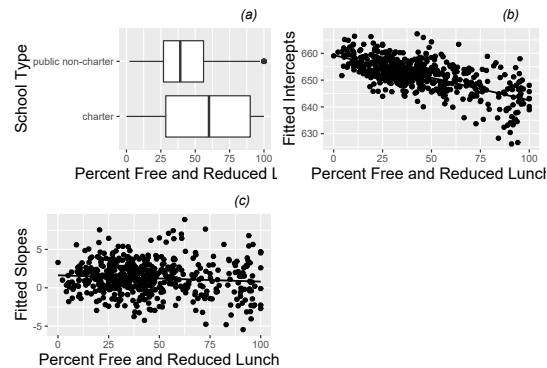
test scores in 2008 but greater improvements between 2008 and 2010, then we might be able to use percentage of free and reduced lunch in statistical modeling of intercepts and slopes, leading to more precise estimates of the charter school effects on these two outcomes. In addition, we should also look for any interaction with school type—any evidence that the difference between charter and non-charter schools changes based on the level of a confounding variable. For example, do charter schools perform better relative to non-charter schools when there is a large percentage of non-white students at a school?

With a confounding variable such as percentage of free and reduced lunch, we will treat this variable as continuous to produce the most powerful exploratory analyses. We can begin by examining boxplots of free and reduced lunch percentage against school type (Figure 1.14). We observe that charter schools tend to have greater percentages of free and reduced lunch students

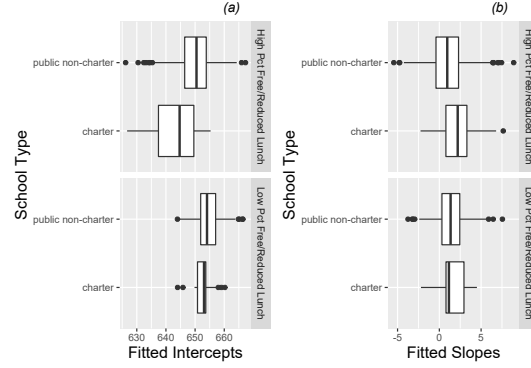
as well as greater school-to-school variability. Next, we can use scatterplots to graphically illustrate the relationships between free and reduced lunch percentages and significant outcomes such as intercept and slope (also Figure 1.14). In this study, it appears that schools with higher levels of free and reduced lunch (i.e., greater poverty) tend to have lower math scores in 2008, but there is little evidence of a relationship between levels of free and reduced lunch and improvements in test scores between 2008 and 2010. These observations are supported with correlation coefficients between percent free and reduced lunch and intercepts ( $r=-0.61$ ) and slopes ( $r=-0.06$ ).

A less powerful but occasionally informative way to look at the effect of a continuous confounder on an outcome variables is by creating a categorical variable out of the confounder. For instance, we could classify any school with a percentage of free and reduced lunch students above the median as having a high percentage of free and reduced lunch students and all other schools as having a low percentage of free and reduced lunch students. Then we could examine a possible interaction between percent free and reduced lunch and school type through a series of four boxplots (Figure 1.15). In fact, these boxplots suggest that the gap between charter and public non-charter schools in 2008 was greater in schools with a high percentage of free and reduced lunch students, while the difference in rate of change in test scores between charter and public non-charter schools appeared similar for high and low levels of free and reduced lunch. We will investigate these trends more thoroughly with statistical modeling.

```
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
```



**FIGURE 1.14:** (a) Boxplot of percent free and reduced lunch by school type (charter vs. public non-charter), along with scatterplots of (b) intercepts and (c) slopes from fitted regression lines by school vs. percent free and reduced lunch.



**FIGURE 1.15:** Boxplots of (a) intercepts and (b) slopes from fitted regression lines by school vs. school type (charter vs. public non-charter), separated by high and low levels of percent free and reduced lunch.

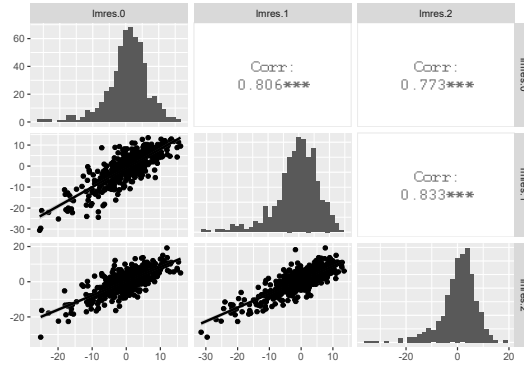
The effect of other confounding variables (e.g., percent non-white, percent special education, urban or rural location) can be investigated in a similar fashion to free and reduced lunch percentage, both in terms of main effect (variability in outcomes such as slope and intercept which can be explained by the confounding variable) and interaction with school type (ability of the confounding variable to explain differences between charter and public non-charter schools). We leave these explorations as an exercise.

### 1.4.3 Error structure within schools

Finally, with longitudinal data it is important to investigate the error variance-covariance structure of data collected within a school (the Level Two observational unit). In multilevel data, as in the examples we introduced in Chapter ??, we suspect observations within group (like a school) to be correlated, and we strive to model that correlation. When the data within group is collected over time, we often see distinct patterns in the residuals that can be modeled—correlations which decrease systematically as the time interval increases, variances that change over time, correlation structure that depends on a covariate, etc. A first step in modeling the error variance-covariance structure is the production of an exploratory plot such as Figure 1.16. To generate this plot, we begin by modeling MCA math score as a linear function of time using all 1733 observations and ignoring the school variable. This population (marginal) trend is illustrated in Figure 1.5 and is given by:

$$\hat{Y}_{ij} = 651.69 + 1.20Time_{ij}, \quad (1.1)$$

where  $\hat{Y}_{ij}$  is the predicted math score of the  $i^{th}$  school at time  $j$ , where time  $j$  is the number of years since 2008. In this model, the predicted math score will be identical for all schools at a given time point  $j$ . Residuals  $Y_{ij} - \hat{Y}_{ij}$  are then calculated for each observation, measuring the difference between actual math score and the average overall time trend. Figure 1.16 then combines three pieces of information: the upper right triangle contains correlation coefficients for residuals between pairs of years, the diagonal contains histograms of residuals at each time point, and the lower left triangle contains scatterplots of residuals from two different years. In our case, we see that correlation between residuals from adjacent years is strongly positive (0.81-0.83) and does not drop off greatly as the time interval between years increases.



**FIGURE 1.16:** Correlation structure within school. The upper right contains correlation coefficients between residuals at pairs of time points, the lower left contains scatterplots of the residuals at time point pairs, and the diagonal contains histograms of residuals at each of the three time points.

## 1.5 Initial models

Throughout the exploratory analysis phase, our original research questions have guided our work, and now with modeling we return to familiar questions such as:

- are differences between charter and public non-charter schools (in intercept, in slope, in 2010 math score) statistically significant?
- are differences between school types statistically significant, even after accounting for school demographics and location?
- do charter schools offer any measurable benefit over non-charter public



schools, either overall or within certain subgroups of schools based on demographics or location?

As you might expect, answers to these questions will arise from proper consideration of variability and properly identified statistical models. As in Chapter ??, we will begin model fitting with some simple, preliminary models, in part to establish a baseline for evaluating larger models. Then, we can build toward a final model for inference by attempting to add important covariates, centering certain variables, and checking assumptions.

### 1.5.1 Unconditional means model

In the multilevel context, we almost always begin with the **unconditional means model**, in which there are no predictors at any level. The purpose of the unconditional means model is to assess the amount of variation at each level, and to compare variability within school to variability between schools. Define  $Y_{ij}$  as the MCA-II math score from school  $i$  and year  $j$ . Using the composite model specification from Chapter ??:

$$Y_{ij} = \alpha_0 + u_i + \epsilon_{ij} \text{ with } u_i \sim N(0, \sigma_u^2) \text{ and } \epsilon_{ij} \sim N(0, \sigma^2) \quad (1.2)$$

the unconditional means model can be fit to the MCA-II data:

```
#Model A (Unconditional means model)
model.a <- lmer(MathAvgScore~ 1 + (1|schoolid),
                REML=T, data=chart.long)
```

```
## Groups   Name                Variance Std.Dev.
## schoolid (Intercept) 41.9         6.47
## Residual              10.6         3.25
```

```
## Number of Level Two groups = 618
```

```
##              Estimate Std. Error t value
## (Intercept)    652.7      0.2726    2395
```

From this output, we obtain estimates of our three model parameters:

- $\hat{\alpha}_0 = 652.7$  = the mean math score across all schools and all years
- $\hat{\sigma}^2 = 10.6$  = the variance in within-school deviations between individual scores and the school mean across all years

- $\hat{\sigma}_u^2 = 41.9$  = the variance in between-school deviations between school means and the overall mean across all schools and all years

Based on the intraclass correlation coefficient:

$$\hat{\rho} = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}^2} = \frac{41.869}{41.869 + 10.571} = 0.798 \quad (1.3)$$

79.8 percent of the total variation in math scores is attributable to difference among schools rather than changes over time within schools. We can also say that the average correlation for any pair of responses from the same school is 0.798.

### 1.5.2 Unconditional growth model

The second model in most multilevel contexts introduces a covariate at Level One (see Model B in Chapter ??). With longitudinal data, this second model introduces time as a predictor at Level One, but there are still no predictors at Level Two. This model is then called the **unconditional growth model**. The unconditional growth model allows us to assess how much of the within-school variability can be attributed to systematic changes over time.

At the lowest level, we can consider building individual growth models over time for each of the 618 schools in our study. First, we must decide upon a form for each of our 618 growth curves. Based on our initial exploratory analyses, assuming that an individual school's MCA-II math scores follow a linear trend seems like a reasonable starting point. Under the assumption of linearity, we must estimate an intercept and a slope for each school, based on their 1-3 test scores over a period of three years. Compared to time series analyses of economic data, most longitudinal data analyses have relatively few time periods for each subject (or school), and the basic patterns within subject are often reasonably described by simpler functional forms.

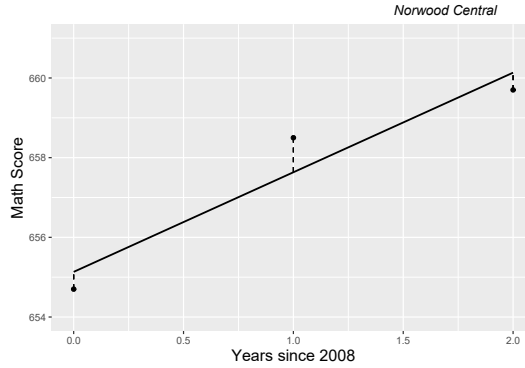
Let  $Y_{ij}$  be the math score of the  $i^{th}$  school in year  $j$ . Then we can model the linear change in math test scores over time for School  $i$  according to Model B:

$$Y_{ij} = a_i + b_i Year_{ij} + \epsilon_{ij} \text{ where } \epsilon_{ij} \sim N(0, \sigma^2) \quad (1.4)$$

The parameters in this model ( $a_i$ ,  $b_i$ , and  $\sigma^2$ ) can be estimated through LLSR methods.  $a_i$  represents the true intercept for School  $i$ —i.e., the expected test score level for School  $i$  when time is zero (2008)—while  $b_i$  represents the true slope for School  $i$ —i.e., the expected yearly rate of change in math score for School  $i$  over the three year observation period. Here we use Roman letters rather than Greek for model parameters since models by school will eventually be a conceptual first step in a multilevel model. The  $\epsilon_{ij}$  terms represent the

deviation of School  $i$ 's actual test scores from the expected results under linear growth—the part of school  $i$ 's test score at time  $j$  that is not explained by linear changes over time. The variability in these deviations from the linear model is given by  $\sigma^2$ . In Figure 1.17, which illustrates a linear growth model for Norwood Central Middle School,  $a_i$  is estimated by the  $y$ -intercept of the fitted regression line,  $b_i$  is estimated by the slope of the fitted regression line, and  $\sigma^2$  is estimated by the variability in the vertical distances between each point (the actual math score in year  $j$ ) and the line (the predicted math score in year  $j$ ).

```
## `geom_smooth()` using formula 'y ~ x'
```



**FIGURE 1.17:** Linear growth model for Norwood Central Middle School

In a multilevel model, we let intercepts ( $a_i$ ) and slopes ( $b_i$ ) vary by school and build models for these intercepts and slopes using school-level variables at Level Two. An unconditional growth model features no predictors at Level Two and can be specified either using formulations at both levels:

- Level One:

$$Y_{ij} = a_i + b_i \text{Year08}_{ij} + \epsilon_{ij} \quad (1.5)$$

- Level Two:

$$\begin{aligned} a_i &= \alpha_0 + u_i \\ b_i &= \beta_0 + v_i \end{aligned}$$

or as a composite model:

$$Y_{ij} = \alpha_0 + \beta_0 \text{Year08}_{ij} + u_i + v_i \text{Year08}_{ij} + \epsilon_{ij} \quad (1.6)$$

where  $\epsilon_{ij} \sim N(0, \sigma^2)$  and

$$\begin{bmatrix} u_i \\ v_i \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{bmatrix} \right).$$

As before,  $\sigma^2$  quantifies the within-school variability (the scatter of points around schools' linear growth trajectories), while now the between-school variability is partitioned into variability in initial status ( $\sigma_u^2$ ) and variability in rates of change ( $\sigma_v^2$ ).

Using the composite model specification, the unconditional growth model can be fit to the MCA-II test data:

```
#Model B (Unconditional growth)
model.b <- lmer(MathAvgScore~ year08 + (year08|schoolid),
  REML=T, data=chart.long)
```

```
## Groups Name Variance Std.Dev. Corr
## schoolid (Intercept) 39.441 6.280
## year08 0.111 0.332 0.72
## Residual 8.820 2.970
```

```
## Number of Level Two groups = 618
```

```
## Estimate Std. Error t value
## (Intercept) 651.408 0.27934 2331.96
## year08 1.265 0.08997 14.06
```

```
## AIC = 10352 ; BIC = 10384
```

From this output, we obtain estimates of our six model parameters:

- $\hat{\alpha}_0 = 651.4$  = the mean math score for the population of schools in 2008.
- $\hat{\beta}_0 = 1.26$  = the mean yearly change in math test scores for the population during the three year observation period.
- $\hat{\sigma}^2 = 8.82$  = the variance in within-school deviations.
- $\hat{\sigma}_u^2 = 39.4$  = the variance between schools in 2008 scores.
- $\hat{\sigma}_v^2 = 0.11$  = the variance between schools in rates of change in math test scores during the three year observation period.
- $\hat{\rho}_{uv} = 0.72$  = the correlation in schools' 2008 math score and their rate of change in scores between 2008 and 2010.

We see that schools had a mean math test score of 651.4 in 2008 and their mean test scores tended to increase by 1.26 points per year over the three year observation period, producing a mean test score at the end of three years of 653.9. According to the t-value (14.1), the increase in mean test scores noted during the three year observation period is statistically significant.

The estimated within-school variance  $\hat{\sigma}^2$  decreased by about 17% from the unconditional means model, implying that 17% of within-school variability in test scores can be explained by a linear increase over time:

$$\begin{aligned} \text{Pseudo } R_{L1}^2 &= \frac{\hat{\sigma}^2(\text{uncond means}) - \hat{\sigma}^2(\text{uncond growth})}{\hat{\sigma}^2(\text{uncond means})} \\ &= \frac{10.571 - 8.820}{10.571} = 0.17 \end{aligned}$$

### 1.5.3 Modeling other trends over time

While modeling linear trends over time is often a good approximation of reality, it is by no means the only way to model the effect of time. One alternative is to model the quadratic effect of time, which implies adding terms for both time and the square of time. Typically, to reduce the correlation between the linear and quadratic components of the time effect, the time variable is often centered first; we have already “centered” on 2008. Modifying Model B to produce an **unconditional quadratic growth model** would take the following form:

- Level One: 
$$Y_{ij} = a_i + b_i \text{Year08}_{ij} + c_i \text{Year08}_{ij}^2 + \epsilon_{ij} \quad (1.7)$$

- Level Two:

$$a_i = \alpha_0 + u_i$$

$$b_i = \beta_0 + v_i$$

$$c_i = \gamma_0 + w_i$$

where  $\epsilon_{ij} \sim N(0, \sigma^2)$  and

$$\begin{bmatrix} u_i \\ v_i \\ w_i \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & & \\ \sigma_{uv} & \sigma_v^2 & \\ \sigma_{uw} & \sigma_{vw} & \sigma_w^2 \end{bmatrix} \right).$$

With the extra term at Level One for the quadratic effect, we now have 3 equations at Level Two, and 6 variance components at Level Two (3 variance terms and 3 covariance terms). However, with only a maximum of 3 observations per school, we lack the data for fitting 3 equations with error

terms at Level Two. Instead, we could model the quadratic time effect with fewer variance components—for instance, by only using an error term on the intercept at Level Two:

$$\begin{aligned}a_i &= \alpha_0 + u_i \\ b_i &= \beta_0 \\ c_i &= \gamma_0\end{aligned}$$

where  $u_i \sim N(0, \sigma_u^2)$ . Models like this are frequently used in practice—they allow for a separate overall effect on test scores for each school while minimizing parameters that must be estimated. The tradeoff is that this model does not allow linear and quadratic effects to differ by school, but we have little choice here without more observations per school. Thus, using the composite model specification, the unconditional quadratic growth model with random intercept for each school can be fit to the MCA-II test data:

```
# Modeling quadratic time trend
model.b2 <- lmer(MathAvgScore~ yearc + yearc2 + (1|schoolid),
  REML=T, data=chart.long)
```

```
## Groups   Name                Variance Std.Dev.
## schoolid (Intercept) 43.05      6.56
## Residual                8.52      2.92

## Number of Level Two groups = 618

##              Estimate Std. Error t value
## (Intercept)  651.942    0.29229 2230.448
## yearc        1.270     0.08758   14.501
## yearc2       1.068     0.15046    7.101

## AIC = 10308 ; BIC = 10335
```

From this output, we see that the quadratic effect is positive and significant ( $t=7.1$ ), in this case indicating that increases in test scores are greater between 2009 and 2010 than between 2008 and 2009. Based on AIC and BIC values, the quadratic growth model outperforms the linear growth model with random intercepts only at level Two (AIC: 10308 vs. 10354; BIC: 10335 vs. 10375).

Another frequently used approach to modeling time effects is the **piecewise linear model**. In this model, the complete time span of the study is divided into two or more segments, with a separate slope relating time to the response in each segment. In our case study there is only one piecewise option—fitting

separate slopes in 2008-09 and 2009-10. With only 3 time points, creating a piecewise linear model is a bit simplified, but this idea can be generalized to segments with more than two years each.

The performance of this model is very similar to the quadratic growth model by AIC and BIC measures, and the story told by fixed effects estimates is also very similar. While the mean yearly increase in math scores was 0.2 points between 2008 and 2009, it was 2.3 points between 2009 and 2010.

Despite the good performances of the quadratic growth and piecewise linear models on our three-year window of data, we will continue to use linear growth assumptions in the remainder of this chapter. Not only is a linear model easier to interpret and explain, but it's probably a more reasonable assumption in years beyond 2010. Predicting future performance is more risky by assuming a steep one year rise or a non-linear rise will continue, rather than by using the average increase over two years.

---

## 1.6 Building to a final model

### 1.6.1 Uncontrolled effects of school type

Initially, we can consider whether or not there are significant differences in individual school growth parameters (intercepts and slopes) based on school type. From a modeling perspective, we would build a system of two Level Two models:

$$\begin{aligned}a_i &= \alpha_0 + \alpha_1 Charter_i + u_i \\ b_i &= \beta_0 + \beta_1 Charter_i + v_i\end{aligned}$$

where  $Charter_i = 1$  if School  $i$  is a charter school and  $Charter_i = 0$  if School  $i$  is a non-charter public school. In addition, the error terms at Level Two are assumed to follow a multivariate normal distribution:

$$\begin{bmatrix} u_i \\ v_i \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{bmatrix} \right).$$

With a binary predictor at Level Two such as school type, we can write out what our Level Two model looks like for public non-charter schools and charter schools.

- Public schools

$$\begin{aligned}a_i &= \alpha_0 + u_i \\ b_i &= \beta_0 + v_i,\end{aligned}$$

- Charter schools

$$\begin{aligned}a_i &= (\alpha_0 + \alpha_1) + u_i \\ b_i &= (\beta_0 + \beta_1) + v_i\end{aligned}$$

Writing the Level Two model in this manner helps us interpret the model parameters from our two-level model. We can use statistical software (such as the `lmer()` function from the `lme4` package in R) to obtain parameter estimates using our 1733 observations, after first converting our Level One and Level Two models into a composite model (Model C) with fixed effects and variance components separated:

$$\begin{aligned}Y_{ij} &= a_i + b_i \text{Year08}_{ij} + \epsilon_{ij} \\ &= (\alpha_0 + \alpha_1 \text{Charter}_i + u_i) + (\beta_0 + \beta_1 \text{Charter}_i + v_i) \text{Year08}_{ij} + \epsilon_{ij} \\ &= [\alpha_0 + \beta_0 \text{Year08}_{ij} + \alpha_1 \text{Charter}_i + \beta_1 \text{Charter}_i \text{Year08}_{ij}] + \\ &\quad [u_i + v_i \text{Year08}_{ij} + \epsilon_{ij}]\end{aligned}$$

```
#Model C (uncontrolled effects of school type on
#  intercept and slope)
model.c <- lmer(MathAvgScore~ charter + year08 +
  charter:year08 + (year08|schoolid),
  REML=T, data=chart.long)
```

```
## Groups   Name      Variance Std.Dev. Corr
## schoolid (Intercept) 35.832   5.986
##          year08      0.131   0.362   0.88
## Residual              8.784   2.964
```

```
## Number of Level Two groups = 618
```

```
##          Estimate Std. Error t value
## (Intercept)  652.0584   0.28449 2291.996
## charter      -6.0184   0.86562  -6.953
## year08        1.1971   0.09427  12.698
## charter:year08  0.8557   0.31430   2.723
```

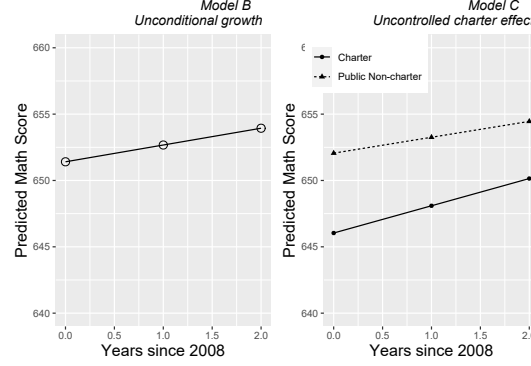


```
## AIC = 10308 ; BIC = 10351
```

Armed with our parameter estimates, we can offer concrete interpretations:

- Fixed effects:
  - $\hat{\alpha}_0 = 652.1$ . The estimated mean test score for 2008 for non-charter public schools is 652.1.
  - $\hat{\alpha}_1 = -6.02$ . Charter schools have an estimated test score in 2008 which is 6.02 points lower than public non-charter schools.
  - $\hat{\beta}_0 = 1.20$ . Public non-charter schools have an estimated mean increase in test scores of 1.20 points per year.
  - $\hat{\beta}_1 = 0.86$ . Charter schools have an estimated mean increase in test scores of 2.06 points per year over the three year observation period, 0.86 points higher than the mean yearly increase among public non-charter schools.
- Variance components:
  - $\hat{\sigma}_u = 5.99$ . The estimated standard deviation of 2008 test scores is 5.99 points, after controlling for school type.
  - $\hat{\sigma}_v = 0.36$ . The estimated standard deviation of yearly changes in test scores during the three year observation period is 0.36 points, after controlling for school type.
  - $\hat{\rho}_{uv} = 0.88$ . The estimated correlation between 2008 test scores and yearly changes in test scores is 0.88, after controlling for school type.
  - $\hat{\sigma} = 2.96$ . The estimated standard deviation in residuals for the individual growth curves is 2.96 points.

Based on t-values reported by R, the effects of `year08` and `charter` both appear to be statistically significant, and there is also significant evidence of an interaction between `year08` and `charter`. Public schools had a significantly higher mean math score in 2008, while charter schools had significantly greater improvement in scores between 2008 and 2010 (although the mean score of charter schools still lagged behind that of public schools in 2010, as indicated in the graphical comparison of models B and C in Figure 1.18). Based on pseudo R-square values, the addition of a charter school indicator to the unconditional growth model has decreased unexplained school-to-school variability in 2008 math scores by 4.7%, while unexplained variability in yearly improvement actually increased slightly. Obviously, it makes little sense that introducing an additional predictor would *reduce* the amount of variability in test scores explained, but this is an example of the limitations in the pseudo  $R^2$  values discussed in Section ??.



**FIGURE 1.18:** Fitted growth curves for Models B and C.

### 1.6.2 Add percent free and reduced lunch as a covariate

Although we will still be primarily interested in the effect of school type on both 2008 test scores and rate of change in test scores (as we observed in Model C), we can try to improve our estimates of school type effects through the introduction of meaningful covariates. In this study, we are particularly interested in Level Two covariates—those variables which differ by school but which remain basically constant for a given school over time—such as urban or rural location, percentage of special education students, and percentage of students with free and reduced lunch. In Section 1.4, we investigated the relationship between percent free and reduced lunch and a school's test score in 2008 and their rate of change from 2008 to 2010.

Based on these analyses, we will begin by adding percent free and reduced lunch as a Level Two predictor for both intercept and slope (Model D):

- Level One: 
$$Y_{ij} = a_i + b_i \text{Year08}_{ij} + \epsilon_{ij} \quad (1.8)$$
- Level Two: 
$$\begin{aligned} a_i &= \alpha_0 + \alpha_1 \text{Charter}_i + \alpha_2 \text{schpctfree}_i + u_i \\ b_i &= \beta_0 + \beta_1 \text{Charter}_i + \beta_2 \text{schpctfree}_i + v_i \end{aligned}$$

The composite model is then:

$$\begin{aligned} Y_{ij} = & [\alpha_0 + \alpha_1 \text{Charter}_i + \alpha_2 \text{schpctfree}_i + \beta_0 \text{Year08}_{ij} \\ & + \beta_1 \text{Charter}_i \text{Year08}_{ij} + \beta_2 \text{schpctfree}_i \text{Year08}_{ij}] \\ & + [u_i + v_i \text{Year08}_{ij} + \epsilon_{ij}] \end{aligned} \quad (1.9)$$

where error terms are defined as in Model C.

```
#Model D2 (Introduce SchPctFree at level 2)
model.d2 <- lmer(MathAvgScore~ charter + SchPctFree + year08 +
  charter:year08 + SchPctFree:year08 + (year08|schoolid),
  REML=T, data=chart.long)
```

```
## Groups Name Variance Std.Dev. Corr
## schoolid (Intercept) 19.13 4.37
## year08 0.16 0.40 0.51
## Residual 8.80 2.97
```

```
## Number of Level Two groups = 618
```

```
## Estimate Std. Error t value
## (Intercept) 659.27848 0.444690 1482.558
## charter -3.43994 0.712836 -4.826
## SchPctFree -0.16654 0.008907 -18.697
## year08 1.64137 0.189499 8.662
## charter:year08 0.98076 0.318583 3.078
## SchPctFree:year08 -0.01041 0.003839 -2.711
```

```
## AIC = 9988 ; BIC = 10043
```

```
drop_in_dev <- anova(model.d2, model.c, test = "Chisq")
```

```
      npar  AIC  BIC logLik  dev Chisq Df
model.c    8 10305 10348 -5144 10289    NA  NA
model.d2   10  9967 10022 -4974  9947 341.5  2
      pval
model.c    NA
model.d2 7.158e-75
```

Compared to Model C, the introduction of school-level poverty based on percentage of students receiving free and reduced lunch in Model D leads to similar conclusions about the significance of the charter school effect on both the intercept and the slope, although the magnitude of these estimates change after controlling for poverty levels. The estimated gap in test scores between charter and non-charter schools in 2008 is smaller in Model D, while estimates of improvement between 2008 and 2010 increase for both types of schools. Inclusion of free and reduced lunch reduces the unexplained variability between

schools in 2008 math scores by 27%, while unexplained variability in rates of change between schools again increases slightly based on pseudo  $R^2$  values. A **likelihood ratio test** using maximum likelihood estimates illustrates that adding free and reduced lunch as a Level Two covariate significantly improves our model ( $\chi^2 = 341.5, df = 2, p < .001$ ). Specific fixed effect parameter estimates are given below:

- $\hat{\alpha}_0 = 659.3$ . The estimated mean math test score for 2008 is 659.3 for non-charter public schools with no students receiving free and reduced lunch.
- $\hat{\alpha}_1 = -3.44$ . Charter schools have an estimated mean math test score in 2008 which is 3.44 points lower than non-charter public schools, controlling for effects of school-level poverty.
- $\hat{\alpha}_2 = -0.17$ . Each 10% increase in the percentage of students at a school receiving free and reduced lunch is associated with a 1.7 point decrease in mean math test scores for 2008, after controlling for school type.
- $\hat{\beta}_0 = 1.64$ . Public non-charter schools with no students receiving free and reduced lunch have an estimated mean increase in math test score of 1.64 points per year during the three years of observation.
- $\hat{\beta}_1 = 0.98$ . Charter schools have an estimated mean yearly increase in math test scores over the three year observation period of 2.62, which is 0.98 points higher than the annual increase for public non-charter schools, after controlling for school-level poverty.
- $\hat{\beta}_2 = -0.010$ . Each 10% increase in the percentage of students at a school receiving free and reduced lunch is associated with a 0.10 point decrease in rate of change over the three years of observation, after controlling for school type.

### 1.6.3 A potential final model with three Level Two covariates

We now begin iterating toward a “final model” for these data, on which we will base conclusions. Being cognizant of typical features of a “final model” as outlined in Chapter ??, we offer one possible final model for this data—Model F:

- Level One:

$$Y_{ij} = a_i + b_i Year08_{ij} + \epsilon_{ij} \quad (1.10)$$

- Level Two:

$$a_i = \alpha_0 + \alpha_1 \text{Charter}_i + \alpha_2 \text{urban}_i + \alpha_3 \text{schpctsped}_i + \alpha_4 \text{schpctfree}_i + u_i$$

$$b_i = \beta_0 + \beta_1 \text{Charter}_i + \beta_2 \text{urban}_i + \beta_3 \text{schpctsped}_i + v_i$$

where we find the effect of charter schools on 2008 test scores after adjusting for urban or rural location, percentage of special education students, and percentage of students that receive free or reduced lunch, and the effect of charter schools on yearly change between 2008 and 2010 after adjusting for urban or rural location and percentage of special education students. We can use AIC and BIC criteria to compare Model F with Model D, since the two models are not nested. By both criteria, Model F is significantly better than Model D: AIC of 9855 vs. 9988, and BIC of 9956 vs. 10043. Based on the R output below, we offer interpretations for estimates of model fixed effects:

```
model.f2 <- lmer(MathAvgScore ~ charter + urban + SchPctFree +
  SchPctSped + charter:year08 + urban:year08 +
  SchPctSped:year08 + year08 +
  (year08|schoolid), REML=T, data=chart.long)
```

```
## Groups Name Variance Std.Dev. Corr
## schoolid (Intercept) 16.94756 4.1167
## year08 0.00475 0.0689 0.85
## Residual 8.82197 2.9702
```

```
## Number of Level Two groups = 618
```

```
## Estimate Std. Error t value
## (Intercept) 661.01042 0.512888 1288.800
## charter -3.22286 0.698547 -4.614
## urban -1.11383 0.427566 -2.605
## SchPctFree -0.15281 0.008096 -18.874
## SchPctSped -0.11770 0.020612 -5.710
## year08 2.14430 0.200867 10.675
## charter:year08 1.03087 0.315159 3.271
## urban:year08 -0.52749 0.186480 -2.829
## SchPctSped:year08 -0.04674 0.010166 -4.598
```

```
## AIC = 9885 ; BIC = 9956
```

- $\hat{\alpha}_0 = 661.0$ . The estimated mean math test score for 2008 is 661.0 for public schools in rural areas with no students qualifying for special education or free and reduced lunch.

- $\hat{\alpha}_1 = -3.22$ . Charter schools have an estimated mean math test score in 2008 which is 3.22 points lower than non-charter public schools, after controlling for urban or rural location, percent special education, and percent free and reduced lunch.
- $\hat{\alpha}_2 = -1.11$ . Schools in urban areas have an estimated mean math score in 2008 which is 1.11 points lower than schools in rural areas, after controlling for school type, percent special education, and percent free and reduced lunch.
- $\hat{\alpha}_3 = -0.118$ . A 10% increase in special education students at a school is associated with a 1.18 point decrease in estimated mean math score for 2008, after controlling for school type, urban or rural location, and percent free and reduced lunch.
- $\hat{\alpha}_4 = -0.153$ . A 10% increase in free and reduced lunch students at a school is associated with a 1.53 point decrease in estimated mean math score for 2008, after controlling for school type, urban or rural location, and percent special education.
- $\hat{\beta}_0 = 2.14$ . Public non-charter schools in rural areas with no students qualifying for special education have an estimated increase in mean math test score of 2.14 points per year over the three year observation period, after controlling for percent of students receiving free and reduced lunch.
- $\hat{\beta}_1 = 1.03$ . Charter schools have an estimated mean annual increase in math score that is 1.03 points higher than public non-charter schools over the three year observation period, after controlling for urban or rural location, percent special education, and percent free and reduced lunch.
- $\hat{\beta}_2 = -0.53$ . Schools in urban areas have an estimated mean annual increase in math score that is 0.53 points lower than schools from rural areas over the three year observation period, after controlling for school type, percent special education, and percent free and reduced lunch.
- $\hat{\beta}_3 = -0.047$ . A 10% increase in special education students at a school is associated with an estimated mean annual increase in math score that is 0.47 points lower over the three year observation period, after controlling for school type, urban or rural location, and percent free and reduced lunch.

From this model, we again see that 2008 sixth grade math test scores from charter schools were significantly lower than similar scores from public non-charter schools, after controlling for school location and demographics. However, charter schools showed significantly greater improvement between 2008 and 2010 compared to public non-charter schools, although charter school test scores were still lower than public school scores in 2010, on average. We also tested several interactions between Level Two covariates and charter schools and found none to be significant, indicating that the 2008 gap between charter schools and public non-charter schools was consistent across demographic subgroups. The faster improvement between 2008 and 2010 for charter schools was also consistent across demographic subgroups (found by testing three-way

interactions). Controlling for school location and demographic variables provided more reliable and nuanced estimates of the effects of charter schools, while also providing interesting insights. For example, schools in rural areas not only had higher test scores than schools in urban areas in 2008, but the gap grew larger over the study period given fixed levels of percent special education, percent free and reduced lunch, and school type. In addition, schools with higher levels of poverty lagged behind other schools and showed no signs of closing the gap, and schools with higher levels of special education students had both lower test scores in 2008 and slower rates of improvement during the study period, again given fixed levels of other covariates.

As we demonstrated in this case study, applying multilevel methods to two-level longitudinal data yields valuable insights about our original research questions while properly accounting for the structure of the data.

---

## 1.7 Covariance structure among observations

Part of our motivation for framing our model for multilevel data was to account for the correlation among observations made on the same school (the Level Two observational unit). Our two-level model, through error terms on both Level One and Level Two variables, actually implies a specific within-school covariance structure among observations, yet we have not (until now) focused on this imposed structure. For example:

- What does our two-level model say about the relative variability of 2008 and 2010 scores from the same school?
- What does it say about the correlation between 2008 and 2009 scores from the same school?

In this section, we will describe the within-school covariance structure imposed by our two-level model and offer alternative covariance structures that we might consider, especially in the context of longitudinal data. In short, we will discuss how we might decide if our implicit covariance structure in our two-level model is satisfactory for the data at hand. Then, in the succeeding optional section, we provide derivations of the imposed within-school covariance structure for our standard two-level model using results from probability theory.

### 1.7.1 Standard covariance structure

We will use Model C (uncontrolled effects of school type) to illustrate covariance structure within subjects. Recall that, in composite form, Model C is:

$$\begin{aligned} Y_{ij} &= a_i + b_i \text{Year08}_{ij} + \epsilon_{ij} \\ &= (\alpha_0 + \alpha_1 \text{Charter}_i + u_i) + (\beta_0 + \beta_1 \text{Charter}_i + v_i) \text{Year08}_{ij} + \epsilon_{ij} \\ &= [\alpha_0 + \alpha_1 \text{Charter}_i + \beta_0 \text{Year08}_{ij} + \beta_1 \text{Charter}_i \text{Year08}_{ij}] + [u_i \\ &\quad + v_i \text{Year08}_{ij} + \epsilon_{ij}] \end{aligned}$$

where  $\epsilon_{ij} \sim N(0, \sigma^2)$  and

$$\begin{bmatrix} u_i \\ v_i \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{bmatrix} \right).$$

For School  $i$ , the covariance structure for the three time points has general form:

$$\text{Cov}(\mathbf{Y}_i) = \begin{bmatrix} \text{Var}(Y_{i1}) & \text{Cov}(Y_{i1}, Y_{i2}) & \text{Cov}(Y_{i1}, Y_{i3}) \\ \text{Cov}(Y_{i1}, Y_{i2}) & \text{Var}(Y_{i2}) & \text{Cov}(Y_{i2}, Y_{i3}) \\ \text{Cov}(Y_{i1}, Y_{i3}) & \text{Cov}(Y_{i2}, Y_{i3}) & \text{Var}(Y_{i3}) \end{bmatrix}$$

where, for instance,  $\text{Var}(Y_{i1})$  is the variability in 2008 test scores (time  $j = 1$ ),  $\text{Cov}(Y_{i1}, Y_{i2})$  is the covariance between 2008 and 2009 test scores (times  $j = 1$  and  $j = 2$ ), etc. Since covariance measures the tendency of two variables to move together, we expect positive values for all three covariance terms in  $\text{Cov}(\mathbf{Y}_i)$ , since schools with relatively high test scores in 2008 are likely to also have relatively high test scores in 2009 or 2010. The correlation between two variables then scales covariance terms to values between -1 and 1, so by the same rationale, we expect correlation coefficients between two years to be near 1. If observations within school were independent—that is, knowing a school had relatively high scores in 2008 tells nothing about whether that school will have relatively high scores in 2009 or 2010—then we would expect covariance and correlation values near 0.

It is important to notice that the error structure at Level Two is *not* the same as the within-school covariance structure among observations. That is, the relationship between  $u_i$  and  $v_i$  from the Level Two equations is not the same as the relationship between test scores from different years at the same school (e.g., the relationship between  $Y_{i1}$  and  $Y_{i2}$ ). In other words,

$$\text{Cov}(\mathbf{Y}_i) \neq \begin{bmatrix} u_i \\ v_i \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{bmatrix} \right).$$

Yet, the error structure and the covariance structure *are* connected to each other, as we will now explore.



Using results from probability theory (see Section 1.7.5), we can show that:

$$\begin{aligned} \text{Var}(Y_{ij}) &= \sigma_u^2 + t_{ij}^2 \sigma_v^2 + \sigma^2 + 2t_{ij} \sigma_{uv}, \\ \text{Cov}(Y_{ij}, Y_{ik}) &= \sigma_u^2 + t_{ij} t_{ik} \sigma_v^2 + (t_{ij} + t_{ik}) \sigma_{uv} \end{aligned}$$

for all  $i$ , where our time variable (**year08**) has values  $t_{i1} = 0$ ,  $t_{i2} = 1$ , and  $t_{i3} = 2$  for every School  $i$ . Intuitively, these formulas are sensible. For instance,  $\text{Var}(Y_{i1})$ , the uncertainty (variability) around a school's score in 2008, increases as the uncertainty in intercepts and slopes increases, as the uncertainty around that school's linear time trend increases, and as the covariance between intercept and slope residuals increases (since if one is off, the other one is likely off as well). Also,  $\text{Cov}(Y_{i1}, Y_{i2})$ , the covariance between 2008 and 2009 scores, does not depend on Level One error. Thus, in the 3-by-3 within-school covariance structure of the charter schools case study, our standard two-level model determines all 6 covariance matrix elements through the estimation of four parameters ( $\sigma_u^2, \sigma_{uv}, \sigma_v^2, \sigma^2$ ) and the imposition of a specific structure related to time.

To obtain estimated variances for individual observations and covariances between two time points from the same school, we can simply plug estimated variance components from our two-level model along with time points from our data collection into the equations above. For instance, in Section 1.6.1, we obtained the following estimates of variance components:  $\hat{\sigma}^2 = 8.69$ ,  $\hat{\sigma}_u^2 = 35.97$ ,  $\hat{\sigma}_v^2 = 0.23$ , and  $\hat{\sigma}_{uv} = \hat{\rho} \hat{\sigma}_u \hat{\sigma}_v = 1.81$ . Therefore, our estimated within-school variances for the three time points would be:

$$\begin{aligned} \hat{\text{Var}}(Y_{i1}) &= 35.97 + 0^2 0.23 + 8.69 + 2(0)1.81 = 44.66 \\ \hat{\text{Var}}(Y_{i2}) &= 35.97 + 1^2 0.23 + 8.69 + 2(1)1.81 = 48.50 \\ \hat{\text{Var}}(Y_{i3}) &= 35.97 + 2^2 0.23 + 8.69 + 2(2)1.81 = 52.80 \end{aligned}$$

and our estimated within-school covariances between different time points would be:

$$\begin{aligned} \hat{\text{Cov}}(Y_{i1}, Y_{i2}) &= 35.97 + (0)(1)0.23 + (0 + 1)1.81 = 37.77 \\ \hat{\text{Cov}}(Y_{i1}, Y_{i3}) &= 35.97 + (0)(2)0.23 + (0 + 2)1.81 = 39.58 \\ \hat{\text{Cov}}(Y_{i2}, Y_{i3}) &= 35.97 + (1)(2)0.23 + (1 + 2)1.81 = 41.84 \end{aligned}$$

In fact, these values will be identical for every School  $i$ , since scores were assessed at the same three time points. Thus, we will drop the  $i$  subscript moving forward.

Written in matrix form, our two-level model implicitly imposes this estimated covariance structure on within-school observations for any specific School  $i$ :

$$\hat{Cov}(\mathbf{Y}) = \begin{bmatrix} 44.66 & & \\ 37.77 & 48.50 & \\ 39.58 & 41.84 & 52.80 \end{bmatrix}$$

and this estimated covariance matrix can be converted into an estimated within-school correlation matrix using the identity  $Corr(Y_1, Y_2) = \frac{Cov(Y_1, Y_2)}{\sqrt{Var(Y_1)Var(Y_2)}}$ :

$$\hat{Corr}(\mathbf{Y}) = \begin{bmatrix} 1 & & \\ .812 & 1 & \\ .815 & .827 & 1 \end{bmatrix}$$

A couple of features of these two matrices can be highlighted that offer insights into implications of our standard two-level model on the covariance structure among observations at Level One from the same school:

- Many longitudinal data sets show higher correlation for observations that are closer in time. In this case, we see that correlation is very consistent between all pairs of observations from the same school; the correlation between test scores separated by two years (.815) is approximately the same as the correlation between test scores separated by a single year (.812 for 2008 and 2009 scores; .827 for 2009 and 2010 scores).
- Many longitudinal data sets show similar variability at all time points. In this case, the variability in 2010 (52.80) is about 18% greater than the variability in 2008 (44.66), while the variability in 2009 is in between (48.50).
- Our two-level model actually imposes a quadratic structure on the relationship between variance and time; note that the equation for  $Var(Y_j)$  contains both  $t_j^2$  and  $t_j$ . The variance is therefore minimized at  $t = \frac{-\sigma_{uv}}{\sigma_v^2}$ . With the charter school data, the variance in test scores is minimized when  $t = \frac{-\sigma_{uv}}{\sigma_v^2} = \frac{-1.81}{0.23} = -8.0$ ; that is, the smallest within-school variance in test scores is expected 8.0 years prior to 2008 (i.e., in 2000), and the variance increases parabolically from there. In general, cases in which  $\sigma_v^2$  and  $\sigma_{uv}$  are relatively small have little curvature and fairly consistent variability over time.
- There is no requirement that time points within school need to be evenly spaced or even that each school has an equal number of measurements over time, which makes the two-level model structure nicely flexible.

### 1.7.2 Alternative covariance structures

The standard covariance structure that's implied by our multilevel modeling structure provides a useful model in a wide variety of situations—it provides a reasonable model for Level One variability with a relatively small number

of parameters, and it has sufficient flexibility to accommodate irregular time intervals as well as subjects with different number of observations over time. However, there may be cases in which a better fitting model requires additional parameters, or when a simpler model with fewer parameters still provides a good fit to the data. Here is an outline of a few alternative error structures:

- *Unstructured* - Every variance and covariance term for observations within a school is a separate parameter and is therefore estimated uniquely; no patterns among variances or correlations are assumed. This structure offers maximum flexibility but is most costly in terms of parameters estimated.
- *Compound symmetry* - Assume variance is constant across all time points and correlation is constant across all pairs of time points. This structure is highly restrictive but least costly in terms of parameters estimated.
- *Autoregressive* - Assume variance is constant across all time points, but correlation drops off in a systematic fashion as the gap in time increases. Autoregressive models expand compound symmetry by allowing for a common structure where points closest in time are most highly correlated.
- *Toeplitz* - Toeplitz is similar to the autoregressive model, except that it does not impose any structure on the decline in correlation as time gaps increase. Thus, it requires more parameters to be estimated than the autoregressive model while providing additional flexibility.
- *Heterogeneous variances* - The assumption that variances are equal across time points found in the compound symmetry, autoregressive, and Toeplitz models can be relaxed by introducing additional parameters to allow unequal (heterogeneous) variances.

When the focus of an analysis is on stochastic parameters (variance components) rather than fixed effects, parameter estimates are typically based on restricted maximum likelihood (REML) methods; model performance statistics then reflect only the stochastic portion of the model. Models with the same fixed effects but different covariance structures can be compared as usual—with AIC and BIC measures when models are not nested and with likelihood ratio tests when models are nested. However, using a chi-square distribution to conduct a likelihood ratio test in these cases can often produce a conservative test, with p-values that are too large and not rejected enough (Raudenbush and Bryk [2002]; Singer and Willett [2003]; J. Faraway [2005]). In Section ??, we introduce the parametric bootstrap as a potentially better way of testing models nested in their random effects.

### 1.7.3 Covariance structure in non-longitudinal multilevel models

Careful modeling and estimation of the Level One covariance matrix is especially important and valuable for longitudinal data (with time at Level

One) and as we've seen, our standard two-level model has several nice properties for this purpose. The standard model is also often appropriate for non-longitudinal multilevel models as discussed in Chapter ??, although we must remain aware of the covariance structure implicitly imposed. In other words, the ideas in this section generalize even if time isn't a Level One covariate.

As an example, in Case Study ?? where Level One observational units are musical performances rather than time points, the standard model implies the following covariance structure for Musician  $i$  in Model C, which uses an indicator for large ensembles as a Level One predictor:

$$\begin{aligned} \text{Var}(Y_{ij}) &= \sigma_u^2 + \text{Large}_{ij}^2 \sigma_v^2 + \sigma^2 + 2\text{Large}_{ij} \sigma_{uv} \\ &= \begin{cases} \sigma^2 + \sigma_u^2 & \text{if } \text{Large}_{ij} = 0 \\ \sigma^2 + \sigma_u^2 + \sigma_v^2 + 2\sigma_{uv} & \text{if } \text{Large}_{ij} = 1 \end{cases} \end{aligned}$$

and

$$\begin{aligned} \text{Cov}(Y_{ij}, Y_{ik}) &= \sigma_u^2 + \text{Large}_{ij} \text{Large}_{ik} \sigma_v^2 + (\text{Large}_{ij} + \text{Large}_{ik}) \sigma_{uv} \\ &= \begin{cases} \sigma_u^2 & \text{if } \text{Large}_{ij} = \text{Large}_{ik} = 0 \\ \sigma_u^2 + \sigma_{uv} & \text{if } \text{Large}_{ij} = 0 \text{ and } \text{Large}_{ik} = 1 \text{ or vice versa} \\ \sigma_u^2 + \sigma_v^2 + 2\sigma_{uv} & \text{if } \text{Large}_{ij} = \text{Large}_{ik} = 1 \end{cases} \end{aligned}$$

Note that, in the Music Performance Anxiety case study, each subject will have a unique Level One variance-covariance structure, since each subject has a different number of performances and a different mix of large ensemble and small ensemble or solo performances.

### 1.7.4 Final thoughts regarding covariance structures

In the charter school example, as is often true in multilevel models, the choice of covariance matrix does not greatly affect estimates of fixed effects. The choice of covariance structure could potentially impact the standard errors of fixed effects, and thus the associated test statistics, but the impact appears minimal in this particular case study. In fact, the standard model typically works very well. So is it worth the time and effort to accurately model the covariance structure? If primary interest is in inference regarding fixed effects, and if the standard errors for the fixed effects appear robust to choice of covariance structure, then extensive time spent modeling the covariance structure is not advised. However, if researchers are interested in predicted random effects and estimated variance components in addition to estimated fixed effects, then choice of covariance structure can make a big difference. For instance, if researchers are interested in drawing conclusions about particular schools rather than charter schools in general, they may more carefully model the covariance structure in this study.

### 1.7.5 Details of covariance structures (Optional)

Using Model C as specified in Section 1.7.1, we specified the general covariance structure for School  $i$  as:

$$Cov(\mathbf{Y}_i) = \begin{bmatrix} Var(Y_{i1}) & Cov(Y_{i1}, Y_{i2}) & Cov(Y_{i1}, Y_{i3}) \\ Cov(Y_{i1}, Y_{i2}) & Var(Y_{i2}) & Cov(Y_{i2}, Y_{i3}) \\ Cov(Y_{i1}, Y_{i3}) & Cov(Y_{i2}, Y_{i3}) & Var(Y_{i3}) \end{bmatrix}$$

If  $Y_1 = a_1X_1 + a_2X_2 + a_3$  and  $Y_2 = b_1X_1 + b_2X_2 + b_3$  where  $X_1$  and  $X_2$  are random variables and  $a_i$  and  $b_i$  are constants for  $i = 1, 2, 3$ , then we know from probability theory that:

$$\begin{aligned} Var(Y_1) &= a_1^2 Var(X_1) + a_2^2 Var(X_2) + 2a_1a_2Cov(X_1, X_2) \\ Cov(Y_1, Y_2) &= a_1b_1Var(X_1) + a_2b_2Var(X_2) + (a_1b_2 + a_2b_1)Cov(X_1, X_2) \end{aligned}$$

Applying these identities to Model C, we first see that we can ignore all fixed effects, since they do not contribute to the variability. Thus,

$$\begin{aligned} Var(Y_{ij}) &= Var(u_i + v_i Year08_{ij} + \epsilon_{ij}) \\ &= Var(u_i) + Year08_{ij}^2 Var(v_i) + Var(\epsilon_{ij}) + 2Year08_{ij}Cov(u_i, v_i) \\ &= \sigma_u^2 + Year08_{ij}^2 \sigma_v^2 + \sigma^2 + 2Year08_{ij}\sigma_{uv} \\ &= \sigma_u^2 + t_j^2 \sigma_v^2 + \sigma^2 + 2t_j \sigma_{uv} \end{aligned}$$

where the last line reflects the fact that observations were taken at the same time points for all schools. We can derive the covariance terms in a similar fashion:

$$\begin{aligned} Cov(Y_{ij}, Y_{ik}) &= Cov(u_i + v_i Year08_{ij} + \epsilon_{ij}, u_i + v_i Year08_{ik} + \epsilon_{ik}) \\ &= Var(u_i) + Year08_{ij}Year08_{ik}Var(v_i) + \\ &\quad (Year08_{ij} + Year08_{ik})Cov(u_i, v_i) \\ &= \sigma_u^2 + t_j t_k \sigma_v^2 + (t_j + t_k)\sigma_{uv} \end{aligned}$$

In Model C, we obtained the following estimates of variance components:  $\hat{\sigma}^2 = 8.69$ ,  $\hat{\sigma}_u^2 = 35.97$ ,  $\hat{\sigma}_v^2 = 0.23$ , and  $\hat{\sigma}_{uv} = \hat{\rho}\hat{\sigma}_u\hat{\sigma}_v = 1.81$ . Therefore, our two level model implicitly imposes this covariance structure on within subject observations:

$$Cov(\mathbf{Y}_i) = \begin{bmatrix} 44.66 & & \\ 37.77 & 48.50 & \\ 39.58 & 41.84 & 52.80 \end{bmatrix}$$

and this covariance matrix can be converted into a within-subject correlation matrix:

$$\text{Corr}(\mathbf{Y}_i) = \begin{bmatrix} 1 & & \\ .812 & 1 & \\ .815 & .827 & 1 \end{bmatrix}$$

## 1.8 Notes on Using R (Optional)

The model below is our final model with  $\sigma_{uv}$  set to 0—i.e., we have added the restriction that Level Two error terms are uncorrelated. Motivation for this restriction came from repeated estimates of correlation in different versions of the final model near 1, when empirically a slightly negative correlation might be expected. As we will describe in Chapter ??, inclusion of the Level Two correlation as a model parameter appears to lead to boundary constraints—maximum likelihood parameter estimates near the maximum or minimum allowable value for a parameter. A likelihood ratio test using full maximum likelihood estimates confirms that the inclusion of a correlation term does not lead to an improved model (LRT test statistic = .223 on 1 df,  $p = .637$ ). Estimates of fixed effects and their standard errors are extremely consistent with the full model in Section 1.6.3; only the estimate of the variability in  $\sigma_1$  is noticeably higher.

```
# Modified final model
model.f2a <- lmer(MathAvgScore ~ charter + urban + SchPctFree +
  SchPctSped + charter:year08 + urban:year08 +
  SchPctSped:year08 + year08 +
  (1|schoolid) + (0+year08|schoolid), REML=T, data=chart.long)
```

```
## Groups      Name      Variance Std.Dev.
## schoolid    (Intercept) 17.355   4.166
## schoolid.1 year08        0.114   0.337
## Residual                        8.716   2.952
```

```
## Number of Level Two groups = 618
```

```
##           Estimate Std. Error  t value
## (Intercept)  661.01770   0.515461 1282.381
## charter      -3.22468   0.703174  -4.586
## urban        -1.11663   0.430422  -2.594
## SchPctFree   -0.15295   0.008096 -18.890
## SchPctSped   -0.11777   0.020739  -5.679
```

```
## year08          2.14271    0.202090    10.603
## charter:year08   1.03341    0.317174     3.258
## urban:year08     -0.52442    0.187678    -2.794
## SchPctSped:year08 -0.04672    0.010219    -4.572

## AIC = 9883 ; BIC = 9948
```

```
# LRT comparing final model in chapter (model.f2ml) with maximum
# likelihood estimates to modified final model (model.f2aml)
# with uncorrelated Level Two errors.
drop_in_dev <- anova(model.f2ml, model.f2aml, test = "Chisq")
```

```
      npar  AIC  BIC logLik  dev  Chisq Df
model.f2aml  12 9855 9921  -4916 9831    NA NA
model.f2ml   13 9857 9928  -4916 9831 0.2231 1
      pval
model.f2aml    NA
model.f2ml 0.6367
```

---

## 1.9 Exercises

### 1.9.1 Conceptual Exercises

1. **Parenting and Gang Activity.** Walker-Barnes and Mason [2001] describe “Ethnic differences in the effect of parenting on gang involvement and gang delinquency: a longitudinal, hierarchical linear modeling perspective”. In this study, 300 ninth graders from one high school in an urban southeastern city were assessed at the beginning of the school year about their gang activity, the gang activity of their peers, behavior of their parents, and their ethnic and cultural heritage. Then, information about their gang activity was collected at 7 additional occasions during the school year. For this study: (a) give the observational units at Level One and Level Two, and (b) list potential explanatory variables at both Level One and Level Two.
2. Describe the difference between the wide and long formats for longitudinal data in this study.

3. Describe scenarios or research questions in which a lattice plot would be more informative than a spaghetti plot, and other scenarios or research questions in which a spaghetti plot would be preferable to a lattice plot.
4. Walker-Barnes and Mason summarize their analytic approach in the following way, where HLM = hierarchical linear models, a synonym for multilevel models:

*The first series [of analyses] tested whether there was overall change and/or significant individual variability in gang [activity] over time, regardless of parenting behavior, peer behavior, or ethnic and cultural heritage. Second, given the well documented relation between peer and adolescent behavior . . . HLM analyses were conducted examining the effect of peer gang [activity] on [initial gang activity and] changes in gang [activity] over time. Finally, four pairs of analyses were conducted examining the role of each of the four parenting variables on [initial gang activity and] changes in gang [activity].*

The last series of analyses controlled for peer gang activity and ethnic and cultural heritage, in addition to examining interactions between parenting and ethnic and cultural heritage.

Although the authors examined four parenting behaviors—behavioral control, lax control, psychological control, and parental warmth—they did so one at a time, using four separate multilevel models. Based on their description, write out a sample model from each of the three steps in the series. For each model, (a) write out the two-level model for predicting gang activity, (b) write out the corresponding composite model, and (c) determine how many model parameters (fixed effects and variance components) must be estimated.

5. Table 1.4 shows a portion of Table 2: Results of Hierarchical Linear Modeling Analyses Modeling Gang Involvement from Walker-Barnes and Mason [2001]. Provide interpretations of significant coefficients in context.
6. **Charter Schools.** Differences exist in both sets of boxplots in Figure 1.12. What do these differences imply for multilevel modeling?
7. What implications do the scatterplots in Figure 1.14 (b) and (c) have for multilevel modeling? What implications does the boxplot in Figure 1.14 (a) have?
8. What are the implications of Figure 1.15 for multilevel modeling?
9. Sketch a set of boxplots to indicate an obvious interaction between percent special education and percent non-white in modeling 2008



**TABLE 1.4:** A portion of Table 2: Results of Hierarchical Linear Modeling Analyses Modeling Gang Involvement from Walker-Barnes and Mason (2001). These columns focus on the parenting behavior of psychological control.

Predictor	Coefficient	SE
<b>Intercept (initial Status)</b>		
Base (intercept for predicting int term)	-.219	.160
Peer behavior	.252**	.026
Black Ethnicity	.671*	.289
White/Other ethnicity	.149	.252
Parenting	.076	.050
Black Ethnicity X Parenting	-.161+	.088
White/Other ethnicity X Parenting	-.026	.082
<b>...Slope(change)</b>		
...Base(intercept for predicting slope term)	.028	.030
...Peer behavior	-.011*	.005
...Black ethnicity	-.132*	.054
...White/Other ethnicity	-.059	.046
...Parenting	-.015+	.009
...Black Ethnicity X Parenting	-.048**	.017
...White/Other ethnicity X Parenting	.016	.015

<sup>a</sup> Table reports values for coefficients in the final model with all variables entered. \*  $p < .05$ ; \*\*  $p < .01$ ; +  $p < .10$

math scores. Where would this interaction appear in the multilevel model?

10. In Model A,  $\sigma^2$  is defined as the variance in within-school deviations and  $\sigma_u^2$  is defined as the variance in between-school deviations. Give potential sources of within-school and between-school deviations.
11. In Chapter ?? Model B is called the “random slopes and intercepts model”, while in this chapter Model B is called the “unconditional growth model”. Are these models essentially the same or systematically different? Explain.
12. In Section 1.5.2, why don’t we examine the pseudo  $R^2$  value for Level Two?
13. If we have test score data from 2001-2010, explain how we’d create new variables to fit a piecewise model.
14. In Section 1.6.2, could we have used percent free and reduced lunch

as a Level One covariate rather than 2010 percent free and reduced lunch as a Level Two covariate? If so, explain how interpretations would have changed. What if we had used average percent free and reduced lunch over all three years or 2008 percent free and reduced lunch instead of 2010 percent free and reduced lunch - how would this have changed the interpretation of this term?

15. In Section 1.6.2, why do we look at a 10 percent increase in the percentage of students receiving free and reduced lunch when interpreting  $\hat{\alpha}_2$ ?
16. In Section 1.6.3, if the gap in 2008 math scores between charter and non-charter schools differed for schools of different poverty levels (as measured by percent free and reduced lunch), how would the final model have differed?
17. Explain in your own words why “the error structure at Level Two is *not* the same as the within-school covariance structure among observations”.
18. Here is the estimated unstructured covariance matrix for Model C:

$$Cov(\mathbf{Y}_i) = \begin{bmatrix} 41.87 & & \\ 36.46 & 48.18 & \\ 35.20 & 39.84 & 45.77 \end{bmatrix}$$

Explain why this matrix cannot represent an estimated covariance matrix with a compound symmetry, autoregressive, or Toeplitz structure. Also explain why it cannot represent our standard two-level model.

### 1.9.2 Guided Exercise

1. **Teen Alcohol Use.** Curran et al. [1997] collected data on 82 adolescents at three time points starting at age 14 to assess factors that affect teen drinking behavior. Key variables in the data set `alcohol.csv` (accessed via Singer and Willett [2003]) are as follows:
  - `id` = numerical identifier for subject
  - `age` = 14, 15, or 16
  - `coa` = 1 if the teen is a child of an alcoholic parent; 0 otherwise
  - `male` = 1 if male; 0 if female
  - `peer` = a measure of peer alcohol use, taken when each subject was 14. This is the square root of the sum of two 6-point items about the proportion of friends who drink occasionally or regularly.

- **alcuse** = the primary response. Four items—(a) drank beer or wine, (b) drank hard liquor, (c) 5 or more drinks in a row, and (d) got drunk—were each scored on an 8-point scale, from 0=“not at all” to 7=“every day”. Then **alcuse** is the square root of the sum of these four items.

Primary research questions included:

- do trajectories of alcohol use differ by parental alcoholism?
  - do trajectories of alcohol use differ by peer alcohol use?
1. Identify Level One and Level Two predictors.
  2. Perform a quick EDA. What can you say about the shape of **alcuse**, and the relationship between **alcuse** and **coa**, **male**, and **peer**? Appeal to plots and summary statistics in making your statements.
  3. Generate a plot as in Figure 1.4 with alcohol use over time for all 82 subjects. Comment.
  4. Generate three spaghetti plots with loess fits similar to Figure 1.7 (one for **coa**, one for **male**, and one after creating a binary variable from **peer**). Comment on what you can conclude from each plot.
  5. Fit a linear trend to the data from each of the 82 subjects using **age** as the time variable. Generate histograms as in Figure 1.6 showing the results of these 82 linear regression lines, and generate pairs of boxplots as in Figure 1.12 for **coa** and **male**. No commentary necessary. [Hint: to produce Figure 1.12, you will need a data frame with one observation per subject.]
  6. Repeat 5 using centered age (**age14** = **age** - 14) as the time variable. Also generate a pair of scatterplots as in Figure 1.14 for peer alcohol use. Comment on trends you observe in these plots. [Hint: after forming **age14**, append it to your current data frame.]
  7. Discuss similarities and differences between (5) and (6). Why does using **age14** as the time variable make more sense in this example?
  8. (Model A) Run an unconditional means model. Report and interpret the intraclass correlation coefficient.
  9. (Model B) Run an unconditional growth model with **age14** as the time variable at Level One. Report and interpret estimated fixed effects, using proper notation. Also report and interpret a pseudo-Rsquare value.
  10. (Model C) Build upon the unconditional growth model by adding the effects of having an alcoholic parent and peer alcohol use in both Level Two equations. Report and interpret all estimated fixed effects, using proper notation.
  11. (Model D) Remove the child of an alcoholic indicator variable

as a predictor of slope in Model C (it will still be a predictor of intercept). Write out Model D as both a two-level and a composite model using proper notation (including error distributions); how many parameters (fixed effects and variance components) must be estimated? Compare Model D to Model C using an appropriate method and state a conclusion.

### 1.9.3 Open-ended Exercises

1. **UCLA Nurse Blood Pressure Study.** A study by [Goldstein and Shapiro \[2000\]](#) collected information from 203 registered nurses in the Los Angeles area between 24 and 50 years of age on blood pressure and potential factors that contribute to hypertension. This information includes family history, including whether the subject had one or two hypertensive parents, as well as a wide range of measures of the physical and emotional condition of each nurse throughout the day. Researchers sought to study the links between blood pressure and family history, personality, mood changes, working status, and menstrual phase.

Data from this study provided by [Weiss \[2005\]](#) includes observations (40-60 per nurse) repeatedly taken on the 203 nurses over the course of a single day. The first blood pressure measurement was taken half an hour before the subject's normal start of work, and BP was then measured approximately every 20 minutes for the rest of the day. At each blood pressure reading, the nurses also rate their mood on several dimensions, including how stressed they feel at the moment the blood pressure is taken. In addition, the activity of each subject during the 10 minutes before each reading was measured using an actigraph worn on the waist. Each of the variables in `nursebp.csv` is described below:

- **SNUM:** subject identification number
- **SYS:** systolic blood pressure (mmHg)
- **DIA:** diastolic blood pressure (mmHg)
- **HRT:** heart rate (beats per minute)
- **MNACT5:** activity level (frequency of movements in 1-minute intervals, over a 10-minute period )
- **PHASE:** menstrual phase (follicular—beginning with the end of menstruation and ending with ovulation, or luteal—beginning with ovulation and ending with pregnancy or menstruation)
- **DAY:** workday or non-workday
- **POSTURE:** position during blood pressure measurement—either sitting, standing, or reclining
- **STR, HAP, TIR:** self-ratings by each nurse of their level of stress,

happiness and tiredness at the time of each blood pressure measurement on a 5-point scale, with 5 being the strongest sensation of that feeling and 1 the weakest

- **AGE**: age in years
- **FH123**: coded as either NO (no family history of hypertension), YES (1 hypertensive parent), or YESYES (both parents hypertensive)
- **time**: in minutes from midnight
- **timept**: number of the measurement that day (approximately 50 for each subject)
- **timepass**: time in minutes beginning with 0 at time point 1

Using systolic blood pressure as the primary response, write a short report detailing factors that are significantly associated with higher systolic blood pressure. Be sure to support your conclusions with appropriate exploratory plots and multilevel models. In particular, how are work conditions—activity level, mood, and work status—related to trends in blood pressure levels? As an appendix to your report, describe your modeling process—how did you arrive at your final model, which covariates are Level One or Level Two covariates, what did you learn from exploratory plots, etc.

Potential alternative directions: consider diastolic blood pressure or heart rate as the primary response variable, or even try modeling emotion rating using a multilevel model.

2. **Completion Rates at US Colleges.** Education researchers wonder which factors most affect the completion rates at US colleges. Using the IPEDS database containing data from 1310 institutions over the years 2002-2009 [[National Center for Education Statistics](#)], the following variables were assembled in `colleges.csv`:

- **id** = unique identification number for each college or university

Response:

- **rate** = completion rate (number of degrees awarded per 100 students enrolled)

Level 1 predictors:

- **year**
- **instpct** = percentage of students who receive an institutional grant
- **instamt** = typical amount of an institutional grant among recipients (in \$1000s)

Level 2 predictors:

- **faculty** = mean number of full-time faculty per 100 students during 2002-2009
- **tuition** = mean yearly tuition during 2002-2009 (in \$1000s)

Perform exploratory analyses and run multilevel models to determine significant predictors of baseline (2002) completion rates and changes in completion rates between 2002 and 2009. In particular, is the percentage of grant recipients or the average institutional grant awarded related to completion rate?

---

## *Bibliography*

---

- Richard Buddin and Ron Zimmer. Student achievement in charter schools: A complex picture. *Journal of Policy Analysis and Management*, 24(2):351–371, 2005. doi: 10.1002/pam.20093. URL <http://dx.doi.org/10.1002/pam.20093>.
- Ann Cannon, George Cobb, Brad Hartlaub, Julie Legler, Robin Lock, Tom Moore, Allan Rossman, and Jeff Witmer. *Stat2*. Macmillan, 2019.
- Patrick J. Curran, Eric Stice, and Laurie Chassin. The relation between adolescent alcohol use and peer alcohol use: A longitudinal random coefficients model. *Journal of Consulting and Clinical Psychology*, 65(1):130–140, 1997. doi: 10.1037/0022-006X.65.1.130. URL <http://dx.doi.org/10.1037/0022-006X.65.1.130>.
- Kara Finnigan, Nancy Adelman, Lee Anderson, Lynnyonne Cotton, Mary Beth Donnelly, and Tiffany Price. Evaluation of public charter schools program: Final evaluation report, 2004.
- I. B. Goldstein and D. Shapiro. Ambulatory blood pressure in women: Family history of hypertension and personality. *Psychology, Health & Medicine*, 5(3):227–240, 2000. doi: 10.1080/713690197. URL <https://doi.org/10.1080/713690197>.
- Preston C. GreenIII, Bruce D. Baker, and Joseph O. Oluwole. Having it both ways: How charter schools try to obtain funding of public schools and the autonomy of private schools. *Emory Law Journal*, 63(2):303–337, 2003.
- Julian J. Faraway. *Extending the Linear Model With R*. Chapman & Hall/CRC, Boca Raton, FL, 2005.
- KIPP. Kipp north star academy, 2017. URL <http://www.kipp.org/school/kipp-north-star-academy/>.
- Nan M. Laird. Missing data in longitudinal studies. *Statistics in Medicine*, 7(1-2):305–315, 1988. doi: 10.1002/sim.4780070131. URL <http://dx.doi.org/10.1002/sim.4780070131>.
- Minnesota Department of Education. Minnesota department of education data center. URL <https://education.mn.gov/MDE/Data/>.

- National Center for Education Statistics. The integrated postsecondary education data system. URL <https://nces.ed.gov/ipeds/>.
- Stephen W. Raudenbush and Anthony S. Bryk. *Hierarchical Linear Models: Applications and Data Analysis Methods*. SAGE Publications, Inc., 2 edition, 2002.
- Judith D. Singer and John B. Willett. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford University Press, Inc., New York, New York, 1 edition, 2003.
- U.S. Department of Education. National charter school resource center. URL <https://charterschoolcenter.ed.gov/faqs>.
- Chanequa J. Walker-Barnes and Craig A. Mason. Ethnic differences in the effect of parenting on gang involvement and gang delinquency: A longitudinal, hierarchical linear modeling perspective. *Child Development*, 72(6):1814–1831, 2001. doi: 10.1111/1467-8624.00380. URL <http://dx.doi.org/10.1111/1467-8624.00380>.
- Robert E. Weiss. *Modeling Longitudinal Data*. Springer-Verlag, New York, NY, 2005.
- John Witte, David Weimer, Arnold Shober, and Paul Schlomer. The performance of charter schools in wisconsin. *Journal of Policy Analysis and Management*, 26(3):557–573, 2007. doi: 10.1002/pam.20265. URL <http://dx.doi.org/10.1002/pam.20265>.