The Report committee for Andrea Jean Kaplan
Certifies that this is the approved version of the following report:

An Overview of Multilevel Regression

APPROVED BY

SUPERVISING COMMITTEE:

Martha Smith, Supervisor

John Luecke, Supervisor

An Overview of Multilevel Regression

by

Andrea Jean Kaplan, B.S.

REPORT

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF ARTS

THE UNIVERSITY OF TEXAS AT AUSTIN

December 2010

For my family.

Acknowledgments

My profound gratitute goes to my supervisors Martha Smith and John Luecke. Their dedicated support and unwavering patience combined with an expert ability to help me learn math has made my graduate school experience unforgettable.

I would also like to thank my family. The support from my mother and father has been unfailing and very appreciated. I would like to acknowledge Carrie Chennault for her tireless encouragement. Thank you for helping me remember the infinite fun in intellectual learning.

My teachers of mathematics have inspired me and they deserve my thanks: Lauri Crestani, Tasha Beretvas, and Mary Parker, to name a few.

Andee Kaplan

The University of Texas at Austin

December 2010

An Overview of Multilevel Regression

Andrea Jean Kaplan, M.A.
The University of Texas at Austin, 2010

Supervisors: Martha Smith
John Luecke

Due to the inherently hierarchical nature of many natural phenomena, data collected rests in nested entities. As an example, students are nested in schools, school are nested in districts, districts are nested in counties, and counties are nested within states. Multilevel models provide a statistical framework for investigating and drawing conclusions regarding the influence of factors at differing hierarchical levels of analysis. The work in this paper serves as an introduction to multilevel models and their comparison to Ordinary Least Squares (OLS) regression. We overview three basic model structures: variable intercept model, variable slope model, and hierarchical linear model and illustrate each model with an example of student data. Then, we contrast the three multilevel models with the OLS model and present a method for producing confidence intervals for the regression coefficients.

Table of Contents

Chapter 1

Introduction & Background

Many data sets inherently possess a hierarchical structure. As an example, think of data on students in a school district or county. Each student is in a particular school; thus we say students are nested in schools. Metrics may exist that describe students, such as GPA or socio-economic status, while other variables describe the school, such as student body size and student-teacher ratio. A researcher interested in the influence of student-teacher ratio on GPA scores needs a technique that can handle multiple levels of a hierarchy.

Multilevel models were developed to analyze data structured in this way, with lower level units, individuals, nested in higher level units, groups. In the previous example, students are the lowest level nested within schools. However, the hierarchy does not have to end here. Data may be collected on the school district, which is within a county, which is within a state. Although hierarchy seemingly can extend forever, this report restricts its scope to a two-level model.

Multilevel models may be thought of as extensions of linear regression in which data is structured in groups and coefficients may vary by group. A primary advantage of multilevel models is that they allow one to investigate relationships within a unit, as well as between units simultaneously [6, pg. 237]. For example, we could investigate the mean and variability of test scores within a school, as well as the variability of test scores between schools.

Data

We will illustrate multilevel models with student examination results obtained from the University of Bristol Centre for Multilevel Modelling [10]. The General Certificate of Secondary Examination (GCSE) scores provides test results from 8,857 students in 74 schools in inner London [7]. After omitting examination results with incomplete intake data, the analysis includes results from 4,059 students in 65 schools. The data set includes a normalized exam score as the outcome variable. The predictors include the standardized London reading test taken when the students were 11 years old, verbal reasoning category intake at 11 years, gender, and school gender (mixed, boys, or girls).

Variables Used

school School ID – a factor with 65 levels.

normexam Normalized exam score.

standLRT Standardized London reading test score.

Models

We will fit the data using the **lm()** [11] and **lmer()** [2] functions in R, which fit a simple linear regression and generalized mixed models with variable coefficients, respectively. The **lmer()** model estimation function has the capability of producing Maximum Likelihood (ML) estimates and Restricted Maximum Likelihood (REML) estimates.

For demonstrative purposes, we will simplify the data. An exploration of the relationship between students' exam scores and their LRT score is shown from the perspective of three multilevel models in comparison with a linear regression model.

Simple Linear Regression

We will overview a simple linear regression of **normexam** on **standLRT** ignoring school from the exam dataset introduced in this chapter. Combining all the schools together and fitting a linear regression to the data is called a complete-pooling method. Looking at the complete pooling method in conjunction with the multilevel modelling method will allow us to present a model comparison later in this paper.

The Ordinary Least Squares (OLS) model equation takes the form:

$$\mathbf{normexam}_i = \beta_0 + \beta_1 \mathbf{standLRT}_i + r_i, \tag{1.1}$$

where r_i are the residual errors that account for the difference between the actual outcomes, $\mathbf{normexam}_i$, and the predicted outcomes, $\beta_0 + \beta_1 \mathbf{standLRT}_i$.

A simple linear regression is fitted in R using the commands in Listing 1.1.

```
##Model 0 - Linear Regression
m0<-lm(normexam ~ standLRT, data=Exam)
```

Listing 1.1: R code for producing OLS model of exam data using lm.

Table 1.1 contains the resulting output. From this, the fitted regression line is $\mathbf{normexam}_i = -0.001 + 0.595\mathbf{standLRT}_i$. Figure 1.1 shows nine of the schools' data overlaid with the simple linear regression line. We truncated the data for ease of viewing, from 4,059 students in 65 schools to 505 students in 9 schools.

	Estimate	St. Error
(Intercept)	-0.001	0.013
standLRT	0.595	0.013

Table 1.1: R (lm) output for OLS model of exam data.

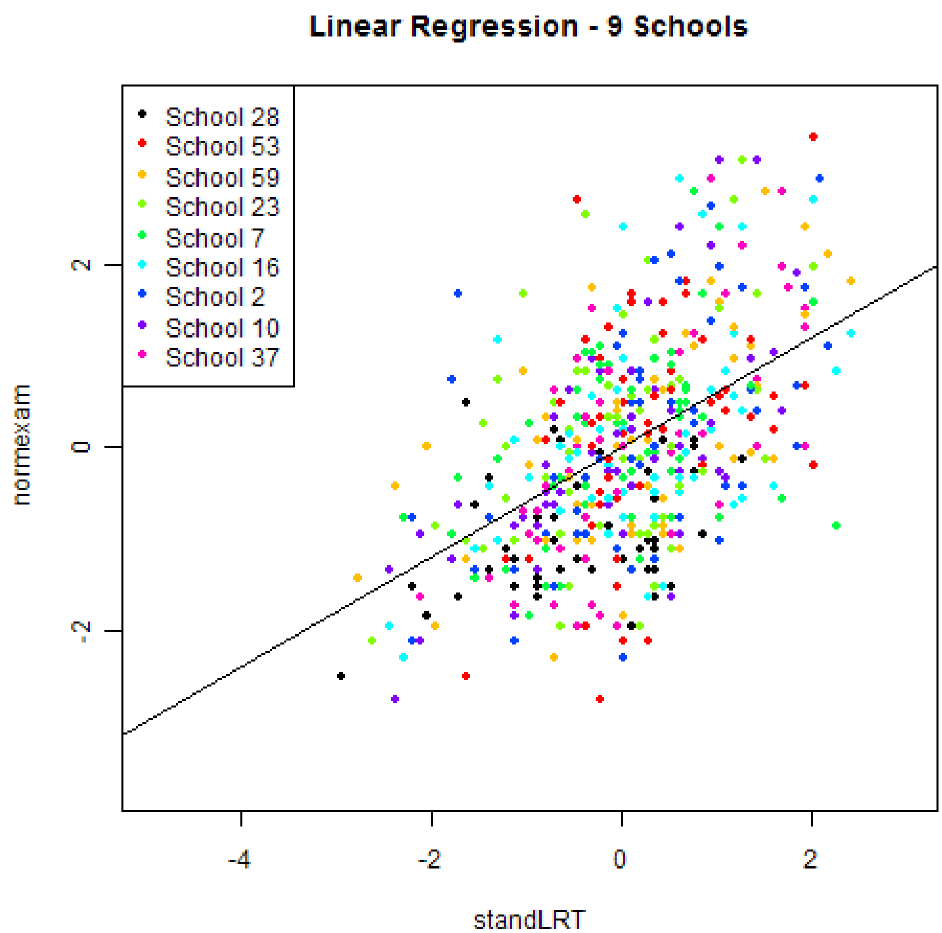

Figure 1.1: OLS model for exam data, limited to nine schools.

Chapter 2

Multilevel Regression

Multiple layers of complexity arise when dealing with multilevel models. In the following sections, these layers of complexity will be defined starting with the simplest case. In order to maintain conventions throughout this paper, we will denote the number of groups as N. We will denote the number of individuals, which will vary between groups, as n_j where $j = 1, ..., N$, and the total number of individuals as $M = \sum_j n_j$.

The multilevel model is a type of regression model that is suitable for data in a hierarchical structure. As with all regression models, there are dependent and explanatory variables. The aim is to identify the relationship between dependent and explanatory variables, e.g. how student-teacher ratios influence GPA scores. Unlike linear regression models, multilevel models contain more than one error term: one for each level. In this paper, we define indices and variables as follows:

j is the index for the groups ($j = 1, ..., N$)

i is the index for the individuals within groups ($i = 1, ..., n_j$).

For the individual i in group j:

Y_{ij} is the dependent variable

x_{ij} is the explanatory variable at the individual level.

For group j:

W_j is the explanatory variable at the group level.

Using the example from Chapter 1, N is the number of schools (65), n_j is the number of students within each school, Y_{ij} are the students' normalized exam scores (**normexam**), and x_{ij} are the students' standardized London reading scores (**standLRT**). There are no group level explanatory variables (W_j) in our example.

An underlying assumption of multilevel modelling is that the variables Y and x have both an individual, as well as a group aspect. For example, the mean of x will differ from group to group. In the following sections, keep in mind that an explanatory variable defined at the individual level will often also contain information about groups.

The multilevel models in the following sections will differ in their specification of the regression coefficients. We will explore three depths of complexity depending on the random or fixed qualities of specific coefficients. The term *fixed effects* denotes regression coefficients that do not vary by group or for group level coefficients. The term *random effects* refers to the group level errors.

2.1 Variable Intercept Model

Consider the first case, in which the intercept varies between groups. This indicates that some groups, on average, tend to have higher responses than others. The regression model contains a variable intercept, but the slope remains fixed:

$$Y_{ij} = \beta_{0j} + \beta_1 x_{ij} + r_{ij}. \tag{2.1}$$

The group dependent intercept, β_{0j} can be split into an average intercept and a group-specific deviation:

$$\beta_{0j} = \gamma_{00} + u_{0j}. \tag{2.2}$$

Substituting (2.2) into (2.1) yields the combined model

$$Y_{ij} = \gamma_{00} + \beta_1 x_{ij} + u_{0j} + r_{ij}. \tag{2.3}$$

As a final step, we change the notation for the regression coefficient from β_j to γ_{10}:

$$Y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + u_{0j} + r_{ij}. \tag{2.4}$$

The reasons for this change in notation will become apparent shortly.

The group specific deviation, u_{0j}, are assumed to be independent, identically distributed variables with mean zero and unknown variance, and are also assumed to be independent of r_{ij}. The random intercept model takes its name from having the group-dependent intercepts vary from group to group.

Example

We illustrate the variable intercept model using the exam data introduced in Chapter 1. Consider the goal of modelling the relationship between **normexam** and **standLRT** with a variable intercept model using (2.1) and (2.2):

$$\mathbf{normexam}_{ij} = \beta_{0j} + \beta_1 \mathbf{standLRT}_{ij} + r_{ij} \tag{2.5}$$

$$\beta_{0j} = \gamma_{00} + u_{0j}. \tag{2.6}$$

In our example, β_1 and γ_{00} are the fixed effects, while u_{0j} are the random effects. Recall, the u_{0j} are assumed to be independent, identically distributed variables with mean zero and unknown variance, and are also assumed to be independent of r_{ij}.

This model can be fit and displayed using the R commands in Listing 2.1

```
##Model 1 - Variable Intercept
m1 <-lmer(normexam ~ standLRT + (1 | school))
```

Listing 2.1: R code for producing variable intercept model of exam data using lmer.

The call to **lmer** starts with a relationship between **normexam** and **standLRT** that does not vary by group and adds **(1 — school)**, which is the intercept allowed to vary by school.

Table 2.1 contains the output of this model. The output of the **lmer**

Linear mixed model fit by **REML**				
Formula: `normexam ~ standLRT + (1 | school)`				
AIC	BIC	logLik	deviance	REMLdev
9377	9402	-4684	9357	9369
Random effects:				
	Groups	Name	Variance	Std.Dev.
	school	(Intercept)	0.093839	0.30633
	Residual		0.565865	0.75224
Number of obs: 4059, **groups:** school, 65				
Fixed effects:				
		Estimate	Std. Error	t value
	(Intercept)	0.002324	0.040349	0.06
	standLRT	0.563308	0.012468	45.18

Table 2.1: R (lmer) output for variable intercept model of exam data.

function is split into six parts: *a*) the estimation technique used; *b*) the formula used; *c*) the model fit criteria used for comparison between models; *d*) the estimates of the variability of the random effects; *e*) the number of individuals and groups; and *f*) the fixed effects estimates. The fixed and random effects estimates correspond to the coefficients in (2.5) and (2.6) as follows:

$$\widehat{\beta}_1 = 0.563308$$
$$\widehat{\gamma}_{00} = 0.002324$$
$$\widehat{\text{Var}}(u_{0j}) \approx 0.093839$$
$$\widehat{\text{Var}}(r_{ij}) \approx 0.565865.$$

Note the $\widehat{}$ notation above indicates these are estimates of the population regression coefficients using the data.

Figure 2.1 shows nine schools in the population and their regression lines. Table 2.2 shows the corresponding coefficients. We find the coefficients

School	(Intercept)	standLRT
28	-0.6088	0.5633
53	0.7269	0.5633
59	-0.6575	0.5633
23	-0.4885	0.5633
7	0.3819	0.5633
16	-0.4081	0.5633
2	0.5054	0.5633
10	-0.3353	0.5633
37	-0.1875	0.5633

Table 2.2: Variable intercept model coefficients for exam data, limited to nine schools.

by using the function **coef(m1)** in R. We can see clearly the intercept coefficients varying by school, while the slope coefficient is the same across all schools because this is what was specified in the model. For example, the fitted

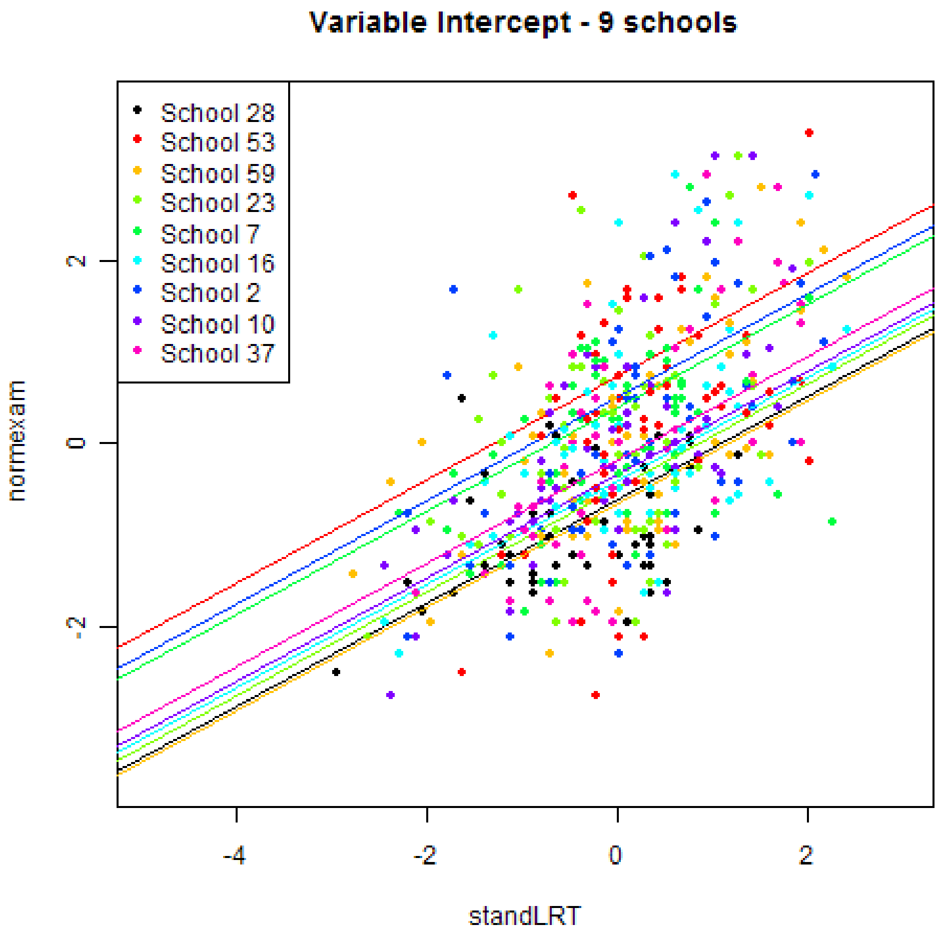

Figure 2.1: Variable intercept model for exam data, limited to nine schools.

regression line for school 28 is $\mathbf{normexam}_i = -0.6088 + 0.5633\mathbf{standLRT}_i$ while the fitted regression line for school 53 is $\mathbf{normexam}_i = 0.7269 + 0.5633\mathbf{standLRT}_i$, and so forth.

By using the commands **fixef(m1)** and **ranef(m1)** in R, we can display the estimated model over the schools – fixed effects – and the school-level errors – random effects. Tables 2.3 and 2.4 show the results of these two commands. Table 2.3 tells us the fitted regression line for the average school is $\mathbf{normexam}_i = 0.0023 + 0.5633\mathbf{standLRT}_i$, while Table 2.4 tells us how much the intercept is shifted up or down in any particular school. For example in school 28, the estimated intercept is 0.6111 lower than average, giving the fitted regression line $\mathbf{normexam}_i = (0.0023 - 0.6111) + 0.5633\mathbf{standLRT}_i = -0.6088 + 0.5633\mathbf{standLRT}_i$. Notice this is the same regression line we found by using the **coef(m1)** function.

`fixef(m1)`	
(Intercept)	standLRT
0.002323916	0.563307940

Table 2.3: Fixed effects of variable intercept model of the exam data.

`ranef(m1)`	
School	(Intercept)
28	-0.6111046
53	0.7245383
59	-0.6597868
23	-0.4908383
7	0.3795565
16	-0.4104365
2	0.5030942
10	-0.3376540
37	-0.1898233

Table 2.4: Random effects of variable intercept model of the exam data – limited to nine schools.

2.2 Variable Slope Model

The second case to consider is one in which the slope varies between groups. This model assumes the same intercept for all groups, but allows the slope to vary:

$$Y_{ij} = \beta_0 + \beta_{1j}x_{1j} + r_{ij}. \tag{2.7}$$

Similar to the variable intercept model (2.2), we can split the group dependent slope,β_{1j}, into an average coefficient and a group-dependent deviation:

$$\beta_{1j} = \gamma_{10} + u_{1j}. \tag{2.8}$$

Substituting (2.8) into (2.7) yields

$$Y_{ij} = \beta_0 + \gamma_{10}x_{ij} + u_{1j}x_{ij} + r_{ij}. \tag{2.9}$$

Again, the final step will be a notation change

$$Y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + u_{1j}x_{ij} + r_{ij}. \tag{2.10}$$

The reason for the notation change becomes clearer when (2.10) is compared to (2.4). A general form for the multilevel model is being fashioned.

The random effects, u_{1j}, are assumed to be independently identically distributed random variables with mean zero and unknown variance, and are also assumed to be independent of the residuals, r_{ij}. The first part of (2.10),$\gamma_{00} + \gamma_{10}x_{ij}$, is the fixed part of the model. The second part, $u_{1j}x_{ij} + r_{ij}$, is the random part. The term $u_{1j}x_{ij}$ shows a variable interaction between x and the groups.

Example

We continue with the exam example from Chapter 1 to illustrate the variable slope model. We again model the relationship between **normexam** and **standLRT**, this time using (2.7) and (2.8):

$$\mathbf{normexam}_{ij} = \beta_0 + \beta_{1j}\mathbf{standLRT}_{ij} + r_{ij} \quad (2.11)$$

$$\beta_{1j} = \gamma_{10} + u_{1j}. \quad (2.12)$$

In this example, β_0 and γ_{10} are the fixed effects, while u_{1j} are the random effects. Recall, the u_{1j} are assumed to be independently identically distributed random variables with mean zero and unknown variance, and are also assumed to be independent of the r_{ij}.

The variable slope model can be fit in R using the commands found in Listing 2.2.

```
#Model 2 - Variable Slope
m2 <- lmer(normexam ~ standLRT + (standLRT - 1 |
   school))
```

Listing 2.2: R code for producing variable slope model of exam data using lmer.

The call to **lmer** again starts with a relationship between **normexam** and **standLRT** and adds **(standLRT - 1 — school)**, which commands R to allow **standLRT** to vary by school, but not the intercept (**1**).

The output of this model is contained in Table 2.5. This is the same output format as found in Section 2.1. The fixed and random effects estimates

Linear mixed model fit by **REML**				
Formula: `normexam ~ standLRT + (standLRT - 1 | school)`				
AIC	BIC	logLik	deviance	REMLdev
9709	9734	-4850	9688	9701
Random effects:				
	Groups	Name	Variance	Std.Dev.
	school	standLRT	0.025902	0.16094
	Residual		0.625091	0.79063
Number of obs: 4059, **groups:** school, 65				
Fixed effects:				
		Estimate	Std. Error	t value
	(Intercept)	-0.01462	0.01284	-1.139
	standLRT	0.58782	0.02423	24.259

Table 2.5: R (lmer) output for variable slope model of exam data.

correspond to the coefficients in (2.11) and (2.12) as follows:

$$\widehat{\beta}_0 = -0.01462$$
$$\widehat{\gamma}_{10} = 0.58782$$
$$\widehat{\text{Var}}(u_{1j}) = 0.025902$$
$$\widehat{\text{Var}}(r_{ij}) = 0.625091.$$

Figure 2.2 shows the same nine schools we examined in Section 2.1 with their variable slope regression lines. Table 2.6 shows their corresponding coefficients. Again, we find the coefficients by using the **coef(m2)** function in R. It is clear in this model that the slopes are varying by school, while the intercepts remain constant. For example, the estimated regression line for school 28 is $\textbf{normexam}_i = -0.0146 + 0.5803\textbf{standLRT}_i$, while the estimate regression line for school 53 is $\textbf{normexam}_i = -0.0146 + 1.1235\textbf{standLRT}_i$.

We find the fixed effects and the random effects again using the **fixef(m2)**

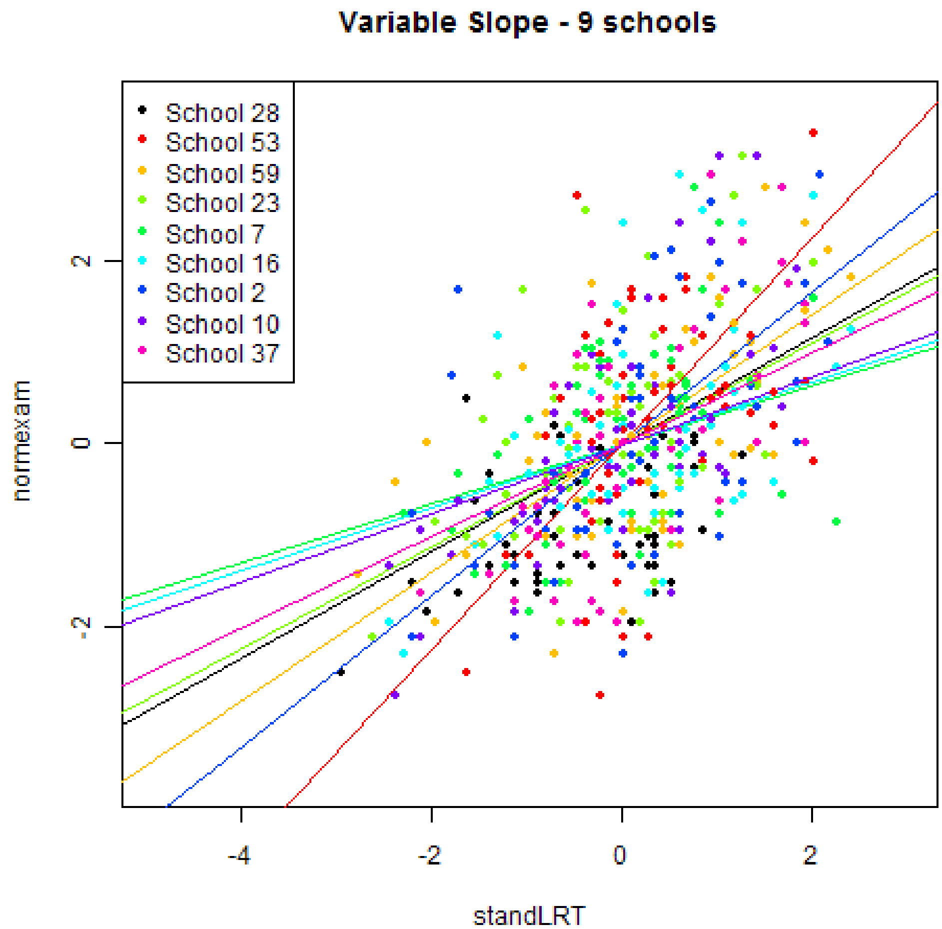

Figure 2.2: Variable slope model for exam data, limited to nine schools.

School	(Intercept)	standLRT
28	-0.01462441	0.5803311
53	-0.01462441	1.1235011
59	-0.01462441	0.7003846
23	-0.01462441	0.5548722
7	-0.01462441	0.3221863
16	-0.01462441	0.3408656
2	-0.01462441	0.8265095
10	-0.01462441	0.3717989
37	-0.01462441	0.4973572

Table 2.6: Variable slope model coefficients for exam data, limited to nine schools.

and **ranef(m2)** functions in R. Table 2.7 and Table 2.8 provide the results. Table 2.7 shows the fitted regression line for the average school is $\textbf{normexam}_i = -0.0146 + 0.5878\textbf{standLRT}_i$, while Table 2.8 shows the deviation of the slope for each school from the average slope. In school 28, the slope is

`fixef(m1)`	
(Intercept)	standLRT
-0.01462441	0.58782063

Table 2.7: Fixed effects of variable slope model of the exam data.

`ranef(m1)`	
School	standLRT
28	-0.0074895
53	0.5356804
59	0.1125639
23	-0.0329484
7	-0.2656343
16	-0.2469550
2	0.2386888
10	-0.2160217
37	-0.0904634

Table 2.8: Random effects of variable slope model of the exam data – limited to nine schools.

0.0074 less than the average, making the fitted regression line $\textbf{normexam}_i = -0.0146 + (0.5878 - 0.0075)\textbf{standLRT}_i = -0.0146 + .5803\textbf{standLRT}_i$. This matches the regression line found using **coef(m2)**.

2.3 Hierarchical Linear Model

In the previous sections, we introduced simpler cases of the hierarchical linear model. In this section a more generalized hierarchical linear model will be defined, where both intercepts and slopes will be allowed to vary.

In the variable intercept model, the groups differ with respect to the average value of the dependent variable; the only random effect is the intercept. In the variable slopes, model the relationship between explanatory and dependent variable differs between groups; the only random effect is the slope. In the hierarchical linear, model the intercept and slope are both random effects. The model equation is now:

$$Y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + r_{ij}. \tag{2.13}$$

We can split the intercepts, β_{0j}, as well as the slopes, β_{1j}, into an average coefficient and a group-dependent deviation:

$$\beta_{0j} = \gamma_{00} + u_{0j} \tag{2.14}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}. \tag{2.15}$$

Substituting (2.14) and (2.15) into (2.13) leads to the model

$$Y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + u_{0j} + u_{1j}x_{ij} + r_{ij}. \tag{2.16}$$

Once again this model can be organized into a fixed part, $\gamma_{00} + \gamma_{10}x_{ij}$, and a random part, $u_{0j} + u_{1j}x_{ij} + r_{ij}$. The residuals ($r_{ij}$, u_{0j} and u_{1j}) are still assumed normal with means equal to zero. It is assumed also that the random effects are identically distributed and between groups j and k each u_{0j} and u_{1j} are independent of each u_{0k} and u_{1k}, where $j \neq k$. Additionally, the random effects are independent of the level-one residuals r_{ij}, and all r_{ij} are independent and identically distributed.

Regression analysis aims to explain variability in the dependent variable from knowledge of the explanatory variables. Regression analysis and its variant, hierarchical (multilevel) modelling, both have the goal of explaining variability in the dependent variable from knowledge of the explanatory variables. However, the additional complexity in hierarchical models permits a more detailed explanation of variability. Hierarchical models take into account group differences (i.e., differences between sub-populations) and thus can be used to help explain differences between groups, as well as differences between individuals. For example, the simple linear regression model (1.1) does not give any information about differences between groups. In contrast, the hierarchical model (2.13), where intercepts as well as slopes are allowed to vary between groups, can help us to describe variability of slopes and intercepts (types of between group variability) as well as variability between individuals.

Example

Again, we continue the exam example from Chapter 1 to illustrate the hierarchical linear model. We model the relationship between **normexam** and **standLRT** using (2.13), (2.14) and (2.15) to give:

$$\mathbf{normexam}_{ij} = \beta_{0j} + \beta_{1j}\mathbf{standLRT}_{ij} + r_{ij} \tag{2.17}$$

$$\beta_{0j} = \gamma_{00} + u_{0j} \tag{2.18}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}. \tag{2.19}$$

In this model, γ_{00} and γ_{10} are the fixed effects, while u_{0j} and u_{1j} are the random effects. Recall, the residuals (r_{ij}, u_{0j} and u_{1j}) are assumed normal with means equal to zero. It is assumed also that the random effects are identically distributed and between groups j and k each u_{0j} and u_{1j} are independent of each u_{0k} and u_{1k}, where $j \neq k$. Additionally, the random effects are assumed independent of the level-one residuals r_{ij}, and all r_{ij} are independent and identically distributed.

The hierarchical linear model is fit in R using the commands in Listing 2.3.

```
##Model 3 - HLM
m3 <-lmer(normexam ~ standLRT + (standLRT | school))
```

Listing 2.3: R code for producing hierarchical linear model of exam data using lmer.

The call to **lmer** again starts with a relationship between **normexam** and **standLRT** and adds **(standLRT — school)**. This addition allows the coefficient of **standLRT** to vary by school. Additionally, since the intercept is not explicitly excluded, this allows it to vary by school.

Table 2.9 displays the output of the call to **lmer**. The output format is the same as displayed Sections 2.1 and 2.2. Table 2.9 gives the fixed and random effects that correspond to the coefficients in (2.17), (2.18), and (2.19)

Linear mixed model fit by **REML**				
Formula: `normexam ~ standLRT + (standLRT | school)`				
AIC	BIC	logLik	deviance	REMLdev
9340	9377	-4664	9317	9328
Random effects:				
	Groups	Name	Variance	Std.Dev.
	school	(Intercept)	0.092117	0.30351
		standLRT	0.014967	0.12234
	Residual		0.553641	0.74407
Number of obs: 4059, **groups:** school, 65				
Fixed effects:				
		Estimate	Std. Error	t value
	(Intercept)	-0.01165	0.04011	-0.29
	standLRT	0.55654	0.02011	27.67

Table 2.9: R (lmer) output for hierarchical linear model of exam data.

as follows:

$$\widehat{\gamma}_{00} = -0.01165$$
$$\widehat{\gamma}_{10} = 0.55654$$
$$\widehat{\text{Var}}(u_{0j}) = 0.092117$$
$$\widehat{\text{Var}}(u_{1j}) = 0.014967$$
$$\widehat{\text{Var}}(r_{ij}) = 0.553641.$$

Figure 2.3 shows the nine schools we examined in Sections 2.1 and 2.2 with their hierarchical linear model regression lines. Table 2.10 gives their corresponding regression coefficients, found using the **coef(m3)** command in R. As designed, the intercept and slope both vary by school. For example, the fitted regression line for school 28 is $\textbf{normexam}_i = -0.6808 + 0.3690\textbf{standLRT}_i$, while the fitted regression line for school 53 is $\textbf{normexam}_i = 0.6337418 + 0.9075559\textbf{standLRT}_i$.

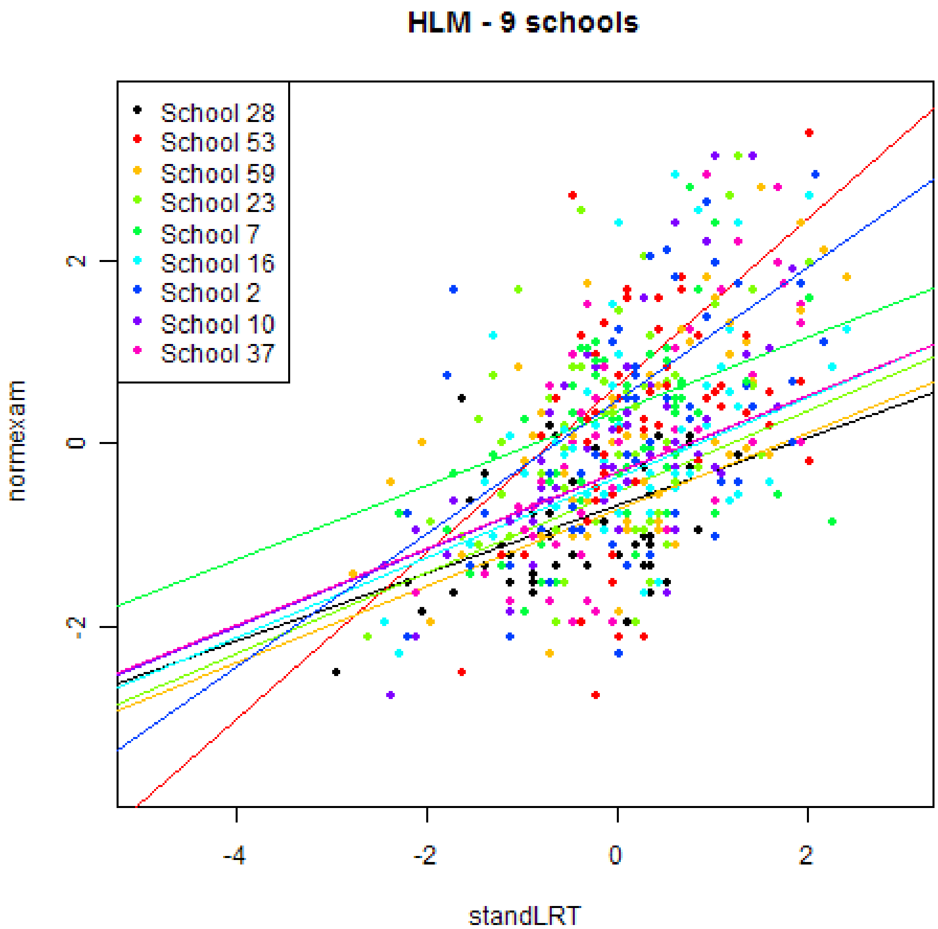

Figure 2.3: Hierarchical linear model for exam data, limited to nine schools.

School	(Intercept)	standLRT
28	-0.6807917	0.3690499
53	0.6337418	0.9075559
59	-0.7299502	0.4159633
23	-0.5211349	0.4405163
7	0.3534431	0.4046149
16	-0.3755861	0.4363625
2	0.4590305	0.7222757
10	-0.3255777	0.4190209
37	-0.3050309	0.4176832

Table 2.10: Hierarchical linear model coefficients for exam data, limited to nine schools.

Again, by using the **fixef(m3)** and **ranef(m3)** functions in R, we are able to display the fixed and random effects of our model. Tables 2.11 and 2.12, respectively, provide the results. Table 2.11 gives the fitted regression line for the average school as $\mathbf{normexam}_i = -0.0116 + 0.5565\mathbf{standLRT}_i$.

`fixef(m1)`	
(Intercept)	standLRT
-0.01164735	0.55653515

Table 2.11: Fixed effects of hierarchical model of the exam data.

`ranef(m1)`		
School	(Intercept)	standLRT
28	-0.6691444	-0.1874853
53	0.6453891	0.3510208
59	-0.7183029	-0.1405719
23	-0.5094876	-0.1160188
7	0.3650904	-0.1519202
16	-0.3639388	-0.1201726
2	0.4706778	0.1657405
10	-0.3139303	-0.1375143
37	-0.2933836	-0.1388519

Table 2.12: Random effects of hierarchical linear model of the exam data – limited to nine schools.

Table 2.12 gives the deviation for the regression coefficients of a particular school from the average. For example, school 28 is 0.6691 lower than the average intercept and 0.1875 lower than the average slope, giving a fitted regression line **normexam** $= (-0.0116 - 0.6691) + (0.5565 - 0.1875)$**standLRT**$_i$ $=$ $-.6807 + .3690$**standLRT**$_i$. Notice this is equal to the regression line found using the **coef(m3)** function (with rounding error).

Chapter 3

Confidence Intervals of Variable Level-1 Coefficients

A confidence interval for the variable level-1 coefficients (βs) will be constructed by looking at the general case of a hierarchical linear model. The purpose of this chapter is to establish a basis by which one can compare the hierarchical linear model with simple linear regression, variable intercept, and variable slope models. With this tool established, we will be able to evaluate whether the models produce different results.

General Model

The general Level-1 model with Q predictors can be expressed in matrix notation as

$$\mathbf{Y}_j = \mathbf{X}_j\beta_j + \mathbf{r}_j, \qquad \mathbf{r}_j \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I}) \tag{3.1}$$

where

$\mathbf{Y}_j$ is an n_j by 1 vector of outcomes,

$\mathbf{X}_j$ is an n_j by $(Q+1)$ matrix of Level-1 predictors,

β_j is a $(Q+1)$ by 1 vector of unknown coefficients,

$\mathbf{I}$ is an n_j by n_j identity matrix, and

$\mathbf{r}_j$ is an n_j by 1 vector of random errors.

For example, we can express the hierarchical linear model (2.13) in the general model notation (3.1),

$$Y_{ij} = \begin{bmatrix} Y_{1j} \\ \vdots \\ Y_{n_j j} \end{bmatrix} = \mathbf{Y}_j, \begin{bmatrix} 1 & x_{1j} \\ \vdots & \vdots \\ 1 & x_{n_j j} \end{bmatrix} = \mathbf{X}_j, \begin{bmatrix} \beta_{0j} \\ \beta_{1j} \end{bmatrix} = \beta_j, r_{ij} = \begin{bmatrix} r_{1j} \\ \vdots \\ r_{n_j j} \end{bmatrix} = \mathbf{r}_j.$$

To illustrate equivalence of the two notations, start with the general model (3.1)

$$\mathbf{Y}_j = \mathbf{X}_j\beta_j + \mathbf{r}_j$$

$$\begin{bmatrix} Y_{1j} \\ \vdots \\ Y_{n_j j} \end{bmatrix} = \begin{bmatrix} 1 & x_{1j} \\ \vdots & \vdots \\ 1 & x_{n_j j} \end{bmatrix} \begin{bmatrix} \beta_{0j} \\ \beta_{1j} \end{bmatrix} + \begin{bmatrix} r_{1j} \\ \vdots \\ r_{n_j j} \end{bmatrix}$$

$$\begin{bmatrix} Y_{1j} \\ \vdots \\ Y_{n_j j} \end{bmatrix} = \begin{bmatrix} \beta_{0j} + \beta_{1j}x_{1j} \\ \vdots \\ \beta_{0j} + \beta_{1j}x_{n_j j} \end{bmatrix} + \begin{bmatrix} r_{1j} \\ \vdots \\ r_{n_j j} \end{bmatrix}$$

$$\begin{bmatrix} Y_{1j} \\ \vdots \\ Y_{n_j j} \end{bmatrix} = \begin{bmatrix} \beta_{0j} + \beta_{1j}x_{1j} + r_{1j} \\ \vdots \\ \beta_{0j} + \beta_{1j}x_{n_j j} + r_{n_j j} \end{bmatrix}$$

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

which is the Level-1 model equation for the hierarchical linear model (2.13).

In (3.1), it is stated that $\mathbf{r}_j \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I})$, meaning $\mathbf{r}_j$ is assumed multivariate normally distributed with mean vector $\mathbf{0}$ and variance-covariance matrix with diagonal elements equal to σ^2 and non-diagonal elements equal to 0 [4, pg. 37].

The OLS estimator of β_j is

$$\widehat{\beta}_j = (\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j'\mathbf{Y}_j[6, \text{pg. } 40] \tag{3.2}$$

with dispersion measured by the variance-covariance matrix

$$\mathbf{V}_j = \text{Var}(\widehat{\beta}_j) = \sigma^2(\mathbf{X}_j'\mathbf{X}_j)^{-1}[6, \text{pg. } 40], \tag{3.3}$$

where $\mathbf{X}'$ denotes the transpose of $\mathbf{X}$. The term *dispersion* refers to the variability or spread in a variable. Multiplying (3.1) by $(\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j'$ yields

$$\begin{aligned}(\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j'\mathbf{Y}_j &= (\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j'\mathbf{X}_j\beta_j + (\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j'\mathbf{r}_j \\ \widehat{\beta}_j &= \beta_j + (\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j'\mathbf{r}_j \\ \widehat{\beta}_j &= \beta_j + \mathbf{e}_j, \qquad (3.4)\end{aligned}$$

where $\mathbf{e}_j = (\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j'\mathbf{r}_j$ is normally distributed with mean $\mathbf{0}$ and variance equal to $\mathbf{V}_j$. In order to show this, a property of variance of random vectors will be presented and proven.

Property 1. *Let A be an $m \times n$ matrix of constants, and Y be an $n \times 1$ random vector. Then $Var[AY] = AVar[Y]A'$.*

Proof.

$$\begin{aligned}\text{Var}[AY] &= \text{E}[(AY - \text{E}[AY])(AY - \text{E}[AY])'][12, \text{pg. } 613, \text{Definition B.13}] \\ &= \text{E}[(AY - A\text{E}[Y])(AY - A\text{E}[Y])'][\text{Lemma 1.a}] \\ &= \text{E}[A(Y - \text{E}[Y])(A(Y - \text{E}[Y])))'] \\ &= A\text{E}[(Y - \text{E}[Y])(Y - \text{E}[Y])']A'[\text{Lemma 1.b}] \\ &= A\text{Var}[Y]A'\end{aligned}$$

□

Lemma 1. *Let A be an $m \times n$ matrix of constants, and Y be an $n \times 1$ random vector. Then a) E[AY] = AE[Y] and b) E[YA] = E[Y]A.*

Proof.

$$\mathrm{E}[AY] = \mathrm{E}\left[\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \ldots & a_{mn} \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}\right]$$

$$= \mathrm{E}\left[\begin{bmatrix} a_{11}y_1 + \cdots + a_{1n}y_n \\ \vdots \\ a_{m1}y_1 + \cdots + a_{mn}y_n \end{bmatrix}\right]$$

$$= \begin{bmatrix} \mathrm{E}[a_{11}y_1 + \cdots + a_{1n}y_n] \\ \vdots \\ \mathrm{E}[a_{m1}y_1 + \cdots + a_{mn}y_n] \end{bmatrix} \text{[14, pg. 45, Definition 2.1.1]}$$

$$= \begin{bmatrix} a_{11}\mathrm{E}[y_1] + \cdots a_{1n}\mathrm{E}[y_n] \\ \vdots \\ a_{m1}\mathrm{E}[y_1] + \cdots a_{mn}\mathrm{E}[y_n] \end{bmatrix} \text{[5, pg. 57]}$$

$$= \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} \mathrm{E}[y_1] \\ \vdots \\ \mathrm{E}[y_n] \end{bmatrix}$$

$$= A\mathrm{E}[Y] \text{[14, pg. 45, Definition 2.1.1]}$$

An analogous calculation shows E[YA] = E[Y]A.

□

Thus,

$$\begin{aligned}
\mathrm{E}[\mathbf{e}_j] &= \mathrm{E}[(\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j'\mathbf{r}_j] \\
&= (\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j'E[\mathbf{r}_j] \\
&= (\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j'\mathbf{0} \\
&= \mathbf{0}
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Var}[\mathbf{e}_j] &= \mathrm{Var}[(\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j'\mathbf{r}_j] \\
&= (\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j'Var[\mathbf{r}_j][(\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j']' \\
&= (\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j'\sigma^2\mathbf{I}[(\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j']' \\
&= \sigma^2(\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j'\mathbf{X}_j(\mathbf{X}_j'\mathbf{X}_j)^{-1} \\
&= \sigma^2(\mathbf{X}_j'\mathbf{X}_j)^{-1} \\
&= \mathbf{V}_j.
\end{aligned}$$

Since normality is preserved under linear transformations [8, pg. 2, Proposition 1.1], $\mathbf{e}_j \sim \mathrm{N}(\mathbf{0}, \mathbf{V}_j)$.

The Level-2 general model for β_j is

$$\beta_j = \mathbf{W}_j\gamma + \mathbf{u}_j \tag{3.5}$$

where

$\mathbf{W}_j$ is a $(Q+1)$ by F matrix of Level-2 predictors,

γ is an F by 1 vector of fixed effects,

$\mathbf{u}_j$ is a $(Q+1)$ by 1 vector of random effects,

$\mathbf{u}_j \sim \mathrm{N}(\mathbf{0}, \mathbf{T})$,

$\mathbf{u}_j$ is independent of $\mathbf{r}_j$.

$\mathbf{W}_j$ is constructed by stacking the $Q+1$ row vectors of predictors in a block diagonal fashion where each row corresponds to one of the $Q+1$ Level-2 outcome variables, β_j [4, pg. 37].

Again, we can express the hierarchical linear model (2.14) and (2.15) in the notation of the general model notation (3.5),

$$\begin{bmatrix} \beta_{0j} \\ \beta_{1j} \end{bmatrix} = \beta_j, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{W}_j, \begin{bmatrix} \gamma_{00} \\ \gamma_{10} \end{bmatrix} = \gamma, \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} = \mathbf{u}_j.$$

To show equivalence of the two notations, start with the general model (3.5)

$$\beta_j = \mathbf{W}_j \gamma + \mathbf{u}_j$$

$$\begin{bmatrix} \beta_{0j} \\ \beta_{1j} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \gamma_{00} \\ \gamma_{10} \end{bmatrix} + \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix}$$

$$\begin{bmatrix} \beta_{0j} \\ \beta_{1j} \end{bmatrix} = \begin{bmatrix} \gamma_{00} \\ \gamma_{10} \end{bmatrix} + \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix}$$

$$\begin{bmatrix} \beta_{0j} \\ \beta_{1j} \end{bmatrix} = \begin{bmatrix} \gamma_{00} + u_{0j} \\ \gamma_{10} + u_{1j} \end{bmatrix}$$

$$\beta_{0j} = \gamma_{00} + u_{0j} \text{ and } \beta_{1j} = \gamma_{10} + u_{1j}$$

which are the Level-2 model equations for the hierarchical linear model (2.14) and (2.15).

Using our exam example from Chapter 1, consider adding a Level-2 predictor of student-teacher ratio. This is a school-level predictor because the ratio is equal for each student within a school. We call our student-teacher ratio **STratio** and by using a hierarchical linear model, obtain the following model equations:

$$\textbf{normexam}_{ij} = \beta_{0j} + \beta_{1j}\textbf{standLRT}_{ij} + r_{ij} \quad (3.6)$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\textbf{STratio}_j + u_{0j} \quad (3.7)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}\textbf{STratio}_j + u_{1j}. \quad (3.8)$$

Using the general model notation in this example yields

$$\beta_j = \begin{bmatrix} \beta_{0j} \\ \beta_{1j} \end{bmatrix}, \gamma = \begin{bmatrix} \gamma_{00} \\ \gamma_{01} \\ \gamma_{10} \\ \gamma_{11} \end{bmatrix}, \mathbf{W}_j = \begin{bmatrix} 1 & \textbf{STratio}_j & 0 & 0 \\ 0 & 0 & 1 & \textbf{STratio}_j \end{bmatrix},$$

and $\mathbf{u}_j = \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix}$.

Thus, the Level-2 general model (3.5) $\beta_j = \mathbf{W}_j\gamma + \mathbf{u}_j$ yields

$$\begin{bmatrix} \beta_{0j} \\ \beta_{1j} \end{bmatrix} = \begin{bmatrix} 1 & \textbf{STratio}_j & 0 & 0 \\ 0 & 0 & 1 & \textbf{STratio}_j \end{bmatrix} \begin{bmatrix} \gamma_{00} \\ \gamma_{01} \\ \gamma_{10} \\ \gamma_{11} \end{bmatrix} + \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix}$$

$$\begin{bmatrix} \beta_{0j} \\ \beta_{1j} \end{bmatrix} = \begin{bmatrix} \gamma_0 0 + \gamma_{01}\textbf{STratio}_j \\ \gamma_1 0 + \gamma_{11}\textbf{STratio}_j \end{bmatrix} + \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix}$$

$$\begin{bmatrix} \beta_{0j} \\ \beta_{1j} \end{bmatrix} = \begin{bmatrix} \gamma_0 0 + \gamma_{01}\textbf{STratio}_j + u_{0j} \\ \gamma_1 0 + \gamma_{11}\textbf{STratio}_j + u_{1j} \end{bmatrix}.$$

Finally, $\beta_{0j} = \gamma_0 0 + \gamma 01\textbf{STratio}_j + u_{0j}$ and $\beta_{1j} = \gamma_1 0 + \gamma_{11}\textbf{STratio}_j + u_{1j}$, which match Level-2 of our model (3.7) and (3.8).

Assuming this 2-Level model, plugging (3.5) into (3.4) yields the combined model

$$\widehat{\beta}_j = \mathbf{W}_j\gamma + \mathbf{u}_j + \mathbf{e}_j, \quad (3.9)$$

with variance-covariance matrix

$$\mathrm{Var}[\widehat{\beta}_j] \quad = \quad \mathrm{Var}[\mathbf{W}_j\gamma + \mathbf{u}_j + \mathbf{e}_j].$$

The following derivation of $\mathrm{Var}[\widehat{\beta}_j]$ depends on Property 2.

Property 2. *Let X and Y be random n × 1 random vectors. Then Var(X + Y) = Var(X) + Var(Y) + 2Cov(X,Y) [3, pg. 504, Property B5.11].*

Since $\mathbf{W}_j$ and γ are constant, $\mathrm{Var}[\widehat{\beta}_j] = \mathrm{Var}[\mathbf{u}_j + \mathbf{e}_j]$, which equals $\mathrm{Var}[\mathbf{u}_j]+\mathrm{Var}[\mathbf{e}_j]+2\mathrm{Cov}[\mathbf{u}_j, \mathbf{e}_j]$ by Property 2. From (3.4), $\mathbf{e}_j$ can be expressed as a linear combination of $\mathbf{r}_j$: $\mathbf{e}_j = (\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j'\mathbf{r}_j$. We know $\mathbf{r}_j$ and $\mathbf{u}_j$ are assumed independent, which implies $\mathrm{Cov}[\mathbf{r}_j, \mathbf{u}_j] = 0$ [14, pg. 49]. It follows:

$$\begin{aligned}\mathrm{Cov}[\mathbf{e}_j, \mathbf{u}_j] &= \mathrm{Cov}[(\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j'\mathbf{r}_j, \mathbf{u}_j] \\ &= (\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j'\mathrm{Cov}[\mathbf{r}_j, \mathbf{u}_j][14, pg.48] \\ &= 0.\end{aligned}$$

Thus,

$$\begin{aligned}\mathrm{Var}[\widehat{\beta}_j] &= \mathrm{Var}[\mathbf{u}_j] + \mathrm{Var}[\mathbf{e}_j] \\ &= \mathbf{T} + \mathbf{V}_j \equiv \mathbf{\Delta}_j \qquad (3.10)\end{aligned}$$

The unique, minimum-variance, unbiased estimator of γ is the generalized least squares (GLS) estimator [4, pg. 38]

$$\widehat{\gamma} = \left(\sum_j \mathbf{W}_j'\mathbf{\Delta}_j^{-1}\mathbf{W}_j\right)^{-1} \sum_j \mathbf{W}_j'\mathbf{\Delta}_j^{-1}\widehat{\beta}_j \qquad (3.11)$$

with variance-covariance matrix

$$\mathrm{Var}(\widehat{\gamma}) = \left(\sum_j \mathbf{W}_j\mathbf{\Delta}_j^{-1}\mathbf{W}_j\right)^{-1}. \qquad (3.12)$$

The GLS estimator weights each group's data by the inverse of its variance-covariance matrix. This weight is called the precision matrix and is denoted $\mathbf{\Delta}_j^{-1}$. By weighting each group's data by its precision matrix, the GLS estimator places more weight on the groups whose data are more closely estimated by $\widehat{\beta}_j$.

Shrinkage Estimator

In order to find an optimal estimator for β_j, two common estimators can be identified. By substituting $\widehat{\gamma}$ for γ in (3.5), we can see that $\mathbf{W}_j\widehat{\gamma}$ is an estimate for β_j. Alternatively, β_j can also be estimated by its OLS estimator, $\widehat{\beta}_j$. We define a third estimator, $\beta_j^\star$ by

$$\beta_j^\star = \mathbf{\Lambda}_j\widehat{\beta}_j + (\mathbf{I} - \mathbf{\Lambda}_j)\mathbf{W}_j\widehat{\gamma} \tag{3.13}$$

where

$$\mathbf{\Lambda}_j = \mathbf{T}(\mathbf{T} + \mathbf{V}_j)^{-1} = \mathbf{T}(\mathbf{\Delta}_j)^{-1}. \tag{3.14}$$

In Bayesian statistics, a parameter is estimated by combining a subjective distribution based on the experimenter's knowledge (prior distribution) and a sample of the population (likelihood function) to form the posterior distribution using Bayes' theorem [5, pg. 324]. By definition, the Bayes estimator is the mean of the posterior distribution for a particular parameter.

A loss function represents the loss associated with an estimate being wrong as a function of a measure of the degree of wrongness. A common loss function is the squared error loss function, which looks at the square of the difference between an estimate and the parameter it estimates. The Bayes Risk of an estimator is defined as the expected value of a loss function evaluated at the estimator taken of the posterior distribution of the parameter. Bayes estimators are optimal according to the criterion of minimizing the Bayes risk with a squared error loss function [5, pg. 352].

$\mathbf{\Lambda}_j$ is the ratio of the dispersion of β_j about $\mathbf{W}_j\gamma$, $\mathbf{T}$, in relation to the dispersion of $\widehat{\beta}_j$, $\mathbf{\Delta}_j$. When $\widehat{\beta}_j$ is a precise estimator of β_j, $\mathbf{V}_j$ is "small", so (3.14) suggests that $\mathbf{\Lambda}_j$ us close to 1. Thus, in this case, (3.13) shows that $\beta_j^\star$ is weighted (or "shrunk") toward $\widehat{\beta}_j$. Similarly, when the precision of $\widehat{\beta}_j$ as an estimator of β_j is large, $\mathbf{V}_j$ will be "large", suggesting from (3.14) that $\mathbf{\Lambda}_j$ is small, and hence from (3.13) that $\beta_j^\star$ is "shrunk" toward $\widehat{\gamma}$. Thus, $\beta_j^\star$ is known as a *shrinkage estimator*.

Confidence Interval

A 95% confidence interval for β_{qj} is given by

$$95\%\ \mathrm{CI}(\beta_{qj}) = \beta_{qj}^\star \pm 1.96(V_{qqj}^\star)^{1/2} \qquad (3.15)$$

where $V_{qqj}^\star$ is the qth diagonal element of $\mathbf{V_j^\star}$ [4, pg. 44] defined as

$$\mathbf{V_j^\star} = (\mathbf{V_j}^{-1} + \mathbf{T}^{-1})^{-1} + (\mathbf{I} - \mathbf{\Lambda}_j)[\mathrm{Var}(\mathbf{W}_j\widehat{\gamma})](\mathbf{I} - \mathbf{\Lambda}_j)'. \qquad (3.16)$$

Example

We continue with the student exam data from Chapter 1 and the hierarchical linear model example from Section 2.3 to find 95% confidence intervals of the Level-1 coefficients β_{0j} and β_{1j}. Using (3.15) and (3.16), we have created a user-defined function (**ci.mer.ak**) in R that computes the confidence intervals for the Level-1 predictors across all groups. The full function definition can be found in Appendix .

Listing 3.1 displays the commands in R used to call **ci.mer.ak**.

```
##Load CI UDF
source("(file␣location)/ci.mer.ak.R")

##Confidence Intervals
m3.ci<-ci.mer.ak(obj=m3,grp="school")
```

Listing 3.1: R code for producing 95% confidence interval of exam data hierarchical linear model Level-1 regression coefficients using user-defined function ci.mer.ak.

Note, by default the function calculates 95% confidence intervals , however by adding "**conf** $= 1 - \alpha$" to the call a confidence level of $1 - \alpha$ is computed.

	β_{0j}		β_{1j}	
School	Lower Bound	Upper Bound	Lower Bound	Upper Bound
28	-0.8878054	-0.47379708	0.2082923	0.5298633
53	0.4517852	0.81569831	0.7667573	1.0482865
59	-0.9631566	-0.49676441	0.2501228	0.5818648
23	-0.7864038	-0.25587192	0.2661476	0.6149199
7	0.1877281	0.51913386	0.2630274	0.5461622
16	-0.5424697	-0.20869503	0.2950651	0.5776969
2	0.2606685	0.65738273	0.5858264	0.8586744
10	-0.5296247	-0.12153315	0.2502329	0.5878300
37	-0.6232708	0.01316396	0.2321296	0.6032463

Table 3.1: R (ci.mer.ak) output for 95% confidence intervals of Level-1 hierarchical linear model regression coefficients for exam data – limited to nine schools.

Table 3.1 displays the output of computed 95% confidence intervals for nine of schools. Figure 3.1 shows the 95% confidence bands plotted for each school compared with $\widehat{\gamma}_{00}$, $\widehat{\beta}_{0j}$, $\widehat{\gamma}_{10}$, and $\widehat{\beta}_{1j}$, respectively. Note that the confidence band for school 48 extends beyond the other schools. This is due to the fact that school 48 only contains two students, which forces a wider confidence interval.

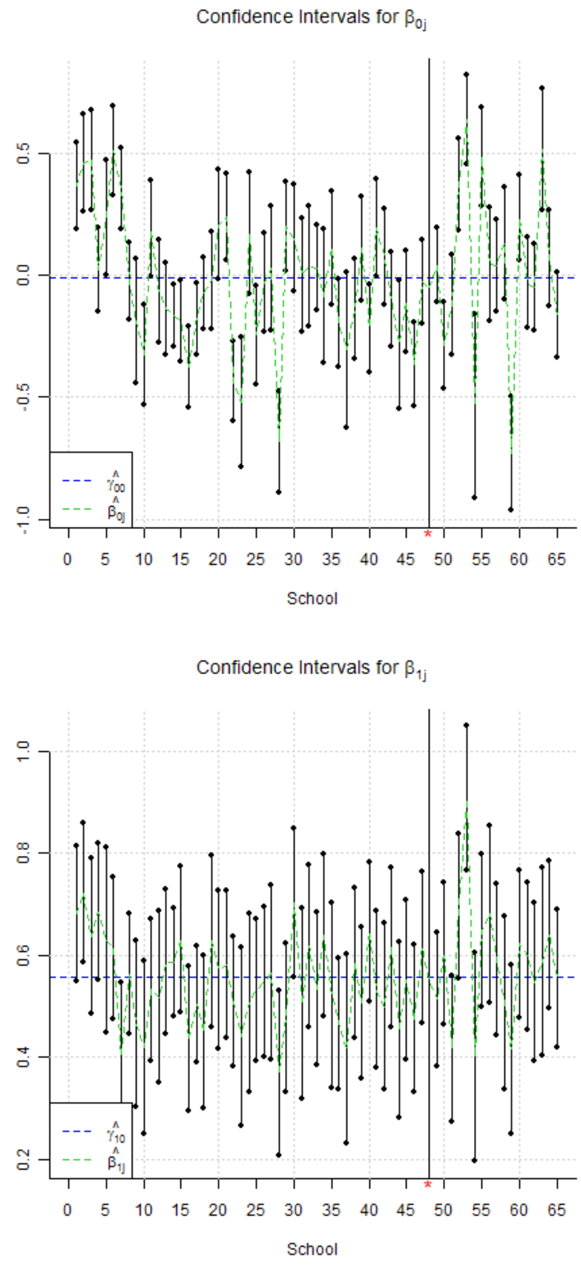

Figure 3.1: Confidence intervals for Level-1 regression coefficients of hierarchical linear model (*Note: for School 48, $n_{48} = 2$).

By definition, the simple linear regression, variable intercept, and variable slope models hold either one or both of β_{0j} and β_{1j} constant. With that in mind, compare the confidence intervals of schools 2 and 28 in Figure 3.1. The confidence intervals for the two schools do not overlap in either coefficient. From this, we can say with 95% confidence that the two schools do not have the same Level-1 coefficients, β_{0j} and β_{1j}. Thus, it stands to reason that the simple linear regression, variable intercept, and variable slope models would not compare favorably with the hierarchical linear model.

Chapter 4

Model Comparison

Chapter 1 and Sections 2.1, 2.2, and 2.3 contain explorations of the simple linear regression, variable intercept, variable slope, and hierarchical linear models, respectively. In each section an illustrative example was presented using the exam dataset from Chapter 1. Now we will highlight the differences between the models using that same example.

Figure 4.1 presents a graphical view of nine of the schools in the population with the four regression models overlaid. The *a*) solid black line represents the hierarchical linear model, (2.13), (2.14), and (2.15); *b*) dotted gray line represents the simple linear regression model, (1.1) *c*) dotted blue line represents the variable intercept model, (2.1) and (2.2); and *d*) dotted red line represents the variable slope model, (2.7) and (2.8). It appears the model most similar to the hierarchical linear model in this subset of the population of schools is the variable intercept model. The simple linear regression model appears to be the least similar.

In order to explore this further, let us examine the 95% confidence intervals calculated in Chapter 3. Table 4.1 contains the hierarchical linear model Level-1 regression coefficient confidence intervals, as well as the Level-1 regression coefficients for the simple linear regression, variable intercept, and variable slope models obtained from the **coef** function. The **lmer** and the **ci.mer.ak** functions provide the same estimates for β_{0j} and β_{1j}. This allows us

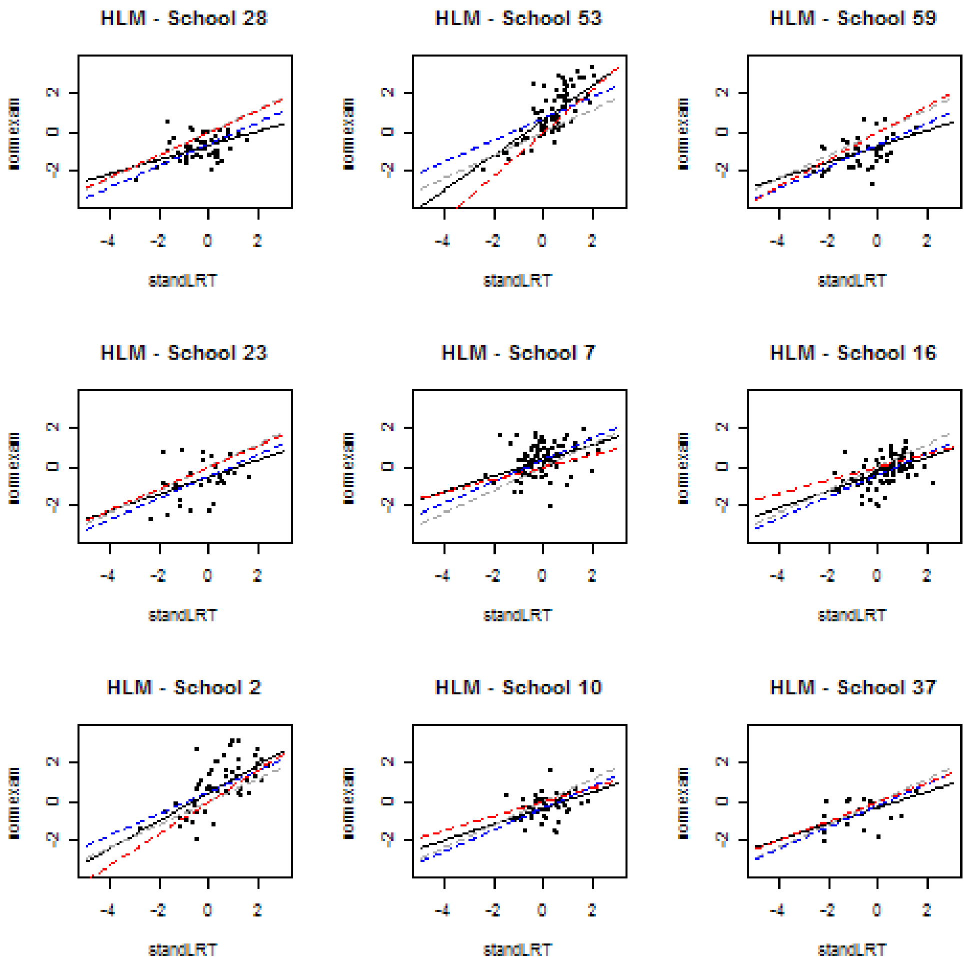

Figure 4.1: Hierarchical linear model (black solid line) compared with OLS (gray dotted line), variable intercept (blue dotted line), and variable slope (red dotted line) models using exam dataset – limited to nine schools.

to evaluate the coefficients from the other models against the 95 % confidence intervals for the hierarchical linear model Level-1 regression coefficients using **ci.mer.ak**. Those coefficients which do not lie within the confidence interval

	School	Lower Bound	Upper Bound	OLS	Var. Interc.	Var. Slope
β_{0j}	28	-0.8878054	-0.47379708	-0.001191	-0.6087807	-0.01462441
	53	0.4517852	0.81569831	-0.001191	0.7268622	-0.01462441
	59	-0.9631566	-0.49676441	-0.001191	-0.6574629	-0.01462441
	23	-0.7864038	-0.25587192	-0.001191	-0.4885144	-0.01462441
	7	0.1877281	0.51913386	-0.001191	0.3818804	-0.01462441
	16	-0.5424697	-0.20869503	-0.001191	-0.4081126	-0.01462441
	2	0.2606685	0.65738273	-0.001191	0.5054181	-0.01462441
	10	-0.5296247	-0.12153315	-0.001191	-0.3353301	-0.01462441
	37	-0.6232708	0.01316396	-0.001191	-0.1874994	-0.01462441
β_{1j}	28	0.2082923	0.52986332	0.595056	0.5633079	0.58033109
	53	0.7667573	1.04828648	0.595056	0.5633079	1.12350108
	59	0.2501228	0.58186483	0.595056	0.5633079	0.70038461
	23	0.2661476	0.61491993	0.595056	0.5633079	0.55487218
	7	0.2630274	0.54616221	0.595056	0.5633079	0.32218627
	16	0.2950651	0.57769689	0.595056	0.5633079	0.34086559
	2	0.5858264	0.85867441	0.595056	0.5633079	0.82650953
	10	0.2502329	0.58782997	0.595056	0.5633079	0.37179892
	37	0.2321296	0.60324631	0.595056	0.5633079	0.49735720

Table 4.1: A comparison of the 95% confidence intervals of Level-1 hierarchical linear model regression coefficients for exam data with OLS, variable intercept, and variable slope Level-1 coefficients – limited to nine schools. Red indicates the coefficient estimates do not fall within the HLM confidence interval.

are highlighted in red. Looking at the coefficients this way allows us to notice for every school except school 37, the simple linear regression model regression coefficients do not lie within the 95% confidence interval for the hierarchical linear model coefficients. With such a disparity between the models, it is clear that applying a multilevel model (in this case, the hierarchical linear model) versus a complete-pooling model will result in substantially different results.

We can examine the standard deviation of the residuals in each model as a measure of how closely each model fits the exam data. The residuals in a model indicate the difference between the data and the regression line. Looking at the standard deviation of the residuals will provide a measure of the average distance between model and data. Thus, a smaller standard deviation of the residuals will indicate a closer fit.

Table 4.2 displays the standard deviation of the residuals for each of our four fitted models. The simple linear regression model has the highest residual

	OLS	Var. Interc.	Var. Slope	HLM
St. Dev of Residuals	0.80534	0.75224	0.79063	0.74407

Table 4.2: A comparison of the standard deviation of the residuals for the simple linear regression, variable intercept, variable slope, and hierarchical linear models applied to the exam data.

standard deviation (0.80534) while the hierarchical linear model has the lowest (0.74407). These findings fit with our earlier conclusions that the two models are very different and goes a step further to indicate the hierarchical linear model is the closest fit to the data. We would expect the multilevel model to have a closer fit because of the ability to investigate relationships within and between groups simultaneously, and our expectations are confirmed in this case.

Appendix

Appendix

Confidence Interval R Function

```
# Function to estimate CI from Bryk & Raudenbush (37-44) method using mer object
# Andee Kaplan
# University of Texas at Austin
# Created November 8, 2010

ci.mer.ak <- function(obj,grp,conf = 0.95) {
  ##Dependent Packages
  require("MASS", quietly=TRUE)

  ##Variable Definitions
  error <- qnorm(1-0.5*(1-conf))

  fix <- fixef(obj)
  ran <- ranef(obj)



  X <- cbind(slot(obj, "X"),slot(obj, "flist")); X <- split(X,slot(obj, "flist"))
  y <- cbind(slot(obj, "y"),slot(obj, "flist")); y <- split(y,slot(obj, "flist"))

  grp.levs <- as.numeric(levels(slot(obj, "flist")[[grp]]))

  Q <- length(X[[1]]) - 2
  N <- length(grp.levs)

```

```
varcor <- VarCorr(obj)[[grp]]
sigma2 <- attr(VarCorr(obj),"sc")^2

bhat <- array(dim=c(Q + 1,1,N))
V <- array(dim=c(Q + 1,Q + 1,N))
T <- matrix(nrow= Q + 1, ncol= Q + 1); colnames(T)
   <- names(fix); rownames(T) <- names(fix)
Tinv <- matrix(nrow= Q + 1, ncol= Q + 1); colnames(
   Tinv) <- names(fix); rownames(Tinv) <- names(fix
   )
D <- array(dim=c(Q + 1,Q + 1,N))
W <- matrix(c(1,0,0,1),2) ## Note, this only works
   if no Level-2 Predictors
temp <- array(dim=c(Q + 1,1,N))
temp.sum <- array(0,dim=c(Q+1,1))
D.sum <- array(0,dim=c(Q + 1,Q + 1))
L <- array(dim=c(Q + 1,Q + 1,N))
I <- diag(dim(L)[1])
bstar <- array(dim=c(Q + 1,1,N))
Vstar <- array(dim=c(Q + 1,Q + 1,N))
output <- array(dim=c(N,3*(Q+1))); colnames(output)
    <- rep(c("L","Bstar","U"),(Q+1))

## OLS Estimator and Error Dispersion
for(j in grp.levs){
  bhat[,,j] <- solve(t(X[[j]][,1:2]) %*% as.matrix(
     X[[j]][,1:2])) %*% t(X[[j]][,1:2]) %*% y[[j
     ]][,1]
  V[,,j] <- sigma2 * solve(t(X[[j]][,1:2]) %*% as.
     matrix(X[[j]][,1:2]))
}

  ## Parameter Dispersion
for(i in colnames(varcor)) {
  for(j in rownames(varcor)) {
    T[j,i] <- varcor[j,i]
  }
```

```
}
T[is.na(T)] <- 0 ##Replace NAs with 0

## Dispersion of OLS Est. Given multilevel and Multivariate Reliability Matrix
for(k in 1:N) {
  D[,,k] <- T + V[,,k]
  D.sum <- D.sum + solve(D[,,k])

  temp[,,k] <- solve(D[,,k]) %*% bhat[,,k]
  temp.sum <- temp.sum + temp[,,k]

  L[,,k] <- T %*% solve(D[,,k])
}

## GLS Estimator
ghat <- solve(D.sum) %*% temp.sum

## Interval Estimation -- Fixed for singular matrices
for(j in 1:N) {
  bstar[,,j] <- L[,,j] %*% bhat[,,j] + (I - L[,,j]) %*% W %*% ghat
  Vstar[,,j] <- solve(solve(V[,,j]) + tryCatch(solve(T), error=function(e) trySolveT(varcor, Tinv))) + (I - D[,,j]) %*% solve(D.sum) %*% t(I - D[,,j])
}

## Define CIs as output
for(q in 1:(Q+1)) {
  for(j in 1:N) {
    output[j,(3*q-2):(3*q)] <- cbind(bstar[,,j][q] - error*sqrt(Vstar[q,q,j]), bstar[,,j][q], bstar[,,j][q] + error*sqrt(Vstar[q,q,j]))
  }
}
```

```
  return(output)
}

trySolveT <- function(varcor,Tinv) {
  for(i in colnames(varcor)) {
    for(j in rownames(varcor)) {
      Tinv[j,i] <- solve(varcor[j,i])
    }
  }
  Tinv[is.na(Tinv)] <- 0 ##Replace NAs with 0

  return(Tinv)
}
```

Listing .1: User defined function for computing confidence intervals of level-1 coefficients.

Bibliography

[1] Douglas Bates. *mlmRev: Examples from Multilevel Modelling Software Review*, 2008. R package version 0.99875-1.

[2] Douglas Bates and Martin Maechler. *lme4: Linear mixed-effects models using S4 classes*, 2010. R package version 0.999375-35.

[3] Peter J. Bickel and Kjella A. Doksum. *Mathematical Statistics*. Prentice Hall, 2nd edition edition, 2001.

[4] Anthony S. Bryk and Stephen W. Raudenbush. *Hierarchical Linear Models*. Sage Publications, London, 1992.

[5] George Casella and Roger L. Berger. *Statistical Inference*. Duxbury Thomson Learning, Pacific Grove, CA, 2002.

[6] Andrew Gelman and Jennifer Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, New York, 2007.

[7] H. Goldstein and J. Rasbash et al. A multilevel analysis of school examination results. *Oxford Review of Education*, 19:425–433, 1993.

[8] Takeaki Kariya and Hiroshi Kurata. *Generalized Least Squares*. John Wiley & Sons, Inc., Chichester, UK, 2004.

[9] Ita G G Kreft and Jan de Leeuw. *Introducing Multilevel Modeling*. Sage Publications, London, 1998.

[10] University of Bristol Centre for Multilevel Modelling. Datasets used in reviews. `http://www.cmm.bristol.ac.uk/learning-training/multilevel-m-software/exam.shtml`.

[11] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2010. ISBN 3-900051-07-0.

[12] Mark J. Schervish. *Theory of Statistics (Springer Series in Statistics)*. Springer-Verlag, New York, 1995.

[13] Tom Snijders and Roel Bosker. *Multilevel Analysis*. Sage Publications, London, 1999.

[14] James H. Stapleton. *Linear Statistical Models*. John Wiley & Sons, Inc., New York, 1995.

Vita

Andrea Jean Kaplan was born in Houston, Texas on 22 October 1985, the daughter of Richard H. Kaplan and Doreen A. Kaplan. She received a Bachelor of Science degree in Mathematics (Option: Applied Mathematics) from the University of Texas at Austin in 2006. She directly enrolled in the Graduate School at the University of Texas at Austin.

Permanent address: 7131 Wood Hollow Dr. Apt. 136
Austin, Texas 78731

This report was typeset with LaTeX[†] by the author.

[†]LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.