Review of the General Linear Model
Introduction to Hierarchical Data

Introduction to Multilevel
Models Workshop

University of Georgia:

Institute for Interdisciplinary Research in
Education and Human Development

Covered this Section

- Introduction to hierarchical data
- Review of general linear models
  - Regression and ANOVA under one umbrella
- Demonstration of hierarchical data

HIERARCHICAL DATA STRUCTURES

Hierarchical Data Structure

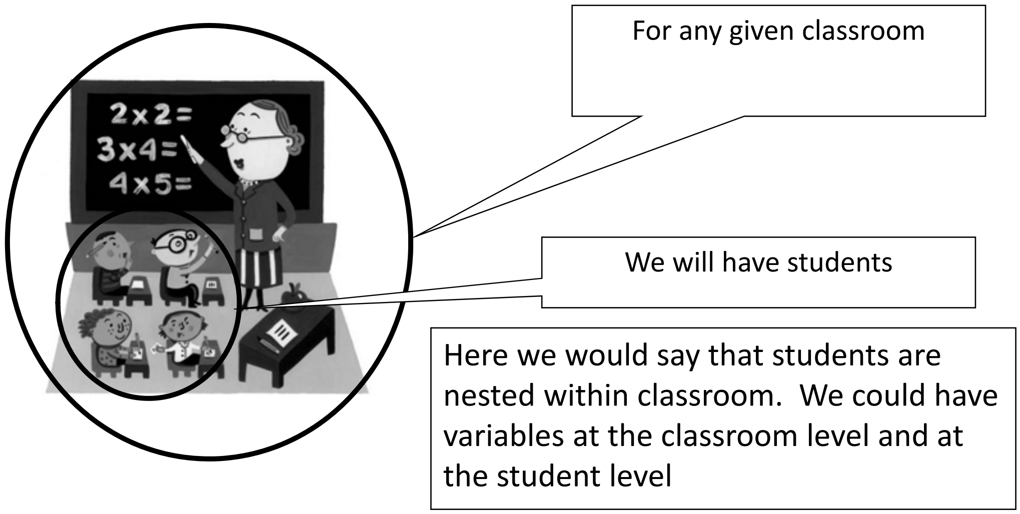
- In the social sciences and education there are many examples of *Hierarchical Data*

- Natural question: What are *Hierarchical Data?*
  - Two different dimensions of sampling units

  - One unit is nested within the other units
    - Later these units will be called levels

Hierarchical Data

- An example is when the two different units are classes and students



Hierarchical Data

- In this particular example we may be interested in
  - Performance of each student as predicted by a set of variables such as gender, socio-economic status (SES), …
- But it also makes sense that different classroom characteristics can have an impact

Hierarchical Data

- So we could have several different classes with different characteristics that impact the students differently

Hierarchical Data

- So lets assume that we can determine the effect of student socio-economic status and how it predicts classroom performance

Describes the student level information

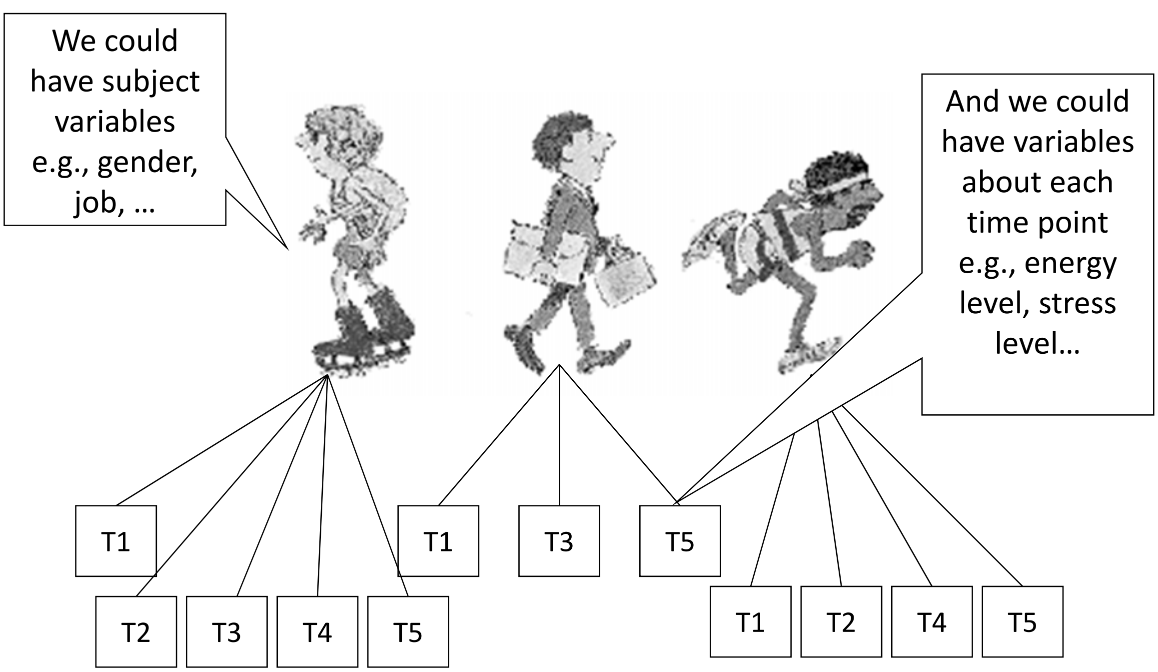
- It is also possible that there are additional effects that depend on how the class is taught

Teaching method is a variable at the classroom level

Hierarchical Data

- We can also use multilevel models (also known as hierarchical linear models or HLM) to model repeated measures (multiple outcomes per unit)
- What might be the sampling units in that case?

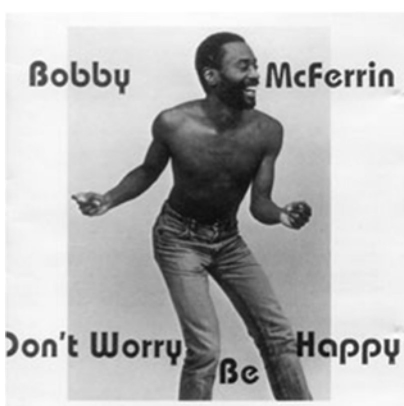
Hierarchical Data



Hierarchical Data

- A repeated measures model or even growth curves can be phrased as an HLM
  - Trials/observations are nested with in subjects

- In this case the subjects may be nested within classes
  - This would have three levels of data
    - Trial (1), Subject (2), Class (3)



WHY WORRY ABOUT HIERARCHICAL DATA?

Why HLM?

- What is the benefit of using such models that incorporate the hierarchical structure of data?
  - Improved estimation within units
  - Test cross-level effects
  - Partitioning of variance components

Improve Estimation

- There is improved estimation
- It is possible to "borrow" information
  - Giving a better or more precise estimate

Cross-Level Effects

- HLM allows us to explore the effects of variables at one level on variables at another level
  - i.e., the effect of classes (level 2) on student performance (level 1)

- This is particularly useful
  - Using regular regression/ANOVA models this capability is generally not possible

Partitioning of Variance

- HLM will allow us to estimate variance components with unbalanced and nested data
  - For example, we measure student performance in a class for many different schools

  - Now we can divide the variability of those observations:
    - Subject variability
    - School variability

Summary

- So hopefully up to this point you are starting to get a feel for what type of data and or how this would be useful in research

- Next, we will review regression and ANOVA under the context of the general linear model
  - These serve as the basis for multilevel models

REGRESSION, ANOVA, AND THE GLM

The General Linear Model

- Regression and Analysis of Variance (ANOVA) fall under the family of statistical techniques known more commonly as the General Linear Model (GLM)

- The GLM is a standard way to analyze data and test hypotheses using observed variables
  - Predictor/independent variables (X)
    - Many of these (multiple regression/multifactor ANOVA)
  - Outcome/criterion/dependent variable (Y)
    - One of these (univariate models – univariate conditional distribution of Y given **X**)

Guiding Example

- To guide us through the ins-and-outs of the GLM, we use a (hopefully relevant) example

- Imagine you are interested in studying the effects of socioeconomic status (SES) on student achievement
  - What do you think the relationship between student achievement and SES happens to be?

- You are interested in predicting achievement from SES
  - Your guiding research question

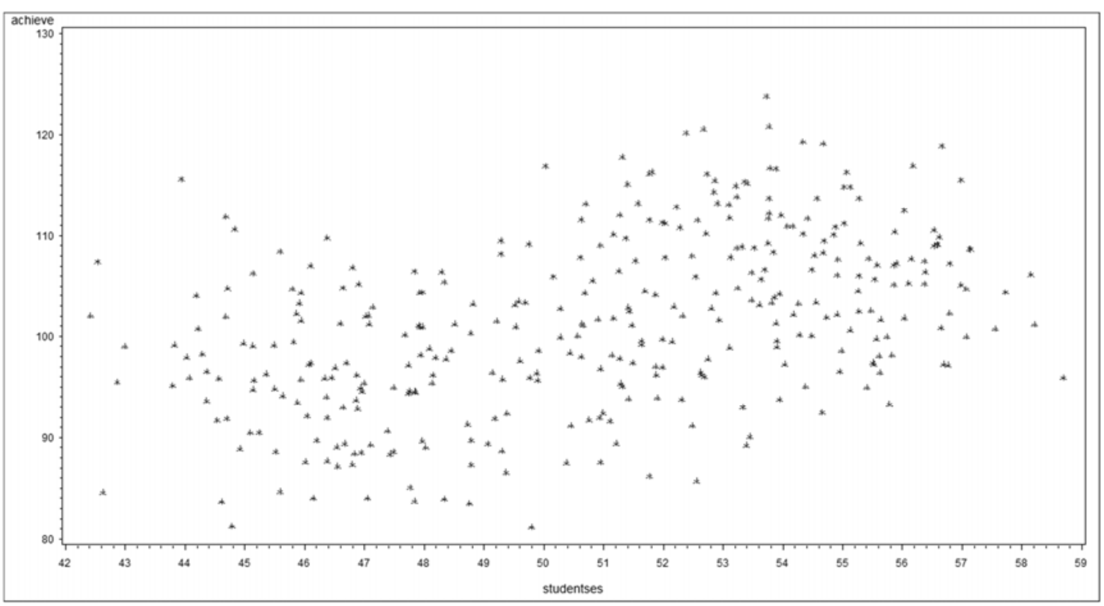
Your Study…

- Let's imagine you are able to get data from 7 elementary schools around Clarke and Oconee counties
  - You sample 50 students from each elementary school
  - You record a measure of their SES (scale with a mean of 50)
  - You record a measure of their achievement (scale with a mean of 100)

- Both scales magically have absolutely perfect reliability

Your Data…Do You See A Trend?

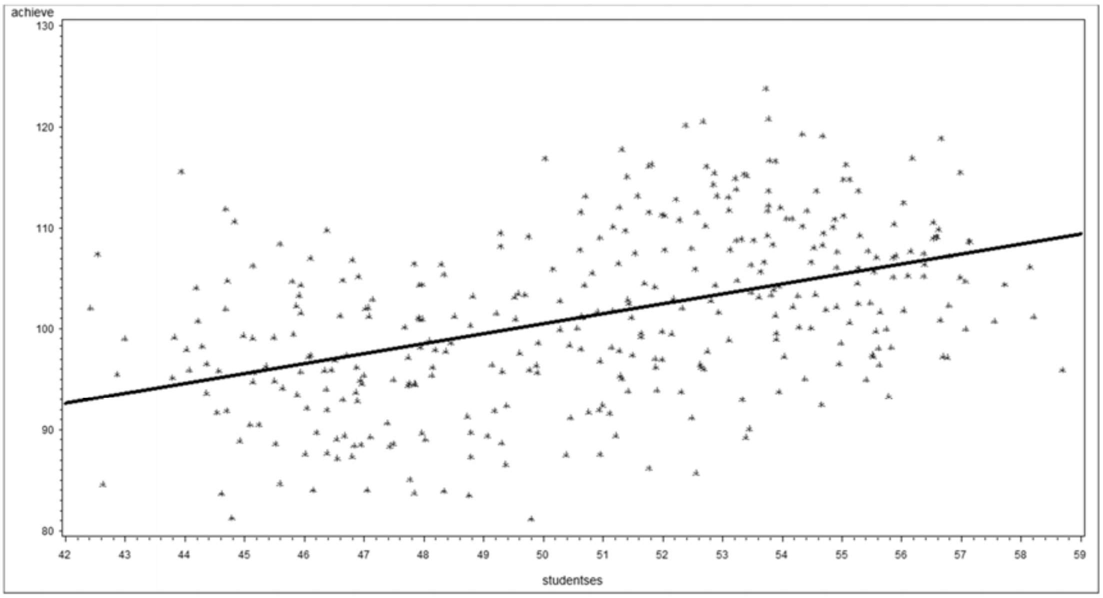
Student Achievement



Student SES

Your Data…Do You See A Trend?

Student Achievement



Student SES

Moving from Qualitative to Quantitative…

- Instead of "feeling" a trend in the data, statistical analyses let us test the hypothesis that a trend exists

- We will use a linear regression model:

$$Y_i = \beta_0 + \beta_1 X_i + r_i$$

This would be student achievement

This would be student SES

Regression

This would be student SES

$$Y_i = \beta_0 + \beta_1 X_i + r_i$$

This would be student achievement

This would be the average performance of a student when X=0 (student SES is zero)

This would be the average increase in performance for every unit increase in SES

The unique effect for student *I*

Usually this is called the error term

Now lets say just a little but more about β_0 and r_i

Regression

- Many times β_0 does not make sense based on the scale (mainly location) of our independent variables (i.e., *X*)
  - In our example what does it mean to have zero SES?
  - In many cases it makes more sense to "**center**" the variable (i.e., use $X_i - \bar{X}$)
  - Now β_0 is the average response for a student with the average SES

Regression

- In most general linear models, like regression, we make a basic set of assumptions about the r_i across all students:
  - They are normally distributed
  - They have mean 0
  - They have variance σ^2

Our Analysis

- First, let's look at the descriptive statistics for our data:

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
achieve	350	101.3811912	8.5143419	81.1617979	123.7855178
studentses	350	50.8501476	3.8808480	42.4146209	58.6925801

Y_i - Achievement Mean(Y) = 101.38 SD(Y) = 8.51

X_i - Student SES Mean(X) = 50.85 SD(X) = 3.88

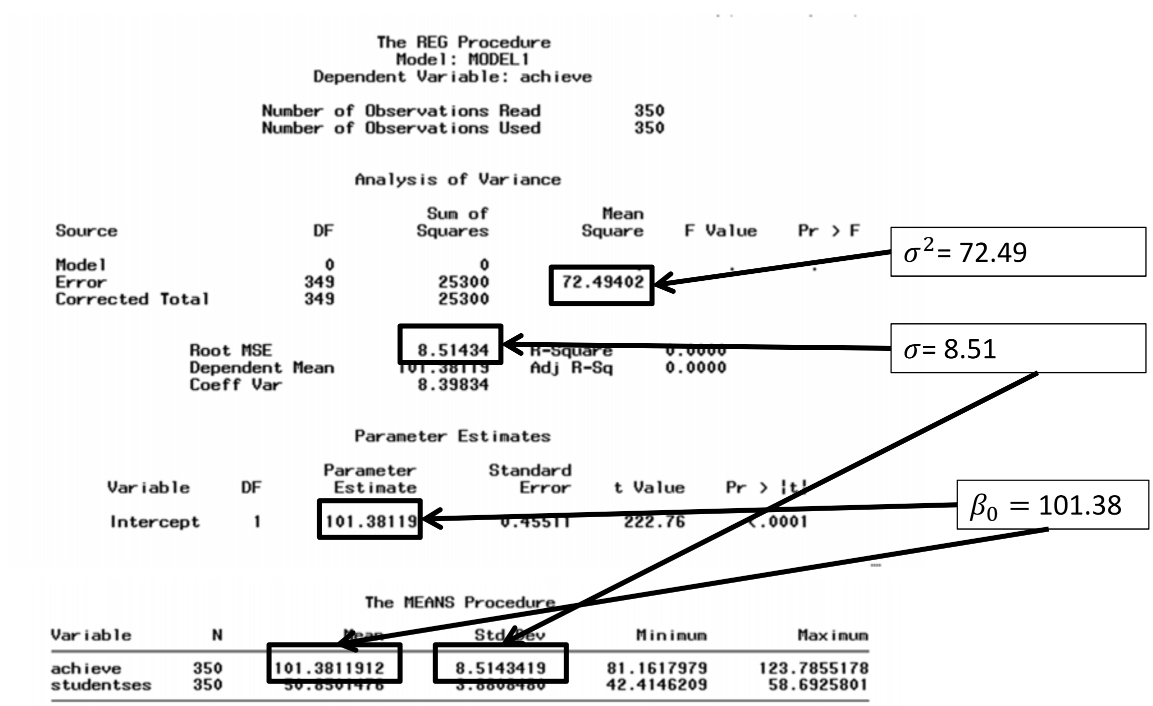
Statistical Inference in Regression

- To walk through the regression analysis, let's start by running a very simple (and unrealistic) analysis:

$$Y_i = \beta_0 + r_i$$

- A model with just the intercept
- What do you think the intercept (β_0) will be?
- What about the error variance (σ^2)?

Empty Model Output



The REG Procedure
Model: MODEL1
Dependent Variable: achieve

Number of Observations Read 350
Number of Observations Used 350

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	0	0		.	.
Error	349	25300	72.49402		
Corrected Total	349	25300			

Root MSE	8.51434	R-Square	0.0000
Dependent Mean	101.38119	Adj R-Sq	0.0000
Coeff Var	8.39834		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	101.38119	0.45511	222.76	<.0001

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
achieve	350	101.3811912	8.5143419	81.1617979	123.7855178
studentses	350	[illegible]	[illegible]	42.4146209	58.6925801

Next…What We Were Interested In…

- Next, let's run the regression model we were interest in running: predicting achievement from SES:

$$Y_i = \beta_0 + \beta_1 X_i + r_i$$

- From this analysis, we can tell a few things:
  - The reduction in error variance because of SES
  - If SES is related to achievement
  - How increases in SES net increases in achievement

And…The Results…

Covariance Parameter Estimates

Cov Parm	Estimate
Residual	58.0604

$\sigma^2 = 58.06$

Empty Model σ^2= 72.49

Fit Statistics

-2 Res Log Likelihood	2415.4
AIC (smaller is better)	2417.4
AICC (smaller is better)	2417.4
BIC (smaller is better)	2421.3

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	51.3152	5.3598	348	9.57	<.0001
studentses	0.9846	0.1051	348	9.37	<.0001

$\beta_0 = 51.32\ (p < 0.0001)$ — Student with 0 SES Scores 51.32

$\beta_1 = 0.98\ (p < 0.0001)$ — Score increases 0.98 for every 1 SES

What would you predict a student with SES of 50 would score?

$$\hat{Y}_i = 51.32 + 0.98 * 50 = 100.32$$

What About ANOVA?

- To describe how ANOVA fits into the GLM with Regression, let's imagine you wanted to see if there were significant differences between your schools with respect to achievement

- Because school is not a quantitative variable, we have to code it into one
  - We will use "dummy coding" where:
    - X_{is} = 1 if student i goes to school s
    - X_{is} = 0 if student i does not go to school s
  - We will not use a code for the 7^{th} school (creates a dependency that cannot be resolved)

- For a categorical variable with C categories, there will be C-1 new variables created
  - The coding system does not assume the categories are ordered – only that they are nominal in nature

The Model

- The ANOVA model for determining if differences between achievement exist between schools attempts to predict a student's achievement (Y) based on the school he or she attends (X):

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_6 X_{i6} + r_i$$

The Results

Covariance Parameter Estimates

Cov Parm	Estimate
Residual	29.0378

$\sigma^2 = 29.04$

Empty Model σ^2= 72.49

Solution for Fixed Effects

Effect	school	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		93.0464	0.7621	343	122.10	<.0001
school	1	-0.5444	1.0777	343	-0.51	0.6138
school	2	15.3386	1.0777	343	14.23	<.0001
school	3	6.5288	1.0777	343	6.06	<.0001
school	4	8.4666	1.0777	343	7.86	<.0001
school	5	18.4930	1.0777	343	17.16	<.0001
school	6	10.0608	1.0777	343	9.34	<.0001
school	7	0	.	.	.	.

$\beta_0 = 93.05$
$\beta_1 = -0.54$
$\beta_2 = 15.34$
$\beta_3 = 6.52$
$\beta_4 = 8.46$
$\beta_5 = 18.49$
$\beta_6 = 10.06$

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
school	6	343	88.05	<.0001

Overall p-value: < 0.0001 – Schools appear to have different achievement scores

More on Results

- The interpretation of the parameters is related to the mean differences in schools:

Least Squares Means

Effect	school	Estimate	Standard Error	DF	t Value	Pr > |t|
school	1	92.5020	0.7621	343	121.38	<.0001
school	2	108.39	0.7621	343	142.22	<.0001
school	3	99.5752	0.7621	343	130.66	<.0001
school	4	101.51	0.7621	343	133.21	<.0001
school	5	111.54	0.7621	343	146.36	<.0001
school	6	103.11	0.7621	343	135.30	<.0001
school	7	93.0464	0.7621	343	122.10	<.0001

- $\beta_0 = 93.05$ = Mean of reference group (here, school 7)
- $\beta_1 = -0.55$ = Difference between mean of school 1 and mean of school 7 (92.5 – 93.05)
- Parameters with dummy codes represent mean differences between schools and reference group

Predictions in ANOVA

- The ANOVA model treats all students within a school **the same**
  - Prediction for a given student is equal to all other students in their school (the school mean):

What do we predict a student from school 1 to score?

$\hat{Y}_i = 93.05 - 0.54 * 1 + 15.3 * 0 + 6.5 * 0 + 8.5 * 0 + 18.5 * 0 + 10 * 0 = 92.5$

What do we predict a student from school 7 to score?

$\hat{Y}_i = 93.05 - 0.54 * 0 + 15.3 * 0 + 6.5 * 0 + 8.5 * 0 + 18.5 * 0 + 10 * 0 = 93.05$

Further Assumptions of the GLM

- Now that we've seen Regression and ANOVA in the GLM, let's now discuss an important assumption: that the residuals are all uncorrelated

- This means that the observations are independent given the predictors in the model

- That means that we expect how wrong we are for a pair of students to be independent regardless of whether or not they come from the same school
  - Do you think that is assumption is likely to hold in our data?

MOVING TO HIERARCHICAL DATA

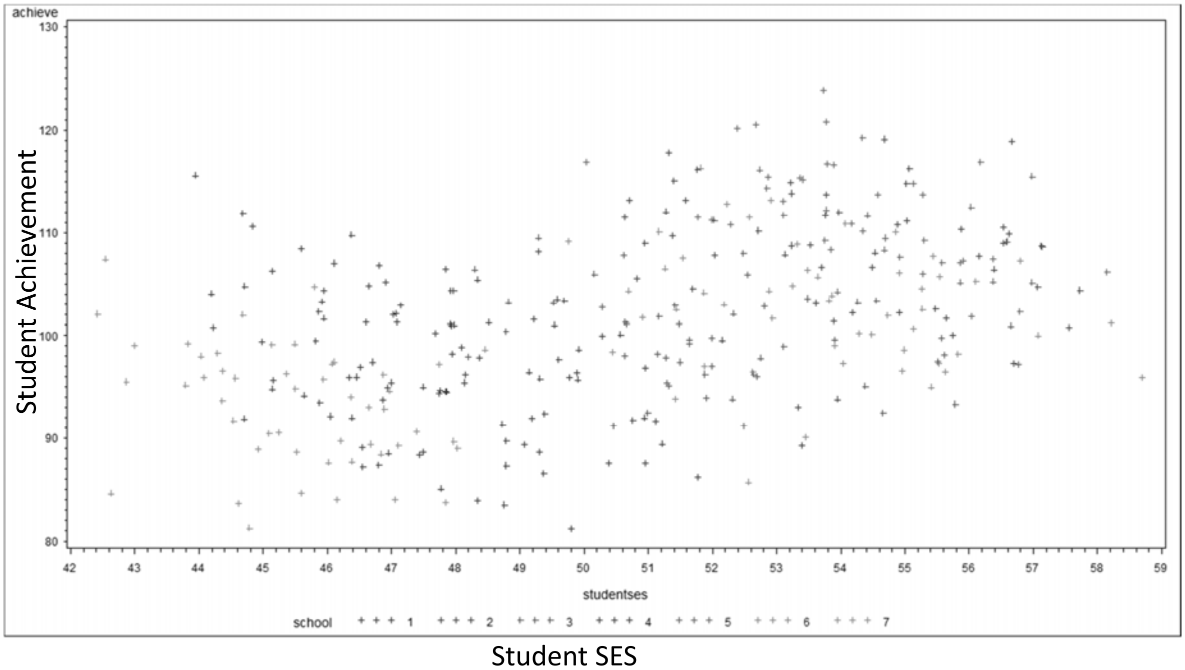
We Have Schools…Don’t They Matter?

- Let’s imagine that you complete these analyses and take them to your advisor (or colleague/co-author)
  - They are happy, but inquisitive

- Clearly the data come from a set of 7 schools
  - Hierarchical in nature
  - School seems to have an effect

- What would our analyses be like if we used a hierarchical approach?

By School Plot



Revisiting our Regression Analysis

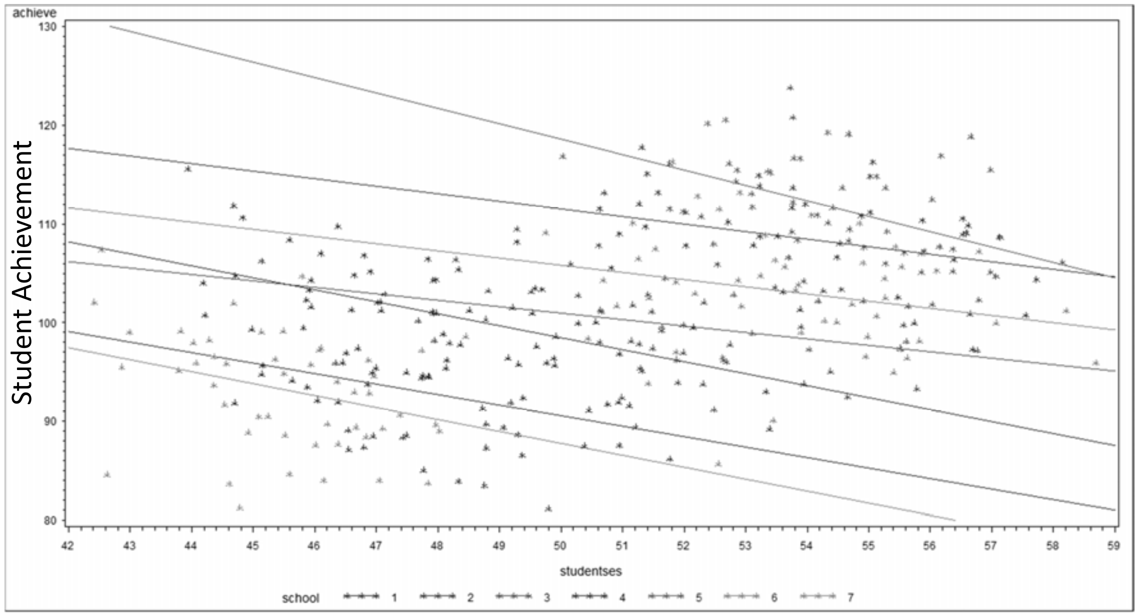
- The idea of HLM (multilevel models) is to look at the different levels of the data and use that information in the analysis

- Let's re-write our original regression model, but this time let's imagine that for each school, we will estimate a regression line:

$$Y_{is} = \beta_{0s} + \beta_{1s}X_{is} + r_{is}$$

- Now, each school has an intercept and a slope
  - The s- subscript appears for each term in the model

By-School Regression Lines



Student SES

Breaking Down Our Analysis By Level

- Now, we have two levels to our analysis:
  - The student level (level 1: where subscript i appears):

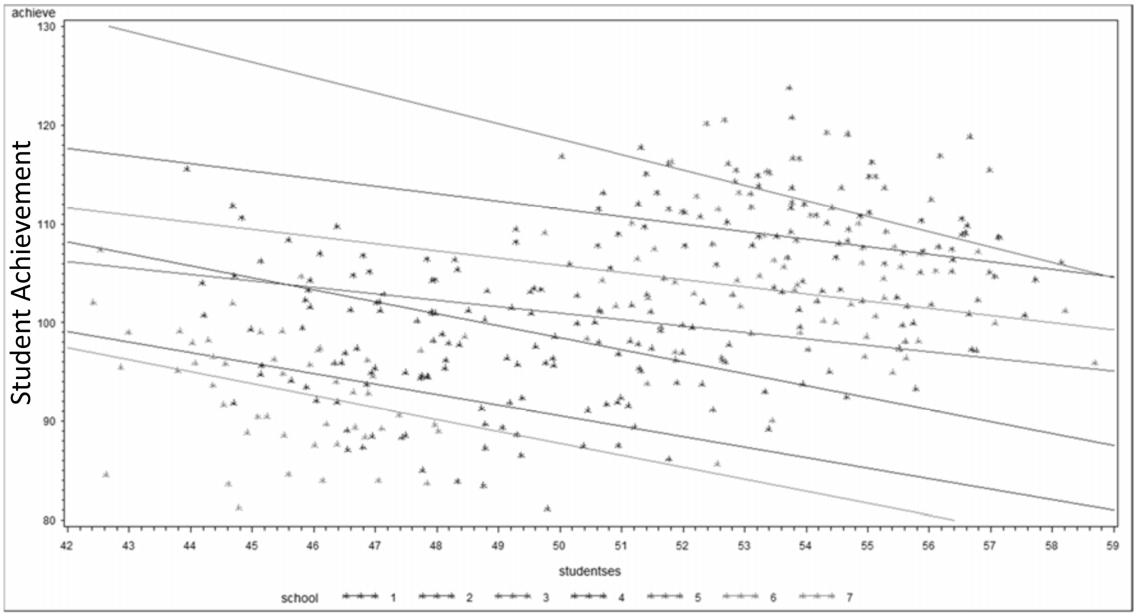
$$Y_{is} = \beta_{0s} + \beta_{1s} X_{is} + r_{is}$$

  - And the school level (level 2: where subscript s appears):

$$\beta_{0s} = ?$$
$$\beta_{1s} = ?$$

- The question is – what predicts the regression function for a school?
  - What variables predict the slope? The intercept?

By-School Regression Lines



Student SES

Revisiting Our Analysis: Adding School

- As is apparent by looking at the data:
  - The slopes for schools seem roughly the same
    - Every school may be able to have a shared slope
  - The intercept for schools seems to differ
    - Higher SES schools seem to have higher intercepts
    - Can use a school's average SES as a predictor of the slope
  - Although hard to see, it seems students within a school have highly related achievement scores
    - Means residuals are likely to be correlated
    - Violates the assumptions of regression
- We will be revisiting each of these questions throughout and seeing how HLM analyses can help

For Now…Let's Model the School Intercept

- Level 1: the student level

$$Y_{is} = \beta_{0s} + \beta_1 X_{is} + r_{is}$$

- Level 2: the school level

$$\beta_{0s} = \gamma_{00} + \gamma_{01}\,\bar{X}_s + U_{0s}$$
$$\beta_{1s} = \gamma_{10}$$

- γ_{00} is the overall intercept (predicted value when all X = 0)
- γ_{01} is the slope for school mean SES (indicates average intercept increase when school mean SES increases by 1.
- γ_{10} is the fixed slope for SES – meaning each school has the same increase (increase in student score when student SES increases by 1)
- U_{0s} - the error associated with school intercepts (called a random intercept)
  - Is assumed to be normally distributed with mean 0 and variance τ_0^2

Putting the Model Together

- We can substitute our level-2 model terms into our level-1 model equation to get an overall regression line:

$$Y_{is} = \beta_{0s} + \beta_1 X_{is} + r_{is} =$$
$$(\gamma_{00} + \gamma_{01}\,\bar{X}_s + U_{0s}) + (\gamma_{10})X_{is} + r_{is} =$$
$$\gamma_{00} + \gamma_{01}\,\bar{X}_s + \gamma_{10}X_{is} + U_{0s} + r_{is}$$

The Analysis Results…

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
Intercept	school	18.8901	12.2688	1.54	0.0618
Residual		25.4289	1.9446	13.08	<.0001

τ_0^2= 18.89

σ^2= 25.42

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	18.6914	25.1981	5	0.74	0.4916
schoolmean	2.6161	0.5140	5	5.09	0.0038
studentses	-0.9900	0.1405	342	-7.05	<.0001

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
schoolmean	1	5	25.90	0.0038
studentses	1	342	49.68	<.0001

γ_{00}= 18.69

γ_{01}= 2.62 (p = 0.0038)

γ_{10}= -0.99 (p < 0.0001)

Analysis Interpretation

- τ_0^2= 18.89
  - The variance of school random intercepts (how much schools vary from each other) after accounting for school SES
  - More on this next

- σ^2= 25.42
  - The variance of residuals for student scores (how much students vary after accounting for school mean SES and student SES)
  - The error variance

Analysis Interpretation

- γ_{00}= 18.69
  - The overall intercept
  - The predicted score for a student who has zero SES (X_{is} = 0)
  - At a school with a mean SES of zero ($\bar{X}_s = 0$)

- γ_{01}= 2.62 (p = 0.0038)
  - The slope for school mean SES
  - The predicted score for a student increases by 2.62 for every one unit increase in the school mean, after controlling for student SES (contextual or incremental effect)
    - Average achievement for a school increases as average SES increases
  - Statistically significant (level-2 degrees of freedom)

Analysis Interpretation

- γ_{10}= -0.99 (p < 0.0001)
  - The slope for student SES
  - The predicted score for a student decreases by 0.99 for every one unit increase in the student's SES
    - Within a school, SES is negatively related to achievement
  - Statistically significant (level-1 degrees of freedom)

- So, what is the nature of the relationship between SES and achievement?
  - Level 1 – SES is negatively related to achievement
  - Level 2 – SES is positively related to achievement

Wrapping Up

- This section was a refresher course on the GLM…
  - ANOVA/Regression – it's all a linear model

- …and an introduction to hierarchical analyses
  - You can see how conclusions can change if hierarchical structures of data are not taken into consideration

Up Next

- We will expand on this example to cover a few more important concepts in hierarchical linear models
  - The importance of centering of variables
    - Distinguishing within from between cluster effects

  - How total variation is partitioned by random effects
    - Implications for how residuals are correlated
    - Implications for hypothesis testing (type 1 and 2 errors)
    - Implications for modeling dependencies

  - More fun with two-level models