Model Building

Edps/Psych/Soc 587

Carolyn J. Anderson

Department of Educational Psychology
ILLINOIS

Outline

- Introduction
- Steps in an analysis:
  1. Selecting Preliminary Fixed Effects Structure
  2. Selecting a Preliminary Random Effects Structure
  3. Model Reduction
  4. Model Diagnostics
  5. Interpretation
- R (and SAS)

Read Chapters on web-site or cull from Verbeke & Molenbergs: Chapters 4, 9 and Snijders & Boskers chapter 6–8

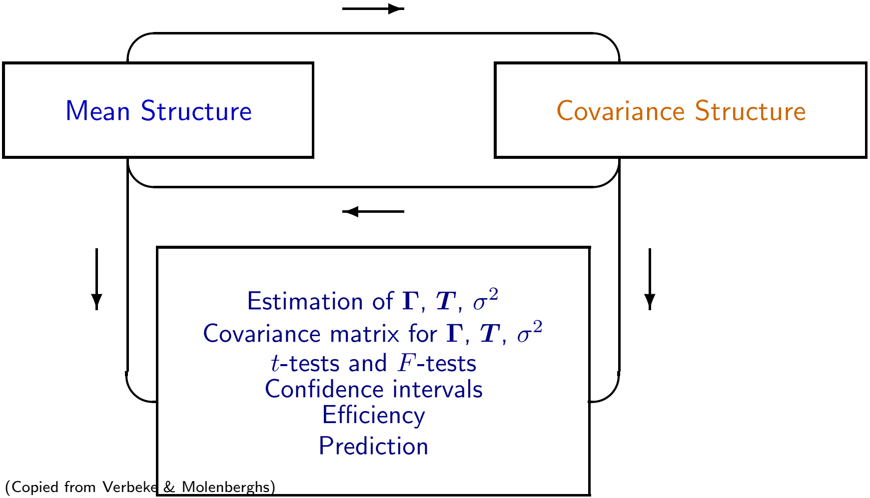
Introduction

When using linear mixed models, we specify

- Mean of responses for group j (i.e., $\boldsymbol{Y}_j$); that is, the fixed effects part of the model, $\boldsymbol{X}_j\boldsymbol{\Gamma}$.
- Covariance matrix for $\boldsymbol{Y}_j$, $\boldsymbol{V}_j = (\boldsymbol{Z}_j\boldsymbol{T}\boldsymbol{Z}_j' + \sigma^2\boldsymbol{I})$.

We must find an appropriate mean structure *and* covariance structure, but the mean and covariance structures are dependent upon each other.

Dependency: Mean and Covariance



(Copied from Verbeke & Molenberghs)

The Covariance Structure

- It explains and helps to understand the random variability in the data; the "unexplained" variance.
- It is highly dependent on the fixed effect structure (i.e., the systematic part of the variability of $\boldsymbol{Y}_j$).
- An appropriate one is required for valid inference regarding the mean structure (unless you use robust estimation).
- Under-parameterized covariance structure invalidates inferences.
- Over-parameterized covariance structure leads to inefficient estimation and poor standard errors.
- Is interesting in helping to understand the random variability in the data.
- An appropriate covariance structure leads to better predictions.

EDA Two Stage Process

In multilevel modeling, we basically have a 2 (or more) stage process:

- Level 1: Specify regression model for individual within group j.
- Level 2: Specify regression models for parameters of level 1 regression model to explain group differences.

If most of the variability in the data is between groups, then this 2 stage approach will often lead to a valid marginal model for the data ; *however, a multilevel approach will sometimes lead to an invalid marginal model.*

If most of the variability is within groups, then you may not need random effects in the model, and $\sigma^2 \boldsymbol{I}$ represents unexplained variability within a group.

Note: With longitudinal data where most of the variability is between individuals (i.e., macro unit) and high dependency within, then need to give more thought to the within individual covariance matrix.

General Guidelines on Model Selection

Finding an appropriate linear mixed model for a specific data set.

The procedures presented here

- Are a combination of general modeling guidelines and possible exploratory data analyses.
- May not yield the most appropriate model.
- Do not guarantee that all distributional assumptions are satisfied.
- Not an exhaustive set of tools.

Basic structure for Model Building

1. Remove the systematic part from the data.
2. Study the residuals in an effort to get a preliminary or reasonable random effects structure that will permit valid inference regarding fixed effects.
3. Remove/revise/add the fixed effects, including testing substantive research hypotheses.
4. Remove/revise/add the random effects, including testing substantive research hypotheses.
5. Cycle through steps 3 and 4.
6. Model diagnostics on potential final model(s).
7. Interpret final model.

Basic structure for Model Building

- This is similar to what others recommend: specify level 1 model and then level 2 model(s).
- It differs in that we start complex rather than simple.

Selecting Preliminary Fixed Effects

- Examine each group graphically.
- Averaging over sub-populations and graph.
- Exploring Group Specific Data
  - Measure each group's goodness of fit.
  - Measure overall goodness of fit.
  - "Testing" for model extension (skip this).

Why Start with Fixed Effects?

- The covariance matrix accounts for all the variability that's not accounted for by the systematic part of the model.
- We start with a complex, preliminary fixed effects (i.e., $\boldsymbol{X}_j\boldsymbol{\Gamma}$) and then remove it from the data leaving data variance due to random effects.
- We can ignore dependencies in the data and use ordinary least squares estimation to estimate the fixed effects...this works for normal models but not others (e.g., multilevel logistic regression).

The justification for using OLS?

Justification for Using OLS

"Generalized Estimation Equation" (GEE) Theory:

The OLS estimate of $\mathbf{\Gamma}$ is consistent.

Therefore, we can use

$$r_{ij} = y_{ij} - \boldsymbol{x}'_{ij}\mathbf{\Gamma}$$

to study the dependencies in the data.

Procedures for Preliminary Fixed

The procedures that we'll cover for selecting a preliminary fixed effects structure,. . .

- Most of them are graphical.
- Looking at the data three different ways
  - Within group.
  - Sub-sets of data.
  - Marginal distributions of response variable.
- Others? Be creative.

Preliminary Fixed by Example

Data: NELS88, N=23 schools.

- Math: Response variable.

- Homework: How much time a student spends doing homework.
- SES: Student's SES.
- Race: Whether a student is white or non-white.
- Gender: Whether a student is male or female.
- Sector: Whether the school is public or private.
- Mean SES: Average SES of students attending a school.

Preliminary Fixed by Example

Data: NELS88, N=23 schools.

Goal of the analysis: Try to account for differences between students' math performance in terms of student characteristics and school characteristics.

Start with some exploratory methods and use the results in our next stage.

Averaging over Sub-populations

Question: Can our response variable (math scores) be modeled by a linear regression model?
Possible graphical displays depend on whether the explanatory variables are

- Discrete.
- Continuous or virtually continuous.

Averaging over Sub-populations (continued)

- Discrete:
  - Nominal — average math scores of students within levels or categories.
  - Ordinal — average math scores of students within levels or categories but look at them in order of the categories (or use numerical values for the categories).
- Continuous or virtually continuous: Create grouping of students' based on their values of this variable and average math scores and explanatory variable.

Question 1

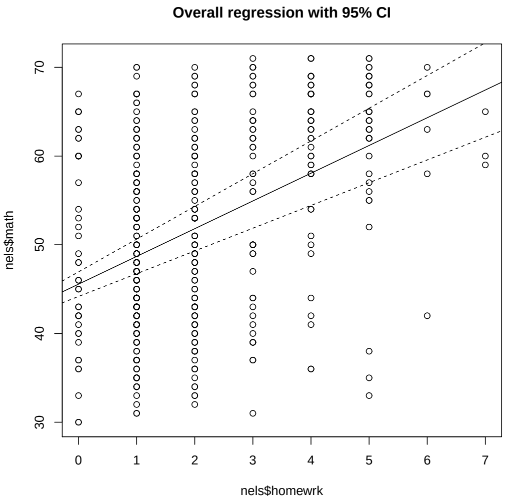
How do the math scores depend on homework?

Ordinal variable that's treated numerically: "homew," time student spends on math homework.

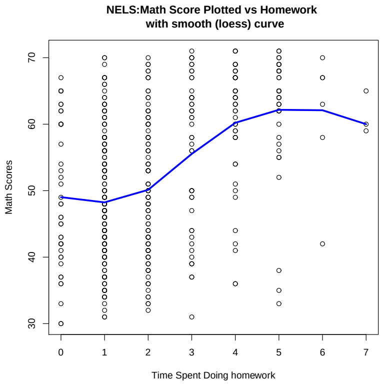
Some possibilities:

- Plot all the math scores by homework and fit a smooth curve to the points.
- Plot the math scores by homework for each school and fit curve.
- Plot the average math scores of students within schools for each level of homework versus homew.

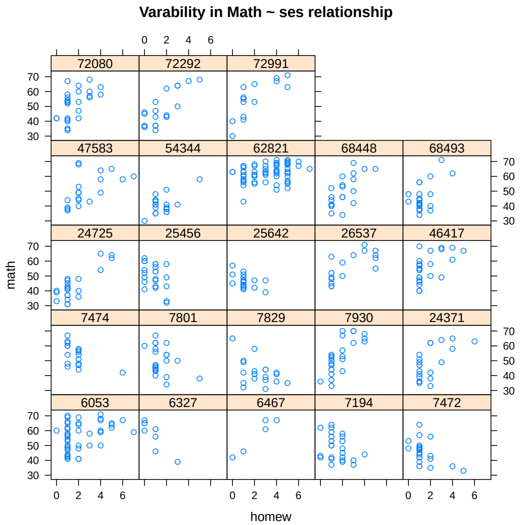
Math Score by Homework: Regression



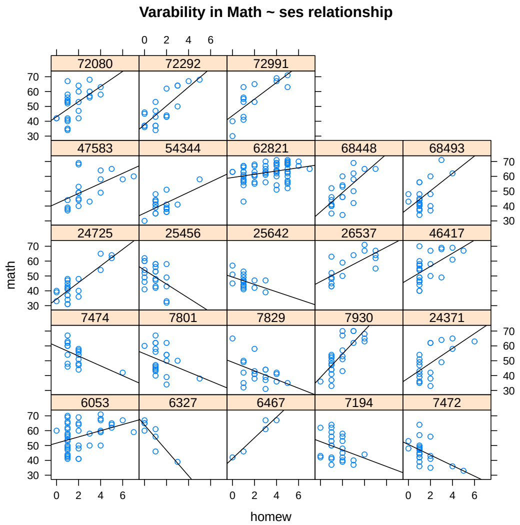
Math Scores by HOMEW: Spline



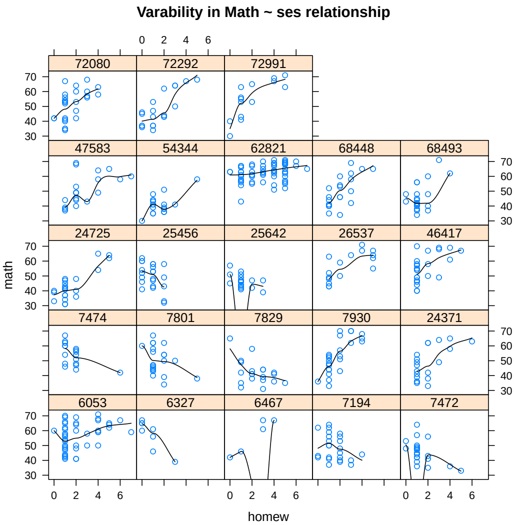
Math Scores by HOMEW: Panel



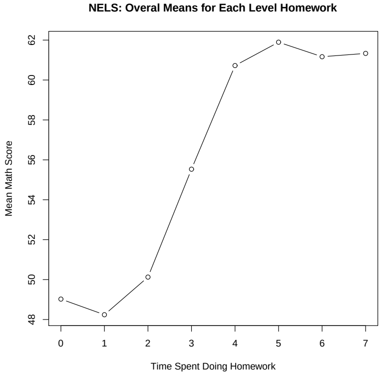
Math Scores by HOMEW: Panel



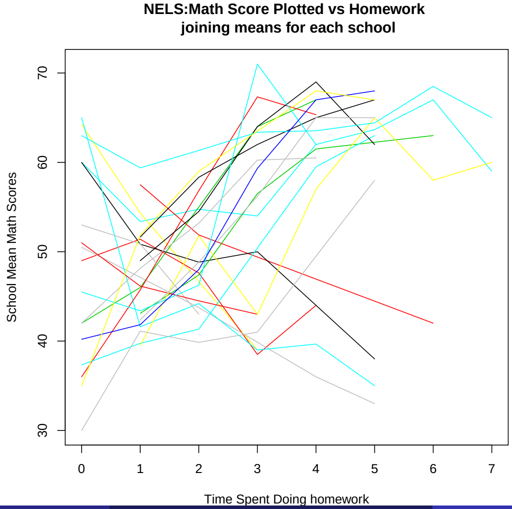
Math Scores by HOMEW: Smooth



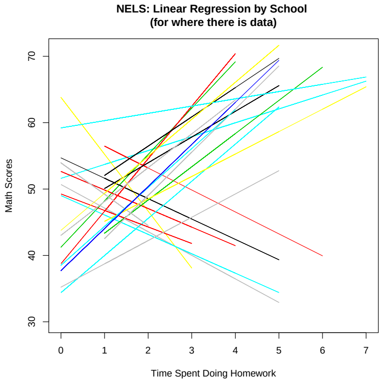
All Schools: Join School Means



Mean Math and Homework (Spaghetti plot)



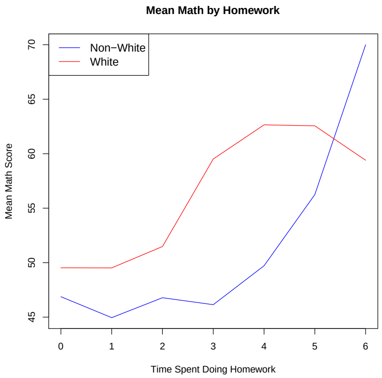
Linear regressions for each School



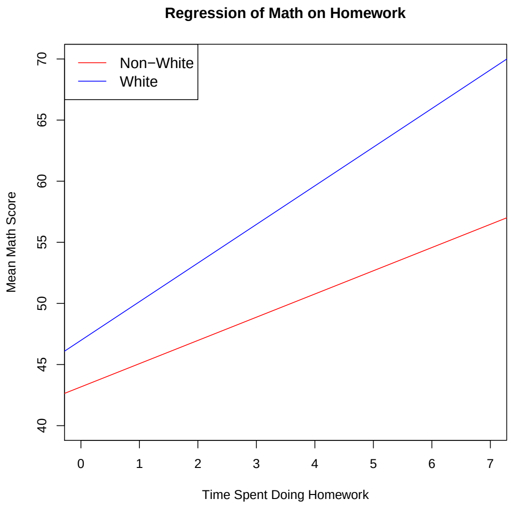
Math Scores by HOMEW

- Relationship between math scores and time spent doing homework appears (could be) linear.
- School differences in the overall level.
- School differences in slope.

Schools Means: HOMEW & WHITE



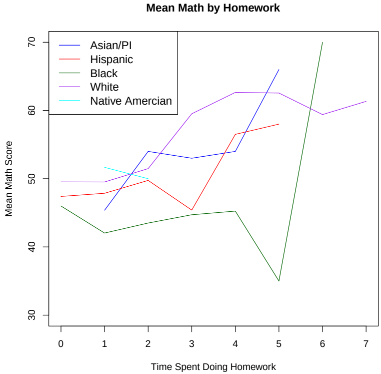
Another look at WHITE: Linear Regressions



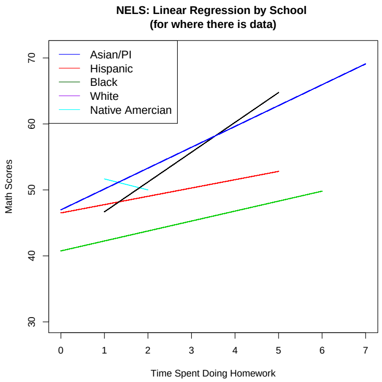
Schools Means: HOMEW & WHITE

- Differences between white and non-white.
- Possible diminishing returns for spending more time on homework for whites? (from plot joining means) However, there isn't much data at high end of HOMEW.
- Probably no interaction between time spent doing homework and `white`.

Schools Means: HOMEW & RACE



Caution. . .


NELS: Linear Regression by School
(for where there is data)
Asian/PI
Hispanic
Black
White
Native Amercian
Math Scores
30
40
50
60
70
0
1
2
3
4
5
6
7
Time Spent Doing Homework

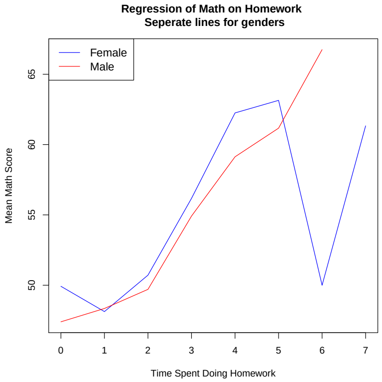
Caution (continued)

Homework	Student Race					
Levels	Asian/PI	Hispanic	Black	White	NatAm	Total
0	0	5	3	34	0	42
1	8	22	30	162	3	225
2	7	4	20	79	1	111
3	2	5	7	33	0	47
4	1	2	4	40	0	47
5	2	1	1	34	0	38
6	0	0	1	5	0	6
7	0	0	0	3	0	3
Total	20	39	66	390	4	519

Schools Means: HOMEW & RACE

- Not much data for Native Americans.
- Not much data for highest 2 levels of time spent doing homework.
- Not much data for races other than white for highest 4 or so levels of time spent doing homework.
- Black and Hispanic similar slopes but different levels.
- Suggests interaction between race and time spend doing homework.
- For now, fall back to test the substantive hypothesis that white differ from others; however, re-consider this later.

Schools Means: HOMEW & Gender

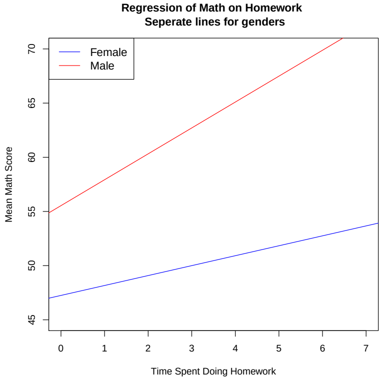


Schools Means: HOMEW & Gender

Number of observations per mean:

	Time Spent Doing Homework								
Gender	0	1	2	3	4	5	6	7	total
MALE	27	110	46	23	24	14	2	3	249
FEMALE	15	115	65	24	23	24	4	0	270
Total	42	225	111	47	47	38	6	3	519

A Smoother Look at Gender



Summary Regarding Gender

- There does not appear to a be an effect due to gender of student.
- There doesn't appear to be an interaction between homework and gender.

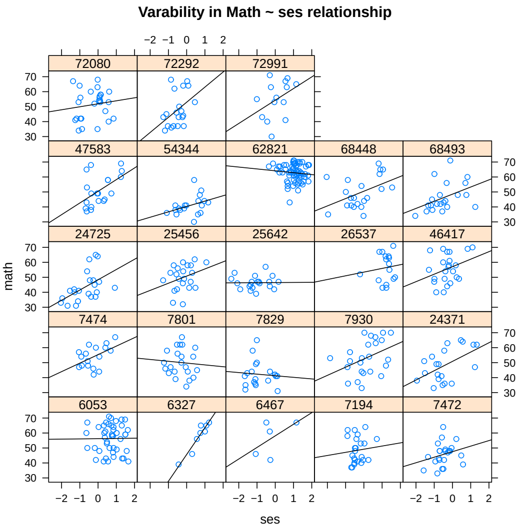
Averages with Continuous x_{ij}'s

Do math scores vary with respect to student SES?

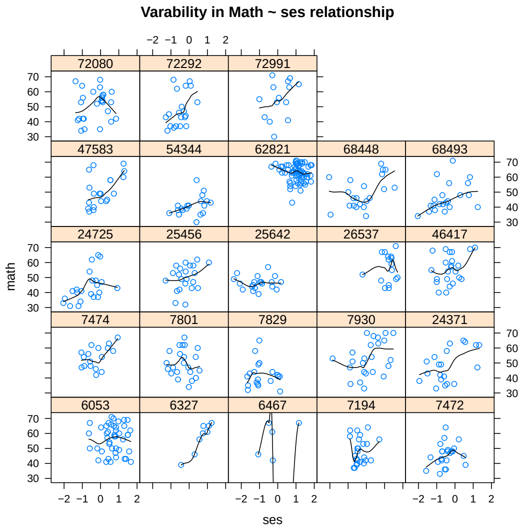
Since SES is "continuous"

- Plot math scores by SES.
- Group data according to SES.
- Compute averages math scores and average SES with the SES grouping.
- Plot the average math scores versus the average of the SES grouping.

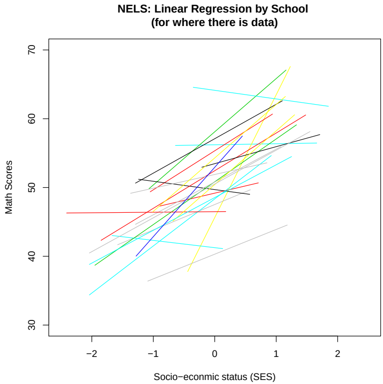
Individual School Plots: SES



Individual School Plots: SES (smooth)



Individual School Plots: SES (schools)

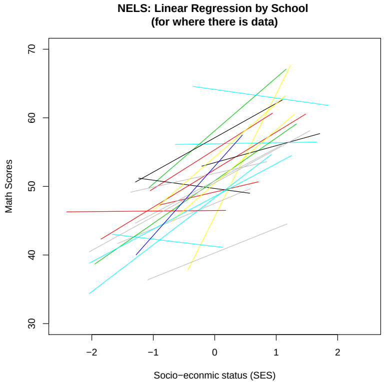


SES (so far)

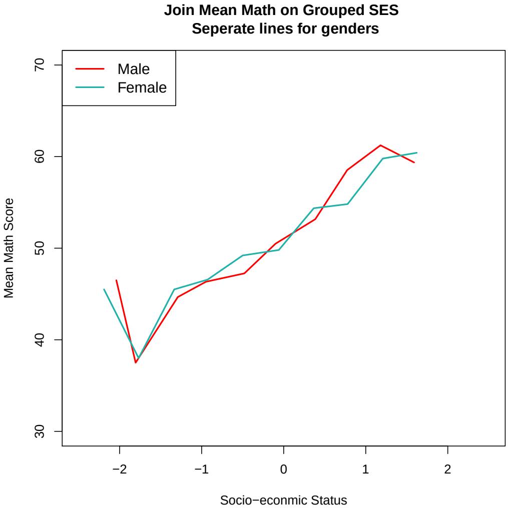
- Generally, math scores increase linearly with increasing SES.
- Variability in the overall level over schools.
- Slopes seem fairly similar.

Schools Means: MATH vs SES

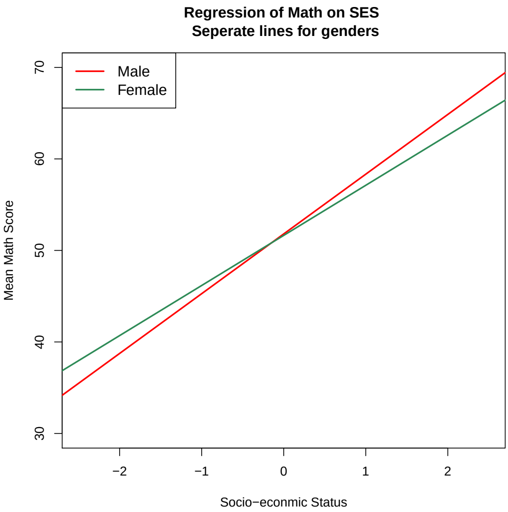
Again, math scores increase with student SES. . . this doesn't add much.



MATH vs SES & Gender (grouped ses)



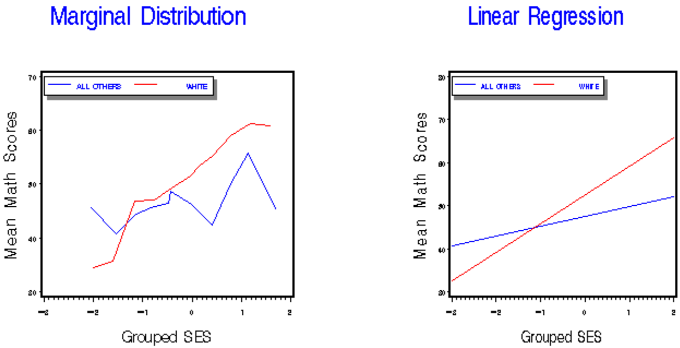
Linear Regression MATH vs SES & Gender



MATH vs SES & Gender

How do math scores vary with respect to SES and is there an interaction with gender?

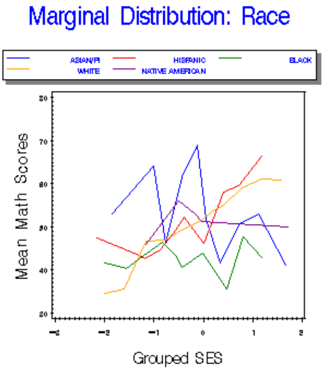
- Basically linear.
- Curves are pretty much on top of each other

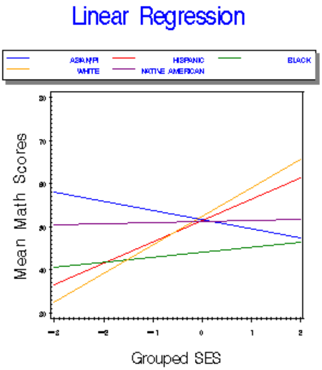

Marginal Distribution
ALL OTHERS
WHITE
Mean Math Scores
Grouped SES
Linear Regression
ALL OTHERS
WHITE
Mean Math Scores
Grouped SES

MATH vs SES & WHITE

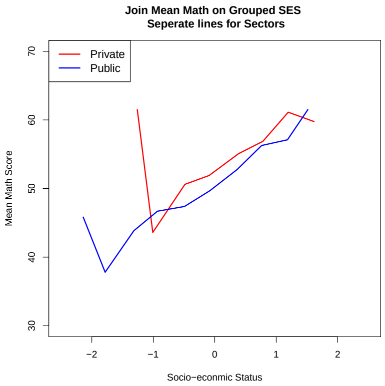
How do math scores vary with respect to `SES` and is there an interaction with `WHITE`?

- Basically linear.
- Maybe a little interaction or just noise?

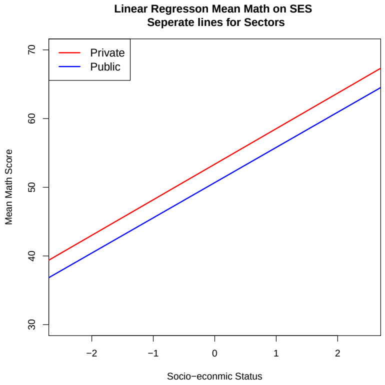

Marginal Distribution: Race
ASIAN/PI
HISPANIC
BLACK
WHITE
NATIVE AMERICAN
Mean Math Scores
Grouped SES


Linear Regression
ASIAN/PI
HISPANIC
BLACK
WHITE
NATIVE AMERICAN
Mean Math Scores
Grouped SES

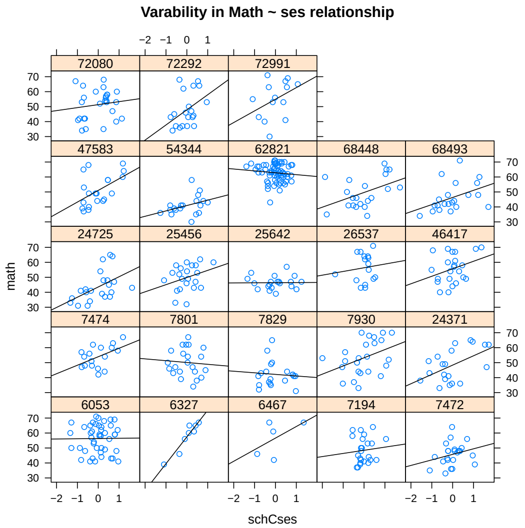
Schools Means: SES & Sector



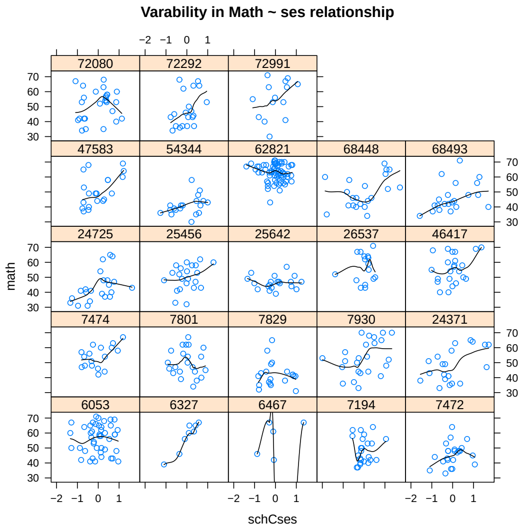
Smoothed Look at SES & Sector



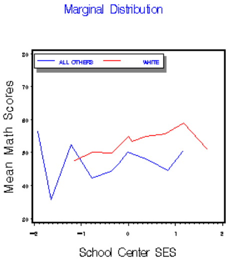
Group Mean Centered SES with Regressions

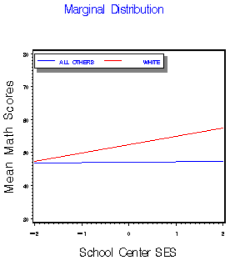


Group Mean Centered SES with Smoothed



Group Mean Centered SES: White

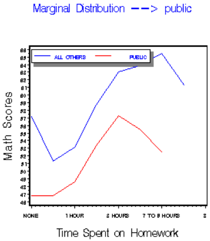


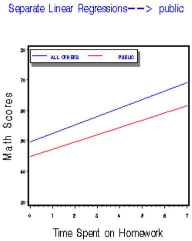


Plots with Macro Level Variables

- So far we have looked at level 1 explanatory variables.
- We can consider macro level explanatory variables.
- e.g., Do math scores tend to differ with respect to sector?
- Sector is a school characteristic.
- This would give information about whether the macro variable may be a potential explanatory variable for a random intercept.

Schools Means: HOMEW & Sector





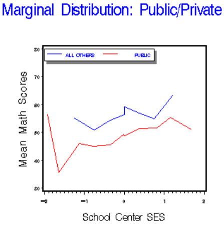
Schools Means: HOMEW & Sector

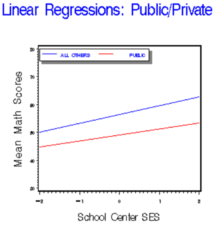
- Average math scores from private are higher than public.
- The curves are basically parallel (i.e., no interaction w/rt `homew`).
- Sector maybe a potential explanatory variable for a random intercept.

Schools Means: SES & Sector

- Not any observations for private at lowest SES levels.
- Linear increase in math scores w/rt to SES.
- Probably no interaction between sector and SES.

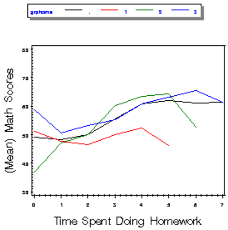
Sector & Group Centered SES

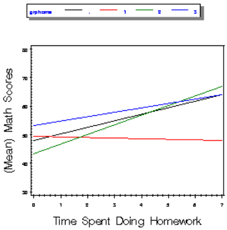




- “Curves” relatively flat (small positive slope).
- Private sector has higher math scores.

Mean Homework as Macro?





Summary of Findings: Level 1

- Overall, Math scores tend to go up with `homew`, but there's lots of variability over schools.
- Differences in math scores w/rt to race and with `homew`.
- Math scores go up with `SES` levels.
- Math scores vary w/rt to group centered `SES`.
- There may or may not be a Race effect when consider `SES`.
- Gender doesn't seem that important and doesn't appear to interaction with any other variables.
- Math scores basically increase with group mean centered `SES`.
- Schools show lots of variability (especially in terms of intercept) when examining math scores vs group mean centered `SES`.
- Others?

Summary of Findings: Level 2

- Math score higher for private vs public sector.
- Basically parallel curves for sector indicating that sector possible explanatory variable for modeling intercept differences between schools but no interaction w/ either `SES`, `cSES` or `homew`.
- Math scores increase with School (mean) `SES`; therefore, if use school mean centered SES, definitely include school mean SES as an explanatory variable for intercept (of level 1 model).
- Maybe investigate using mean homework as a predictor for intercept (& use school centered homework)?

Preliminary Models for level 1

Simple:

$$
\begin{aligned}
(\text{math})_{ij} &= \beta_0 + \beta_1(\text{homework})_{ij} + \beta_2(\text{cSES})_{ij} \\
&\quad + \beta_3(\text{white})_{ij} + \beta_4(\text{female})_{ij} + \beta_5(\text{cSES})_{ij}(\text{white})_{ij} + R_{ij}
\end{aligned}
$$

A more complex model to consider (for illustration & pedological reasons):

$$
\begin{aligned}
(\text{math})_{ij} &= \beta_0 + \beta_1(\text{homework})_{ij} + \beta_2(\text{cSES})_{ij} \\
&\quad + \beta_3(\text{white})_{ij} + \beta_4(\text{female})_{ij} \\
&\quad + \beta_5(\text{cSES})_{ij}(\text{white})_{ij} \\
&\quad + \beta_6(\text{homework})_{ij}(\text{female})_{ij} \\
&\quad + \beta_7(\text{homework})_{ij}(\text{white})_{ij} + R_{ij}
\end{aligned}
$$

Exploring Group Specific Data

We'll now try to see how well potential level 1 model actually fits the data.

- Measure fit of preliminary model(s) to each group's data.
- Measure overall fit to all groups.
- "Testing" for model extensions.

Measure Each Group's Fit

Fit the preliminary regression models (using standard multiple regression) to each of the group's (school's) data and see how well they fit.

Measure of goodness-of-fit: R_j^2 where

$$R_j^2 = \text{corr}(Y_j, \hat{Y}_j)^2 = \frac{\text{SStotal}_j - \text{SSerror}_j}{\text{SStotal}_j} = \frac{\text{SSmodel}_j}{\text{SStotal}_j}$$

i.e., The proportion of variance of Y_{ij} accounted for by the model for group j.

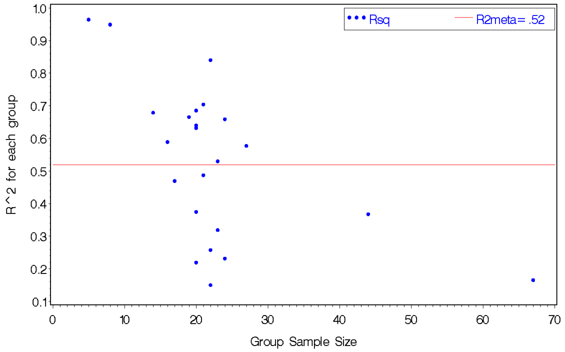
Measure Each Group's Fit (continued)

Rather than examining a table of R_j^2's values, visual displays much nicer.

- Histogram/bar chart — Ok but groups/schools with small to moderate n_j's will tend to have larger R_j^2 values.
- Scatter plot — R_j^2 versus n_j.

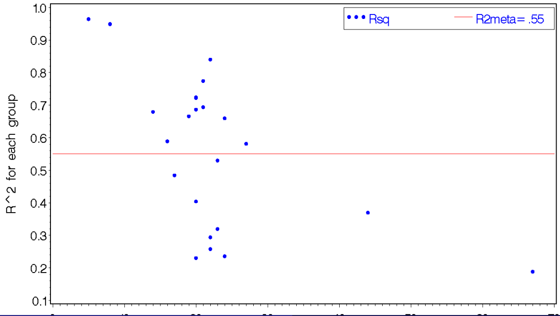
Scatter Plot: Simpler Model

$$\begin{aligned}(\text{math})_{ij} &= \beta_o + \beta_1(\text{homework})_{ij} + \beta_2(\text{cSES})_{ij} + \beta_3(\text{white})_{ij} \\ &\quad + \beta_4(\text{female})_{ij} + \beta_5(\text{cSES})_{ij}(\text{white})_{ij} + R_{ij}\end{aligned}$$



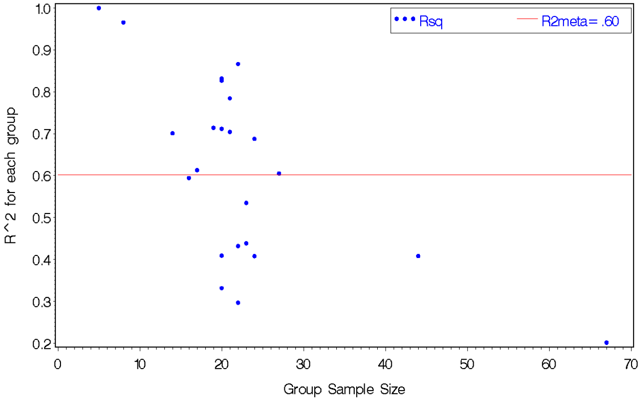
Scatter Plot: More Complex Model

$$\begin{aligned}(\text{math})_{ij} \quad = \quad & \beta_o + \beta_1(\text{homework})_{ij} + \beta_2(\text{cSES})_{ij} + \beta_3(\text{white})_{ij} \\ & +\beta_4(\text{female})_{ij} + +\beta_5(\text{cSES})_{ij}(\text{white})_{ij} \\ & +\beta_6(\text{homework})_{ij}(\text{female})_{ij} + \beta_7(\text{homework})_{ij}(\text{white})_{ij} \\ & +R_{ij}\end{aligned}$$

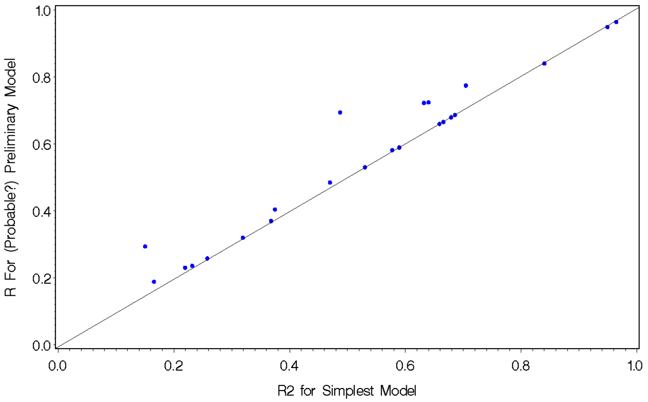


Even More Complex Model

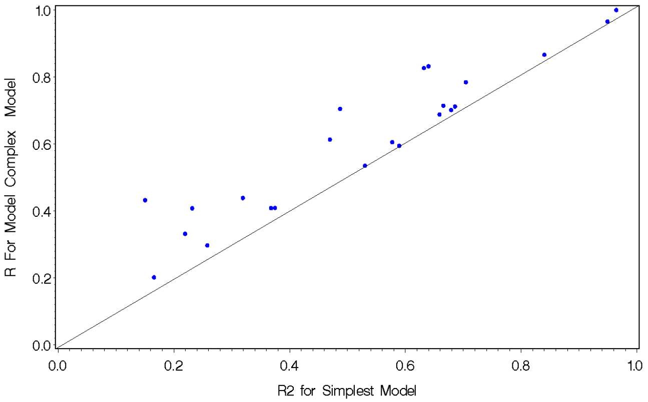
Model with all 2-way interactions. . . Looks pretty similar to previous ones.



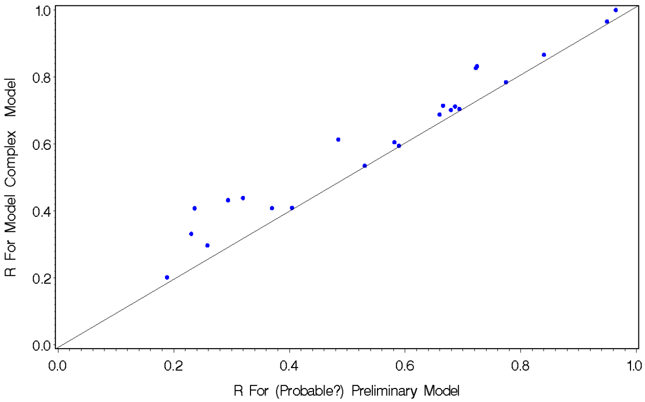
Comparing Models: Simplest vs Next



Comparing Models: Simplest vs Most



Comparing Models: Next vs Most



Measure Over-all-goodness of Fit

- R_j^2 gives us a measure for each group/school; however, it would be nice to have a global measure of fit; that is an overall-goodness of fit statistic. . . .
- $$R^2_{meta} = \frac{\sum_{j=1}^{N}(\text{SStotal}_j - \text{SSerror}_j)}{\sum_{j=1}^{N}\text{SStotal}_j} = \frac{\sum_{j=1}^{N}\text{SSmodel}_j}{\sum_{j=1}^{N}\text{SStotal}_j}$$
- Interpretation: the proportion of total within groups variability that can be explained by the level 1 linear regression model.
- This is what is plotted in previous figures.

Measure Over-all-goodness of Fit

- For our simple, preliminary model:

$$R^2_{meta} = \frac{19932.60}{40329.40} = .52$$

- For our more complex preliminary model:

$$R^2_{meta} = \frac{22221.76}{40329.40} = .55$$

- For our really over-parameterized model:

$$R^2_{meta} = \frac{24263.99}{40329.40} = .60$$

No Testing for Model Extension

- We have now examined the fit of three models.
- Is our simple model OK? Do we really need the extra interactions?
- In standard multiple regression, you can test to see whether additional terms are required by performing an F–test
- BUT assumptions of independence have been violated.

Conclusion: EDA Level 1 Model

- Preliminary Level 1 Model (too complex?)

$$
\begin{aligned}
(\text{math})_{ij} &= \beta_{0j} + \beta_{1j}(\text{homework})_{ij} + \beta_{2j}(\text{cSES})_{ij} + \beta_{3j}(\text{white})_{ij} \\
&\quad + \beta_{4j}(\text{female})_{ij} + \beta_{5j}(\text{cSES})_{ij}(\text{white})_{ij} + R_{ij}
\end{aligned}
$$

- We also plan to include
    - Sector and the group mean of SES as explanatory variables for the intercept.
    - An interaction between homework and sector.
    - What about an interaction between group mean of SES and other variables? Need to investigate this. . .

Preliminary Structural Model

$$
\begin{aligned}
\beta_{0j} &= \gamma_{00} + \gamma_{01}(\text{sector})_j + \gamma_{02}(\overline{\text{SES}})_j \\
\beta_{1j} &= \gamma_{10} + \gamma_{11}(\text{sector})_j \\
\beta_{2j} &= \gamma_{20} \\
\beta_{3j} &= \gamma_{30} \\
\beta_{4j} &= \gamma_{40} \\
\beta_{rj} &= \gamma_{50}
\end{aligned}
$$

Preliminary fixed/structural model:

$$
\begin{aligned}
(\text{math})_{ij} = \; & \gamma_{00} + \gamma_{01}(\text{sector})_j + \gamma_{02}(\overline{\text{SES}})_j \\
& + \gamma_{10}(\text{homework})_{ij} + \gamma_{11}(\text{sector})_j(\text{homework})_{ij} \\
& + \gamma_{20}(\text{cSES})_{ij} + \gamma_{30}(\text{white})_{ij} \\
& + \gamma_{40}(\text{female})_{ij} + \gamma_{50}(\text{cSES})_{ij}(\text{white})_{ij} \\
& + R_{ij}
\end{aligned}
$$

Preliminary Random Effects Structure

Two features of HLM's that result from random effects:

- Variance of the response variable, Y_{ij}, can be broken down into parts:
  - Between group differences and
  - Within group differences.
- Correlation between individuals withinthe same group (macro unit) is not equal to zero.

The Variance of Y_{ij}

- For a random intercept model:

$$\text{var}(Y_{ij}) = \tau_o^2 + \sigma^2$$

- For a random intercept and slope model:

$$\text{var}(Y_{ij}) = \tau_o^2 + 2\tau_{10}x_{ij} + \tau_1^2 x_{ij}^2 + \sigma^2$$

  - “Heteroscedasticity”
  - Quadratic function of x_{ij}.

- More generally, for a random intercept and slope model:

$$\text{var}(Y_{ij}) = \sum_{k=0}^{p} \tau_k^2 x_{k,ij}^2 + 2\sum_{k<l} \tau_{kl} x_{k,ij} x_{l,ij} + \sigma^2.$$

The Correlation Between Y_{ij} and $Y_{i'j}$

- For a random intercept model,

$$\text{corr}(Y_{ij}, Y_{i'j}) = \frac{\tau_0^2}{\tau_0^2 + \sigma^2}$$

- For a random intercept and slope model,

$$\frac{\tau_0^2 + \tau_{10}(x_{ij} - x_{i'j}) + \tau_1^2}{\sqrt{\tau_o^2 + 2\tau_{10}x_{ij} + \tau_1^2 x_{ij}^2 + \sigma^2}\sqrt{\tau_o^2 + 2\tau_{10}x_{i'j} + \tau_1^2 x_{i'j}^2 + \sigma^2}}$$

The Variance Function

- Remove the fixed effects structure from the data and examine the residuals.
- Use ordinary least squares to get estimates of the parameters of the preliminary model,

$$Y_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\Gamma}$$

- Compute the residuals,

$$\hat{e}_{ij} = Y_{ij} - \boldsymbol{x}'_{ij}\hat{\boldsymbol{\Gamma}}$$

- Plot and study residuals.
  - Raw residuals (or mean) versus explanatory variables.
  - Square residuals versus explanatory variables.

Residuals of Random Intercept Model

- If the "true" model is a random intercept model, then the residuals should equal

$$e_{ij} = U_{0j} + R_{ij}$$

- Further implications:
  - $e_{ij}|\boldsymbol{x}_{ij} = U_{0j} + R_{ij} \longrightarrow e_{ij}.$
  - $E[e_{ij}] = E[U_{0j} + R_{ij}] = 0 \longrightarrow \bar{e}_{++}$
  - $E[e_{ij}|\text{macro unit } = j] = U_{0j} \longrightarrow \bar{e}_{+j}.$
  - $\text{var}(e_{ij}) = E[(e_{ij})^2] = \tau_0^2 + \sigma^2 \longrightarrow e_{ij}^2$
  - $\text{var}(e_{ij}|\boldsymbol{x}_{ij}) = \tau_0^2 + \sigma^2 \longrightarrow e_{ij}^2$
  - $\text{var}(e_{ij}|\text{macro unit } = j) = \sigma^2.$

Random Intercept & Slope Model

- If the "true" model is a random intercept & slope model, then the residuals should equal

$$e_{ij} = U_{0j} + U_{1j}x_{ij} + R_{ij}$$

- Implications:
  - $e_{ij}|x_{ij} = U_{0j} + U_{1j}x_{ij} + R_{ij} \quad \rightarrow e_{ij}.$
  - $E[e_{ij}] = E[U_{0j} + U_{1j}x_{ij} + R_{ij}] = 0 \quad \rightarrow \bar{e}_{++}.$
  - $E[e_{ij}|\text{macro unit} = j] = U_{0j} + U_{1j}\mu_{x_{ij}} \quad \rightarrow \bar{e}_{+j}.$
  - $\text{var}(e_{ij}) = E[(e_{ij})^2] = \tau_0^2 + 2\tau_{10}x_{ij} + \tau_1^2 x_{ij}^2 + \sigma^2 \quad \rightarrow e_{ij}^2.$
  - $\text{var}(e_{ij}|x_{ij}) = \tau_0^2 + 2\tau_{10}x_{ij} + \tau_1^2 x_{ij}^2 + \sigma^2 \quad \rightarrow e_{ij}^2.$
  - $\text{var}(e_{ij}|\text{macro unit} = j) = \sigma^2.$

Comparison

Statistic	Expectation of Model w/ a Random Intercept	And Slope
$\bar{e}_{++}$	0	0
e_{ij}	$U_{0j} + R_{ij}$	$U_{0j} + U_{1j}x_{ij} + R_{ij}$
$\bar{e}_{+j}$	U_{0j}	$U_{0j} + U_{1j}\bar{x}_{+j}$
e_{ij}^2	$\tau_0^2 + \sigma^2$	$\tau_0^2 + 2\tau_{10}x_{ij} + \tau_1^2 x_{ij}^2 + \sigma^2$
$\text{Var}(e_{ij}\mid\text{macro} = j)$	σ^2	σ^2

Tools to Study & Examine e_{ij}'s

	"Prediction" of Model w/ a Random	
Plot	Intercept	And Slope
e_{ij} vs x_{ij} for j	flat	linear w/rt x_{ij}
e_{ij} vs x_{ij} over j	parallel	not parallel
		linear trend
$\bar{e}_{+j}$ vs $\bar{x}_j$ over j	differences but	linear trend
	not systematic	
e_{ij}^2 vs x_{ij} for j	flat	polynomial

Simulation and then Real Data

- Random Intercept Model:

$$Y_{ij} = 5 + 2x_{ij} + U_{0j} + R_{ij}$$

where $U_{0j} \sim \mathcal{N}(0,1)$, $R_{ij} \sim \mathcal{N}(0,4)$, and $x_{ij} \sim \mathcal{N}(0,2)$.

- Random Intercept and Slope Model:

$$Y_{ij} = 5 + 2x_{ij} + U_{0j} + U_{1j}x_{ij} + R_{ij}$$

where $U_{0j} \sim \mathcal{N}(0,1)$, $U_{1j} \sim \mathcal{N}(0,1)$, $\tau_{10} = 0$, $R_{ij} \sim \mathcal{N}(0,4)$, and $x_{ij} \sim \mathcal{N}(0,2)$.

- $N = 100$, and $n_j = 10$.

Estimated Fixed for Simulations

Using ordinary least squares (i.e., PROC/GLM or R/lm for fixed effect model).

		Random Intercept		. . . and Slope	
Parameter	"Actual"	Estimate	S.E.	Estimate	S.E.
intercept	5	4.9578	.0683	4.9795	.1086
slope	2	1.9611	.0229	1.8641	.0734
variance Y_{ij}	5*	4.6653	.2088	11.8014	.5283

For Random Intercept: var$(Y_{ij}) = \tau_0^2 + \sigma^2 = 1 + 4 = 5$.

For Random Intercept & Slope: more complicated.

Digression: Correct Models

The correct random effects model fit by MLE. . .

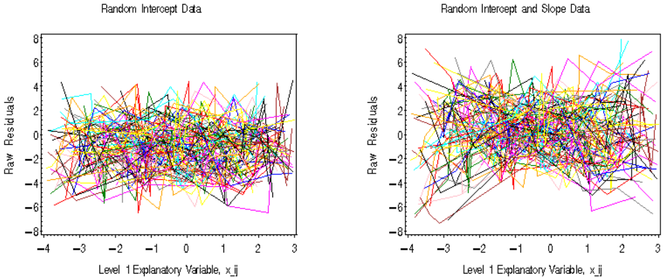
		Random Intercept		. . . and Slope	
Parameter	"Actual"	Estimate	S.E.	Estimate	S.E.
intercept	5	4.9582	.1106	4.9505	.1123
slope	2	1.9689	.0215	1.7973	.0930
τ_0^2	1	.8392	—	.8419	—
τ_{01}	0	—	—	.2189	—
τ_1^2	1	—	—	.8091	—
σ^2	4	3.8337	—	3.8144	—

Fixed effects estimates are similar, but their S.E.'s tend to be smaller with the "wrong" model.

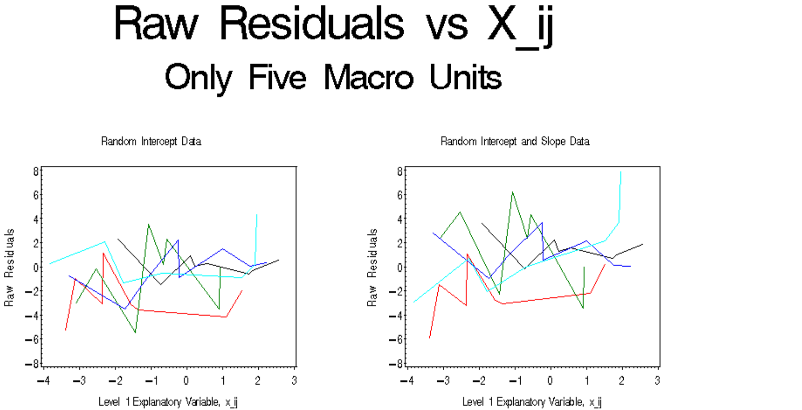
$(U_{0j} + R_{ij})$ and $(U_{0j} + U_{1j}x_{ij} + R_{ij})$

Raw Residuals vs X_ij

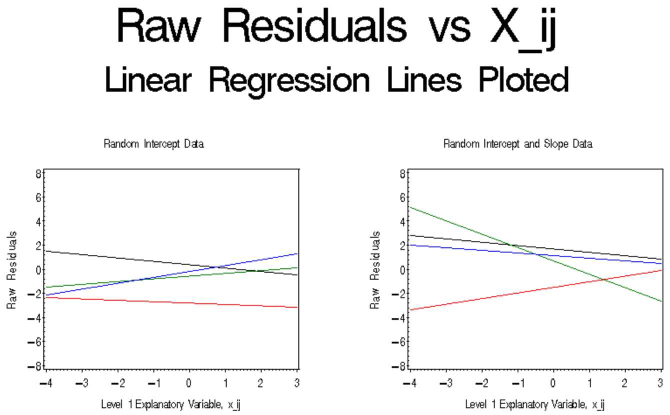
Separate Curvers per Macro



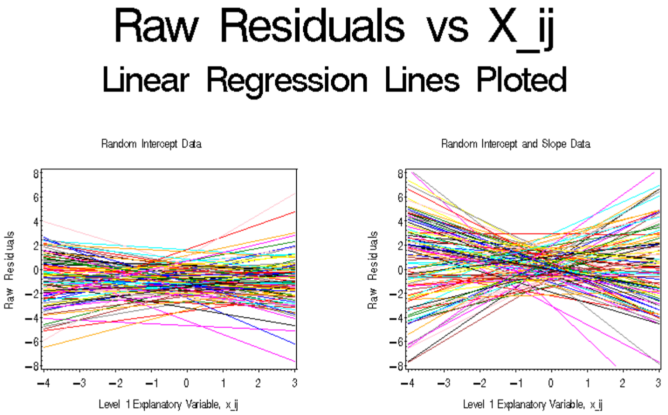
$(U_{0j} + R_{ij})$ and $(U_{0j} + U_{1j}x_{ij} + R_{ij})$



$(U_{0j} + R_{ij})$ and $(U_{0j} + U_{1j}x_{ij} + R_{ij})$



$(U_{0j} + R_{ij})$ and $(U_{0j} + U_{1j}x_{ij} + R_{ij})$



$(U_{0j} + R_{ij})$ and $(U_{0j} + U_{1j}x_{ij} + R_{ij})$

For each macro unit & each simulated data set, I fit the model

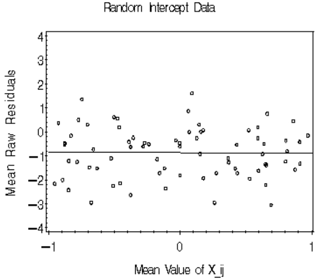
$$e_{ij} = \omega_{0j} + \omega_{1j}x_{ij}$$

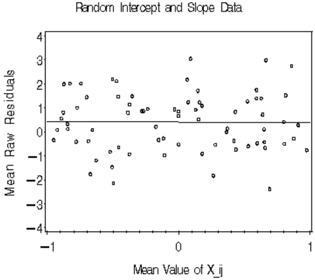
using ordinary least squares regression.

- For the random intercept model, the slope parameter was only "significant" 5 times while the intercept was 37 times (out of 100).
- Random intercept & slope data, the slope parameter was "significant" about 60 times while the intercept was 28 times (out of 100).

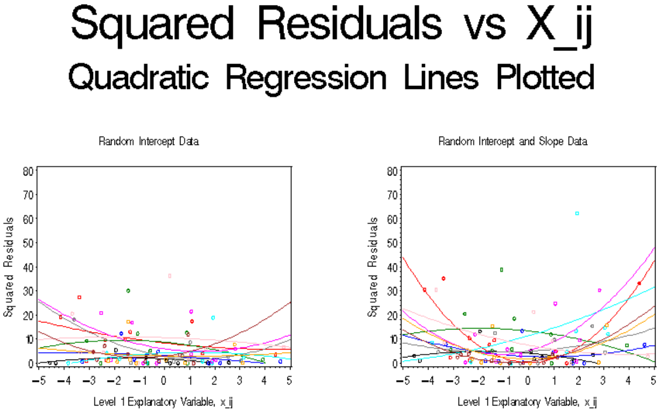
$(U_{0j} + \bar{R}_{+j})$ and $(U_{0j} + U_{1j}\bar{x}_{+j} + \bar{R}_{+j})$

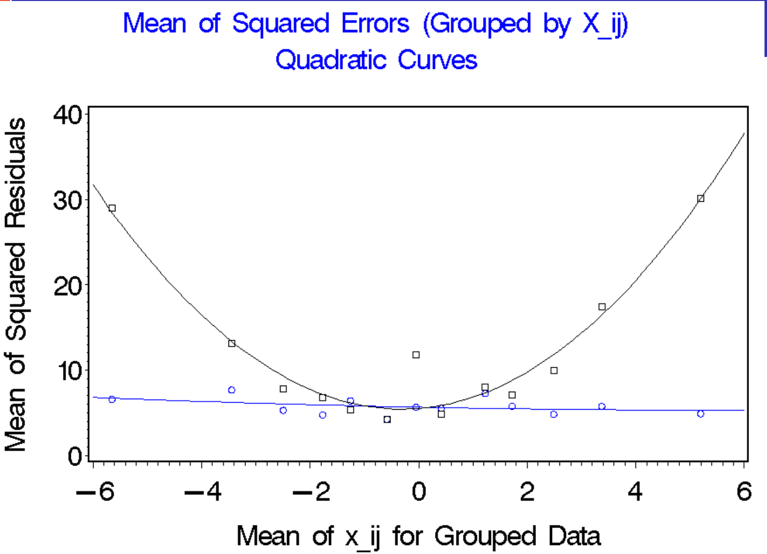
Mean Raw Residuals vs Mean X_+j





Variance Function for 10 Macro




Mean of Squared Errors (Grouped by X_ij)
Quadratic Curves
Mean of Squared Residuals
40
30
20
10
0
−6
−4
−2
0
2
4
6
Mean of x_ij for Grouped Data

Simulation 2: "Longitudinal"

- In the first simulation, within variance 4 times larger than between macro unit variance.
- In the next simulation, between variance 4 times larger than between variance.
- Same fixed effects model, except now:
  - $R_{ij} \sim \mathcal{N}(0, 1)$.
  - $U_{0j} \sim \mathcal{N}(0, 4)$ and $U_{1j} \sim \mathcal{N}(0, 4)$.
  - cov$(U_{oj}, U_{1j}) = 0$.

Preliminary Fixed Estimates

Using ordinary least squares (i.e., PROC/GLM or the `lm` package in R for fixed effect model).

		Random Intercept		. . . and Slope	
Parameter	Actual	Estimate	S.E.	Estimate	S.E.
intercept	5	4.9754	.0696	4.8168	.1881
slope	2	1.9645	.0234	1.7705	.0635
variance Y_{ij}	5^*	4.8360	.2165	35.3773	1.5837

Note: $\sigma^2 = 1$, $\tau_0^2 = 4$, $\tau_1^2 = 4$, and $\tau_{01} = 0$

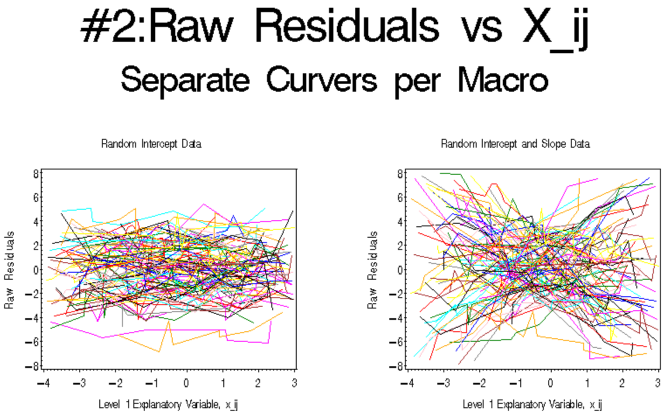
#2 Digression: Correct Models

MLE & the appropriate random effects model...

		Random Intercept		...and Slope	
Parameter	"Actual"	Estimate	S.E.	Estimate	S.E.
intercept	5	4.9763	.2002	4.9730	.2006
slope	2	1.9859	.0109	1.6372	.1858
τ_0^2	4	3.9131	—	3.9190	—
τ_{01}	0	—	—	.9206	—
τ_1^2	4	—	—	3.4356	—
σ^2	1	.9584	.0452	.9555	—

Fixed effects estimates are similar, but their S.E.'s tend to be smaller with the "wrong" model.

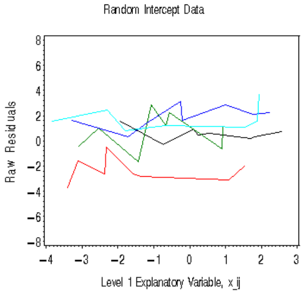
$(U_{0j} + R_{ij})$ and $(U_{0j} + U_{1j}x_{ij} + R_{ij})$

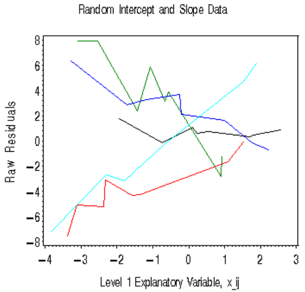


$(U_{0j} + R_{ij})$ and $(U_{0j} + U_{1j}x_{ij} + R_{ij})$

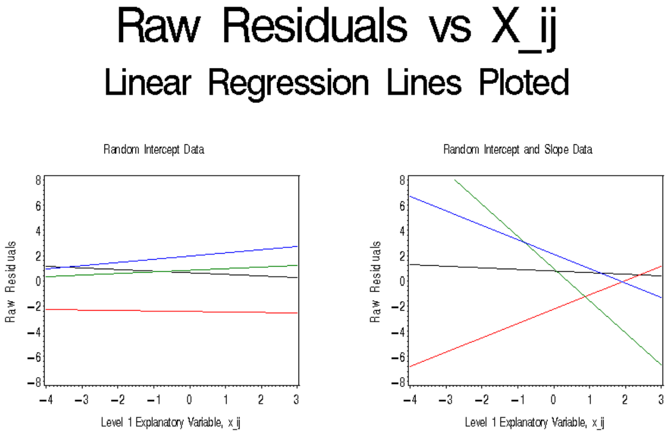
Raw Residuals vs X_ij

Only Five Macro Units





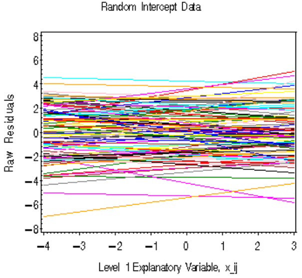
$(U_{0j} + R_{ij})$ and $(U_{0j} + U_{1j}x_{ij} + R_{ij})$

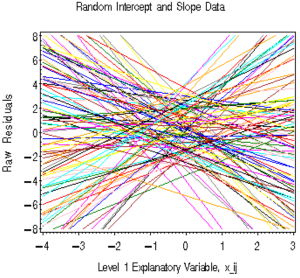


$(U_{0j} + R_{ij})$ and $(U_{0j} + U_{1j}x_{ij} + R_{ij})$

#2: Raw Residuals vs X_ij

Linear Regression Lines Ploted





$(U_{0j} + R_{ij})$ and $(U_{0j} + U_{1j}x_{ij} + R_{ij})$

For each macro unit & data set,

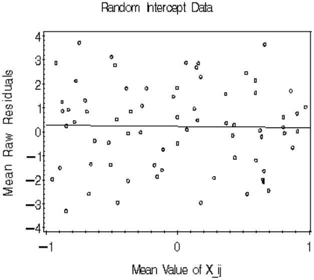
$$e_{ij} = \omega_{0j} + \omega_{1j}x_{ij}$$

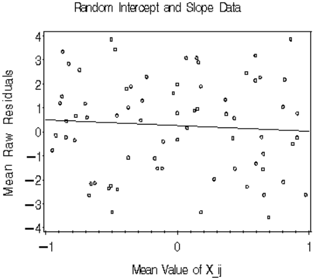
using ordinary least squares regression.

- For the random intercept data, the slope parameter was only "significant" 4 times whereas the intercept was 78 times (out of 100).
- Random intercept & slope data, the slope parameter was "significant" about 92 times whereas the intercept was 77 times (out of 100).

$(U_{0j} + \bar{R}_{+j})$ and $(U_{0j} + U_{1j}\bar{x}_{+j} + \bar{R}_{+j})$

#2: Mean Raw Residuals vs Mean X_+j

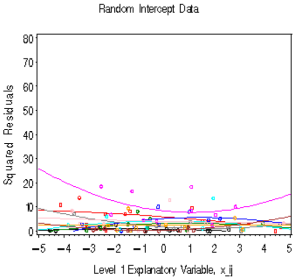


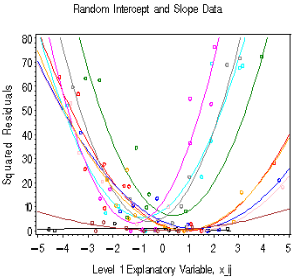


Variance Function for 10 Macro

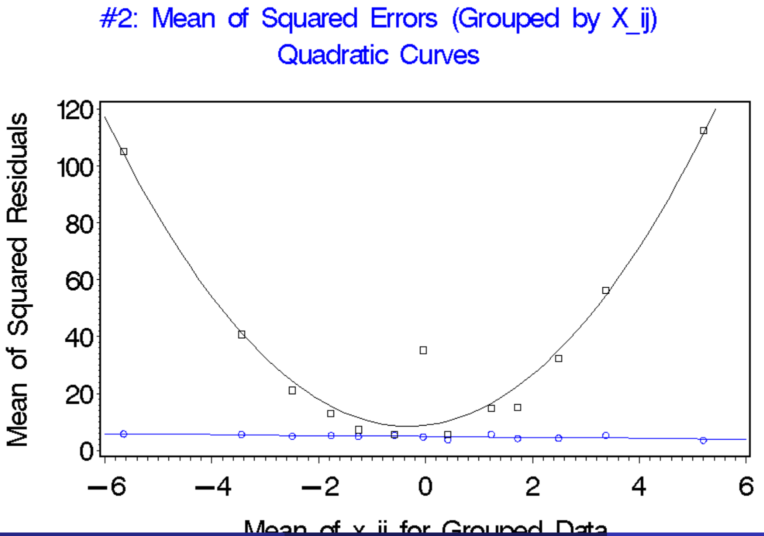
#2: Squared Residuals vs X_ij

Quadratic Regression Lines Plotted





Variance Function



Summary of Findings

- The larger τ^2 is relative to σ, the easier to "see" the expected patterns.
- Fitted curves help reduce noise.
- The variance function (squared residuals) by macro unit useful.
- The most clear-cut results when descretizing x_{ij} into k groups, and computing mean x_k^* and mean of squared residuals within the groups.
- With real data, patterns expected to be much noisier, which is where fitted curves becomes even more useful.

Now for Real Data: NELS88 $N = 23$

- Recall preliminary fixed model

$$\begin{aligned}(\text{math})_{ij} \;=\;& \gamma_{00} + \gamma_{01}(\text{sector})_j + \gamma_{02}(\overline{\text{SES}})_j \\ &+\gamma_{10}(\text{homework})_{ij} + \gamma_{11}(\text{sector})_j(\text{homework})_{ij} \\ &+\gamma_{20}(\text{cSES})_{ij} + \gamma_{30}(\text{white})_{ij} + \gamma_{40}(\text{female})_{ij} \\ &+\gamma_{50}(\text{cSES})_{ij}(\text{white})_{ij} + R_{ij}\end{aligned}$$

- I used GLM (or PROC/MIXED without RANDOM) and fit this regression model and saved $\hat{e}_{ij}$ to a SAS data file

$$\hat{e}_{ij} = (\text{math})_{ij} - \widehat{(\text{math})}_{ij}$$

SAS/GLM and NELS88 $N = 23$

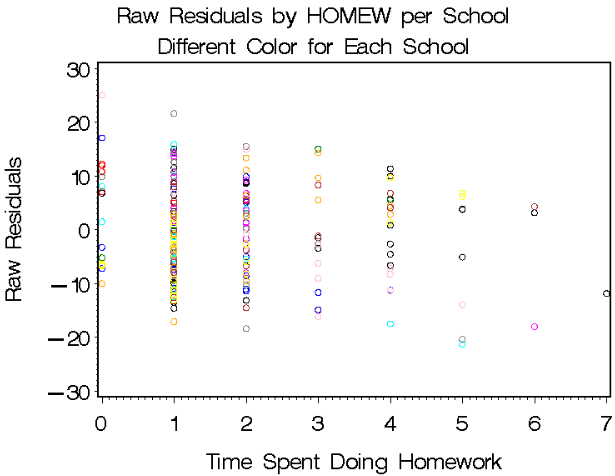
```
PROC glm data=school23;
   CLASS school white sex public;
   MODEL math = homew cses white sex public
      public*homew cses*white gmeanses;
   OUTPUT out=model1 r=rmath p=pmath student=stdres;
   TITLE 'Fit Preliminary Fixed Effects Model';

DATA school23;
   SET model1;
   sqrmath = rmath*rmath;
RUN;
```

Plan

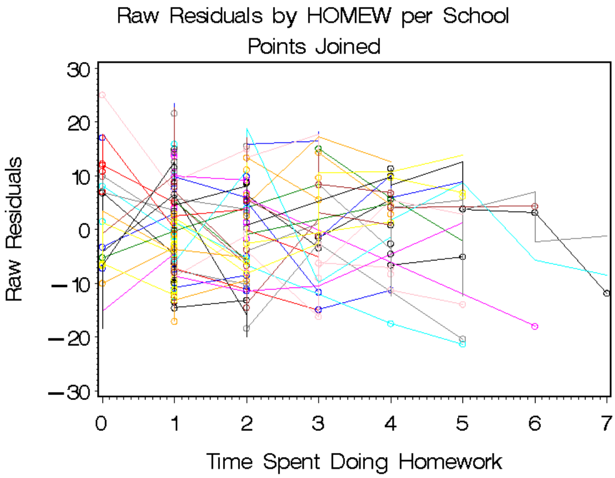
For each level one variable

- Study raw residuals.
- Study squared residuals.
- Does analysis indicate need a random intercept?
- Does analysis indicate possible random slope?
- If need random slope and/or random intercept, study possible level 2 explanatory variables.

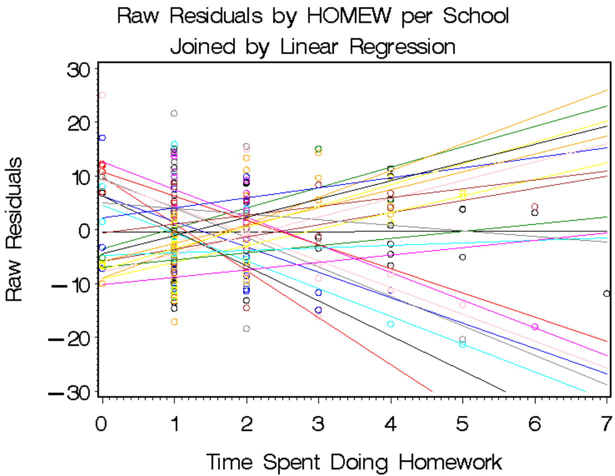
Raw Residuals by HOMEW



Raw Residuals by HOMEW

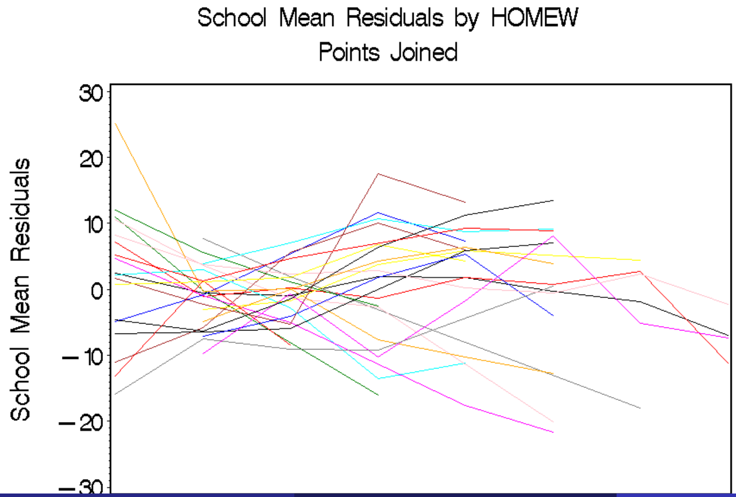


Raw Residuals by HOMEW



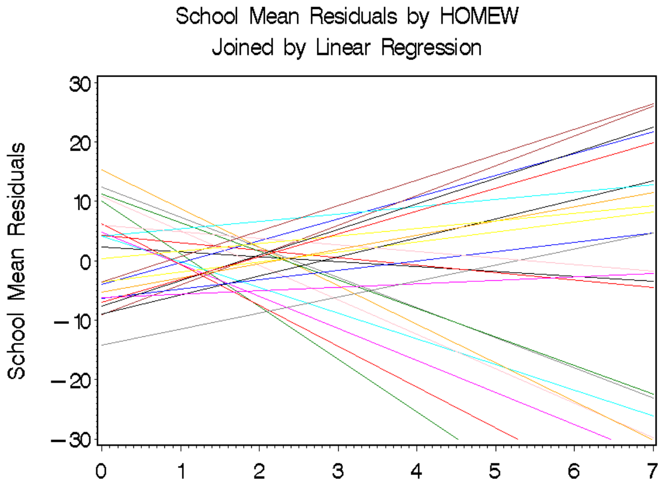
Mean Raw Residuals by HOMEW

Means computed for each school and homework level.

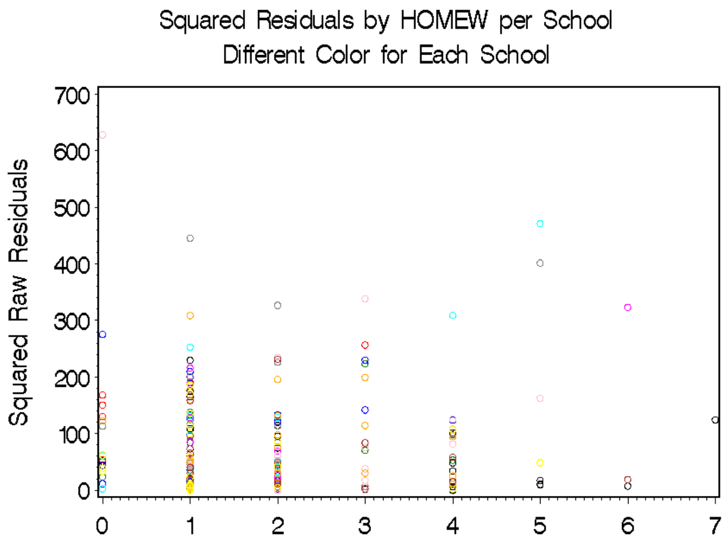


Mean Raw Residuals by HOMEW

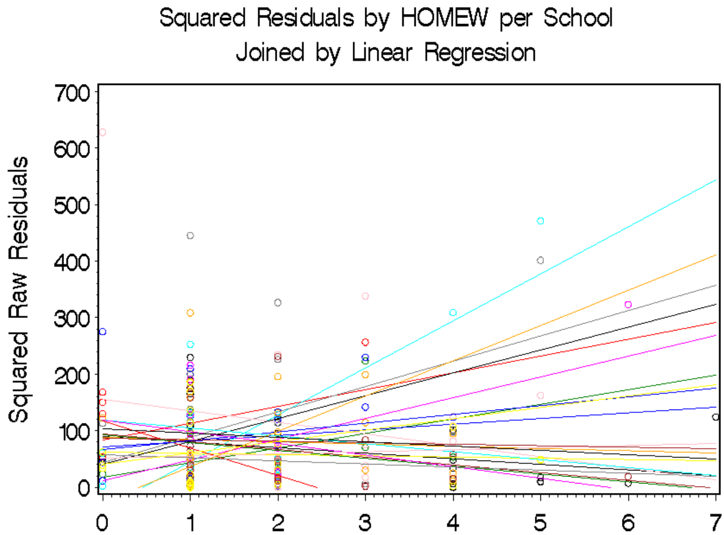
Means computed for each school and homework level.



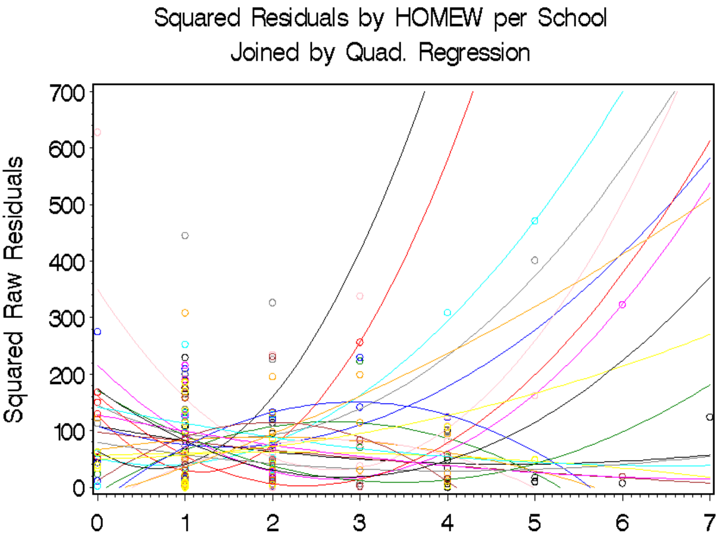
Variance Function for HOMEW



Variance Function for HOMEW

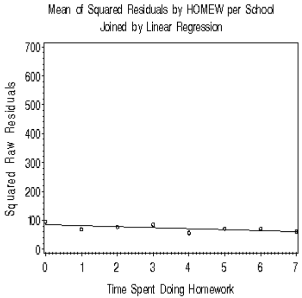


Variance Function for HOMEW



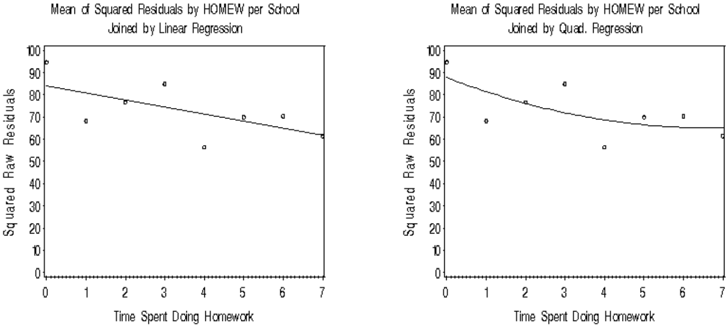
Variance Function for HOMEW

Mean of Squared Residuals
Linear versus Quadratic



Variance Function for HOMEW

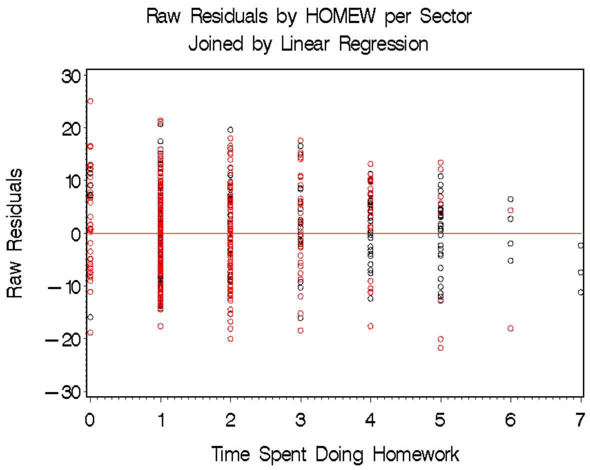
Mean of Squared Residuals
Linear versus Quadratic



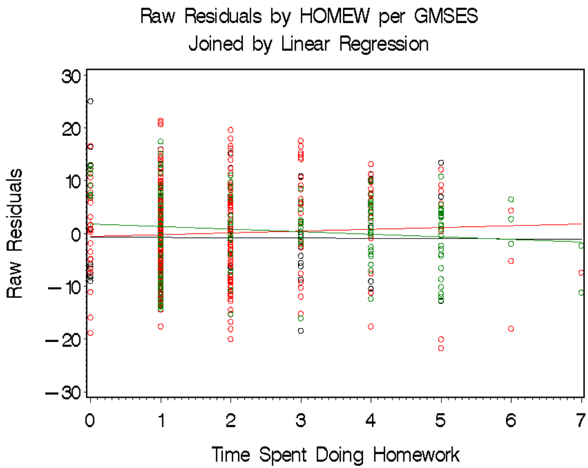
Conclusions So Far

- Evidence that we need a random intercept.
- Evidence that we need a random slope for HOMEW.
- Potential explanatory variables for random effects?
  - Sector (public/private).
  - School mean SES.

Sector as an Explanatory Variable



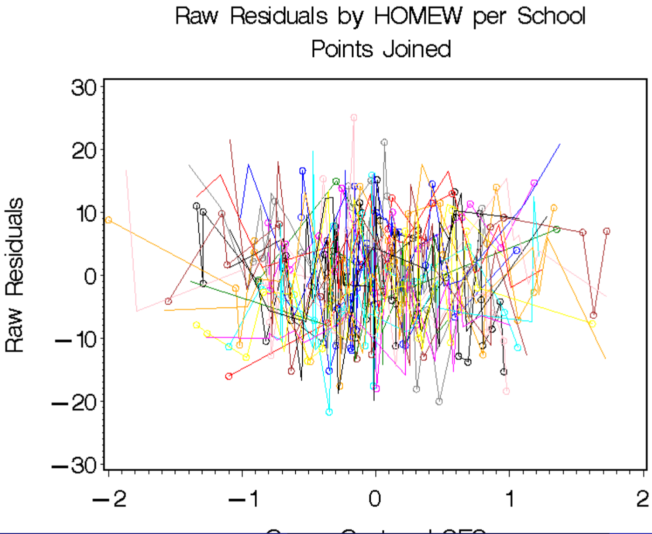
School Mean SES



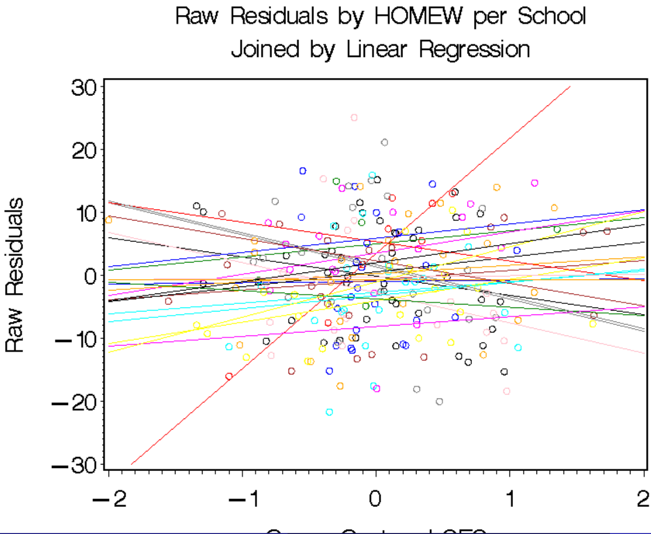
Conclusions So Far

- Need a random intercept and sector & group mean SES are potential level 2 explanatory variables (this is from studying mean structure and random structure).
- Evidence that we need a random slope for HOMEW but neither sector nor school mean SES useful predictors of slope.
- Study School Mean SES next...a more “continuous” variable.

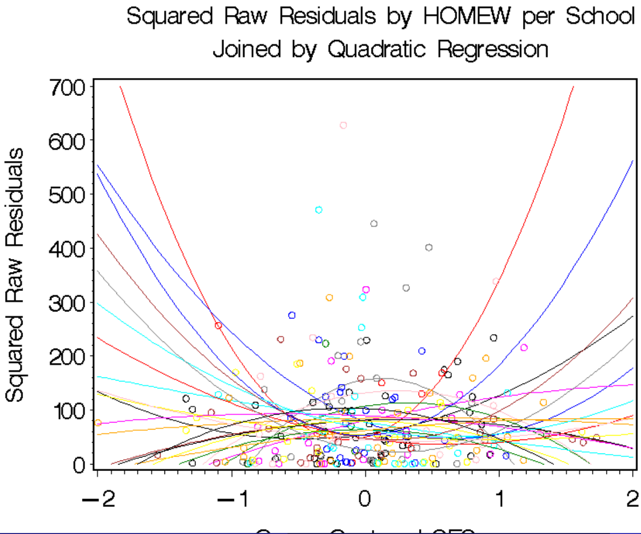
Raw Residuals and cSES



Raw Residuals and cSES

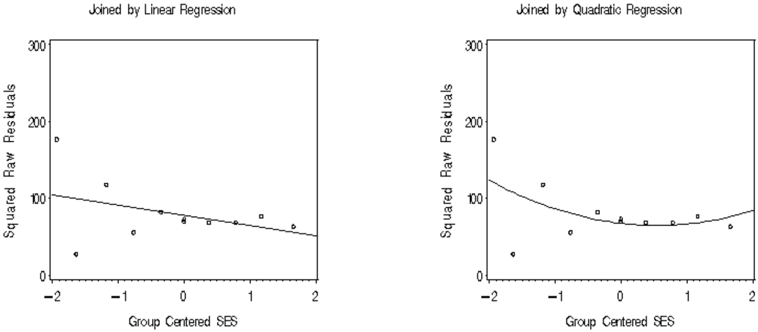


Squared Residuals and cSES

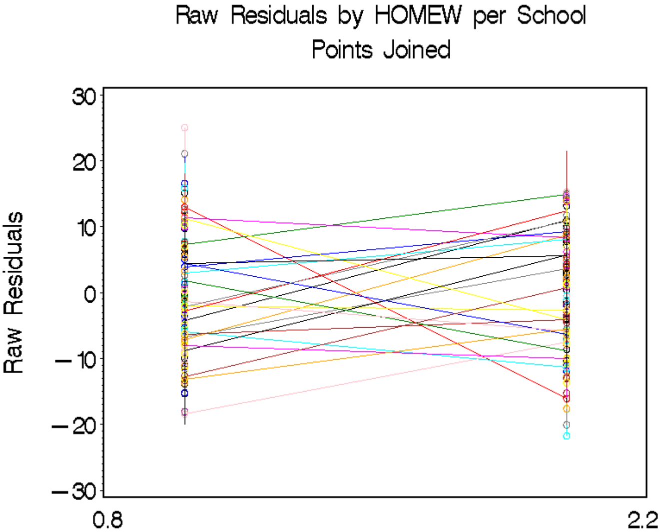


Squared Residuals and cSES

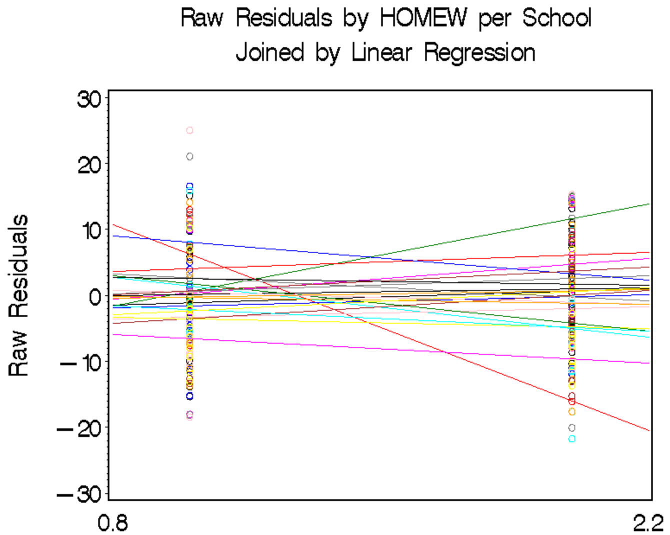
Mean of Squared Residuals
Linear versus Quadratic



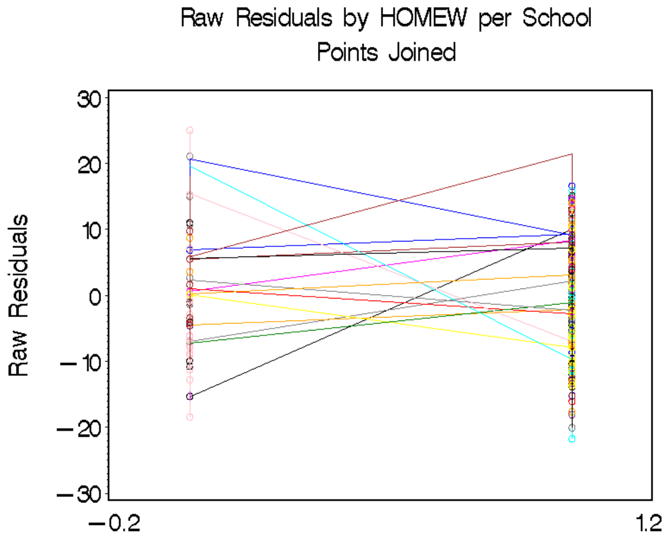
Raw Residuals and Gender



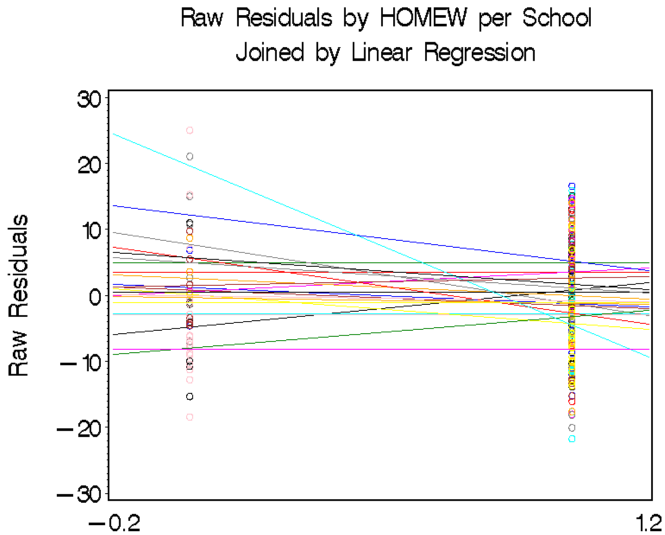
Raw Residuals and Gender



Raw Residuals and White



Raw Residuals and White



Conclusions Regarding Random Effects

- Need a random intercept (school mean SES and sectors are possible explanatory variables).
- Random slope for time spent doing homework.
- Do not need a random slope for school mean centered SES, gender, or race.

Preliminary HLM

Level 1:

$$\begin{aligned}
(\text{math})_{ij} &= \beta_{0j} + \beta_{1j}(\text{homework})_{ij} + \beta_{2j}(\text{cSES})_{ij} \\
&\quad + \beta_{3j}(\text{white})_{ij} + \beta_{4j}(\text{female})_{ij} \\
&\quad + \beta_{5j}(\text{cSES})_{ij}(\text{white})_{ij} + R_{ij}
\end{aligned}$$

Level 2:

$$\begin{aligned}
\beta_{0j} &= \gamma_{00} + \gamma_{01}(\overline{\text{SES}})_j + \gamma_{02}(\text{sector})_j + U_{01j} \\
\beta_{1j} &= \gamma_{10} + U_{1j} \\
\beta_{2j} &= \gamma_{20} \\
\beta_{3j} &= \gamma_{30} \\
\beta_{4j} &= \gamma_{40} \\
\beta_{5j} &= \gamma_{50}
\end{aligned}$$

Preliminary Linear Mixed Model

$$
\begin{aligned}
(\text{math})_{ij} \quad = \quad & \gamma_{00} + \gamma_{10}(\text{homework})_{ij} + \gamma_{20}(\text{cSES})_{ij} \\
& +\gamma_{30}(\text{white})_{ij} + \gamma_{40}(\text{female})_{ij} \\
& +\gamma_{50}(\text{cSES})_{ij}(\text{white})_{ij} \\
& +\gamma_{01}(\overline{\text{SES}})_{j} + \gamma_{02}(\text{sector})_{j} \\
& +U_{0j} + U_{1j}(\text{homework})_{ij} + R_{ij}
\end{aligned}
$$

We also try random slopes for other level 1 variables.

Another model to try: school mean centered homework and mean homework per school.?

Model Information

Model Information

Data Set	WORK.SCHOOL23C
Dependent Variable	MATH
Covariance Structure	Unstructured
Subject Effect	SCHOOL
Estimation Method	ML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Satterthwaite

Dimensions

Covariance Parameters	4
Columns in X	10
Columns in Z Per Subject	2
Subjects	23
Max Obs Per Subject	67

Number of Observations

Number of Observations Read	519
Number of Observations Used	519
Number of Observations Not Used	0

Convergence criteria met.

Comparison

		Parameters		Fit Statistics			
						BIC	
Model	#	$\hat{\tau}$'s	$\hat{\sigma}^2$	-2LnLike	AIC	n_{++}	new
Baseline	3	24.85	81.24	3800.7764	3806.8	3810.2	3813.3
Preliminary	11	45.69	51.05	3598.2708	3622.3	3635.9	3660.8
		-22.39					
		13.30					

Note:

$$\hat{\rho}_I = \frac{24.8503}{24.8503 + 81.2374} = .23$$

Solution for Fixed Effects

Effect		Estimate	Standard Error	DF	t Value	Pr > $\|t\|$
Intercept		48.1265	1.7849	33.7	26.96	< .0001
HOMEW		1.8534	0.8151	20.7	2.27	0.0338
cSES		2.6483	0.6545	489	4.05	< .0001
gender	Fema	−0.4705	0.6641	491	−0.71	0.4790
gender	Male	0	.	.	.	.
whitec	Not W	2.4595	0.9843	373	−2.50	0.0129
whitec	White	0	.	.	.	.
cSES*whitec	Not W	−1.4039	1.1613	493	−1.21	0.2273
cSES*whitec	White	0	.	.	.	.
meanSES		5.2698	1.8133	25	2.91	0.0076
PUBLIC	0	−0.09328	2.1126	25.5	−0.04	0.9651
PUBLIC	1	0	.	.	.	.

R Summary (Nelder-Mead, MLE)

Effect	est.	se	
(Intercept)	45.57	(2.11)	***
homew	1.85	(0.82)	*
schCses	1.24	(0.95)	
sex2	-0.47	(0.66)	
white1	2.46	(0.98)	*
schMses	5.27	(1.81)	**
public1	0.09	(2.11)	
schCses:white1	1.40	(1.16)	
AIC	3622.27		
BIC (new)	3660.83		
Log Likelihood	-1799.14		
Num. obs.	519		
Num. groups: school	23		
Var: school (Intercept)	45.69		
Var: school homew	13.30		
Cov: school (Intercept) homew	-22.40		

SAS: Additional Random Effects?

- I also tried random effects for each of the other level 1 explanatory variables (along with random intercept and random slope for homework); however, all of these models yields "bad".
  - cses: "Estimated G matrix is not positive definite."
  - gender: "WARNING: Did not converge."
    R: boundary (singular) fit: see ?isSingular
  - white: It converges to an OK solution, but $AIC = 3623.45$ BIC.new $= 3665.41$, which are larger than model w/o white
- If we used all the data, results could differ.

R: Additional Random Effects?

- I also tried random effects in R. The following messages were given and also yields "bad".
  - cses: "boundary (singular) fit: see ?isSingular
  - gender: "boundary (singular) fit: see ?isSingular "
  - white: It converges to an OK solution, but $AIC = 3623.5$ and $BIC.new = 3665.41$. These are a only a littler larger than prelimnary model.

Model Reduction

Plan:

- See if we can simply the covariance structure
- Test fixed effects.

Testing Random Slope

- $\boldsymbol{T}$ is "positive semi-definite"…it's a proper covariance matrix; therefore, the following test is valid.
- Hypothesis test:

$$H_o : \tau_{10} = \tau_1^2 = 0 \quad \text{vs} \quad H_a : \text{ not } H_o$$

- Fit

$$\begin{aligned}(\text{math})_{ij} \quad = \quad & \gamma_{00} + \gamma_{10}(\text{homework})_{ij} + \gamma_{20}(\text{cSES})_{ij} \\ & +\gamma_{30}(\text{white})_{ij} + \gamma_{40}(\text{female})_{ij} \\ & +\gamma_{01}(\overline{\text{SES}})_j + \gamma_{02}(\text{sector})_j \\ & +U_{0j} + R_{ij}\end{aligned}$$

Testing Random Slope (continued)

Model	No. of param	-2lnLike	Test stat.	p-value
Null	9	3673.6011	75.33	$<< .0001$
Prelim	11	3598.2708		

Note: the p-value is a mixture of "p-values" from χ^2_1 and χ^2_2, both of which are tiny.

We need a random slope for `homew`.

Testing Fixed Effects Structure

From the t-tests,

- `homew` has statistically significant random slope, so we keep it in the model regardless of t-test result.
- Gender is not significant (not surprising given exploratory analyses).
- Sector is not significant (somewhat surprising given exploratory analyses).

Explanation why sector not significant.

Why Sector Not Significant

If we remove $(\overline{\text{SES}})_j$ from the model:

Effect	estimate	s.e.
(Intercept)	48.10	$(2.29)^{***}$
homew	1.83	$(0.83)^{*}$
schCses	1.15	(0.95)
sex	-0.41	(0.66)
white	2.66	$(1.00)^{**}$
public	-4.19	$(1.73)^{*}$
schCses:white	1.48	(1.16)
AIC	3627.30	
BIC.new	3661.61	
Log Likelihood	-1802.65	
Var: school (Intercept)	48.00	
Var: school homew	13.65	
Cov: school (Intercept) homew	-22.16	
Var: Residual	50.98	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Comments Regarding Sector and $(\overline{\text{SES}})_j$

- The $\hat{\gamma}$ for public goes from -0.09 (s.e. 2.11) in the preliminary model to 4.19 (s.e. 1.73) in model without mean SES.

Model	#	-2LnLike	AIC	$BIC.new$
Preliminary	11	3598.27	3622.3	3660.8
No $(\overline{\text{SES}})_j$	10	3605.30	3627.3	3661.6

- Likelihood ratio test statistic $(\overline{\text{SES}})_j$
  lr $= = 3605.3011 - 3598.2708 = 7.03$ on $df = 1$, p-value$< .001$.

- Information criteria support preliminary model.

Testing Fixed Effects

- F-test for fixed effects for gender and sector:

$$H_o : \begin{pmatrix} \gamma_{40} \\ \gamma_{02} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{vs} \quad H_o : \begin{pmatrix} \gamma_{40} \\ \gamma_{02} \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

- SAS command

```
contrast 'gender & sector'
    sex 1 -1 ,
   public 1 -1;
```

- Result: $F = .25$ with $df = 2$ and 46.9, p-value$= .78$.

Testing Fixed Effects

- Likelihood ratio test for no gender and no sector hypothesis, i.e.,

$$H_o : \begin{pmatrix} \gamma_{40} \\ \gamma_{02} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{vs} \quad H_o : \begin{pmatrix} \gamma_{40} \\ \gamma_{02} \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

- Test statistic$= 3598.7723 - 3598.2708 = .5$ with $df = 2$. Comparing this to χ^2_2 distribution gives$p$-value$= .78$.

Testing Fixed Effects (continued)

- Summary of global fits:

Model	#	-2LnLike	AIC	$BIC.new$
Preliminary	11	3598.2708	3622.3	3660.8
No $(\overline{\text{SES}})_j$	10	3605.3011	3627.3	3661.6
No sector & no gender	9	3598.7723	3618.8	3648.8

- Conclusion: Remove gender and sector from our model.

Fixed Effects

	Simpler Model	
Effect	estimate	s.e.
(Intercept)	45.43	$(1.71)^{***}$
homew	1.83	$(0.82)^{*}$
schCses	1.32	(0.94)
white	2.44	$(0.98)^{*}$
schMses	5.19	$(1.27)^{***}$
schCses:white	1.33	(1.15)

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Remove `cSES*whitec`?

Testing Removal of interaction

- Summary of global fits:

Model	#	-2LnLike	AIC	$BIC.new$
Preliminary	12	3598.27	3622.3	3660.8
No $(\overline{\text{SES}})_j$	11	3605.30	3627.3	3661.6
No sector	10	3598.77	3618.8	3648.8
& no gender				
& no ses$\times$white	9	3600.08	3618.1	3643.9

- $LR = 3600.0829 - 3598.7723 = 1.31$, $df = 1$, $p = .25$.
- Conclusion: Remove interaction.

Possible Final Model)

	Final Model	
(Intercept)	45.65	$(1.71)^{***}$
homew	1.83	$(0.83)^{*}$
schCses	2.21	$(0.53)^{***}$
white	2.22	$(0.96)^{*}$
schMses	5.18	$(1.28)^{***}$
AIC	3618.1	
BIC.new	3643.9	
Deviance	3600.1	
Var: school (Intercept)	46.61	
Var: school homew	13.80	
Cov: school (Intercept) homew	-23.04	
Var: Residual	51.12	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Interpretation

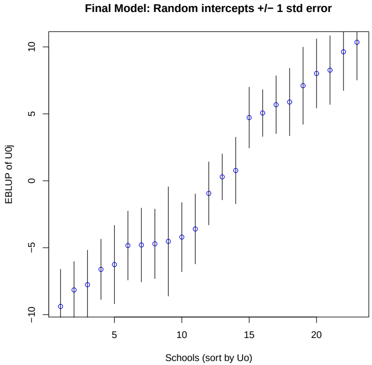
$$\begin{aligned}
(\mathsf{math})_{ij} \quad = \quad & (45.65 + 5.18(\overline{\mathsf{SES}})_j + U_{oj}) \\
& +(1.83 + U_{1j})(\mathsf{homework})_{ij} \\
& +2.65(\mathsf{cSES})_{ij} \\
& -2.22(\mathsf{white})_{ij} \\
& +R_{ij}
\end{aligned}$$

Final Model

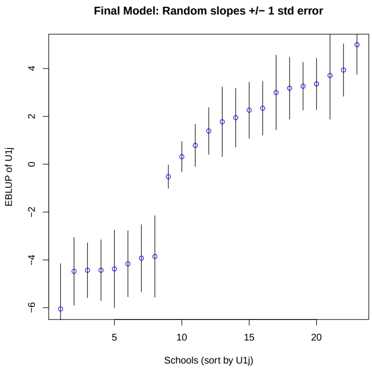
Covariance Parameter Estimates

	Final Model
Var: school (Intercept)	46.61
Var: school homew	13.80
Cov: school (Intercept) homew	-23.04
Var: Residual	51.12

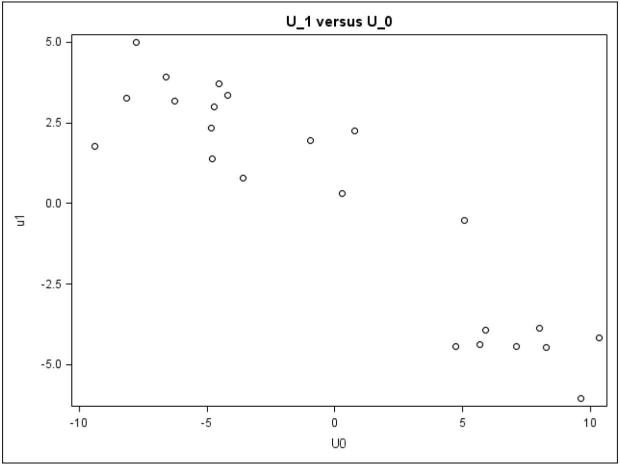
Final Model: $\hat{U}_{0j}$



Final Model: $\hat{U}_{1j}$



Final Model: $\hat{U}_0$ vs $\hat{U}_{1j}$



Final Model: Evidence of Fit

$$r((\text{math})_{ij}, \boldsymbol{X}'_{ij}\hat{\boldsymbol{\Gamma}}) = .60$$

$$r((\text{math})_{ij}, \boldsymbol{X}'_{ij}\hat{\boldsymbol{\Gamma}} + \boldsymbol{Z}_j\hat{\boldsymbol{U}}_j) = .77$$

$$R_1^2 = .14$$

$$R_2^2 = .32$$

Note: harmonic mean $= 17.73$

Model Diagnostics

- Model Assumptions:
  - Linear structure and explanatory variables.
  - $R_{ij} \sim \mathcal{N}(0, \sigma^2)$.
  - $\boldsymbol{U}_j \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{T})$.
- Influence of observations on
  - Model fit to data.
  - Estimated fixed effects.
  - Estimates of standard errors.
  - Estimates of covariances and variances.
  - Fitted and predicted values

σ^2 Same for All?

- Will illustrate using a Random Intercept Model model which I fit I using SAS/PROC NLMIXED.
- Since Leckie uses HSB data, we'll use it here.
- See (among others): Leckie, et al. (2014). Modeling heterogeneous variance-covariance components in two-level models. *JEBS, 39*, 307-332.
- Use a stand alone MIXREGLS program, which can be called from R or Stata or it can be run in a cmd window. See Hedeker & Nordgren (2013) and for an extended version see Nordgren, Hedeker, Dunton, & Yang (2019). Or use SAS/PROC NLMIXED.
- For random intercept & slope $\longrightarrow$ could also use Bayesian estimation.
- more on estimation after looking at these models.

Random Intercept and Heterogeneous σ^2

Level 1

$$Y_{ij} = \beta_{0j} + \beta_{1j} x_{1ij} + \ldots + R_{ij}$$

where $R_{ij} \sim N(0, \sigma^2)$
Level 2:

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01} z_{1j} + \ldots + U_{0j} \\ \beta_{1j} &= \gamma_{10} \\ &\vdots \end{aligned}$$

$U_{0j} \sim N(0, \tau^2)$ i.i.d and independent of R_{ij}
For now, we will consider σ^2 to be a random variable.

Modeling Heterogeneous σ^2

- Since $\sigma^2 \geq 0$, we'll model $\log(\sigma^2)$, which can be $-\infty$ to $+\infty$.
- $\log(\sigma^2_{ij})$ could be itself random; that is,

$$\log(\sigma^2_{ij}) = \omega_0 + V_{0j}, \qquad \text{or} \qquad \sigma^2_{ij} = \exp(\omega_0 + V_{0j})$$

  where $V_{0j} \sim N(0, i\psi^2)$ i.i.d.
- Can also add predictors to explain heterogeneity,

$$\log(\sigma^2_{ij}) = \omega_0 + \omega_1 w_{1j} + \ldots + V_{0j}.$$

Example: Modeling Heterogeneous of σ^2

HSB data, Level 1:

$$\text{math}_{ij} = \beta_{0j} + R_{ij} \qquad \text{where} \qquad R_{ij} \sim N(0, \sigma^2_{ij})\ i.i.d.$$

and

$$\log(\sigma^2_{ij}) = \omega_0 + V_{0j} \qquad \text{where} \qquad V_{0j} \sim N(0, \psi^2)\ i.i.d.$$

Level 2:

$$\beta_{0j} = \gamma_{00} + \ldots + U_{0j}$$

$U_{0j} \sim N(0, \tau^2)$ i.i.d and independent of R_{ij}

Null HLM and Simple Heterogeneity

	Null		Simple	
Effect	est	(se)	est	(se)
γ_{00}	12.64	(0.24)	12.645	(0.246)
ω_0	—	—	3.657	(0.022)
σ^2	39.15	(0.66)	—	—
ψ^2	—	—	0.032	(0.009)
τ^2	8.55	(1.07)	8.694	(1.078)
-2lnlike	47116		47093	
AIC	47122		47101	
BIC	47131		47113	

$$\sigma_j^2 = \exp(\omega_0) = \exp(3.657) = 38.7601$$

Example: Add predictors for μ_{ij}

HSB data, Level 1:

$$\text{math}_{ij} = \beta_{0j} + \beta_{1j}\text{female}_{ij} + \beta_{2ij}(cSES)_{ij} + R_{ij}$$

where $R_{ij} \sim N(0, \sigma^2_{ij})$ $i.i.d.$ and

$$\log(\sigma^2_{ij}) = \omega_0 + V_{0j} \qquad \text{where} \qquad V_{0j} \sim N(0, \psi^2)\ i.i.d.$$

Level 2:

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01}\text{sector}_j + \gamma_{02}\overline{\text{SES}}_j + U_{0j} \\ \beta_{1j} &= \gamma_{10} \\ \beta_{2j} &= \gamma_{20} \end{aligned}$$

$U_{0j} \sim N(0, \tau^2)$ i.i.d and independent of R_{ij}

More Complex μ_{ij}

	Null		Simple		Complex μ_{ij}		
Effect	est	(se)	est	(se)	est	(se)	
γ_{00}	12.64	(0.24)	12.645	(0.246)	12.724	(0.209)	intercept
γ_{01}					-1.198	(0.163)	sector
γ_{02}					5.230	(0.354)	mean ses
γ_{10}					-1.198	(0.163)	female
γ_{20}					2.127	(0.109)	cses
ω_0			3.657	(0.022)	3.600	(0.020)	
σ^2	39.15	(0.66)					
ψ^2			0.032	(0.009)	0.014	(0.007)	
τ^2	8.55	(1.07)	8.694	(1.078)	2.122	(0.338)	
-2lnlike	47116		47093		46505		
AIC	47122		47101		46505		
BIC	47131		47113		46529		

$$\sigma_j^2 = \exp(\omega_0) = \exp(3.600) = 36.6063$$

Example: Add predictors for σ^2

HSB data, Level 1:

$$\text{math}_{ij} = \beta_{0j} + \beta_{1j}\text{female}_{ij} + \beta_{2ij}(cSES)_{ij} + R_{ij}$$

where $R_{ij} \sim N(0, \sigma^2_{ij})$ $i.i.d.$ and

$$\log(\sigma^2_{ij}) = \omega_0 + \omega_1\text{sector}_j + \omega_2\overline{\text{SES}}_j + V_{0j}$$

where $V_{0j} \sim N(0, \psi^2)$ $i.i.d.$
Level 2:

$$\begin{aligned}\beta_{0j} &= \gamma_{00} + \gamma_{01}\text{sector}_j + \gamma_{02}\overline{\text{SES}}_j + U_{0j}\\ \beta_{1j} &= \gamma_{10}\\ \beta_{2j} &= \gamma_{20}\end{aligned}$$

$U_{0j} \sim N(0, \tau^2)$ i.i.d and independent of R_{ij}

More Complex σ^2

	Null		Simple		Complex μ_{ij}	Complex σ_j^2	
Effect	est	(se)	est	(se)	est (se)	est (se)	
intercept	12.64	(0.24)	12.65	(0.25)	12.724 (0.209)	12.719 (0.210)	
sector					1.258 (0.294)	1.259 (0.294)	
mean ses					5.230 (0.354)	5.212 (0.355)	
female					-1.198 (0.163)	-1.189 (0.164)	
cses					2.127 (0.109)	2.097 (0.110)	
ω_0					3.66 (0.020)	3.600 (0.020)	3
sector						-0.162 (0.039)	
mean ses						-0.072 (0.048)	
σ^2	39.15	(0.66)					
ψ^2			0.03	(0.01)	0.014 (0.007)	0.006 (0.006)	
τ^2	8.55	(1.07)	8.69	(1.08)	2.122 (0.338)	2.114 (0.337)	

More Complex σ^2

	Null	Simple	Complex μ_{ij}	Complex σ^2
-2lnlike	47116	47093	46505	46463
AIC	47122	47101	46505	46483
BIC	47131	47113	46529	46514

$$\hat{\sigma}_j^2 = \begin{cases} \exp(3.60 - 0.162 + 0) = 33.6440 & \text{Catholic \& } \overline{\text{SES}}_j = 0 \\ \exp(3.6) = 39.575 & \text{Public \& } \overline{\text{SES}}_j = 0 \end{cases}$$

For 1 unit increase in mean SES,

$$\hat{\sigma}_j^2 = \begin{cases} \exp(3.60 - 0.162 - .072) = 31.320 & \text{Catholic + 1 unit } \overline{\text{SES}}_j \\ \exp(3.6 - .072) = 36.842 & \text{Public + 1 unit } \overline{\text{SES}}_j \end{cases}$$

Example: Final Tweaking

HSB data, Level 1:

$$\text{math}_{ij} = \beta_{0j} + \beta_{1j}\text{female}_{ij} + \beta_{2ij}(cSES)_{ij} + R_{ij}$$

where $R_{ij} \sim N(0, \sigma^2_{ij})$ $i.i.d.$ and

$$\log(\sigma^2_{ij}) = \omega_0 + \omega_1\text{sector}_j$$

There is no V_{0j}, so $\sigma^2_j = \exp(\omega_0 + \omega_1\text{sector}_j)$
Level 2:

$$\begin{aligned}
\beta_{0j} &= \gamma_{00} + \gamma_{01}\text{sector}_j + \gamma_{02}\overline{\text{SES}}_j + U_{0j} \\
\beta_{1j} &= \gamma_{10} \\
\beta_{2j} &= \gamma_{20}
\end{aligned}$$

$U_{0j} \sim N(0, \tau^2)$ i.i.d and independent of R_{ij}

More Complex σ^2

Effect	Null est	Simple est	Complex μ_{ij} est	Complex σ_j^2 est	Final est (se)
intercept	12.64	12.65	12.724	12.719	12.718 (0.210)
sector			1.258	1.259	1.254 (0.293)
mean ses			5.230	5.212	5.206 (0.354)
female			-1.198	-1.189	-1.190 (0.164)
cses			2.127	2.097	2.095 (0.109)
ω_0		3.66	3.600	3.678	3.689 (0.024)
sector				-0.162	-0.177 (0.034)
mean ses				-0.072	
σ^2	39.15				
ψ^2		0.03	0.014	0.006	
τ^2	8.55	8.69	2.122	2.114	2.101 (0.336)

More Complex σ^2

	Null	Simple	Complex μ_{ij}	Complex σ_j^2	Final?
-2lnlike	47116	47093	46505	46463	46467
AIC	47122	47101	46505	46482	46483
BIC	47131	47113	46529	46514	46507

$$\hat{\sigma}_j^2 = \begin{cases} \exp(3.689 - 0.177) = 33.5077 & \text{Catholic} \\ \exp(3.689) = 39.9953 & \text{Public} \end{cases}$$

Recall: $\hat{\sigma^2} = 39.15$ from Null HLM.

Random intercept, slope and σ^2

HSB data, Level 1:

$$\text{math}_{ij} = \beta_{0j} + \beta_{1j}(cSES)_{ij} + R_{ij}$$

where $R_{ij} \sim N(0, \sigma^2_{ij})$ $i.i.d.$ and

$$\log(\sigma^2_{ij}) = \omega_0 + \omega_1 \text{sector}_j + V_{ij}$$

$\sigma^2_j = \exp(\omega_0 + \omega_1 \text{sector}_j + V_{ij})$ and $V_{ij} \sim N(0, \psi^2)$.
Level 2:

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01}\text{sector}_j + \gamma_{02}\overline{\text{SES}}_j + U_{0j} \\ \beta_{1j} &= \gamma_{10} + U_{1j} \end{aligned}$$

$(U_{0j}, U_{1j})) \sim MVN((0,0)', \boldsymbol{T})$ i.i.d and independent of R_{ij}

Random intercept, slope and σ^2

I switched to LaPlace approximation (i.e., quadrature points=1)

Parameter	estimate	se	df	t	Pr $< t$
γ's					
intercept	12.0659	.2133	157	56.57	<.0001
cses	2.1443	.1285	157	16.68	<.0001
mean.ses	5.2752	.3816	157	13.82	<.0001
sector	1.3285	.3477	157	3.82	.0002
ω					
intercept	3.6878	.02573	157	143.33	<.0001
sector	-0.1884	.03708	157	-5.08	<.0001
Variances and covariance random effects					
tau00	2.3641	.3635			
tau11	0.6817	.2763			
tau10	0.1550	.2593	157	0.60	.5509
psi2	0.007383	0.006086			

More Complex σ^2

	Null	Simple	Complex μ_{ij}	Complex σ_j^2	Final	Last one
-2lnlike	47116	47093	46505	46463	46467	46507
AIC	47122	47101	46505	46482	46483	46527
BIC	47131	47113	46529	46514	46507	46558

SAS: Random intercept & Modeling Heterogeneous ofσ^2

Simple model in PROC NLMIXED:

```
title 'Null model in nlmixed';
proc nlmixed data=hsball method=gauss gconv=0;
parms g0=0 s2=1 tau2=1;
mu = g0 + U0j;
model mathach ~ normal(mu,s2);
random U0j ~ normal(0,tau2) subject=id;
estimate 'ICC' tau2/(s2+tau2);
run;
```

Check GRADIENT and log (& compare results with PROC MIXED)!!!

SAS: More Complex Model

```
title 'Random intercept: complex mean & random variance';
proc nlmixed data=hsball method=gauss gconv=0;
parms g0=12.6 gfemale=-1 gcses=2 gsector=2 gmeanses=2
      w0=3.6 wsector=-.2 wmeanses=0 tau2=2 psi2=0;
mu = g0 + gfemale*female + gcses*cSES + gsector*sector
   + gmeanses*meanses + U;
s2 = exp(w0 + wsector*sector + wmeanses*meanses + V);
model mathach ~  normal(mu,s2);
random U V ~ normal([0,0],[tau2,0,psi2]) subject=id;
estimate 'Catholic & mean SES=0' exp(w0+wsector);
estimate 'Public & mean SES' exp(w0);
estimate 'Catholic + 1 mean SES=0'
   exp(w0+wsector+wmeanses);
estimate 'Public + 1 mean SES=0' exp(w0+wmeanses);
run;
```

R: Location and Scale Model

- For this we'll use Hedeker & Nordren's MIXRELGS run through R. Use the R function "R_mixregls.txt" on course web-site.
- We'll also use the HSB data set so that we can compare results. The results are very similar but MIXRELGS and what I did in SAS are slightly different models.
- Estimation is done using Newton-Raphson and integration over random effects is done using numerical quadrature. The program starts by running 20 EM steps and uses this as input to Newton-Raphson (w/ ridge stabilization).
- The program estimates 3 models in sequential order using output from previous models as starting values for next model.
  - Model 1 is an HLM model with homogenous variances; that is, only location $E(Y_{ij})$ is random.
  - Model 2 adds estimates coefficients of the within variance effects; that is $\log(\sigma^2) = \omega + \ldots$ all fixed effects for variance.
  - Model 3 adds parameters for model for $\log(\sigma^2)$ for random locations (i.e., our U_js), random within (i.e., R_{ij}s)

MIXREGLS in R: Step by Step

Do not follow the instructions in Hedeker & Nordgren. Instead use the function that I wrote: R_mixregls.R

- Step 1: From https://www.jstatsoft.org/article/view/v052i12 download zip file and extract contents. You will only need "mixreglsb.exe"
- Step 2: Save your data and mixreglsb.exe in the same directory, which should be your working directory
- Step 3: We'll need these: library(formula.tool) and library(stringr)
- Step 4: Define function.
- Step 5: Run the function

See detailed example in web-site

R_mixregls Function

The formula has the following general form

$$\text{response} \sim \text{fixed effects} \mid \text{Between} \mid \text{Within}$$

```
library(formula.tools)
library(stringr)
setwd("D:/Dropbox/edps587/lectures/8
modelbuilding/MIXREGLS/hsb_example")
source("R_mixregls.txt")
indata <- read.table("hsball.txt", header=TRUE)
fo ← formula(mathach ∼ female + cSES + meanses + sector |
meanses + sector | meanses + sector)
R_mixregls(fo, indata, idname="id",
       outdata="hsb_example1.dat",
       outresults="hsb_example1.out",
       save_def="hsb_example1.def"
```

Output from MIXREGLS

Model 1: Just our regular HLM:

```
Dependent variable:  mathach
-2 ln L: 46494.07
```

		Estimate	AsymStdErr	z-value	p-va
γ_{00}	beta Intercept	12.7237	0.20728	61.384	0.00e
γ_{01}	beta meanses	5.2183	0.35266	14.797	0.00e
γ_{10}	beta female	-1.1982	0.16207	-7.393	0.00e
γ_{20}	beta cSES	2.1521	0.10847	19.841	0.00e
γ_{02}	beta sector	1.2514	0.29221	4.283	4.15e
$\log(\tau_0^2)$	alpha Intercept	0.7368	0.16073	4.584	1.09e
$\log(\sigma^2)$	tau Intercept	3.6053	0.01688	213.604	0.00e

$\sigma^2 = \exp(3.6053) = 36.7927$ and $\tau_0^2 = \exp(0.7368) = 2.0892$

Output MIXREGLS Model 2

```
-2 ln L: 46464.34
```

		Estimate	AsymStdErr	z-value	p-v
γ_{00}	beta Intercept	12.71839	0.20978	60.628	0.000
γ_{01}	beta meanses	5.20369	0.35421	14.691	0.000
γ_{10}	beta female	-1.18751	0.16361	-7.258	0.000
γ_{20}	beta cSES	2.10463	0.10932	19.252	0.000
γ_{02}	beta sector	1.25129	0.29256	4.277	4.251
$\log(\tau_0^2)$	alpha Intercept	0.74073	0.15999	4.630	8.837
ω_0	tau Intercept	3.67932	0.02444	150.533	0.000
ω_1	tau sector	-0.15763	0.03605	-4.373	2.814
ω_2	tau meanses	-0.07114	0.04490	-1.585	1.137

$$\sigma_{ij}^2 = \exp(3.67932 - 0.1576(\text{sector})_j - 0.0714(\text{meanses})_j)$$

Output MIXREGLS Model 3

-2 ln L: 46458.23

		Estimate	AsymStdErr	z-value
γ_{00}	beta Intercept	1.272e+01	0.20884	6.088e+01
γ_{01}	beta meanses	5.187e+00	0.35262	1.471e+01
γ_{10}	beta female	-1.192e+00	0.16325	-7.303e+00
γ_{20}	beta cSES	2.107e+00	0.10933	1.927e+01
γ_{02}	beta sector	1.266e+00	0.29090	4.352e+00
$\log(\tau_0^2)$	alpha Intercept	7.261e-01	0.16003	4.538e+00
ω_0	tau Intercept	3.680e+00	0.02516	1.462e+02
ω_1	tau sector	-1.587e-01	0.03722	-4.265e+00
ω_2	tau meanses	-8.549e-02	0.04674	-1.829e+00
ξ_ℓ	S1	-5.500e-02	0.02230	-2.466e+00
ψ	S2	1.563e-15	0.05810	2.690e-14

Output MIXREGLS Model 3

$$\text{var}(R_{ij}) = \sigma^2_{ij} = \exp(3.680 - 1.587(\text{sector})_j - 0.845(\text{meanses})_j + \frac{1}{2}\left(\xi^2_\ell + \psi^2\right))$$

Note: The model reported was fit to the data by mixreglsb.exe. The function handles a model without variables in the BS or WS section using the word "none".

How I Obtained Reported Results

An example of mixregles.def should look like this

```
Only and intercept for BS
Change BS to 0 and dropped from following lines
hsb_no_BS.dat
hsb_no_BS.out
hsb_no_BS.def
6 4 0 2 0 0 0 1.00000000000000E-05 11 1 200 0 0 1
1 2
3 4 5 6
5 6
mathach
female cSES meanses sector
meanses sector
```

More estimation options

- An extended version of MIXREGLS that allows for random slopes as well. There is SAS and Stata code in appendix of Nordgren R, Hedeker D, Dunton G, Yang C-H. Extending the mixed-effects model to consider within-subject variance for ecological momentary assessment data. *Statistics in Medicine*. 2020;39:577590. https://doi.org/10.1002/sim.8429. An R version of this is underdevelopment at the time of the writing of this paper.
- brms. I tried this out but results didn't correspond to those of SAS and MIXREGLS. I look at Stan code and it seems to work on standard deviation rather than variances.
- Some R packages that use Bayesian estimation but I didn't try them out, e.g., LMMELSM fits Latent Multivariate Mixed Effects Location Scale Model.
- Seems like an active area of development

Location and Scale Model for NELS

	Variable	Estimate	AsymStdError	z-value	p-value
BETA (regression coefficients)					
γ's	Intercept	50.17732	2.39892	20.91664	0.00000
	homew	2.06480	0.25794	8.00509	0.00000
	sex	-0.47739	0.72904	-0.65482	0.51258
	schCses	3.17400	0.58292	5.44502	0.00000
	schMses	8.19304	1.96733	4.16456	0.00003
	schtype	-0.76985	1.24651	-0.61760	0.53684
ALPHA (BS variance parameters: log-linear model)					
$\log(\tau^2)$	Intercept	0.12962	1.15725	0.11201	0.91082
	schtype	0.86771	0.44266	1.96022	0.04997
TAU (WS variance parameters: log-linear model)					
ω_0	Intercept	4.45617	0.11453	38.90860	0.00000
ω_1	schtype	-0.13854	0.04825	-2.87140	0.00409

Location and Scale Model for NELS

So for a random INTERCEPT model

$$\begin{aligned}\hat{\tau}^2_{0,j} &= \exp(0.12963 + 0.86771(\text{schtype})_j) \\ &= \begin{cases} \exp(0.12963 + 0.86771) = 2.71 & \text{private} \\ \exp(0.12963) = 0.13 & \text{public} \end{cases}\end{aligned}$$

$$\begin{aligned}\hat{\sigma}^2_j &= \exp(4.456617 - 0.13854(\text{schtype})_j) \\ &= \begin{cases} \exp(4.456617 - 0.13854) = 75.04 & \text{private} \\ \exp(4.456617) = 86.19542 & \text{public} \end{cases}\end{aligned}$$

Using Cholsky root

I used two different parameterizations in SAS/NLMIXED.

- Same as algebraic model given on previous slide.
- One that uses a Cholsky Root, which makes dependencies clearer (and estimation easier).

Let $\boldsymbol{\Sigma}$ be a square symmetric matrix (e.g., a covariance matrix), the Cholsky root of $\boldsymbol{\Sigma}$ is

$$\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}' = \begin{pmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} a_{11} & a_{21} & a_{31} \\ 0 & a_{22} & a_{32} \\ 0 & 0 & a_{33} \end{pmatrix} = \boldsymbol{A}\boldsymbol{A}'$$

Using Cholsky root (continued)

Let's suppose that we have a vector $\boldsymbol{U}$ such that

$$\boldsymbol{U} = \begin{pmatrix} U_{0j} \\ U_{1j} \\ R_{ij} \end{pmatrix} \sim N\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$$

Let $\boldsymbol{\theta}$ be a (3×3) vector that follow a $N(\mathbf{0}, \boldsymbol{I})$, and take

$$\boldsymbol{U} = \boldsymbol{A}\boldsymbol{\theta} = \begin{pmatrix} a_{11}\theta_1 \\ a_{21}\theta_1 + a_{22}\theta_2 \\ a_{31}\theta_1 + a_{32}\theta_2 + a_{33}\theta_3 \end{pmatrix}$$

So $\boldsymbol{\mu_U} = \boldsymbol{A}\boldsymbol{\mu_\theta} = \mathbf{0}$ and $\boldsymbol{\Sigma_U} = \boldsymbol{A}\boldsymbol{I}\boldsymbol{A'} = \boldsymbol{A}\boldsymbol{A'}$

Using Cholsky root (continued)

If you have estimated the elements of $\boldsymbol{A}$, then

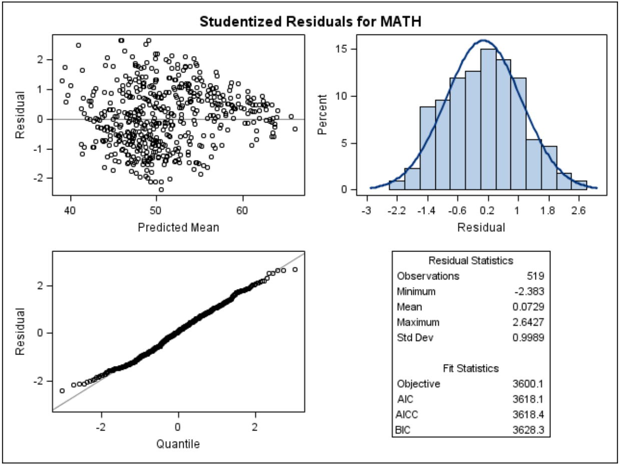
$$\boldsymbol{\Sigma_U} = \boldsymbol{A}\boldsymbol{A}' = \begin{pmatrix} a_{11}^2 & a_{11}a_{21} & a_{11}a_{31} \\ a_{21}^2 + a_{22}^2 & a_{21}^2 + a_{22}^2 & a_{21}a_{31} + a_{22}a_{32} \\ a_{11}a_{31} & a_{31}a_{21} + a_{22}a_{32} & a_{31}^2 + a_{32}^2 + a_{33}^2 \end{pmatrix}$$

$$= \begin{pmatrix} \text{var}(U_{0j}) & \text{cov}(U_{0j}, U_{1j}) & \text{cov}(U_{0j}, R_{ij}) \\ \text{cov}(U_{0j}, U_{1j}) & \text{var}(U_{1j}) & \text{cov}(U_{0j}, R_{ij}) \\ \text{cov}(U_{0j}, R_{ij}) & \text{cov}(U_{1j}, R_{ij}) & \text{var}(R_{ij}) \end{pmatrix}$$

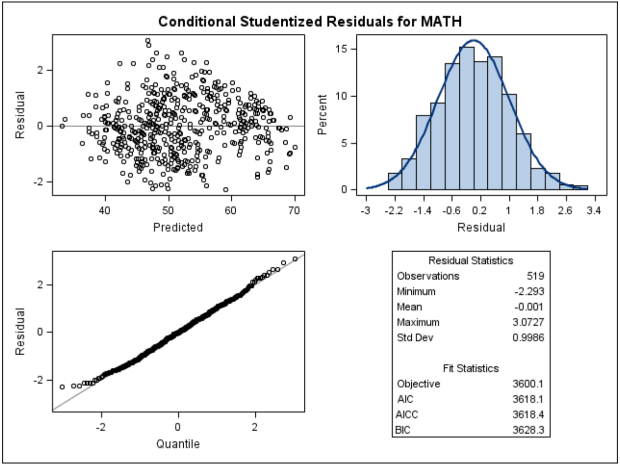
Residuals (back to NELS)

- Level 1 residuals for assessing $R_{ij} \sim \mathcal{N}(0, \sigma^2)$.
  - Raw and/or standardized.
  - Graphical displays.
  - Test for homogeneous variance (see Snijders & Bosker, p 126–128, rather vague. Modeling σ^2 as above).
- Level 2 residuals, $\hat{\boldsymbol{U}}_j$, are confounded with $\hat{R}_{ij}$.
  - If normal, then maybe OK.
  - If non-normal, then problem.
  - Try alternative distribution — NLMIXED, Bayesian, or MIXED MACRO.
- Marginal Residuals: $R_{ij} + \boldsymbol{z}_{ij}\boldsymbol{U}_j = y_{ij} - \boldsymbol{x}_{ij}\boldsymbol{\Gamma}$
- Conditional Residuals: $R_{ij} = y_{ij} - \boldsymbol{x}_{ij}\boldsymbol{\Gamma} - \boldsymbol{z}_{ij}\boldsymbol{U}_j$

Studentized Marginal Residuals (SAS)



Studentized Conditional Residuals (SAS)

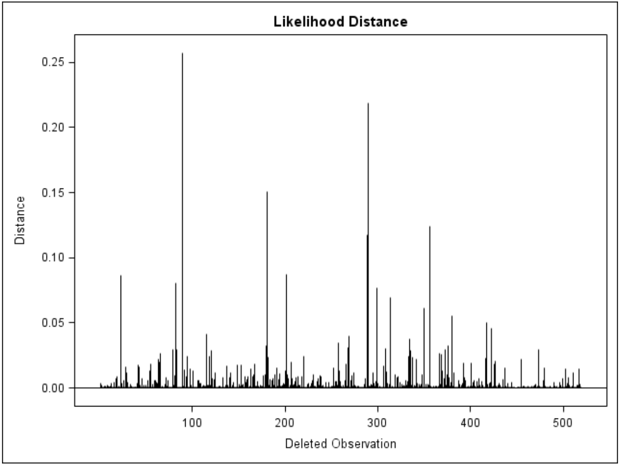


Influence (SAS)

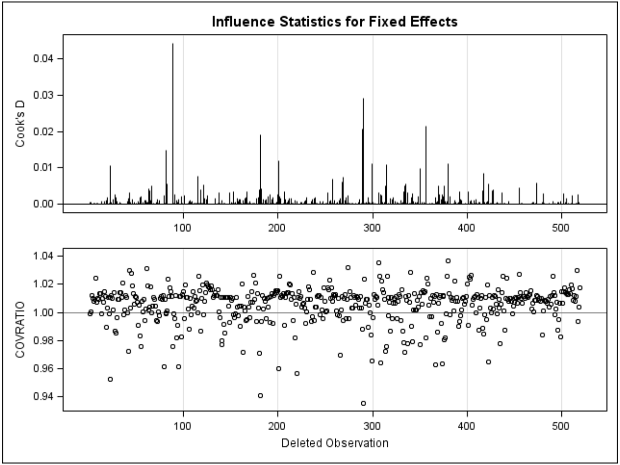
Quantify the influence of 1 or more observation on

- Overall measures of fit (i.e., likelihood ratio).
- Parameter estimates (i.e., `Cook's D`, `MDFFITS`)
- Precision of estimates (i.e., `CovRatio`, `CovTrace`).
- Fitted & predicted values (i.e., `PRESS` residuals, `PRESS` statistic).
- Outliers (internally and externally studentized residuals, leverage).

Distance Plot



Influence Plots

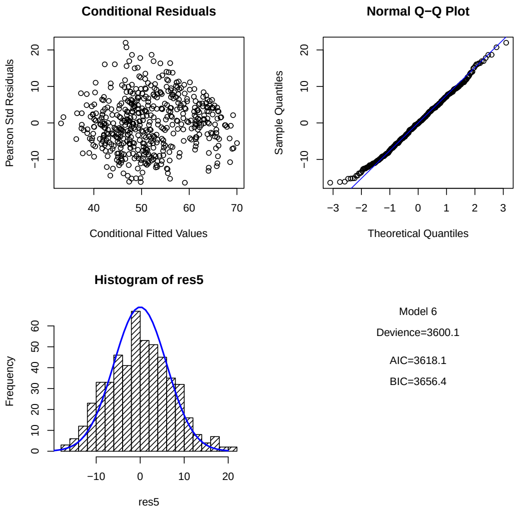


Diagnostics in SAS

```
ODS graphics on / imagefmt=jpeg;
PROC MIXED data=school23 method=REML noclprint covtest ic
        plots (maxpoints=6000) =all;
   CLASS school white ;
   MODEL math = homew cses white gmeanses /
        solution residual influence ;
   RANDOM intercept homew / subject=school type=un;
   ODS output Influence=inf;
   TITLE 'Final Model with Diagnostics';
ODS graphics off;
```

The option "(maxpoints=$<$ number $>$)" is needed if number of cases is larger than $5,000$ (e.g., TIMSS).

$\hat{U}_{0j}$ and $\hat{U}_{1j}$



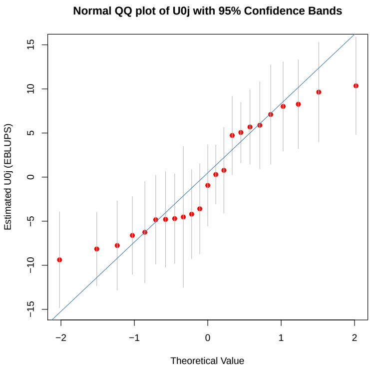
Model 6

Devience=3600.1

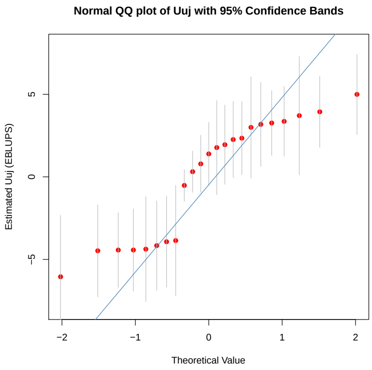
AIC=3618.1

BIC=3656.4

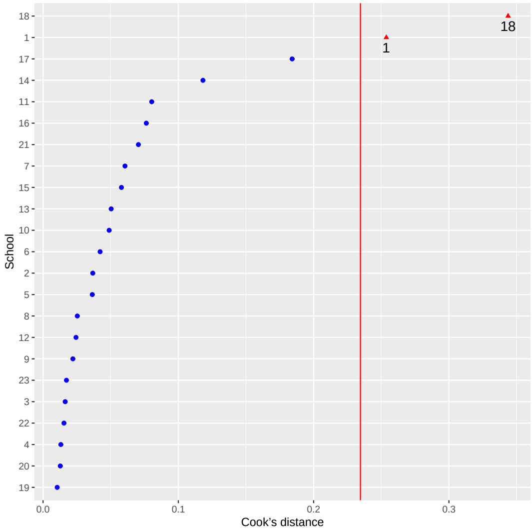
QQ plot: This doesn't look so good



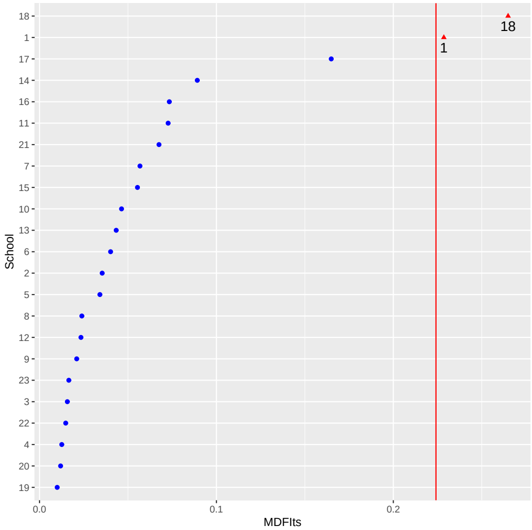
QQ plot: This doesn't look so good



Diagnostics from R lme4: Cook Distance

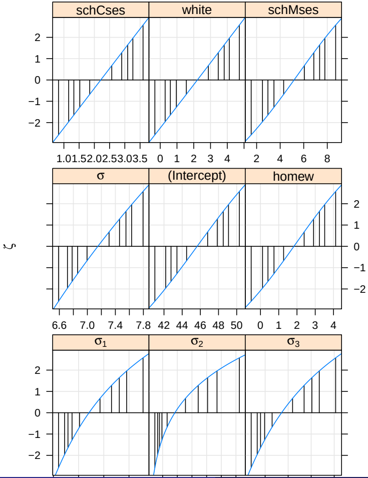


Diagnostics from R lme4: MDfits



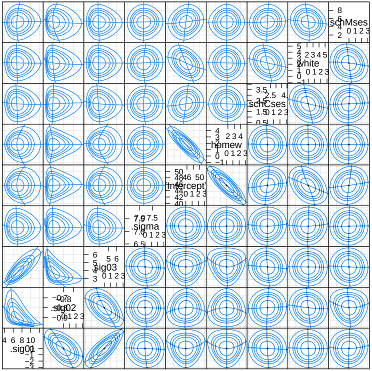
Diagnostics from R lme4: Zeta plots

Linear blue lines good:



Diagnostics from R lme4: Profile Pairs

How parameters depend on each other



Scatter Plot Matrix

Confidence Limits

```
> pr5 ← profile(model5)
> round(confint(pr5, level=.99),digits=2)
```

	0.5%	99.5%
.sig01	4.41	11.11
.sig02	-0.98	-0.69
.sig03	2.39	6.06
.sigma	6.59	7.80
(Intercept)	41.00	50.29
homew	-0.51	4.12
schCses	0.83	3.60
white	-0.32	4.72
schMses	1.54	8.71

SAS and R

- Code that goes with this lecture
  - SAS: everything except extra diagnostics that lme4 gives.
  - R: almost everything except on graphics for preliminary random effects.
- Next Lab (last one)