Outline

- Notation.
- NELS88 data
- Fixed Effects ANOVA
- Random Effects ANOVA
- Multiple Regression
- HLM and the general linear mixed model

Reading: Snijders & Bosker, chapter 3
Snijders & Bosker’s notation.

$j$ index for groups (macro-level units).

$N = \text{number of groups.}$

$j = 1, \ldots, N.$

$i$ index for individuals (micro-level units).

$n_j = \text{number of individuals in group } j.$

$i = 1, \ldots, n_j.$
Notation (continued)

$Y_{ij}$ response or dependent variable.

$y_{ij}$ is a value on the response/dependent variable.

$x_{ij}$ explanatory variable.

$n_+ = \text{total number of level 1 observations},$

$n_+ = \sum_{j=1}^{N} n_j$.

This is different from Snijder & Bosker (they use $M$).
National Education Longitudinal Study:

Conducted by National Center for Education Statistics of the US department of Education. These data constitute the first in a series of longitudinal measurements of students starting in 8th grade. These data were collected Spring 1988.
NELS88: The Data for 10 Schools

Data for All 10 Schools

Ignoring Hierarchical Structure

Math Scores

Time Spent Doing Homework

NONE 1 HOUR 3 HOURS 7 TO 9 HOURS 8
# Descriptive Statistics — Math scores

Total sample: \( \sum_j^N n_j = n_+ = 260, \bar{Y} = 51.30, s = 11.14 \)

<table>
<thead>
<tr>
<th>By School:</th>
<th>( \text{sch}_{id} )</th>
<th>( n_j )</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>7472</td>
<td>23</td>
<td>45.73</td>
<td>7.53</td>
<td>33.00</td>
<td>64.00</td>
<td></td>
</tr>
<tr>
<td>7829</td>
<td>20</td>
<td>42.15</td>
<td>8.31</td>
<td>31.00</td>
<td>65.00</td>
<td></td>
</tr>
<tr>
<td>7930</td>
<td>24</td>
<td>53.25</td>
<td>11.52</td>
<td>33.00</td>
<td>70.00</td>
<td></td>
</tr>
<tr>
<td>24725</td>
<td>22</td>
<td>43.54</td>
<td>10.00</td>
<td>31.00</td>
<td>65.00</td>
<td></td>
</tr>
<tr>
<td>25456</td>
<td>22</td>
<td>49.86</td>
<td>8.44</td>
<td>32.00</td>
<td>62.00</td>
<td></td>
</tr>
<tr>
<td>25642</td>
<td>20</td>
<td>46.40</td>
<td>4.32</td>
<td>39.00</td>
<td>57.00</td>
<td></td>
</tr>
<tr>
<td>62821</td>
<td>67</td>
<td>62.82</td>
<td>5.67</td>
<td>43.00</td>
<td>71.00</td>
<td></td>
</tr>
<tr>
<td>68448</td>
<td>21</td>
<td>49.66</td>
<td>10.33</td>
<td>34.00</td>
<td>69.00</td>
<td></td>
</tr>
<tr>
<td>68493</td>
<td>21</td>
<td>46.33</td>
<td>9.55</td>
<td>34.00</td>
<td>71.00</td>
<td></td>
</tr>
<tr>
<td>72292</td>
<td>20</td>
<td>47.85</td>
<td>11.30</td>
<td>34.00</td>
<td>68.00</td>
<td></td>
</tr>
</tbody>
</table>
Another Look at the Data

Mean math scores ±1 standard deviation for the $N = 10$ schools.

Mean $±1$ standard deviation
Review of Fixed Effects ANOVA

The model:

\[ Y_{ij} = \mu + \alpha_j + R_{ij} \]

where

- \( \mu \) is the overall mean (constant).
- \( \alpha_j \) is group \( j \) effect (constant).
- \( R_{ij} \) is the residual or error (random).
- \( R_{ij} \sim \mathcal{N}(0, \sigma^2) \) and independent.

- Accounts for all group differences.
- Only interested in the \( N \) groups.
Example: NELS88 using 10 schools

Math scores as response variable.

ANOVA Summary Table
From SAS/GLM and MIXED (method=REML)

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model (school)</td>
<td>9</td>
<td>14030.54</td>
<td>1558.95</td>
<td>21.55</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Error (residual)</td>
<td>250</td>
<td>18086.06</td>
<td>72.34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected total</td>
<td>259</td>
<td>32116.60</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Example (continued)

Only from SAS/MIXED:

- Maximum Likelihood Estimation

<table>
<thead>
<tr>
<th>Effect</th>
<th>Num df</th>
<th>Den df</th>
<th>$F$ value</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>school</td>
<td>9</td>
<td>250</td>
<td>22.41</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>

- Expected mean squares

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of squares</th>
<th>Mean square</th>
<th>Expected mean square</th>
<th>Error term</th>
</tr>
</thead>
<tbody>
<tr>
<td>sch$_{id}$</td>
<td>9</td>
<td>14031</td>
<td>1558.95</td>
<td>Var(Res) + Q(sch$_{id}$)</td>
<td>MS(Res)</td>
</tr>
<tr>
<td>Residual (Res)</td>
<td>250</td>
<td>18086</td>
<td>72.34</td>
<td>Var(Res)</td>
<td></td>
</tr>
</tbody>
</table>
Linear Regression

Formulation of fixed effects ANOVA

\[ Y_{ij} = \mu_j + R_{ij} \]
\[ = \mu + \alpha_j + R_{ij} \]
\[ = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \ldots + \beta_{(N-1)} x_{(N-1),ij} + R_{ij} \]

where \( x_{kij} = 1 \) if \( j = k \), and 0 otherwise (i.e., dummy codes).
### Linear Regression in Matrix Notation

\[
\begin{pmatrix}
Y_{11} \\
Y_{21} \\
\vdots \\
Y_{23,1} \\
Y_{1,2} \\
Y_{2,2} \\
\vdots \\
Y_{20,2} \\
\vdots \\
Y_{1,10} \\
\vdots \\
Y_{20,10}
\end{pmatrix}
= 
\begin{pmatrix}
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\end{pmatrix}
\begin{pmatrix}
\beta_0 \\
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4 \\
\beta_5 \\
\beta_6 \\
\beta_7 \\
\beta_8 \\
\beta_9 \\
\end{pmatrix}
+ 
\begin{pmatrix}
R_{11} \\
R_{21} \\
\vdots \\
R_{23,1} \\
R_{1,2} \\
R_{2,2} \\
\vdots \\
R_{20,2} \\
\vdots \\
R_{1,10} \\
R_{20,10}
\end{pmatrix}
\]
Linear Regression in Matrix Notation

\[
\begin{pmatrix}
  Y_1 \\
  Y_2 \\
  \vdots \\
  Y_{10}
\end{pmatrix}
= \begin{pmatrix}
  X_1 \\
  X_2 \\
  \vdots \\
  X_{10}
\end{pmatrix} \beta + \begin{pmatrix}
  R_1 \\
  R_2 \\
  \vdots \\
  R_{10}
\end{pmatrix}
\]

\[Y = X \beta + R\]
Fixed Effects Model (continued)

\[ Y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \ldots + \beta_{(N-1)} x_{(N-1)ij} + R_{ij} \]

- The parameters \( \beta_0, \beta_1, \ldots, \beta_{N-1} \) are considered fixed.

- The \( R_{ij} \)'s are random: \( R_{ij} \sim \mathcal{N}(0, \sigma^2) \) and independent.

- Therefore,

\[ Y_{ij} \sim \mathcal{N}(\beta_0 + \beta_1 x_{1ij} + \ldots + \beta_{(N-1)} x_{(N-1)ij}, \sigma^2) \]
In Matrix Terms... 

The model:

$$Y = X\beta + R$$

- where $\beta$ is an $(N \times 1)$ vector of fixed parameters.
- $R$ is an $(n_+ \times 1)$ vectors of random variables:

$$R \sim \mathcal{N}_{n_+}(0, \sigma^2 I)$$

- $Y \sim \mathcal{N}_{n_+}(X\beta, \sigma^2 I)$

- Covariance matrix,

$$\sigma^2 I = \begin{pmatrix}
\sigma^2 & 0 & \ldots & 0 \\
0 & \sigma^2 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \sigma^2
\end{pmatrix}$$
Estimated model parameters, $\sigma$ and $\beta$

From the ANOVA table: $\hat{\sigma}^2 = 72.34$

Solution for Fixed Effects

<table>
<thead>
<tr>
<th>Effect</th>
<th>$sch_{id}$</th>
<th>Estimate</th>
<th>Error</th>
<th>df</th>
<th>$t$</th>
<th>Pr &gt;</th>
<th>$t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>47.85</td>
<td>1.90</td>
<td>250</td>
<td>25.16</td>
<td>&lt; .0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$sch_{id}$</td>
<td>7472</td>
<td>−2.11</td>
<td>2.60</td>
<td>250</td>
<td>−0.81</td>
<td>.42</td>
<td></td>
</tr>
<tr>
<td>$sch_{id}$</td>
<td>7829</td>
<td>−5.70</td>
<td>2.69</td>
<td>250</td>
<td>−2.12</td>
<td>.04</td>
<td></td>
</tr>
<tr>
<td>$sch_{id}$</td>
<td>7930</td>
<td>5.40</td>
<td>2.58</td>
<td>250</td>
<td>2.10</td>
<td>.04</td>
<td></td>
</tr>
<tr>
<td>$sch_{id}$</td>
<td>24725</td>
<td>−4.31</td>
<td>2.63</td>
<td>250</td>
<td>−1.64</td>
<td>.10</td>
<td></td>
</tr>
<tr>
<td>$sch_{id}$</td>
<td>25456</td>
<td>2.01</td>
<td>2.63</td>
<td>250</td>
<td>0.77</td>
<td>.44</td>
<td></td>
</tr>
<tr>
<td>$sch_{id}$</td>
<td>25642</td>
<td>−1.45</td>
<td>2.69</td>
<td>250</td>
<td>−0.54</td>
<td>.59</td>
<td></td>
</tr>
<tr>
<td>$sch_{id}$</td>
<td>62821</td>
<td>14.97</td>
<td>2.17</td>
<td>250</td>
<td>6.91</td>
<td>&lt; .0001</td>
<td></td>
</tr>
<tr>
<td>$sch_{id}$</td>
<td>68448</td>
<td>1.82</td>
<td>2.66</td>
<td>250</td>
<td>0.68</td>
<td>0.49</td>
<td></td>
</tr>
<tr>
<td>$sch_{id}$</td>
<td>68493</td>
<td>−1.52</td>
<td>2.66</td>
<td>250</td>
<td>−0.57</td>
<td>0.57</td>
<td></td>
</tr>
<tr>
<td>$sch_{id}$</td>
<td>72292</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Random Effects ANOVA

**Situation**: When groups are a random sample from some larger population of groups and you want to make inferences about this population.

Data are hierarchically structured data with no explanatory variables.
Random Effects ANOVA: The Model

\[ Y_{ij} = \underbrace{\mu}_{\text{fixed}} + \underbrace{U_j + R_{ij}}_{\text{stochastic}} \]

where

- \( \mu \) is the overall mean (constant).
- \( U_j \) is group \( j \) effect (random).
- \( U_j \sim \mathcal{N}(0, \tau^2) \) and independent.
- \( R_{ij} \) is the w/in groups residual or error (random).
- \( R_{ij} \sim \mathcal{N}(0, \sigma^2) \) and independent.
- \( U_j \) and \( R_{ij} \) are independent, i.e., \( \text{cov}(U_j, R_{ij}) = 0 \).
Notes Regarding Random Effects ANOVA

- If $\tau^2 > 0$, group differences exist.
- Instead of estimating $N$ parameters for groups, we estimate a single parameter, $\tau^2$.
- The variance of $Y_{ij} = (\tau^2 + \sigma^2)$.
- Random effects ANOVA can be represented as a linear regression model with a random intercept.
Random Effects ANOVA

As Linear Regression Model:

\[ Y_{ij} = \beta_{0j} + R_{ij} \]

- No explanatory variables.
- The intercept: \( \beta_{0j} = \gamma + U_j \)
  where \( \gamma \) is Constant (overall mean) and \( U_j \) is Random, \( \mathcal{N}(0, \tau^2) \)
- \( \beta_{0j} \) is random with \( \beta_{0j} \sim \mathcal{N}(\gamma, \tau^2) \)
- \( U_j \) are Level 2 residuals,
  \( R_{ij} \) are Level 1 residuals,
  and they’re independent.
- The “empty” or “null” HLM.
Random Effects ANOVA: NELS88

Using SAS/MIXED with Maximum likelihood estimation:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>std error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed effect(s)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma$</td>
<td>48.87</td>
<td>1.83</td>
</tr>
<tr>
<td>Variance parameters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>intercept $\tau^2$</td>
<td>30.52</td>
<td>14.48</td>
</tr>
<tr>
<td>residual $\sigma^2$</td>
<td>72.23</td>
<td>11.20</td>
</tr>
</tbody>
</table>
A Closer Look (Intuitive)

\[ Y_{ij} = \beta_{0j} + R_{ij} \]
\[ = \gamma + U_j + R_{ij} \]

- \( Y_{ij} \) is random due to \( U_j \) and \( R_{ij} \).

- Since \( U_j \) and \( R_{ij} \) are independent and normally distributed,

  (i) \( Y_{ij} \sim \mathcal{N}(\gamma, (\tau^2 + \sigma^2)) \)

  (ii) \( \text{cov}(Y_{ij}, Y_{kj'}) = 0 \) (between groups)

  e.g., \( Y_{ij} = \gamma + U_j + R_{ij} \)
  \( Y_{i'j'} = \gamma + U_{j'} + R_{i'j'} \)
A Closer Look (continued)

\[
Y_{ij} = \beta_0 + R_{ij} = \gamma + U_j + R_{ij}
\]

Within groups, observations are dependent.

\[
Y_{ij} = \gamma + U_j + R_{ij} \\
Y_{i'j} = \gamma + U_j + R_{i'j}
\]

\(\gamma\) is fixed.

\(U_j\) is random but the same value for both \(i\) & \(i'\).

\(R_{ij}\) and \(R_{i'j}\) are random and different values for \(i\) & \(i'\).
Within Group Dependency (formal)

\[
\text{cov}(Y_{ij}, Y_{i'j}) \equiv E[(Y_{ij} - E(Y_{ij}))(Y_{i'j} - E(Y_{i'j}))] \\
= E[((\gamma + U_j + R_{ij}) - \gamma)((\gamma + U_j + R_{i'j}) - \gamma)] \\
= E[(U_j + R_{ij})(U_j + R_{i'j})] \\
= E[U_j^2 + U_jR_{i'j} + R_{ij}U_j + R_{ij}R_{i'j}] \\
= E[U_j^2] + E[U_jR_{i'j}] + E[R_{ij}U_j] + E[R_{ij}R_{i'j}] \\
= E[U_j^2] \\
= E[(U_j - 0)^2] \\
\equiv \text{var}(U_j) \\
= \tau^2
\]
Intra-class Correlation

Measure of within group dependency.

\[ \rho_I \equiv \text{corr}(Y_{ij}, Y_{i'j}) \]

\[ = \frac{\text{cov}(Y_{ij}, Y_{i'j})}{\sqrt{\text{var}(Y_{ij})} \sqrt{\text{var}(Y_{i'j})}} \]

\[ = \frac{\tau^2}{\sqrt{\tau^2 + \sigma^2} \sqrt{\tau^2 + \sigma^2}} \]

\[ = \frac{\tau^2}{\tau^2 + \sigma^2} \]
Intra-class Correlation: $\rho_I$

\[
\rho_I = \frac{\tau^2}{\tau^2 + \sigma^2}
\]

- Assuming within group dependency is the same in all groups.
- If $\tau^2 = 0$, then don’t need HLM.
- Could be used as a measure of how much variance accounted for by the model,

  e.g., Our NELS88 example,

  \[
  \hat{\rho}_I = \frac{30.52}{30.52 + 72.24} = .30
  \]
Matrix representation of $Y_{ij} = \gamma + U_j + R_{ij}$

\[
\begin{pmatrix}
  Y_{11} \\
  Y_{21} \\
  \vdots \\
  Y_{23,1} \\
  Y_{12} \\
  Y_{22} \\
  \vdots \\
  Y_{20,2} \\
  \vdots \\
  Y_{1,10} \\
  \vdots \\
  Y_{20,10}
\end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{pmatrix} \begin{pmatrix} \gamma \end{pmatrix} + \begin{pmatrix} U_1 \\ U_1 \\ \vdots \\ U_1 \\ U_2 \\ \vdots \\ U_2 \\ \vdots \\ U_{10} \end{pmatrix} + \begin{pmatrix} R_{11} \\ R_{21} \\ \vdots \\ R_{23,1} \\ R_{1,2} \\ \vdots \\ R_{2,2} \\ \vdots \\ R_{20,10} \end{pmatrix}
\]
Matrix representation (continued)

\[
\begin{pmatrix}
Y_1 \\
Y_2 \\
\vdots \\
Y_{10}
\end{pmatrix} =
\begin{pmatrix}
X_1 \\
X_2 \\
\vdots \\
X_{10}
\end{pmatrix} \beta
+ \begin{pmatrix}
U_1 \\
U_2 \\
\vdots \\
U_{10}
\end{pmatrix}
+ \begin{pmatrix}
R_1 \\
R_2 \\
\vdots \\
R_{10}
\end{pmatrix}
\]

\[
Y = X \beta + U + R
\]
Matrix Representation (continued)

For group $j$, the distribution of $Y_j$ is

$$Y_j \sim N_{n_j}(X_j \beta, \Sigma_j)$$

where

$$X_j = 1_j$$

is the an $(n_j \times 1)$ column vector with elements equal to one.

$$\beta = \gamma$$

The covariance matrix for the $j^{th}$ group is an $(n_j \times n_j)$ symmetric matrix with elements

$$\Sigma_j = \sigma^2 I_j + \tau^2 1_j 1_j'$$
Covariance Matrix

\[
\Sigma_j = \sigma^2 I_j + \tau^2 1_j 1'_j
\]

\[
= \sigma^2 \begin{pmatrix}
1 & 0 & \ldots & 0 \\
0 & 1 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & 1
\end{pmatrix} + \tau^2 \begin{pmatrix}
1 & 1 & \ldots & 1 \\
1 & 1 & \ldots & 1 \\
\vdots & \vdots & \ddots & \vdots \\
1 & 1 & \ldots & 1
\end{pmatrix}
\]

\[
= \begin{pmatrix}
(\tau^2 + \sigma^2) & \tau^2 & \ldots & \tau^2 \\
\tau^2 & (\tau^2 + \sigma^2) & \ldots & \tau^2 \\
\vdots & \vdots & \ddots & \vdots \\
\tau^2 & \tau^2 & \ldots & (\tau^2 + \sigma^2)
\end{pmatrix}
\]
Covariance Matrix (continued)

\[ \Sigma_j = \sigma^2 l_j + \tau^2 l_j l_j' \]

Notes:

- “Compound symmetry”.

- Only difference between groups is the size of this matrix (i.e., \((n_j \times n_j)\)). For simplicity, drop the sub-script from \(\Sigma\).
**Covariance Matrix: \( \mathbf{Y} \)**

For the whole data vector \( \mathbf{Y} \), the covariance matrix is of size \((n_+ \times n_+)\) and equals

\[
\begin{pmatrix}
\Sigma & 0 & \ldots & 0 \\
0 & \Sigma & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \Sigma & (n_N \times n_N)
\end{pmatrix}
\]
Summary: Random Effects ANOVA

Model isn’t too interesting.

- No explanatory variables.
- If we add variables to the regression, we only have a shift intercept in the intercept.
Multiple Regression

Intercepts and Slopes as outcomes.

Two stages:

**Stage 1**: Fit multiple regression model to each group (level 1 analysis).

i.e., Summarize groups by their regression parameters.

**Stage 2**: Use the estimated intercept and slope coefficients as outcomes (level 2 analysis).

i.e., Analysis of the summary statistics.
NELS88: Overall Regression

Before we do the two stage analysis, ...

Regression Ignoring Hierarchical Structure

math = 44.074 + 3.5719 homework

N: 260
Rsq: 0.2470
AdjRsq: 0.2441
RMSE: 9.6815
Individual Schools’ Data
Individual Schools’ Regressions

Separate Regressions for Each School

Math Scores

Time Spent Doing Homework

NONE 4 TO 6 HOURS 10
Two Stage Example: NELS88

Stage 1: Within groups regressions —

- $Y_{ij} =$ student’s math score.
- $x_{ij} =$ time spent doing homework.
- Proposition: More time spent doing homework, the higher the math score.
- Appropriate model:

$$
(m\text{ath})_{ij} = \beta_0 + \beta_1 (\text{homew})_{ij} + R_{ij}
$$

$$
Y_{ij} = \beta_0 + \beta_1 x_{ij} + R_{ij}
$$

where $R_{ij} \sim \mathcal{N}(0, \sigma^2)$.
Stage One

Model fit to each of the 10 NELS88 schools:

<table>
<thead>
<tr>
<th>School ID</th>
<th>$\hat{\beta}_{0j}$</th>
<th>$\hat{\beta}_{1j}$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>7272</td>
<td>50.68354</td>
<td>-3.55380</td>
<td>.2780</td>
</tr>
<tr>
<td>7829</td>
<td>49.01229</td>
<td>-2.92012</td>
<td>.2111</td>
</tr>
<tr>
<td>7930</td>
<td>38.75000</td>
<td>7.90909</td>
<td>.6007</td>
</tr>
<tr>
<td>24725</td>
<td>34.39382</td>
<td>5.59266</td>
<td>.7002</td>
</tr>
<tr>
<td>25456</td>
<td>53.93863</td>
<td>-4.71841</td>
<td>.1873</td>
</tr>
<tr>
<td>25642</td>
<td>49.25896</td>
<td>-2.48606</td>
<td>.2186</td>
</tr>
<tr>
<td>62821</td>
<td>59.21022</td>
<td>1.09464</td>
<td>.1105</td>
</tr>
<tr>
<td>68448</td>
<td>36.05535</td>
<td>6.49631</td>
<td>.5098</td>
</tr>
<tr>
<td>68493</td>
<td>38.52000</td>
<td>5.86000</td>
<td>.3137</td>
</tr>
<tr>
<td>72292</td>
<td>37.71392</td>
<td>6.3350</td>
<td>.6417</td>
</tr>
</tbody>
</table>
### Stage One (continued)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$n$</th>
<th>mean</th>
<th>std dev</th>
<th>s.e.</th>
<th>variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}_{0j}$</td>
<td>10</td>
<td>44.75</td>
<td>8.66</td>
<td>2.74</td>
<td>74.93</td>
</tr>
<tr>
<td>$\hat{\beta}_{1j}$</td>
<td>10</td>
<td>1.96</td>
<td>4.98</td>
<td>1.58</td>
<td>24.76</td>
</tr>
</tbody>
</table>
Stage Two: NELS88

We use the estimated regression parameters as response variables and try to model the parameters.

We’ll start with the intercepts.

Example — simple model for the $\hat{\beta}_{0j}$’s:

$$\hat{\beta}_{0j} = \gamma_0 + U_{0j}$$
### Stage Two: Intercepts

The REG Procedure

Model: MODEL1

Dependent Variable: alpha

#### Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Error</td>
<td>9</td>
<td>674.40021</td>
<td>74.93336</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>9</td>
<td>674.40021</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Stage Two: Intercepts

Root MSE 8.65641
R-Square 0.0000
Dependent Mean 44.75367
Adj R-Sq 0.0000
Coeff Var 19.34234

| Parameter | Estimate | Error   | t   | Pr > |t| |
|-----------|----------|---------|-----|------|-----|
| Intercept | 44.75367 | 2.73740 | 16.35 | < .0001 |

Descriptives statistics
A little more complex model

Model for the level 2 analysis:

\[ \hat{\beta}_{0j} = \gamma_0 + \gamma_1 (\text{mean } SES_j) + U_{0j} \]
\[ = \gamma_0 + \gamma_1 Z_j + U_{0j} \]

where \( U_{0j} \sim \mathcal{N}(0, \tau_0^2) \).
Regression Model for Intercepts

\[ \text{intercept}_j = 42.425 - 8.2817(\text{meanSES})_j \]
SAS PROC REG (intercepts)

The REG Procedure Model: MODEL2
Dependent Variable: alpha

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>219.10398</td>
<td>219.10398</td>
<td>3.85</td>
<td>.0854</td>
</tr>
<tr>
<td>Error</td>
<td>8</td>
<td>455.29623</td>
<td>56.91203</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>9</td>
<td>674.40021</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
SAS PROC REG (intercepts)

Root MSE 7.54401  R-Square 0.3249
Dependent Mean 44.75367  Adj $R^2$ 0.2405
Coeff Var 16.85673

Parameter Estimates

| Variable     | DF | Parameter Estimate | Standard Error | t    | Pr > |t| |
|--------------|----|--------------------|----------------|------|------|---|
| Intercept    | 1  | 42.42466           | 2.66461        | 15.92| < .0001 |
| meanSES      | 1  | -8.28170           | 4.22082        | -1.96| 0.0854 |

Conclusions?

Note regarding p-value decimal places
Stage Two: Slopes

If the estimated slopes differ over groups, the effect of time spent doing homework on math scores changes depending on what school a student is in.

Simplest case: no explanatory variable.

The model:

$$\hat{\beta}_{1j} = \gamma_{10} + U_{1j}$$

where $$U_{1j} \sim \mathcal{N}(0, \tau_1^2)$$. 
### Slopes: No Explanatory Variables

The REG Procedure
Model: MODEL1 Dependent Variable: beta

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>9</td>
<td>222.85163</td>
<td>24.76129</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>9</td>
<td>222.85163</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Slopes: No Explanatory Variables

| Parameter        | Estimate | Error  | t     | Pr > |t| |
|------------------|----------|--------|-------|------|---|
| Intercept        | 1.96093  | 1.57357| 1.25  | 0.2442| |

Descriptives statistics: see page 41
Slopes: More Complex Model

Proposition: The higher the SES, the less of an effect time spent doing homework has on math achievement.

Reasoning for this: Schools with higher mean SES, more wealthy area, more money for schools, better school.

Diagram for this:

\[
\begin{array}{c}
\uparrow \\
Z \\
\downarrow \\
x \rightarrow y
\end{array}
\]
SAS PROC REG

Regression Model for Slopes

\[ \text{slopedj} = 2.5577 + 2.1221(\text{meanSES}) \]

![Graph showing the relationship between school mean SES and estimated slopes.](image-url)
Slopes: SAS PROC REG Output

The model:

$$\hat{\beta}_{1j} = \gamma_{10} + \gamma_{11}(\text{mean SES})_j + U_{1j}$$

where $U_{1j} \sim \mathcal{N}(0, \tau_{1}^2)$.

Dependent Variable: beta

### Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>14.38612</td>
<td>14.38612</td>
<td>0.55</td>
<td>.48</td>
</tr>
<tr>
<td>Error</td>
<td>8</td>
<td>208.46552</td>
<td>26.05819</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>9</td>
<td>222.85163</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Slopes: SAS PROC REG Output

| Parameter   | DF | Estimate | Error   | t     | Pr>|t| |
|-------------|----|----------|---------|-------|-----|
| Intercept   | 1  | 2.55772  | 1.80304 | 1.42  | .1938|
| meanSES     | 1  | 2.12210  | 2.85605 | 0.74  | .4787|

Conclusions?

Problems with this analysis?
Possible Improvements

To the two stage analysis of NELS88

- Center SES.
- Use multivariate (multiple) regression so that the intercepts and slopes are modeled simultaneously.

Would get an estimate of $\tau_{01}$, covariance between $U_{0j}$ and $U_{1j}$. 
Serious problems remain

- Lost information because regression coefficients are used to summarize groups.
- The number of observations per group is not taken into account when analyzing the estimated regression coefficients (in the 2nd stage).
- Random variability is introduced by using estimated regression coefficients.
- How can model fit to data be assessed?
- Incorrect estimation of standard errors (level 2).
A Simple Hierarchical Linear Models

**Level 1:**

\[ Y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + R_{ij} \]

where \( R_{ij} \sim N(0, \sigma^2) \) and independent.

...and more compactly,

\[ Y_j = X_j\beta_j + R_j \]

where \( Y_j \sim N(X_j\beta_j, \sigma^2I) \) i.i.d.
HLM: Level 2

Level 2:

\[ \beta_{0j} = \gamma_{00} + U_{0j} \]

\[ \beta_{1j} = \gamma_{10} + U_{1j} \]

where

\[
\begin{pmatrix}
U_{0j} \\
U_{1j}
\end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_0^2 & \tau_{01} \\ \tau_{01} & \tau_1^2 \end{pmatrix} \right) \quad \text{i.i.d.}
\]

\ldots or more compactly,

\[ \beta_j = Z_j \Gamma + U_j \]

where \( U_j \sim \mathcal{N}(0, T) \) i.i.d.
Linear Mixed Model

Put models for $\beta_{0j}$ and $\beta_{1j}$ into the level 1 model:

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

$$= (\gamma_{0o} + U_{0j}) + (\gamma_{1o} + U_{1j})x_{ij} + R_{ij}$$

$$= \underbrace{\gamma_{00} + \gamma_{10}x_{ij}} + \underbrace{U_{0j} + U_{1j}x_{ij} + R_{ij}}$$

fixed part                  random part
Given all the assumptions of the hierarchical model,

\[ Y_{ij} \sim \mathcal{N}((\gamma_0 + \gamma_{10}x_{ij}), \text{var}(Y_{ij})) \quad i.i.d. \]

where \( \text{var}(Y_{ij}) = (\tau_0^2 + 2\tau_{01}x_{ij} + \tau_1^2x_{ij}^2) + \sigma^2 \).

This is the **Marginal Model** for \( Y_{ij} \).
In Matrix Form

The Model:

\[ Y_j = X_j \Gamma + Z_j U_j + R_j \]

where

\[ Y_j \sim N_{n_j}(X_j \Gamma, V_j) \text{i.i.d.} \]

and the covariance matrix equals,

\[ V_j = Z_j T Z_j' + \sigma^2 I \]

and later for longitudinal data

\[ V_j = Z_j T Z_j' + \Sigma_j. \]
The Marginal Model

The distinction between the hierarchical formulation (and linear mixed model) and the marginal model are important with respect to

- Interpretation.
- Estimation.
- Inference.

...but more on this later.