Models for Clustered Data

Edps/Psych/Soc 589

Carolyn J. Anderson

ILLINOIS

Outline

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

Reading: Snijders & Bosker, chapter 3

Notation

Snijders & Bosker's notation.

j index for groups (macro-level units).

$N =$ number of groups.

$j = 1, \ldots, N$.

i index for individuals (micro-level units).

$n_j =$ 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.

$n_+ =$ total number of level 1 observations,
$n_+ = \sum_{j=1}^{N} n_j$.

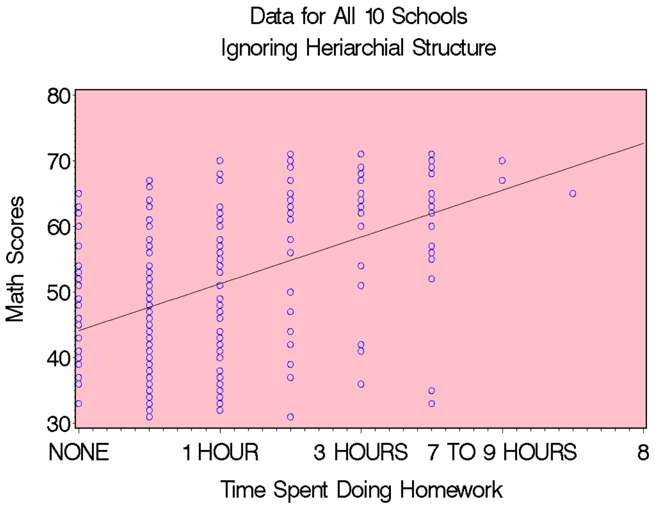
This is different from Snijder & Bosker
(they use M).

NELS88

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



Descriptive Statistics — Math scores

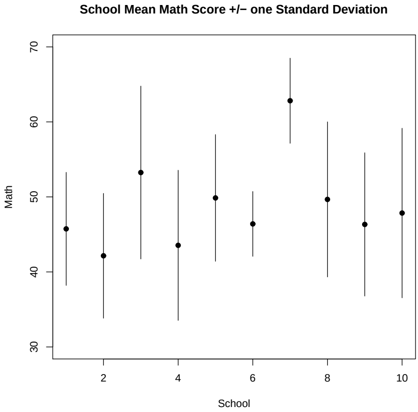
Total sample: $\sum_j^N n_j = n_+ = 260,\ \bar{Y} = 51.30,\ s = 11.14$

By School:

sch$_{id}$	n_j	Mean	Std Dev	Min	Max
7472	23	45.73	7.53	33.00	64.00
7829	20	42.15	8.31	31.00	65.00
7930	24	53.25	11.52	33.00	70.00
24725	22	43.54	10.00	31.00	65.00
25456	22	49.86	8.44	32.00	62.00
25642	20	46.40	4.32	39.00	57.00
62821	67	62.82	5.67	43.00	71.00
68448	21	49.66	10.33	34.00	69.00
68493	21	46.33	9.55	34.00	71.00
72292	20	47.85	11.30	34.00	68.00

Another Look at the Data

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



Review of Fixed Effects ANOVA

The model:

$$Y_{ij} = \mu + \alpha_j + R_{ij}$$

where

μ is the overall mean (constant).

α_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)

Source	df	Sum of Squares	Mean Square	F	p
Model (school)	9	14030.54	1558.95	21.55	$< .0001$
Error (residual)	250	18086.06	72.34		
Corrected total	259	32116.60			

Example (continued)

Only from SAS/MIXED:

- Maximum Likelihood Estimation

Effect	Num *df*	Den *df*	F value	p
school	9	250	22.41	$< .0001$

- Expected mean squares

Source	*df*	Sum of squares	Mean square	Expected mean square	Error term
sch_{id}	9	14031	1558.95	Var(Res) + Q(sch_{id})	MS(Res)
Residual (Res)	250	18086	72.34	Var(Res)	.

Linear Regression

Formulation of fixed effects ANOVA

$$\begin{aligned} Y_{ij} &= \mu_j + R_{ij} \\ &= \mu + \alpha_j + R_{ij} \\ &= \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \ldots + \beta_{(N-1)} x_{(N-1),ij} + R_{ij} \end{aligned}$$

where $x_{kij} = 1$ if $j = k$, and 0 otherwise
(i.e., dummy codes).

Linear Regression in Matrix Notation

$$
\begin{pmatrix} Y_{11} \\ Y_{21} \\ \vdots \\ Y_{23,1} \\ \hline Y_{1,2} \\ Y_{2,2} \\ \vdots \\ Y_{20,2} \\ \hline \vdots \\ \hline Y_{1,10} \\ \vdots \\ Y_{20,10} \end{pmatrix}
=
\begin{pmatrix}
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \hline
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}
\begin{pmatrix} \beta_o \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \\ \beta_7 \\ \beta_8 \\ \beta_9 \end{pmatrix}
+
\begin{pmatrix} R_{11} \\ R_{21} \\ \vdots \\ R_{23,1} \\ \hline R_{1,2} \\ R_{2,2} \\ \vdots \\ R_{20,2} \\ \hline \vdots \\ \hline R_{1,10} \\ \vdots \\ R_{20,10} \end{pmatrix}
$$

Linear Regression in Matrix Notation

$$\begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \\ \vdots \\ \boldsymbol{Y}_{10} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_{10} \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \boldsymbol{R}_1 \\ \boldsymbol{R}_2 \\ \vdots \\ \boldsymbol{R}_{10} \end{pmatrix}$$

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{R}$$

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 :

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{R}$$

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

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

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

$$\sigma^2\boldsymbol{I} = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix}$$

Estimated model parameters, σ and β

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

Solution for Fixed Effects

Effect	sch_{id}	$\boldsymbol{\beta}$ Estimate	Standard Error	df	t	Pr $> \lvert t \rvert$
Intercept		47.85	1.90	250	25.16	$< .0001$
sch_{id}	7472	-2.11	2.60	250	-0.81	.42
sch_{id}	7829	-5.70	2.69	250	-2.12	.04
sch_{id}	7930	5.40	2.58	250	2.10	.04
sch_{id}	24725	-4.31	2.63	250	-1.64	.10
sch_{id}	25456	2.01	2.63	250	0.77	.44
sch_{id}	25642	-1.45	2.69	250	-0.54	.59
sch_{id}	62821	14.97	2.17	250	6.91	$< .0001$
sch_{id}	68448	1.82	2.66	250	0.68	0.49
sch_{id}	68493	-1.52	2.66	250	-0.57	0.57
sch_{id}	72292	0	.	.	.	.

Estimated model parameters, σ and β

From the table from R: $\hat{\sigma}^2 = 72.34$ . . . Are these different from SAS?

Solution for Fixed Effects

Coefficients:

	Estimate	Std. Error	t value	Pr($>|t|$)
(Intercept)	45.7391	1.7735	25.790	$< 2e-16$
id7829	-3.5891	2.6005	-1.380	0.16877
id7930	7.5109	2.4819	3.026	.00273
id24725	-2.1937	2.5365	-0.865	0.38795
id25456	4.1245	2.5365	1.626	0.10519
id25642	0.6609	2.6005	0.254	0.79960
id62821	17.0818	2.0555	8.310	$6.13e-15$
id68448	3.9275	2.5672	1.530	0.12730
id68493	0.5942	2.5672	0.231	0.81715
id72292	2.1109	2.6005	0.812	0.41773

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.

$$\begin{array}{c} \cdots\cdots\cdots \\ y \end{array}$$

Random Effects ANOVA: The Model

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

where

μ 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, ie., 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, τ^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 γ is Constant (overall mean) and U_j is Random, $\mathcal{N}(0, \tau^2)$
- β_{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:

Parameter		Estimate	std error
Fixed effect(s)			
	γ	48.87	1.83
Variance parameters			
intercept	τ^2	30.52	14.48
residual	σ^2	72.23	11.20

R does not give "std error", but it gives "Std Dev" $= \sqrt{\tau^2}$

A Closer Look (Intuitive)

$$
\begin{aligned}
Y_{ij} &= \beta_{0j} + R_{ij} \\
&= \gamma + U_j + R_{ij}
\end{aligned}
$$

- 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)

$$\begin{aligned} Y_{ij} &= \beta_{0j} + R_{ij} \\ &= \gamma + U_j + R_{ij} \end{aligned}$$

Within groups, observations are dependent.

$$\begin{aligned} Y_{ij} &= \gamma + U_j + R_{ij} \\ Y_{i'j} &= \gamma + U_j + R_{i'j} \end{aligned}$$

γ 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)

$$
\begin{aligned}
\text{cov}(Y_{ij}, Y_{i'j}) &\equiv \text{E}[(Y_{ij} - \text{E}(Y_{ij}))(Y_{i'j} - \text{E}(Y_{i'j}))] \\
&= \text{E}[((\gamma + U_j + R_{ij}) - \gamma)((\gamma + U_j + R_{i'j}) - \gamma)] \\
&= \text{E}[(U_j + R_{ij})(U_j + R_{i'j})] \\
&= \text{E}[U_j^2 + U_j R_{i'j} + R_{ij} U_j + R_{ij} R_{i'j}] \\
&= \text{E}[U_j^2] + \text{E}[U_j R_{i'j}] + \text{E}[R_{ij} U_j] + \text{E}[R_{ij} R_{i'j}] \\
&= \text{E}[U_j^2] \\
&= \text{E}[(U_j - 0)^2] \\
&\equiv \text{var}(U_j) \\
&= \tau^2
\end{aligned}
$$

Intra-class Correlation

Measure of within group dependency.

$$
\begin{aligned}
\rho_I &\equiv \text{corr}(Y_{ij}, Y_{i'j}) \\
&= \frac{\text{cov}(Y_{ij}, Y_{i'j})}{\sqrt{\text{var}(Y_{ij})}\sqrt{\text{var}(Y_{i'j})}} \\
&= \frac{\tau^2}{\sqrt{\tau^2+\sigma^2}\sqrt{\tau^2+\sigma^2}} \\
&= \frac{\tau^2}{\tau^2+\sigma^2}
\end{aligned}
$$

Intra-class Correlation: ρ_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 $Y_{ij} = \gamma + U_j + R_{ij}$

$$
\begin{pmatrix} Y_{11} \\ Y_{21} \\ \vdots \\ Y_{23,1} \\ Y_{1,2} \\ Y_{2,2} \\ \vdots \\ Y_{20,2} \\ \vdots \\ Y_{1,10} \\ \vdots \\ Y_{20,10} \end{pmatrix}
=
\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{pmatrix}
\begin{pmatrix} \gamma \end{pmatrix}
+
\begin{pmatrix} U_1 \\ U_1 \\ \vdots \\ U_1 \\ U_2 \\ U_2 \\ \vdots \\ U_2 \\ \vdots \\ U_{10} \\ \vdots \\ U_{10} \end{pmatrix}
+
\begin{pmatrix} R_{11} \\ R_{21} \\ \vdots \\ R_{23,1} \\ R_{1,2} \\ R_{2,2} \\ \vdots \\ R_{20,2} \\ \vdots \\ R_{1,10} \\ \vdots \\ R_{20,10} \end{pmatrix}
$$

Matrix representation (continued)

$$\begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \\ \vdots \\ \boldsymbol{Y}_{10} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_{10} \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \boldsymbol{U}_1 \\ \boldsymbol{U}_2 \\ \vdots \\ \boldsymbol{U}_{10} \end{pmatrix} + \begin{pmatrix} \boldsymbol{R}_1 \\ \boldsymbol{R}_2 \\ \vdots \\ \boldsymbol{R}_{10} \end{pmatrix}$$

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{U} + \boldsymbol{R}$$

Matrix Representation (continued)

For group j, the distribution of $\boldsymbol{Y}_j$ is

$$\boldsymbol{Y}_j \sim \mathcal{N}_{n_j}(\boldsymbol{X}_j\boldsymbol{\beta}, \boldsymbol{\Sigma}_j)$$

where

$\boldsymbol{X}_j = \mathbf{1}_j$ is the an $(n_j \times 1)$ column vector with elements equal to one.

$\boldsymbol{\beta} = \gamma$

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

$$\boldsymbol{\Sigma}_j = \sigma^2 \boldsymbol{I}_j + \tau^2 \mathbf{1}_j \mathbf{1}'_j$$

Covariance Matrix

$$\underset{(n_j \times n_j)}{\boldsymbol{\Sigma}_j} = \sigma^2 \boldsymbol{I}_j + \tau^2 \mathbf{1}_j \mathbf{1}_j'$$

$$= \sigma^2 \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} + \tau^2 \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix}$$

$$= \begin{pmatrix} (\tau^2 + \sigma^2) & \tau^2 & \dots & \tau^2 \\ \tau^2 & (\tau^2 + \sigma^2) & \dots & \tau^2 \\ \vdots & \vdots & \ddots & \vdots \\ \tau^2 & \tau^2 & \dots & (\tau^2 + \sigma^2) \end{pmatrix}$$

Covariance Matrix (continued)

$$\boldsymbol{\Sigma}_j = \sigma^2 \boldsymbol{I}_j + \tau^2 \mathbf{1}_j \mathbf{1}_j'$$

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 $\boldsymbol{\Sigma}$.

Covariance Matrix: Y

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

$$\begin{pmatrix} \underset{(n_1 \times n_1)}{\boldsymbol{\Sigma}} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \underset{(n_2 \times n_2)}{\boldsymbol{\Sigma}} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \underset{(n_N \times n_N)}{\boldsymbol{\Sigma}} \end{pmatrix}$$

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).

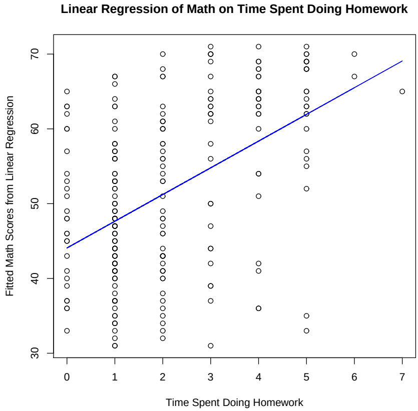
i.e., Summarize groups by their regression parameters.

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

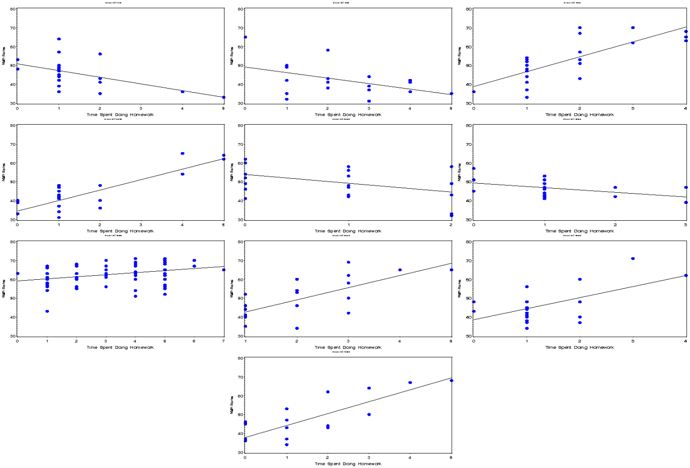
i.e., Analysis of the summary statistics.

NELS88: Overall Regression

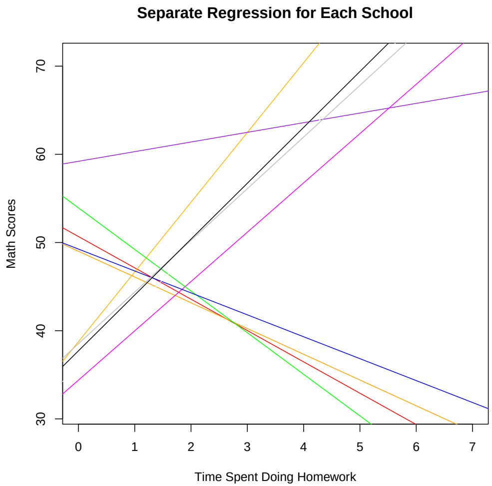
Before we do the two stage analysis,. . .



Individual Schools' Data



Individual Schools' Regressions



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:

$$\begin{aligned} (\text{math})_{ij} &= \beta_{0j} + \beta_{1j}(\text{homew})_{ij} + R_{ij} \\ Y_{ij} &= \beta_{0j} + \beta_{1j}x_{ij} + R_{ij} \end{aligned}$$

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

Stage One

Model fit to each of the 10 NELS88 schools:

School ID	$\hat{\beta}_{0j}$	$\hat{\beta}_{1j}$	R^2
7272	50.68354	-3.55380	.2780
7829	49.01229	-2.92012	.2111
7930	38.75000	7.90909	.6007
24725	34.39382	5.59266	.7002
25456	53.93863	-4.71841	.1873
25642	49.25896	-2.48606	.2186
62821	59.21022	1.09464	.1105
68448	36.05535	6.49631	.5098
68493	38.52000	5.86000	.3137
72292	37.71392	6.3350	.6417

Stage One (continued)

Parameter	n	mean	std dev	s.e.	variance
$\hat{\beta}_{0j}$	10	44.75	8.66	2.74	74.93
$\hat{\beta}_{1j}$	10	1.96	4.98	1.58	24.76

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

Source	DF	Sum of Squares	Mean Square	F	Pr > F
Model	0	0		.	.
Error	9	674.40021	74.93336		
Corrected Total	9	674.40021			

Stage Two: Intercepts

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

Variable	DF	Parameter Estimate	Standard Error	t	Pr $> \|t\|$
Intercept	1	44.75367	2.73740	16.35	$< .0001$

Descriptives statistics

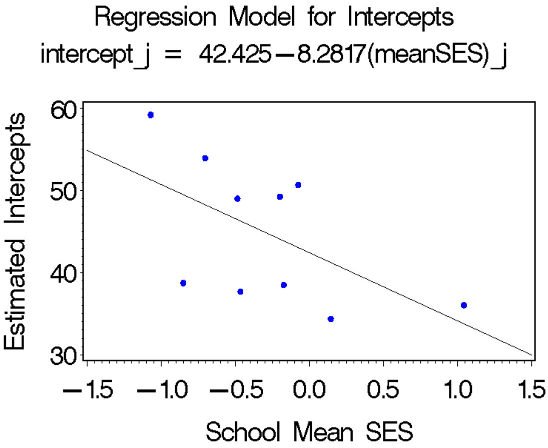
A little more complex model

Model for the level 2 analysis:

$$\begin{aligned}\hat{\beta}_{0j} &= \gamma_{00} + \gamma_{01}(\text{mean SES}_j) + U_{0j} \\ &= \gamma_{00} + \gamma_{01} Z_j + U_{0j}\end{aligned}$$

where $U_{0j} \sim \mathcal{N}(0, \tau_0^2)$.

SAS PROC REG


Regression Model for Intercepts
intercept_j = 42.425−8.2817(meanSES)_j
Estimated Intercepts
60
50
40
30
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
School Mean SES

SAS PROC REG (intercepts)

The REG Procedure Model: MODEL2
Dependent Variable: alpha

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F	Pr > F
Model	1	219.10398	219.10398	3.85	.0854
Error	8	455.29623	56.91203		
Corrected Total	9	674.40021			

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	$\Pr > \lvert t \rvert$
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

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F	Pr > F
Model	0	0	.	.	.
Error	9	222.85163	24.76129		
Corrected Total	9	222.85163			

Slopes: No Explanatory Variables

Root MSE	4.97607	R-Square	0.0000
Dependent Mean	1.96093	Adj R-Sq	0.0000
Coeff Var	253.76069		

Parameter Estimates

Variable	F	Parameter Estimate	Standard Error	t	Pr $> \|t\|$
Intercept	1	1.96093	1.57357	1.25	0.2442

Descriptives statistics: see page 41

Slopes: More Complex Model

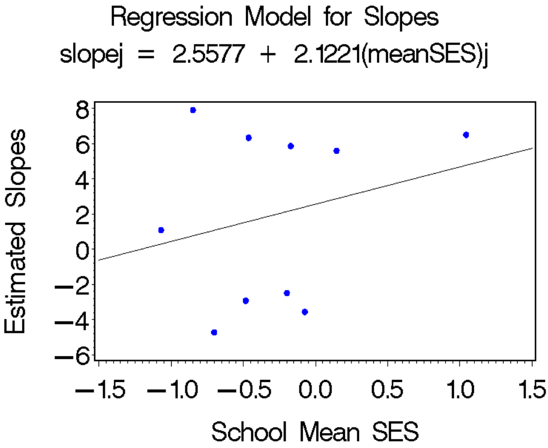
Proposition: The higher the SES, the less of an effect time spent doing homework has on math achievement.

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

Diagram for this:

$$\begin{array}{ccc} & Z & \\ \ldots & \downarrow & \ldots\ldots \\ x & \rightarrow & y \end{array}$$

SAS PROC REG


Regression Model for Slopes
slopej = 2.5577 + 2.1221(meanSES)j
8
6
4
2
0
−2
−4
−6
Estimated Slopes
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
School Mean SES

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

Source	DF	Sum of Squares	Mean Square	F	Pr>F
Model	1	14.38612	14.38612	0.55	.48
Error	8	208.46552	26.05819		
Corrected Total	9	222.85163			

Slopes: SAS PROC REG Output

Root MSE	5.10472	Dependent Mean	1.96093
R-Square	0.0646	Adj R^2	-0.0524

		Parameter	Standard		
Variable	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 τ_{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 $R_{ij} \sim \mathcal{N}(0, \sigma^2)$ and independent.

. . . and more compactly,

$$\boldsymbol{Y}_j = \boldsymbol{X}_j\boldsymbol{\beta}_j + \boldsymbol{R}_j$$

where $\boldsymbol{Y}_j \sim \mathcal{N}(\boldsymbol{X}_j\boldsymbol{\beta}_j, \sigma^2\boldsymbol{I}) i.i.d.$.

HLM: Level 2

Level 2:

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

where

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

. . . or more compactly,

$$\boldsymbol{\beta}_j = \boldsymbol{Z}_j\boldsymbol{\Gamma} + \boldsymbol{U}_j$$

where $\boldsymbol{U}_j \sim \mathcal{N}(\mathbf{0}, \boldsymbol{T})\, i.i.d.$

Linear Mixed Model

Put models for β_{0j} and β_{1j} into the level 1 model:

$$
\begin{aligned}
Y_{ij} &= \beta_{0j} + \beta_{1j} x_{ij} + R_{ij} \\
&= (\gamma_{0o} + U_{0j}) + (\gamma_{1o} + U_{1j}) x_{ij} + R_{ij} \\
&= \underbrace{\gamma_{00} + \gamma_{10} x_{ij}}_{\text{fixed part}} + \underbrace{U_{0j} + U_{1j} x_{ij} + R_{ij}}_{\text{random part}}
\end{aligned}
$$

Linear Mixed Model (continued)

Given all the assumptions of the hierarchical model,

$$Y_{ij} \sim \mathcal{N}((\gamma_{00} + \gamma_{10}x_{ij}), \text{var}(Y_{ij})) \quad i.i.d.$$

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

This is the **Marginal Model** for Y_{ij}.

In Matrix Form

The Model:

$$\boldsymbol{Y}_j = \boldsymbol{X}_j\boldsymbol{\Gamma} + \boldsymbol{Z}_j\boldsymbol{U}_j + \boldsymbol{R}_j$$

where

$$\boldsymbol{Y}_j \sim \mathcal{N}_{n_j}(\boldsymbol{X}_j\boldsymbol{\Gamma}, \boldsymbol{V}_j) i.i.d.$$

and the covariance matrix equals,

$$\boldsymbol{V}_j = \boldsymbol{Z}_j\boldsymbol{T}\boldsymbol{Z}_j' + \sigma^2\boldsymbol{I}$$

and later for longitudinal data

$$\boldsymbol{V}_j = \boldsymbol{Z}_j\boldsymbol{T}\boldsymbol{Z}_j' + \boldsymbol{\Sigma}_j.$$

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.