Hierarchical Linear Models Using SAS®

Iowa SAS Users Group Conference

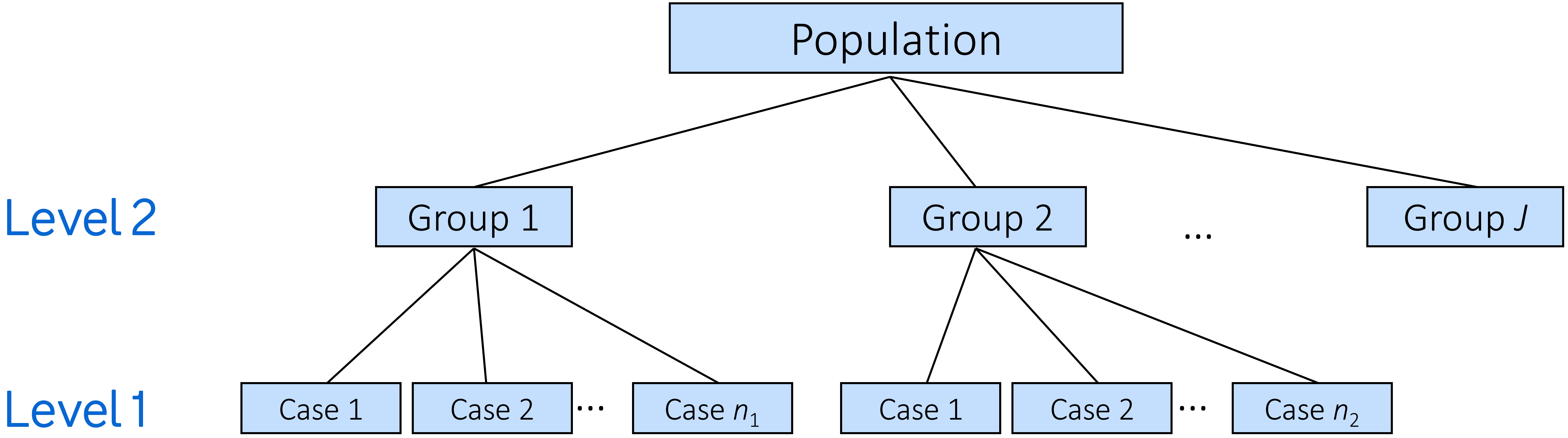
May 13, 2024

Jacqueline Johnson, DrPH, SAS Global Academic Programs

Hierarchical Data Structures

– Hierarchical data structures are those in which multiple micro-level units are sampled for each macro-level unit.

– A common hierarchical data structure is when individuals (micro-units) are sampled from naturally occurring groups (macro-units).

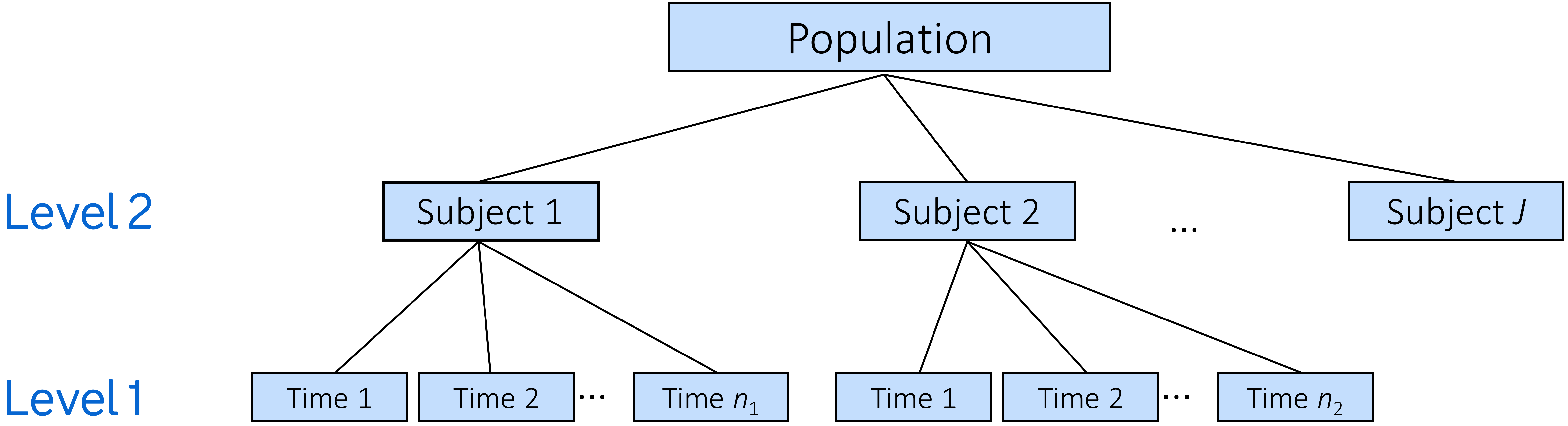


Dependence in Hierarchical Data

- Because many micro-level observations come from the same macro-level unit, this produces dependence in the data.
  - Students attending the same school might have more similar academic outcomes than students attending different schools.
  - Employees working with the same manager might have more similar problem-solving strategies than employees working with different managers.
- Multilevel models provide a way to model this dependence, whereas more traditional models do not.

Longitudinal Data Structures

- Longitudinal data structures arise when the same units are sampled repeatedly over time.
- Longitudinal data are useful for tracking change in an outcome over time (for example, response to a drug).



Dependence in Longitudinal Data

- Because repeated measures are collected on the same unit, this produces dependence in the data.
- Example: Employee job performance is tracked over a period of four years.
  - Some employees perform at consistently higher levels compared to other employees.
  - Some employees increase in performance at a steeper rate over time compared to other employees.
- Again, multilevel models provide a way to model this dependence, whereas more traditional models do not.

The High School and Beyond (hsb) Data Set

- Data are from a 1982 survey of US public and catholic high schools.
- 7,185 students from 160 schools.
- 90 public schools, 60 catholic schools.
- 14 to 67 students per school.
- Variables:
  - Math achievement score for student
  - Socio-economic status of student's family, centered at zero

The High School and Beyond (hsb) Data Set

Questions of Interest

- How much do US high schools vary in mean math achievement?
- Is math achievement related to student SES?
- Is the strength of the relationship between SES and math scores similar across schools? Or is SES a more important predictor of some schools and not others?
- How do public and Catholic high schools compare in mean math achievement and in the strength of the SES-math achievement relationship?

Multilevel Modeling

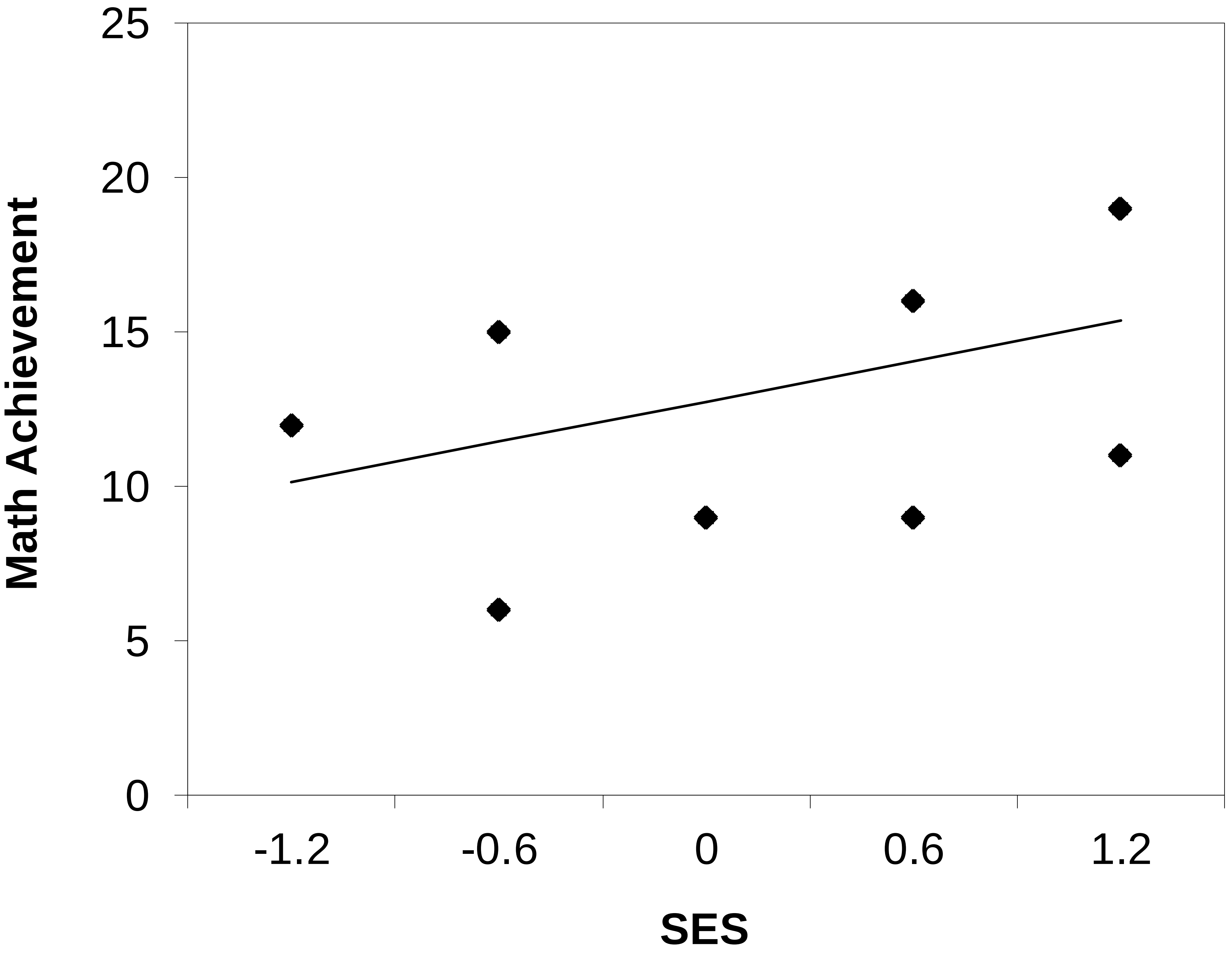
- Multilevel modeling does not incorporate schools as a fixed effects predictor, but rather treats schools as randomly sampled from a population.
- Effects are not estimated individually for each school but are assumed to have a particular distribution across the population of schools.
- Nested data structure
- Level 1 = students
- Level 2 = schools

The MIXED Procedure

General form of the MIXED procedure:

```
PROC MIXED options;
      CLASS classification variables;
      MODEL outcome = fixed-effects / options;
      RANDOM random effects / options;
RUN;
```

Fixed Effects Regression Model



Fixed Effects Regression Model

Level 1 Equation:

$$\text{Math}_{ij} = b_{0j} + b_{1j}\text{SES}_{ij} + \varepsilon_{ij} \qquad \varepsilon_{ij} \sim N(0, \sigma^2)$$

Level 2 Equations:

$$b_{0j} = \beta_{00}$$

$$b_{1j} = \beta_{10}$$

Reduced-Form Equation:

$$\text{Math}_{ij} = \underbrace{\beta_{00} + \beta_{10}\text{SES}_{ij}}_{\text{Fixed-Effects}} + \varepsilon_{ij}$$

Fixed Effects Regression Model

Reduced-Form Equation:

$$\mathrm{Math}_{ij} = \underbrace{\beta_{00} + \beta_{10}\mathrm{SES}_{ij}}_{\text{Fixed-Effects}} + \varepsilon_{ij}$$

```
proc mixed data=mixed.hsb cl covtest;
  model student_mathach = student_ses / solution;
run;
```

Fixed Effects Regression Model

Assumes Independence of All Observations

Covariance Parameter Estimates							
Cov Parm	**Estimate**	**Standard Error**	**Z Value**	**Pr > Z**	**Alpha**	**Lower**	**Upper**
Residual	41.1588	0.6868	59.93	<.0001	0.05	39.8452	42.5388

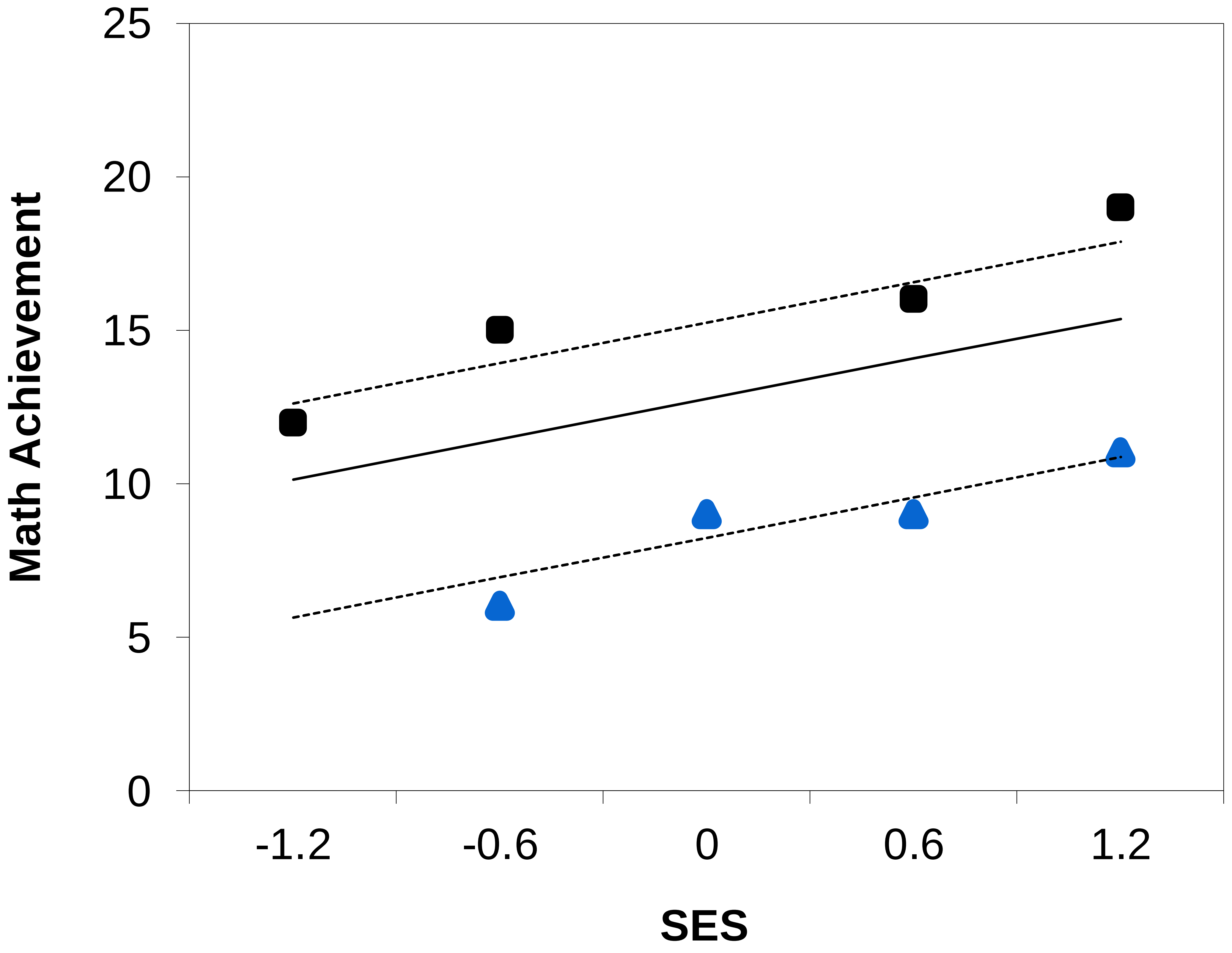
41.16 = Level 1 variability of math achievement scores

Solution for Fixed Effects					
Effect	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
Intercept	12.7474	0.07569	7183	168.42	<.0001
student_ses	3.1839	0.09712	7183	32.78	<.0001

12.75 = expected math achievement of a student from an average SES family

3.18 = expected increase in math achievement per one-unit increase in SES

Random Intercepts Model



Random Intercepts Model

Level 1 Equation:

$$\text{Math}_{ij} = b_{0j} + b_{1j}\text{SES}_{ij} + \varepsilon_{ij} \qquad \varepsilon_{ij} \sim N(0, \sigma^2)$$

Level 2 Equations:

$$b_{0j} = \beta_{00} + b*_{0j} \qquad b*_{0j} \sim N(0, \sigma^2_{00})$$

$$b_{1j} = \beta_{10}$$

Reduced-Form Equation:

$$\text{Math}_{ij} = \left(\beta_{00} + b*_{0j}\right) + \beta_{10}\text{SES}_{ij} + \varepsilon_{ij}$$

$$= \underbrace{\left(\beta_{00} + \beta_{10}\text{SES}_{ij}\right)}_{\text{Fixed-Effects}} + \underbrace{b*_{0j}}_{\text{Random-Effect}} + \varepsilon_{ij}$$

Random Intercepts Model

Reduced-Form Equation:

$$\text{Math}_{ij} = \left(\beta_{00} + b^*_{0j}\right) + \beta_{10}\text{SES}_{ij} + \varepsilon_{ij}$$

$$= \underbrace{\left(\beta_{00} + \beta_{10}\text{SES}_{ij}\right)}_{\text{Fixed-Effects}} + \underbrace{b^*_{0j}}_{\text{Random-Effect}} + \varepsilon_{ij}$$

```
proc mixed data=mixed.hsb cl covtest;
  model student_mathach = student_ses / solution ddfm=bw;
  random intercept / subject=school_id;
run;
```

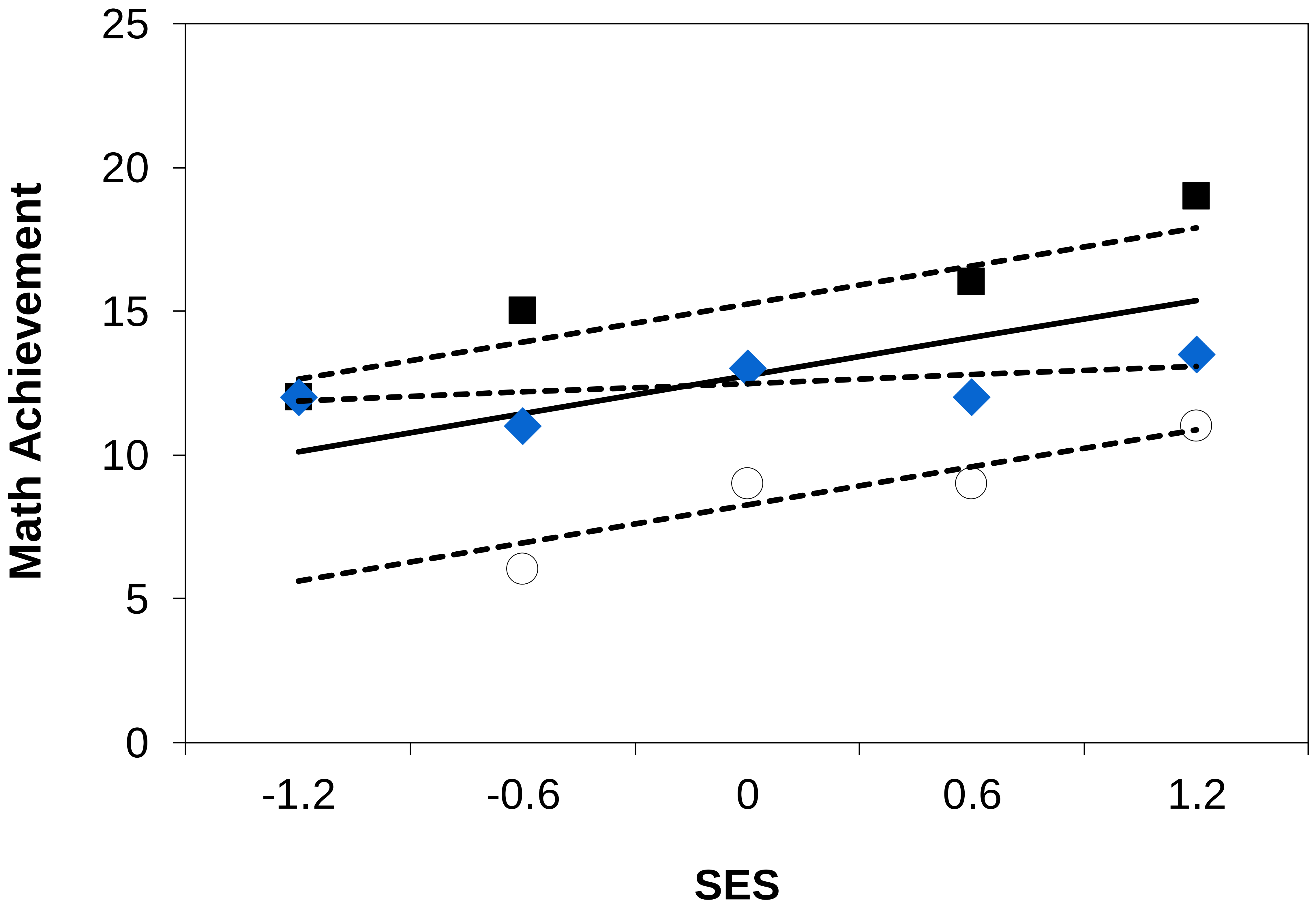
Random Intercepts Model

Covariance Parameter Estimates								
Cov Parm	**Subject**	**Estimate**	**Standard Error**	**Z Value**	**Pr > Z**	**Alpha**	**Lower**	**Upper**
Intercept	school_ID	4.7665	0.6549	7.28	<.0001	0.05	3.7045	6.3636
Residual		37.0346	0.6254	59.22	<.0001	0.05	35.8388	38.2916

4.77 = Variability of the school intercepts – significantly different from zero
37.03 = Level 1 variability of math achievement scores

Solution for Fixed Effects					
Effect	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
Intercept	12.6575	0.1880	159	67.34	<.0001
student_ses	2.3903	0.1057	7024	22.61	<.0001

Random Slopes and Intercepts Model



Random Slopes and Intercepts Model

Level 1 Equation:

$$\text{Math}_{ij} = b_{0j} + b_{1j}\text{SES}_{ij} + \varepsilon_{ij} \qquad \varepsilon_{ij} \sim N(0, \sigma^2)$$

Level 2 Equations:

$$b_{0j} = \beta_{00} + b^*_{0j}$$
$$b_{1j} = \beta_{10} + b^*_{1j}$$

$$\begin{pmatrix} b^*_{0j} \\ b^*_{1j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{00} & \\ \sigma_{10} & \sigma^2_{11} \end{pmatrix}\right]$$

Reduced-Form Equation:

$$\text{Math}_{ij} = \left(\beta_{00} + b^*_{0j}\right) + \left(\beta_{10} + b^*_{1j}\right)\text{SES}_{ij} + \varepsilon_{ij}$$
$$= \underbrace{\left(\beta_{00} + \beta_{10}\text{SES}_{ij}\right)}_{\text{Fixed-Effects}} + \underbrace{\left(b^*_{0j} + b^*_{1j}\,\text{SES}_{ij}\right)}_{\text{Random-Effects}} + \varepsilon_{ij}$$

Random Slopes and Intercepts Model

Reduced-Form Equation:

$$\text{Math}_{ij} = (\beta_{00} + b^*_{0j}) + (\beta_{10} + b^*_{1j})\text{SES}_{ij} + \varepsilon_{ij}$$

$$= \underbrace{(\beta_{00} + \beta_{10}\text{SES}_{ij})}_{\text{Fixed-Effects}} + \underbrace{(b^*_{0j} + b^*_{1j}\,\text{SES}_{ij})}_{\text{Random-Effects}} + \varepsilon_{ij}$$

$$\begin{pmatrix} b^*_{0j} \\ b^*_{1j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{00} & \\ \sigma_{10} & \sigma^2_{11} \end{pmatrix}\right]$$

G matrix

```
proc mixed data=mixed.hsb cl covtest;
  model student_mathach = student_ses / solution ddfm=bw;
  random intercept student_ses / subject=school_id  type=un g gcorr;
run;
```

Random Slopes and Intercepts Model

Covariance Parameter Estimates								
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z	Alpha	Lower	Upper
UN(1,1)	school_ID	4.8278	0.6719	7.18	<.0001	0.05	3.7406	6.4716
UN(2,1)	school_ID	-0.1547	0.2988	-0.52	0.6046	0.05	-0.7403	0.4308
UN(2,2)	school_ID	0.4127	0.2350	1.76	0.0395	0.05	0.1730	1.9418
Residual		36.8304	0.6293	58.52	<.0001	0.05	35.6274	38.0956

4.82 = Variability of the school intercepts — significantly different from zero
-0.15 = Covariance of intercepts and slopes — negative, not different from zero
0.41 = Variability of the school slopes — significantly different from zero
36.83 = Level 1 variability of math achievement scores

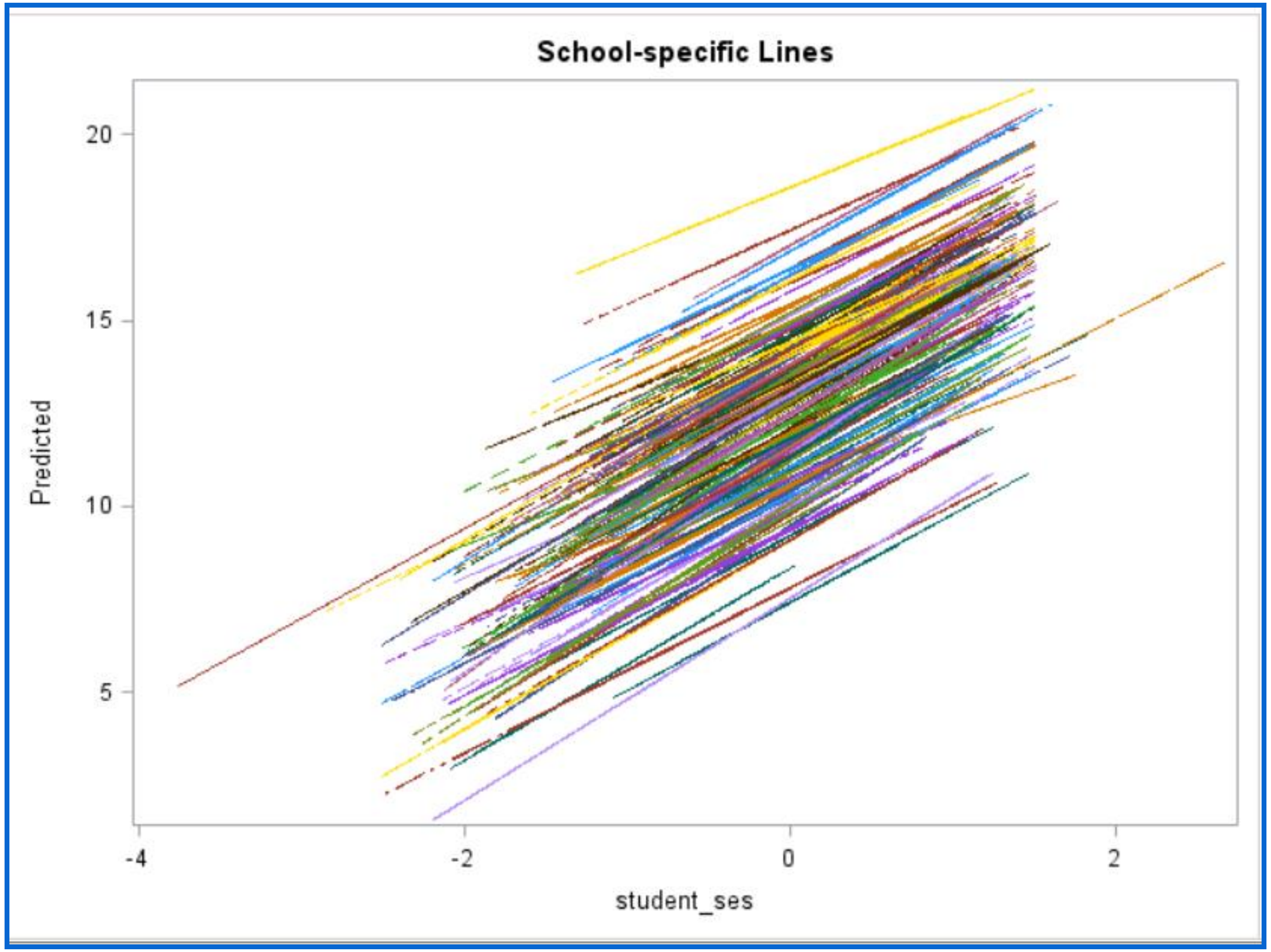
Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	12.6651	0.1898	159	66.72	<.0001
student_ses	2.3938	0.1181	7024	20.27	<.0001

Comparison of Models

Model Effects		Variance Parameter Estimates				Fixed Effect Parameter Estimates		Model Fit
Intercept	Student SES Slope	Residual variance	Variance of Random Intercept	Variance of Random SES Slope	Covariance of Intercept and Slope	Intercept	Slope Student SES	AIC
Fixed	Fixed	41.20				12.74	3.18	47,106
Random	Fixed	37.03	4.77			12.66	2.39	46,649
Random	Random	36.83	4.83	0.41	-0.15	12.67	2.39	46,648

Estimated School Relationships

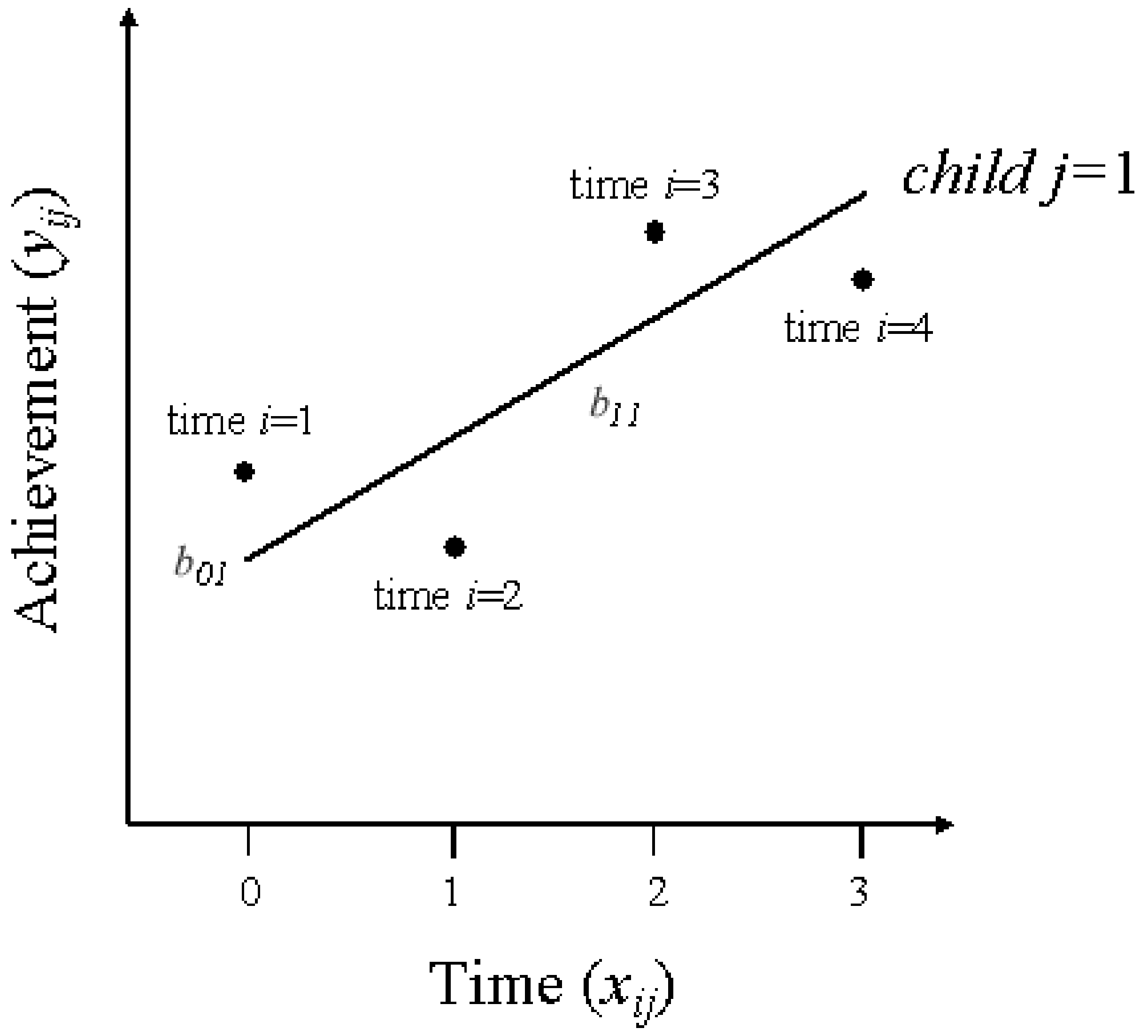
```
proc mixed data=mixed.hsb cl covtest;
  model student_mathach = student_ses /
     solution
     ddfm=bw
     outpred=predicted;
  random intercept student_ses /
     subject=school_id
     type=un;
run;

title 'School-specific Lines';
proc sgplot data=predicted;
  series y=pred x=student_ses /
     group=school_id;
run; quit;
```

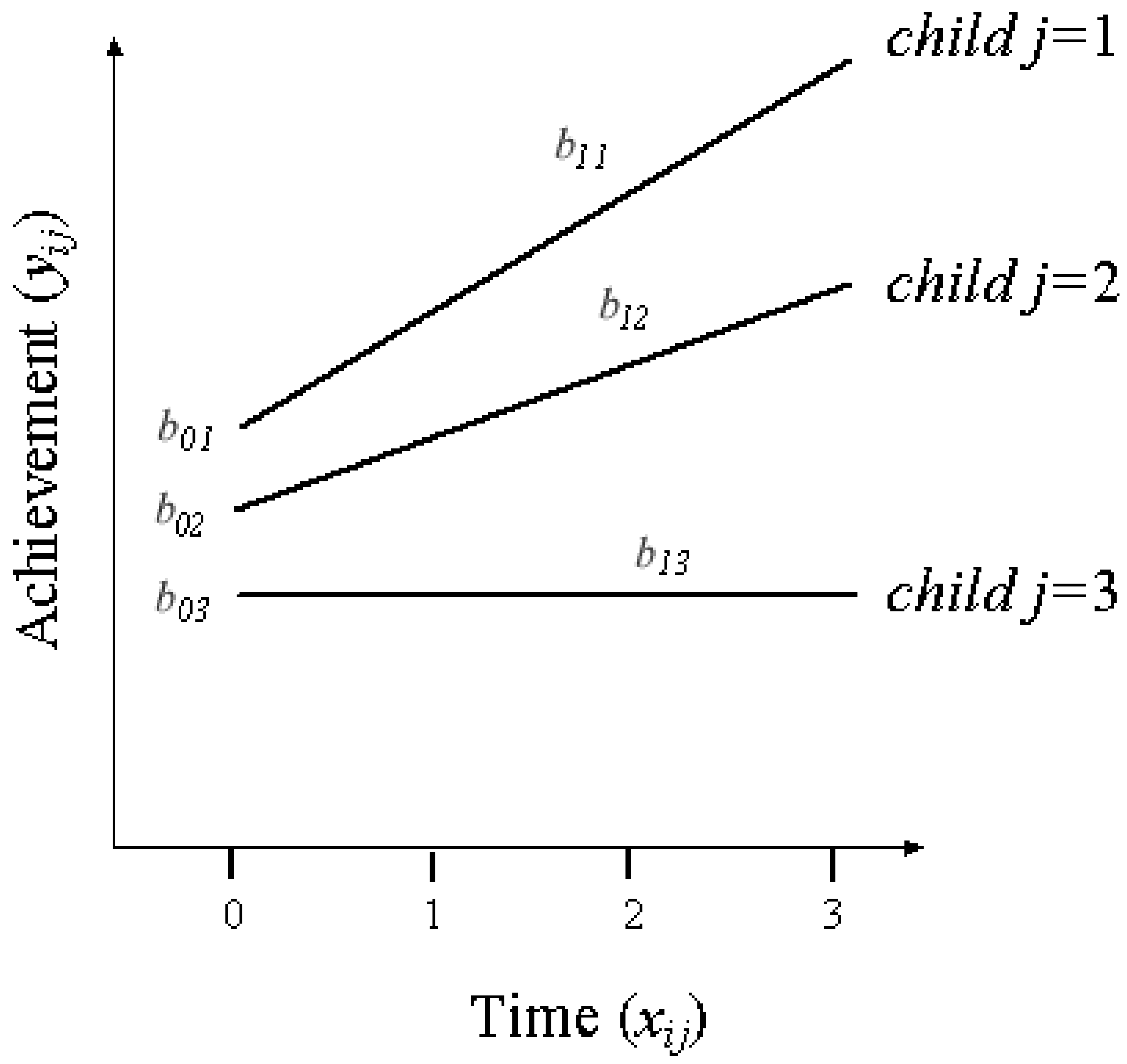


A Comment on Notation

	Popular social science textbooks	SAS and statistics texts
Level 1:	$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + r_{ij}$	$y_{ij} = b_{0j} + b_{1j} x_{ij} + \varepsilon_{ij}$
Level 2:	$\beta_{0j} = \gamma_{00} + \gamma_{01} w_j + u_{0j}$ $\beta_{1j} = \gamma_{10} + \gamma_{11} w_j + u_{1j}$	$b_{0j} = \beta_{00} + \beta_{01} w_j + b*_{0j}$ $b_{1j} = \beta_{10} + \beta_{11} w_j + b*_{1j}$
Variance:	$r_{ij} \sim N(0, \sigma^2)$ $\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \\ \tau_{10} & \tau_{11} \end{bmatrix} \right)$	$\varepsilon_{ij} \sim N(0, \sigma^2)$ $\begin{bmatrix} b*_{0j} \\ b*_{1j} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{00} & \\ \sigma_{10} & \sigma^2_{11} \end{bmatrix} \right)$

Longitudinal Data: Time Nested within a Child





Level 1

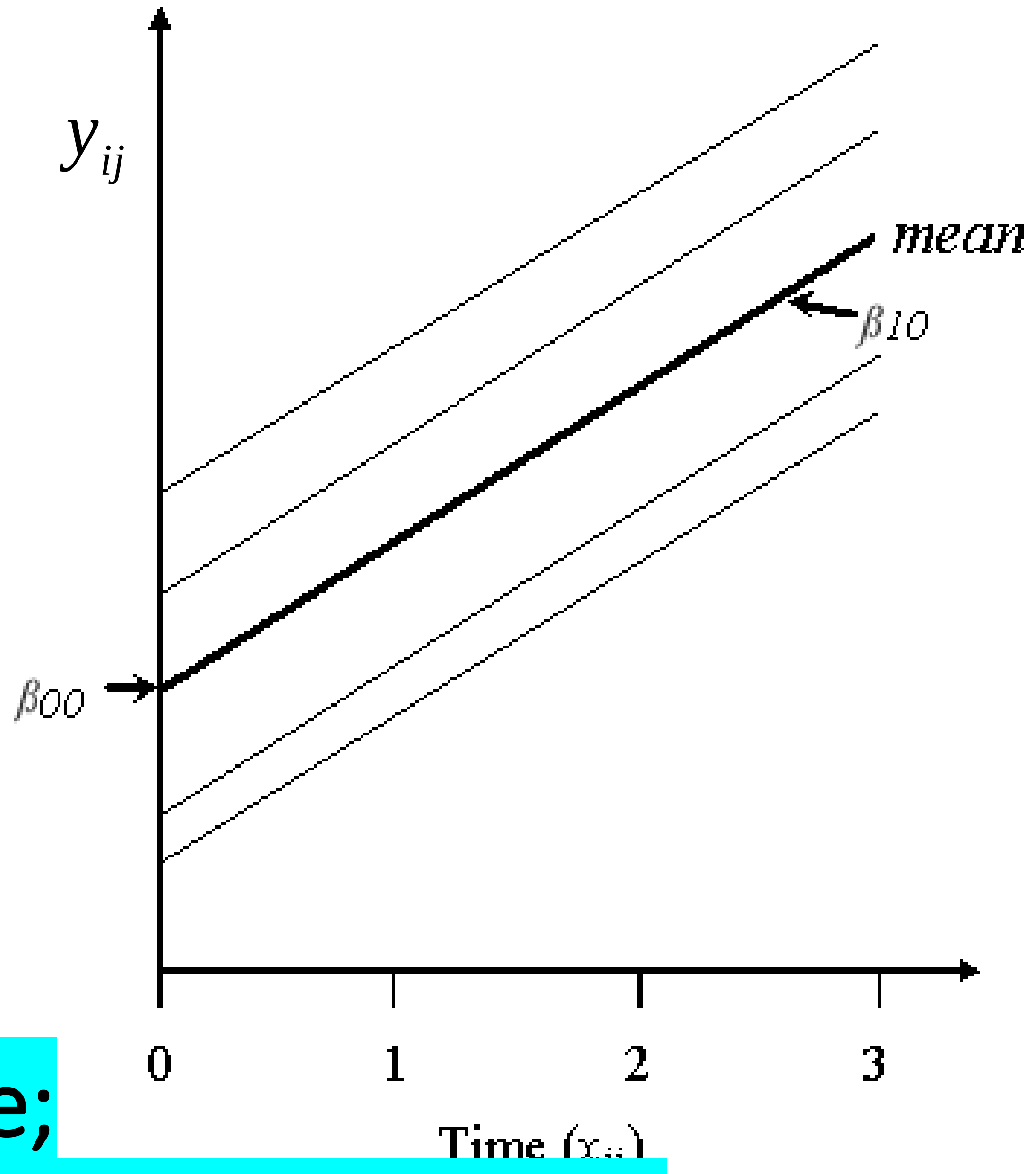
$$y_{ij} = b_{0j} + b_{1j} x_{ij} + \varepsilon_{ij}$$

Level 2

$$b_{0j} = \beta_{00} + b^*_{0j}$$
$$b_{1j} = \beta_{10} + b^*_{1j}$$

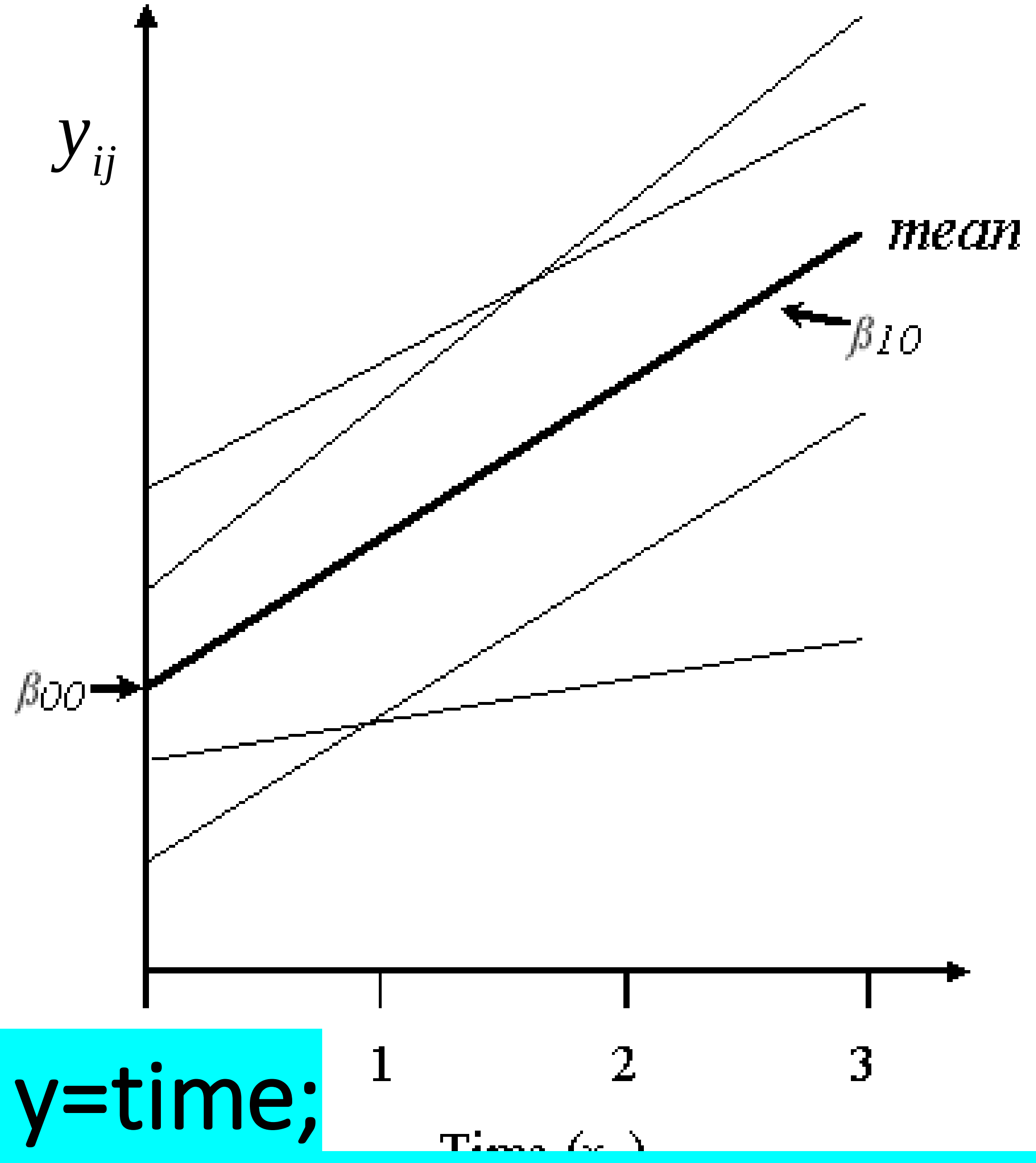
Combinations of Fixed and Random Effects

– random intercept
– fixed slope




```
model y=time;
random intercept /subject=id;
```

– random intercept
– random slope




```
model y=time;
random intercept time/subject=id type=un;
```

Homoscedastic versus Heteroscedastic Level 1

$$
\mathrm{var}(\varepsilon_{ij}) =
\begin{array}{c}
 \\ \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T
\end{array}
\begin{array}{c}
\begin{array}{cccc} \varepsilon_1 & \varepsilon_2 & \cdots & \varepsilon_T \end{array} \\
\begin{pmatrix}
\sigma^2 & & & 0 \\
 & \sigma^2 & & \\
 & & \ddots & \\
0 & & & \sigma^2
\end{pmatrix}
\end{array}
$$

– Homoscedastic Level 1 residual variance matrix for T time points (equal variance at all time points)

$$
\mathrm{var}(\varepsilon_{ij}) =
\begin{array}{c}
 \\ \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T
\end{array}
\begin{array}{c}
\begin{array}{cccc} \varepsilon_1 & \varepsilon_2 & \cdots & \varepsilon_T \end{array} \\
\begin{pmatrix}
\sigma_1^2 & & & 0 \\
 & \sigma_2^2 & & \\
 & & \ddots & \\
0 & & & \sigma_T^2
\end{pmatrix}
\end{array}
$$

– Heteroscedastic Level 1 residual variance matrix for T time points (unequal variance at each time point)

Example: Antisocial Behavior

- N=405 cases drawn from the National Longitudinal Survey of Youth (NLSY)
- Age 6 to 8 years at first assessment; reassessed a maximum of three more times every other year
- Mother's report of child antisocial behavior on six items; each has a 0,1,2 response scale; sum score ranges from 0 to 12
- Two predictors: child gender and level of cognitive support of child in the home at initial assessment

Research question:

- What are the characteristics of trajectories of antisocial behavior, and can these trajectories be predicted by child-level measures?

Homoscedastic Level 1 Variance

```
proc mixed data=mixed.antilong covtest ;
 class id;
 model anti = age/ solution ddfm=bw;
 random intercept age / subject=id type=un g gcorr;
run;
```

Fit Statistics	
-2 Res Log Likelihood	5290.5
AIC (smaller is better)	5298.5
AICC (smaller is better)	5298.6
BIC (smaller is better)	5314.6

Heteroscedastic Level 1 Variances

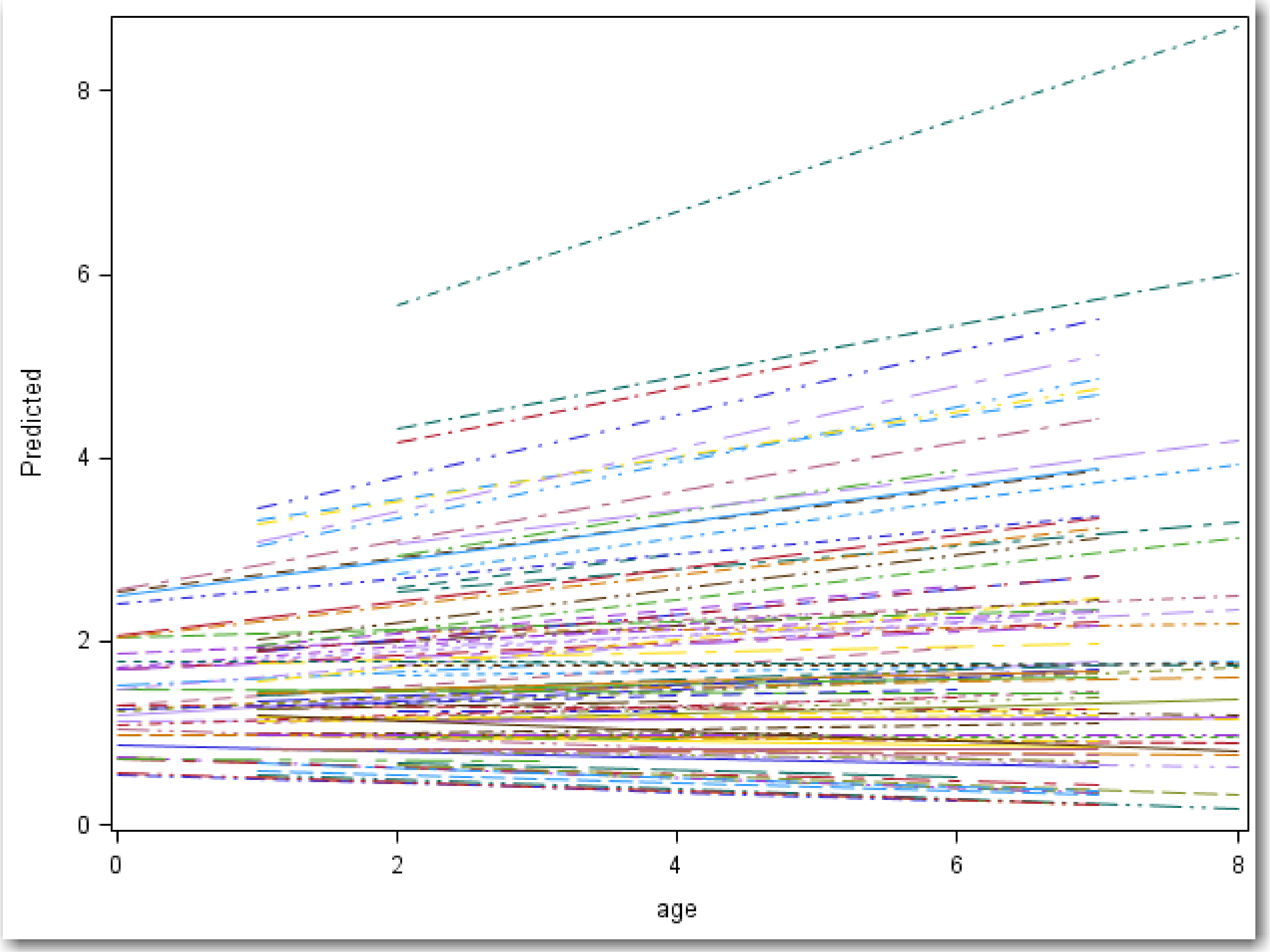
```
proc mixed data=mixed.antilong covtest;
 class id ageclas;
 model anti = age / solution ddfm=bw;
 random intercept age / subject=id type=un g gcorr;
 repeated ageclas / type=un(1) subject=id;
run;
```

Fit Statistics	
-2 Res Log Likelihood	5284.0
AIC (smaller is better)	5308.0
AICC (smaller is better)	5308.3
BIC (smaller is better)	5356.1

Example: Antisocial Behavior

Covariance Parameter Estimates					
Cov Parm	**Subject**	**Estimate**	**Standard Error**	**Z Value**	**Pr Z**
UN(1,1)	id	1.0134	0.2312	4.38	<.0001
UN(2,1)	id	0.07303	0.03831	1.91	0.0566
UN(2,2)	id	0.02295	0.01009	2.27	0.0115
Residual		1.7518	0.1017	17.23	<.0001

Solution for Fixed Effects					
Effect	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
Intercept	1.6289	0.08573	404	19.00	<.0001
age	0.07425	0.01774	956	4.18	<.0001



Conclusions: Antisocial Behavior

- The significant fixed effects reflect that the mean level of antisocial behavior at age six (coded 0 in these models) is 1.63, and antisocial behavior is increasing by 0.07 units per year.
- The significant random effects reflect individual variability among the intercepts (1.01) and among the slopes (0.023). This is also seen in the plot of predicted values for each individual.
- There is a marginally significant (p=0.056) covariance between the intercepts and slopes, and the correlation is 0.48. This indicates that, on average, children reporting higher initial levels of antisocial behavior tend to increase more steeply over time.

Examples of Three-Level Data

- Three-level data might occur because individuals are nested within groups and groups are in turn clustered within still higher-level units.
  - residents within neighborhoods within counties
  - patients within physicians within hospitals
- Three-level data often also occurs in studies of individual change over time when individuals are nested within groups.
  - performance over time nested within person nested within department
  - symptoms over time nested within patient nested within physician

Variance Partitioning

Variance Decomposition: $V(y_{ijk}) = \sigma^{2(3)}_{00} + \sigma^{2(2)}_{00} + \sigma^2$

Suppose *i* is patient, *j* is physician, and *k* is hospital:

$\sigma^2 / \left(\sigma^{2(3)}_{00} + \sigma^{2(2)}_{00} + \sigma^2\right)$ is the proportion of variance due to patients within physicians.

$\sigma^{2(2)}_{00} / \left(\sigma^{2(3)}_{00} + \sigma^{2(2)}_{00} + \sigma^2\right)$ is the proportion of variance due to physicians within hospitals.

$\sigma^{2(3)}_{00} / \left(\sigma^{2(3)}_{00} + \sigma^{2(2)}_{00} + \sigma^2\right)$ is the proportion of variance due to hospitals.

Specification in the MIXED Procedure

```
proc mixed data=mydata;
   class L3_ID L2_ID;
   model y = x w1 w2 / solution ddfm=kr2;
   random intercept / subject=L2_ID(L3_ID);
   random intercept / subject=L3_ID;
run;
```

- Notice the use of two RANDOM statements to define the random intercepts at Level 2 and Level 3.
- Notice the use of the CLASS statement for the Level 2 and Level 3 ID variables.

Random Slopes for Three-Level Data

- In three-level models, you can have the following predictors:
  - predictors at Level 1 with random effects that vary over the second or third levels (or both) of the model
  - predictors at Level 2 with random effects that vary over the third level of the model
- When the three levels consist of repeated measures within persons within groups, random slopes for time are common at both Levels 2 and 3.
  - These reflect that rates of change over time might differ across groups as well as across individuals within groups.
- Random slopes are most common with repeated measures.

Random Slopes for Three Level Data

- Notice that in this model, there are random intercepts and slopes for x (time) at both Levels 1 and 2.

Level 1: $y_{ijk} = b_{0jk} + b_{1jk} x_{ijk} + \varepsilon_{ijk} \qquad \varepsilon_{ijk} \sim N(0, \sigma^2)$

Level 2: $b_{0jk} = b_{00k} + b*_{0jk}$

$b_{1jk} = b_{10k} + b*_{1jk}$

$$\begin{pmatrix} b*_{0jk} \\ b*_{1jk} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^{2(2)}_{00} & \\ \sigma^{(2)}_{10} & \sigma^{2(2)}_{11} \end{pmatrix}\right]$$

Level 3: $b_{00k} = \beta_{000} + b*_{00k}$

$b_{10k} = \beta_{100} + b*_{10k}$

$$\begin{pmatrix} b*_{00k} \\ b*_{10k} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^{2(3)}_{00} & \\ \sigma^{(3)}_{10} & \sigma^{2(3)}_{11} \end{pmatrix}\right]$$

Reduced Form:

$$y_{ijk} = \left(\beta_{000} + \beta_{100} x_{ijk}\right) + \left(b*_{00k} + b*_{10k}\, x_{ijk}\right) + \left(b*_{0jk} + b*_{1jk}\, x_{ijk}\right) + \varepsilon_{ijk}$$

Interpretations of Multilevel Equations

- For a three-level, linear growth model, the equations for the three levels can be conceptualized as follows:

Level 1: $y_{ijk} = b_{0jk} + b_{1jk} x_{ijk} + \varepsilon_{ijk}$ } Individual trajectory + time-specific residuals

Level 2:
$$b_{0jk} = b_{00k} + b^*_{0jk}$$
$$b_{1jk} = b_{10k} + b^*_{1jk}$$
} Group mean trajectory + individual variability around group trajectory

Level 3:
$$b_{00k} = \beta_{000} + b^*_{00k}$$
$$b_{10k} = \beta_{100} + b^*_{10k}$$
} Mean trajectory over all groups + variability across the group trajectories

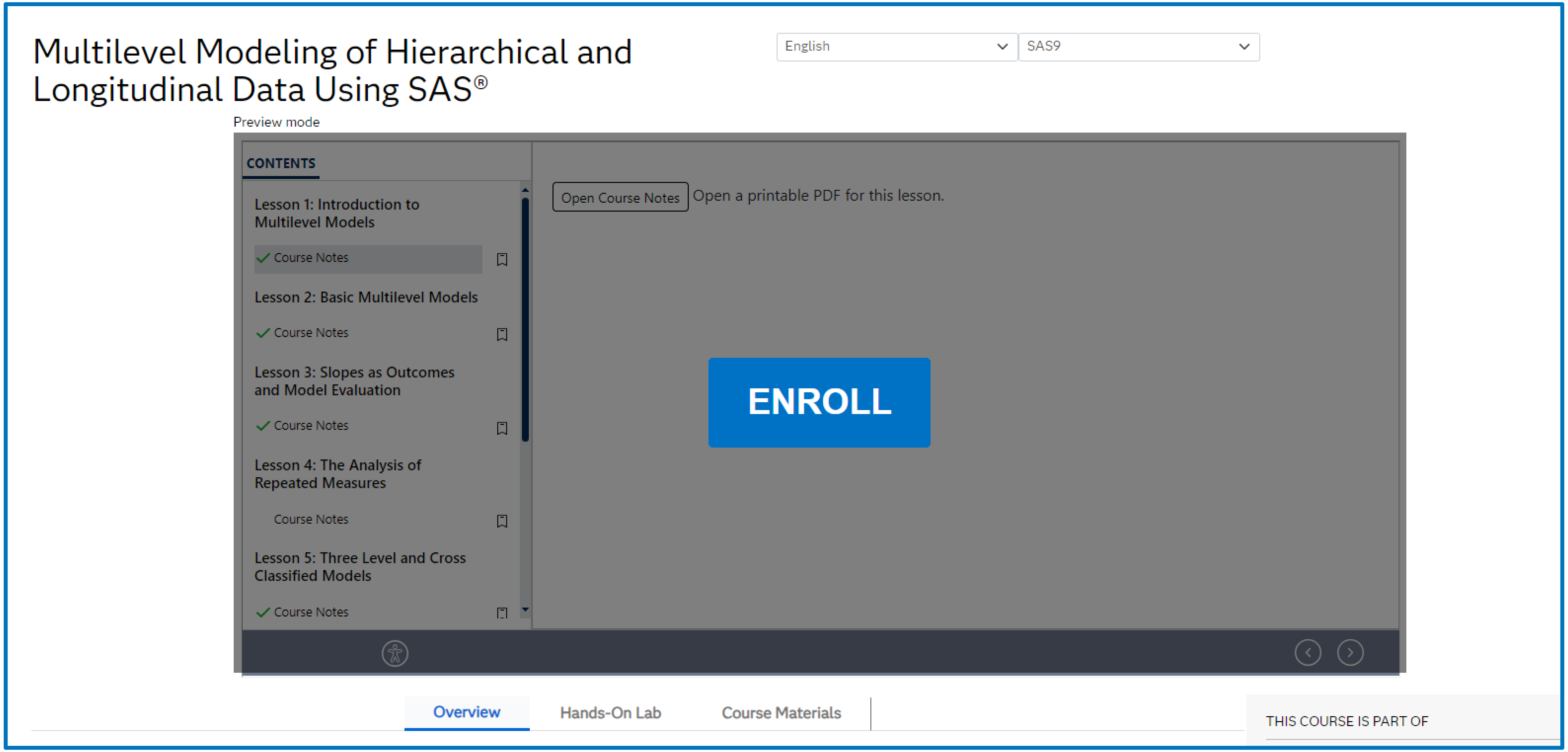
Specification in the MIXED Procedure

```
proc mixed data=mydata;
   class L3_ID L2_ID;
   model y = x / solution ddfm=kr2;
   random intercept x / subject=L2_ID(L3_ID)
                        type=un;
   random intercept x / subject=L3_ID
                        type=un;
run;
```

- All fixed effects are entered in the MODEL statement.
- Two RANDOM statements are used to designate the random intercepts and slopes at Levels 2 and 3.

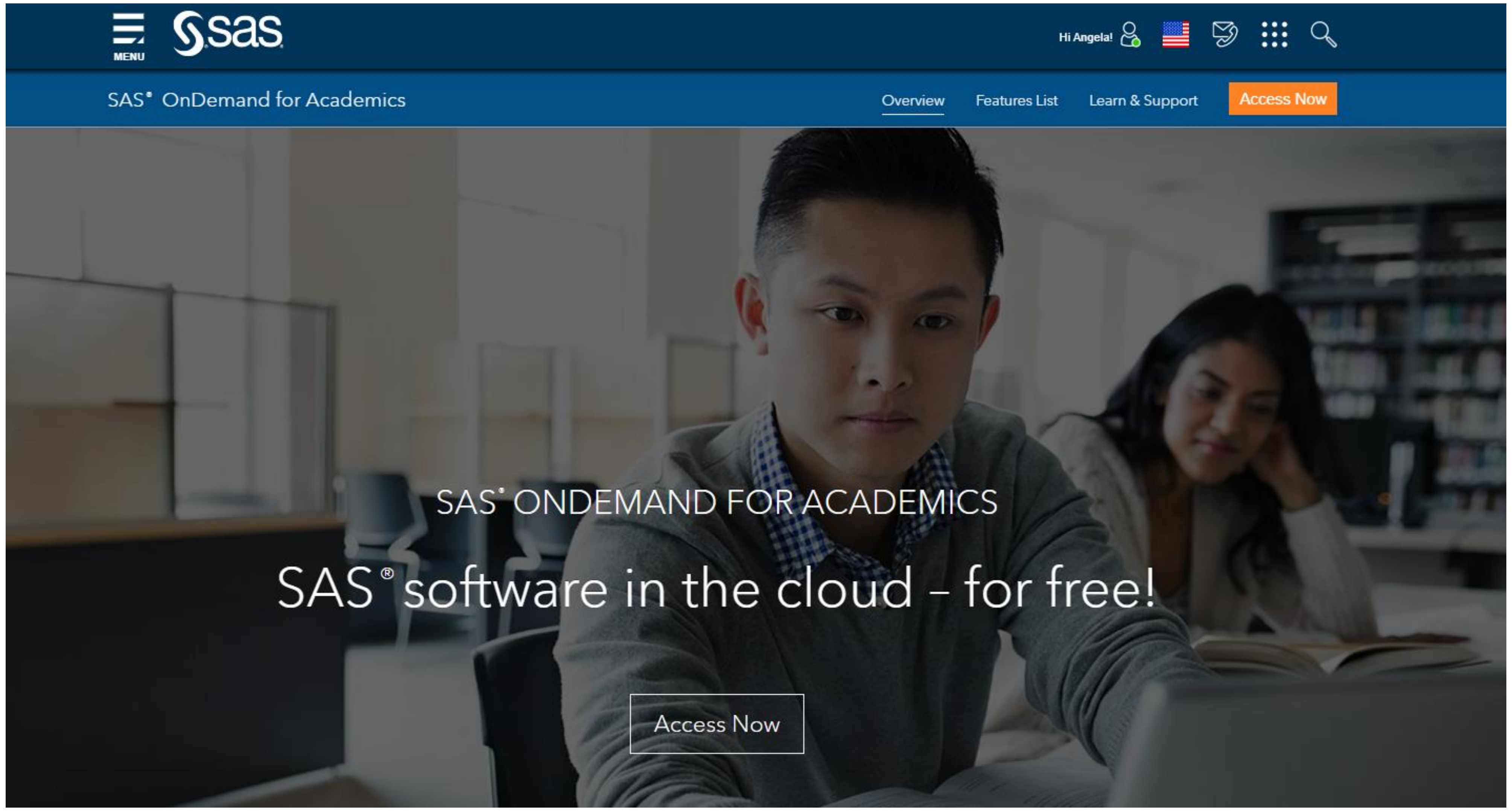
Continue Your Learning with a SAS Course

https://learn.sas.com/course/view.php?id=268 Free 7-day trial!



Continue Your Learning with SAS Programming

SAS OnDemand for Academics



- Free SAS software for students, educators, and independent learners.
- Register at: www.sas.com/ondemand
- Launch at: welcome.oda.sas.com

Questions?

Thank you!

Jacqueline.Johnson@sas.com