An introduction to (Generalized) Linear and Nonlinear Mixed Models

Dealing with Hierarchical, Nested and Temporal Dependencies in Data:

I. Introduction to random effects in statistical models

- Fixed vs. Random effects
- Random intercepts, random slopes and covariance matrix structures
- Cross-level interactions

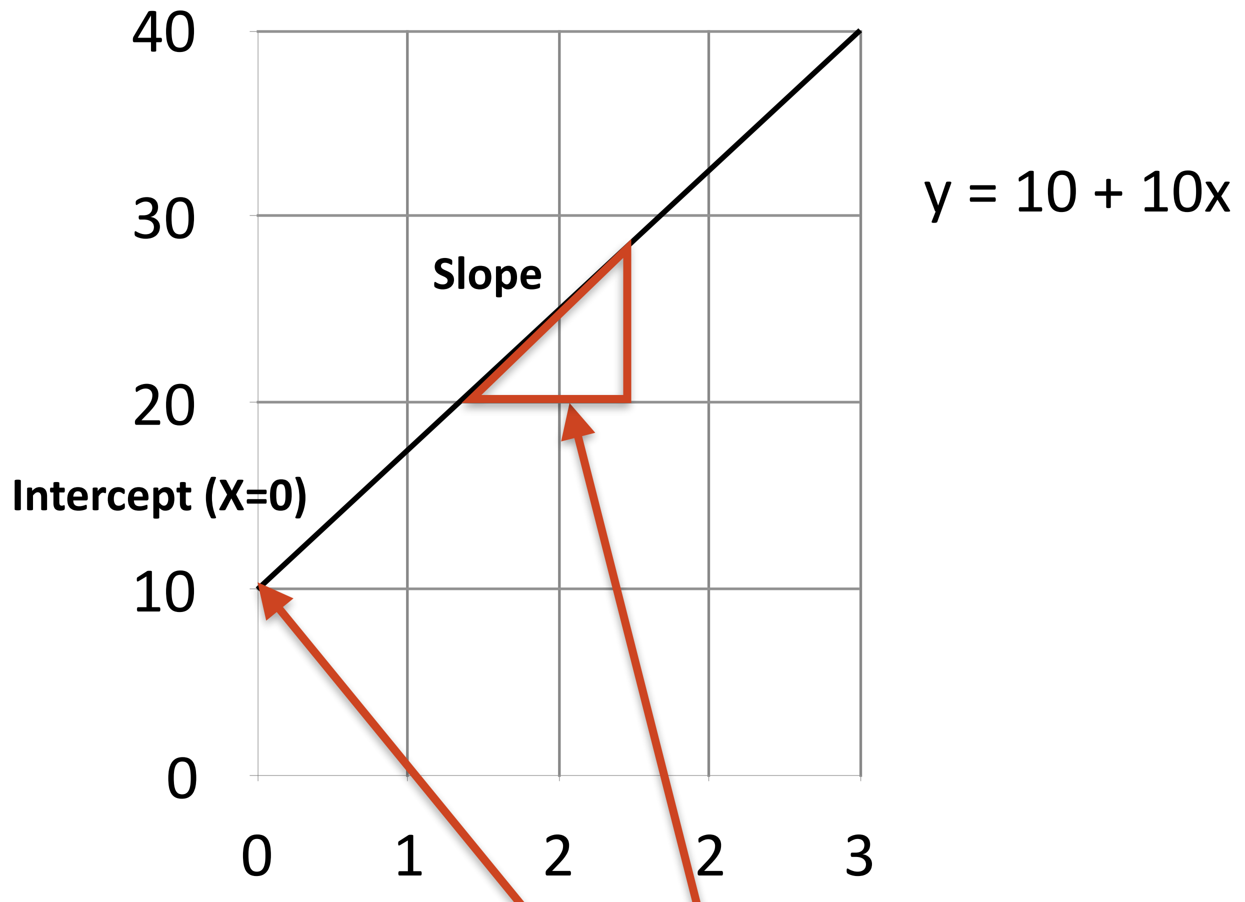
II. The multilevel model for change

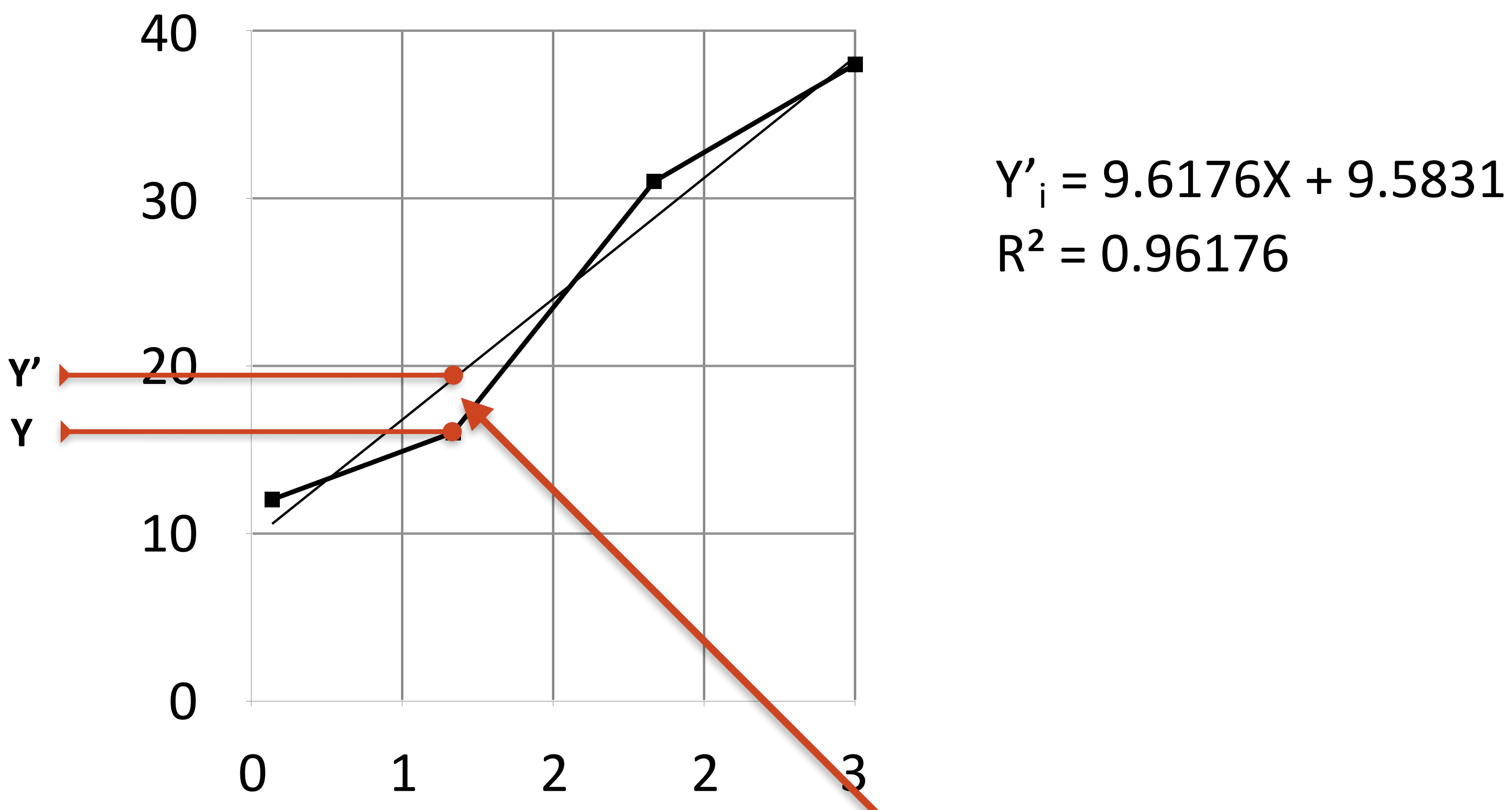
- Repeated measurements as a clustering level within individuals
- Growth curve models
- Models of piece-wise and nonlinear growth

III. Advanced models

- The generalized linear mixed model for binary outcomes and count data
- Cross-classified and multiple membership models
- Multivariate-multilevel models

Statistical models versus equations





$$y = \boxed{a} + \boxed{b}x$$

$$Y_i = \beta_0 + \beta_1 X_{1i} + \boxed{\varepsilon_i}$$

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

The line as a model: Multiple linear regression

- X_i as linear additive independent source(s) of variance in Y_i
  Linear prediction: Y' >> Y_i from X_i

- Variance *not* 'explained' by X_i is captured by an error term ε_i
  Residual variance: Y' – Y_i

- Model parameters are estimated using a 'least-squares' method:
  Minimise residual variance: smallest squared differences between observed and predicted scores.

- Remember Ordinary Least Squares (OLS) assumptions?

OLS regression: Assumptions for being BLUE

Hypothesised model: $$Y_i = \sum_{j=1}^{k} \beta_j X_{ij} + \varepsilon_i \qquad for \quad i = 1, \ldots, n$$

β_j : unobserved, non-random parameter

X_{ij} : observed, (non-) random variable

Y_i : observed, random variable

ε_i : unobserved random statistical error ($E(Y_i)$ - Y_i , or: 'true' - observation, population - sample)

Weak set of assumptions (Gauss-Markov Theorem):

1. $E(Y_i) = \sum_{j=1}^{k} \beta_j X_{ij} \qquad E(\varepsilon_i) = 0 \qquad for \quad i = 1, \ldots, n$

   The model represents a TRUE linear relationship so the expected value of the statistical errors is 0.

2. $V(\varepsilon_i) = \sigma^2 < \infty \qquad for \quad i = 1, \ldots, n$

   Assumption 1 + 2 = **homoskedastic error variance** (also for variance of Y_i)

3. **No dependencies (correlations) among errors** (also no correlations between Y_i and ε_i)

Strong assumptions (Gauss-Markov Theorem + Central Limit Theorem):

1 + 2 + 3 + Y_i is normally distributed

Assumptions for simple OLS regression

Single level model

	1	2	3	4	5	6	7	8	9	10	11	12	13	⋯
1	1	0	0	0	0	0	0	0	0	0	0	0	0	⋯
2	0	1	0	0	0	0	0	0	0	0	0	0	0	⋯
3	0	0	1	0	0	0	0	0	0	0	0	0	0	⋯
4	0	0	0	1	0	0	0	0	0	0	0	0	0	⋯
5	0	0	0	0	1	0	0	0	0	0	0	0	0	⋯
6	0	0	0	0	0	1	0	0	0	0	0	0	0	⋯
7	0	0	0	0	0	0	1	0	0	0	0	0	0	⋯
8	0	0	0	0	0	0	0	1	0	0	0	0	0	⋯
9	0	0	0	0	0	0	0	0	1	0	0	0	0	⋯
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋱

Correlation matrix V

Example: Sources of variability

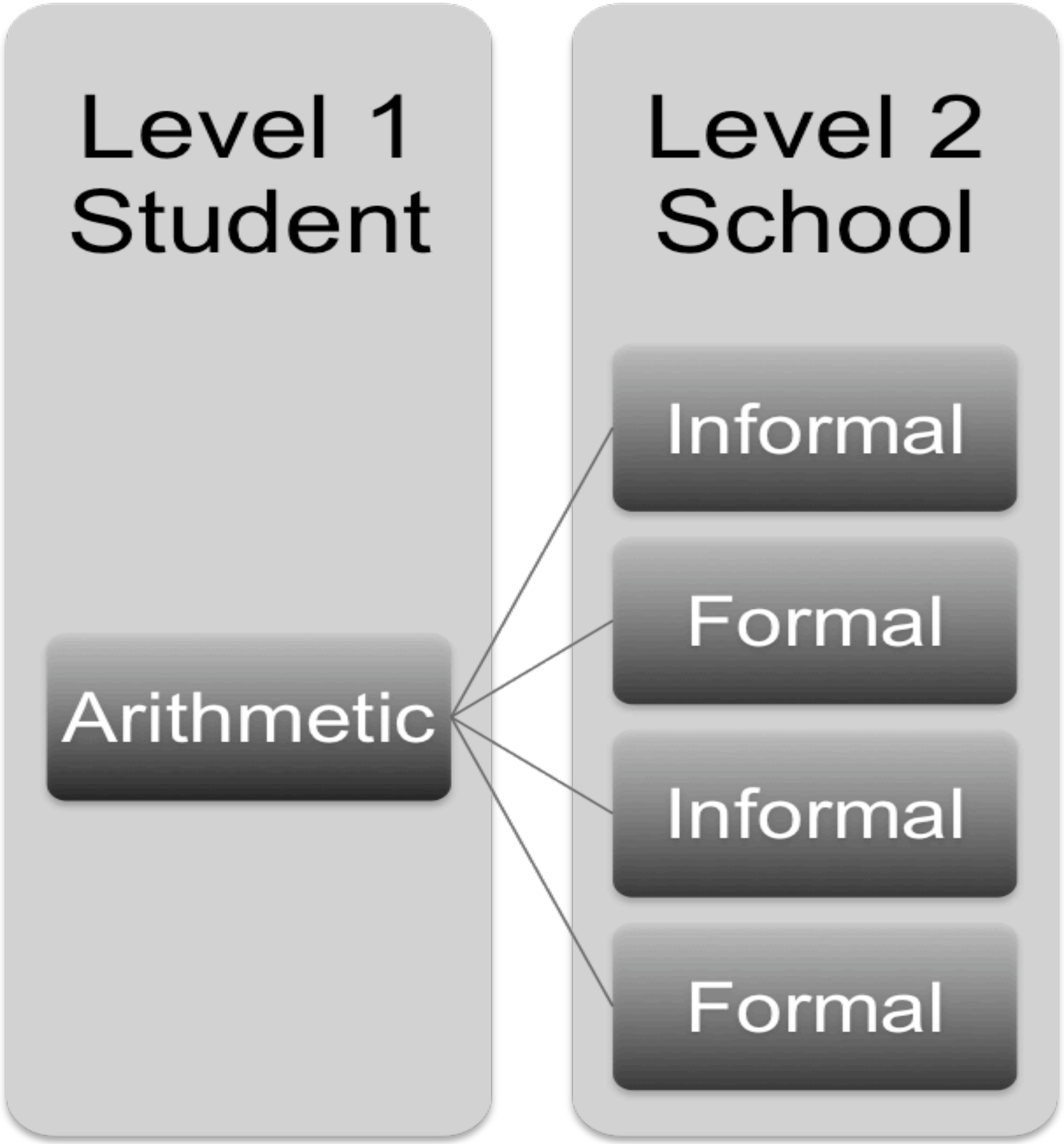
- Consider the following dataset:

  - 20 students from 4 different schools
  - X = arithmetic test at the beginning of the school year
  - Y = arithmetic test at the end of the school year
  - 2 schools have a formal teaching style
  - 2 schools have an informal teaching style

- Suppose we want to predict Y from X, what sources of variance are there?

- Students - Arithmetic scores
  Schools - Teaching style

Example: Sources of variability



Data file

	student	school	x	y	Formal Teaching Style
1	1	1	3	10	0
2	2	1	3	10	0
3	3	1	4	14	0
4	4	1	5	16	0
5	5	1	6	17	0
6	6	2	2	13	1
7	6	2	3	17	1
8	8	2	4	16	1
9	9	2	4	19	1
10	10	2	5	21	1
11	11	3	4	22	0
12	12	3	5	26	0
13	13	3	6	24	0
14	14	3	7	30	0
15	15	3	8	33	0
16	16	4	5	27	1
17	17	4	6	33	1
18	18	4	7	35	1
19	19	4	8	38	1
20	20	4	9	36	1

Example from Voeten & Van den Bercken, Lineaire regressieanalyse. Groningen: Wolters-Noordhoff, 2003, chapter 14.

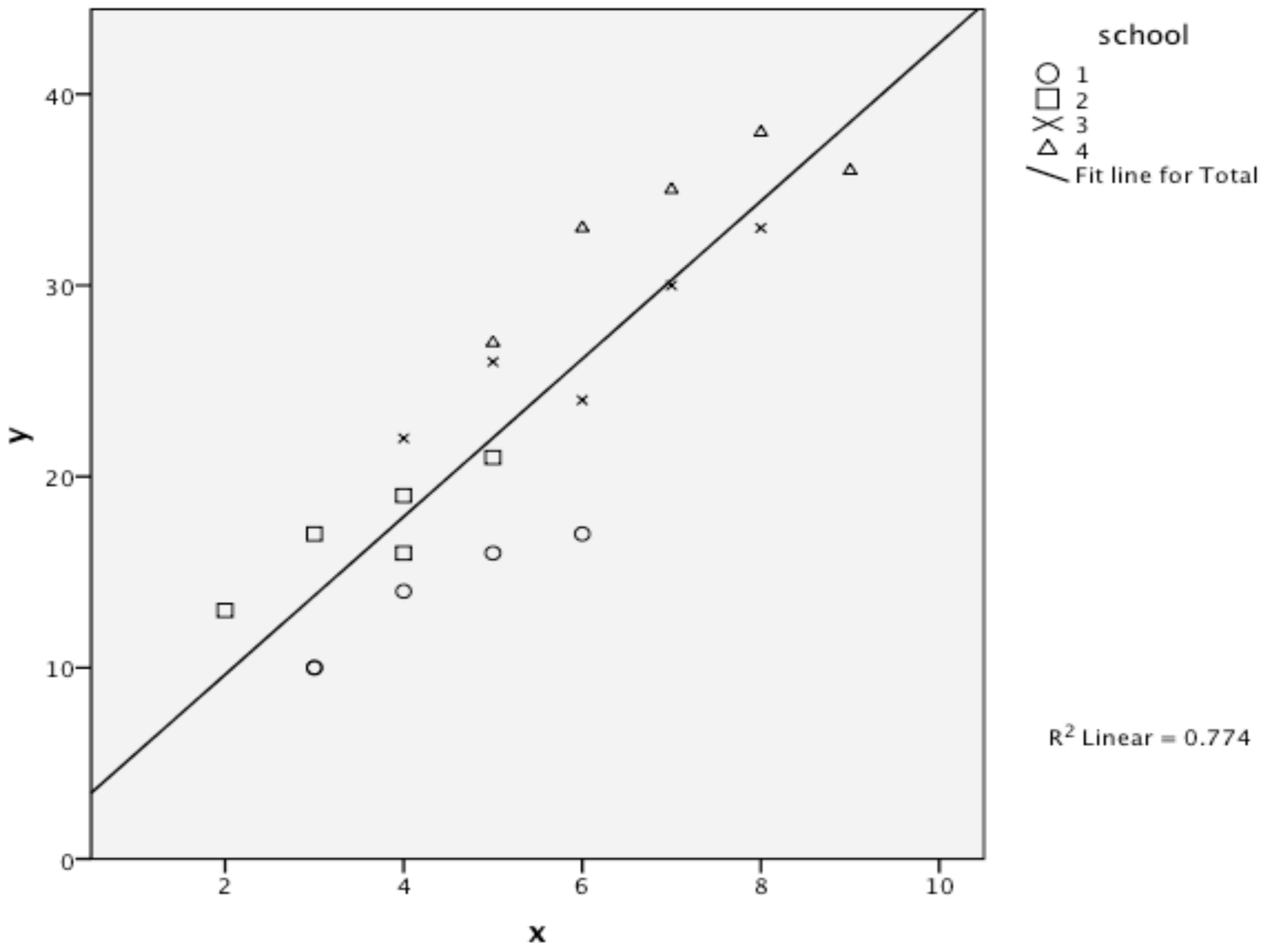
Example: Analysis using multiple regression

- These data clearly vary on multiple levels

- To understand how multilevel analysis works we'll start by taking a classical multiple regression approach to analyse these data

- Start with the entire group, pretend there are no levels

Example: Analysis using multiple regression



Example: Analysis using multiple regression

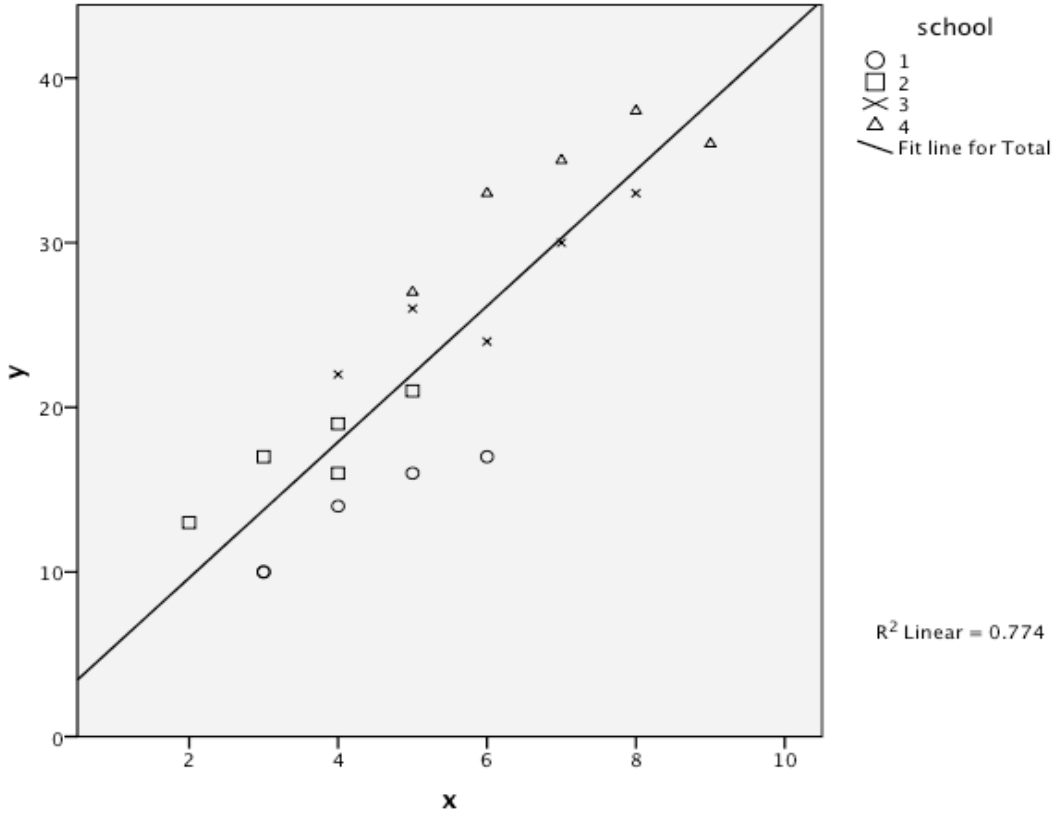
Regression analysis at the student level, ignoring school:

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

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	1.389	2.902		.479	.638
	x	4.127	.525	.880	7.855	.000

a. Dependent Variable: y



$$Y'_i = 1.389 + 4.127X$$

Example: Analysis using multiple regression

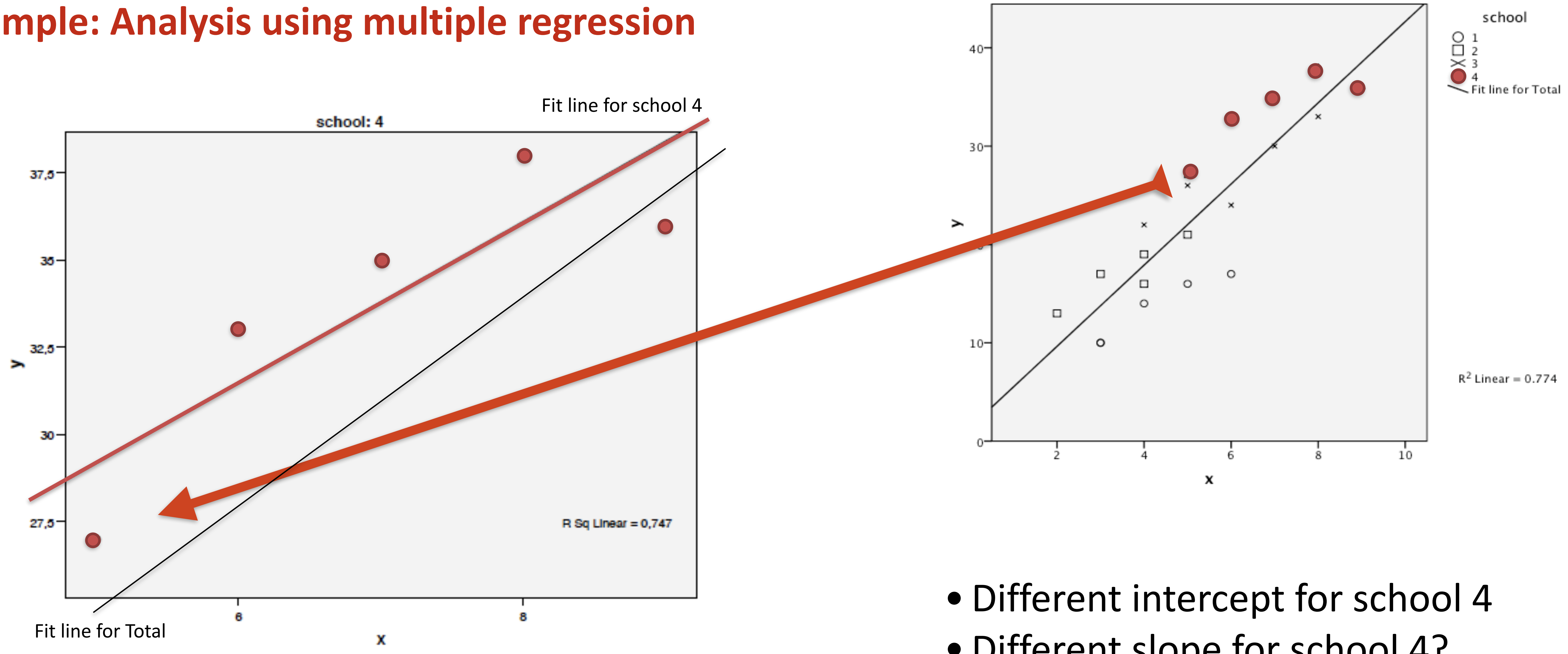
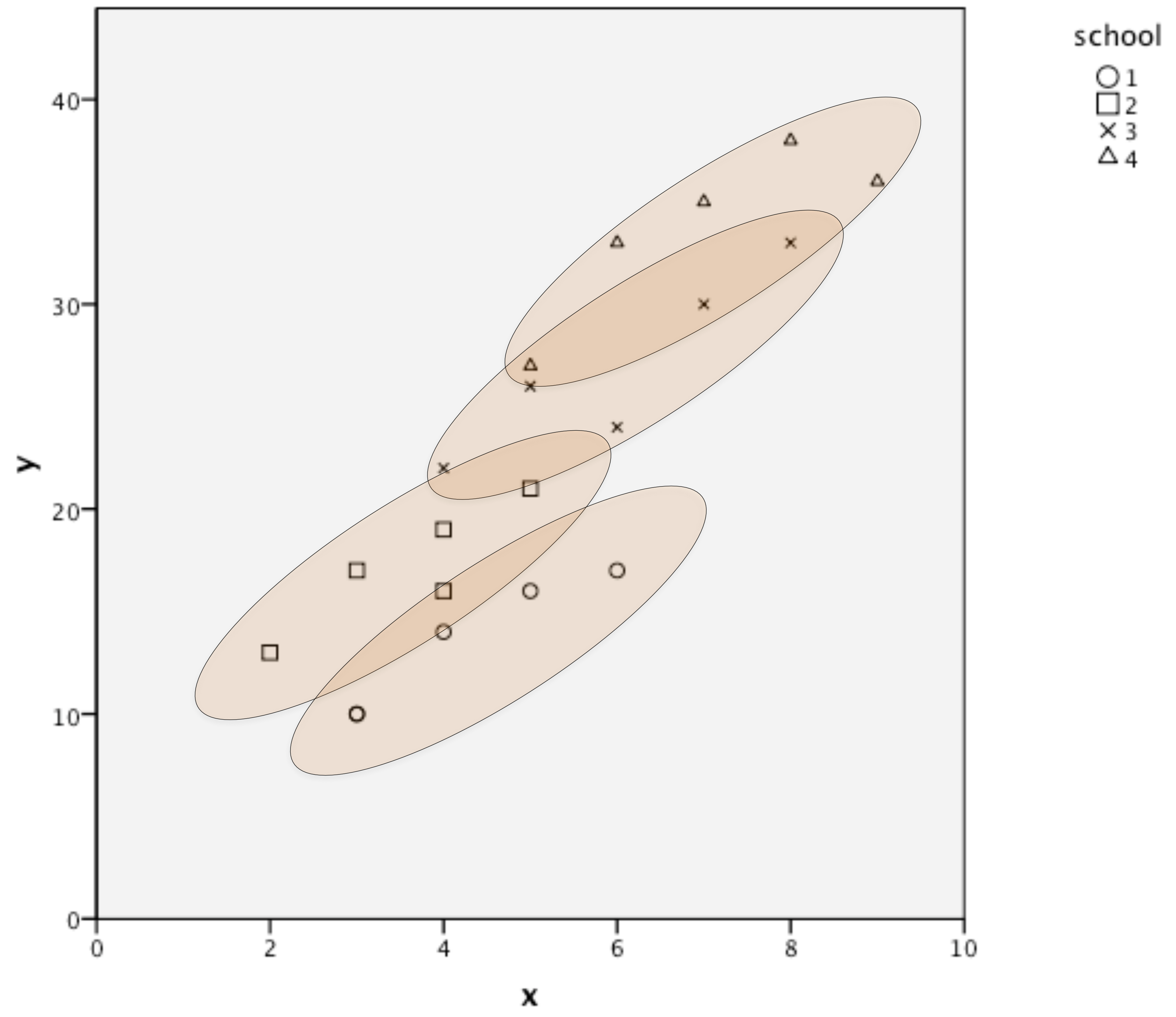


Figure 2: Variation at the student level within a particular school.

$$Y_{i4} = \beta_{04} + \beta_{14}X_{i4} + \varepsilon_{i4}$$

- Different intercept for school 4
- Different slope for school 4?
- Different residuals for pupils relative best fit for school 4

Example: Analysis using multiple regression



Example: Analysis using multiple regression

Regression analysis including a possibly different intercept for each school, but a common slope:

$$Y_{ij} = \beta_{0j} + \beta_1 X_{ij} + \varepsilon_{ij}$$

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.987[a]	.973	.966	1.647

a. Predictors: (Constant), school3, x, school1, school2

b. Dependent Variable: y

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	16.738	2.167		7.725	.000
	x	2.438	.291	.520	8.373	.000
	school1	-13.575	1.323	-.674	-10.265	.000
	school2	-8.313	1.437	-.413	-5.786	.000
	school3	-4.363	1.081	-.217	-4.034	.001

a. Dependent Variable: y

Variables in dataset:

Shool1 = 1, rest 0

Shool2 = 1, rest 0

Shool3 = 1, rest 0

Consequence:

All relative to school 4

5 parameters to be estimated

Example: Analysis using multiple regression

The analysis includes three dummy variables to represent the four schools; school 4 used as the reference category.

The regression equations for separate schools are:

School 4: Y' = 16.738 + 2.438X

School 1: Y' = (16.378 – 13.575) + 2.438X = 3.163 + 2.438X

School 2: Y' = (16.378 – 8.313) + 2.438X = 8.425 + 2.438X

School 3: Y' = (16.378 – 4.363) + 2.438X = 12.375 + 2.438X.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	16.738	2.167		7.725	.000
	x	2.438	.291	.520	8.373	.000
	school1	-13.575	1.323	-.674	-10.265	.000
	school2	-8.313	1.437	-.413	-5.786	.000
	school3	-4.363	1.081	-.217	-4.034	.001

a. Dependent Variable: y

Note:

1. The slope of X is now 2.438, but when ignoring school the slope is 4.127.
2. Rsquare increased from .774 to .973. School differences do matter.

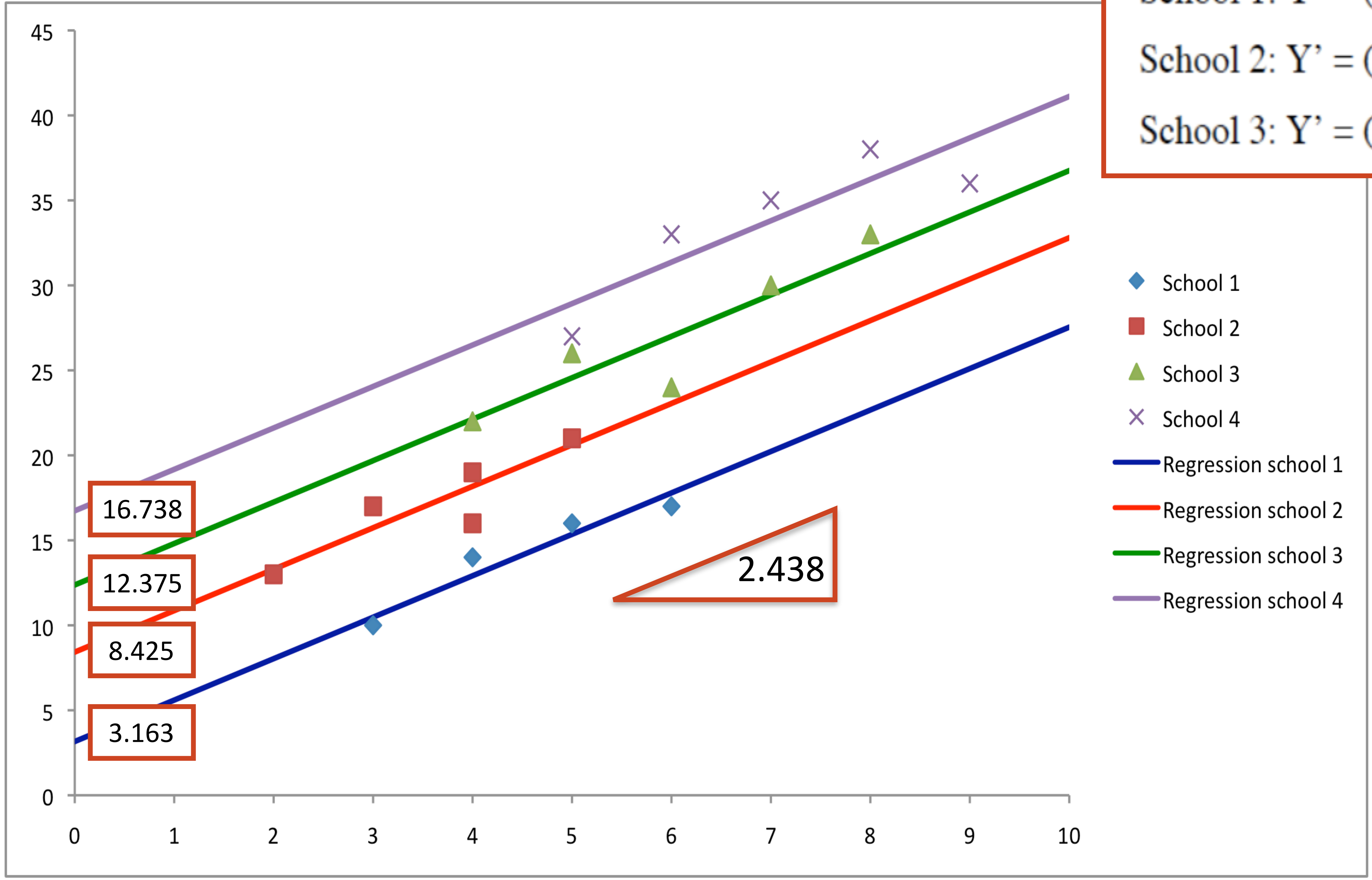
Variations between schools (intercept)

School 4: Y' = 16.738 + 2.438X

School 1: Y' = (16.378 – 13.575) + 2.438X = 3.163 + 2.438X

School 2: Y' = (16.378 – 8.313) + 2.438X = 8.425 + 2.438X

School 3: Y' = (16.378 – 4.363) + 2.438X = 12.375 + 2.438X.



5 parameters:

4 intercepts
1 slope

Schools also vary in teaching style

- Disregard previously found school differences.

- Same procedure:
  - 2 levels, so 1 dummy variable: Formal style = 1, rest 0.
  - Results relative to Informal style

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	20.606	1.198		17.208	.000
	Formal style	4.488	1.695	.257	2.648	.017
	x centered	4.062	.456	.866	8.917	.000

a. Dependent Variable: y

Schools with informal style: Y' = 20.606 + 4.062X

Schools with formal style: Y' = (20.606 + 4.488) + 4.062X.

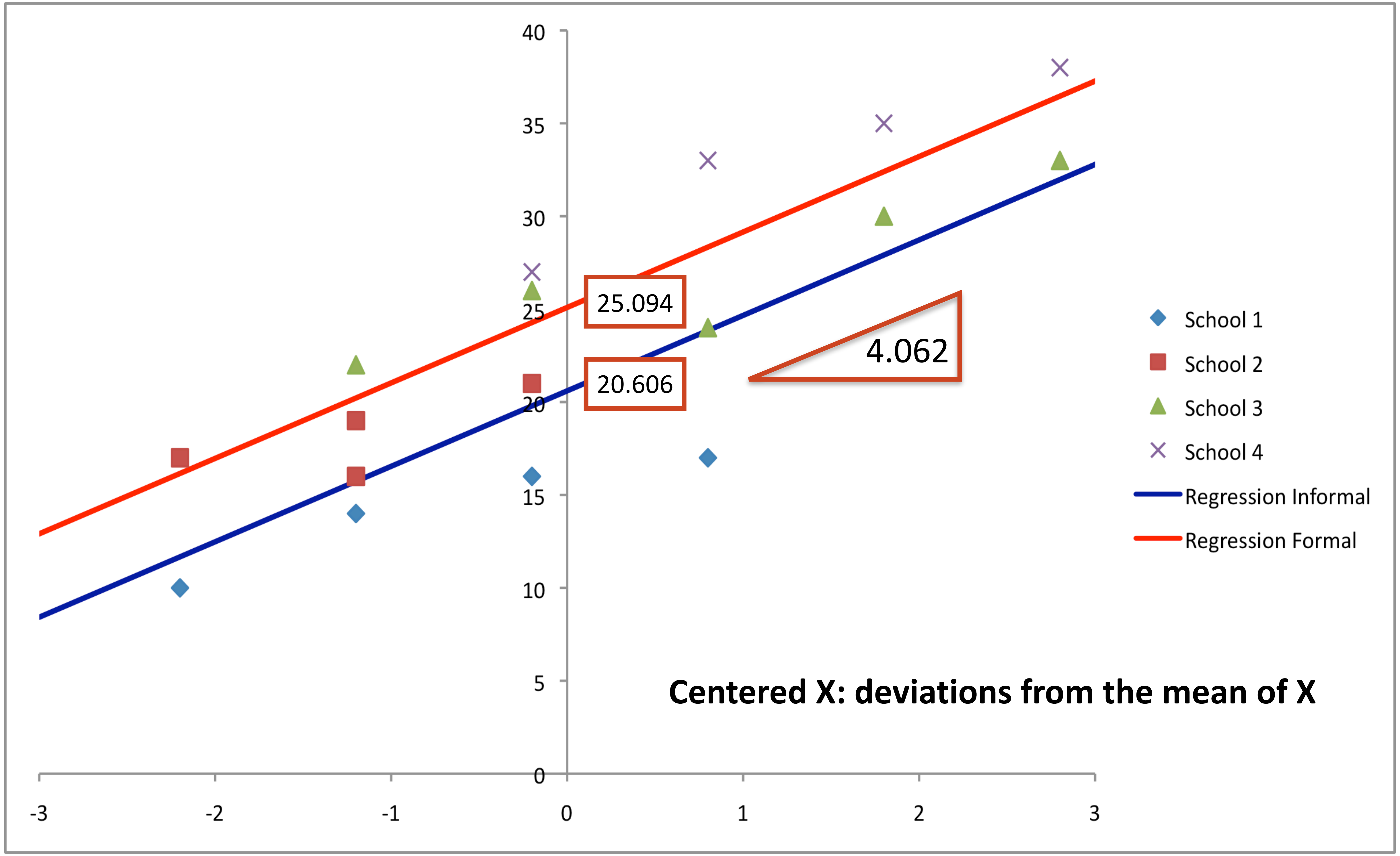
3 parameters to be estimated

$R^2_{general} = .77$

$R^2_{schools} = .93$

$R^2_{style} = .84$

Schools also vary in teaching style



Teaching style and school differences in 1 model

- Add two dummy variables:
  - D1 = difference between informal schools 1 (-1) and 3 (1), rest 0
  - D2 = difference between formal schools 2 (-1) and 4 (1), rest 0

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	20.444	.522		39.198	.000
	x centered	2.438	.291	.520	8.373	.000
	Formal style	4.813	.739	.276	6.515	.000
	d1	4.606	.583	.373	7.902	.000
	d2	4.156	.718	.337	5.786	.000

a. Dependent Variable: y

5 parameters

$R^2_{general} = .77$

$R^2_{schools} = .93$

$R^2_{style} = .84$

$R^2_{both} = .97$

Interpretation is not straightforward (rearrange equations)

Teaching style and school differences: Fixed vs. Random effects

Compare the multiple regression results with the results of a multilevel analysis (fixed parameter estimates only):

	b	*SE(b)*	*t*	*p*
Intercept	20.45	3.02	8.57	.000
X	2.53	0.28	9.11	.000
Formal style	4.79	4.27	1.12	.378

Coefficient

	Unstandardized Coefficients	
	B	Std. Error
(Constant)	20.444	.522
x centered	2.438	.291
Formal style	4.813	.739
d1	4.606	.583
d2	4.156	.718

The parameter estimates don't differ very much; the difference is mainly in the standard errors.

Both analyses take school differences into account. The multiple regression with dummy variables to represent schools treats 'school' as a fixed effect; the multilevel analysis treats 'school' as a random effect.

Random intercepts, random slopes and covariance matrix structures

Multilevel model as a regression model

Random intercepts

Random Slopes

Intercept-Slope covariance

Why multilevel models?

- Take clustered structure of data into account… what does that mean?

- Assumption of independent observations is not satisfied… what does that mean?

- SE are underestimated... DANGER: infer a relationship exists when none is present… (check SE formula)

- In standard regression: only individual level random variability

- <u>Solution</u>: Treat variability at higher levels as random variability

- <u>Assumption</u>: Level 2 variable is a random selection of groups/clusters out of a population who are normallly distributed around the population mean.

Why multilevel models

- Dependencies in the data: $$\frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_e^2}$$

- Similarity between individuals in the same group: **Intra class correlation (ICC)** / Proportion of <u>residual</u> variation due to differences between groups: **variance partition coefficient (VPC)**

- Between 0 and 1: 0.3 is large!

The multilevel model as a regression model

- The multiple regression model includes only 1 random source of variability. Individual differences expressed as residuals around the regression line:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2\,\mathrm{Style}_i + \beta_3\,\mathrm{d1}_i + \beta_4\,\mathrm{d2}_i + \varepsilon_i$$

- In a multilevel model we can include random sources at two or more levels:

$$Y_{ij} = \beta_{0j} + \beta_1 X_{ij} + \beta_2\,\mathrm{Style}_j + \boldsymbol{\varepsilon}_{\mathbf{ij}}$$ **-> Residuals around regression line within a school**

$$\beta_{0j} = \beta_{00} + \mathbf{u_j}$$ **-> Variation in school intercepts**

Or: $$Y_{ij} = \beta_{00} + \beta_1 X_{ij} + \beta_2\,\mathrm{Style}_j + (\boldsymbol{\varepsilon}_{\mathbf{ij}} + \mathbf{u_j})$$

In words: variance to be explained

- All multilevel modelling starts by defining an "empty model" in which you define the levels you think are present in the data.

- In the case of our school example:

  The empty model divides the total variance in arithmetic performance into two components:

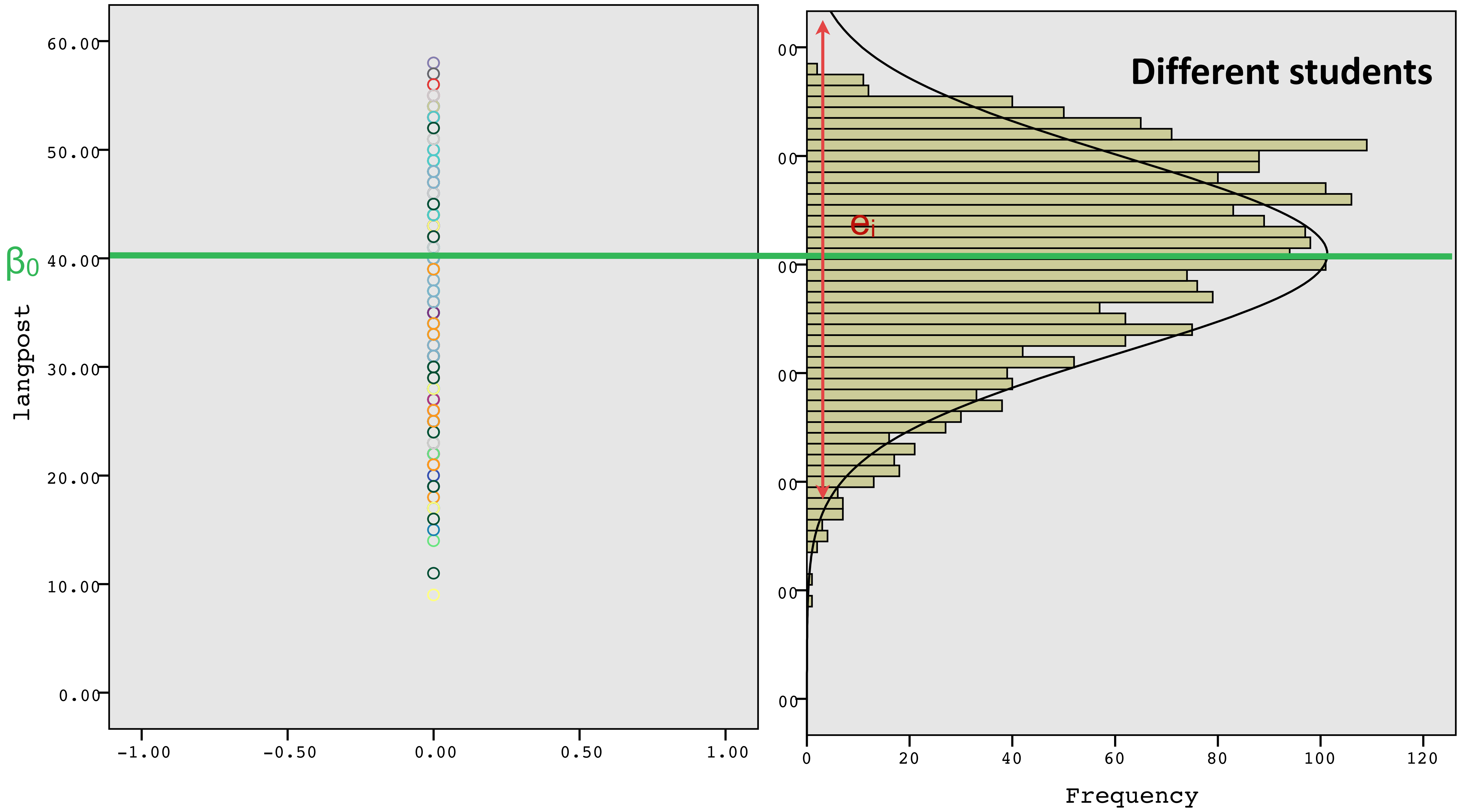
  Level 2: variance of school means (around the grand mean)
  Level 1: variance of individual scores within a school

- This is random variance which needs to be explained, by explanatory variables. For instance, the variable teaching style might be able to explain variance between school means (and hence individuals)

Or: $$Y_{ij} = \beta_{00} + \beta_1 X_{ij} + \beta_2 \, Style_j + (\boldsymbol{\varepsilon_{ij}} + \mathbf{u_j})$$

The single level model, no covariate:

$$Y_i = \beta_{0i} X_0 + e_{0i}$$



The single level model, no covariate:

$$Y_i = \beta_{0i}X_0 + e_{0i}$$

Single level model

	1	2	3	4	5	6	7	8	9	10	11	12	13	⋯
1	1	0	0	0	0	0	0	0	0	0	0	0	0	⋯
2	0	1	0	0	0	0	0	0	0	0	0	0	0	⋯
3	0	0	1	0	0	0	0	0	0	0	0	0	0	⋯
4	0	0	0	1	0	0	0	0	0	0	0	0	0	⋯
5	0	0	0	0	1	0	0	0	0	0	0	0	0	⋯
6	0	0	0	0	0	1	0	0	0	0	0	0	0	⋯
7	0	0	0	0	0	0	1	0	0	0	0	0	0	⋯
8	0	0	0	0	0	0	0	1	0	0	0	0	0	⋯
9	0	0	0	0	0	0	0	0	1	0	0	0	0	⋯
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋱

Correlation matrix V

The 'empty' model / Variance components

Equations *Note:* This is the **MLwiN** interface

$$y_{ij} \sim \mathrm{N}(XB, \Omega)$$

$$y_{ij} = \beta_{0ij} x_0$$

$$\beta_{0ij} = \beta_0 + u_{0j} + e_{0ij}$$

$$\begin{bmatrix} u_{0j} \end{bmatrix} \sim \mathrm{N}(0, \Omega_u) : \Omega_u = \begin{bmatrix} \sigma^2_{u0} \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim \mathrm{N}(0, \Omega_e) : \Omega_e = \begin{bmatrix} \sigma^2_{e0} \end{bmatrix}$$

Estimation of one common variance at each level: Distribution of level units around a mean fixed at 0

Only 3 parameters to be estimated:

β_0: Mean of dependent variable

σ^2_{u0}: Between-group differences

σ^2_{e0}: Within-group differences

The ‘empty’ model / Variance components

L2		1	1	1	1	2	2	3	3	3	3	…
	L1	1	2	3	4	1	2	1	2	3	4	…
1	1	1	ρ	ρ	ρ	0	0	0	0	0	0	…
1	2	ρ	1	ρ	ρ	0	0	0	0	0	0	…
1	3	ρ	ρ	1	ρ	0	0	0	0	0	0	…
1	4	ρ	ρ	ρ	1	0	0	0	0	0	0	…
2	1	0	0	0	0	1	ρ	0	0	0	0	…
2	2	0	0	0	0	ρ	1	0	0	0	0	…
3	1	0	0	0	0	0	0	1	ρ	ρ	ρ	…
3	2	0	0	0	0	0	0	ρ	1	ρ	ρ	…
3	3	0	0	0	0	0	0	ρ	ρ	1	ρ	…
3	4	0	0	0	0	0	0	ρ	ρ	ρ	1	…
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋱

Values belonging to the same level (school) can be correlated.

The correlation is the same for each level: ρ

Intra Class Correlation (ICC)
Variance Partition Coefficient (VPC)

This is a simple ratio of the variances estimated at each level

$$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$$

Correlation matrix V

Is a multilevel model necessary?

1. Variance Components / intercept only model; how much variance exists at the various levels? ICC / VPC. Is a multilevel model necessary?

Equations

$$\text{langpost}_{ij} = \beta_{0j} + e_{ij}$$

$$\beta_{0j} = 40.364(0.426) + u_{0j}$$

$$u_{0j} \sim N(0, \sigma^2_{u0}) \quad \sigma^2_{u0} = 19.419(2.921)$$

$$e_{ij} \sim N(0, \sigma^2_{e}) \quad \sigma^2_{e} = 64.569(1.967)$$

-2*loglikelihood = 16253.219(2287 of 2287 cases in use)

Equations

$$\text{langpost}_{ij} = 40.935(0.188) + e_{ij}$$

$$e_{ij} \sim N(0, \sigma^2_{e}) \quad \sigma^2_{e} = 81.031(2.396)$$

-2*loglikelihood = 16541.198(2287 of 2287 cases in use)

intercept

intercept random at

- [ ] j(schoolnr)

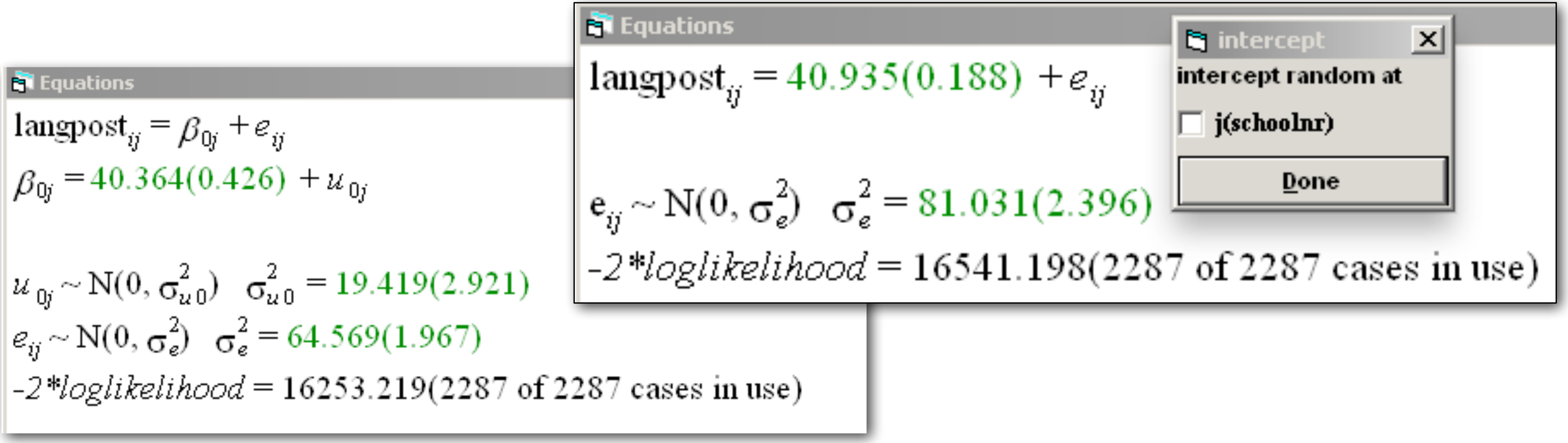
Done

$$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$$

ICC = VPC = 19.4 / (64.6 + 19.4) = 0.23

Is a multilevel model necessary?

1. Variance Components / intercept only model; how much variance exists at the various levels? ICC / VPC.



Deviance test for nested models: *1 level - 2 levels*

16541.198 - 16253.219 = $\chi^2(1)$= 287.97 [df = # of extra model parameters]

Variance components model:
Estimating random variance “to be explained”

• ‘Empty’ model: Analyze variance of a dependent variable into two variance components. Level 1: Individuals, Level 2: Groups

• Compare to ANOVA: Groups as a random factor

• Three parameters:

- Fixed: Grand mean
- Random Level 2: Variance of group means
- Random Level 1: Variance of individuals within groups

The ‘empty’ model: 1 mean + 2 random effects

$$Y_{ij} = \beta_{0ij}X_0 + (u_{0j} + e_{0ij})$$

X_0: A variable containing values of 1 for each case (constant)

Equations

$$y_{ij} \sim \mathrm{N}(XB, \Omega)$$

$$y_{ij} = \beta_{0ij}x_0$$

$$\beta_{0ij} = \beta_0 + u_{0j} + e_{0ij}$$

$$\begin{bmatrix} u_{0j} \end{bmatrix} \sim \mathrm{N}(0, \Omega_u) : \Omega_u = \begin{bmatrix} \sigma^2_{u0} \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim \mathrm{N}(0, \Omega_e) : \Omega_e = \begin{bmatrix} \sigma^2_{e0} \end{bmatrix}$$

β_0: Mean of dependent variable

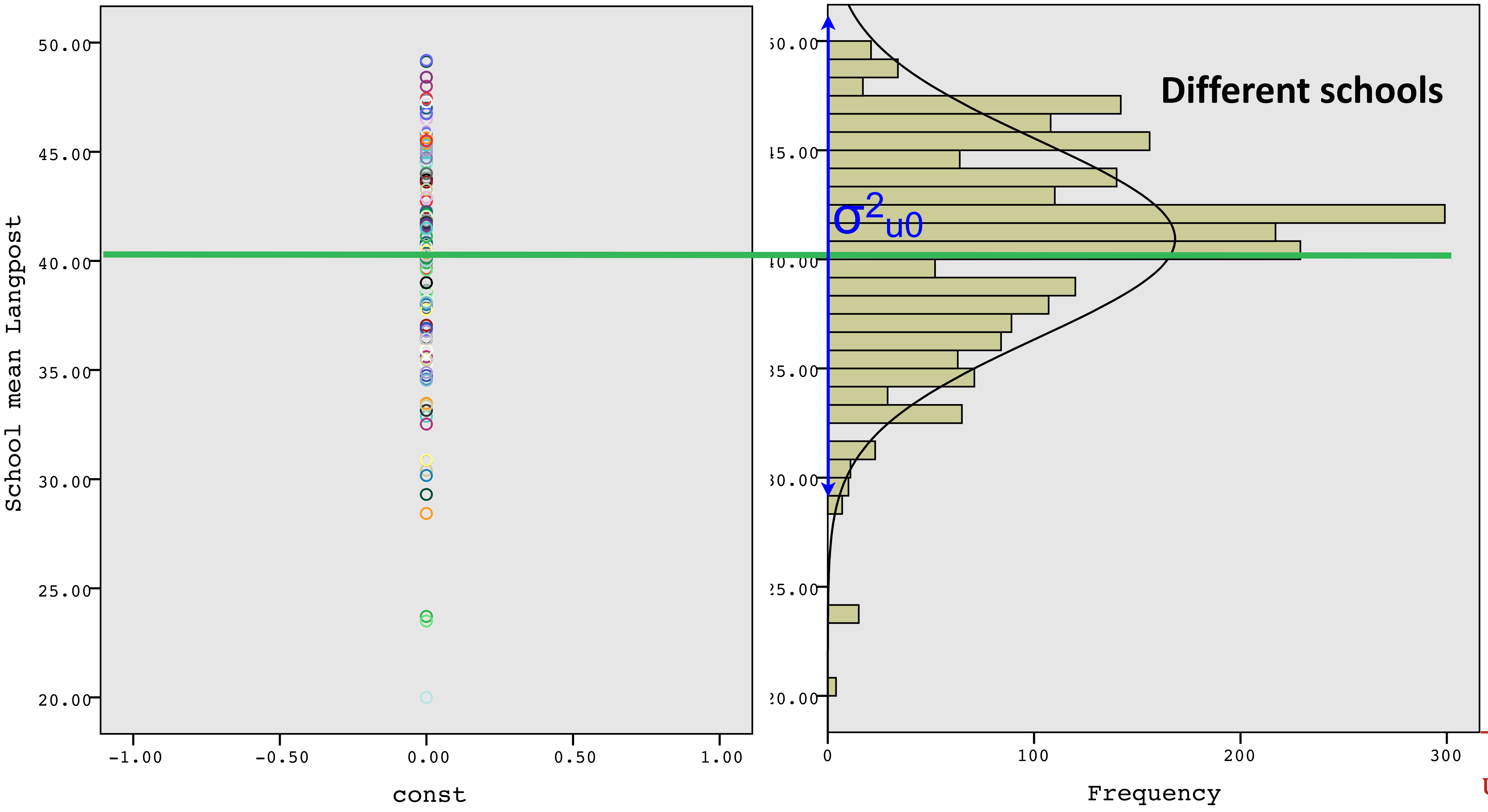
σ^2_{u0}: Between-group differences

σ^2_{e0}: Within-group differences

The Variance components at 2 levels: Between schools

$$Y_{ij} = \beta_{0ij}X_0 + (u_{0j} + e_{0ij})$$

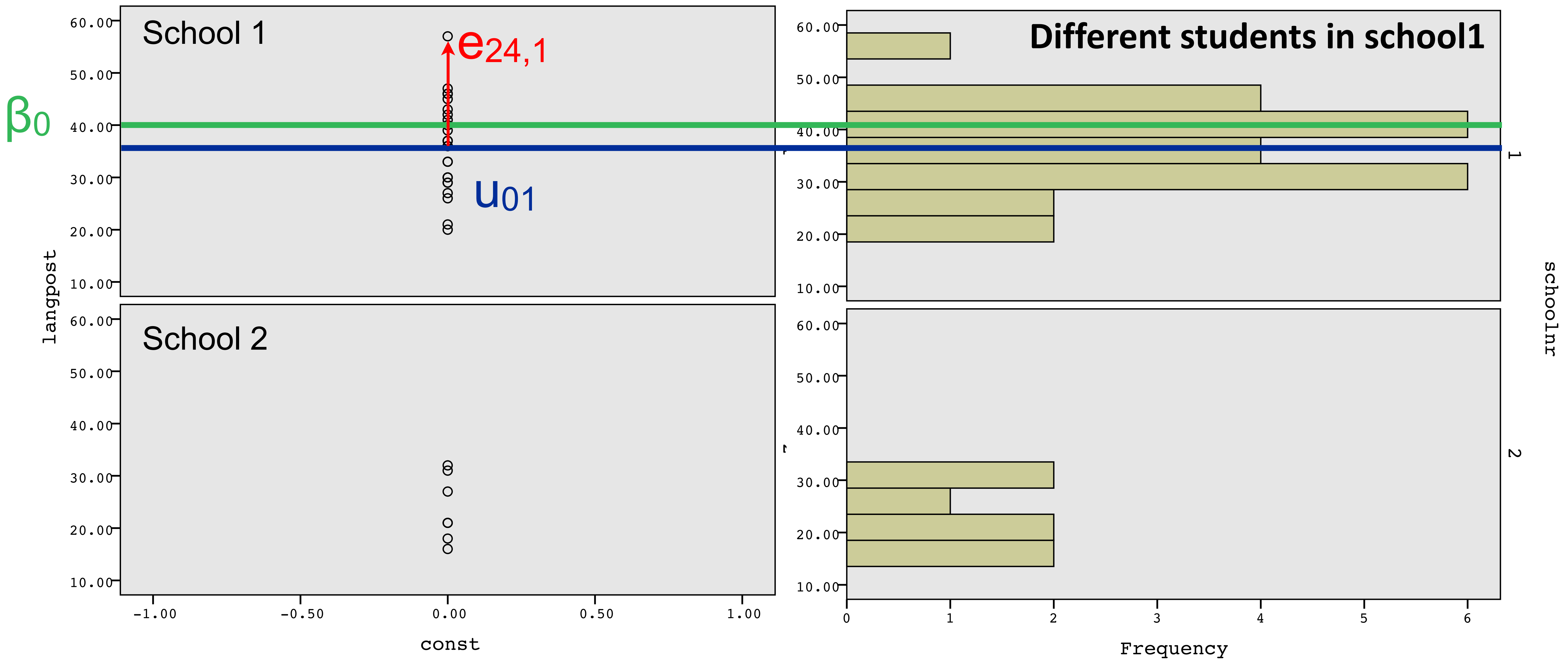
β_0



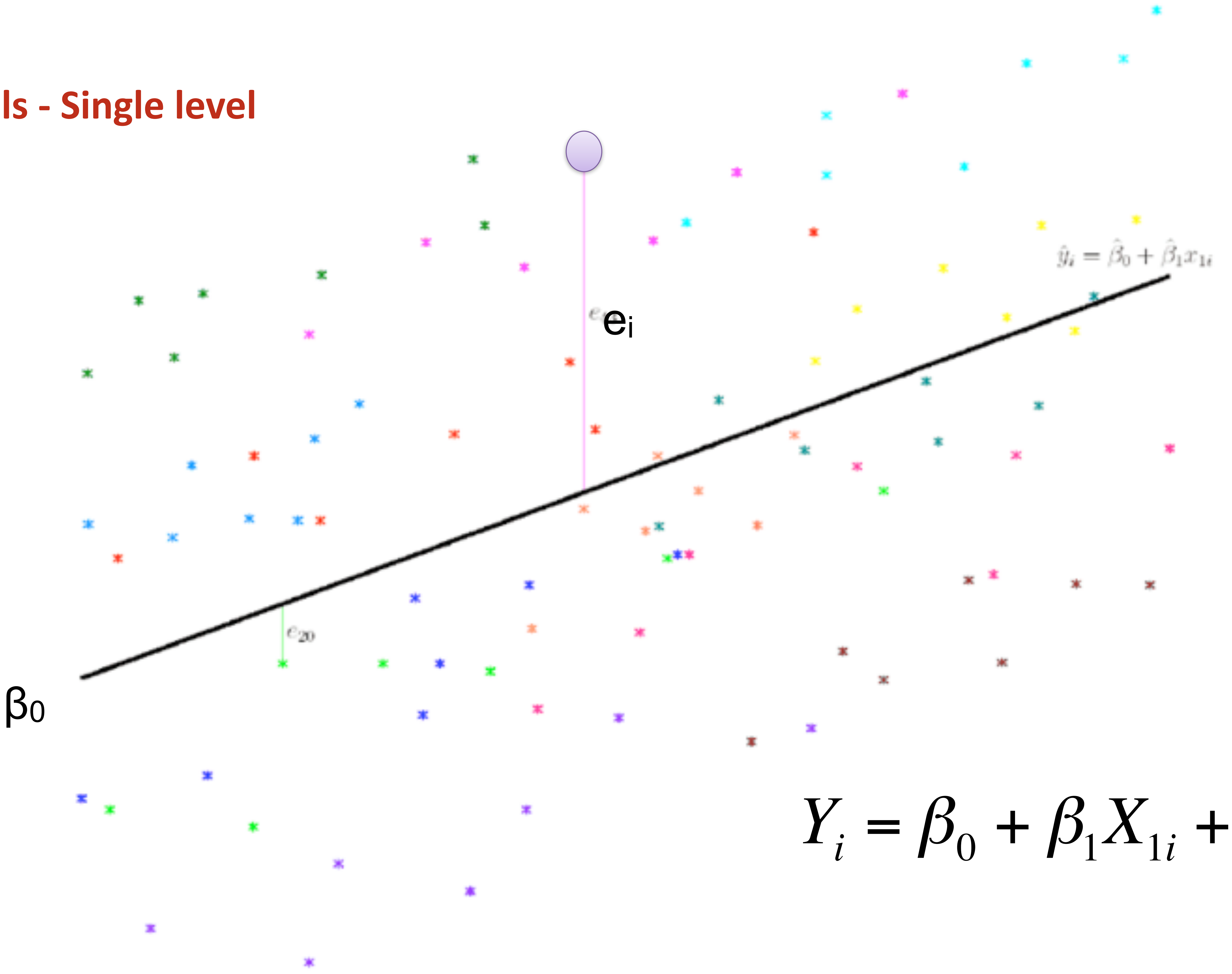
University

Variance components at 2 levels: Within schools

$$Y_{ij} = \beta_{0ij}X_0 + (u_{0j} + e_{0ij})$$

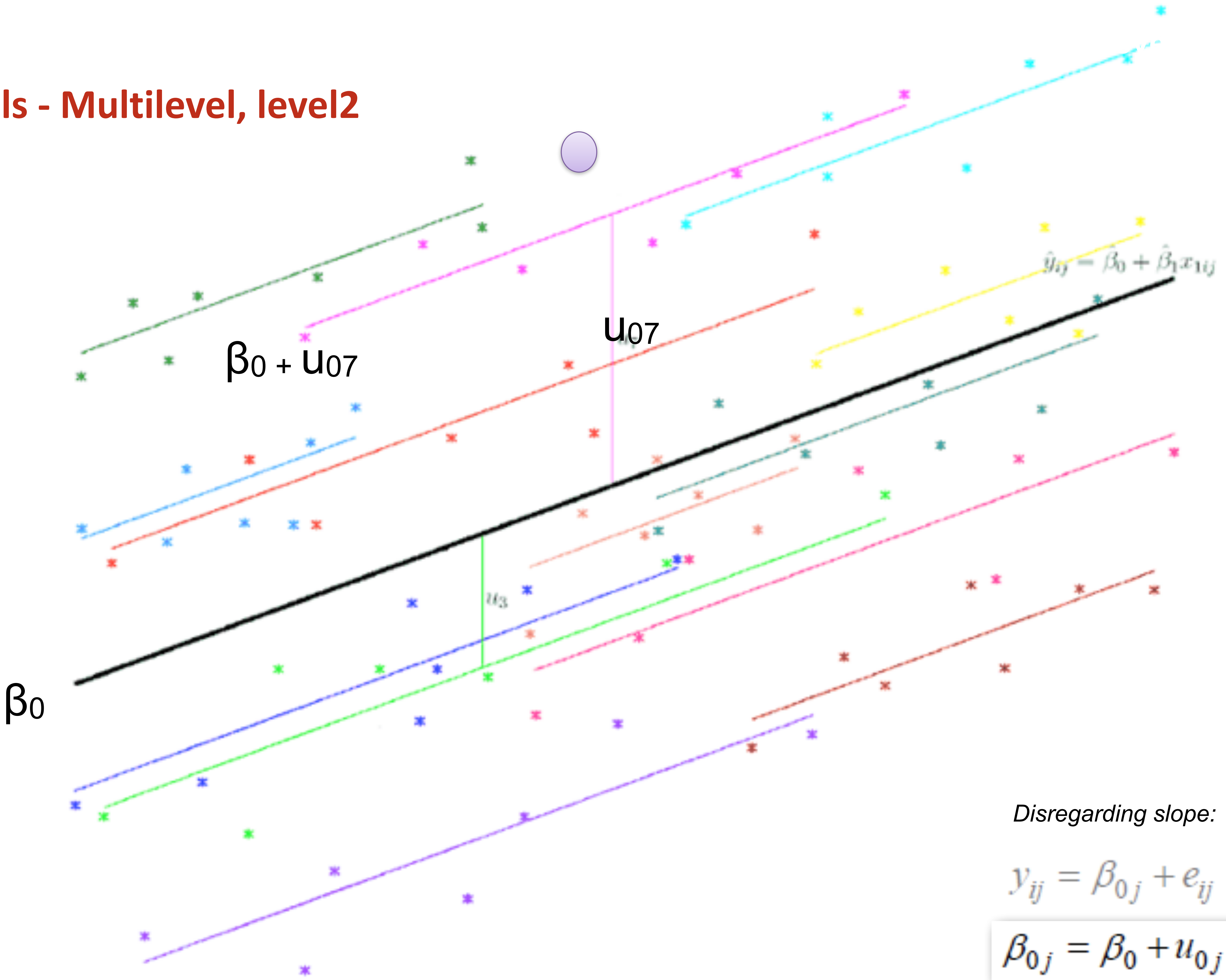


Residuals - Single level



$$Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i$$

Residuals - Multilevel, level2

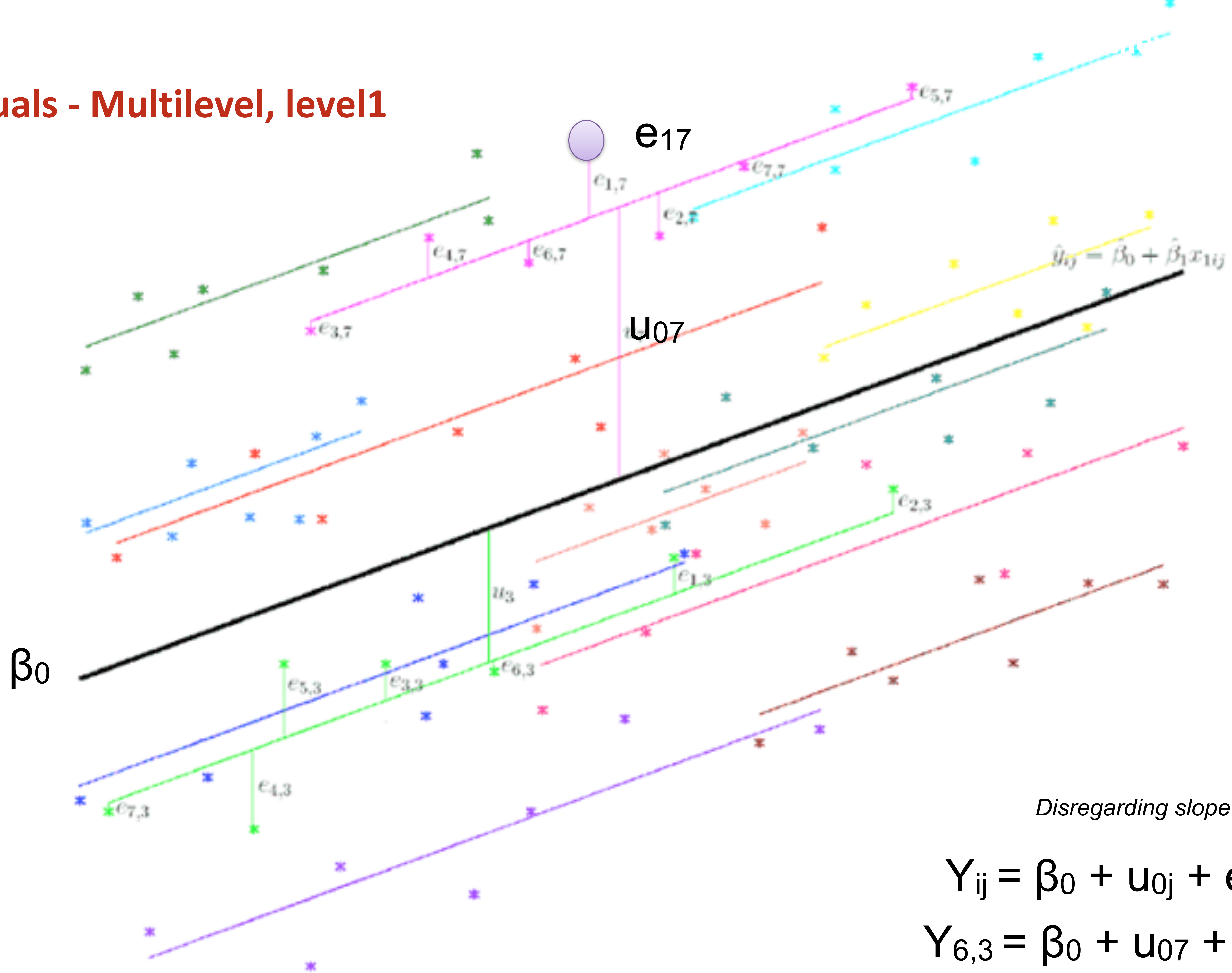


Disregarding slope:

$$y_{ij} = \beta_{0j} + e_{ij}$$

$$\boxed{\beta_{0j} = \beta_0 + u_{0j}}$$

Residuals - Multilevel, level1

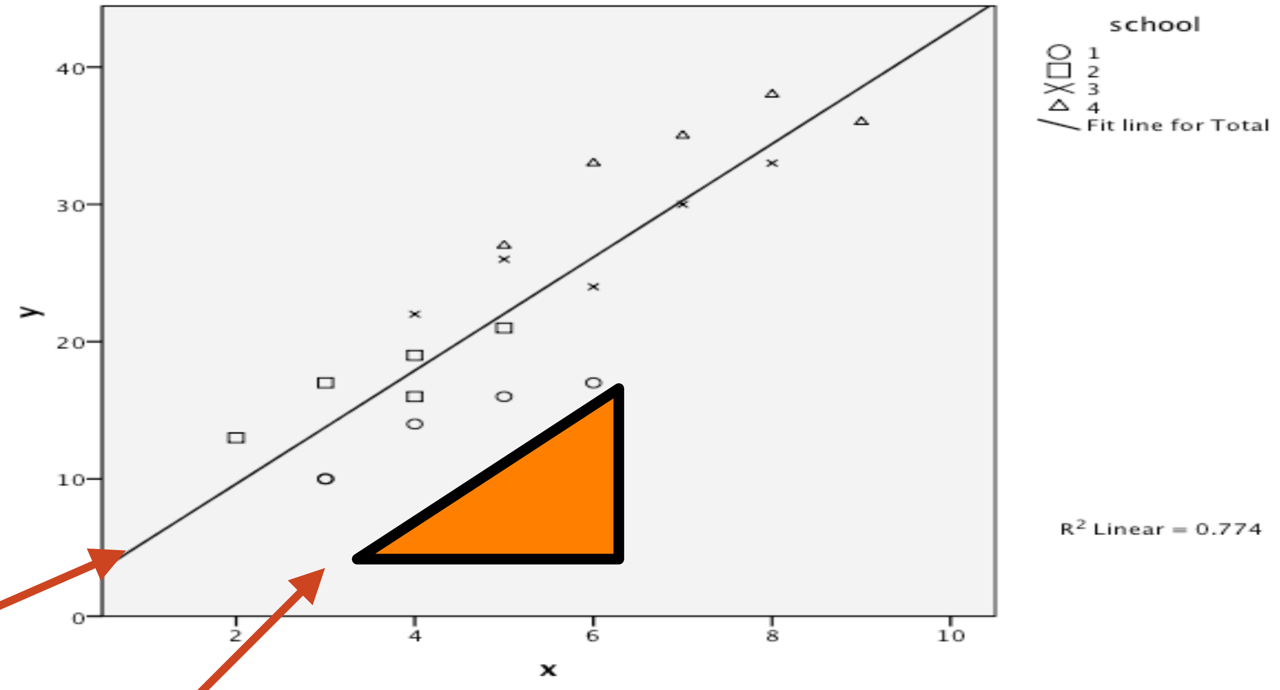


Disregarding slope:

$$Y_{ij} = \beta_0 + u_{0j} + e_{ij}$$

$$Y_{6,3} = \beta_0 + u_{07} + e_{17}$$

Random intercept model, 1 covariate, fixed slope

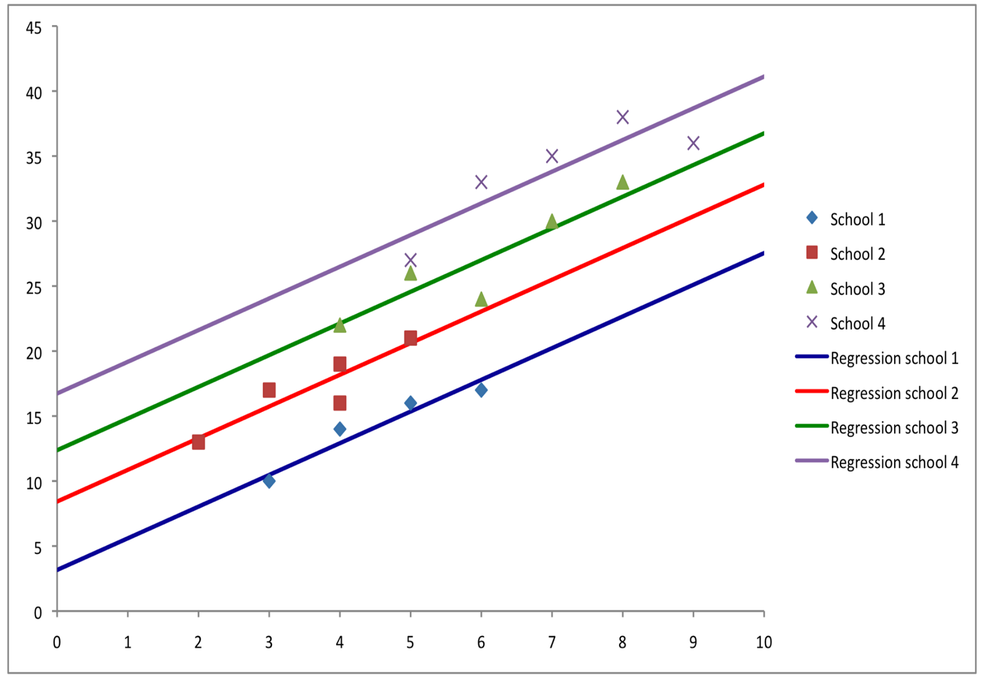


- Random intercepts (for each school) plus a covariate with a fixed slope

- Compare to ANCOVA: Groups still a random factor

- Four parameters:

- Fixed: Average intercept
- Fixed: Pooled within-group slope of covariate
- Random Level 2: Variance of intercepts
- Random Level 1: Residual variance within groups



Random intercept model, 1 covariate, fixed slope

$$Y_{ij} = \beta_{0ij} X_0 + \beta_1 X_{1ij} + (u_{0j} + e_{0ij})$$

Equations

$$y_{ij} \sim \mathrm{N}(XB, \Omega)$$

$$y_{ij} = \beta_{0ij} x_0 + \beta_1 x_{1ij}$$

$$\beta_{0ij} = \beta_0 + u_{0j} + e_{0ij}$$

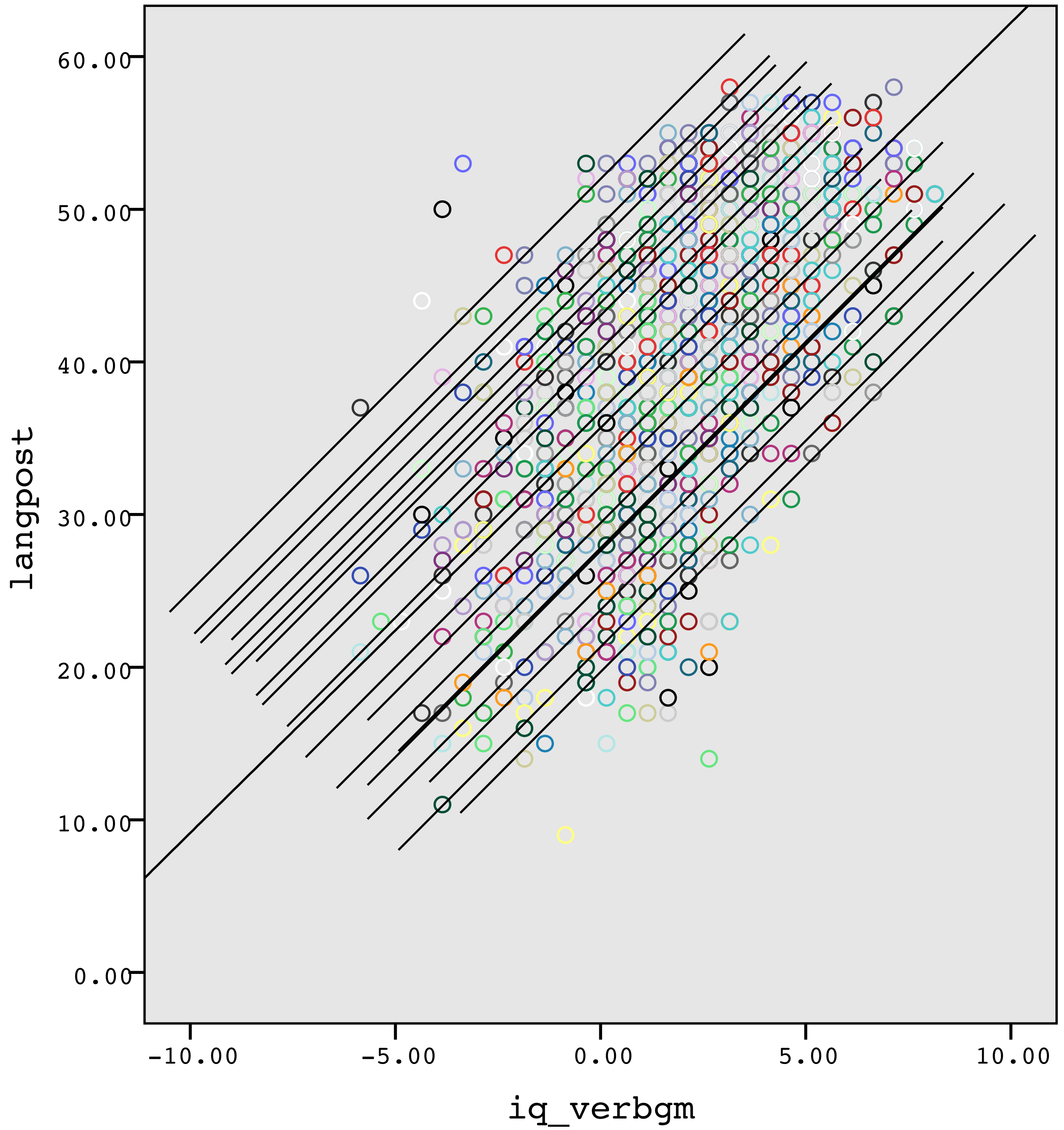
$$\left[u_{0j}\right] \sim \mathrm{N}(0, \Omega_u) \; : \; \Omega_u = \left[\sigma^2_{u0}\right]$$

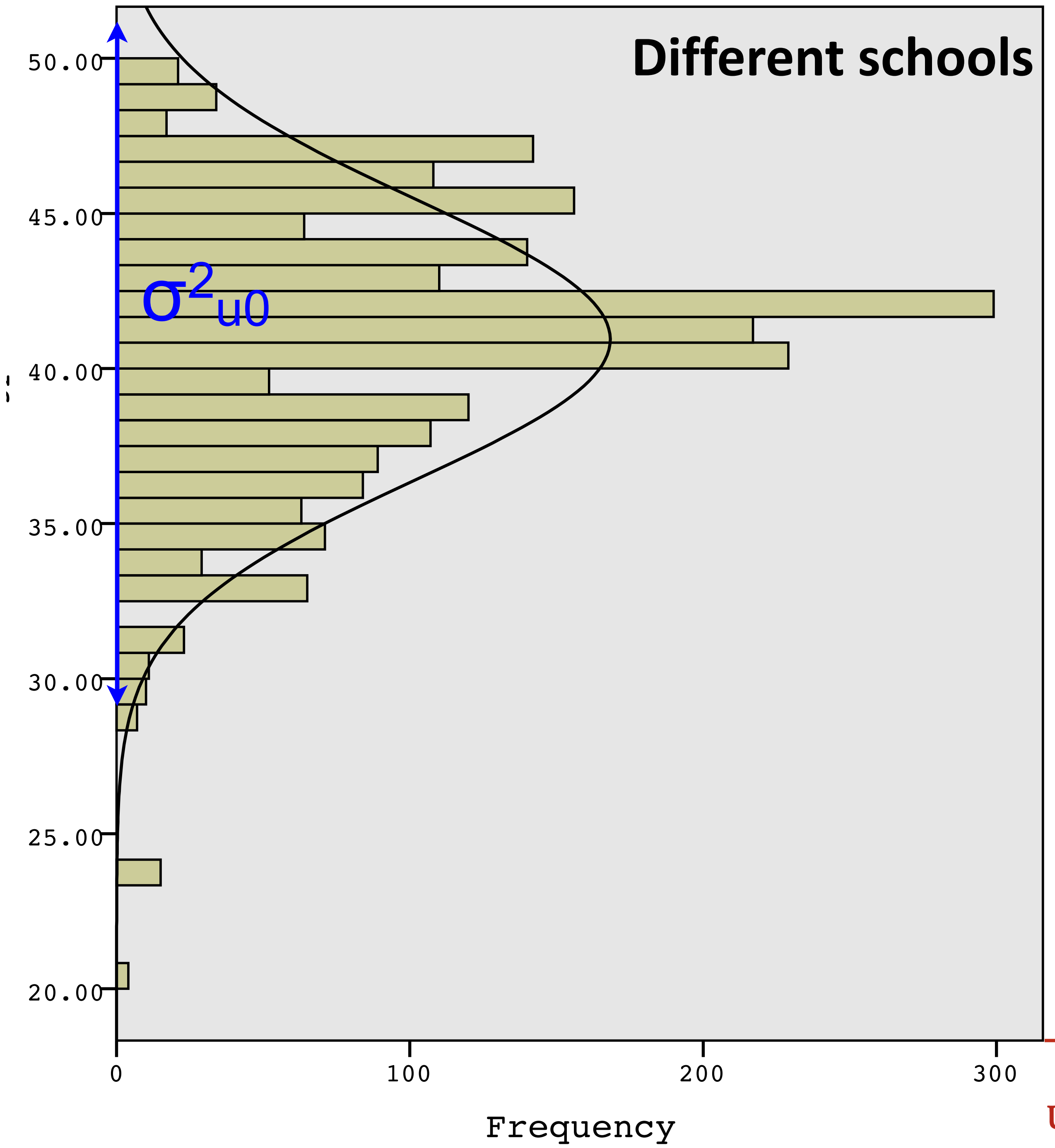
$$\left[e_{0ij}\right] \sim \mathrm{N}(0, \Omega_e) \; : \; \Omega_e = \left[\sigma^2_{e0}\right]$$

β_1: Fixed slope

Random intercept model, 1 covariate, fixed slope

$$Y_{ij} = \beta_{0ij}X_0 + \beta_1 X_{1ij} + (u_{0j} + e_{0ij})$$





Random intercept, 1 covariate, random slope

- Random intercepts plus a covariate with a random slope

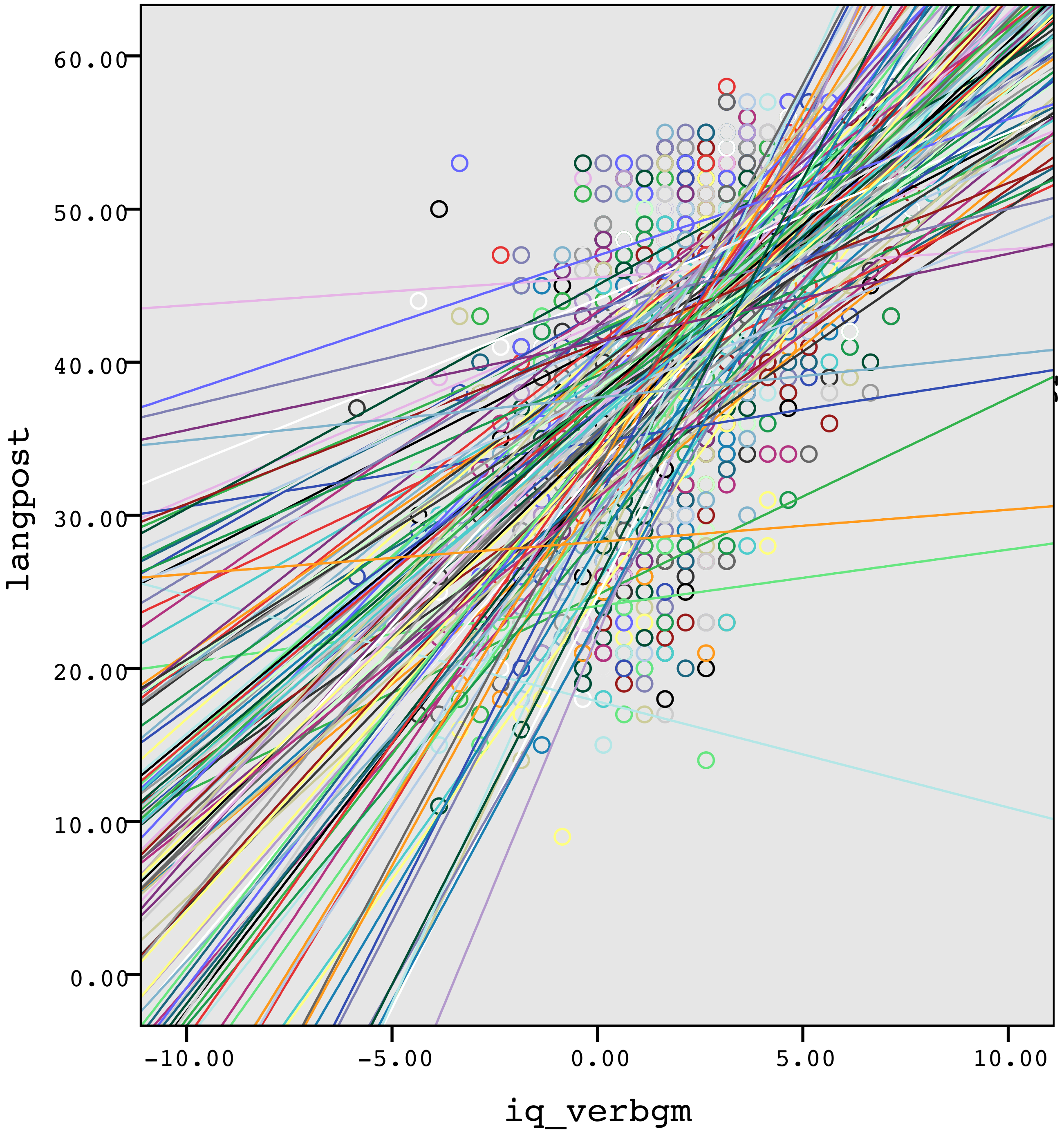
- Compare to heterogeneous (non-parallel) regression (or factor – covariate interaction)

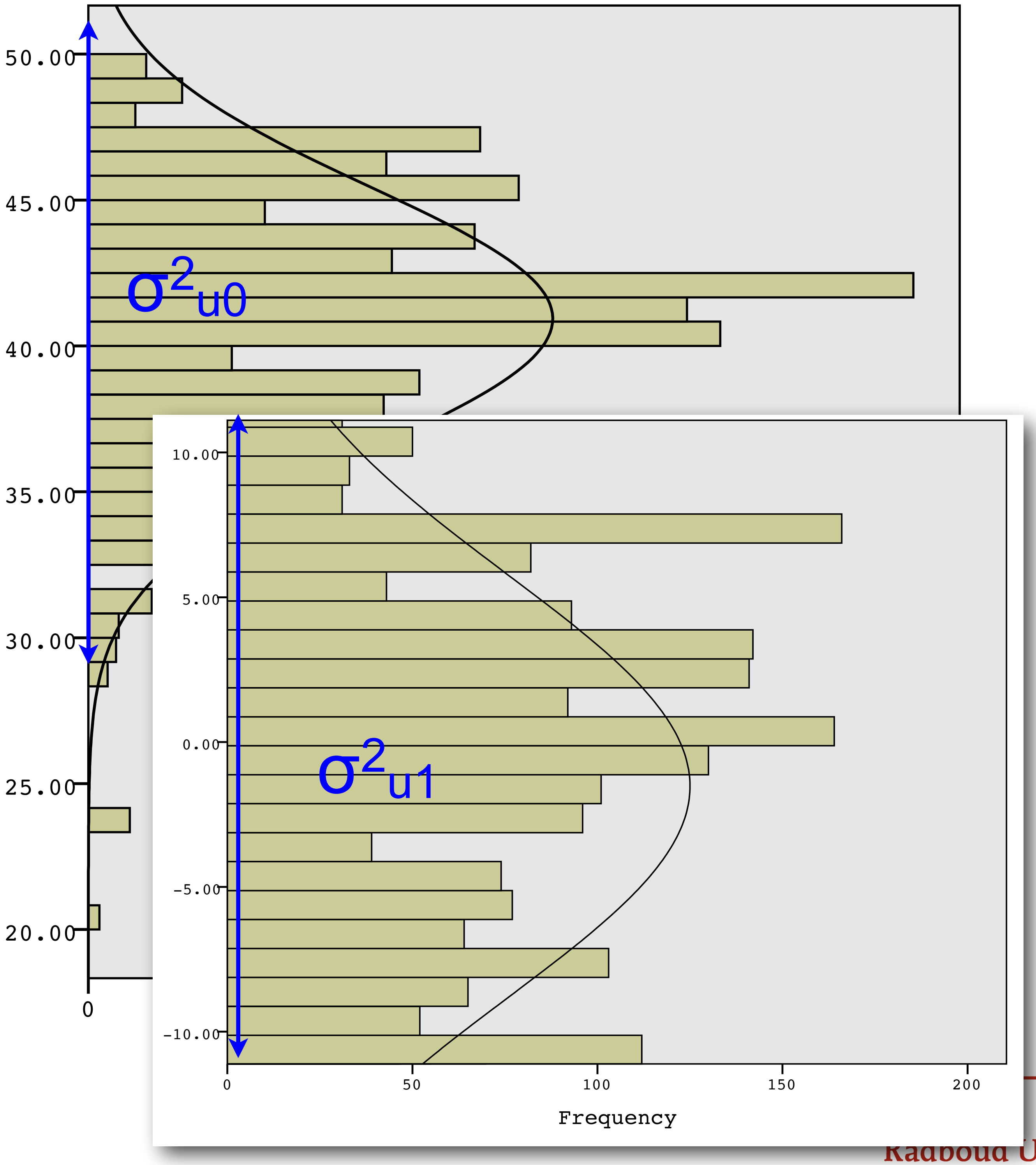
- Six parameters:

- Fixed: Average intercept
- Fixed: Average pooled within-group slope of covariate
- Random Level 2: Variance of intercepts
- Random Level 2: Variance of slopes
- Random Level 2: Intercept-slope covariance
- Random Level 1: Residual variance within groups

Random intercept, 1 covariate, random slope

$$Y_{ij} = \beta_{0ij}X_0 + \beta_{1j}X_{1ij} + (u_{0j} + u_{1j} + e_{0ij})$$





Random intercept, 1 covariate, random slope

$$Y_{ij} = \beta_{0ij}X_0 + \beta_{1j}X_{1ij} + (u_{0j} + u_{1j} + e_{0ij})$$

Equations

$$y_{ij} \sim \mathrm{N}(XB, \Omega)$$

$$y_{ij} = \beta_{0ij}x_0 + \beta_{1j}x_{1ij}$$

$$\beta_{0ij} = \beta_0 + u_{0j} + e_{0ij}$$

$$\beta_{1j} = \beta_1 + u_{1j}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim \mathrm{N}(0, \Omega_u) : \Omega_u = \begin{bmatrix} \sigma^2_{u0} & \\ \sigma_{u01} & \sigma^2_{u1} \end{bmatrix}$$

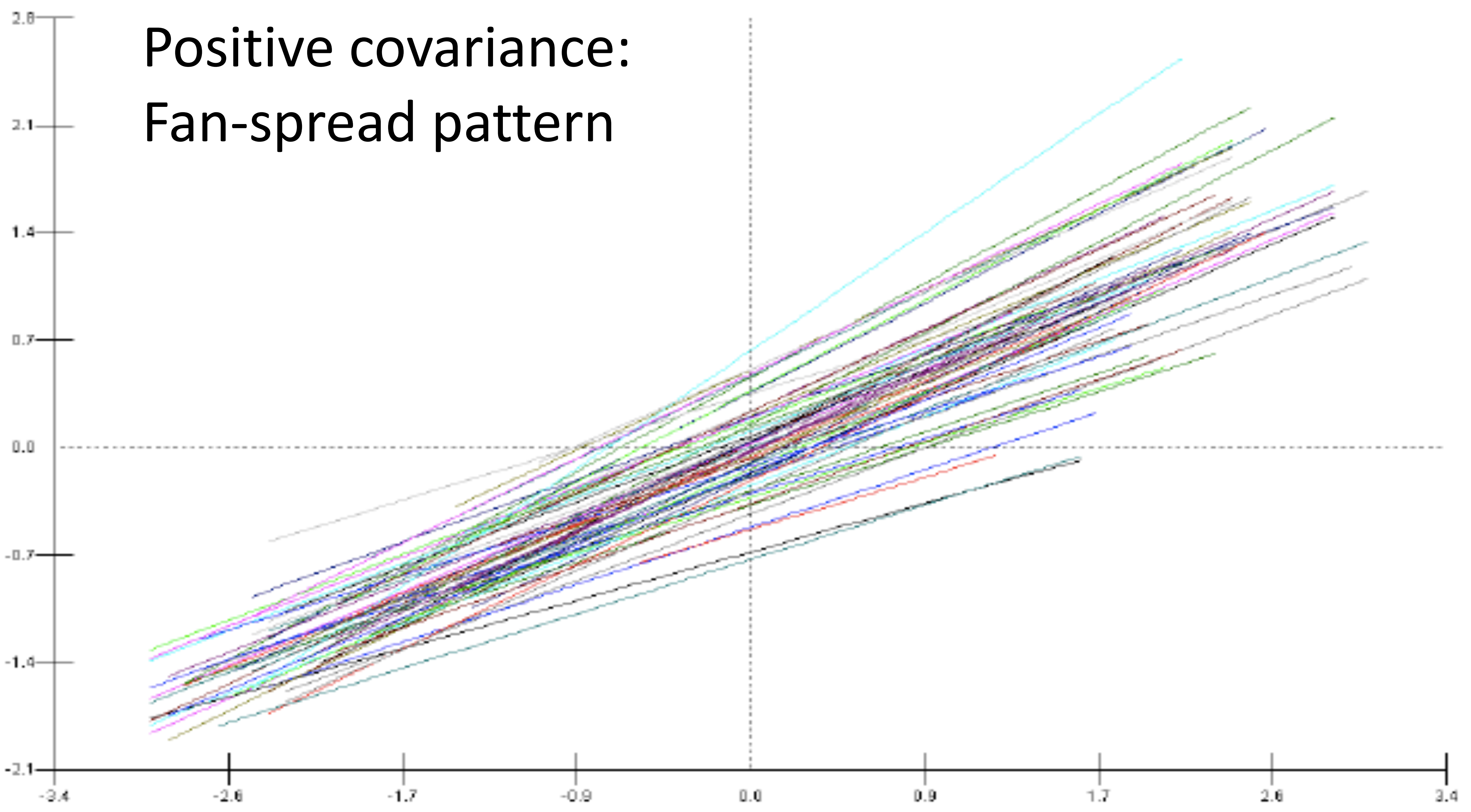
$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim \mathrm{N}(0, \Omega_e) : \Omega_e = \begin{bmatrix} \sigma^2_{e0} \end{bmatrix}$$

β_{1j}: Mean pooled within-group slope

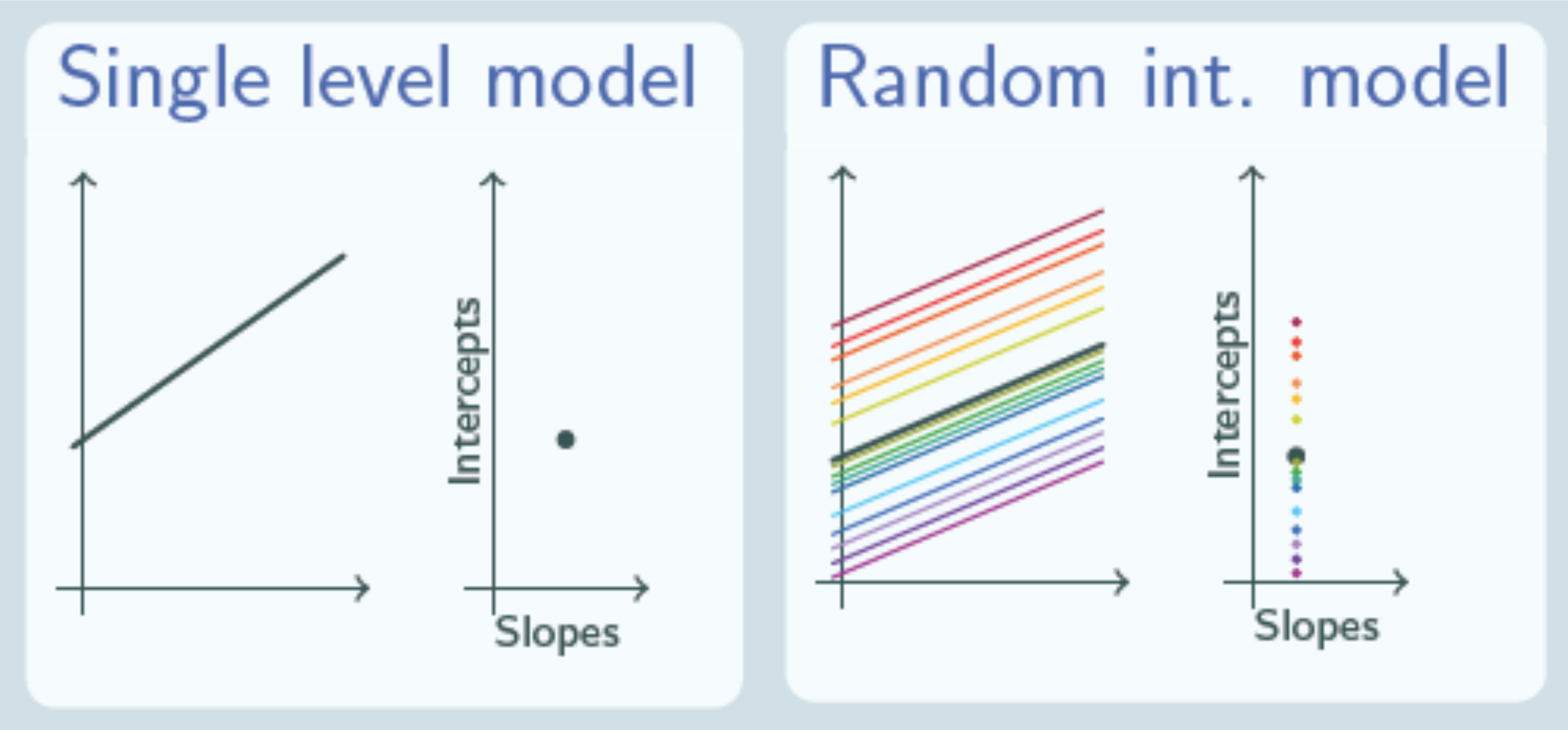
σ^2_{u1}: Variance of slopes

σ^2_{u01}: Intercept-slope covariance

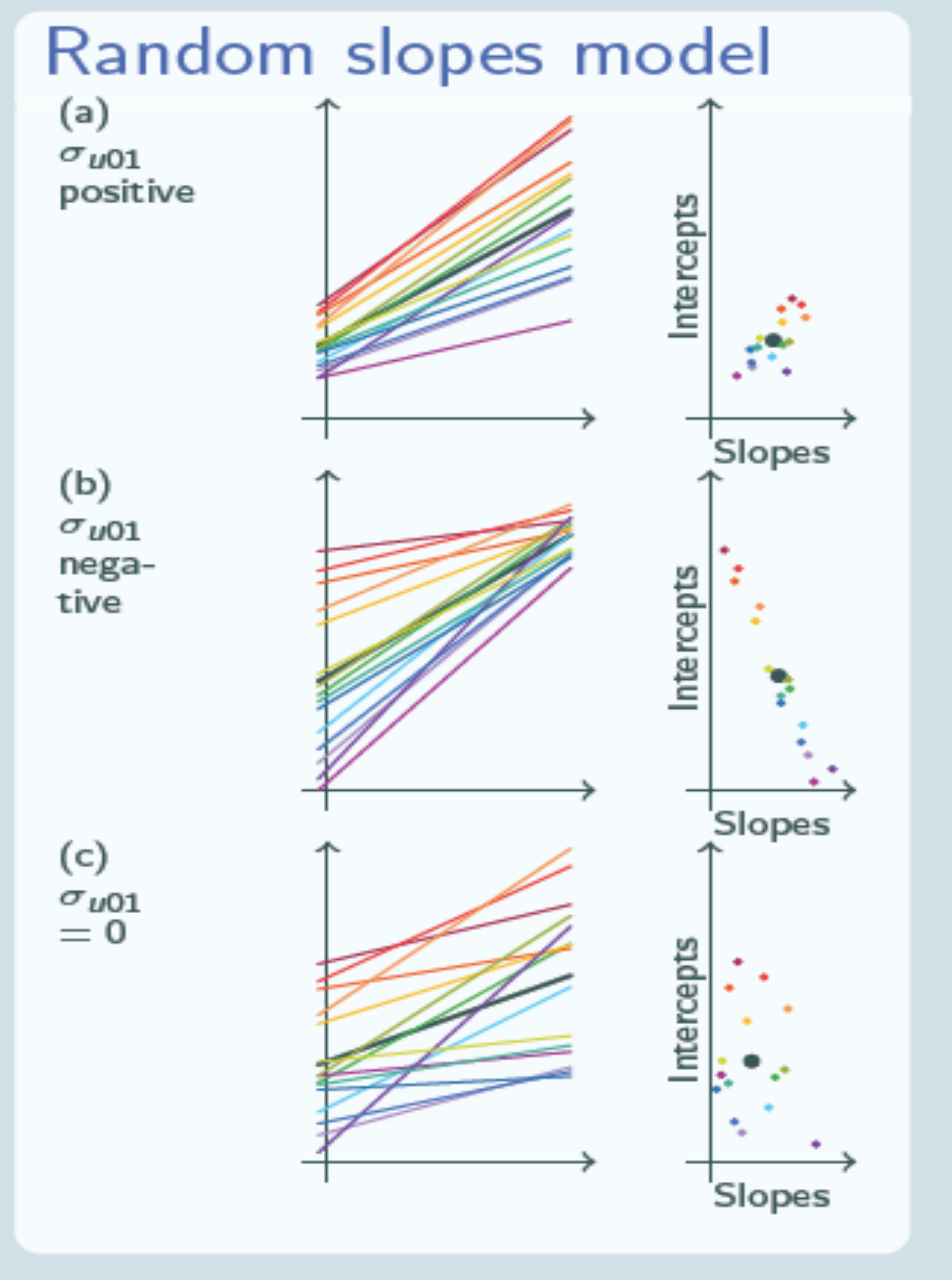
Intercept-Slope covariance



Interpret σ_{u0}^2 , σ_{u1}^2 and σ_{u01} together



For single level or random intercept models, σ_{u01} is not defined (there is no variation in slopes)



For random slope models,

- σ_{u01} positive means a pattern of fanning out
- σ_{u01} negative means a pattern of fanning in
- $\sigma_{u01} = 0$ means no pattern

Residuals

In multilevel models, residuals exist at every level

random-intercept model:

$$Y_{ij} = \beta_{0j} + \beta_1 X_{ij} + \varepsilon_{ij}$$

$$= (\beta_0 + u_{0j}) + \beta_1 X_{ij} + \varepsilon_{ij}$$

$$= (\beta_0 + \beta_1 X_{ij}) + u_{0j} + \varepsilon_{ij}$$

$$= \text{predicted value} + \text{level-2 residual} + \text{level-1 residual}$$

Caterpillar plot