What is Multilevel Modelling?

Mark Tranmer
CCSR / Mitchell Centre for Social Network Analysis
Social Statistics
University of Manchester
mark.tranmer@manchester.ac.uk

methods@manchester
14th October 2010

population structure and dependencies

- Notion of randomness, and independence makes sense sometimes e.g. in natural sciences, and keeps statistical theory relatively simple
- **But** people are not raffle tickets - population has structure, and people therefore have things in common; living in same area, going to the same school ... people cannot be regarded as 'independent units'.

Examples

- Pupils in schools
- Individuals in areas
- Workers in organisations

Brief History

- Problems of single level analysis, cross level inferences and ecological fallacy, highlighted in 1950s.
- Ecological fallacy also Demonstrated in geography in 1970s and 80s e.g. work by Stan Openshaw

Brief History

- In the 1980s then much discussion of school league tables (based on single level aggregate data), and the need to take pupil exam score variations into account when comparing schools, not just rely on a single level analysis of school means.
- Key papers and books by Aitkin and Longford, Goldstein and Raudenbush

Brief History

- Advances in computing power and estimation methods such as IGLS allowed models to be fitted with specialist software
- VarCL, HLM, ML2 > ML3 > MLn > MLwiN

Brief History

- Focus was initially on hierarchical structures and especially pupils in schools
- Also longitudinal, geographical studies
- More recently moved to non hierarchical situations such as cross-classified models.
- Also methods such as MCMC and ever increasing computing power have allowed more realistically complex models to be estimated

Substantive applications in social statistics: non exhaustive list

- Education
- Longitudinal studies
- Geography
- Health
- Social Networks
- Psychology

Extensions: more levels

- Individuals in households in areas
- Pupils in classes in schools in regions

Extensions: people not at level 1

- Longitudinal studies, where the **occasion** is the first level of a hierarchy
- Multivariate studies with several y variables per individual to capture a latent variable: e.g. various test scores for maths based subjects all taken by the individual may indicate numeracy

Realistically complex structures

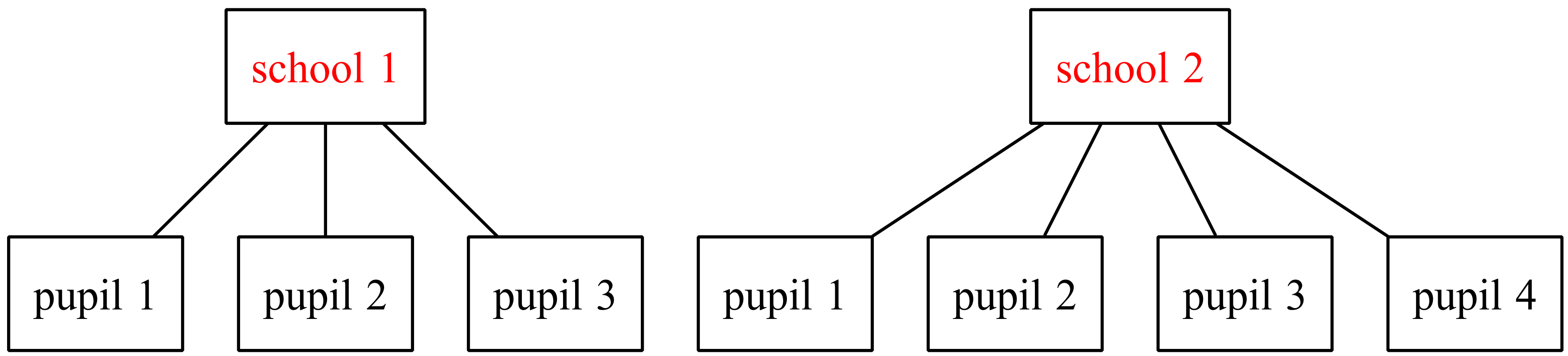
- Cross classifications: two pupils that sit next to each other in a school each live in a different local area of a city, but two people who live in the same local area each go to different schools
- Influence of neighbourhood and school on educational performance of individual

Realistically complex structures

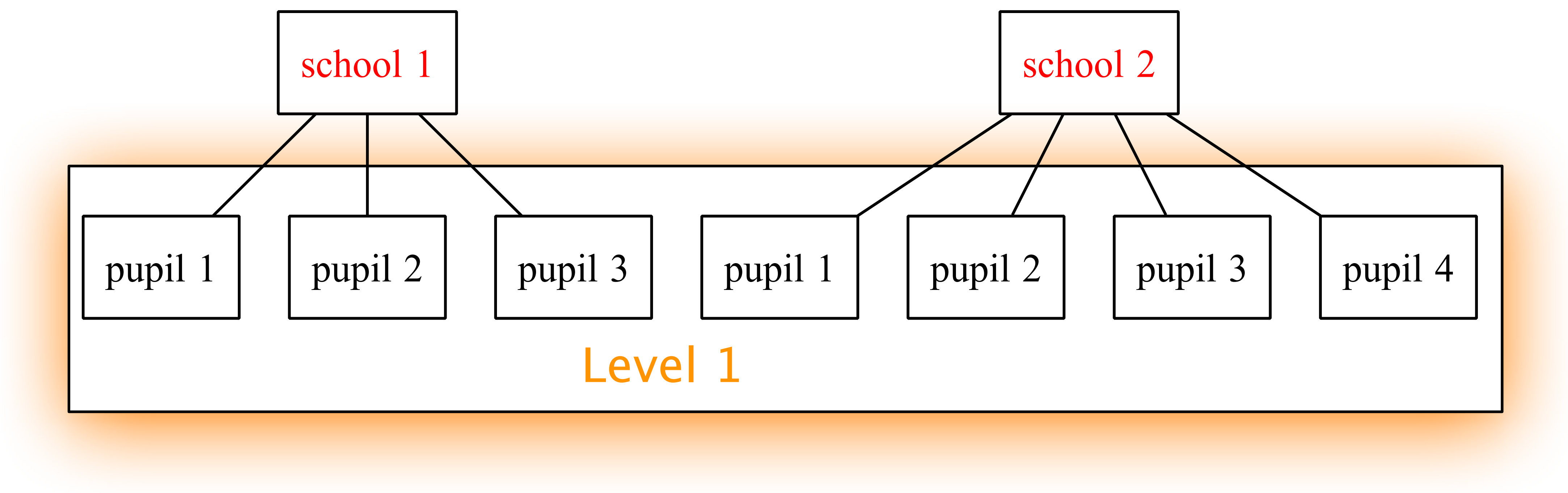
- Multiple membership models during the course of secondary education, some pupils attend more than one school, perhaps because their parents move. Some pupils therefore members of more than one group. Weights reflect this - number of years in each school. Most pupils stay at same school.

Let's focus on the two level situation for the rest of the talk

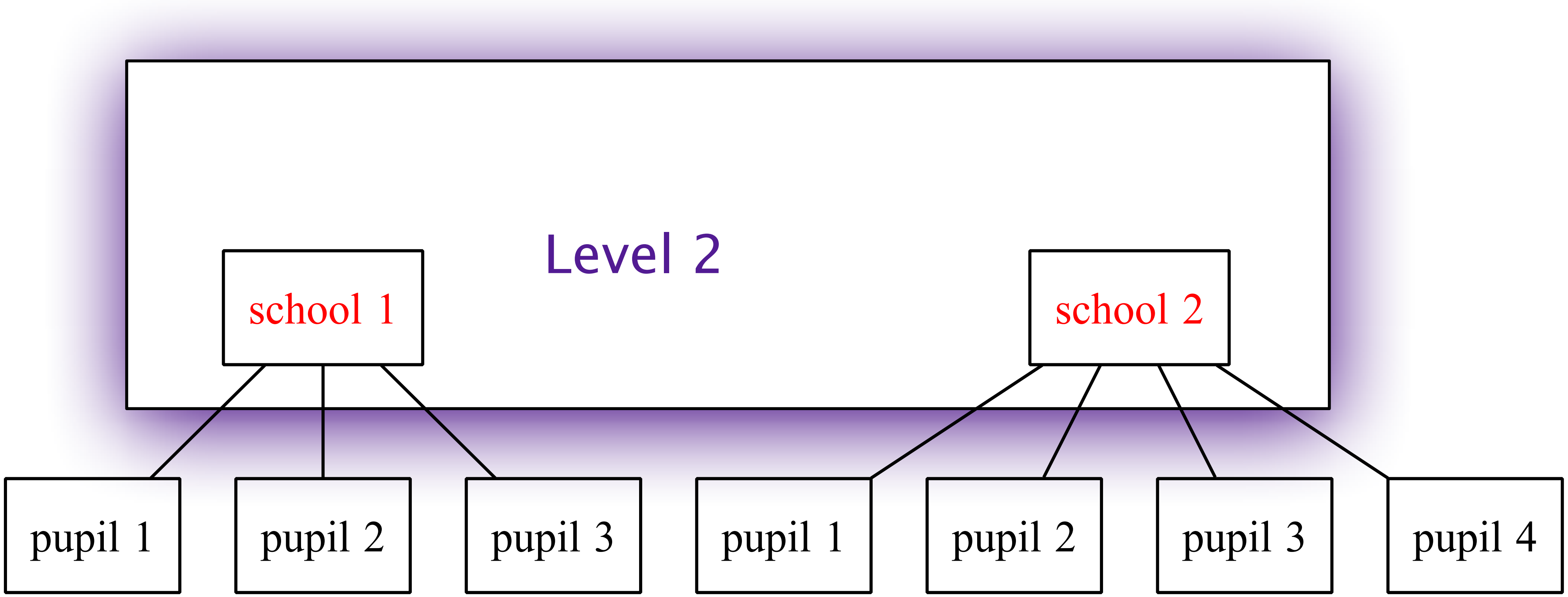
A simple 2-level hierarchy



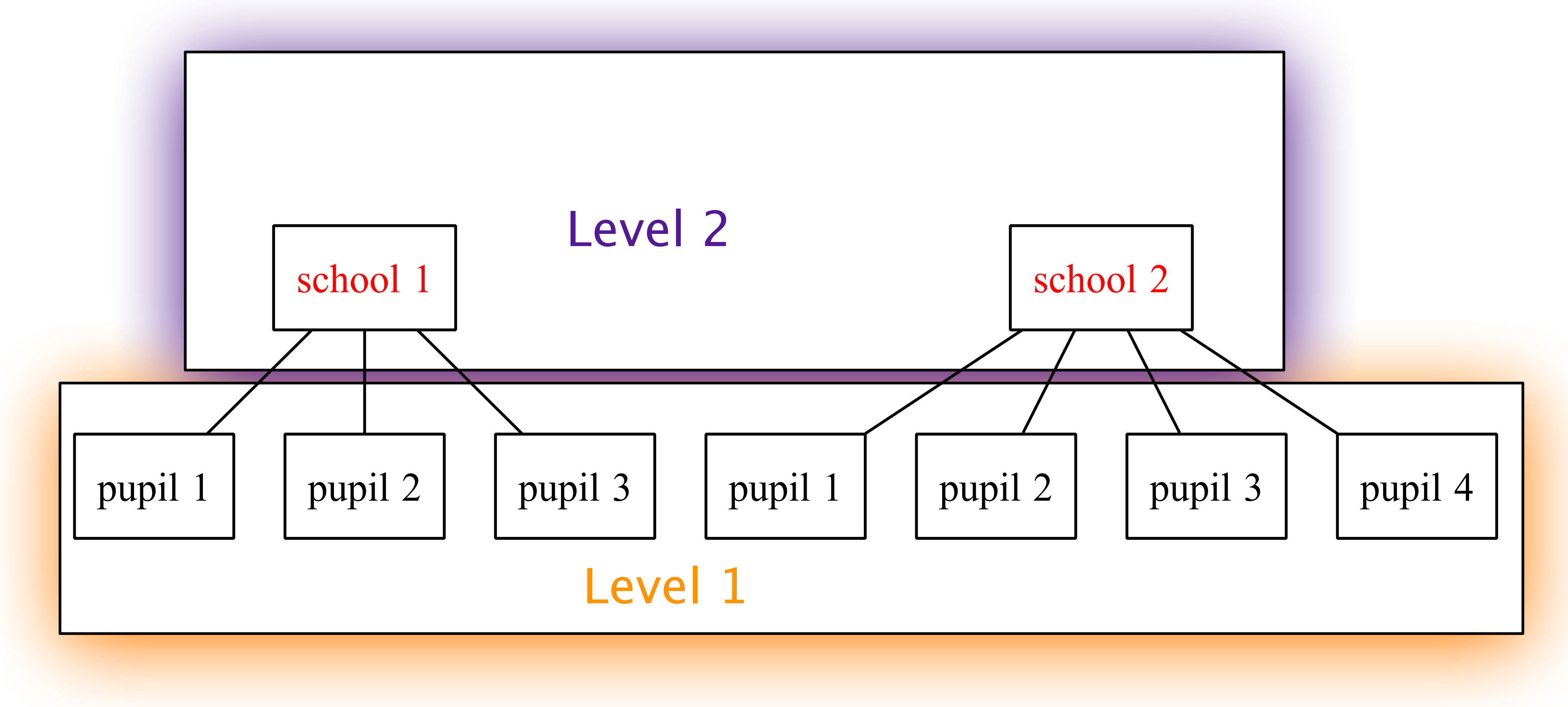
A simple 2-level hierarchy



A simple 2-level hierarchy



A simple 2-level hierarchy



What data do we need?

- Individual units (often people), with their group indicators (e.g School, Area).
- One or more response variable(s)
- In general we need more than one person per group
- In general we would expect to have at least 10 groups, 20 or more even better. Partly depends on what we want to do.

Two level example: pupils in schools

- Suppose we have data for 4000 pupils in 60 schools
- Including a measure of exam performance at 16 (*y*) and exam performance at 11(x)
- perhaps also other explanatory variables: gender, age of school buildings, % pupils on free school meals.
- Suppose we want to relate y to x, what can we do.

Aggregate to school level

- We could aggregate the exam score at 11 and exam score at school level, so that we have 60 pairs of school means, rather than 4000 pairs of exam scores.
- We could regress school mean y on school mean x
- However if we make inferences from that school level regression back to individual, we run into problems.
- “Ecological Fallacy” (Robinson, 1950).

Problems of single level analysis

- We could work at the pupil level, and ignore the schools.
- Then we are ignoring the context: each pupil goes to a particular school
- We could add the 60 schools to the model as 59 dummy variables: fixed effects model
- But that's a lot of dummy variables - model quickly becomes very full of parameters.

Multilevel models

- Multilevel models allow us to look at different levels simultaneously: e.g. the pupil level and the school level
- Doesn't require a huge number of parameters.
- Also allows flexibility: e.g .the relationship between exam score at 11 and exam score at 16 can be different in different schools
- Takes into account different group sizes through the idea of 'shrinkage'.

Inferences and assumptions

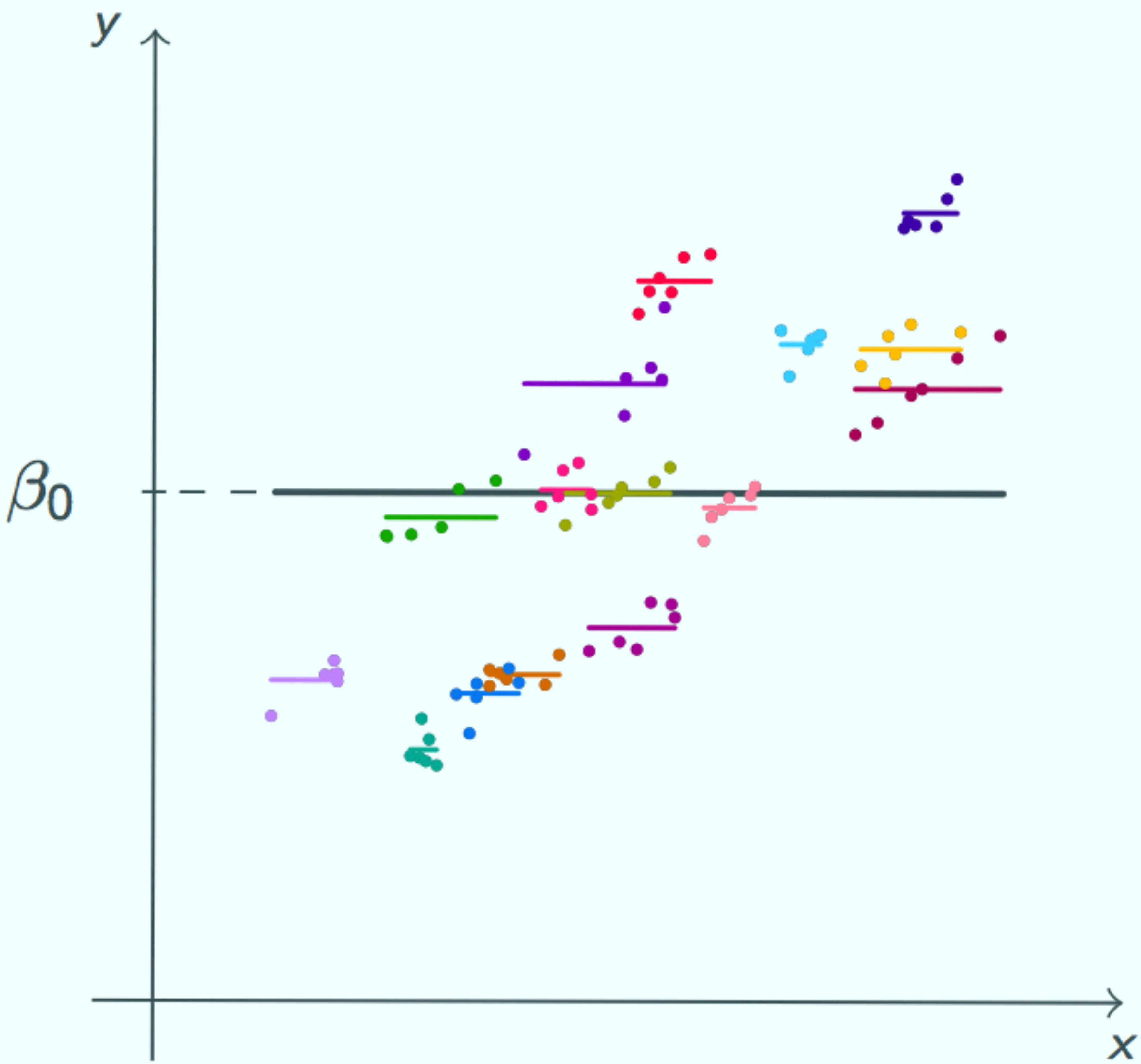
- Multilevel models sometimes called random effects models: partly because groups are themselves regarded as a random sample
- Can make inferences about groups not in sample
- If we have all groups in population can still regard these as sample; realisation of underlying population generating process. In short, can use multilevel models even if all the groups in our data.
- We can use multilevel models regardless of whether the population structure is directly of interest. or not. E.g. we can apply a model based approach to reflect the way that the data were collected.

variance components model

$$y_{ij} = \beta_0 + u_j + e_{ij}$$

- Multilevel model with no explanatory variables
- i,j subscripts for pupil i in school j
- variance of u school component; variance of e pupil component; *u* and e assumed uncorrelated
- Hence allows us to see how much variation in the response is at level 2 and how much at level 1 prior to adding x variables to model
- level 1 and level 2 variance add up to total variance

Variance components model

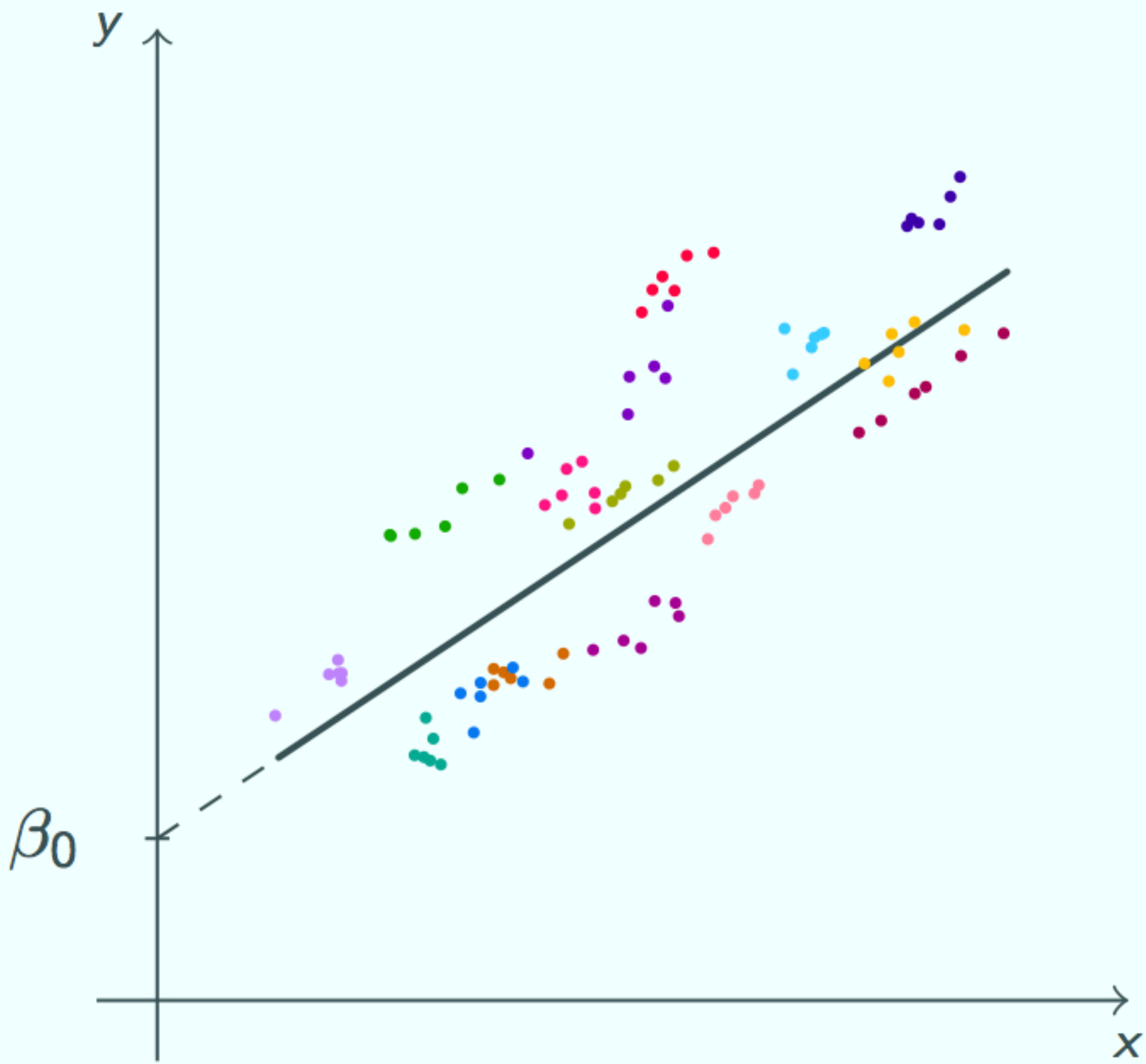


Simple regression model

$$y_i = \beta_0 + \beta_1 x_i + e_i$$

- Single level model
- relates response (y) for pupil i to explanatory variable (x) for pupil i
- Doesn’t take school into account

Single level regression model

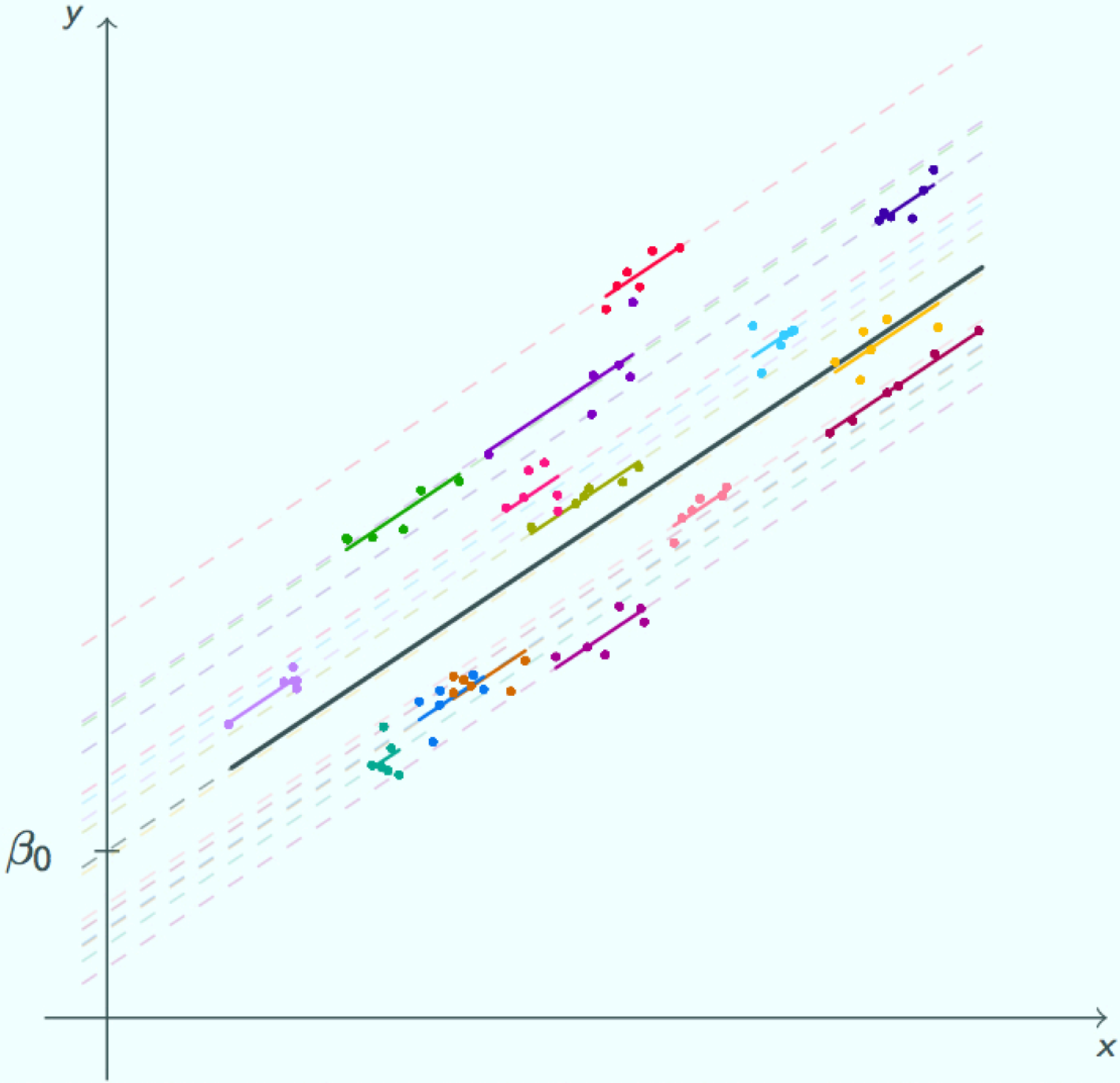


Random intercepts model

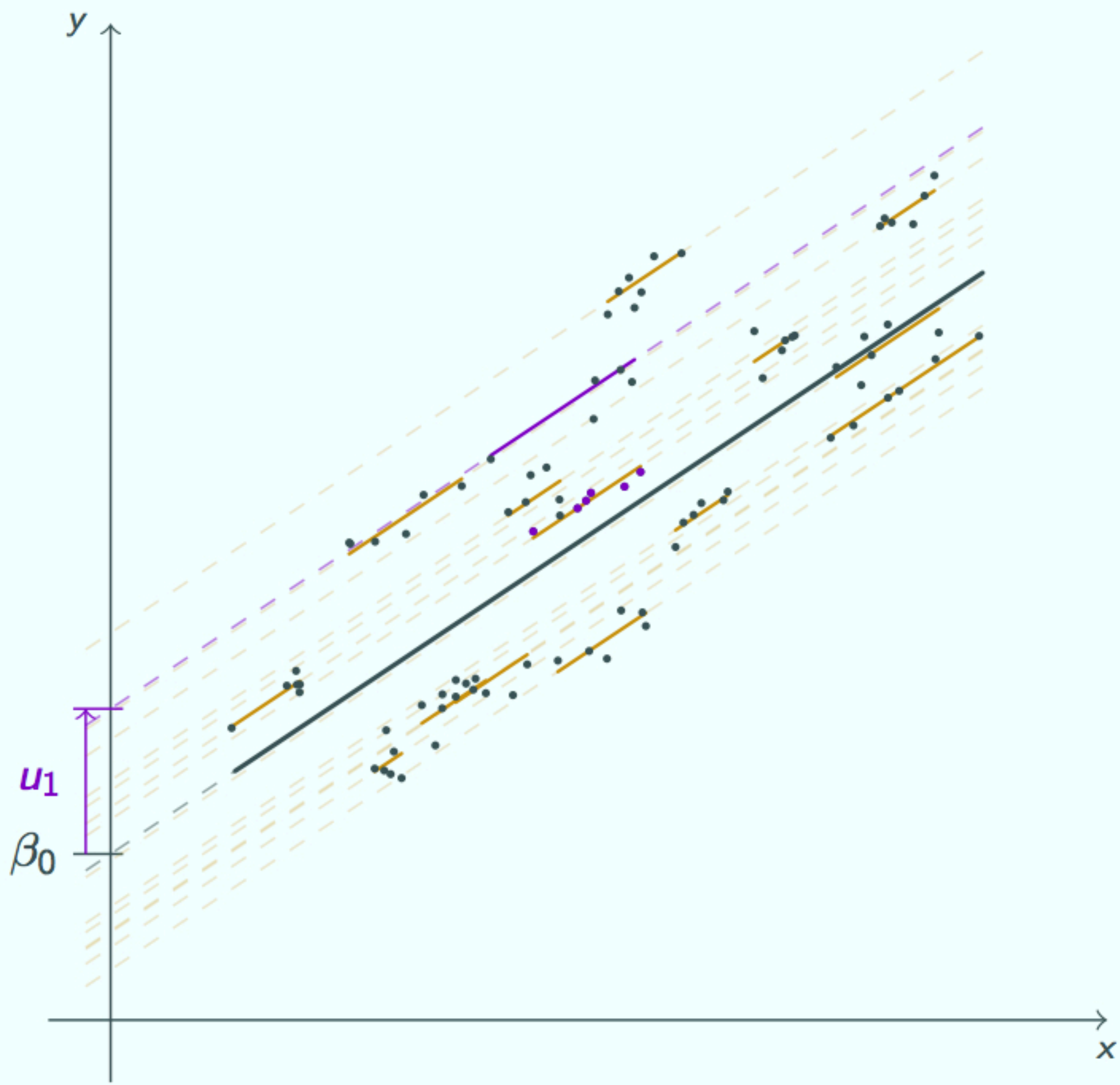
$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_j + e_{ij}$$

- Multilevel model: combines variance components with single level model
- Relates response (y) for pupil i in school j to explanatory variable (x) for pupil i in school j
- Also allows the school mean performance to vary
- Can plot school level residuals (u j) and their confidence intervals to fairly compare schools. “caterpillar plots”.

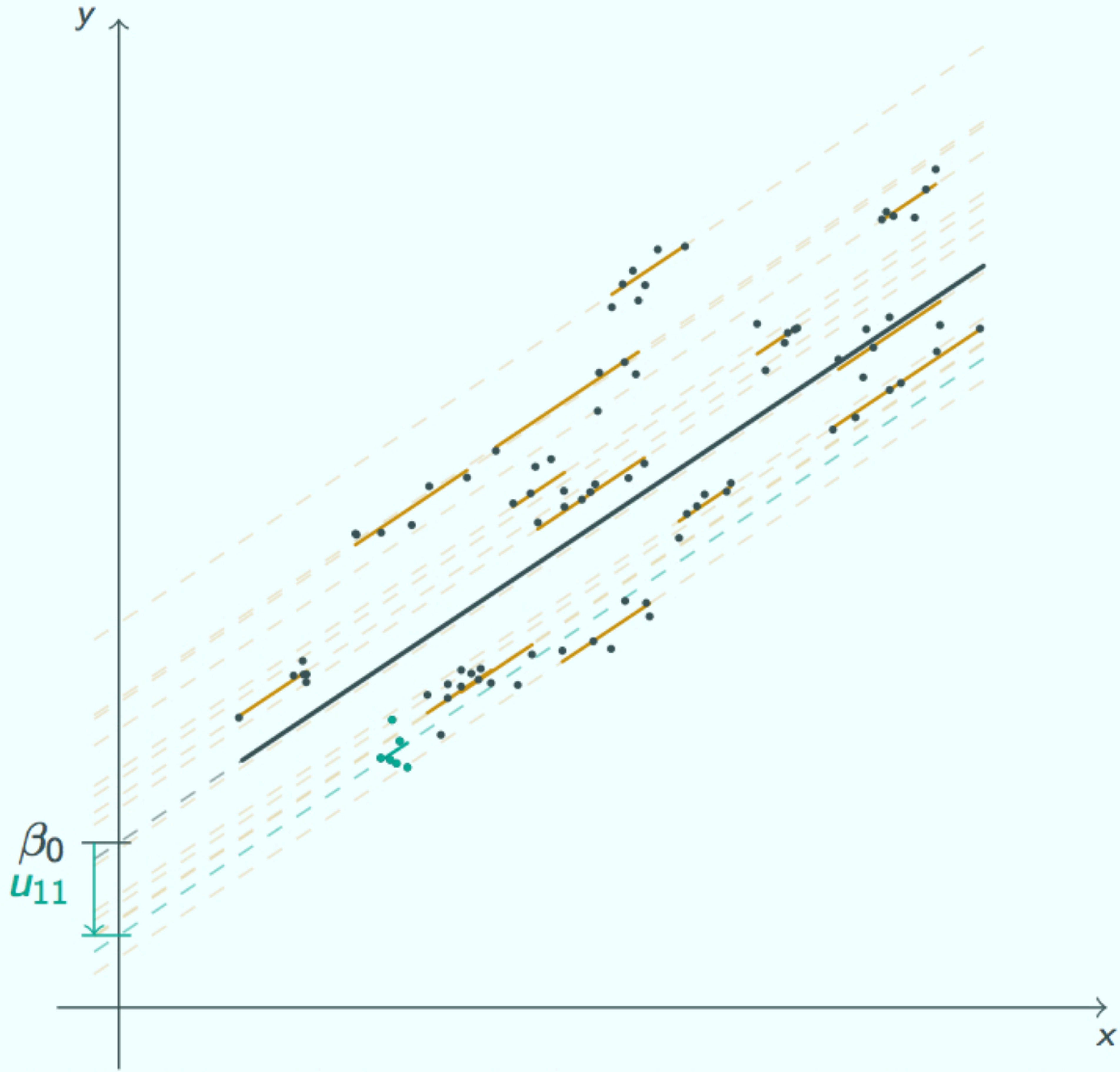
Random intercept model



Random intercept model



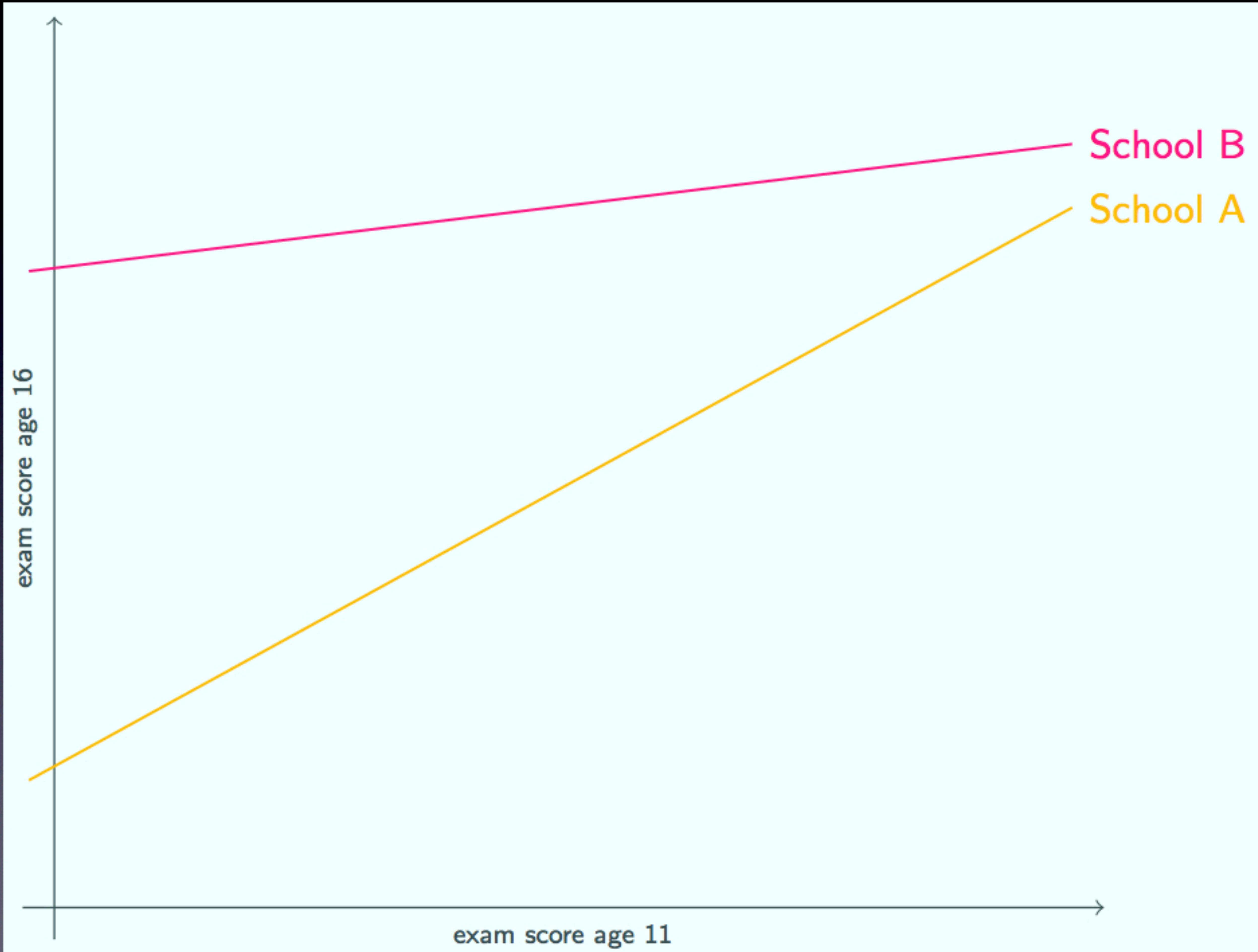
Random intercept model



Random slopes model

$$y_{ij} = \beta_0 + \beta_{1j} x_{ij} + u_j + e_{ij}$$

- Multilevel model
- Relates response (y) for pupil i in school j to explanatory variable (x) for pupil i in school j
- Also allows the school mean performance to vary
- Also allows relationship between y and x to be different from school to school


School B
School A
exam score age 16
exam score age 11

Adding level 2 variables

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 \bar{x}_{2j} + \beta_3 x_{3j} + u_j + e_{ij}$$

- Multilevel model: extending the random intercepts model
- Adds two level 2 (school level variables)
- An aggregate variable: "x 2 bar j" is the % of pupils on free school meals in each school
- A true school level variable: "x 3 j" is whether the school built in the last 50 years.

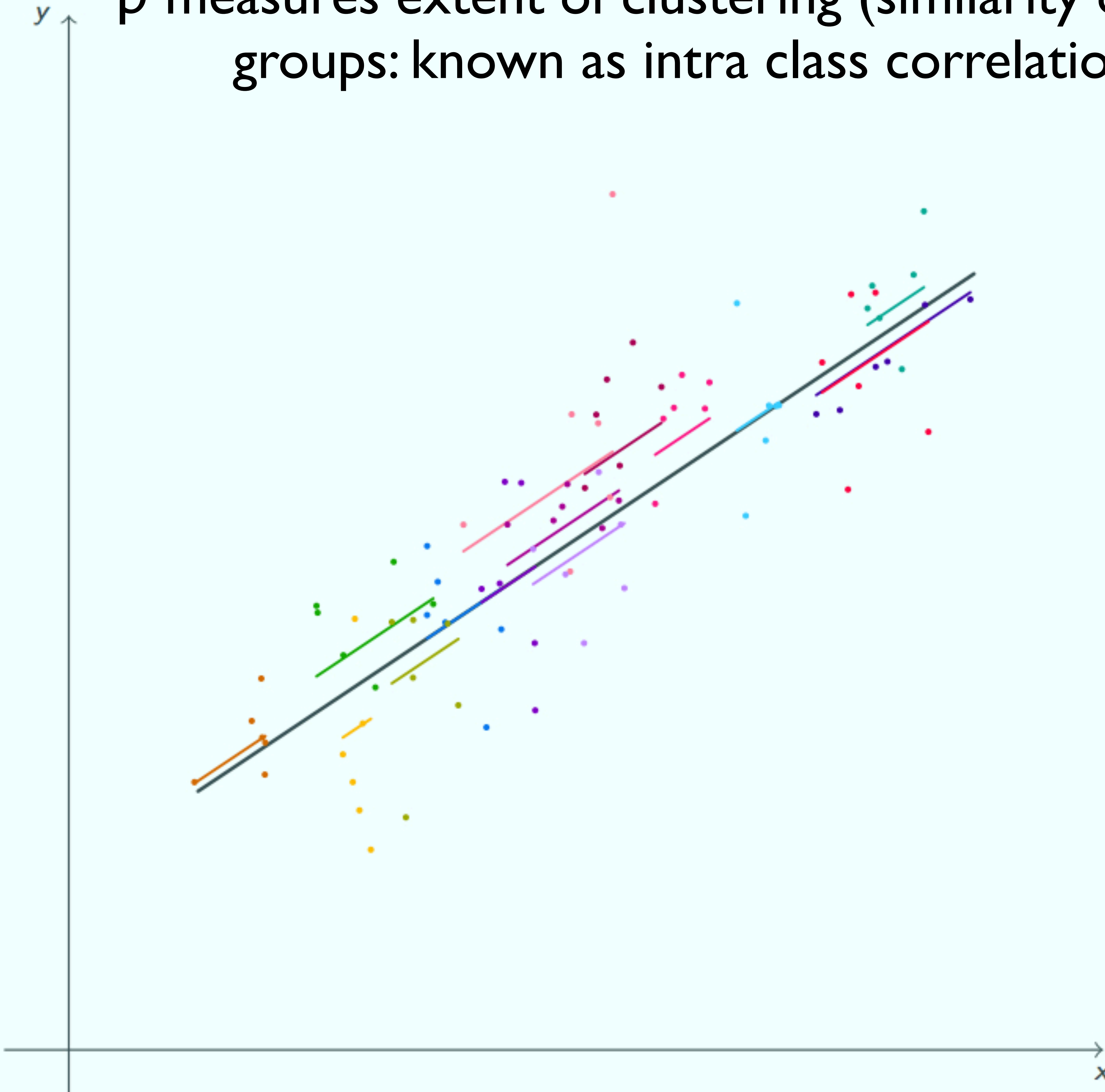
Cross level interactions

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 \bar{x}_{2j} + \beta_3 x_{1ij} \times \bar{x}_{2j} + u_j + e_{ij}$$

- Cross level interactions allow us to investigate the effect of an explanatory variable on the response in the context of another explanatory variable
- e.g the relationship between exam score at 11 and exam score at 16 in the context of % free school meals in the school
- Multilevel framework has powerful substantive use

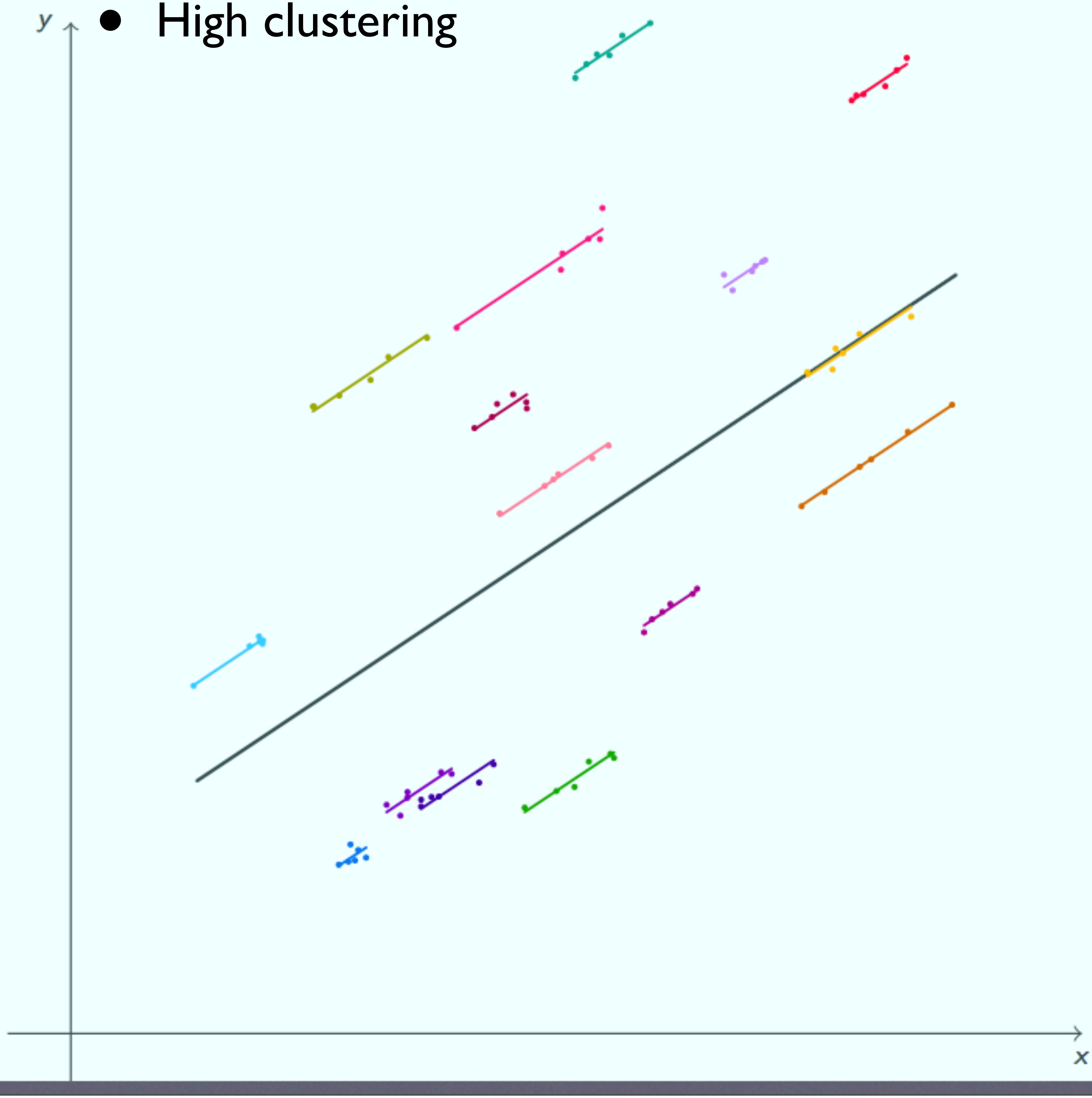
A small value of ρ

ρ measures extent of clustering (similarity of y)
groups: known as intra class correlation

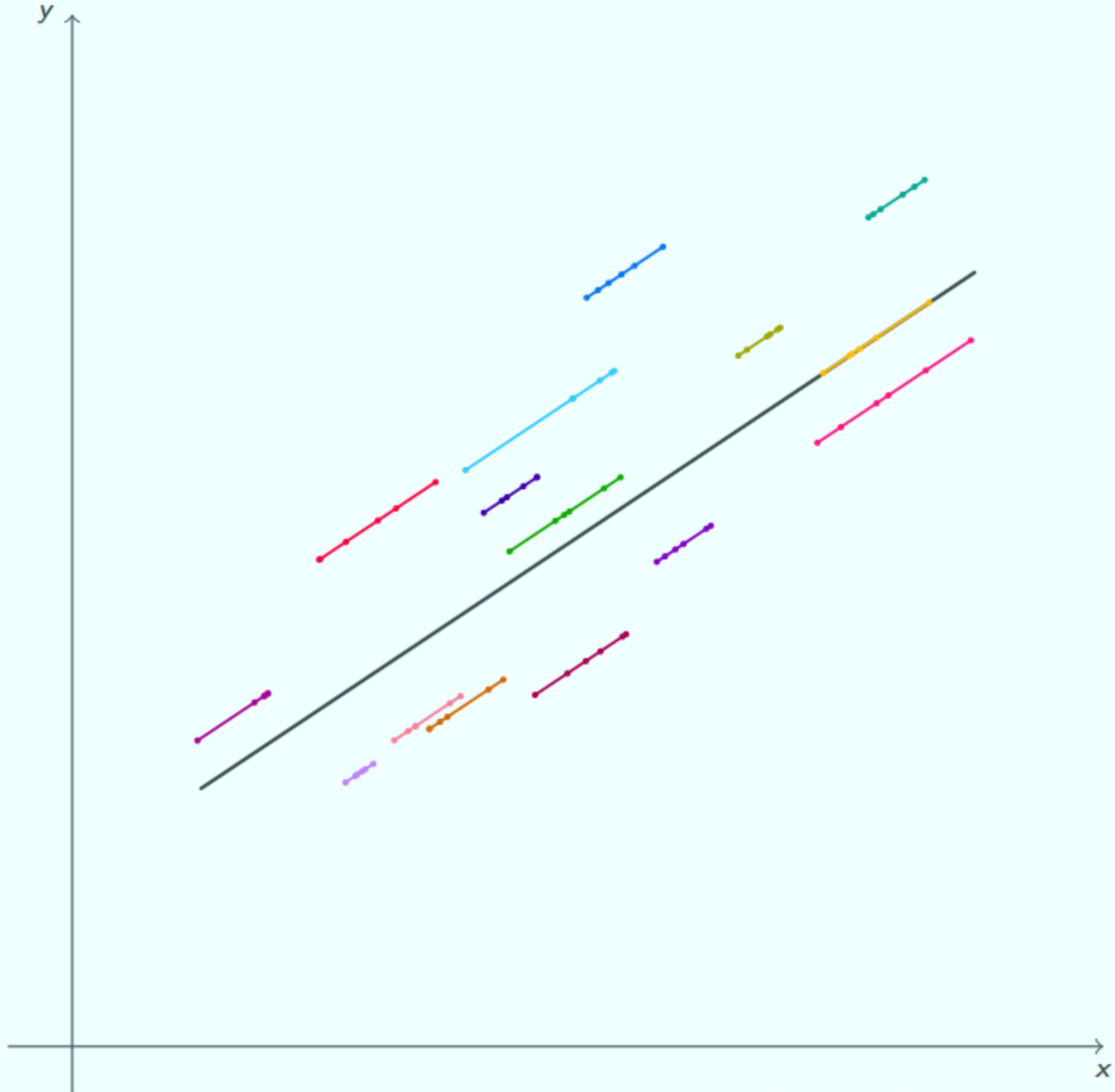


A large value of ρ

- High clustering

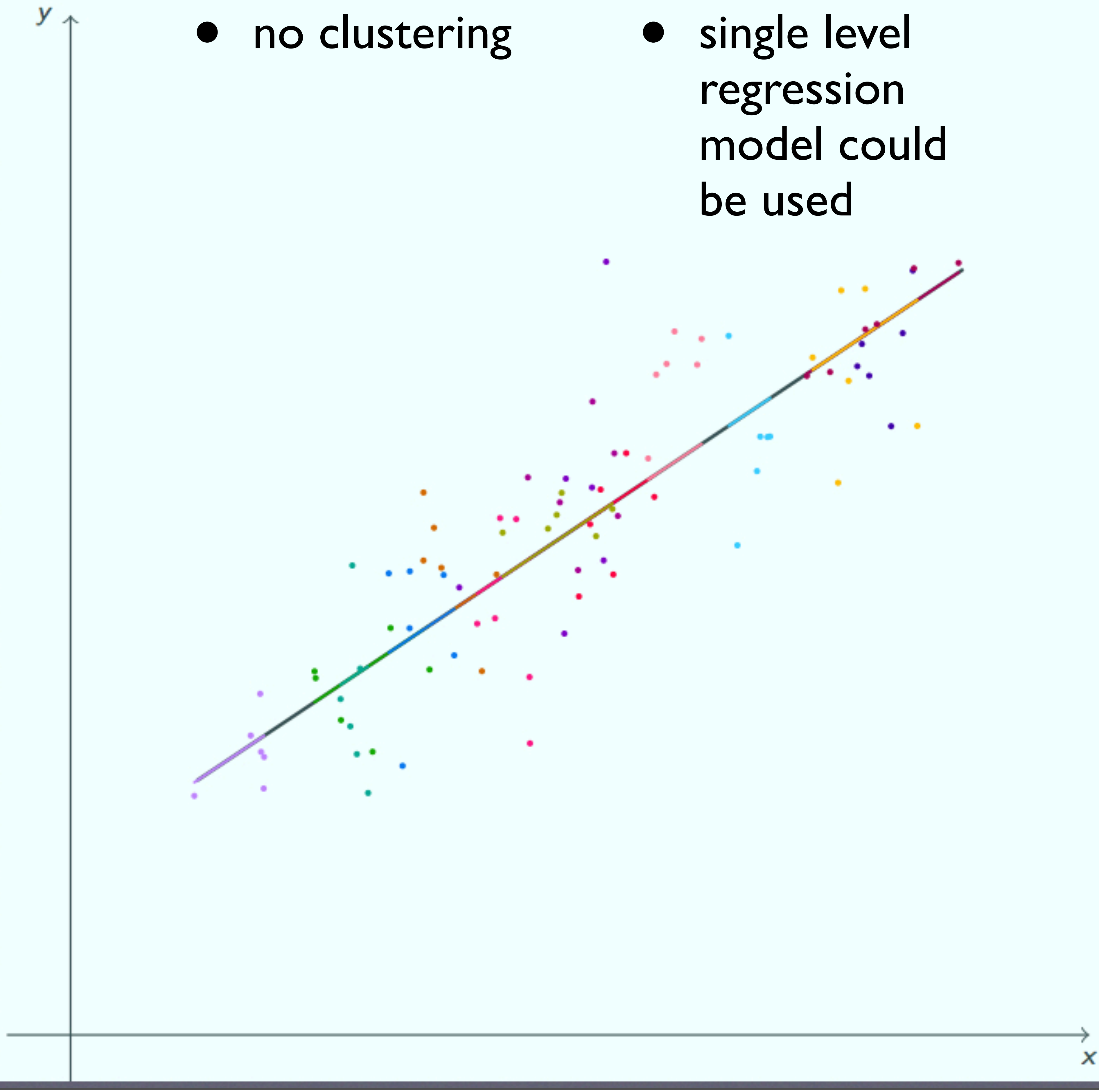


$\rho = 1$

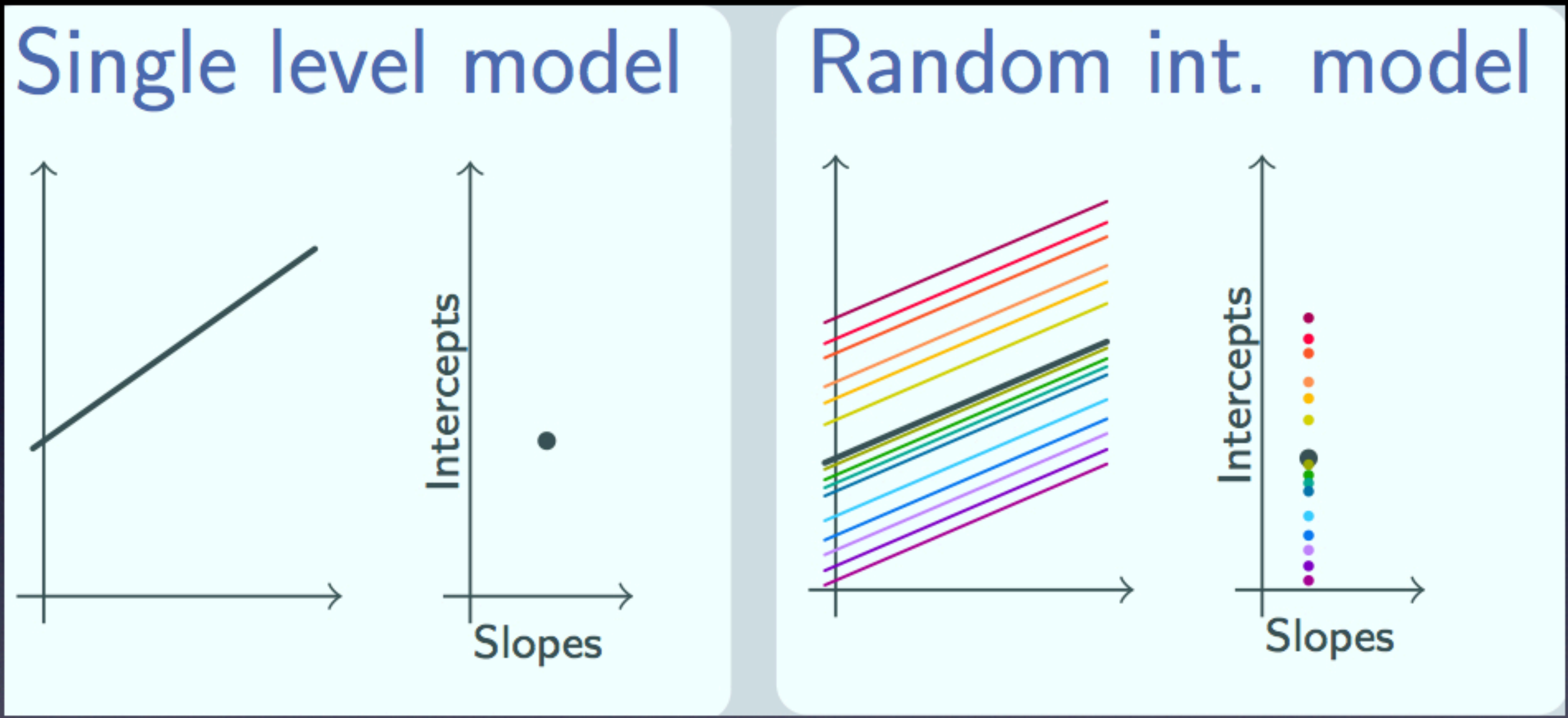


$\rho = 0$

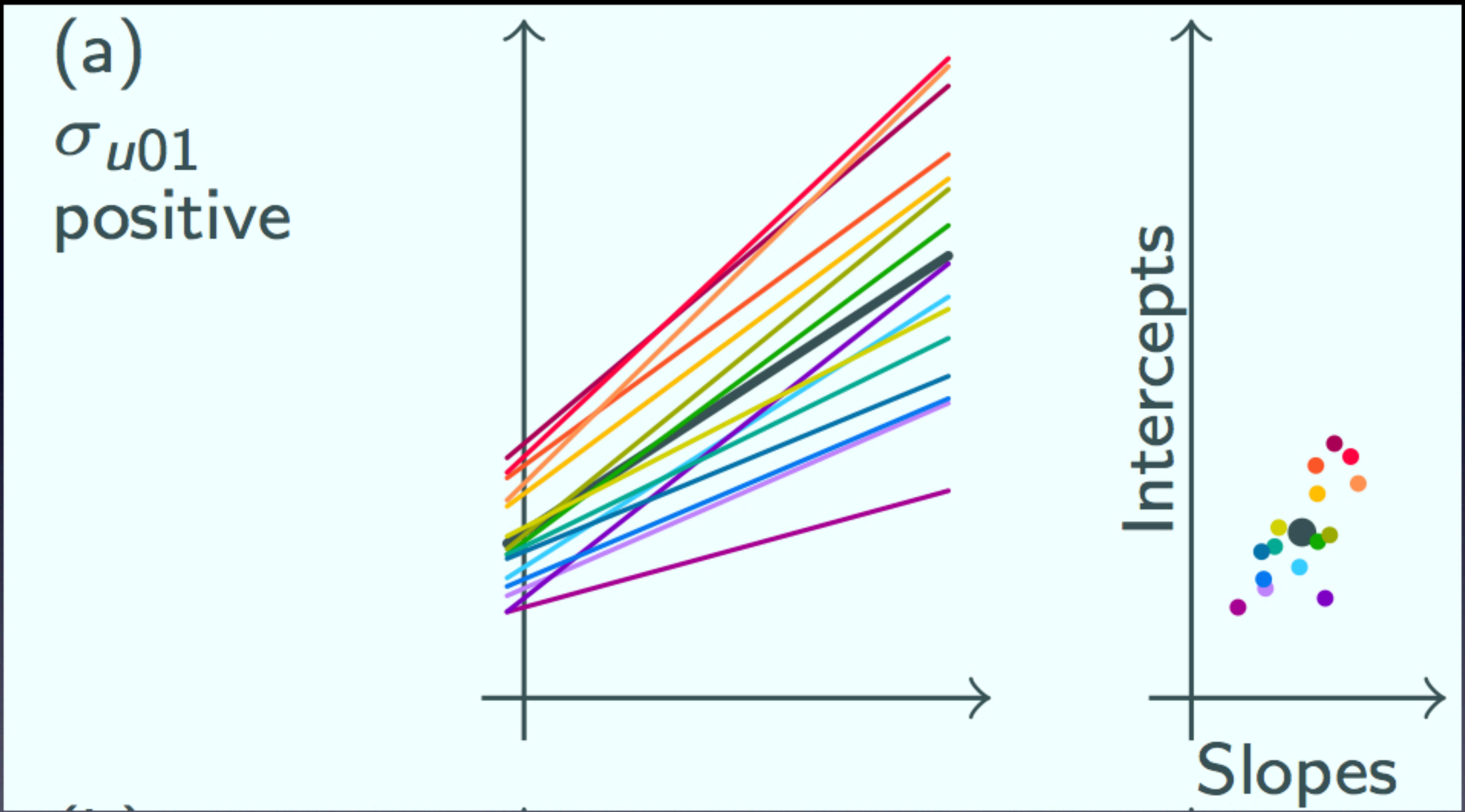
- no clustering
- single level regression model could be used



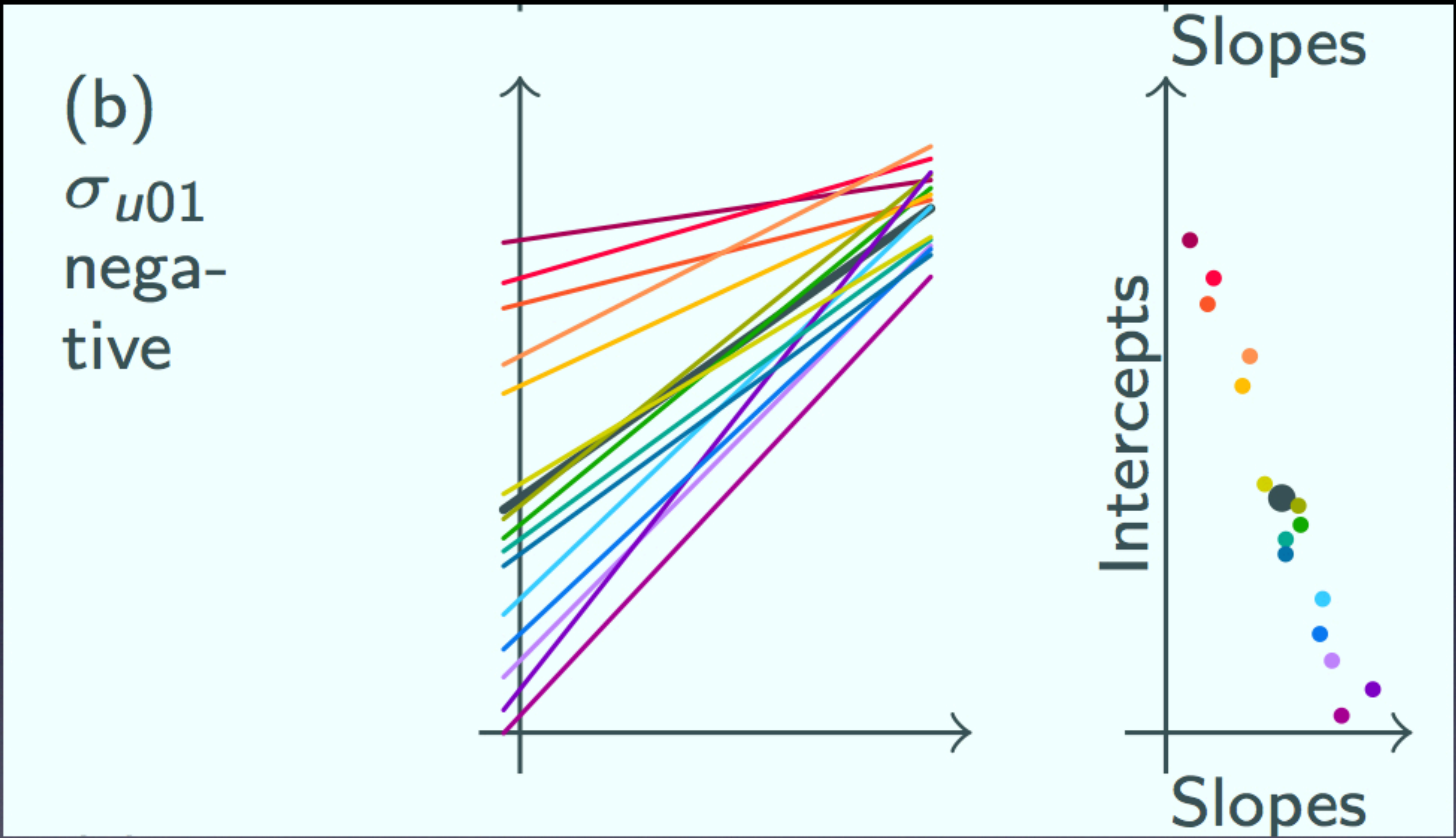
- Patterns of intercepts and slopes



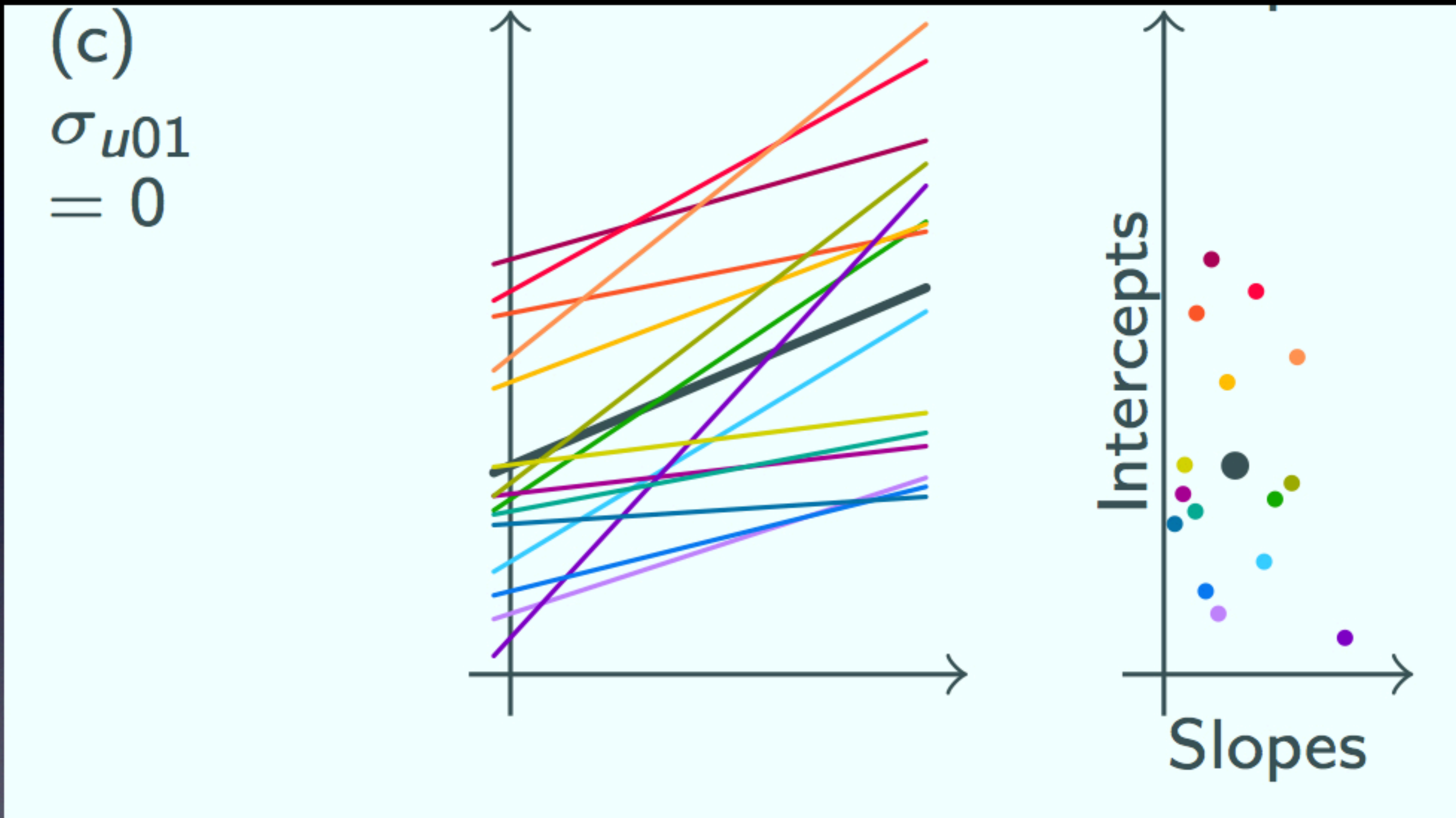
- Patterns of intercepts and slopes: random slopes model (1)



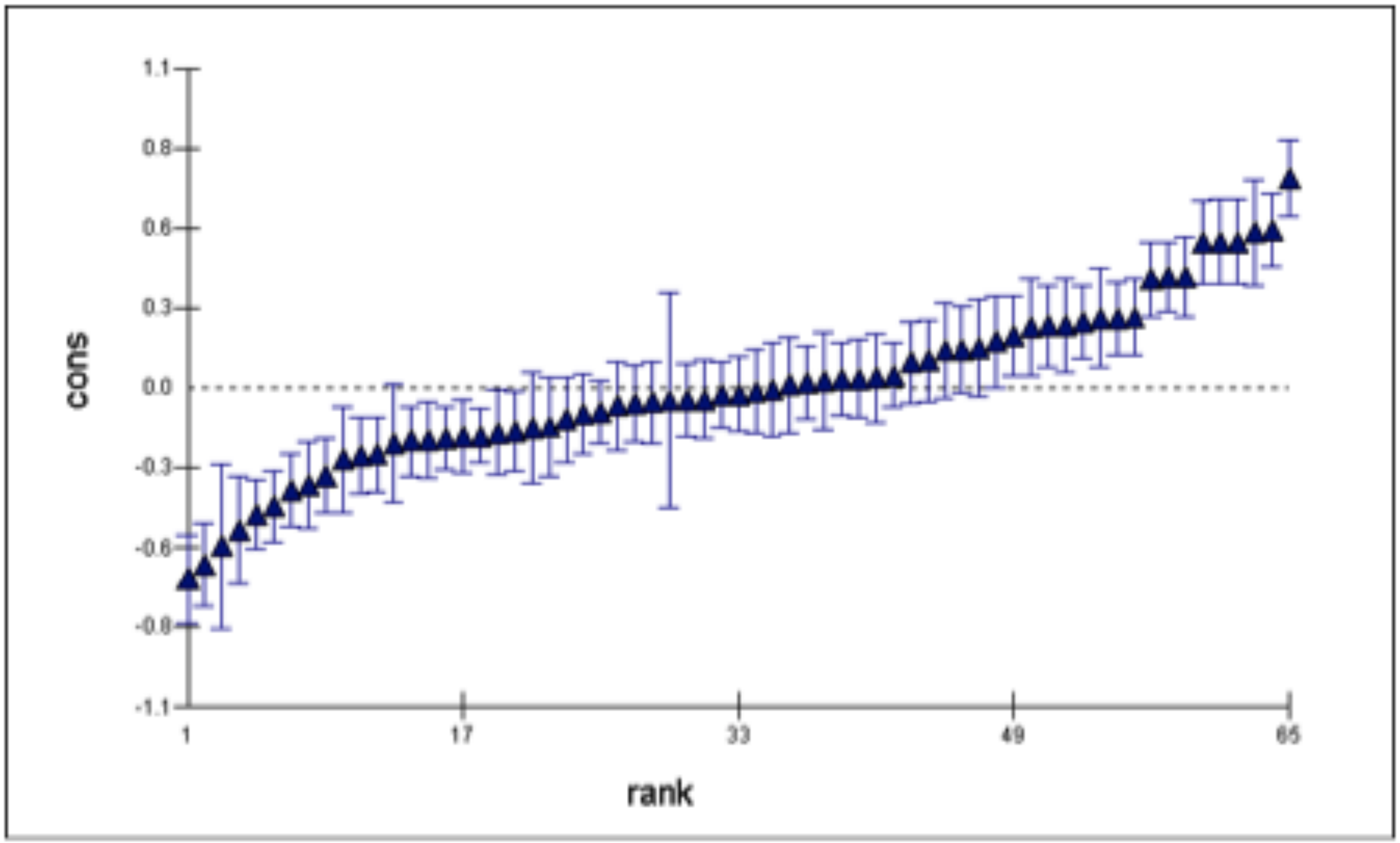
- Patterns of intercepts and slopes: random slopes model (2)



- Patterns of intercepts and slopes: random slopes model (3)



Plot of school level residuals with their confidence intervals: pupil level variations and group size accounted for.



Conclusion: what is multilevel modelling?

- Way to investigate variations and relationships for variables of interest, taking into account population structure and dependencies, even if these not of primary substantive interest.
- Flexible framework for testing sophisticated social (or other) theories, looking at individuals in context, including change over time.

Software

- MLwiN
- HLM
- R - e.g. lme4 package
- Stata - e.g xtreg, xtlogit
- SAS - e.g. Proc Mixed
- SPSS (limited range of models)

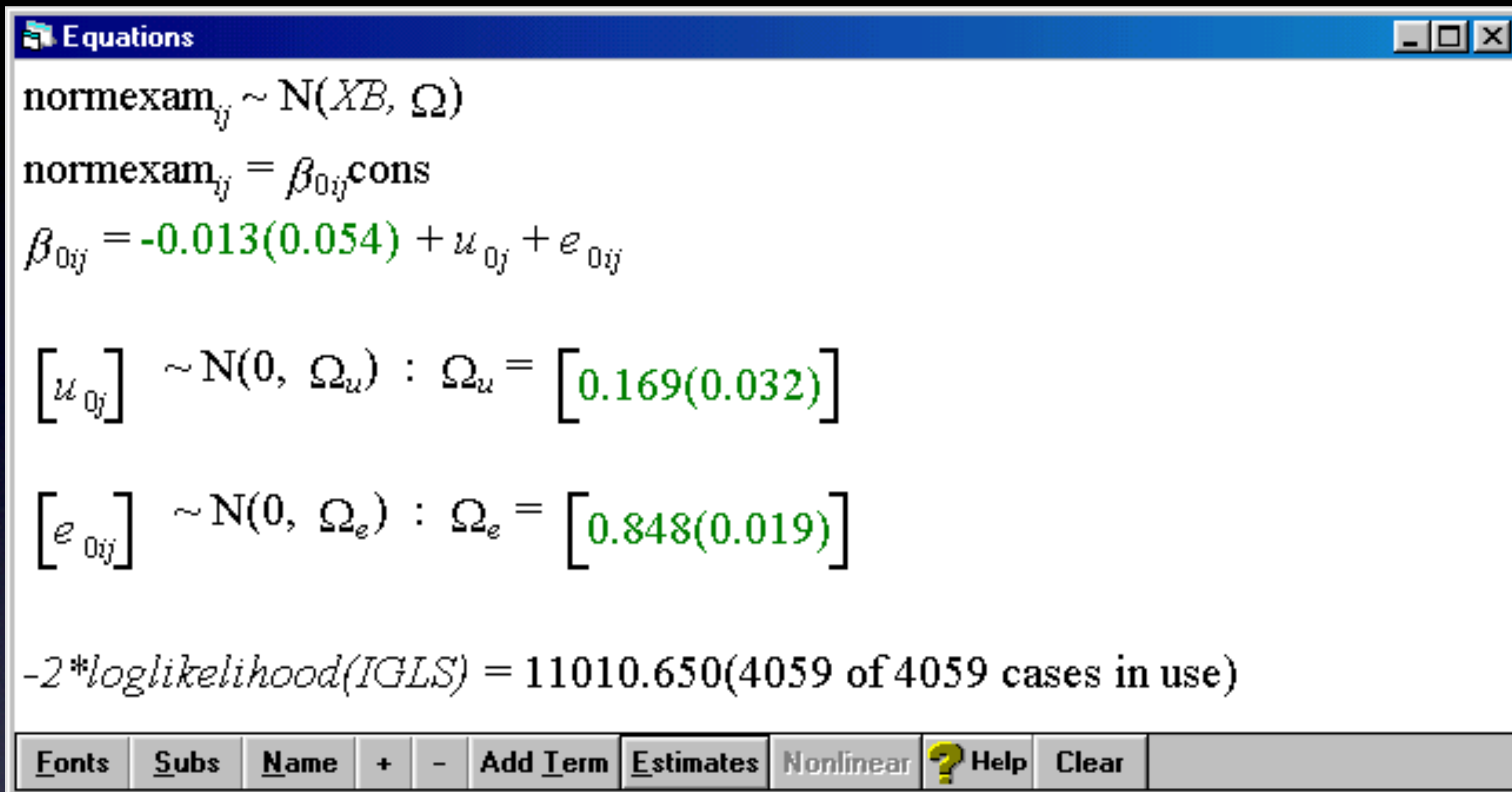
Web / Books

- http://www.cmm.bristol.ac.uk
- Snijders and Bosker (1999) ‘Multilevel Analysis’. Sage.
- Goldstein * (2003)
- Luke D (2008) Multilevel Models. Sage.
- Singer and Willett (2003) Applied Longitudinal Data Analysis.

Courses

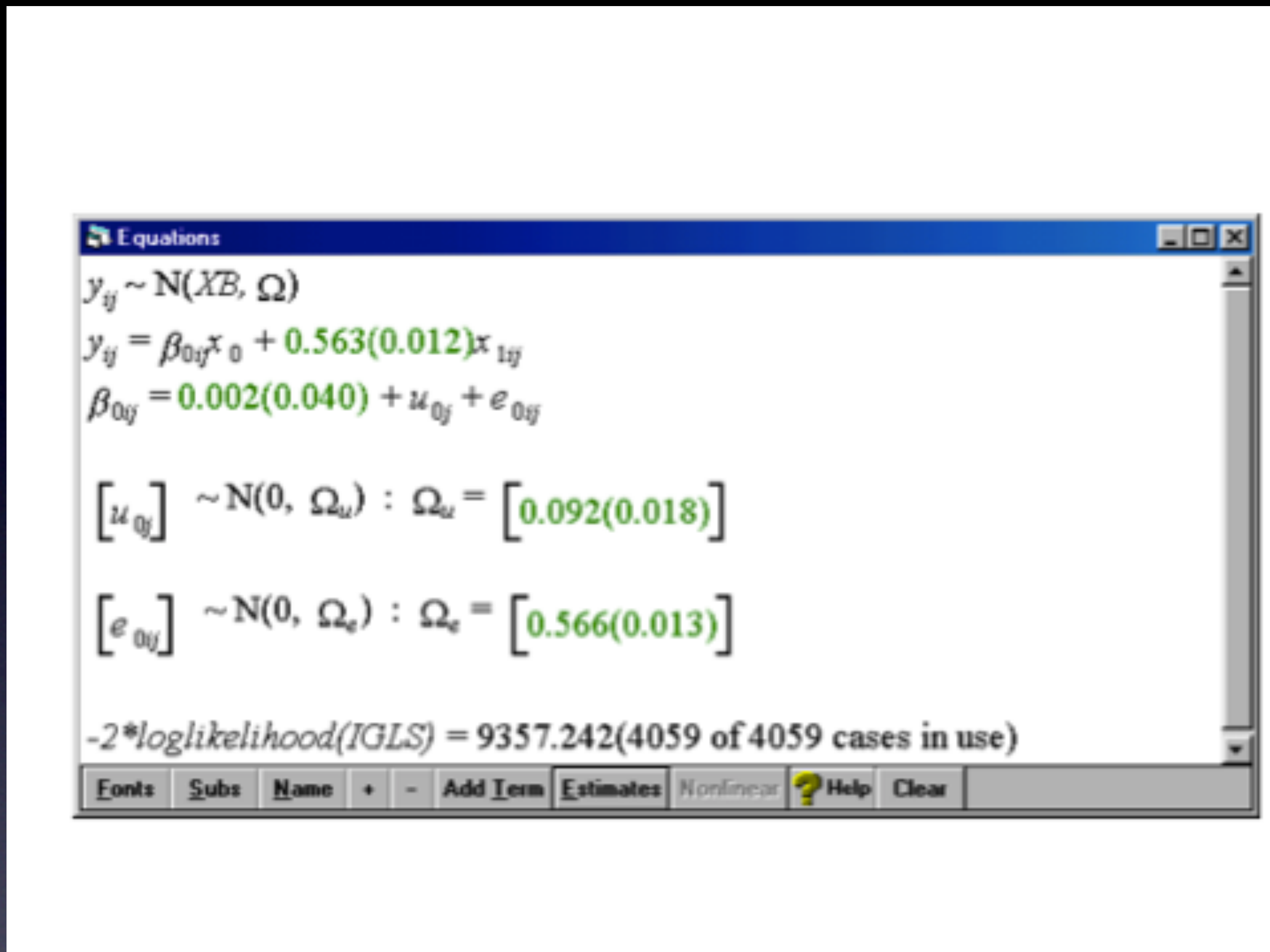
- Here are just two places where you can learn more:
- CCSR / Social Statistics (Manchester) - MLM, Longitudinal, Social Networks
- Bristol - Comprehensive courses based on use of MLwiN; multilevel event history analysis etc.
- *Acknowledgement: graphs from slides 25 to 43 adapted from CMM Bristol teaching materials.*

Example: variance components model



- school level variance = 0.169
- pupil level variance = 0.848
- intra class correlation = (0.169) / (0.169 + 0.848) = 0.166
- About 17% of variation of exam score at school level

Example: random intercepts model



- positive coefficient on x - previous exam score - makes sense
- some of the variation explained at each level, but not all.

Example: random slopes model

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

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

$$\beta_{0ij} = -0.012(0.040) + u_{0j} + e_{0ij}$$

$$\beta_{1j} = 0.557(0.020) + u_{1j}$$

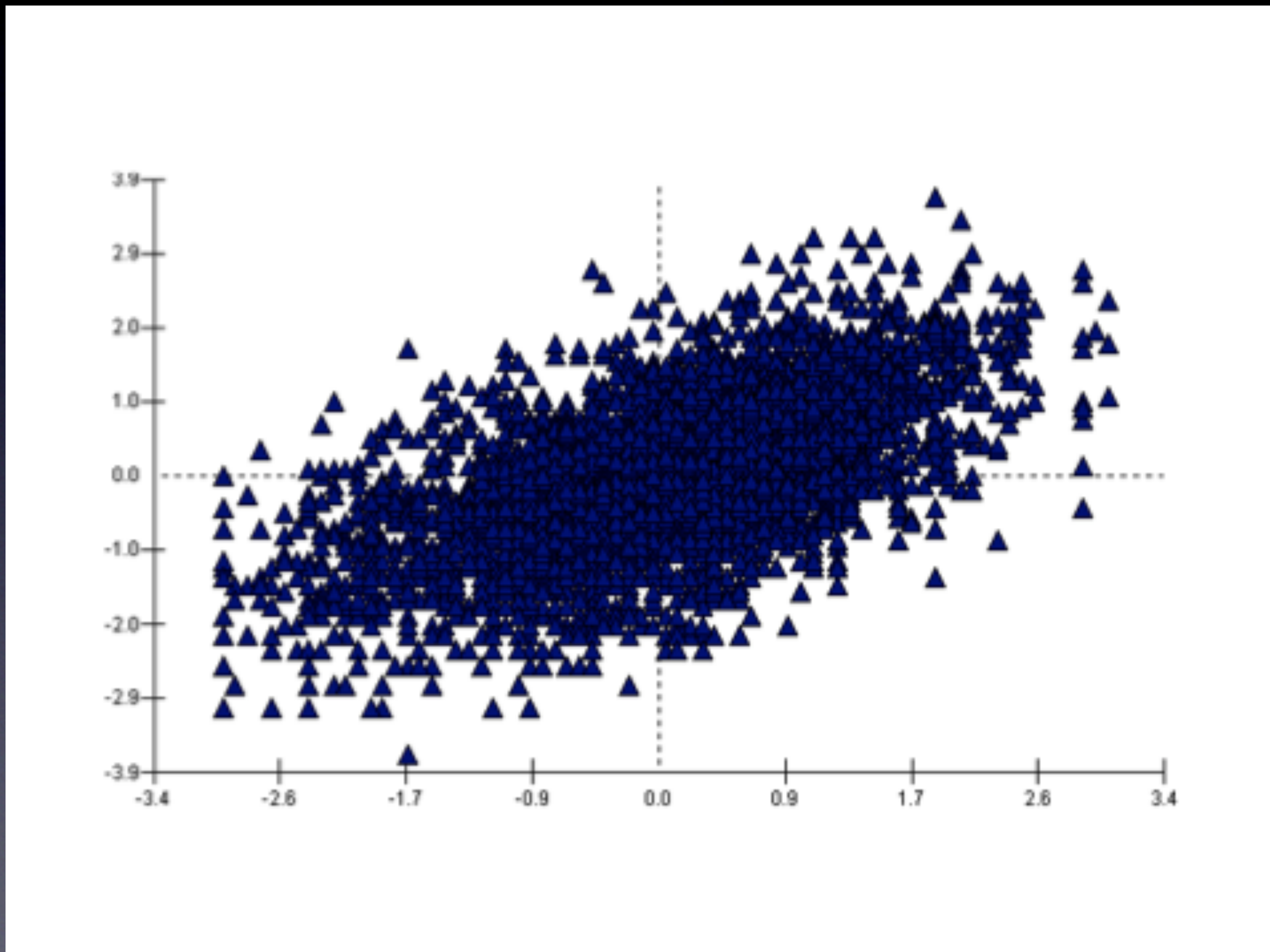
$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_u) \ : \ \Omega_u = \begin{bmatrix} 0.090(0.018) & \\ 0.018(0.007) & 0.015(0.004) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_e) \ : \ \Omega_e = \begin{bmatrix} 0.554(0.012) \end{bmatrix}$$

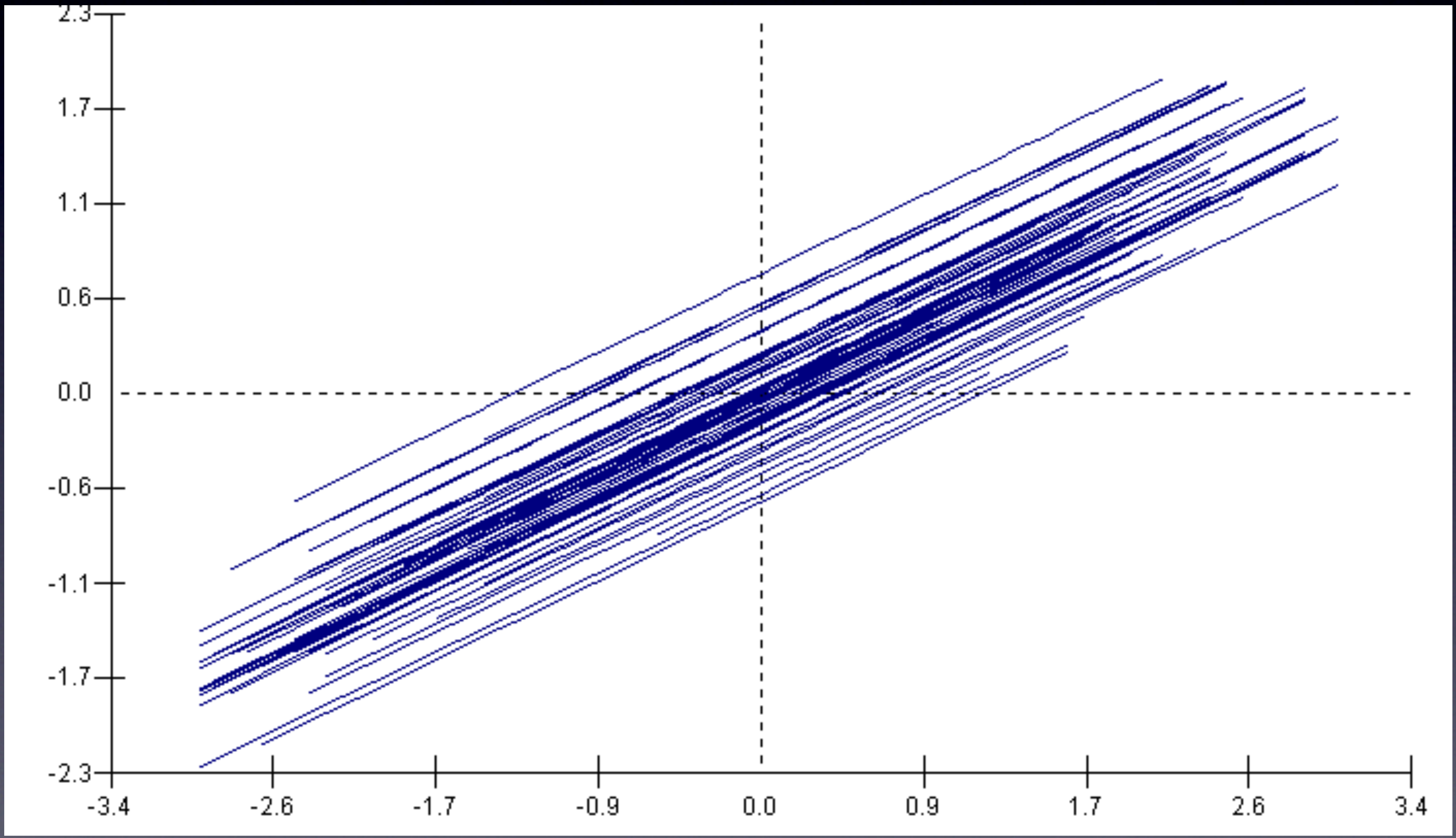
$$-2{*}loglikelihood(IGLS) = 9316.870(4059 \text{ of } 4059 \text{ cases in use})$$

- some school level variation in intercepts and slopes
- positive relationship between intercept and slope
- higher intercept tends to be associated with steeper slope

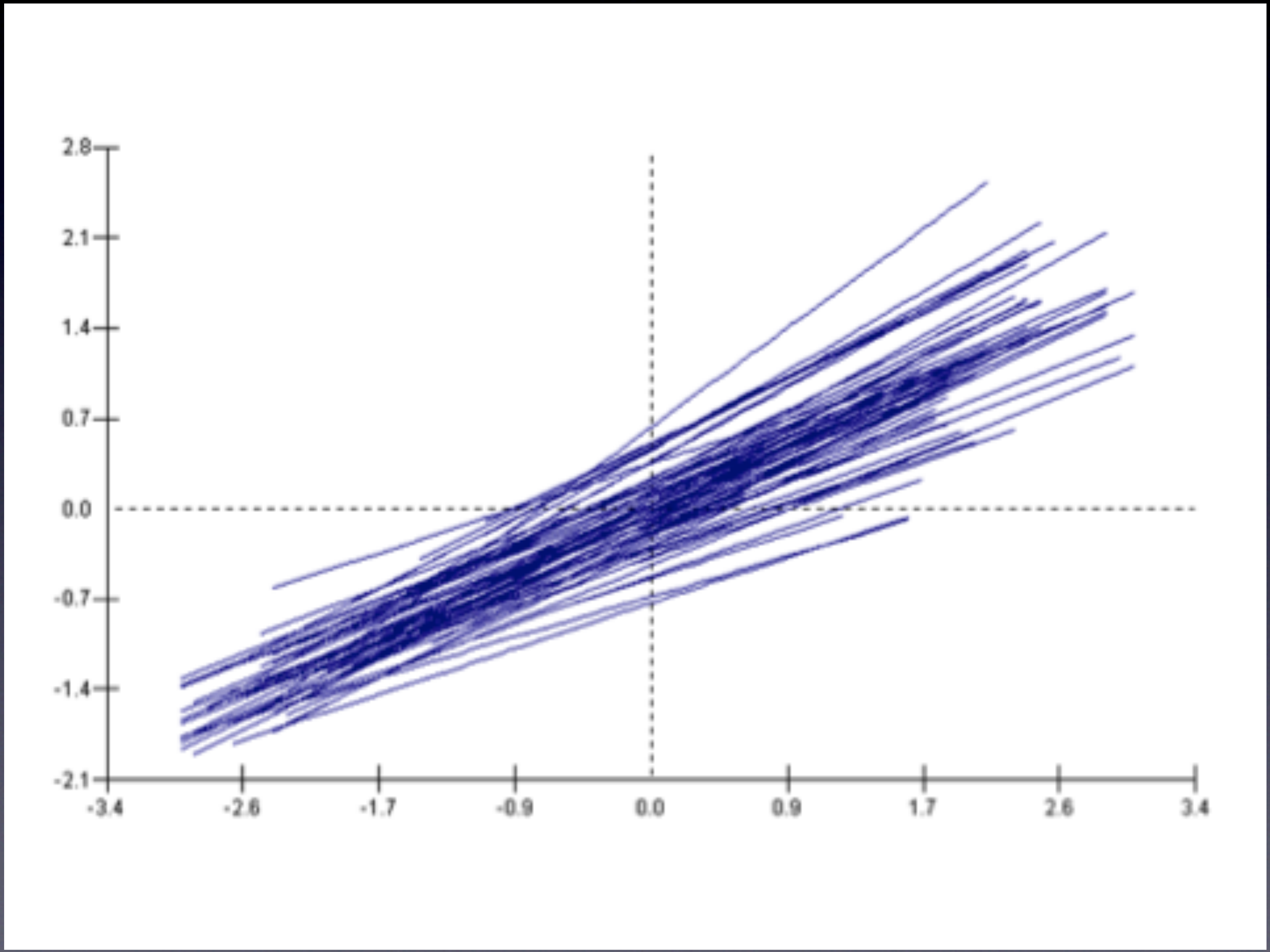
Scatterplot of exam 16 on exam 11 for 4059 pupils in 65 schools



Predicted lines from random intercepts model: 65 lines one for each school



Predicted lines from random slopes model: 65 lines one for each school



Plot of school level residuals with their confidence intervals: pupil level variations and group size accounted for.