Random Slope Models

Hedonism example

Our questions in the last session

- Do differences between countries in hedonism remain after controlling for individual age?
- How much of the variation in hedonism is due to individual age?
- What is the relationship between an individual's hedonism and their age?

Our questions in this session

- Are there differences between countries in the relationship between an individual's hedonism and their age?
- How does the amount of variation in hedonism due to country differences change as a function of individual age?
- How does the proportion of variation in hedonism due to country differences change as a function of individual age?

Exam scores example

Our questions in the last session

- Are there differences between schools in pupils' progress between age 11 and 16?
- How much of the variation in pupils' progress between age 11 and 16 is due to school differences?

Our questions in this session

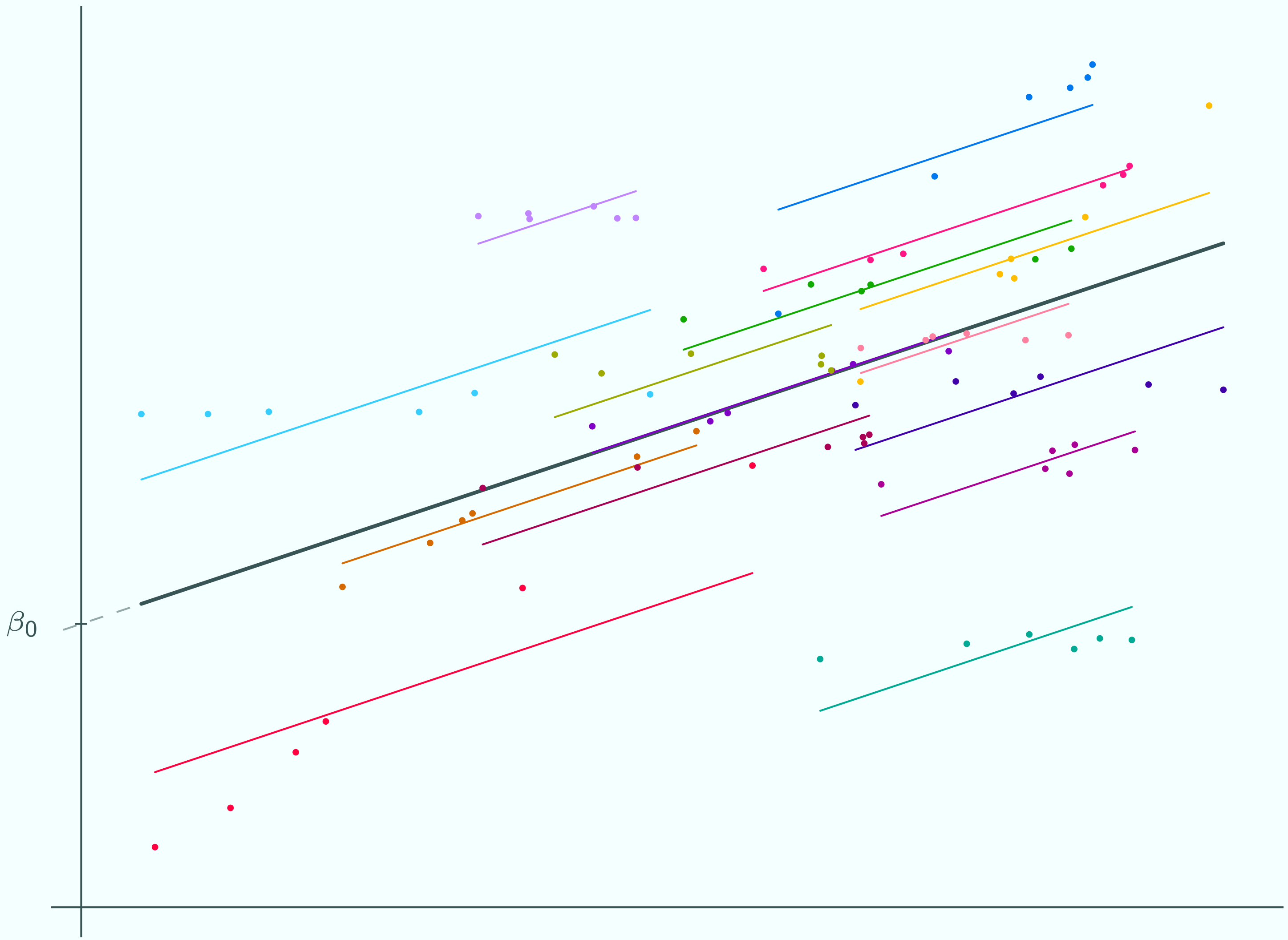
- Are there differences between schools in the relationship between a pupil's exam scores at age 11 and 16?
- Are there differences between schools in the relationship between a pupil's exam score age 16 and their gender?
- How does the amount of variation in exam scores at age 16 due to school differences change across exam score age 11?
- How does the proportion of variation in exam scores at age 16 due to school differences change across exam score age 11?

Allowing for different slopes between groups

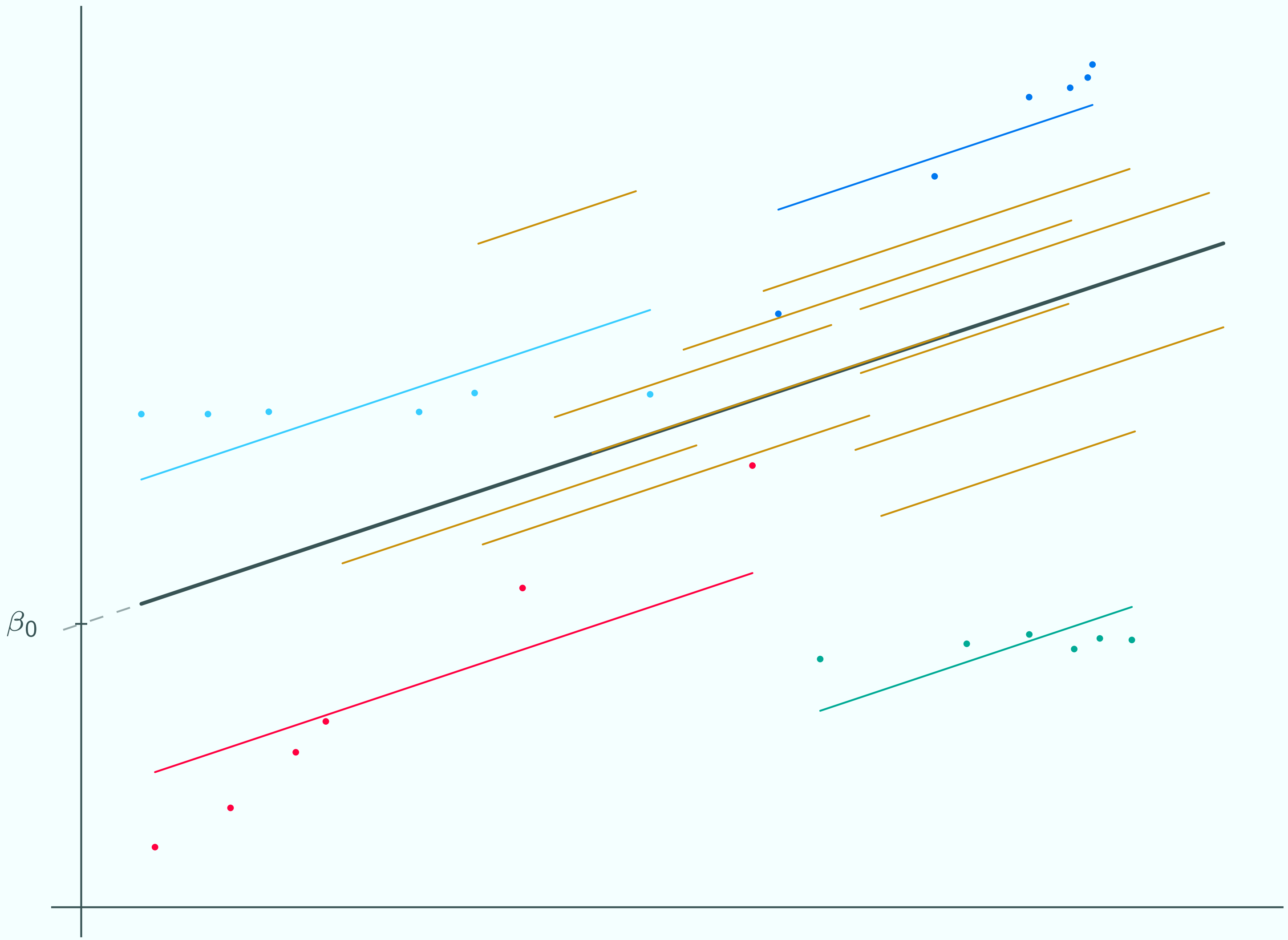
Group lines

- We have seen how random intercept models allow us to include explanatory variables
- We saw that, just like variance components models, each group has a line
- We also saw that the group lines all have the same slope as the overall regression line
  - Recall that for the variance components model all lines were flat i.e. they had slope 0
- So in every group, the relationship between the explanatory variable and the response is the same
- This is one of the assumptions of the random intercept model
- However, sometimes the effect of the explanatory variable may differ from group to group and this may be of interest

A possible situation



A possible situation



Questioning the assumption

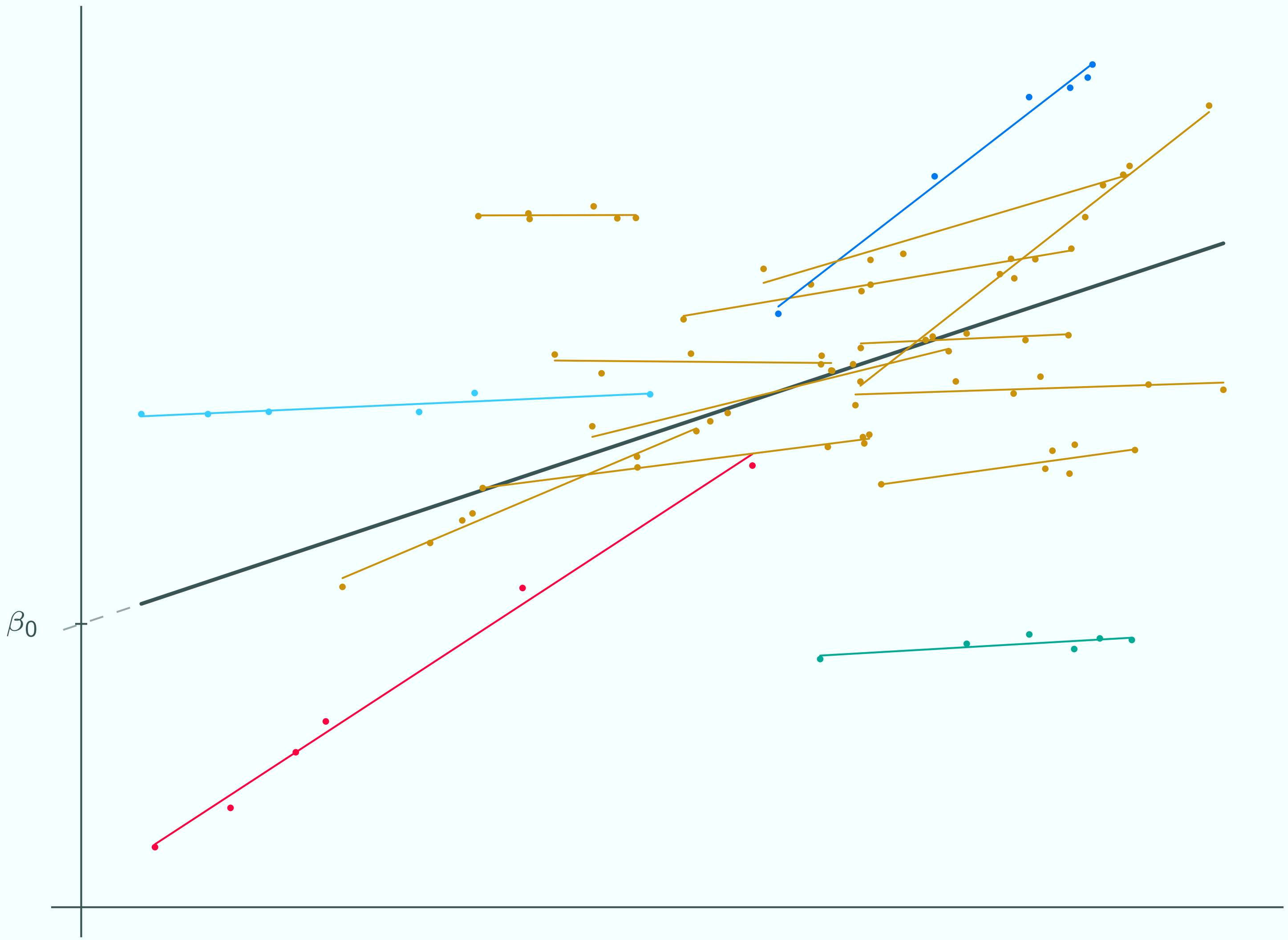
For this data,

- For some groups, the explanatory variable has a large effect on the response; for others it has a small effect
- Clearly the random intercepts model, with its parallel group lines, is not doing a very good job of fitting the data

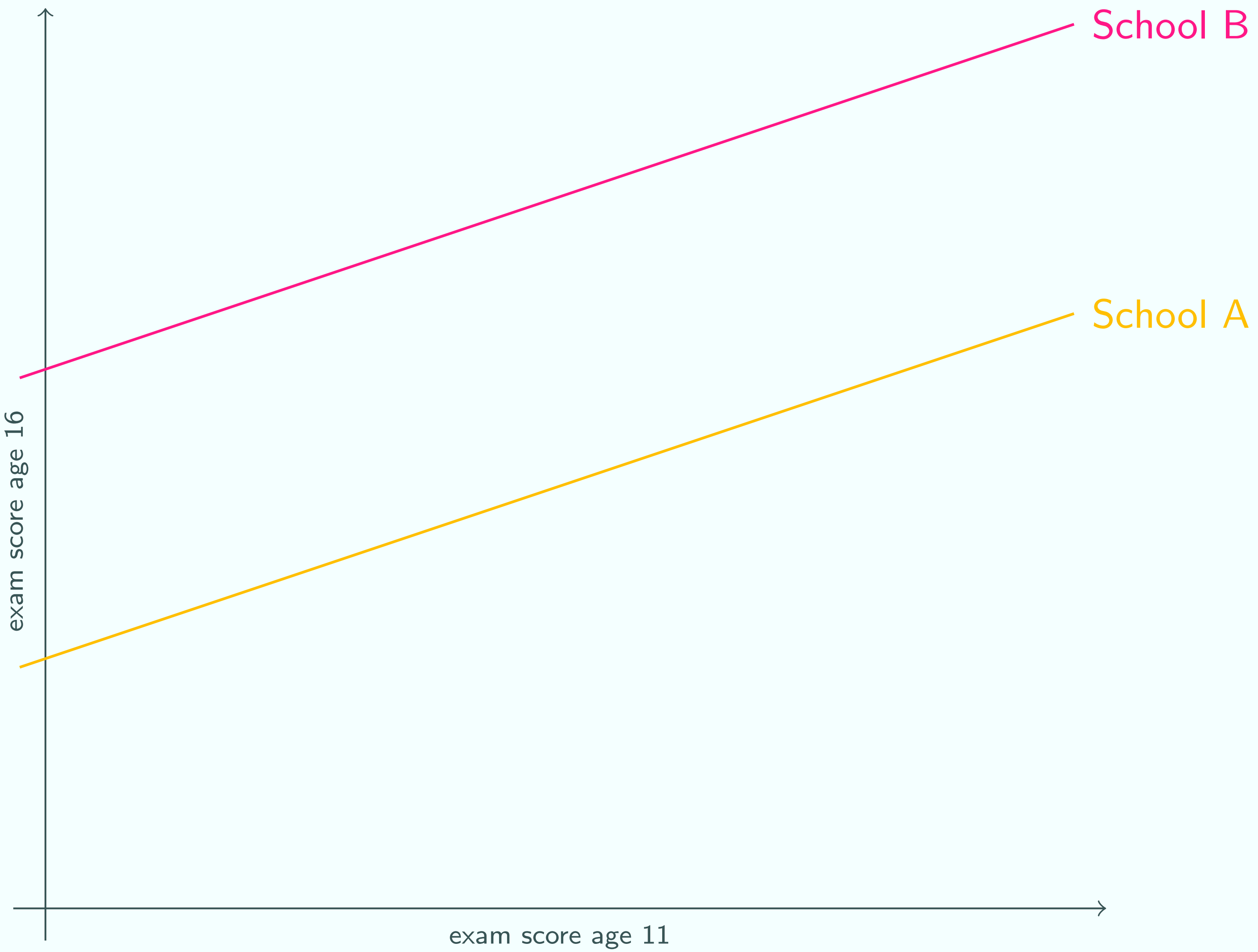
Real examples

- Some investigators have found that data for pupils within schools (response: exam score; explanatory variable: previous exam score) behaves like this:
  - for some schools pretest has a large effect on the response
  - while for others the effect is smaller
- Others have found that this is not the case and that the random intercepts model is an adequate fit to the data
  - For some datasets there is only enough power to fit a random intercepts model in any case

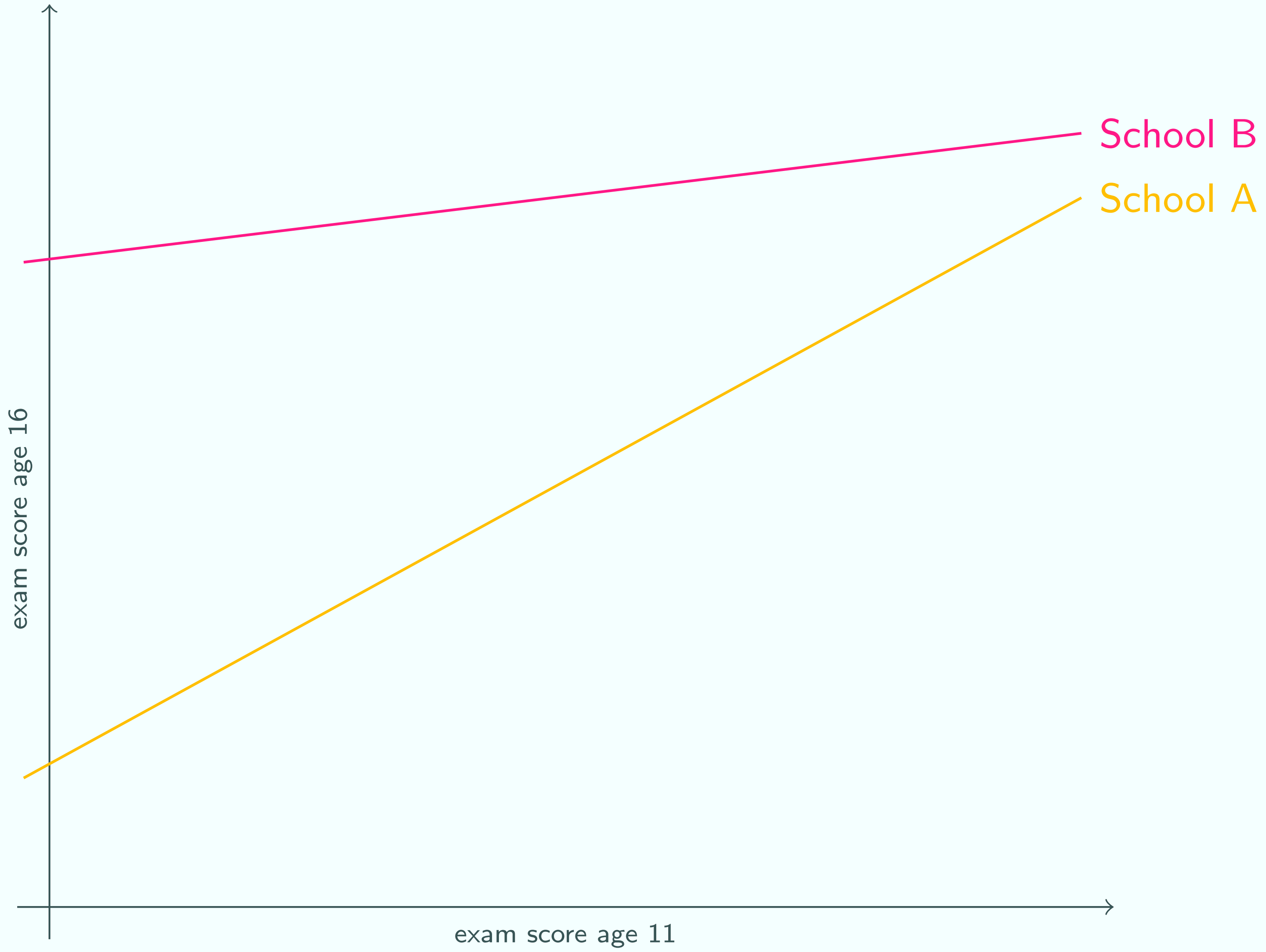
Solution: Random Slopes Model



Using dummy variables



Using dummy variables



Solution: Random Slopes Model

Difference from a random intercept model

- Unlike a random intercept model, a random slope model allows each group line to have a different slope
- So the random slope model allows the explanatory variable to have a different effect for each group

How do we achieve this?

By adding a random term to the coefficient of x_{1ij}, so it can be different for each group:

$$y_{ij} = \beta_0 + (\beta_1 + u_{1ij})x_{1ij} + u_{0j} + e_{0ij}$$

Rearrange to give:

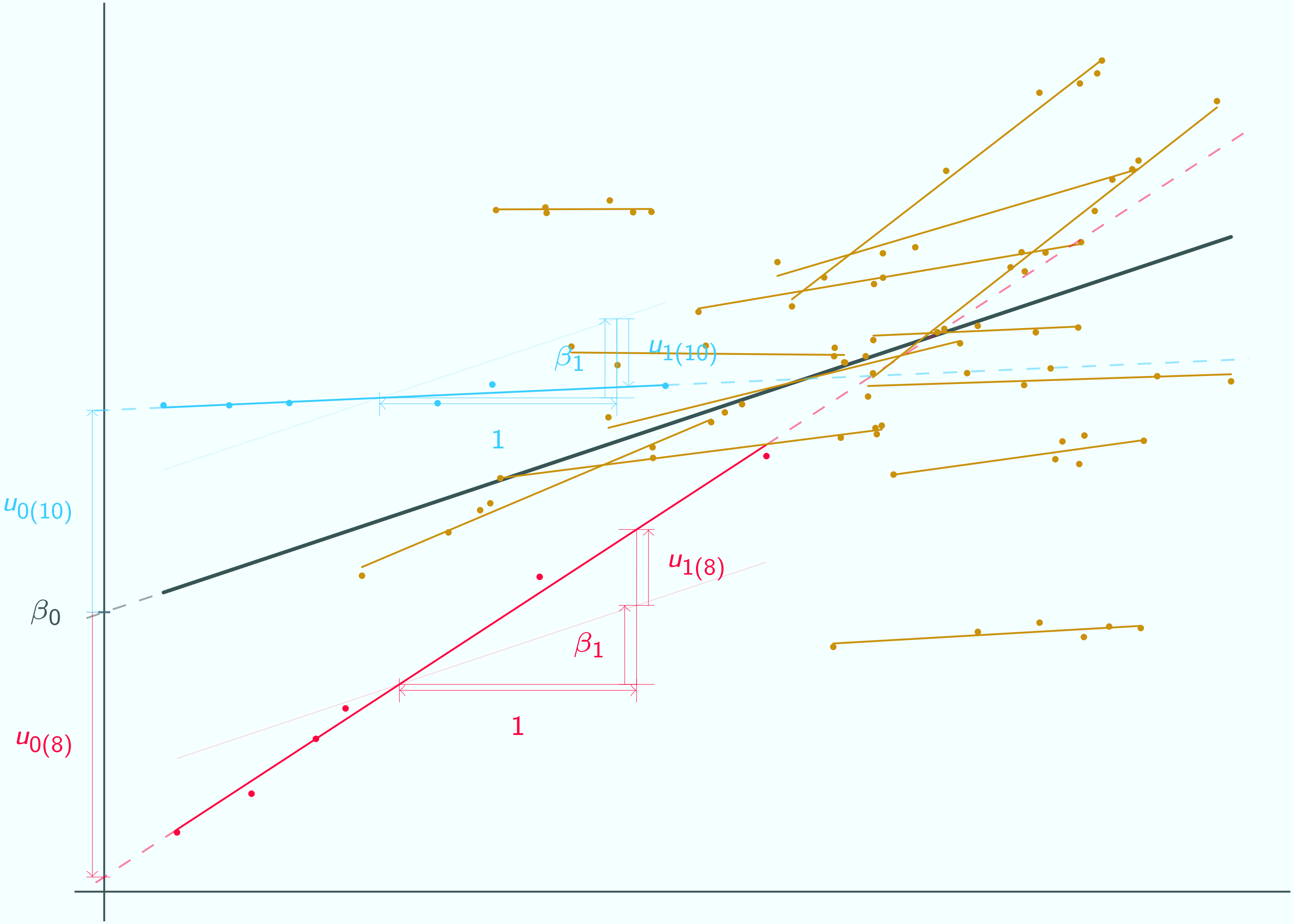
$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + u_{0j} + u_{1j}x_{1ij} + e_{0ij}$$

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

$$e_{0ij} \sim \mathrm{N}(0, \sigma^2_{e0})$$

Note that although we have only introduced one extra thing into our model, u_{1j}, we have 2 extra parameters, σ^2_{u1} and σ_{u01}. We will come back to this shortly.

Solution: Random Slopes Model



Interpreting the parameters

$\beta_0, \beta_1, \sigma_e^2$

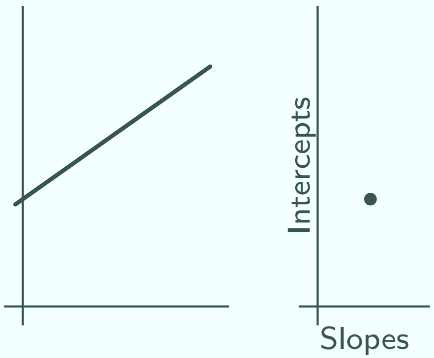
- β_0 and σ_e^2 can be interpreted as for the random intercepts model
- β_1 is the slope of the average line: the average increase (across all groups) in y for a 1 unit change in x_1

$\sigma_{u0}^2, \sigma_{u1}^2, \sigma_{u01}$

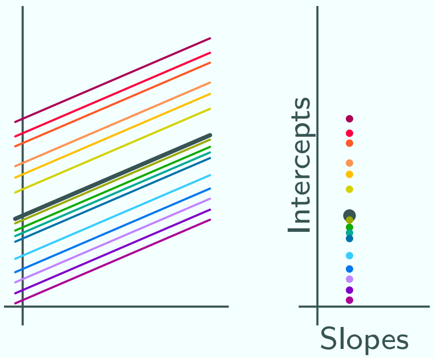
- Interpretation of these parameters is a bit more complicated
  - σ_{u1}^2 is the variance in slopes between groups
  - σ_{u0}^2 is the variance in intercepts between groups (and the level 2 variance at $x_1 = 0$)
  - σ_{u01} is the covariance between intercepts and slopes
- BUT the estimates of σ_{u1}^2 and σ_{u0}^2 are not very meaningful in themselves.
- We will explain why that is after having a look at what the 'covariance between intercepts and slopes' means

Covariance between intercepts and slopes

Single level model

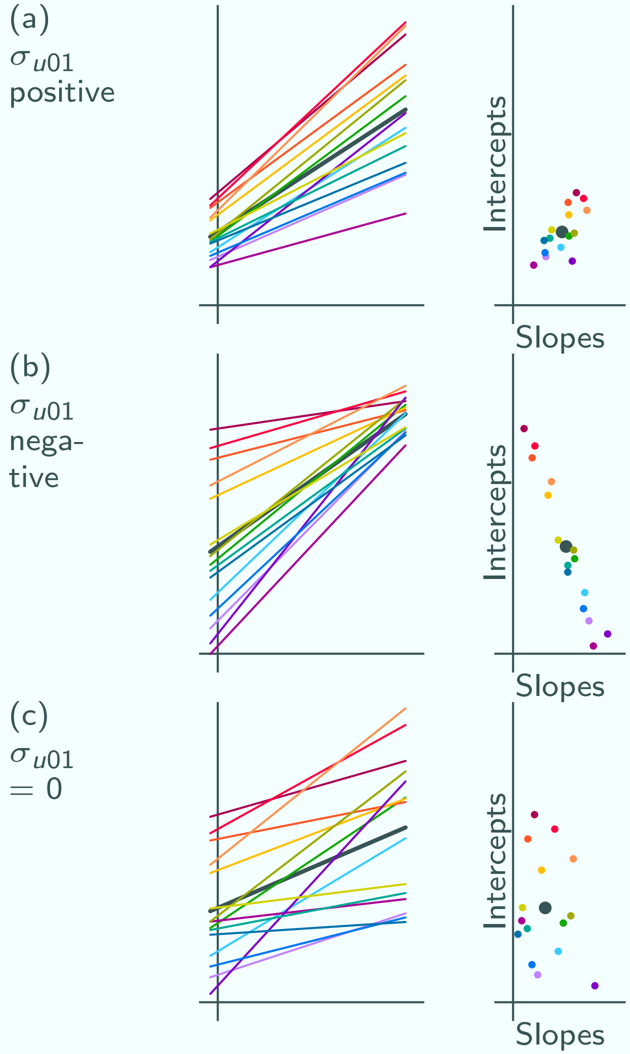


Random int. model



- 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

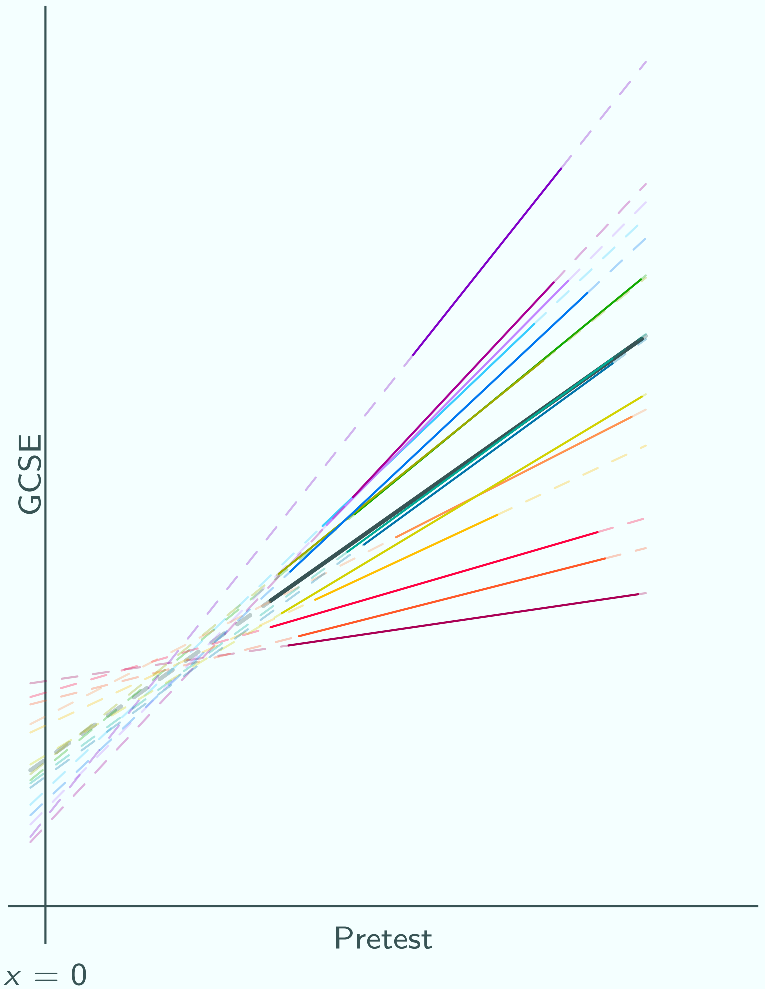
Random slopes model



σ_{u01} and the scale of x

Example

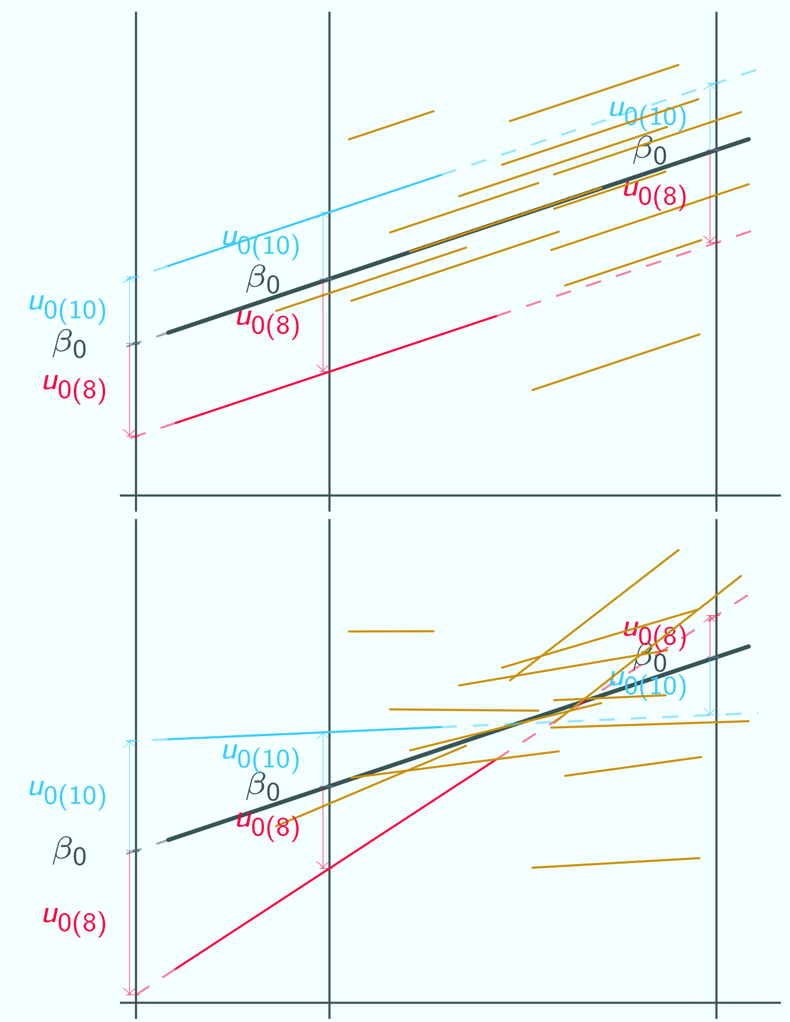
- We fit a random slopes model:
  - response: **GCSE**
  - explanatory variable: previous exam score (%)
  - units: students within schools
- If we only look at the value of σ_{u01} we will think the pattern is of fanning in
- Actually over the range of our data the pattern is of fanning out
- We can see this if we look at the graph



$\sigma_{u01} < 0$

u_{0j} and the scale of x

- For a random intercepts model, where $x = 0$ occurs makes no difference to the value of u_{0j}
- For a random slopes model, it makes no difference to the value of u_{1j}, but it does make a difference to the value of u_{0j}
- The variance σ^2_{u0} will also be affected
- as will the covariance σ_{u01}
- This is why we have to interpret σ^2_{u1}, σ^2_{u0} and σ_{u01}
  - together
  - and in light of where we have put $x = 0$



Hypothesis testing

Hypothesis testing is the same as for the random intercept model

Fixed part

β_k is significant at the 5% level if $|z_k| \geqslant 1.96$

Random part

We use a likelihood ratio test

- Fit the model with $u_{1j}x_{1ij}$ (①)
- and without $u_{1j}x_{1ij}$ (⓪)

- In other words we are comparing the random slope model to a random intercept model
- The test statistic is again 2(log(likelihood(①)) − log(likelihood(⓪)))
- This time there are 2 degrees of freedom because there are 2 extra parameters in ① compared to ⓪
- So we compare the test statistic against the $\chi^2_{(2)}$ distribution
- The null hypothesis is that σ^2_{u1} and σ_{u01} are both 0 and hence that a random intercept model is more appropriate than a random slope model

Exam scores example

Question

Are there differences between schools in the relationship between a pupil's exam scores at age 11 and 16?

Answer

1. Fit a model with a random slope on exam score age 11 and note the $-2 \times \log(\text{likelihood})$ value: 9316.870
2. Fit a model without a random slope on exam score age 11 and note the $-2 \times \log(\text{likelihood})$ value: 9357.242
3. Calculate the test statistic: $9357.242 - 9316.870 = 40.372$
4. Compare to the χ^2 distribution with 2 degrees of freedom $p = 1.7113 \times 10^{-9}$
5. We conclude that there are differences between schools in the relationship between a pupil's exam scores at age 11 and 16

Exam scores example

1

Equations

$$\text{normexam}_{ij} = \beta_{0j} + \beta_{1j}\text{standlrt}_{ij} + e_{ij}$$

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

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

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

$$e_{ij} \sim \text{N}(0, \sigma_e^2) \quad \sigma_e^2 = 0.554(0.012)$$

*-2*loglikelihood* = 9316.870(4059 of 4059 cases in use)

Name | + | - | Add Term | Estimates | Nonlinear | Clear | Notation | Responses | Store | Help | Zoom 100

2

Equations

$$\text{normexam}_{ij} = \beta_{0j} + 0.563(0.012)\text{standlrt}_{ij} + e_{ij}$$

$$\beta_{0j} = 0.002(0.040) + u_{0j}$$

$$u_{0j} \sim \text{N}(0, \sigma_{u0}^2) \quad \sigma_{u0}^2 = 0.092(0.018)$$

$$e_{ij} \sim \text{N}(0, \sigma_e^2) \quad \sigma_e^2 = 0.566(0.013)$$

*-2*loglikelihood* = 9357.242(4059 of 4059 cases in use)

Name | + | - | Add Term | Estimates | Nonlinear | Clear | Notation | Responses | Store | Help | Zoom 100

Exam scores example

Question

Are there differences between schools in the relationship between a pupil's exam score at age 16 and their gender?

Answer

1. Fit a model with a random slope on gender and note the $-2 \times \log(\text{likelihood})$ value: 10967.750
2. Fit a model without a random slope on gender and note the $-2 \times \log(\text{likelihood})$ value: 10968.689
3. Calculate the test statistic: $10968.689 - 10967.750 = 0.939$
4. Compare to the χ^2 distribution with 2 degrees of freedom $p = 0.63$
5. We conclude that there are no differences between schools in the relationship between a pupil's gender and their exam score at age 16

Exam scores example

1

Equations

$$\text{normexam}_{ij} = \beta_{0j} + \beta_{1j}\text{girl}_{ij} + e_{ij}$$

$$\beta_{0j} = -0.164(0.054) + u_{0j}$$

$$\beta_{1j} = 0.267(0.035) + u_{1j}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim \text{N}(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0.146(0.033) & \\ 0.017(0.014) & -0.009(0.008) \end{bmatrix}$$

$$e_{ij} \sim \text{N}(0, \sigma_e^2) \quad \sigma_e^2 = 0.841(0.019)$$

-2*loglikelihood = 10967.750(4059 of 4059 cases in use)

Name | + | - | Add Term | Estimates | Nonlinear | Clear | Notation | Responses | Store | Help | Zoom 100

2

Equations

$$\text{normexam}_{ij} = \beta_{0j} + 0.262(0.040)\text{girl}_{ij} + e_{ij}$$

$$\beta_{0j} = -0.161(0.057) + u_{0j}$$

$$u_{0j} \sim \text{N}(0, \sigma_{u0}^2) \quad \sigma_{u0}^2 = 0.161(0.031)$$

$$e_{ij} \sim \text{N}(0, \sigma_e^2) \quad \sigma_e^2 = 0.839(0.019)$$

-2*loglikelihood = 10968.689(4059 of 4059 cases in use)

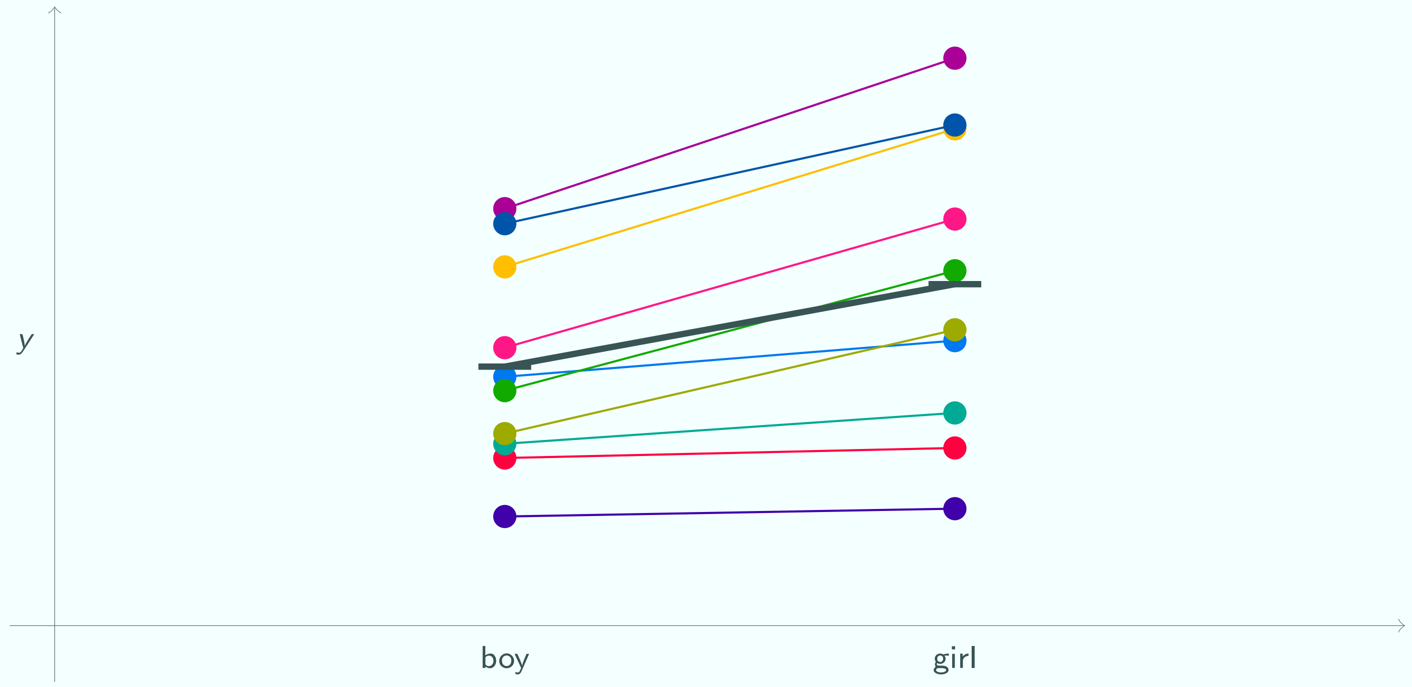
Name | + | - | Add Term | Estimates | Nonlinear | Clear | Notation | Responses | Store | Help | Zoom 100

Negative variances

- In the second example, we had to allow negative variances in MLwiN
- When we did this, we got an negative estimate for the variance of the random slopes
- This seems strange because a variance can't be negative
- In this case, the variance (and covariance) were not significant, so we don't need to worry
- In general, it is actually possible to get a significant and negative variance
- This actually does make sense due to the second interpretation of random slope models which we will see later
- Sometimes the final estimated variance is not negative but it goes negative during estimation so we need to allow negative variances

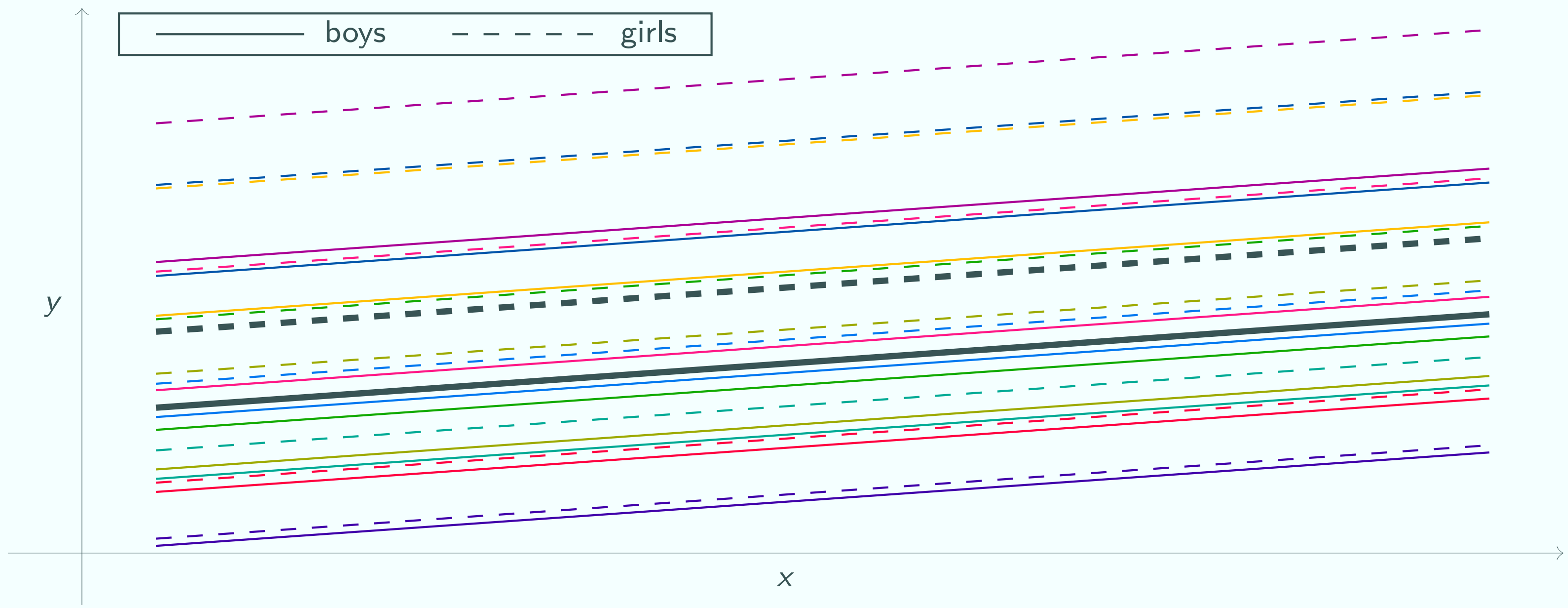
Random slopes and categorical variables

- It is possible to put a random slope on a categorical variable such as gender
- Often called a random coefficient rather than random slope
- Random slopes on continuous variables can also be called random coefficients



Random slopes and categorical variables

Model with a continous explanatory variable and gender; random coefficient on gender only; plotting against the continuous variable

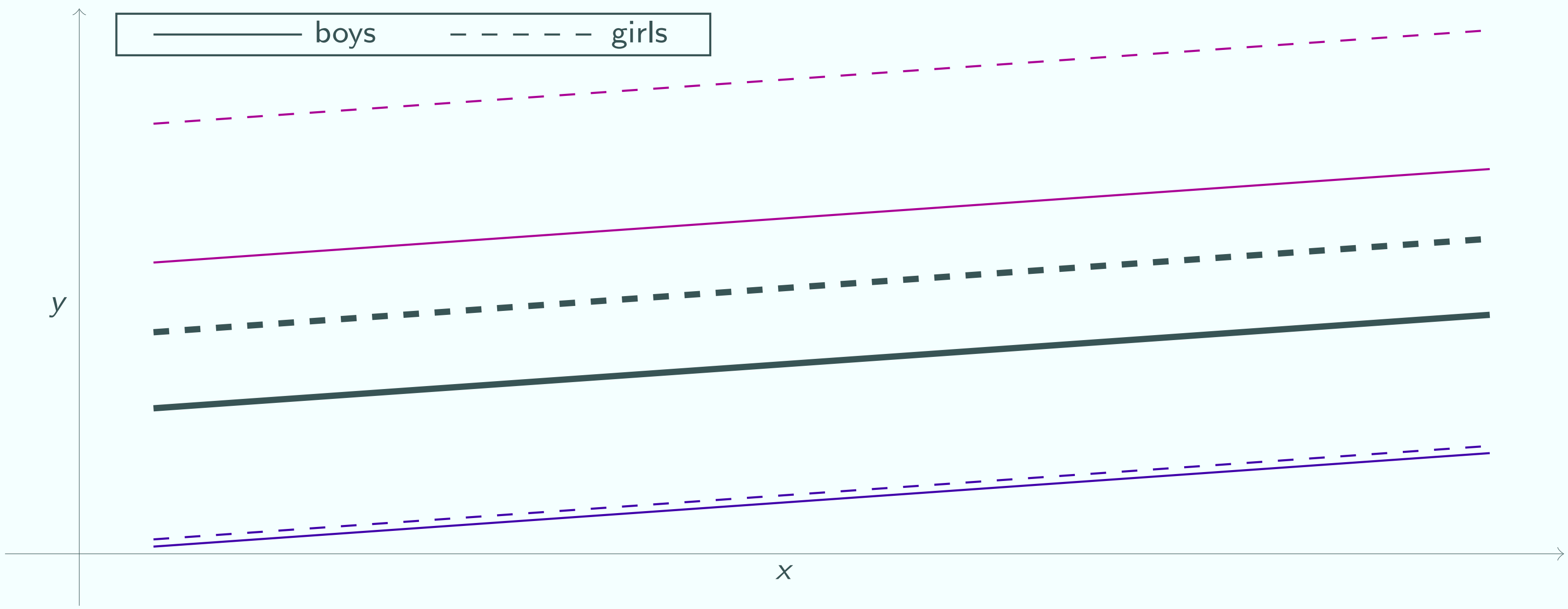


- Random coefficient means distance between the line for girls and the line for boys differs from group to group
- In other words, difference between boys' and girls' predicted values differs from group to group

Random slopes and categorical variables

Model with a continous explanatory variable and gender; random coefficient on gender only; plotting against the continuous variable

Picking out 2 groups for clarity



- Random coefficient means distance between the line for girls and the line for boys differs from group to group
- In other words, difference between boys' and girls' predicted values differs from group to group

Examples of research questions

Clark et al. (1999)

Is there a large amount of variability between subjects in the rate of change in MMSE score?

Levels:	2	subject	Random slope on:	year
	1	occasion	Answer:	Yes

Correlation between slopes and intercepts was 0.33 so subjects with higher intercepts have less decline in MMSE

MMSE is used to assess patients diagnosed with Alzheimer's and the authors were interested in whether it is a good measure. They decided not, largely because of the variability in slope between subjects

Tymms et al. (1997)

Does the effect of attainment at the start of school on attainment after the first year of school vary across schools?

Levels:	2	school	Random slope on:	pretest
	1	pupil	Answer:	Yes

The authors comment that the variability in slopes that they find could be due to ceiling effects of the post-test

Examples of research questions

Polsky and Easterling (2001)

Do districts vary in their sensitivity of land value to climate (mean maximum July temperature over 30 years, JULTMX)?

Levels:	2	district	Random slope on:	JULTMX
	1	county	Answer:	Yes

The authors go on to fit a model that shows that counties in districts with more variability in temperature from year to year benefit more from high July temperatures

Jex and Bliese (1999)

Is there variability across army companies in the relationship between hours worked and psychological strain?

Levels:	2	company	Random slope on:	hours worked
	1	soldier	Answer:	Yes

The authors go on to examine whether the variabilty could be explained by differing beliefs in the efficacy of the company across companies, but conclude that it cannot

See also the Gallery of Multilevel Papers on the CMM website!

Calculating the total variance

Level 1

- We only have one random term at level 1, e_{0ij}
- So the level 1 variance is easy to calculate: it is σ^2_{e0}

Level 2

- We have two random terms at level 2: u_{0j} and $u_{1j}x_{1ij}$
- So the level 2 variance is

$$\begin{aligned}\mathrm{Var}(u_{0j} + u_{1j}x_{1ij}) &= \mathrm{Var}(u_{0j}) + 2\mathrm{Cov}(u_{0j}, u_{1j}x_{1ij}) + \mathrm{Var}(u_{1j}x_{1ij}) \\ &= \sigma^2_{u0} + 2\sigma_{u01}x_{1ij} + \sigma^2_{u1}x^2_{1ij}\end{aligned}$$

- Notice the level 2 variance is now a quadratic function of x_{1ij}

The variance partitioning coefficient now also depends on x_{1ij}

$$\mathrm{VPC} = \frac{\text{level 2 variance}}{\text{total residual variance}} = \frac{\sigma^2_{u0} + 2\sigma_{u01}x_{1ij} + \sigma^2_{u1}x^2_{1ij}}{\sigma^2_{u0} + 2\sigma_{u01}x_{1ij} + \sigma^2_{u1}x^2_{1ij} + \sigma^2_{e0}}$$

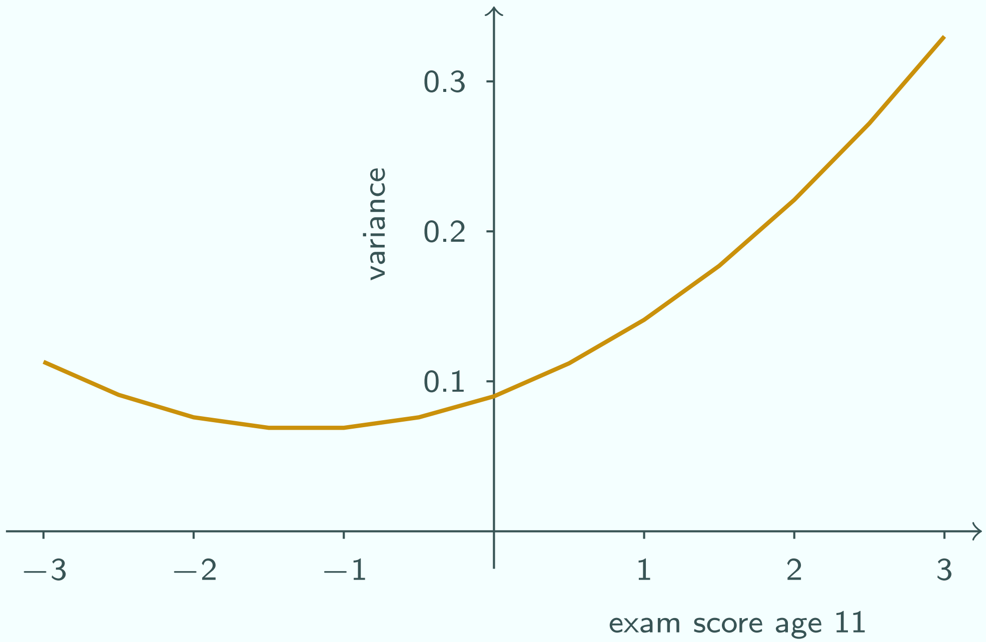
Exam scores example

Question

How does the amount of variation in exam scores at age 16 due to school differences change as a function of exam score age 11?

Answer

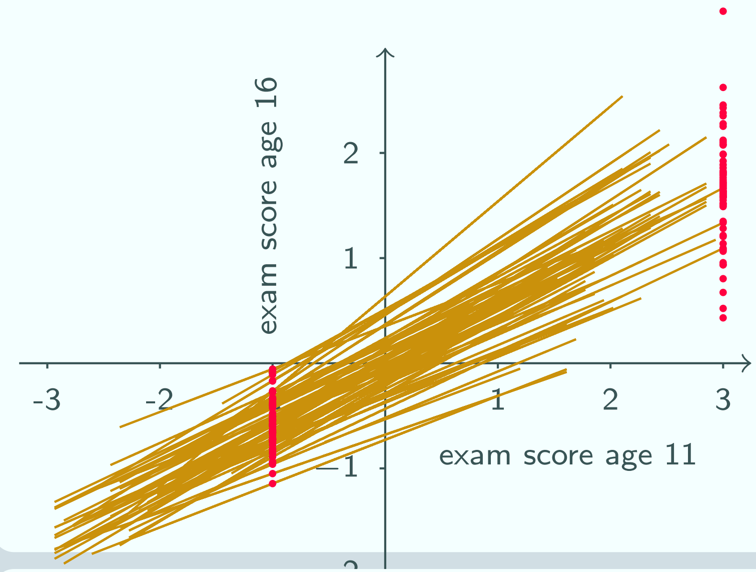
1. Fit a model with a random slope on exam score age 11
2. Calculate the level 2 variance
3. Plot:



Why does the variance depend on x?

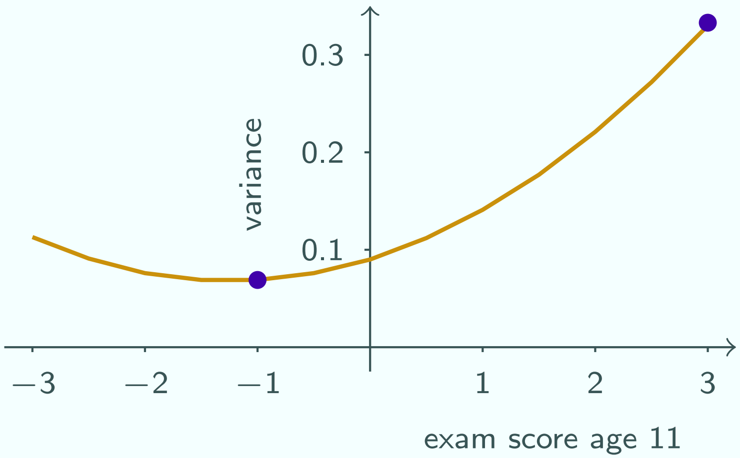
At $x = 3$

- The school lines are spread out
- There are greater differences between schools
- The school level variance is higher



At $x = -1$

- The school lines are closer together
- There are smaller differences between schools
- The school level variance is lower



Assumptions of random part

Random slope model

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + u_{0j} + u_{1j} x_{1ij} + e_{0ij}$$

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

$$e_{0ij} \sim \mathrm{N}(0, \sigma^2_{e0})$$

We have all the same assumptions as for the random intercept model, plus:

$$\mathrm{Cov}(u_{1j_1}, u_{1j_2}) = 0 \qquad \mathrm{Cov}(u_{0j_1}, u_{1j_1}) = \sigma_{u01} \qquad \mathrm{Cov}(u_{1j_1}, e_{0i_1j_1}) = 0$$

$$\mathrm{Cov}(u_{0j_1}, u_{1j_2}) = 0 \qquad \mathrm{Cov}(u_{1_{j_1}}, e_{0i_1j_2}) = 0$$

$$\mathrm{Cov}(u_{1j}, x_{1ij}) = 0$$

V, the correlation matrix

Random slope model

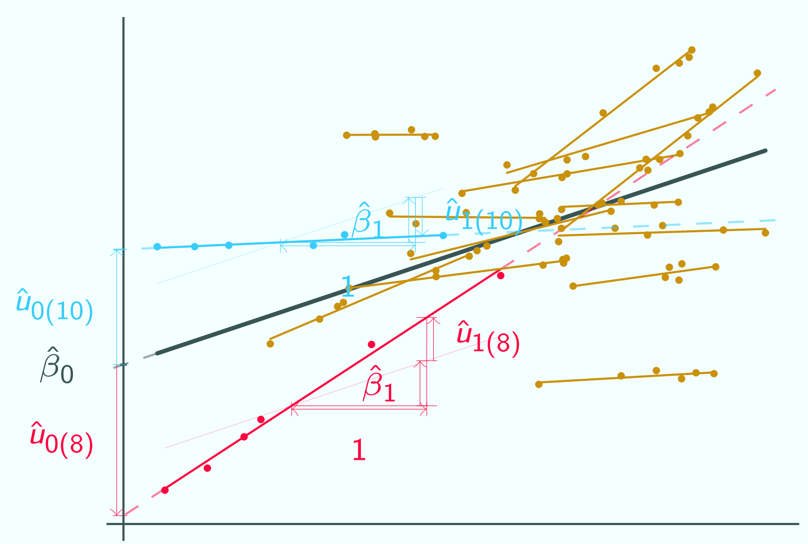
- Now $y_{ij} - \hat{y}_{ij} = u_{0j} + u_{1j}x_{1ij} + e_{0ij}$
- $\text{Cov}(y_{i_1j_1} - \hat{y}_{i_1j_1}, y_{i_2j_2} - \hat{y}_{i_2j_2}) =$
  - $\sigma^2_{u0} + 2\sigma_{u01}x_{1ij} + \sigma^2_{u1}x^2_{1ij} + \sigma^2_{e0}$
    for the same element $(i_1 = i_2 = i; j_1 = j_2 = j)$
  - 0 for two elements from different groups $(j_1 \neq j_2)$
- For a random intercept model, the intraclass correlation was identical to the variance partitioning coefficient
- For a random slopes model, it's not equal to the VPC:
- the intraclass correlation will depend on the value of x_1 for each of the two elements in question
- The exact expression for the intraclass correlation is complicated, and we will not give it here
- The important thing is to recognise that it depends on the two values of x_1, as well as σ^2_{u1}, σ^2_{u0} and σ_{u01}

Residuals

With a random slope model we have several sets of level 2 residuals:

- a set of intercept residuals
- and a set of residuals for each set of random slopes

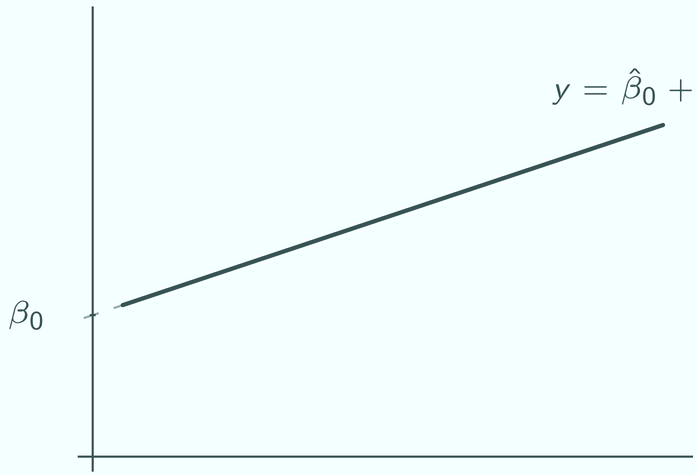
Each set of residuals is shrunk (using very complicated formulae!)



Prediction: visualising the model

Overall regression line

- Prediction from the fixed part gives the overall regression line
- Prediction: $\hat{y}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 x_{1ij}$

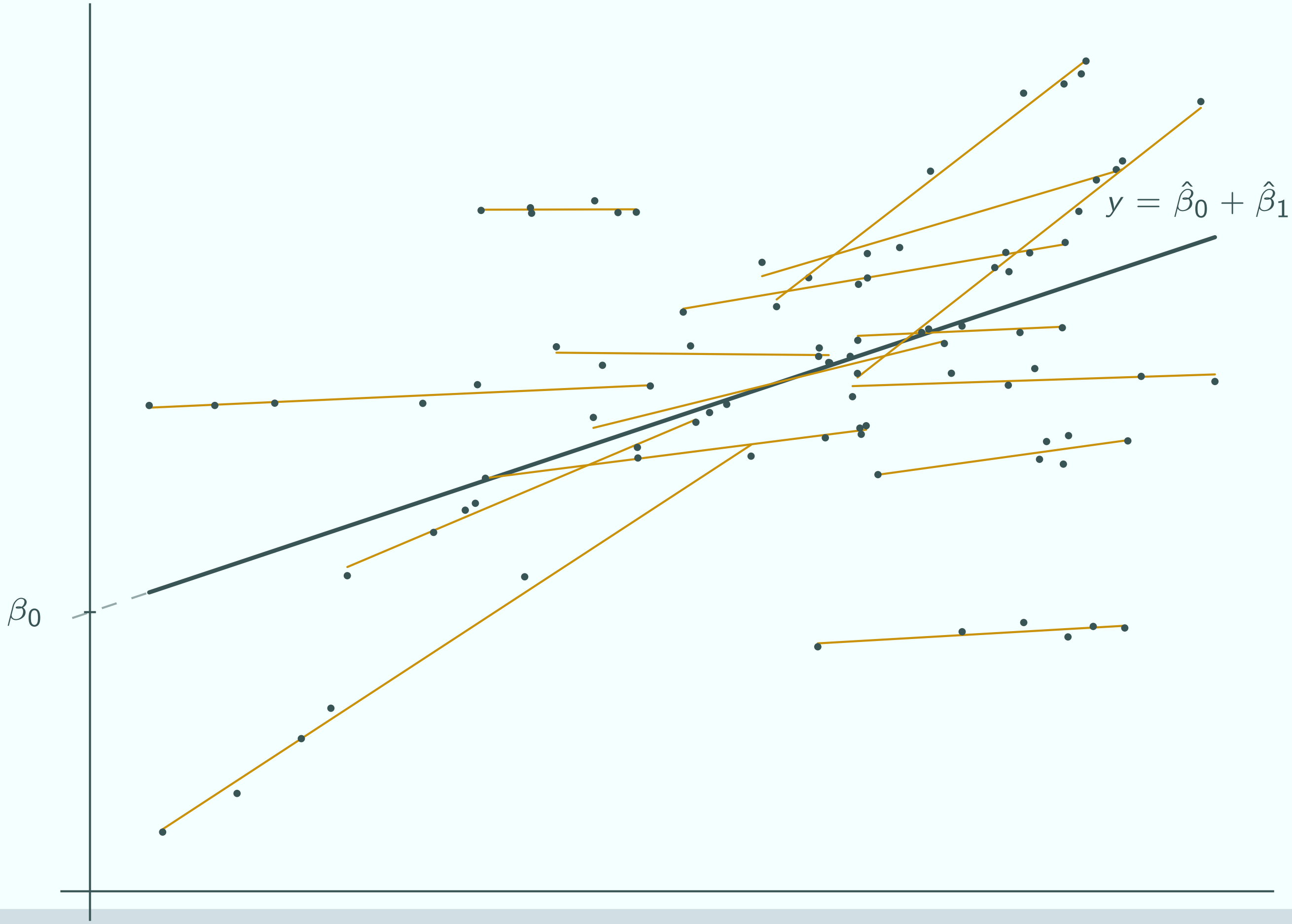


Group lines

- Adding in the level 2 residuals u_{0j} and u_{1j} gives the group lines
- Prediction: $\hat{y}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 x_{1ij} + \hat{u}_{0j} + \hat{u}_{1j} x_{1ij}$

Combined predictions

Putting the prediction from the fixed part and the prediction from the random part together on the same graph, we get:



Random slope models and random intercepts

Terminology

- The random slope model

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + u_{0j} + u_{1j} x_{1ij} + e_{0ij}$$

  has a random intercept as well as a random slope
- So technically it is also a random intercept model
- However, usually when we use the term ‘random intercept model’ we mean a model that has only a random intercept, and no random slope

Do we always add a random intercept?

- We have so far always shown a random intercept in our random slope model
- Leaving out the random intercept means that all group lines cross at $x = 0$
- If we have a good reason to believe this is so, we can fit a model without random intercepts
- Usually there is no reason to believe this and so we put the random intercept in

Multiple explanatory variables

- We can in theory have a random slope on just one of our explanatory variables:

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3ij} + u_{0j} + u_{1j} x_{1ij} + e_{0ij}$$

- or on several of them:

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3ij} + u_{0j} + u_{1j} x_{1ij} + u_{3j} x_{3ij} + e_{0ij}$$

- or even on all of them.
- However, depending on the number of level 2 units in our dataset, we may not in practice have enough power to fit a random slope to more than one explanatory variable
- Random slopes can be fitted to interaction terms as well

Bibliography

- Clark, C., Sheppard, L., Fillenbaum, G., Galasko, D., Morris, J., Koss, E., Mohs, R., Heyman, A., the CERAD Investigators (1999) *Variability in annual mini-mental sstate examination score in patients with probable Alzheimer's disease* Archives of Neurology, **56** pp857 – 862
- Jex, S. and Bliese, P. (1999) *Efficacy beliefs as a moderator of the impact of work-related stressors: A multilevel study* Journal of Applied Psychology **84** 3 pp349 – 361
- Polsky, C. and Easterling, W.E. (2001) *Adaptation to climate variability and change in the US Great Plains: A multi-scale analysis of Ricardian climate sensitivities* Agriculture, Ecosystems and Environment **85** pp133 – 144
- Tymms, P., Merrell, C., Henderson, B. (1997) *The first year at school: A quantitative investigation of the attainment and progress of pupils* Educational Research and Evaluation **3**:2 pp101 – 118