Random Intercept Models

Hedonism example

Our questions in the last session

Hedonism example

Our questions in the last session

- Are there differences between countries in hedonism?

Hedonism example

Our questions in the last session

- Are there differences between countries in hedonism?
- Are some countries more hedonistic than others?

Hedonism example

Our questions in the last session

- Are there differences between countries in hedonism?
- Are some countries more hedonistic than others?
- How much of the variation in hedonism is due to these country differences?

Hedonism example

Our questions in the last session

- Are there differences between countries in hedonism?
- Are some countries more hedonistic than others?
- How much of the variation in hedonism is due to these country differences?

Our questions in this session

Hedonism example

Our questions in the last session

- Are there differences between countries in hedonism?
- Are some countries more hedonistic than others?
- How much of the variation in hedonism is due to these country differences?

Our questions in this session

- Do differences between countries in hedonism remain after controlling for individual age?

Hedonism example

Our questions in the last session

- Are there differences between countries in hedonism?
- Are some countries more hedonistic than others?
- How much of the variation in hedonism is due to these country differences?

Our questions in this session

- Do differences between countries in hedonism remain after controlling for individual age?
- Are some countries more hedonistic than others after controlling for individual age?

Hedonism example

Our questions in the last session

- Are there differences between countries in hedonism?
- Are some countries more hedonistic than others?
- How much of the variation in hedonism is due to these country differences?

Our questions in this session

- Do differences between countries in hedonism remain after controlling for individual age?
- Are some countries more hedonistic than others after controlling for individual age?
- How much of the variation in hedonism is due to country differences after controlling for individual age?

Hedonism example

Our questions in the last session

- Are there differences between countries in hedonism?
- Are some countries more hedonistic than others?
- How much of the variation in hedonism is due to these country differences?

Our questions in this session

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

Exam scores example

Our questions in the last session

Exam scores example

Our questions in the last session

- Are there differences between schools in exam scores at age 16?

Exam scores example

Our questions in the last session

- Are there differences between schools in exam scores at age 16?
- How much of the variation in exam scores at age 16 is due to these school differences?

Exam scores example

Our questions in the last session

- Are there differences between schools in exam scores at age 16?
- How much of the variation in exam scores at age 16 is due to these school differences?

Our questions in this session

Exam scores example

Our questions in the last session

- Are there differences between schools in exam scores at age 16?
- How much of the variation in exam scores at age 16 is due to these school differences?

Our questions in this session

- Do differences between schools in exam scores at age 16 remain after controlling for exam score at age 11?

Exam scores example

Our questions in the last session

- Are there differences between schools in exam scores at age 16?
- How much of the variation in exam scores at age 16 is due to these school differences?

Our questions in this session

- Do differences between schools in exam scores at age 16 remain after controlling for exam score at age 11?
- Are there differences between schools in pupils' progress between age 11 and 16?

Exam scores example

Our questions in the last session

- Are there differences between schools in exam scores at age 16?
- How much of the variation in exam scores at age 16 is due to these school differences?

Our questions in this session

- Do differences between schools in exam scores at age 16 remain after controlling for exam score at age 11?
- Are there differences between schools in pupils' progress between age 11 and 16?
- How much of the variation in exam scores at age 16 is due to school differences after controlling for exam score at age 11?

Exam scores example

Our questions in the last session

- Are there differences between schools in exam scores at age 16?
- How much of the variation in exam scores at age 16 is due to these school differences?

Our questions in this session

- Do differences between schools in exam scores at age 16 remain after controlling for exam score at age 11?
- Are there differences between schools in pupils' progress between age 11 and 16?
- How much of the variation in exam scores at age 16 is due to school differences after controlling for exam score at age 11?
- How much of the variation in pupils' progress between age 11 and 16 is due to school differences?

Introducing explanatory variables

- We’ve seen how to fit a variance components model

Introducing explanatory variables

- We've seen how to fit a variance components model
- This lets us see how much of the variance in our response is at each level

Introducing explanatory variables

- We've seen how to fit a variance components model
- This lets us see how much of the variance in our response is at each level
- But what if we want to look at or control for the effects of explanatory variables?

Introducing explanatory variables

- We've seen how to fit a variance components model
- This lets us see how much of the variance in our response is at each level
- But what if we want to look at or control for the effects of explanatory variables?

Example

Suppose we have data on exam results for pupils within schools

Introducing explanatory variables

- We've seen how to fit a variance components model
- This lets us see how much of the variance in our response is at each level
- But what if we want to look at or control for the effects of explanatory variables?

Example

Suppose we have data on exam results for pupils within schools

- We fit a variance components model and we find 20% of the variance is at the school level.

Introducing explanatory variables

- We've seen how to fit a variance components model
- This lets us see how much of the variance in our response is at each level
- But what if we want to look at or control for the effects of explanatory variables?

Example

Suppose we have data on exam results for pupils within schools

- We fit a variance components model and we find 20% of the variance is at the school level.
- But can we interpret this as
  "20% of the variance in exam scores is caused by schools"?

Introducing explanatory variables

- We've seen how to fit a variance components model
- This lets us see how much of the variance in our response is at each level
- But what if we want to look at or control for the effects of explanatory variables?

Example

Suppose we have data on exam results for pupils within schools

- We fit a variance components model and we find 20% of the variance is at the school level.
- But can we interpret this as "20% of the variance in exam scores is caused by schools"?
- Schools differ in their intake policy and in the pupils who apply

Introducing explanatory variables

- We've seen how to fit a variance components model
- This lets us see how much of the variance in our response is at each level
- But what if we want to look at or control for the effects of explanatory variables?

Example

Suppose we have data on exam results for pupils within schools

- We fit a variance components model and we find 20% of the variance is at the school level.
- But can we interpret this as "20% of the variance in exam scores is caused by schools"?
- Schools differ in their intake policy and in the pupils who apply
- These differences also contribute to school-level variance

Introducing explanatory variables

- We've seen how to fit a variance components model
- This lets us see how much of the variance in our response is at each level
- But what if we want to look at or control for the effects of explanatory variables?

Example

Suppose we have data on exam results for pupils within schools

- We fit a variance components model and we find 20% of the variance is at the school level.
- But can we interpret this as "20% of the variance in exam scores is caused by schools"?
- Schools differ in their intake policy and in the pupils who apply
- These differences also contribute to school-level variance
- So we would like to control for previous exam score

Using a single-level regression model

- We usually do this by fitting a regression model:

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad e_i \sim \mathrm{N}(0, \sigma_e^2)$$

Using a single-level regression model

- We usually do this by fitting a regression model:

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad e_i \sim \mathrm{N}(0, \sigma_e^2)$$

- When we have clustered data, using this model causes problems.

Using a single-level regression model

- We usually do this by fitting a regression model:

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad e_i \sim \mathrm{N}(0, \sigma_e^2)$$

- When we have clustered data, using this model causes problems.

Clustered data

Data where observations in the same group are related, e.g.

Using a single-level regression model

- We usually do this by fitting a regression model:

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad e_i \sim \mathrm{N}(0, \sigma_e^2)$$

- When we have clustered data, using this model causes problems.

Clustered data

Data where observations in the same group are related, e.g.

- exam results of pupils within schools

Using a single-level regression model

- We usually do this by fitting a regression model:

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad e_i \sim \mathrm{N}(0, \sigma_e^2)$$

- When we have clustered data, using this model causes problems.

Clustered data

Data where observations in the same group are related, e.g.

- exam results of pupils within schools
- heights of children within families

Using a single-level regression model

- We usually do this by fitting a regression model:

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad e_i \sim \mathrm{N}(0, \sigma_e^2)$$

- When we have clustered data, using this model causes problems.

Clustered data

Data where observations in the same group are related, e.g.

- exam results of pupils within schools
- heights of children within families

Problem 1

Using a single-level regression model

- We usually do this by fitting a regression model:

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad e_i \sim \mathrm{N}(0, \sigma_e^2)$$

- When we have clustered data, using this model causes problems.

Clustered data

Data where observations in the same group are related, e.g.

- exam results of pupils within schools
- heights of children within families

Problem 1

If we fit this model to clustered data we get the wrong answers

Using a single-level regression model

- We usually do this by fitting a regression model:

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad e_i \sim \mathrm{N}(0, \sigma_e^2)$$

- When we have clustered data, using this model causes problems.

Clustered data

Data where observations in the same group are related, e.g.

- exam results of pupils within schools
- heights of children within families

Problem 1

If we fit this model to clustered data we get the wrong answers

Problem 2

Using a single-level regression model

- We usually do this by fitting a regression model:

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad e_i \sim \mathrm{N}(0, \sigma_e^2)$$

- When we have clustered data, using this model causes problems.

Clustered data

Data where observations in the same group are related, e.g.

- exam results of pupils within schools
- heights of children within families

Problem 1

If we fit this model to clustered data we get the wrong answers

Problem 2

- This model doesn't show us how much variation is at each level

Using a single-level regression model

- We usually do this by fitting a regression model:

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad e_i \sim \mathrm{N}(0, \sigma_e^2)$$

- When we have clustered data, using this model causes problems.

Clustered data

Data where observations in the same group are related, e.g.

- exam results of pupils within schools
- heights of children within families

Problem 1

If we fit this model to clustered data we get the wrong answers

Problem 2

- This model doesn't show us how much variation is at each level
- So we can't find out by using this model how much of an effect school has on exam score after controlling for intake score

Using a single-level regression model

- We usually do this by fitting a regression model:

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad e_i \sim \mathrm{N}(0, \sigma_e^2)$$

- When we have clustered data, using this model causes problems.

Clustered data

Data where observations in the same group are related, e.g.

- exam results of pupils within schools
- heights of children within families

Problem 1

If we fit this model to clustered data we get the wrong answers

Problem 2

- This model doesn't show us how much variation is at each level
- So we can't find out by using this model how much of an effect school has on exam score after controlling for intake score
- This is what we're interested in

Using a single-level regression model

- We usually do this by fitting a regression model:

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad e_i \sim \mathrm{N}(0, \sigma_e^2)$$

- When we have clustered data, using this model causes problems.

Clustered data

Data where observations in the same group are related, e.g.

- exam results of pupils within schools
- heights of children within families

Problem 1

If we fit this model to clustered data we get the wrong answers

Problem 2

- This model doesn't show us how much variation is at each level
- So we can't find out by using this model how much of an effect school has on exam score after controlling for intake score
- This is what we're interested in $\rightarrow$ problem

Solution: Random Intercept Model

We combine the variance components and the regression models

Solution: Random Intercept Model

We combine the variance components and the regression models

Variance components model

$$y_{ij} = \beta_0 + u_j + e_{ij} \quad e_{ij} \sim \mathrm{N}(0, \sigma_e^2)$$
$$u_j \sim \mathrm{N}(0, \sigma_u^2)$$

Single level regression model

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad e_i \sim \mathrm{N}(0, \sigma^2)$$

Solution: Random Intercept Model

We combine the variance components and the regression models

Variance components model

$$y_{ij} = \beta_0 + u_j + e_{ij} \quad e_{ij} \sim \mathrm{N}(0, \sigma_e^2)$$
$$u_j \sim \mathrm{N}(0, \sigma_u^2)$$

Single level regression model

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad e_i \sim \mathrm{N}(0, \sigma^2)$$

Solution: Random Intercept Model

We combine the variance components and the regression models

Variance components model

$$y_{ij} = \beta_0 + u_j + e_{ij} \quad e_{ij} \sim \mathrm{N}(0, \sigma_e^2)$$
$$u_j \sim \mathrm{N}(0, \sigma_u^2)$$

Single level regression model

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad e_i \sim \mathrm{N}(0, \sigma^2)$$

$+$

Solution: Random Intercept Model

We combine the variance components and the regression models

Variance components model

$$y_{ij} = \beta_0 + u_j + e_{ij} \quad e_{ij} \sim \mathrm{N}(0, \sigma_e^2)$$
$$u_j \sim \mathrm{N}(0, \sigma_u^2)$$

Single level regression model

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad e_i \sim \mathrm{N}(0, \sigma^2)$$

$+$

Random intercept model

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_j + e_{ij} \quad e_{ij} \sim \mathrm{N}(0, \sigma_e^2)$$
$$u_j \sim \mathrm{N}(0, \sigma_u^2)$$

Fixed and random part

The random intercept model has two parts:

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

Fixed and random part

The random intercept model has two parts:

- a "fixed part"

$$y_{ij} = \overbrace{\beta_0 + \beta_1 x_{ij}}^{\text{fixed part}} + u_j + e_{ij}$$

Fixed part

Fixed and random part

The random intercept model has two parts:

- a "fixed part"
- and a "random part"

$$y_{ij} = \overbrace{\beta_0 + \beta_1 x_{ij}}^{\text{fixed part}} + \overbrace{u_j + e_{ij}}^{\text{random part}}$$

Fixed part

Random part

Fixed and random part

The random intercept model has two parts:

- a "fixed part"
- and a "random part"

$$y_{ij} = \overbrace{\beta_0 + \beta_1 x_{ij}}^{\text{fixed part}} + \overbrace{u_j + e_{ij}}^{\text{random part}}$$

Fixed part

- Parameters that we estimate are the coefficients

$$\beta_0, \beta_1, \ldots$$

Random part

Fixed and random part

The random intercept model has two parts:

- a "fixed part"
- and a "random part"

$$y_{ij} = \overbrace{\beta_0 + \beta_1 x_{ij}}^{\text{fixed part}} + \overbrace{u_j + e_{ij}}^{\text{random part}}$$

Fixed part

- Parameters that we estimate are the coefficients

$$\beta_0, \beta_1, \ldots$$

Random part

- Parameters that we estimate are the variances

$$\sigma_u^2 \text{ and } \sigma_e^2$$

Fixed and random part

The random intercept model has two parts:

- a “fixed part”
- and a “random part”

$$y_{ij} = \overbrace{\beta_0 + \beta_1 x_{ij}}^{\text{fixed part}} + \overbrace{u_j + e_{ij}}^{\text{random part}}$$

Fixed part

- Parameters that we estimate are the coefficients

$$\beta_0, \beta_1, \ldots$$

Random part

- Parameters that we estimate are the variances

$$\sigma_u^2 \text{ and } \sigma_e^2$$

- The “random part” is random in the same way that the error term e_i of the single level regression model is random:

Fixed and random part

The random intercept model has two parts:

- a "fixed part"
- and a "random part"

$$y_{ij} = \overbrace{\beta_0 + \beta_1 x_{ij}}^{\text{fixed part}} + \overbrace{u_j + e_{ij}}^{\text{random part}}$$

Fixed part

- Parameters that we estimate are the coefficients

$$\beta_0, \beta_1, \ldots$$

Random part

- Parameters that we estimate are the variances

$$\sigma_u^2 \text{ and } \sigma_e^2$$

- The "random part" is random in the same way that the error term e_i of the single level regression model is random:
  - the u_j and e_{ij} are allowed to vary

Fixed and random part

The random intercept model has two parts:

- a "fixed part"
- and a "random part"

$$y_{ij} = \overbrace{\beta_0 + \beta_1 x_{ij}}^{\text{fixed part}} + \overbrace{u_j + e_{ij}}^{\text{random part}}$$

Fixed part

- Parameters that we estimate are the coefficients

$$\beta_0, \beta_1, \ldots$$

Random part

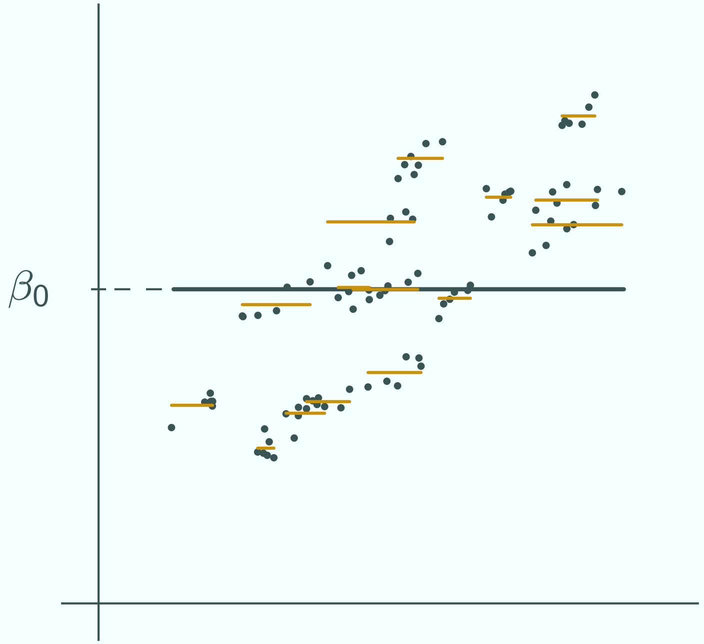
- Parameters that we estimate are the variances

$$\sigma_u^2 \text{ and } \sigma_e^2$$

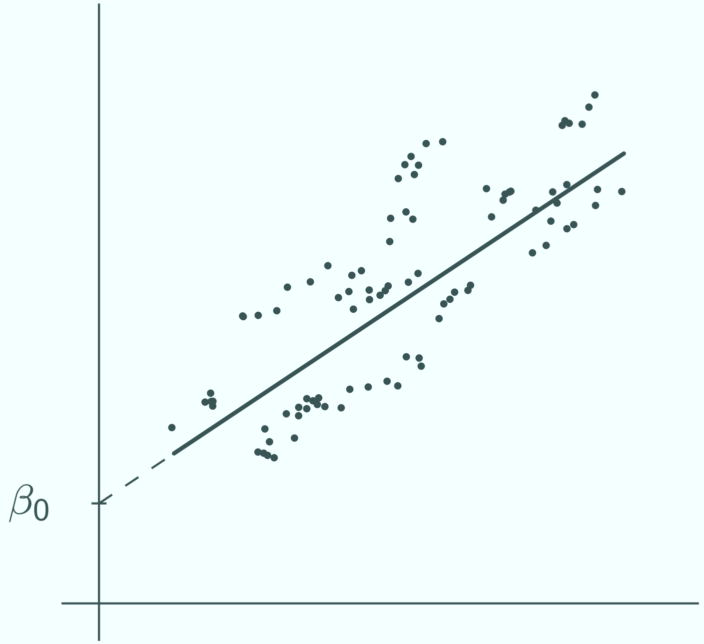
- The "random part" is random in the same way that the error term e_i of the single level regression model is random:
  - the u_j and e_{ij} are allowed to vary
  - some unmeasured processes are generating the u_j and e_{ij}

What does the model look like?

Variance components model

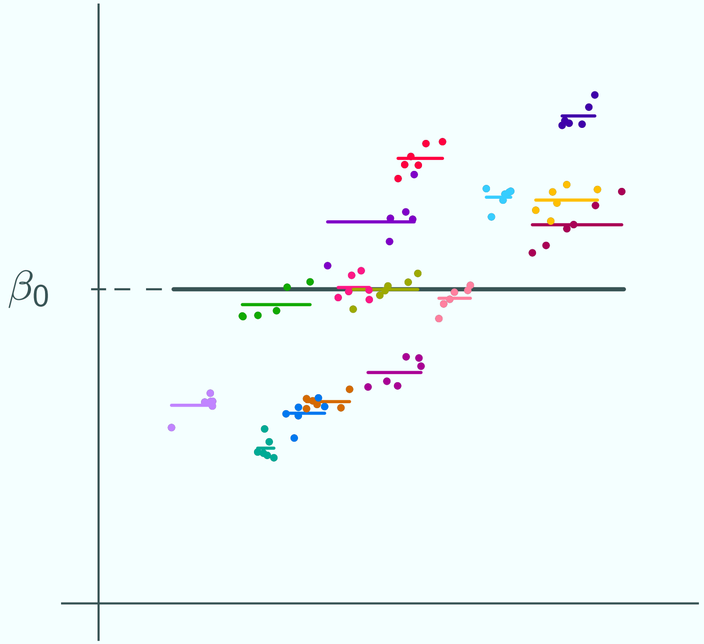


Single level regression model

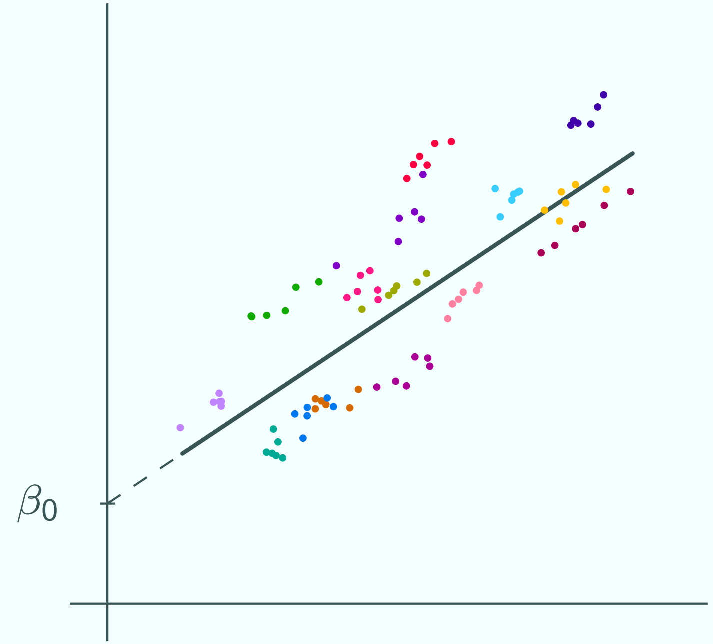


What does the model look like?

Variance components model

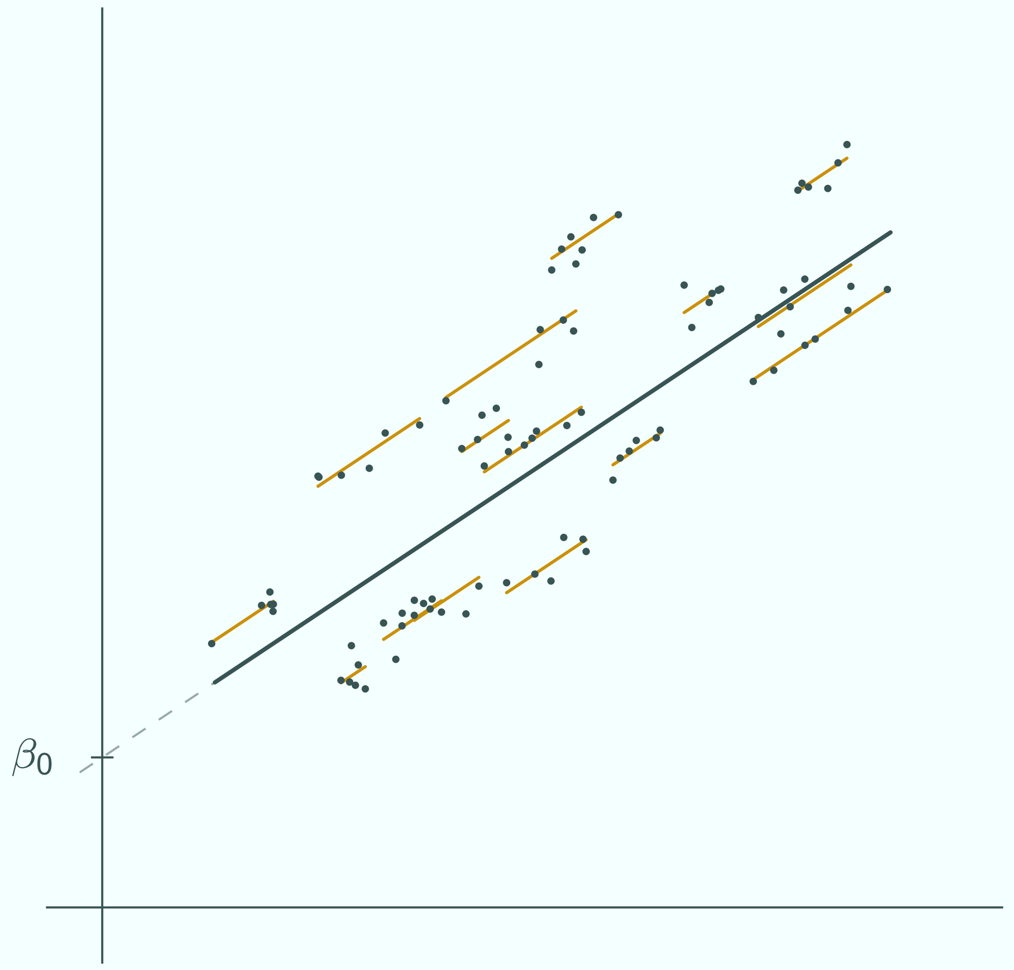


Single level regression model



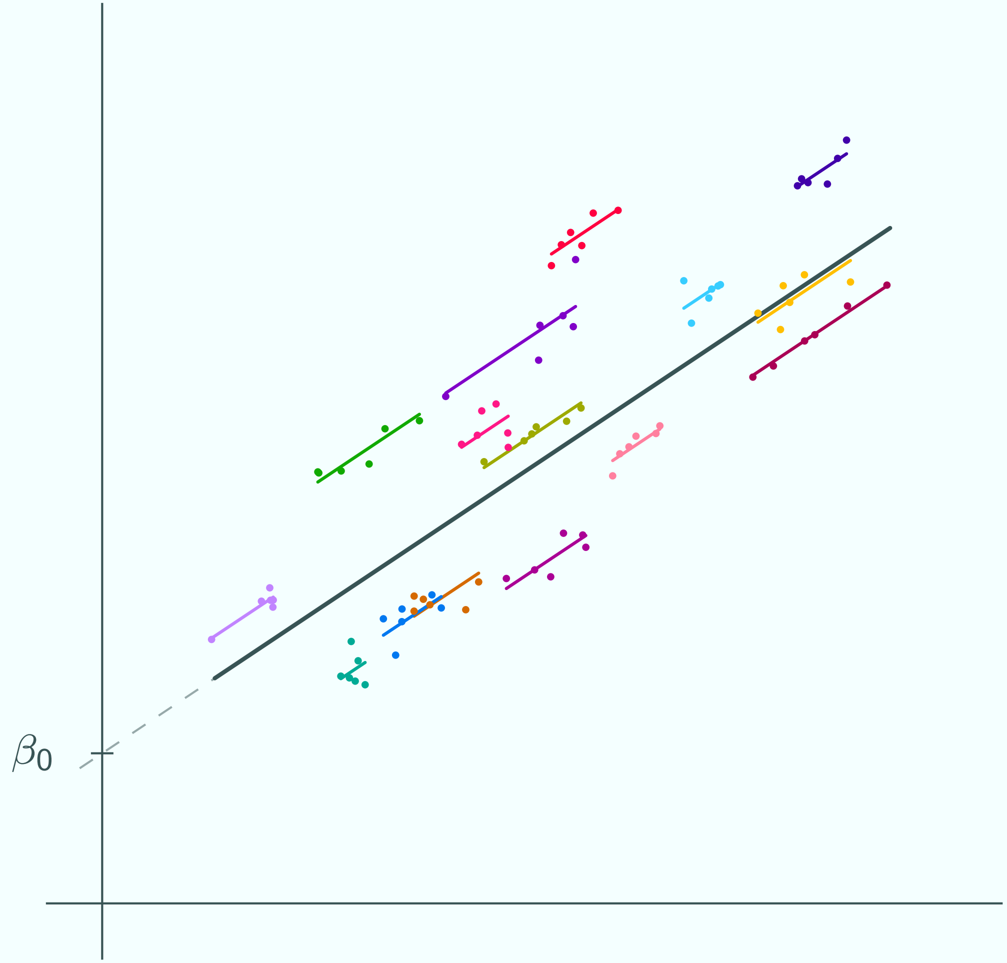
What does the model look like?

Random intercept model



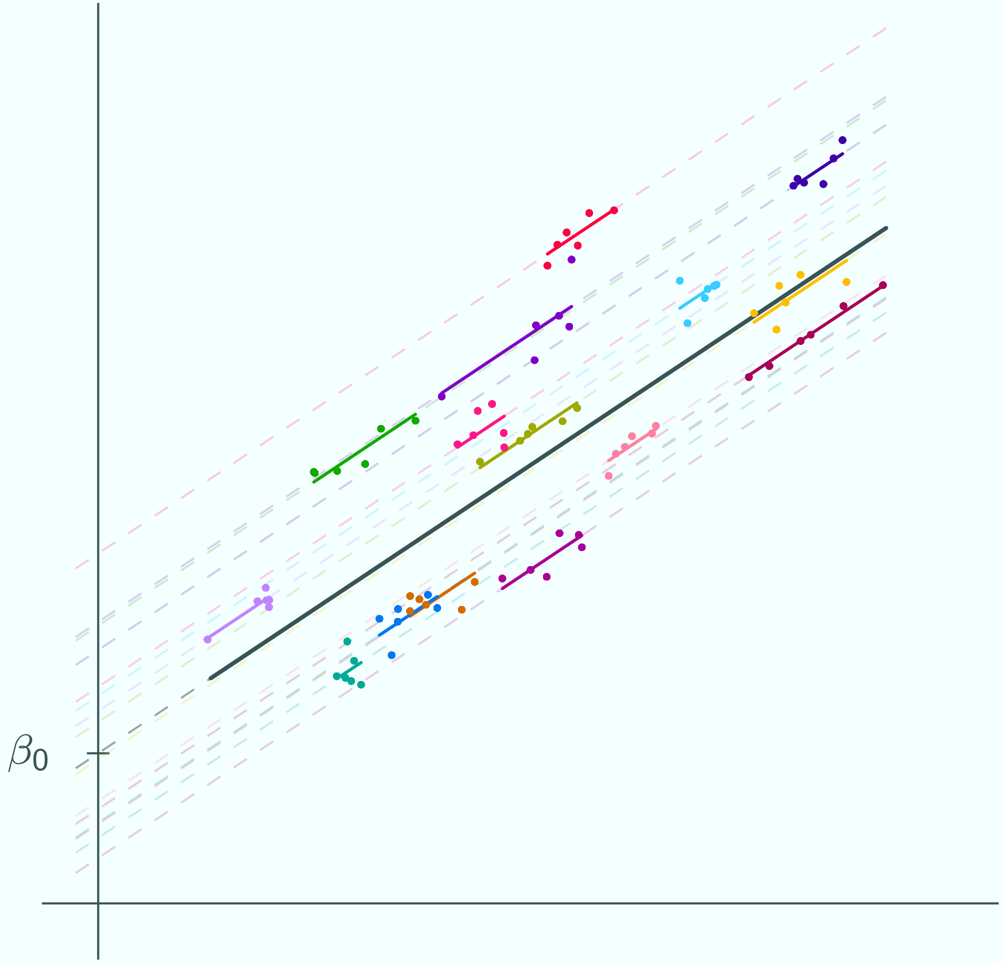
What does the model look like?

Random intercept model



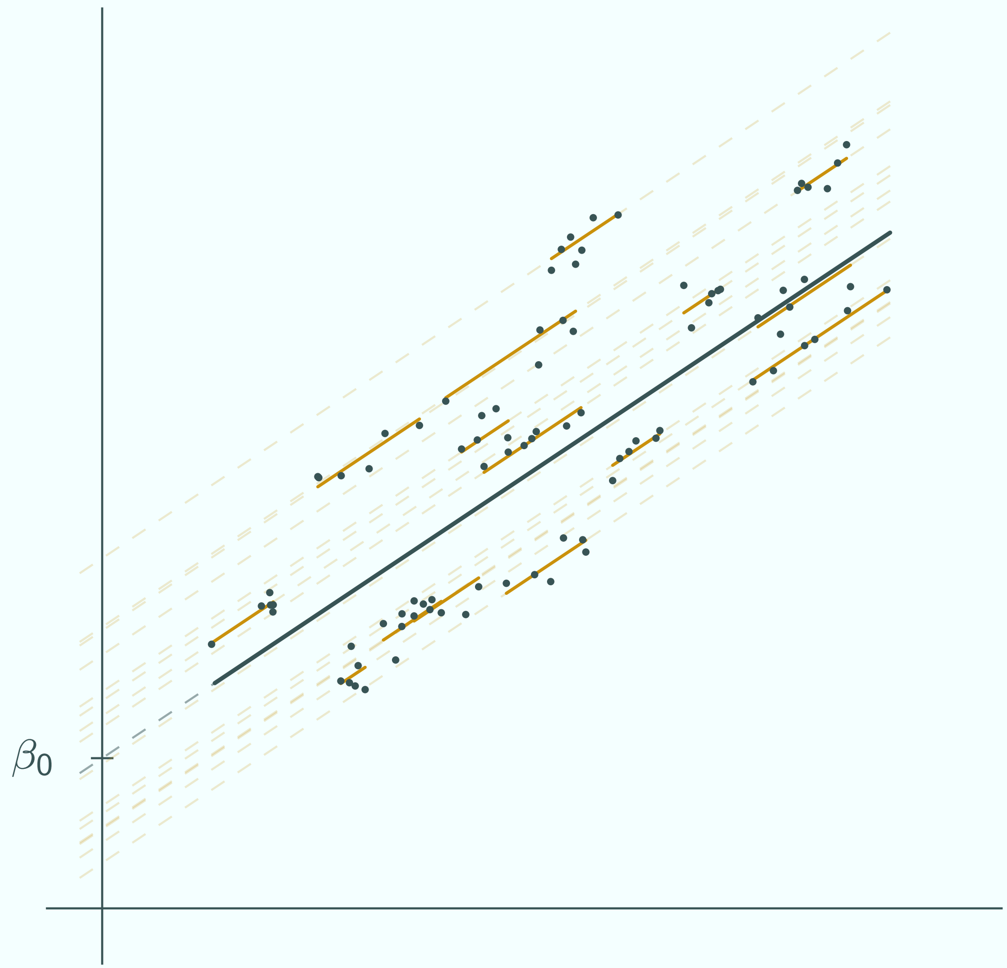
What does the model look like?

Random intercept model



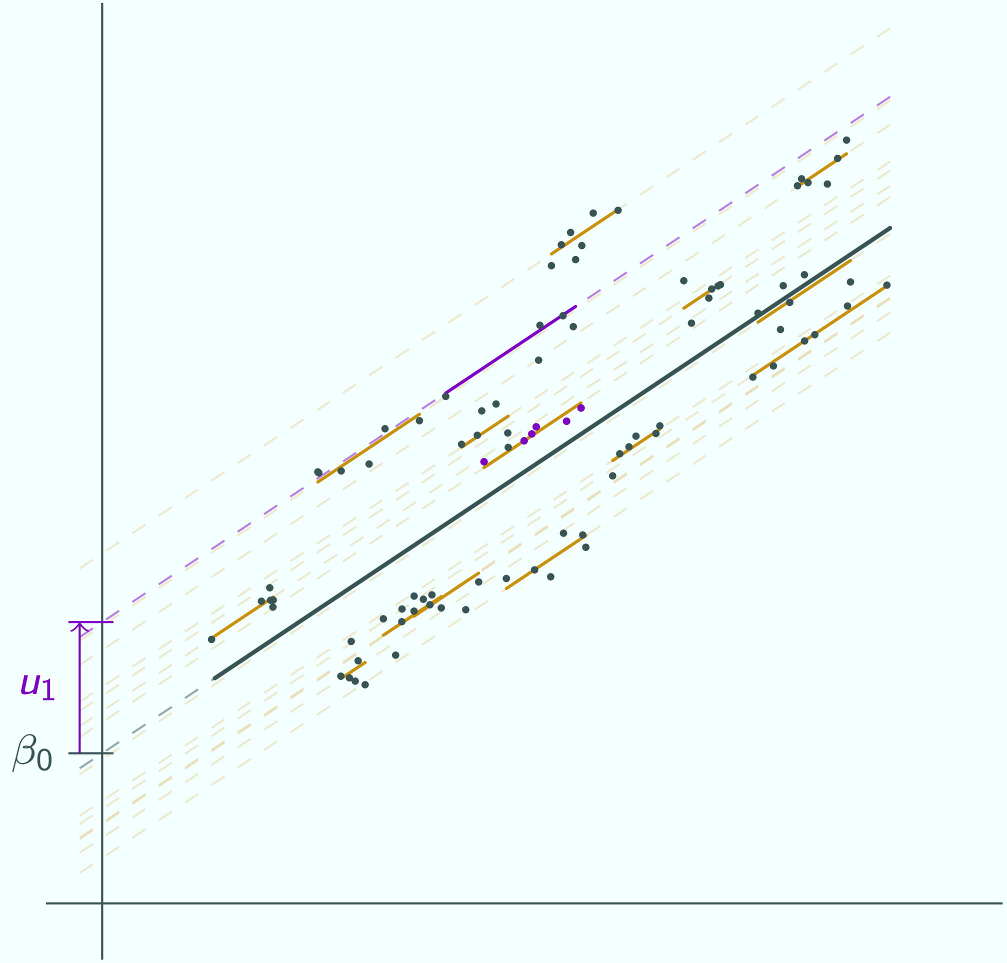
What does the model look like?

Random intercept model



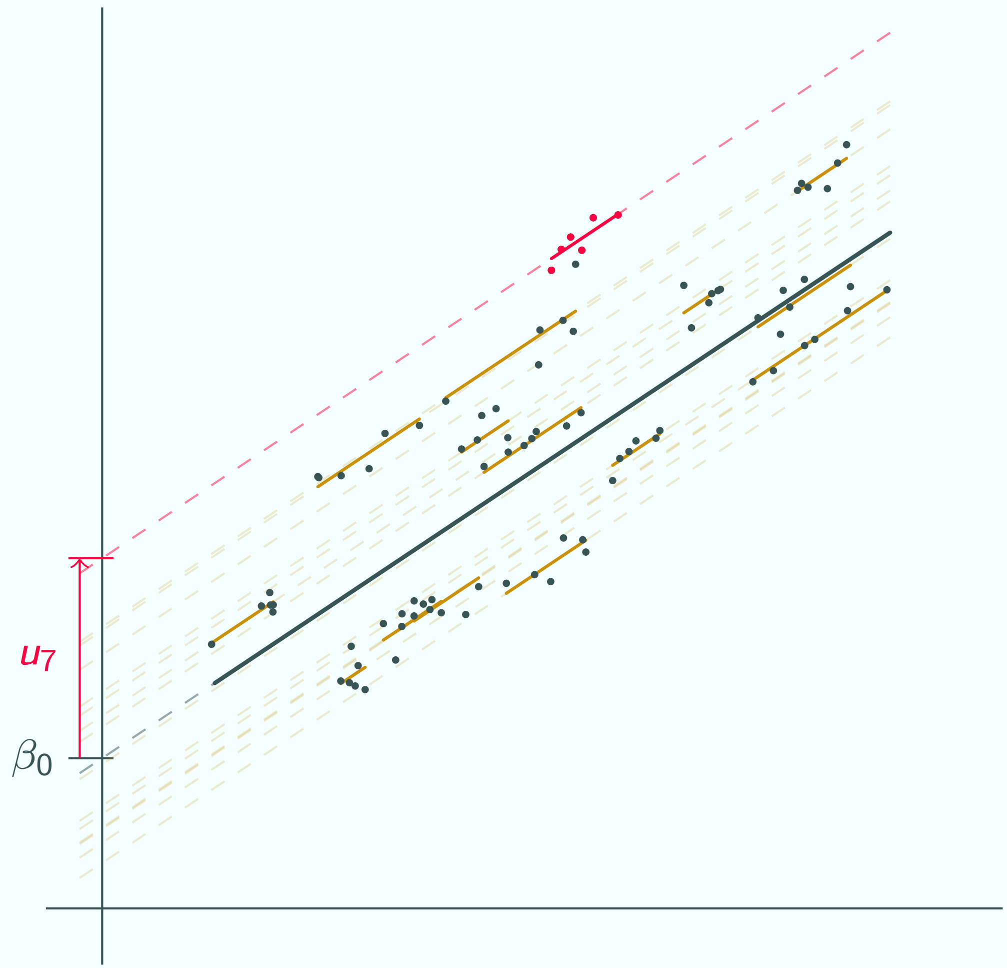
What does the model look like?

Random intercept model



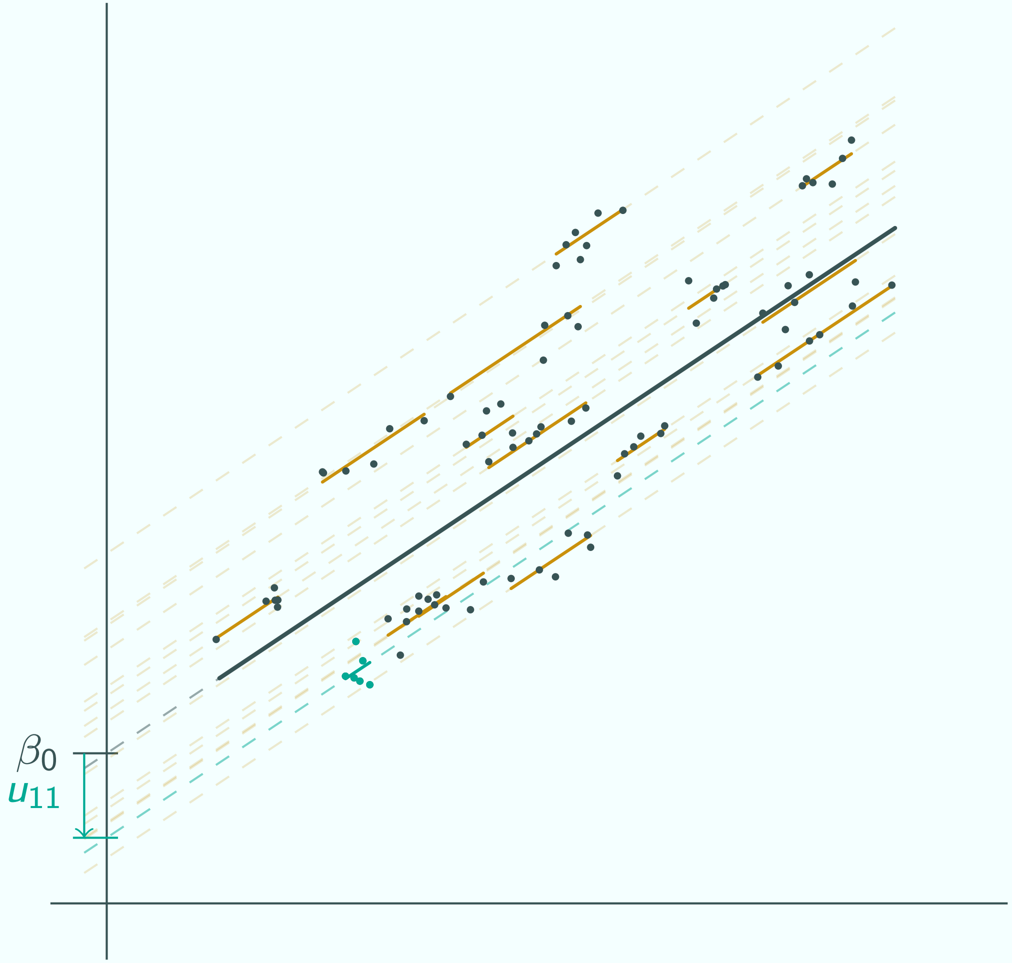
What does the model look like?

Random intercept model



What does the model look like?

Random intercept model



What does the model look like?

Overall line

Like the single level regression model, the overall average line has equation $\beta_0 + \beta_1 x_{ij}$

Group lines

The 'random intercept'

What does the model look like?

Overall line

Like the single level regression model, the overall average line has equation $\beta_0 + \beta_1 x_{ij}$

Group lines

Like the variance components model, each group has its own line, parallel to the overall average line

The ‘random intercept’

What does the model look like?

Overall line

Like the single level regression model, the overall average line has equation $\beta_0 + \beta_1 x_{ij}$

Group lines

Like the variance components model, each group has its own line, parallel to the overall average line

The 'random intercept'

- For the single level regression model, the intercept is just β_0

What does the model look like?

Overall line

Like the single level regression model, the overall average line has equation $\beta_0 + \beta_1 x_{ij}$

Group lines

Like the variance components model, each group has its own line, parallel to the overall average line

The 'random intercept'

- For the single level regression model, the intercept is just β_0
- This is a parameter from the fixed part of the model

What does the model look like?

Overall line

Like the single level regression model, the overall average line has equation $\beta_0 + \beta_1 x_{ij}$

Group lines

Like the variance components model, each group has its own line, parallel to the overall average line

The 'random intercept'

- For the single level regression model, the intercept is just β_0
- This is a parameter from the fixed part of the model
- For the random intercept model, the intercept for the overall regression line is still β_0

What does the model look like?

Overall line

Like the single level regression model, the overall average line has equation $\beta_0 + \beta_1 x_{ij}$

Group lines

Like the variance components model, each group has its own line, parallel to the overall average line

The 'random intercept'

- For the single level regression model, the intercept is just β_0
- This is a parameter from the fixed part of the model
- For the random intercept model, the intercept for the overall regression line is still β_0
- For each group line the intercept is $\beta_0 + u_j$

What does the model look like?

Overall line

Like the single level regression model, the overall average line has equation $\beta_0 + \beta_1 x_{ij}$

Group lines

Like the variance components model, each group has its own line, parallel to the overall average line

The 'random intercept'

- For the single level regression model, the intercept is just β_0
- This is a parameter from the fixed part of the model
- For the random intercept model, the intercept for the overall regression line is still β_0
- For each group line the intercept is $\beta_0 + u_j$
- This involves a parameter from the random part and so it is allowed to vary

What does the model look like?

Overall line

Like the single level regression model, the overall average line has equation $\beta_0 + \beta_1 x_{ij}$

Group lines

Like the variance components model, each group has its own line, parallel to the overall average line

The 'random intercept'

- For the single level regression model, the intercept is just β_0
- This is a parameter from the fixed part of the model
- For the random intercept model, the intercept for the overall regression line is still β_0
- For each group line the intercept is $\beta_0 + u_j$
- This involves a parameter from the random part and so it is allowed to vary $\rightarrow$ 'random intercept'

Interpreting the parameters

Fixed part

Interpretation is as for a single level regression model

Random part

Interpreting the parameters

Fixed part

Interpretation is as for a single level regression model

- β_1 is the increase in the response for a 1 unit increase in x

Random part

Interpreting the parameters

Fixed part

Interpretation is as for a single level regression model

- β_1 is the increase in the response for a 1 unit increase in x
  - e.g. the increase in hedonism for a 1 year increase in age

Random part

Interpreting the parameters

Fixed part

Interpretation is as for a single level regression model

- β_1 is the increase in the response for a 1 unit increase in x
  - e.g. the increase in hedonism for a 1 year increase in age

Random part

Interpretation is as for a variance components model

Interpreting the parameters

Fixed part

Interpretation is as for a single level regression model

- β_1 is the increase in the response for a 1 unit increase in x
  - e.g. the increase in hedonism for a 1 year increase in age

Random part

Interpretation is as for a variance components model

Note that again the parameters we estimate are σ_u^2 and σ_e^2, not u_j and e_{ij}

Interpreting the parameters

Fixed part

Interpretation is as for a single level regression model

- β_1 is the increase in the response for a 1 unit increase in x
  - e.g. the increase in hedonism for a 1 year increase in age

Random part

Interpretation is as for a variance components model
Note that again the parameters we estimate are σ_u^2 and σ_e^2, not u_j and e_{ij}

- σ_u^2 is the unexplained variation at level 2

Interpreting the parameters

Fixed part

Interpretation is as for a single level regression model

- β_1 is the increase in the response for a 1 unit increase in x
  - e.g. the increase in hedonism for a 1 year increase in age

Random part

Interpretation is as for a variance components model

Note that again the parameters we estimate are σ_u^2 and σ_e^2, not u_j and e_{ij}

- σ_u^2 is the unexplained variation at level 2
  - e.g. the variation in hedonism due to differences between countries after controlling for age

Interpreting the parameters

Fixed part

Interpretation is as for a single level regression model

- β_1 is the increase in the response for a 1 unit increase in x
    - e.g. the increase in hedonism for a 1 year increase in age

Random part

Interpretation is as for a variance components model
Note that again the parameters we estimate are σ_u^2 and σ_e^2, not u_j and e_{ij}

- σ_u^2 is the unexplained variation at level 2
    - e.g. the variation in hedonism due to differences between countries after controlling for age
- σ_e^2 is the unexplained variation at level 1

Interpreting the parameters

Fixed part

Interpretation is as for a single level regression model

- β_1 is the increase in the response for a 1 unit increase in x
  - e.g. the increase in hedonism for a 1 year increase in age

Random part

Interpretation is as for a variance components model

Note that again the parameters we estimate are σ_u^2 and σ_e^2, not u_j and e_{ij}

- σ_u^2 is the unexplained variation at level 2
  - e.g. the variation in hedonism due to differences between countries after controlling for age
- σ_e^2 is the unexplained variation at level 1
  - e.g. the variation in hedonism due to differences between individuals after controlling for age

Hypothesis testing

- Hypothesis testing is an important part of interpretation

Fixed part

Random part

Hypothesis testing

- Hypothesis testing is an important part of interpretation
- We don't only want to know the size of the fixed effects and the amount of variance at each level

Fixed part

Random part

Hypothesis testing

- Hypothesis testing is an important part of interpretation
- We don't only want to know the size of the fixed effects and the amount of variance at each level
- We also want to know whether the fixed effects are significant

Fixed part

Random part

Hypothesis testing

- Hypothesis testing is an important part of interpretation
- We don't only want to know the size of the fixed effects and the amount of variance at each level
- We also want to know whether the fixed effects are significant
- and whether there is a significant amount of variance at level 2

Fixed part

Random part

Hypothesis testing

- Hypothesis testing is an important part of interpretation
- We don't only want to know the size of the fixed effects and the amount of variance at each level
- We also want to know whether the fixed effects are significant
- and whether there is a significant amount of variance at level 2

Fixed part

- Divide the coefficient by its standard error to get

$$z = \frac{\beta_1}{\text{s.e.}(\beta_1)}$$

Random part

Hypothesis testing

- Hypothesis testing is an important part of interpretation
- We don't only want to know the size of the fixed effects and the amount of variance at each level
- We also want to know whether the fixed effects are significant
- and whether there is a significant amount of variance at level 2

Fixed part

- Divide the coefficient by its standard error to get

$$z = \frac{\beta_1}{\text{s.e.}(\beta_1)}$$

- If $|z| \geqslant 1.96$ (or informally if $|z| \geqslant 2$), then β_1 is significant at the 5% level

Random part

Hypothesis testing

- Hypothesis testing is an important part of interpretation
- We don't only want to know the size of the fixed effects and the amount of variance at each level
- We also want to know whether the fixed effects are significant
- and whether there is a significant amount of variance at level 2

Fixed part

- Divide the coefficient by its standard error to get

$$z = \frac{\beta_1}{\text{s.e.}(\beta_1)}$$

- If $|z| \geqslant 1.96$ (or informally if $|z| \geqslant 2$), then β_1 is significant at the 5% level

Random part

- We CAN'T just divide σ_u^2 by s.e.(σ_u^2) and compare the modulus with 1.96

Hypothesis testing

- Hypothesis testing is an important part of interpretation
- We don't only want to know the size of the fixed effects and the amount of variance at each level
- We also want to know whether the fixed effects are significant
- and whether there is a significant amount of variance at level 2

Fixed part

- Divide the coefficient by its standard error to get

$$z = \frac{\beta_1}{\text{s.e.}(\beta_1)}$$

- If $|z| \geqslant 1.96$ (or informally if $|z| \geqslant 2$), then β_1 is significant at the 5% level

Random part

- We CAN'T just divide σ_u^2 by s.e.(σ_u^2) and compare the modulus with 1.96
- Instead we have to fit the model with and without u_j and do a likelihood ratio test to see whether σ_u^2 is significant

Likelihood ratio test

- $$\begin{array}{rll} \text{We fit} & y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \ldots + u_j + e_{ij} & (1) \\ \text{and} & y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \ldots + e_{ij} & (0) \end{array}$$
  and note the likelihoods

Likelihood ratio test

- We fit $y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \ldots + u_j + e_{ij}$ ①
  and $y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \ldots + e_{ij}$ ⓪
  and note the likelihoods
- The test statistic is 2(log(likelihood(①)) − log(likelihood(⓪)))

Likelihood ratio test

- $$\begin{array}{rl} \text{We fit} & y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \ldots + u_j + e_{ij} \quad ① \\ \text{and} & y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \ldots + e_{ij} \quad ⓪ \end{array}$$
  and note the likelihoods
- The test statistic is 2(log(likelihood(①)) − log(likelihood(⓪)))
  - MLwiN gives $-2 \times$ log(likelihood) in the **Equations** window

Likelihood ratio test

- We fit $y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \ldots + u_j + e_{ij}$ ①
  and $y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \ldots + e_{ij}$ ⓪
  and note the likelihoods
- The test statistic is 2(log(likelihood(①)) − log(likelihood(⓪)))
  - MLwiN gives $-2 \times \log(\text{likelihood})$ in the **Equations** window
  - So we just take
    (MLwiN's value for ⓪) − (MLwiN's value for ①)

Likelihood ratio test

- $$\begin{aligned} \text{We fit} \quad & y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \ldots + u_j + e_{ij} \quad & (1) \\ \text{and} \quad & y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \ldots + e_{ij} & (0) \end{aligned}$$
  and note the likelihoods
- The test statistic is 2(log(likelihood((1))) − log(likelihood((0))))
  - MLwiN gives $-2 \times$ log(likelihood) in the **Equations** window
  - So we just take
    (MLwiN's value for (0)) − (MLwiN's value for (1))
- The null hypothesis is that $\sigma_u^2 = 0$ and so we don't need u_j in the model

Likelihood ratio test

- We fit $y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \ldots + u_j + e_{ij}$ ①
  and $y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \ldots + e_{ij}$ ⓪
  and note the likelihoods
- The test statistic is $2(\log(\text{likelihood}(①)) - \log(\text{likelihood}(⓪)))$
  - MLwiN gives $-2 \times \log(\text{likelihood})$ in the **Equations** window
  - So we just take
    (MLwiN's value for ⓪) − (MLwiN's value for ①)
- The null hypothesis is that $\sigma_u^2 = 0$ and so we don't need u_j in the model
- We compare the test statistic to the $\chi^2_{(1)}$ distribution

Likelihood ratio test

- We fit $y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \ldots + u_j + e_{ij}$ ①
  and $y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \ldots + e_{ij}$ ⓪
  and note the likelihoods
- The test statistic is 2(log(likelihood(①)) − log(likelihood(⓪)))
  - MLwiN gives $-2 \times$ log(likelihood) in the **Equations** window
  - So we just take
    (MLwiN's value for ⓪) − (MLwiN's value for ①)
- The null hypothesis is that $\sigma_u^2 = 0$ and so we don't need u_j in the model
- We compare the test statistic to the $\chi^2_{(1)}$ distribution
  - Some people divide the corresponding p-value by 2 (since $\sigma_u^2 \geqslant 0$)

Likelihood ratio test

- We fit $y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \ldots + u_j + e_{ij}$ ①
  and $y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \ldots + e_{ij}$ ⓪
  and note the likelihoods
- The test statistic is 2(log(likelihood(①)) − log(likelihood(⓪)))
  - MLwiN gives $-2 \times$ log(likelihood) in the **Equations** window
  - So we just take
    (MLwiN's value for ⓪) − (MLwiN's value for ①)
- The null hypothesis is that $\sigma_u^2 = 0$ and so we don't need u_j in the model
- We compare the test statistic to the $\chi^2_{(1)}$ distribution
  - Some people divide the corresponding *p*-value by 2 (since $\sigma_u^2 \geqslant 0$)
- There is 1 degree of freedom because there is one more parameter, σ_u^2, in ① compared to ⓪

Exam scores example

Question

Do differences between schools in exam scores at age 16 remain after controlling for exam score at age 11?

Answer

Exam scores example

Question

Do differences between schools in exam scores at age 16 remain after controlling for exam score at age 11?

Answer

1. Fit model with random intercept and note the $-2 \times \log(\text{likelihood})$ value:

Exam scores example

1

Equations

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

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

$$u_{0j} \sim \mathrm{N}(0, \sigma^2_{u0}) \quad \sigma^2_{u0} = 0.092(0.018)$$

$$e_{ij} \sim \mathrm{N}(0, \sigma^2_e) \quad \sigma^2_e = 0.566(0.013)$$

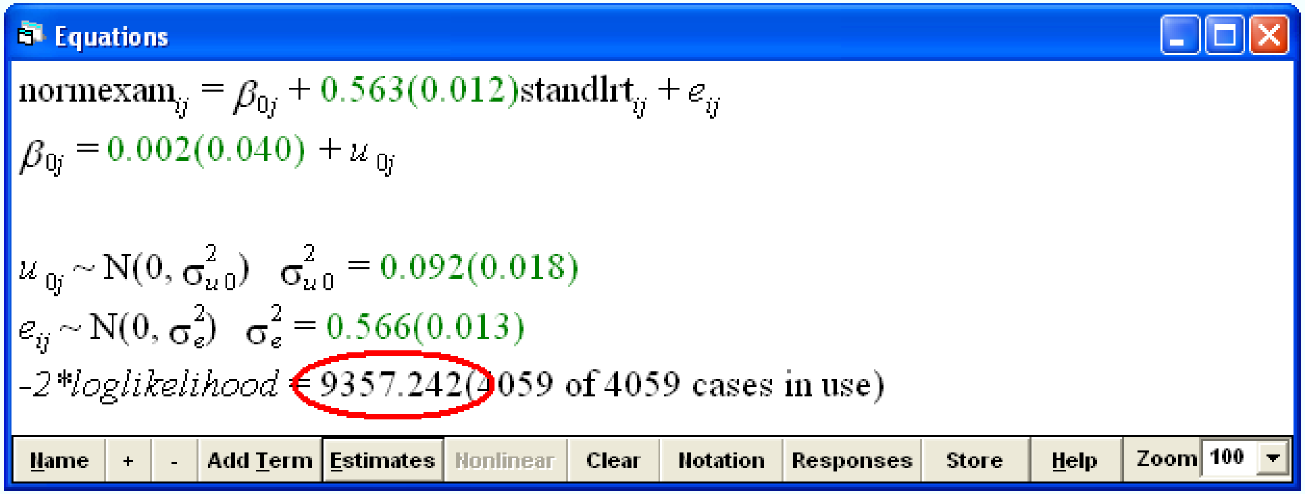
$-2*loglikelihood = 9357.242$(4059 of 4059 cases in use)

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

2

Exam scores example

1



2

Exam scores example

Question

Do differences between schools in exam scores at age 16 remain after controlling for exam score at age 11?

Answer

1. Fit model with random intercept and note the $-2 \times \log(\text{likelihood})$ value: 9357.242

Exam scores example

Question

Do differences between schools in exam scores at age 16 remain after controlling for exam score at age 11?

Answer

1. Fit model with random intercept and note the $-2 \times \log(\text{likelihood})$ value: 9357.242
2. Fit model without random intercept and note the $-2 \times \log(\text{likelihood})$ value:

Exam scores example

1

Equations

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

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

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

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

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

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

2

Equations

$$\text{normexam}_{ij} = -0.001(0.013) + 0.595(0.013)\text{standlrt}_{ij} + e_{ij}$$

$$e_{ij} \sim \text{N}(0, \sigma^2_e) \quad \sigma^2_e = 0.648(0.014)$$

$-2*loglikelihood = 9760.509$(4059 of 4059 cases in use)

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

Exam scores example

1

Equations

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

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

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

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

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

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

2

Equations

$$\text{normexam}_{ij} = -0.001(0.013) + 0.595(0.013)\text{standlrt}_{ij} + e_{ij}$$

$$\text{e}_{ij} \sim \text{N}(0, \sigma^2_{e}) \quad \sigma^2_{e} = 0.648(0.014)$$

*-2*loglikelihood* = 9760.509(4059 of 4059 cases in use)

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

Exam scores example

Question

Do differences between schools in exam scores at age 16 remain after controlling for exam score at age 11?

Answer

1. Fit model with random intercept and note the $-2 \times \log(\text{likelihood})$ value: 9357.242
2. Fit model without random intercept and note the $-2 \times \log(\text{likelihood})$ value: 9760.509

Exam scores example

Question

Do differences between schools in exam scores at age 16 remain after controlling for exam score at age 11?

Answer

1. Fit model with random intercept and note the $-2 \times \log(\text{likelihood})$ value: 9357.242
2. Fit model without random intercept and note the $-2 \times \log(\text{likelihood})$ value: 9760.509
3. Form the test statistic:

Exam scores example

Question

Do differences between schools in exam scores at age 16 remain after controlling for exam score at age 11?

Answer

1. Fit model with random intercept and note the $-2 \times \log(\text{likelihood})$ value: 9357.242
2. Fit model without random intercept and note the $-2 \times \log(\text{likelihood})$ value: 9760.509
3. Form the test statistic: $9760.509 - 9357.242 = 403.267$

Exam scores example

Question

Do differences between schools in exam scores at age 16 remain after controlling for exam score at age 11?

Answer

1. Fit model with random intercept and note the $-2 \times \log(\text{likelihood})$ value: 9357.242
2. Fit model without random intercept and note the $-2 \times \log(\text{likelihood})$ value: 9760.509
3. Form the test statistic: $9760.509 - 9357.242 = 403.267$
4. Compare against the χ^2 distribution with 1 degree of freedom

Exam scores example

Question

Do differences between schools in exam scores at age 16 remain after controlling for exam score at age 11?

Answer

1. Fit model with random intercept and note the $-2 \times \log(\text{likelihood})$ value: 9357.242
2. Fit model without random intercept and note the $-2 \times \log(\text{likelihood})$ value: 9760.509
3. Form the test statistic: $9760.509 - 9357.242 = 403.267$
4. Compare against the χ^2 distribution with 1 degree of freedom $p = 1.0709 \times 10^{-89}$

Exam scores example

Question

Do differences between schools in exam scores at age 16 remain after controlling for exam score at age 11?

Answer

1. Fit model with random intercept and note the $-2 \times \log(\text{likelihood})$ value: 9357.242
2. Fit model without random intercept and note the $-2 \times \log(\text{likelihood})$ value: 9760.509
3. Form the test statistic: $9760.509 - 9357.242 = 403.267$
4. Compare against the χ^2 distribution with 1 degree of freedom $p = 1.0709 \times 10^{-89}$
5. We conclude that there are differences between schools in exam scores at age 16 after controlling for exam score at age 11

What questions can we answer?

We can use the model to answer two kinds of question:

About variables

About levels

Often, we will look at both kinds of question, even if our main focus is on one or the other

What questions can we answer?

We can use the model to answer two kinds of question:

About variables

- e.g. "What is the relationship between an individual's hedonism and their age?"

About levels

Often, we will look at both kinds of question, even if our main focus is on one or the other

What questions can we answer?

We can use the model to answer two kinds of question:

About variables

- e.g. "What is the relationship between an individual's hedonism and their age?"
- This is a question about means

About levels

Often, we will look at both kinds of question, even if our main focus is on one or the other

What questions can we answer?

We can use the model to answer two kinds of question:

About variables

- e.g. "What is the relationship between an individual's hedonism and their age?"
- This is a question about means
- It is answered using the fixed part: the slope β_1 of the overall regression line

About levels

Often, we will look at both kinds of question, even if our main focus is on one or the other

What questions can we answer?

We can use the model to answer two kinds of question:

About variables

- e.g. “What is the relationship between an individual’s hedonism and their age?”
- This is a question about means
- It is answered using the fixed part: the slope β_1 of the overall regression line
- σ_u^2 may be regarded as a ‘nuisance parameter’

About levels

Often, we will look at both kinds of question, even if our main focus is on one or the other

What questions can we answer?

We can use the model to answer two kinds of question:

About variables

- e.g. "What is the relationship between an individual's hedonism and their age?"
- This is a question about means
- It is answered using the fixed part: the slope β_1 of the overall regression line
- σ_u^2 may be regarded as a 'nuisance parameter'

About levels

- e.g. "How much variation in hedonism is due to country differences (after controlling for age)?"

Often, we will look at both kinds of question, even if our main focus is on one or the other

What questions can we answer?

We can use the model to answer two kinds of question:

About variables

- e.g. "What is the relationship between an individual's hedonism and their age?"
- This is a question about means
- It is answered using the fixed part: the slope β_1 of the overall regression line
- σ_u^2 may be regarded as a 'nuisance parameter'

About levels

- e.g. "How much variation in hedonism is due to country differences (after controlling for age)?"
- This is a question about variances

Often, we will look at both kinds of question, even if our main focus is on one or the other

What questions can we answer?

We can use the model to answer two kinds of question:

About variables

- e.g. "What is the relationship between an individual's hedonism and their age?"
- This is a question about means
- It is answered using the fixed part: the slope β_1 of the overall regression line
- σ_u^2 may be regarded as a 'nuisance parameter'

About levels

- e.g. "How much variation in hedonism is due to country differences (after controlling for age)?"
- This is a question about variances
- It is answered using the random part: the level 2 variance σ_u^2

Often, we will look at both kinds of question, even if our main focus is on one or the other

What questions can we answer?

We can use the model to answer two kinds of question:

About variables

- e.g. "What is the relationship between an individual's hedonism and their age?"
- This is a question about means
- It is answered using the fixed part: the slope β_1 of the overall regression line
- σ_u^2 may be regarded as a 'nuisance parameter'

About levels

- e.g. "How much variation in hedonism is due to country differences (after controlling for age)?"
- This is a question about variances
- It is answered using the random part: the level 2 variance σ_u^2
- β_1 may be regarded as a 'nuisance parameter'

Often, we will look at both kinds of question, even if our main focus is on one or the other

Examples of research

Muijs (2003)

Examples of research

Muijs (2003)

Question: Do pupils make more progress in maths when receiving support from numeracy support assistants?

Examples of research

Muijs (2003)

Question: Do pupils make more progress in maths when receiving support from numeracy support assistants? Answer: No

Examples of research

Muijs (2003)

Question: Do pupils make more progress in maths when receiving support from numeracy support assistants? Answer: No

Levels		Variance	
2	school	15.82	51.4%
1	pupil	14.93	48.6%

Examples of research

Muijs (2003)

Question: Do pupils make more progress in maths when receiving support from numeracy support assistants? Answer: No

Levels		Variance	
2	school	15.82	51.4%
1	pupil	14.93	48.6%

Adding in a variable for receiving support from an assistant to a model including pupil background characteristics and prior achievement made almost no difference to the school or pupil level variance, but adding in teacher effectiveness reduced the school level variance by 1.91 (12%) (while pupil level variance remained roughly the same)

Examples of research

Judge et al. (2006)

Examples of research

Judge et al. (2006)

Question: Is job satisfaction negatively related to workplace deviance (misbehaviour)?

Examples of research

Judge et al. (2006)

Question: Is job satisfaction negatively related to workplace deviance (misbehaviour)? Answer: Yes

Examples of research

Judge et al. (2006)

Question: Is job satisfaction negatively related to workplace deviance (misbehaviour)? Answer: Yes

	Levels	Variance
2	individual	Not reported
1	occasion	Not reported

Examples of research

Judge et al. (2006)

Question: Is job satisfaction negatively related to workplace deviance (misbehaviour)? Answer: Yes

	Levels	Variance
2	individual	Not reported
1	occasion	Not reported

The multilevel structure arises because we expect two measurements from the same person at different times to be similar (more similar than two measurements from different people)

Examples of research

Goldstein et al. (2007)

See also the Gallery of Multilevel papers on CMM's website
`http://www.bris.ac.uk/cmm/gallery`

Examples of research

Goldstein et al. (2007)

Question: Does which school a pupil attends affect their progress in maths between KS1 and KS2?

See also the Gallery of Multilevel papers on CMM's website
`http://www.bris.ac.uk/cmm/gallery`

Examples of research

Goldstein et al. (2007)

Question: Does which school a pupil attends affect their progress in maths between KS1 and KS2? Answer: Yes

See also the Gallery of Multilevel papers on CMM's website
`http://www.bris.ac.uk/cmm/gallery`

Examples of research

Goldstein et al. (2007)

Question: Does which school a pupil attends affect their progress in maths between KS1 and KS2? Answer: Yes

Levels		Variance	
2	junior school	0.05	14.9%
1	pupil	0.30	85.0%

See also the Gallery of Multilevel papers on CMM's website
`http://www.bris.ac.uk/cmm/gallery`

Examples of research

Goldstein et al. (2007)

Question: Does which school a pupil attends affect their progress in maths between KS1 and KS2? Answer: Yes

Levels		Variance	
2	junior school	0.05	14.9%
1	pupil	0.30	85.0%

The school level variance is found to be significant, so the authors conclude that which junior school a pupil attends does affect their progress in maths between KS1 and KS2

See also the Gallery of Multilevel papers on CMM's website
`http://www.bris.ac.uk/cmm/gallery`

Adding more explanatory variables

Extending our questions

- How much variation in pupils' progress is due to differences between schools after we control for pupil background characteristics such as SES, gender and ethnicity?

Adding more explanatory variables

Extending our questions

- How much variation in pupils' progress is due to differences between schools after we control for pupil background characteristics such as SES, gender and ethnicity?
- What are the relationships between hedonism and an individual's age, income, education and gender?

Adding more explanatory variables

Extending our questions

- How much variation in pupils' progress is due to differences between schools after we control for pupil background characteristics such as SES, gender and ethnicity?
- What are the relationships between hedonism and an individual's age, income, education and gender?

- We can easily add in more explanatory variables, as for a single level regression model

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3ij} + \beta_4 x_{2ij} x_{3ij} + \ldots + u_j + e_{ij}$$

Adding more explanatory variables

Extending our questions

- How much variation in pupils' progress is due to differences between schools after we control for pupil background characteristics such as SES, gender and ethnicity?
- What are the relationships between hedonism and an individual's age, income, education and gender?

- We can easily add in more explanatory variables, as for a single level regression model

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3ij} + \beta_4 x_{2ij} x_{3ij} + \ldots + u_j + e_{ij}$$

- Notice that we can also include interactions in the usual way

Adding more explanatory variables

Extending our questions

- How much variation in pupils' progress is due to differences between schools after we control for pupil background characteristics such as SES, gender and ethnicity?
- What are the relationships between hedonism and an individual's age, income, education and gender?

- We can easily add in more explanatory variables, as for a single level regression model

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3ij} + \beta_4 x_{2ij} x_{3ij} + \ldots + u_j + e_{ij}$$

- Notice that we can also include interactions in the usual way
- Variables can be continuous or categorical

Adding more explanatory variables

Extending our questions

- How much variation in pupils' progress is due to differences between schools after we control for pupil background characteristics such as SES, gender and ethnicity?
- What are the relationships between hedonism and an individual's age, income, education and gender?

- We can easily add in more explanatory variables, as for a single level regression model

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3ij} + \beta_4 x_{2ij} x_{3ij} + \ldots + u_j + e_{ij}$$

- Notice that we can also include interactions in the usual way
- Variables can be continuous or categorical
- Variables can be defined at a higher level (e.g. school mean intake score) (see Contextual effects session)

Adding more explanatory variables

Level 2 variance can increase

- When we add in a (level 1) variable, the variation at level 2 may decrease or increase (or stay the same)

Adding more explanatory variables

Level 2 variance can increase

- When we add in a (level 1) variable, the variation at level 2 may decrease or increase (or stay the same)
- However, the level 1 variation and the total residual variation will both decrease (or stay the same)

Adding more explanatory variables

Level 2 variance can increase

- When we add in a (level 1) variable, the variation at level 2 may decrease or increase (or stay the same)
- However, the level 1 variation and the total residual variation will both decrease (or stay the same)

This also applies when we add a variable to a variance components model to get a random intercept model.

Adding more explanatory variables

Level 2 variance can increase

- When we add in a (level 1) variable, the variation at level 2 may decrease or increase (or stay the same)
- However, the level 1 variation and the total residual variation will both decrease (or stay the same)

This also applies when we add a variable to a variance components model to get a random intercept model. An example is in modelling house prices (with area as level 2 and house as level 1).

Adding more explanatory variables

Level 2 variance can increase

- When we add in a (level 1) variable, the variation at level 2 may decrease or increase (or stay the same)
- However, the level 1 variation and the total residual variation will both decrease (or stay the same)

This also applies when we add a variable to a variance components model to get a random intercept model. An example is in modelling house prices (with area as level 2 and house as level 1).

- We start with a variance components model

Adding more explanatory variables

Level 2 variance can increase

- When we add in a (level 1) variable, the variation at level 2 may decrease or increase (or stay the same)
- However, the level 1 variation and the total residual variation will both decrease (or stay the same)

This also applies when we add a variable to a variance components model to get a random intercept model. An example is in modelling house prices (with area as level 2 and house as level 1).

- We start with a variance components model
- We add house size as an explanatory variable

Adding more explanatory variables

Level 2 variance can increase

- When we add in a (level 1) variable, the variation at level 2 may decrease or increase (or stay the same)
- However, the level 1 variation and the total residual variation will both decrease (or stay the same)

This also applies when we add a variable to a variance components model to get a random intercept model. An example is in modelling house prices (with area as level 2 and house as level 1).

- We start with a variance components model
- We add house size as an explanatory variable
- More expensive areas tend to have smaller houses

Adding more explanatory variables

Level 2 variance can increase

- When we add in a (level 1) variable, the variation at level 2 may decrease or increase (or stay the same)
- However, the level 1 variation and the total residual variation will both decrease (or stay the same)

This also applies when we add a variable to a variance components model to get a random intercept model. An example is in modelling house prices (with area as level 2 and house as level 1).

- We start with a variance components model
- We add house size as an explanatory variable
- More expensive areas tend to have smaller houses
- So before we add size, prices appear more similar across areas

Adding more explanatory variables

Level 2 variance can increase

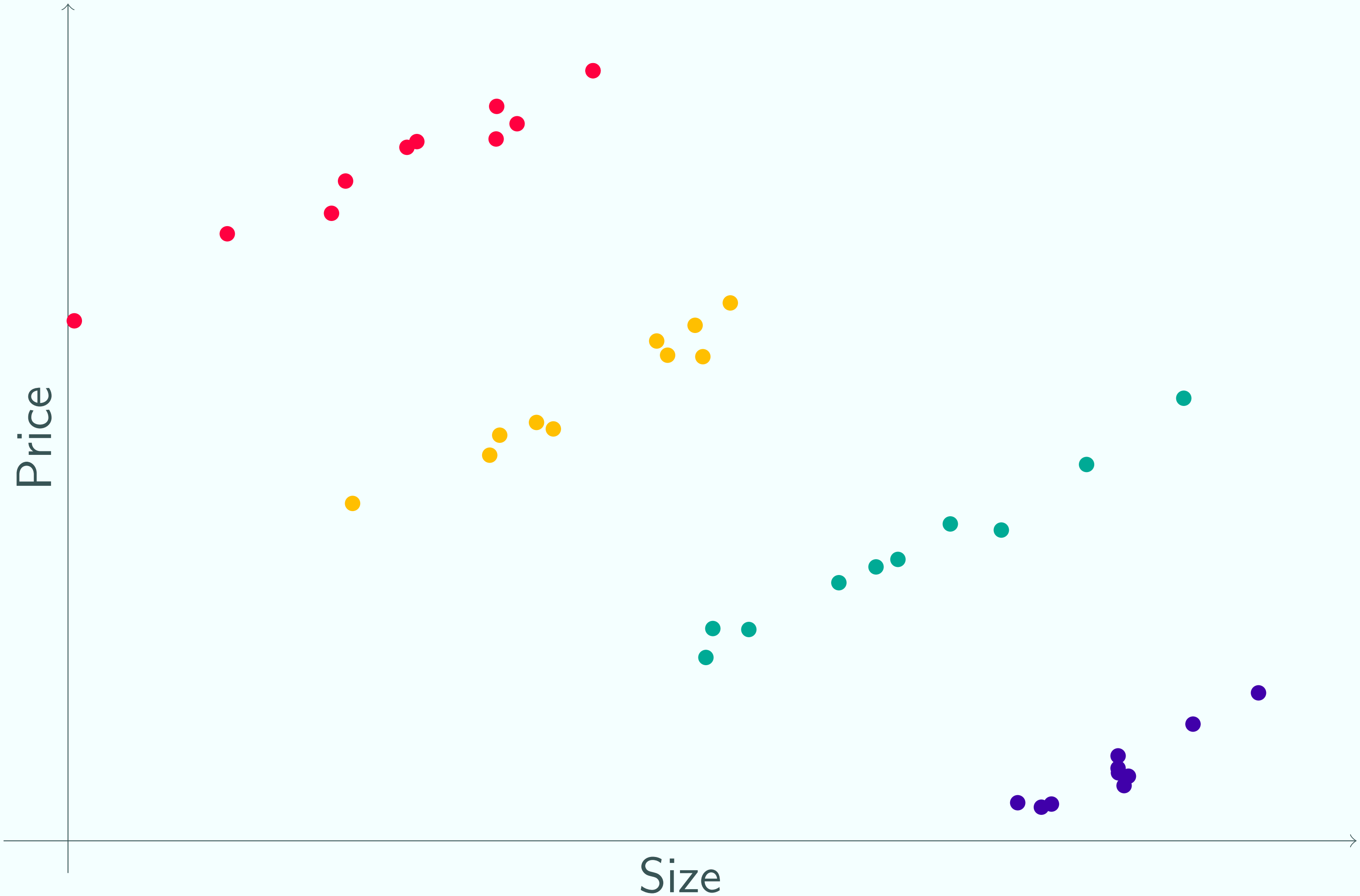
- When we add in a (level 1) variable, the variation at level 2 may decrease or increase (or stay the same)
- However, the level 1 variation and the total residual variation will both decrease (or stay the same)

This also applies when we add a variable to a variance components model to get a random intercept model. An example is in modelling house prices (with area as level 2 and house as level 1).

- We start with a variance components model
- We add house size as an explanatory variable
- More expensive areas tend to have smaller houses
- So before we add size, prices appear more similar across areas
- The variance components model has less variation at level 2 than the random intercept model

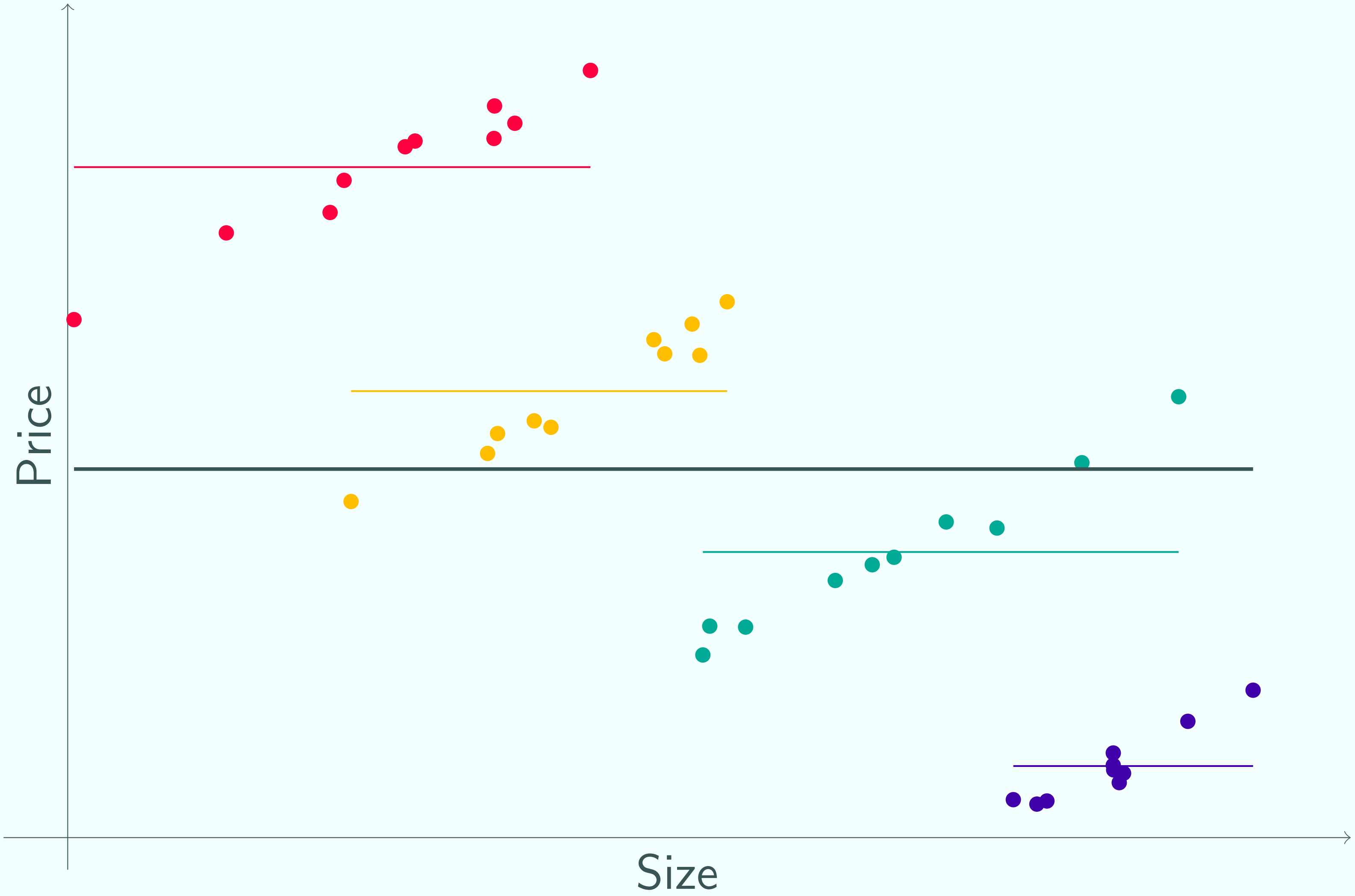
Level 2 variance can increase

House prices in 4 different areas of a city



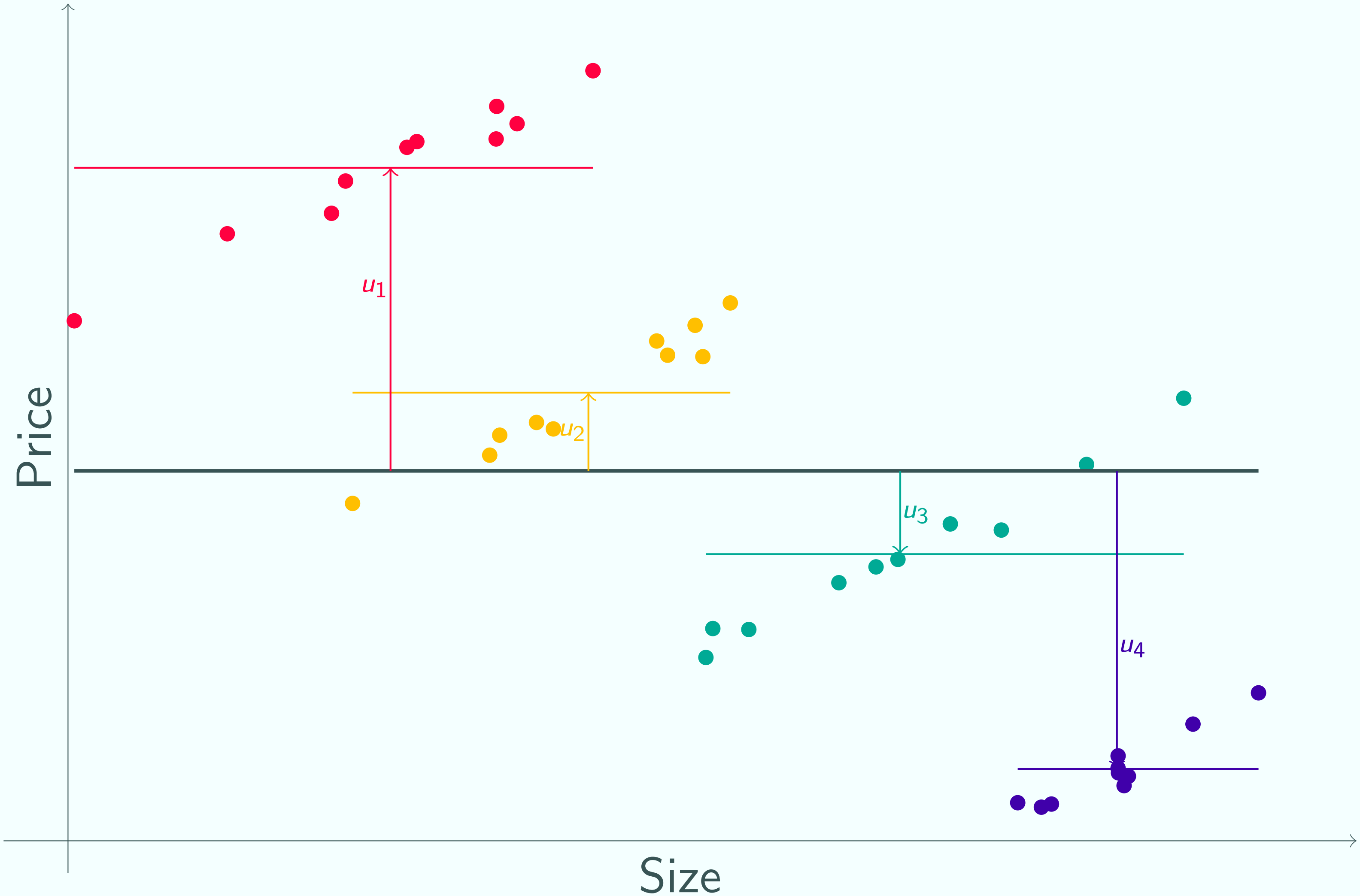
Level 2 variance can increase

House prices in 4 different areas of a city



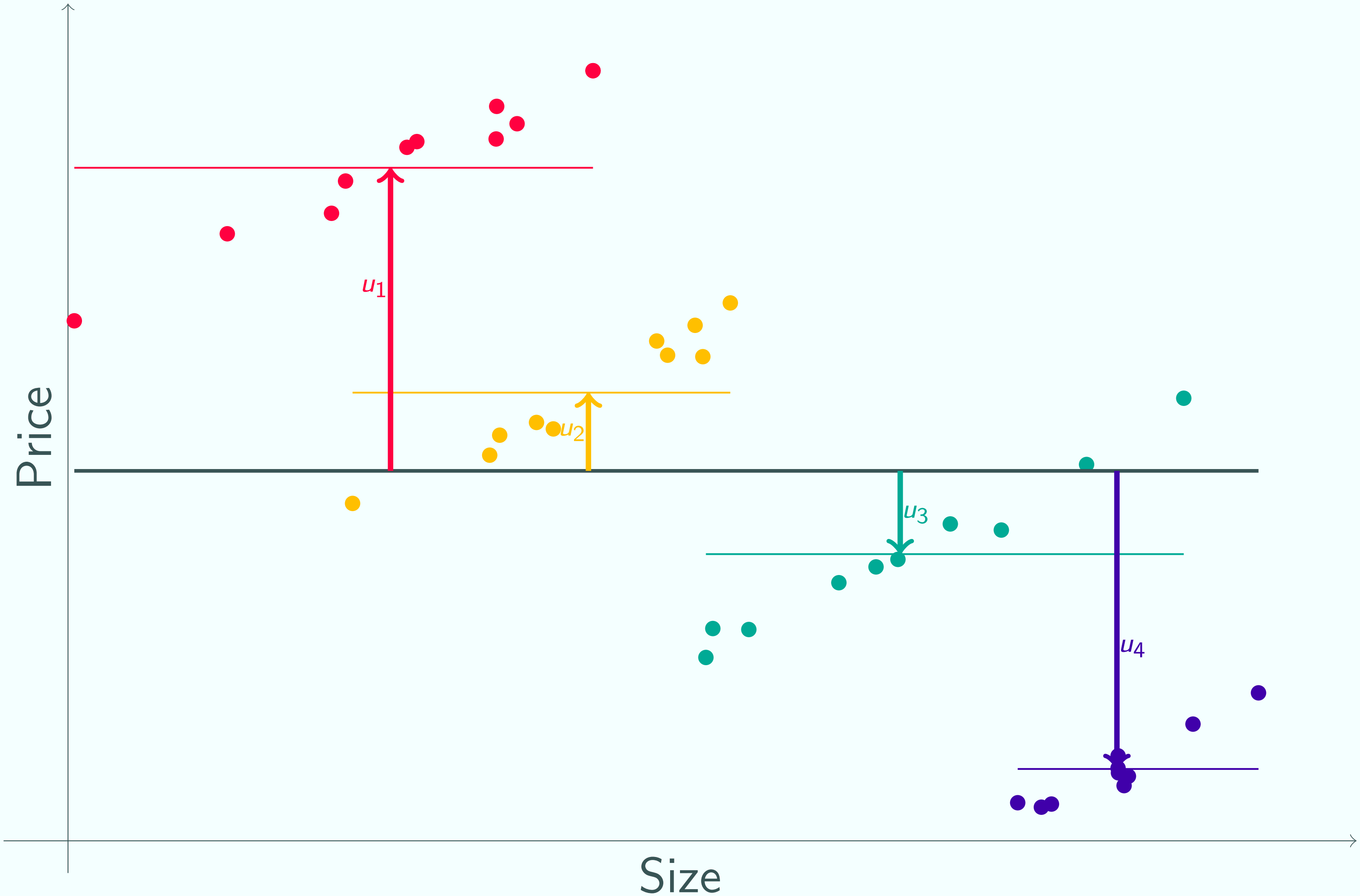
Level 2 variance can increase

House prices in 4 different areas of a city



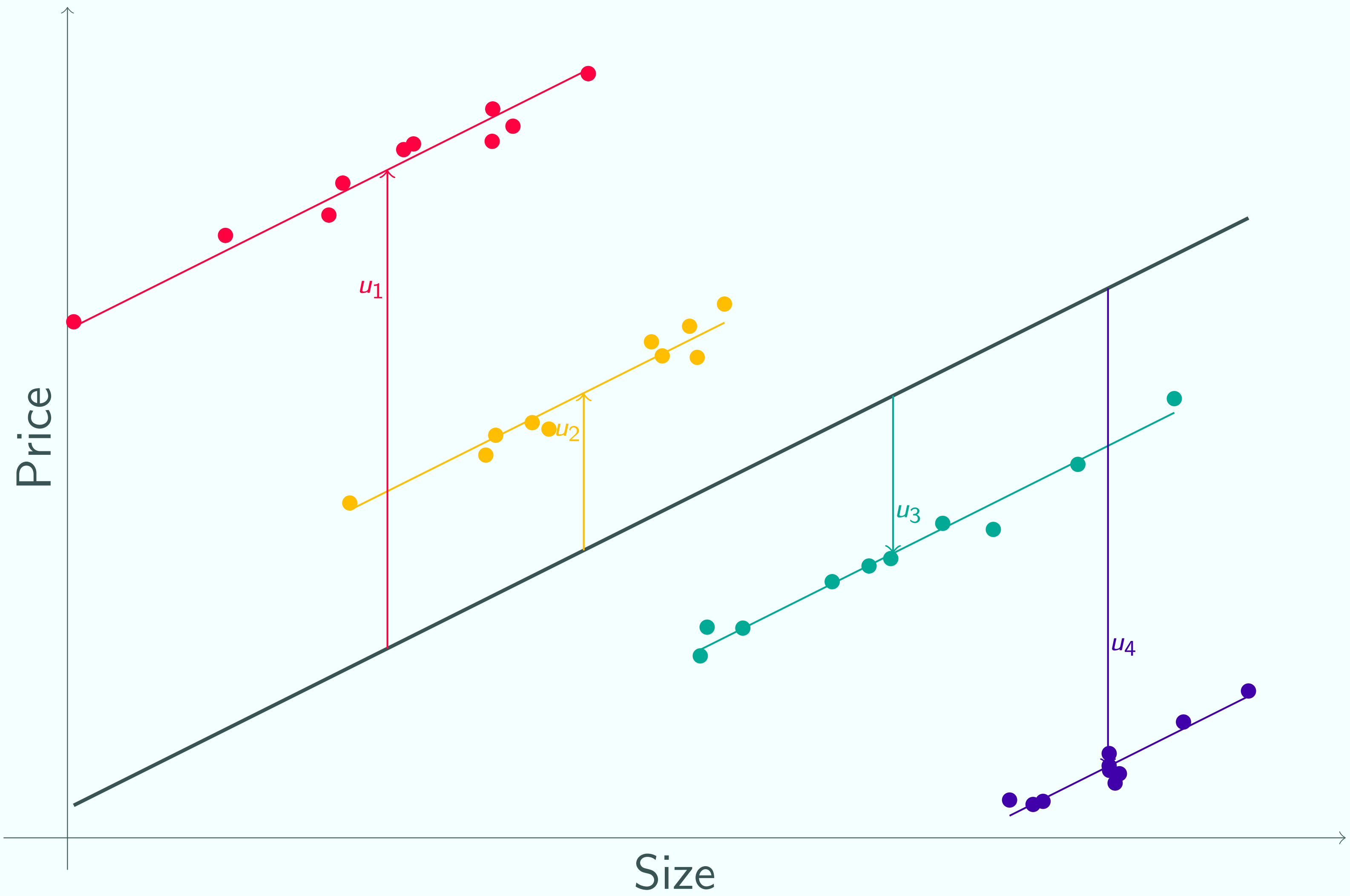
Level 2 variance can increase

House prices in 4 different areas of a city



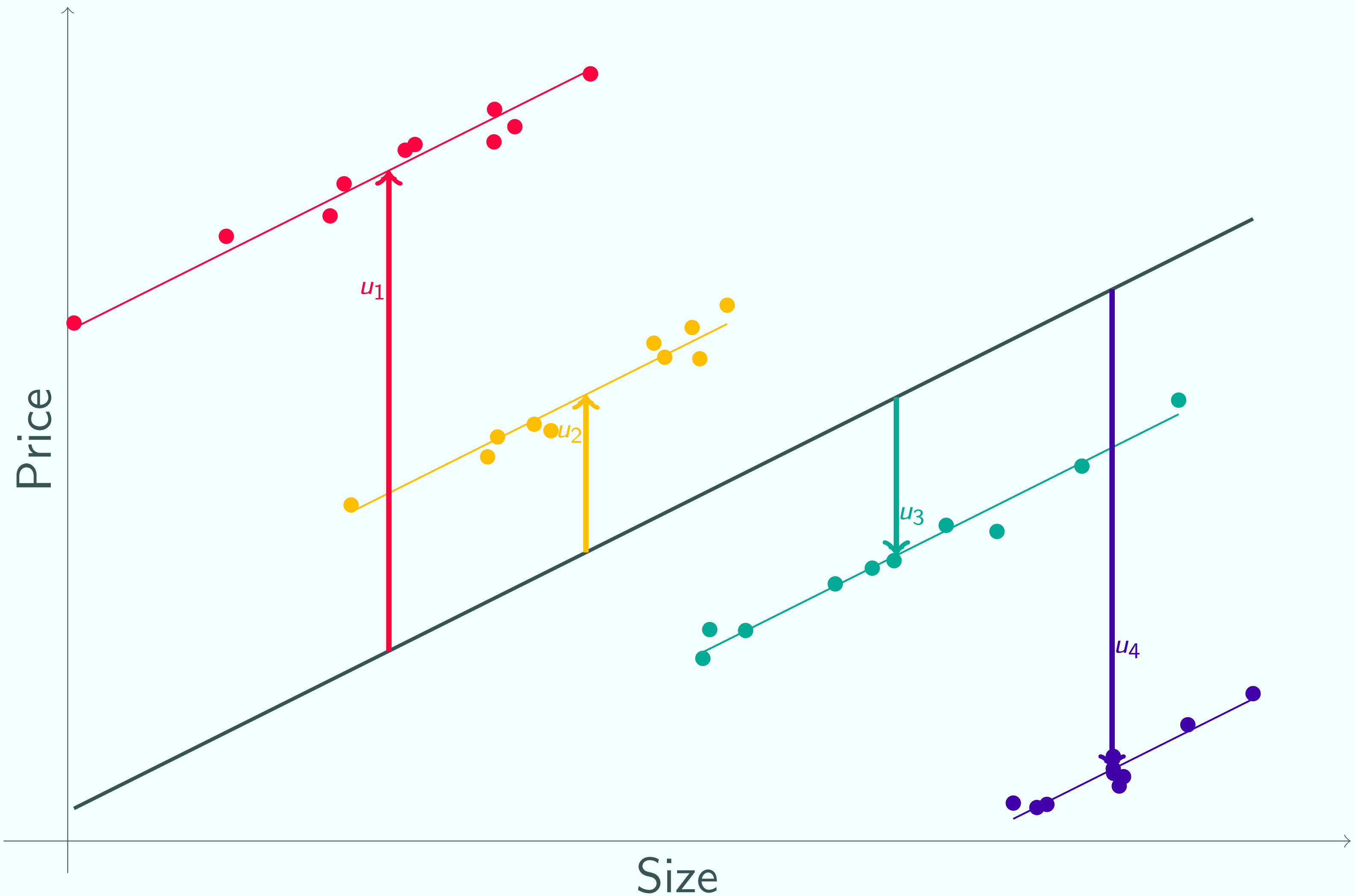
Level 2 variance can increase

House prices in 4 different areas of a city



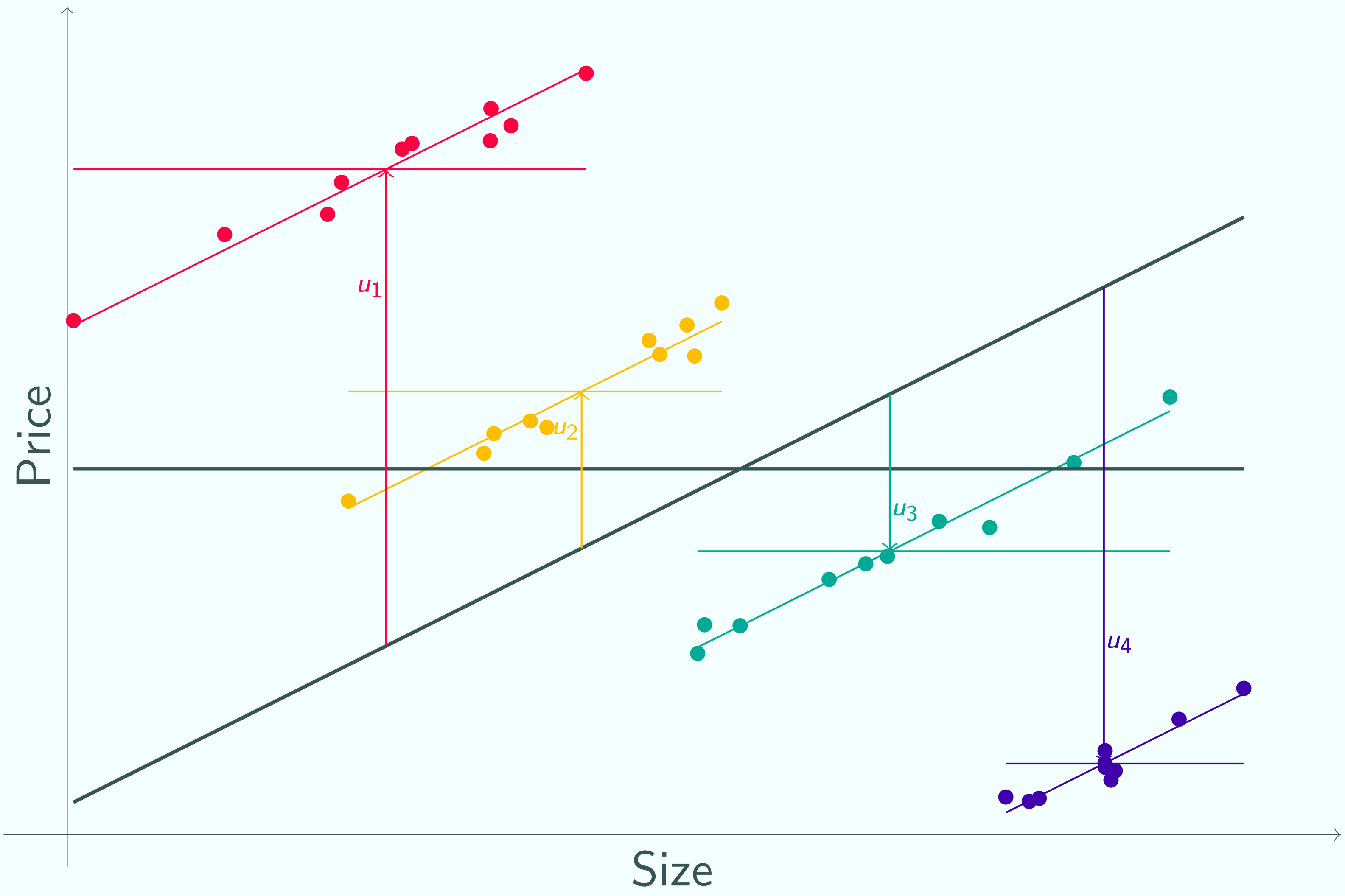
Level 2 variance can increase

House prices in 4 different areas of a city



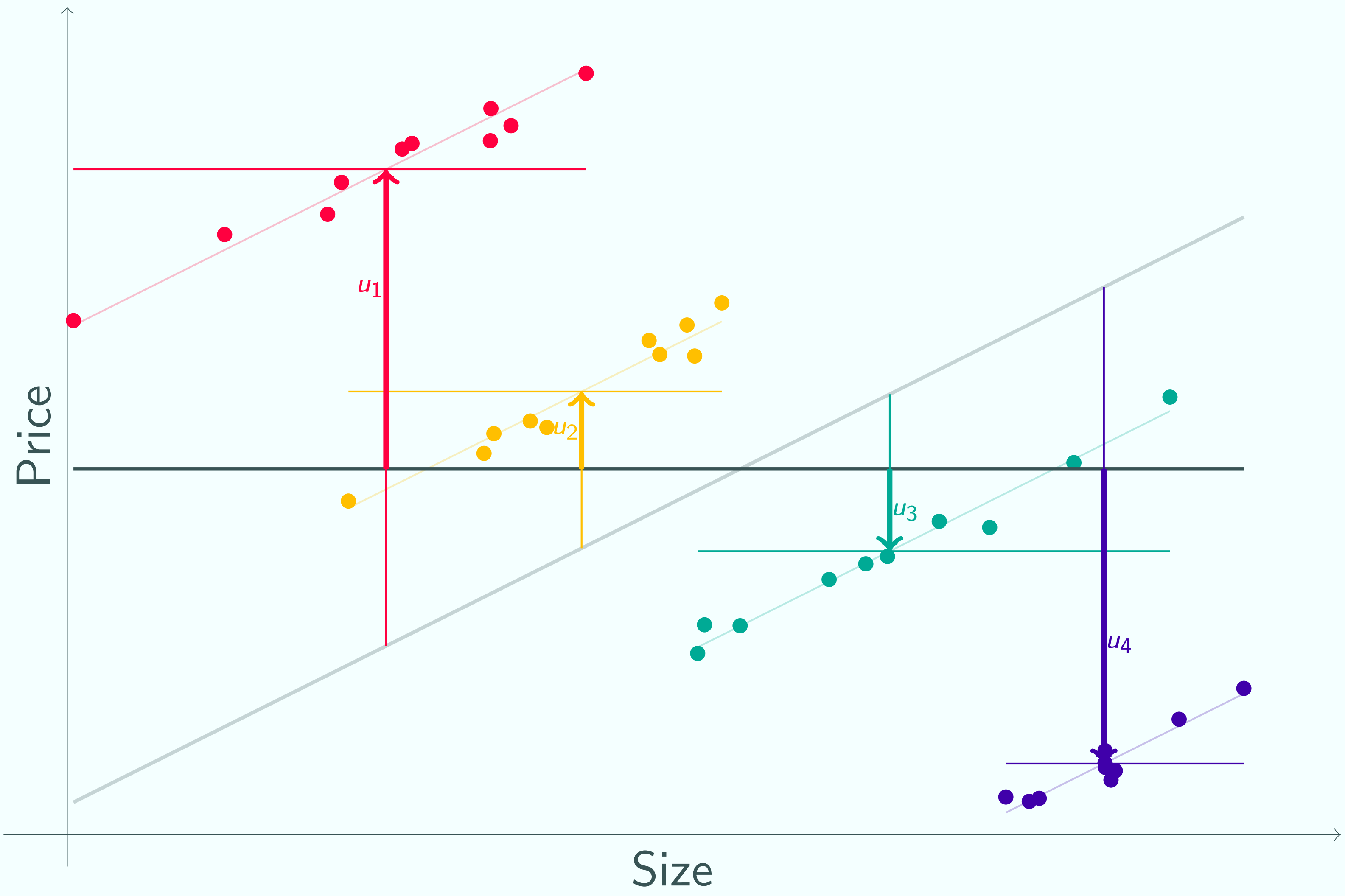
Level 2 variance can increase

House prices in 4 different areas of a city



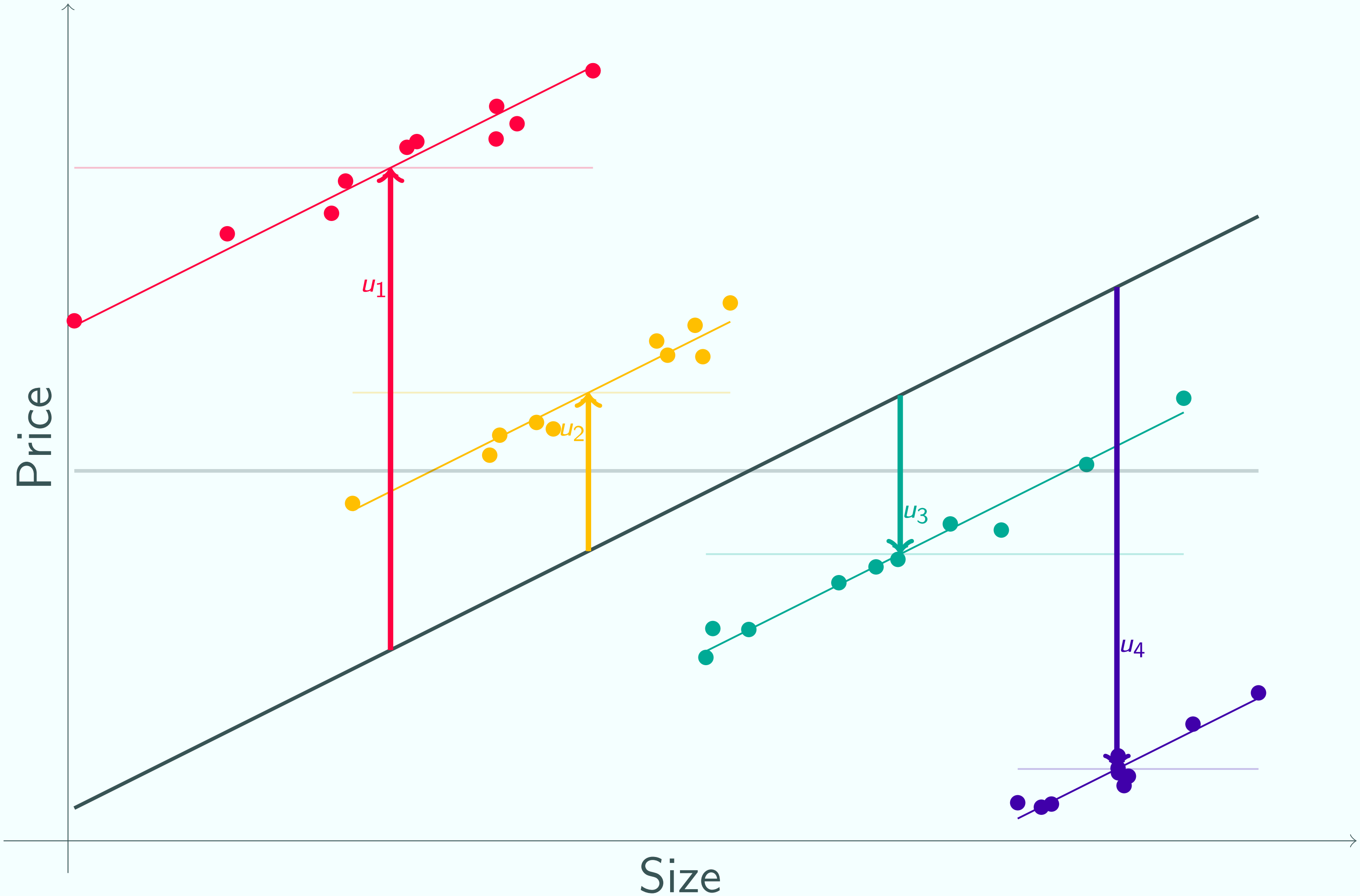
Level 2 variance can increase

House prices in 4 different areas of a city



Level 2 variance can increase

House prices in 4 different areas of a city



When is a variable a level?

- Sometimes it's very clear-cut

Exchangeability

When is a variable a level?

- Sometimes it's very clear-cut
  - we would not consider using a variable as a level

Exchangeability

When is a variable a level?

- Sometimes it's very clear-cut
  - we would not consider using a variable as a level
  - or else we would definitely want to put it in as a level

Exchangeability

When is a variable a level?

- Sometimes it's very clear-cut
  - we would not consider using a variable as a level
  - or else we would definitely want to put it in as a level
- In other cases, it's less obvious what we should do

Exchangeability

When is a variable a level?

- Sometimes it's very clear-cut
  - we would not consider using a variable as a level
  - or else we would definitely want to put it in as a level
- In other cases, it's less obvious what we should do
  - often when we have a relatively small number of units

Exchangeability

When is a variable a level?

- Sometimes it's very clear-cut
  - we would not consider using a variable as a level
  - or else we would definitely want to put it in as a level
- In other cases, it's less obvious what we should do
  - often when we have a relatively small number of units
- Basically, two things govern the decision:

Exchangeability

When is a variable a level?

- Sometimes it's very clear-cut
  - we would not consider using a variable as a level
  - or else we would definitely want to put it in as a level
- In other cases, it's less obvious what we should do
  - often when we have a relatively small number of units
- Basically, two things govern the decision:
  - number of units and

Exchangeability

When is a variable a level?

- Sometimes it's very clear-cut
  - we would not consider using a variable as a level
  - or else we would definitely want to put it in as a level
- In other cases, it's less obvious what we should do
  - often when we have a relatively small number of units
- Basically, two things govern the decision:
  - number of units and
  - exchangeability

Exchangeability

The units are exchangeable if we could randomly reassign their codes without losing any information

When is a variable a level?

- Sometimes it's very clear-cut
  - we would not consider using a variable as a level
  - or else we would definitely want to put it in as a level
- In other cases, it's less obvious what we should do
  - often when we have a relatively small number of units
- Basically, two things govern the decision:
  - number of units and
  - exchangeability
- The units should be a representative draw from a population or superpopulation (process)

Exchangeability

The units are exchangeable if we could randomly reassign their codes without losing any information

When is a variable a level?

- Sometimes it's very clear-cut
  - we would not consider using a variable as a level
  - or else we would definitely want to put it in as a level
- In other cases, it's less obvious what we should do
  - often when we have a relatively small number of units
- Basically, two things govern the decision:
  - number of units and
  - exchangeability
- The units should be a representative draw from a population or superpopulation (process)
- Exchangeability is a tricky concept, so it helps to draw up some guidelines

Exchangeability

The units are exchangeable if we could randomly reassign their codes without losing any information

Example

Hospital is probably a level

If we have data on treatment outcomes for patients in 100 UK hospitals:

Ethnicity is not a level

Example

Hospital is probably a level

If we have data on treatment outcomes for patients in 100 UK hospitals:

- You might naïvely use hospital as the units in a single level model (if you didn't know about multilevel modelling)

Ethnicity is not a level

Example

Hospital is probably a level

If we have data on treatment outcomes for patients in 100 UK hospitals:

- You might naïvely use hospital as the units in a single level model (if you didn't know about multilevel modelling)
- Hospital ID is a nominal variable whose categories have no special meanings

Ethnicity is not a level

Example

Hospital is probably a level

If we have data on treatment outcomes for patients in 100 UK hospitals:

- You might naïvely use hospital as the units in a single level model (if you didn't know about multilevel modelling)
- Hospital ID is a nominal variable whose categories have no special meanings
- We would not expect any particular hospital to give different results beforehand

Ethnicity is not a level

Example

Hospital is probably a level

If we have data on treatment outcomes for patients in 100 UK hospitals:

- You might naïvely use hospital as the units in a single level model (if you didn't know about multilevel modelling)
- Hospital ID is a nominal variable whose categories have no special meanings
- We would not expect any particular hospital to give different results beforehand

Ethnicity is not a level

If we have data on height for people of a variety of ethnicities:

Example

Hospital is probably a level

If we have data on treatment outcomes for patients in 100 UK hospitals:

- You might naïvely use hospital as the units in a single level model (if you didn't know about multilevel modelling)
- Hospital ID is a nominal variable whose categories have no special meanings
- We would not expect any particular hospital to give different results beforehand

Ethnicity is not a level

If we have data on height for people of a variety of ethnicities:

- Even if we could measure ethnicity very finely using many categories, we would not use it as the units in a single level model

Example

Hospital is probably a level

If we have data on treatment outcomes for patients in 100 UK hospitals:

- You might naïvely use hospital as the units in a single level model (if you didn't know about multilevel modelling)
- Hospital ID is a nominal variable whose categories have no special meanings
- We would not expect any particular hospital to give different results beforehand

Ethnicity is not a level

If we have data on height for people of a variety of ethnicities:

- Even if we could measure ethnicity very finely using many categories, we would not use it as the units in a single level model
- The categories of ethnicity have a special meaning

Example

Hospital is probably a level

If we have data on treatment outcomes for patients in 100 UK hospitals:

- You might naïvely use hospital as the units in a single level model (if you didn't know about multilevel modelling)
- Hospital ID is a nominal variable whose categories have no special meanings
- We would not expect any particular hospital to give different results beforehand

Ethnicity is not a level

If we have data on height for people of a variety of ethnicities:

- Even if we could measure ethnicity very finely using many categories, we would not use it as the units in a single level model
- The categories of ethnicity have a special meaning
- We expect different results for different categories

Example

Hospital is probably a level

If we have data on treatment outcomes for patients in 100 UK hospitals:

- You might naïvely use hospital as the units in a single level model (if you didn't know about multilevel modelling)
- Hospital ID is a nominal variable whose categories have no special meanings
- We would not expect any particular hospital to give different results beforehand

Ethnicity is not a level

If we have data on height for people of a variety of ethnicities:

- Even if we could measure ethnicity very finely using many categories, we would not use it as the units in a single level model
- The categories of ethnicity have a special meaning
- We expect different results for different categories
- We probably don't have many different categories in any case

Exercises: Session 1

Variance partitioning coefficients

- For variance components models, we saw that the VPC is a useful way to see how the variance divides up

Calculating the VPC

Variance partitioning coefficients

- For variance components models, we saw that the VPC is a useful way to see how the variance divides up
- This is even more true for random intercept models, since the total amount of variance may change as we add explanatory variables, making comparison hard

Calculating the VPC

Variance partitioning coefficients

- For variance components models, we saw that the VPC is a useful way to see how the variance divides up
- This is even more true for random intercept models, since the total amount of variance may change as we add explanatory variables, making comparison hard

Note

We most often use ‘Level 1 variance’ to mean ‘Residual variance at level 1’ – not ‘Variance of y at level 1’. Similarly for ‘Level 2 variance’.

Calculating the VPC

Variance partitioning coefficients

- For variance components models, we saw that the VPC is a useful way to see how the variance divides up
- This is even more true for random intercept models, since the total amount of variance may change as we add explanatory variables, making comparison hard

Note

We most often use 'Level 1 variance' to mean 'Residual variance at level 1' – not 'Variance of y at level 1'. Similarly for 'Level 2 variance'.

Calculating the VPC

- Recall that

$$\rho = \frac{\text{Level 2 variance}}{\text{Total residual variance}}$$

Variance partitioning coefficients

- For variance components models, we saw that the VPC is a useful way to see how the variance divides up
- This is even more true for random intercept models, since the total amount of variance may change as we add explanatory variables, making comparison hard

Note

We most often use 'Level 1 variance' to mean 'Residual variance at level 1' – not 'Variance of y at level 1'. Similarly for 'Level 2 variance'.

Calculating the VPC

- Recall that

$$\rho = \frac{\text{Level 2 variance}}{\text{Total residual variance}}$$

- For a random intercept model,

Variance partitioning coefficients

- For variance components models, we saw that the VPC is a useful way to see how the variance divides up
- This is even more true for random intercept models, since the total amount of variance may change as we add explanatory variables, making comparison hard

Note

We most often use 'Level 1 variance' to mean 'Residual variance at level 1' – not 'Variance of y at level 1'. Similarly for 'Level 2 variance'.

Calculating the VPC

- Recall that

$$\rho = \frac{\text{Level 2 variance}}{\text{Total residual variance}}$$

- For a random intercept model,
  - Level 1 variance $= \sigma_e^2$

Variance partitioning coefficients

- For variance components models, we saw that the VPC is a useful way to see how the variance divides up
- This is even more true for random intercept models, since the total amount of variance may change as we add explanatory variables, making comparison hard

Note

We most often use 'Level 1 variance' to mean 'Residual variance at level 1' – not 'Variance of y at level 1'. Similarly for 'Level 2 variance'.

Calculating the VPC

- Recall that

$$\rho = \frac{\text{Level 2 variance}}{\text{Total residual variance}}$$

- For a random intercept model,
  - Level 1 variance $= \sigma_e^2$
  - Level 2 variance $= \sigma_u^2$

Variance partitioning coefficients

- For variance components models, we saw that the VPC is a useful way to see how the variance divides up
- This is even more true for random intercept models, since the total amount of variance may change as we add explanatory variables, making comparison hard

Note

We most often use 'Level 1 variance' to mean 'Residual variance at level 1' – not 'Variance of y at level 1'. Similarly for 'Level 2 variance'.

Calculating the VPC

- Recall that

$$\rho = \frac{\text{Level 2 variance}}{\text{Total residual variance}}$$

- For a random intercept model,
  - Level 1 variance $= \sigma_e^2$
  - Level 2 variance $= \sigma_u^2$
  - Total residual variance $= \sigma_e^2 + \sigma_u^2$

Variance partitioning coefficients

- For variance components models, we saw that the VPC is a useful way to see how the variance divides up
- This is even more true for random intercept models, since the total amount of variance may change as we add explanatory variables, making comparison hard

Note

We most often use 'Level 1 variance' to mean 'Residual variance at level 1' – not 'Variance of y at level 1'. Similarly for 'Level 2 variance'.

Calculating the VPC

- Recall that

$$\rho = \frac{\text{Level 2 variance}}{\text{Total residual variance}}$$

- For a random intercept model,
  - Level 1 variance $= \sigma_e^2$
  - Level 2 variance $= \sigma_u^2$
  - Total residual variance $= \sigma_e^2 + \sigma_u^2$
- So

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

Variance partitioning coefficients

- For variance components models, we saw that the VPC is a useful way to see how the variance divides up
- This is even more true for random intercept models, since the total amount of variance may change as we add explanatory variables, making comparison hard

Note

We most often use 'Level 1 variance' to mean 'Residual variance at level 1' – not 'Variance of y at level 1'. Similarly for 'Level 2 variance'.

Calculating the VPC

- Recall that

$$\rho = \frac{\text{Level 2 variance}}{\text{Total residual variance}}$$

- For a random intercept model,
  - Level 1 variance $= \sigma_e^2$
  - Level 2 variance $= \sigma_u^2$
  - Total residual variance $= \sigma_e^2 + \sigma_u^2$
- So

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

- This is just the same as for the variance components model

Exam scores example

Question

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

Answer

Exam scores example

Question

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

Answer

1. Fit our random intercepts model and note the variances

Equations

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

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

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

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

$$-2{*}loglikelihood = 9357.242(4059 \text{ of } 4059 \text{ cases in use})$$

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

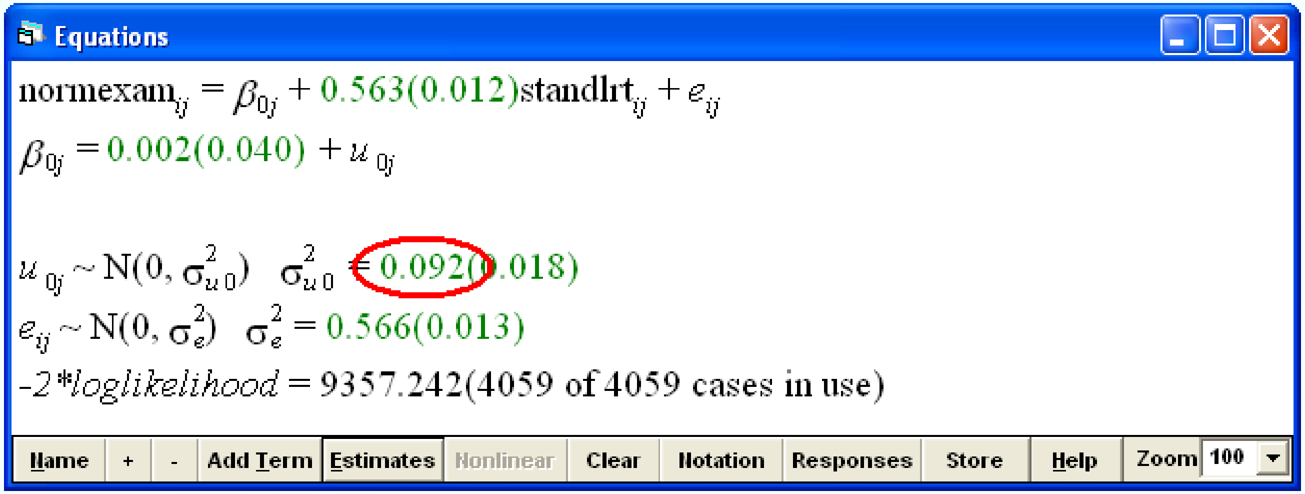
Exam scores example

Question

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

Answer

1. Fit our random intercepts model and note the variances
   Level 2: 0.092



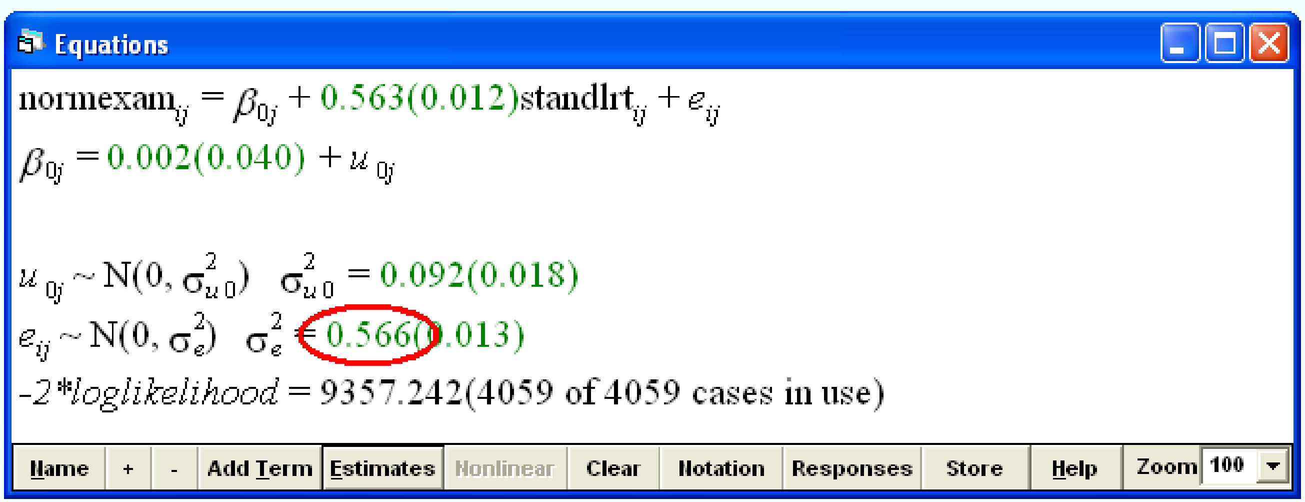
Exam scores example

Question

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

Answer

1. Fit our random intercepts model and note the variances
   Level 2: 0.092, Level 1: 0.566



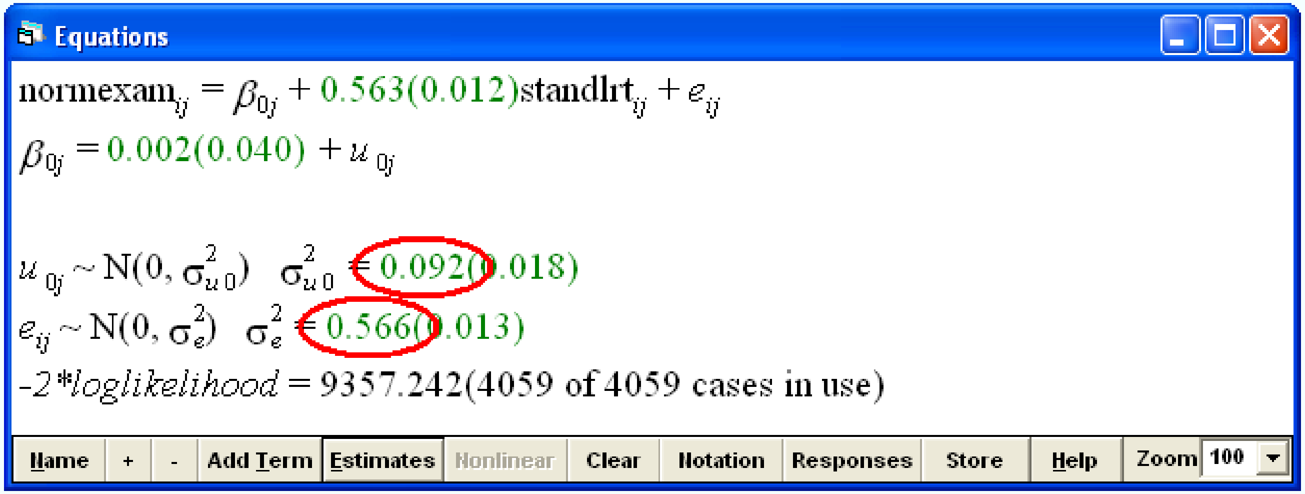
Exam scores example

Question

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

Answer

1. Fit our random intercepts model and note the variances
   Level 2: 0.092, Level 1: 0.566
2. Calculate proportion of variance at level 2



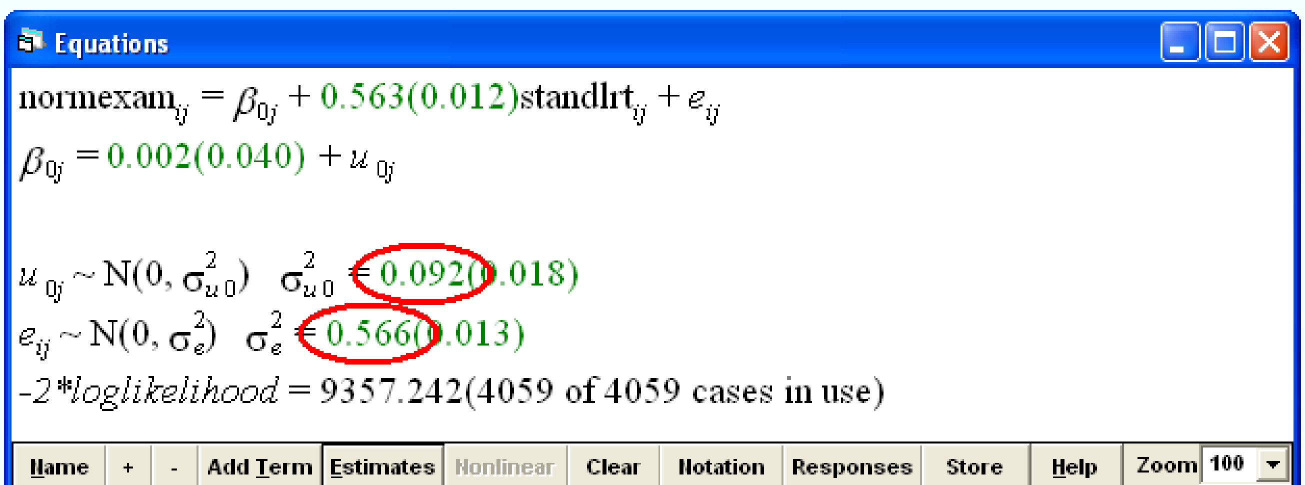
Exam scores example

Question

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

Answer

1. Fit our random intercepts model and note the variances
   Level 2: 0.092, Level 1: 0.566
2. Calculate proportion of variance at level 2
   0.092/(0.092+0.566) = 13.9%



ρ and clustering

Another way to think of ρ is that it measures the clustering

Large ρ

Large $\rho \Rightarrow$ a lot of clustering

Small ρ

Small $\rho \Rightarrow$ little clustering

ρ and clustering

Another way to think of ρ is that it measures the clustering

Large ρ

- When ρ is large, a lot of the variance is at level 2

Large $\rho \Rightarrow$ a lot of clustering

Small ρ

Small $\rho \Rightarrow$ little clustering

ρ and clustering

Another way to think of ρ is that it measures the clustering

Large ρ

- When ρ is large, a lot of the variance is at level 2
- so units within each group are quite similar

Large $\rho \Rightarrow$ a lot of clustering

Small ρ

Small $\rho \Rightarrow$ little clustering

ρ and clustering

Another way to think of ρ is that it measures the clustering

Large ρ

- When ρ is large, a lot of the variance is at level 2
- so units within each group are quite similar
- but there is a lot of difference between groups

Large $\rho \Rightarrow$ a lot of clustering

Small ρ

Small $\rho \Rightarrow$ little clustering

ρ and clustering

Another way to think of ρ is that it measures the clustering

Large ρ

- When ρ is large, a lot of the variance is at level 2
- so units within each group are quite similar
- but there is a lot of difference between groups
- Values of the response are largely determined by which group the unit belongs to

Large $\rho \Rightarrow$ a lot of clustering

Small ρ

Small $\rho \Rightarrow$ little clustering

ρ and clustering

Another way to think of ρ is that it measures the clustering

Large ρ

- When ρ is large, a lot of the variance is at level 2
- so units within each group are quite similar
- but there is a lot of difference between groups
- Values of the response are largely determined by which group the unit belongs to
- So the data are very clustered

Large $\rho \Rightarrow$ a lot of clustering

Small ρ

Small $\rho \Rightarrow$ little clustering

ρ and clustering

Another way to think of ρ is that it measures the clustering

Large ρ

- When ρ is large, a lot of the variance is at level 2
- so units within each group are quite similar
- but there is a lot of difference between groups
- Values of the response are largely determined by which group the unit belongs to
- So the data are very clustered

Large $\rho \Rightarrow$ a lot of clustering

Small ρ

- When ρ is small, not much variance is at level 2

Small $\rho \Rightarrow$ little clustering

ρ and clustering

Another way to think of ρ is that it measures the clustering

Large ρ

- When ρ is large, a lot of the variance is at level 2
- so units within each group are quite similar
- but there is a lot of difference between groups
- Values of the response are largely determined by which group the unit belongs to
- So the data are very clustered

Large $\rho \Rightarrow$ a lot of clustering

Small ρ

- When ρ is small, not much variance is at level 2
- so units within each group may be quite dissimilar

Small $\rho \Rightarrow$ little clustering

ρ and clustering

Another way to think of ρ is that it measures the clustering

Large ρ

- When ρ is large, a lot of the variance is at level 2
- so units within each group are quite similar
- but there is a lot of difference between groups
- Values of the response are largely determined by which group the unit belongs to
- So the data are very clustered

Large $\rho \Rightarrow$ a lot of clustering

Small ρ

- When ρ is small, not much variance is at level 2
- so units within each group may be quite dissimilar
- but there is not much difference between groups

Small $\rho \Rightarrow$ little clustering

ρ and clustering

Another way to think of ρ is that it measures the clustering

Large ρ

- When ρ is large, a lot of the variance is at level 2
- so units within each group are quite similar
- but there is a lot of difference between groups
- Values of the response are largely determined by which group the unit belongs to
- So the data are very clustered

Large $\rho \Rightarrow$ a lot of clustering

Small ρ

- When ρ is small, not much variance is at level 2
- so units within each group may be quite dissimilar
- but there is not much difference between groups
- Which group a unit belongs to does not have much impact on the response

Small $\rho \Rightarrow$ little clustering

ρ and clustering

Another way to think of ρ is that it measures the clustering

Large ρ

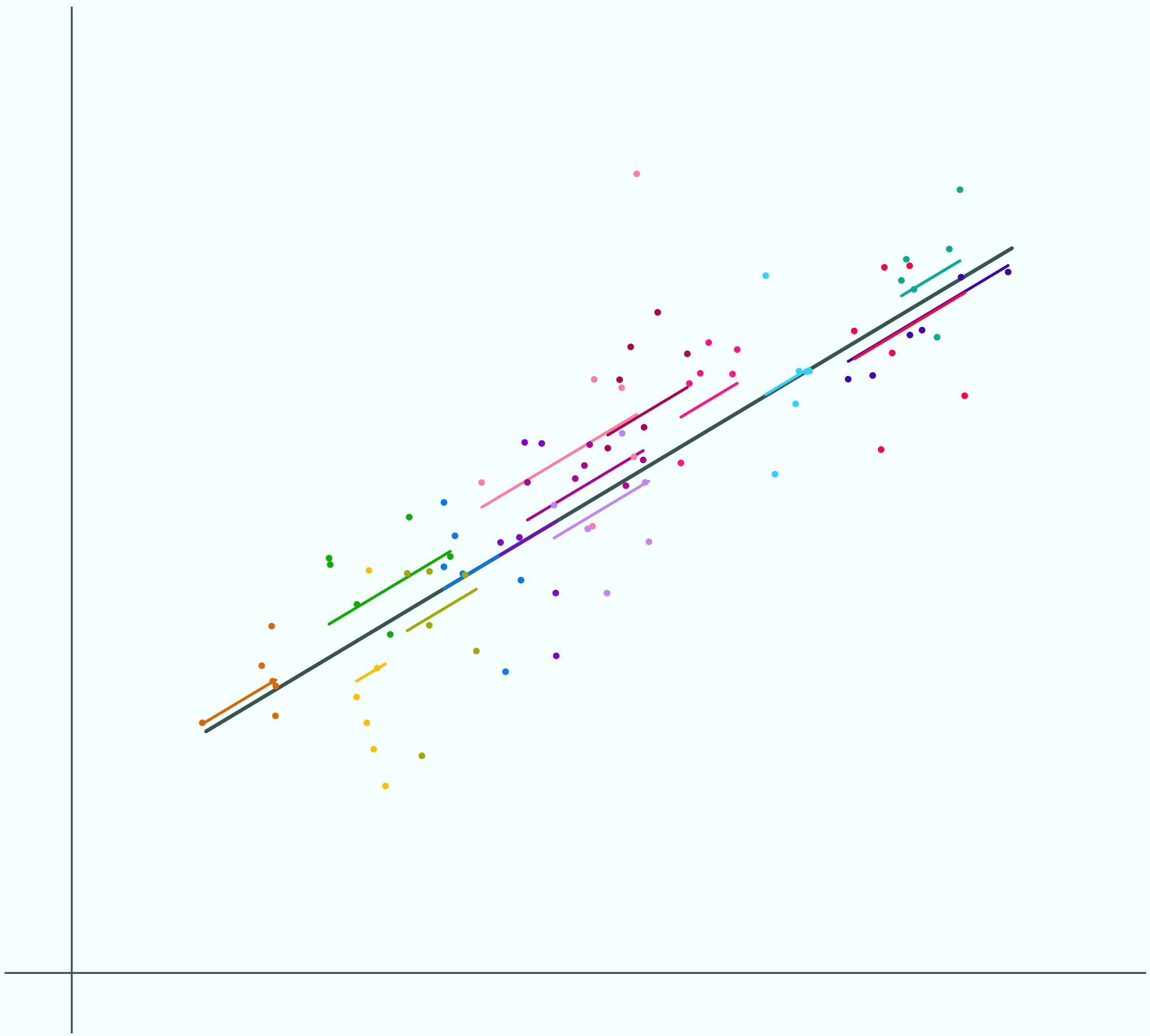
- When ρ is large, a lot of the variance is at level 2
- so units within each group are quite similar
- but there is a lot of difference between groups
- Values of the response are largely determined by which group the unit belongs to
- So the data are very clustered

Large $\rho \Rightarrow$ a lot of clustering

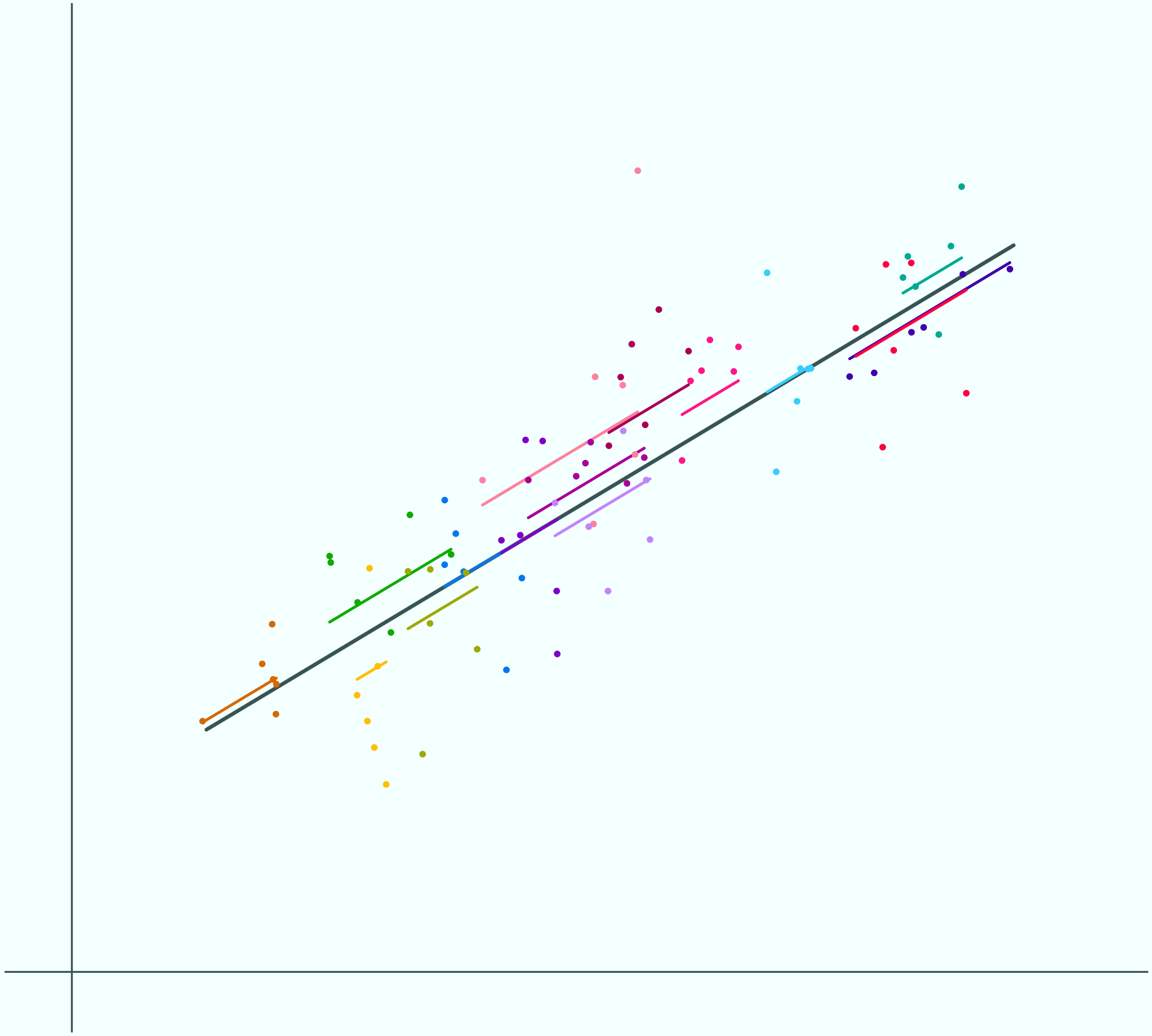
Small ρ

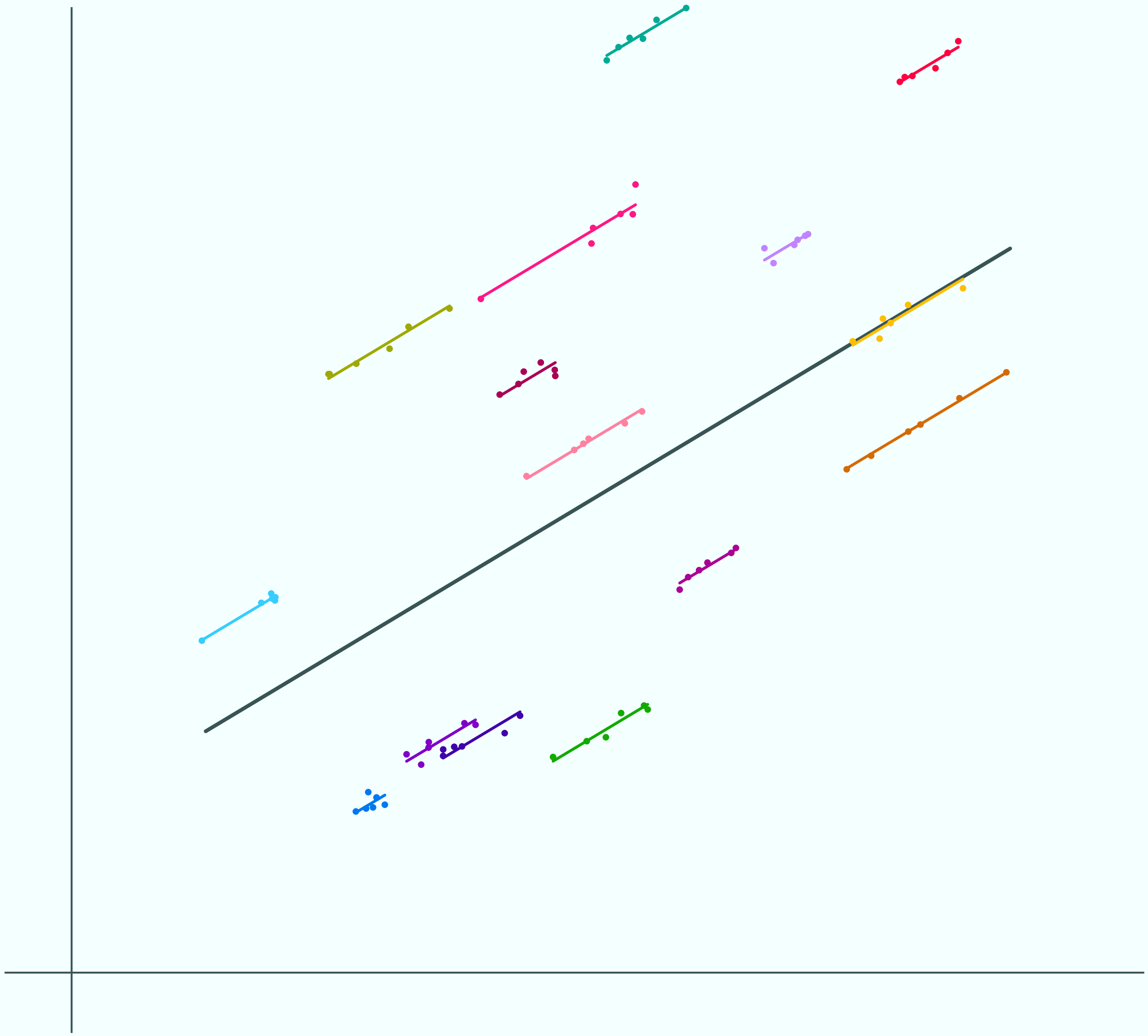
- When ρ is small, not much variance is at level 2
- so units within each group may be quite dissimilar
- but there is not much difference between groups
- Which group a unit belongs to does not have much impact on the response
- So the data are not very clustered

Small $\rho \Rightarrow$ little clustering

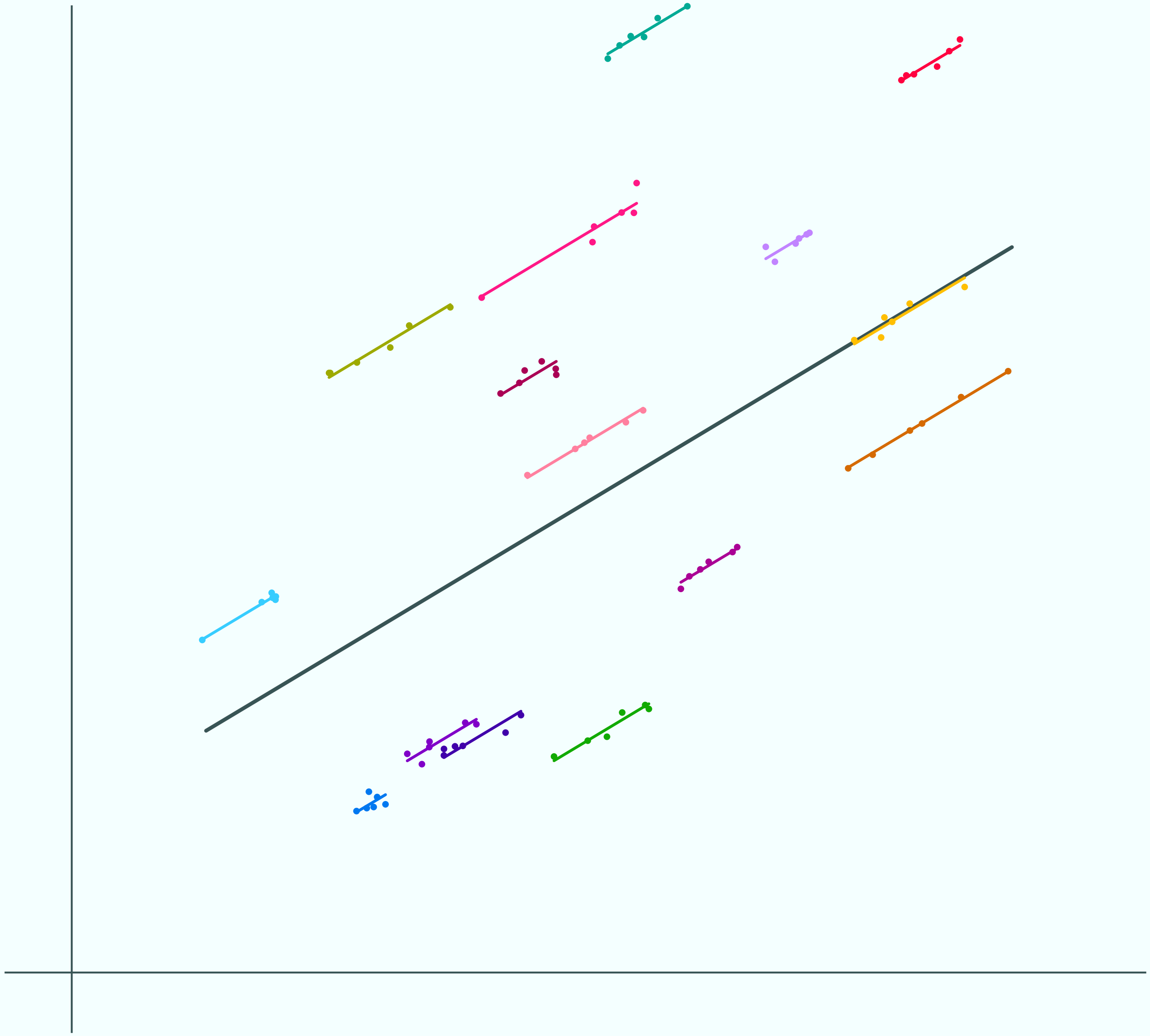


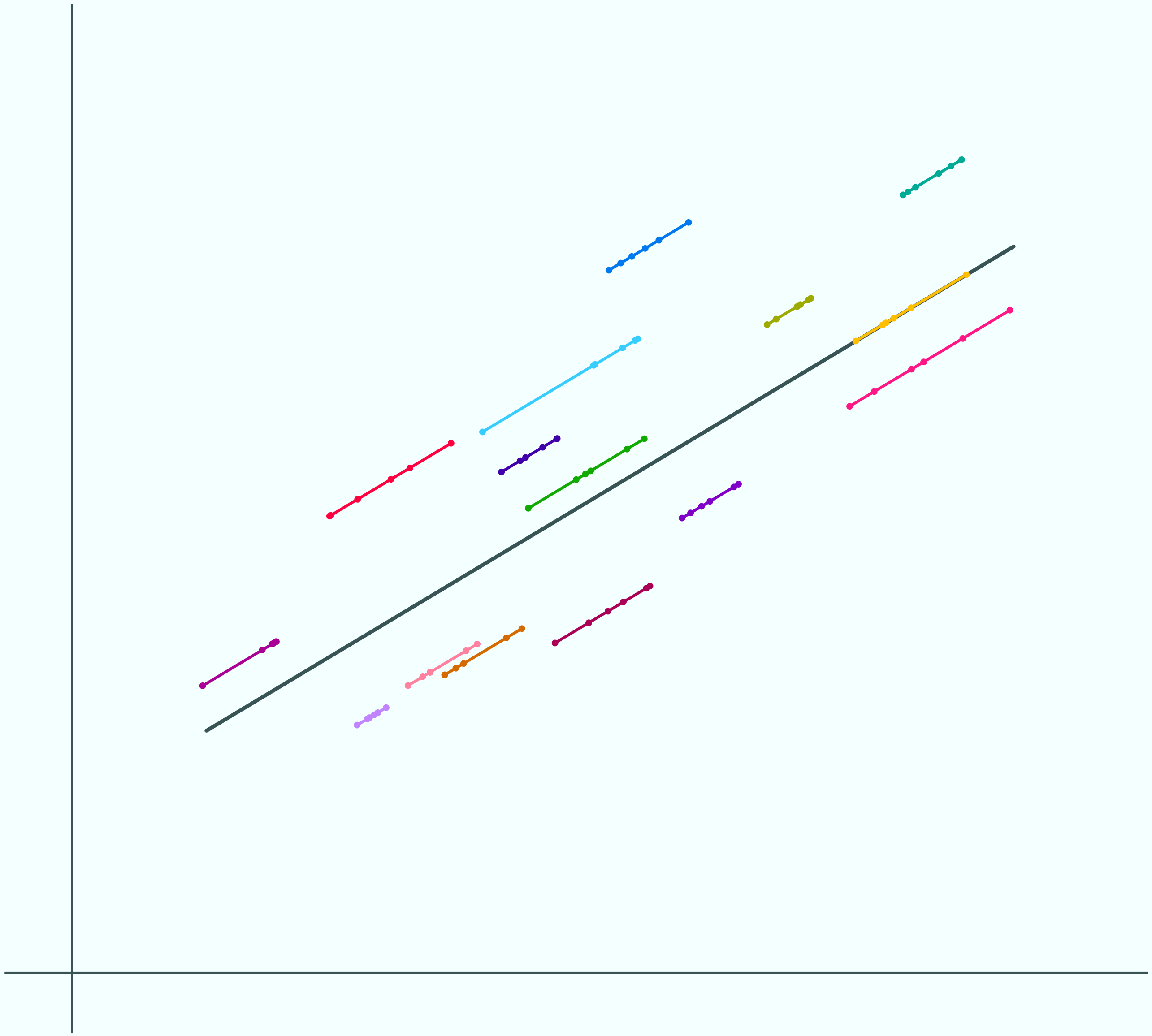
A small value of ρ



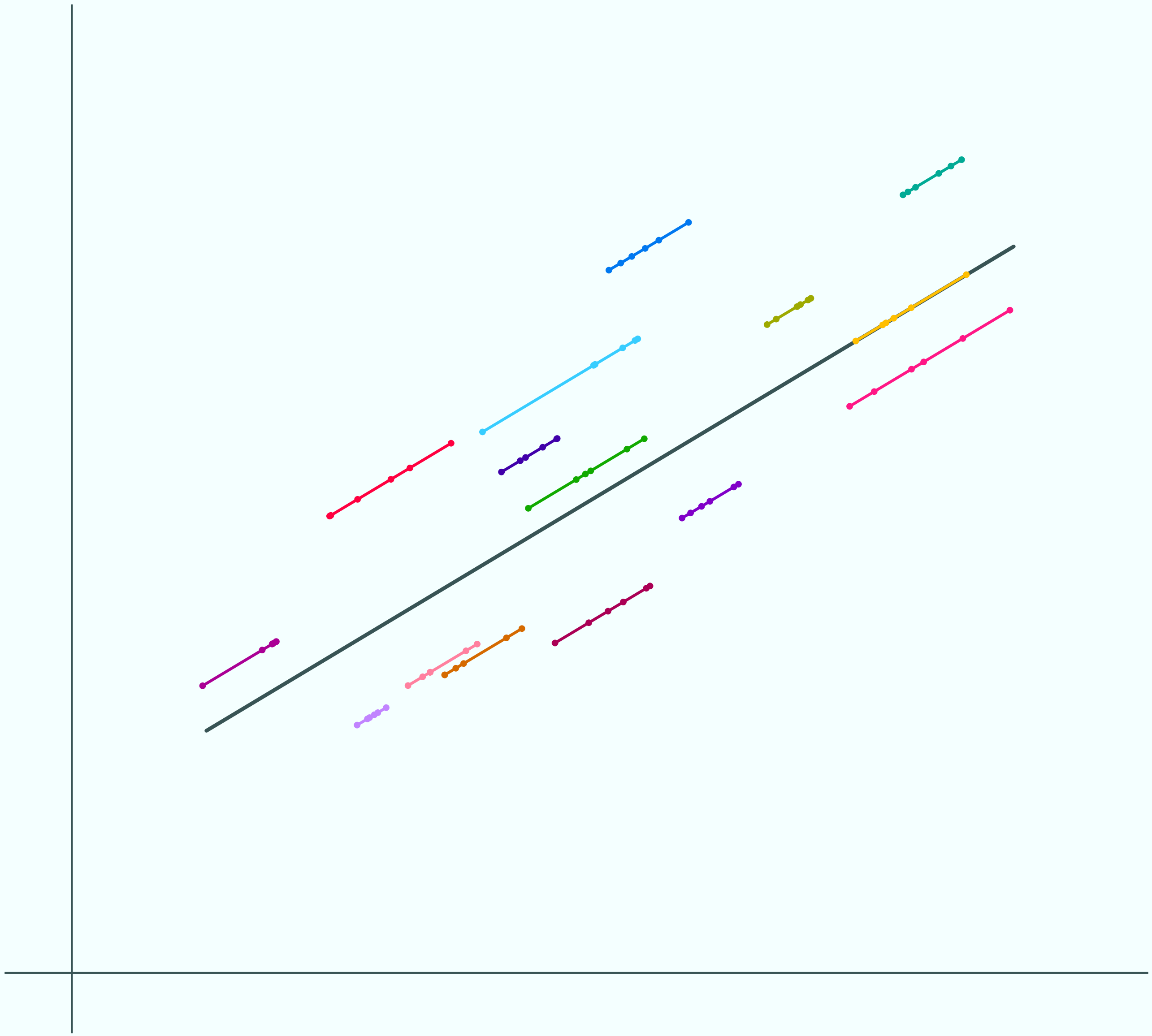


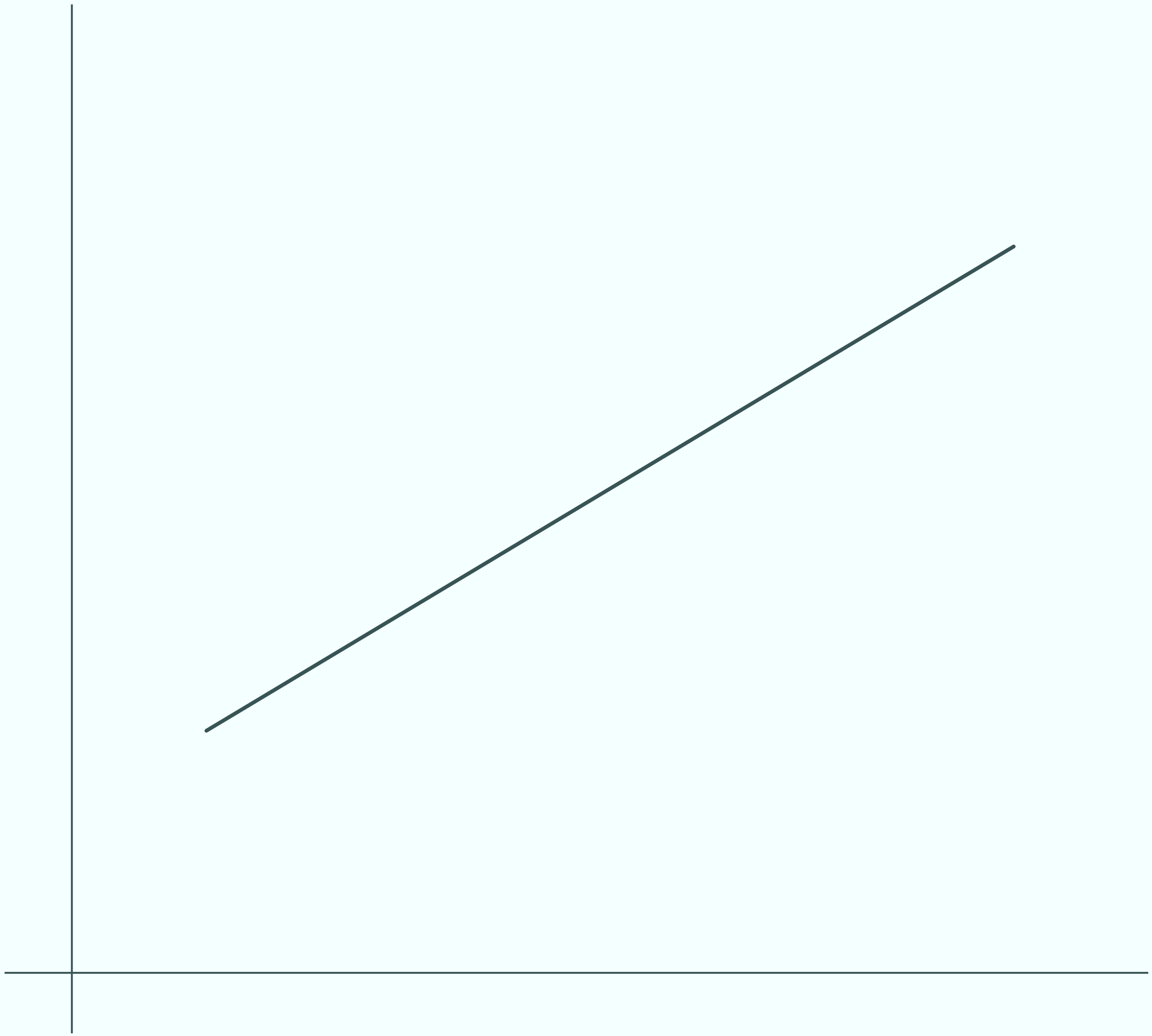
A large value of ρ

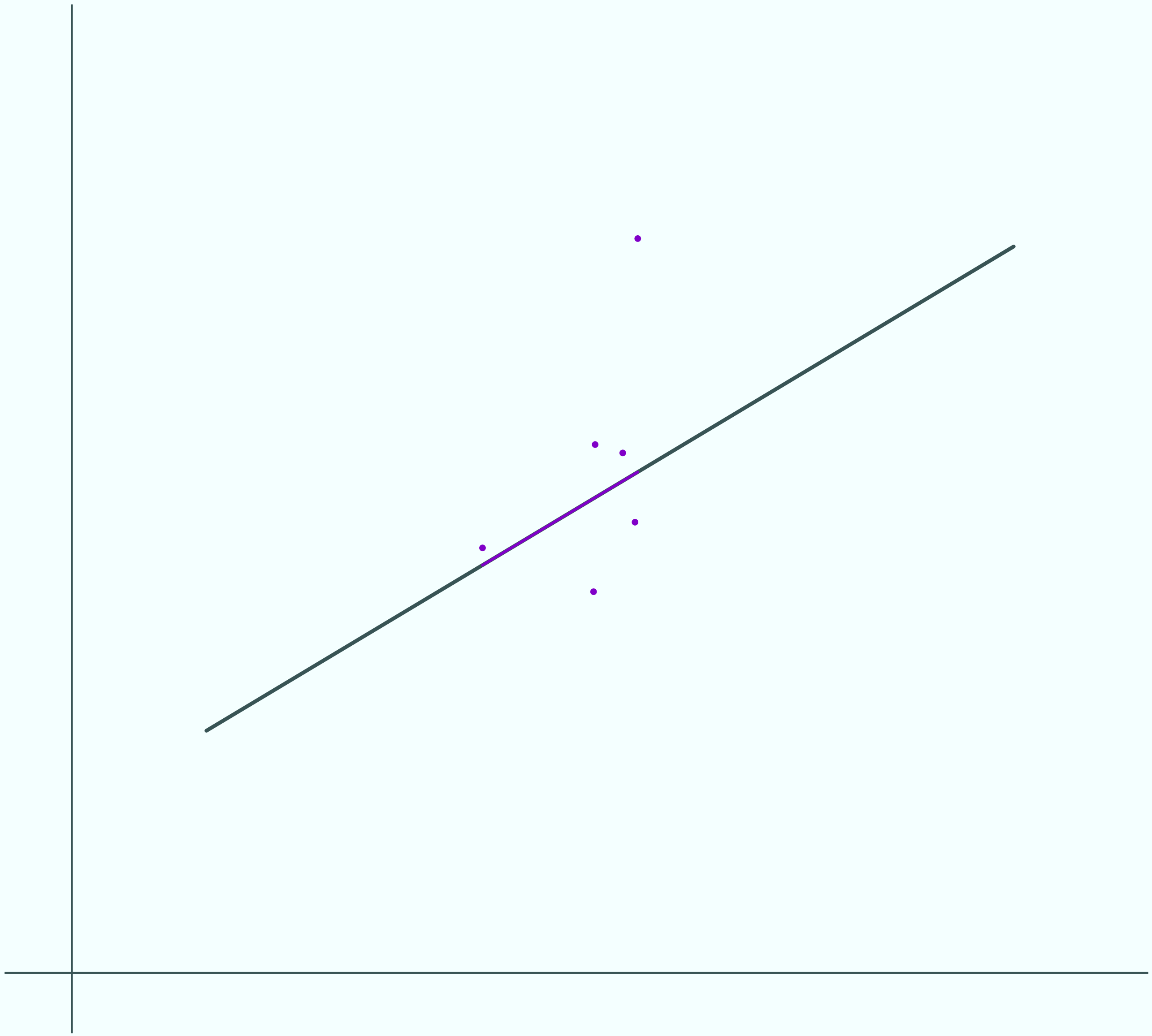


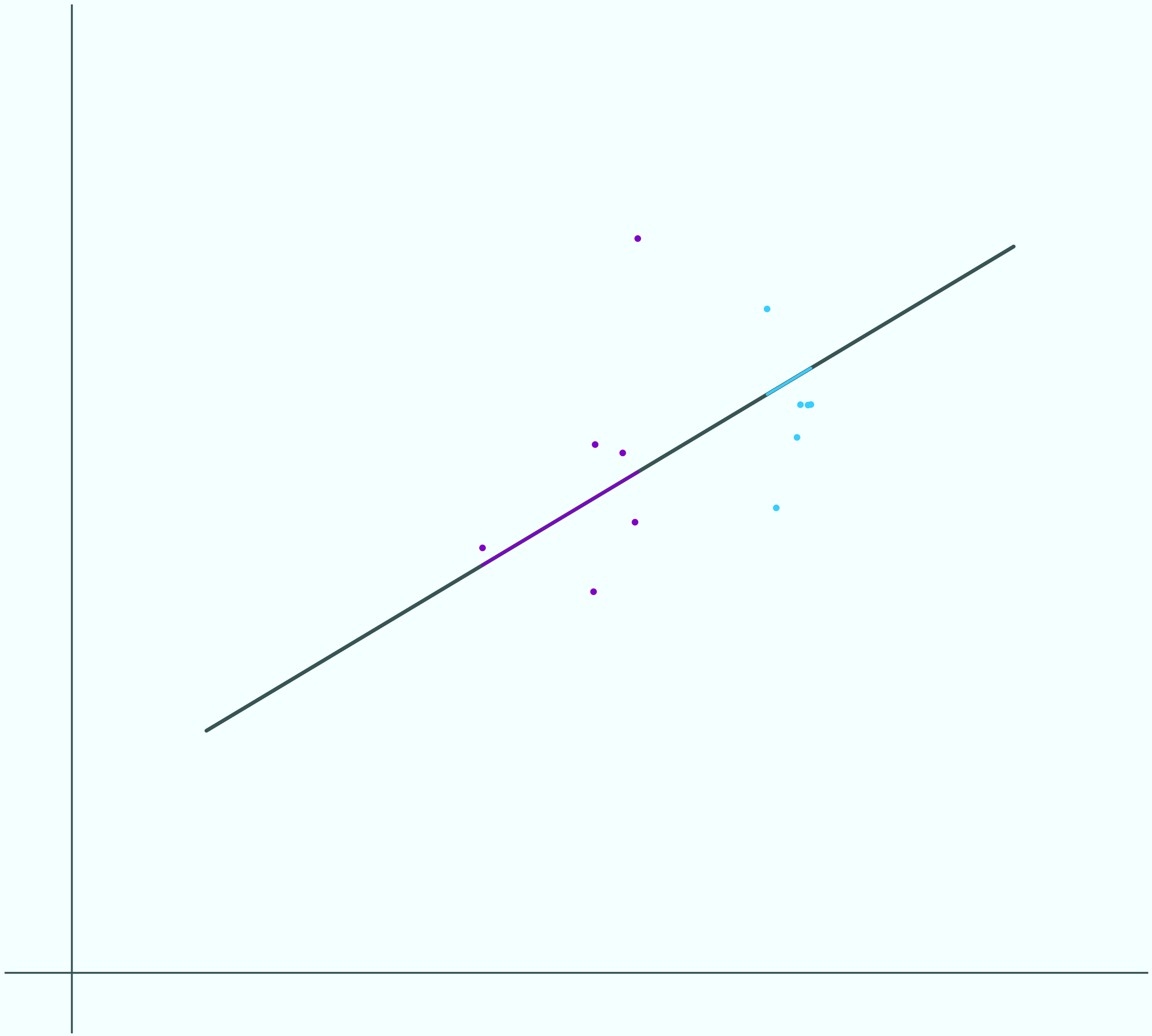


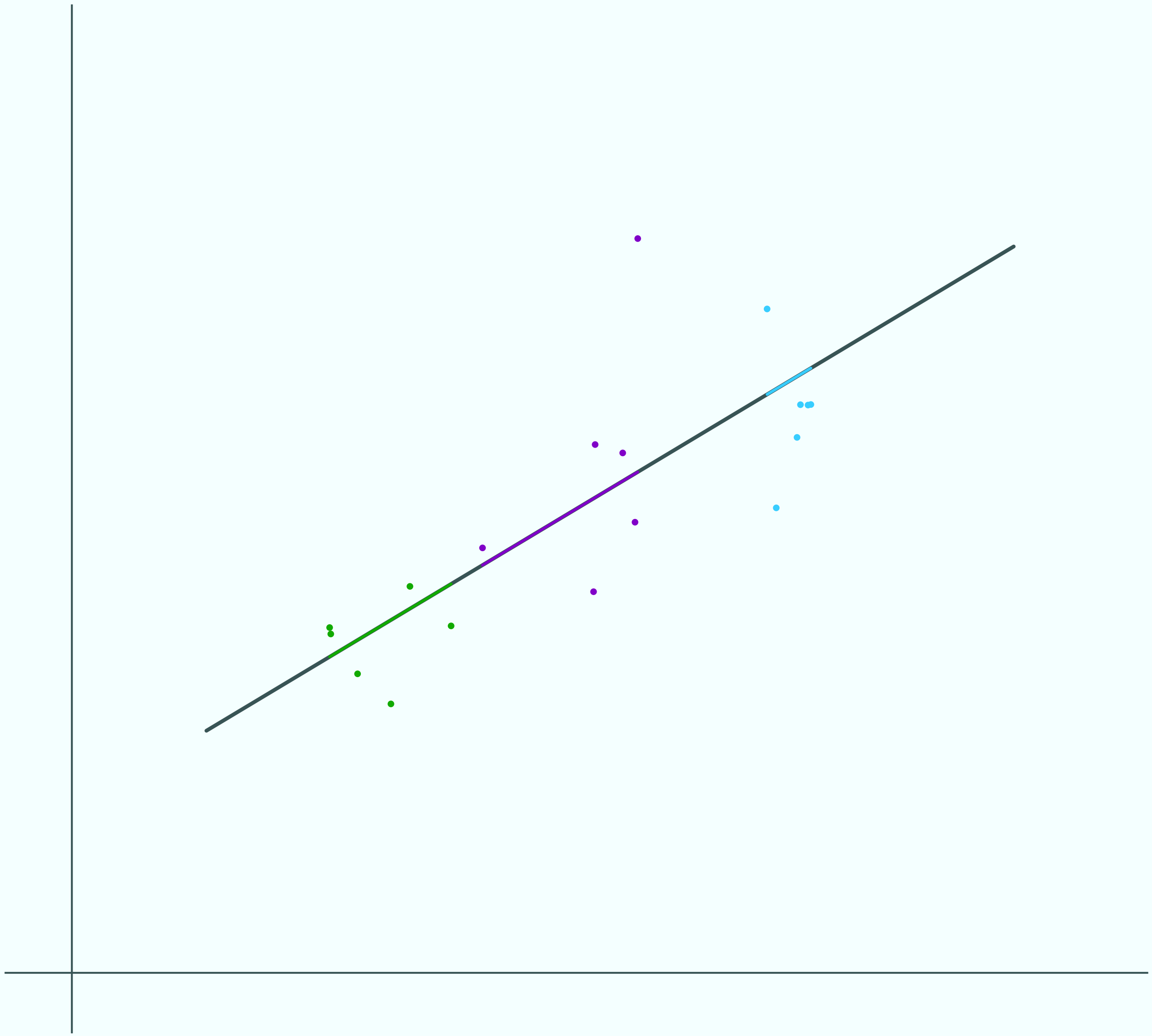
$\rho = 1$

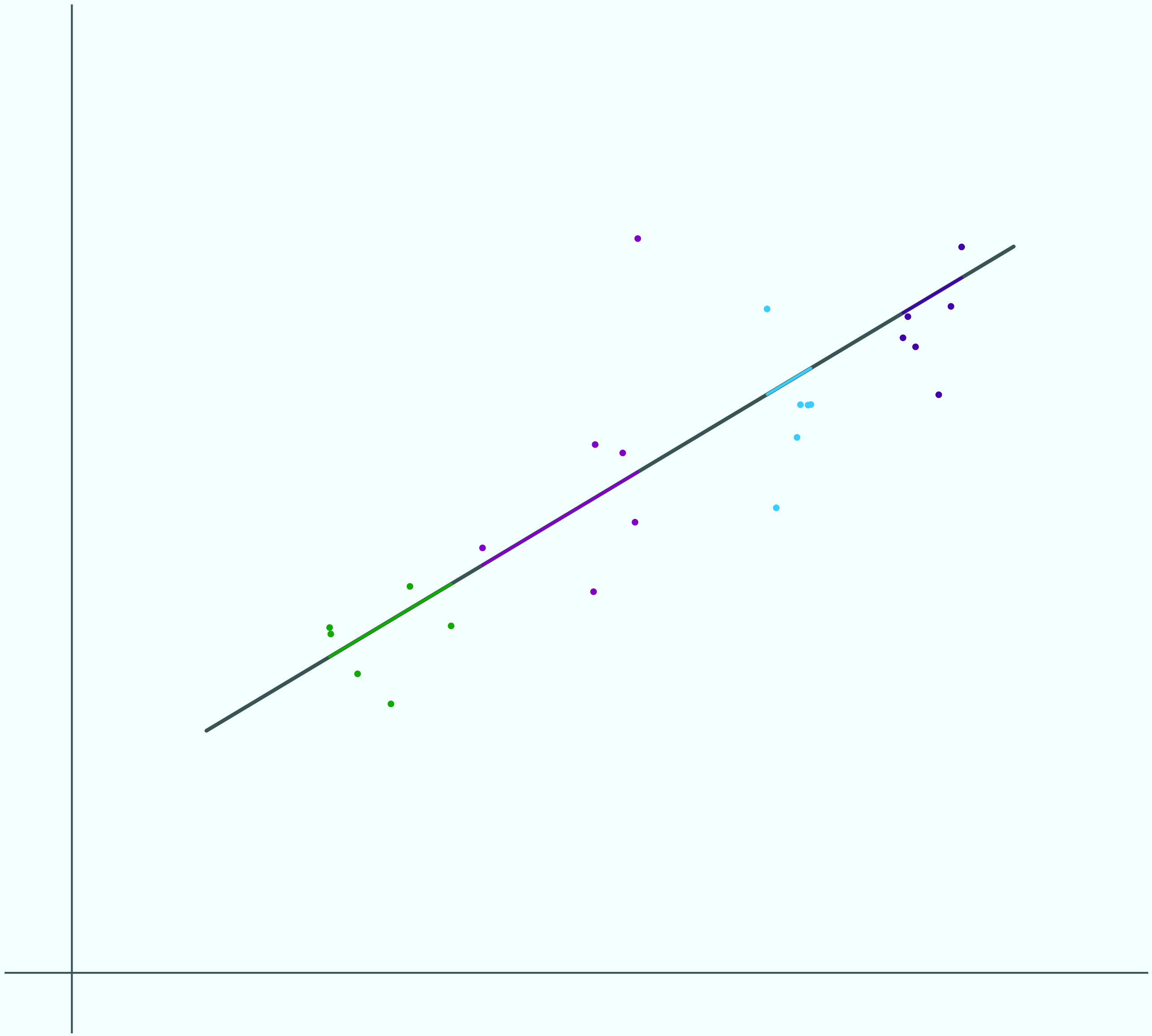


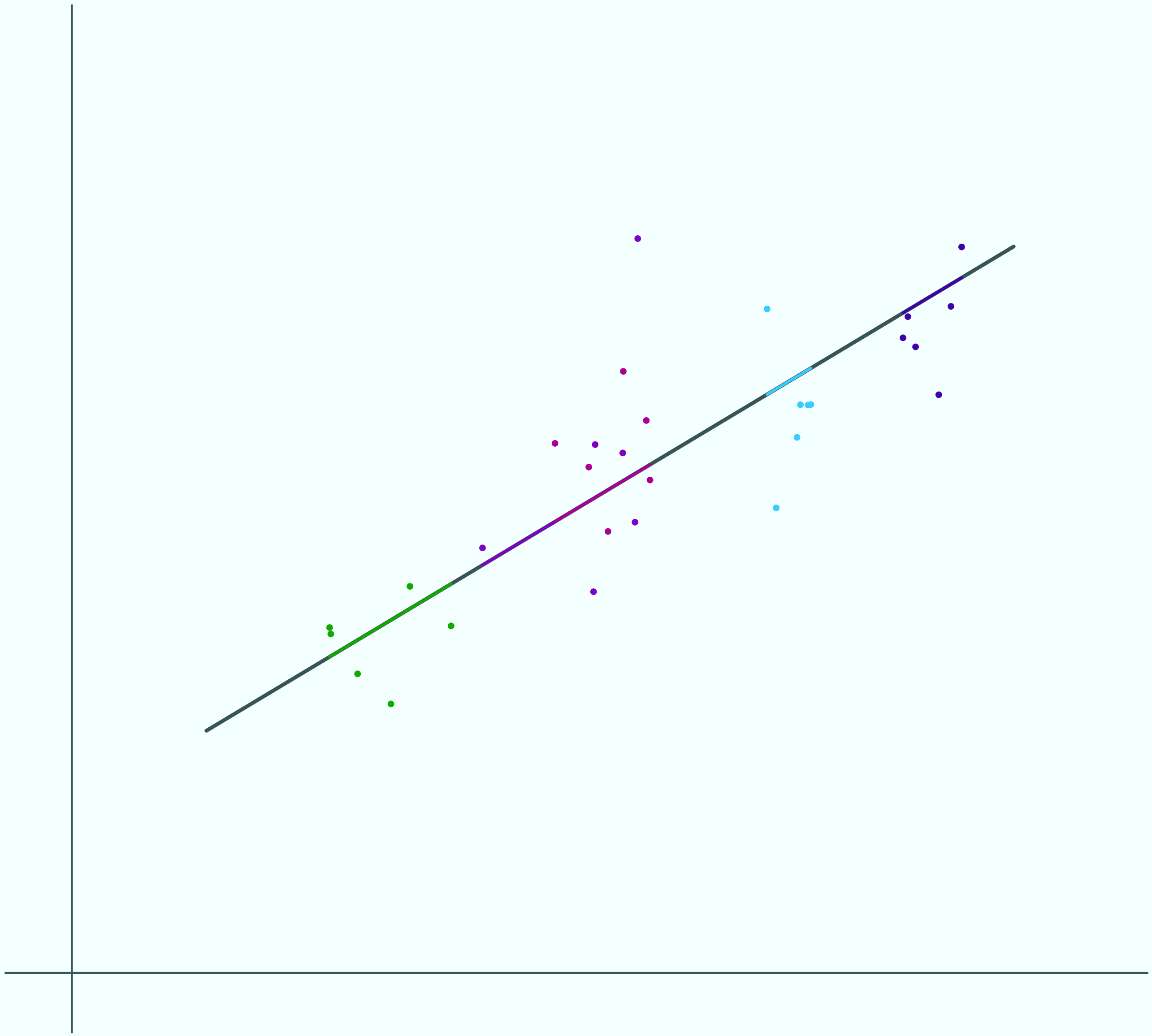


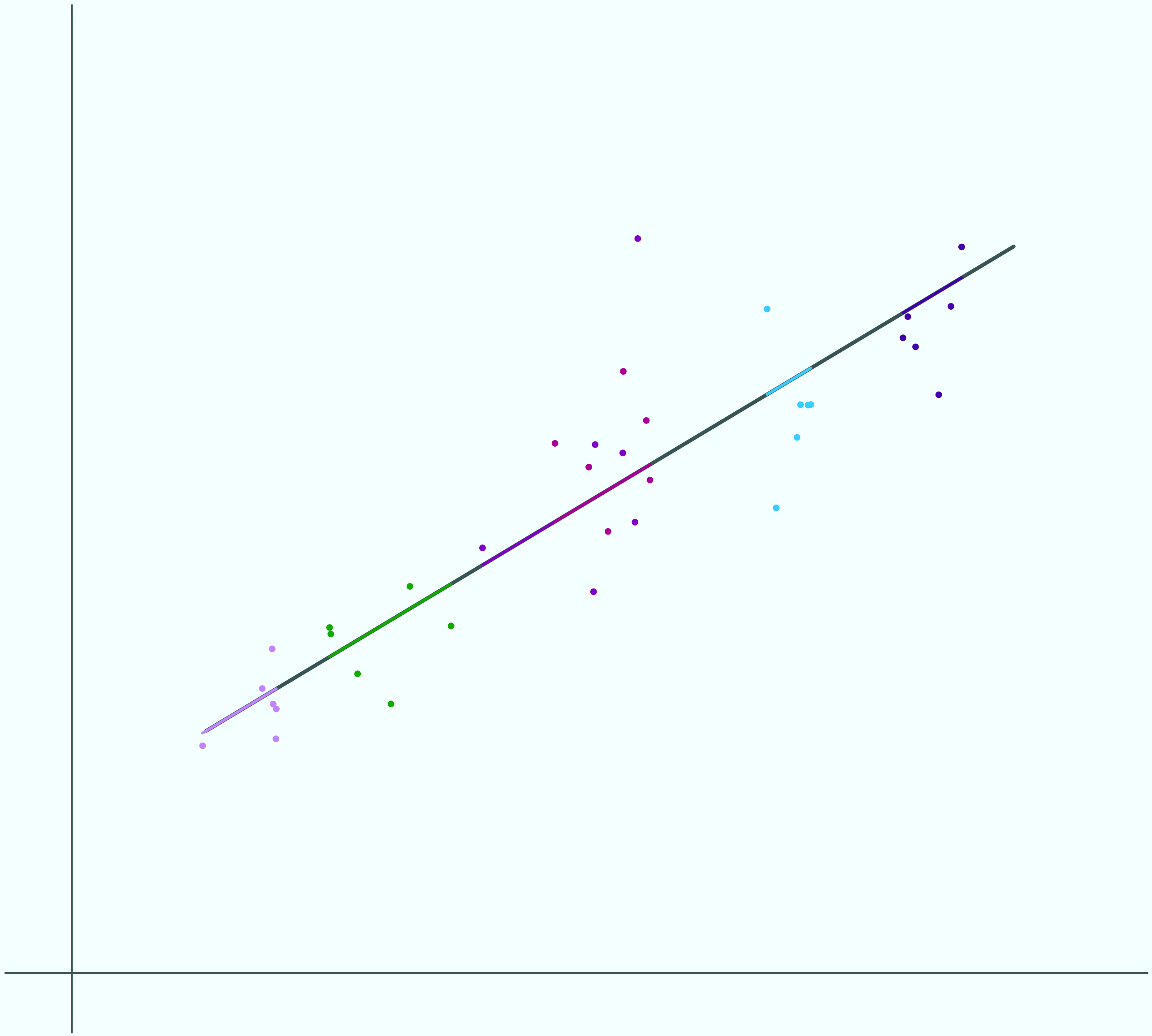


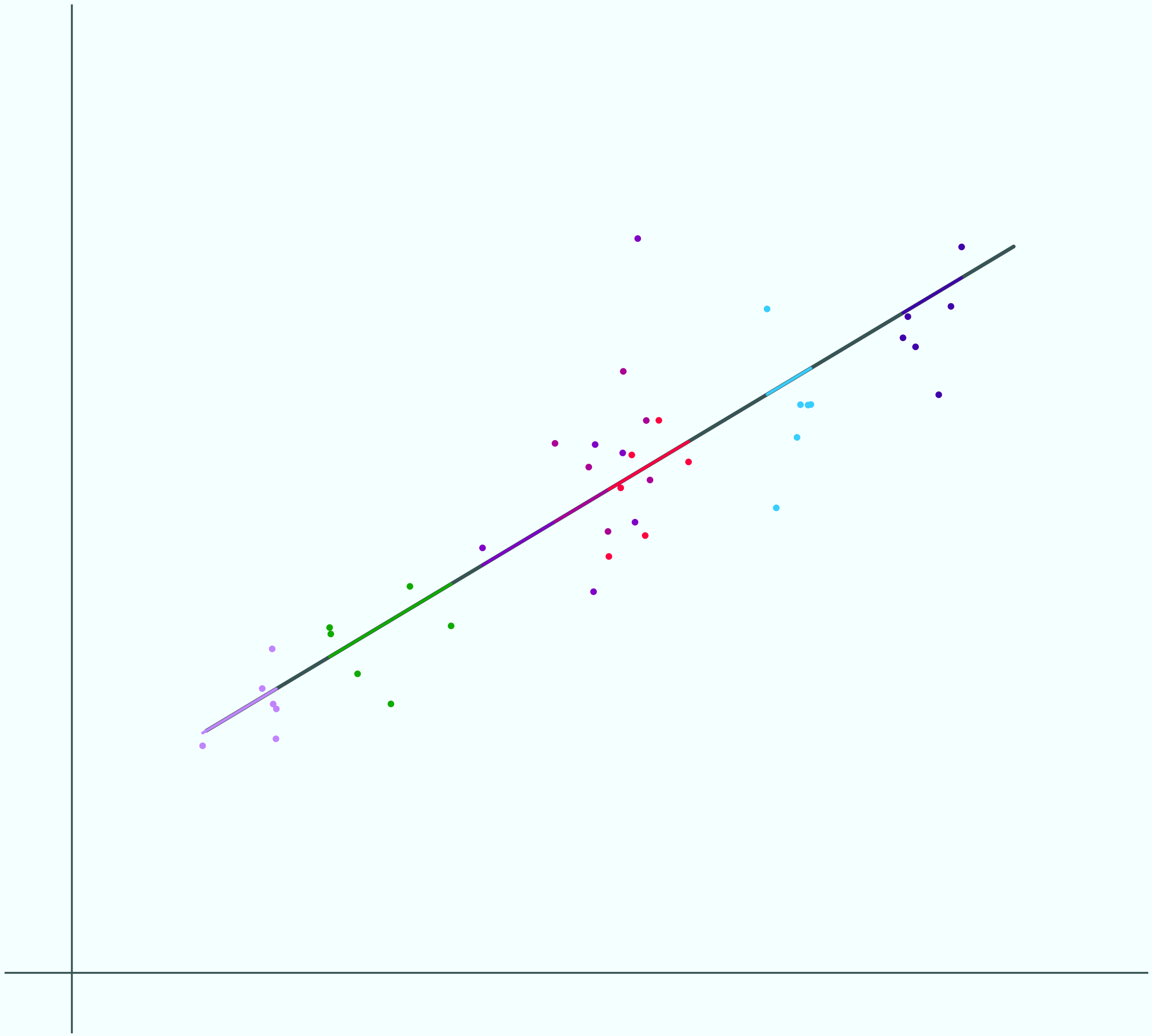


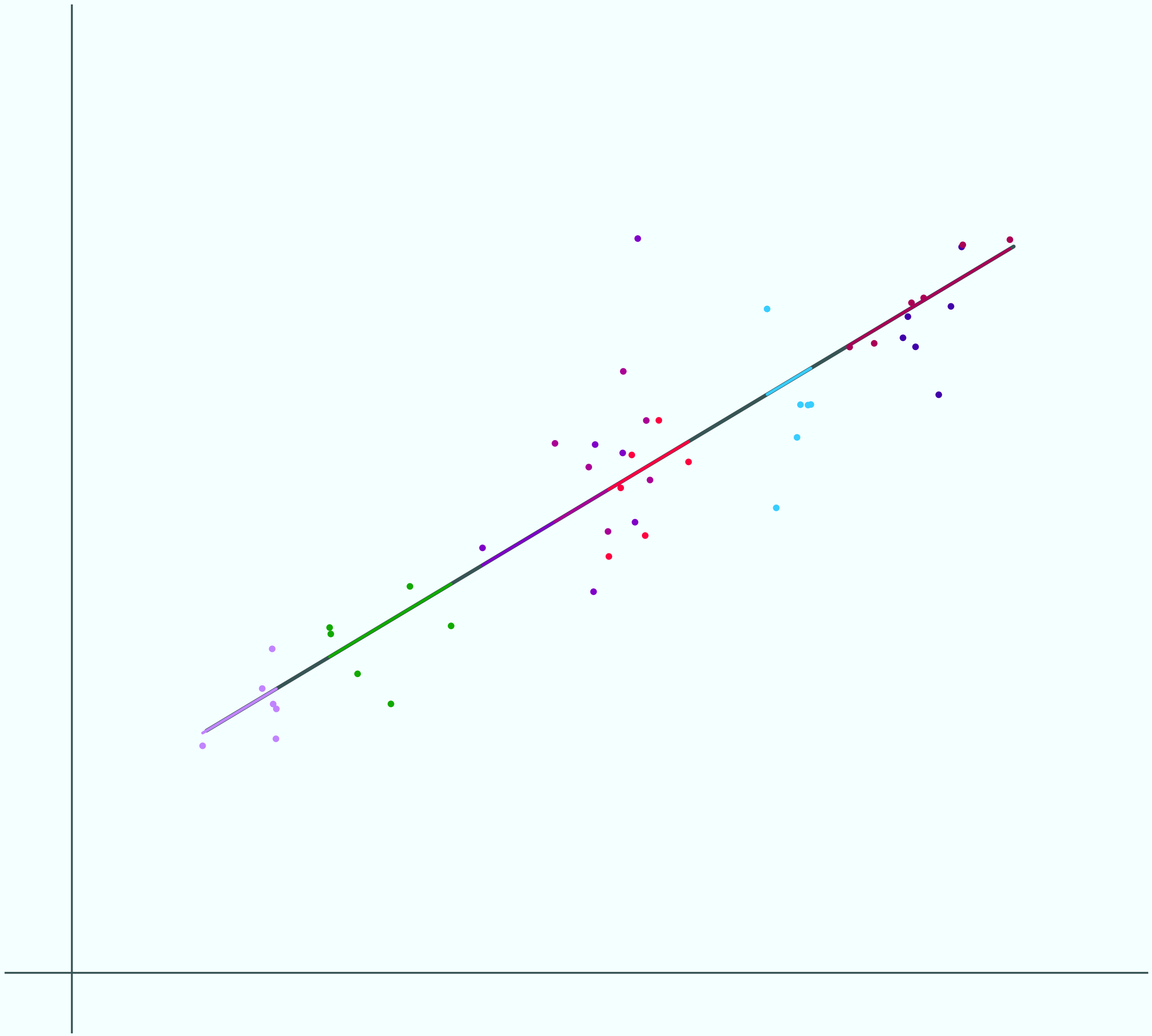


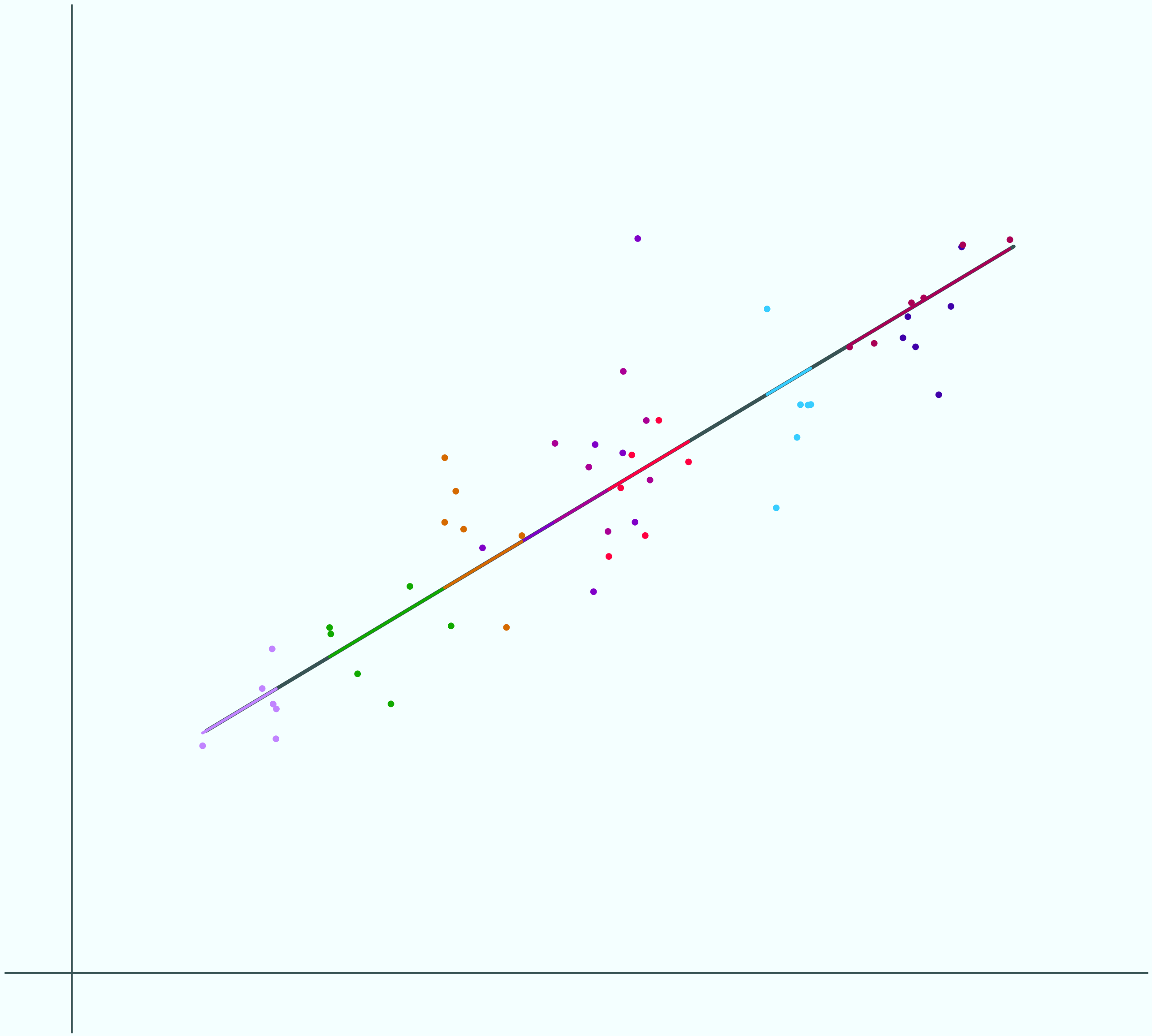


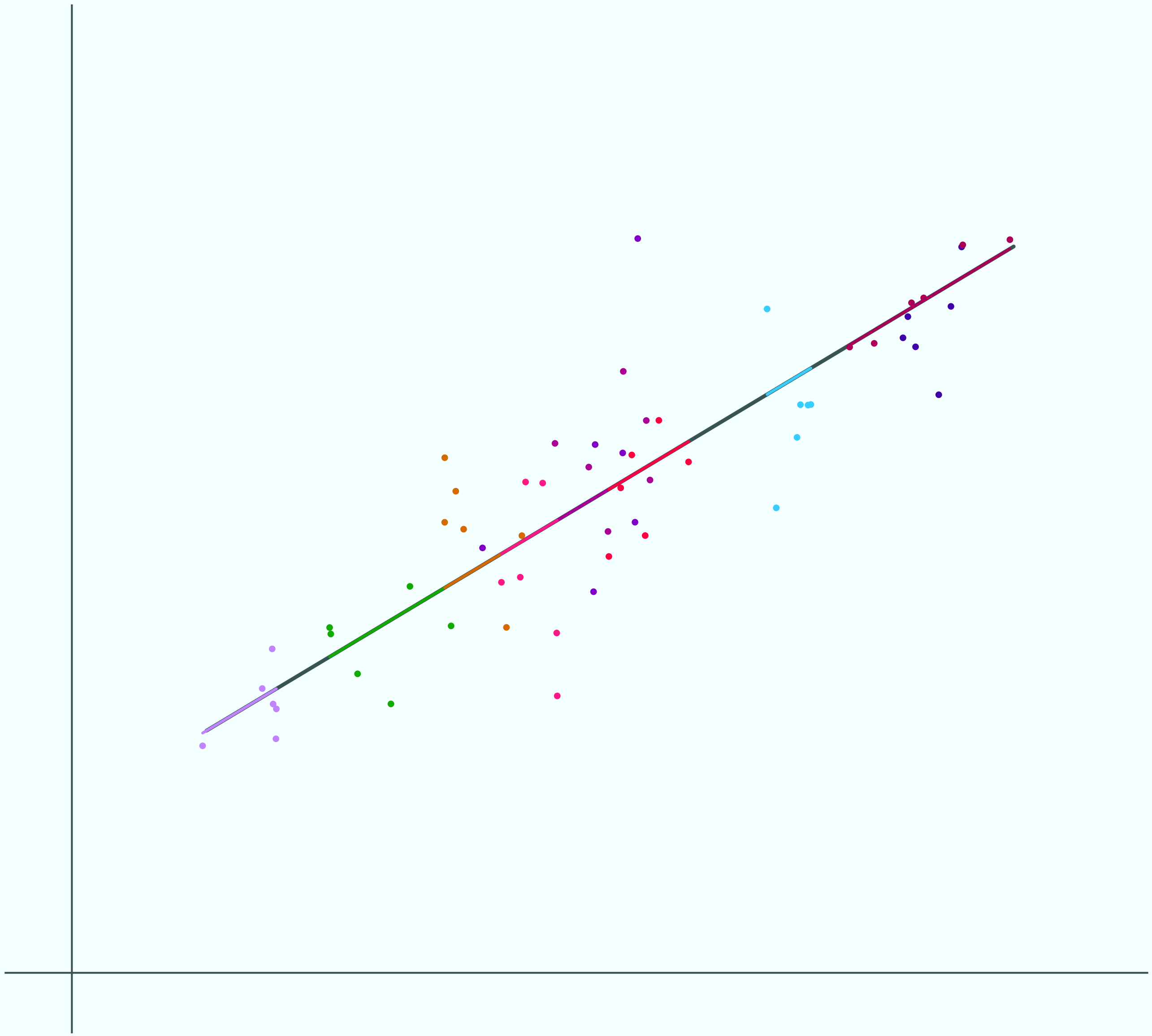


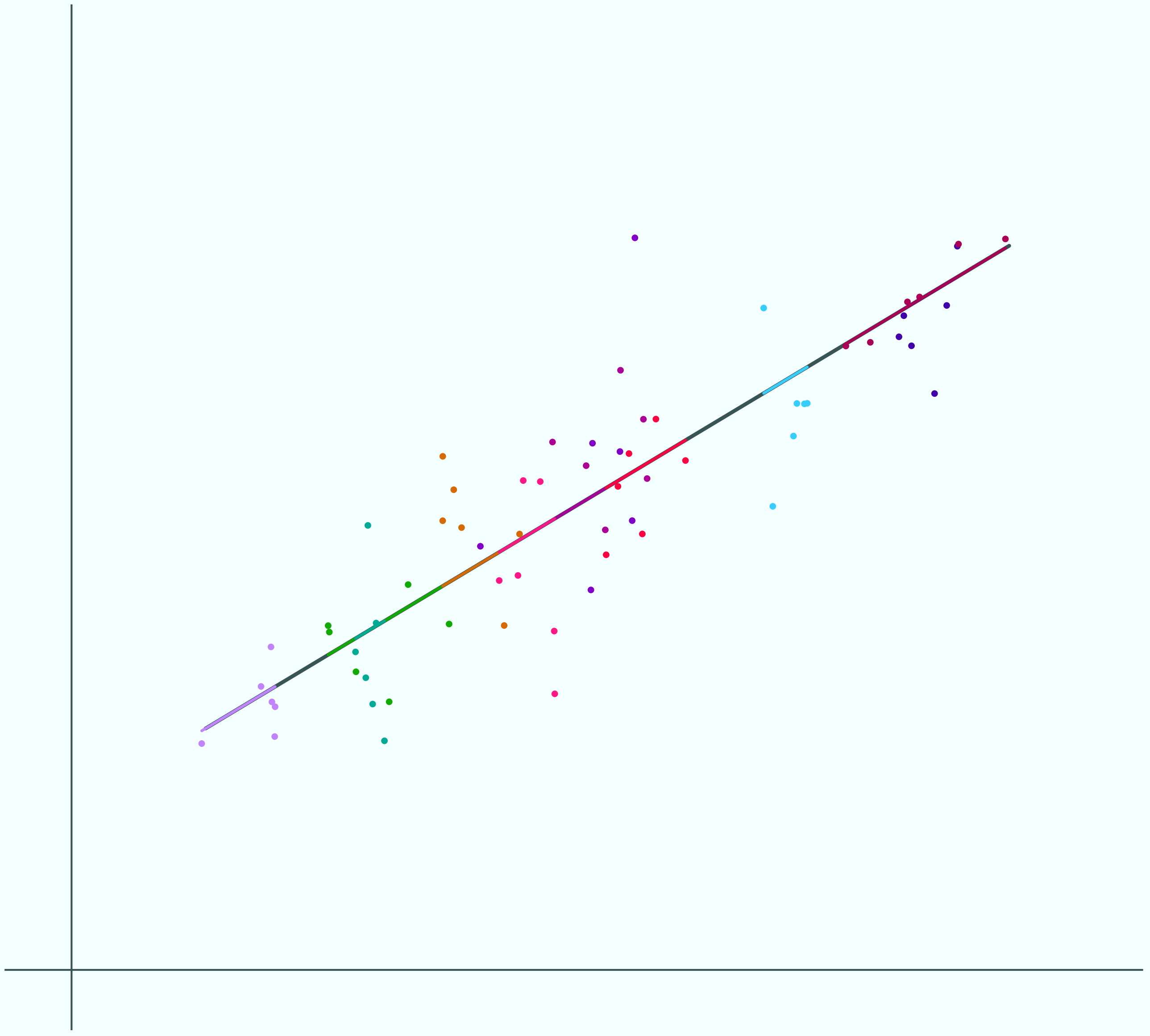


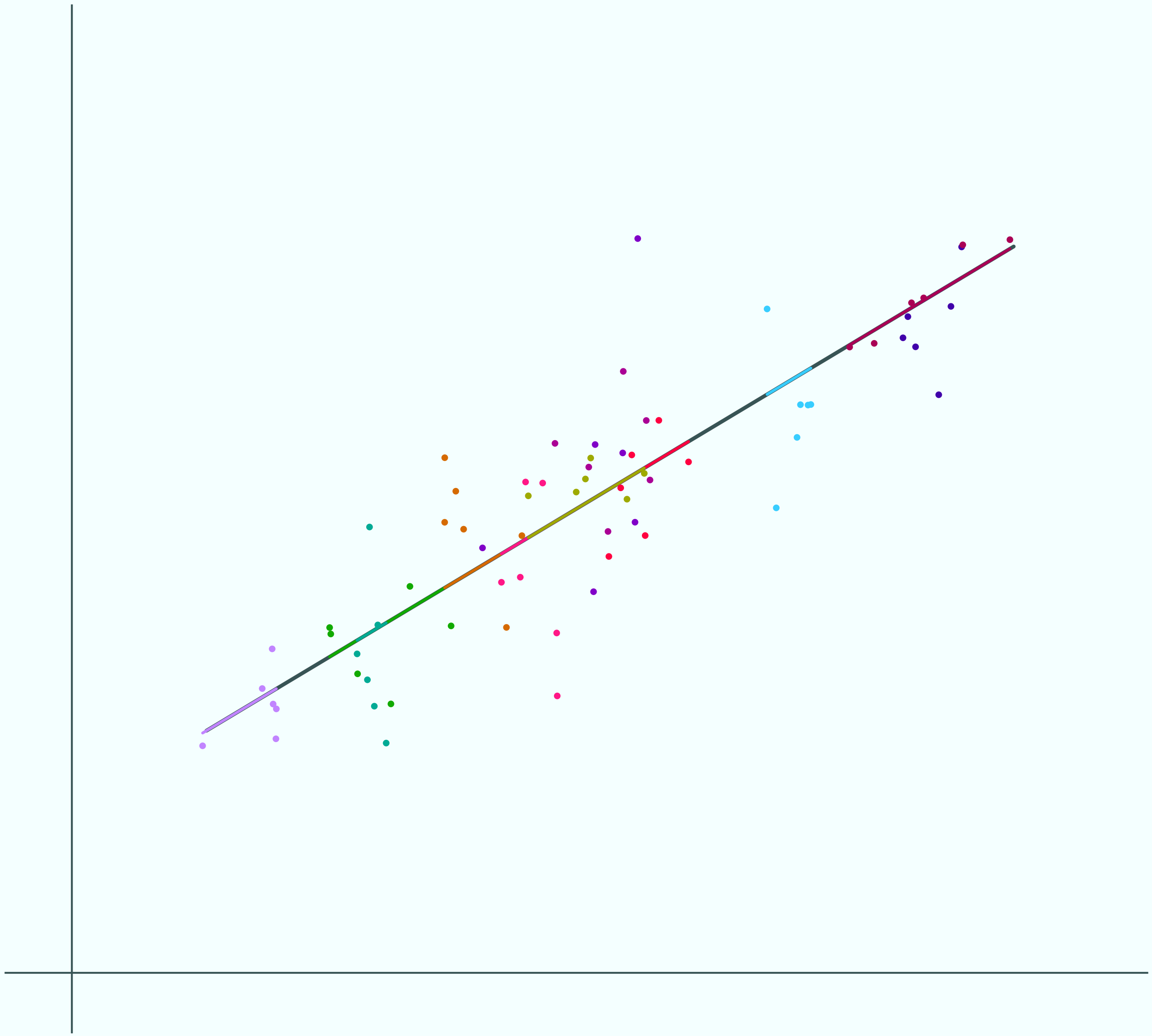


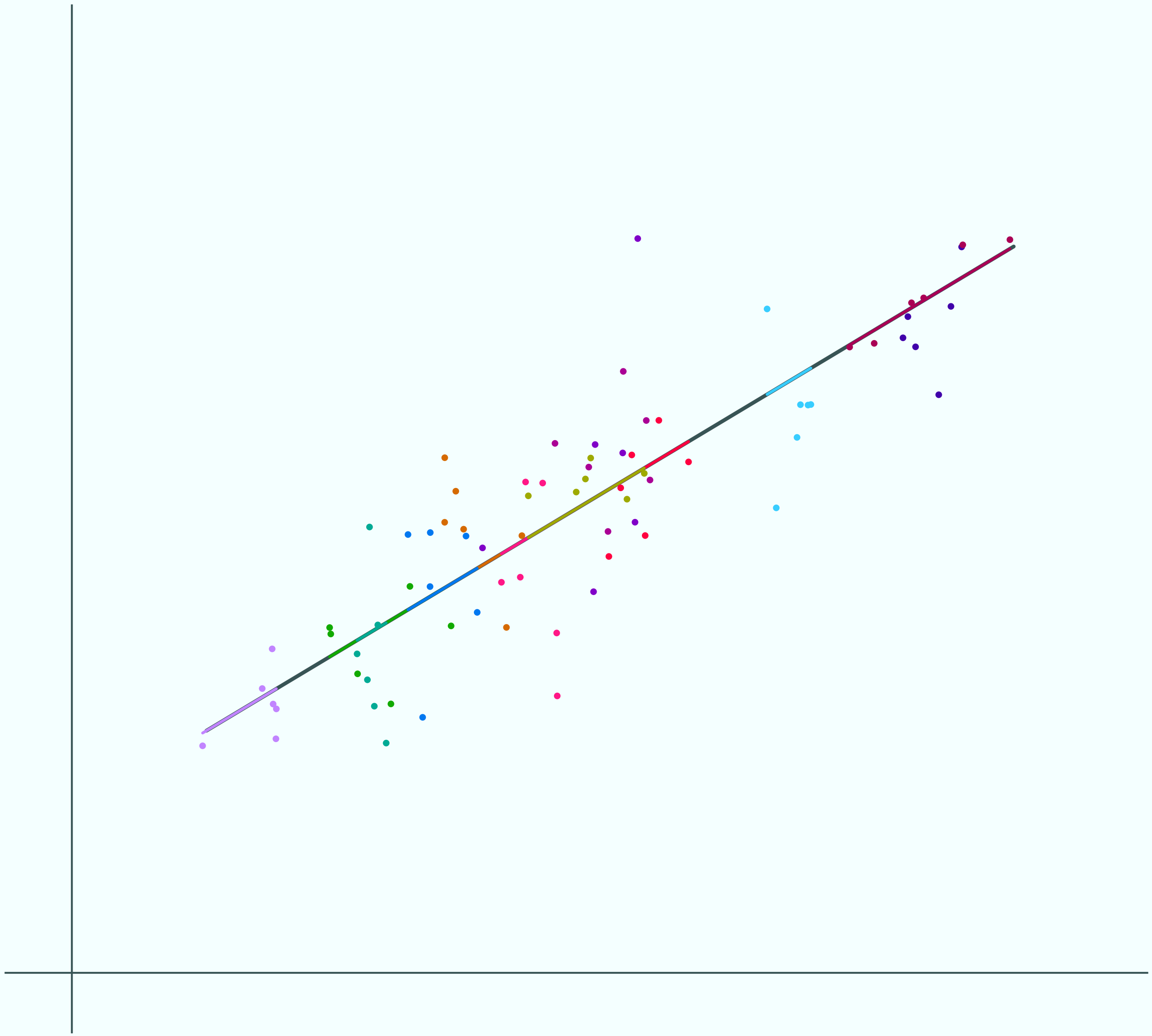


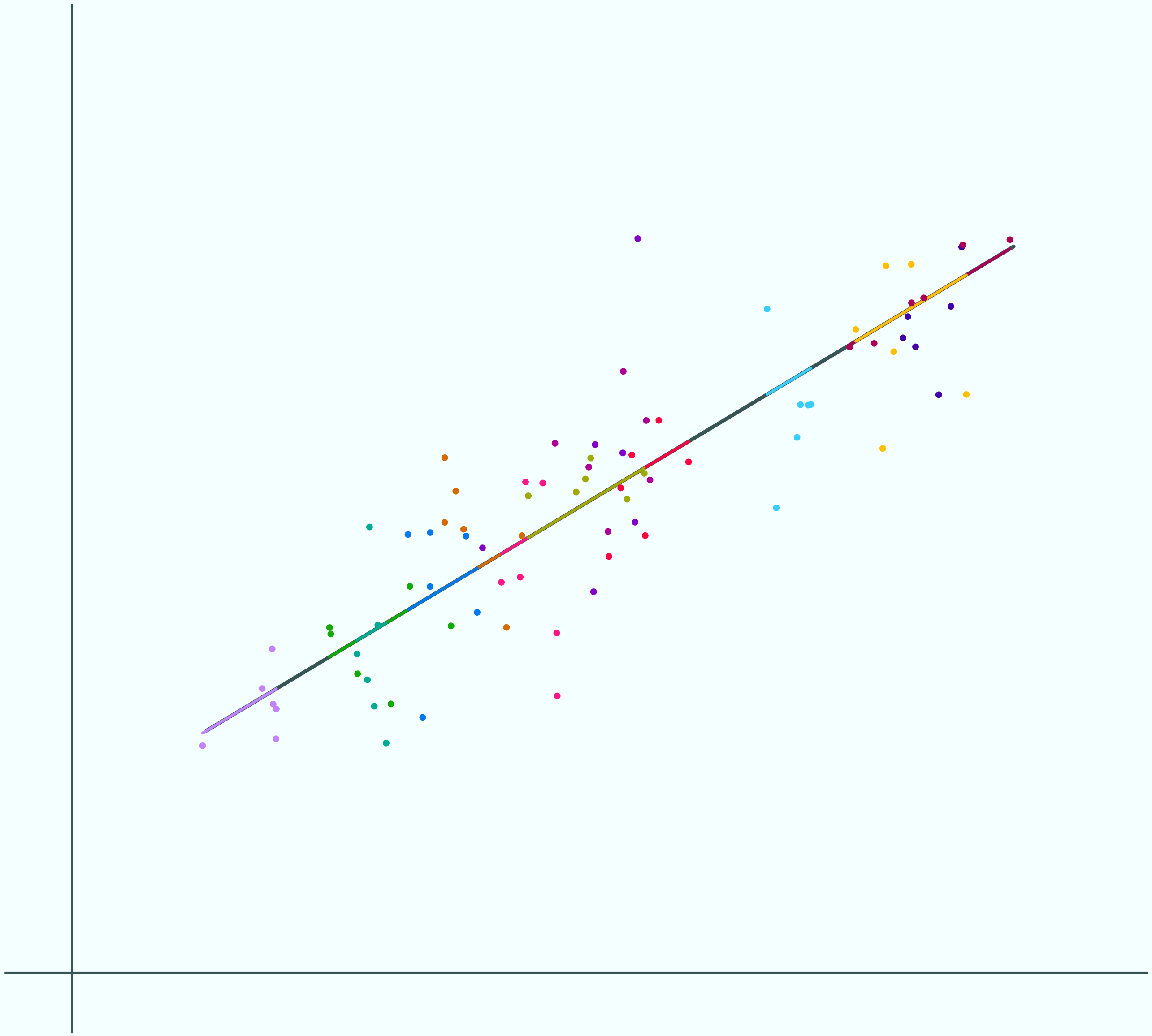




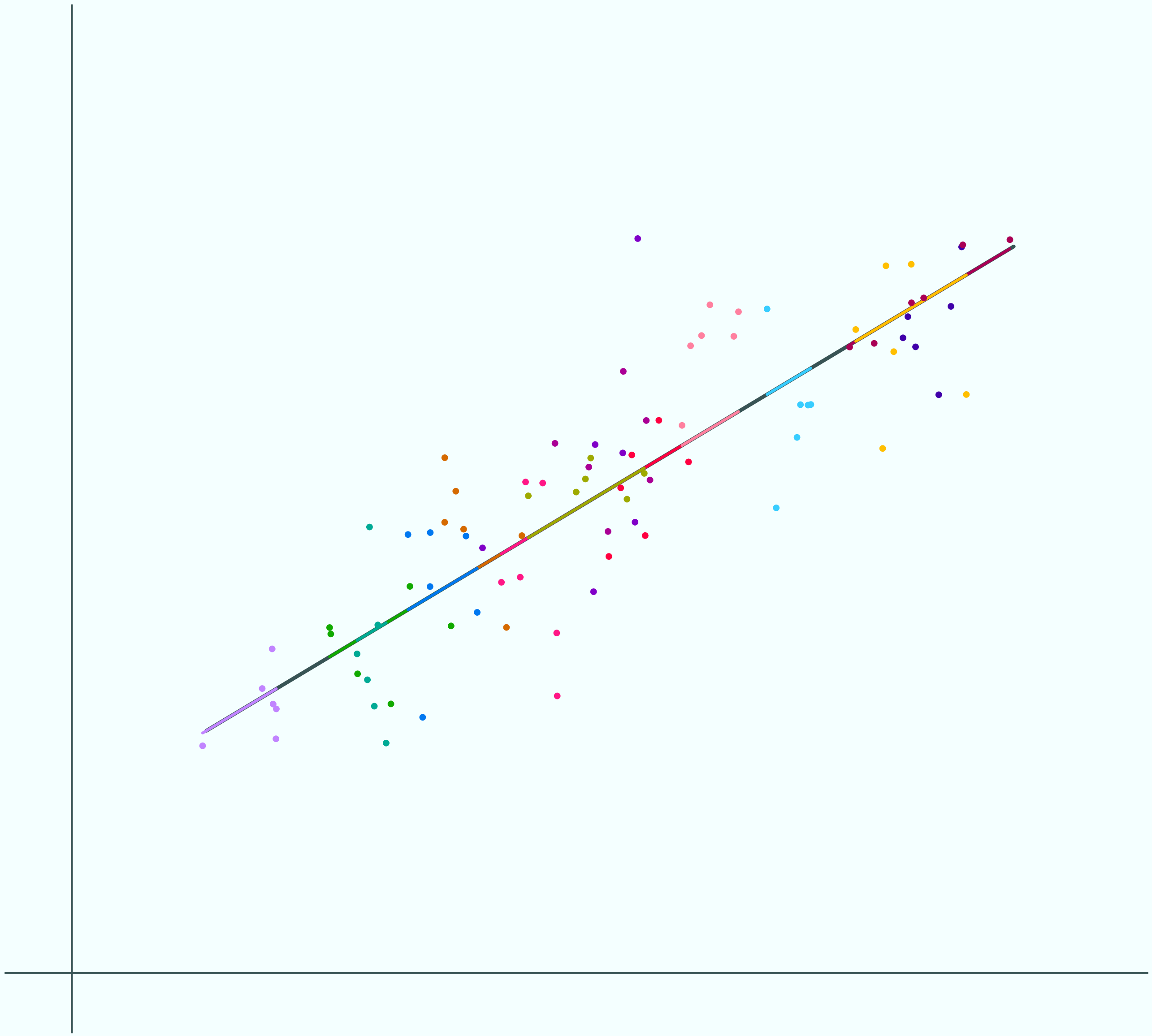








$\rho = 0$



Interpreting the value of ρ

Theoretical limits for ρ

- Looking at the formula for ρ, we can see that in theory the smallest it can be is 0 and the largest it can be is 1

What is a large value for ρ?

Interpreting the value of ρ

Theoretical limits for ρ

- Looking at the formula for ρ, we can see that in theory the smallest it can be is 0 and the largest it can be is 1
  - 1 indicates maximum clustering

What is a large value for ρ?

Interpreting the value of ρ

Theoretical limits for ρ

- Looking at the formula for ρ, we can see that in theory the smallest it can be is 0 and the largest it can be is 1
    - 1 indicates maximum clustering
    - 0 indicates no clustering (single level data structure)

What is a large value for ρ?

Interpreting the value of ρ

Theoretical limits for ρ

- Looking at the formula for ρ, we can see that in theory the smallest it can be is 0 and the largest it can be is 1
  - 1 indicates maximum clustering
  - 0 indicates no clustering (single level data structure)
- In practice, we never expect to see a value of 0 or 1 for ρ

What is a large value for ρ?

Interpreting the value of ρ

Theoretical limits for ρ

- Looking at the formula for ρ, we can see that in theory the smallest it can be is 0 and the largest it can be is 1
  - 1 indicates maximum clustering
  - 0 indicates no clustering (single level data structure)
- In practice, we never expect to see a value of 0 or 1 for ρ
- If ρ is small enough we can use a single level model

What is a large value for ρ?

Interpreting the value of ρ

Theoretical limits for ρ

- Looking at the formula for ρ, we can see that in theory the smallest it can be is 0 and the largest it can be is 1
  - 1 indicates maximum clustering
  - 0 indicates no clustering (single level data structure)
- In practice, we never expect to see a value of 0 or 1 for ρ
- If ρ is small enough we can use a single level model
- But we don't make that decision by looking at the value of ρ, we use the likelihood ratio test described earlier

What is a large value for ρ?

Interpreting the value of ρ

Theoretical limits for ρ

- Looking at the formula for ρ, we can see that in theory the smallest it can be is 0 and the largest it can be is 1
  - 1 indicates maximum clustering
  - 0 indicates no clustering (single level data structure)
- In practice, we never expect to see a value of 0 or 1 for ρ
- If ρ is small enough we can use a single level model
- But we don't make that decision by looking at the value of ρ, we use the likelihood ratio test described earlier

What is a large value for ρ?

It depends on the subject area and what the units of each level are.

Interpreting the value of ρ

Theoretical limits for ρ

- Looking at the formula for ρ, we can see that in theory the smallest it can be is 0 and the largest it can be is 1
  - 1 indicates maximum clustering
  - 0 indicates no clustering (single level data structure)
- In practice, we never expect to see a value of 0 or 1 for ρ
- If ρ is small enough we can use a single level model
- But we don't make that decision by looking at the value of ρ, we use the likelihood ratio test described earlier

What is a large value for ρ?

It depends on the subject area and what the units of each level are.

- We expect more clustering for observations on occasions within individuals than observations on people within families

Interpreting the value of ρ

Theoretical limits for ρ

- Looking at the formula for ρ, we can see that in theory the smallest it can be is 0 and the largest it can be is 1
  - 1 indicates maximum clustering
  - 0 indicates no clustering (single level data structure)
- In practice, we never expect to see a value of 0 or 1 for ρ
- If ρ is small enough we can use a single level model
- But we don't make that decision by looking at the value of ρ, we use the likelihood ratio test described earlier

What is a large value for ρ?

It depends on the subject area and what the units of each level are.

- We expect more clustering for observations on occasions within individuals than observations on people within families
- and more clustering for observations on people within families than pupils within schools, for example

Clustering in the model

Clustering

We've talked a lot about clustering:

Incorporating the clustering

Clustering in the model

Clustering

We've talked a lot about clustering:

- Clustering is the reason we can't use a single level regression model

Incorporating the clustering

Clustering in the model

Clustering

We've talked a lot about clustering:

- Clustering is the reason we can't use a single level regression model
- Clustering is why we have variance at both level 1 and level 2– why our response is determined at two levels

Incorporating the clustering

Clustering in the model

Clustering

We've talked a lot about clustering:

- Clustering is the reason we can't use a single level regression model
- Clustering is why we have variance at both level 1 and level 2– why our response is determined at two levels
- We've seen how to measure this clustering

Incorporating the clustering

Clustering in the model

Clustering

We've talked a lot about clustering:

- Clustering is the reason we can't use a single level regression model
- Clustering is why we have variance at both level 1 and level 2– why our response is determined at two levels
- We've seen how to measure this clustering
- and how to interpret it

Incorporating the clustering

Clustering in the model

Clustering

We've talked a lot about clustering:

- Clustering is the reason we can't use a single level regression model
- Clustering is why we have variance at both level 1 and level 2– why our response is determined at two levels
- We've seen how to measure this clustering
- and how to interpret it

Incorporating the clustering

- We haven't seen yet how the clustering is incorporated into the model:

Clustering in the model

Clustering

We've talked a lot about clustering:

- Clustering is the reason we can't use a single level regression model
- Clustering is why we have variance at both level 1 and level 2– why our response is determined at two levels
- We've seen how to measure this clustering
- and how to interpret it

Incorporating the clustering

- We haven't seen yet how the clustering is incorporated into the model:
- how does the random intercepts model allow for similarities between different observations from the same group?

Clustering in the model

Clustering

We've talked a lot about clustering:

- Clustering is the reason we can't use a single level regression model
- Clustering is why we have variance at both level 1 and level 2– why our response is determined at two levels
- We've seen how to measure this clustering
- and how to interpret it

Incorporating the clustering

- We haven't seen yet how the clustering is incorporated into the model:
- how does the random intercepts model allow for similarities between different observations from the same group?
- To discover this, we need to look at the correlation matrix V.

Clustering in the model

Clustering

We've talked a lot about clustering:

- Clustering is the reason we can't use a single level regression model
- Clustering is why we have variance at both level 1 and level 2– why our response is determined at two levels
- We've seen how to measure this clustering
- and how to interpret it

Incorporating the clustering

- We haven't seen yet how the clustering is incorporated into the model:
- how does the random intercepts model allow for similarities between different observations from the same group?
- To discover this, we need to look at the correlation matrix V.
- And to do that we need to return to the technicalities of the model

Assumptions of random part

Let's first recall the assumptions for a single level model and for a variance components model:

Assumptions of random part

Let's first recall the assumptions for a single level model and for a variance components model:

Single level model

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad e_i \sim \mathrm{N}(0, \sigma_e^2)$$

Assumptions of random part

Let's first recall the assumptions for a single level model and for a variance components model:

Single level model

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad e_i \sim \mathrm{N}(0, \sigma_e^2)$$

$$\mathrm{Cov}(e_i, x_i) = 0$$

Assumptions of random part

Let's first recall the assumptions for a single level model and for a variance components model:

Single level model

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad e_i \sim \mathrm{N}(0, \sigma_e^2)$$

$$\mathrm{Cov}(e_i, x_i) = 0 \qquad \mathrm{Cov}(e_{i_1}, e_{i_2}) = 0$$

Assumptions of random part

Let's first recall the assumptions for a single level model and for a variance components model:

Single level model

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad e_i \sim \mathrm{N}(0, \sigma_e^2)$$

$$\mathrm{Cov}(e_i, x_i) = 0 \qquad\qquad \mathrm{Cov}(e_{i_1}, e_{i_2}) = 0$$

Variance components model

$$\begin{aligned} y_{ij} = \beta_0 + u_j + e_{ij} \quad & u_j \sim \mathrm{N}(0, \sigma_u^2) \\ & e_{ij} \sim \mathrm{N}(0, \sigma_e^2) \end{aligned}$$

Assumptions of random part

Let's first recall the assumptions for a single level model and for a variance components model:

Single level model

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad e_i \sim \mathrm{N}(0, \sigma_e^2)$$

$$\mathrm{Cov}(e_i, x_i) = 0 \qquad \mathrm{Cov}(e_{i_1}, e_{i_2}) = 0$$

Variance components model

$$\begin{aligned} y_{ij} = \beta_0 + u_j + e_{ij} \quad & u_j \sim \mathrm{N}(0, \sigma_u^2) \\ & e_{ij} \sim \mathrm{N}(0, \sigma_e^2) \end{aligned}$$

$$\mathrm{Cov}(u_{j_1}, u_{j_2}) = 0$$

Assumptions of random part

Let's first recall the assumptions for a single level model and for a variance components model:

Single level model

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad e_i \sim \mathrm{N}(0, \sigma_e^2)$$

$$\mathrm{Cov}(e_i, x_i) = 0 \qquad\qquad \mathrm{Cov}(e_{i_1}, e_{i_2}) = 0$$

Variance components model

$$\begin{aligned} y_{ij} = \beta_0 + u_j + e_{ij} \quad u_j &\sim \mathrm{N}(0, \sigma_u^2) \\ e_{ij} &\sim \mathrm{N}(0, \sigma_e^2) \end{aligned}$$

$$\mathrm{Cov}(u_{j_1}, u_{j_2}) = 0 \qquad \mathrm{Cov}(u_{j_1}, e_{i_1 j_1}) = 0$$

$$\mathrm{Cov}(u_{j_1}, e_{i_1 j_2}) = 0$$

Assumptions of random part

Let's first recall the assumptions for a single level model and for a variance components model:

Single level model

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad e_i \sim \mathrm{N}(0, \sigma_e^2)$$

$$\mathrm{Cov}(e_i, x_i) = 0 \qquad \mathrm{Cov}(e_{i_1}, e_{i_2}) = 0$$

Variance components model

$$\begin{aligned} y_{ij} = \beta_0 + u_j + e_{ij} \quad u_j &\sim \mathrm{N}(0, \sigma_u^2) \\ e_{ij} &\sim \mathrm{N}(0, \sigma_e^2) \end{aligned}$$

$$\mathrm{Cov}(u_{j_1}, u_{j_2}) = 0 \qquad \mathrm{Cov}(u_{j_1}, e_{i_1 j_1}) = 0 \qquad \mathrm{Cov}(e_{i_1 j_1}, e_{i_2 j_1}) = 0$$

$$\mathrm{Cov}(u_{j_1}, e_{i_1 j_2}) = 0 \qquad \mathrm{Cov}(e_{i_1 j_1}, e_{i_2 j_2}) = 0$$

Assumptions of random part

Assumptions for a random intercept model are a mixture:

Random intercept model

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_j + e_{ij} \quad u_j \sim \mathrm{N}(0, \sigma_u^2)$$
$$e_{ij} \sim \mathrm{N}(0, \sigma_e^2)$$

Assumptions of random part

Assumptions for a random intercept model are a mixture:

Random intercept model

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_j + e_{ij} \quad \begin{aligned} u_j &\sim \mathrm{N}(0, \sigma_u^2) \\ e_{ij} &\sim \mathrm{N}(0, \sigma_e^2) \end{aligned}$$

$\mathrm{Cov}(u_{j_1}, u_{j_2}) = 0$

- Level 2 residuals for different groups are uncorrelated

Assumptions of random part

Assumptions for a random intercept model are a mixture:

Random intercept model

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_j + e_{ij} \quad \begin{aligned} u_j &\sim \mathrm{N}(0, \sigma_u^2) \\ e_{ij} &\sim \mathrm{N}(0, \sigma_e^2) \end{aligned}$$

$$\mathrm{Cov}(u_{j_1}, u_{j_2}) = 0 \qquad \mathrm{Cov}(u_{j_1}, e_{i_1 j_1}) = 0$$

$$\mathrm{Cov}(u_{j_1}, e_{i_1 j_2}) = 0$$

- Level 2 residuals for different groups are uncorrelated
- Level 2 and level 1 residuals are uncorrelated

Assumptions of random part

Assumptions for a random intercept model are a mixture:

Random intercept model

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_j + e_{ij} \quad \begin{aligned} u_j &\sim \mathrm{N}(0, \sigma_u^2) \\ e_{ij} &\sim \mathrm{N}(0, \sigma_e^2) \end{aligned}$$

$$\mathrm{Cov}(u_{j_1}, u_{j_2}) = 0 \qquad \mathrm{Cov}(u_{j_1}, e_{i_1 j_1}) = 0 \qquad \mathrm{Cov}(e_{i_1 j_1}, e_{i_2 j_1}) = 0$$

$$\mathrm{Cov}(u_{j_1}, e_{i_1 j_2}) = 0 \qquad \mathrm{Cov}(e_{i_1 j_1}, e_{i_2 j_2}) = 0$$

- Level 2 residuals for different groups are uncorrelated
- Level 1 residuals for different observations are uncorrelated
- Level 2 and level 1 residuals are uncorrelated

Assumptions of random part

Assumptions for a random intercept model are a mixture:

Random intercept model

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_j + e_{ij} \quad u_j \sim \mathrm{N}(0, \sigma_u^2)$$
$$e_{ij} \sim \mathrm{N}(0, \sigma_e^2)$$

$$\mathrm{Cov}(u_{j_1}, u_{j_2}) = 0 \qquad \mathrm{Cov}(u_{j_1}, e_{i_1 j_1}) = 0 \qquad \mathrm{Cov}(e_{i_1 j_1}, e_{i_2 j_1}) = 0$$

$$\mathrm{Cov}(u_{j_1}, e_{i_1 j_2}) = 0 \qquad \mathrm{Cov}(e_{i_1 j_1}, e_{i_2 j_2}) = 0$$

$$\mathrm{Cov}(u_j, x_{ij}) = 0$$

$$\mathrm{Cov}(e_{ij}, x_{ij}) = 0$$

- Level 2 residuals for different groups are uncorrelated
- Level 1 residuals for different observations are uncorrelated
- Level 2 and level 1 residuals are uncorrelated
- Residuals and covariates are uncorrelated

V, the correlation matrix

The correlation matrix gives the correlation between every pair of level 1 units in our dataset after controlling for the explanatory variables

Single level model

	1	2	3	4	5	6	7	8	9	10	11	12	13	$\cdots$
1	1	0	0	0	0	0	0	0	0	0	0	0	0	$\cdots$
2	0	1	0	0	0	0	0	0	0	0	0	0	0	$\cdots$
3	0	0	1	0	0	0	0	0	0	0	0	0	0	$\cdots$
4	0	0	0	1	0	0	0	0	0	0	0	0	0	$\cdots$
5	0	0	0	0	1	0	0	0	0	0	0	0	0	$\cdots$
6	0	0	0	0	0	1	0	0	0	0	0	0	0	$\cdots$
7	0	0	0	0	0	0	1	0	0	0	0	0	0	$\cdots$
8	0	0	0	0	0	0	0	1	0	0	0	0	0	$\cdots$
9	0	0	0	0	0	0	0	0	1	0	0	0	0	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

V for random intercepts model

L2		1	1	1	1	2	2	3	3	3	3	$\cdots$
	L1	1	2	3	4	1	2	1	2	3	4	$\cdots$
1	1	1	ρ	ρ	ρ	0	0	0	0	0	0	$\cdots$
1	2	ρ	1	ρ	ρ	0	0	0	0	0	0	$\cdots$
1	3	ρ	ρ	1	ρ	0	0	0	0	0	0	$\cdots$
1	4	ρ	ρ	ρ	1	0	0	0	0	0	0	$\cdots$
2	1	0	0	0	0	1	ρ	0	0	0	0	$\cdots$
2	2	0	0	0	0	ρ	1	0	0	0	0	$\cdots$
3	1	0	0	0	0	0	0	1	ρ	ρ	ρ	$\cdots$
3	2	0	0	0	0	0	0	ρ	1	ρ	ρ	$\cdots$
3	3	0	0	0	0	0	0	ρ	ρ	1	ρ	$\cdots$
3	4	0	0	0	0	0	0	ρ	ρ	ρ	1	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

- The correlation matrix is identical to the matrix for the variance components model

V for random intercepts model

L2		1	1	1	1	2	2	3	3	3	3	$\cdots$
	L1	1	2	3	4	1	2	1	2	3	4	$\cdots$
1	1	1	ρ	ρ	ρ	0	0	0	0	0	0	$\cdots$
1	2	ρ	1	ρ	ρ	0	0	0	0	0	0	$\cdots$
1	3	ρ	ρ	1	ρ	0	0	0	0	0	0	$\cdots$
1	4	ρ	ρ	ρ	1	0	0	0	0	0	0	$\cdots$
2	1	0	0	0	0	1	ρ	0	0	0	0	$\cdots$
2	2	0	0	0	0	ρ	1	0	0	0	0	$\cdots$
3	1	0	0	0	0	0	0	1	ρ	ρ	ρ	$\cdots$
3	2	0	0	0	0	0	0	ρ	1	ρ	ρ	$\cdots$
3	3	0	0	0	0	0	0	ρ	ρ	1	ρ	$\cdots$
3	4	0	0	0	0	0	0	ρ	ρ	ρ	1	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

- The correlation matrix is identical to the matrix for the variance components model
- As expected, observations within the same group are correlated but observations from different groups are uncorrelated

More on correlation matrices

See also the audio presentation on our website at

http://www.cmm.bristol.ac.uk/learning-training/videos/index.shtml#correlation
(which gives details of how we derive the entries of these correlation matrices)

Residuals

- Having looked at the correlation between residuals, let's look at the residuals themselves in more detail.

Residuals

- Having looked at the correlation between residuals, let's look at the residuals themselves in more detail.
- Residuals are estimates for the random part

Residuals

- Having looked at the correlation between residuals, let's look at the residuals themselves in more detail.
- Residuals are estimates for the random part
  - For a single level model, we often say 'estimates for the error term'

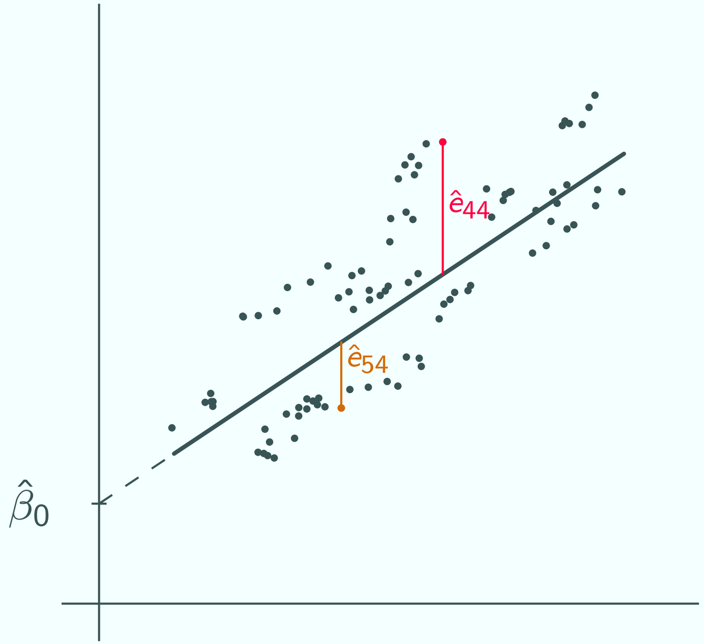
Residuals

- Having looked at the correlation between residuals, let's look at the residuals themselves in more detail.
- Residuals are estimates for the random part
  - For a single level model, we often say 'estimates for the error term'

Reminder

For a single level model, the residual for an observation is an estimate for e_i

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



Residuals

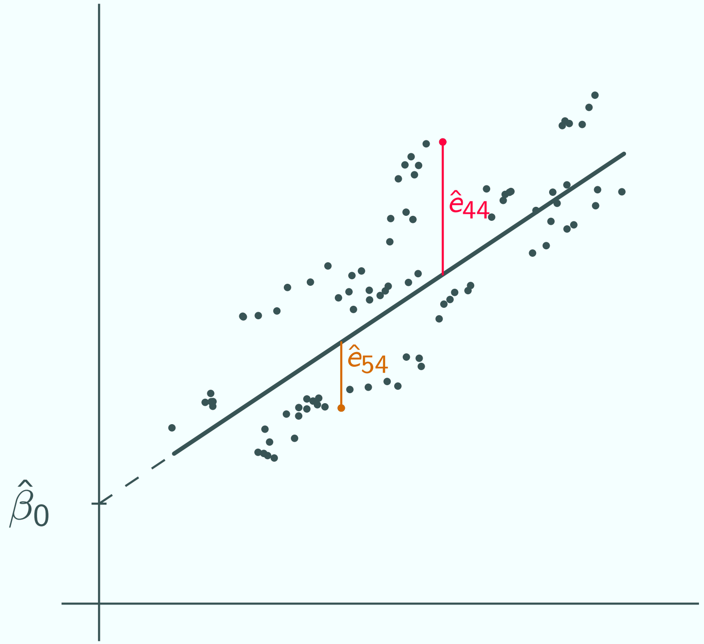
- Having looked at the correlation between residuals, let's look at the residuals themselves in more detail.
- Residuals are estimates for the random part
  - For a single level model, we often say 'estimates for the error term'

Reminder

For a single level model, the residual for an observation is an estimate for e_i

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

If we write $\hat{y}_i = \beta_0 + \beta_1 x_i$ then $\hat{e}_i = y_i - \hat{y}_i$:
the observed value − the value predicted by the regression line



Why are we interested in the residuals?

Often we're not, but they can be useful in some cases:

Diagnostics

- We can plot the residuals to check their Normality
- This is part of checking how well the model fits

Rankings

Interest in a unit

Prediction/ visualisation

Why are we interested in the residuals?

Often we're not, but they can be useful in some cases:

Diagnostics

- We can plot the residuals to check their Normality
- This is part of checking how well the model fits

Rankings

- We can rank level 2 units by u_j
- e.g. school league tables

Interest in a unit

Prediction/ visualisation

Why are we interested in the residuals?

Often we're not, but they can be useful in some cases:

Diagnostics

- We can plot the residuals to check their Normality
- This is part of checking how well the model fits

Rankings

- We can rank level 2 units by u_j
- e.g. school league tables

Interest in a unit

- We can find out how a particular unit compares to the average

Prediction/ visualisation

Why are we interested in the residuals?

Often we're not, but they can be useful in some cases:

Diagnostics

- We can plot the residuals to check their Normality
- This is part of checking how well the model fits

Rankings

- We can rank level 2 units by u_j
- e.g. school league tables

Interest in a unit

- We can find out how a particular unit compares to the average

Prediction/ visualisation

- The level 2 residuals are needed to make predictions for individuals in a particular level 2 unit
- We need them to graph the group lines

Why are we interested in the residuals?

Often we're not, but they can be useful in some cases:

Diagnostics

- 'Was our Normality assumption justified?'

Rankings

Interest in a unit

Prediction/ visualisation

Why are we interested in the residuals?

Often we're not, but they can be useful in some cases:

Diagnostics

- 'Was our Normality assumption justified?'

Rankings

- 'After accounting for intake, which school performs the best?'

Interest in a unit

Prediction/ visualisation

Why are we interested in the residuals?

Often we're not, but they can be useful in some cases:

Diagnostics

- 'Was our Normality assumption justified?'

Rankings

- 'After accounting for intake, which school performs the best?'

Interest in a unit

- 'How is Hospital 18 doing?'
- 'How is Pupil 6 doing compared to the rest of their school?'

Prediction/ visualisation

Why are we interested in the residuals?

Often we're not, but they can be useful in some cases:

Diagnostics

- 'Was our Normality assumption justified?'

Rankings

- 'After accounting for intake, which school performs the best?'

Interest in a unit

- 'How is Hospital 18 doing?'
- 'How is Pupil 6 doing compared to the rest of their school?'

Prediction/ visualisation

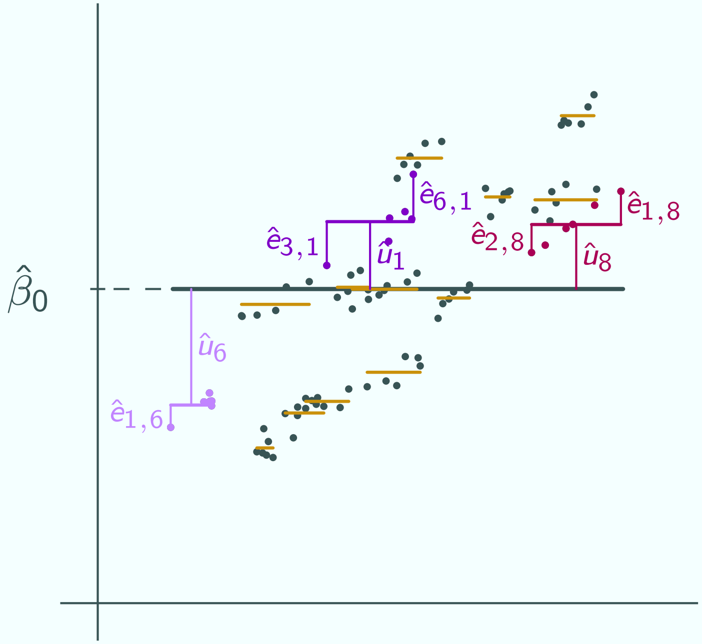
- 'What is the expected weight of a salmon from Fish Farm 28?'
- 'What does our model look like?'

Multilevel residuals

Variance components model

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

Recall that now that we have 2 random terms, we have 2 kinds of residual:



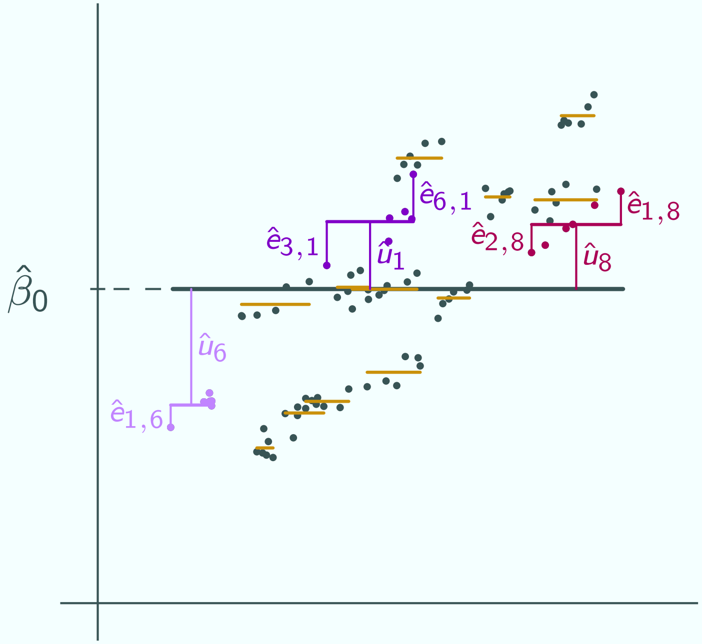
Multilevel residuals

Variance components model

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

Recall that now that we have 2 random terms, we have 2 kinds of residual:

- the level 2 residual, an estimate for u_j



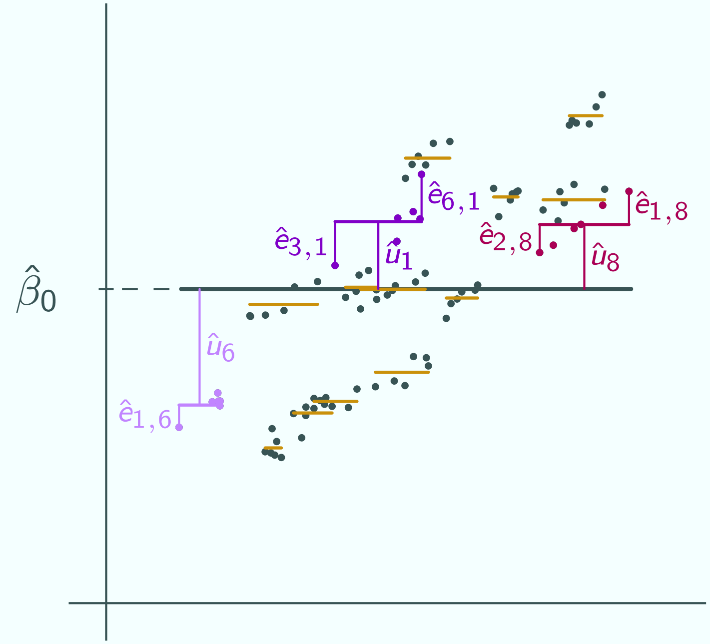
Multilevel residuals

Variance components model

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

Recall that now that we have 2 random terms, we have 2 kinds of residual:

- the level 2 residual, an estimate for u_j
- the level 1 residual, an estimate for e_{ij}

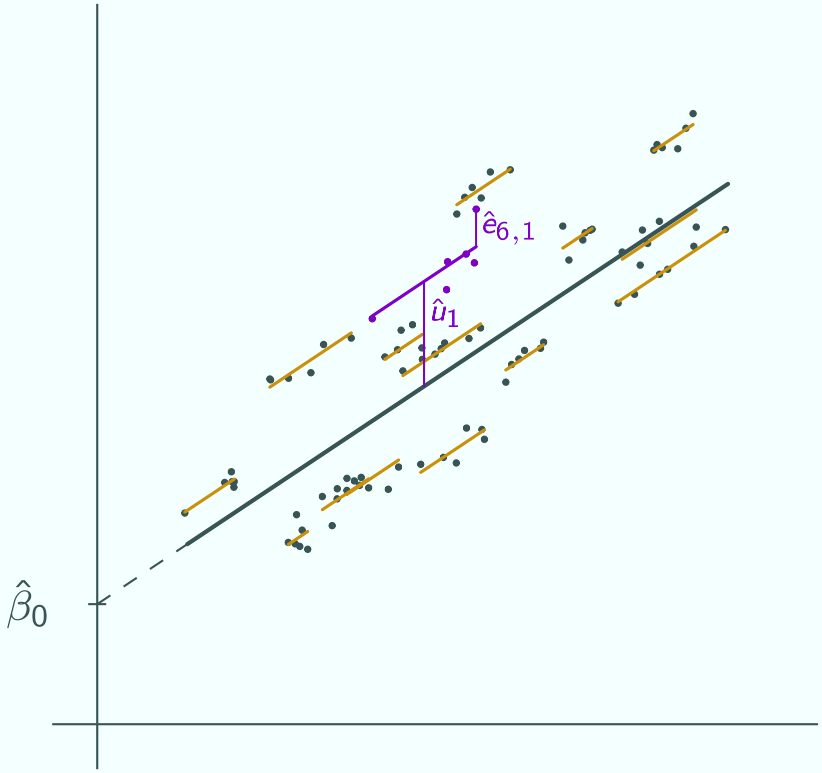


Multilevel residuals

Random intercept model

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

Again now that we have 2 random terms, we have 2 kinds of residual:



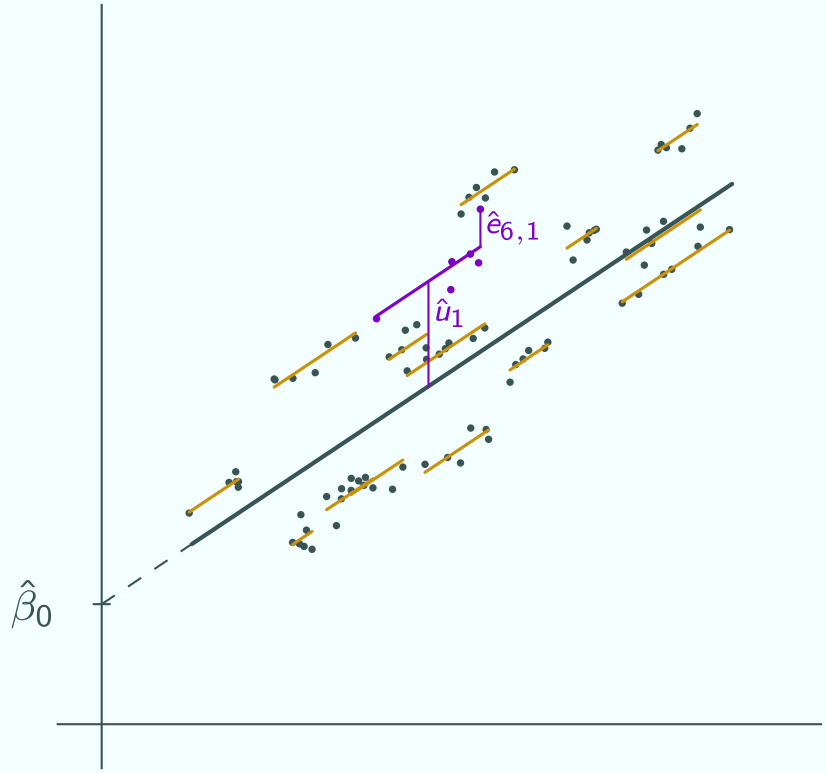
Multilevel residuals

Random intercept model

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

Again now that we have 2 random terms, we have 2 kinds of residual:

- the level 2 residual, an estimate for u_j



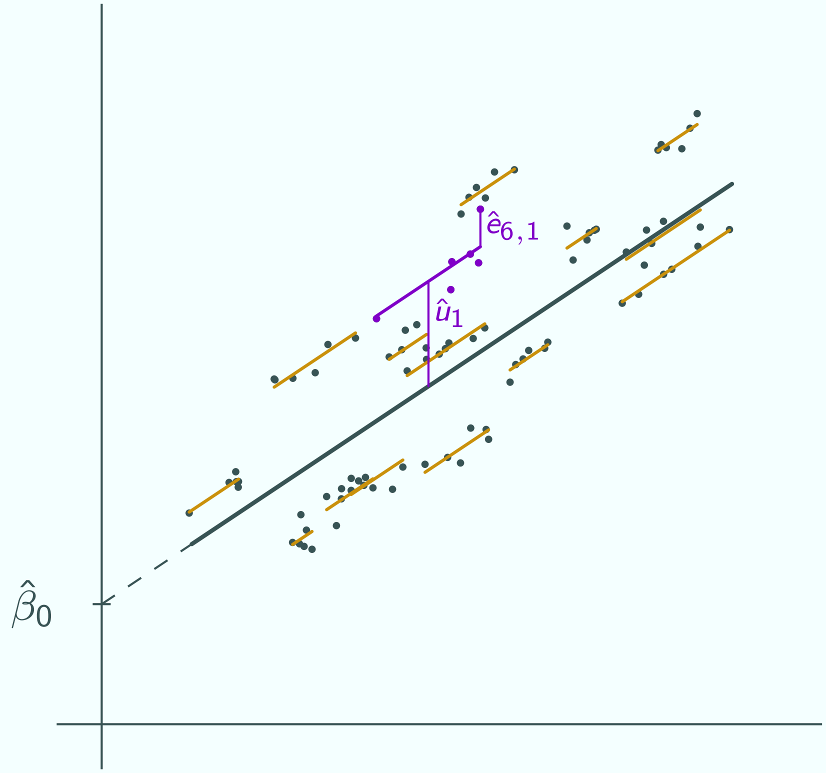
Multilevel residuals

Random intercept model

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

Again now that we have 2 random terms, we have 2 kinds of residual:

- the level 2 residual, an estimate for u_j
- the level 1 residual, an estimate for e_{ij}



Calculation of residuals

Raw mean residual

- Recall $r_{ij} = y_{ij} - \hat{y}_{ij}$ and

Shrinkage factor

Level 2 residual

Level 1 residual

Calculation of residuals

Raw mean residual

- Recall $r_{ij} = y_{ij} - \hat{y}_{ij}$ and
- the raw mean residual $\bar{r}_j$ is the mean of r_{ij} for group j

Shrinkage factor

Level 2 residual

Level 1 residual

Calculation of residuals

Raw mean residual

- Recall $r_{ij} = y_{ij} - \hat{y}_{ij}$ and
- the raw mean residual $\bar{r}_j$ is the mean of r_{ij} for group j

Shrinkage factor

$$k = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_e^2 / n_j}$$

Level 2 residual

Level 1 residual

Calculation of residuals

Raw mean residual

- Recall $r_{ij} = y_{ij} - \hat{y}_{ij}$ and
- the raw mean residual $\bar{r}_j$ is the mean of r_{ij} for group j

Shrinkage factor

$$k = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_e^2/n_j}$$

Be careful! The shrinkage factor is similar to the VPC ρ

Level 2 residual

Level 1 residual

Calculation of residuals

Raw mean residual

- Recall $r_{ij} = y_{ij} - \hat{y}_{ij}$ and
- the raw mean residual $\bar{r}_j$ is the mean of r_{ij} for group j

Shrinkage factor

$$k = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_e^2 / n_j}$$

Be careful! The shrinkage factor is similar to the VPC ρ

Level 2 residual

- Just as for the variance components model, we shrink the raw residuals towards the overall mean

Level 1 residual

Calculation of residuals

Raw mean residual

- Recall $r_{ij} = y_{ij} - \hat{y}_{ij}$ and
- the raw mean residual $\bar{r}_j$ is the mean of r_{ij} for group j

Shrinkage factor

$$k = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_e^2 / n_j}$$

Be careful! The shrinkage factor is similar to the VPC ρ

Level 2 residual

- Just as for the variance components model, we shrink the raw residuals towards the overall mean

$$\hat{u}_j = \bar{r}_j \times k = \bar{r}_j \times \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_e^2 / n_j}$$

Level 1 residual

Calculation of residuals

Raw mean residual

- Recall $r_{ij} = y_{ij} - \hat{y}_{ij}$ and
- the raw mean residual $\bar{r}_j$ is the mean of r_{ij} for group j

Shrinkage factor

$$k = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_e^2 / n_j}$$

Be careful! The shrinkage factor is similar to the VPC ρ

Level 2 residual

- Just as for the variance components model, we shrink the raw residuals towards the overall mean

$$\hat{u}_j = \bar{r}_j \times k = \bar{r}_j \times \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_e^2 / n_j}$$

- Once again, the fixed effects model uses the raw residual r_j

Level 1 residual

Calculation of residuals

Raw mean residual

- Recall $r_{ij} = y_{ij} - \hat{y}_{ij}$ and
- the raw mean residual $\bar{r}_j$ is the mean of r_{ij} for group j

Shrinkage factor

$$k = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_e^2 / n_j}$$

Be careful! The shrinkage factor is similar to the VPC ρ

Level 2 residual

- Just as for the variance components model, we shrink the raw residuals towards the overall mean

$$\hat{u}_j = \bar{r}_j \times k = \bar{r}_j \times \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_e^2 / n_j}$$

- Once again, the fixed effects model uses the raw residual r_j

Level 1 residual

- $\hat{e}_{ij} = y_{ij} - \hat{y}_{ij} - \hat{u}_j = r_{ij} - \hat{u}_j$

Why do we shrink? A thought experiment

The situation

- Suppose we have exam results for 6 pupils each from a number of schools

Why do we shrink? A thought experiment

The situation

- Suppose we have exam results for 6 pupils each from a number of schools
- Now suppose we drop 4 pupils from one school, just to see what will happen

Why do we shrink? A thought experiment

The situation

- Suppose we have exam results for 6 pupils each from a number of schools
- Now suppose we drop 4 pupils from one school, just to see what will happen
- How close will the school line using the remaining 2 pupils be to the "correct" school line using all 6 pupils?

Why do we shrink? A thought experiment

The situation

- Suppose we have exam results for 6 pupils each from a number of schools
- Now suppose we drop 4 pupils from one school, just to see what will happen
- How close will the school line using the remaining 2 pupils be to the "correct" school line using all 6 pupils?

How close is the line?

Why do we shrink? A thought experiment

The situation

- Suppose we have exam results for 6 pupils each from a number of schools
- Now suppose we drop 4 pupils from one school, just to see what will happen
- How close will the school line using the remaining 2 pupils be to the "correct" school line using all 6 pupils?

How close is the line?

- It depends how "typical" those 2 pupils are of the full set of 6.

Why do we shrink? A thought experiment

The situation

- Suppose we have exam results for 6 pupils each from a number of schools
- Now suppose we drop 4 pupils from one school, just to see what will happen
- How close will the school line using the remaining 2 pupils be to the "correct" school line using all 6 pupils?

How close is the line?

- It depends how "typical" those 2 pupils are of the full set of 6.
- Since we're picking 2 pupils from 6, it's quite likely that we might pick 2 untypical pupils

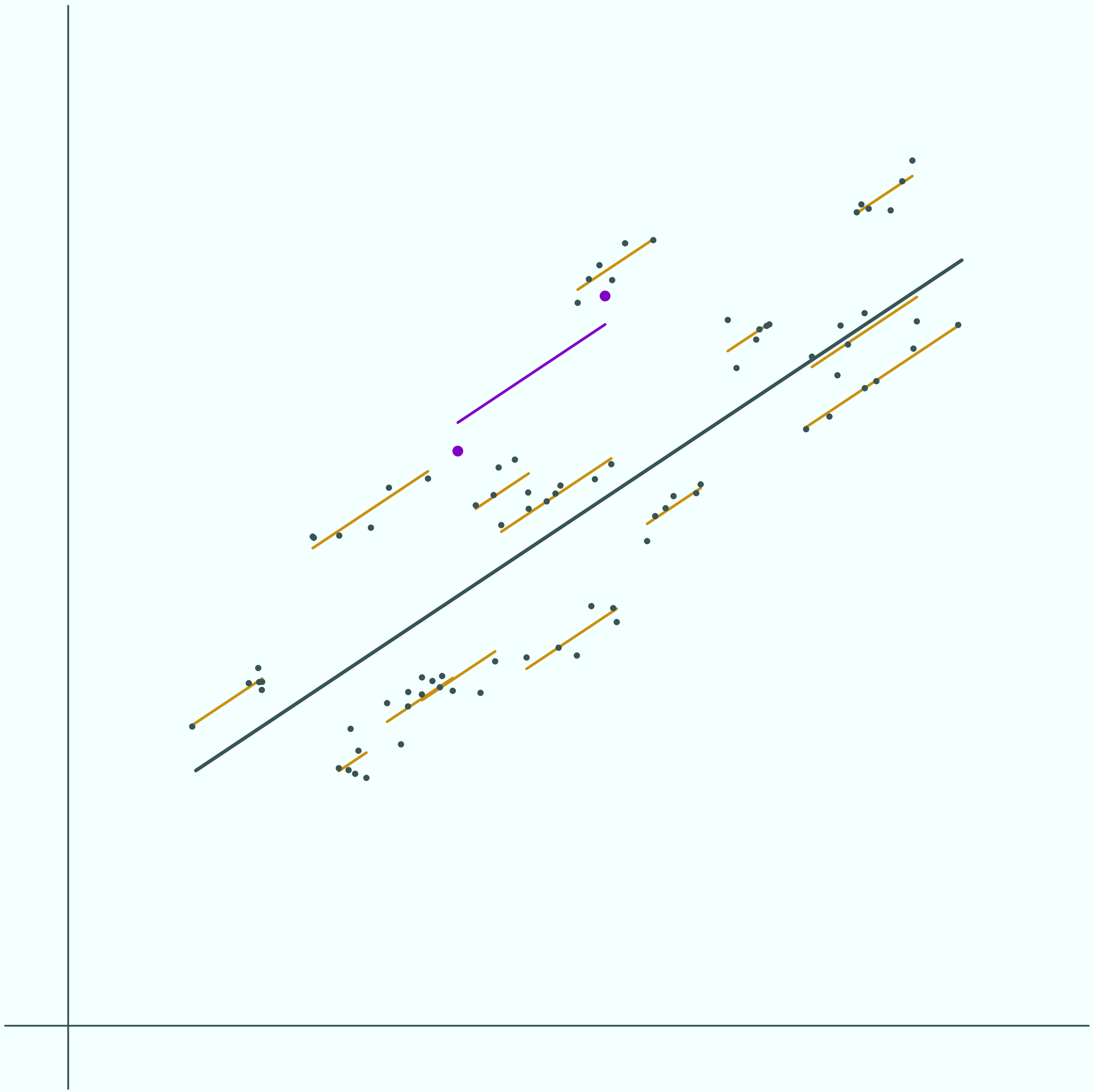
Why do we shrink? A thought experiment

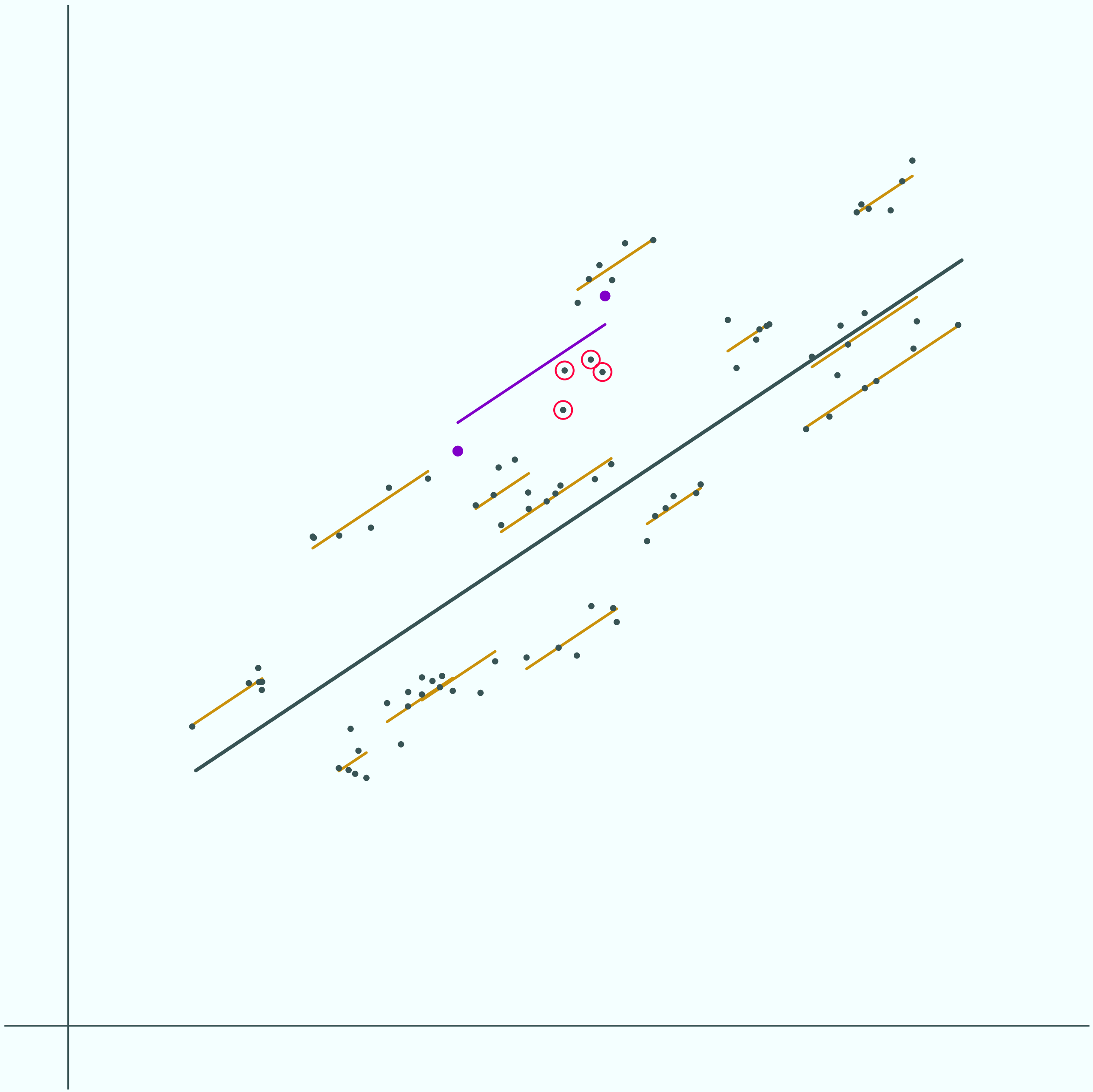
The situation

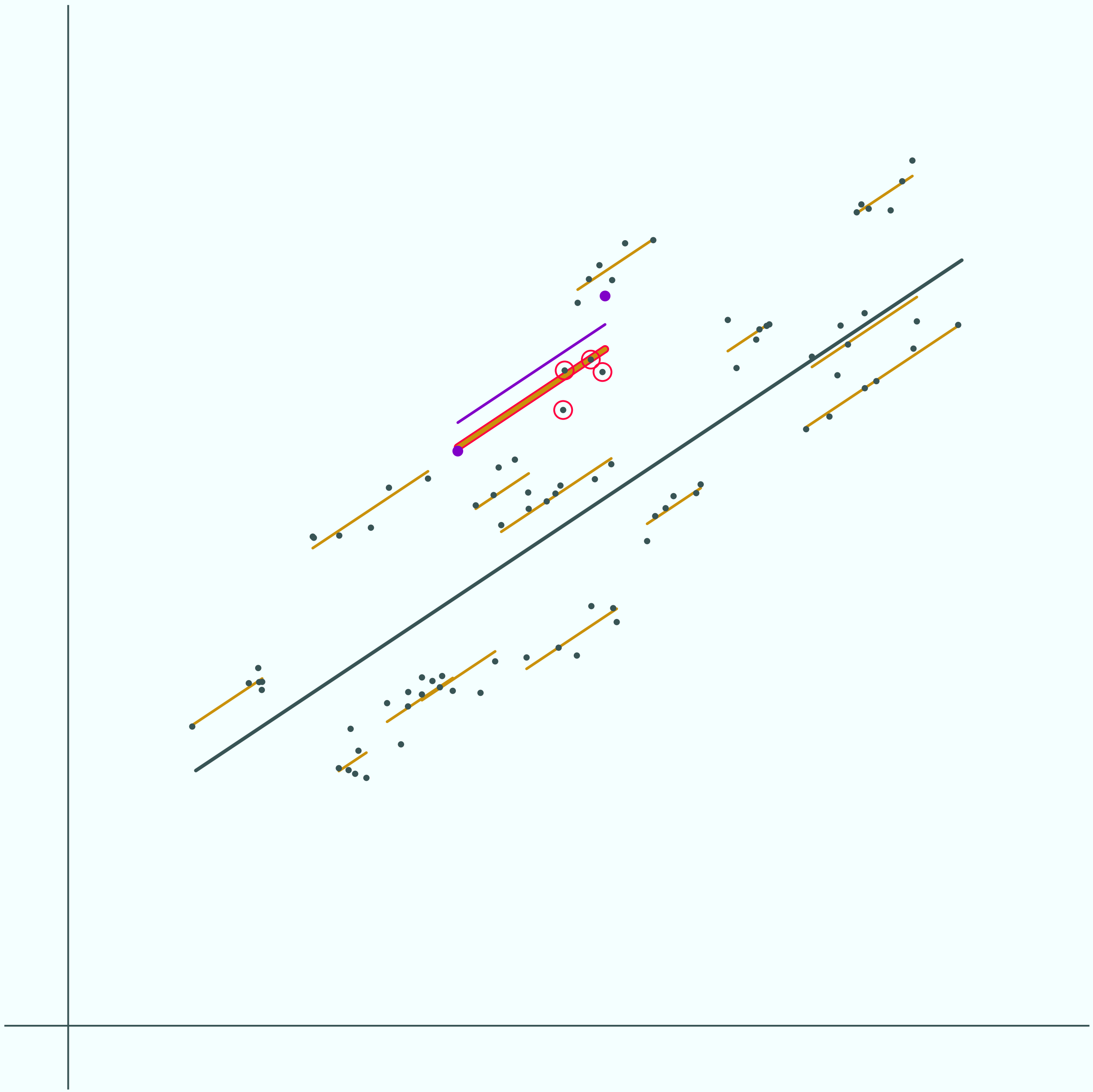
- Suppose we have exam results for 6 pupils each from a number of schools
- Now suppose we drop 4 pupils from one school, just to see what will happen
- How close will the school line using the remaining 2 pupils be to the "correct" school line using all 6 pupils?

How close is the line?

- It depends how "typical" those 2 pupils are of the full set of 6.
- Since we're picking 2 pupils from 6, it's quite likely that we might pick 2 untypical pupils
- Then the school line drawn using those 2 will be quite far from the school line using all 6 pupils- as happens in this example







Thought experiment cont.

Drawing the group line

- We're interested in the line for all 6 pupils

Thought experiment cont.

Drawing the group line

- We're interested in the line for all 6 pupils
- We want our line when we have 2 pupils

Thought experiment cont.

Drawing the group line

- We're interested in the line for all 6 pupils
- We want our line when we have 2 pupils
  - to stay close to the line for all 6 pupils

Thought experiment cont.

Drawing the group line

- We're interested in the line for all 6 pupils
- We want our line when we have 2 pupils
  - to stay close to the line for all 6 pupils
  - not to move around a lot according to which 2 pupils we pick

Thought experiment cont.

Drawing the group line

- We're interested in the line for all 6 pupils
- We want our line when we have 2 pupils
  - to stay close to the line for all 6 pupils
  - not to move around a lot according to which 2 pupils we pick
- How can we draw a line that does those things?

Thought experiment cont.

Drawing the group line

- We're interested in the line for all 6 pupils
- We want our line when we have 2 pupils
  - to stay close to the line for all 6 pupils
  - not to move around a lot according to which 2 pupils we pick
- How can we draw a line that does those things?

Information about the dropped points

Thought experiment cont.

Drawing the group line

- We're interested in the line for all 6 pupils
- We want our line when we have 2 pupils
  - to stay close to the line for all 6 pupils
  - not to move around a lot according to which 2 pupils we pick
- How can we draw a line that does those things?

Information about the dropped points

- To draw a line like that, we need some information on the 4 pupils we dropped

Thought experiment cont.

Drawing the group line

- We're interested in the line for all 6 pupils
- We want our line when we have 2 pupils
    - to stay close to the line for all 6 pupils
    - not to move around a lot according to which 2 pupils we pick
- How can we draw a line that does those things?

Information about the dropped points

- To draw a line like that, we need some information on the 4 pupils we dropped
- It seems we don't have any, but actually that's not true

Thought experiment cont.

Drawing the group line

- We're interested in the line for all 6 pupils
- We want our line when we have 2 pupils
  - to stay close to the line for all 6 pupils
  - not to move around a lot according to which 2 pupils we pick
- How can we draw a line that does those things?

Information about the dropped points

- To draw a line like that, we need some information on the 4 pupils we dropped
- It seems we don't have any, but actually that's not true
- The position of the other school lines tells us something about where the 4 pupils and the line for this school are likely to be:

Thought experiment cont.

Drawing the group line

- We're interested in the line for all 6 pupils
- We want our line when we have 2 pupils
  - to stay close to the line for all 6 pupils
  - not to move around a lot according to which 2 pupils we pick
- How can we draw a line that does those things?

Information about the dropped points

- To draw a line like that, we need some information on the 4 pupils we dropped
- It seems we don't have any, but actually that's not true
- The position of the other school lines tells us something about where the 4 pupils and the line for this school are likely to be:
- they are more likely to be closer to the overall average line.

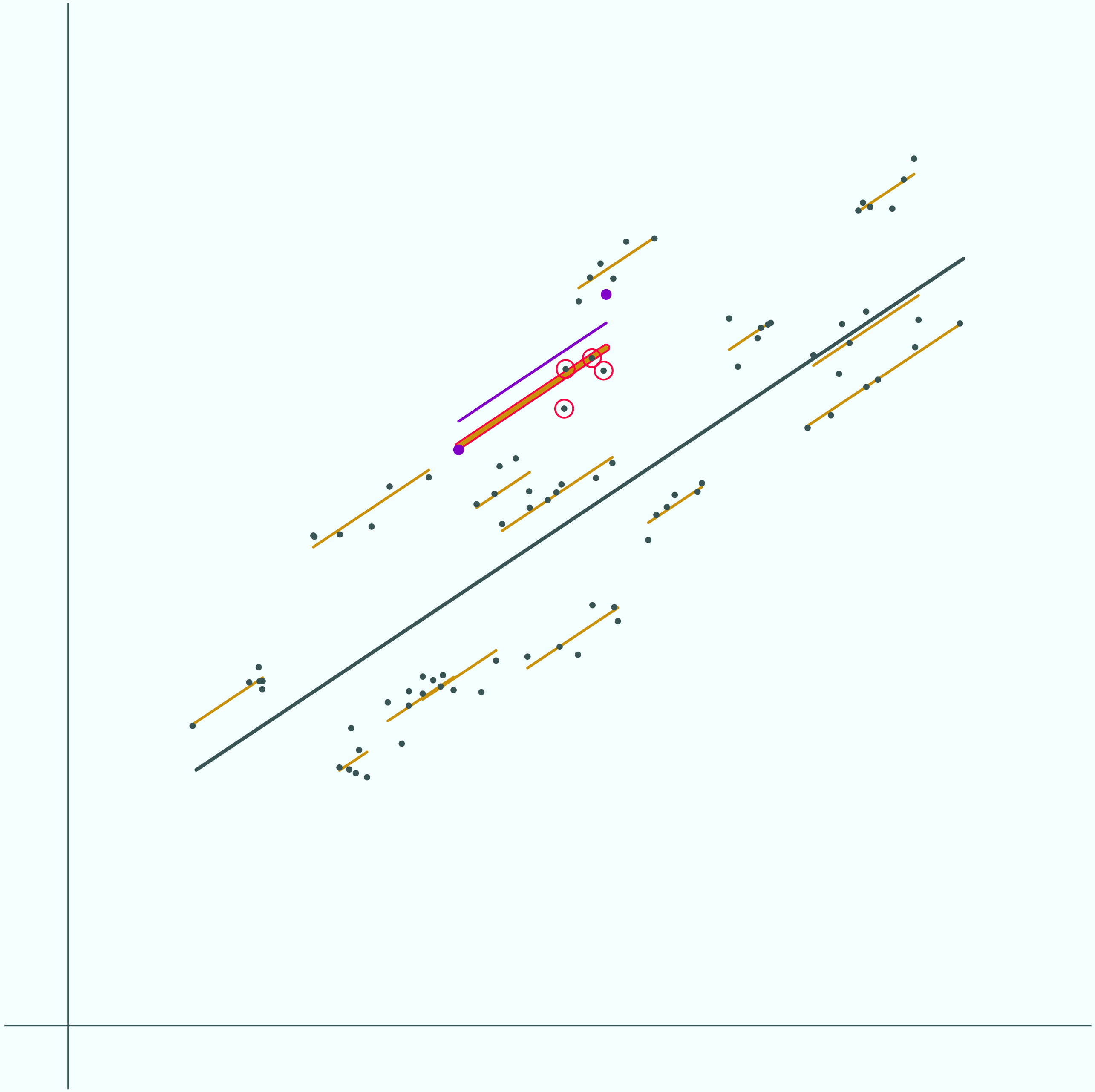
Thought experiment cont.

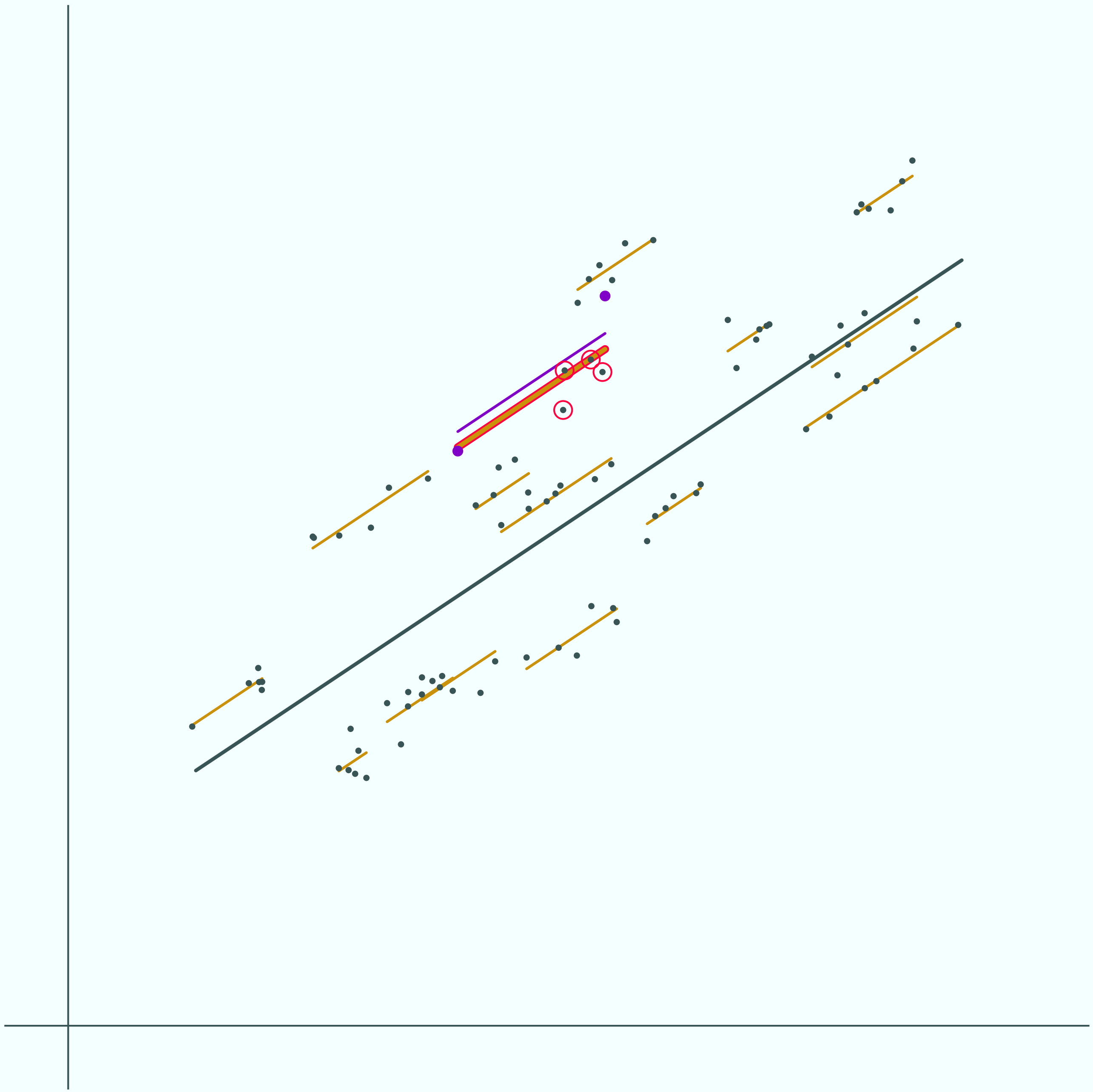
Drawing the group line

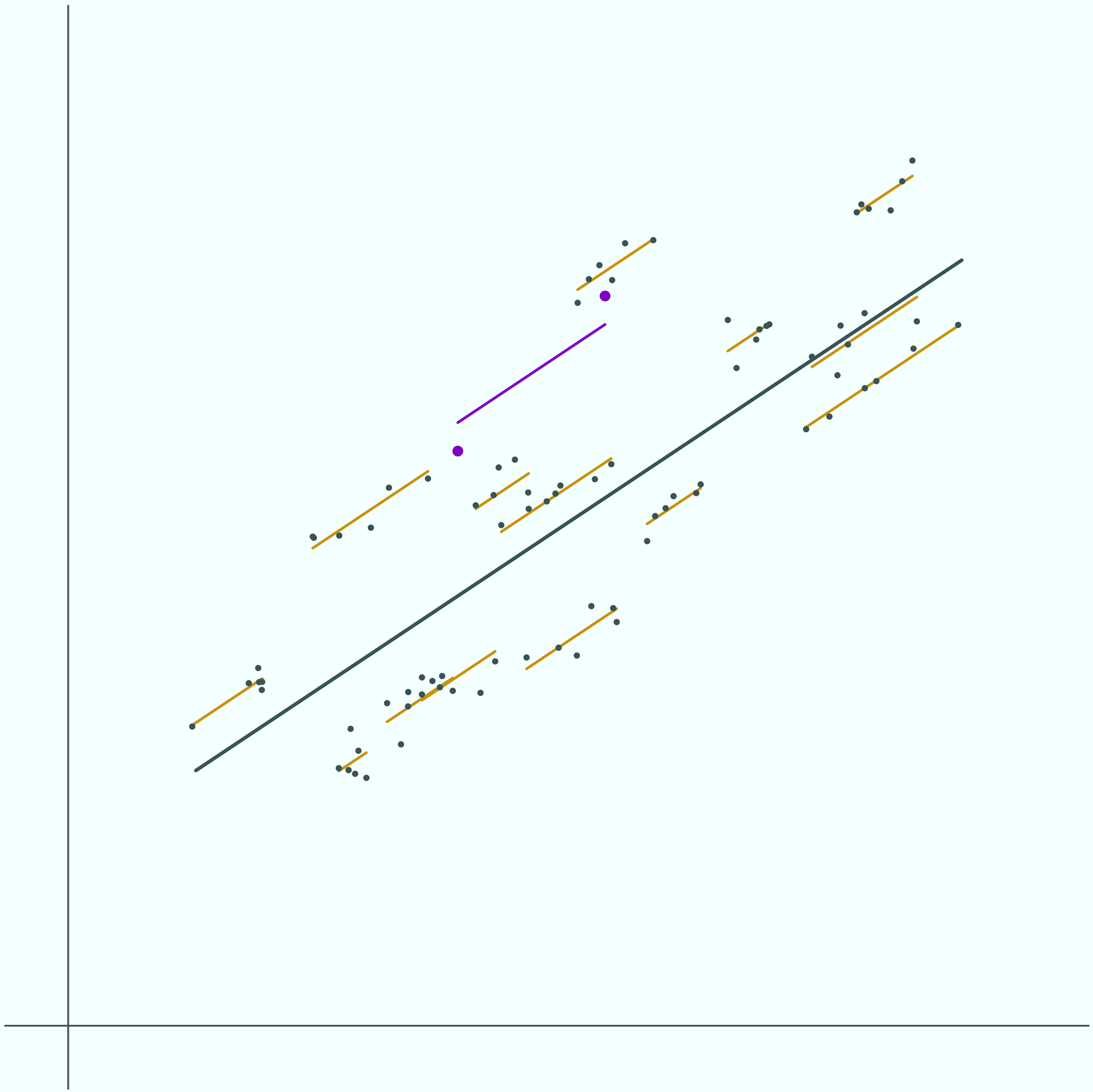
- We're interested in the line for all 6 pupils
- We want our line when we have 2 pupils
  - to stay close to the line for all 6 pupils
  - not to move around a lot according to which 2 pupils we pick
- How can we draw a line that does those things?

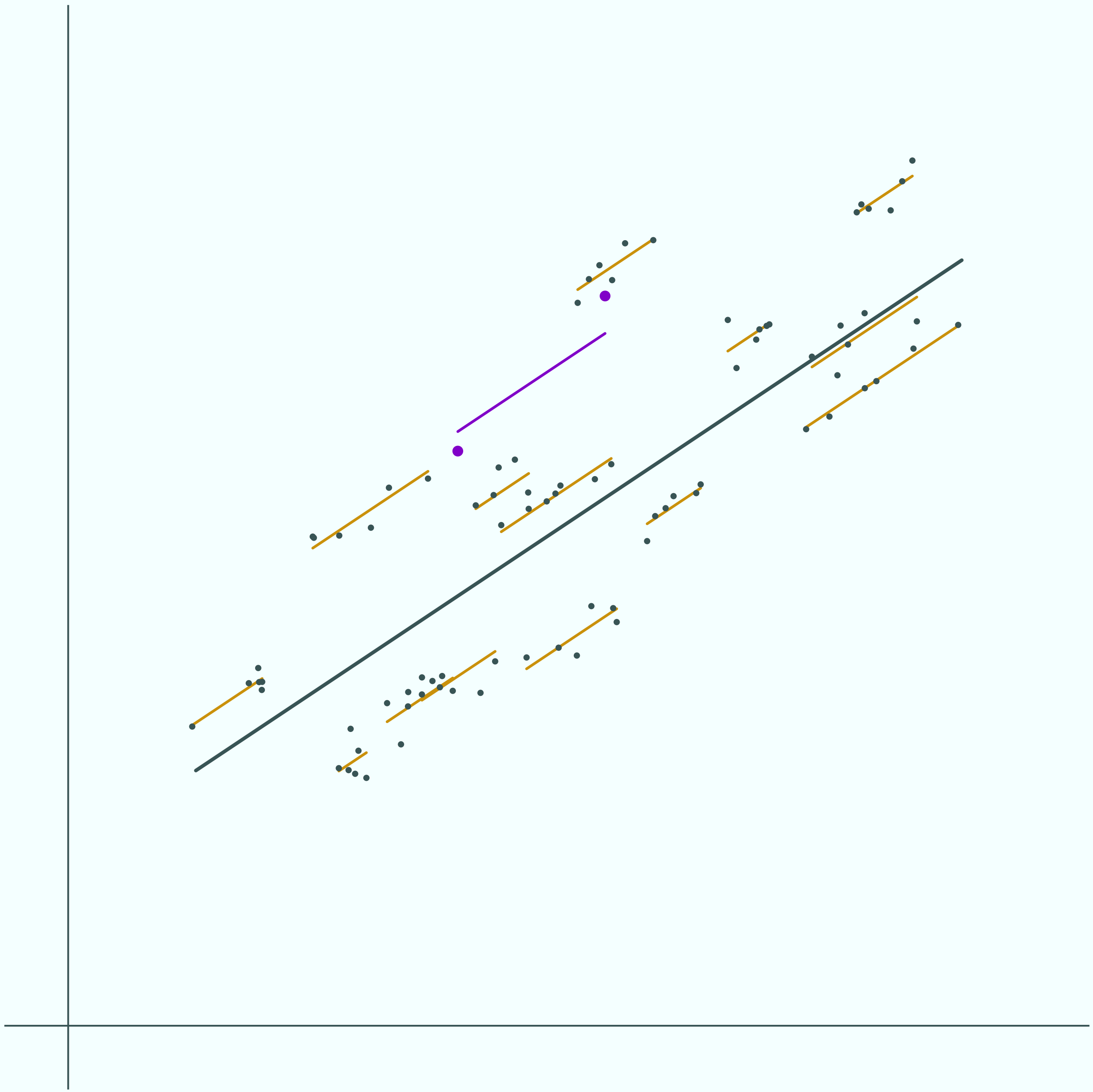
Information about the dropped points

- To draw a line like that, we need some information on the 4 pupils we dropped
- It seems we don't have any, but actually that's not true
- The position of the other school lines tells us something about where the 4 pupils and the line for this school are likely to be:
- they are more likely to be closer to the overall average line.
- So we can improve our positioning of the line by shrinking it in towards the overall average









Thought experiment cont.

Is shrinkage always better?

- It is possible that shrinking the line will move it away from the line using all pupils

Thought experiment cont.

Is shrinkage always better?

- It is possible that shrinking the line will move it away from the line using all pupils
- If we don't know where the other 4 pupils are we don't know if shrinking moves the line closer or further away

Thought experiment cont.

Is shrinkage always better?

- It is possible that shrinking the line will move it away from the line using all pupils
- If we don't know where the other 4 pupils are we don't know if shrinking moves the line closer or further away
- But we do know that shrinking will move the line closer more often than it will move it further away

Thought experiment cont.

Is shrinkage always better?

- It is possible that shrinking the line will move it away from the line using all pupils
- If we don't know where the other 4 pupils are we don't know if shrinking moves the line closer or further away
- But we do know that shrinking will move the line closer more often than it will move it further away
- So shrinking the line is always our best chance to get the line closer

Thought experiment cont.

Is shrinkage always better?

- It is possible that shrinking the line will move it away from the line using all pupils
- If we don't know where the other 4 pupils are we don't know if shrinking moves the line closer or further away
- But we do know that shrinking will move the line closer more often than it will move it further away
- So shrinking the line is always our best chance to get the line closer

How does this generalise?

Thought experiment cont.

Is shrinkage always better?

- It is possible that shrinking the line will move it away from the line using all pupils
- If we don't know where the other 4 pupils are we don't know if shrinking moves the line closer or further away
- But we do know that shrinking will move the line closer more often than it will move it further away
- So shrinking the line is always our best chance to get the line closer

How does this generalise?

- We are in the same situation with our dataset as a whole

Thought experiment cont.

Is shrinkage always better?

- It is possible that shrinking the line will move it away from the line using all pupils
- If we don't know where the other 4 pupils are we don't know if shrinking moves the line closer or further away
- But we do know that shrinking will move the line closer more often than it will move it further away
- So shrinking the line is always our best chance to get the line closer

How does this generalise?

- We are in the same situation with our dataset as a whole
- We have just 6 pupils from each school

Thought experiment cont.

Is shrinkage always better?

- It is possible that shrinking the line will move it away from the line using all pupils
- If we don't know where the other 4 pupils are we don't know if shrinking moves the line closer or further away
- But we do know that shrinking will move the line closer more often than it will move it further away
- So shrinking the line is always our best chance to get the line closer

How does this generalise?

- We are in the same situation with our dataset as a whole
- We have just 6 pupils from each school
- There are really many more

Thought experiment cont.

Is shrinkage always better?

- It is possible that shrinking the line will move it away from the line using all pupils
- If we don't know where the other 4 pupils are we don't know if shrinking moves the line closer or further away
- But we do know that shrinking will move the line closer more often than it will move it further away
- So shrinking the line is always our best chance to get the line closer

How does this generalise?

- We are in the same situation with our dataset as a whole
- We have just 6 pupils from each school
- There are really many more
- The position of the other school lines gives us information about the likely position of the pupils not included in the dataset

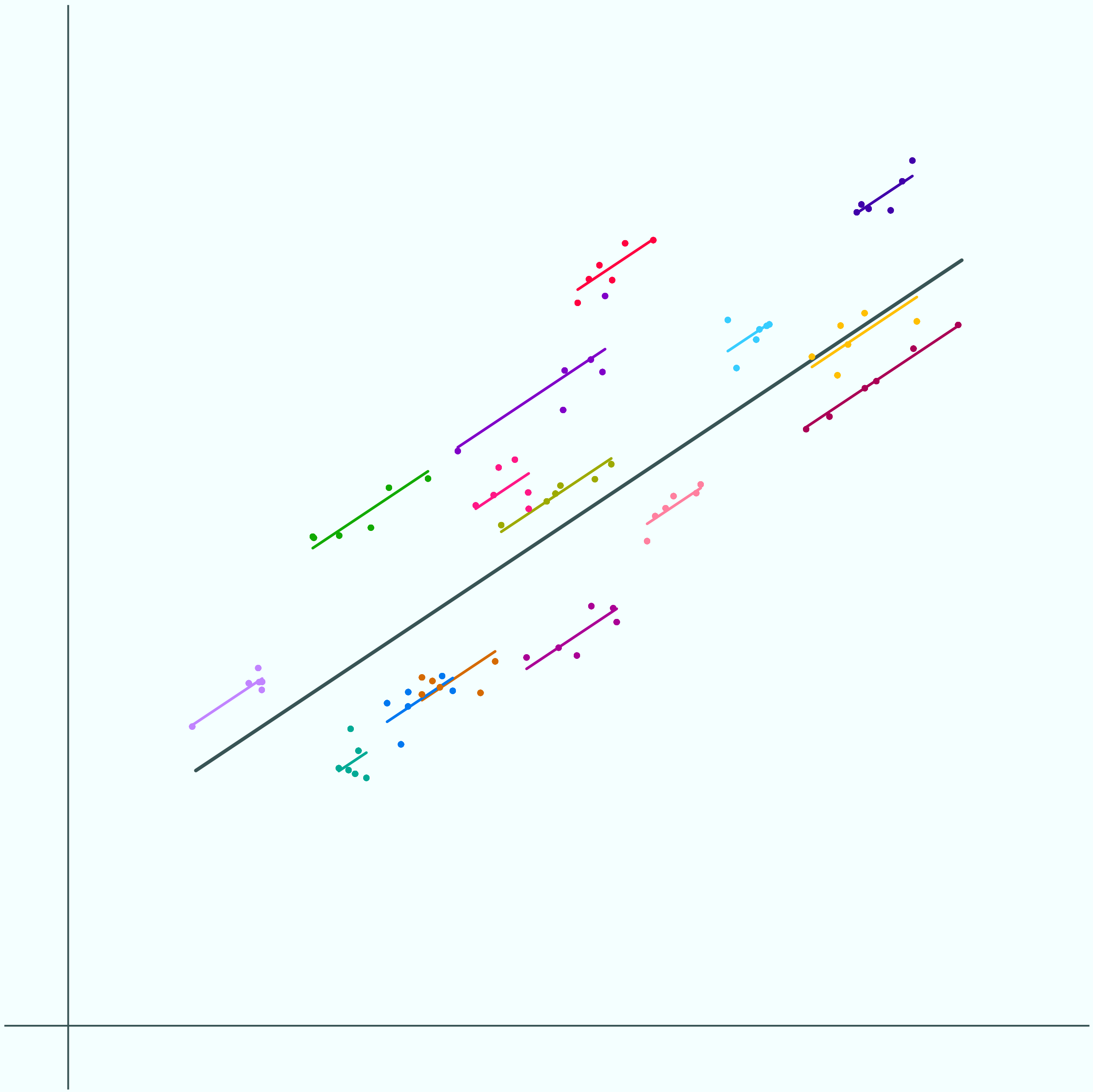
Thought experiment cont.

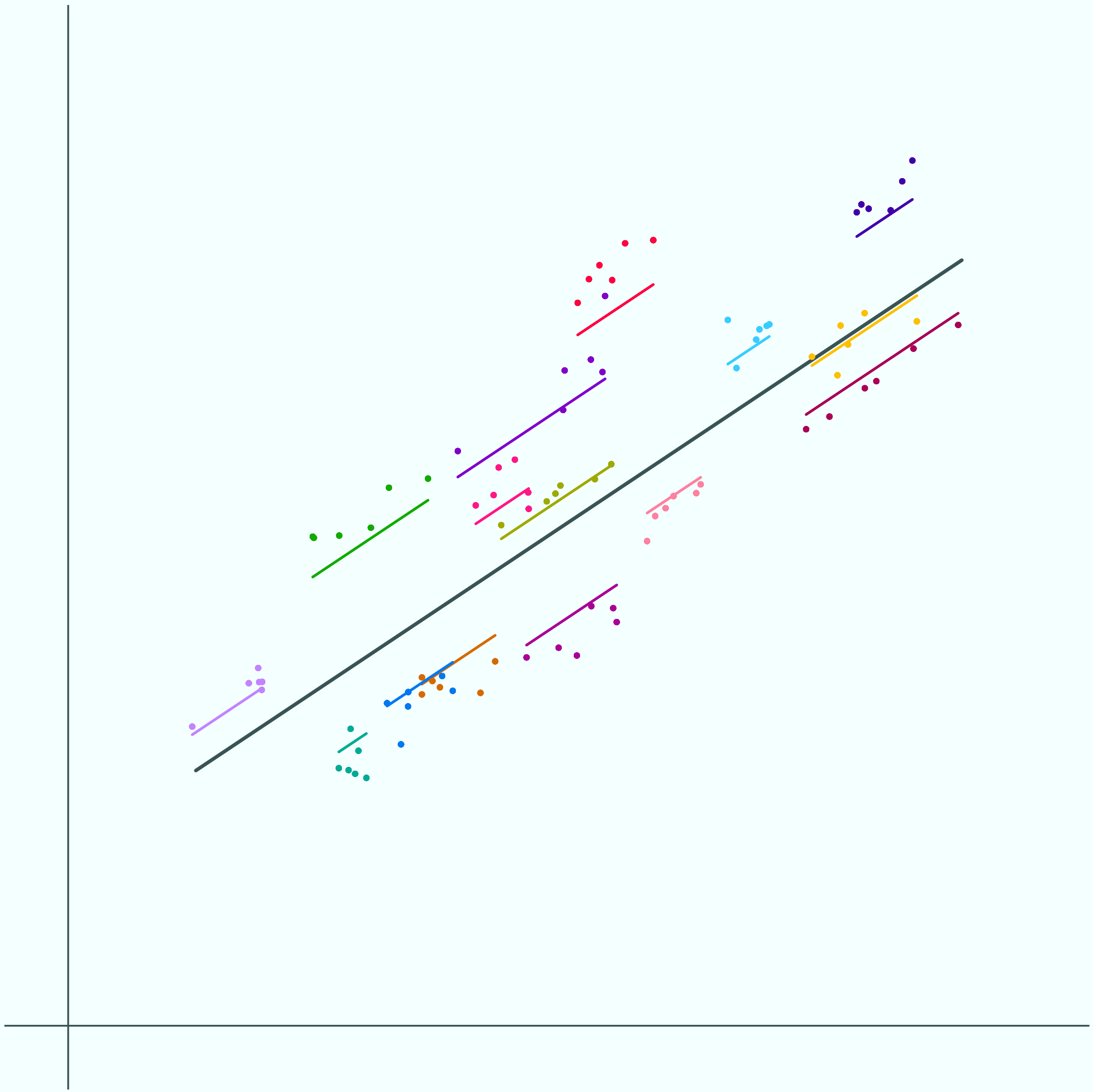
Is shrinkage always better?

- It is possible that shrinking the line will move it away from the line using all pupils
- If we don't know where the other 4 pupils are we don't know if shrinking moves the line closer or further away
- But we do know that shrinking will move the line closer more often than it will move it further away
- So shrinking the line is always our best chance to get the line closer

How does this generalise?

- We are in the same situation with our dataset as a whole
- We have just 6 pupils from each school
- There are really many more
- The position of the other school lines gives us information about the likely position of the pupils not included in the dataset
- Again, we get a better estimate of the group line by shrinking it in towards the overall average





Points to note about shrinkage

When do we shrink?

Always!

- We always shrink the residuals because we always have a sample from each level 2 unit

How much do we shrink by?

Points to note about shrinkage

When do we shrink?

Always!

- We always shrink the residuals because we always have a sample from each level 2 unit
  - even if we have 499 out of the 500 pupils attending a school

How much do we shrink by?

Points to note about shrinkage

When do we shrink?

Always!

- We always shrink the residuals because we always have a sample from each level 2 unit
  - even if we have 499 out of the 500 pupils attending a school
  - even if we have all the pupils attending all the schools in the UK

How much do we shrink by?

Points to note about shrinkage

When do we shrink?

Always!

- We always shrink the residuals because we always have a sample from each level 2 unit
  - even if we have 499 out of the 500 pupils attending a school
  - even if we have all the pupils attending all the schools in the UK
- Even if our dataset contains the whole population, we regard that as a sample from the superpopulation and shrink

How much do we shrink by?

Points to note about shrinkage

When do we shrink?

Always!

- We always shrink the residuals because we always have a sample from each level 2 unit
  - even if we have 499 out of the 500 pupils attending a school
  - even if we have all the pupils attending all the schools in the UK
- Even if our dataset contains the whole population, we regard that as a sample from the superpopulation and shrink

How much do we shrink by?

Recall that:

Points to note about shrinkage

When do we shrink?

Always!

- We always shrink the residuals because we always have a sample from each level 2 unit
  - even if we have 499 out of the 500 pupils attending a school
  - even if we have all the pupils attending all the schools in the UK
- Even if our dataset contains the whole population, we regard that as a sample from the superpopulation and shrink

How much do we shrink by?

Recall that:

- We don't shrink by a fixed amount

Points to note about shrinkage

When do we shrink?

Always!

- We always shrink the residuals because we always have a sample from each level 2 unit
  - even if we have 499 out of the 500 pupils attending a school
  - even if we have all the pupils attending all the schools in the UK
- Even if our dataset contains the whole population, we regard that as a sample from the superpopulation and shrink

How much do we shrink by?

Recall that:

- We don't shrink by a fixed amount
  - If we have 500 pupils from a school we shrink less than if we have 7

Points to note about shrinkage

When do we shrink?

Always!

- We always shrink the residuals because we always have a sample from each level 2 unit
  - even if we have 499 out of the 500 pupils attending a school
  - even if we have all the pupils attending all the schools in the UK
- Even if our dataset contains the whole population, we regard that as a sample from the superpopulation and shrink

How much do we shrink by?

Recall that:

- We don't shrink by a fixed amount
  - If we have 500 pupils from a school we shrink less than if we have 7
- The amount we shrink by depends on the absolute number of level 1 units, not the proportion of the total for that level 2 unit

Points to note about shrinkage

When do we shrink?

Always!

- We always shrink the residuals because we always have a sample from each level 2 unit
  - even if we have 499 out of the 500 pupils attending a school
  - even if we have all the pupils attending all the schools in the UK
- Even if our dataset contains the whole population, we regard that as a sample from the superpopulation and shrink

How much do we shrink by?

Recall that:

- We don't shrink by a fixed amount
  - If we have 500 pupils from a school we shrink less than if we have 7
- The amount we shrink by depends on the absolute number of level 1 units, not the proportion of the total for that level 2 unit
- We can also see that the amount of shrinkage depends on the variances σ_u^2 and σ_e^2

Points to note about shrinkage

- Our estimate of the level 2 variance, $\hat{\sigma}_u^2$, is not the variance of the shrunken residuals

Points to note about shrinkage

- Our estimate of the level 2 variance, $\hat{\sigma}_u^2$, is not the variance of the shrunken residuals
- It is the variance of the residuals before they are shrunk, i.e. the variance of the raw mean residuals

Points to note about shrinkage

- Our estimate of the level 2 variance, $\hat{\sigma}_u^2$, is not the variance of the shrunken residuals
- It is the variance of the residuals before they are shrunk, i.e. the variance of the raw mean residuals
- This variance is then used to calculate the shrinkage factor

Points to note about shrinkage

- Our estimate of the level 2 variance, $\hat{\sigma}_u^2$, is not the variance of the shrunken residuals
- It is the variance of the residuals before they are shrunk, i.e. the variance of the raw mean residuals
- This variance is then used to calculate the shrinkage factor
- $\hat{\sigma}_u^2$ is the estimated variance of the level 2 units in the population not in our sample

Predictions

There are several reasons for making predictions:

Model testing

Model visualisation

Estimates for units not in the dataset

Predictions

There are several reasons for making predictions:

Model testing

To see how close predictions from the model are to the values we observe

Model visualisation

Estimates for units not in the dataset

Predictions

There are several reasons for making predictions:

Model testing

To see how close predictions from the model are to the values we observe

Model visualisation

To try to understand what happens to our response when we change the values of the explanatory variables, using graphs

Estimates for units not in the dataset

Predictions

There are several reasons for making predictions:

Model testing

To see how close predictions from the model are to the values we observe

Model visualisation

To try to understand what happens to our response when we change the values of the explanatory variables, using graphs

Estimates for units not in the dataset

To obtain an estimate for the value of the response for units not in the dataset

Predictions

There are several reasons for making predictions:

Model testing

To see how close predictions from the model are to the values we observe

Model visualisation

To try to understand what happens to our response when we change the values of the explanatory variables, using graphs

Estimates for units not in the dataset

To obtain an estimate for the value of the response for units not in the dataset

- that you have values of the explanatory variables for

Predictions

There are several reasons for making predictions:

Model testing

To see how close predictions from the model are to the values we observe

Model visualisation

To try to understand what happens to our response when we change the values of the explanatory variables, using graphs

Estimates for units not in the dataset

To obtain an estimate for the value of the response for units not in the dataset

- that you have values of the explanatory variables for
- existing units or hypothetical ones

Predictions

There are several reasons for making predictions:

Model testing

To see how close predictions from the model are to the values we observe

Model visualisation

To try to understand what happens to our response when we change the values of the explanatory variables, using graphs

Estimates for units not in the dataset

To obtain an estimate for the value of the response for units not in the dataset

- that you have values of the explanatory variables for
- existing units or hypothetical ones

We focus on the second use

Visualising the model

This should already be familiar from single level models

Visualising the model

This should already be familiar from single level models

Visualising a single level model

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

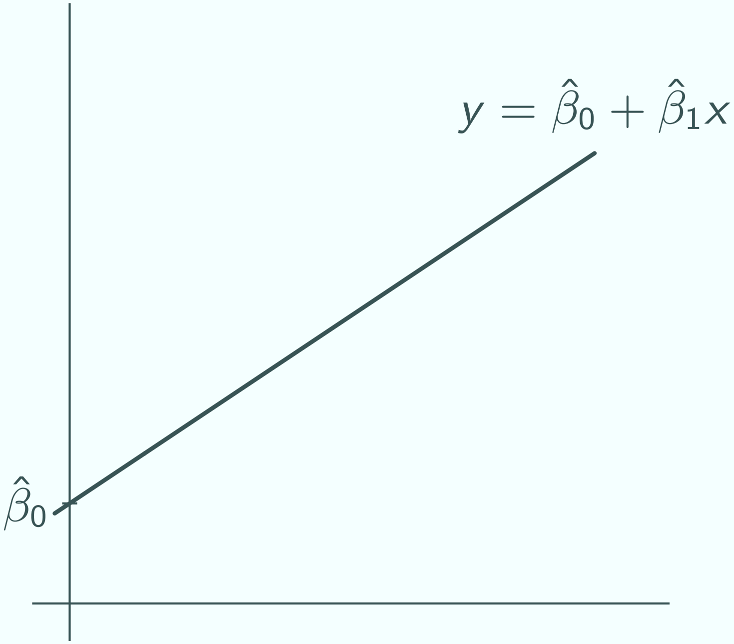
Visualising the model

This should already be familiar from single level models

Visualising a single level model

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

- We plot $y = \hat{\beta}_0 + \hat{\beta}_1 x$ to get our graph



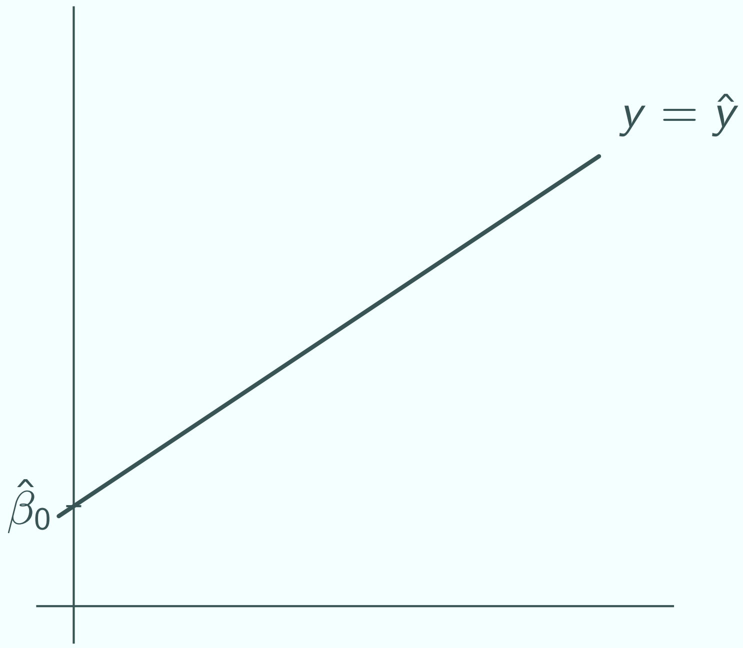
Visualising the model

This slide should already look familiar from single level models

Visualising a single level model

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

- We plot $y = \hat{\beta}_0 + \hat{\beta}_1 x$ to get our graph
- $\hat{\beta}_0 + \hat{\beta}_1 x$ is actually $\hat{y}$, our predicted value.



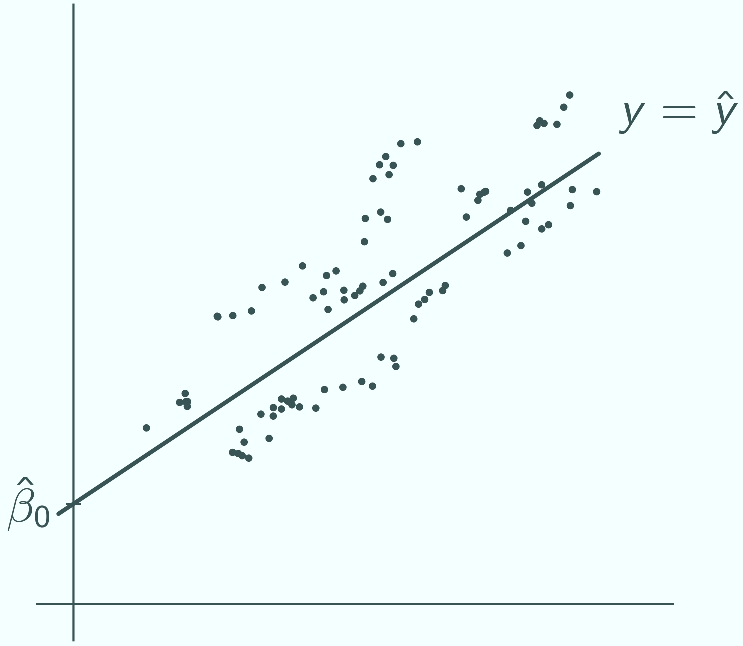
Visualising the model

This should already be familiar from single level models

Visualising a single level model

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

- We plot $y = \hat{\beta}_0 + \hat{\beta}_1 x$ to get our graph
- $\hat{\beta}_0 + \hat{\beta}_1 x$ is actually $\hat{y}$, our predicted value.
- We can add on the actual data points



Visualising the random intercepts model

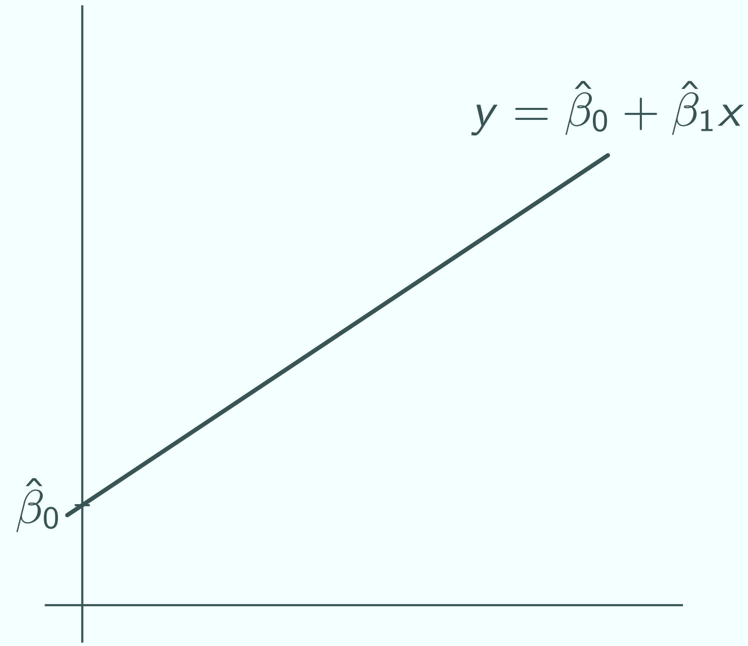
Overall regression line

- Prediction from the fixed part gives the overall regression line

Visualising the random intercepts model

Overall regression line

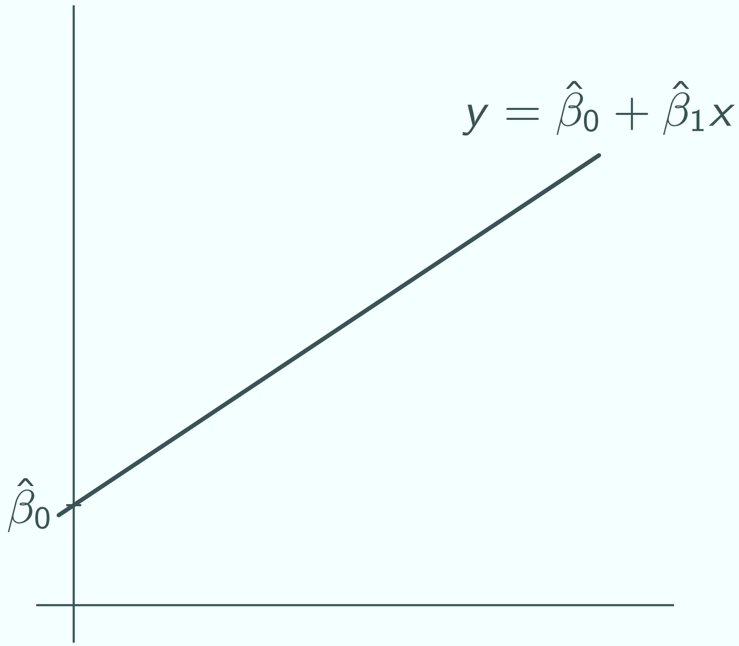
- Prediction from the fixed part gives the overall regression line
- Prediction: $\hat{y}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 x_{1ij}$



Visualising the random intercepts model

Overall regression line

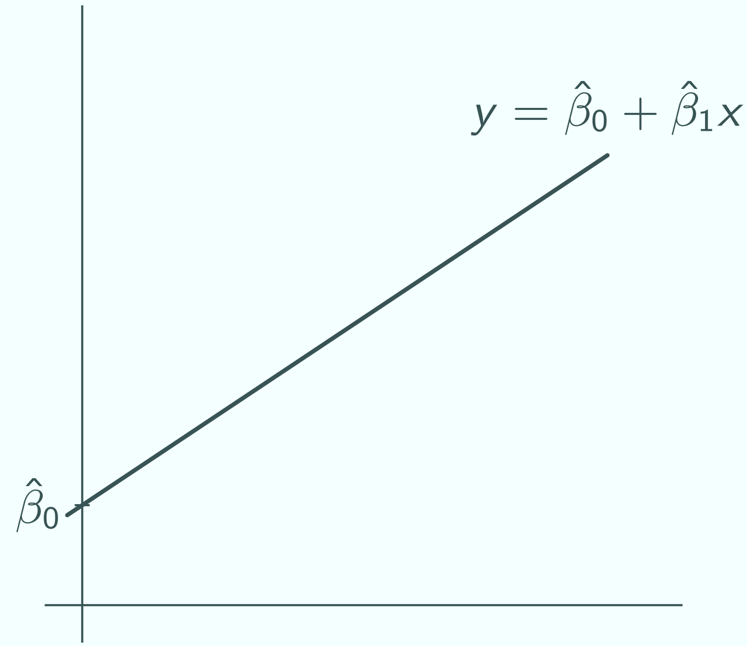
- Prediction from the fixed part gives the overall regression line
- Prediction: $\hat{y}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 x_{1ij}$
- This is just the same as the line for the single level model



Visualising the random intercepts model

Overall regression line

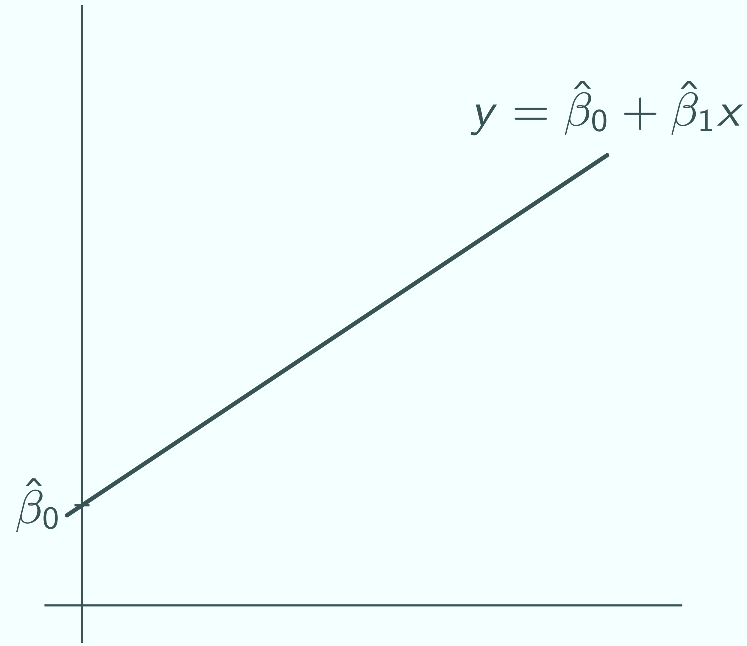
- Prediction from the fixed part gives the overall regression line
- Prediction: $\hat{y}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 x_{1ij}$
- This is just the same as the line for the single level model
- The value of $\hat{y}_{ij}$ does not depend on the group j in this case, only the explanatory variables



Visualising the random intercepts model

Overall regression line

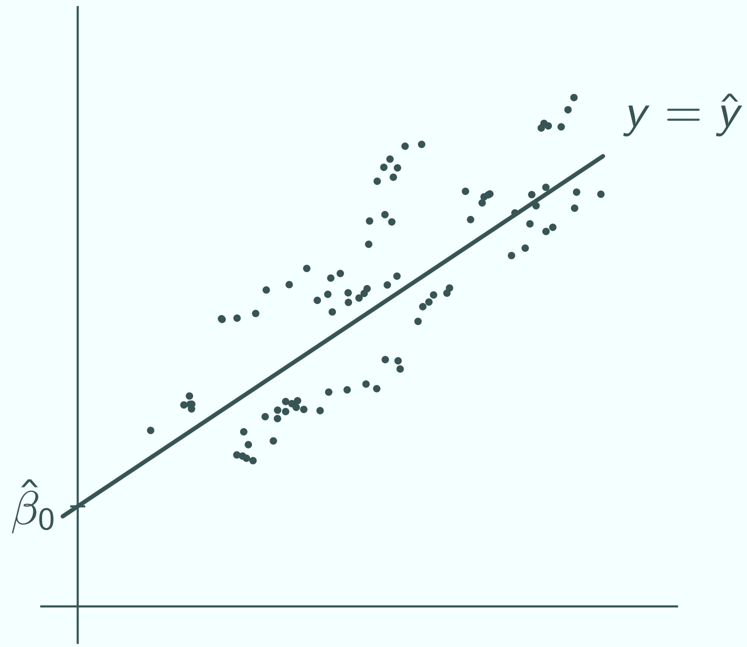
- Prediction from the fixed part gives the overall regression line
- Prediction: $\hat{y}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 x_{1ij}$
- This is just the same as the line for the single level model
- The value of $\hat{y}_{ij}$ does not depend on the group j in this case, only the explanatory variables
- So this prediction only produces one line



Visualising the random intercepts model

Overall regression line

- Prediction from the fixed part gives the overall regression line
- Prediction: $\hat{y}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 x_{1ij}$
- This is just the same as the line for the single level model

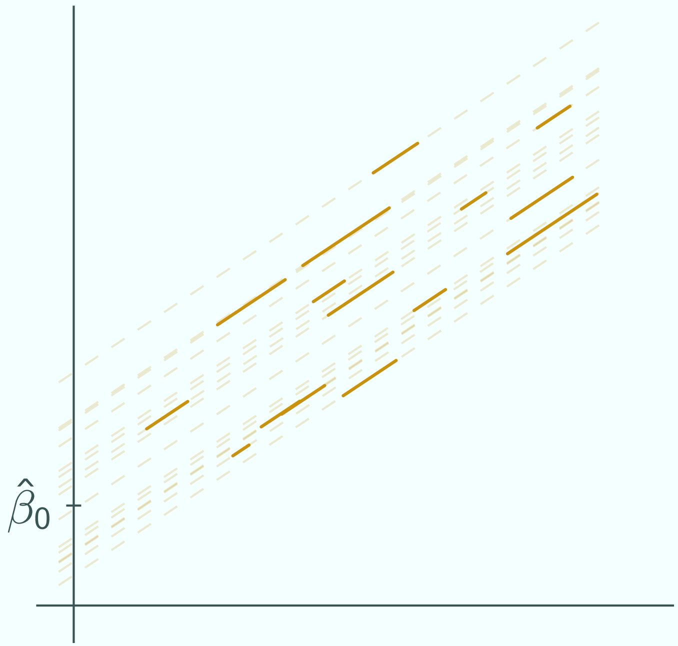


- The value of $\hat{y}_{ij}$ does not depend on the group j in this case, only the explanatory variables
- So this prediction only produces one line
- This is what we would predict if we didn't know which group a data point belonged to

Visualising the random intercepts model

Group lines

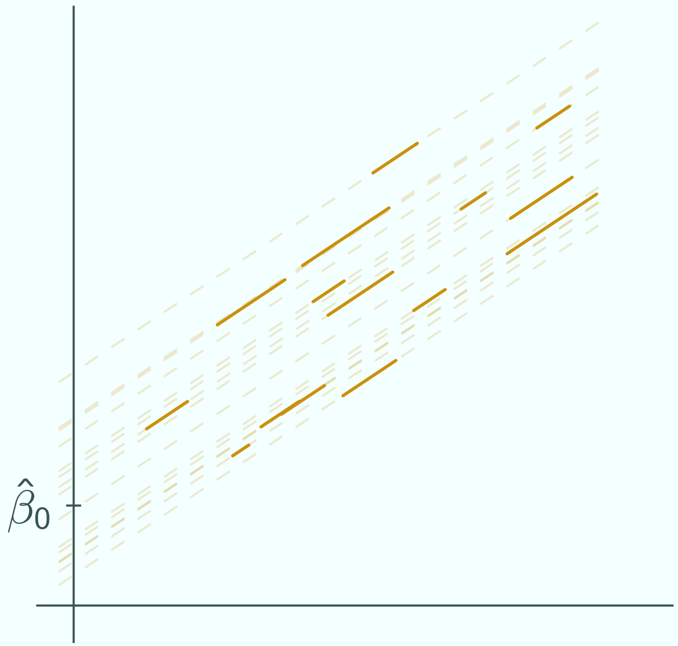
- Adding in the group residual u_j gives the group lines



Visualising the random intercepts model

Group lines

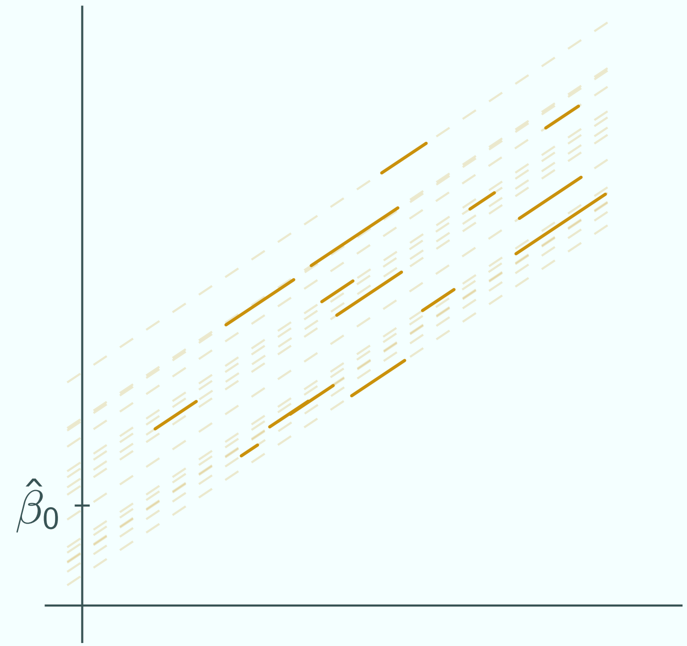
- Adding in the group residual u_j gives the group lines
- Prediction: $\hat{y}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 x_{ij} + \hat{u}_j$



Visualising the random intercepts model

Group lines

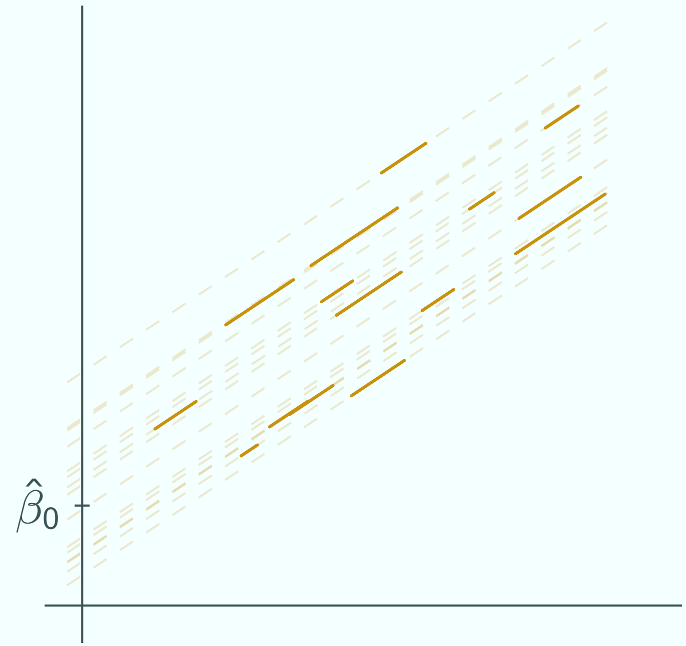
- Adding in the group residual u_j gives the group lines
- Prediction:
  $\hat{y}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 x_{ij} + \hat{u}_j$
- Now the value of $\hat{y}_{ij}$ depends on the group j as well as the explanatory variables



Visualising the random intercepts model

Group lines

- Adding in the group residual u_j gives the group lines
- Prediction: $\hat{y}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 x_{ij} + \hat{u}_j$
- Now the value of $\hat{y}_{ij}$ depends on the group j as well as the explanatory variables

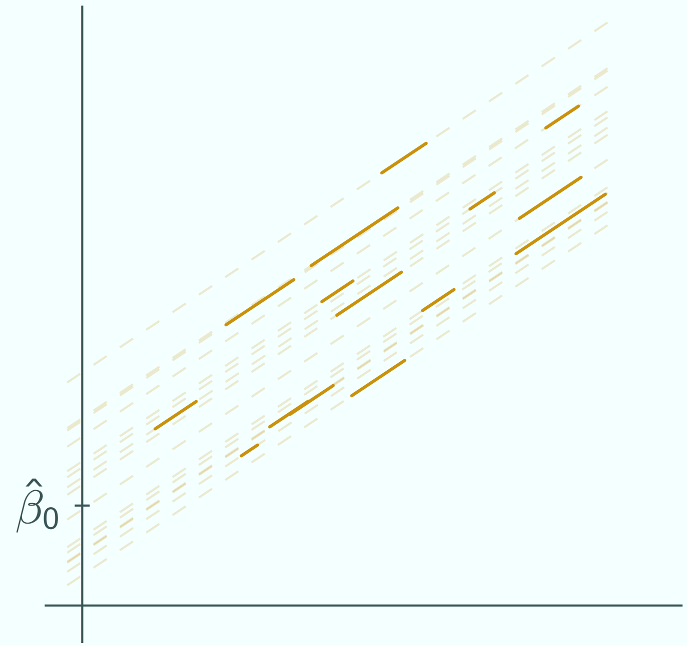


- So there is a different line for each group

Visualising the random intercepts model

Group lines

- Adding in the group residual u_j gives the group lines
- Prediction: $\hat{y}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 x_{ij} + \hat{u}_j$
- Now the value of $\hat{y}_{ij}$ depends on the group j as well as the explanatory variables

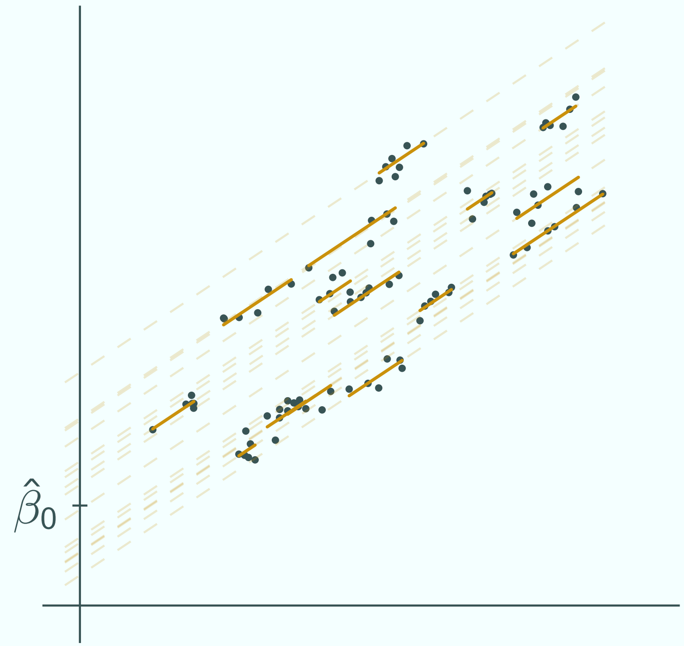


- So there is a different line for each group
- The group line is what we would predict if we did know which group a data point belonged to

Visualising the random intercepts model

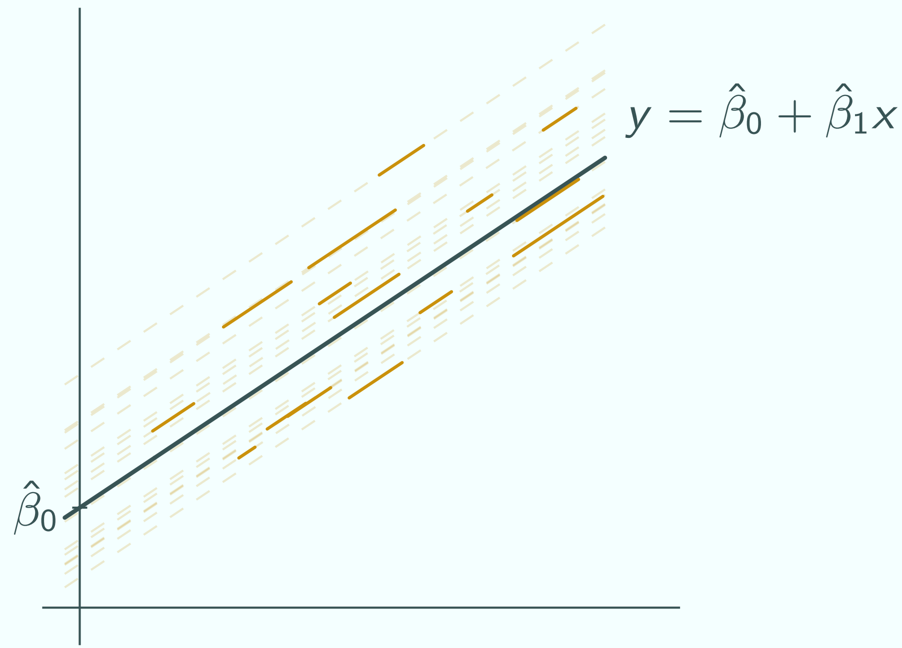
Group lines

- Adding in the group residual u_j gives the group lines
- Prediction:
  $\hat{y}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 x_{ij} + \hat{u}_j$
- Now the value of $\hat{y}_{ij}$ depends on the group j as well as the explanatory variables



- So there is a different line for each group
- The group line is what we would predict if we did know which group a data point belonged to

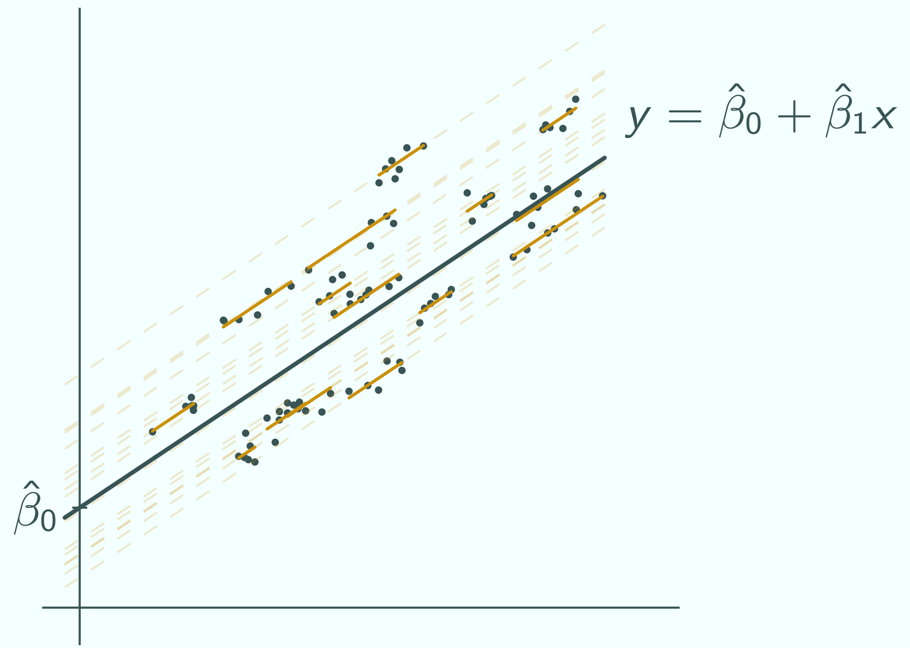
Visualisng the random intercepts model



Complete model

- We can combine the predictions from the fixed and random part in one graph to get a complete visualisation of the model

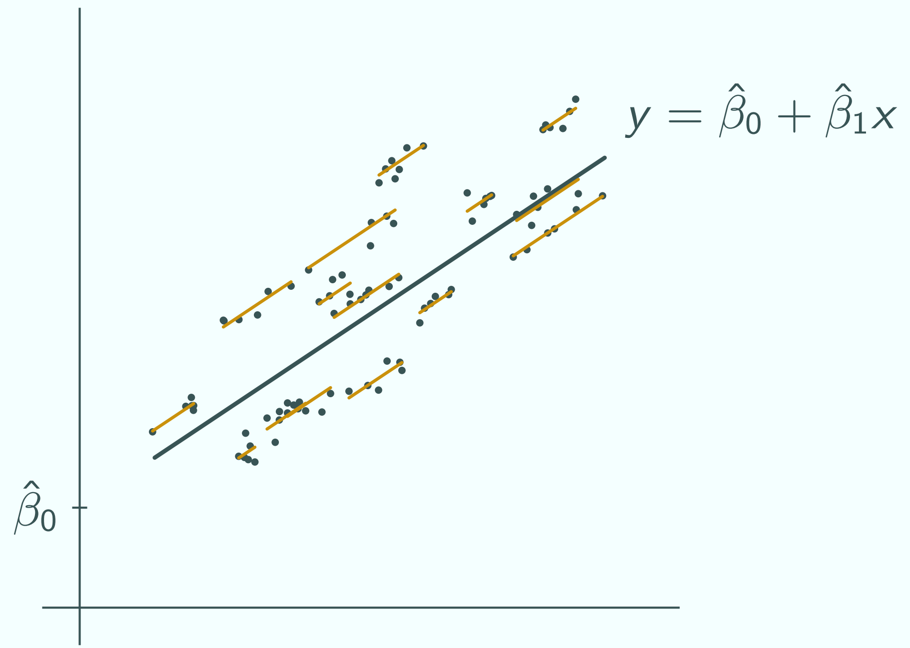
Visualisng the random intercepts model



Complete model

- We can combine the predictions from the fixed and random part in one graph to get a complete visualisation of the model
- and we can add in the actual data points for comparison

Visualisng the random intercepts model



Complete model

- We can combine the predictions from the fixed and random part in one graph to get a complete visualisation of the model
- and we can add in the actual data points for comparison
- Usually we only plot predictions for the range of values we have in our dataset

Exercises: Session 2

Bibliography

- Buckley, Y., Briese, D. and Rees, M. (2008) *Demography and management of the invasive plant species Hypericum perforatum. I. Using multi-level mixed-effects models for characterizing growth, survival and fecundity in a long-term data set* Journal of Applied Ecology, **40** pp 481 – 493
- Goldstein, H., Burgess, S. and McConnell, B. (2007) *Modelling the effect of pupil mobility on school differences in educational achievement* Journal of the Royal Statistical Society Series A, **170** 4 pp 941 – 954
- Judge, T., Scott, B. and Ilies, R. (2006) *Hostility, job attitudes and workplace deviance: Test of a multilevel model* Journal of Applied Psychology **91** 1 pp 126 – 138
- Muijs, D. (2003) *The effectiveness of learning support assistants in improvind the mathematices achievement of low achieving pupils in primary school* Educational Research **45**:3, pp 219 – 230
- Southwell, B. (2005) *Between messages and people: A multilevel model of memory for television content* Communication Research **32** 1 112 – 140