A brush-up on residuals

Going back to a single level model...

$$y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i$$

- The residual for each observation is the difference between the value of y predicted by the equation and the actual value of y

A brush-up on residuals

Going back to a single level model...

$$y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i$$

- The residual for each observation is the difference between the value of y predicted by the equation and the actual value of y
- In symbols, $y_i - \hat{y}_i$

A brush-up on residuals

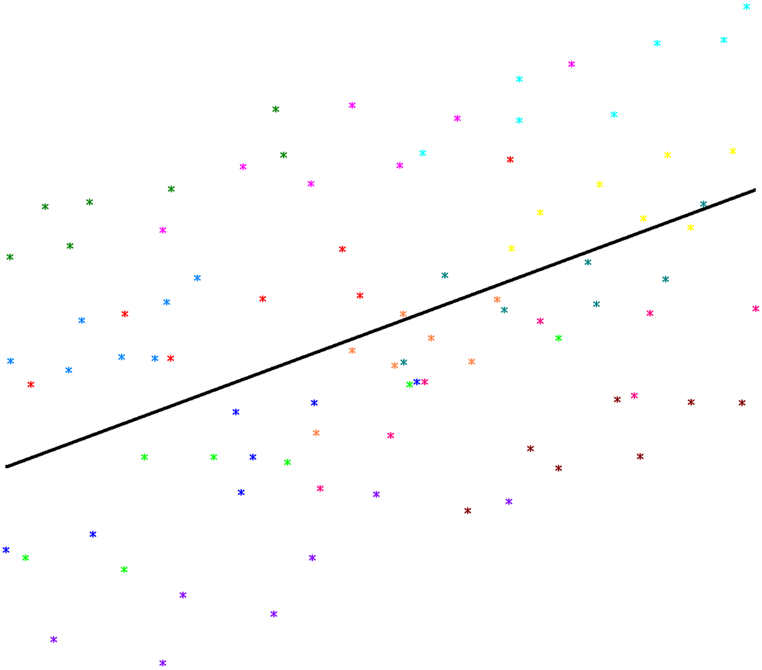
Going back to a single level model...

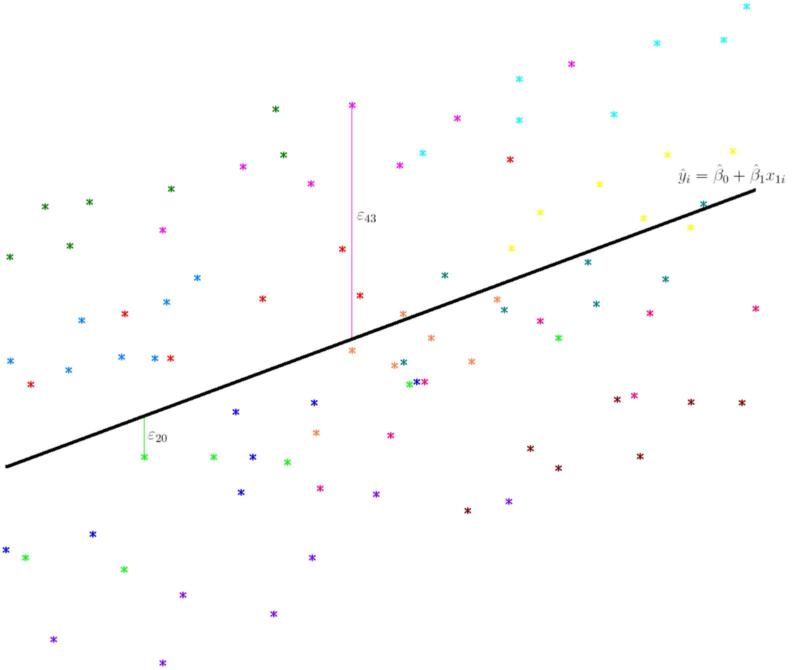
$$y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i$$

- The residual for each observation is the difference between the value of y predicted by the equation and the actual value of y
- In symbols, $y_i - \hat{y}_i$
- So we can calculate the residuals by:

$$r_i = y_i - \hat{y}_i = y_i - (\hat{\beta}_0 - \hat{\beta}_1 x_{1i})$$

- Graphically, we can think of the residual as the distance between the data point and the regression line




$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i}$
ε_{43}
ε_{20}

Multilevel residuals

Back to multilevel modelling...

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + u_j + e_{ij}$$

- Multilevel residuals follow the same basic idea, but now that we have two error terms, we have two residuals

Multilevel residuals

Back to multilevel modelling...

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + u_j + e_{ij}$$

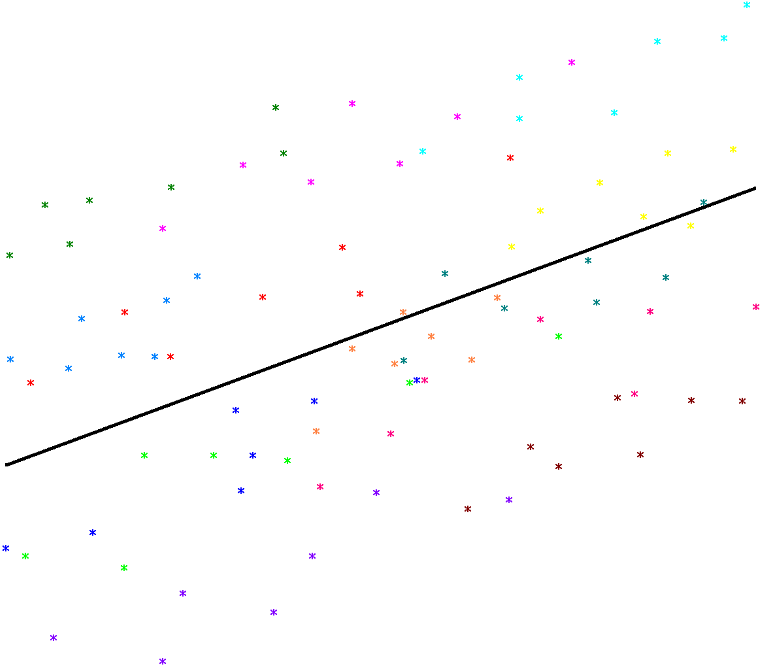
- Multilevel residuals follow the same basic idea, but now that we have two error terms, we have two residuals:
  - the distance from the overall regression line to the line for the group (the level 2 residual)

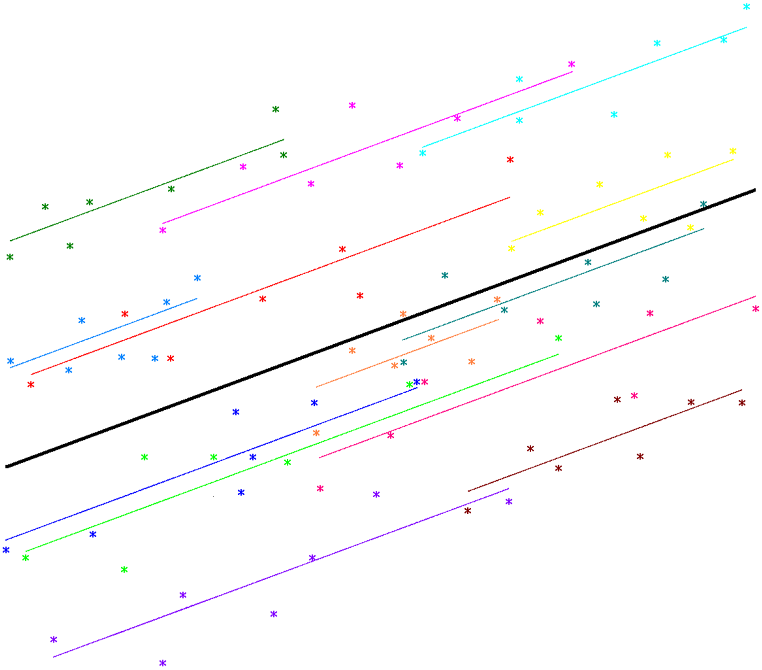
Multilevel residuals

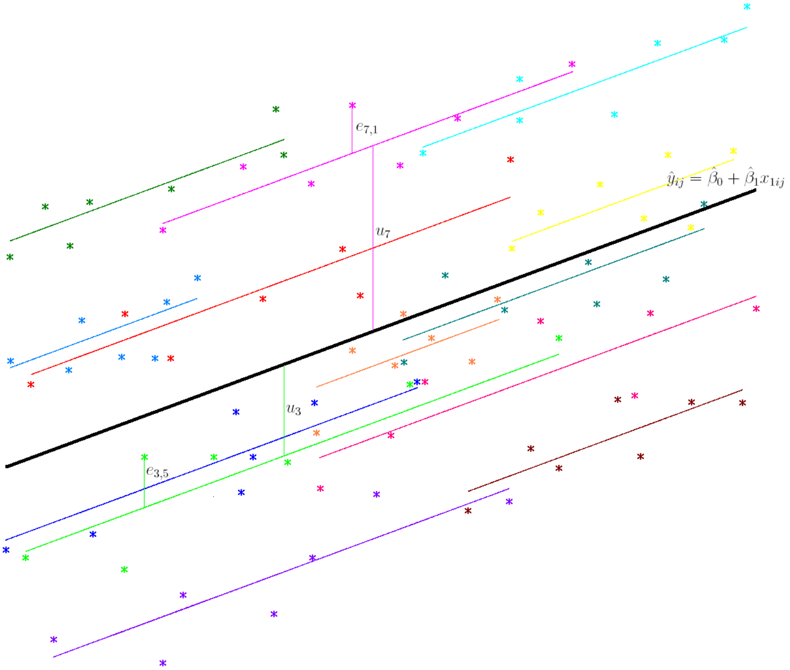
Back to multilevel modelling...

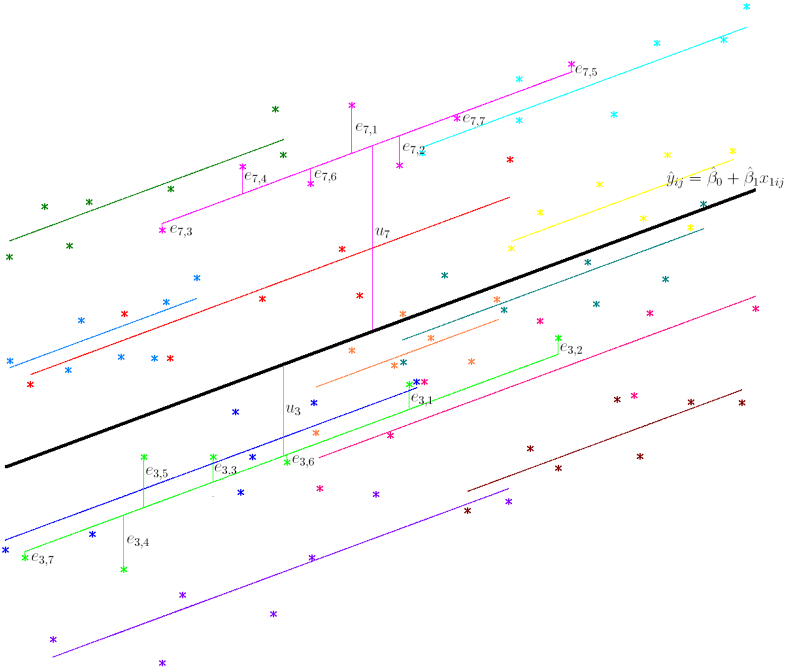
$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + u_j + e_{ij}$$

- Multilevel residuals follow the same basic idea, but now that we have two error terms, we have two residuals:
  - the distance from the overall regression line to the line for the group (the level 2 residual)
  - and the distance from the line for the group to the data point (the level 1 residual)






$e_{7,1}$
u_7
$\hat{y}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 x_{1ij}$
u_3
$e_{3,5}$



$\hat{y}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 x_{1ij}$
$e_{7,5}$
$e_{7,1}$
$e_{7,7}$
$e_{7,2}$
$e_{7,4}$
$e_{7,6}$
$e_{7,3}$
u_7
$e_{3,2}$
$e_{3,1}$
u_3
$e_{3,6}$
$e_{3,5}$
$e_{3,3}$
$e_{3,4}$
$e_{3,7}$

Calculation of multilevel residuals

Raw residuals

- First, we must calculate the raw residual r_j

Calculation of multilevel residuals

Raw residuals

- First, we must calculate the raw residual r_j
- Define $r_{ij} = y_{ij} - \hat{y}_{ij}$
  - This is just as if we were treating it as a single-level model

Calculation of multilevel residuals

Raw residuals

- First, we must calculate the raw residual r_j
- Define $r_{ij} = y_{ij} - \hat{y}_{ij}$
  - This is just as if we were treating it as a single-level model
- Now r_j is the mean of r_{ij} for group j

Why can't we just use these raw residuals?

- Suppose we have just two data points for one of our groups

Why can't we just use these raw residuals?

- Suppose we have just two data points for one of our groups
- These two data points might be unusual for their group

Why can't we just use these raw residuals?

- Suppose we have just two data points for one of our groups
- These two data points might be unusual for their group
- Then the r_j calculated using them will be 'wrong'

Why can't we just use these raw residuals?

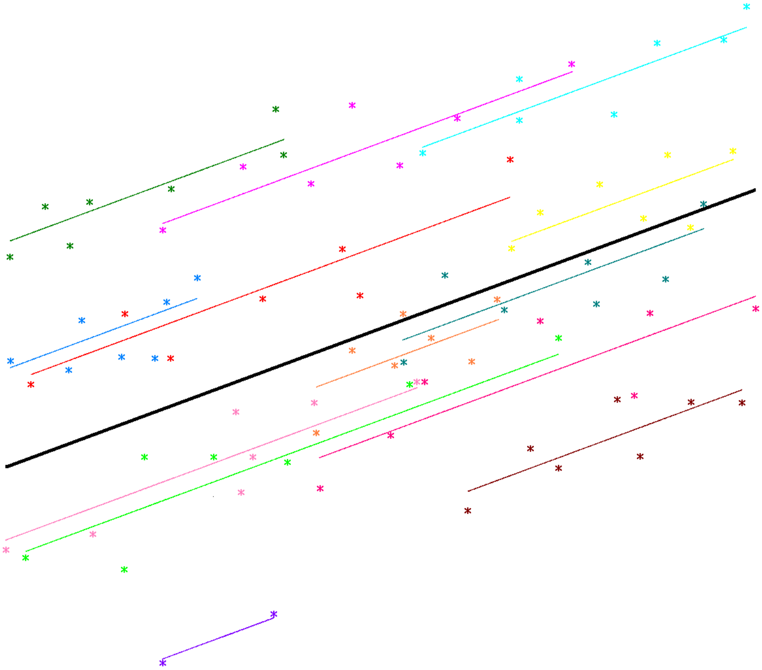
- Suppose we have just two data points for one of our groups
- These two data points might be unusual for their group
- Then the r_j calculated using them will be 'wrong'
- But our points are always a sample and we can't know if they're typical of their group

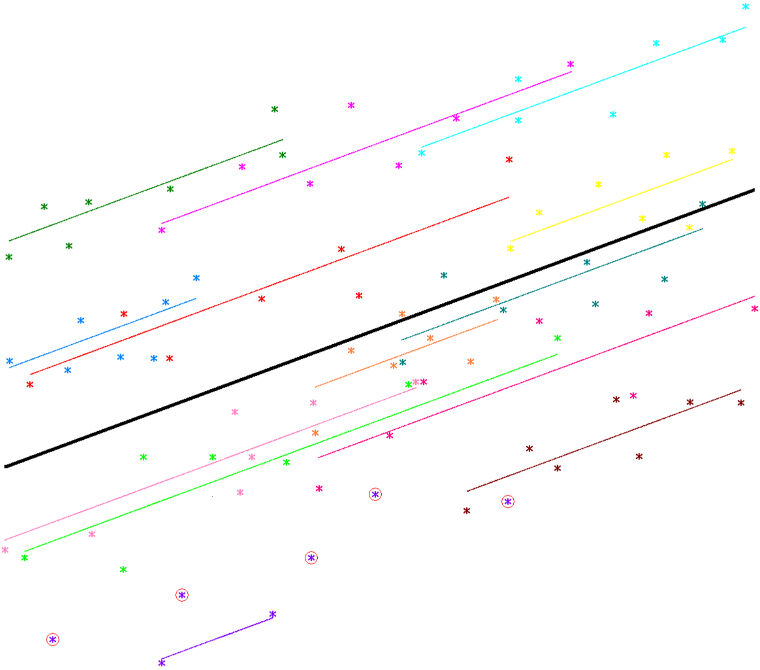
Why can't we just use these raw residuals?

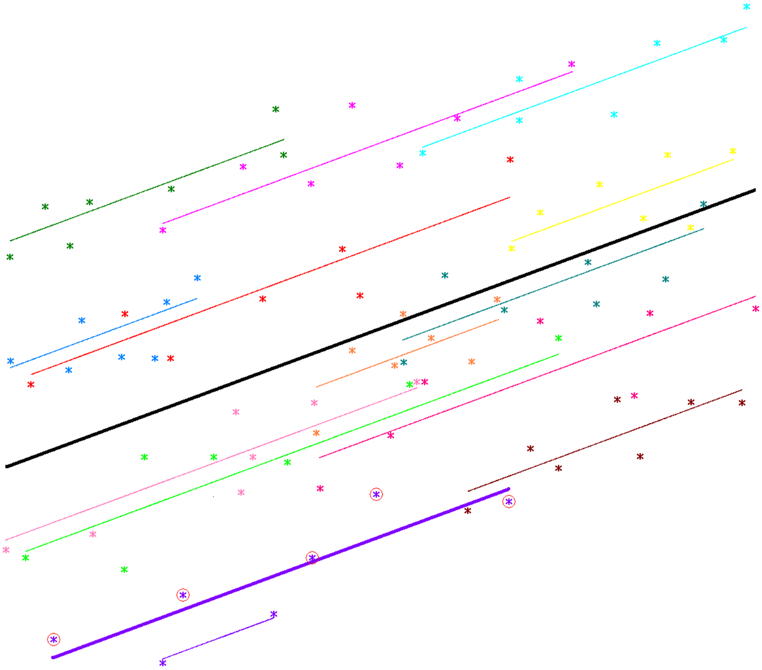
- Suppose we have just two data points for one of our groups
- These two data points might be unusual for their group
- Then the r_j calculated using them will be 'wrong'
- But our points are always a sample and we can't know if they're typical of their group
- We can combine the data from the group with information from the other groups to bring the residuals towards the overall average

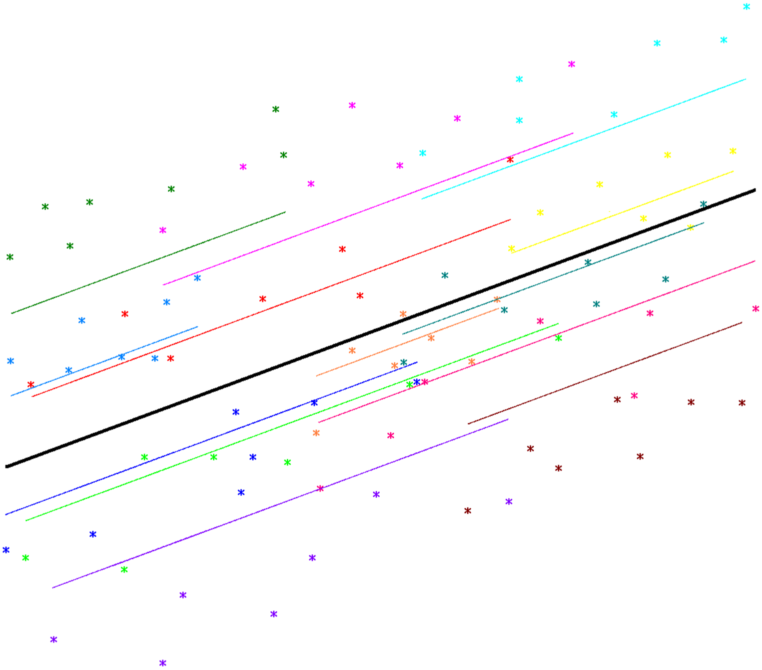
Why can't we just use these raw residuals?

- Suppose we have just two data points for one of our groups
- These two data points might be unusual for their group
- Then the r_j calculated using them will be 'wrong'
- But our points are always a sample and we can't know if they're typical of their group
- We can combine the data from the group with information from the other groups to bring the residuals towards the overall average
- Then the level 2 residuals will be less sensitive to outlying elements of the group









Calculation of multilevel residuals

Raw residuals

- $r_{ij} = y_{ij} - \hat{y}_{ij}$
- r_j is the mean of r_{ij} for group j

Shrinkage factor

- $\frac{\sigma^2_{u0}}{\sigma^2_{u0} + \frac{\sigma^2_{e0}}{n_j}}$

Level 2 residual

- The estimated level 2 residual is the shrinkage factor times the raw residual

$$\hat{u}_{0j} = \frac{\sigma^2_{u0}}{\sigma^2_{u0} + \frac{\sigma^2_{e0}}{n_j}} \cdot r_j$$

Level 1 residual

- The level 1 residual is the observed value, minus the predicted value from the overall regression line, minus the level 2 residual

$$\hat{e}_{0ij} = y_{ij} - (\hat{\beta}_0 + \hat{\beta}_1 x_{1ij}) - \hat{u}_{0j}$$

How much shrinkage occurs?

A lot of shrinkage	Not much shrinkage

How much shrinkage occurs?

	A lot of shrinkage	Not much shrinkage
n_j		

How much shrinkage occurs?

	A lot of shrinkage	Not much shrinkage
n_j	When there are not many level 1 units in the group	When there are a lot of level 1 units in the group

How much shrinkage occurs?

	A lot of shrinkage	Not much shrinkage
n_j	When there are not many level 1 units in the group	When there are a lot of level 1 units in the group
$\sigma^2_{e_0}$		

How much shrinkage occurs?

	A lot of shrinkage	Not much shrinkage
n_j	When there are not many level 1 units in the group	When there are a lot of level 1 units in the group
$\sigma^2_{e_0}$	When the level 1 variance is big	When the level 1 variance is small

How much shrinkage occurs?

	A lot of shrinkage	Not much shrinkage
n_j	When there are not many level 1 units in the group	When there are a lot of level 1 units in the group
$\sigma^2_{e_0}$	When the level 1 variance is big	When the level 1 variance is small
$\sigma^2_{u_0}$		

How much shrinkage occurs?

	A lot of shrinkage	Not much shrinkage
n_j	When there are not many level 1 units in the group	When there are a lot of level 1 units in the group
$\sigma^2_{e_0}$	When the level 1 variance is big	When the level 1 variance is small
$\sigma^2_{u_0}$	When the level 2 variance is small	When the level 2 variance is big

Measuring dependency

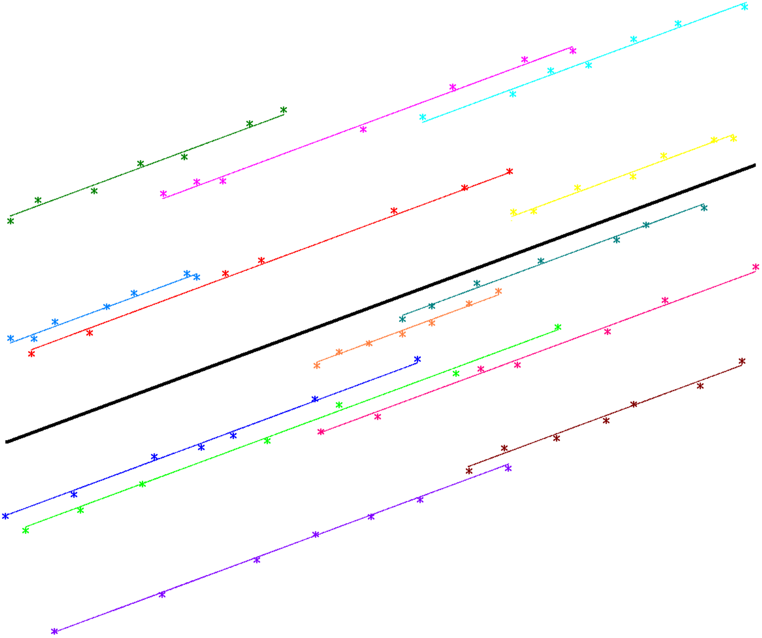
- The question of the relative sizes of $\sigma^2_{u_0}$ and $\sigma^2_{e_0}$ is actually quite important.
- The relative sizes change according to how much dependency we have in our data.
- The dependency arises because observations from the same group are likely to be more similar than those from different groups.
- The fact that we have dependent data is the whole reason that we are using multilevel modelling.
  - We use multilevel modelling partly in order to correctly estimate standard errors
  - If we use a single-level model for dependent data, standard errors will be overestimated
- So we would like to know how much dependency we have in our data

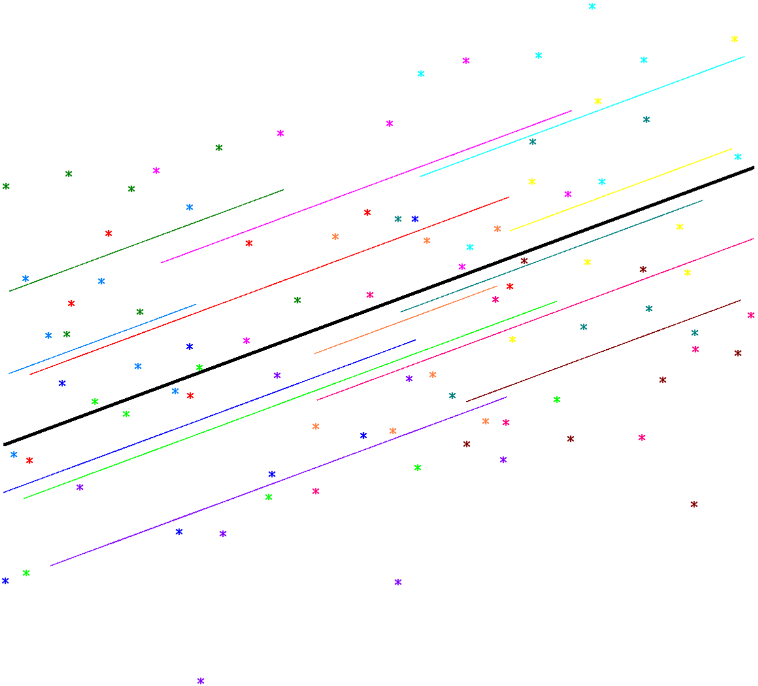
Example

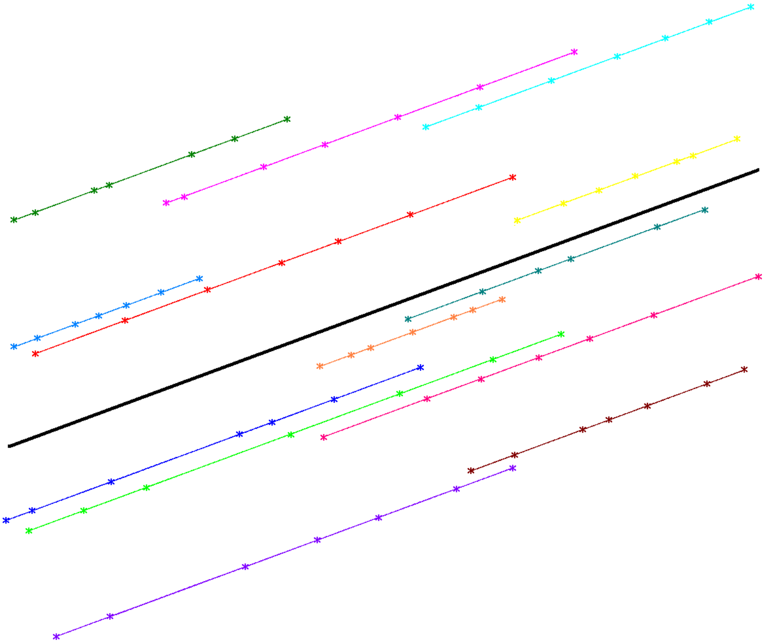
Parameter	Single level	Multilevel
Intercept	-0.098(0.021)	-0.101(0.070)
Boy school	0.122(0.049)	0.120(0.149)
Girl school	0.244(0.034)	0.258(0.117)
Between school variance (σ_u^2)	. (.)	0.155(0.030)
Between student variance (σ_e^2)	0.985(0.022)	0.848(0.019)

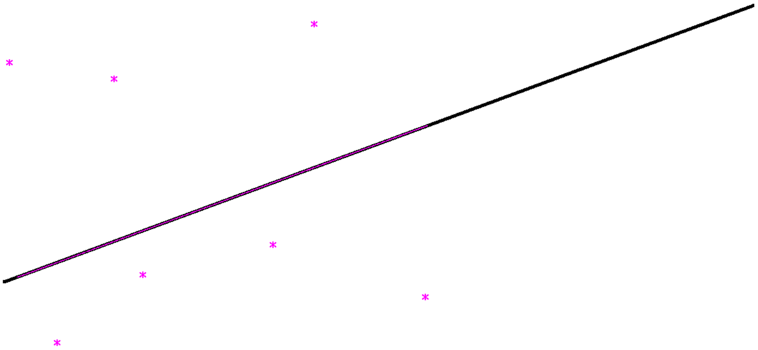
Measuring dependency

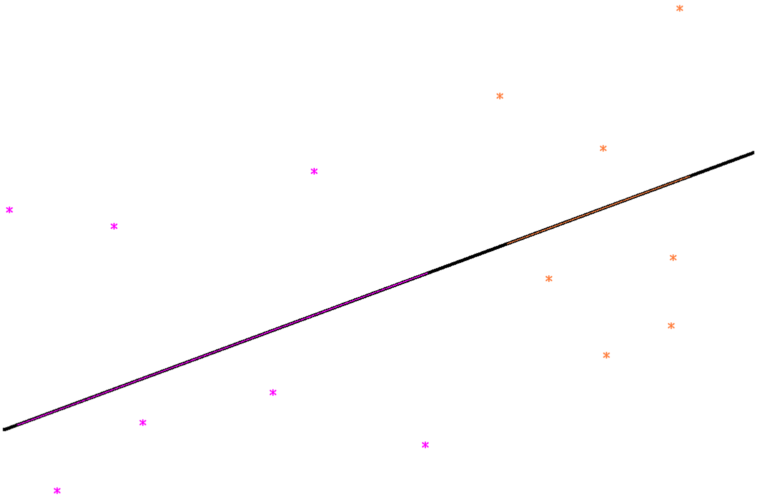
- We can use the variance partitioning coefficient to measure dependency
  - also called VPC or ρ or intraclass correlation
  - For the two-level random intercepts case, $\rho = \frac{\sigma_{u_0}^2}{\sigma_{u_0}^2 + \sigma_{e_0}^2}$
  - Beware! Note how the VPC is similar to, but not identical to, the shrinkage factor
- The VPC is the proportion of the total variance that is at level 2
- How can we interpret it?

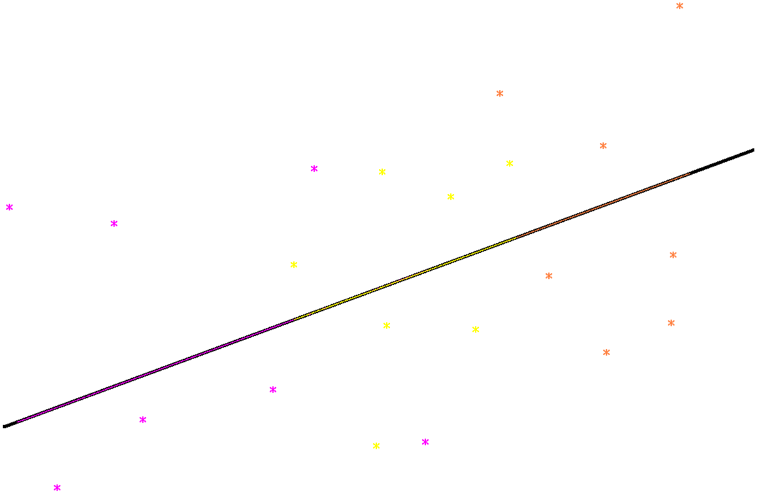


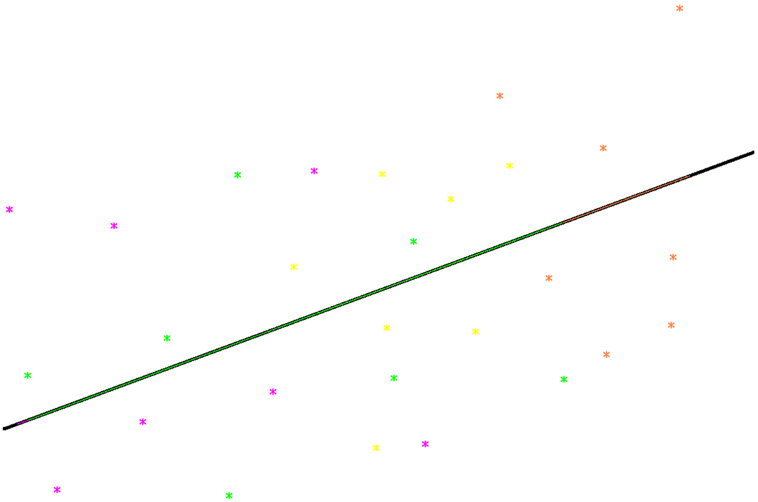


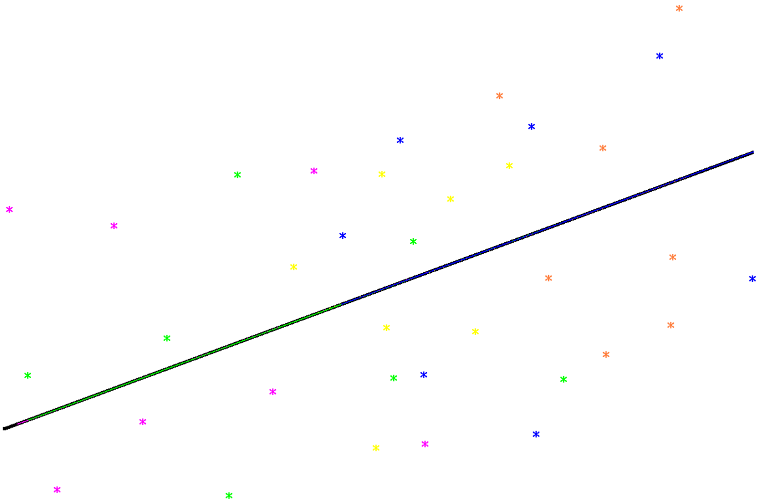


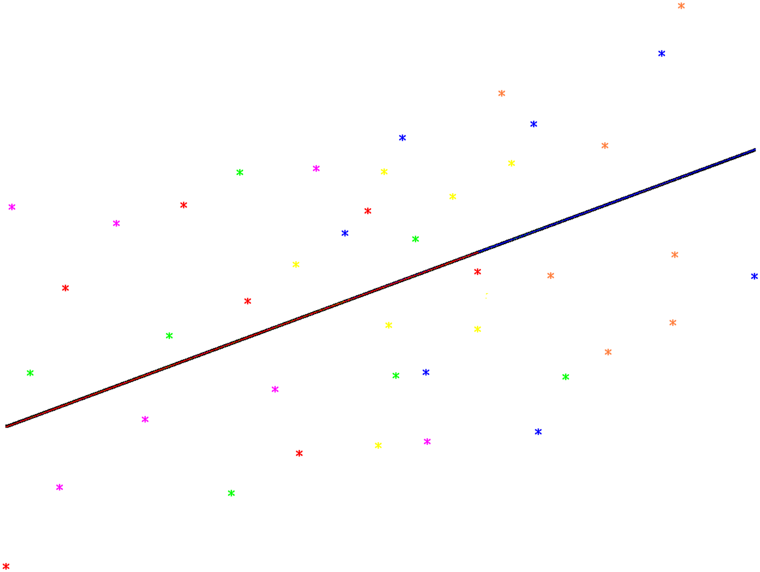












Correlation structure of single level model

$$y_i = \beta_0 + \beta_1 x_{1i} + e_{0i}$$

S		1	1	1	2	2	2	2	3	3	3
	P	1	2	3	1	2	3	4	1	2	3
1	1	1	0	0	0	0	0	0	0	0	0
1	2	0	1	0	0	0	0	0	0	0	0
1	3	0	0	1	0	0	0	0	0	0	0
2	1	0	0	0	1	0	0	0	0	0	0
2	2	0	0	0	0	1	0	0	0	0	0
2	3	0	0	0	0	0	1	0	0	0	0
2	4	0	0	0	0	0	0	1	0	0	0
3	1	0	0	0	0	0	0	0	1	0	0
3	2	0	0	0	0	0	0	0	0	1	0
3	3	0	0	0	0	0	0	0	0	0	1

Correlation structure of two level random intercept model

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

S		1	1	1	2	2	2	2	3	3	3
	P	1	2	3	1	2	3	4	1	2	3
1	1	1	ρ	ρ	0	0	0	0	0	0	0
1	2	ρ	1	ρ	0	0	0	0	0	0	0
1	3	ρ	ρ	1	0	0	0	0	0	0	0
2	1	0	0	0	1	ρ	ρ	ρ	0	0	0
2	2	0	0	0	ρ	1	ρ	ρ	0	0	0
2	3	0	0	0	ρ	ρ	1	ρ	0	0	0
2	4	0	0	0	ρ	ρ	ρ	1	0	0	0
3	1	0	0	0	0	0	0	0	1	ρ	ρ
3	2	0	0	0	0	0	0	0	ρ	1	ρ
3	3	0	0	0	0	0	0	0	ρ	ρ	1