FACULTEIT DER SOCIALE WETENSCHAPPEN

Behavioural Science Institute

Workshop Multilevel Analysis

Contextual effects and cross-level interaction

Fred Hasselman

f.hasselman@bsi.ru.nl

Radboud Universiteit Nijmegen

What we learned so far...

- Take the hierarchical structure of the data into account by allowing individuals within groups to be clustered around their group mean. The group means are random deviations from the grand mean of the sample.

- If we don't take clustering into account the SE will be underestimated.

- Comparison of nested models

Comparing nested models: Random intercepts only / no covariate

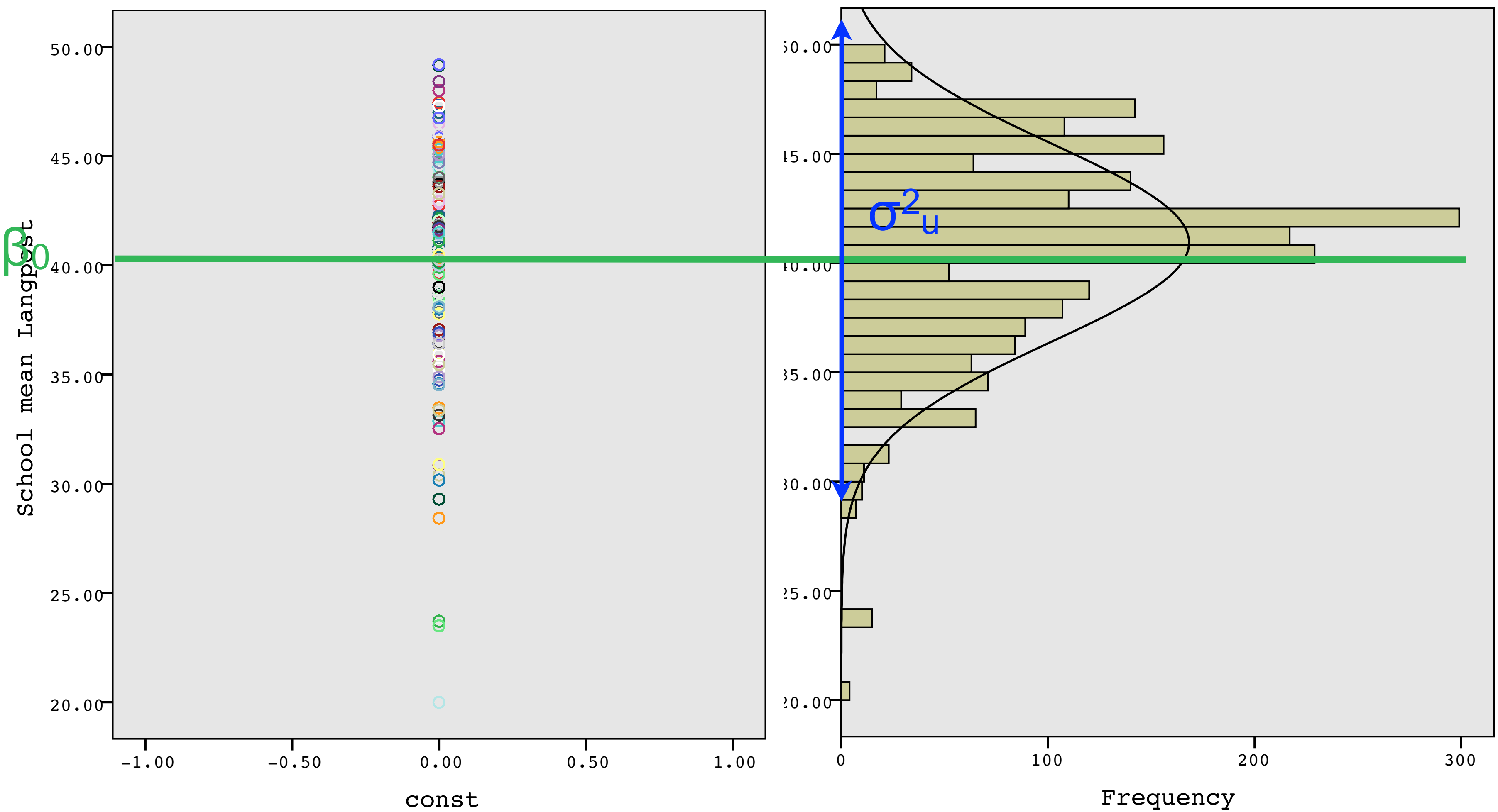
- ‘Empty’ model: Analyze variance of a dependent variable into two variance components. Level 1: Individuals, Level 2: Groups

- Compare to ANOVA: Groups as a random factor

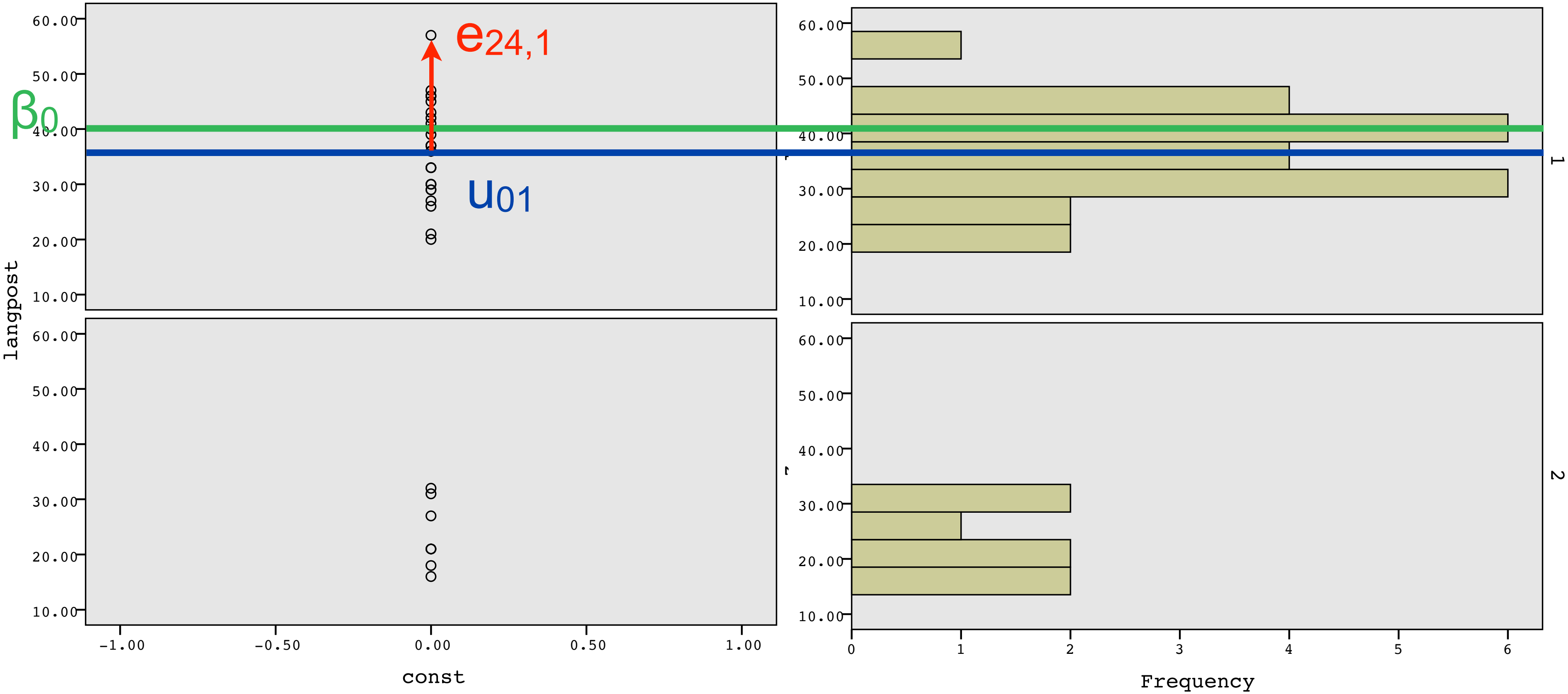
- Three parameters:

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

The ‘empty’ model: Between groups



The ‘empty’ model: Within groups (school 1 & 2)



The ‘empty’ model: Random intercepts, no predictor / Variance components

Equations

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

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

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

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

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

β_0: Mean of dependent variable

σ^2_{u0}: Between-group differences

σ^2_{e0}: Within-group differences

Random intercept model, 1 covariate, fixed slope

- Random intercepts plus a covariate with a fixed slope
- Compare to ANCOVA: Groups still a random factor
- Four parameters:

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

Random intercept model, 1 covariate, fixed slope

Equations

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

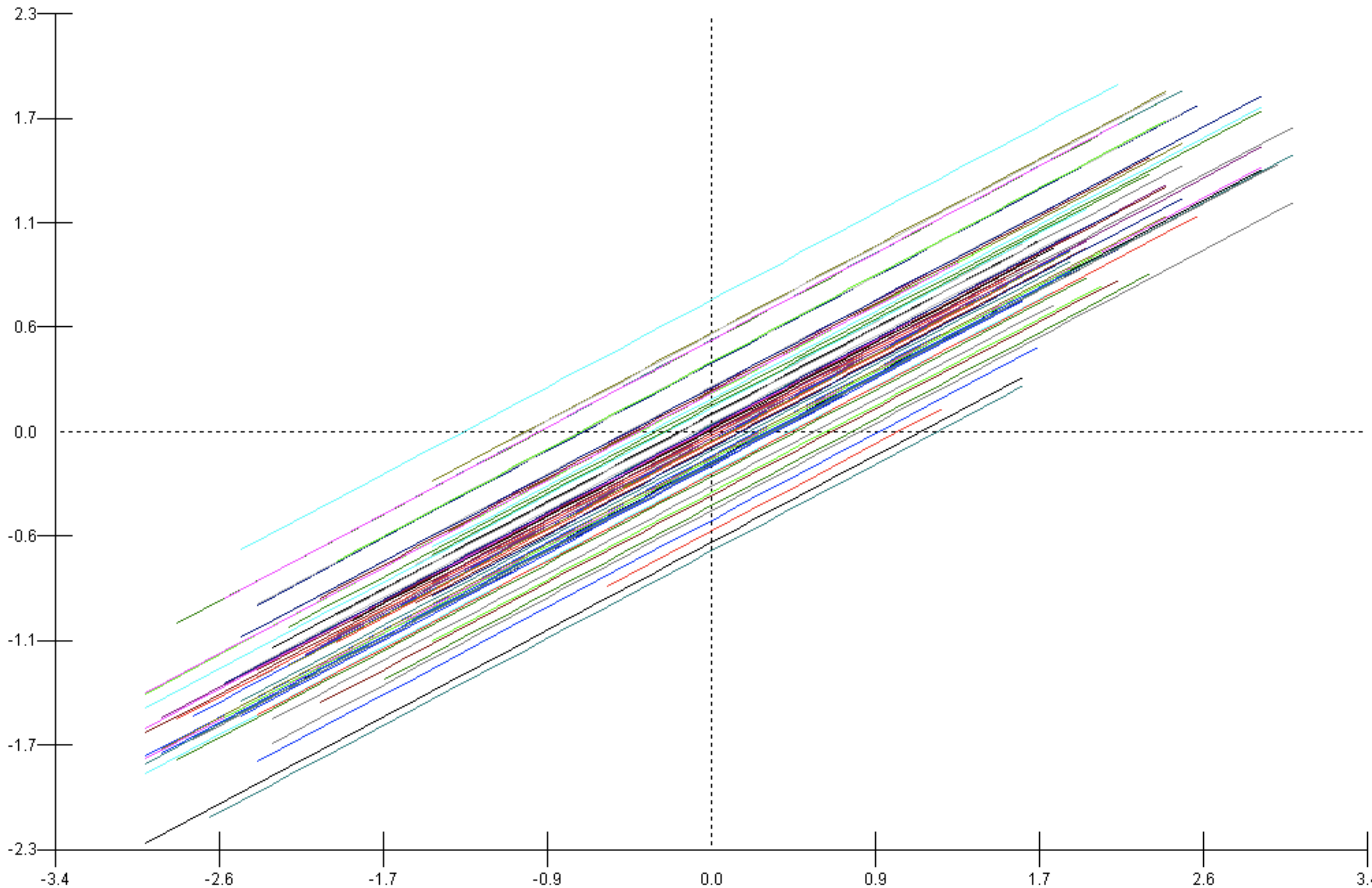
$$y_{ij} = \beta_{0ij} x_{0} + \beta_1 x_{1ij}$$

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

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

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

β_1: Fixed slope


2.3
1.7
1.1
0.6
0.0
-0.6
-1.1
-1.7
-2.3
-3.4
-2.6
-1.7
-0.9
0.0
0.9
1.7
2.6
3.4

Random intercept, 1 covariate, random slope

- Random intercepts plus a covariate with a random slope

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

- Six parameters:

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

Random intercept, 1 covariate, random slope

Equations

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

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

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

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

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

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

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

σ^2_{u1}: Variance of slopes

σ_{u01}: Intercept-slope covariance

Random intercept, 1 covariate, random slope

Equations

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

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

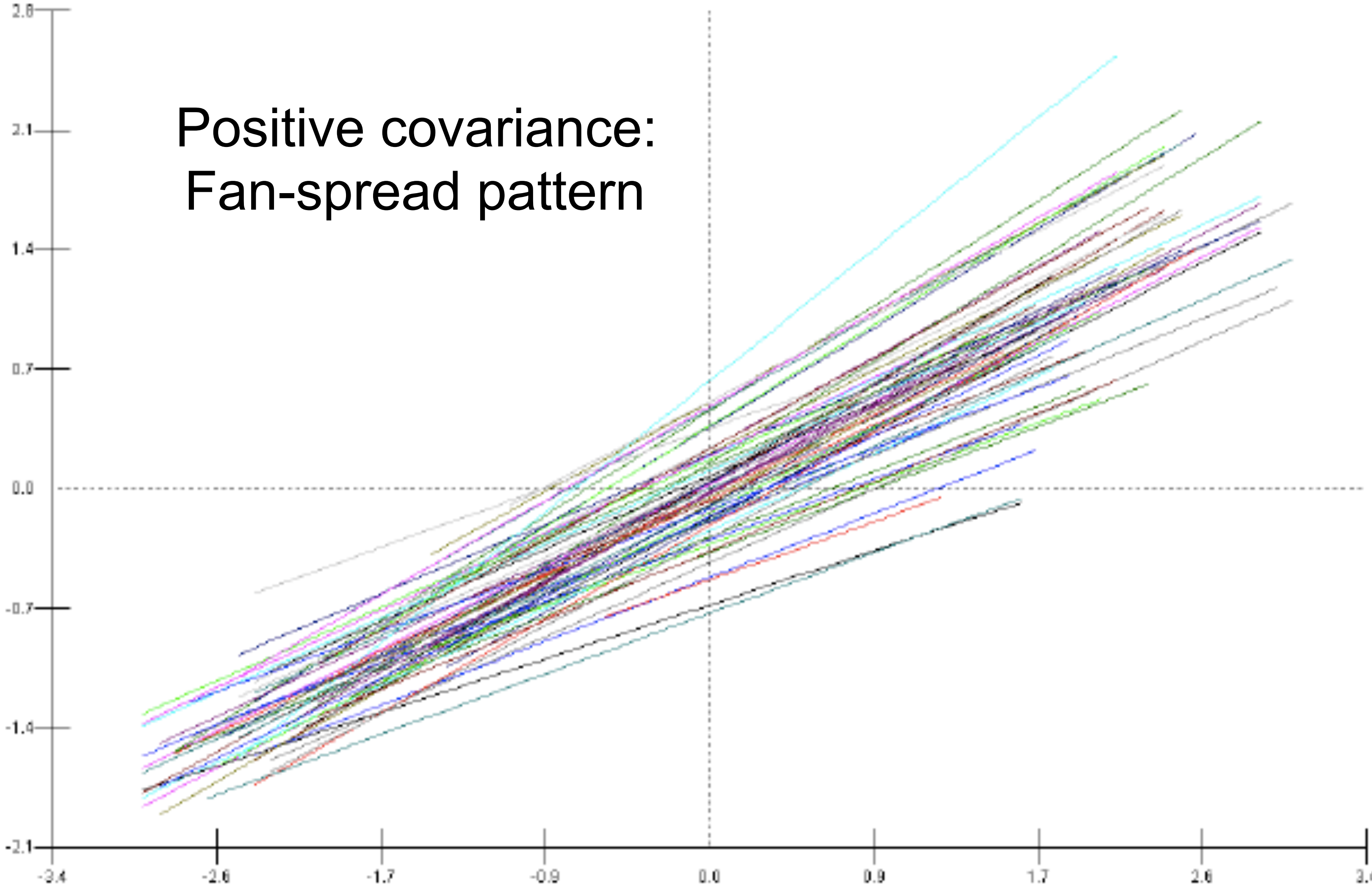
$$\beta_{0ij} = -0.012(0.040) + u_{0j} + e_{0ij}$$

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

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

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

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

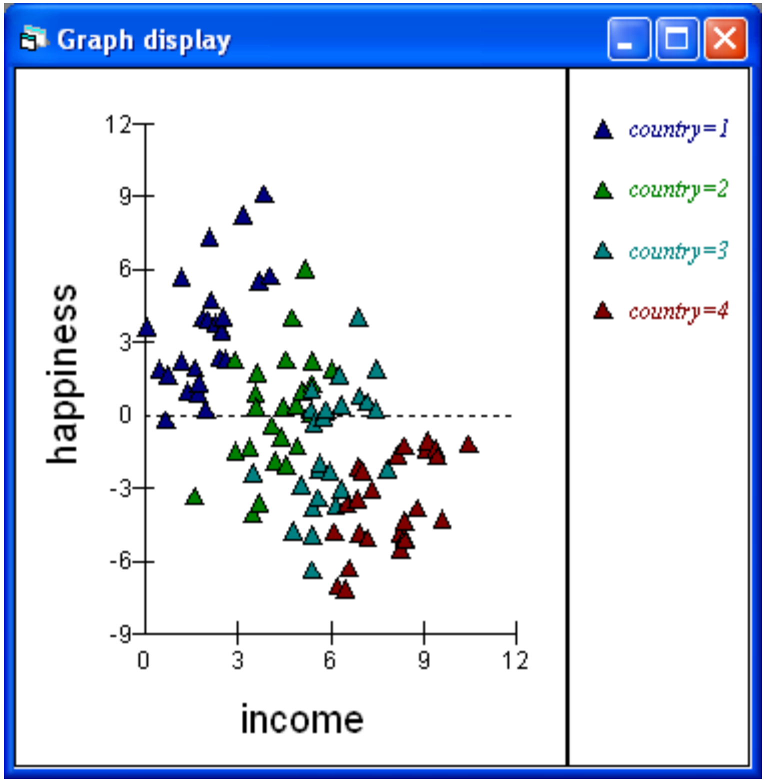

Positive covariance:
Fan-spread pattern
2.8
2.1
1.4
0.7
0.0
-0.7
-1.4
-2.1
-3.4
-2.6
-1.7
-0.9
0.0
0.9
1.7
2.6
3.4

Within group, between group and contextual effects of for level 1 predictors

Often contextual variables are constructed by aggregating level 1 predictors, country level income, family level aggression etc

We will simulate a small data set with 100 individuals from four groups, for the sake of argument lets say that our response is happiness, we have a predictor variable that is income and the groups are 4 different countries. Lets simulate the mean income in each country as 2,4,6,8. We then simulate our response as

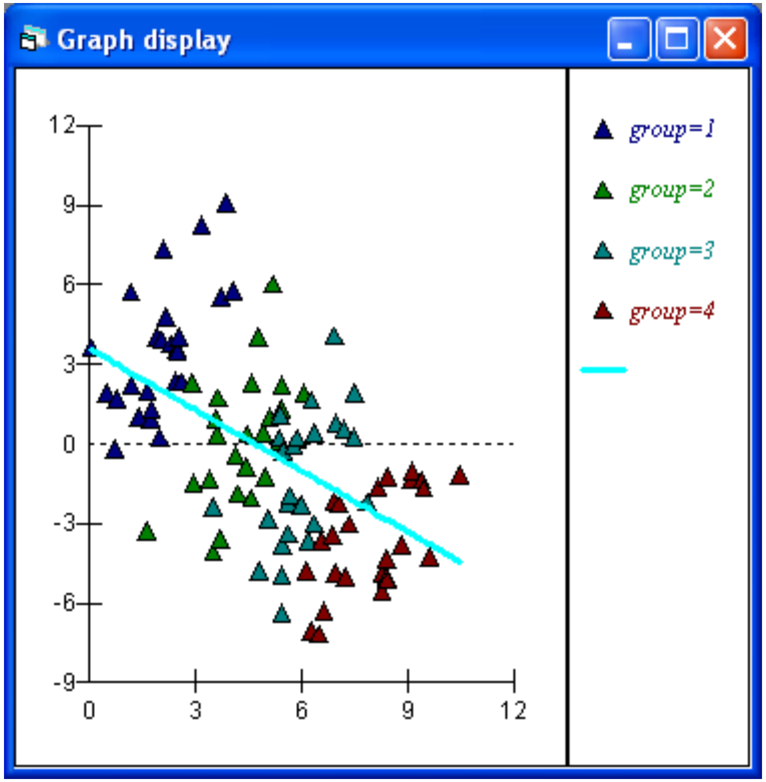
$$y_{ij} = 5 + 1 \times \text{income}_{ij} + (-2) \times \text{av\_income}_j + u_j + e_{ij}$$
$$e_{ij} \sim N(0,2)$$



Note that this model does not include a country level random effect. Therefore the only differences between country happiness levels are produced by differing incomes in the countries.

Fitting a regression model ignoring average income

$$happiness_i = \beta_0 + \beta_1 income_i + e_i$$
$$e_i \sim N(0, \sigma_e^2)$$
$$\hat{\beta}_0 = 3.577(0.680)$$
$$\hat{\beta}_1 = -0.770(0.122)$$



And we might conclude as income rises happiness reduces

Including the average income variable

$$\text{happiness}_{ij} = \beta_0 + \beta_1 \text{income}_{ij} + \beta_2 \text{av\_income}_j + e_{ij}$$

$$e_{ij} \sim N(0, \sigma_e^2)$$

$$\hat{\beta}_0 = 5.772(0.492)$$

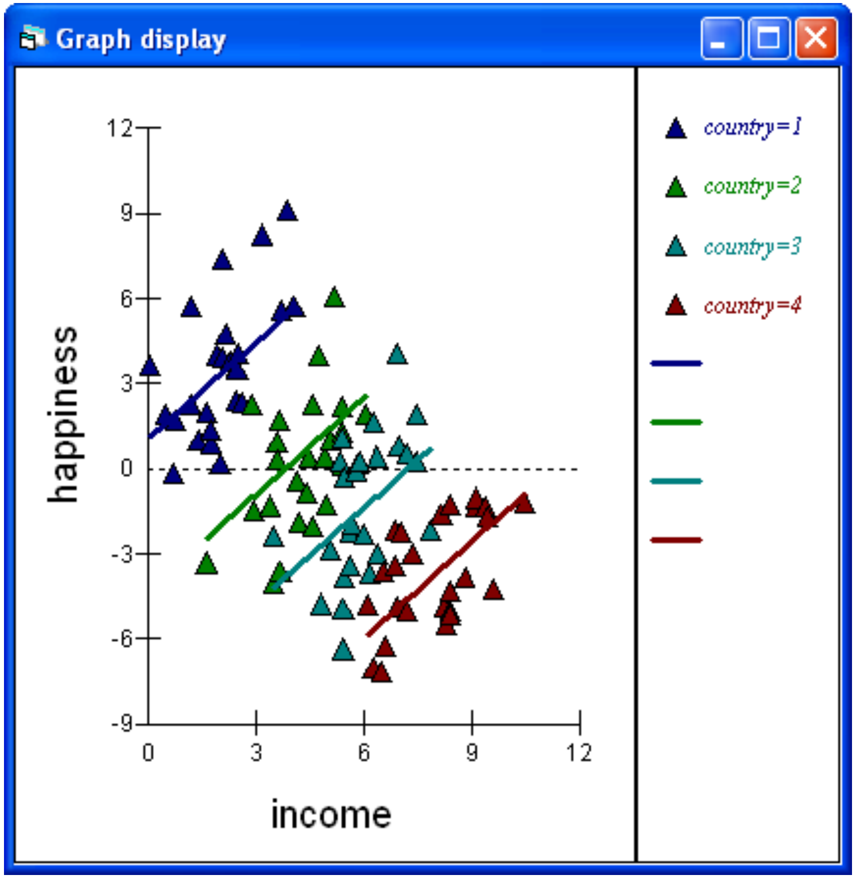
$$\hat{\beta}_1 = 1.128(0.122)$$

$$\hat{\beta}_2 = -2.332(0.207)$$

Now we have a different picture. Within all countries the relationship between income and happiness is positive. This is the ***within*** country slope(1.128)

However the intercept for country j is 5.772 - 2.32 $\times$av_income$_j$

-2.32 is the ***contextual*** effect.



Recall the country average incomes are {2,4,6,8}, which gives intercepts of {1.13,-3.51, -8.15, -12.8}

Now within all countries we have a positive relationship between income and happiness, however people in richer countries tend to be less happy than poorer ones.

What about the between group effect?

We have identified the within group and contextual effects of the income predictor.

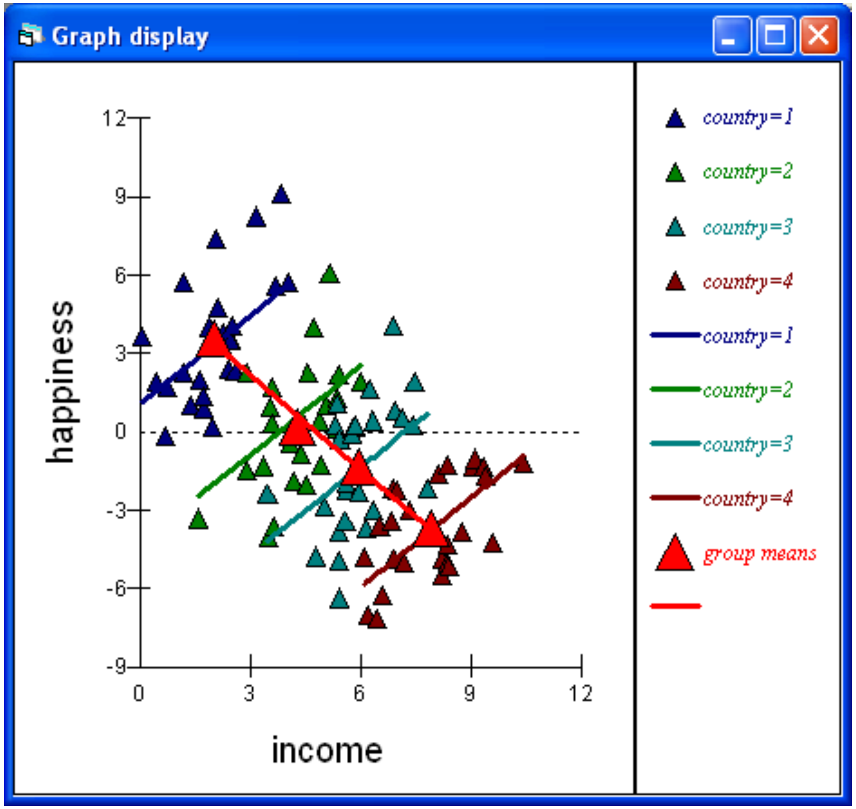
The between group effect is the slope of the regression of the country mean for happiness on the country mean for income.

That is the slope of the regression line through the four red points and is -1.20

within group = 1.128

contextual = -2.332

between group = -1.204



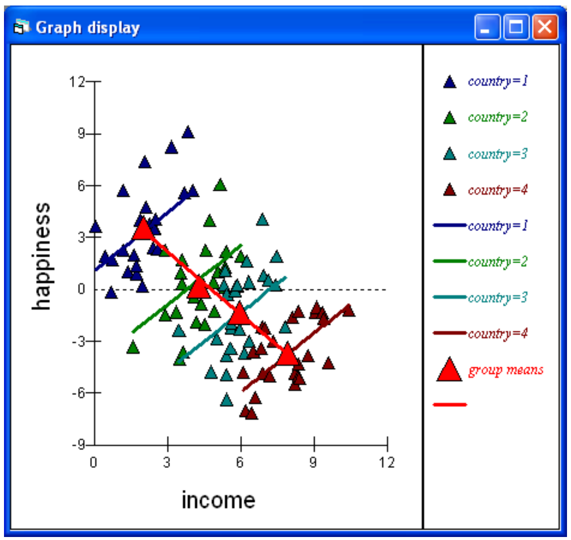
Notice that the between group = the within group + the contextual

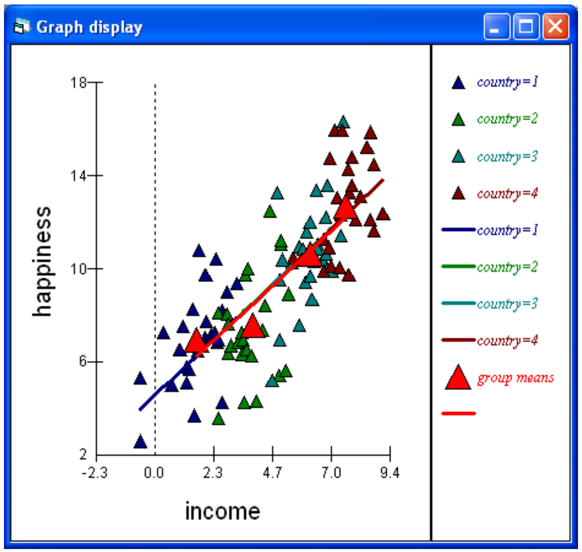
B = W + C

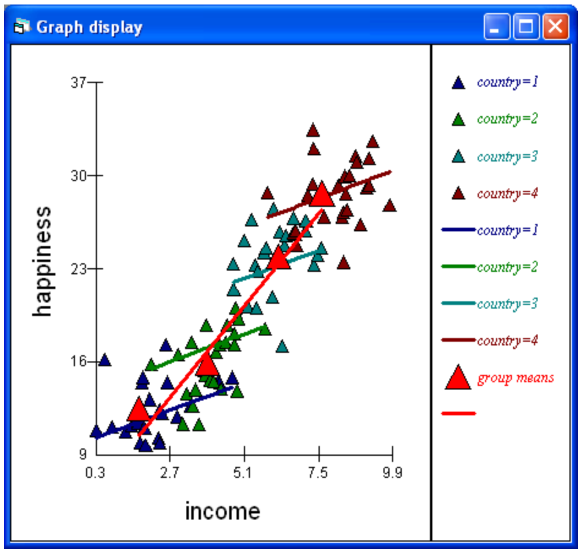
-1.204=1.128+(-1.204)

Or C = B – W that is the contextual effect is the the difference of the between and within regressions

Examples of different contextual effects







W=1,B= -1,C=-2

Contextual effect pulls down intercept as av_income increases

W = 1, B=1,C=0

Contextual effect 0 so intercept unchanged as av_income increases

W=1,B=3,C=2

Contextual effect pushes up intercept as av_income increases

Interactions in regular regression

- Interactions between predictors in multiple regression models (or moderator effects):

- Categorical and continuous variable
- Two continuous variables

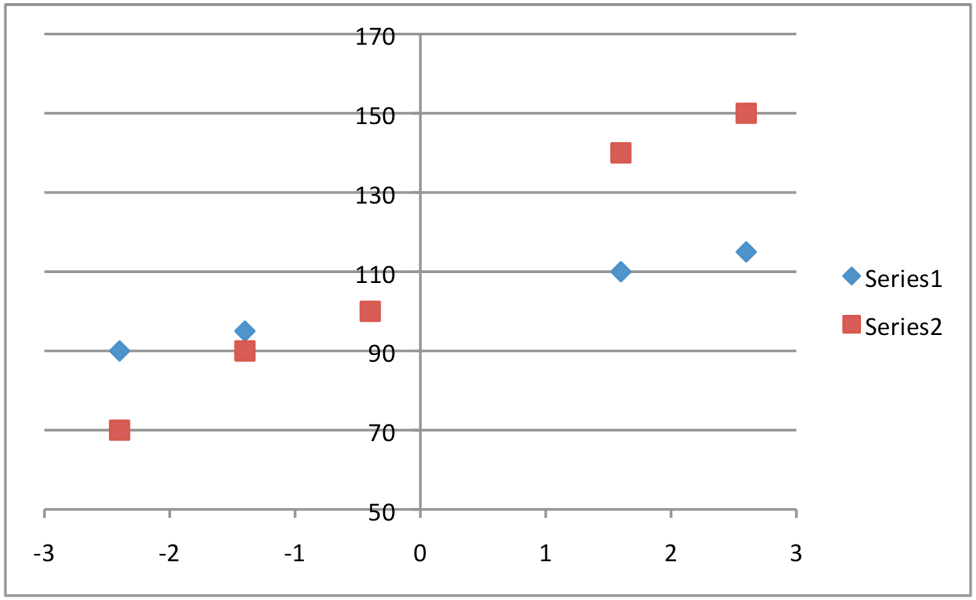
- General form: $Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \beta_3 XZ_i + \varepsilon_i$

- General approach: Start without the interactions, add them in a second model

- Center variables

Categorical / Nominal x Continuous

- Create dummy variable(s) for nominal variable
- Center other predictor(s)
- Create interaction term(s): Series1 x centered predictor



Categorical / Nominal x Continuous

- Model 1: $Y_i = \beta_0 + \beta_1 \text{ xcent}_i + \beta_2 \text{ dummy}_i + \varepsilon_i$

- Model 2: $Y_i = \beta_0 + \beta_1 \text{ xcent}_i + \beta_2 \text{ dummy}_i + \beta_3 \text{ xcent*dummy}_i + \varepsilon_i$

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	102.000	5.697		17.904	.000
	xcent	10.640	2.172	.866	4.899	.002
	dummy	8.000	8.057	.176	.993	.354
2	(Constant)	102.000	1.181		86.349	.000
	xcent	5.000	.637	.407	7.851	.000
	dummy	8.000	1.671	.176	4.789	.003
	interaction	11.279	.901	.649	12.523	.000

a. Dependent Variable: y

Categorical / Nominal x Continuous

- Regression equations, rearrange terms for easier interpretation:
  $\beta_0 + \beta_1$ xcent$_i$ + β_2 dummy$_i$ + β_3 xcent*dummy$_i$ =
  $\beta_0 + \beta_2$ dummy$_i$ + ($\beta_1 + \beta_3$ dummy)*xcent

- Dummy = 0: **Y' = 102 + 5 * xcent**

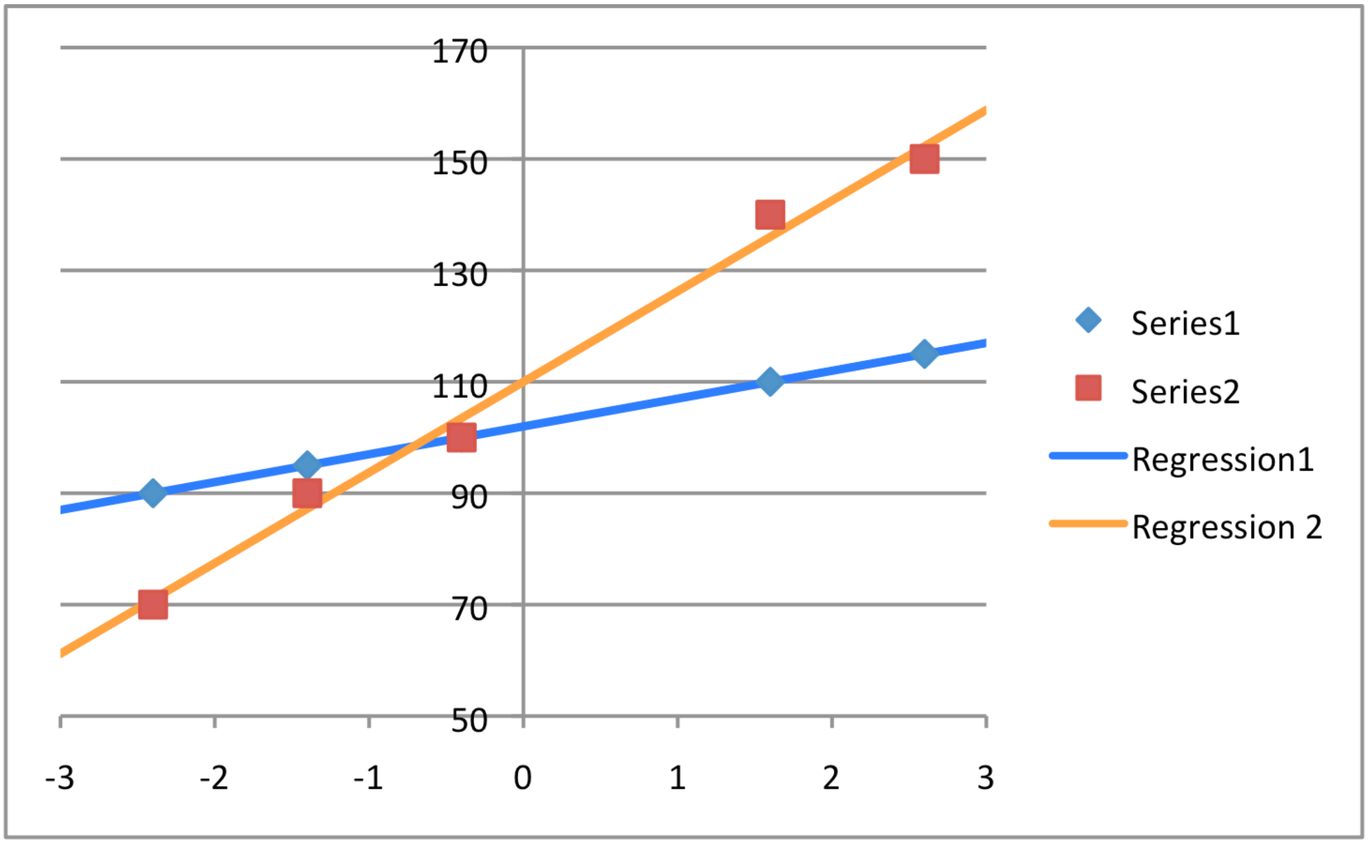
- Dummy = 1: Y' = 102 + 8 + (5 + 11.279) * xcent = **110 + 16.270 * xcent**

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	102.000	5.697		17.904	.000
	xcent	10.640	2.172	.866	4.899	.002
	dummy	8.000	8.057	.176	.993	.354
2	(Constant)	102.000	1.181		86.349	.000
	xcent	5.000	.637	.407	7.851	.000
	dummy	8.000	1.671	.176	4.789	.003
	interaction	11.279	.901	.649	12.523	.000

a. Dependent Variable: y

Categorical / Nominal x Continuous



Recall the definition of a moderator variable:

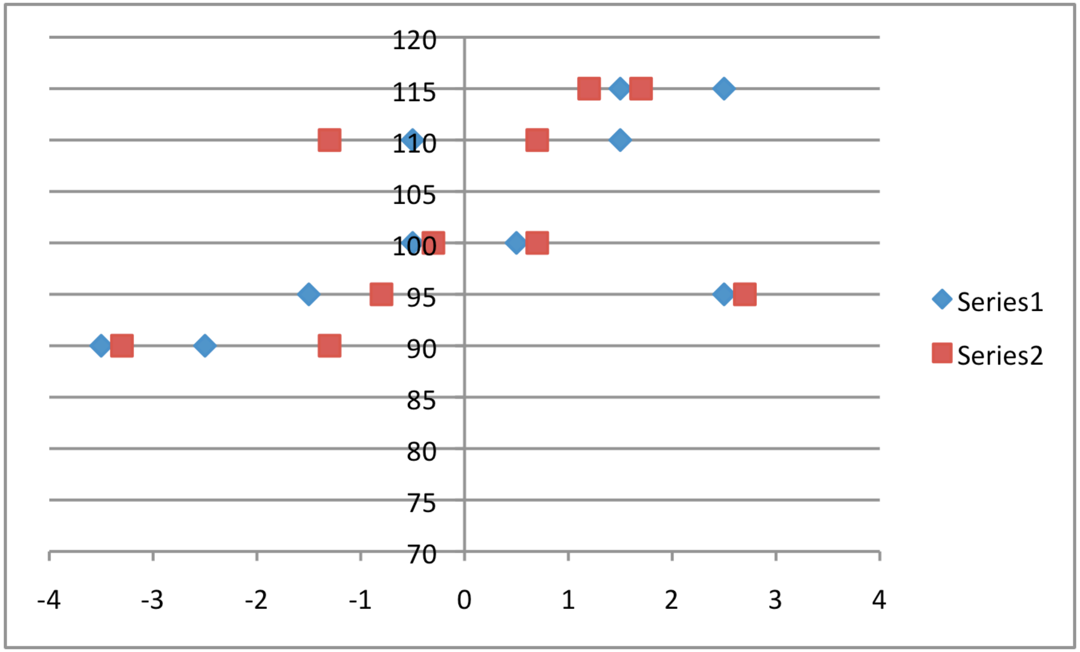
The relationship between X and Y changes for different levels of the moderator:

The slope

Test whether slopes are significant for each group

Continuous x Continuous

- Center predictors
- Create interaction term(s): x1cent * x2cent



Continuous x Continuous

- Model 1: $Y_i = \beta_0 + \beta_1 \text{ x1cent}_i + \beta_2 \text{ x2cent}_i + \varepsilon_i$
- Model 2: $Y_i = \beta_0 + \beta_1 \text{ x1cent}_i + \beta_2 \text{ x2cent}_i + \beta_3 \text{ x1cent*x2cent} + \varepsilon_i$

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	102.000	1.971		51.748	.000
	x1cent	9.014	2.846	1.907	3.167	.016
	x2cent	-7.280	3.362	-1.304	-2.165	.067
2	(Constant)	104.502	2.528		41.345	.000
	x1cent	8.679	2.661	1.836	3.262	.017
	x2cent	-7.416	3.132	-1.329	-2.368	.056
	interaction	-.820	.570	-.300	-1.440	.200

a. Dependent Variable: y

Continuous x Continuous

- Regression equations, rearrange terms for easier interpretation, take x2 as moderator:

$$\beta_0 + \beta_1\ x1cent_i + \beta_2\ x2cent_i + \beta_3\ x1cent{*}x2cent =$$
$$\beta_0 + \beta_2\ x2cent_i + (\beta_1 + \beta_3\ x2cent_i){*}x1cent_i$$

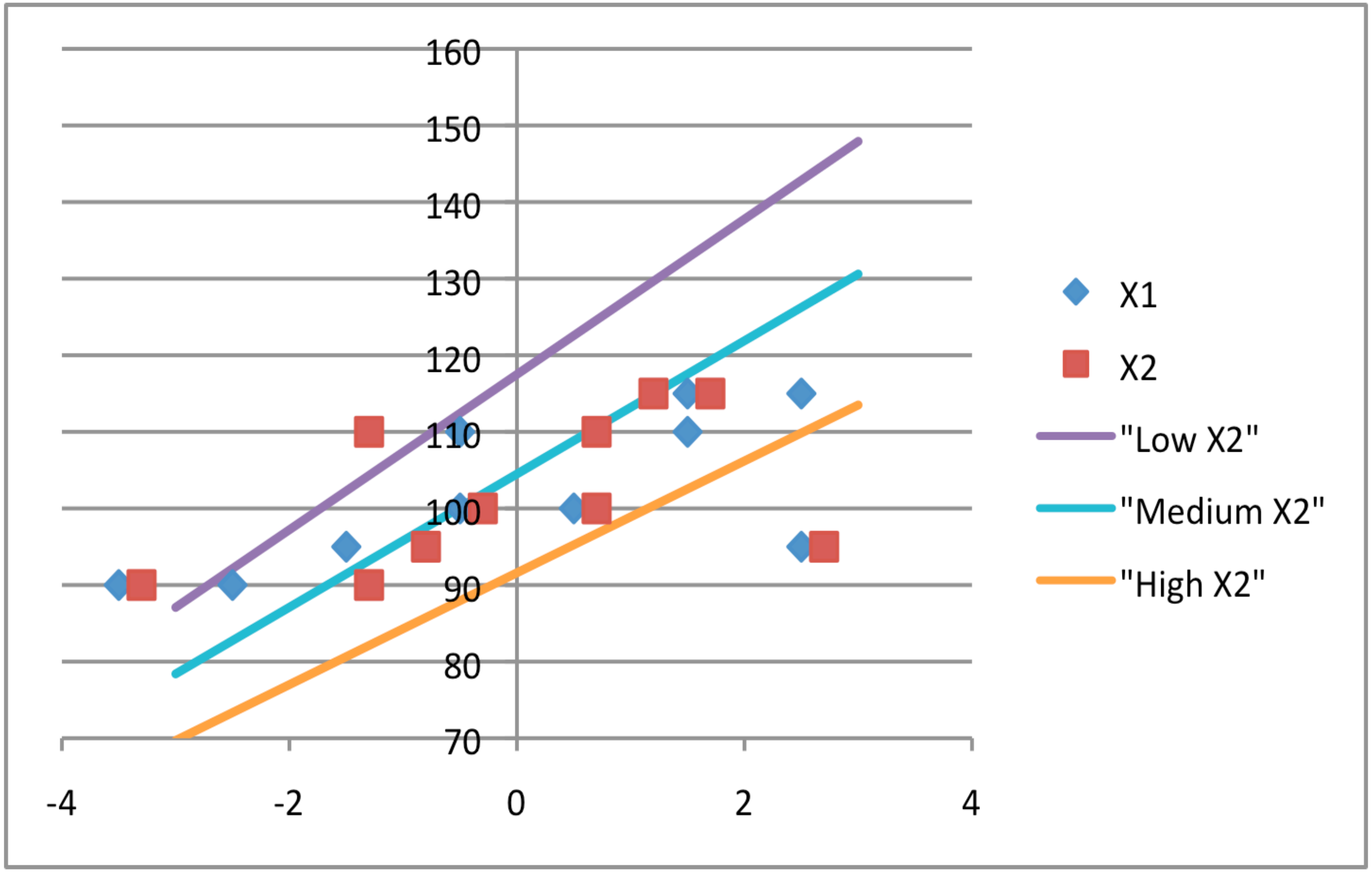
- -1 SD of x2cent (-1.75): Y' = 104.5 + -7.4* -1.75 + (8.7 -0.82*-1.75) * x1cent = **117.5 + 10.14 * x1cent**
- Mean of x2cent (0): Y' = **104.5 + 8.7 * x1cent**
- +1 SD of x2cent (1.75): Y' = 104.5 + -7.4*1.75 + (8.7 -0.82*1.75) * x1cent = **91.6 + 7.3 * x1cent**

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	102.000	1.971		51.748	.000
	x1cent	9.014	2.846	1.907	3.167	.016
	x2cent	-7.280	3.362	-1.304	-2.165	.067
2	(Constant)	104.502	2.528		41.345	.000
	x1cent	8.679	2.661	1.836	3.262	.017
	x2cent	-7.416	3.132	-1.329	-2.368	.056
	interaction	-.820	.570	-.300	-1.440	.200

a. Dependent Variable: y

Continuous x Continuous



Interaction was not significant

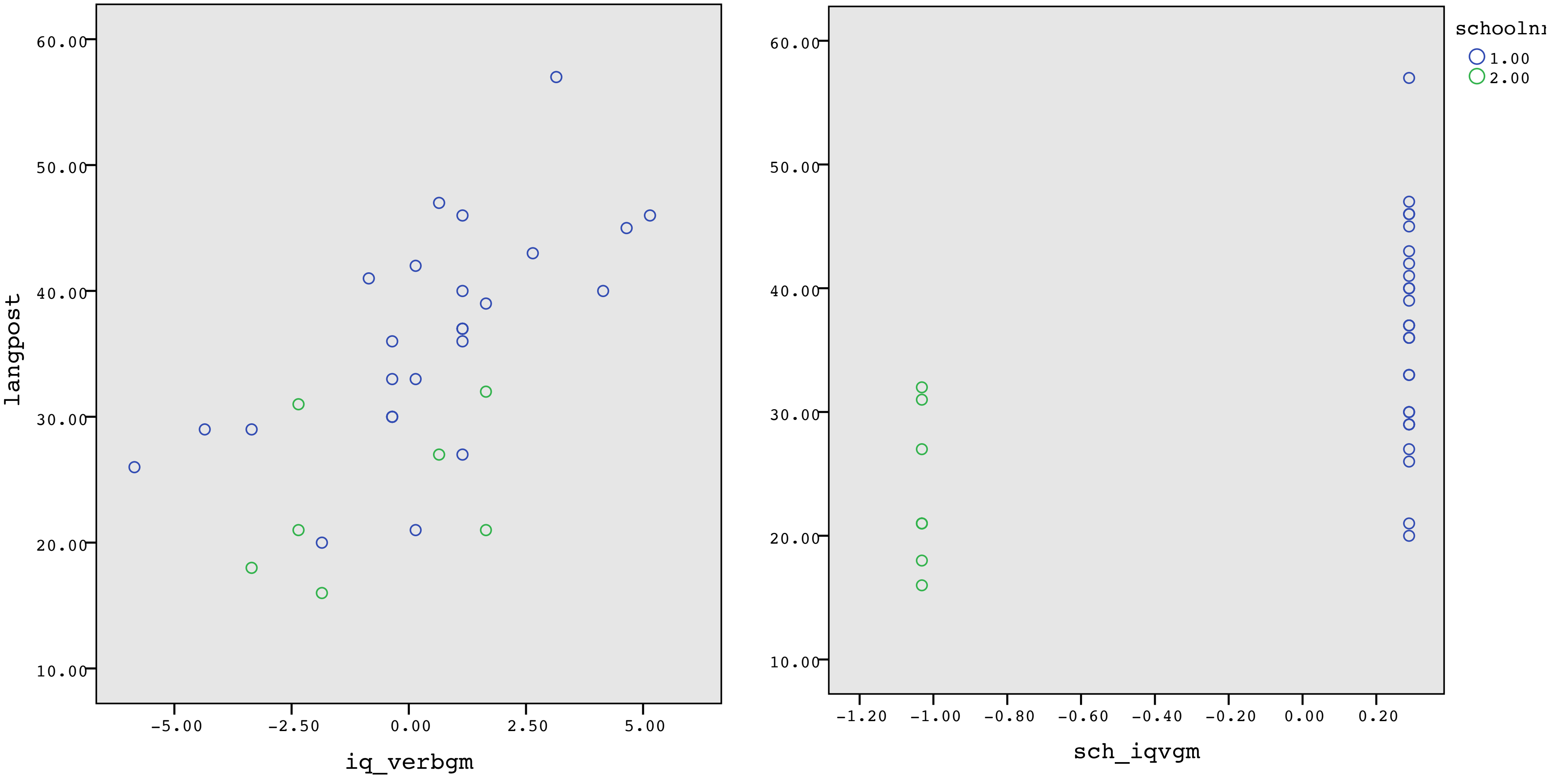
Cross-level interaction in multilevel models

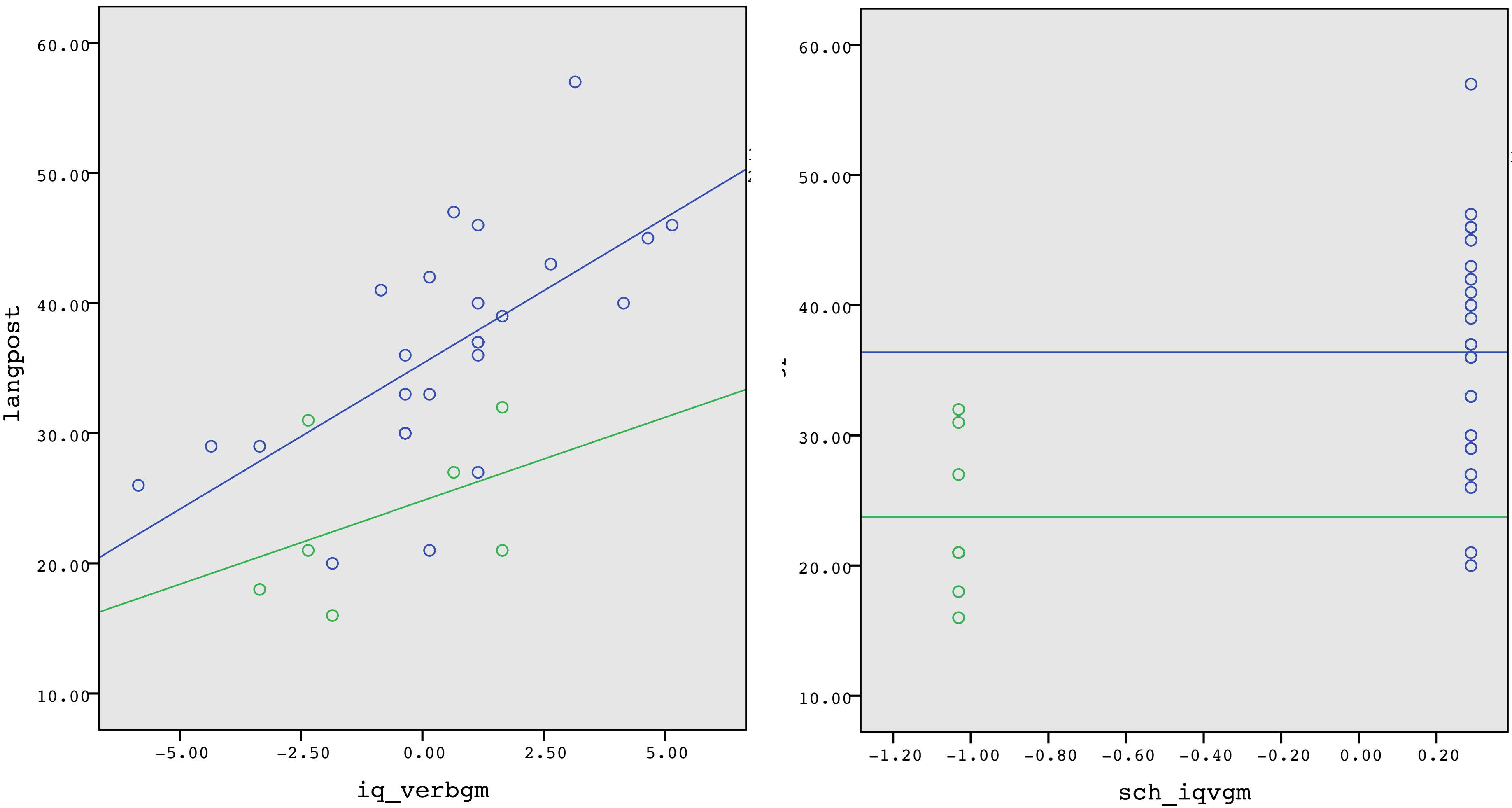
Recall the definition of a moderator variable:

The relationship between X and Y is different for different levels of the moderator: The slope

At which level do values of a cross-level interaction vary?

Level 1, so adding cross-level interactions = explain slope variance


60.00
50.00
40.00
30.00
20.00
10.00
langpost
-5.00
-2.50
0.00
2.50
5.00
iq_verbgm
-1.20
-1.00
-0.80
-0.60
-0.40
-0.20
0.00
0.20
sch_iqvgm
schoolnı
1.00
2.00


60.00
50.00
40.00
30.00
20.00
10.00
langpost
-5.00
-2.50
0.00
2.50
5.00
iq_verbgm
60.00
50.00
40.00
30.00
20.00
10.00
-1.20
-1.00
-0.80
-0.60
-0.40
-0.20
0.00
0.20
sch_iqvgm

Contextual effects explaining level 2 variance

Given a random slope model

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

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

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

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

Contextual variables explain level 2 variation in the intercept and cross-level interactions of contextual variables with x_1 explain level variability in the slope coefficient

This can be seen by re-arranging

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + \beta_2 x_{2j} + \beta_3(x_2 x_1)_{ij} + e_{ij}$$

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

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

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

$\longrightarrow$

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

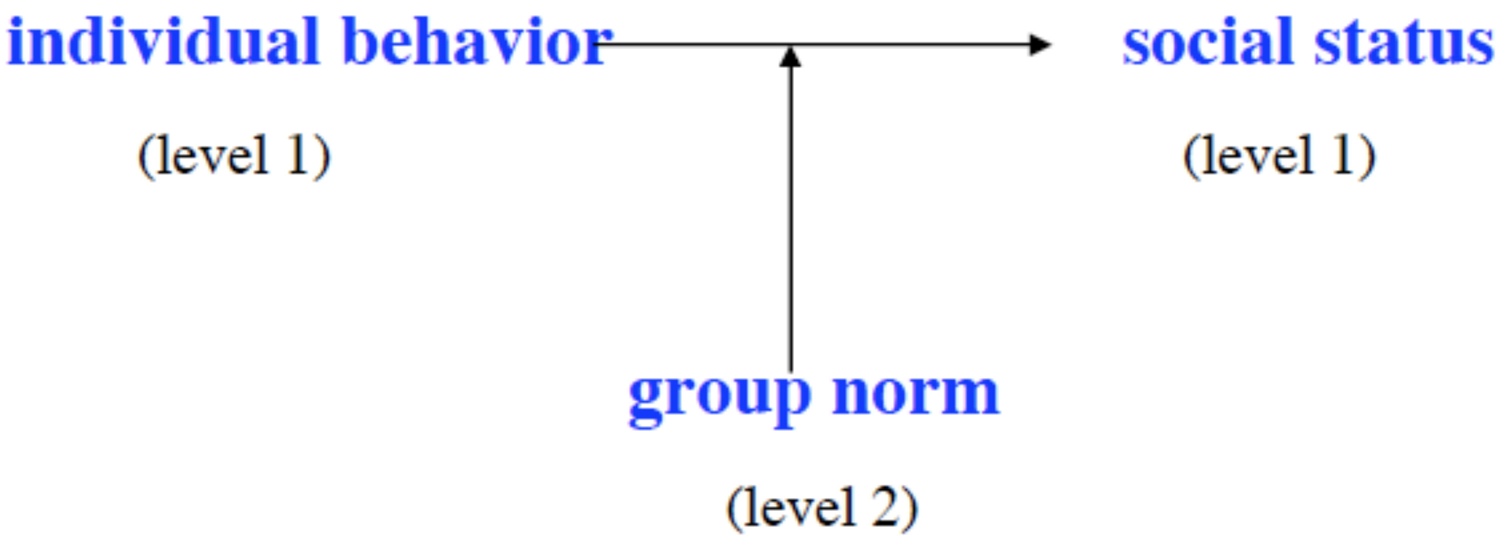
$$\beta_{0j} = \beta_0 + \beta_2 x_{2j} + u_{0j}$$

$$\beta_{1j} = \beta_1 + \beta_3 x_{2j} + u_{1j}$$

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

What might explain the inconsistent findings?

the person-group dissimilarity model (Wright *et al*. 1986)



Aggression ➔ peer rejection, when a person does *not fit in.*

This is a cross-level interaction.

Estimated final model for Bullying

$$Y' = 0.057 - 0.108\text{Boys} - 0.103\text{Bullying}$$
$$+ 0.011\text{ClassroomBullying} + 0.399\text{Bullying}*\text{CLBullying}$$

$$Y' = 0.057 - 0.108\text{Boys} + 0.011\text{CLBullying}$$
$$+ (-0.103 + 0.399\text{CLBullying})\text{Bullying}$$

- In classrooms with an average level of Bullying, the effect of individual Bullying equals -0.103.
- In classrooms with below-average levels of Bullying, the effect of individual Bullying is even more negative.
- But, in classrooms with above-average levels of Bullying, the effect of individual Bullying goes to zero and can even become positive.

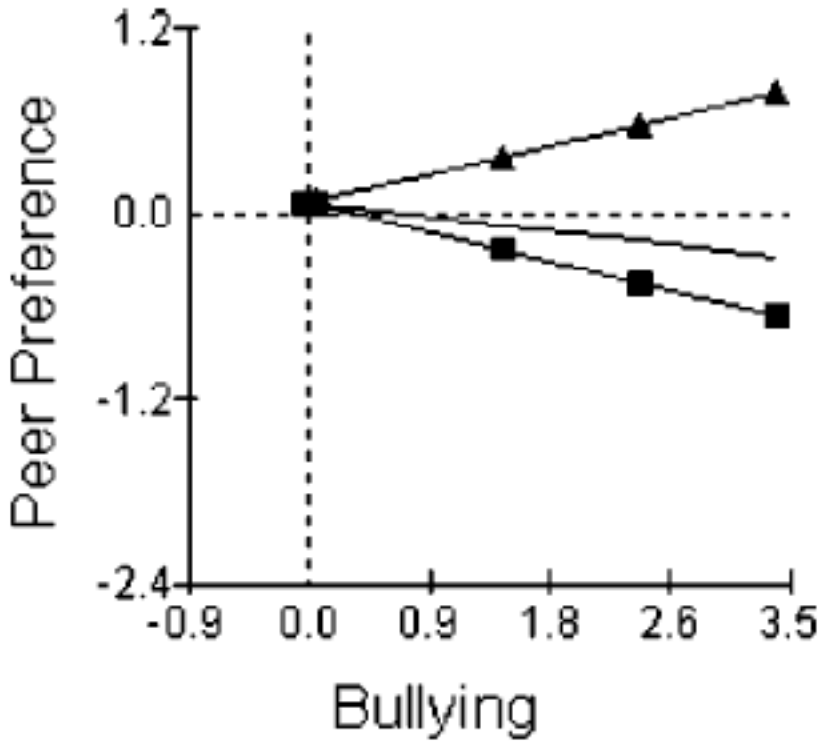


Fig. 1 Regression lines of bullying predicting peer preference in classrooms with low (*filled squares*), average (*line*) and high (*filled triangles*) levels of bullying

Modeling strategy summary (available on Github)

1) Focus on the **research questions and hypotheses**. It is self-evident that all effects stated in the research questions and the hypotheses should be in the model. On top of that, there may be effects that should be controlled for.

2) In a multilevel model one can distinguish ‘explaining’ the scores of a dependent variable at the first level, and ‘explaining’ variation between intercepts and slopes at the second level.
Both level-1 and level-2 predictors can be used to explain intercept variance. Cross-level interactions (interactions between level-2 and level-1 predictors) serve to explain slope variance. Level-2 predictors, of course, cannot explain level-1 variance.

3) **In general, when a model has an interaction effect make sure that the corresponding ‘main effects’ are also in the model.**

4) Allow a **predictor always a fixed effect.** In addition, level-1 predictors may be given a random effect at the second level (a random slope).

5) For a predictor at level 1, it may be important to question whether the effect of the predictor at the individual level within groups (within level-2 units) is the same as the effect of the same predictor at the group level. This can be checked by entering at level 2 a variable holding the group means of this predictor. **If these group means have a significant effect on the dependent variable, then it is clear that the betweengroups regression differs from the within-groups regression.**

6) When two or more random effects are present in the model, then one should consider whether or not to include the covariance(s) between these random effects as a free parameter in the model. **MLwiN automatically includes all covariances when random slopes are defined.** Maybe, some of these should be deleted from the model. Always keep the covariance between a random slope and the intercept in the model. When there are two or more random slopes, then there may be reasons to leave out (some) slope by slope covariances. In case of random slopes for a set of dummy variables that together represent one categorical variable, one should keep all covariances between random slopes of these dummies in the model.

7) It often makes sense to **simplify a model by removing non-significant effects.** But there may be reasons to keep a non-significant effect in the model. These reasons may be related for instance with issue 1), issue 3), and issue 6). A set of dummy variables representing one categorical variable is in general best treated as a conceptual unity; so don't remove non-significant dummies from a set that includes a dummy variable with a significant coefficient, or test the set as a whole for significance using the deviance test.

8) Multicollinearity (too high correlations between effects in a model) can be a big problem in multilevel analysis, especially when the number of predictors is relatively high compared with N. This can easily happen at a higher level.

9) **Cross-level interactions are meant to explain variation in slopes.** This does not necessarily imply that you should always first show significant slope variance before allowing a cross-level interaction in the model. There might be theoretical considerations to expect or to test for a specific cross-level interaction.

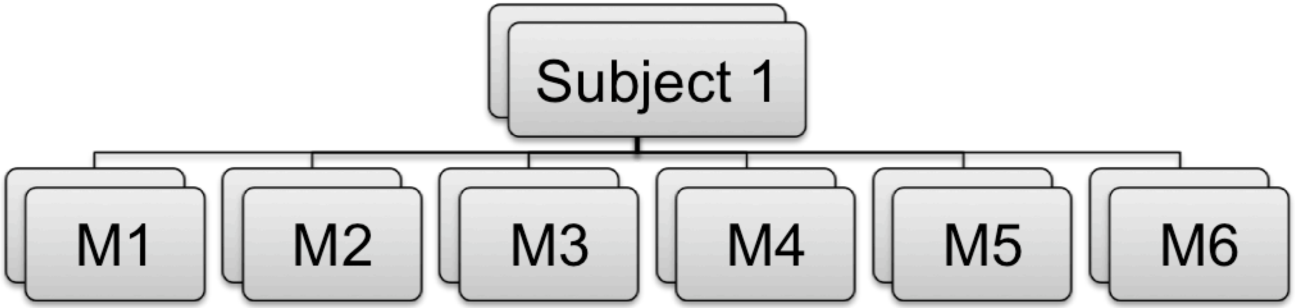
10) Test results concerning a certain effect may depend on other effects in the model; one should be aware of that when deciding to include or exclude an effect.

What's next?

We keep our focus on 2-level models, but now to analyse repeated measures data:

Level 2

Level 1



Of course this subject can again be nested within a classroom or a school making it a three-or-more-level model

Multilevel models can also be used for non-nested models.
What follows is just to show you what else you could do.

Logistic growth

If we combine these **linear** models we get **nonlinear** restricted (logistic) growth

$$\boxed{Y_{i+1} = rY_i(K - Y_i)}$$

$$\boxed{Y_{i+1} = rY_i(K - Y_i)}$$

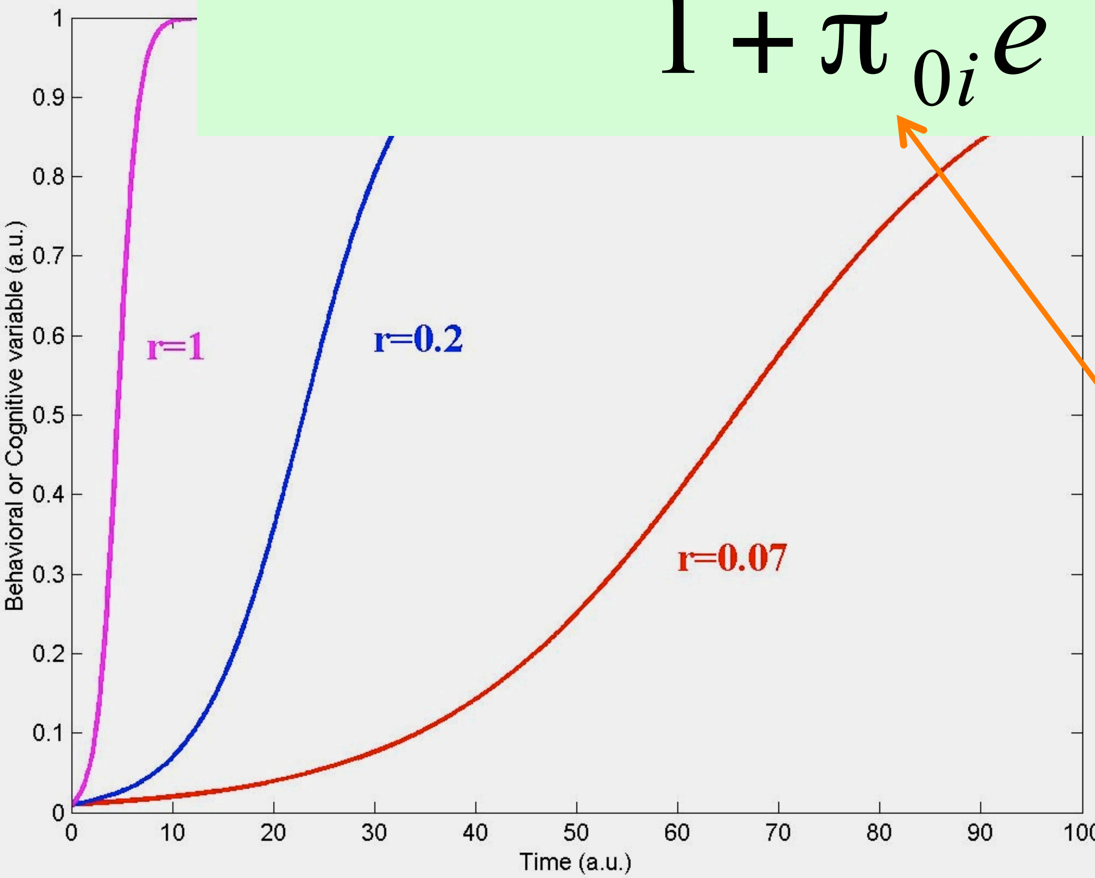
no analytic solution

$$\boxed{\frac{dY}{dt} = rY(K - Y)}$$

$$\boxed{Y(t) = \frac{KY_0}{Y_0 + (K - Y_0)e^{-Krt}}}$$

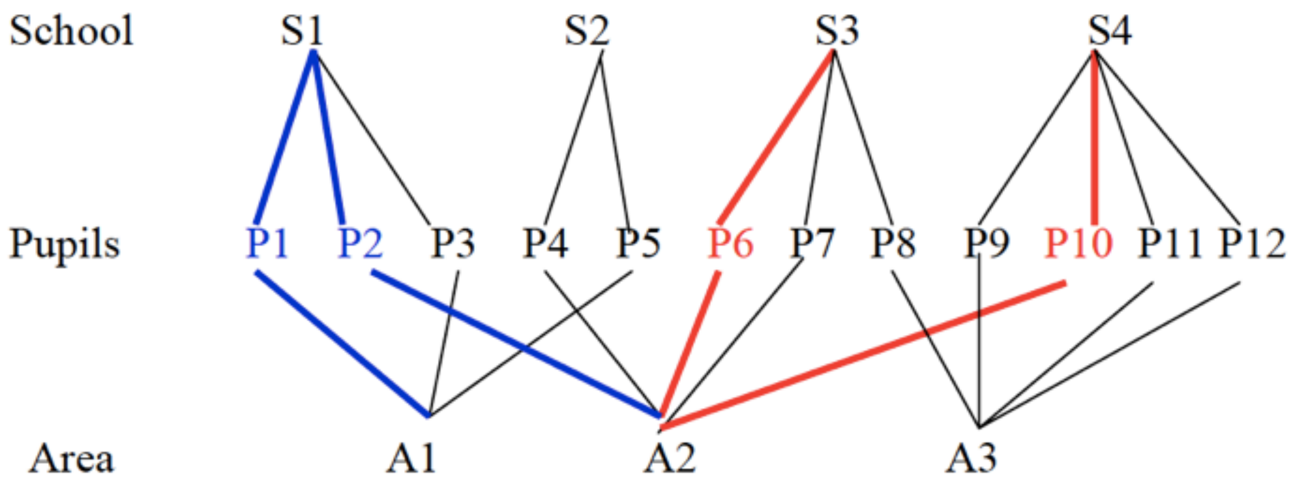
$$Y_{ij} = 1 + \frac{19}{1 + \pi_{0i} e^{-(\pi_{1i} TIME_{ij})}} + \varepsilon_{ij}$$

Singer & Willett example



$$\frac{dY}{dt} = rY(K - Y)$$

$$Y(t) = \frac{KY_0}{Y_0 + (K - Y_0)e^{-Krt}}$$

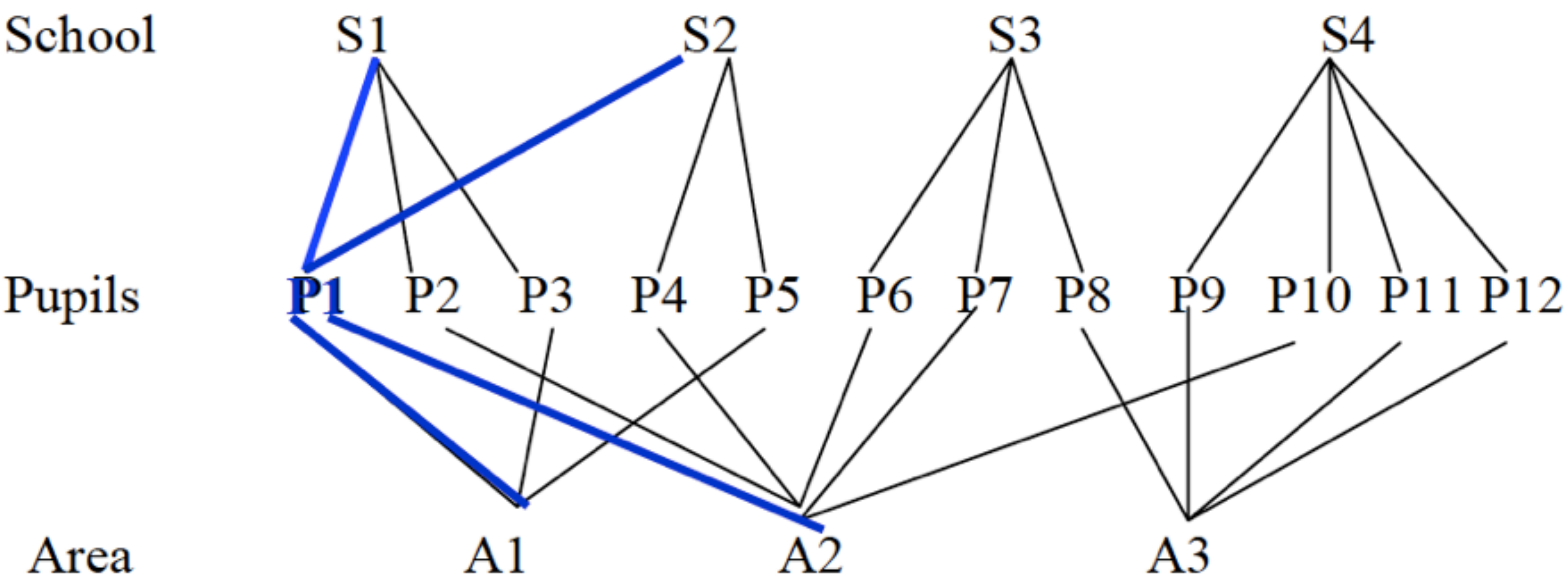


In this structure schools are not nested within areas. For example

Pupils 1 and 2 attend school 1 but come from different areas
Pupils 6 and 10 come from the same area but attend different schools

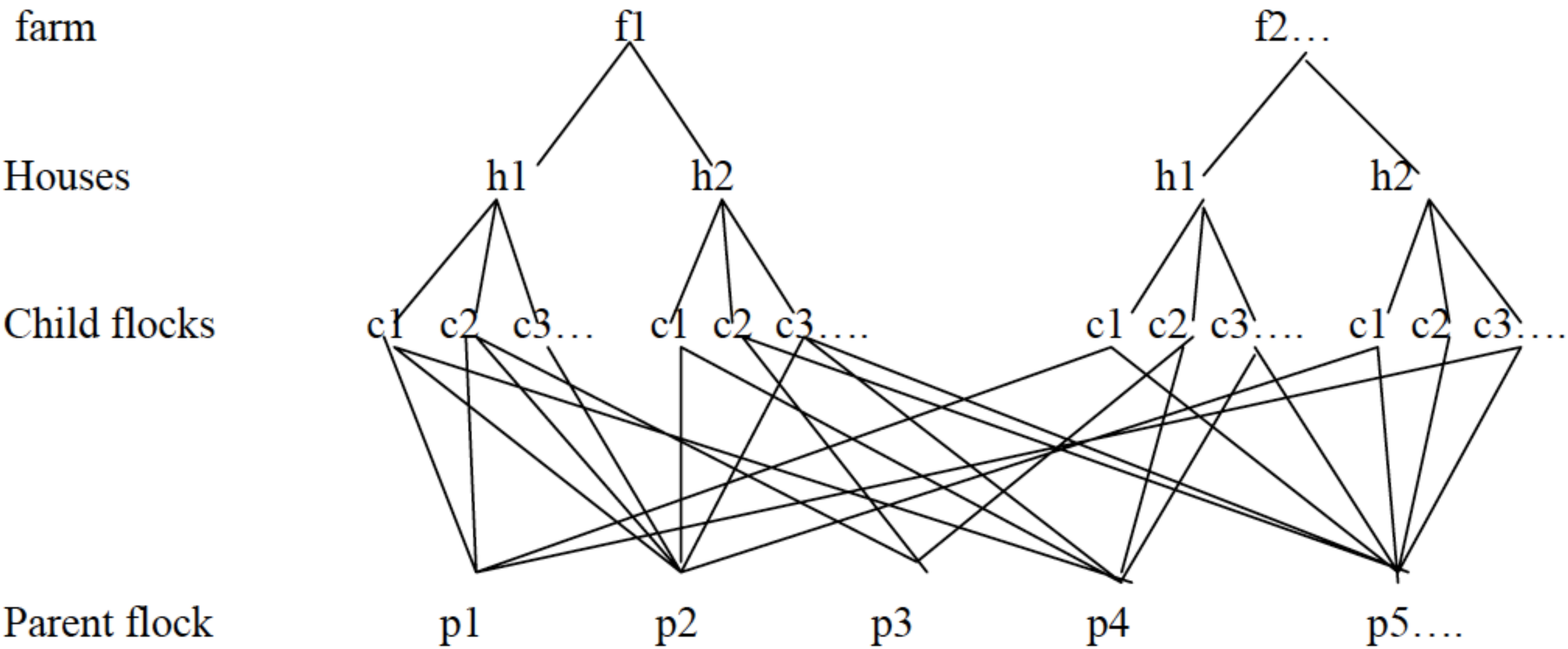
Area is not nested within school and school is not nested within area.
Pupils lie within a cross classification of school by area.

cross-classification models



Lets take the structure of pupils within a cross-classification of school by area. But now suppose pupil 1 moves in the course of the study from residential area 1 to 2 and from school 1 to 2.

multiple membership models



(complex) mixtures of structures

What's next?

When your dependent variable is not continuous and on a ratio scale... see chapters about (logistic) multilevel regression for:

- nominal data (unordered categorical data),
- ordinal data (ordered categorical data),
- count data (Poisson distributed data),

There is also a chapter on multivariate response data