Modelling Hierarchically Structured Data with MLwiN Software

Antero Malin
Finnish Institute for Educational Research
University of Jyväskylä

Helsinki 22.-23.5.2014

CONTENTS

1.

Introduction

Multilevel model

Hierarchical model

- Starting point: Population and the sample have a natural multilevel or hierarchical structure
- Data has observational units (units of analyses, cases) at every level of data
- Units at lower level are clustered, and these clusters (or groups) are units at higher level
- Sample is representative at every level of data

Examples of two-level data structure

Level 1	Level 2
Students	Schools
Persons	Households/Families
Employees	Employers
Athletes	Athletic club
Patients	Clinics/Hospitals
Peers	Peer groups

In longitudinal study:

Level 1: Measurement occasions

Level 2: Persons

Multilevel analyses should be used when:

- Data have a natural hierarchical/multilevel/nested structure
- Data include variables at different levels
- Level 1 units are not statistically independent
- We are interested in the effects of data structure

If multilevel analyses are not used when they should be used

We may get incorrect results in:

- Coefficient estimates
- Standard errors of estimates
- Statistical tests
- Conclusions

And we loose information of how the data structure influences the results.

=> We have to include the data structure in the statistical model and analyses

Applications of multilevel analyses

- Variance component model
- Random intercept model
- Random coefficient model
- Modeling contextual (compositional) variables
- Complex level 1 and 2 variation
- Model for repeated measures
- Growth curve model
- Model for multivariate response data
- Model for binary responses and proportions
- Multiple membership model

Artificial example

Data:

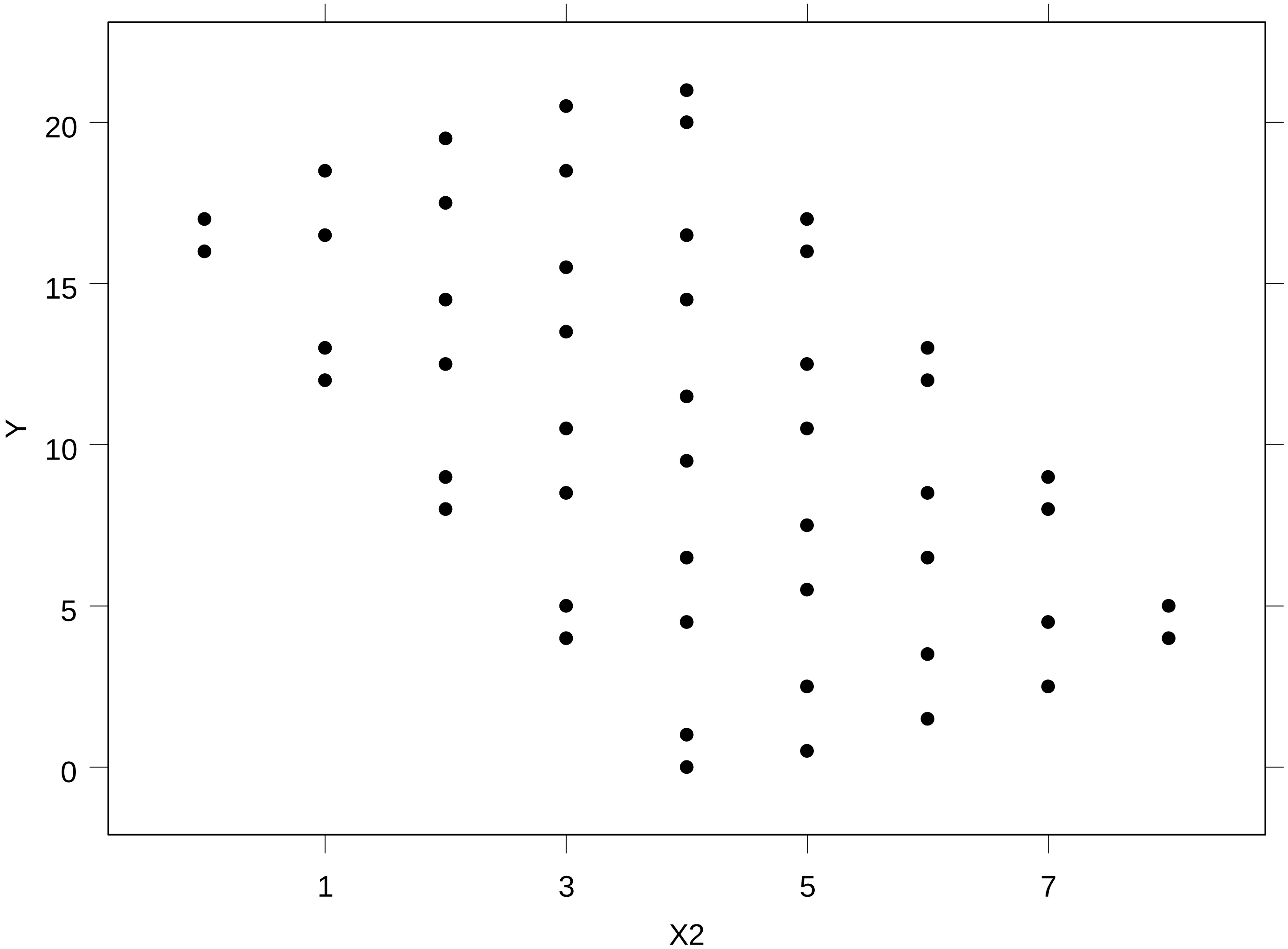
Schools: 5
Students: 5 x 10 = 50
Dependent variable: y
Explanatory variable: x2
Overall correlation: r(y,x2) = -0.51
Within-schools correlations: r(y,x2) = 0.86

Three regression lines:

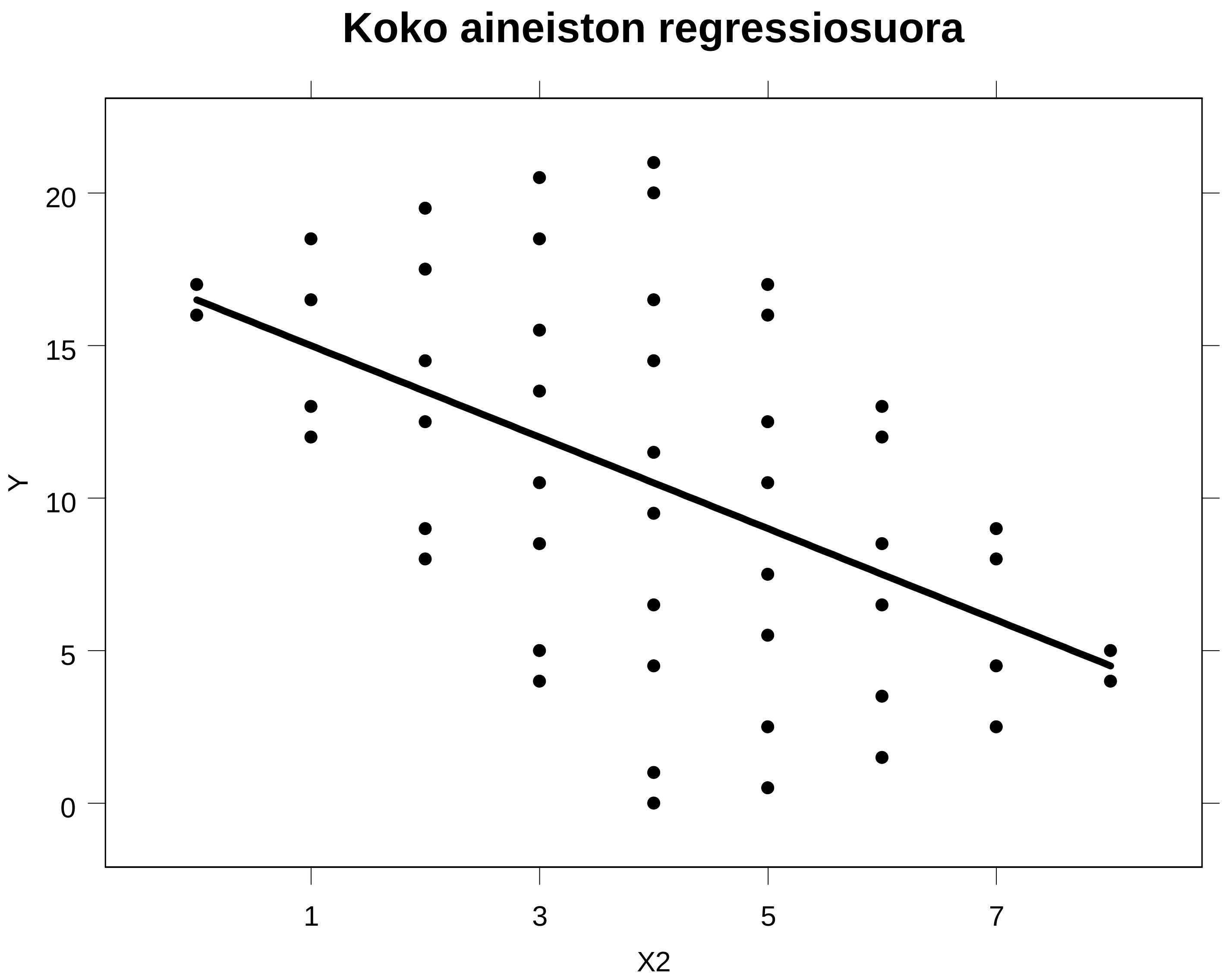
1. “Normal” regression
2. Within-schools regression
3. Between-schools regression

Overall scatterplot of data

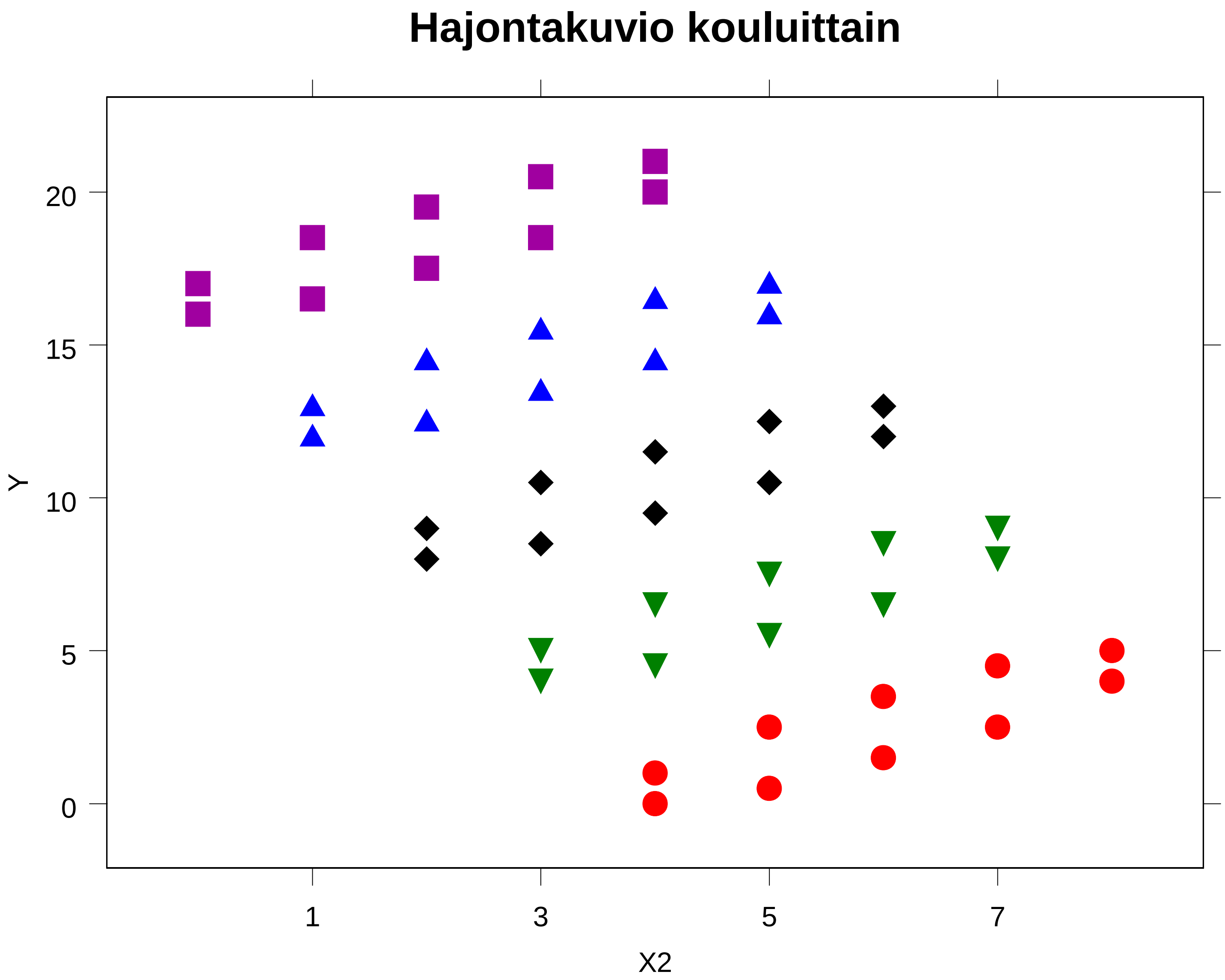
Koko aineiston hajontakuvio



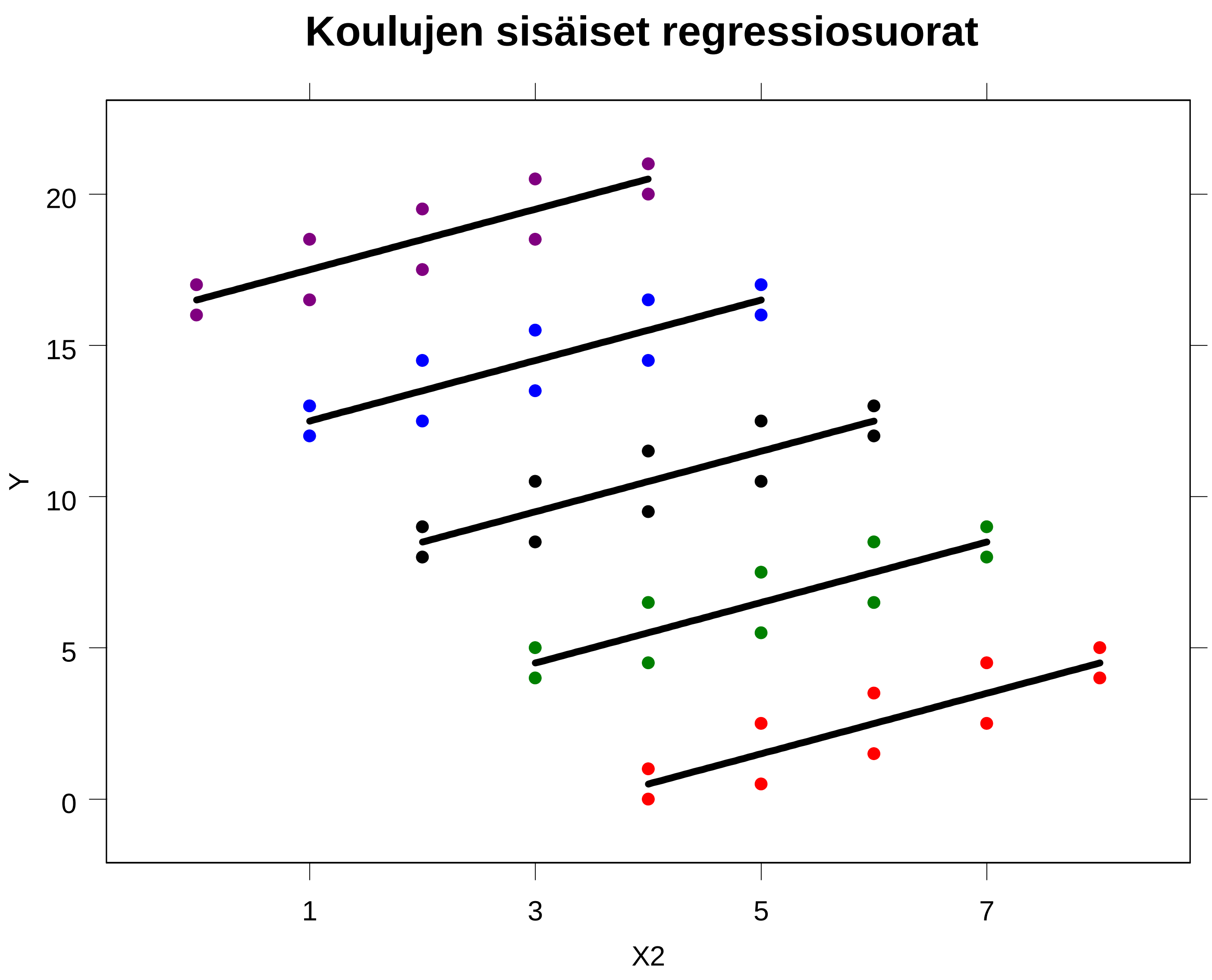
Regression line of Y on X2



Scatterplot by schools

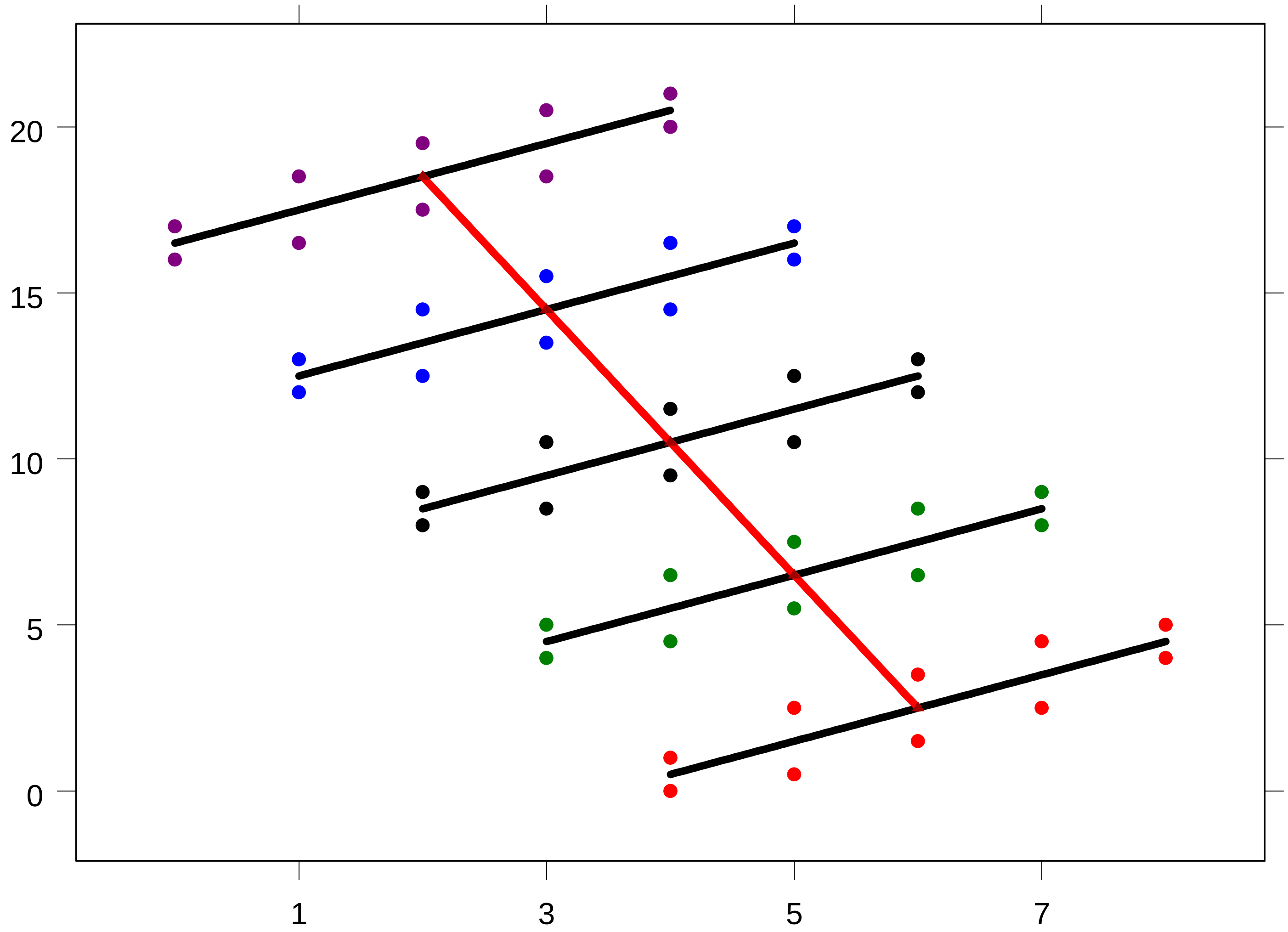


Within-schools regression lines

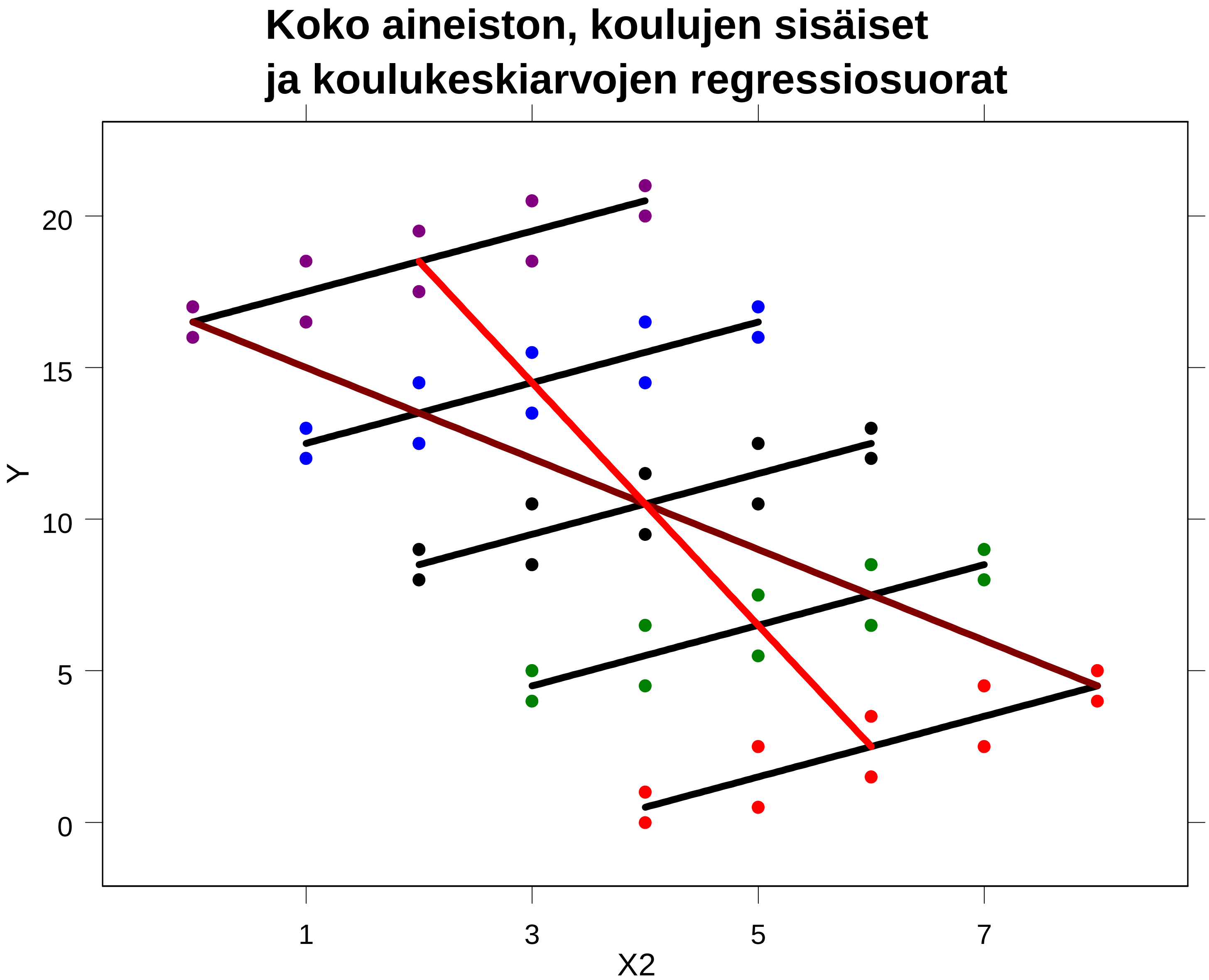


Within-schools and between-schools regression lines

Koulujen sisäiset ja koulukeskiarvojen regressiosuorat

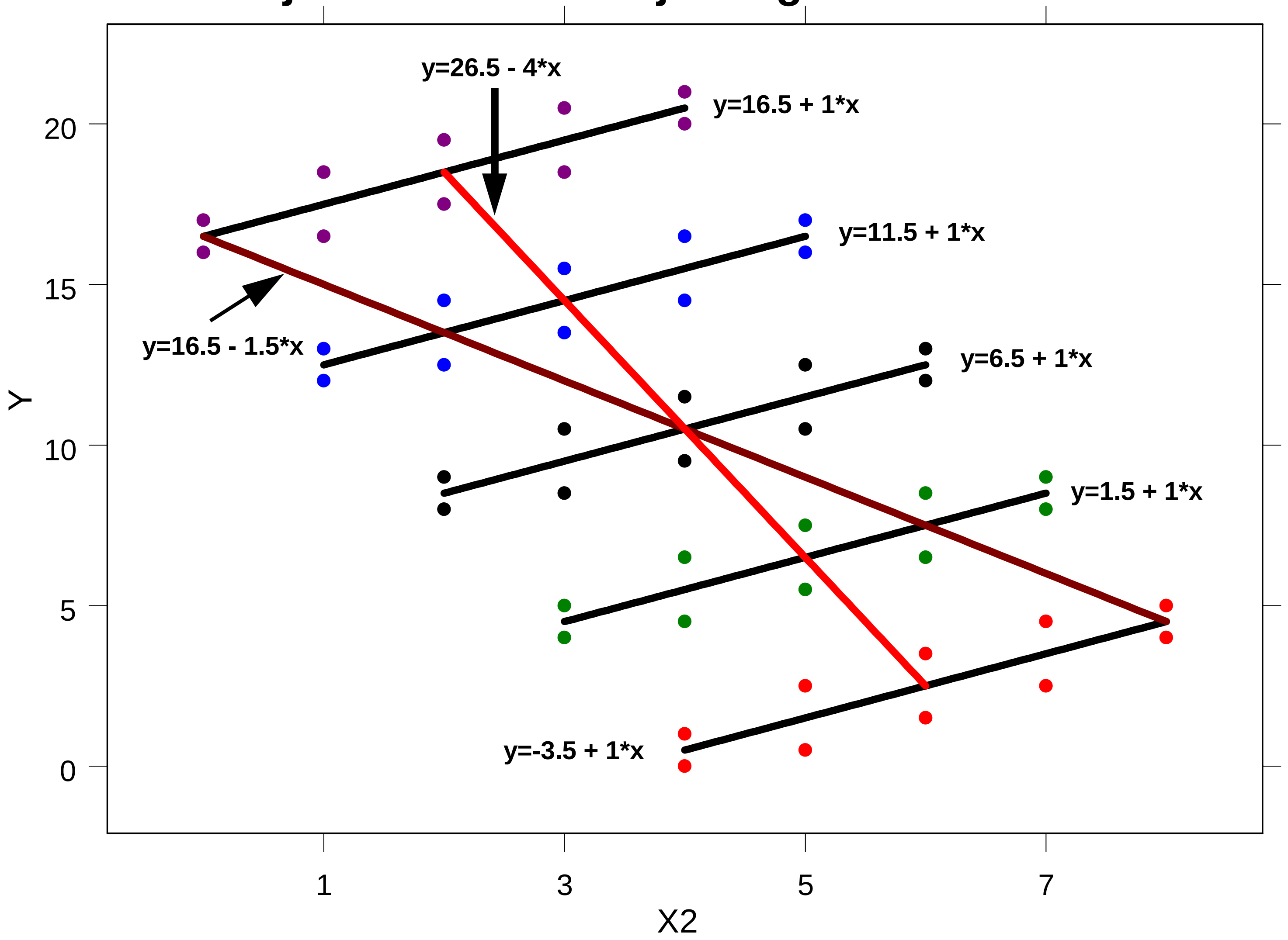


Within-schools, between-schools and total regression lines



Within-schools, between-schools and total regression lines

Koko aineiston, koulujen ja koulukeskiarvojen regressiosuorat



We can calculate three regression lines:

Within-schools, between-schools and total regression lines

Which one is the correct one?

All are correct, but:

- we have to know which one we have calculated,
- to make the correct interpretations

2. Simple variance component model

Antero Malin

Helsinki 22.-23.5.2014

Simple variance component model

- Variation between higher level units in mean performance
- No explanatory variables
- ICC is estimate of the between-group variation (and of statistical dependence between level-one units)
  - Intra-class correlation
  - Intra-cluster correlation
  - Intra-level-2 correlation
  - Intra-school correlation
  - Variance Partition Coefficient (VPC)

Statistical model

$$y_{ij} = \quad _0 + u_j + e_{ij}$$

y_{ij} *is outcome variable*

i is level 1 unit, $i=1,\ldots,n_j$

j is level 2 unit, j=$1,\ldots,J$

$_0$ is intercept of the model

e_{ij} is level 1 residual

u_j is level 2 residual

$$u_j \sim N(0, \pm_u^2) \quad e_{ij} \sim N(0, \pm_e^2) \quad \mathrm{cov}(u_j e_{ij}) = 0$$

ICC (VPC)

Total variance $\pm^{2}_{tot}$

is divided in two variance components:

between clusters $\pm^{2}_{u}$

and within clusters $\pm^{2}_{e}$

$$\pm^{2}_{tot} = \pm^{2} = \pm^{2}_{u} + \pm^{2}_{e}$$

ICC:

Between-clusters variance component is divided by total variance

$$\circ = \frac{\pm_u^2}{\pm_u^2 + \pm_e^2} = \frac{\pm_u^2}{\pm^2}$$

$$\text{IF}: \pm_u^2 = 0 \quad \Rightarrow \quad \circ = 0$$

$$\text{IF}: \pm_e^2 = 0 \quad \Rightarrow \quad \circ = 1$$

ICC and sample size

If ICC>0, efficient sample size is smaller than number of observations (nominal sample size):

$$N_{eff} = \frac{J \times n}{1 + (n-1)\circ}$$

N_{eff} efficient sample size

J number of cluster

n group size (constant in this example)

ρ ICC

J x n number of observations

$$\text{If} \quad \circ = 0 \text{ then } N_{eff} = J \times n$$

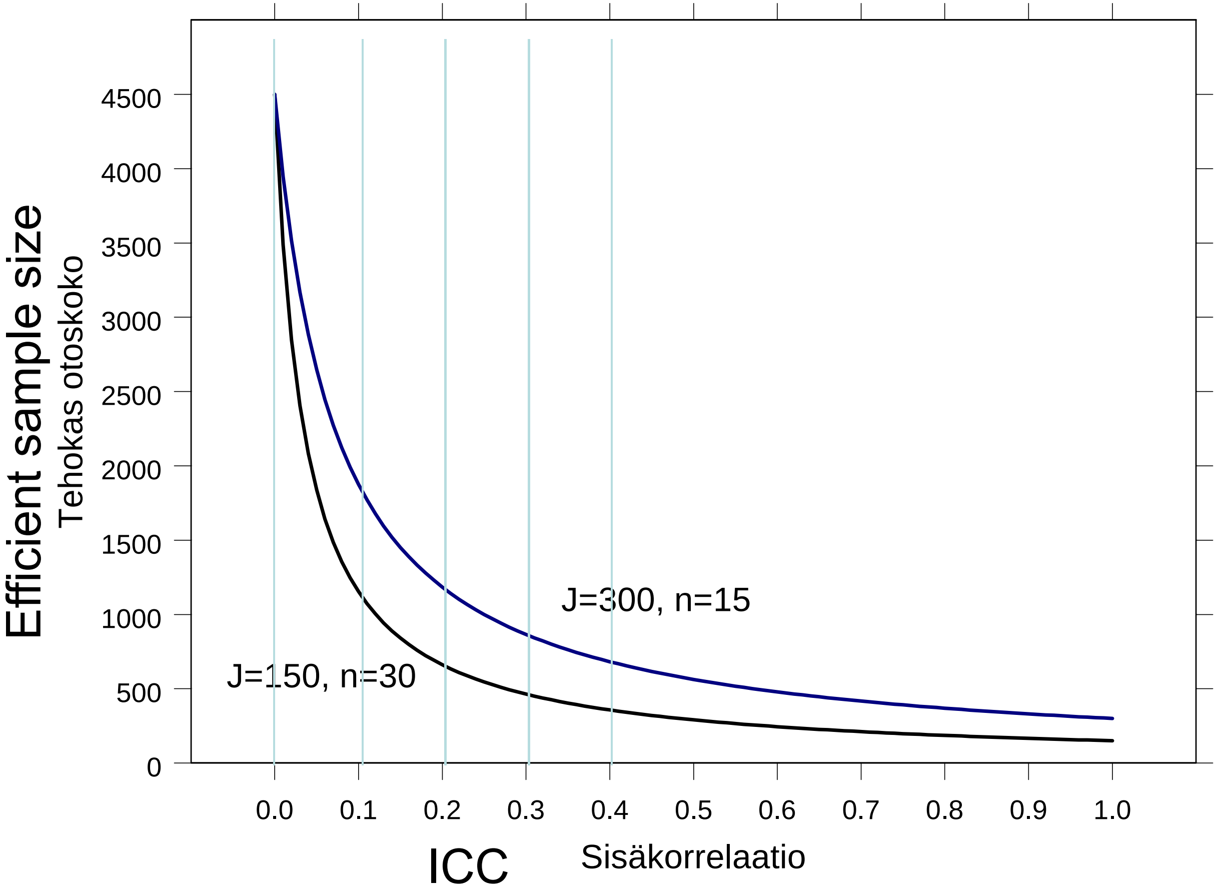
$$\text{If} \quad \circ = 1 \text{ then } N_{eff} = J$$

Example: ICC and efficient sample size

Number of level 1 observations N=4500

A. Schools 150=J, students in every school 30=n

B. Schools 300=J, students in every school 15=n



Example: ICC and efficient sample size

Number of level 1 observations N=4500

A. Schools 150=J, students in every school 30=n

B. Schools 300=J, students in every school 15=n

	Efficient sample size	
	J=150	J=300
ICC	n=30	n=15
0	4500	4500
0.1	1154	1875
0.2	662	1184
0.3	464	865
0.4	357	682
0.5	290	563
0.6	245	479
0.7	211	417
0.8	186	369
0.9	166	331
1.0	150	300

Intra-cluster correlation ICC

- Is the correlation between the lower level units within higher level units. It is not correlation between two variables.
- Range: $0 \leq \circ \leq 1$
- If there is ICC, the lower level units are statistically dependent, not statistically independent which is one of the assumptions of ”traditional” statistical methods
- ICC means that lower level units are more or less homogeneous
- ICC tells how much of the total variance is attributable to the variation between higher level units
- If ICC increases, efficient sample size decreases

Example 1: Shool differences in reading literacy in Finland (PISA 2000). Simple variance component model.

Statistics of reading literacy score:

Mean	549
SD	87.0
Min	204
Max	838
Students	4859
Schools	154

Statistical model

$$y_{ij} = {}_0 + u_j + e_{ij}$$

y_{ij} is outcome variable: *reading literacy score*

i is level 1 unit: *student*, $i=1,\dots,n_j$

j is level 2 unit: *school*, $j=1,\dots,J$

${}_0$ is constant of the model: *overall mean of reading literacy score*

u_j is level 2 residual: *school's j deviation from the estimated overall mean* $\hat{}_0$

e_{ij} is level 1 residual: *student's deviation from the school mean* $\hat{}_0 + \hat{u}_j$

$$u_j \sim N(0, \pm_u^2) \quad e_{ij} \sim N(0, \pm_e^2) \quad \mathrm{cov}(u_j e_{ij}) = 0$$

Example 1 (cont.): Finland

PISA 2000 data, Combined reading literacy score

Equations

$$\mathrm{read1}_{ij} \sim \mathrm{N}(XB,\ \Omega)$$

$$\mathrm{read1}_{ij} = \beta_{0ij}\mathrm{intercept}$$

$$\beta_{0ij} = 548.803(2.160) + u_{0j} + e_{0ij}$$

$$\begin{bmatrix} u_{0j} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_u)\ :\ \Omega_u = \begin{bmatrix} 490.768(81.799) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_e)\ :\ \Omega_e = \begin{bmatrix} 7087.543(146.125) \end{bmatrix}$$

$-2{*}loglikelihood(IGLS) = 57047.410$(4859 of 4859 cases in use)

Fonts | Subs | Name | + | - | Add Term | Estimates | Nonlinear | Help | Clear

Example 1 (cont.): Finland

PISA 2000 data, Combined reading literacy score

$$\hat{\gamma}_0 = 548.8$$

$$\hat{\sigma}_u^2 = 490.8 \quad \hat{\sigma}_e^2 = 7087.5$$

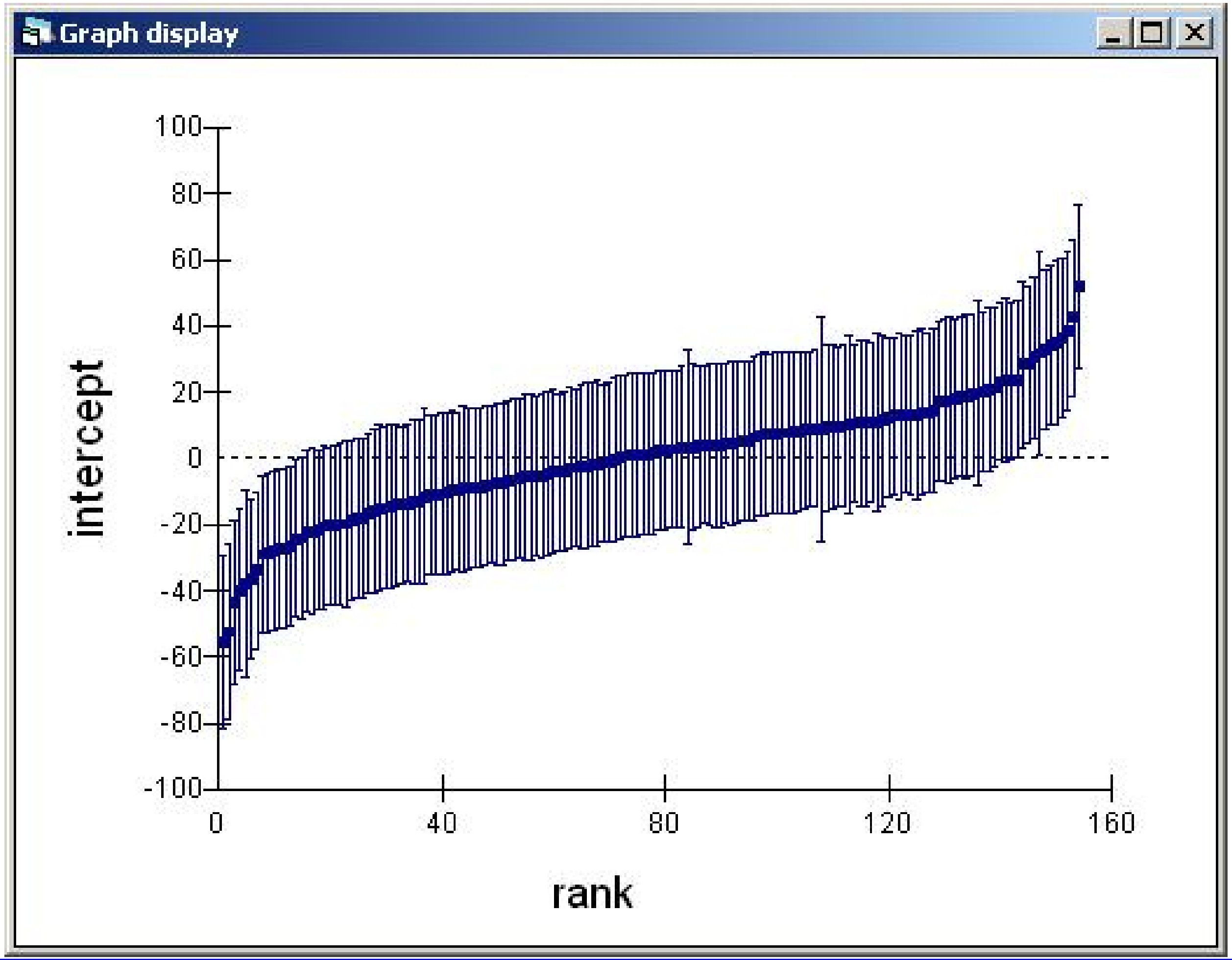
$$\hat{\rho} = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_e^2} = \frac{490.8}{490.8 + 7087.5}$$

$$= \frac{490.8}{7578.3} = 0.065$$

About 6.5 % of the total student variance in the reading literacy score is attributable to the between-school variation

Example 1 (cont.): Finland

School means (deviations from overall mean)of the combined reading literacy score with 95 % confidence intervals in rank order



Example 2: Shool differences in reading literacy in Germany (PISA 2000). Simple variance component model.

Statistics of reading literacy score:

Mean	504
SD	98.9
Min	143.1
Max	779.4
Students	4108
Schools	183

Example 2 (cont.): Germany

PISA 2000 data, Combined reading literacy score

Equations

$$\text{read1}_{ij} \sim \mathrm{N}(XB, \Omega)$$

$$\text{read1}_{ij} = \beta_{0ij}\text{intercept}$$

$$\beta_{0ij} = 496.324(5.610) + u_{0j} + e_{0ij}$$

$$\left[u_{0j}\right] \sim \mathrm{N}(0, \Omega_u) : \Omega_u = \left[5541.795(601.573)\right]$$

$$\left[e_{0ij}\right] \sim \mathrm{N}(0, \Omega_e) : \Omega_e = \left[4512.242(101.859)\right]$$

$-2*loglikelihood(IGLS) = 46832.950$(4108 of 4108 cases in use)

Fonts | Subs | Name | + | - | Add Term | Estimates | Nonlinear | Help | Clear

Example 2 (cont.): Germany

PISA 2000 data, Combined reading literacy score

$$\hat{\gamma}_0 = 496.3$$

$$\hat{\sigma}_u^2 = 5541.8 \quad \hat{\sigma}_e^2 = 4512.2$$

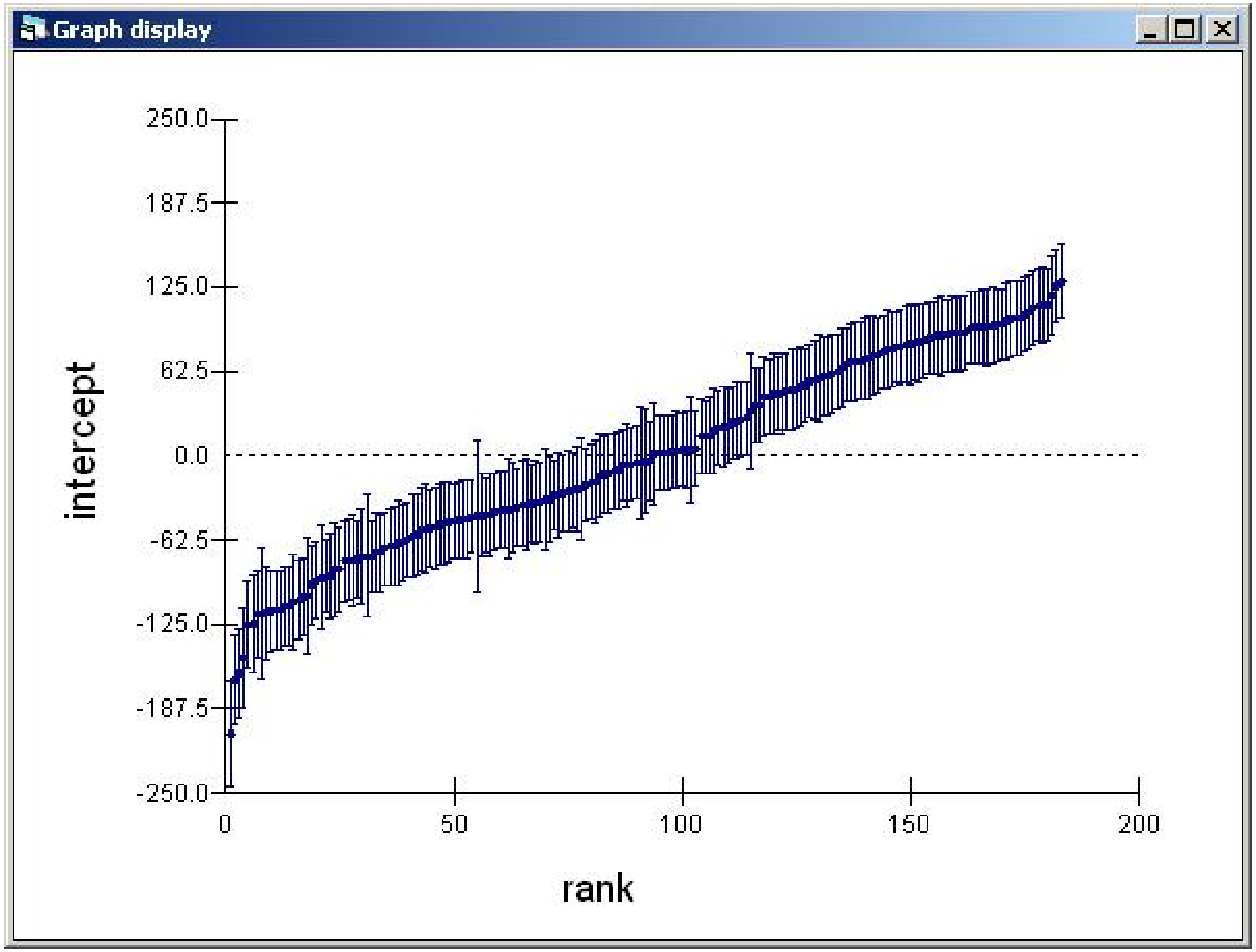
$$\hat{\rho} = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_e^2} = \frac{5541.8}{5541.8 + 4512.2}$$

$$= \frac{5541.8}{10054.0} = 0.551$$

About 55.1 % of the total student variance in the reading literacy score is attributable to the between-school variation

Example 2 (cont.): Germany

School residuals of the combined reading literacy score with 95 % confidence intervals in rank order



3.

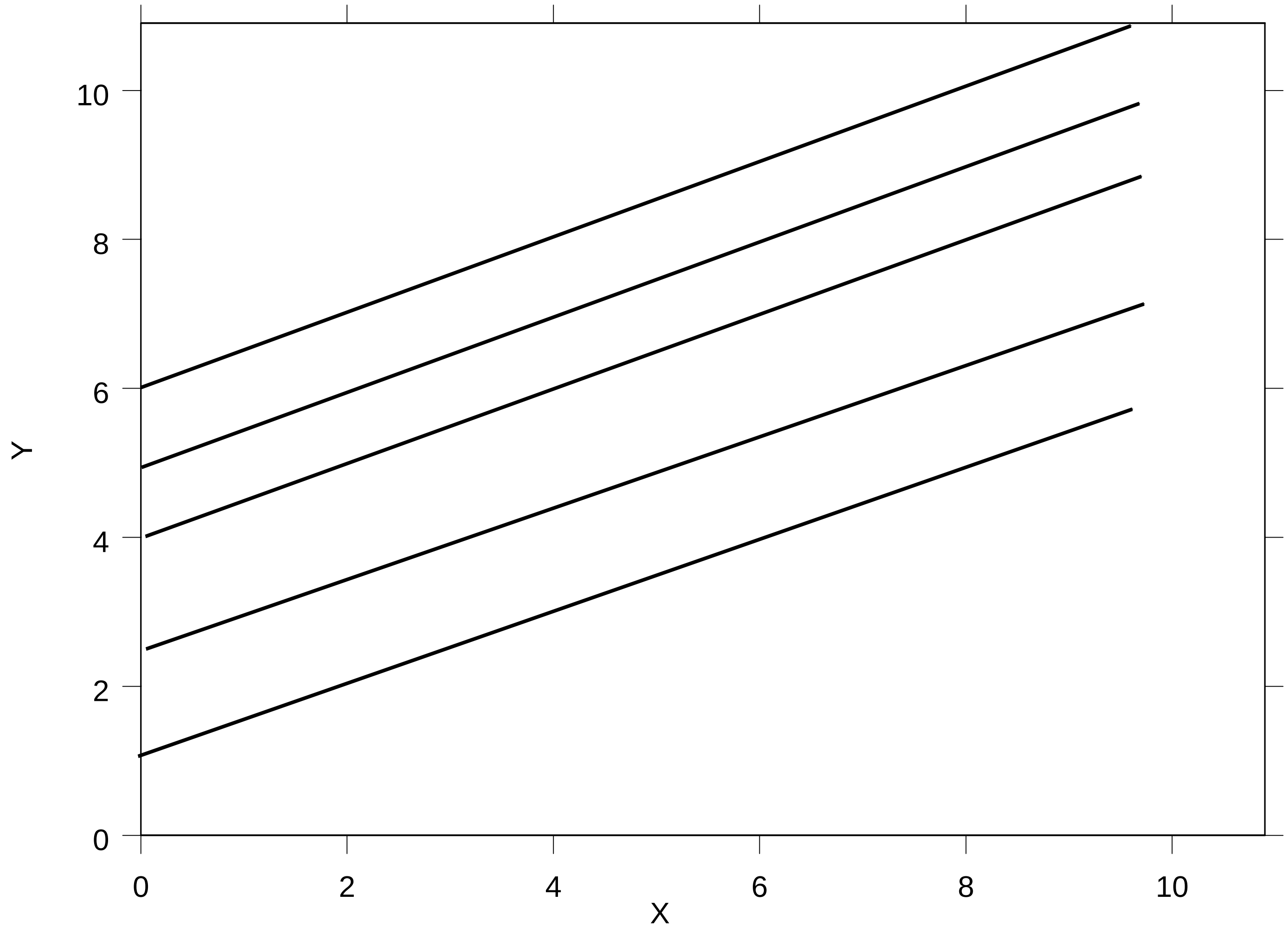
Random intercept model

Antero Malin
Helsinki 22.-23.5.2014

Random intercept model

- Intercept is varying between clusters
- Regression lines are parallel, i.e. slope (regression coefficient) is not varying between clusters

Random intercept model



Random intercept model:

$$y_{ij} = {}_0 + {}_1 \times x_{ij} + u_j + e_{ij}$$
$$= ({}_0 + u_j) + {}_1 \times x_{ij} + e_{ij}$$

y_{ij} is outcome variable

i is level 1 unit, $i=1,\ldots,n_j$

j is level 2 unit, $j=1,\ldots,J$

${}_0$ is intercept of the model

${}_1$ is coefficient

e_{ij} is level 1 residual

u_j is level 2 residual

$$u_j \sim N(0,\pm_u^2) \quad e_{ij} \sim N(0,\pm_e^2) \quad \mathrm{cov}(u_j e_{ij}) = 0$$

Example 3. Reading literacy and socio-economic index (hisei) in Finland (PISA 2000).

Statistics of reading literacy score and hisei:

	Reading score	Hisei
Mean	549	50.1
SD	87.0	16.2
Min	204	16
Max	838	90
Students	4859	4859
Schools	154	154

Example 3 (cont.): Finland

PISA 2000 data, Combined reading literacy score

$$y_{ij} = \;{}_0 + \;{}_1 \times Hisei_{ij} + u_j + e_{ij}$$
$$= (\;{}_0 + u_j) + \;{}_1 \times Hisei_{ij} + e_{ij}$$
$$= (\;{}_0 + u_j + e_{ij}) + \;{}_1 \times Hisei_{ij}$$

Equations

$$\mathrm{read1}_{ij} \sim \mathrm{N}(XB, \Omega)$$
$$\mathrm{read1}_{ij} = \beta_{0ij}\mathrm{intercept} + 1.174(0.077)\mathrm{hisei}_{ij}$$
$$\beta_{0ij} = 491.006(4.343) + u_{0j} + e_{0ij}$$

$$\left[u_{0j}\right] \sim \mathrm{N}(0,\ \Omega_u)\ :\ \Omega_u = \left[396.150(69.887)\right]$$

$$\left[e_{0ij}\right] \sim \mathrm{N}(0,\ \Omega_e)\ :\ \Omega_e = \left[6665.530(138.822)\right]$$

-2*loglikelihood(IGLS) = 55637.020(4765 of 4859 cases in use)

Testing statistical significance

Example 3 (cont): Finland

$$\hat{\beta}_1 = 1.174 \quad s.e. = 0.077$$

$$H_0 : \beta_1 = 0 \quad ; \quad H_1 : \beta_1 \neq 0$$

$$\chi^2_1 = 233.4 \quad ; \quad p < .001$$

$$\hat{\sigma}^2_u = 396.150, \quad with\ s.e. = 69.887$$

$$H_0 : \sigma^2_u = 0 \quad ; \quad H_1 : \sigma^2_u > 0$$

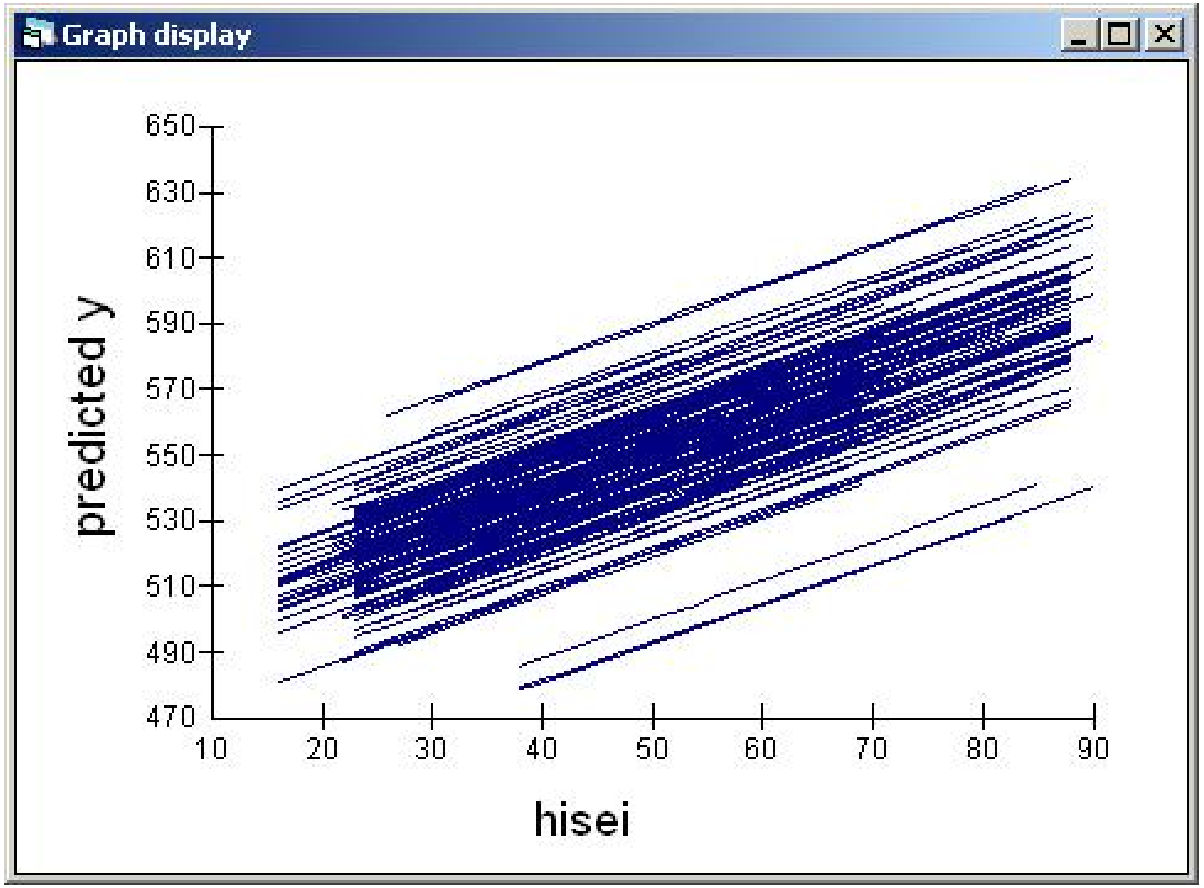
$$\chi^2_1 = 32.1 \quad ; \quad p < .001$$

Example 3 (cont.): Finland

Separate regression lines for sample schools

are due to variation in the intercept: $(\hat{\beta}_0 + \hat{u}_j)$

Regression lines are parallel: $(\hat{\beta}_0 + \hat{u}_j) + \hat{\beta}_1 \times Hisei_{ij}$



Proportional reduction in variance component

Estimated variance components for the null model are: $\hat{\sigma}^2_{u(0)} \; and \; \hat{\sigma}^2_{u(0)}$

Estimated variance components for the model including predictors are: $\hat{\sigma}^2_{u(x)} \; and \; \hat{\sigma}^2_{u(x)}$

Proportional reduction in level 2 variance component is:

$$R_B^2 = \frac{(\hat{\sigma}^2_{u(0)} - \hat{\sigma}^2_{u(x)})}{\hat{\sigma}^2_{u(0)}} = 1 - \frac{\hat{\sigma}^2_{u(x)}}{\hat{\sigma}^2_{u(0)}}$$

Proportional reduction in level 1 variance component is:

$$R_B^2 = \frac{(\hat{\sigma}^2_{e(0)} - \hat{\sigma}^2_{e(x)})}{\hat{\sigma}^2_{e(0)}} = 1 - \frac{\hat{\sigma}^2_{e(x)}}{\hat{\sigma}^2_{e(0)}}$$

Proportional reduction in total variance is:

$$R_{Tot}^2 = 1 - \frac{\hat{\sigma}^2_{tot(x)}}{\hat{\sigma}^2_{tot(0)}}$$

Example 3 (cont.): How much does the model (or *hisei* alone in this example) explain of the variance of reading literacy score?

	Null model	Model with predictors	Variance explained %
Between-school variance	490.8	396.2	19.3
Between-student variance	7087.5	6665.5	6.0
Total variance	7578.5	7061.7	6.8

Example 4: Reading literacy and socio-economic index (hisei) in Germany (PISA 2000).

Statistics of reading literacy score and hisei:

	Reading score	Hisei
Mean	504	49.9
SD	98.9	15.6
Min	143.1	16
Max	779.4	90
Students	4108	4108
Schools	183	183

Example 4 (cont.): Germany

PISA 2000 data, Combined reading literacy score

$$
\begin{aligned}
y_{ij} &= {}_0 + {}_1 \times Hisei_{ij} + u_j + e_{ij} \\
&= ({}_0 + u_j) + {}_1 \times Hisei_{ij} + e_{ij} \\
&= ({}_0 + u_j + e_{ij}) + {}_1 \times Hisei_{ij}
\end{aligned}
$$

Equations

$\mathrm{read1}_{ij} \sim \mathrm{N}(XB,\ \Omega)$

$\mathrm{read1}_{ij} = \beta_{0ij}\mathrm{intercept} + 0.726(0.076)\mathrm{hisei}_{ij}$

$\beta_{0ij} = 460.582(6.465) + u_{0j} + e_{0ij}$

$\begin{bmatrix} u_{0j} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_u)\ :\ \Omega_u = \begin{bmatrix} 4843.148(528.299) \end{bmatrix}$

$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_e)\ :\ \Omega_e = \begin{bmatrix} 4438.686(100.198) \end{bmatrix}$

$-2{*}loglikelihood(IGLS) = 46744.610$(4108 of 4108 cases in use)

Fonts | Subs | Name | + | - | Add Term | Estimates | Nonlinear | Help | Clear

Testing statistical significance

Example 4 (cont.): Germany

$$\hat{\beta}_1 = 0.726 \quad with\ s.e. = 0.076$$

$$H_0 : \beta_1 = 0 \quad ; \quad H_1 : \beta_1 \neq 0$$

$$\chi^2_1 = 90.5 \quad ; \quad p < .001$$

$$\hat{\sigma}^2_u = 4843.1 \quad s.e. = 528.3$$

$$H_0 : \sigma^2_u = 0 \quad ; \quad H_1 : \sigma^2_u > 0$$

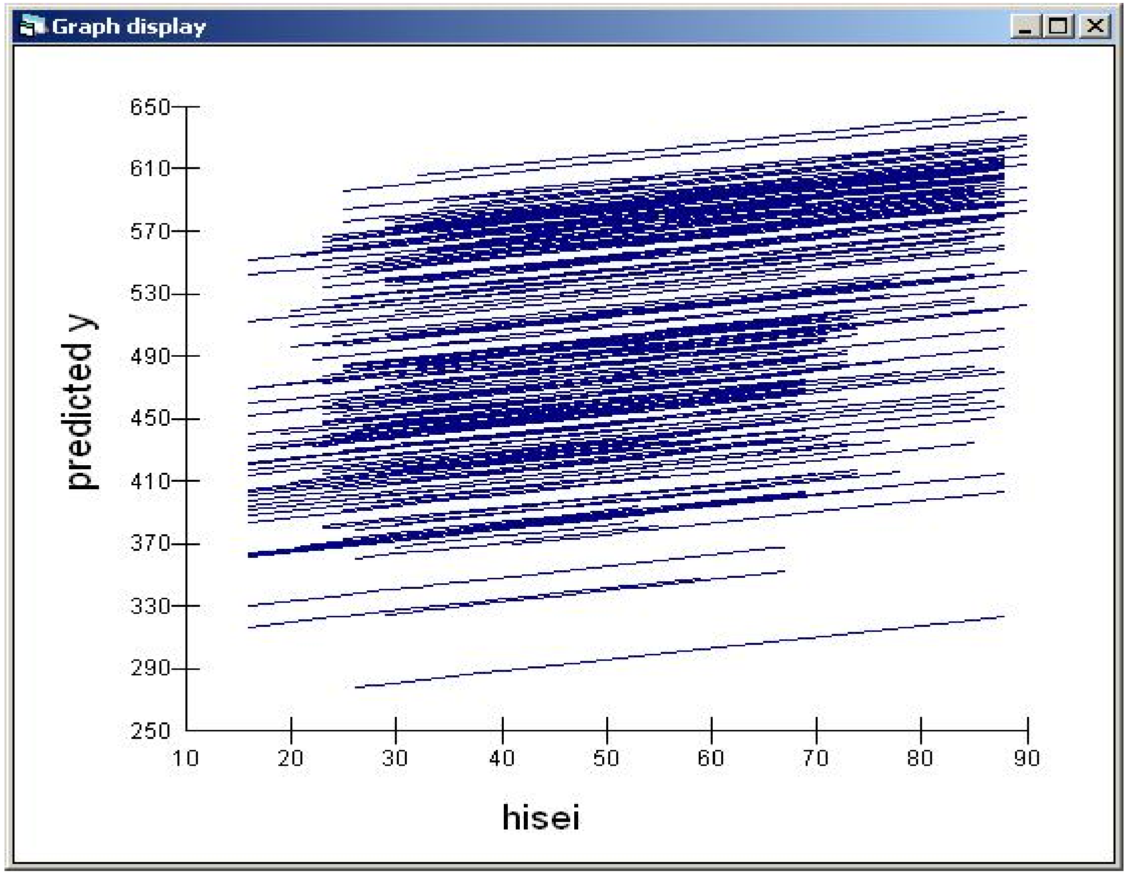
$$\chi^2_1 = 84.0 \quad ; \quad p < .001$$

Example 4 (cont.): Germany

Separate regression lines for sample schools

are due to variation in the intercept: $(\hat{\ }_{0}+\hat{u}_j)$

Regression lines are parallel: $(\ _{0}+u_j)+\ _{1}\times Hisei_{ij}$



Example 4 **(cont.):** How much does the model (or *hisei* in this case) explain of the variance of reading literacy score?

	Null model	Model with predictors	Variance explained %
Between-school variance	5541.8	4843.1	12.6
Between-student variance	4512.2	4438.7	1.6
Total variance	10054.0	9281.8	7.7

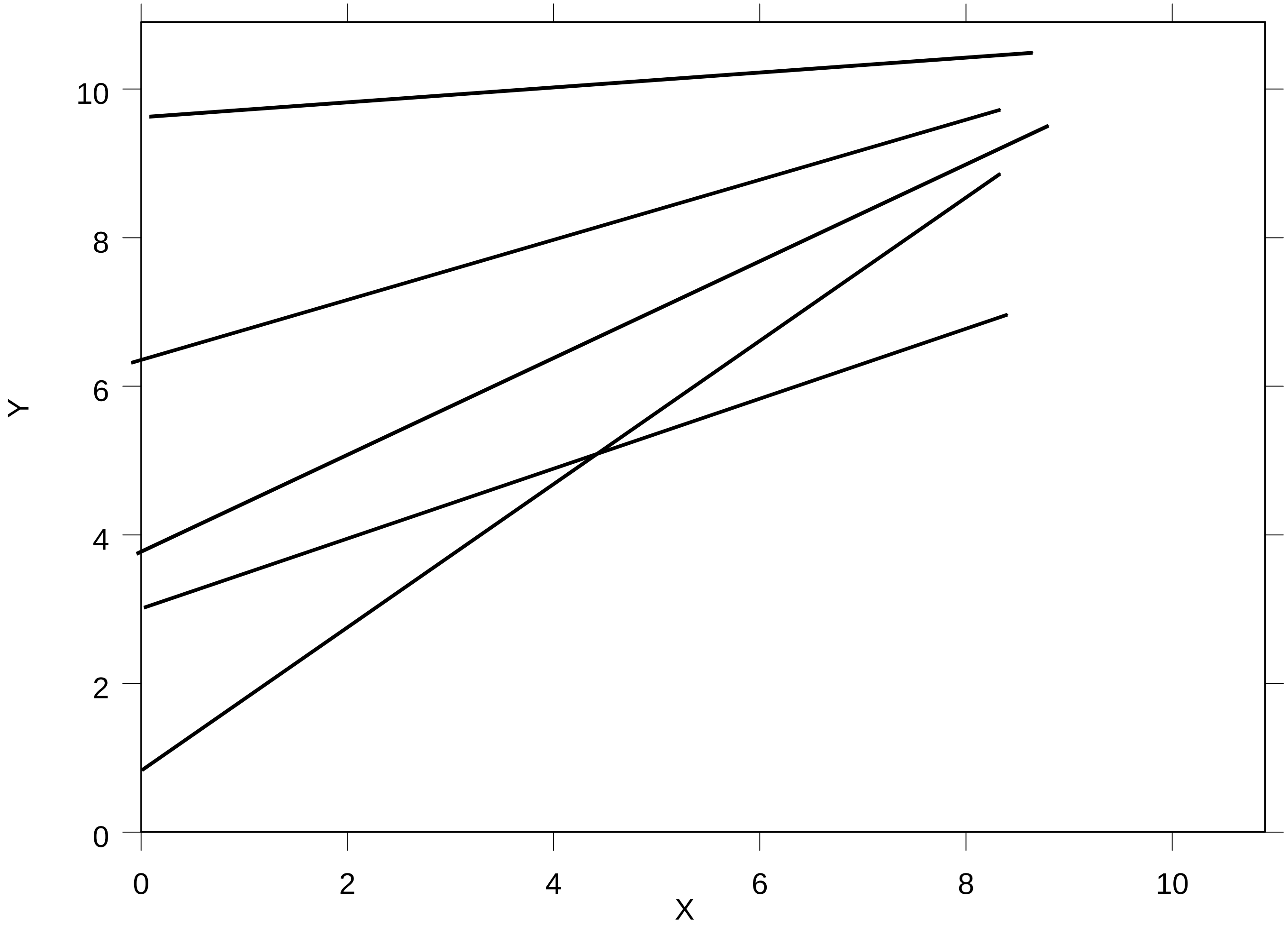
4. Random coefficient model

Antero Malin
Helsinki 22.-23.5.2014

Random coefficient model

- Slopes (regression coefficients) are varying between clusters
  - Regression lines are not parallel!
- Usually the intercept is also varying between clusters

Random coefficient model



Random coefficient model

$$y_{ij} = ({}_0 + u_{0j}) + ({}_1 + u_{1j}) \times x_{1ij} + e_{ij}$$

$$= {}_0 + {}_1 \times x_{1ij} + u_{0j} + u_{1j} \times x_{1ij} + e_{ij}$$

y_{ij} *is outcome variable*

i is level 1 unit, $i=1,\ldots,n_j$; j is level 2 unit, $j=1,\ldots,J$

${}_0$ is intercept of the model

${}_1$ is slope

e_{ij} is level 1 residual

u_{0j} is level 2 residual associated with intercept ${}_0$

u_{1j} is level 2 residual associated with slope ${}_1$

$$u_{0j} \sim N(0, \pm^2_{u0}); \quad u_{1j} \sim N(0, \pm^2_{u1}); \quad e_{ij} \sim N(0, \pm^2_{e})$$

Example 5: Finland

PISA 2000 data, Combined reading literacy score

$$y_{ij} = (\;_{0} + u_{0j}) + (\;_{1} + u_{1j}) \times Hisei_{1ij} + e_{ij}$$

$$= \;_{0} + \;_{1} \times Hisei_{1ij} + u_{0j} + u_{1j} \times Hisei_{1ij} + e_{ij}$$

Equations

$$\text{read1}_{ij} \sim \mathrm{N}(XB, \Omega)$$

$$\text{read1}_{ij} = \beta_{0ij}\text{intercept} + \beta_{1j}\text{hisei}_{ij}$$

$$\beta_{0ij} = 491.423(4.307) + u_{0j} + e_{0ij}$$

$$\beta_{1j} = 1.162(0.083) + u_{1j}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim \mathrm{N}(0, \Omega_u) : \Omega_u = \begin{bmatrix} 368.669(316.641) & \\ -3.886(5.884) & 0.158(0.120) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim \mathrm{N}(0, \Omega_e) : \Omega_e = \begin{bmatrix} 6635.175(140.186) \end{bmatrix}$$

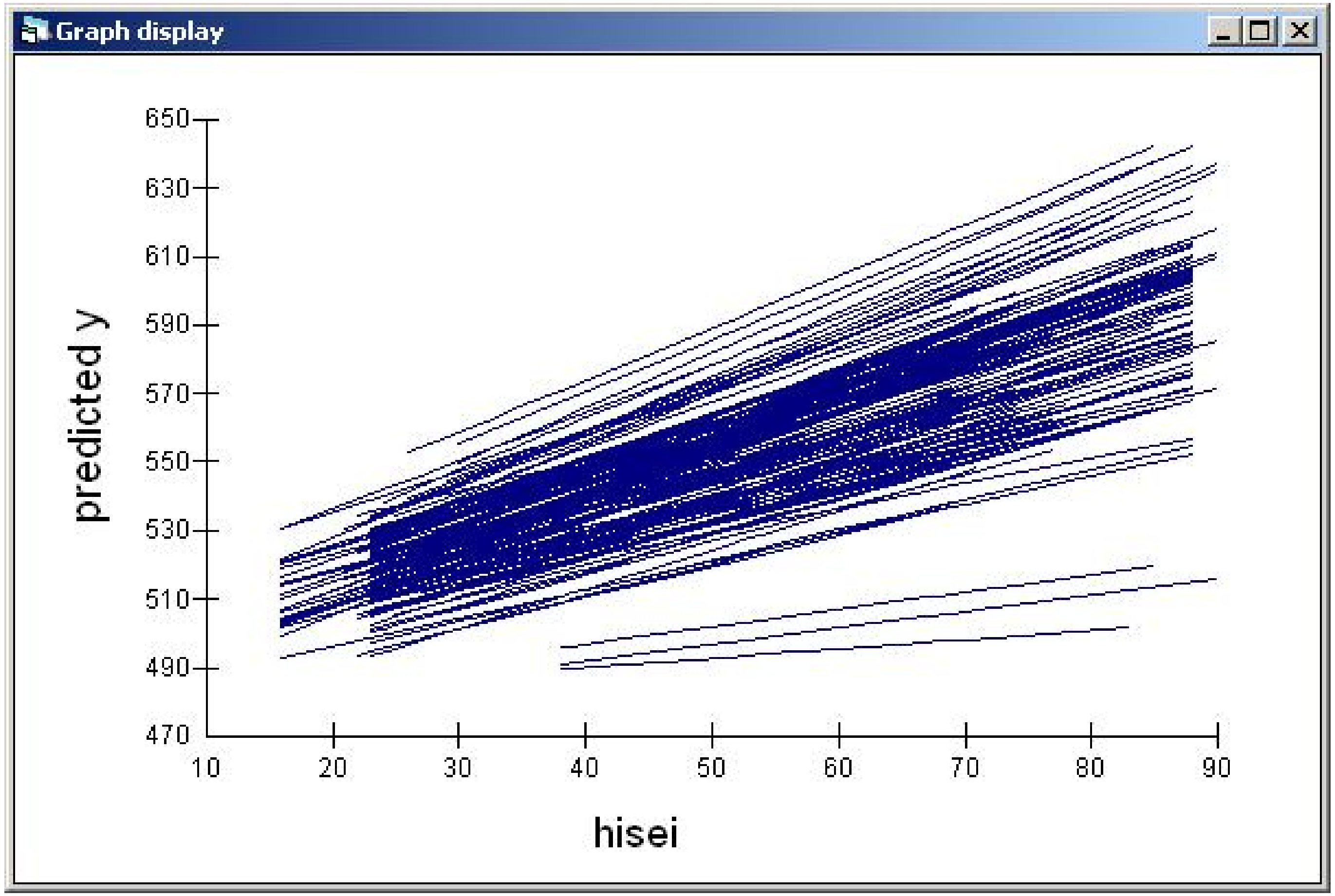
-2*loglikelihood(IGLS) = 55631.030(4765 of 4859 cases in use)

Fonts | Subs | Name | + | - | Add Term | Estimates | Nonlinear | Help | Clear

Example 5 (cont.): Finland

Separate regression lines of sample schools

$$\hat{y}_{ij} = (\hat{\ }_{0} + \hat{u}_{0j}) + (\hat{\ }_{1} + \hat{u}_{1j}) \times Hisei_{1ij}$$



Testing statistical significance

Example 5 **(cont.):** Finland

$$\hat{\sigma}^2_{u1} = 0.158$$

$$H_0 : \sigma^2_{u1} = 0 \quad ; \quad H_1 : \sigma^2_{u1} > 0$$

$$\chi^2_1 = 1.72 \quad ; \quad p = .190$$

$\Rightarrow$ Variance estimate is not statistically significant.

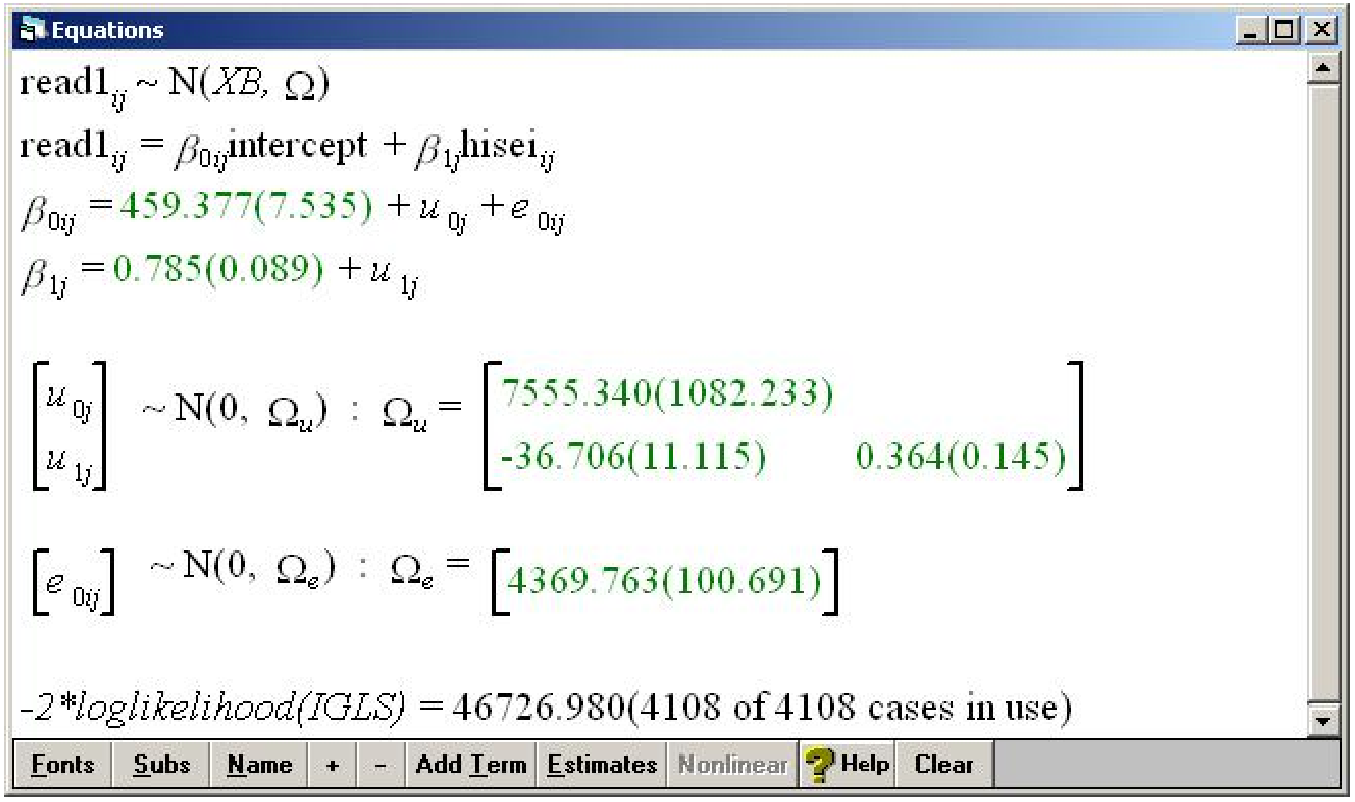
$\Rightarrow$ We leave it out of the model

Example 6: Germany

PISA 2000 data, Combined reading literacy score

$$y_{ij} = (\;{}_0 + u_{0j}) + (\;{}_1 + u_{1j}) \times Hisei_{1ij} + e_{ij}$$

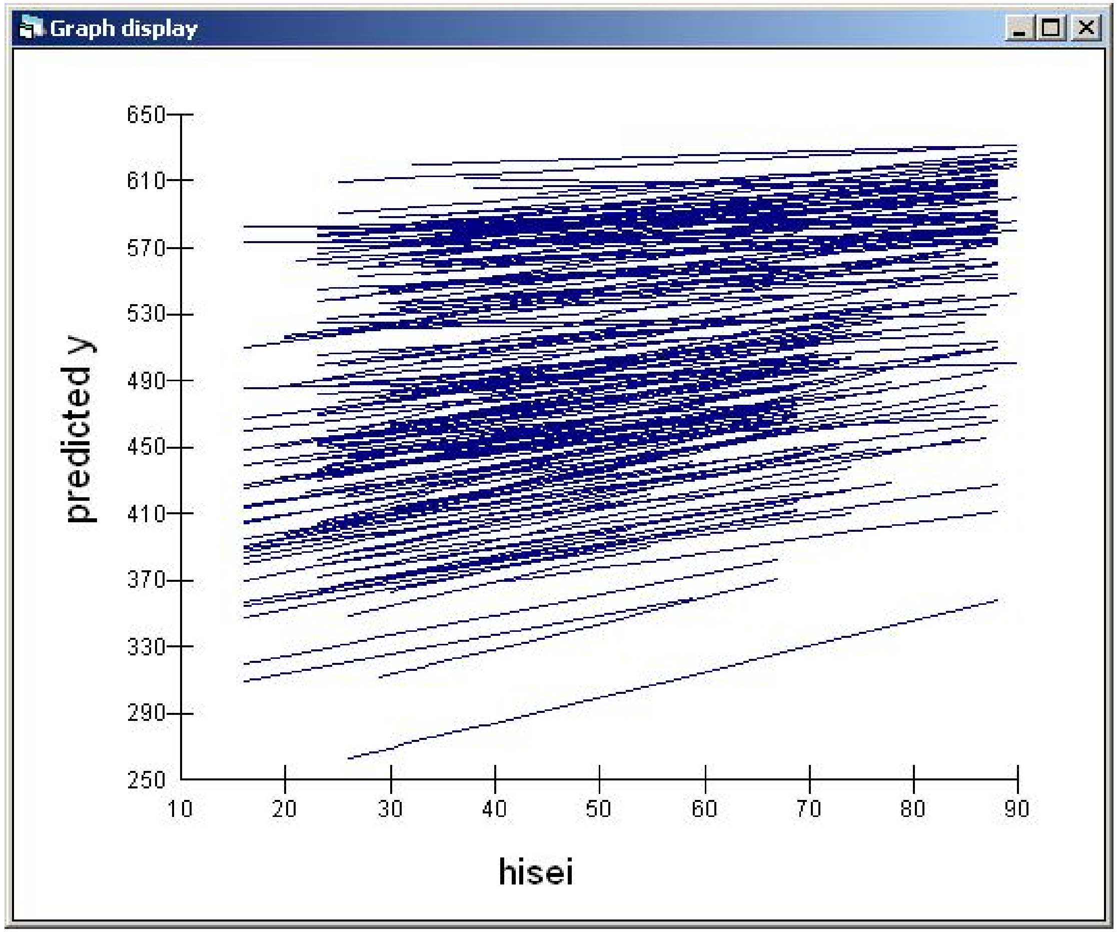
$$= \;{}_0 + \;{}_1 \times Hisei_{1ij} + u_{0j} + u_{1j} \times Hisei_{1ij} + e_{ij}$$



Example 6 (cont.): Germany

Separate regression lines of sample schools

$$\hat{y}_{ij} = (\hat{\ }_{0} + \hat{u}_{0j}) + (\hat{\ }_{1} + \hat{u}_{1j}) \times Hisei_{1ij}$$



Testing statistical significance

Example 6 **(cont.): Germany**

$$\hat{\sigma}^2_{u1} = 0.158 \quad and \quad \hat{\sigma}_{u01} = -36.7$$

$$H_0 : \sigma^2_{u1} = 0 \quad and \quad \sigma_{u01} = 0$$

$$\chi^2_2 = 78.2 \quad ; \quad p < .001$$

$\Rightarrow$ Variance estimates are statistically significant.

$\Rightarrow$ We keep them in the model

5.
Modelling contextual and compositional effect

Antero Malin
Helsinki 22.-23.5.2014

Group level variables:

1. Aggregated variables: Value of the group level variable is calculated from the individual values of group members, i.e. group mean

Interesting question: Is there group effect, in addition to the individual effect?

2. ”Real” group level variables, like shool’s georaphical location or school size

NOTE: The value of the group level variable is the same for all members of the group

Compositional/Contextual effect

Compositional variable:

- Group level variable
- Usually aggregated variable
- It measures an aspect of the composition of level 2 unit, i.e. of the school to which the individual students belong.

How the outcome of individuals is affected by their social contexts?

How the outcome of students is affected by the social composition of the school?

Example 7: The effect of socio-economic background in Germany

Compositional variable: School mean of Highest socio-economic index (*AHisei*)

Mean	49.9
SD	15.6
Min	29.2
Max	71.1
Schools	183

Statistical model

$$y_{ij} = \quad {}_0 + \quad {}_1 \times Hisei_{ij} + \text{¥}_1 \times AHisei_j + u_j + e_{ij}$$

y_{ij} is outcome variable

i is level 1 unit, $i=1,\dots,n_j$

j is level 2 unit, j=$1,\dots,J$

${}_0$ is intercept of the model

${}_1$ is individual level coefficient of socio-economic background

¥_1 is group level coefficient of socio-economic background

e_{ij} is level 1 residual

u_j is level 2 residual

$$u_j \sim N(0,\pm_u^2) \quad e_{ij} \sim N(0,\pm_e^2) \quad \mathrm{cov}(u_j e_{ij}) = 0$$

Example 7 (cont.): Germany

PISA 2000 data, Combined reading literacy score

$$y_{ij} = \;_{0} + \;_{1} \times Hisei_{ij} + ¥_{1} \times AHisei_{j} + u_{j} + e_{ij}$$

Equations

$$\text{read1}_{ij} \sim \text{N}(XB,\ \Omega)$$

$$\text{read1}_{ij} = \beta_{0ij}\text{intercept} + 0.628(0.077)\text{hisei}_{ij} + 7.366(0.385)\text{ahisei}_{j}$$

$$\beta_{0ij} = 102.468(18.847) + u_{0j} + e_{0ij}$$

$$\begin{bmatrix} u_{0j} \end{bmatrix} \sim \text{N}(0,\ \Omega_u)\ :\ \Omega_u = \begin{bmatrix} 1427.413(170.978) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim \text{N}(0,\ \Omega_e)\ :\ \Omega_e = \begin{bmatrix} 4444.690(100.333) \end{bmatrix}$$

$-2*loglikelihood(IGLS) = 46543.900$(4108 of 4108 cases in use)

Fonts | Subs | Name | + | - | Add Term | Estimates | Nonlinear | Help | Clear

Example 7 (cont.):

Statistical significance

$$\hat{¥}_1 = 7.366 \quad with\ s.e. = 0.385$$

$$H_0 : ¥_1 = 0 \quad ; \quad H_1 : ¥_1 \neq 0$$

$$p < 0.0001$$

=> Socio-economic background has statistically significant contextual effect

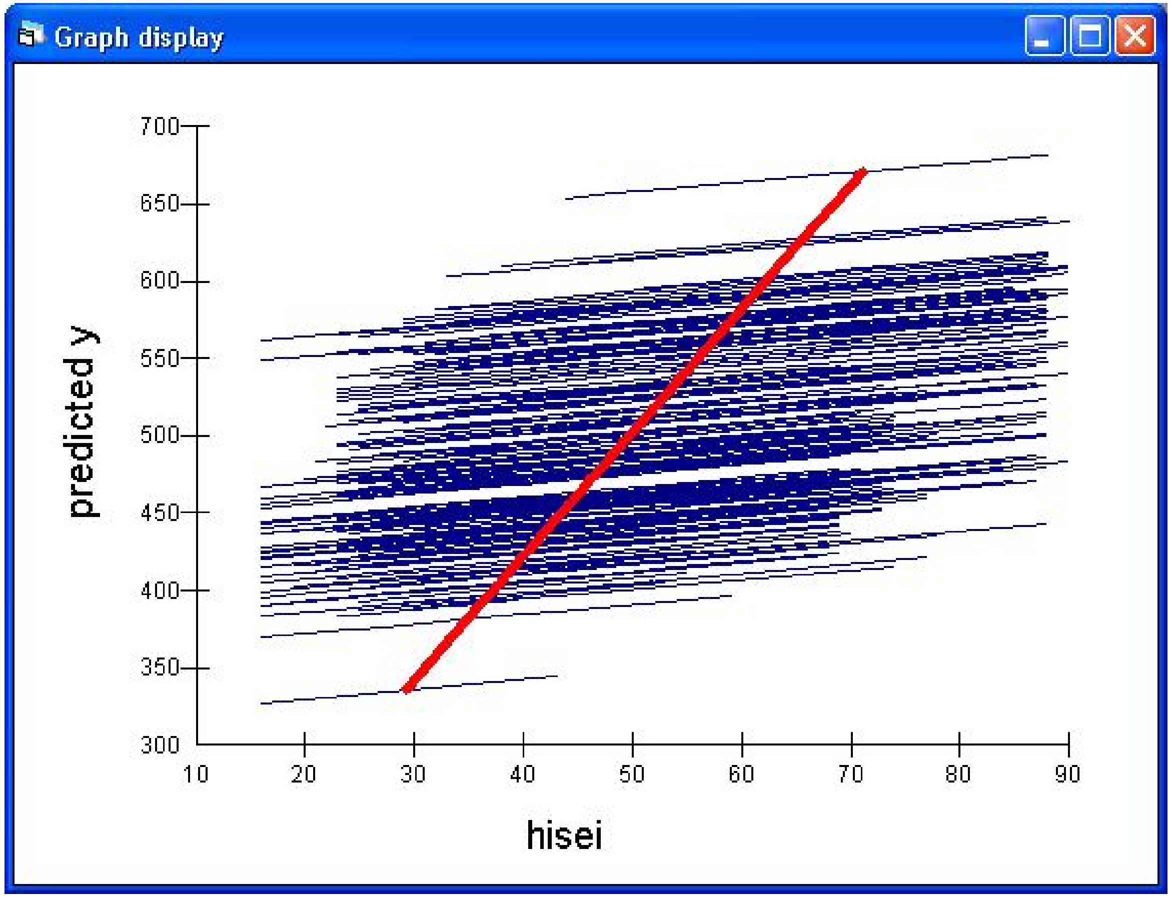
Example 7 (cont.): How much does the model (or *hisei* and *ahisei* in this case) explain of the variance of reading literacy score?

	Null model	Model with predictors	Variance explained %
Between-school variance	5541.8	1427.4	74.2
Between-student variance	4512.2	4444.7	1.5
Total variance	10054.0	5872.1	41.6

Example 7 (cont.): Germany

PISA 2000 data, Combined reading literacy score

Red line: Regression of school mean literacy score on shool mean *hisei* (*ahisei*)
Blue lines: Within-school regression lines of literacy score on *hisei*

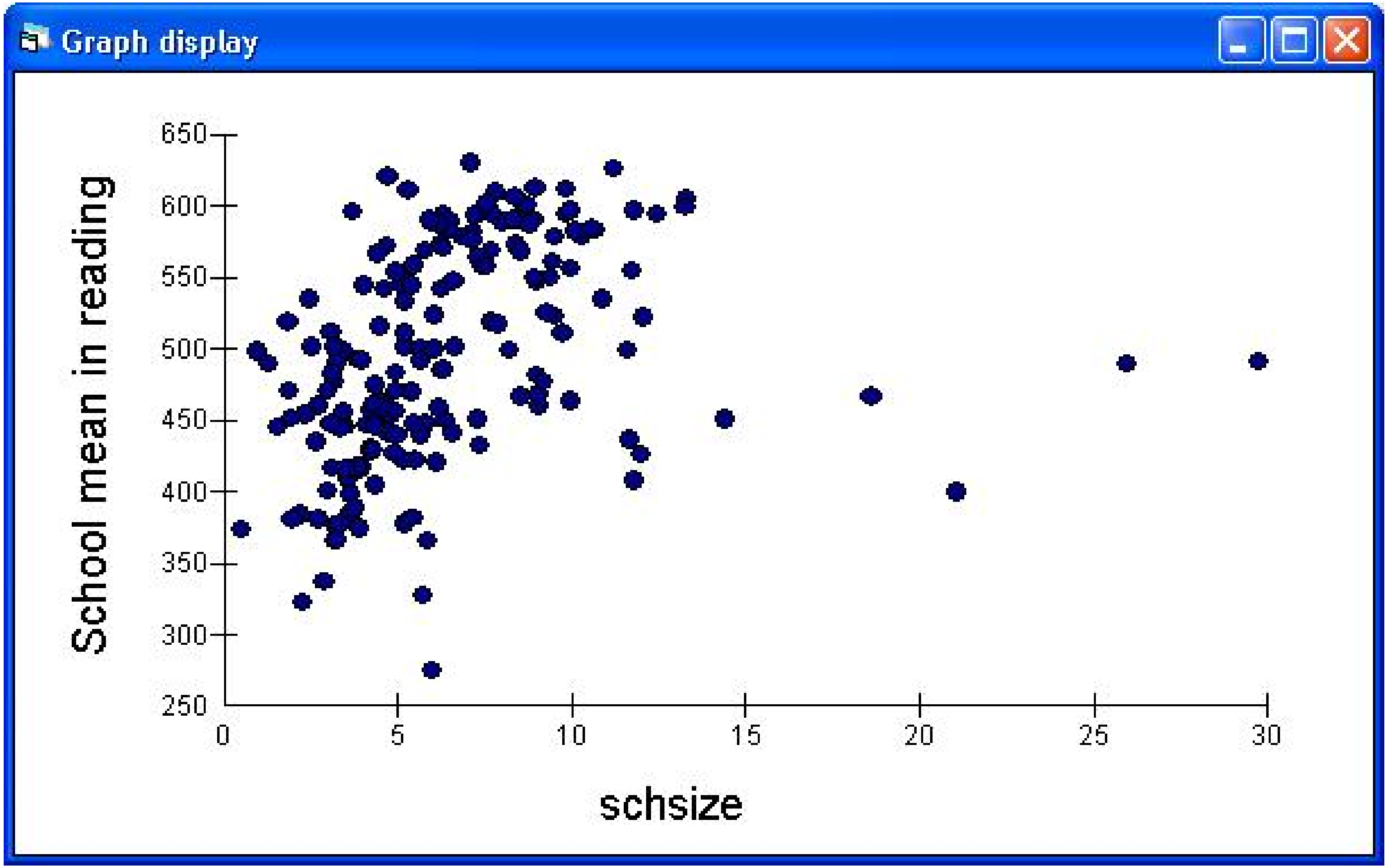


Example 8: Effect of school size in Germany

School size variable (Schsize) is scaled by dividing the number of students in each school by 100

Mean	6.79
SD	3.7
Min	0.5
Max	29.8
Schools	183

Example 8 (cont.): Effect of school size on literacy score



Example 8 (cont.): Effect of school size on literacy score

$$y_{ij} = \quad {}_0 + ¥_1 \times schsize_j + u_j + e_{ij}$$

Equations

$$\mathrm{read1}_{ij} \sim \mathrm{N}(XB,\ \Omega)$$

$$\mathrm{read1}_{ij} = \beta_{0ij}\mathrm{intercept} + 6.574(1.362)\mathrm{schsize}_j$$

$$\beta_{0ij} = 453.423(10.349) + u_{0j} + e_{0ij}$$

$$\left[u_{0j}\right] \sim \mathrm{N}(0,\ \Omega_u)\ :\ \Omega_u = \left[4884.445(532.614)\right]$$

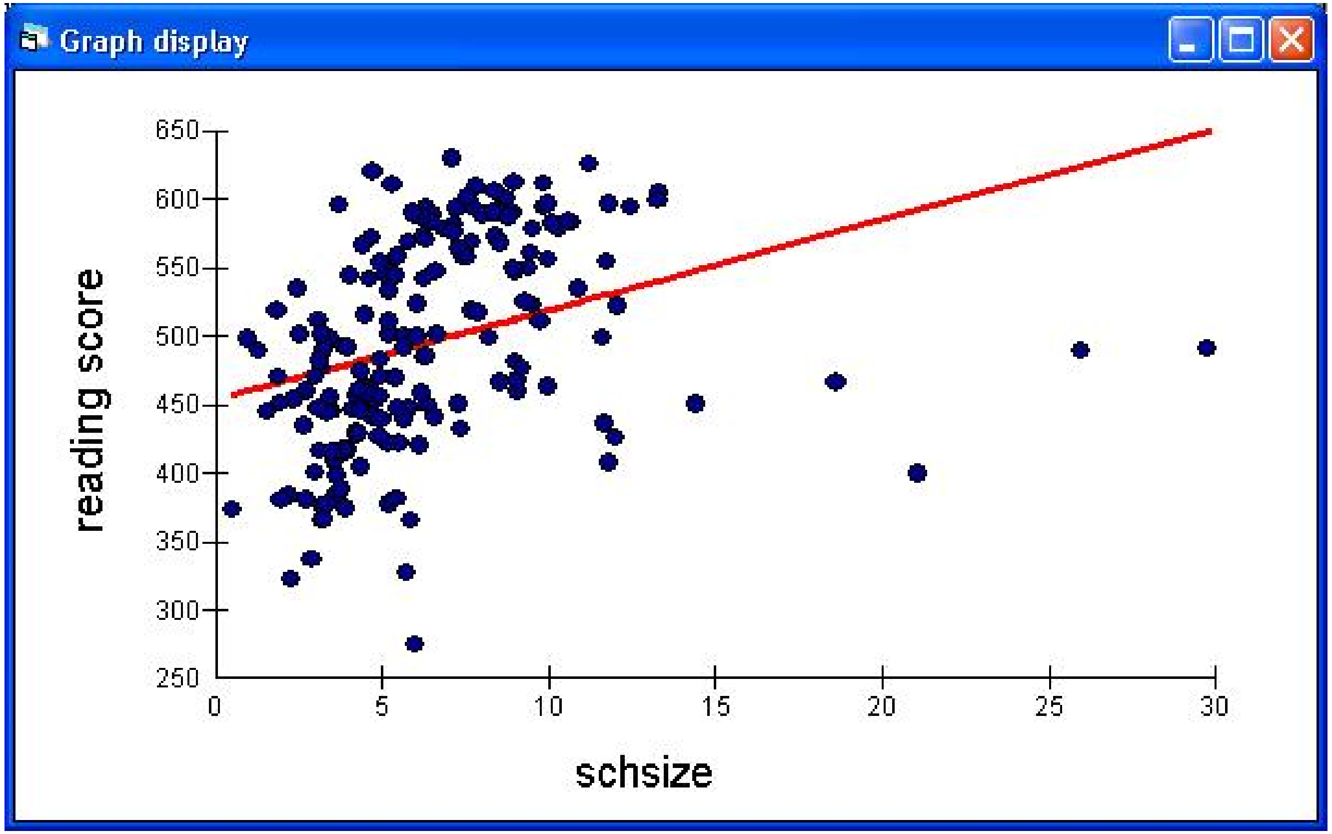
$$\left[e_{0ij}\right] \sim \mathrm{N}(0,\ \Omega_e)\ :\ \Omega_e = \left[4512.568(101.869)\right]$$

*-2*loglikelihood(IGLS)* = 46811.030(4108 of 4108 cases in use)

Fonts | Subs | Name | + | - | Add Term | Estimates | Nonlinear | Help | Clear

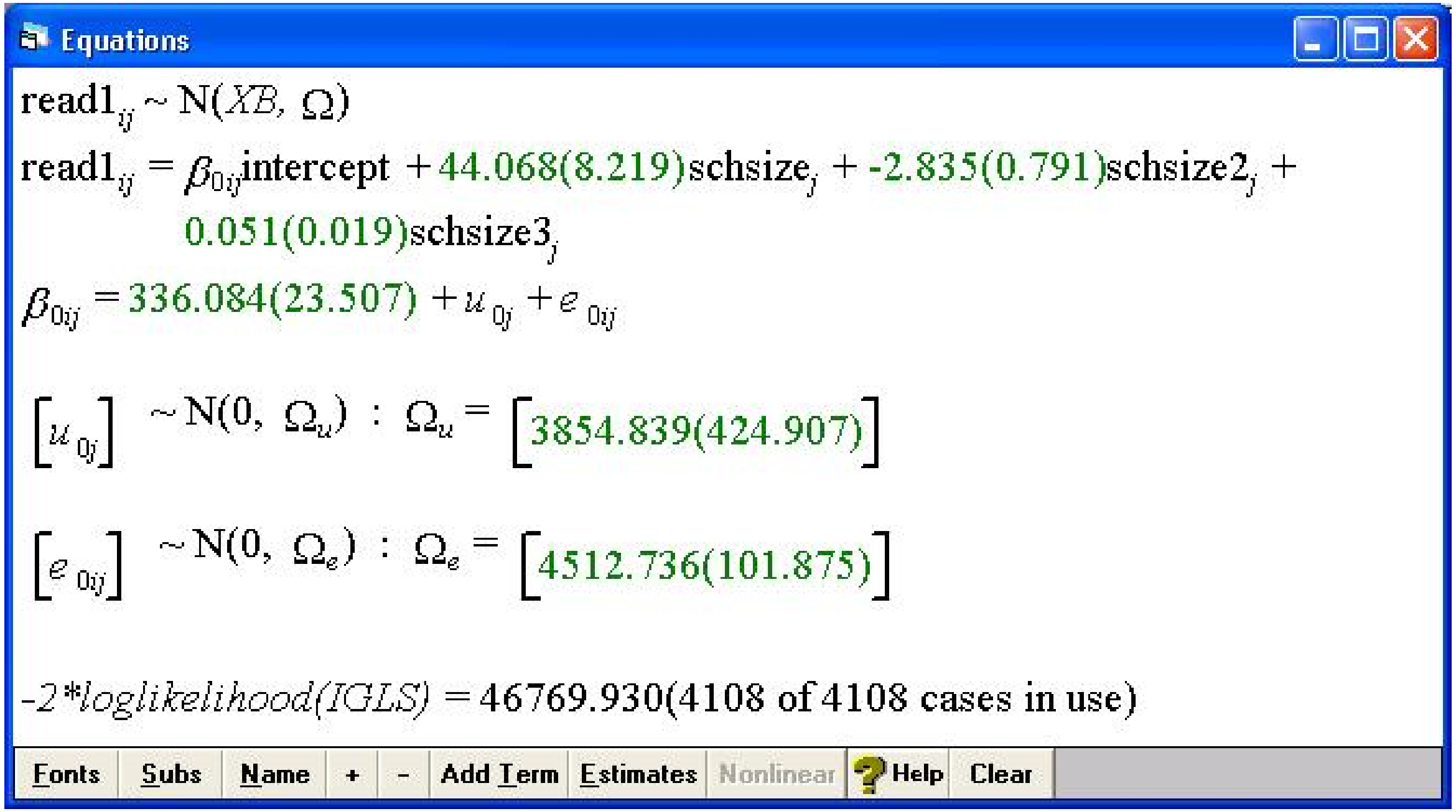
Example 8 (cont.): Effect of school size on literacy score

Red line: Model prediction
Blue dots: Empirical means



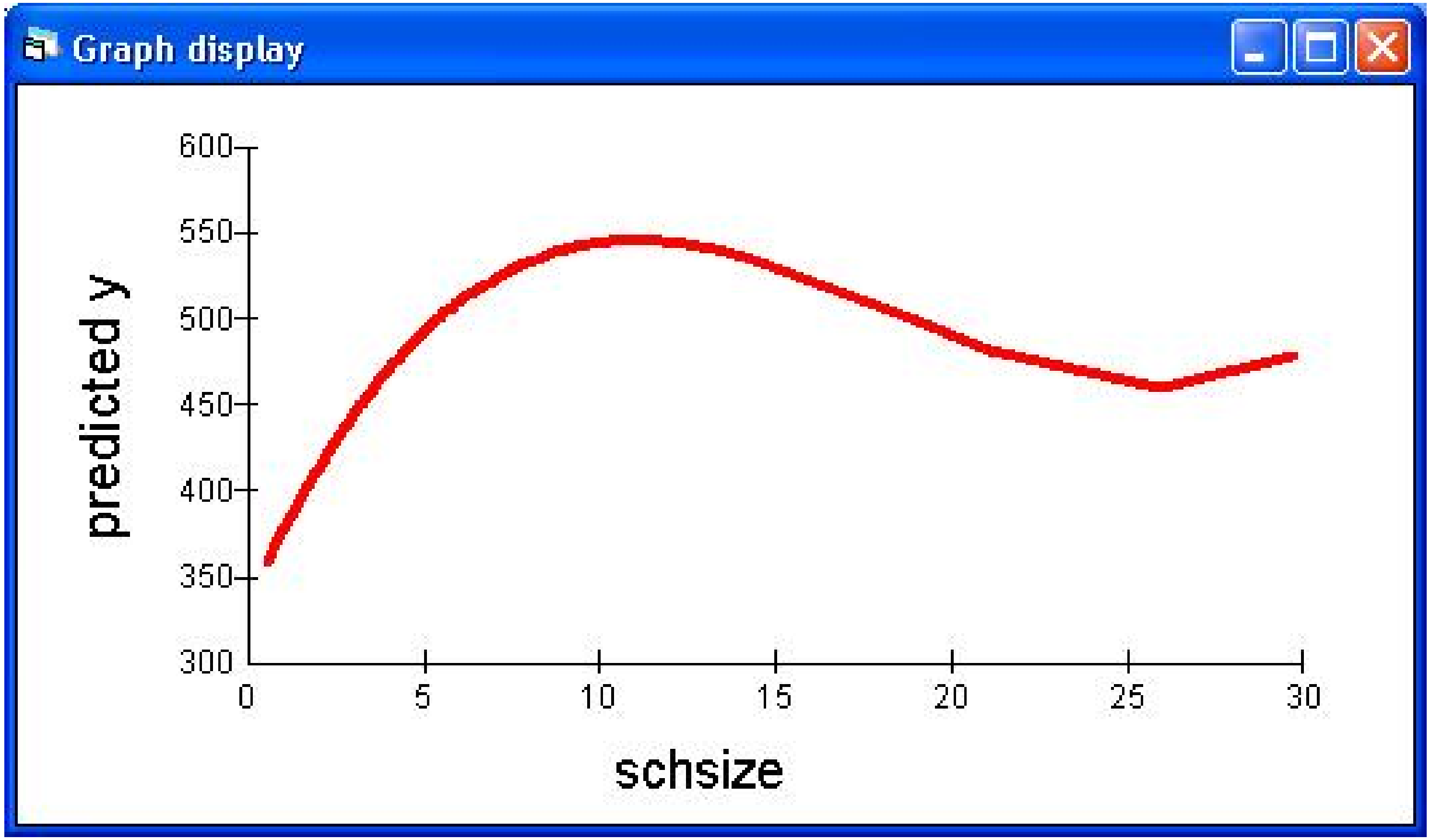
Example 8 (cont): Effect of school size on literacy score

$$y_{ij} = \quad {}_{0} + \text{¥}_1 \times schsize_j + \text{¥}_2 \times schsize_j^2 + \text{¥}_3 \times schsize_j^3 + u_j + e_{ij}$$



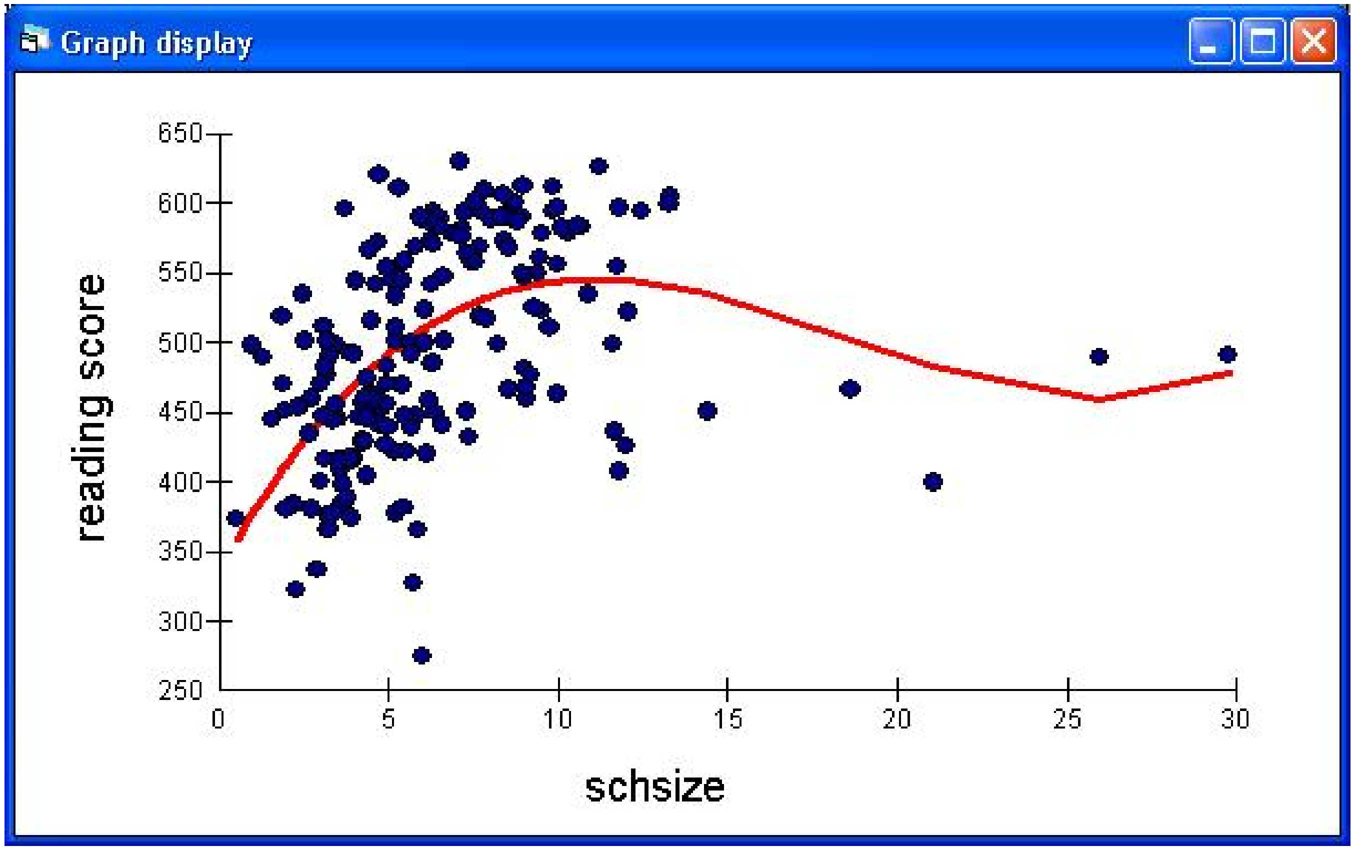
Example 8 (cont.): Effect of school size on literacy score

$$y_{ij} = \quad {}_0 + ¥_1 \times schsize_j + ¥_2 \times schsize_j^2 + ¥_3 \times schsize_j^3 + u_j + e_{ij}$$



Example 8 (cont.): Effect of school size on literacy score

Red line: Model prediction
Blue dots: Empirical means



6.

Modelling variance structure

Antero Malin

Helsinki 22.-23.5.2014

Modelling variance structure

- We can model the variation explicitly as a function of explanatory variables.
- Variance does not have to be constant

Two examples:

1. Separate variances for subgroups at level 1
2. Variance is a function of continuous level 1 explanatory variable

Example 9: Separate variances for subgroups

Gender difference in reading literacy in Germany

Statistics of reading literacy score:

	Girls	Boys	All
Mean	518	490	504
SD	99.0	96.8	97.9
Min	143	172	143
Max	779	765	779
Students	2104	2004	4108
Schools	183	183	183

Example 9 (cont.): Gender difference in reading literacy in Germany

$$y_{ij} = \quad {}_0 + \quad {}_1 \times female_{ij} + u_j + e_{ij}$$

Equations

$$\text{read1}_{ij} \sim \text{N}(XB,\ \Omega)$$

$$\text{read1}_{ij} = \beta_{0ij}\text{intercept} + 19.658(2.173)\text{female}_{ij}$$

$$\beta_{0ij} = 486.354(5.671) + u_{0j} + e_{0ij}$$

$$\begin{bmatrix} u_{0j} \end{bmatrix} \sim \text{N}(0,\ \Omega_u)\ :\ \Omega_u = \begin{bmatrix} 5450.360(591.499) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim \text{N}(0,\ \Omega_e)\ :\ \Omega_e = \begin{bmatrix} 4423.501(99.857) \end{bmatrix}$$

$$-2*loglikelihood(IGLS) = 46751.890(4108 \text{ of } 4108 \text{ cases in use})$$

Fonts | Subs | Name | + | - | Add Term | Estimates | Nonlinear | Help | Clear

Girls are better in reading than boys

Example 9 (cont.): We estimate separate variances for girls and boys at both levels

Variable *female*=1 if girl, 0 if boy

Variable *male*=1 if boy, 0 if girl

$$y_{ij} = \;{}_0 + \;{}_1 \times female_{ij} + (u_{1j} + e_{1ij}) \times female + (u_{2j} + e_{2ij}) \times male$$

$$\text{Total variance for girls}: \; \sigma^2_{u1} + \sigma^2_{e1}$$

$$\text{Total variance for boys}: \; \sigma^2_{u2} + \sigma^2_{e2}$$

Example 9 (cont.): Separate variances for girls and boys at both levels

$$y_{ij} = {}_0 + {}_1 \times female_{ij} + (u_{1j} + e_{1ij}) \times female + (u_{2j} + e_{2ij}) \times male$$

Equations

$$\text{read1}_{ij} \sim \text{N}(XB,\ \Omega)$$

$$\text{read1}_{ij} = 486.477(5.493)\text{intercept} + \beta_{1ij}\text{female}_{ij} + e_{2ij}\text{male}_{ij} + u_{2j}\text{male}_{ij}$$

$$\beta_{1ij} = 19.196(2.545) + u_{1j} + e_{1ij}$$

$$\begin{bmatrix} u_{1j} \\ u_{2j} \end{bmatrix} \sim \text{N}(0,\ \Omega_u)\ :\ \Omega_u = \begin{bmatrix} 5949.436(664.712) & \\ 5351.134(590.129) & 5065.300(576.993) \end{bmatrix}$$

$$\begin{bmatrix} e_{1ij} \\ e_{2ij} \end{bmatrix} \sim \text{N}(0,\ \Omega_e)\ :\ \Omega_e = \begin{bmatrix} 4145.215(133.541) & \\ 0 & 4565.706(150.988) \end{bmatrix}$$

*-2*loglikelihood(IGLS)* = 46732.900(4108 of 4108 cases in use)

Fonts | Subs | Name | + | - | Add Term | Estimates | Nonlinear | Help | Clear

Example 9 (cont.): Testing the equality of variances

Between-school variances:

$$\hat{\sigma}_{u1}^2 = \hat{\sigma}_{uF}^2 = 5949.4 \quad ; \quad \hat{\sigma}_{u2}^2 = \hat{\sigma}_{uM}^2 = 5065.3$$

$$H_0 : \sigma_{uF}^2 = \sigma_{uM}^2 \quad ; \quad H_1 : \sigma_{uF}^2 \neq \sigma_{uM}^2$$

$$\chi_1^2 = 5.23 \quad ; \quad p = .022$$

Between-student variances within schools:

$$\hat{\sigma}_{e1}^2 = \hat{\sigma}_{eF}^2 = 4145.2 \quad ; \quad \hat{\sigma}_{e2}^2 = \hat{\sigma}_{eM}^2 = 4565.7$$

$$H_0 : \sigma_{eF}^2 = \sigma_{eM}^2 \quad ; \quad H_1 : \sigma_{eF}^2 \neq \sigma_{eM}^2$$

$$\chi_1^2 = 4.35 \quad ; \quad p = .037$$

Results: Variances are not equal

Example 10: Germany Variance is a function of continuous level 1 explanatory variable

Level 2 variance depends on the individual level predictor *hisei*

Statistics of *hisei*:

Mean	49.9
SD	15.6
Min	16
Max	90
n	4108

Statistical model:

$$
\begin{aligned}
y_{ij} &= {}_0 + ({}_1 + u_{1j}) \times Hisei_{ij} + ¥_1 \times AHisei_j + u_{0j} + e_{0ij} \\
&= {}_0 + {}_1 \times Hisei_{ij} + ¥_1 \times AHisei_j + u_{0j} + u_{1j} \times Hisei_{ij} + e_{0ij} \\
&= {}_0 + {}_1 \times Hisei_{ij} + ¥_1 \times AHisei_j + (u_{0j} + u_{1j} \times Hisei_{ij}) + e_{0ij}
\end{aligned}
$$

Level 2 variance function:

$$
\begin{aligned}
\pm_u^2 &= \mathrm{var}(u_{0j} + u_{1j} \times Hisei_{ij}) \\
&= \pm_{u0}^2 + 2 \times \pm_{u0} \times hisei_{ij} + \pm_{u1}^2 \times hisei_{ij}^2
\end{aligned}
$$

Example 10 (cont): Level 2 variance depends on the individual level predictor *hisei*

$$y_{ij} = \quad _0 + (\quad _1 + u_{1j}) \times Hisei_{ij} + \text{¥}_1 \times AHisei_j + u_{0j} + e_{0ij}$$

Equations

$$\text{read1}_{ij} \sim \mathrm{N}(XB,\ \Omega)$$

$$\text{read1}_{ij} = \beta_{0ij}\text{intercept} + \beta_{1j}\text{hisei}_{ij} + 7.260(0.381)\text{ahisei}_j$$

$$\beta_{0ij} = 107.323(18.852) + u_{0j} + e_{0ij}$$

$$\beta_{1j} = 0.646(0.089) + u_{1j}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_u)\ :\ \Omega_u = \begin{bmatrix} 2908.145(588.967) & \\ -23.811(8.530) & 0.351(0.141) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_e)\ :\ \Omega_e = \begin{bmatrix} 4376.907(100.808) \end{bmatrix}$$

$-2{*}loglikelihood(IGLS) = 46531.250$(4108 of 4108 cases in use)

Fonts | Subs | Name | + | - | Add Term | Estimates | Nonlinear | Help | Clear

Example 10 (cont):

Testing statistical significance of variance component:

$$\hat{\sigma}^2_{u1} = 0.351$$

$$H_0 : \sigma^2_{u1} = 0 \quad ; \quad H_1 : \sigma^2_{u1} > 0$$

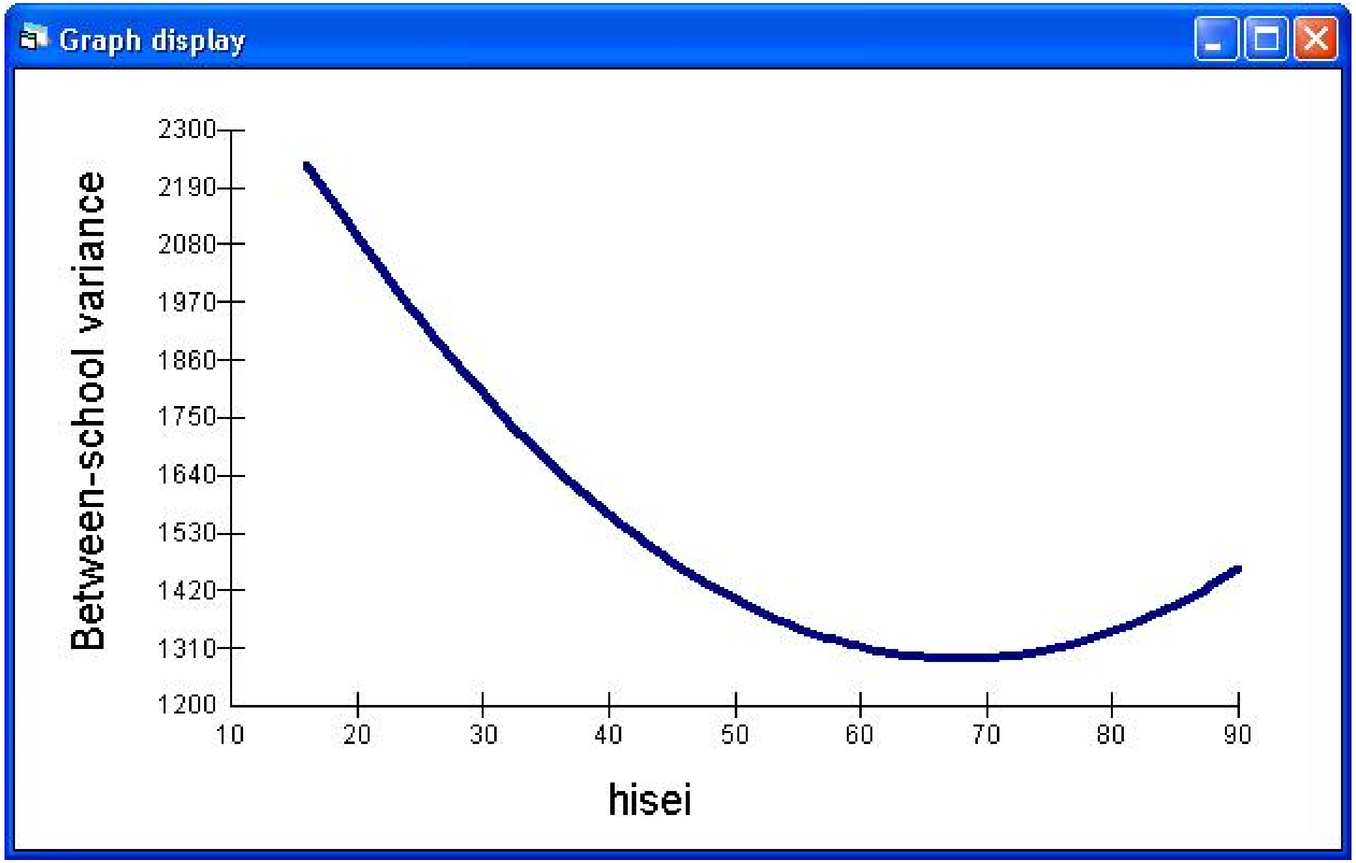
$$\chi^2_1 = 6.14 \quad ; \quad p = .013$$

Result: Variance component is statistically significantly different from 0

Example 10 (cont): Level 2 variance depends on the individual level predictor *hisei*

$$\hat{\sigma}_u^2 = \hat{\sigma}_{u0}^2 + 2 \times \hat{\sigma}_{u0} \times hisei_{ij} + \hat{\sigma}_{u1}^2 \times hisei_{ij}^2$$

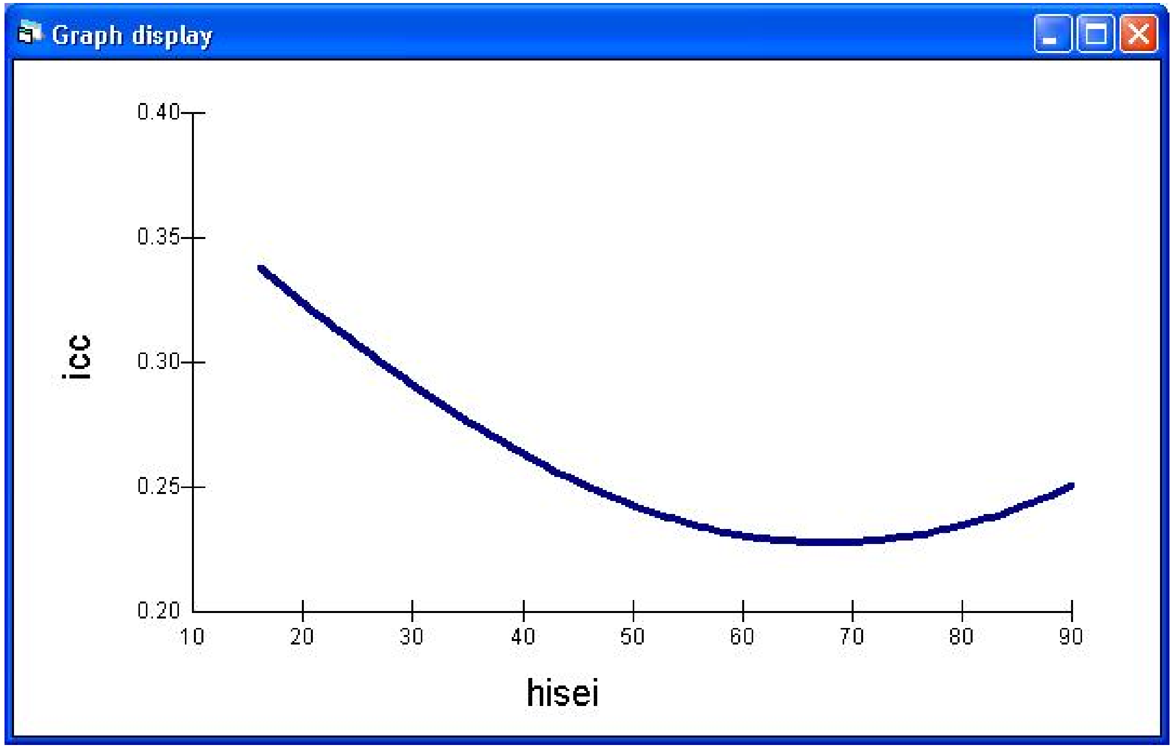
$$= 2908.1 - 2 \times 23.8 \times hisei_{ij} + 0.351 \times hisei_{ij}^2$$



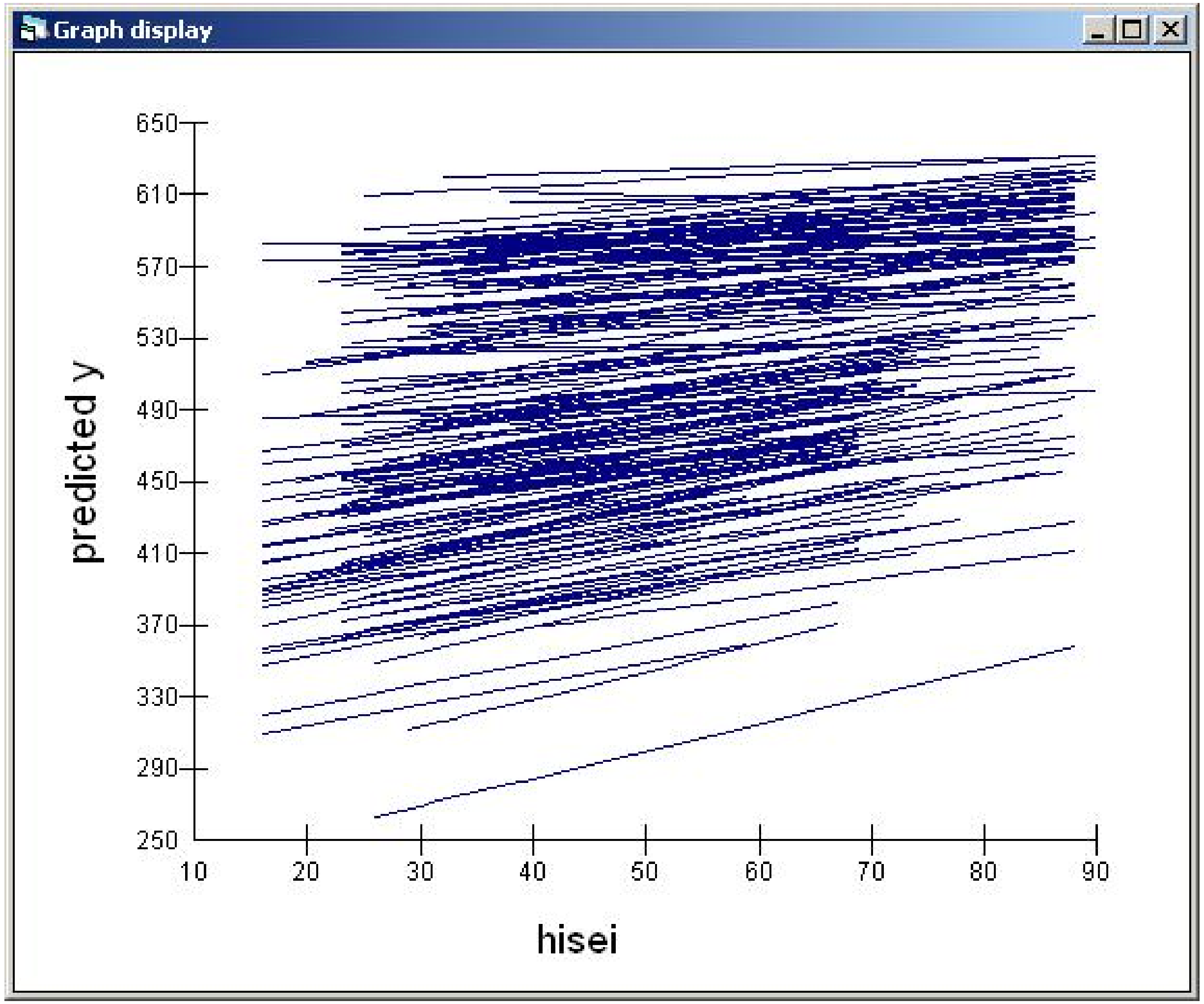
Example 10 (cont.): ICC (or VPC, and now conditional) is a function of *hisei*

$$\hat{\rho} = \frac{\hat{\sigma}_{u0}^2 + \hat{\sigma}_{u01} x_{ij} + \hat{\sigma}_{u1}^2 x_{ij}^2}{\hat{\sigma}_{u0}^2 + \hat{\sigma}_{u01} x_{ij} + \hat{\sigma}_{u1}^2 x_{ij}^2 + \hat{\sigma}_{e}^2}$$

$$= \frac{2908.1 - 23.8 \times hisei_{ij} + 0.351 \times hisei_{ij}^2}{2908.1 - 23.8 \times hisei_{ij} + 0.351 \times hisei_{ij}^2 + 4376.9}$$



Example 10 (cont.): Within-school regression lines in Germany



7.

Interaction models

Antero Malin
Helsinki 22.-23.5.2014

2 x 2 ANOVA model

Example 11: Is there any interaction between gender and place of residence in Finland?

Predictors:
Female: Boys =0, Girls = 1
Urban: Rural areas = 0, Urban areas= 1

Empirical mean reading score:

	Urban = 1	Rural = 0	All
Girls = 1	574	574	574
Boys =0	527	509	522
All	551	542	549

Example 11 (cont.): Is there any interaction between gender and place of residence in Finland?

Model with only main effects:

$$\mathrm{read1}_{ij} \sim \mathrm{N}(XB,\ \Omega)$$

$$\mathrm{read1}_{ij} = \beta_{0ij}\mathrm{intercept} + 52.115(2.326)\mathrm{female\_1}_{ij} + 8.628(5.012)\mathrm{urban\_1}_{j}$$

$$\beta_{0ij} = 515.401(4.521) + u_{0j} + e_{0ij}$$

$$\begin{bmatrix} u_{0j} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_u)\ :\ \Omega_u = \begin{bmatrix} 504.700(80.820) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_e)\ :\ \Omega_e = \begin{bmatrix} 6402.254(132.001) \end{bmatrix}$$

$-2{*}loglikelihood(IGLS\ Deviance)$ = 56567.216(4859 of 4859 cases in use)

Gender difference is statistically significant.

Place of residence is not statistically significant.

Example 11 (cont.): Interaction model, all effects are statistically significant

$$\text{read1}_{ij} \sim \mathrm{N}(XB, \Omega)$$

$$\text{read1}_{ij} = \beta_{0ij}\text{intercept} + 64.494(4.675)\text{female\_1}_{ij} + 17.034(5.708)\text{urban\_1}_{j} + -16.442(5.388)\text{female\_1.urban\_1}_{ij}$$

$$\beta_{0ij} = 509.098(4.960) + u_{0j} + e_{0ij}$$

$$\begin{bmatrix} u_{0j} \end{bmatrix} \sim \mathrm{N}(0, \Omega_u) : \Omega_u = \begin{bmatrix} 501.286(80.388) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim \mathrm{N}(0, \Omega_e) : \Omega_e = \begin{bmatrix} 6390.725(131.763) \end{bmatrix}$$

*-2*loglikelihood(IGLS Deviance)* = 56557.914(4859 of 4859 cases in use)

Model predictions:

Rural boys: 509.098

Rural girls: $509.098 + 64.494 = 573.592$

Urban boys: $509.098 + 17.034 = 526.132$

Urban girls: $509.098 + 64.494 + 17.034 - 16.442 = 574.184$

Example 12: Is there gender difference in the effect of *hisei* in Finland? Interaction of categorical and continuous variables.

$$\text{read1}_{ij} \sim \mathrm{N}(XB, \Omega)$$

$$\text{read1}_{ij} = \beta_{0ij}\text{intercept} + 60.020(7.359)\text{female\_1}_{ij} + 1.274(0.103)\text{hisei}_{ij} + -0.160(0.140)\text{female\_1.hisei}_{ij}$$

$$\beta_{0ij} = 458.946(5.696) + u_{0j} + e_{0ij}$$

$$\begin{bmatrix} u_{0j} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_u) \ : \ \Omega_u = \begin{bmatrix} 432.483(71.401) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_e) \ : \ \Omega_e = \begin{bmatrix} 5978.839(124.526) \end{bmatrix}$$

*-2*loglikelihood(IGLS Deviance)* = 55139.151(4765 of 4859 cases in use)

Coefficient estimates:

Effect of *hisei* for boys: 1.274

Effect of *hisei* for girls: 1.274 - 0.160 = 1.114

However: Difference between boys' and girls' estimates, i.e. -0.160, is *not* statistically significant!

About interactions:

1. Both variables are level 1 predictors
2. Both variables are level 2 predictors
3. One variable is level 1 predictor and another is level 2 predictor
4. Interaction of more than 2 factors

8.

Logistic regression model

Antero Malin
Helsinki 22.-23.5.2014

In logistic regression model:

- Reponse variable y_{ij} is binary (0 or 1)
- Predictors are continuous or categorical
- Level 1 variation is binomially distributed
- Level 2 variation is normally distributed

Outcome variable in the statistical model is logit-transformation of the probability ®_{ij} that the response variable y_{ij} is 1:

$$\text{logit}(\text{®}_{ij}) = \log(\text{®}_{ij}/(1-\text{®}_{ij}))$$

$$= \;{}_0 + \;{}_1 \times x_{ij} + \ldots + u_j$$

$$y_{ij} \sim \text{Bin}(\text{®}_{ij}, n_{ij})$$

$$\text{var}(y_{ij} \mid \text{®}_{ij}) = \text{®}_{ij}(1-\text{®}_{ij})/n_{ij}$$

When we have estimated the model, we can calculate the propabilities for each individual on belonging to group with $y_{ij} = 1$

Distributional assumption at level 1 can be:

- Binomial
- Extra Binomial, when Binomial distribution assumption is relaxed. Use this if necessary!

Linearisation:

- 1st order approximation in the Taylor expansion
- 2nd order approximation in the Taylor expansion

  This one gives greater accuracy. Use this!

Estimation type:

- MQL – marginal quasi-likelihood

  This tend to underestimate the values of fixed and random parameters , especially when n_{ij} is small
- PQL – predictive quasi-likelihood. Use this!

Example 13, German data

Read1bin is dichotomous variable:

Read1bin	Description	N
1	Good readers (about 34 %)	1412
0	Other readers (about 66 %)	2696

Read1bin = 1 if read1 > 550
Read1bin = 0 if read1 < 550

Binomial distribution, 1st order linearisation, MQL

$$\text{read1bin}_{ij} \sim \text{Binomial}(\text{denom}_{ij}, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = \beta_{0j}\text{intercept} + 0.666(0.084)\text{female\_1}_{ij} + 0.015(0.003)\text{hisei}_{ij} + 0.170(0.011)\text{ahisei}_{j}$$

$$\beta_{0j} = -10.594(0.536) + u_{0j}$$

$$\begin{bmatrix} u_{0j} \end{bmatrix} \sim \text{N}(0,\ \Omega_u)\ :\ \Omega_u = \begin{bmatrix} 0.665(0.106) \end{bmatrix}$$

$$\text{var}(\text{read1bin}_{ij} | \pi_{ij}) = \pi_{ij}(1 - \pi_{ij})/\text{denom}_{ij}$$

Binomial distribution, 2nd order linearisation, PQL

$$\text{read1bin}_{ij} \sim \text{Binomial}(\text{denom}_{ij},\ \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = \beta_{0j}\text{intercept} + 0.746(0.090)\text{female\_1}_{ij} + 0.016(0.003)\text{hisei}_{ij} + 0.192(0.012)\text{ahisei}_{j}$$

$$\beta_{0j} = -11.943(0.590) + u_{0j}$$

$$\begin{bmatrix} u_{0j} \end{bmatrix} \sim \text{N}(0,\ \Omega_u)\ :\ \Omega_u = \begin{bmatrix} 0.743(0.122) \end{bmatrix}$$

$$\text{var}(\text{read1bin}_{ij}|\pi_{ij}) = \pi_{ij}(1 - \pi_{ij})/\text{denom}_{ij}$$

Extra binomial distribution, 2nd order linearisation, PQL

$$\text{read1bin}_{ij} \sim \text{Binomial}(\text{denom}_{ij}, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = \beta_{0j}\text{intercept} + 0.747(0.086)\text{female\_1}_{ij} + 0.016(0.003)\text{hisei}_{ij} + 0.193(0.011)\text{ahisei}_{j}$$

$$\beta_{0j} = -12.004(0.589) + u_{0j}$$

$$\begin{bmatrix} u_{0j} \end{bmatrix} \sim \text{N}(0,\ \Omega_u) \ :\ \Omega_u = \begin{bmatrix} 0.785(0.124) \end{bmatrix}$$

$$\text{var}(\text{read1bin}_{ij} | \pi_{ij}) = 0.907(0.020)\pi_{ij}(1 - \pi_{ij})/\text{denom}_{ij}$$

Literature

Bryk, A., & Raudenbush, S. (1992) *Hierarchical Linear Models: Application and Data Analysis Methods*. Newbury Park: Sage Publications.

Goldstein, H. (1995) *Multilevel Statistical Models*. London: Edward Arnold. [http://www.cmm.bristol.ac.uk/team/HG_Personal/multbook1995.pdf]

Goldstein, H. (2003) *Multilevel Statistical Models*. Third Edition. London: Edward Arnold.

Goldstein, H. (2011) *Multilevel Statistical Models*. 4th Edition. Wiley Series in Probability and Statistics.

Hox, J. J. (1995) *Applied Multilevel Analysis*. Amsterdam: TT-Publikaties. [http://www.soziologie.uni-halle.de/langer/multilevel/books/hox95mla.pdf]

Snijders, T. & Bosker, R. (2002). *Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage Publications.