Introduction to Multilevel Modelling and the software *MLwiN*

Lecture at China National Institute for Educational Research, China
7 March 2008

Wen-Jung Peng
Graduate School of Education
University of Bristol

Review some concepts

The data set used to address the related issues in this lecture – *MLwiN* tutorial sample

(see Rasbash, et al, 2005 – A User's Guide to *MLwiN*)

Basic information of the data set

1 school | Refresh | Categories | Help

	Name	n	missing	min	max
1	school	4059	0	1	65
2	student	4059	0	1	198
3	normexam	4059	0	-3.666072	3.666091
4	cons	4059	0	1	1
5	standlrt	4059	0	-2.934953	3.015952
6	girl	4059	0	0	1
7	schgend	4059	0	1	3
8	avslrt	4059	0	-0.7559605	0.6376559
9	schav	4059	0	1	3
10	vrband	4059	0	1	3

Data source: *MLwiN* tutorial sample

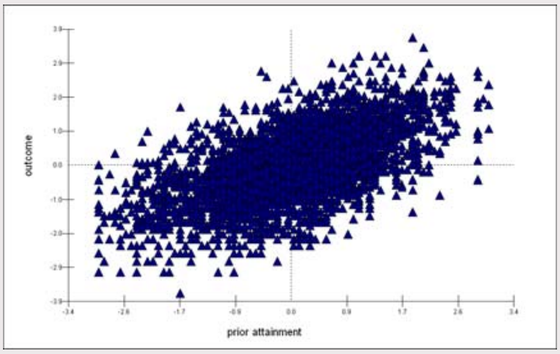
Number of schools: 65
Number of pupils: 4059

Normexam: pupil's exam score at age 16

Standlrt: pupil's score at age 11 on the London Reading Test

We are interested in

Q1: **The relationship between 'pupil's exam score at age 16' and 'pupil's score at age 11 on the London Reading Test' - the effect of 'standlrt' (prior attainment) on 'normexam' (outcome)**



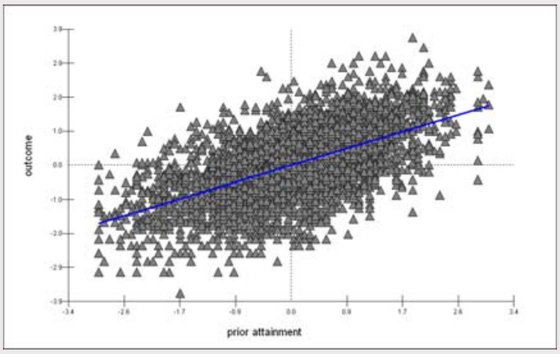
Data source: *MLwiN* tutorial sample

Simple scatter plot

All 4059 pupils

WJ Peng

University of BRISTOL



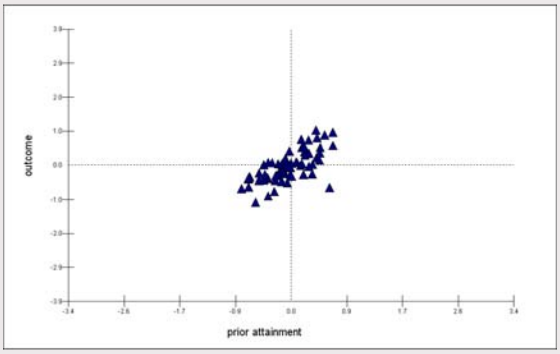
Data source: *MLwiN* tutorial sample

Ordinary linear regression method

Unit of analysis – pupil
One regression line

WJ Peng

University of BRISTOL

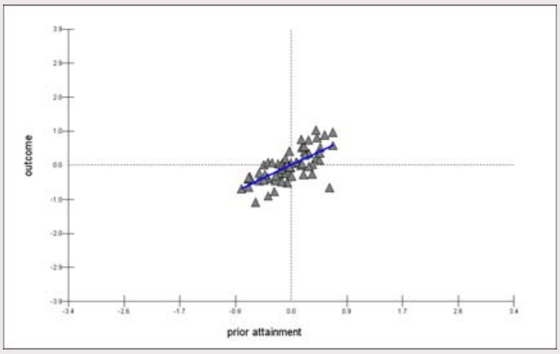


Data source: *MLwiN* tutorial sample

Simple scatter plot

Aggregated data for 65 schools

University of BRISTOL



Data source: *MLwiN* tutorial sample

Ordinary linear regression method

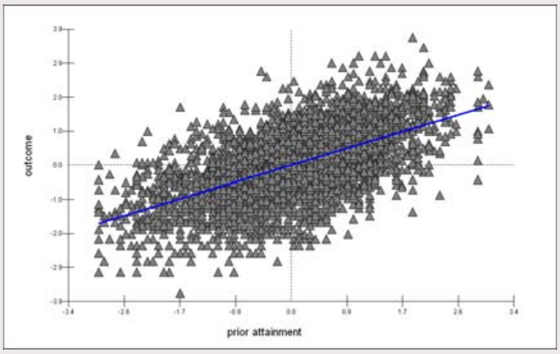
Unit of analysis – school
One regression line

University of BRISTOL

What went missing in the analysis?

The hierarchical structure of the data set

WJ Peng



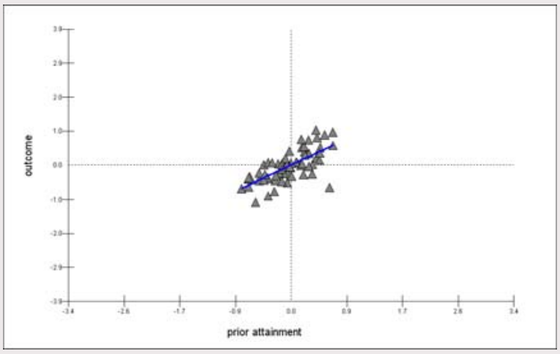
Data source: *MLwiN* tutorial sample

- The fact that those 4059 pupils were from 65 schools was ignored

Ordinary linear regression method

Unit of analysis – pupil
One regression line

WJ Peng



Data source: *MLwiN* tutorial sample

- The information about individual pupils was discarded

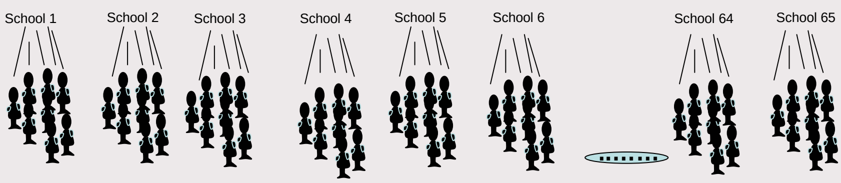
Ordinary linear regression method

Unit of analysis – school
One regression line

WJ Peng

University of BRISTOL

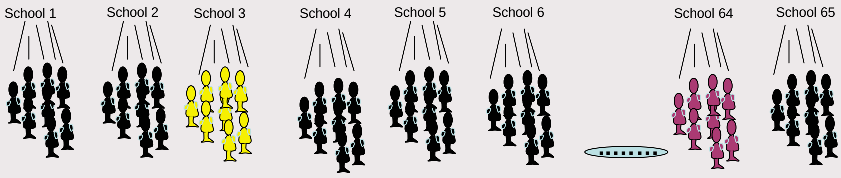
How this structure affects the measurement of interest?

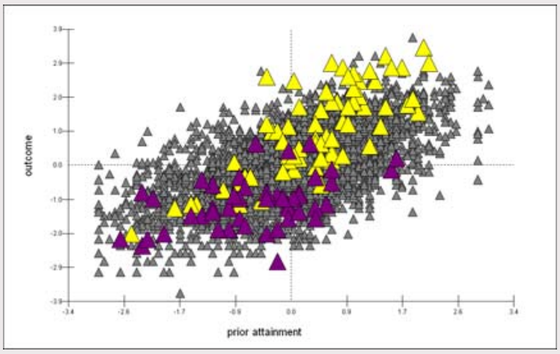


WJ Peng

University of BRISTOL

How this structure affects the measurement of interest?

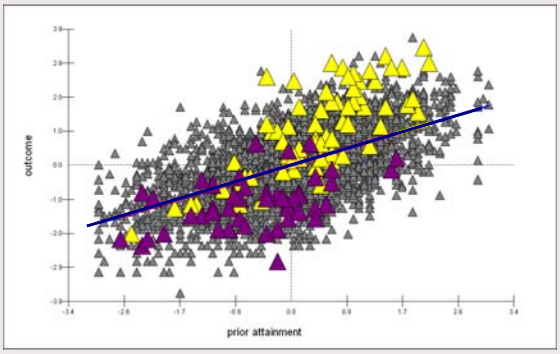




Data source: *MLwiN* tutorial sample

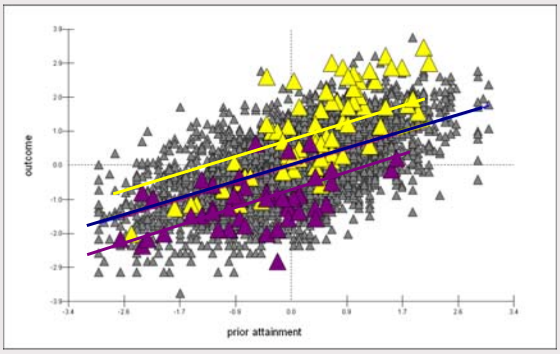
Simple scatter plots

For School 3 and School 64



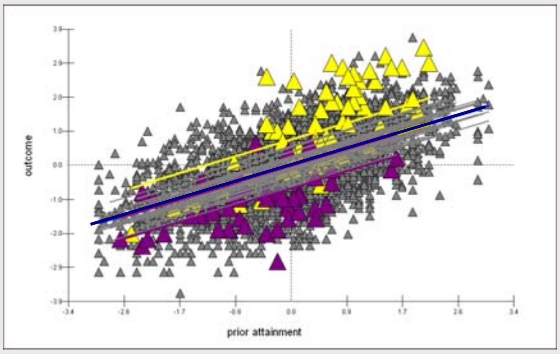
Data source: *MLwiN* tutorial sample

WJ Peng



Data source: *MLwiN* tutorial sample

WJ Peng



Data source: *MLwiN* tutorial sample

Underlying Meaning

In this case, pupils within a school will be more alike, on average, than pupils from different schools.

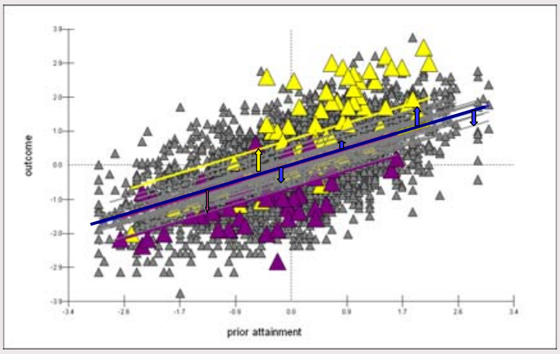
We are also interested in

Q1: The relationship between 'pupil's exam score at age 16' and 'pupil's score at age 11 on the London Reading Test' - the effect of 'standlrt' (prior attainment) on 'normexam' (outcome)

Q2: How different the relationship is across schools? - the variability of the effect of 'standlrt' (prior attainment) on 'normexam' (outcome) across schools

WJ Peng

University of BRISTOL



Data source: *MLwiN* tutorial sample

The variation between schools

WJ Peng

University of BRISTOL

Can ordinary linear regression method estimate the variation between schools?

It is possible that "The variation between schools could be modelled by incorporating separate terms for each school…"

(Rasbash, et al., 2005)

For example, to fit 64 school dummy variables in a model using school 1 as the reference school

WJ Peng

University of BRISTOL

$$
\begin{aligned}
\text{normexam}_i = \; & a + \beta_1\text{standlrt}_i + \beta_2\text{school\_2}_i + \beta_3\text{school\_3}_i + \beta_4\text{school\_4}_i + \beta_5\text{school\_5}_i + \\
& \beta_6\text{school\_6}_i + \beta_7\text{school\_7}_i + \beta_8\text{school\_8}_i + \beta_9\text{school\_9}_i + \beta_{10}\text{school\_10}_i + \\
& \beta_{11}\text{school\_11}_i + \beta_{12}\text{school\_12}_i + \beta_{13}\text{school\_13}_i + \beta_{14}\text{school\_14}_i + \beta_{15}\text{school\_15}_i + \\
& \beta_{16}\text{school\_16}_i + \beta_{17}\text{school\_17}_i + \beta_{18}\text{school\_18}_i + \beta_{19}\text{school\_19}_i + \beta_{20}\text{school\_20}_i + \\
& \beta_{21}\text{school\_21}_i + \beta_{22}\text{school\_22}_i + \beta_{23}\text{school\_23}_i + \beta_{24}\text{school\_24}_i + \beta_{25}\text{school\_25}_i + \\
& \beta_{26}\text{school\_26}_i + \beta_{27}\text{school\_27}_i + \beta_{28}\text{school\_28}_i + \beta_{29}\text{school\_29}_i + \beta_{30}\text{school\_30}_i + \\
& \beta_{31}\text{school\_31}_i + \beta_{32}\text{school\_32}_i + \beta_{33}\text{school\_33}_i + \beta_{34}\text{school\_34}_i + \beta_{35}\text{school\_35}_i + \\
& \beta_{36}\text{school\_36}_i + \beta_{37}\text{school\_37}_i + \beta_{38}\text{school\_38}_i + \beta_{39}\text{school\_39}_i + \beta_{40}\text{school\_40}_i + \\
& \beta_{41}\text{school\_41}_i + \beta_{42}\text{school\_42}_i + \beta_{43}\text{school\_43}_i + \beta_{44}\text{school\_44}_i + \beta_{45}\text{school\_45}_i + \\
& \beta_{46}\text{school\_46}_i + \beta_{47}\text{school\_47}_i + \beta_{48}\text{school\_48}_i + \beta_{49}\text{school\_49}_i + \beta_{50}\text{school\_50}_i + \\
& \beta_{51}\text{school\_51}_i + \beta_{52}\text{school\_52}_i + \beta_{53}\text{school\_53}_i + \beta_{54}\text{school\_54}_i + \beta_{55}\text{school\_55}_i + \\
& \beta_{56}\text{school\_56}_i + \beta_{57}\text{school\_57}_i + \beta_{58}\text{school\_58}_i + \beta_{59}\text{school\_59}_i + \beta_{60}\text{school\_60}_i + \\
& \beta_{61}\text{school\_61}_i + \beta_{62}\text{school\_62}_i + \beta_{63}\text{school\_63}_i + \beta_{64}\text{school\_64}_i + \beta_{65}\text{school\_65}_i
\end{aligned}
$$

WJ Peng

University of BRISTOL

Can ordinary linear regression method estimate the variation between schools?

However, it is inefficient and inadequate " because it involves estimating many times coefficients…because it does not treat schools as a random sample..."

(Rasbash, et al., 2005)

Think about a national data set with hundreds of schools……

WJ Peng

University of BRISTOL

Multilevel modelling

A statistical technique that allows an analysis to take account of the levels of hierarchical structure in the population so that we can

- treat sample as random
- specify and fit a wide range of multilevel models
- understand where and how effects are occurring

(Rasbash, et al., 2005)

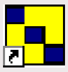
WJ Peng

University of BRISTOL

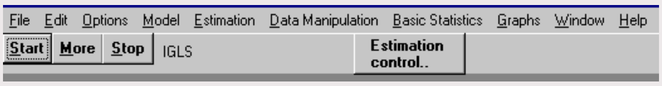
Statistical software packages

There are some statistical packages have the function

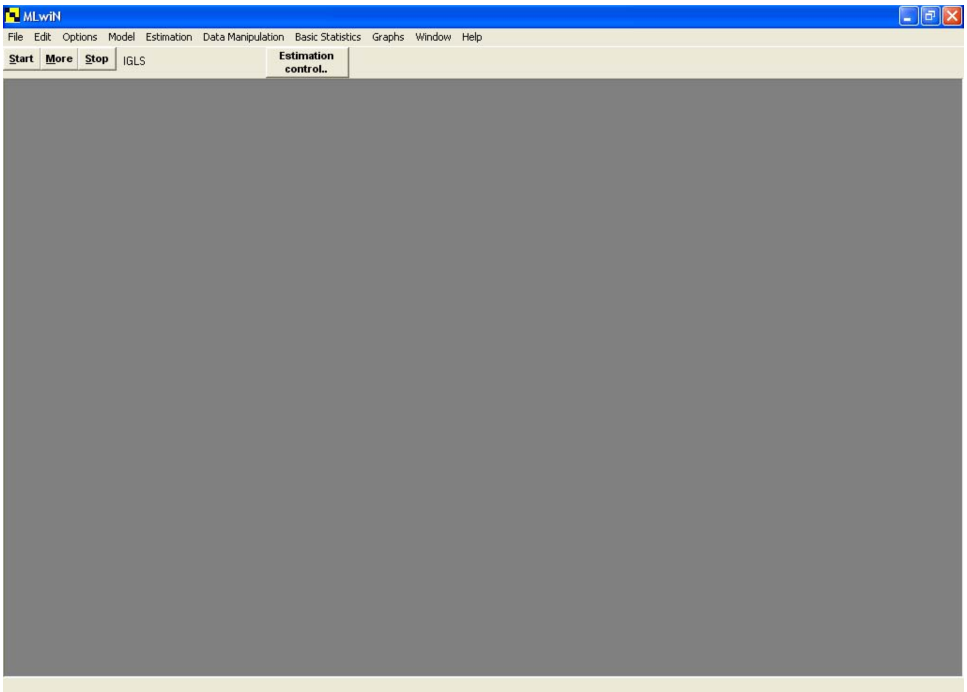
- ***MLwiN*** is one of them



Menu bar



WJ Peng

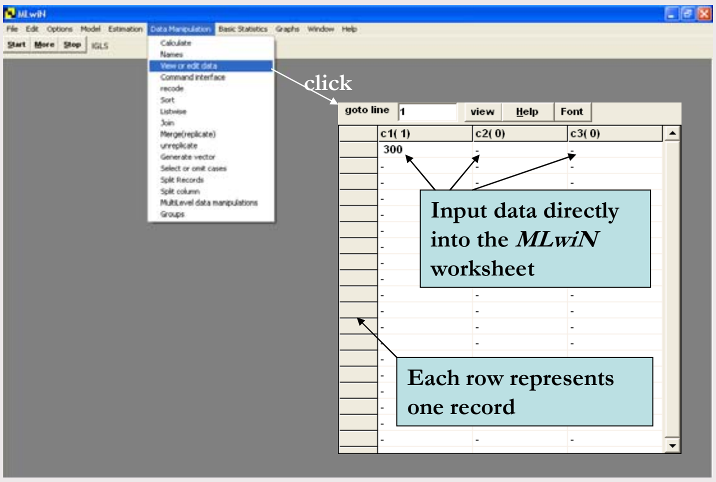


Get started with creating a *MLwiN worksheet*

MLwiN **can only input and output numerical data**

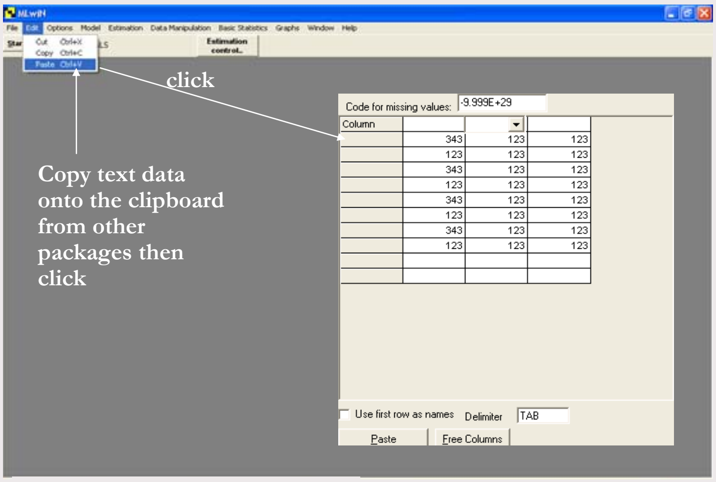
- **code data numerically**
- **assign an identical numerical code to all missing data**
- **three ways of creating a *MLwiN* worksheet:**
  - **input data into a *MLwiN* worksheet**
  - **copy and paste data into a *MLwiN* worksheet**
  - **import ASCII data from a text file**

Input data into a *MLwiN* worksheet



Data Manipulation/View or edit data

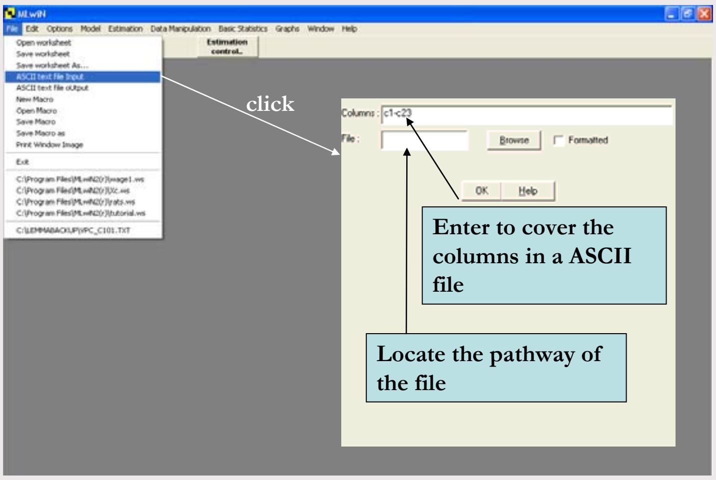
Paste data into a *MLwiN* worksheet



Edit/Paste

University of BRISTOL

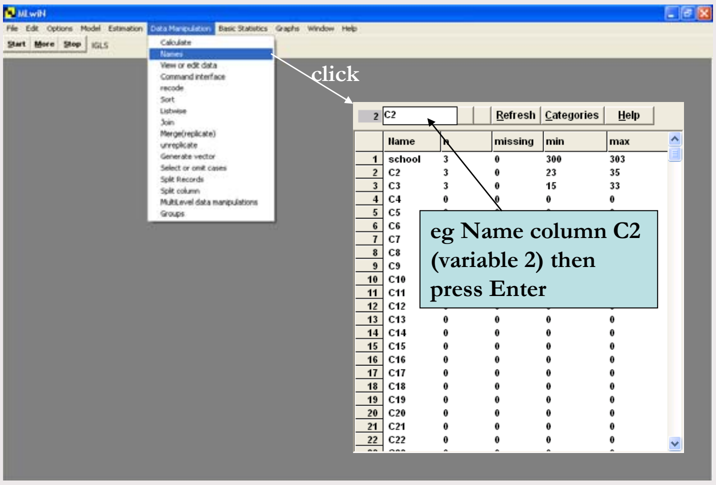
Import ASCII data from a text file



File/ASCII text file Input

University of BRISTOL

Name columns - variables

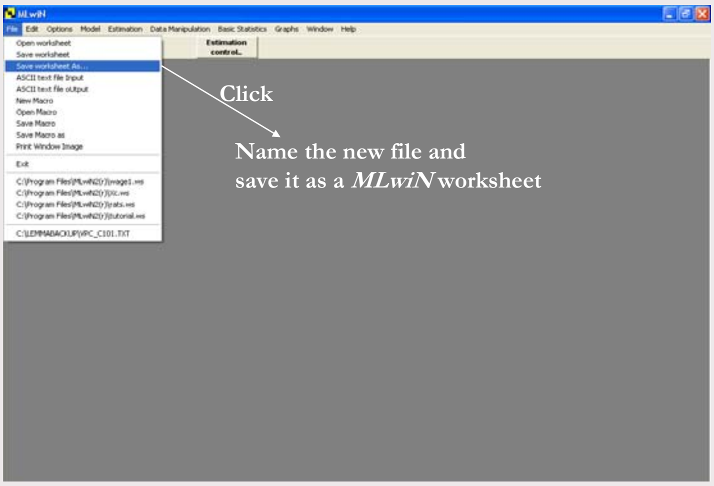


Data Manipulation/Names

WJ Peng

University of BRISTOL

Save the file as a *MLwiN* worksheet

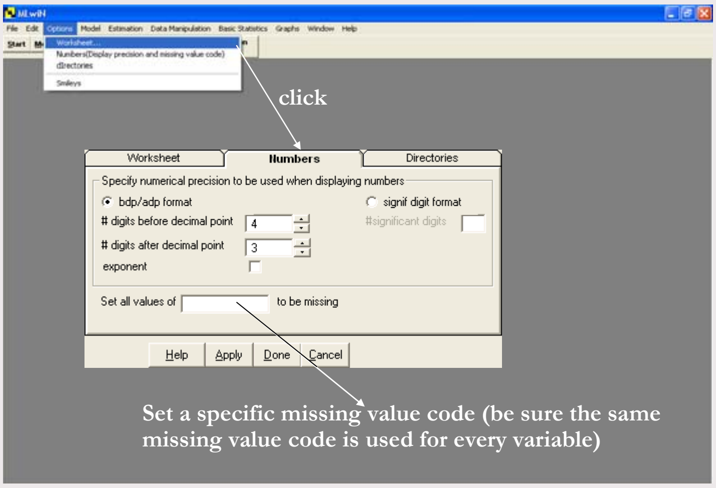


File/Save worksheet As…

WJ Peng

University of BRISTOL

Declare the missing data



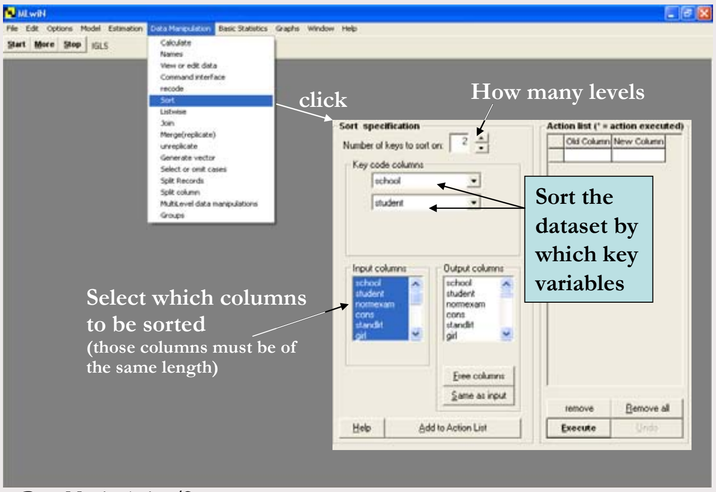
Options/Worksheet/Numbers

Sort a dataset to reflect its hierarchical structure

"before trying to fit a multilevel model to a dataset……the dataset must be sorted so that all records for the same highest-level unit are grouped together and within this group, all records for a particular lower level unit are contiguous"

(Rasbash, et al., 2005)

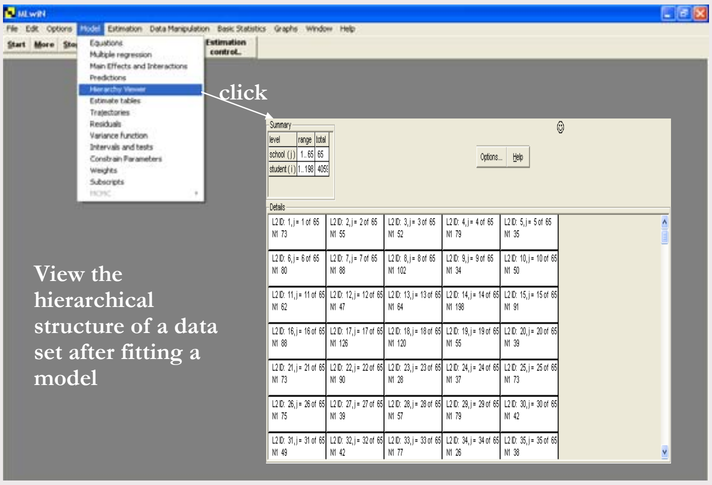
Sort the dataset



Data Manipulation/Sort

University of BRISTOL

View hierarchical structure



Models/Hierarchy Viewer **(Note: to View after fitting a model)**

University of BRISTOL

Checklist

All value codes are numerical?	√
An identical missing value code?	√
The dataset has been sorted?	√
The dataset is a MLwin worksheet?	√

WJ Peng

University of BRISTOL

Understand the notation used in *MLwiN*

An example – linear regression with continuous variables ***x*** and ***y*** for one school with ***i*** number of pupils

$\hat{y}_i = a + bx_i$ — i = 1, 2, 3…the number of pupils

$y_i = \hat{y}_i + e_i$
$\quad = a + bx_i + e_i$

e_i = residual (or error) ie, the difference between y_i and $\hat{y}_i$ – pupil level
a = intercept (average across all pupils)
b = slope (coefficient – the effect of x)

a (intercept) and b (slope of ***x***) define the average line across all pupils in the school.

WJ Peng

University of BRISTOL

Understand the notation used in *MLwiN*

For one school	$y_i = a + bx_i + e_i$
For a number of schools	$y_{i1} = a_1 + bx_{i1} + e_{i1}$ $y_{i2} = a_2 + bx_{i2} + e_{i2}$ .... $y_{ij} = a_j + bx_{ij} + e_{ij}$
There is	$a_j = a + u_j$ – **school level**
Thus	$y_{ij} = a + bx_{ij} + u_j + e_{ij}$

WJ Peng

University of BRISTOL

Understand the notation used in *MLwiN*

For a number of schools	$y_{ij} = a \quad + bx_{ij} \quad + u_j \quad + e_{ij}$
Introduce x_0 (=1) and symbols β_0 and β_1 to denote a, b	$y_{ij} = \beta_0 x_0 + \beta_1 x_{ij} + u_{0j} x_0 + e_{0ij} x_0$ x_0 **called cons in *MLwiN***
$y_{ij} = \beta_{0ij} x_0 + \beta_1 x_{ij}$	**general notation**
$\beta_{0ij} = \beta_0 + u_{0j} + e_{0ij}$	***i* = pupil level, *j* = school level**

β_0 and β_1 **define the average line across all pupils in all schools.**

WJ Peng

University of BRISTOL

The fixed and random parts in *MLwiN*

The fixed part of the model:

β_0, β_1 – multilevel modelling regression coefficients
– explained in the model

The random part of the model:

σ^2_{u0j} – the variance of the school level random effects u_{0j}
σ^2_{e0ij} – the variance of the pupil level random effects e_{0ij}
– unexplained in the model

(http://tramss.data-archive.ac.uk/documentation/MLwiN/chapter1.pdf)

Fit a multilevel model in *MLwiN*

- start with simple models -

"Multilevel modelling is like any other type of statistical modelling and a useful strategy is to start by fitting simple models and slowly increase the complexity."

"It is important…to know as much as possible about your data and to establish what questions you are trying to answer."

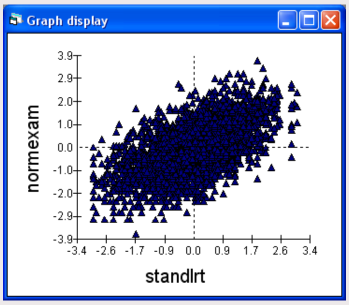
(Rasbash, et al., 2005)

Research questions

We are interested in exploring – via data modelling – the size, nature and extent of the school effect on progress in normexam.

Q1 – What the relationship between the outcome attainment measure normexam and the intake ability measure standlrt would be?

Q2 – How this relationship varies across schools (what the proportions of the overall variability shown in the plot attributable to schools and to student)?



(http://tramss.data-archive.ac.uk/documentation/MLwiN/chapter1.pdf)

Considering the following 3 models

Cons Model – a random intercepts Null model with 'normexam' as the response variable, no predictor/explanatory variables apart from the Constant (ie representing the intercept) which is allowed to vary randomly across schools and with the levels defined as pupils (level 1) in schools (level 2)

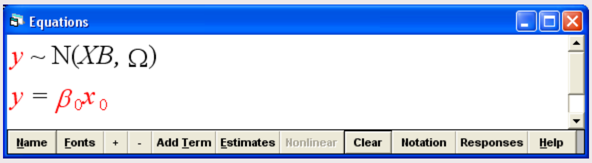
Model A – a random intercepts/variance components model – Cons Model with also an explanatory variable (standlrt)

Model B – a random intercepts/slopes model – Model A with also the parameter associated with standlrt being allowed to vary randomly across schools (ie random slopes as well as intercepts)

(Thomas, 2007, http://www.cmm.bristol.ac.uk/research/Lemma/two-level.pdf)

Fitting Cons Model

Select Model/Equations from Menu bar



Notice the red colour parts? – indicating that the variable and the parameter associated with it has not yet been specified

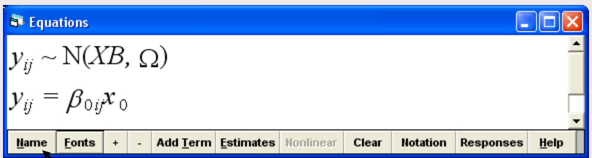
y ~ N(XB, Ω) – the default distributional assumption:

"The response vector has a mean specified in matrix notation by the *fixed part XB* , and a random part consisting of a set of random variables described by the *covariance matrix* Ω."

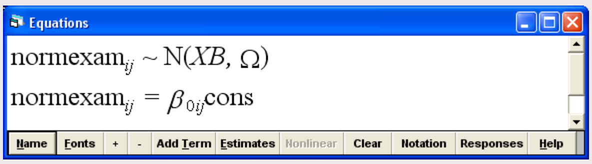
(http://tramss.data-archive.ac.uk/documentation/MLwiN/chapter1.pdf)

Click either of the *y's* to specify the response variable – normexam, assign *i* and *j* at pupil and school levels respectively.

Click either x_0 or β_0 to specify the explanatory variable – cons, assign *i* and *j* at pupil and school levels respectively to model the intercept.



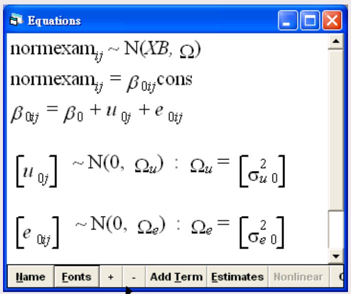
Click Name to show the names of the variables.



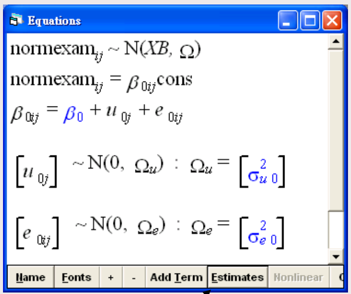
Notice the variables and parameters have changed from red to black? – indicating that specification is completed.

(http://www.cmm.bristol.ac.uk/research/Lemma/two-level.pdf)

Cons Model has now been specified

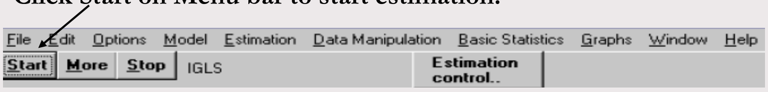


Click the + or - buttons to see the composition of β_{0ij}.



Click Estimates to see the parameters highlighted in blue that are to be estimated.

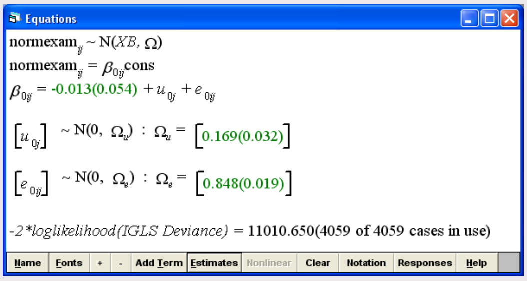
Click Start on Menu bar to start estimation.



WJ Peng

University of BRISTOL

Completion of the parameters estimation



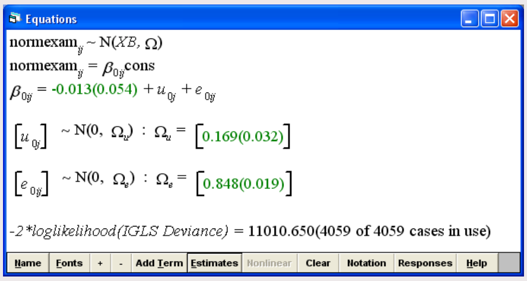
Note that the default method of estimation is iterative generalised least squares (IGLS).

The blue highlighted parameters in the Equations window change to green to indicate convergence.

WJ Peng

University of BRISTOL

What the parameter estimates tell us?



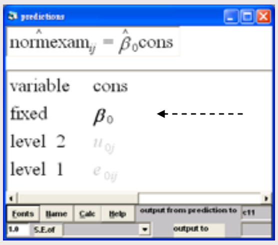
Note that normexam scores were normalised to have a proximately standard normal distribution

Overall mean	**-0.013 (approach to zero)**
Total variance	**0.169 + 0.848 = 1.017 (approach to 1) (if children were taken from the whole population at random the variance would be)**
Intra-school correlation	**0.169/(0.169 + 0.848) = 16.6% (the proportion of the total variance attributable to the school)**

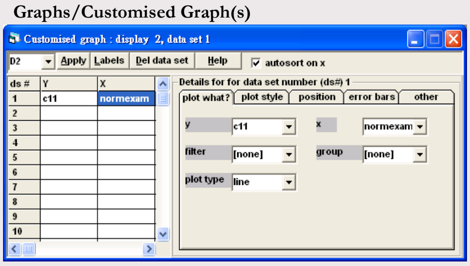
University of BRISTOL

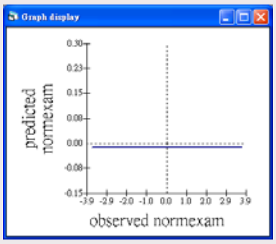
Graphing prediction

Models/Prediction



Graphs/Customised Graph(s)



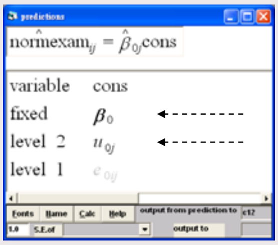


In Prediction window, click β_0 to calculate the average predicted line produced from the intercept coefficient β_0 - this is the predicted overall mean normexam (= -0.013) line for all pupils in all schools.

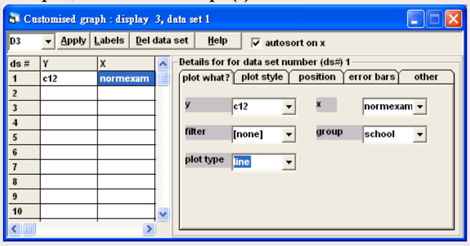
University of BRISTOL

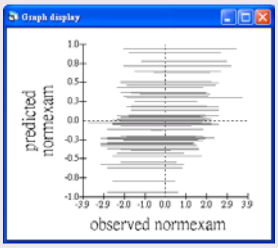
Graphing prediction

Models/Prediction



Graphs/Customised Graph(s)





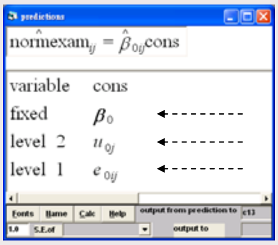
Click also u_{0j} to include the estimated school level intercept residuals in the prediction function and produce the predicted lines for all 65 schools. The line for school j departs from the average prediction line by an amount u_{0j}.

WJ Peng

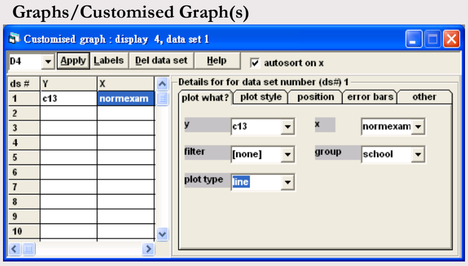
University of BRISTOL

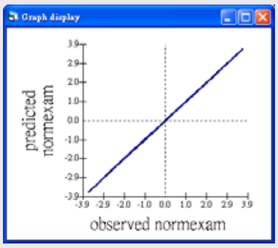
Graphing prediction

Models/Prediction



Graphs/Customised Graph(s)





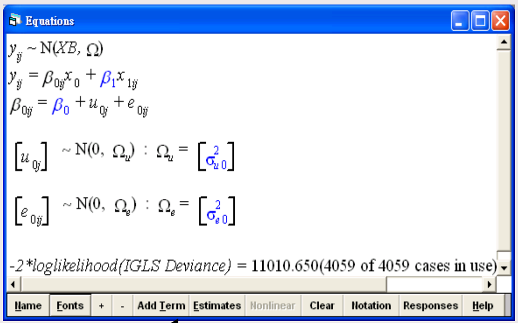
Click e_{0ij} to include the estimated pupil level intercept residuals in the prediction function too. Plot shows identical predicted and observed normexam (r = 1). Pupil i in school j departs from the predicted line for school j by an amount e_{0ij}.

WJ Peng

University of BRISTOL

Fitting Model A

- an random intercepts model -



"Note that x_0 has no other subscript but that x_1 has collected subscripts ***ij***.

MLwiN detects that cons is constant over the whole data set, whereas the values of standlrt change at both level 1 and level 2."

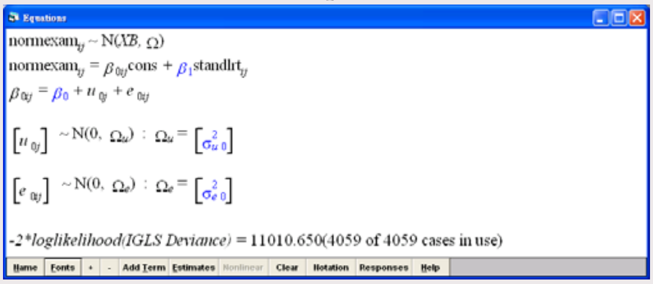
Click Add Term to add an explanatory variable – standlrt.

(http://tramss.data-archive.ac.uk/documentation/MLwiN/chapter1.pdf)

WJ Peng

University of BRISTOL

Click the +, -, and Name buttons to see how much the detail of the model is displayed.



β_0 (the intercept) and β_1 (the slope of standlrt) define the average line across all pupils in all schools.

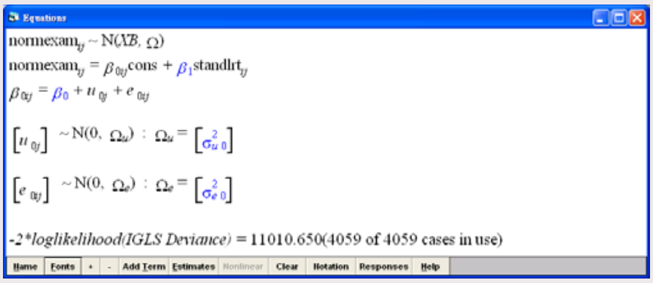
"The model is made multilevel by allowing each school's summary line to depart (be raised or lowered) from the average line by an amount u_{0j}." Pupil *i* in the school *j* departs from its school's summary line by an amount e_{0ij} .

(http://tramss.data-archive.ac.uk/documentation/MLwiN/chapter1.pdf)

WJ Peng

University of BRISTOL

In other words……



u_{0j} – the level 2 or school level *residuals* (one for each school); distributed Normally with mean 0 and variance σ^2_{u0}

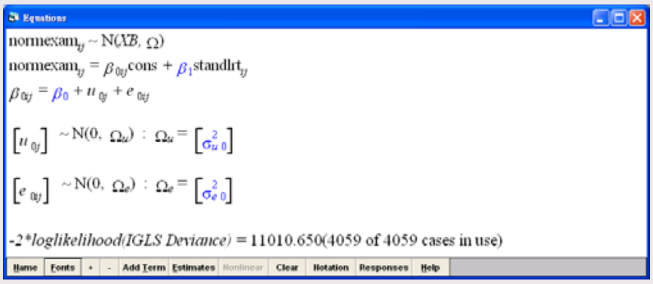
e_{0ij} – the level 1 or pupil level residuals (one for each pupil); distributed Normally with mean 0 and variance σ^2_{e0ij}

(http://tramss.data-archive.ac.uk/documentation/MLwiN/chapter1.pdf)

WJ Peng

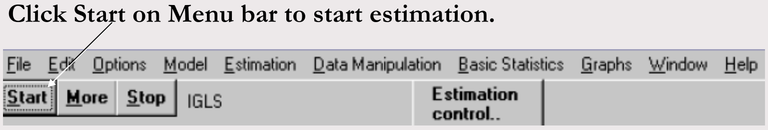
University of BRISTOL

Estimate the parameters of the specified model



The parameters highlighted in blue are to be estimated.

Click Start on Menu bar to start estimation.

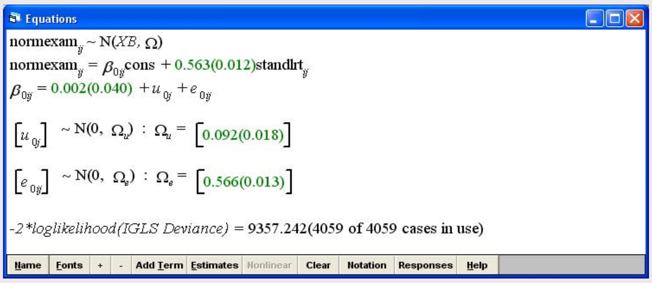


(http://tramss.data-archive.ac.uk/documentation/MLwiN/chapter1.pdf)

WJ Peng

University of BRISTOL

Completion of the parameters estimation



Slope – the slopes of the lines across schools are all the same, of which the common slope is 0.563 with SE = 0.012

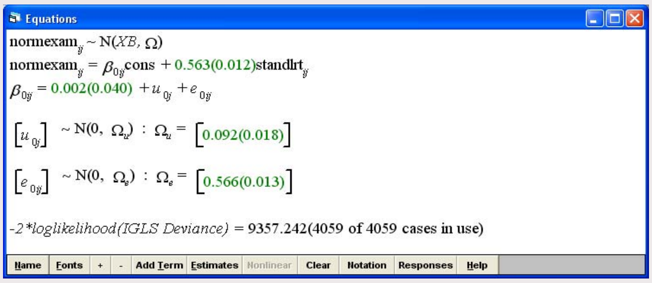
Intercept – the intercepts of the lines vary across schools. Their mean is 0.002 with SE = 0.040. The intercept of school *j* is 0.002 + u_{0j} with a variance of 0.092 and SE = 0.018.

(http://tramss.data-archive.ac.uk/documentation/MLwiN/chapter1.pdf)

WJ Peng

University of BRISTOL

What the parameter estimates tell us?



Total variance (0.092 + 0.566 = 0.658) – the sum of the level 2 and level 1 variances

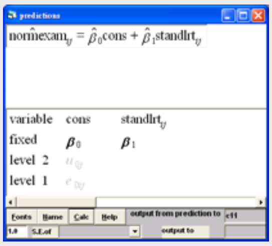
Intra-school correlation (0.092/0.658 = 0.140) – measuring the extent to which pupils' scores in the same school are more alike as compared with those from pupils at different schools

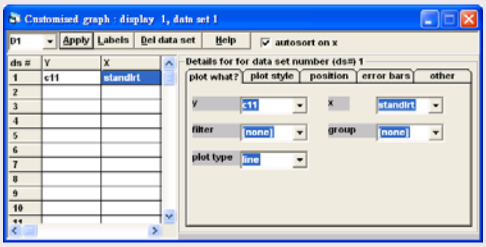
(http://tramss.data-archive.ac.uk/documentation/MLwiN/chapter1.pdf)

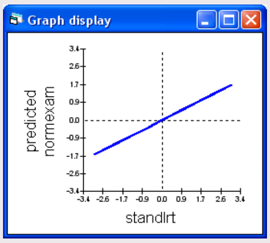
WJ Peng

University of BRISTOL

Graphing predication







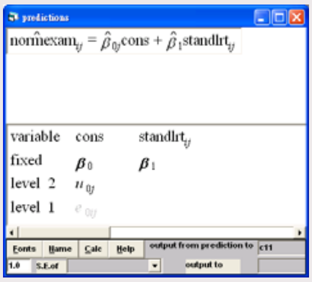
$\hat{y} = \beta_0 \text{cons} + \beta_1 x$

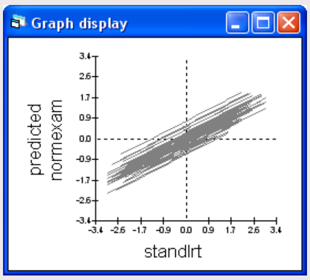
$\hat{y} = 0.002 + 0.563\text{standlrt}$

The average line across all pupils in all schools

WJ Peng

Graphing predication





One line for each school

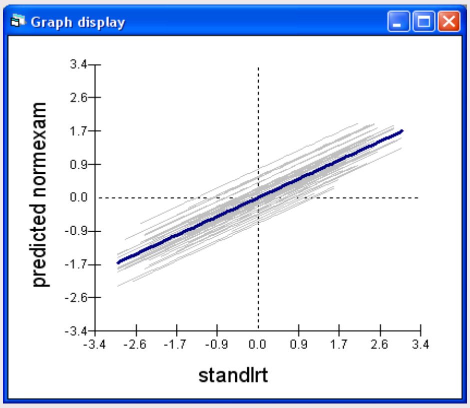
$\hat{y} = \beta_{0j} \text{cons} + \beta_1 x$

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

$= 0.002 + u_{0j}$

WJ Peng

Graphing predication

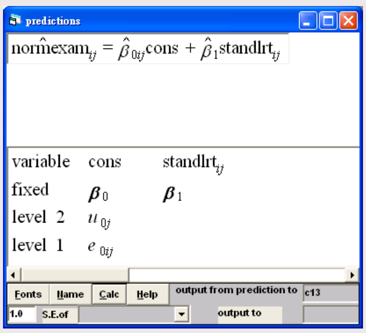


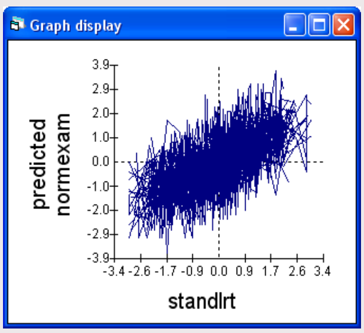
The line for school j departs from the average prediction blue line by an amount u_{0j} .

WJ Peng

University of BRISTOL

Graphing predication





$\hat{y} = \beta_{0ij}\text{cons} + \beta_1 x$

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

$\beta_{0ij} = 0.002 + u_{0j} + e_{0ij}$

Pupil i in school j departs from the school j summary line by an amount e_{0ij} .

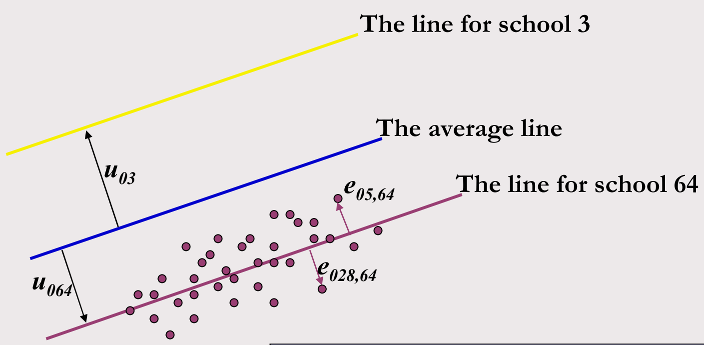
WJ Peng

University of BRISTOL

What all these about – Model A?

WJ Peng

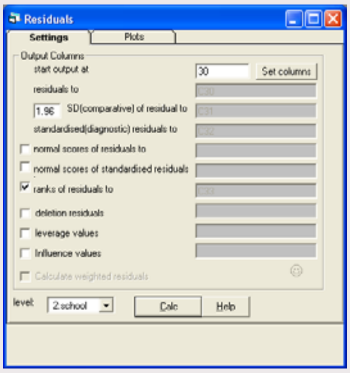
In brief, by employing multilevel modelling approach…



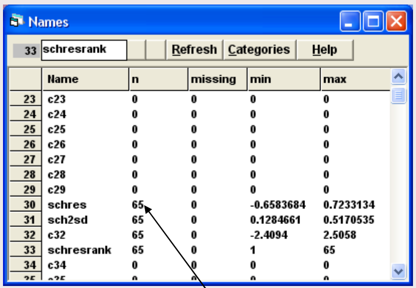
u_{03} – residual for School 3
u_{064} – residual for School 64
$e_{05,64}$ – residual for pupil 5 in school 64
$e_{028,64}$ – residual for pupil 28 in school 64

WJ Peng

Graphing residuals



Model/Residuals/Settings



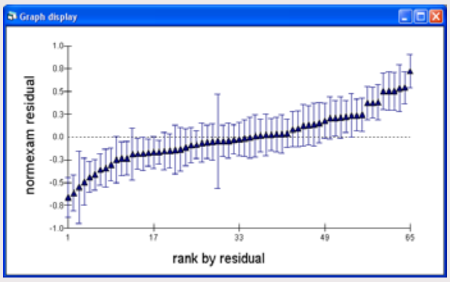
	Name	n	missing	min	max
23	c23	0	0	0	0
24	c24	0	0	0	0
25	c25	0	0	0	0
26	c26	0	0	0	0
27	c27	0	0	0	0
28	c28	0	0	0	0
29	c29	0	0	0	0
30	schres	65	0	-0.6583684	0.7233134
31	sch2sd	65	0	0.1284661	0.5170535
32	c32	65	0	-2.4094	2.5058
33	schresrank	65	0	1	65
34	c34	0	0	0	0

Residuals for individual schools, of which their mean is 0 and their estimated variance of 0.092

University of BRISTOL

Graphing residuals

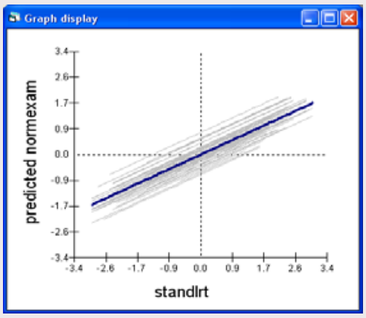
Model/Residuals/Plots

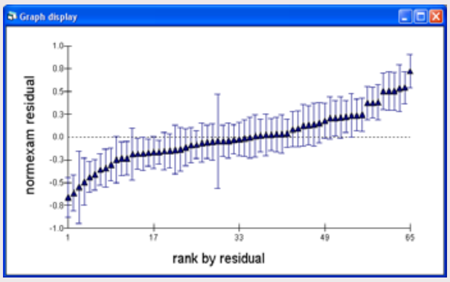


Each vertical line represents a residual with 95% confidence interval estimated for each school.

University of BRISTOL

What is meant by residual?





School residual
– the departure of a school (grey) line from the average (blue) line

These school residuals might be regarded as school effect – expressed by the term 'value added' in school effectiveness and improvement research.

What is meant by value added?

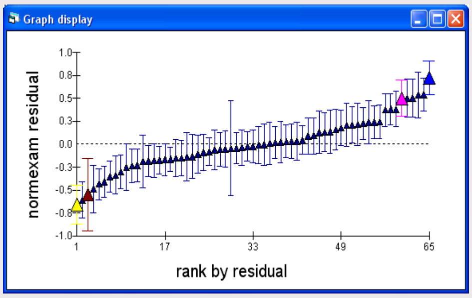
In this case, value added (or residual) for each school represents the differences between the *observed* level of school performance (pupil normexam scores taken at age 16) and what would be *expected* on the basis of pupils' prior attainment (pupul standlrt scores taken at age 11).

In other words "value added is a measure of the relative progress made by pupil in a school over a particular period of time (usually from entry to the school until public examinations in the case of secondary schools, or over particular years in primary schools – in this case, between age 11 and 16) in comparison to pupils in others schools in the same sample."

(Thomas, 2005)

(See Thomas (2005) Using indicators of value added to evaluation school performance in UK. Educational Research Journal. 2005 September 2005. China National Institute of Educational Research: Beijing – in Chinese)

Were some schools doing better than others?



- a positive value added score (i.e. residual) indicating a school may be performing above expectation.

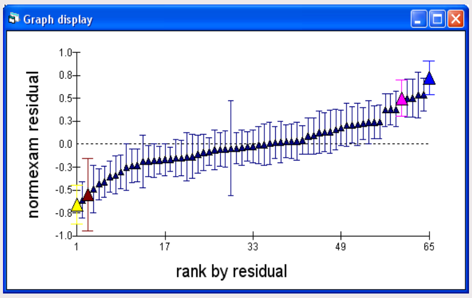
- a negative value added score indicating a school may be performing below expectation

"However, information about the 95% confidence interval (CI) is required to evaluate whether an individual school's value added performance is likely to have occurred by chance.

In other words, the confidence interval is vital to judge whether a school's performance above or below expectation is *statistically significant* ." (Thomas, 2005)

University of BRISTOL

Were some schools doing better than others?

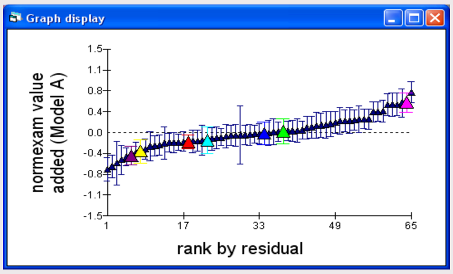


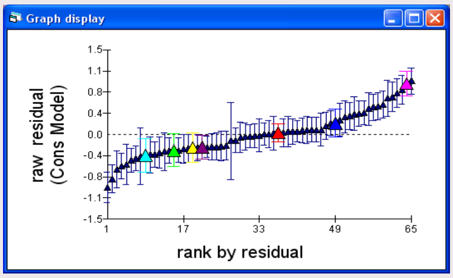
Information is also needed about the statistical uncertainty of performance measures when different schools are compared.

So…
Was School▲ doing better than school△ ?
Was School△ doing worse than school▲ ?
How about school▲ and school▲ ?

University of BRISTOL

Compare raw (Cons Model) and value added residuals

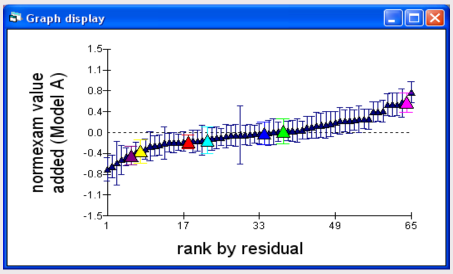


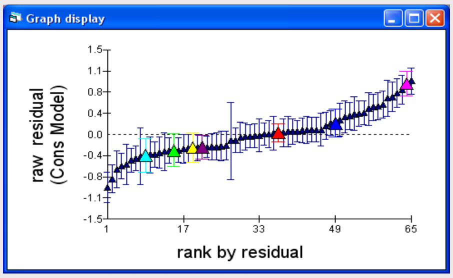


On average, how were these schools performing in their raw normexam and value added (VA) scores as compared to other schools?

As compared to other schools:
- △ performed higher than expected in both scores
- ▲ performed as expected in both scores
- △ performed lower than expected in both scores
- ▲ performed as expected in raw but lower than expected in VA
- how about other schools?

Compare level 2 (and level 1) variance between two models



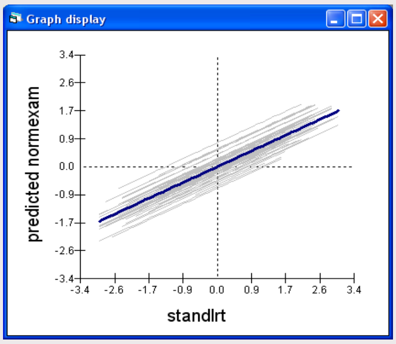


Variance between schools
- which residual curve ~ is steeper
- what the implication of it

Variance within schools
- the likely bounds (95%CI) of variation on schools for raw residuals wider than the ones for value added
- What the implication of it

Fitting Model B

- random intercepts/slopes model -



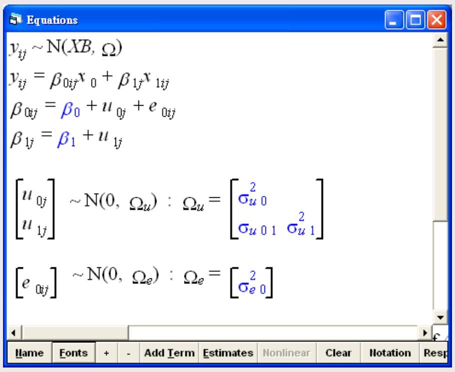
Graph of Model A - variance components model

Model A which we have just specified and estimated assumes that the only variation between schools is in their intercepts. "However, there is a possibility that the school lines have different slopes. This implies that the coefficient of standlrt will vary from school to school."

(http://tramss.data-archive.ac.uk/documentation/MLwiN/chapter1.pdf)

University of BRISTOL

Specifying Model B



Click β_1 to specify the coefficient of standlrt which is random at level 2.

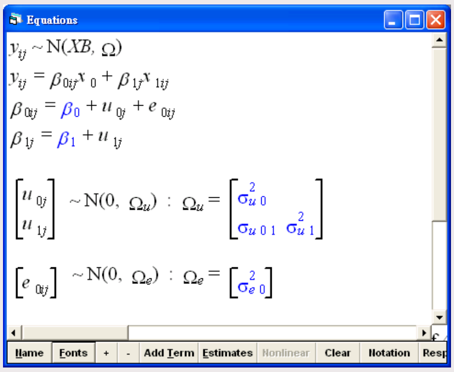
"The terms u_{0j} and u_{1j} are random departures or 'residuals' at the school level from β_0 and β_1. They allow the j'th school's summary line to differ from the average line in both its slope and its intercept."

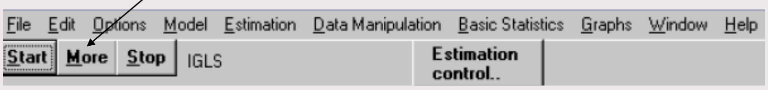
(http://tramss.data-archive.ac.uk/documentation/MLwiN/chapter1.pdf)

University of BRISTOL

Specifying Model B

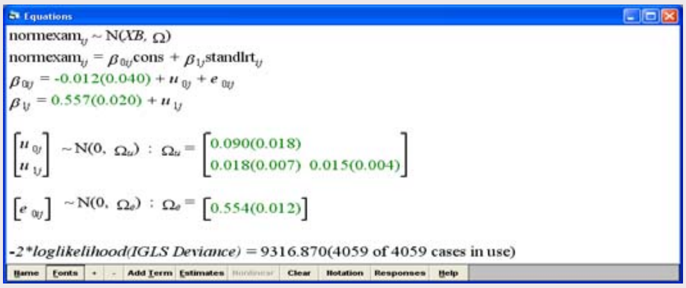
To fit this new model we could click Start as before, but it will probably be quicker to use the estimates already obtained from the 1st model as initial values for the iterative calculations. Click More.





(http://tramss.data-archive.ac.uk/documentation/MLwiN/chapter1.pdf)

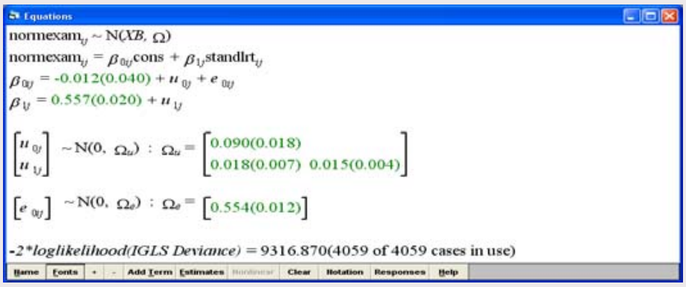
Completion of the parameters estimation



	mean (SE)	variance (SE)
individual school slopes vary	0.557 (0.020)	0.015 (0.004)
school line intercepts vary	-0.012 (0.040)	0.090 (0.018)

(http://tramss.data-archive.ac.uk/documentation/MLwiN/chapter1.pdf)

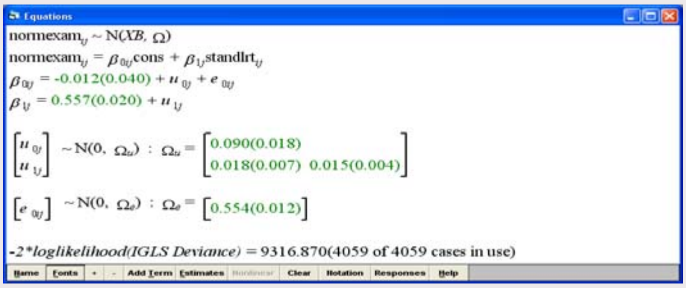
Completion of the parameters estimation



"The positive covariance between intercepts and slopes estimated as +0.018 (SE = 0.007) suggests that schools with higher intercepts tend to some extent to have steeper slopes and this corresponds to a correlation between the intercept and slope (across schools) of 0.49."

(http://tramss.data-archive.ac.uk/documentation/MLwiN/chapter1.pdf)

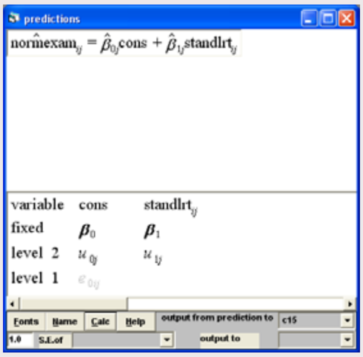
Completion of the parameters estimation

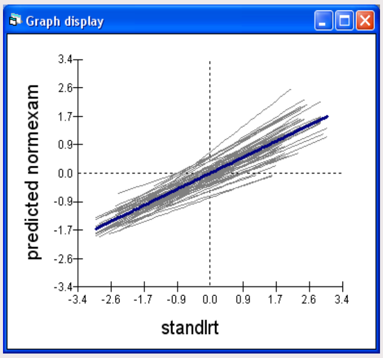


"The pupils' individual scores vary around their schools' lines by quantities e_{0ij}, the level 1 residuals, whose variance is estimated as 0.554 (SE = 0.012)."

(http://tramss.data-archive.ac.uk/documentation/MLwiN/chapter1.pdf)

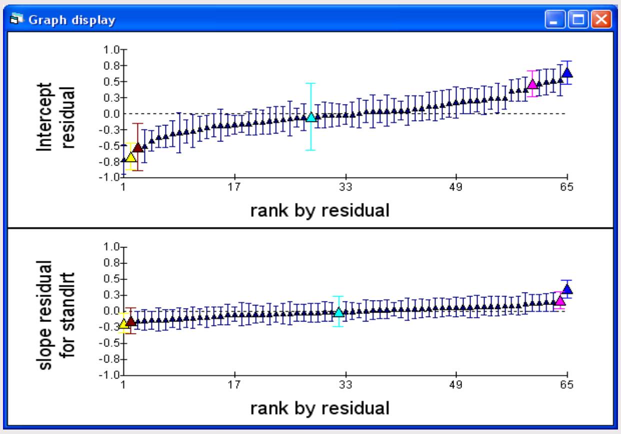
Graphing prediction





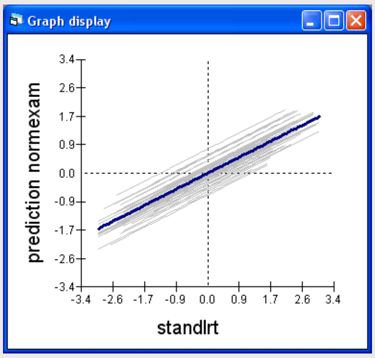
The positive covariance between slopes and intercepts leading to a fanning out pattern when plotting the schools predicted lines (the average line = blue line)

Graphing residual

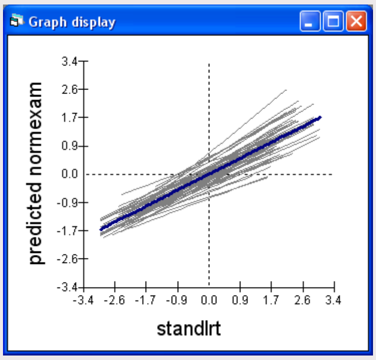


One residual plot for the intercepts of individual school lines
One residual plot for the individual line slopes

Compare the two models



Model A

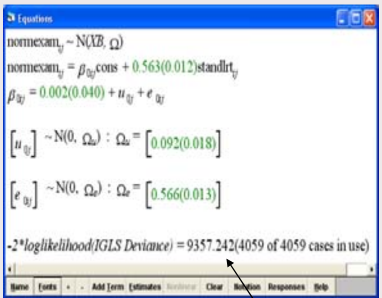


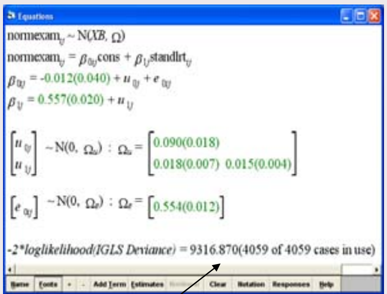
Model B

Which model is better fit?

University of BRISTOL

Which model was better fit?



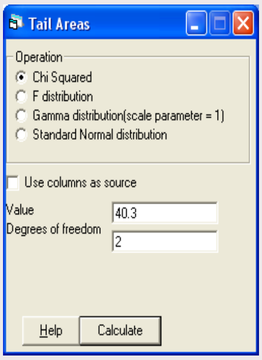


"A -2log-likelihood value is the probability of obtaining the observed data if the model were true and can be used in the comparison of two different models." **(Rasbash, et al., 2005)**

University of BRISTOL

Which model is better fit – the likelihood ratio test

Basic Statistsics/Tail Areas



- the change of the two -2*log-likelihood values 9357.2 - 9316.9 = 40.3.
- the change in the –2log-likelihood value (which is also the change in deviance) has a chi-squared distribution on 2 degrees of freedom under the null hypothesis that the extra parameters have population values of zero.
- two extra parameters involved in the 2nd model
  (1) the variance of the slope residuals u_{1j}
  (2) their covariance with the intercept residuals u_{0j}.

(Rasbash, et al., 2005)

The change is very highly significant, confirming the better fit of the 2nd model, a more elaborate model to the data.

WJ Peng

University of BRISTOL

Examples of other modelling

Gender effects
- Do girls make more progress than boys? (F)
- Are boys more or less variable in their progress than girls? (R)

Contextual effects
- Are pupils in key schools less variable in their progress? (R)
- Do pupils do better in urban schools (or key schools)? (F)
- Does gender gap vary across schools? (R)

Cross-level interaction
- Do boys learn more effectively in a boys' or mixed sex school? (F)
- Do low ability pupils fare better when educated alongside higher ability pupils? (F)

(Jones, 2007; Rasbash, et al., 2005)

WJ Peng

University of BRISTOL

Examples of other hierarchical structures in education settings

Level 1	level 2	level 3	level 4
Pupils	classes	schools	
Pupils	schools	regions	
Pupils	schools	regions	countries
Pupils	neighbourhoods	schools	regions

NB:
What happens if we have other types of data (eg ordered/unordered categorical data, binomial/multinomial data, repeated data) or non-hierachical structure (eg pupils changing schools)?

Useful references/links

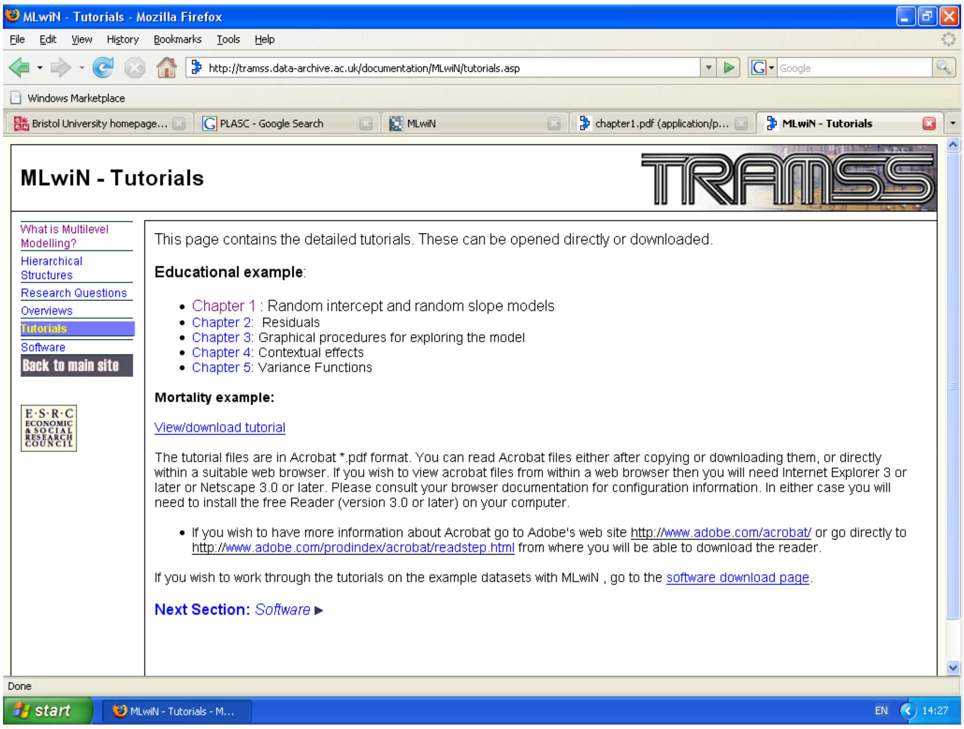
Getting start with the concept of value added in school effectiveness and improvement research:

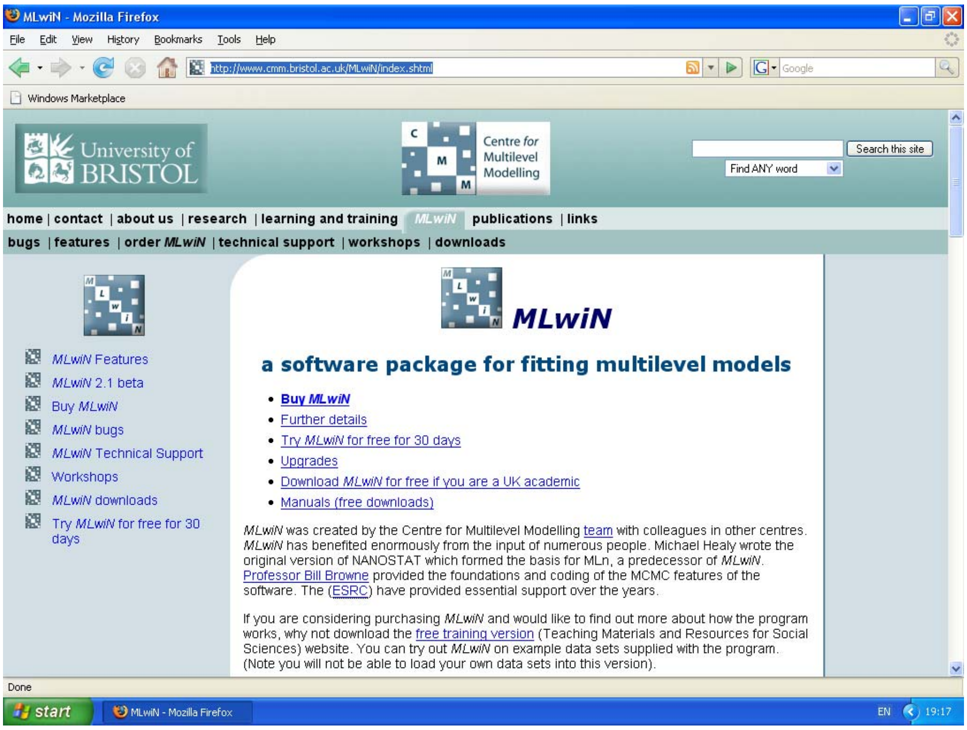
Thomas (2005) Using indicators of value added to evaluation school performance in UK. Educational Research Journal, September 2005, CNIER: Beijing. (translated into Chinese)

Getting start with how to fit a model in MLwiN:

Centre for Multilevel Modelling, Graduate School of Education, University of Bristol
http://www.cmm.bristol.ac.uk/MLwiN/index.shtml

Teaching Resources and Materials for Social Scientists, ESRC
http://tramss.data-archive.ac.uk/documentation/MLwiN/what-is.asp


MLwiN - Tutorials - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://tramss.data-archive.ac.uk/documentation/MLwiN/tutorials.asp
Windows Marketplace
Bristol University homepage...
PLASC - Google Search
MLwiN
chapter1.pdf (application/p...
MLwiN - Tutorials
MLwiN - Tutorials
TRAMSS
What is Multilevel Modelling?
Hierarchical Structures
Research Questions
Overviews
Tutorials
Software
Back to main site
E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL
This page contains the detailed tutorials. These can be opened directly or downloaded.
Educational example:
Chapter 1 : Random intercept and random slope models
Chapter 2: Residuals
Chapter 3: Graphical procedures for exploring the model
Chapter 4: Contextual effects
Chapter 5: Variance Functions
Mortality example:
View/download tutorial
The tutorial files are in Acrobat *.pdf format. You can read Acrobat files either after copying or downloading them, or directly within a suitable web browser. If you wish to view acrobat files from within a web browser then you will need Internet Explorer 3 or later or Netscape 3.0 or later. Please consult your browser documentation for configuration information. In either case you will need to install the free Reader (version 3.0 or later) on your computer.
If you wish to have more information about Acrobat go to Adobe's web site http://www.adobe.com/acrobat/ or go directly to http://www.adobe.com/prodindex/acrobat/readstep.html from where you will be able to download the reader.
If you wish to work through the tutorials on the example datasets with MLwiN , go to the software download page.
Next Section: Software ▸
Done
start
MLwiN - Tutorials - M...
EN
14:27


MLwiN - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://www.cmm.bristol.ac.uk/MLwiN/index.shtml
Windows Marketplace
University of BRISTOL
Centre for Multilevel Modelling
Search this site
Find ANY word
home | contact | about us | research | learning and training | MLwiN | publications | links
bugs | features | order MLwiN | technical support | workshops | downloads
MLwiN
a software package for fitting multilevel models
Buy MLwiN
Further details
Try MLwiN for free for 30 days
Upgrades
Download MLwiN for free if you are a UK academic
Manuals (free downloads)
MLwiN Features
MLwiN 2.1 beta
Buy MLwiN
MLwiN bugs
MLwiN Technical Support
Workshops
MLwiN downloads
Try MLwiN for free for 30 days
MLwiN was created by the Centre for Multilevel Modelling team with colleagues in other centres. MLwiN has benefited enormously from the input of numerous people. Michael Healy wrote the original version of NANOSTAT which formed the basis for MLn, a predecessor of MLwiN. Professor Bill Browne provided the foundations and coding of the MCMC features of the software. The (ESRC) have provided essential support over the years.
If you are considering purchasing MLwiN and would like to find out more about how the program works, why not download the free training version (Teaching Materials and Resources for Social Sciences) website. You can try out MLwiN on example data sets supplied with the program. (Note you will not be able to load your own data sets into this version).
Done
start
MLwiN - Mozilla Firefox
EN
19:17