Visualising multilevel models: the Initial Analysis of Data

John F Bell
Research and Evaluation Division
University of Cambridge Local Examinations Syndicate
1 Hills Road
Cambridge
CB1 2EU

20th March 2001

Abstract

This paper considers the use of the freeware package XLISP-STAT (Tierney, 1990) and the add-on package ARC (Cook and Weisberg, 1999) to explore multilevel data structures before more formal modelling. Exploratory data analysis (Chatfield, 1988) is regarded as an essential before further analysis. XLISP-STAT is highly interactive and with ARC can be used to investigate multilevel structures graphically. This feature is used to demonstrate how exploratory analysis can be carried out for a range of multilevel models including those with binary or ordinal outcome variables. The advantages of this software are that it can be used to identify potential outliers before models are fitted and it is also possible to decide on suitable models for a more formal multilevel analysis.

Note: Many of the plots in this paper use colour. This paper can be viewed on the web at the following location: f:\data\jfb\visualising multilevel models.doc (*To be changed to a web location eventually)*

Introduction

Multilevel models are complex and need careful specification. Multilevel models are used for the analysis of data with complex patterns of variability, with a focus on nested sources of variability: e.g., pupils in classes, patient in hospitals, longitudinal measurements of subjects. For such data, it is usually necessary to consider the variability associated with each level of nesting. In this paper, the objective is to identify how graphical displays can be used to provide guidance in the multilevel modelling process. This is sometimes described as an Initial Analysis of Data (IDA). Chatfield (1985) identified the need for the use of techniques to carry out an IDA. He argued that the objective of IDA is to clarify the structure of the data, obtain a simple descriptive summary, and also to plan a more sophisticated analysis. He further argued that IDA should be carried out before attempting formal inference because this would avoid the use of inappropriate techniques or models. Chatfield and Schimek (1987) give an example of IDA for time series analysis. Chatfield proposed that, although the general approach for a regression analysis is determined a priori, IDA is crucial when making assumptions about the model to be fitted. For example, in regression, a scatterplot should be produced to indicate the shape of the curve (linear or non-linear) and on other assumptions such as normality and homogeneity of variance. (It is ironic that Chatfield's original paper was used to criticise research that led to the development of one of the seminal papers about multilevel modelling (Aitken, Anderson and Hinde, 1981)).

IDA is useful for multilevel models because of the need to give some thought to model formulation. Multilevel modelling uses complex algorithms to estimate the parameters. Sometimes these algorithms take a long time to converge or fail to converge at all. Experience would suggest that this might occur when a model includes non-significant terms in it. An IDA allows investigators to identify when this is likely to happen.

Since Chatfield introduced the term IDA there have been considerable developments in statistical computing. In particular, there has been a growth in the use and availability of highly interactive statistical software. Another development that is relevant to the analysis is the growth in the use of smoothing techniques to supplement more formal regression analysis. For ordinary regression, one of the most useful packages that has these features is *ARC* (http://www.stat.umn.edu/arc/software.html). This program is free software that can be distributed and/or modified under the terms of the General Public License. This package is described in Cook and Weisberg (1999). It is a computer package written in *Xlisp-Stat* language (Tierney, 1990). Although this software was not originally designed for multilevel models, it is possible to demonstrate some of the features that make it useful for exploratory analysis of multilevel data. An important feature of *XLISP-STAT* is that it is easily extendable by users who understand the programming language. This means that new developments in statistics can be incorporated. In this paper, additional code produced by other authors will be used.

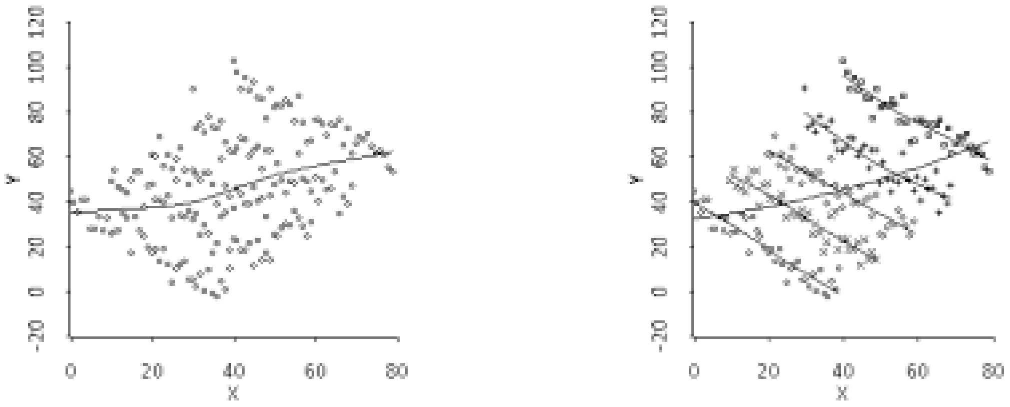
Although this paper will not consider the fitting of multilevel models, there is research into using *Xlisp-Stat* to fit multilevel models. *Terrace-Two* (Hilden-Minton, 1994; Afsharthous and Hilden-Minton, 1994) can be used to fit two-level regression models and has some facilities for generating diagnostics. There is also *glmer*, a majorization method for mixed model fitting, which is at http://www.xlispstat.org/code/statistics/regression/ with accompanying documentation at http://www.stat.ucla.edu/papers/preprints/115.ps.gz. At the time of writing, fitting multilevel models is not possible with *ARC* but there are plans for an implementation. However, given that not all multilevel software have graphics facilities, the ability to import data into *ARC* provides a powerful to tool for carrying out graphical analysis of this data. In the paper, the freeware mix suite of programs (http://www.uic.edu/~hedeker/mix.html) will be used. These programs only fit multilevel models and so it is useful to use a more general statistical tool with them.

In the next section, the limitations of applying techniques developed for ordinary single level regression to multilevel data will be considered using a hypothetical example. This is followed by an analysis of a set of real educational data that considers the progress made by pupils from different ethnic groups. Then another example involving exploratory plots with binary response variables is considered. Finally, there is a discussion of the issues raised by this paper.

Exploratory plots and the ecology problem

One advantage of using preliminary plots for the IDA of multilevel data is that it can aid the understanding of the need for multilevel models. Although Aitkin and Longford (1986) proved the need for multilevel model by setting out the theory in a series of equations, this is only accessible to readers with a reasonable understanding of mathematical statistics. However, it is easy to demonstrate the need for multilevel models by considering some simple plots of hypothetical data sets.

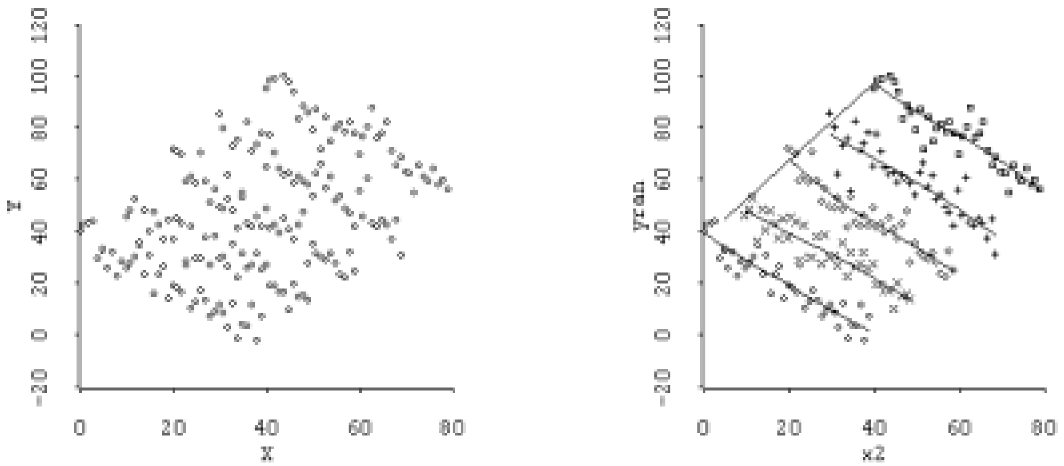
Ordinary scatterplots may be misleading for structured data. For example, Preece (1987) in a paper on statistical good practice gave an example of this problem with repeated measures data from a standard textbook. For multilevel data this can be illustrated with a rather extreme hypothetical example. A dependent variable Y has been plotted against an independent variable X in Figure 1(a). A LOWESS smooth has been added to the plot. This plot would suggest that there is a non-linear increasing relationship between X and Y. This plot, however, ignores the fact that the data come from five distinct groups and is misleading. (Note that this small data set was created for demonstration purposes. For formal modelling it would be more sensible to treat the groups as fixed effects and use analysis of covariance.) This is obvious when group information is added on to the plot as in Figure 1(b). In this plot, different symbols have been used for each group. The quadratic relationship suggested by the previous plot has also been added. However, after fitting a LOWESS smooth to each group, it is clear that within each group there is a negative linear relationship between X and Y.



(1a)) A LOWESS smooth ignoring groups **1(b) With a LOWESS smooth by group**

Figure 1: Example of hypothetical data

This hypothetical example demonstrates the need for considering the multilevel structure when carrying out preliminary analysis of multilevel data. This type of plot is also useful for identifying outliers. Consider a new hypothetical data set given in Figure 2(a). There is no evidence of any anomalies. In fact, this data set has been created with a large outlier. By adding the group information and plotting the LOWESS fits as in Figure 2(b), it is evident that there is a problem with group 5 (represented by the square symbols) because an outlier has distorted the smooth.



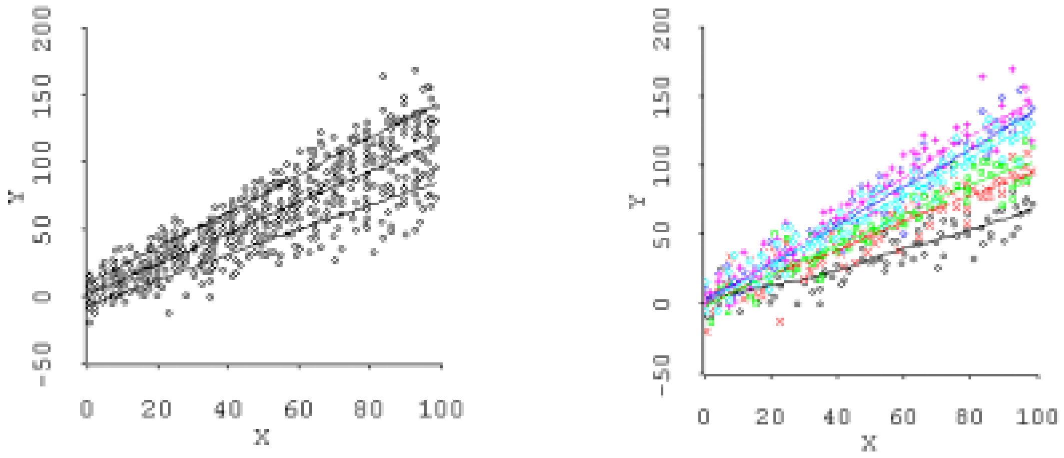
2(a) Hypothetical data set with 'hidden' outlier

2(b) 'Hidden' outlier data set with multilevel information added

Figure 2: Outliers in multilevel data

Another problem associated with ignoring structure when producing scatterplots is illustrated in Figure 3. In Figure 3(a) the multilevel structure has been ignored. A LOWESS smooth line has been added together with the variance smooths (Cook and Weisberg, 1999, pp. 51-52). This plot would suggest that the data are heteroscedastic, i.e., the variance function is changing with X. However, when smoothed lines are added by group it is clear that the shape of the cloud is determined by variation in the group-level slope.

When fitting models with random slopes, it can be useful to consider how to scale the independent variable. The magnitude of the intercept variance component varies depending on the choice of origin of the scaled variable. Sometimes it is useful to centre the independent variable on the mean or other meaningful value. Another option would be to select a point where it is at a minimum using an exploratory plot.



3(a) Ignoring multilevel structure

(3b) With multilevel structure

Figure 3: Example of spurious heteroscedascity

These hypothetical examples illustrate the need to consider the multilevel structure when plotting data. In the next section, an analysis of a real data set will be considered.

An example with data on school progress

This example uses data gathered by Haque (1999) as part of a PhD. A detailed multilevel analysis of these data can be found in Haque and Bell (in press). This is a small data set consisting of 958 pupils nested in twelve schools. These schools were selected because they had a high proportion of Bangladeshi pupils and are not representative of the population of English schools. The pupils in the study were assessed on two occasions. At fourteen years of age, the pupils involved in this study took national tests in English, mathematics and science. The results of these tests were converted into a single score. Two years later the pupils sat formal examinations in a range of subjects. The grades in the examinations were converted into points and a numeric mean grade was calculated (although this is the standard method for producing an aggregate measure of GCSE performance it is not necessarily the best (Bell, 2000)). This latter measure is the dependent variable in this analysis and the former is one of the explanatory variables used in the study. Although there was a considerable number of potential explanatory variables in the data set, this example only includes three: gender, ethnic origin and national test score. In particular, it is only intended to demonstrate the techniques and not to provide a definitive analysis of the data.

In all the examples described in this paper, the data sets were initially created in and manipulated using SAS. These data sets were subsequently converted in to *ARC* format using the SAS2ARC macro which is available from Andrzej Galecki's website (http://www-personal.umich.edu/~agalecki/). This data set was then analysed with *ARC.*

One the first things that can be done with this software is that the within group variation in the dependent and the explanatory variable can be investigated. In the screen dump below (Figure 3), four windows have been displayed which show boxplots for mean grade by school and ethnic origin and for national test score by school and ethnic origin. It is also possible to investigate the effect of ethnic origin (in this case, it is determined by place of birth of the pupil's mother). It is intended that, in the subsequent models, schools are to be treated as random effects and the ethnic origin is to be modelled as a series of dummy variables. In *ARC* the boxplot window normally includes a set of interactive controls. These have been turned off. (There are two potentially useful modifications that could be made to *ARC.* Firstly, it would be useful if the boxplots could be arranged in order of ascending medians, and, secondly, if the minimal boxplots suggested in Bell (1998) could be used.)

From the boxplots in Figure 4, it is clear that there is some variation between the schools. It is also apparent that there are differences in relative progress between the schools. For example, pupils from school S tended to perform better on the National tests than on the GCSE examinations relative to the other schools. It is also clear that there is considerable overlap in performance between the ethnic groups. There is also evidence that there are differences in relative progress between the groups. For example, the median performance for the UK group is highest on the National tests but not for the GCSE examinations. It is clear from these plots that there may be some interesting relationships that need to be investigated.

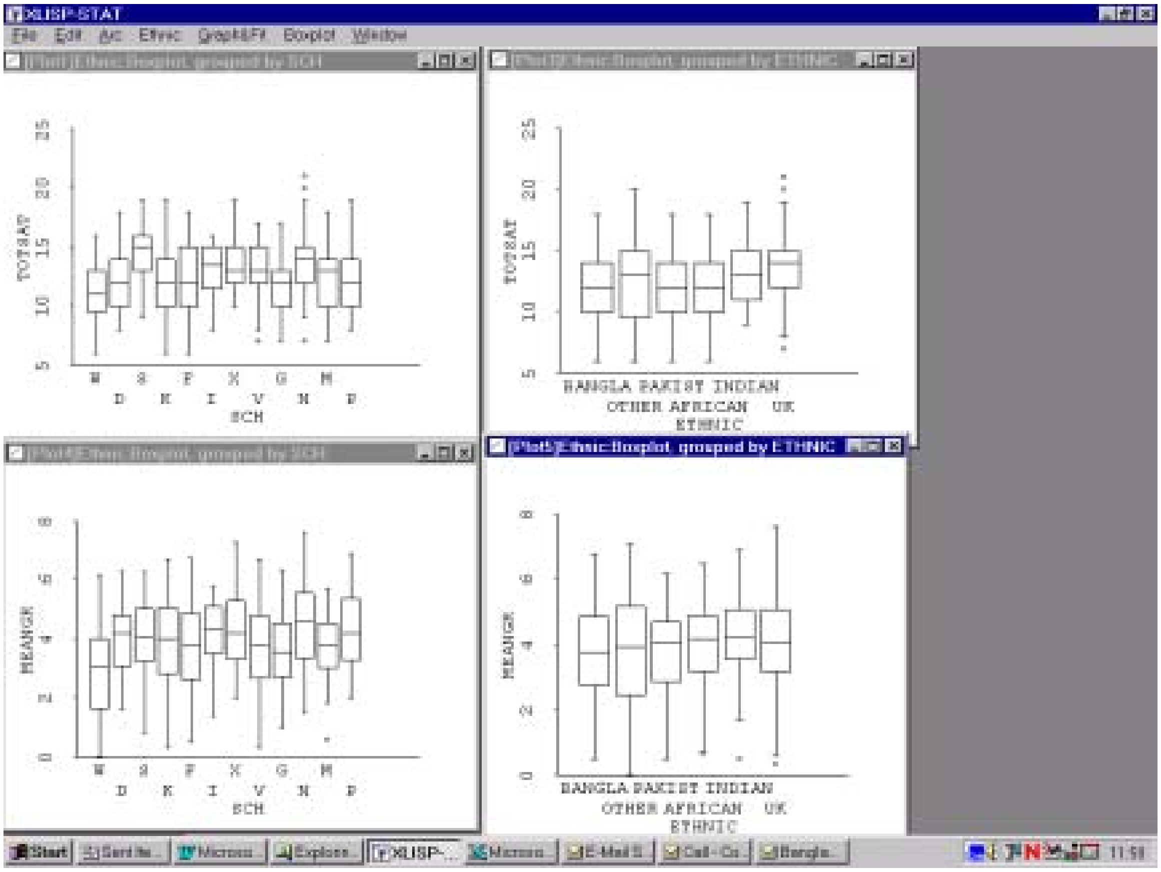


Figure 4: Boxplots for the ethnic origin data

The next step is to investigate the scatterplots for the relationships between mean GCSE grade and National test scores. Firstly, it is possible to identify the relationship between the two measures, i.e., whether it is linear or something more complex. To investigate this, a LOWESS smooth (Cook and Weisberg, 1999) has been added for each school. This smooth is calculated as follows:

For a particular point x_l , fitted value $\hat{y}_l$ is obtained by carrying out the following sequence of steps:

1. A smoothing parameter, *h*, a number between zero and one, is selected (i.e. by moving the slider on the plot).
2. The *hn* closest points to x_l are identified.
3. These points are then used to compute weighted least square estimates for the regression of *y* on *x*. ARC uses a triangular weight function. This function linearly decreases from a maximum value at x_l to zero at the end of the neighbourhood.
4. The fitted value $\hat{y}_l$ is calculated.
5. Steps (1)-(4) are then repeated for lots of values of x_l and the resulting coordinates are joined.

Secondly, it is also possible to investigate whether the variation is constant between groups. It should be recognised that these initial plots exaggerate the variability between groups. However, it should be noted that plots based on the shrunken estimates obtained after multilevel models have been fitted understate the variability. This is because all the school level parameters are shrunk. This will be discussed further later in the paper. From Figure 5, it is clear that there is a considerable degree of overlap between the points from different schools. This would suggest that the amount of school level variation is low.

A number of options can be used to manipulate the plots generated by *ARC*. Firstly, it is clear from the plot in Figure 5 that the TOTSAT score can only take integer values. This means that there are many overlapping points in the plot so some of the variation in the data is masked. To spread out the overlapping points, a technique known as jittering can be used. This involves adding a small amount of random variation to each of the points. Figure 6 is the same data as presented in Figure 5 with a small amount of jittering. In *ARC* moving the slide bar with the mouse controls this process. It is also possible by clicking on the triangle next to the slide bar to access three options that allow the jittering to be either vertical or horizontal or both. This plot gives a much clearer indication of the distribution of the points. Of less importance is that an option to rescale the plot has been used to make better use of the available space.

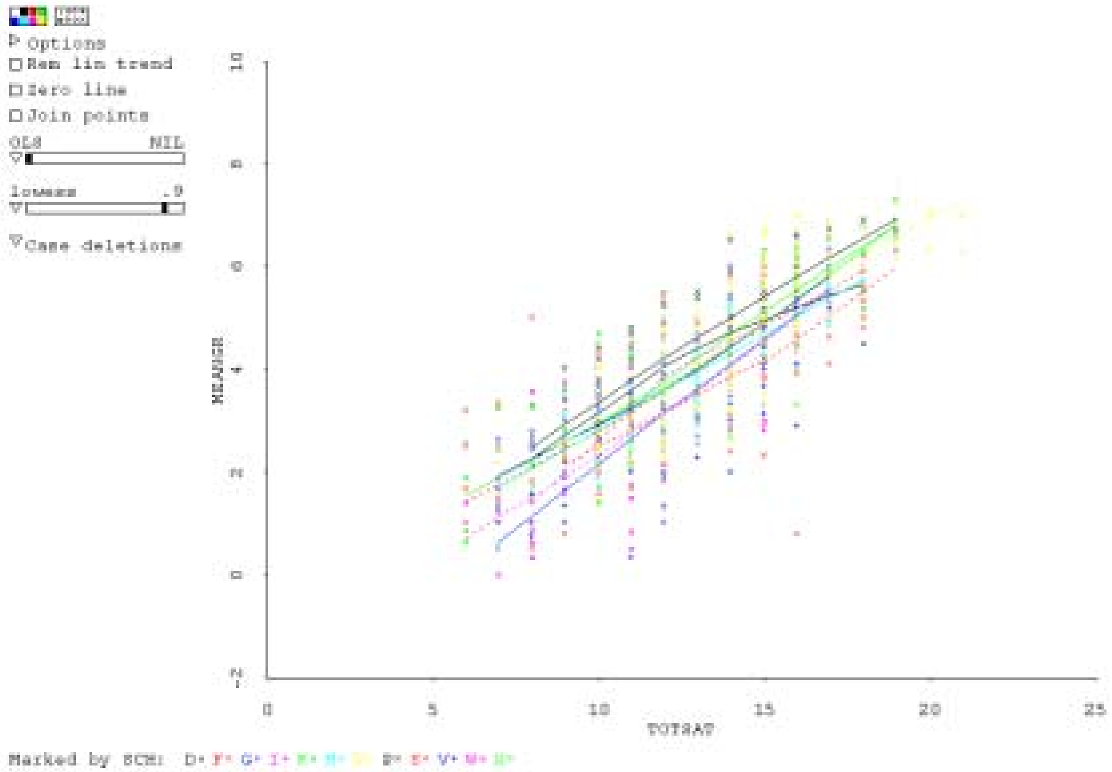
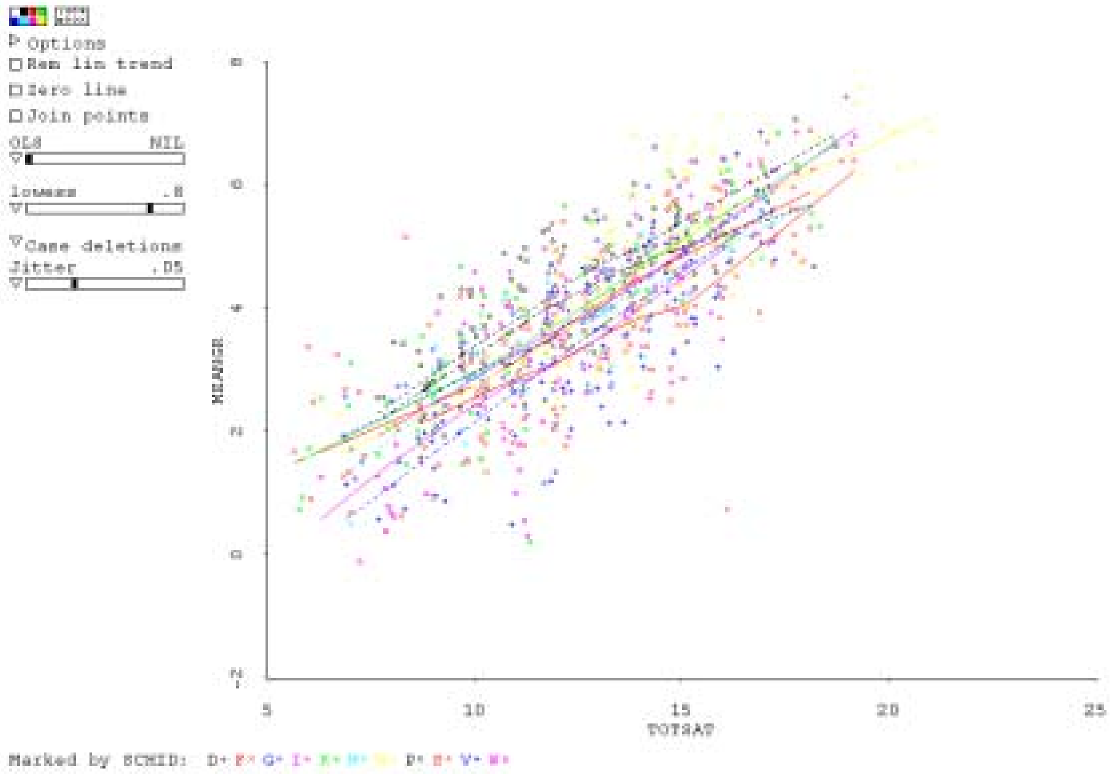


Figure 5: Scatterplot of mean GCSE grade by national test score grouped by school



OLS NIL
lowess .8
Jitter .05

Figure 6: Scatter plot of mean grade by national test score (TOTSAT) with jittering

By clicking on the labels on the legend, it is possible to produce a plot for just one school. This process is quick and if it is necessary to select a particular school because of the shape of the smoothed line it is easy to click on the labels that are the same colour as the line until the required centre is identified. Some of the colours on the screen may seem a little faint. It is possible to change the background to black by clicking on the “Options” option of the “2Dplot” menu. In the plot above, the legend is at the bottom of the plot and the number of labels that are visible is restricted. There are plans to put these labels into a separate window in a future version of *ARC.* (At the moment, it is best to use as short an ID as is possible). For example, one of the two schools (F or S) that have been coloured red has a line that seems to be deviating from the linear. Clicking on both F and then S reveals that it is school S. This plot is given below as Figure 7. For this plot, an OLS fit has also been added to provide a reference line to give some indication of the deviation from linearity (this is not the line that would be fitted by the multilevel model). It is possible to add reference lines to a plot of all schools but this is too confusing. The difference between this line and the smoothed line is relatively small.

In Figure 7, there is obviously one outlier. This individual performed well on the national tests and very badly on the GCSE examinations. This situation sometimes occurs in educational data when a candidate has been ill or affected by serious external problems. A number of options can be applied to this point. These options can be found in the ‘2DPLOT’ menu. In this plot, the point has been identified using the ‘Show label’ option. In this case, the label is the case number in the data set. It is also possible to set the label to a particular variable (e.g. a pupil identifier) using the ‘Set case names’ option in the data set menu. It is also possible to remove the point. This will be demonstrated later in this paper.

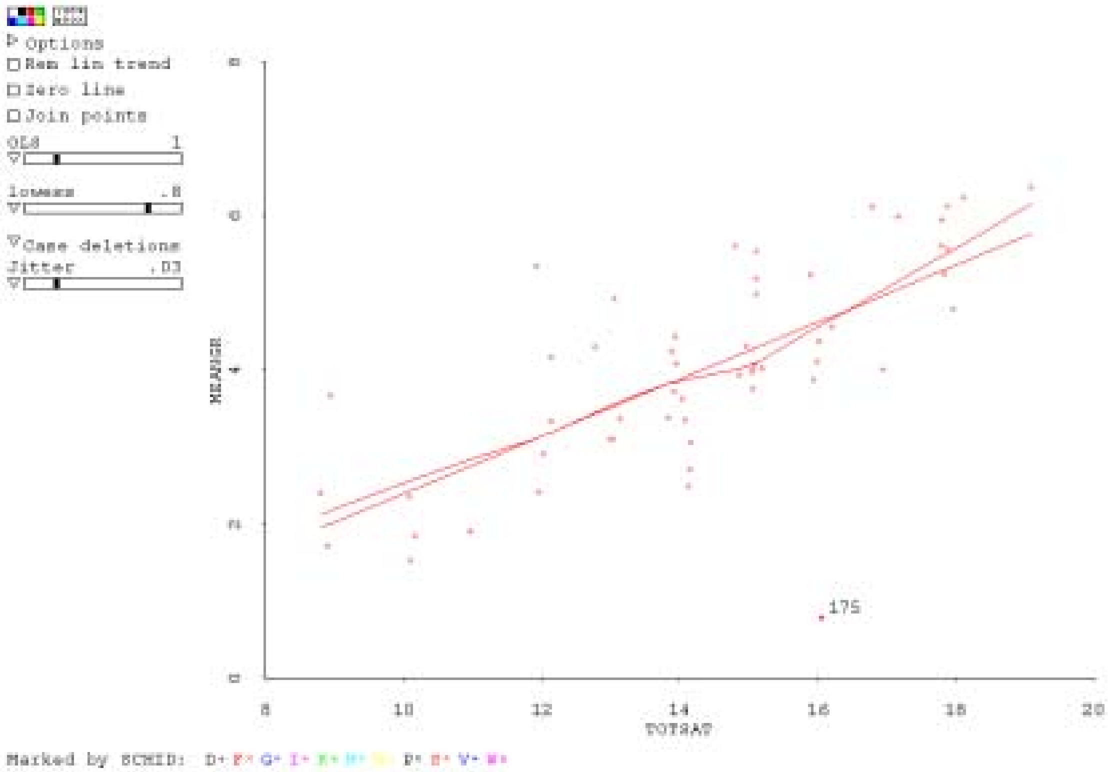
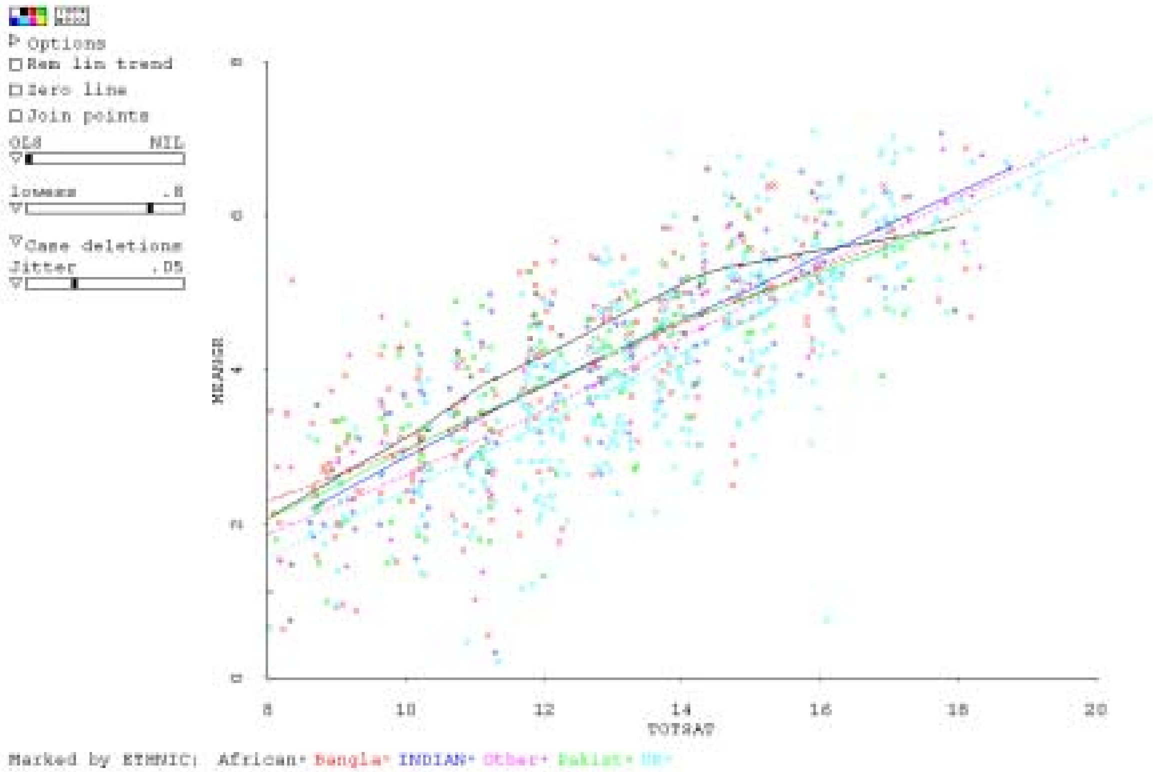


Figure 7: Scatterplot of mean grade by national test score for school s

The conclusions from all these plots would suggest that a multilevel model with a simple linear relationship between mean grade and national test score is appropriate. (It is not unknown for linear fits to be inappropriate for examination data because of floor and ceiling effects). In addition, the amount of school level variation is relatively low and there is no evidence of a large amount of variation in the slopes for different schools, which suggests that a random slopes model will not be appropriate.

The variable, ETHNIC, will be modelled using a series of dummy variables in the multilevel model. However, for the initial plots, it is useful to use a single text variable for marking the points, this text variable taking the values: African, Bangla, Indian, other, Pakist, and UK. Clicking the "Set marks" option of the "Graph&Fit" menu changes the variable used to mark the points in *ARC*. A window appears with a list of variables and the current selection for the marking variable. This is changed by clicking on this selection and then clicking anywhere on the list of variables. The selection box is then emptied. The new marking variable is then selected by clicking on it and then clicking on the selection box. The result of this change is given in Figure 8.



▷ Options

□ Rem lin trend

□ Zero line

□ Join points

OLS NIL

lowess .8

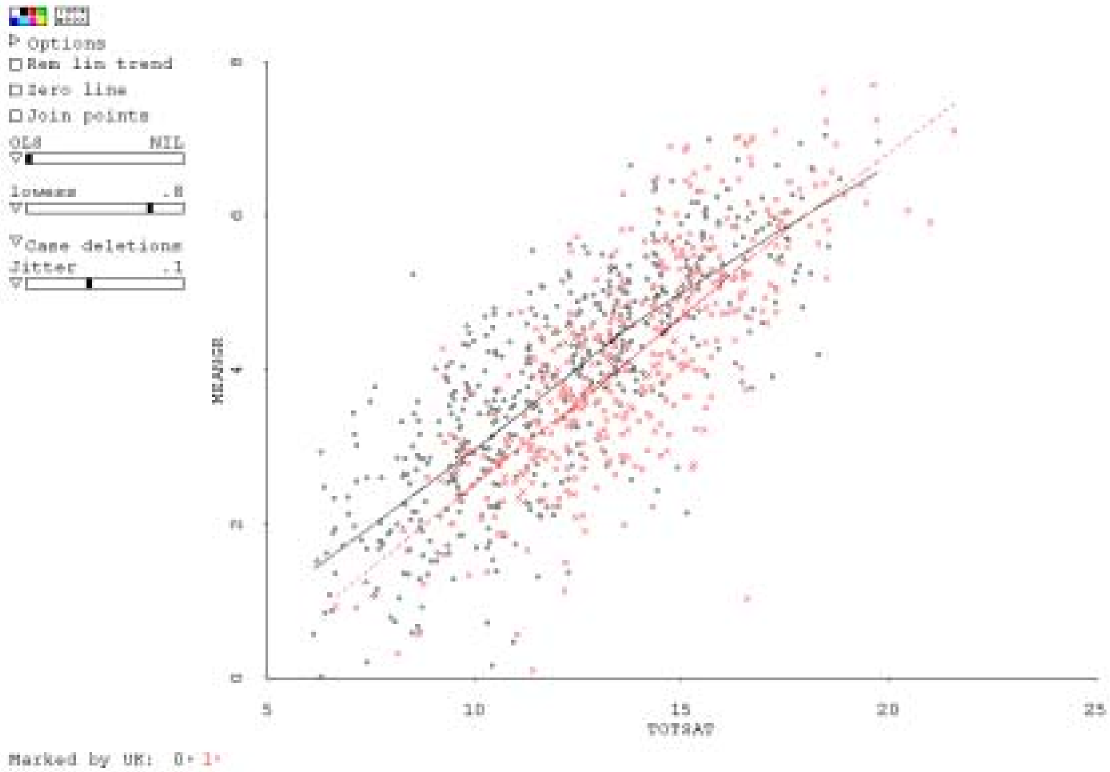
▽ Case deletions

Jitter .05

Figure 8: Scatterplot of mean grade by national test score marked by ethnic group

In Figure 8, LOWESS smoothed lines have been fitted for each ethnic group. It clear from the plot that there are more pupils whose mothers were born in Bangladesh in the bottom left of the plot compared with pupils born in the UK, i.e. pupils with a low level of performance on both the national tests and the GCSE examinations. The relationship between mean grade and national test score seems to be approximately linear for each of the Ethnic groups except those pupils who mother was born in Africa.

By using the dummy variables, it is also possible to look at how the distribution of points compare for one sub-group compared with all others. This is demonstrated in Figure 9 that uses a dummy variable UK that takes the value 1 if the pupil's mother was born in the UK and 0 otherwise. The plot clearly indicates that there are more 'UK pupils' in the top right of the plot. This indicates that they tended to start with higher scores on the national tests and went on to obtain higher grades on their GCSE examinations. This is a feature of the data that would not have been detected solely by looking at the parameters of a multilevel model.



▷ Options

□ Rem lin trend

□ Zero line

□ Join points

OLS NIL

lowess .8

▽Case deletions

Jitter .1

Marked by UK: 0 · 1 ·

Figure 9: Scatterplot of mean grade by national test score marked by UK dummy variable

Although the Figure 8 is confusing, by clicking on each label in the legend, it is possible to look at the results for each ethnic group separately. The plot for the African group is given in Figure 10. The shape of this plot seems to be determined by the results for one pupil who obtained a relatively good score on the national tests but did not perform as expected on the GCSE examinations. Unusual points at the extreme ends of axes have an undue influence on the shape of locally smoothed curves. A useful modification to *ARC* would be for a linked window to open containing only the data for that group.

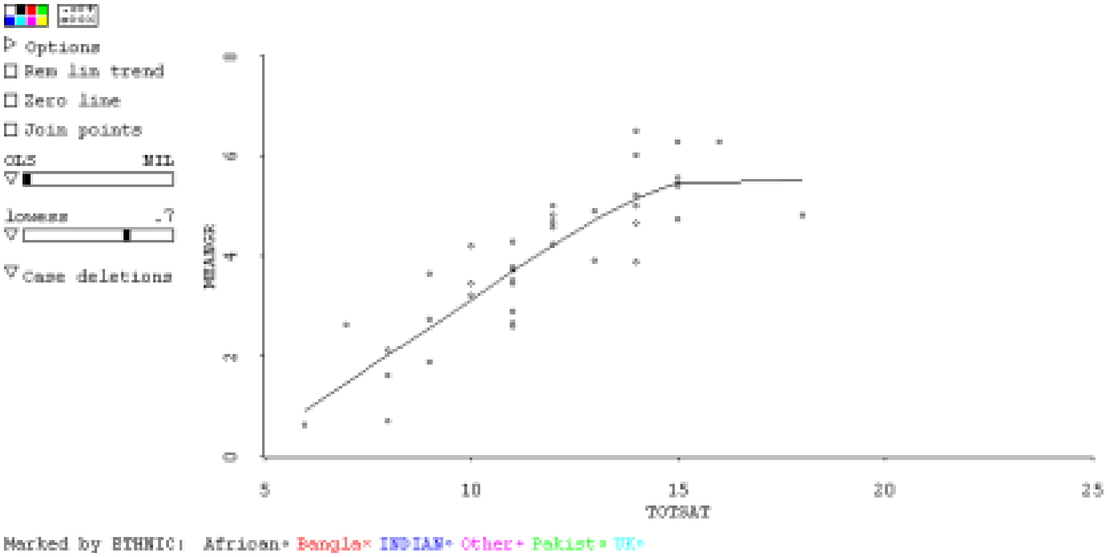


Figure 10: Scatterplot for the African group only

The effect of this candidate can be investigated by selecting this candidate with the mouse and removing the selection using the appropriate command in the "2DPLOT" menu. The LOWESS smooth can be refreshed by moving the slider. The result of this action is illustrated in Figure 11. The resulting smooth is approximately straight.

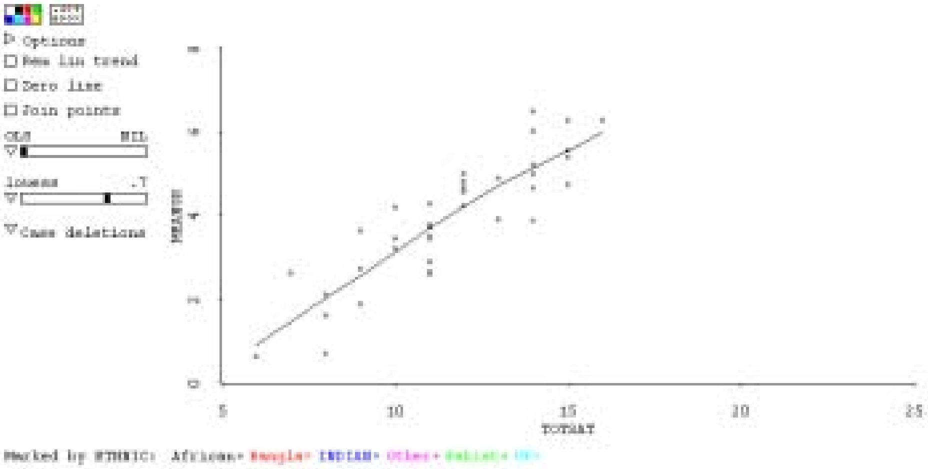


Figure 11: Scatterplot for African group with one point removed

Figure 8 is confusing with both lines and points presented. In Figure 12, the points have been hidden. This was achieved by selecting all the points with the mouse. In the "2DPLOT" menu there is an option that allows the colour of the selection to be changed. By setting the colour to white (i.e., the same as the background) the points can be hidden. In this plot, the LOWESS smooths have been removed by setting the slider to zero and the OLS pointer has been moved to one to give a linear fit. By clicking on the triangle next to the OLS slider and selecting "Fit by marks – general", lines for each ethnic group have been added. This plot suggests that it might be necessary to fit an interaction term for the Bangladeshi pupils to model the variation in the slopes. If a situation like this occurs with a group that is going to be treated as random then a parsimonious model might be fitted by creating one dummy variable rather than random slopes.

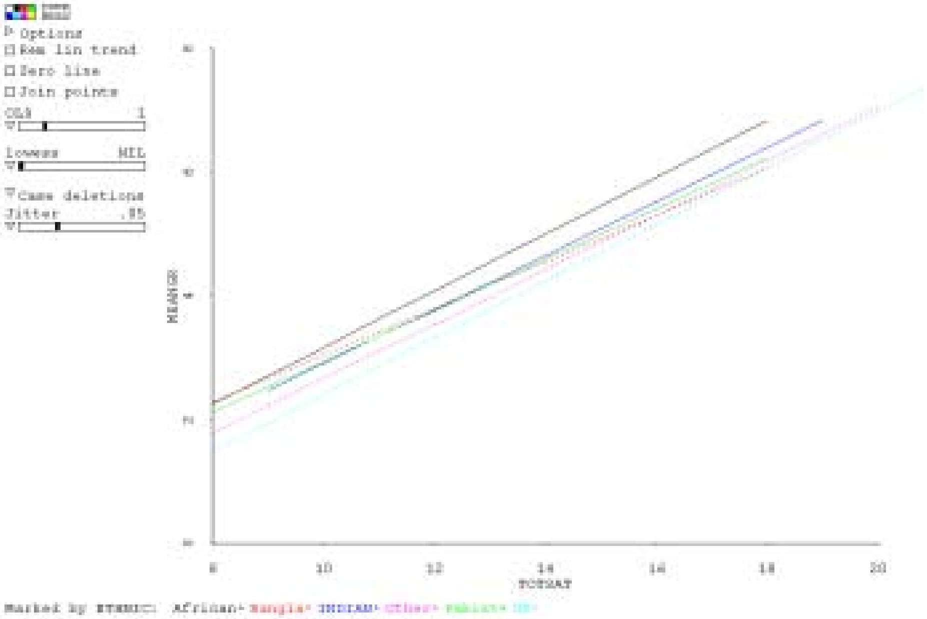


Figure 12: Scatterplot of mean grade by national test score marked by ethnic group (points hidden and OLS lines fitted)

Given that the results of the initial analysis of data, it is now possible to fit an appropriate multilevel model. Because *Xlisp-Stat* and *ARC* are freely available packages, it was decided to analyse the data using *MIXREG*, another freely available package. Any multilevel modelling package could be used for this stage. For example, the more complex analysis of these data described in Haque and Bell (in press) was carried using the *SAS* procedure Mixed (Singer, 1998). The analyses presented here are intended to demonstrate the use of an IDA with *ARC* and are not intended to provide a definitive interpretation of the data. The results of this analysis are presented in Table 1.

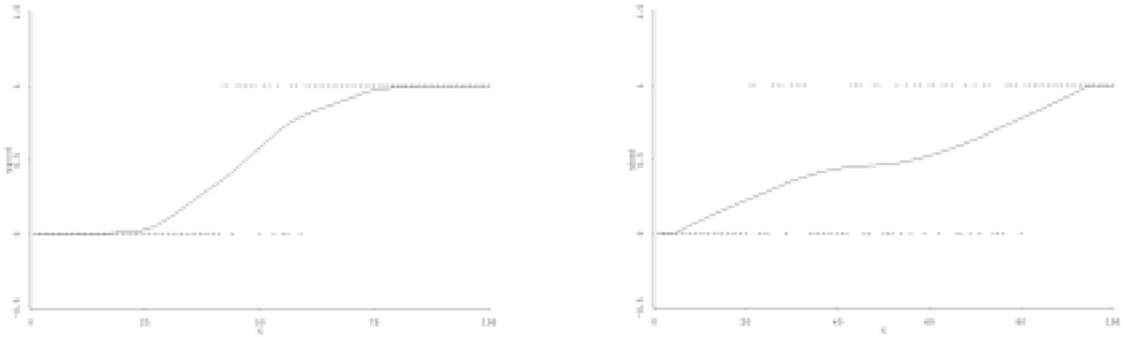
Table 1: Multilevel analysis of minority ethnic data set

Parameter	Model 1	Model 2	Model 3	Model 4
Fixed				
Constant	0.3857 (0.112)*	-1.763 (0.150)*	-2.224 (0.237)*	-2.018(0.170)*
SAT score	-	0.420 (0.010)*	0.454 (0.017)*	0.439 (0.012)*
African	0.052 (0.236)	0.743 (0.138)*	0.843 (0.635)	0.778 (0.138)*
Bangladeshi	-0.149 (0.140)	0.504 (0.084)*	1.558 (0.330)*	1.352 (0.285)*
Indian	0.359 (0.170)*	0.468 (0.100)*	0.675 (0.517)	0.474 (0.101)*
Pakistan	0.489 (0.096)*	1.031 (0.410)*	0.513 (0.097)*	-0.061 (0.164)
Other	0.242 (0.118)*	0.657 (0.457)	0.264 (0.012)*	-0.104 (0.198)
African*SAT	-	-	-0.003 (0.005)	-
Bangladeshi*SAT	-	-	-0.084 (0.025)*	0.069 (0.022)*
Indian*SAT	-	-	-0.022 (0.038)	-
Pakistan*SAT	-	-	-0.029 (0.034)	-
Other*SAT	-	-	-0.041 (0.031)	-
Random				
School	0.192 (0.070)*	0.048 (0.021)*	0.051 (0.021)*	0.051 (0.021)
Pupil	1.780 (0.087)*	0.643 (0.031)*	0.634 (0.031)*	0.635 (0.031)

Model 4, which was predicted by the initial analysis of the data, is satisfactory in that the significant terms have been identified. The residuals from this model could be analysed with diagnostic levels. The problem with analysing multilevel data as described in this section is that when the amount of data becomes large the plots can become unwieldy. In the next section, the *ARC* package will be used to perform an initial analysis of data with a binary response variable.

Exploratory plots for multilevel logistic regression

In this section, the use of *ARC* with binary data will be considered. It is worth considering how to use scatterplots with such data. It is still possible to fit LOWESS smooths with such data. If these curves look like logistic regression curves, then a linear logistic regression model is appropriate. It is also worth noting that unlike in ordinary regression the slope of the line indicates how good the fit is. Consider Figure 13 below. In the Figure 13(a) there is not much overlap between the 1's and the 0's and in the Figure 13(b) there is a great deal of overlap between the points. The LOWESS smooth is steeper for 13(a) compared with 13(b).



(a) **(b)**

Figure 13: Hypothetical example of LOWESS fits for binary data

For the second example, a multilevel data set was downloaded from the Multilevel Models website (http://www.ioe.ac.uk/multilevel/datapref.html). The data chosen for this section come from the Junior School Project (Mortimore et al, 1988). The original data set consisted of tests results for over 1000 students measured over three school years with 3236 records. For the purposes of this example, data from only the first year will be considered and two measures have been used. This reduced set of data consists of 1,129 pupils nested in 49 schools. The first is the Ravens test in year 1 and is an ability measure. The second is the English mark at the end of the year. The latter was converted into a binary variable by arbitrarily setting a pass mark of 72. In addition, before importing the data into *ARC* an additional variable that consisted of the number of pupils for each school was created. The relationship between the RAVENS test score and the binary variable, ENGLPASS, can be investigated using box plots (Figure 14). This figure shows that pupils with higher RAVENS test scores tend be more likely to pass the English test suggesting that that a logistic regression relationship might be appropriate.

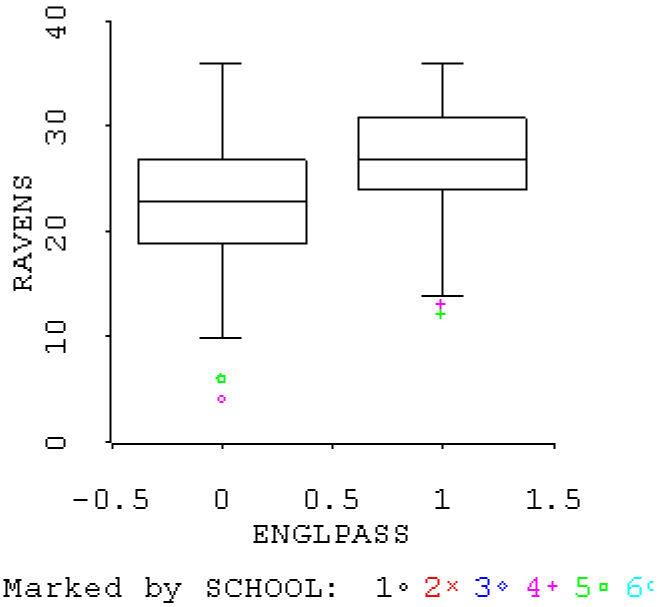


Figure 14: Boxplots of Ravens Test scores by ENGLPASS

Binary data can also be investigated using a scatterplot. In an ordinary scatterplot, there is obviously a large degree of overlap because the dependent variable only takes two values. The solution to this problem is to use vertical jittering. This has been done in Figure 15. Because there is a large number of points, the level of jittering is high. Although it can be useful to jitter plots to investigate the spread of points, this process does influence the LOWESS smooth. In the previous example with a continuous independent variable, the effect was negligible for the amount of jittering used. However, this is not necessarily the case with a binary dependent variable because the amount of jittering required to separate the points tends to be larger for binary data and the jitter slidebar should be set to zero before producing the LOWESS smooths. It is possible to look at individual schools by clicking on the identification numbers for individual schools (unfortunately only the first 26 schools can be viewed in this way because the rest of the legend extends beyond the edge of the window).

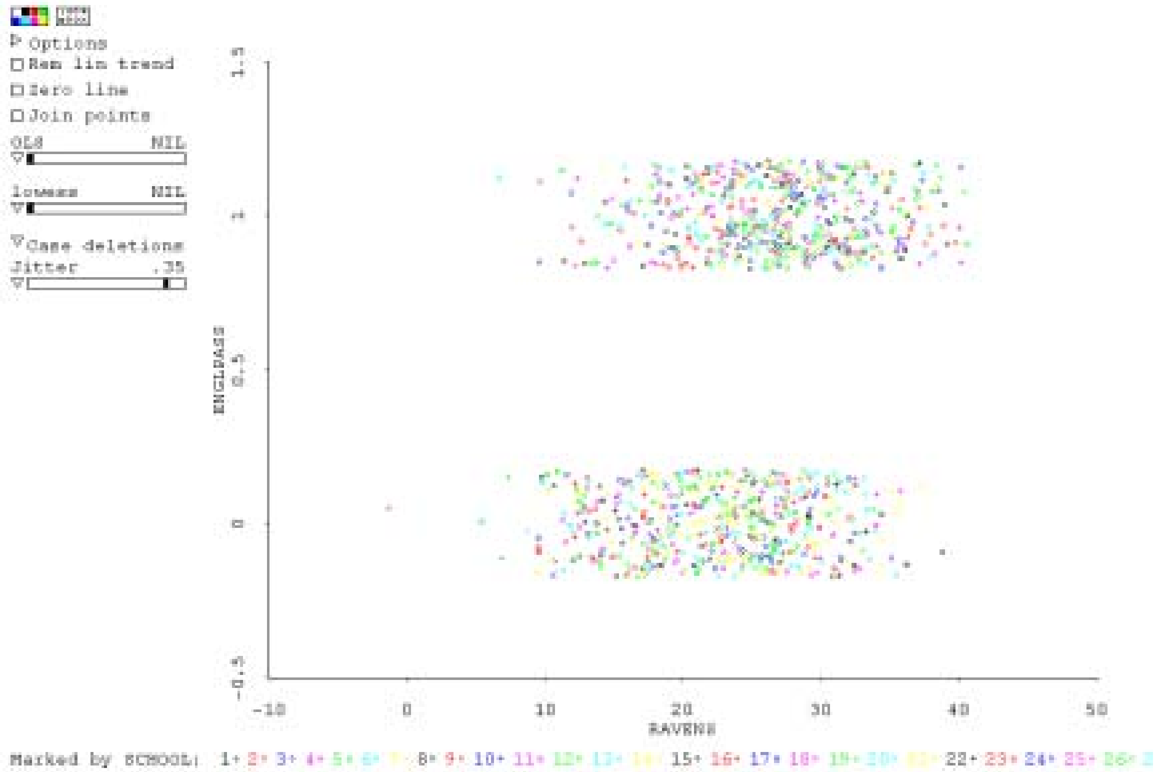


Figure 15: Scatterplot of ENGLPASS against RAVENS with vertical jittering

In Figure 16, ENGLPASS (the new binary variable) has been plotted against the RAVENS score and a LOWESS curve has been added for each school. There are two problems with this plot. Firstly, there is not enough room for all the schools in the legend and, secondly, this plot is frankly a confused mess. From this plot, it would seem that there is an inverse relationship for some schools. However, some of the schools are very small. LOWESS curves are not really meaningful when they are based on a small amount of data. It is worth noting that when a multilevel model is fitted, the predictors for groups with only small numbers of individuals will be shrunk most. This is also a particular problem with binary data because a misfitting point is much more influential given the dependent variable can only take two values. This is illustrated in Figure 17 that consists of the data for school 19. Only 13 pupils attended this school. It is clearly not sensible to try and interpret any form of locally smoothed line using so few data points.

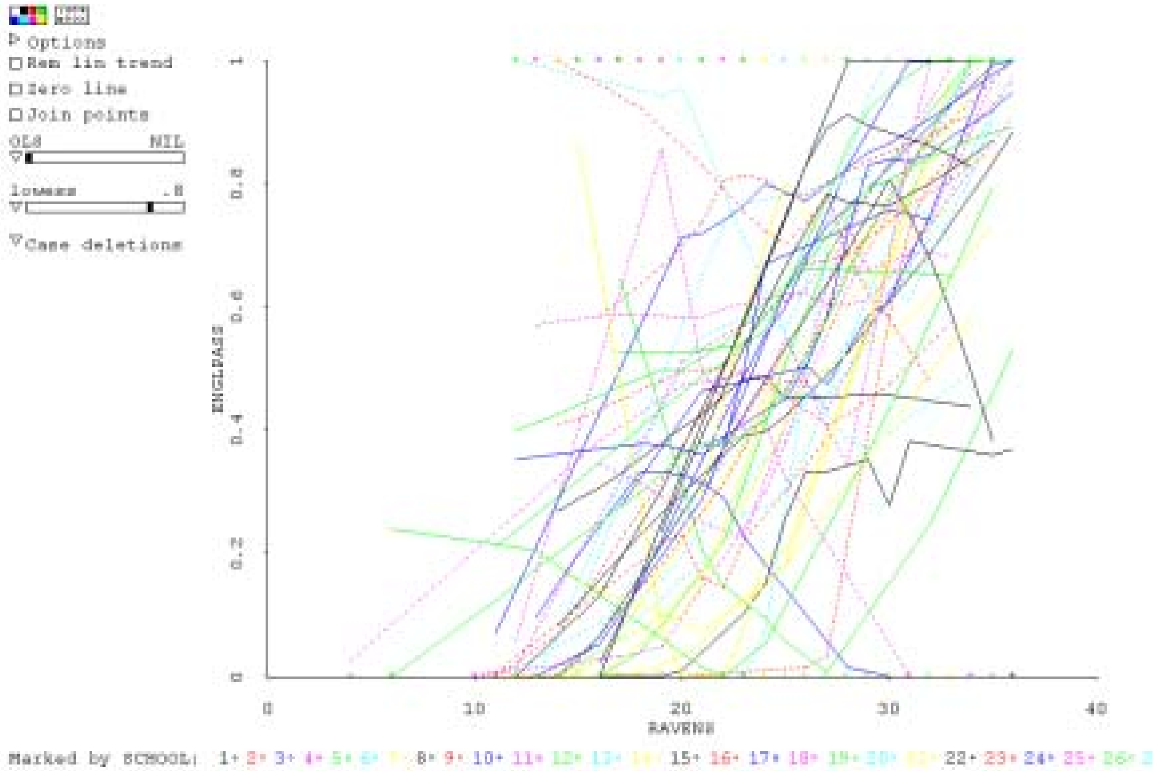


Figure 16: Plot of ENGLPASS by RAVENS marked by school

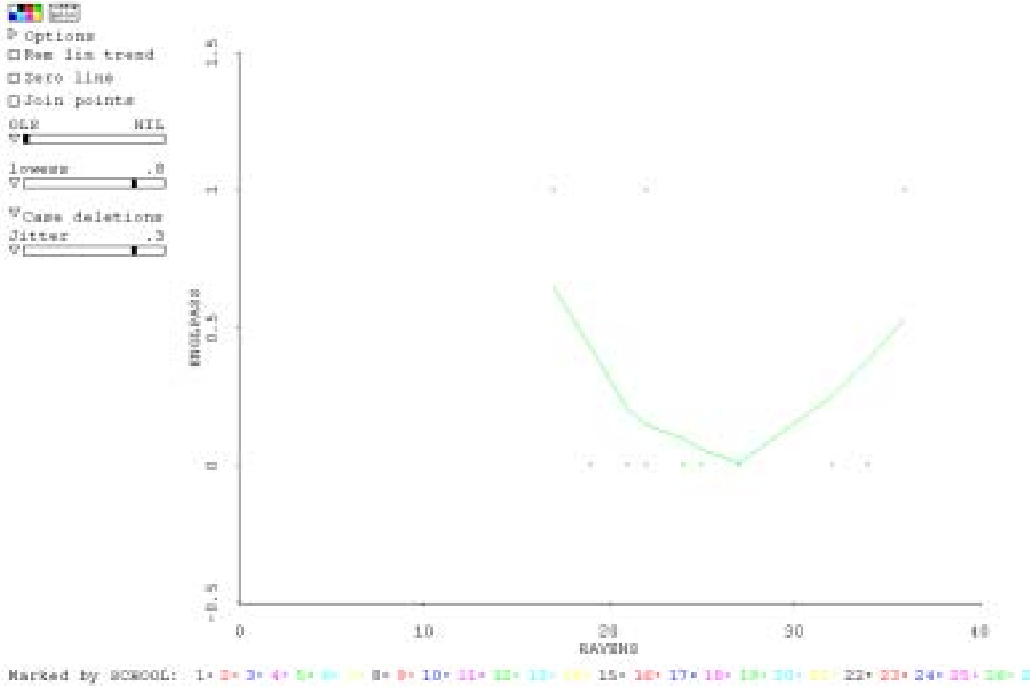


Figure 17: Plot for school 19

Given that the reason for producing this plot is to investigate whether there is variation in the slopes, it would seem sensible to consider only those that are well defined because they are generated by the larger schools. A subset of the data was created by selecting only those schools with at least 25 pupils. This reduces the number of schools under consideration to 19. These data have been plotted in Figure 18. This plot is not as confused as Figure 16. From this plot, it would seem that there is evidence that a random slopes model might be appropriate. There is also one potential problematic school. This is school 29 and the results for this school have been given in Figure 19. Plots for all the schools were examined by clicking on each school identifier. These plots support the hypothesis that random slopes should be considered for this set of data.

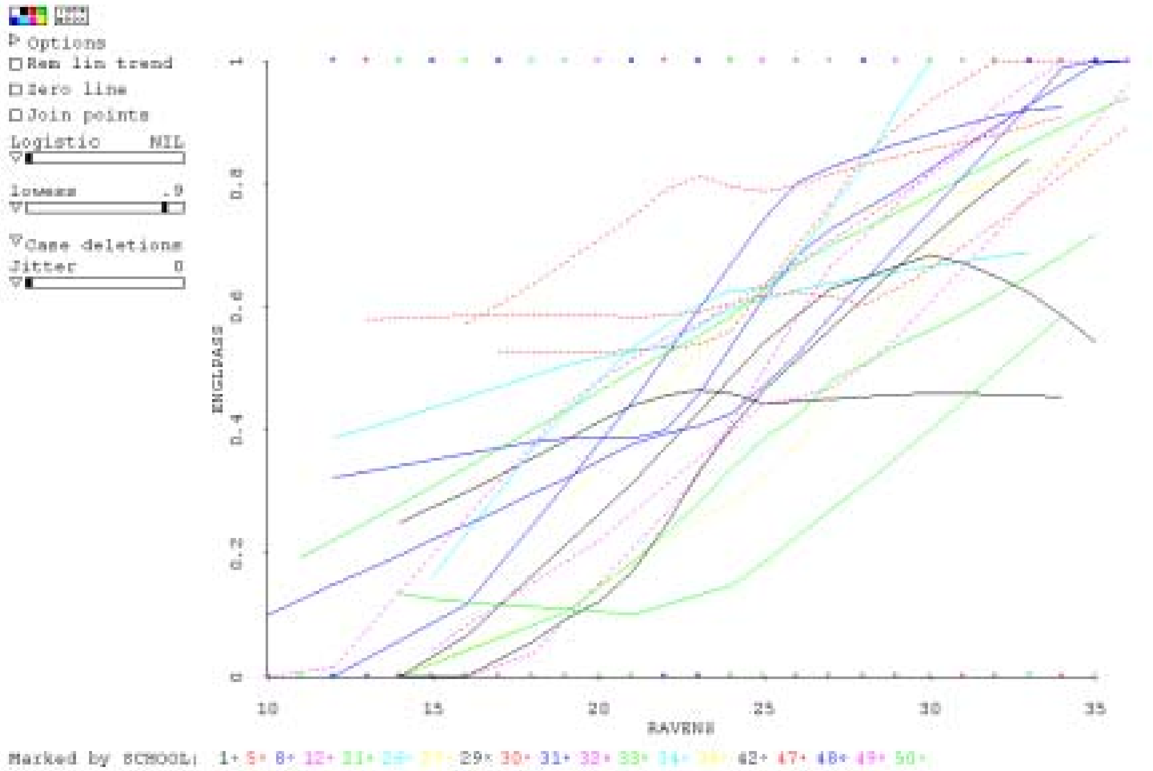

▷ Options
□ Rem lin trend
□ Zero line
□ Join points
Logistic NIL
lowess .9
▽Case deletions
Jitter 0

Figure 18: Plot of ENGLPASS against RAVEN for school with entry larger than 25

The plot for school 29 demonstrates the problem with LOWESS smoothing and binary data. A relatively small number of points has produced an unusual relationship. It is clear that this type of initial data analysis is only appropriate if there is a relatively large number of individuals in a group.

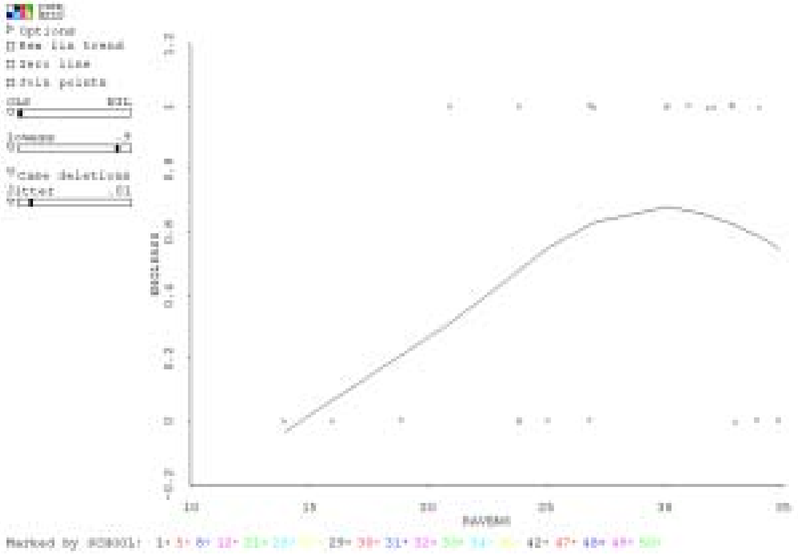
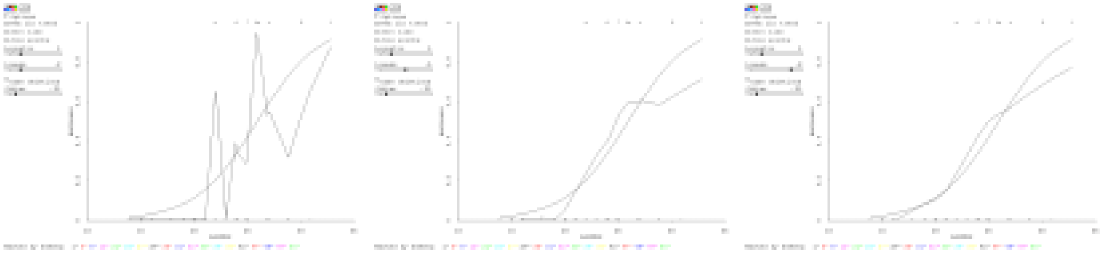

▷ Options
□ Rem lin trend
□ Zero line
□ Join points
OLS NIL
lowess .9
▽Case deletions
Jitter .01

Figure 19: Lowess smooth for school 29

When using LOWESS smoothing with binary data it is more useful to use higher values of the smoothing parameter than would be necessary for continuous variables. This is illustrated by Figure 20. This shows the effect of choosing different values of the smoothing parameter for one school. Given the relatively small group size and the binary nature of the data, it is clear that larger values of the parameter are more appropriate.



(a) 0.2 (b) 0.5 (c) (0.8)

Figure 20: Examples of different values of the smoothing parameter with binary data

There is a problem with using the LOWESS smooth for binary data. The smoothed line can take values less than zero and greater than one. However, there is a better method of smoothing for such data called local likelihood logistic (LLL) regression described Bowman and Azzalini (1997). They produced code for the statistical package *S*. This has been ported to *ARC* by Luca Scrucca (http://www.stat.unipg.it/luca/xlisp-stat/index.html).

This form of smoothing involves apply weights to a log-likelihood of the form

$$l_{[h,x]}(\alpha,\beta)=\sum_i l_i(\alpha,\beta)w(x_i-x;h)$$

where $l_i(\alpha,\beta)$ is the contribution to the usual log-likelihood from the *i*th observation, i.e.

$$l_i(\alpha,\beta)=y_i\log\left(\frac{p_i}{1-p_i}\right)+\log(1-p_i)$$

and the function $w(y-y_i;h)=\phi(y-y_i;h)$ where $\phi(z;h)$ denotes normal density function in z with mean 0 and standard deviation *h*. (Note that this is different from the LOWESS smooth used by *ARC*. However, the exact shape is not regarded as critical provided it is symmetric about 0). p_i denotes the probability of a 1 at x_i and the logit link function is assumed, i.e.

$$\text{logit}(p_i)=\log\left(\frac{p_i}{1-p_i}\right)=\alpha+\beta x_i \qquad (i=1,\ldots,n).$$

Local estimates $(\hat{\alpha},\hat{\beta})$ are obtained by maximising $l_{(h,x)}(\alpha,\beta)$ and the fitted value $\hat{m}(x)$ is calculated as follows:

$$\hat{m}(x)=\frac{\exp(\hat{\alpha}+\hat{\beta}x)}{1+\exp(\hat{\alpha}+\hat{\beta}x)}.$$

This process is then repeated for a sequence of *x* values and resulting pairs of $(x,\hat{m}(x))$ are plotted. Obviously this procedure is computer intensive. The method can also be applied to other link functions.

This *ARC* add-in was used to produce Figure 21. LLL is a much more complex procedure than LOWESS is and there was a noticeable delay in calculating the smoothed lines. However, this is less of a problem when plots for individual schools are considered.

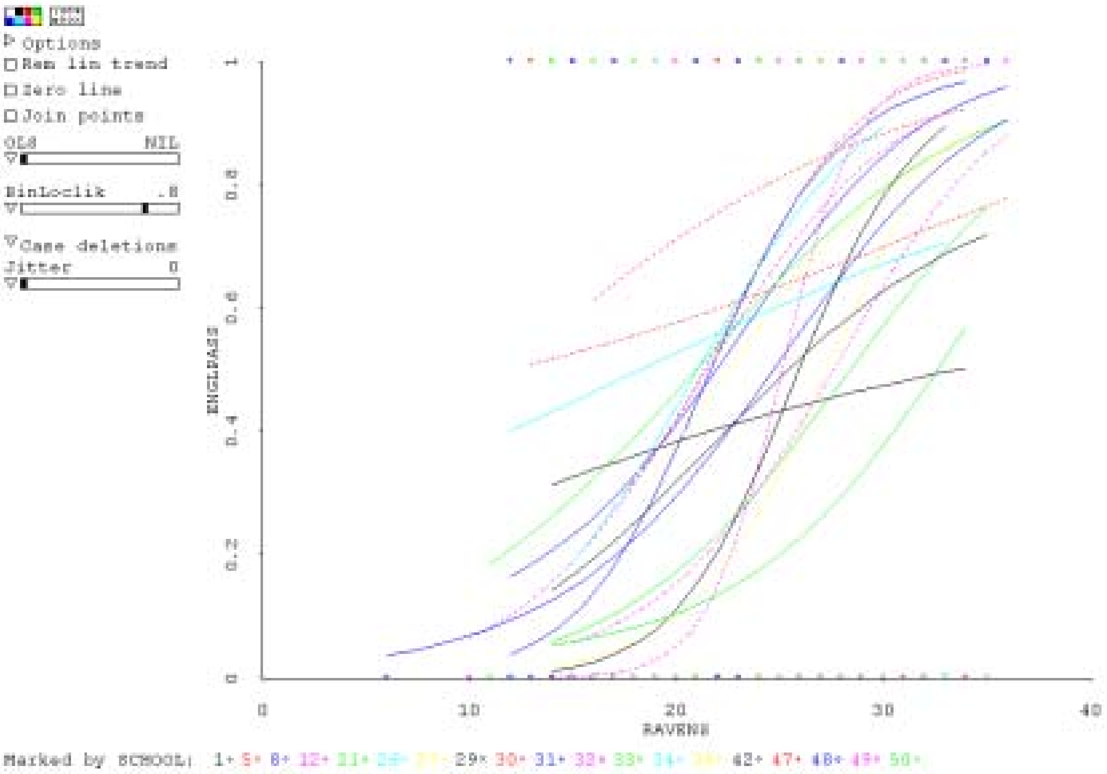

▷ Options
□ Rem lin trend
□ Zero line
□ Join points
OLS NIL
BinLoclik .8
▽ Case deletions
Jitter 0
1
0.8
0.6
ENGLPASS
0.4
0.2
0
0
10
20
30
40
RAVENS
Marked by SCHOOL: 1+ 5+ 8+ 12+ 13+ 26+ 27+ 29× 30+ 31+ 32+ 33+ 34+ 38+ 42+ 47+ 48+ 49+ 50+

Figure 21: Local likelihood logistic (LLL) regression smooths for the reduced data set

It is easier to compare the local likelihood regression smooth with the LOWESS smooth by considering the results for individual schools. The two smoothed lines are presented in Figure 22. Confidence intervals for the two smooth lines have been added. For the LLL the interval is ±2 SD and for the LOWESS smooth the interval is ± 1 SD. As expected, the LLL smooth is better than the LOWESS smooth. The spread of the variance function for the LOWESS smooth is to be expected given that a procedure for a continuous dependent variable has been applied to a binary variable. However, given that in this application, the objective is explore possible relationships before fitting a more formal model it is arguable that the LOWESS smooth is sufficient to gain a reasonable idea of the quality of the fit.

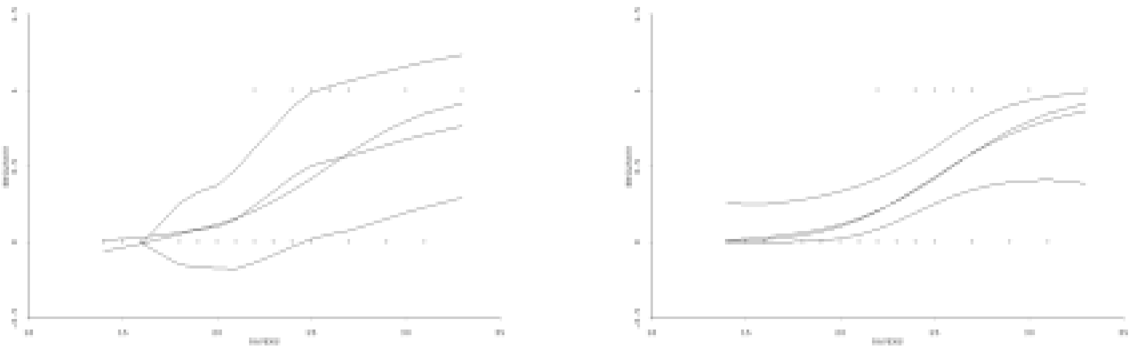


Figure 22: Comparison of the LOWESS smooth and LLL smooth

A series of multilevel models were fitted with MIXOR (Hedeker and Gibbons, 1996) and MLwin. The results of these analyses are presented in Table 2. For the purposes of analyses, the Ravens score was standardized. Results 1 and 2 were generated using MIXOR. This program uses numerical integration and this method produces a deviation statistic that can be used in hypothesis testing. Results 2 come from a model that includes random slopes while results 1 come from a model that does not. For the random slopes model MIXOR is sensitive to the scale of the independent variable and would not produce estimates for the original RAVENS variable). The test for the random slope is not significant (note the z statistics and the p-value for the estimated standard deviations given by MIXOR should be ignored). The program MIXOR calculates the estimated standard deviations ($\hat{\sigma}$) and the associated standard error as random parameters. For comparative purposes, the variance components ($\hat{\sigma}^2$) were calculated and the following approximation was used to calculate the associated standard errors (Snijders and Bosker, 1999):

$$\text{s.e.}\left(\hat{\sigma}^2\right) \approx 2\hat{\sigma}\,\text{s.e.}\left(\hat{\sigma}\right)$$

Unfortunately this is only a valid approximation if the relative standard error $\left(\text{s.e.}\left(\hat{\sigma}\right)/\hat{\sigma}\right)$ is small (e.g., less than ¼).

For the MLwin, the same models were fitted. Results 3 and 4 were obtained using second PQL and results 5 and 6 were obtained using the nonparametric bootstrap procedure. For the final analysis, results 6, the bootstrap diagnostics for the random slopes parameter were calculated. The kernel density plot is given as Figure 23 and the diagnostic statistics were as follows:
Posterior mean = 0.104 (0.0089), SD ~0.076, mode ~ 0.094
Quantiles: 2.5% ~ 0.000, 5% ~ 0.000, 50% ~ 0.100, 95% ~0.222, 97.5% ~0.228
It is clear that the random slopes parameter is not significantly different from zero but also that there is a lack of power for detecting such a parameter.

For the same models, the results of the three estimation methods give very similar results with the differences in estimates being of no practical significance.

Table 2: Parameters estimate for JSP model

Program Parameter	MIXOR Results 1	MIXOR Results 2	MLwin Results 3	MLwin Results 4	MLwin Results 5	MLwin Results 6
Fixed						
Constant	0.03 (0.10)	0.02 (0.11)	0.04 (0.10)	0.02 (0.11)	0.04 (0.12)	0.02 (0.01)
Raven (st.)	0.92 (0.07)	0.97 (0.11)	0.93 (0.08)	0.97 (0.10)	0.92 (0.09)	0.96 (0.11)
Random						
School	0.27 (0.11)	0.29 (0.12)	0.28 (0.10)	0.29 (0.11)	0.29 (0.12)	0.28 (0.09)
Cov.	-	-.11 (0.15)		-.07 (0.07)		-.06 (0.08)
Raven (st)	-	0.10 (0.09)		0.12 (0.09)		0.10 (0.08)
-2log-Lik.	1358.088	1355.145				

The explained variance was calculated using the procedure given in Snijders and Bosker (1999, p. 225-226). This quantity ($R^2_{dicho} = 0.21$) was relatively small. This is as expected from the exploratory plots.

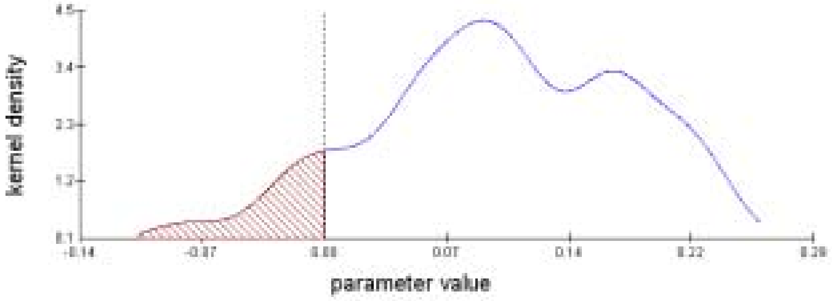


Figure 23: Kernel density plot from MLwin for random slopes parameter

This second example demonstrates that exploratory data analysis can be carried out for multilevel logistic regression. The IDA suggest that a random slopes model might have been appropriate, However, in the more formal analysis the random slopes parameter was not significantly different from zero.

The fitted values for the non-significant random-slope model have been plotted using MLwin in Figure 24. This shows that for the majority of the schools the fitted slopes are very similar. However, given that some of the slopes are for very small schools this is to be expected given the amount of shrinkage. It can be useful to present the results in terms of estimated probabilities rather than fitted values of logits. This is shown in Figure 25. Although the random slopes model is not significant, there might be a case for investigating the two extreme schools.

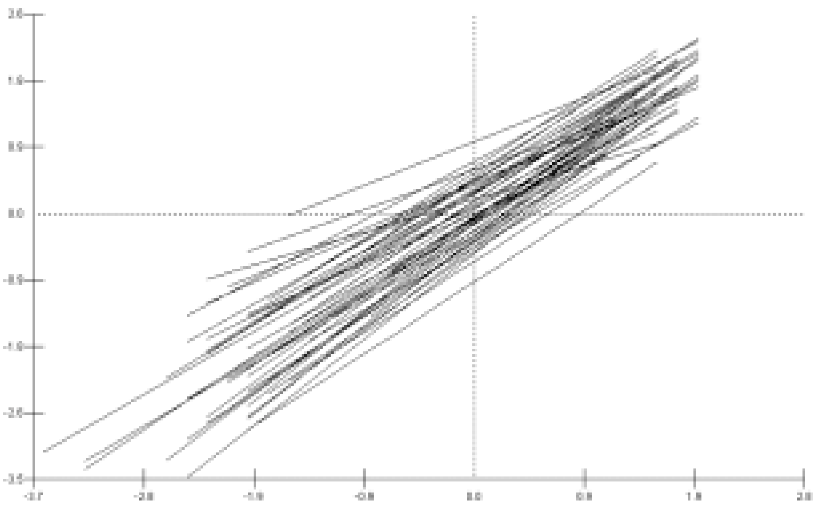


Figure 24: Fitted values for the random slopes model as generated by MLwin

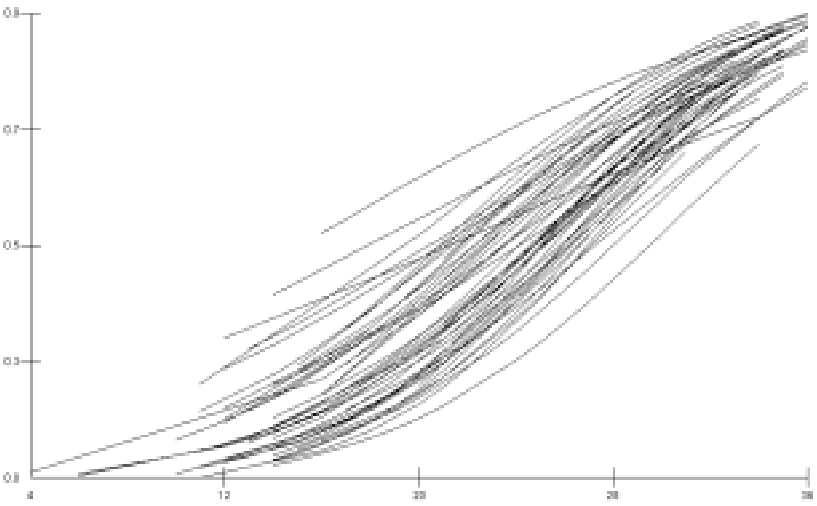


Figure 25: Estimated probabilities by Ravens score as generated by MLwin

For comparative purposes, the estimated probabilities for the larger schools based on the analysis carried out with MIXOR and plotted with *ARC* are presented in Figure 26. Note that only a small degree of smoothing is necessary for joining up the points. It can be argued that including the lines for small schools which are very highly shrunk is not helpful in visualising the data. Using this plot it is possible to identify particular schools by using the colours and the symbols or by clicking on the legend to get plots for individual schools.

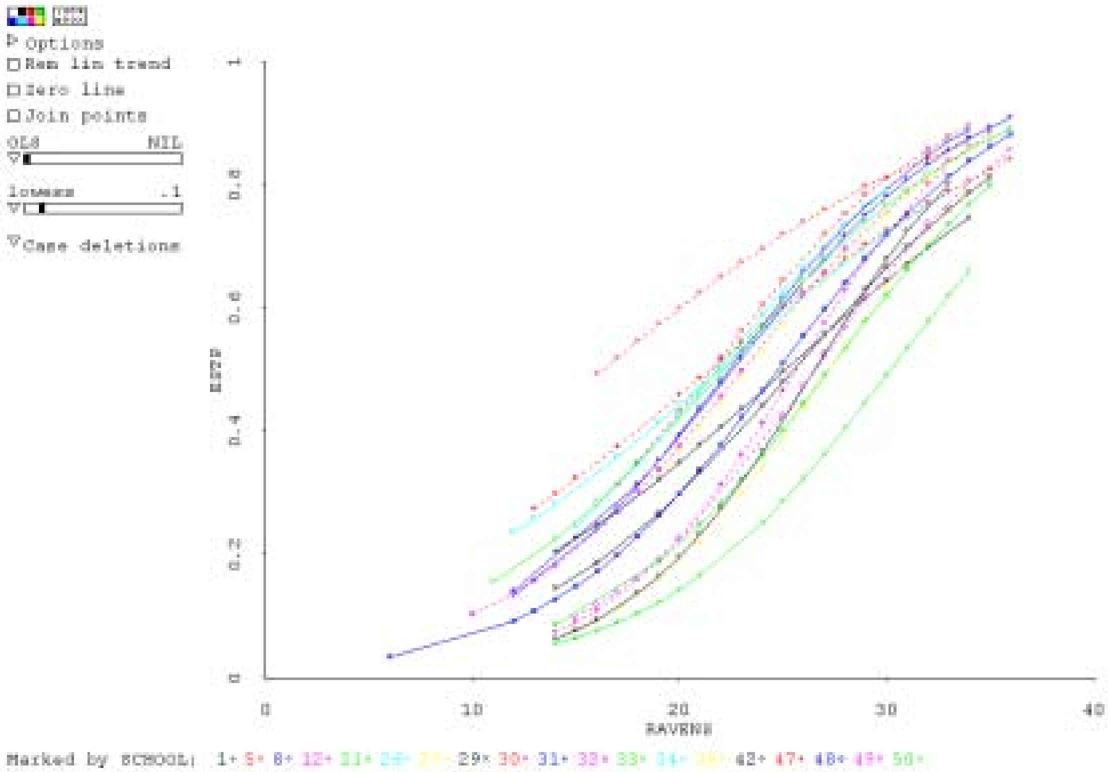


Figure 26: Estimated probabilities for school with at least 25 pupils plotted by XLISP-STAT

Figures 24–26 do not give a true impression of the variability of the schools. This can be illustrated by considering the magnitude of the random slopes. One way of visualising the random slopes is a ‘bundle’ plot (which resembles a bundle of sticks tied in the middle). The vertical axis of this plot is a predictor of the form $\text{logit}(\hat{y}) = (\beta_2 + u_{2j})x_{ij}$, i.e. based on the slope terms only and the x-axis is standardised x_{ij}. This results in Figure 27. On this figure lines have been added to indicate the 95% confidence interval using the estimate of the random slope variance component and assuming a normal distribution. These are the solid black lines with no points marked on them. This demonstrates the problem of shrinkage. Although shrinkage estimators are better for predicting individual values they are not satisfactory for generating plots that give a reasonable representation of the variability slope.

This plot indicates that the magnitude of the estimated parameter is large enough to be of practical importance. This was the conclusion from the IDA. However, the significance test results from the formal model was not significant. This indicates that the data is not large enough to give sufficient power to detect potentially important variation. This is an important point. While an IDA can be sufficient for demonstrating the lack of variation between groups, it is not sufficient for determining whether there are statistical differences. Data analysis is about finding useful and important patterns in data. The role of statistical inference is to determine whether the observed patterns have occurred by chance.

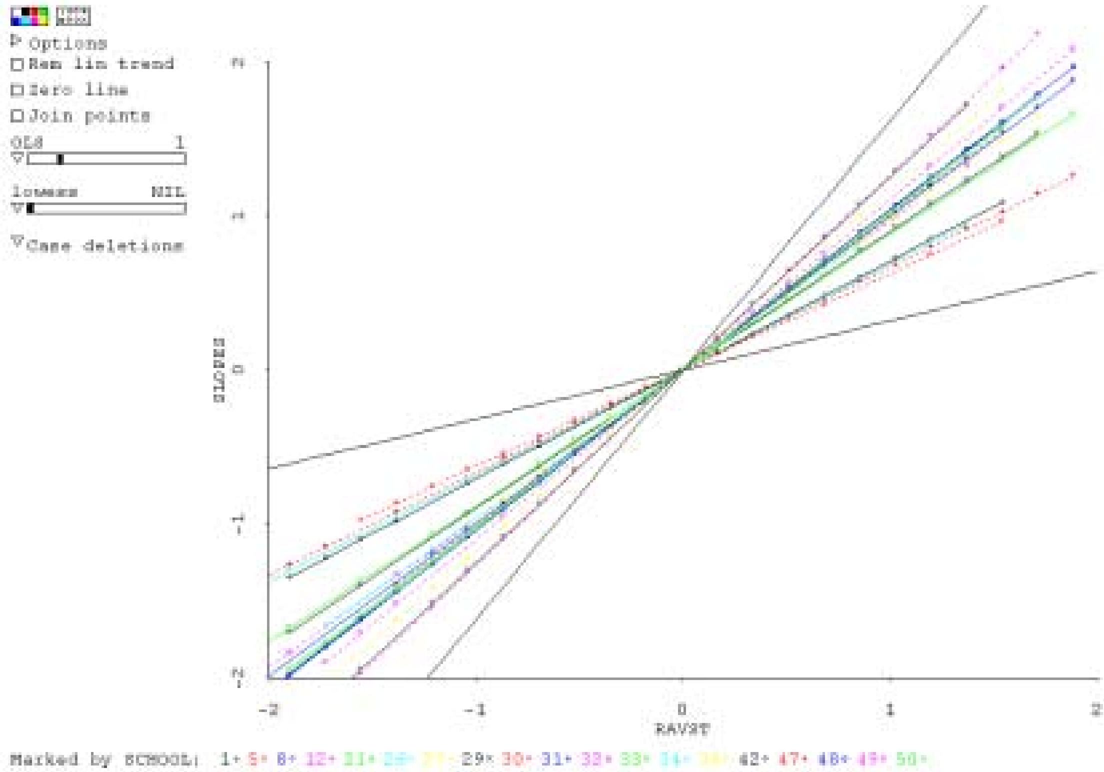
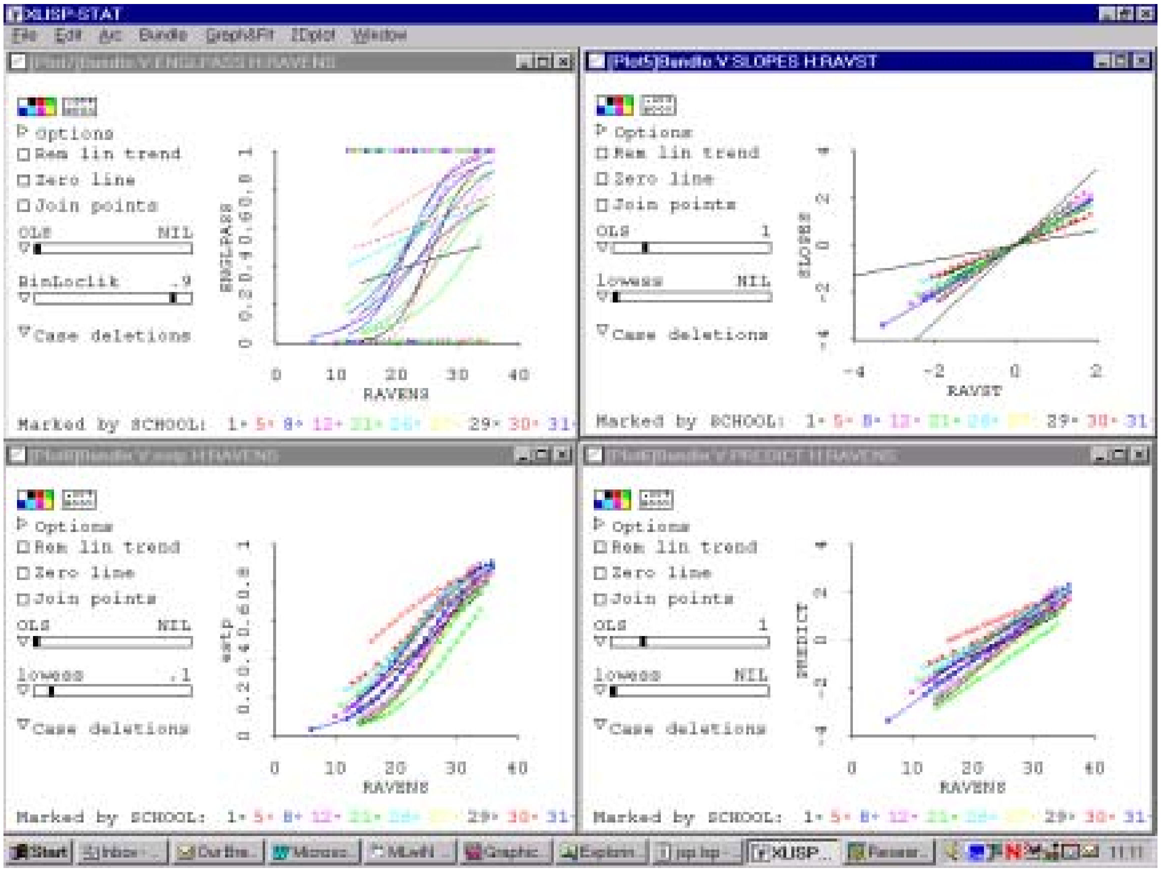


Figure 27: Bundle plot for random slopes

It is possible to investigate several features of the variation simultaneously. Figure 28 is a screen dump of four plot windows that have been opened simultaneously. In the top left plot [plot7], LLL smooths have been added to a plot of the original data (notice how the size of the plot influences the amount of the legend that is visible). In the top right plot [plot5], the bundle plot used in Figure 27 is presented. The bottom left plot [plot8] is the estimated probabilities from the multilevel model against the RAVEN score. This can be compared with the top left plot. The true level of variation in the curves is somewhere between the extremes represented by these two plots. Finally in the bottom right [plot6] there is a plot of predicted values against RAVENS score. Comparing this with the top right plot indicates the advantage of the bundle plot for assessing variation in the slopes.

The plots in Figure 28 are linked. This means that a change in one of them can change all the others. In Figure 29, school 29 has been selected. The slope of the smooth for this school is relative shallow. However, the fitted slope has been heavily shrunk. The plots in this Figure would be improved by adding appropriate confidence intervals to each plot and reference lines indicating zero variation from the constant slope parameter.


XLISP-STAT
File Edit Arc Bundle Graph&Fit 2Dplot Window
[Plot5]Bundle V:SLOPES H:RAVST
Options
Case deletions
ENGLPASS
RAVENS
0 0.2 0.4 0.6 0.8 1
0 10 20 30 40
SLOPES
RAVST
-4 -2 0 2
estp
PREDICT
Marked by SCHOOL: 1 5 8 12 21 26 27 29 30 31
Start Inbox Our fine Microsc MLwiN Graphic Explorin jsp.lsp XLISP... Paramr 11:11

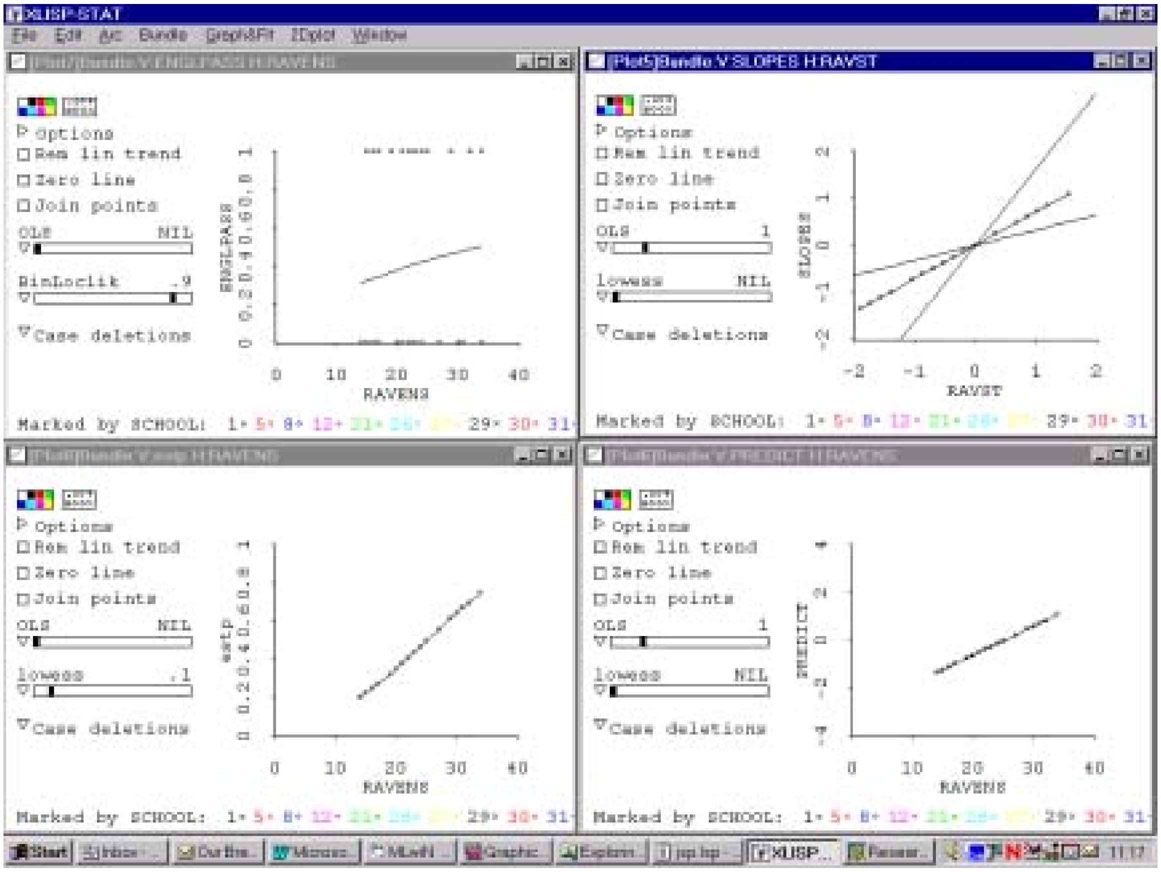
□ Rem lin trend
□ Zero line
□ Join points
OLS NIL
BinLoclik .9

□ Rem lin trend
□ Zero line
□ Join points
OLS 1
lowess NIL

□ Rem lin trend
□ Zero line
□ Join points
OLS NIL
lowess .1

□ Rem lin trend
□ Zero line
□ Join points
OLS 1
lowess NIL

Figure 28: An example of linked plots for the Junior School Project data


XLISP-STAT
File Edit Arc Bundle Graph&Fit 2Dplot Window
Plot7]Bundle V ENGLPASS H RAVENS
[Plot5]Bundle V SLOPES H RAVST
Plot6]Bundle V estp H RAVENS
Plot8]Bundle V PREDICT H RAVENS
▷ Options
□ Rem lin trend
□ Zero line
□ Join points
▽ Case deletions
ENGLPASS
0 0.2 0.4 0.6 0.8 1
RAVENS
0 10 20 30 40
SLOPES
-2 -1 0 1 2
RAVST
estp
PREDICT
-4 -2 0 2 4
Start Inbox Our the Micros MLwiN Graphic Explorin jsp bp XLISP Paesar 11:17

OLS NIL
BinLoclik .9

OLS 1
lowess NIL

OLS NIL
lowess .1

OLS 1
lowess NIL

Marked by SCHOOL: 1 5 8 12 21 26 27 29 30 31

Figure 29: An example of the selecting one school in the linked plots

It has been stated that plots of fitted values to not represent the true level of variation. This can be explained by considering Figures 30 and 31 which are based on the estimates for the fixed slopes model. Figure 31 is a caterpillar plot generated by the program MLwin. This has been modified by adding horizontal lines which indicate ±1 standard deviations and ±2 standard deviations based on the estimate of the school level variance component. If it assumed that the school level effects are normally distributed then 66% of the school effects should lie within the first set of limits and 95% should lie in the second set of limits. Most of the shrunken estimates lie in the first set of limits. However, a large number of confidence intervals enter the region between either +1 and +2 standard deviations or –1 or –2 standard deviations. This means that at least some of the true values lie in that region.

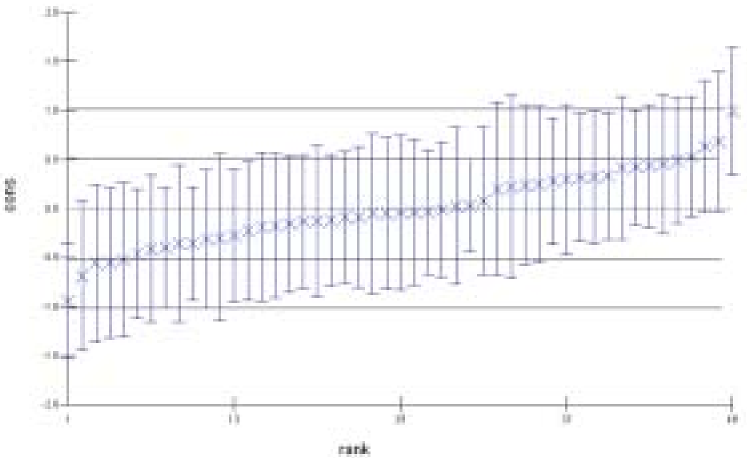


Figure 30: "Caterpillar" plot for model 3

When the shrunken parameter estimates are used in plots, there is not enough variability in the lines. Figure 31 (which was generated with the function plot of SYSTAT) illustrates this. The lines based on the shrunken parameter estimates are the solid blue lines and the 95% and 99% confidence intervals based on the variance components are based on the dashed red lines. Consider the predictions for candidates with a standardised Ravens score of 1. Using the variance component estimate, a 99% confidence interval would be between 0.35 and 0.92. However all the fitted lines lie between 0.50 and 0.87.

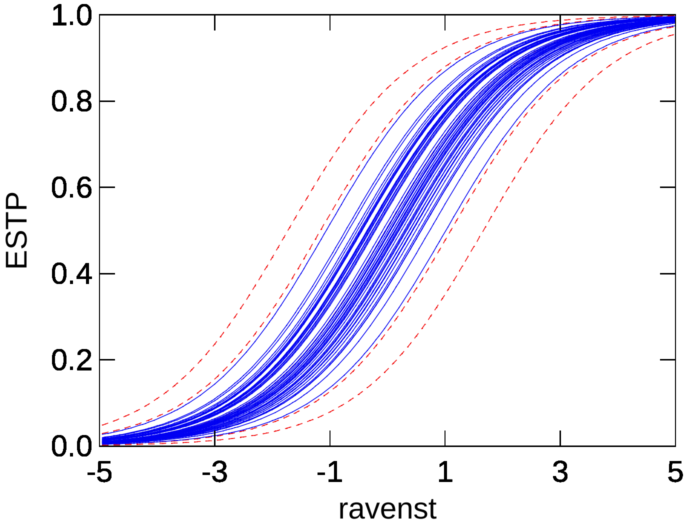


Figure 31: Estimated probabilities for the fixed slopes with 95% and 99% confidence intervals

Although it is easy to see how confidence intervals for fixed slope models can be presented, further research is needed into how the results of a random slopes model should be presented. It should be recognised that this problem depends on the amount of shrinkage.

Discussion and conclusions

This paper has demonstrated the usefulness of *ARC* for exploring multilevel data prior to a more formal analysis. The use of IDA means that it is possible to identify potential models before fitting multilevel models. It also, provides some insights that would not be immediately apparent from the multilevel analysis results.

Although this paper contains a large number of figures, they represent only a tiny proportion of those used in the preparation of this paper. The highly interactive nature of *ARC* made it easy to change and modify the plots and so thoroughly explore the nature of the data.

XLISP-STAT and *ARC* are in a process of continual development and there is ongoing work into multilevel modelling. All the analyses described in this paper used unmodified *ARC* and *ARC* plug-ins. It should also be noted that *Arc* includes a powerful set of tools for regression diagnostics for ordinary non-multilevel regression and there are plans to extend these for multilevel models. Obviously some types of residual analyses are the same for multilevel regression as they are for ordinary regression. These analyses can be carried out in *ARC* if a suitable file is created for the program.

Smoothing techniques are very useful for the preliminary analysis of logistic analysis. It is debatable whether it is better to use the quick LOWESS smoothing or the more appropriate LLL smoothing. The use of smoothing techniques is useful for explaining what is involved in logistic regression. The use of *ARC* as an initial analysis tool for multilevel modelling tool does have an advantage for teaching. *ARC* is already an excellent tool for teaching conventional single level regression techniques. This means that when it is used for this purpose it allows the arguments for fitting multilevel models to be developed easily.

In the example given in this paper, the number of groups was relatively small. Using *ARC* with a greater number of second level units could lead to plots that are unwieldy and difficult to interpret. In this situation, there are two potentially useful approaches. Firstly, a random selection of groups could be analysed. This would give some insight into the relationships of the various variables under consideration. Secondly, the groups in the data could be randomly partitioned into a series of subsamples and plots produced for all the subsamples. When investigating random slopes, data from small groups just serves to confuse plots with smoothed lines based on a few points and it is sensible to consider only the larger groups for the initial data analysis. Of course, all the data should be used in the subsequent analysis. It is possible to develop procedures to do this in XLISP-STAT.

IDA is important. It allows researchers to gain a feel for the data. For example, some researchers (e.g., Kreft, 1996) have questioned whether multilevel analyses are really necessary in certain circumstances. It is clear that their arguments could be usefully illustrated by plotting some of the simulation data.

There are a number of modifications that could be considered for *ARC.* It would be useful if a distinction could be drawn between using a variable to group the data and a variable to mark the data. This would allow the visualisation of categorical variables at the group level (e.g. school type). It would also be useful if rather than selecting one group for a plot it were possible to select a small subset of groups.

Acknowledgement

The author is grateful to Dr Zubaida Haque for making the minority ethnic data available for this paper.

Any views presented in this paper are solely those of the author and do not necessarily represent those of the University of Cambridge Local Examinations Syndicate.

References

Afsharthous, D., and Hilden-Minton, J. (1994) *Terrace-two user's guide: An Xlisp-Stat package for estimating multilevel models.* (http://www.xlispstat.org/code/statistics/regression/terrace/ter-guide.ps).
Aitkin, M.A., Anderson, D., and Hinde, J. (1981) Statistical modeling of data on teaching styles. *Journal of the Royal Statistical Society, Ser. A.*, 144, 419-461.
Aitkin, M.A., and Longford, N.T. (1986) Statistical modelling issues in school effectiveness studies. *Journal of the Royal Statistical Society, Series A*, 149,
Bell, J.F., (1998) Contribution to discussion of Langford, I., and Lewis, T. Outliers in Multilevel data. *Journal of the Royal Statistical Society Series A - Statistics in Society*, 161, 2, 153-160.
Bell, J.F. (2000) *Methods of aggregating GCSE results to predict A-level performance*. Paper presented at the British Educational Research Conference, University of Wales, Cardiff. (http://www.leeds.ac.uk/educol/documents/00001506.htm).
Bowman, A.W., and Azzalini, A. *(*1997) *Applied smoothing techniques for data analysis. The kernel approach with S-plus illustrations*. Oxford: Oxford University Press.
Chatfield, C. (1985) The initial examination of data (with discussion). *Journal of the Royal Statistical Society, Series A.*, 148, 214-53.
Chatfield, C. (1988) *Problem solving: A statistician's guide.* London: Chapman and Hall.
Chatfield, C., and Schimek, M.G. (1987) An example of model-formulation using IDA. *The Statistician,* 36, 357-363.
Cook, R.D., and Weisberg, S. (1999) *Applied regression including computing and graphics.* New York: Wiley.
Haque, Z. (1999) *Exploring the validity and the possible causes of the apparently poor performances of Bangladeshi pupils in British secondary schools.* Ph.D. thesis. University of Cambridge.
Haque, Z., and Bell, J.F. (in press) Evaluating the performances of minority ethnic pupils in secondary schools. *Oxford Review of Education.*
Hedeker, D., and Gibbons, R.D. (1996) MIXOR: A computer program for mixed-effect ordinal regression analysis. *Computer Methods and Programs in Biomedicine*, 49, 157-176.
Hilden-Minton, J. (1994) *Terrace-Two: a new Xlisp-stat package for multilevel modeling with diagnostics.* (Available at http://www.stat.ucla.edu/papers/preprints/146.ps.gz).
Kreft, I.G.G. (1996) *Are multilevel techniques necessary? An overview, including simulation studies*. Los Angeles: California State University. (Available at http://www.ioe.ac.uk/multilevel/kreft.pdf.)
Langford, I., and Lewis, T. (1998). Outliers in multilevel data. *Journal of the Royal Statistical Society, Series A*. 161(2).
Mortimore, P., Sammons, P., Stoll, L., Lewis, D., and Ecob, R. (1988). *School matters, the junior years*. Wells, Open Books.
Preece, D.A. (1987) Good statistical practice. *The Statistician*, 36, 397-408.
Singer, J.D. (1998) Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*. 24(4):323-355.
Snijders, T.A.B., and Bosker, R.J. (1999) *Multilevel analysis. An introduction top basic and advanced multilevel modeling*. London: Sage.
Tierney, L. (1990) *Lisp-Stat: an object-oriented environment for statistical computing and dynamic graphics*. New York: Wiley.