

A review on SYSTAT (10) on random effects modelling

*Min Yang
Centre for Multilevel Modelling
Institute of Education, University of London*

1. A general introduction

1.1 Background

SYSTAT is one of the products from SYSTAT Software Inc. (SSI). It is a statistics and graphics package for technical professionals in the areas of Data Analysis and Modelling. Launched in the mid eighties during last century, it has been used in the fields of Life Sciences, Bio-medical, Environmental Sciences, Automobile, Pharmaceuticals, etc. In recent years, it was acquired from SPSS Science.

SYSTAT 10 was released in Spring 2001. The Mixed Regression tool for random effects analysis of hierarchical data was one of the new features in this release coming out with full documentations and examples. By the end of 2002, SYSTAT 10.2 has been released but with little added to the Mixed Regression facilities.

1.2 Software and hardware requirements

It is recommended that one should install the latest program in Windows 95, 98, 2000 or NT on a machine with 32 MB or more memory, 30 MB hard disk space and a Pentium processor. Installation of SYSTAT is very friendly with clear instructions. The latest version supports Windows XP too.

1.3 Data input/output functionality

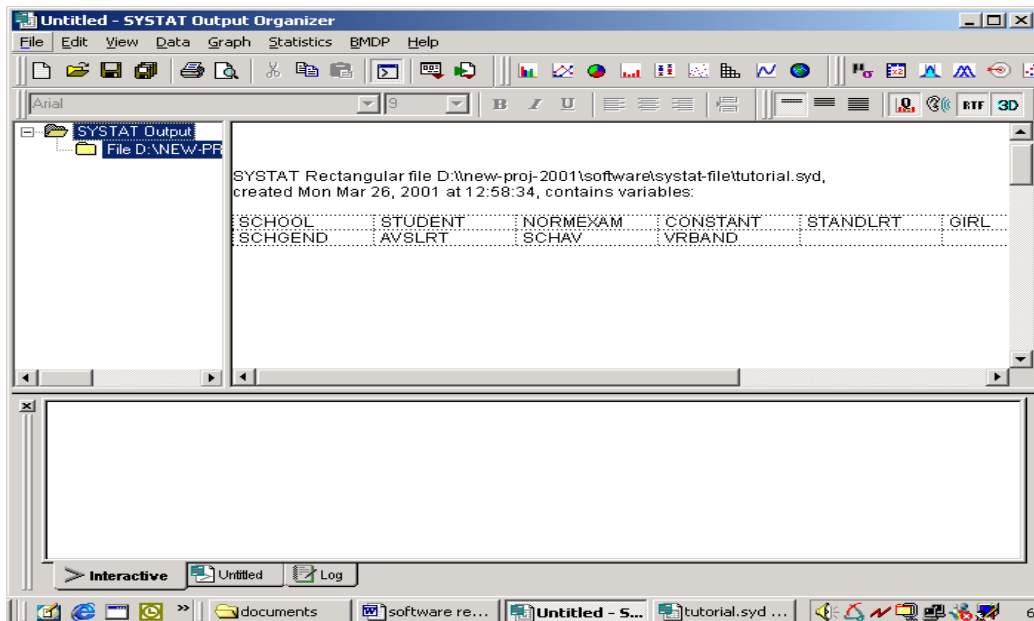
The package has friendly and flexible data input and output facilities. In the drop down list of **File** tool in SYSTAT, there are **New** and **Open** options for input of Data, Command and Output. Small datasets can be keyed in from the data window by selecting the **New** data option, and new command or syntax files can be created by selecting the **New** command option.

The option **Open** brings in existing SYSTAT data files with extension names *.syd* saved from the newer versions and *.sys* saved from old versions of SYSTAT. It also reads data files directly saved from Excel (*.xls*), SPSS (*.sav*), SAS(*.sd2*, *.xpt*, *.tpt*), Dbase (*.dbf*), ASCII (*.txt*) and BMDP(*.por*, *.sav*). The option **Save As** exports a SYSTAT data file in all the formats available for import.

The option **Open** reads saved SYSTAT command files (*.syc*, *.cmd*) and BMDP command files (*.inp*) into the syntax window.

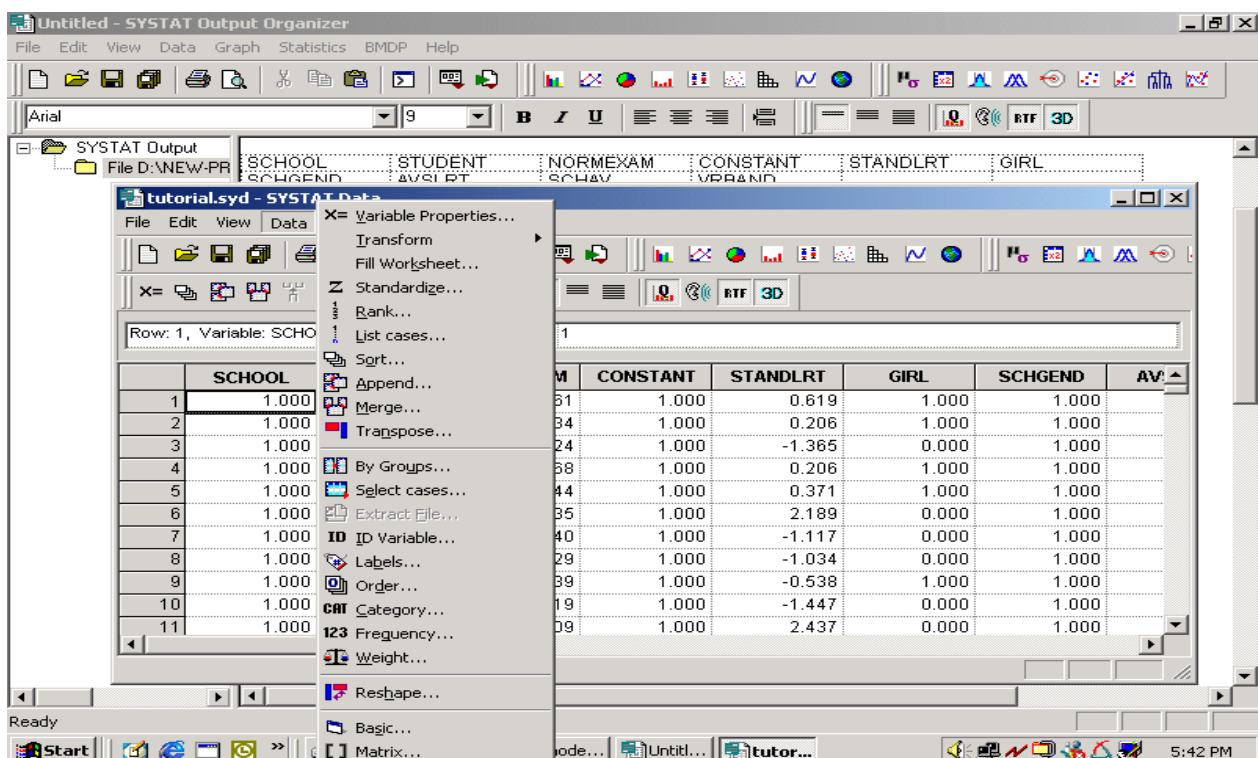
1.4 Interface features

Invoking SYSTAT brings the following **Output Organiser** window on the screen. Results from any operation will be displayed in the upper right panel with the output file folders in the upper left panel. The bottom panel is a place for displaying the syntax corresponding to the window operation history by clicking on the **Log** button, and for using of syntax in model fitting by clicking on the **Interactive** button with the prompt **>** in front of each syntax line.



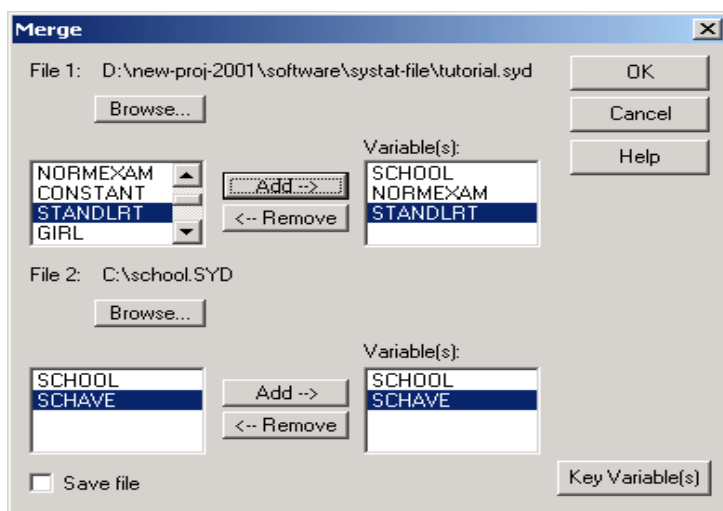
The tool **View Data** shows the data matrix opened. It only works on one data file at a time.

As a general statistical package, SYSTAT has comprehensive and easy-to-use window tools for data manipulation, descriptive analysis, statistical tests or inferences, classical linear and nonlinear modeling, survival analysis as well as good graphical facilities. Options for these are in the dropdown list of the **Data**, **Statistics** and **Graph** menus. The list of options for data manipulation is displayed below.



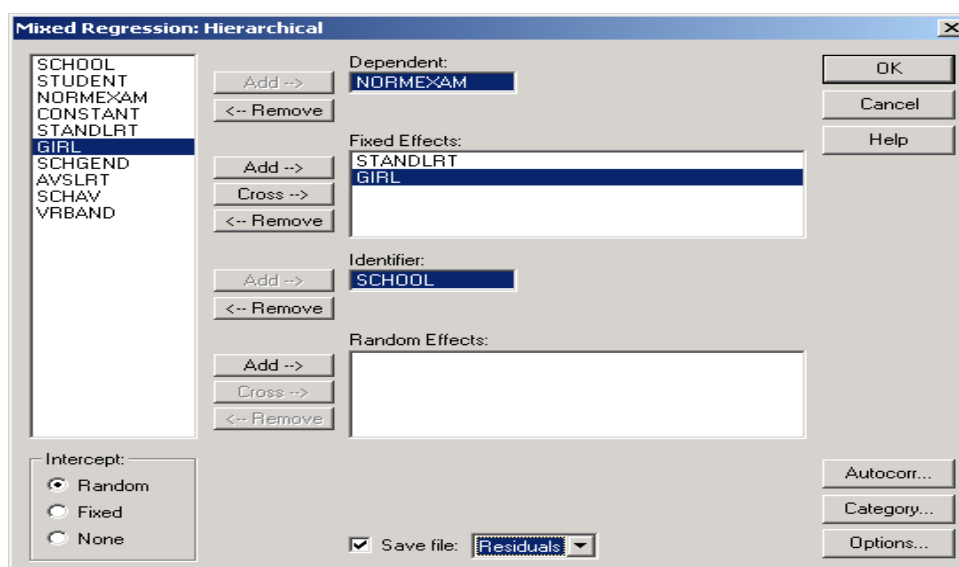
For multilevel data are concerned, several interface features in SYSTAT are useful for data manipulation and analysis.

- Merging a level 2 data file with a level 1 data file

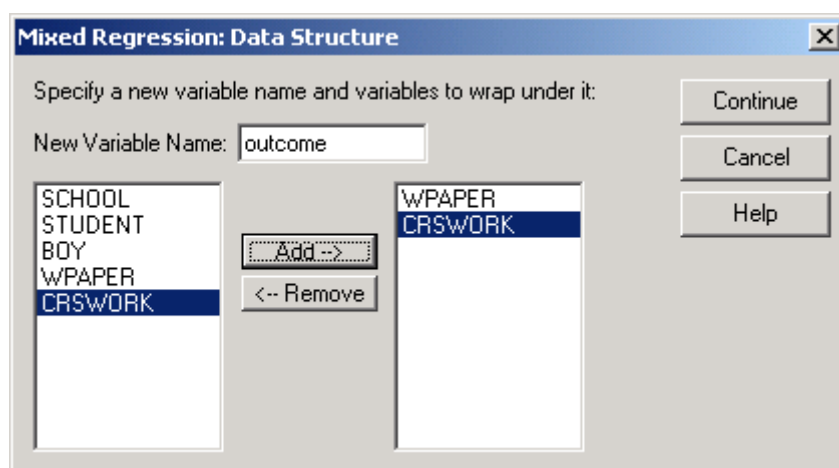


This *Merge* window in the **Data** menu can be used in combining level 2 and level 1 data files. Both files to be merged have to be SYSTAT system data files with the same name for the key variable. In the example shown in this window, File 1 contains student level data, File 2 contains school level data, and the Key variable for the merging is 'school', the school id.

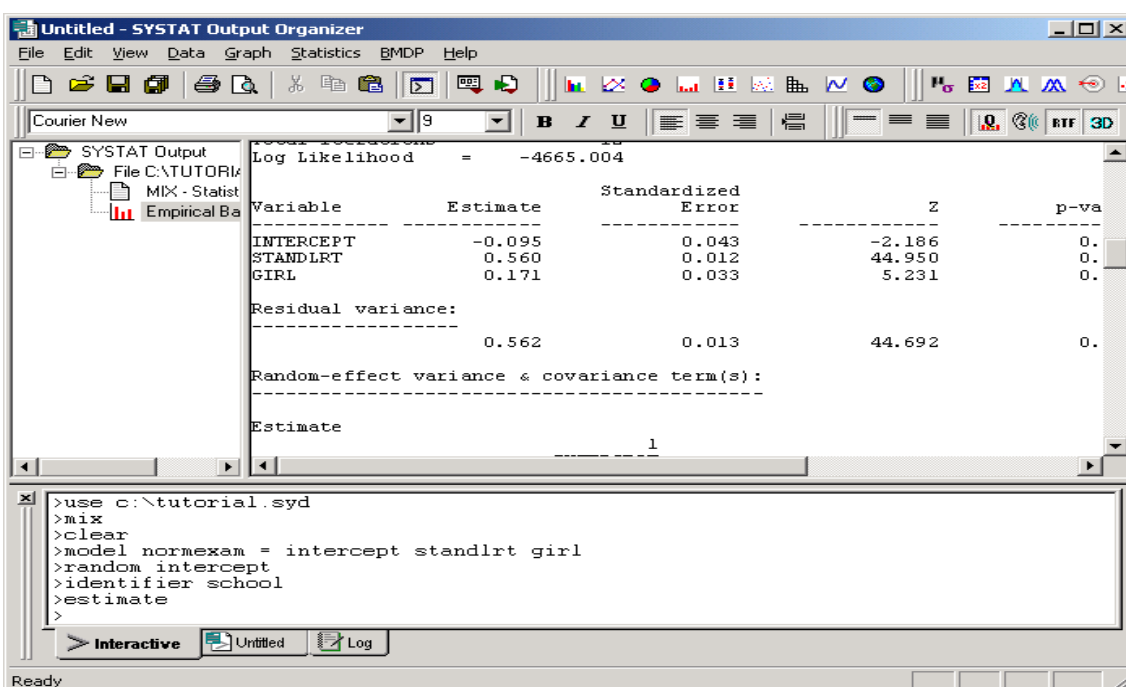
- The *Sort* window can sort on several keys in ascending or descending orders. This allows multilevel data being arranged as level 1 units nested within level 2 units nested within level 3 units.
- The general data handling tools such as transformation, standardization, case selection, and grouping are convenient for data exploration before fitting multilevel models.
- The basic multilevel model specification is via the window of *Mixed Regression* in the **Statistics** menu. This window consists two sub-options: *Hierarchical data* and *Multivariate data*. The *Hierarchical* sub-option allows the specification of single level models, two-level models with either random intercept or random slopes. Dummy variable(s) can be produced and fitted in the model via the *Category* button in the window. Several autocorrelation structures for residuals can be specified via the *Autocorr* button. A tick in the *Save* file box saves a data file containing predictions and estimates of residuals at both levels 1 and 2 by default based on the model specified. Users can then explore the residuals and predictions graphically or numerically.



The *Multivariate* sub-option brings up a new window to stack several responses into a single outcome. The following shows the window for a case of two responses to be stacked. While stacking responses, other variables in the data list will be repeated to match the length of the new response, and a new categorical variable 'TRIAL' will be created to indicate the original responses.



- Using of the syntax interactively in the *Interactive* panel below is the standard way to specify and fit a model. In the example a simple two level model with two covariates in the fixed part of the model and the random effects of intercepts was specified by commands *Use*, *Mix*, *Model*, *Random* and *Identifier*, and fitted by the command *Estimate*. The command *Use* opens the system file. Among the commands, *Use*, *Mix*, *Model* and *Estimate* are hot commands that are executed immediately. Others are cold commands, and executed only after a hot command appears.



- The *Log* panel by default contains both fixed and random parameter estimates, their standard errors, z-score statistic and p-value, Likelihood value, correlation coefficients between parameter estimates, intra-cluster correlation and distribution of level 2 residual estimates. For categorical variables, SYSTAT names them 'FXDn(m)', standing for FIXED term, with n indicating the reference category value and m numbers of categories contrasting to the reference category. Apart from these, numbers of units at both levels are reported and basic descriptive statistics (mean, SD, min and max) for each variable are provided, although they may not be meaningful for categorical variables.

2 Standard modelling tools for multilevel analysis

2.1 A brief check of model list

Since the Mixed Regression tool is new in SYSTAT, the range of multilevel models it can fit is limited. Table 1 lists what it can do and what it cannot.

Table 1 Multilevel models in SYSTAT

Data / model type	Estimation procedures: EM + Fisher scoring	Limitation on levels in data	Covariates	Random slopes	Weighting	Fitting variance function	Other comments/limitations
Normal response	Marginal ML	2	Any	Yes	No	Time series only	
Binary/Binomial							N/A
Poisson							N/A
Negative Binomial							N/A
Repeated measures	Marginal ML	2	Any	YES	No	Time series only	
Nominal multinomial							N/A
Ordered multinomial							N/A
Cross-classified							N/A
Multiple membership							N/A
Survival data							N/A
Time series	Marginal ML	2	Any	Yes	No	1 st order AR NS 1 st order AR 1 st order MA ARMA(1,1) Toeplitz	Level 1 residuals only
Multivariate Normal	Marginal ML	2	Any	Yes			Single residual variance at level 1
Multivariate mixed responses							N/A
Nonlinear							N/A

Although the modeling tool is termed as Mixed Regression after the MIXREG package, the general terminology for random effects models used in the package is similar to what in HLM 5 and MLwiN.

2.2 Tools for statistical inference and model diagnostics

The combination of EM and Fisher scoring algorithm used in SYSTAT produces marginal ML estimates. The test available for overall goodness of fitting is the Chi-squared statistic of deviance based on likelihood values of two nested models. In the model output, a Z score statistic and its p-value from Normal distribution is attached for each parameter estimate (fixed and random coefficients). Users have to make their own judgment whether the test results are adequate or not based on their own data, in particular for random parameter coefficients when sample size at level 2 be small. The *Mixed Regression* window outputs the value of log-likelihood for each model fitted. The likelihood ratio test can then be used to test joint effects of more than one parameter estimate added or removed from a model.

From the *Mixed Hierarchical* model window, residuals at both levels and predicted variables can be saved for further diagnostic analysis readily available in the package. Good graphical tools in windows or syntax for general data analysis in SYSTAT are useful for diagnostics of residuals. Examples of graphs are used throughout the manuals of SYSTAT. The multi dimensional plot of marginal and multivariate Normal distributions for random effects is impressive.

2 Model specifications —basic models

There are only two types of data that can be fitted under the *Mixed Regression* tool: hierarchical and multivariate data. Basic models that can be fitted only under this title are considered in the review.

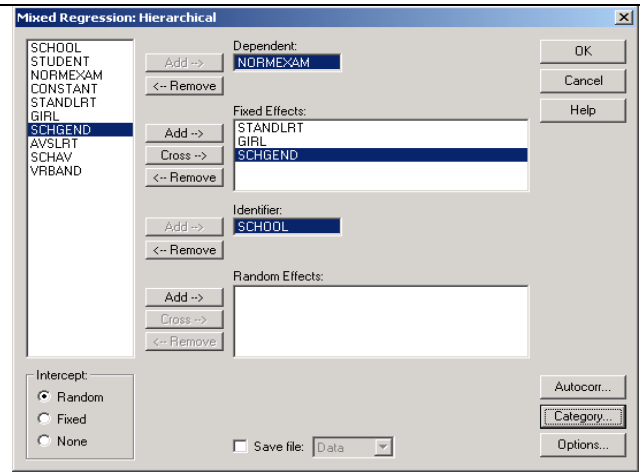
3.1 Two-level Normal models

The data consist of 4,059 students nested with 65 schools with their examination score at age 16 as the outcome variable standardised. A SYSTAT system data file is created for the data. The explanatory variables used are the pre-school reading test score (STANDLRT, x_1), student gender (GIRL, x_2 : 0 for boy and 1 for girl) and the school gender (SCHGEN: 1 for mixed school, 2 for boy school and 3 for girl school). Two dummy variables are required for the school gender effects in models.

Model **A** is a variance components model with fixed effects for all the three covariates. Only the intercepts are allowed to vary randomly among schools. The model specification is in Table 2.

The specification of this model is very easy in the user-friendly window of *Mixed Regression: Hierarchical* below. The commands corresponding to the window specification are simple, easy to understand and straightforward. By default, the command CATEGORY var / DUMMY treats the last category to be the reference, in the same way that PROC MIXED does in SAS. The default setting for the ESTIMATE procedure is used for the model, i.e., REPARAMeterise variance terms automatically if zero estimates occurred, listing one level 2 unit (NREC), number of EM iteration is 10 (NEM) and CONvergence criterion is 0.0001. These are all options that users can change before fitting a model.

Table 2 Model **A** specification using the window tool and commands

Specification in the window	Commands from the <i>Log</i> window
	<pre>USE "D:\tutorial.syd" REM -- Following commands were produced by the MIXHIER dialog: MIX CLEAR CATEGORY SCHGEN / DUMMY MODEL NORMEXAM = STANDLRT GIRL SCHGEN IDENTIFIER SCHOOL RANDOM INTERCEPT ESTIMATE / REPAR=ON NREC=1 NEM=10 CONV=.0001 REM -- End of commands from the MIXHIER dialog</pre>

Model **B** extends Model **A** in the fixed part by including in an interaction term between x_1 and x_2 . This job is easily done in three steps from the tool window in Table 2: adding the variable ‘standlrt’ again in the box of *Fixed Effects*, highlighting the variable ‘girl’ in the variable list, and clicking at the *Cross button* for *Fixed Effects*. The only change in the command panel corresponding to the front-end action is in MODEL command, i.e.

```
MODEL NORMEXAM = STANDLRT GIRL SCHGEN GIRL*STANDLRT
```

Model **C** fits random slopes of the reading score ‘standlrt’ between schools. Two steps from the tool window in Table 2 lead to specify this model from Model **B**: selecting ‘standlrt’ from the variable list and adding it to the box of *Random Effects*, removing this variable from the box of *Fixed Effects*. This action results changes in commands MODEL and RANDOM as below:

```
MODEL NORMEXAM = GIRL SCHGEN GIRL*STANDLRT
RANDOM INTERCEPT STANDLRT
```

The estimates of the three models are displayed in Table 3. The results are identical in comparison to SAS estimates of ML. But the algorithm seems less efficient than that in PROC MIXED of SAS.

Table 3 Two-level Normal model estimates and their S.E. by Mixed Regression (Hierarchical) tool

	Symbol	Model A	Model B	Model C	Model C, SAS
Fixed Effects					
Intercept	$\hat{\beta}_0$	-0.009 (0.076)	-0.009 (0.076)	-0.011 (0.073)	-0.011 (0.073)
Standlrt	$\hat{\beta}_1$	0.560 (0.012)	0.563 (0.018)	0.551 (0.026)	0.551 (0.026)
Girl-Boy	$\hat{\beta}_2$	0.167 (0.034)	0.167 (0.034)	0.168 (0.034)	0.168 (0.034)
Mixed - Girl school	$\hat{\beta}_3$	-0.159 (0.087)	-0.159 (0.087)	-0.178 (0.080)	-0.178 (0.080)
Boy - Girl school	$\hat{\beta}_4$	0.019 (0.123)	0.019 (0.123)	-0.001 (0.114)	-0.001 (0.114)
Standlrt*(Girl-Boy)	$\hat{\beta}_5$		-0.005 (0.025)	0.007 (0.029)	0.007 (0.029)
Random coefficients					
Level 2 variance of intercepts	$\hat{\sigma}_{u_0}^2$	0.081 (0.016)	0.081 (0.018)	0.080 (0.016)	0.084 (0.016)
Level 2 covariance of intercepts and slopes	$\hat{\sigma}_{u_{01}}$			0.029 (0.006)	0.021 (0.007)
Level 2 variance of slopes	$\hat{\sigma}_{u_1}^2$			0.015 (0.004)	0.015 (0.004)
Level 1 residual variance	$\hat{\sigma}_{e_0}^2$	0.562 (0.013)	0.562 (0.013)	0.550 (0.012)	0.550 (0.012)
-2Loglikelihood		9,325.4	9,325.4	9,281.1	9,281.1
Approximate time		< 6 secs	< 6 secs	< 7 secs	< 1 secs
Note: Models were run on a computer of Pentium 1.6 GHz with 520 Mb RAM under Window 2000.					

This example data has heterogeneous level 1 variance by student gender (Rasbash, etc 2000) and being a nonlinear function of the reading score of student fitted by SAS as well as MLwiN (Yang, 2002). However, these models apparently cannot be fitted in SYSTAT.

3.2 Two-level Normal model for repeated measures data

Using the same **Mixed Regression** tool, repeated measures data with either balanced or unbalanced design and adjusting for any number of covariates can be fitted as ordinary two-level models with both random intercepts and random slopes. In addition to that, five types of time series functions for the residual variance are available in SYSTAT. They are *1st order stationary auto-regression AR(1)*, *1st order non-stationary auto-regression AR(1)*, *stationary Moving Average MA(1)*, *stationary ARMA(1,1)* and *general (toeplitz) structure*. These are subset of the time series structures of residuals that Mixed PROC in SAS 8.2 fits.

The data we tried to fit consist of the height (cm) of 26 boys between the ages of 11-13 over 9 occasions approximately 0.25 year apart. The overall growth model is polynomial function of age up to 4th order with both the linear and quadratic terms varying at level 2 among boys. Specifying and fitting this model is straightforward with the syntax as below. The higher order polynomial terms can be formed from the **Interactive** panel.

```
USE [filename]
MIX
CLEAR
MODEL HT = CAGE*CAGE*CAGE CAGE*CAGE*CAGE*CAGE
IDENTIFIER BOYS
RANDOM INTERCEPT CAGE CAGE*CAGE
ESTIMATE
```

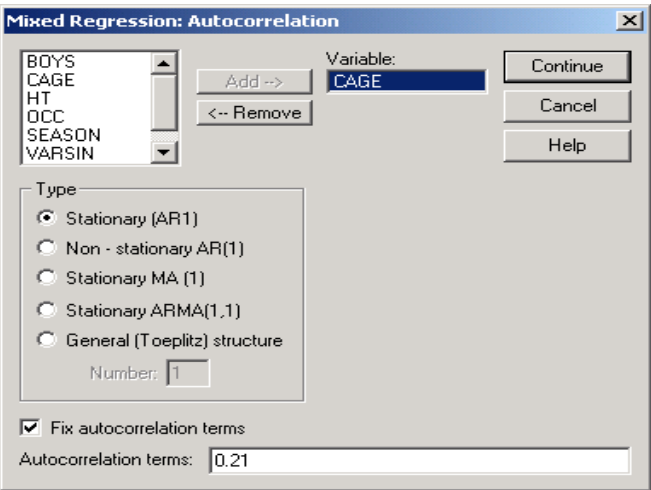
These results of Model 1 are displayed in Table 4. The estimates are identical to the ML estimates by other packages such as SAS, MLwiN and HLM.

Based on the first model fitted, we tried to fit the stationary AR1 model to the level 1 residuals. For the well balanced data with approximately equal spaced age interval between measurements, we expected the stationary AR1 model to estimate an auto-regression coefficient analogical to that reported by Goldstein, Healy and Rasbash (1994) on the same dataset.

It is easy to specify the AR1 model with only one more command line, `AUTO CAGE / type = AR`. This is added after the commands `CLEAR` or `RANDOM` based on the last model fitted, and adds the terms sin and cos in the command `MODEL`. However, the program reported estimation difficulties and stopped the procedure.

Fitting the same AR1 model on the same data, PROC MIXED of SAS and MLwiN estimated the autocorrelation term as 0.21. Fixing the auto correlation term to be 0.21 in the following commands or from the *Autocorrelation* window, we were able to make Model 2 converge only when the random effects of the quadratic term of age at level 2 were removed.

Table 4 Fitting AR1 model to the residual variance

Hierarchical window specification	Commands from the <i>Log</i> window
	<pre>USE [filename] MIX CLEAR AUTO CAGE / TYPE = AR NUMBER = 1 FIX=0.21 MODEL HT = CAGE*CAGE CAGE*CAGE*CAGE CAGE*CAGE*CAGE*CAGE IDENTIFIER BOYS RANDOM INTERCEPT CAGE ESTIMATE</pre>

In the SYSTAT manual (Statistics II) a repeated measures data from a clinical trial was used to illustrate the growth model and the model with autocorrelation of AR1. It was founded that the autocorrelation coefficient only became significant when the quadratic parameter of the time variable was not fitted in the fixed part nor in the random part at level 2. The two examples on the autocorrelation model make the impression that the autocorrelation coefficient for the residual variance structure seems only to absorb the variance and covariance that would have been fitted by the fixed and random effects of some higher order coefficients of the time variable if had they not been fitted in the model!

Table 5 MML estimates (SE in brackets) for 2-level repeated measures data
— Height of 26 boys measured at 9 occasions

	Symbol	Model 1	Model 2
Fixed effects			
Intercept	β_0	148.98 (1.540)	148.04 (1.183)
Age	β_1	6.166 (0.351)	5.983 (0.290)
Age ²	β_2	1.091 (0.349)	1.770 (0.616)
Age ³	β_3	0.468 (0.162)	0.553 (0.231)
Age ⁴	β_4	-0.340 (0.300)	-1.406 (0.630)
Sin	β_5		-0.075 (0.073)

Cos	β_6		-0.250 (0.096)
Random coefficients			
Level 2 var. of intercepts	$\sigma_{u_0}^2$	61.55 (17.09)	36.25 (12.10)
Level 2 var. of Age effects	$\sigma_{u_1}^2$	2.763 (0.782)	3.982 (2.137)
Level 2 var of age^2 effects	$\sigma_{u_2}^2$	0.630 (0.225)	1.712 (0.609)
Level 2 cov (Int. Age)	$\sigma_{u_{01}}$	7.992 (3.023)	
Level 2 cov (Int. Age^2)	$\sigma_{u_{02}}$	1.363 (1.406)	
Level 2 cov (Age. Age^2)	$\sigma_{u_{12}}$	0.875 (0.342)	
Level 1 residual variance	$\sigma_{e_0}^2$	0.217 (0.025)	0.326 (0.041)
Autocorrelation term (fixed)			0.21
-2 Loglikelihood		627.28	619.42
Approximate time to run		<5 secs	<2 secs
Note: Models were run on a computer of Pentium 1.6 GHz with 520 Mb RAM under Window 2000.			

3.3 Multivariate Normal response model

The **Mixed Regression** tool in SYSTAT specifies and fits two-level multivariate Normal models either through the **Hierarchical** or **Multivariate** data window.

Using the **Hierarchical** window, users should stack the responses and organize the data into the structure required before specify and fitting models illustrated in Section 3.1. Considering 3 individuals from two clusters with up to 2 responses, the structure for the bivariate model should be:

Cluster ID	Individual ID	Outcome	Trial	Gender
1	1	y_{11}	1	1
1	1	y_{12}	2	1
1	2	y_{21}	1	2
1	2	•	2	2
2	3	y_{31}	1	1
2	3	y_{32}	2	1

The missing data from the individual 2 on the second response is kept and indicated by dot as system missing.

Using **Multivariate** window, users can simply open the data in the following form.

Cluster ID	Individual ID	Y1	Y2	Gender
1	1	y_{11}	y_{12}	1
1	2	y_{21}	•	2
2	3	y_{31}	y_{32}	1

To convert them into the structure above, two operations are involved:

- (1) Click to choose **Multivariate** data tool to bring up the **Mixed Regression: Data Structure** window;
- (2) Type in a name for the new outcome in stacking, choose responses to be stacked;

Clicking on OK to finish the data reconstruction to show the **Hierarchical** window for model specification and estimating. Alternatively, command CONVERT in syntax does the data reconstruction.

The example used here consists of scores on a science examination tested by a traditional written paper (Y_1) and assessed by course work (Y_2) from 1905 pupils in 73 schools in England. Gender (x) is the covariate coded 0 for girls and 1 for boys. Both the Multivariate and Hierarchical data window are used to fit the same bivariate 2-level model.

The model we are interested to fit is:

$$\begin{aligned}
 y_{ijk} &= z y_{1,jk} + (1-z) y_{2,jk} \\
 y_{1,jk} &= \beta_0 + \beta_1 x_{jk} + v_{1k} + u_{1,jk} \\
 y_{2,jk} &= \alpha_0 + \alpha_1 x_{jk} + v_{2k} + u_{2,jk} \\
 \begin{pmatrix} v_{1k} \\ v_{2k} \end{pmatrix} &\sim MVN(0, \Omega_v) : \Omega_v = \begin{pmatrix} \sigma_{v_1}^2 & \\ \sigma_{v_{12}} & \sigma_{v_2}^2 \end{pmatrix} \\
 \begin{pmatrix} u_{1,jk} \\ u_{2,jk} \end{pmatrix} &\sim MVN(0, \Omega_u) : \Omega_u = \begin{pmatrix} \sigma_{u_1}^2 & \\ \sigma_{u_{12}} & \sigma_{u_2}^2 \end{pmatrix}
 \end{aligned}$$

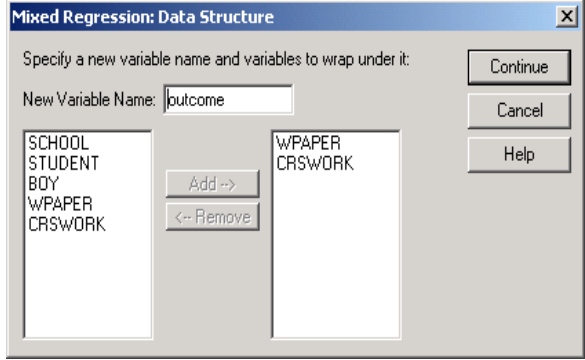
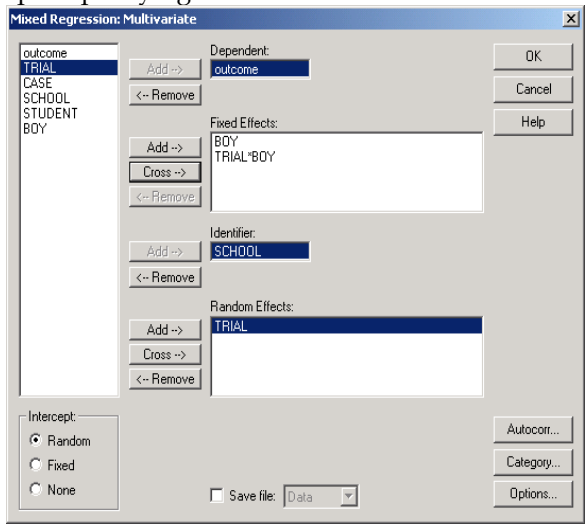
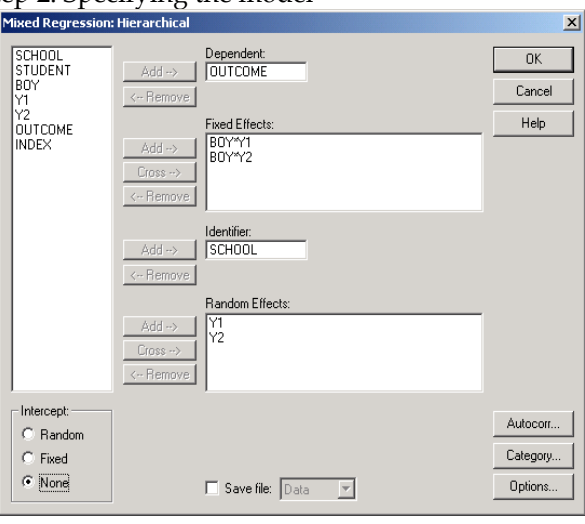
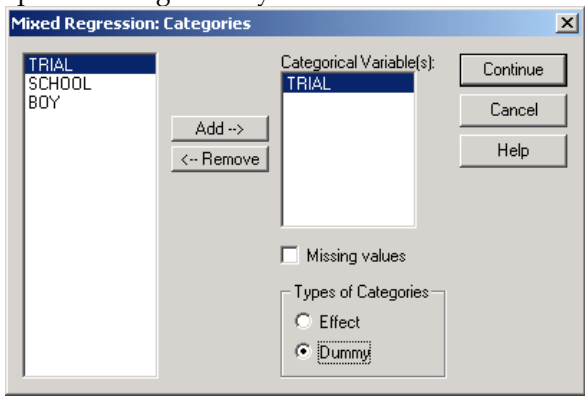
Where $i = 1, 2$ indicates the responses, j the students, and k the schools. An index variable z (1 for written paper, 2 for coursework) is required and is the new variable *TRIAL* from the **Multivariate** window.

However, the current **Mixed Regression** tools cannot fit the bottom level variance-covariance matrix. Only a single term is allowed at the student level. No covariance or correlation between the residuals of the two scores at student level can be estimated directly. So instead of fitting the model interested we fit a slightly different model in which terms $u_{1,jk}$ and $u_{2,jk}$ are replaced by a single term

e_{jk} .

The model specifications and the corresponding commands of the two tools are displayed in Table 6.

Table 6 Steps of specifying and fitting a two-level multivariate model by the *Multivariate data* window and by the *Hierarchical data* window respectively

Multivariate data window	Hierarchical data window
<p>Step 1: Data conversion</p> 	<p>Step 1: Preparing data structure</p> <p>This can be done in other packages such as Excel, SAS or SPSS. SYSTAT reads those files directly.</p>
<p>Step 2: Specifying the model</p> 	<p>Step 2: Specifying the model</p> 
<p>Step 3: Defining dummy variable for TRIAL</p> 	<p>Step 3: Click OK to fit the model</p>
<p>Step 4: Click Continue, then OK to fit the model</p> <pre>Rem MIX CLEAR CONVERT outcome = WPAPER CRSWORK CATEGORY TRIAL / DUMMY MODEL outcome = BOY TRIAL*BOY IDENTIFIER SCHOOL RANDOM INTERCEPT TRIAL ESTIMATE REM -- End of commands from the MIXMULTI dialog</pre>	<pre>Rem MIX CLEAR MODEL OUTCOME = BOY*Y1 BOY*Y2 IDENTIFIER SCHOOL RANDOM Y1 Y2 ESTIMATE REM -- End of commands from the MIXHIER dialog</pre>

In the commands of *Multivariate data* window, the fixed part of the model includes ‘intercept’, ‘trial’, ‘boy’ and ‘trial*’boy’. The variable ‘boy’ is coded 0/1, and treated as dummy variable by default. The ‘trial’ should be defined as dummy variable too. This specification produces fixed and random parameter estimates as below, with -20.922 the difference of mean scores between written paper and coursework for girls, -9.227 the gender difference of mean scores between the two responses. To obtain estimates of the intercept and the gender slope effects for each response with their standard error shown in Table 7 for easy interpretation, one needs to do some algebra based on the model outputs as below. Many users might be put off by this task, in particular doing this for the random parameters.

Variable	Estimate	Standardized Error	Z	p-value
INTERCEPT	69.793	1.177	59.273	0.000
TRIAL	-20.922	1.198	-17.459	0.000
BOY	6.790	0.624	10.876	0.000
FXD1	-9.227	0.883	-10.454	0.000
Residual variance:				
	152.597	3.764	40.537	0.000
Random-effect variance & covariance term(s):				
Estimate				
		1	2	
		INTERCEPT	TRIAL	
1	INTERCEPT	79.315		
2	TRIAL	-49.051	63.892	
Standardized Error				
		1	2	
		INTERCEPT	TRIAL	
1	INTERCEPT	14.993		
2	TRIAL	12.347	13.821	

In contrast, the *Hierarchical data* window reports estimates of the intercept and gender slope for each response as well as their covariance directly, so long the required data structure has been specified before using the window. The results are in the last column of Table 7.

Table 7 Estimates (s.e.) of the two-level multivariate model of Normal distribution

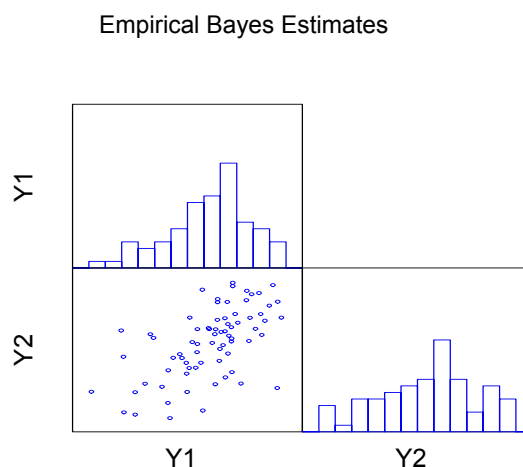
Parameter: variable	From <i>Multivariate</i> window	From <i>Hierarchical</i> window
Intercept: Written paper	48.87	49.27 (0.956)
Intercept: Coursework	69.79 (1.177)	69.68 (1.173)
Slope of written paper: boy-girl	-2.437	-2.474 (0.630)
Slope of coursework: boy-girl	6.790 (0.624)	6.891 (0.627)
School var.: Written paper	45.11	46.03 (9.326)
School var.: Coursework	79.32 (14.99)	78.42 (14.86)
School cov.: (Written, Course)	30.26	30.99 (9.115)
Residual variance	152.60 (3.764)	154.11 (3.801)
-2log-likelihood	27252.3	27300.7
Approximate time to run	<6 secs	<5 secs
Note: Models were run on a computer of Pentium 1.6 GHz with 520 Mb RAM under Window 2000.		

Estimates of the two models listed in Table 7 are somehow different. This could be because that the multivariate estimation procedure imputes missing data assuming their missing at random whilst the hierarchical procedure ignores the missing data points. To help the explanation of the difference in the model estimates from the two windows further, we fitted the same model on the balanced data, i.e., deleting 20% of individuals who missed a record in any of the two outcomes from the data. Both windows gave identical estimates on the same model.

As we see in the model output that only one residual variance could be estimated at level 1. It is a pooled variance of the two outcomes at that level. This is a serious restriction and implies that the use of the multivariate fitting option cannot be recommended.

A more general structure for level 1 residuals in the *Multivariate* data tool allowing the full variance-covariance form should be an obvious way to please many users of SYSTAT in near future.

Compared to the *Multivariate* data tool, the *Hierarchical* window reports results of each response directly with easier interpretation. The graphical distributions of empirical Bayes estimates of random effects of the two responses at school level from the *Hierarchical* window are as the default output for simple diagnostic purpose on the assumptions of level 2 random effects.



4 Documentation and user support

The package is supplied with two volumes of Statistics manual (I and II). The *Mixed Regression* is documented in Chapter 2 of Statistics II, contributed by D. Hedeker, R. Marcantonio and M. Pechnyo. The chapter of 71 pages is written in an easy understand form, with every major button, box and option in the GUI explained clearly and effectively. For each type of model, a working example is presented to illustrate the modelling process with collection of syntax to specify the main model. The default outcomes in full for the main model are presented. Some detailed graphical explorations before and after fitting a model are helpful in understanding what trends or statistics the fitted model were trying to get out of. They are also helpful in the model interpretation. This chapter would be a good tutorial of basic random effects modelling for beginners.

Since five types of autocorrelation structure of residuals are available for longitudinal data, more worked examples in fitting those structures would be welcomed.

SYSTAT used to be one of the SPSS Science products. Last November, it was sold to the new software company under its own name SYSTAT Software Inc. A new UK based Europe office has been set up in Hounslow in addition to other two offices in US and India.

The SYSTAT home website <http://www.systat.com/> has full details to all three contacts.

Technical support is via the website <http://www.systat.com/support/?sec=t004> where an entry page is displayed for users to fill in details of their purchase and problems that need supports.

Currently, a single user copy of SYSTAT 10.2 is priced at \$1,299.00 for commercial users and \$799.00 for academic users. For multiple, site or network licensing, for upgrade and special purchase or teaching, one can send email to sales@systat.com.

Conclusion

For two-level structured data, SYSTAT has a basic window tool for efficiently fitting Normal response data including growth models with time series structures for level 1 residuals, random intercepts and random slopes models, as well as multivariate models, although the latter was problematic. Although the range of models it can handle is rather limited, the window interfaces for data input/output, data manipulation and model specification are user friendly and easy to follow. With a clearly written, well-illustrated document for the basic models, beginners in the multilevel modeling field may find it a good tutorial package. Excellent graphical tools in the package for general statistical analysis come handy for diagnostic analysis of random effects at level 2 and residuals at level 1.

References

Goldstein H., Healy, M.J.R. and Rasbash, J. 1994: Multilevel time series models with applications to repeated measures data. *Statistics in Medicine* **13**, 1643-55.

Hedeker, D., Marcantonio, R. and Pechnyo, M. *Chapter 2: Mixed Regression*, in SYSTAT 10, Statistics II (2000).

Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I. etc, 2000: *A user's guide to MLwiN*, Multilevel Models Project, Institute of Education, University of London.