

Three-level analysis of simulated data

Contents

1. The data.....	1
2. The model.....	2
3. Fitting the model.....	3
4. Discussion of results.....	6

Unlike real data sets, simulated data sets have the advantage that the true population parameters are known. Consequently, it is possible to evaluate how closely a particular model approaches these values.

1. The data

A linear growth curve model with two dummy-coded covariates (Lang1 and Lang2) is fitted to a simulated dataset **surveyhlm.isf** in the **Multilevel Examples** folder. It is assumed that the level-3 units are institutions. Within each of 100 institutions, 10 patients are selected on the basis of their initial achievement in a test of short term memory (Score1) and measurements were repeated over six time intervals for five patients from each institution and over 4 time intervals for the remaining 5. In the table below, (Weight3) shows the level-3 weight calculations based on standardized initial scores.

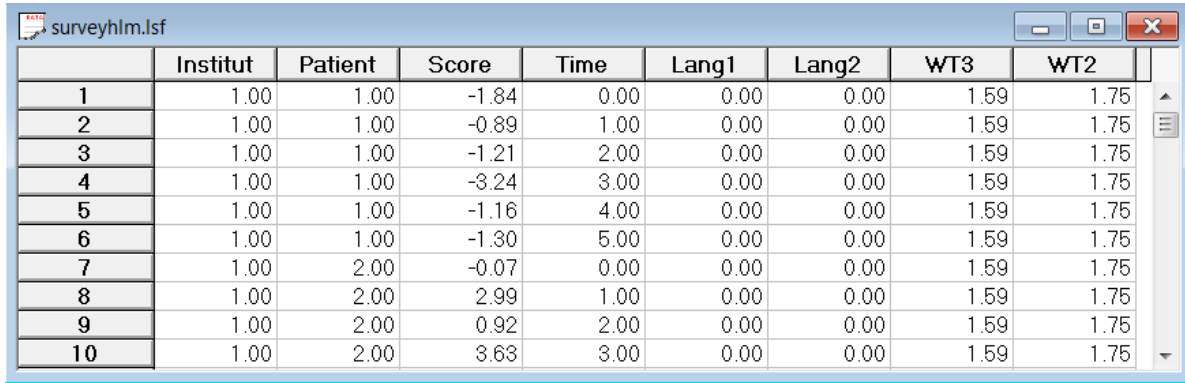
Interval	Lower	Upper	% Expected	% Selected	Weight3
1	-Inf	-1.00	15.87	10.00	1.587
2	-1.00	-0.70	8.33	10.00	0.833
3	-0.70	-0.20	17.88	10.00	1.788
4	-0.20	0.00	7.93	10.00	0.793
5	0.00	0.30	11.79	10.00	1.179
6	0.30	1.00	22.34	10.00	2.234
7	1.00	1.30	6.19	10.00	0.619
8	1.30	1.80	6.09	10.00	0.609
9	1.80	2.30	2.52	10.00	0.252
10	2.30	Inf	1.07	10.00	0.107

Ten patients were selected from each institution as follows:

- Four from ethnic group 1 with Weight2 = 7.0/4.0
- Three from ethnic group 2 with Weight2 = 2.0/3.0

- Three from ethnic group 3 with Weight2 = 1.0/3.0

The first 10 records of the dataset in **surveyhlm.isf** are shown below.



	Institut	Patient	Score	Time	Lang1	Lang2	WT3	WT2
1	1.00	1.00	-1.84	0.00	0.00	0.00	1.59	1.75
2	1.00	1.00	-0.89	1.00	0.00	0.00	1.59	1.75
3	1.00	1.00	-1.21	2.00	0.00	0.00	1.59	1.75
4	1.00	1.00	-3.24	3.00	0.00	0.00	1.59	1.75
5	1.00	1.00	-1.16	4.00	0.00	0.00	1.59	1.75
6	1.00	1.00	-1.30	5.00	0.00	0.00	1.59	1.75
7	1.00	2.00	-0.07	0.00	0.00	0.00	1.59	1.75
8	1.00	2.00	2.99	1.00	0.00	0.00	1.59	1.75
9	1.00	2.00	0.92	2.00	0.00	0.00	1.59	1.75
10	1.00	2.00	3.63	3.00	0.00	0.00	1.59	1.75

Note that the data were simulated in such a way that odd-numbered patients have six score measurements at time points 0, 1, 2, 3, 4, 5. The even-numbered patients have only four score measurements.

2. The model

The three-level model used here is similar to that described in the previous section. Within that hierarchical framework, 500 data sets, **surveyhlm.isf** being the first, were simulated according to the following hypothetical model

$$Score_{ijk} = \beta_0 + \beta_1 * Time + \gamma_1 * Lang1 + \gamma_2 * Lang2 + v_{i0} + Time * v_{i1} + u_{ij0} + Time * u_{ij1} + e_{ijk}$$

where i denotes institution i , ($i = 1, 2, \dots, 100$), ij patient j ($j = 1, 2, \dots, 10$) in institution i and ijk the k -th measurement ($k = 1, 2, \dots, 6$) on patient j in institution i . The outcome variable Score denotes a patient's measurement on some test of interest, Time the time of measurement, and Lang1 and Lang2 are indicator variables indicating a patient's first or home language as being English or another language.

In this model, β_0 denotes the average expected score, while β_1 indicates the estimated coefficients associated with the time of measurement as represented by the fixed effect Time. The fixed part of the model also includes the predictor variables Lang1 and Lang2. The random part of the model is represented by v_{i0} , u_{ij0} and e_{ijk} , which denote the variation in score over institutions, between patients (or, in other words, over patients nested within institutions) and between measurements at the lowest level of the hierarchy.

The data were simulated under the assumption that

$$\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} 1.0 \\ 0.5 \end{pmatrix}, \quad \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} = \begin{pmatrix} 0.5 \\ -1.0 \end{pmatrix}$$

$$\phi_2 = \text{Cov}(u_{ij0}, u_{ij1}) = \begin{pmatrix} 1 & \\ 0.3 & 0.2 \end{pmatrix}$$

$$\phi_3 = \text{Cov}(v_{i0}, v_{i1}) = \begin{pmatrix} 1 & \\ 0.3 & 0.2 \end{pmatrix}$$

and

$$\sigma^2 = \text{Var}(e_{ijk}) = 1.0.$$

3. Fitting the model

The first step is to open the LSF shown above, which is accomplished as follows: use the **File, Open** option to activate the display of an **Open** dialog box. Set the **Files of type** drop-down list box to **LISREL System Data (*.lsf)** and browse for the file **surveyhlm.lsf**. Select the file and click the **Open** button to return to the main LISREL window, where the contents of the LSF are displayed.

The next step is to describe the model to be fitted using the multilevel module in LISREL. To fit a growth model to the data, we proceed as follows. From the main menu bar, select **Multilevel, Linear Model, Title and Options**. Type in a title, and change the number of iterations to 20 and the convergence criterion to 0.0001 as shown below. Click the **Next** button to activate the **Identification Variables** dialog box.

Title and Options

Title (Maximum 70 characters):
Level-3 model (simulated data) with design weights

Maximum Number of Iterations: 20

Convergence Criterion: 0.0001

Missing Data Value: -999999 Nfree: 0

Missing Dep Value: -999999 Deviance:

Use OLS for starting values Calculate effect sizes

Additional Output

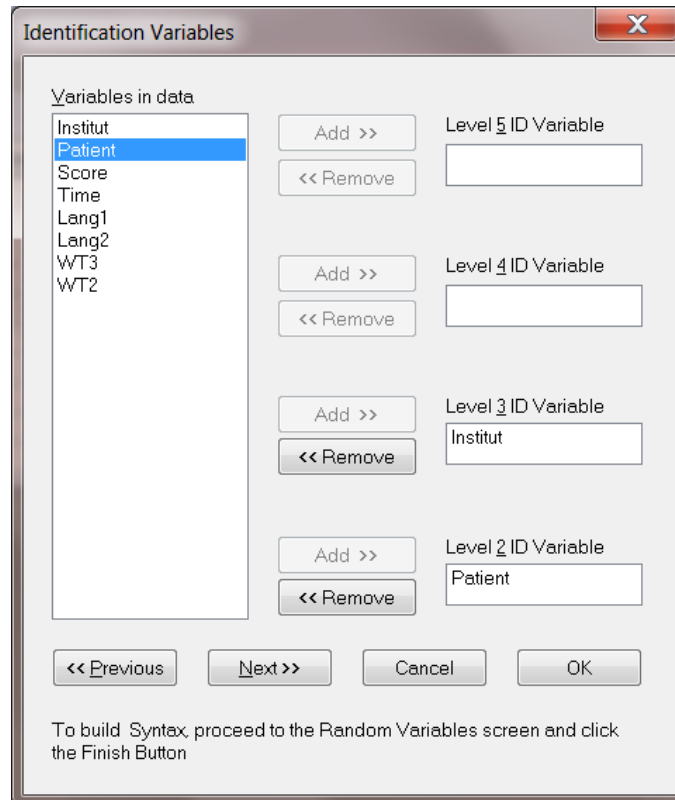
Asymptotic Covariances Residuals

Empirical Bayes Estimates No Data Summary

Between and Within Covariance Matrices

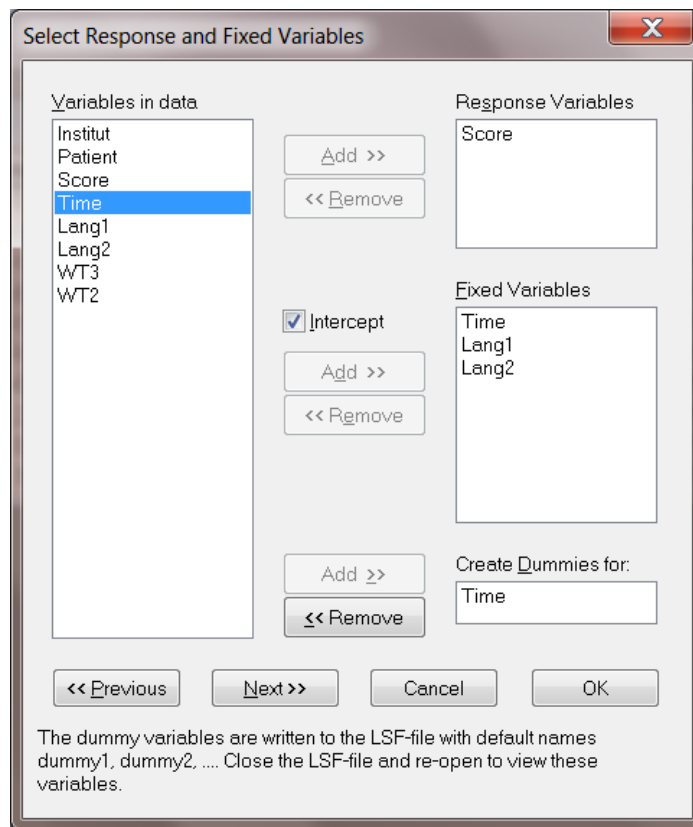
Next >> Cancel OK

To build Syntax, proceed to the Random Variables screen and click the Finish Button

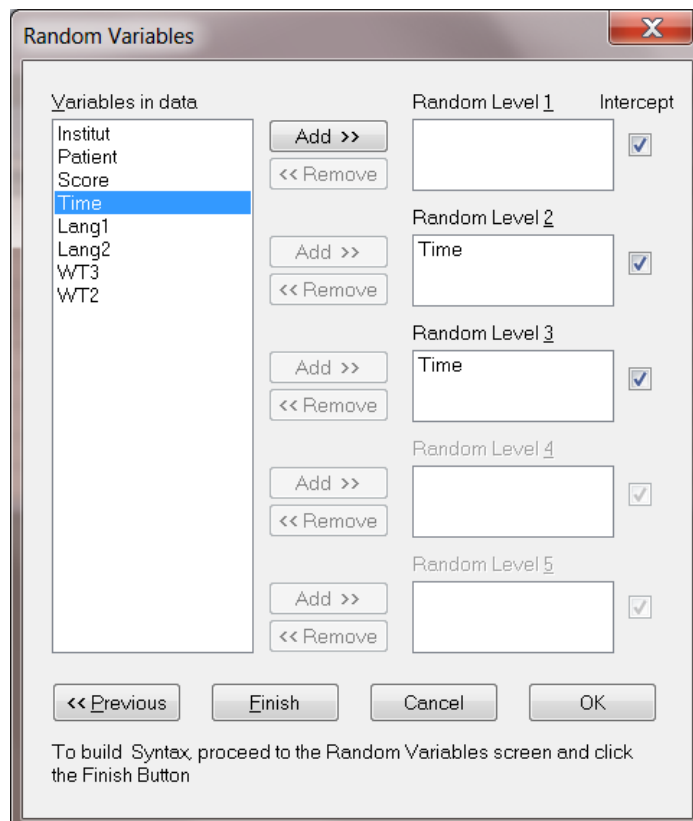


Add the level-3 and level-2 identification variables (Institut and Patient). Use the **Weight Variables** dialog box to select the variables WT3 and WT2 as the level-3 and level-2 weights respectively. To continue to the **Response and Fixed variables** dialog box click **Next**.

Select Score as the dependent (response) variable and Time, Lang1 and Lang2 as the fixed variables (predictors). Note that an intercept term is automatically included unless the **Intercept** check box is unchecked. For illustrative purposes, Time was added to the **Create Dummies for:** text box.



Click **Next** to go to the **Random Variables** dialog box and add Time as level-2 and level-3 random components (the variances are denoted by $Var(u_{ij1}) = \phi_{(2),22}$ and $Var(v_{i1}) = \phi_{(3),22}$ respectively. Note that by default, intercept terms are included at the different levels of the hierarchy. The level-1, level-2 and level-3 variance components for the intercept are denoted by σ_e^2 , $\phi_{(2),11}$ and $\phi_{(3),11}$ respectively.



When done, click the **Finish** button to obtain the PRELIS syntax file **surveyhlm.prl**. Click the **Run PRELIS** icon button to invoke the multilevel module.

```

surveyhlm.PRL
|
|  OPTIONS OLS=YES CONVERGE=0.000100 MAXITER=20 OUTPUT=STANDARD ;
|  TITLE=Level-3 model (simulated data) with design weights;
|  SY='C:\LISREL9 Examples\MLEVELEX\surveyhlm.1sf';
|  ID3=Institut;
|  ID2=Patient;
|  WEIGHT3=WT3;
|  WEIGHT2=WT2;
|  RESPONSE=Score;
|  FIXED=intcept Time Lang1 Lang2;
|  DUMMY=Time;
|  RANDOM1=intcept;
|  RANDOM2=intcept Time;
|  RANDOM3=intcept Time;

```

4. Discussion of results

The output for the **fixed part** of the model is given first, as shown below.

```

surveyhlm.OUT
Level-3 model (simulated data) with design weights

ITERATION NUMBER      4

      +-----+
      | FIXED PART OF MODEL |
      +-----+

-----+-----+-----+-----+-----+
| COEFFICIENTS      | BETA-HAT | STD.ERR. | Z-VALUE | PR > |Z| |
|-----+-----+-----+-----+-----+
| intcept           | 0.92865  | 0.12985  | 7.15145 | 0.00000 |
| Time              | 0.51411  | 0.05250  | 9.79254 | 0.00000 |
| Lang1             | 0.39340  | 0.11381  | 3.45676 | 0.00055 |
| Lang2             | -1.03430 | 0.14107  | -7.33172| 0.00000 |
|-----+-----+-----+-----+-----+

      +-----+
      | -2 LOG-LIKELIHOOD |
      +-----+

DEVIANCE= -2*LOG(LIKELIHOOD) =    23121.9339178070
NUMBER OF FREE PARAMETERS =                11

CHI-SQUARE SCALE FACTOR =      0.45957

```

Recall that the "true" values of the intcept, Time, Lang1 and Lang2 parameters were 1.0, 0.5, 0.5, and -1.0 respectively. To obtain 95% confidence intervals for these estimates, we calculate

$$\text{Estimate} \pm 1.96(\text{std. error})$$

and find that the confidence intervals for the estimated intcept, Time, Lang1 and Lang2 parameters are (0.7009; 1.1565), (0.4220; 0.6062), (0.1937; 0.5931) and (-1.2818; -0.7868) respectively. In all four cases, the

confidence intervals include the "true" values of the corresponding parameter. Note that a χ^2 scale factor of 0.45957 is reported. This value is used to obtain a corrected χ^2 -statistic for testing one model against another model, as will be shown in the next example.

The output for the **random part** of the model is given next. Note that the parameter estimates reported in the output are generally close to the population values which were used to simulate the data. The "true" values for both the level-3 and level-2 variance-covariance components are 1.0, 0.3, and 0.2 respectively. The standard error estimates shown have been corrected as described in the theoretical section (see Section 2.6 of the *Multilevel Modeling Guide*).

RANDOM PART OF MODEL				
LEVEL 3	TAU-HAT	STD. ERR.	Z-VALUE	PR > Z
intcept /intcept	0.93174	0.20698	4.50164	0.00001
Time /intcept	0.25568	0.07490	3.41361	0.00064
Time /Time	0.17510	0.04203	4.16637	0.00003
LEVEL 2	TAU-HAT	STD. ERR.	Z-VALUE	PR > Z
intcept /intcept	0.96301	0.23237	4.14424	0.00003
Time /intcept	0.36079	0.06511	5.54100	0.00000
Time /Time	0.20039	0.04226	4.74169	0.00000
LEVEL 1	TAU-HAT	STD. ERR.	Z-VALUE	PR > Z
intcept /intcept	1.02326	0.04431	23.09147	0.00000

For the level-3 variance components, 95% confidence intervals can be obtained as shown previously. The confidence intervals corresponding to intcept/intcept, Time/intcept, and Time/Time are (0.6002; 1.2633), (0.1367; 0.3747) and (0.1083; 0.2419) respectively. Again, the "true" values fall within these intervals. This conclusion also holds for confidence intervals for the level-1 and level-2 variance-covariance components, which are calculated in the same way.

Note that the spreadsheet presentation of **surveyhlm.lsf** will only show the variables Institut, Patient, ..., WT2, although dummy variables corresponding to the six measurement occasions were written to the actual LSF file. To see these dummy variables, close the LSF file (**without** saving it) and then use **File, Open** to display the modified LSF file.

surveyhlm.lsf

	Lang2	WT3	WT2	dummy1	dummy2	dummy3	dummy4	dummy5	dummy6
1	0.00	1.59	1.75	1.00	0.00	0.00	0.00	0.00	0.00
2	0.00	1.59	1.75	0.00	1.00	0.00	0.00	0.00	0.00
3	0.00	1.59	1.75	0.00	0.00	1.00	0.00	0.00	0.00
4	0.00	1.59	1.75	0.00	0.00	0.00	1.00	0.00	0.00
5	0.00	1.59	1.75	0.00	0.00	0.00	0.00	1.00	0.00
6	0.00	1.59	1.75	0.00	0.00	0.00	0.00	0.00	1.00
7	0.00	1.59	1.75	1.00	0.00	0.00	0.00	0.00	0.00
8	0.00	1.59	1.75	0.00	1.00	0.00	0.00	0.00	0.00
9	0.00	1.59	1.75	0.00	0.00	1.00	0.00	0.00	0.00
10	0.00	1.59	1.75	0.00	0.00	0.00	1.00	0.00	0.00