MIXED EFFECTS MODELING USING SPSS®

Instructor: Manolo ROMERO ESCOBAR

http://mixedreviews.wordpress.com

12 January 2011

At the Research and Outcome Measurement Branch. Ministry of Children and Youth Services
"Two facts - that children's growth is typically the object of inquiry, and that such growth occurs in organizational settings - correspond to two of the most troublesome and persistent methodological problems in the social sciences: the measurement of change and the assessment of multi-level effects“ (Bryk & Raudenbush, 1989, p. 160).
AGENDA

- Review of general linear model
- Mixed Effects Model for Clustered/Grouped Data
  - Lab 1
- Lunch Break
- Longitudinal Outcome Analysis with Mixed Effects Modeling
  - Lab 2 and 3
“Nested” data

- Common in behavioural and social sciences
- Students-classes-schools-boards-etc…
- Repeated observations
- Violation of assumptions (e.g. students in same classroom might share characteristics, non-independent, non-random)
Different names: Same idea

- Multilevel linear models (cf. Goldstein, 1995)
- Multilevel analysis (cf. Hox, 1995)
- Mixed effects models
- Slopes as Outcomes
General Linear Model (GLM)

- A continuous outcome (dependent) variable
- Do not confuse with “generalized linear model” in which DV is not continuous (e.g. Logistic regression)
- Predictors can be continuous (multiple regression) or categorical (ANOVA) or a combination of both (ANCOVA).
- Estimation via Ordinary Least Squares
GLM formulation

\[ y_i = \beta_0 + \beta_1 x_{1i} + r_i, \]

where \( r_i \sim N(0, \sigma^2) \)
A GLM example

- Simulated data
- DV: Self Esteem
- Factors: Psycho-therapy treatment and Gender
- Covariates: Intelligence, Resilience, and Socio-economic Status.
Observed Data

Predicted Data
Unconditional Means Model

\[ y_{ij} = \beta_{0j} + r_{ij}, \]

where \( r_{ij} \sim N(0, \sigma^2) \)

\[ \beta_{0j} = \gamma_{00} + u_{0j}, \]

where \( u_{0j} \sim N(0, \tau_{00}) \)

Fixed Effect \hspace{1cm} Random Effects

\[ Y_{ij} = \gamma_{00} + u_{0j} + r_{ij} \]
Level 2 (Group) Predictors

\[ y_{ij} = \beta_{0j} + r_{ij} \]

\[ \beta_{0j} = \gamma_{00} + \gamma_{01}z_1 + \gamma_{02}z_2 + u_{0j} \]

\[ Y_{ij} = [\gamma_{00} + \gamma_{01}z_1 + \gamma_{02}z_2] + [u_{0j} + r_{ij}] \]

\[ r_{ij} \sim N(0, \sigma^2) \quad \text{and} \quad u_{0j} \sim N(0, \tau_{00}) \]
Level 1 (individual) variables

\[ y_{ij} = \beta_{0j} + \beta_{1j} x_1 + r_{ij} \]

\[ \beta_{0j} = \gamma_{00} + u_{0j}; \quad \beta_{1j} = \gamma_{10} + u_{1j} \]

\[ y_{ij} = [\gamma_{00} + \gamma_{10} x_1] + [u_{0j} + u_{1j} x_1 + r_{ij}] \]
Level 1 and Level 2

\[ y_{ij} = \beta_{0j} + \beta_{1j} x_{1ij} + r_{ij} \]

\[ \beta_{0j} = \gamma_{00} + \gamma_{01} z_{1j} + u_{0j} \]

\[ \beta_{1j} = \gamma_{10} + \gamma_{11} z_{1j} + u_{1j} \]

\[ y_{ij} = \begin{bmatrix} \gamma_{00} + \gamma_{01} z_{1j} + \gamma_{10} x_{1ij} + \gamma_{11} x_{1ij} z_{1j} \\
+ u_{0j} + u_{1j} x_{1ij} + r_{ij} \end{bmatrix} \]

where \( r_{ij} \sim N(0, \sigma^2) \)

and \( \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N \begin{bmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix} \end{bmatrix} \)
# Grouped data predictors

<table>
<thead>
<tr>
<th>Level</th>
<th>Type/Field</th>
<th>Fixed Effect Interaction</th>
<th>Random Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level 1 (Within groups)</strong></td>
<td>Continuous/Covariate</td>
<td>With Level 2 predictor by default (can be taken off if n.s.)</td>
<td>Yes, but test if it is necessary</td>
</tr>
<tr>
<td><strong>Level 1 (Within groups)</strong></td>
<td>Categorical/Factors</td>
<td>With Level 1 predictor if it is of research interest.</td>
<td></td>
</tr>
<tr>
<td><strong>Level 2 (Between groups)</strong></td>
<td>Continuous/Covariate</td>
<td>With Level 1 predictor by default (can be taken off if n.s.)</td>
<td>No.</td>
</tr>
<tr>
<td><strong>Level 2 (Between groups)</strong></td>
<td>Categorical/Factors</td>
<td>With Level 2 predictor if it is of research interest.</td>
<td></td>
</tr>
<tr>
<td><strong>Grouping Variable</strong></td>
<td>Continuous/Subjects on first dialog window</td>
<td>No. Unless it is also a level 1 predictor</td>
<td>As grouping variable</td>
</tr>
</tbody>
</table>
A clustered data example

- High school and beyond
- Example of HLM par excellence
- Students within schools
- Schools by sector: Catholic-Public
- DV: Math achievement
- Predictors: SES, MEANSES, school size, minority, sex, others.
- From Bryk & Raudenbush via nlme package
Click Continue for models with uncorrelated terms.

Specify Subject variable for models with correlated random effects.

Specify both Repeated and Subject variables for models with correlated residuals within the random effects.

Subjects:
- School

Repeated Covariance Type: Diagonal
Unconditional means model

![Linear Mixed Models dialog box with MathAch selected as the dependent variable]
Adding level 2 predictor
Linear Mixed Models: Random Effects

Random Effect 1 of 1

Covariance Type: Variance Components

Random Effects

- Build terms
- Build nested terms
- Include intercept

Factors and Covariates:

- MEANSES

Model:

Factorial

Build Term:

Subject Groupings

Subjects:
- School

Combinations:
- School

Continue | Cancel | Help
Adding level 1 predictor
Level 1 and 2 predictors

[Image: A screenshot of SPSS(R) MIXED MODELS software interface showing Level 1 and 2 predictors for a linear mixed model. The selected variables include: Dependent Variable: MathAch, Factor(s): Sector, Covariate(s): CSES, MEANSES.]
Lab 1: Clustered Data

- File: BrykRauden_HighSchool&Beyond.sav inside data folder
- Open on SPSS
- Fit Unconditional means model:
- Find the parameters and estimate the intra-class correlation given by this formula:

\[ \hat{\rho} = \frac{\hat{\tau}_{00}}{\hat{\tau}_{00} + \hat{\sigma}^2} \]
Lab 1: continuation

- Add a level 2 and a level 1 predictor one at a time
- Why is Sector not a grouping variable?
- What would happen if there were more than two levels of Sector?
- Try other predictors, one or two at a time and note the results
Part II

LONGITUDINAL OUTCOME ANALYSIS
Time variable as level 1 predictor

- Time variable is continuous
- Adds a random effect

\[ y_{ij} = \beta_{0j} + \beta_{1j} \text{time} + r_{ij} \]

\[ \beta_{0j} = \gamma_{00} + u_{0j}; \quad \beta_{1j} = \gamma_{10} + u_{1j} \]

\[ y_{ij} = [\gamma_{00} + \gamma_{10} \text{time}] + [u_{0j} + u_{1j} \text{time} + r_{ij}] \]
Adding time varying predictor

- **Level 1 predictor**

\[ y_{ij} = \left[ \gamma_{00} + \gamma_{10} \text{time} + \gamma_{20} x_1 \right] + \left[ u_{0j} + u_{1j} \text{time} + u_{2j} x_1 + r_{ij} \right] \]

- **Complex random effects covariance structure:**

where \( r_{ij} \sim N(0, \sigma^2) \)

\[
\begin{pmatrix}
  u_{0j} \\
  u_{1j} \\
  u_{2j}
\end{pmatrix}
\sim N\left(\begin{pmatrix}
  0 \\
  0 \\
  0
\end{pmatrix},\begin{pmatrix}
  \tau_{00} & \tau_{01} & \tau_{02} \\
  \tau_{10} & \tau_{11} & \tau_{12} \\
  \tau_{20} & \tau_{21} & \tau_{22}
\end{pmatrix}\right)
\]
Adding a time constant predictor

- Level 2 predictor

\[ y_{ij} = \left[ \gamma_{00} + \gamma_{01} z_{1j} + \gamma_{10} \text{time} + \gamma_{11} \text{time} \times z_{1j} \right] \\
+ \left[ u_{0j} + u_{1j} \text{time} + r_{ij} \right] \]
<table>
<thead>
<tr>
<th>Time relationship (level)</th>
<th>Type/Field</th>
<th>Fixed Effect Interaction</th>
<th>Random Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time variant (Level 1 - within subjects)</td>
<td>Continuous/ Covariate</td>
<td>With Level 2 predictor by default (can be taken off if n.s.)</td>
<td>Yes, but test if it is necessary</td>
</tr>
<tr>
<td>Time variant (Level 1 - within subjects)</td>
<td>Categorical/ Factors</td>
<td>With other Level 1 predictors if it is of research interest. It is unlikely an interaction with time would be of interest</td>
<td></td>
</tr>
<tr>
<td>Time constant (Level 2 – between subjects)</td>
<td>Continuous/ Covariate</td>
<td>With time predictor by default. With Level 2 predictor if it is of research interest.</td>
<td>No.</td>
</tr>
<tr>
<td>Time constant (Level 2 - Between subjects)</td>
<td>Categorical/ Factors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grouping Variable</td>
<td>Continuous/Subjects on first dialog window</td>
<td>No. Unless it is also a level 1 predictor</td>
<td>As grouping variable</td>
</tr>
<tr>
<td>Time variable</td>
<td>Continuous/Covariate</td>
<td>With level 2 predictors</td>
<td>Yes, but test if it is necessary</td>
</tr>
</tbody>
</table>
Covariance Structures

- AR(1)

\[
\begin{pmatrix}
1 & \rho & \rho^2 & \rho^3 \\
\rho & 1 & \rho & \rho^2 \\
\rho^2 & \rho & 1 & \rho \\
\rho^3 & \rho^2 & \rho & 1
\end{pmatrix}
\]

- Compound symmetry

\[
\begin{pmatrix}
\sigma^2 + \sigma_1^2 & \sigma_1 & \sigma_1 & \sigma_1 \\
\sigma_1 & \sigma^2 + \sigma_1^2 & \sigma_1 & \sigma_1 \\
\sigma_1 & \sigma_1 & \sigma^2 + \sigma_1^2 & \sigma_1 \\
\sigma_1 & \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1^2
\end{pmatrix}
\]
Wide vs Long format

- Open the wide.sav file inside the data folder
  - There are columns for each time point for each time varying variable, including the DV (DV is wm)
- Open the long.sav file inside the data folder
  - There is one column per predictor and a “wave” variable to indicate time. Notice that Centered Age could also be taken as a time variable
Re-structuring datasets in SPSS
Welcome to the Restructure Data Wizard!

This wizard helps you to restructure your data from multiple variables (columns) in a single case to groups of related cases (rows) or vice versa, or you can choose to transpose your data.

The wizard replaces the current data set with the restructured data. Note that data restructuring cannot be undone.

What do you want to do?

- **Restructure selected variables into cases**
  Use this when each case in your current data has some variables that you would like to rearrange into groups of related cases in the new data set.

- **Restructure selected cases into variables**
  Use this when you have groups of related cases that you want to rearrange so that data from each group are represented as a single case in the new data set.

- **Transpose all data**
  All cases will become variables and selected variables will become cases in the new data set. (Choosing this option will end the wizard, and the Transpose dialog will appear.)
Variables to Cases: Number of Variable Groups

You have chosen to restructure selected variables into groups of related cases in the new file.

A group of related variables, called a variable group, represents measurements on one variable.

For example, the variable may be width. If it is recorded in three separate measurements, each one representing a different point in time—w1, w2, and w3, then the data are arranged in a group of variables.

If there is more than one variable in the file often it is also recorded in a variable group, for example height, recorded in h1, h2, and h3.

How many variable groups do you want to restructure?

- One (for example, w1, w2, and w3)

- More than one (for example, w1, w2, w3 and h1, h2, h3, etc.)

How Many? 4
Variables to Cases: Select Variables

For each variable group you have in the current data the restructured file will have one target variable. In this step, choose how to identify case groups in the restructured data, and choose which variables belong with each target variable.

Optionally, you can also choose variables to copy to the new file as Fixed Variables.

Variables in the Current File:

- Participant Number [id]
- Time constant predictor [c...]
- Age in months Time 1 [age...]
- Age in months Time 2 [age...]
- Age in months Time 3 [age...]
- Time variant pred t1 [cov1.1]
- Time variant pred t2 [cov1.2]
- Time variant pred t3 [cov1.3]
- WM t1 [wm2.1]
- WM t2 [wm2.2]
- WM t3 [wm2.3]
- Age in years t1 [aly.1]
- Age in years t2 [aly.2]
- Age in years t3 [aly.3]

Case Group Identification

- Use selected variable
- Variable: Participant Number [id]

Variables to be Transposed

- Target Variable: Age

- Age in months Time 1 [age.1]
- Age in months Time 2 [age.2]
- Age in months Time 3 [age.3]

Fixed Variable(s):

- 

[Options for Back, Next, Finish, Cancel, Help]
For each variable group you have in the current data, the restructured file will have one target variable.
In this step, choose how to identify case groups in the restructured data, and choose which variables belong with each target variable.

 Optionally, you can also choose variables to copy to the new file as Fixed Variables.

### Variables in the Current File:
- Participant Number [id]
- Time constant predictor [c.]
- Age in months Time 1 [ag.]
- Age in months Time 2 [ag.]
- Age in months Time 3 [ag.]
- Time variant pred1 [cov1.1]
- Time variant pred2 [cov1.2]
- Time variant pred3 [cov1.3]
  - WM11 [wm2.1]
  - WM12 [wm2.2]
  - WM13 [wm2.3]
- Age in years Time1 [a.y.1]
- Age in years Time2 [a.y.2]
- Age in years Time3 [a.y.3]

### Case Group Identification
- Use selected variable
  - Variable: Participant Number [id]

### Variables to be Transposed
- Target Variable: trans2
  - Age
  - trans2
  - trans3
  - trans4

### Fixed Variables:
Variables to Cases: Select Variables

For each variable group you have in the current data the restructured file will have one target variable.

In this step, choose how to identify case groups in the restructured data, and choose which variables belong with each target variable.

Optionally, you can also choose variables to copy to the new file as Fixed Variables.

Variables in the Current File:

- Participant Number [id]
- Time constant predictor [const]
- Age in months Time 1 [age1]
- Age in months Time 2 [age2]
- Age in months Time 3 [age3]
- Time variant pred t1 [cov1.1]
- Time variant pred t2 [cov1.2]
- Time variant pred t3 [cov1.3]
- WM t1 [wm2.1]
- WM t2 [wm2.2]
- WM t3 [wm2.3]
- Age in years t1 [a1y.1]
- Age in years t2 [a1y.2]
- Age in years t3 [a1y.3]

Case Group Identification:
- Use selected variable
- Variable: Participant Number [id]

Variables to be Transposed:

- Target Variable: a1y
  - Age in years t1 [a1y.1]
  - Age in years t2 [a1y.2]
  - Age in years t3 [a1y.3]

Fixed Variable(s):
- Time constant predictor [const]
Variables to Cases: Create Index Variables

In the current data, values for a variable group appear in a single case in multiple variables. For example, a single case contains the values for w1, w2, and w3.

In the new data, values for a variable group will appear in multiple cases in a single variable. For example, there will be three cases, one each for w1, w2, and w3.

An index is a new variable that identifies the group of new cases that was created from the original case. For example, an index named "w" would have the values 1, 2, and 3.

How many index variables do you want to create?

- **One**
  Use this when a variable group records the effects of a single factor, treatment or condition.

- **More than one**
  How many? 2
  Use this when a variable group records the effects of more than one factor, treatment or condition.

- **None**
  Use this if index information is stored in one of the sets of variables to be transposed.
Variables to Cases: Create One Index Variable

You have chosen to create one index variable. The variable’s values can be sequential numbers or the names of variables in a group.

In the table you can specify the name and label for the index variable.

What kind of index values?
- Sequential numbers
  - Index Values: 1, 2, 3

- Variable names
  - Index Values: age.1, age.2, age.3

Edit the Index Variable Name and Label:

<table>
<thead>
<tr>
<th>Name</th>
<th>Label</th>
<th>Levels</th>
<th>Index Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>wave</td>
<td></td>
<td>3</td>
<td>1, 2, 3</td>
</tr>
</tbody>
</table>
Variables to Cases: Options

In this step you can set options that will be applied to the restructured data file.

- Handling of Variables not Selected
  - Drop variable(s) from the new data file
  - **Keep and treat as fixed variable(s)**

- System Missing or Blank Values in all Transposed Variables
  - Create a case in the new file
  - Discard the data

- Case Count Variable
  - Count the number of new cases created by the case in the current data
    - Name: 
    - Label: 

Options:
- < Back
- Next >
- Finish
- Cancel
- Help
Lab 2: Longitudinal data 3 time points

- Simulated data: Development of Working Memory from 10 to 12 years of age
- Somewhat accelerated longitudinal design
- Variables in the long.sav file ID, a time constant variable (const), Age in months, a time varying covariate (cov), WM.
- Also: Age in years (aiy), Age in years rounded (aiy_r), Age in years centered (aiy_c) and Age in months centered (aim_c)
Click Continue for models with uncorrelated terms.

Specify Subject variable for models with correlated random effects.

Specify both Repeated and Subject variables for models with correlated residuals within the random effects.

Subjects:
- Participant Number [id]

Repeated Covariance Type: Diagonal

Options:
- Continue
- Reset
- Cancel
- Help
The image shows the 'Linear Mixed Models: Statistics' window in SPSS. The selected statistics include:

- Parameter estimates
- Tests for covariance parameters
- Covariances of random effects

The confidence interval is set to 95%.
Lab continuation

- Try different time variables and different covariance structures
- Also add the time constant predictor and the time varying predictor accordingly
Lab 3: Unbalanced data

- Simulated data inspired by consulting case
- Longitudinal study on depression and treatment techniques
- Up to five timepoints: baseline, one week, one month, three months, and six months follow ups
- Not all participants with observations at all points, at least two timepoints
Variables in dep.sav

- **id**: Subject ID; **time**: in months 0, .25, 1,3,6, centered at baseline
- **cond**: treatment condition
- **Changes over time:**
  - LDep3: Depression scale from 0 to 5
  - selfest: Self-esteem scale from 1 to 7
- **Time constant**: Age, anxiety, avoidance, extraversion openness, Frequency (number of observations for the subject)
Click Continue for models with uncorrelated terms.

Specify Subject variable for models with correlated random effects.

Specify both Repeated and Subject variables for models with correlated residuals within the random effects.

Repeated Covariance Type: Diagonal

Subjects:
- Part Name [id]

Repeated:
- 

Continue  Reset  Cancel  Help
If random effect of time not needed...
Adding treatment condition
Adding time constant predictor

![SPSS Mixed Models interface with variables selected: Part Name [id], Self-Esteem [selfest], Age in years [age], Anxiety, Avoidance, Extraversion [extra], No. of assessments [Freq], Dependent Variable: Depression [LDep3], Factor(s): Treatment Condition [cond], Covariate(s): Time in months [time], Openness [open].]
Adding time varying predictor

[Image of SPSS(R) MIXED MODELS interface showing dependent and covariate variables, including Self-Esteem [selfest] as a covariate]
SPSS(R) MIXED MODELS

Random Effect 1 of 1
Covariance Type: Unstructured

Random Effects
Build terms
Build nested terms
Include intercept

Factors and Covariates:
- cond
- time
- open
- selfest

Model:
- time
- selfest

Subject Groupings
Subjects:
- Part Name [id]

Combinations:
- Part Name [id]

Continue Cancel Help
Test for different covariance structures
Lab 3: Fit different models

- Try adding different predictors
- Change DV from depression to self-esteem
- Center the time variable at different timepoints:
  - At 1 month assuming that was the length of the treatment
  - At 6 months to test for differences among conditions in maintaining results
- What if time is not a significant predictor of the DV?
THANK YOU

Manolo ROMERO ESCOBAR: emromesco@gmail.com
Example

In this example, the original data contain separate variables for two measures taken at three separate times for each case. This is the correct data structure for most procedures that compare related observations—but there is one important exception: Linear Mixed Models (available in the Advanced Statistics add-on module) requires a data structure in which related observations are recorded as separate cases.
Figure 4-21
Related observations recorded as separate variables

*varstocases2.sps.
GET FILE = 'c:\examples\data\varstocases.sav'.
VARSTOCASES /MAKE V1 FROM V1_Time1 V1_Time2 V1_Time3
/MAKE V2 FROM V2_Time1 V2_Time2 V2_Time3
/INDEX = Time
/KEEP = ID Age.
Figure 4-22
Related variables restructured into cases