Almost 31 Flavors of Multi-level Modeling in SAS

Presentation to the New York Area SAS Users’ Group

Brian F. Patterson
Assistant Research Scientist

December 9, 2008

www.collegeboard.com
Overview

1. Introduction (10 min.)
2. Review of Statistics (10 min.)
3. SAS for Multi-level Models (20 min.)
4. Research Using MMs (5 min.)
5. Question / Answer Period (15 min.)

Multi-level, NYASUG, Dec. 2008
1.1 Introduction

- Why multi-level models?
  - Avoid aggregation bias and inflated std. err.
  - Disentangle group- and individual-level effects
  - Strengthen generalizability to group-level pop.
  - Model changes over time
1.2 Definition of Terms

- **SAT**
  - College admissions test consisting of Critical Reading, Math & Writing sections on 200-800 scale

- **Advanced Placement**
  - High school program created to give students the opportunity to take college-level courses

1.3 Multi-level Modeling Opportunities

• Model first-year grade point average (FYGPA) at 110 colleges and universities

• Model prob. of participation in Advanced Placement Program® at many high schools

• Model FYGPA with high-school- and college-variability with cross-classified data
1.4 SAS Requirements for Multi-level Modeling

- Many observations
  - Large sample size within each group
  - Sufficient number of groups
- Computing power
  - Procedures (and optional statements within some procedures) are memory-intensive
- Data structure
1.5 Nested Data Structure in SAS

<table>
<thead>
<tr>
<th>Student_ID</th>
<th>Group</th>
<th>Group_Size</th>
<th>Group_Mean</th>
<th>Group_Variance</th>
<th>Y</th>
<th>X1</th>
<th>X2</th>
<th>X3</th>
<th>X4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0001</td>
<td>0001</td>
<td>11</td>
<td>0.450</td>
<td>0.200</td>
<td>2.530</td>
<td>4.330</td>
<td>0.800</td>
<td>0.510</td>
<td>0.710</td>
</tr>
<tr>
<td>0002</td>
<td>0001</td>
<td>11</td>
<td>0.450</td>
<td>0.200</td>
<td>2.717</td>
<td>4.330</td>
<td>0.800</td>
<td>0.800</td>
<td>0.400</td>
</tr>
<tr>
<td>0003</td>
<td>0001</td>
<td>11</td>
<td>0.450</td>
<td>0.200</td>
<td>2.706</td>
<td>4.330</td>
<td>0.600</td>
<td>0.800</td>
<td>0.800</td>
</tr>
<tr>
<td>0004</td>
<td>0001</td>
<td>11</td>
<td>0.450</td>
<td>0.200</td>
<td>2.821</td>
<td>3.650</td>
<td>0.800</td>
<td>0.800</td>
<td>0.800</td>
</tr>
<tr>
<td>0005</td>
<td>0001</td>
<td>11</td>
<td>0.450</td>
<td>0.200</td>
<td>1.886</td>
<td>0.000</td>
<td>0.800</td>
<td>0.800</td>
<td>0.800</td>
</tr>
<tr>
<td>0006</td>
<td>0001</td>
<td>11</td>
<td>0.450</td>
<td>0.200</td>
<td>2.732</td>
<td>4.330</td>
<td>0.800</td>
<td>0.800</td>
<td>0.800</td>
</tr>
<tr>
<td>0007</td>
<td>0001</td>
<td>11</td>
<td>0.450</td>
<td>0.200</td>
<td>2.713</td>
<td>4.330</td>
<td>0.800</td>
<td>0.800</td>
<td>0.800</td>
</tr>
<tr>
<td>0008</td>
<td>0001</td>
<td>11</td>
<td>0.450</td>
<td>0.200</td>
<td>2.713</td>
<td>4.330</td>
<td>0.800</td>
<td>0.800</td>
<td>0.210</td>
</tr>
<tr>
<td>0009</td>
<td>0001</td>
<td>11</td>
<td>0.450</td>
<td>0.200</td>
<td>2.818</td>
<td>4.330</td>
<td>0.800</td>
<td>0.800</td>
<td>0.390</td>
</tr>
<tr>
<td>0010</td>
<td>0001</td>
<td>11</td>
<td>0.450</td>
<td>0.200</td>
<td>3.289</td>
<td>4.330</td>
<td>0.800</td>
<td>0.440</td>
<td>0.800</td>
</tr>
<tr>
<td>0011</td>
<td>0001</td>
<td>11</td>
<td>0.450</td>
<td>0.200</td>
<td>2.026</td>
<td>1.310</td>
<td>0.340</td>
<td>0.800</td>
<td>0.800</td>
</tr>
<tr>
<td>0012</td>
<td>0002</td>
<td>8</td>
<td>0.700</td>
<td>0.200</td>
<td>3.309</td>
<td>4.330</td>
<td>0.800</td>
<td>0.800</td>
<td>0.800</td>
</tr>
<tr>
<td>0013</td>
<td>0002</td>
<td>8</td>
<td>0.700</td>
<td>0.200</td>
<td>3.250</td>
<td>4.330</td>
<td>0.800</td>
<td>0.800</td>
<td>0.800</td>
</tr>
<tr>
<td>0014</td>
<td>0002</td>
<td>8</td>
<td>0.700</td>
<td>0.200</td>
<td>3.275</td>
<td>4.330</td>
<td>0.800</td>
<td>0.800</td>
<td>0.800</td>
</tr>
<tr>
<td>0015</td>
<td>0002</td>
<td>8</td>
<td>0.700</td>
<td>0.200</td>
<td>2.977</td>
<td>4.330</td>
<td>0.800</td>
<td>0.800</td>
<td>0.800</td>
</tr>
<tr>
<td>0016</td>
<td>0002</td>
<td>8</td>
<td>0.700</td>
<td>0.200</td>
<td>2.572</td>
<td>4.330</td>
<td>0.800</td>
<td>0.800</td>
<td>0.800</td>
</tr>
<tr>
<td>0017</td>
<td>0002</td>
<td>8</td>
<td>0.700</td>
<td>0.200</td>
<td>3.261</td>
<td>4.330</td>
<td>0.800</td>
<td>0.800</td>
<td>0.800</td>
</tr>
<tr>
<td>0018</td>
<td>0002</td>
<td>8</td>
<td>0.700</td>
<td>0.200</td>
<td>3.247</td>
<td>4.330</td>
<td>0.800</td>
<td>0.800</td>
<td>0.800</td>
</tr>
</tbody>
</table>
1.6 Repeated Measures Data in SAS

![SAS Interface Image]

<table>
<thead>
<tr>
<th>Student_ID</th>
<th>Group_ID</th>
<th>Group_Size</th>
<th>Group_Mean</th>
<th>Group_Variance</th>
<th>Test_ID</th>
<th>Test_Date</th>
<th>Test_Mean</th>
<th>Y</th>
<th>X1</th>
<th>X3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0001</td>
<td>0001</td>
<td>75</td>
<td>0.70000000</td>
<td>0.50000000</td>
<td>0001</td>
<td>2008-09-01</td>
<td>500.0</td>
<td>0.00</td>
<td>1.19</td>
<td>389</td>
</tr>
<tr>
<td>0002</td>
<td>0001</td>
<td>75</td>
<td>0.70000000</td>
<td>0.50000000</td>
<td>0002</td>
<td>2008-11-01</td>
<td>500.0</td>
<td>106.0</td>
<td>1.19</td>
<td>343</td>
</tr>
<tr>
<td>0003</td>
<td>0001</td>
<td>75</td>
<td>0.70000000</td>
<td>0.50000000</td>
<td>0003</td>
<td>2008-11-15</td>
<td>500.0</td>
<td>362.3</td>
<td>1.19</td>
<td>644</td>
</tr>
<tr>
<td>0004</td>
<td>0001</td>
<td>75</td>
<td>0.70000000</td>
<td>0.50000000</td>
<td>0004</td>
<td>2008-09-01</td>
<td>500.0</td>
<td>0.00</td>
<td>1.19</td>
<td>389</td>
</tr>
<tr>
<td>0005</td>
<td>0001</td>
<td>75</td>
<td>0.70000000</td>
<td>0.50000000</td>
<td>0005</td>
<td>2008-11-01</td>
<td>500.0</td>
<td>263.9</td>
<td>2.68</td>
<td>504</td>
</tr>
<tr>
<td>0006</td>
<td>0001</td>
<td>75</td>
<td>0.70000000</td>
<td>0.50000000</td>
<td>0006</td>
<td>2008-11-15</td>
<td>500.0</td>
<td>318.4</td>
<td>2.68</td>
<td>665</td>
</tr>
<tr>
<td>0007</td>
<td>0001</td>
<td>75</td>
<td>0.70000000</td>
<td>0.50000000</td>
<td>0007</td>
<td>2008-09-01</td>
<td>500.0</td>
<td>147.6</td>
<td>1.15</td>
<td>339</td>
</tr>
<tr>
<td>0008</td>
<td>0001</td>
<td>75</td>
<td>0.70000000</td>
<td>0.50000000</td>
<td>0008</td>
<td>2008-11-01</td>
<td>500.0</td>
<td>390.4</td>
<td>1.15</td>
<td>200</td>
</tr>
<tr>
<td>0009</td>
<td>0001</td>
<td>75</td>
<td>0.70000000</td>
<td>0.50000000</td>
<td>0009</td>
<td>2008-11-15</td>
<td>500.0</td>
<td>282.7</td>
<td>1.15</td>
<td>513</td>
</tr>
<tr>
<td>0010</td>
<td>0002</td>
<td>75</td>
<td>0.40000000</td>
<td>1.50000000</td>
<td>0010</td>
<td>2008-09-01</td>
<td>500.0</td>
<td>0.00</td>
<td>0.83</td>
<td>200</td>
</tr>
<tr>
<td>0011</td>
<td>0002</td>
<td>75</td>
<td>0.40000000</td>
<td>1.50000000</td>
<td>0011</td>
<td>2008-11-01</td>
<td>500.0</td>
<td>0.00</td>
<td>0.83</td>
<td>200</td>
</tr>
<tr>
<td>0012</td>
<td>0002</td>
<td>75</td>
<td>0.40000000</td>
<td>1.50000000</td>
<td>0012</td>
<td>2008-11-15</td>
<td>500.0</td>
<td>102.2</td>
<td>0.83</td>
<td>200</td>
</tr>
<tr>
<td>0013</td>
<td>0002</td>
<td>75</td>
<td>0.40000000</td>
<td>1.50000000</td>
<td>0013</td>
<td>2008-09-01</td>
<td>500.0</td>
<td>85.46</td>
<td>0.00</td>
<td>449</td>
</tr>
<tr>
<td>0014</td>
<td>0002</td>
<td>75</td>
<td>0.40000000</td>
<td>1.50000000</td>
<td>0014</td>
<td>2008-11-01</td>
<td>500.0</td>
<td>500.0</td>
<td>0.00</td>
<td>200</td>
</tr>
<tr>
<td>0015</td>
<td>0002</td>
<td>75</td>
<td>0.40000000</td>
<td>1.50000000</td>
<td>0015</td>
<td>2008-11-15</td>
<td>500.0</td>
<td>500.0</td>
<td>0.00</td>
<td>200</td>
</tr>
<tr>
<td>0016</td>
<td>0002</td>
<td>75</td>
<td>0.40000000</td>
<td>1.50000000</td>
<td>0016</td>
<td>2008-09-01</td>
<td>500.0</td>
<td>500.0</td>
<td>0.00</td>
<td>200</td>
</tr>
<tr>
<td>0017</td>
<td>0002</td>
<td>75</td>
<td>0.40000000</td>
<td>1.50000000</td>
<td>0017</td>
<td>2008-11-01</td>
<td>500.0</td>
<td>500.0</td>
<td>0.00</td>
<td>200</td>
</tr>
<tr>
<td>0018</td>
<td>0002</td>
<td>75</td>
<td>0.40000000</td>
<td>1.50000000</td>
<td>0018</td>
<td>2008-11-15</td>
<td>500.0</td>
<td>500.0</td>
<td>0.00</td>
<td>200</td>
</tr>
</tbody>
</table>
1.7 SAS Mixed Effects Procedures

• Main procedures discussed:
  • PROC MIXED
  • PROC GLIMMIX (experimental in v9.1.3; v9.2)

• Others with mixed model capabilities:
  • PROC GLM; PROC HPMIXED (v9.2); PROC LATTICE; PROC NESTED; PROC NLMIXED; PROC VARCOMP
2.1 From Linear to Multi-level Models

• Multi-level models (MMs) as generalization of linear and generalized linear models

• Assumption of independent error terms

\[ Y_{ij} = \beta_0 + \beta_kX_{ijk} + r_{ij} \text{ with } r_{ij} \sim N(0, \sigma^2) \]

PROC REG DATA= DATA_001;
MODEL Y = X1-Xp;
RUN;

PROC MIXED DATA= DATA_001;
MODEL Y = X1-Xp;
RUN;
2.2 Null Multi-level Model

- Two-level data; individuals within groups
- Intercepts vary by group; no other pred.
- Notation from Raudenbush & Bryk (2001).

Model

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

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

Variance Components

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

\[ T = [\tau_{00}] \]

SAS Code

```sas
PROC MIXED DATA= DATA_001
  COVTEST;
  MODEL Y = / SOLUTION;
  RANDOM INTERCEPT / SUBJECT= Group TYPE= VC;
RUN;
```
2.3 Null MM as a Single Equation

- Substituting the expression for $\beta_{0j}$ from the group-level into the individual-level equation:

$$
\beta_{0j} = \gamma_{00} + u_{0j}
$$
$$
Y_{ij} = \beta_{0j} + r_{ij}
$$

Grand mean

Group random effect

Residual
2.4 Graph of Null MM

- Individual variation about group mean
- Group variation about grand mean
2.5 Do we Need a Multi-level Model?

• Based on the null model (or an ANOVA table) we compute the intra-class correlation coef. (ICC) for linear models

• Indicates the proportion of error variance in the outcome relative to overall error variance.

\[
\rho = \frac{\tau_{00}}{\tau_{00} + \sigma^2} = \frac{\text{Group-level Error Var.}}{\text{Total Error Var.}}
\]
2.6 SAS Output: Null Model

- ICC = 0.014 / 0.10 = 0.14
- ICC not meaningful for GLMM
- COVTEST option in PROC MIXED
2.7 SAS Output: Null Model

• Random intercept specified
• No other predictors included in the model
3.1 MM with Level-1 Random Intercepts & Fixed Slopes

- Intercepts vary by group; other predictors are fixed

**Model**

\[ Y_{ij} = \beta_{0j} + \sum_{k=1}^{p} \beta_{kj}X_{kj} + r_{ij} \]

- \( \beta_{0j} = \gamma_{00} + u_{0j} \)

- \( \beta_{kj} = \gamma_{k0} \)

**Variance Components**

\[ r_{ij} \sim N(0, \sigma^2) \quad T = [\tau_{00}] \]

**SAS Code**

```sas
PROC MIXED DATA= DATA_001
   COVTEST;
   MODEL Y= X1-Xp / SOLUTION;
   RANDOM INTERCEPT / SUBJECT= Group TYPE= VC;
RUN;
```

Multi-level, NYASUG, Dec. 2008
3.2 Graph of MM with Level-1 Fixed X1 Slope
3.3 Graph of Null MM
3.4 SAS Output: Fixed Slopes

- Same random effects as null
- Slopes are constrained to be equal across groups
3.5 MM with Fixed- & Random-Intercept and Slope Effects

- Intercepts and other parameters vary

\[ Y_{ij} = \beta_0 + \sum_{k=1}^{p} \beta_{kj} X_{kj} + r_{ij} \]

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

\[ \beta_{kj} = \gamma_{k0} + u_{kj} \]

Variance Components

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

**SAS Code**

```
PROC MIXED DATA= DATA_001
   COVTEST;
   MODEL Y= X1-Xp / SOLUTION;
   RANDOM INTERCEPT X1-Xp /
     SUBJECT= Group TYPE= VC;
RUN;
```
3.6 Graph of MM with Level-1 Fixed- and Random-Effects
3.7 SAS Output: Random Slopes

- Same random effects as null
- Slope for $X_1$ allowed to be vary across groups
- Notice $\tau_{11}$ in CovParms table
3.8 Generalized Linear Multi-level Models

- With a non-normal outcome, it may be appropriate to use a generalized linear multi-level model (GLMM)
  - PROC GLIMMIX will meet most needs; PROC NLMIXED may be appropriate otherwise
  - Recall ICC not meaningful; rely on size and standard error of cov. parameter estimates
  - Prediction sluggish in exp. GLIMMIX
3.9 SAS Specification of GLMM

PROC GLIMMIX DATA = Data_004
   METHOD= RSPL INITGLM IC= PQ
   NAMELEN= 65 NOCLPRINT;
   CLASS Group_ID;
   NLOPTIONS TECHNIQUE= NRRIDG;
   MODEL Event= X1-Xp / DIST= BINARY
      LINK= LOGIT SOLUTION CL;
   RANDOM INTERCEPT / SUBJECT= Group_ID;
RUN;
3.10 PROC GLIMMIX for Propensity Score Matching

• Add the following to our PROC GLIMMIX:

```plaintext
OUTPUT OUT= Data_004_Pred
   PREDICTED(BLUP ILINK)= Treat_Prob_Pred
   STDERR(BLUP ILINK)= Treat_Prob_SE;
```

• Use the predicted probabilities to perform propensity score matching.
3.11 Notes on Covariance Structure

• Theory and subject matter expertise should dictate the covariance structure

• For hierarchically structured data:
  • Generally most restrictive ($J$ parm.): `TYPE= VC`
  • Generally least restrictive ($J^2$ parm.): `TYPE= UN`

• Many other options; check OnlineDoc
3.12 What Can Go Wrong?

• Convergence problems
  • PROC HPMIXED? Need SAS 9.2.
  • Another estimation method?
  • Fewer fixed- and random-effects?
• Variance components not estimated
  • Consider modifying RANDOM statement.
3.13 What *Else* Can Go Wrong?

- “Out of memory” Error
  - Some mixed procedures are prone to this error
  - “PROC MIXED is looking for a contiguous memory space and cannot find one large enough due to the relocation of the operating system modules.” -SAS Problem Note 15060.

- **Solution**: Try to defragment the target drive.
3.14 Diagnosing and Solving Problems

- Diagnose what’s going wrong and when
  - Use the LOGNOTE option on PROC MIXED for estimation progress step-by-step
  - Remove statements / options to find the issue
  - If convergence is a problem, consider re-parameterizing the model
    - Especially in the case of generalized linear MMs
3.15 Example of LOGNOTE Output

NOTE: Levelizing effects.  
NOTE: Processing subject and group effects.  
NOTE: Setting up data.  
NOTE: Loading data.  
NOTE: Computing likelihood in iteration 0.  
NOTE: Computing G derivatives in iteration 1.  
NOTE: Computing likelihood in iteration 1.  
NOTE: Computing G derivatives in iteration 1.  
NOTE: Convergence criteria met but final hessian is not positive definite.  
NOTE: Computing likelihood in iteration 2.  

NOTE: A linear combination of covariance parameters is confounded with the residual variance.  
NOTE: Computing Cholesky root of cross-products matrix.  
NOTE: Computing H matrix.  
NOTE: Computing Type 3 sums of squares.  
NOTE: PROCEDURE MIXED used (Total process time):  
real time 0.35 seconds  
cpu time 0.14 seconds
4.1 Research & SAS Applications of Multi-level Models

• Model first-year grade point average at 110 colleges and universities

• Model probability participation in Advanced Placement Program® at many high schools

• Model high-school- and college-variability in the effect of AP® participation
4.2 Model FYGPA Across 110 Colleges

- **Outcome**: FYGPA, as normal continuous
- **Levels**: student and college (nested)
- **Possible SAS Procedures**:
  - PROC MIXED
  - PROC HPMIXED (new in 9.2)
    - For large number of fixed- or random-effects
4.3 Model AP Participation Across Many High Schools

• **Outcome**: Indicator of AP Participation

• **Levels**: student and high school (nested)
  - High schools may be nested within districts

• **Possible SAS Procedures**:
  - PROC GLIMMIX (9.2; exp. in 9.1.3)
  - PROC NLMIXED
4.4 Model Effect of AP® Participation across HS & Colleges

- **Outcome**: FYGPA, as normal continuous
- **Levels**: student, high school and college
  - Data are cross-classified by HS and college
- **Possible SAS Procedures**:
  - PROC MIXED
  - PROC HPMIXED (new in 9.2)
4.5 HS & College Cross-classified Data

• Data are not strictly nested
  • Students from high schools A, B and C attend colleges I and II
  • Random effects: high school and college levels
4.6 SAS Specification of CCMM

PROC MIXED DATA = Data_002 COVTEST;
    CLASS HS_ID College_ID;
    MODEL Y= X1-Xp / SOLUTION CL;
    RANDOM INTERCEPT X1 /
        SUBJECT= HS_ID;
    RANDOM INTERCEPT X1-Xp /
        SUBJECT= College_ID;
RUN;
4.7 Non-Educational Applications

- Multi-site clinical trials
  - Patients strictly nested within sites or cross-classified across sites

- Retail applications
  - PROC GLIMMIX to estimate same-store sales over quarterly observations
4.8 General References

• Multi-level Modeling Texts:
4.9 SAS-Oriented Reference

- Multi-level Modeling Articles for SAS:
5.1 Question / Answer Period

• Any questions? Statistical, SAS-oriented or otherwise?
Thank You

• Thank you all very much for having me.
• Special thanks to Martin Feuerman of NYASUG and Mary-Margaret Kerns of The College Board.
Contact Information

Brian F. Patterson
The College Board, Research & Development
Assistant Research Scientist

212-713-7714
bpatterson@collegeboard.org

http://www.collegeboard.com/research