Cross-Sectional and Longitudinal Effects with Repeated Measures

Overview

There are many observational studies that collect repeated measures over time on a sample of subjects. We discuss some simple examples, and explore how to analyze data from such examples by simulating data for the example, and using the parameters from the simulation to check that we have specified an appropriate analysis. There is a danger in such settings in not including a random effect for a subject. Without a random effect, the slope estimates are a weighted average of cross-sectional and longitudinal slopes. Normally, this result is not interpretable. This problem can be avoided by including random subject effects. If subject-specific means (for the independent variables) are included, both cross-sectional and longitudinal effects can be estimated.

Introductory Example

Example 1. We consider as an example a random coefficient model for growth of children. Suppose that children are measured at five ages (8, 9, 10, 11, and 12), with two measures made of height at each year. Suppose that we assume growth is linear for each child, but that the slope and intercept may differ between children. Let us assume that the slopes and intercepts for subjects in the population follow a normal distribution, where \( E(B_i) = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} 50 \\ 2 \end{pmatrix} \), while the variance is given by \( \text{var}(B_i) = \begin{pmatrix} \sigma_{00} & \sigma_{01} \\ \sigma_{01} & \sigma_{11} \end{pmatrix} = \begin{pmatrix} 5 & 2 \\ 2 & 3 \end{pmatrix} \). Notice that the intercept and slope of the \( i^{th} \) selected subject is given by \( \begin{pmatrix} \beta_0 + B_{0i} \\ \beta_1 + B_{1i} \end{pmatrix} \). Finally, suppose that each subject is measured twice at each age, with the response variance given by \( \sigma_r^2 = 0.4 \). We represent a mixed model for the \( i^{th} \) selected subject at age \( j \) as

\[ Y_{ij} = \beta_0 + x_{ij} \beta_1 + B_{0i} + x_{ij} B_{1i} + \epsilon_{ij}. \]

The following SAS code generates a set of data that corresponds to a simple random sample from such a population.

Source: mm06p16.sas on 05/12/06 by ejs

DATA d1;
%LET n_id=20;   * Number of subjects;
%LET n_rep=1;   * Number of replications at a time;
%LET beta0=50;  * Mean height;
%LET age_m=10;  * Average age in population;
%LET v_agem=2; * Variance in average age for subjects;
%LET xb1=-4;    * Cross-sectional change in ht with age;
xb0=&beta0-&xb1*&age_m;   *Intercept for crosssectional model;
%LET beta1=2;   * Within height growth;
%LET v_00=5; * Variance of intercepts for Random coefficients centered at 0;
%LET v_11=3; * Variance of slopes for random coefficients centered at 0;
%LET v_01=2; * Covariance of intercepts and slopes;
%LET v_rep=0.05;

DO id=1 to &n_id;
  ********************************************;
  *** Generate slope and intercept for subject;
  ********************************************;
  b0_star=sqrt(&v_00)*rannor(321123);
  b1_star=sqrt((&v_00*&v_11-&v_01*&v_01)/&v_00)*rannor(45667);
  ***********************************************;
  ** With Ave(age)=0, the intercept(height)=b0, and the ;
  ** within subject slope(height)=b1          ;
  ***********************************************;
  b0=&beta0+b0_star;
  b1=&beta1+(&v_01/&v_00)*b0_star+b1_star;
  ***********************************************;
  ** Determine the subjects average age=age_m by adding the random intercept ;
  ** Intercept(ht)=b0+b1*age_m                  ;
  ***********************************************;
  age_idm=&age_m+v_agem*rannor(2323);    *average age for subject in the study;
  ht_m=xb0+xb1*age_idm;                  *true Height of subject at average age;
  DO age=-2 to 2;             *Yearly increments for measures relative age to average age;
    true_ht=ht_m+b1*(age);    * True height of subject at deviation from average age;
    tage=age+age_idm;           *true age;
  DO rep=1 to &n_rep;
    ht=true_ht+sqrt(&v_rep)*rannor(234355);;
    OUTPUT;
  END;
  END;
END;
RUN;
PROC PRINT DATA=d1 (OBS=20) NOOBS;
BY id;
VAR id age age_idm tage ht true_ht b0 b1 ;
TITLE2 'Table 16-1. (2006) List of simulated data of height at 5 ages ;';
TITLE3 ' # Subjects=&n_id, #reps=&n_rep Intercept=&beta0 Longitudinal Slope=&beta1 ;';
TITLE4 ' Cross-sectional Slope=&xb1 Response error=&v_rep;';
TITLE5 ' var(interc)=&v_00  var(slope)=&v_11  cov(int, slope)=&v_01  '; 
RUN;

An example of the type of data generated is given in Table 16-1.

Source: mm06p16.sas on 05/12/06 by ejs
Table 16-1. (2006) List of simulated data of height at 5 ages
  # Subjects=20, #reps=1 Intercept=50 Longitudinal Slope=2
  Cross-sectional Slope=-4 Response error=0.05
  var(interc)=5 var(slope)=3 cov(int, slope)=2

<table>
<thead>
<tr>
<th>id</th>
<th>age</th>
<th>age_idm</th>
<th>tage</th>
<th>ht</th>
<th>true_ht</th>
<th>b0</th>
<th>b1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-2</td>
<td>9.99664</td>
<td>7.9966</td>
<td>43.7045</td>
<td>44.1509</td>
<td>50.5275</td>
<td>2.93128</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>9.99664</td>
<td>8.9966</td>
<td>47.7503</td>
<td>47.0822</td>
<td>50.5275</td>
<td>2.93128</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>9.99664</td>
<td>9.9966</td>
<td>50.4150</td>
<td>50.0135</td>
<td>50.5275</td>
<td>2.93128</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>9.99664</td>
<td>10.9966</td>
<td>52.6978</td>
<td>52.9447</td>
<td>50.5275</td>
<td>2.93128</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>9.99664</td>
<td>11.9966</td>
<td>56.0051</td>
<td>55.8760</td>
<td>50.5275</td>
<td>2.93128</td>
</tr>
</tbody>
</table>
A plot of the data for 20 subjects is given in Figure 1.

symboll i=join r=200;
PROC GPLOT DATA=d1;
   PLOT ht*tage=id;
   TITLE1 "Figure 1. Summary of Height for 20 subjects";
   FOOTNOTE1 "&prg";
RUN;
FOOTNOTE1 "";
TITLE1 "&prg";
We fit a simple regression model to these data to illustrate the problem of not including random effects. The slope coefficient in this model can not readily be interpreted.

```
PROC MIXED DATA=d1;
   CLASS id;
   MODEL ht=tage/SOLUTION;
   TITLE2 'Table 16-2. (2006) Model fit without accounting for random subject effects';
   TITLE3 ' # Subjects=&n_id, #reps=&n_rep Intercept=&beta0 Slope=&beta1 Response error=&v_rep';
   TITLE4 ' Average age=&age_m Cross-sectional Slope=&xb1 Response error=&v_rep';
   TITLE5 ' var(interc)=&v_00 var(slope)=&v_11 cov(int, slope)=&v_01 '
RUN;
```
We fit three other models including random subject effects (Table 16-3), a random coefficient model (Table 16-4), and a random coefficient model including both cross-sectional and longitudinal effects (Table 16.5) (where we increased the simulation to include 200 subjects). The model that includes cross-sectional and longitudinal effects includes the mean age for each subject.

PROC MIXED DATA=d1;
   CLASS id;
   MODEL ht=tage/SOLUTION;
   RANDOM id;
   TITLE2 'Table 16-3. (2006) Mixed model with random subject effect';
   TITLE3 ' # Subjects=&n_id, #reps=&n_rep Intercept=&beta0 Slope=&beta1 Response error=&v_rep';
   TITLE4 ' Average age=&age_m Cross-sectional Slope=&xb1 Response error=&v_rep';
   TITLE5 ' var(interc)=&v_00 var(slope)=&v_11 cov(int, slope)=&v_01 ';
RUN;

PROC MIXED DATA=d1;
   CLASS id;
   MODEL ht=tage/SOLUTION;
   RANDOM INTERCEPT tage/TYPE=UN SUBJECT=id;
   TITLE2 'Table 16-4. (2006) Mixed model with random coefficients for subjects ';
   TITLE3 ' # Subjects=&n_id, #reps=&n_rep Intercept=&beta0 Slope=&beta1 Response error=&v_rep';
   TITLE4 ' Average age=&age_m Cross-sectional Slope=&xb1 Response error=&v_rep';
   TITLE5 ' var(interc)=&v_00 var(slope)=&v_11 cov(int, slope)=&v_01 ';
RUN;
PROC MIXED DATA=d1;
  CLASS id;
  MODEL ht= age_idm tage/SOLUTION;
  RANDOM INTERCEPT tage/TYPE=UN SUBJECT=id;
  TITLE2 "Table 16-5. (2006) Mixed model with random subject effect and Xsec and Longit. slopes";
  TITLE3 "  \# Subjects=\n_id, \#reps=\n_rep Intercept=&beta0 Slope=&beta1 Response error=&v_rep";
  TITLE4 "  Average age=&age_m Cross-sectional Slope=&xb1 Response error=&v_rep";
  TITLE5 "  var(interc)=&v_00 var(slope)=&v_11 cov(int, slope)=&v_01 ";
RUN;

Source: mm06p16.sas on 05/12/06 by ejs
Table 16-3. (2006) Mixed model with random subject effect
  \# Subjects=20, \#reps=1 Intercept=50 Slope=2 Response error=0.05
  Average age=10 Cross-sectional Slope=-4 Response error=0.05
  var(interc)=5 var(slope)=3 cov(int, slope)=2

  Convergence criteria met.

  Covariance Parameter
  Cov Parm     Estimate
    id             139.79
  Residual       5.0376

  Fit Statistics
  -2 Res Log Likelihood           540.3
  AIC (smaller is better)         544.3
  AICC (smaller is better)        544.5
  BIC (smaller is better)         546.3

  Solution for Fixed Effects
    Standard
  Effect       Estimate       Error      DF    t Value    Pr > |t|
    Intercept     32.6814      3.0636      19      10.67      <.0001

Source: mm06p16.sas on 05/12/06 by ejs
Table 16-4. (2006) Mixed model with random coefficients for subjects
  \# Subjects=20, \#reps=1 Intercept=50 Slope=2 Response error=0.05
  Average age=10 Cross-sectional Slope=-4 Response error=0.05
  var(interc)=5 var(slope)=3 cov(int, slope)=2

  Convergence criteria met.

  Covariance Parameter Estimates
  Cov Parm     Subject    Estimate
    UN(1,1)      id           338.45
    UN(2,1)      id         -19.1077
    UN(2,2)      id           2.0580
  Residual                 0.06935

  Fit Statistics
  -2 Res Log Likelihood           312.6
  AIC (smaller is better)         320.6
  AICC (smaller is better)        321.0
  BIC (smaller is better)         324.5
Solution for Fixed Effects

| Effect  | Estimate | Error | DF  | t Value | Pr > |t| |
|---------|----------|-------|-----|---------|-------|
| Intercept | 31.9246  | 4.1179 | 19  | 7.75    | <.0001|
| tage     | 1.9808   | 0.3213 | 19  | 6.16    | <.0001|

Source: mm06p16.sas on 05/12/06 by ejs
Table 16-5. (2006) Mixed model with random subject effect and Xsec and Longit. slopes
# Subjects=200, #reps=1  Intercept=50 Slope=2 Response error=0.05
Average age=10  Cross-sectional Slope=-4 Response error=0.05
var(interc)=5  var(slope)=3  cov(int, slope)=2
Convergence criteria met.

Covariance Parameter Estimates

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Subject</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>UN(1,1)</td>
<td>id</td>
<td>0</td>
</tr>
<tr>
<td>UN(2,1)</td>
<td>id</td>
<td>-0.1287</td>
</tr>
<tr>
<td>UN(2,2)</td>
<td>id</td>
<td>0.01227</td>
</tr>
<tr>
<td>Residual</td>
<td>id</td>
<td>6.7865</td>
</tr>
</tbody>
</table>

Fit Statistics

-2 Res Log Likelihood 4031.3
AIC (smaller is better) 4037.3
AICC (smaller is better) 4037.3
BIC (smaller is better) 4047.2

Solution for Fixed Effects

| Effect     | Estimate | Error  | DF  | t Value | Pr > |t| |
|------------|----------|--------|-----|---------|-------|
| Intercept  | 89.9952  | 0.1031 | 198 | 872.73  | <.0001|
| age_idm    | -6.0751  | 0.05856| 600 | -103.74 | <.0001|
| tage       | 2.0751   | 0.05878| 199 | 35.31   | <.0001|