Sampling, WLS, and Mixed Models

II ESAMP Meetings
Nov 6, 2009
Natal, Brazil

Ed Stanek and Julio Singer
U of Mass, Amherst, and
U of Sao Paulo, Brazil

Finite Population Mixed Models
Research Group

Luz Mery Gonzalez, Columbia; Viviana Lencina, Argentina; Julio Singer,
Brazil; Silvina San Martino, Argentina; Wenjun Li, US; and Ed Stanek US

Background

- Motivation:
  - 2-stage cluster sample of hospitals
    - m Hospitals
    - m Appendectomy operations per hospital
  - What is the average cost of an operation at a selected hospital (latent value)?

- Choices:
  - Use average cost of m operations for selected hospital
  - Use ‘shrunk’ cost- regressing to the mean for other sample hospitals

- Which should we use?

How do we make up models to get better insight from limited information?

- Consider/Account for:
  - Study Design
  - Sampling
  - Response Error

- Model Assumptions

An Example

What is a subject’s saturated fat intake?

Seasons Study UMASS Worc

Figure 1. Mean vs Standard Deviation of Saturated Fat Intake (g/m/day) for 12854 subjects with 0 or more 24 recall measures in Seasons Study

Source: Data from the 1989-91 DSS by US
The Problem-Simplified

- Observe:
  - 1 Measure of SFat on each Subject
- Assume:
  - Response Error (RE) Variance known
- Question:
  - How do we estimate Subject’s True Sat Fat intake?

Begin with a Response Error Model ... which leads to:
- Mixed Model
- Finite Population Mixed Model

Daisy  Lily  Rose

Population

Population Set

Response

Response
Response Error Model for Set

\[ Y_{jk} = E_R(Y_{jk}) + E_{jk} \]
\[ E_R(Y_{jk}) = y_j \]
\[ \text{var}(Y_{jk}) = \sigma_j^2 \]
\[ Y_{jk} = y_j + E_{jk} \]

Summary Response Error Model

\[ \sigma_1^2 = 1 \]
\[ \sigma_2^2 = 4 \]
Re-parameterized RE Model
\[
Y_1 = y_1 + E_1 \\
= \mu + \beta_1 + E_1 \\
Y_2 = y_2 + E_2 \\
= \mu + \beta_2 + E_2
\]
Mean Latent Value-of what?: or the Population
\[
\mu = \frac{1}{N} \sum_{j=1}^{N} y_j
\]

Generating Response in the RE Model
\[
\text{Sample Space}
\]

Generating Response in the RE Model
\[
\sigma_i^2 = 1 \\
\sigma_j^2 = 4
\]

Generating an Observed Response in the RE Model
\[
Y_i = \mu + \beta_1 + E_i \\
Y_j = \mu + \beta_2 + E_j
\]

Generating an Observed Response in the RE Model
\[
9 = 6 + 4 + -1 \\
0 = 6 + 4 + -2
\]

Response Error Model
\[
Y_j = \mu + \beta_j + E_j
\]

Response Error Model
\[
j = 1 \\
j = 2
\]
Mixed Model (MM)

\[ Y_j^* = \mu + a_j + E_j \]

Random Effect

Mixed Model (MM)

\[ Y_j^* = \mu + a_j + E_j \]

Latent Value

\[ E_j(a_j) = 0 \]
\[ \text{var}(a_j) = \gamma^2 \]

Mixed Model (MM) in action

\[ y^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \mu)^2 \]

Mixed Model (MM) in action

\[ \beta \]
\[ \beta \]

Mixed Model (MM) in action

\[ \beta \]
\[ \beta \]

Mixed Model (MM) in action

\[ \beta \]
\[ \beta \]

Who Are They??

Mixed Model (MM) Mixed Model (MM) Mixed Model (MM) Mixed Model (MM)

\[ Y_1^* = \mu + \beta_1 E_1 \]
\[ Y_2^* = \mu + \beta_2 E_2 \]

What Does it Mean??
### Sample Space (MM)

<table>
<thead>
<tr>
<th>Artificial</th>
<th>Real</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 8 3 8</td>
<td>11 9</td>
</tr>
<tr>
<td>1 12 3 12</td>
<td>11 9</td>
</tr>
</tbody>
</table>

### MM-Latent Values?

<table>
<thead>
<tr>
<th>Samples</th>
<th>Daisy (j=1)</th>
<th>Rose (j=2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y'</td>
<td>P'</td>
<td>E'</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>-1</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>-1</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>10</td>
<td>1</td>
</tr>
</tbody>
</table>

### What are they (for Daisy)?

\[
Y'_j = P'_j + E'_j
\]

### What are they (for Rose)?

\[
Y'_j = P'_j + E'_j
\]

### BLUPs of the MM-Latent Value

\[
\hat{Y}'_j = (\mu + a_j) + E_j
\]

\[
\text{Var}(a_j) = 0
\]

\[
\text{Var}(E_j) = \sigma_j^2
\]

\[
\hat{Y}'_j = \hat{\mu} + k_j \left( Y'_j - \hat{\mu} \right)
\]

\[
\hat{\mu} = \frac{\sum w_j Y'_j}{\sum w_j}
\]

\[
k_j = \frac{\sum w_j^2}{\sum w_j^2 + \sum w_j^2}
\]

\[
\text{Var}(\hat{Y}'_j) = \sigma_j^2 k_j \left( 1 + \frac{1}{nk} \right)
\]

### MSE of BLUPs for MM-Latent Values

<table>
<thead>
<tr>
<th>Samples</th>
<th>Daisy (j=1)</th>
<th>Rose (j=2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P'</td>
<td>P'</td>
<td>P'</td>
</tr>
<tr>
<td>3.1</td>
<td>2</td>
<td>0.81</td>
</tr>
<tr>
<td>1.2</td>
<td>2</td>
<td>1.15</td>
</tr>
<tr>
<td>3.1</td>
<td>2</td>
<td>0.71</td>
</tr>
<tr>
<td>1.1</td>
<td>2</td>
<td>1.28</td>
</tr>
<tr>
<td>10.9</td>
<td>10</td>
<td>1.28</td>
</tr>
<tr>
<td>8.9</td>
<td>10</td>
<td>0.71</td>
</tr>
<tr>
<td>10.8</td>
<td>10</td>
<td>1.15</td>
</tr>
<tr>
<td>8.9</td>
<td>10</td>
<td>0.81</td>
</tr>
</tbody>
</table>

\[
\text{Ave} = 0.986
\]

\[
\text{Ave} = 3.768
\]
MSE of BLUPs $| P = y $

<table>
<thead>
<tr>
<th>Samples</th>
<th>Daisy ($j=1$)</th>
<th>Rose ($j=2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_1$</td>
<td>10.90</td>
<td>10.84</td>
</tr>
<tr>
<td>$P_2$</td>
<td>10.81</td>
<td>10.82</td>
</tr>
<tr>
<td>$P_3$</td>
<td>4.41</td>
<td>9.64</td>
</tr>
<tr>
<td>$P_4$</td>
<td>2.49</td>
<td>2.29</td>
</tr>
<tr>
<td>$P_5$</td>
<td>2.59</td>
<td>2.18</td>
</tr>
<tr>
<td>$P_6$</td>
<td>2.24</td>
<td>2.15</td>
</tr>
</tbody>
</table>

MSE = Ave = 0.986  MSE = Ave = 3.768

Response Error Model

$Y = y + E$  
$= \mu + \beta_s + E_s$

Latent Value

$Y_1 = y_1 + E_1$  
$Y_2 = y_2 + E_2$  
$Y_3 = y_3 + E_3$

$
\mu = \frac{1}{N} \sum_{s=1}^{S} y_s$

$y_1 = 10$  
$y_2 = 3$  
$y_3 = 2$

$\sigma^2_1 = 1$  
$\sigma^2_2 = 100$  
$\sigma^2_3 = 4$

Accounting for Sampling

Indicator random variable, 1 if ith selected sample subject is subject “s”

$Y^*_i = \sum_{s=1}^{S} U_{i,s} Y_s$

$Y_{i,s} = \mu + \beta_s + E_s$

$U_{i,s} = \begin{cases} 
1 & \text{if } i \text{th selected sample subject is subject } “s” \\
0 & \text{otherwise}
\end{cases}$

$Y^*_i = \mu + b_i + E_i$

$E^*_i = \sum_{s=1}^{S} U_{i,s} E_s$

Finite Population Mixed Model (FPMM)

$Y_n^* = \mu + b_i + E_i$  
$i = 1, \ldots, n$

$Y^*_1 = \mu + b_1 + E_1$

$Y^*_2 = \mu + b_2 + E_2$
Finite Population Mixed Model (FPMM)

\[ Y^*_n = \mu + \beta_1 E_1 + \beta_2 + \beta_3 E_3 \]

\[ Y^*_n = \mu + \beta_1 E_1 \]

\[ Y^*_n = \mu + \beta_2 + E_i \]

FPMM - Sample Space ...

FPMM - Sample Space ...

FPMM - Sample Space ...

FPMM - Sample Space ...

FPMM - Sample Space ...

FPMM - Sample Space ...
FPMM- Sample Space ...

**All sample points are Potentially Observable**

FPMM- BLUPs of Realized Latent Values

\[ Y^*_i = \mu + b_i + E_i \]

\[ i = 1 \]

\[ Y^*_1 = \mu + \beta \]

\[ i = 2 \]

\[ Y^*_2 = \mu + \beta \]

FPMM- BLUPs of Realized Latent Values

\[ \text{var}_i(b) = \sigma^2 \]

\[ \text{var}_i(E) = \sigma^2 = \frac{1}{N-1} \sum_{i=1}^{N}(Y_i - \mu)^2 \]
FPMM-BLUPs of Realized Latent Values

\[ Y''_i = (\mu + b_i) + E''_i \]

\[ = P''_i + E''_i \]

\[ \hat{P}_i = \bar{Y} + k(Y''_i - \bar{Y}) \]

\[ \bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y'_i \]

\[ k = \frac{\gamma}{\gamma' + \sigma'} \]

\[ \text{var}_{X'}(\hat{P}_i - P_i) = \frac{\sigma'^2}{n} \left[ 1 + k(n - 1) \right] \]

\[ \text{MSE}_{X'}(\hat{P}_i - P_i) = \frac{1}{n} \left[ (n - 2)k^2 + \frac{1}{n} \frac{1}{n - 1} \left( \sum_{j=1}^{n} (Y_j - \mu_j)^2 \right) \right] \]

Comparison of MM-BLUP and FPMM-BLUP

**MM-BLUP**

Predictor

\[ \hat{Y}_j = \bar{Y} + k(Y'_j - \bar{Y}) \]

\[ \bar{Y} = \frac{1}{n} \sum_{j=1}^{n} Y'_j \]

\[ k = \frac{\gamma}{\gamma' + \sigma'} \]

**FPMM-BLUP**

\[ \hat{Y}_j = \hat{\mu} + k_j \left( Y''_j - \hat{\mu} \right) \]

\[ \hat{\mu} = \frac{1}{n} \sum_{j=1}^{n} Y'_j \]

\[ k_j = \frac{\gamma_j}{\gamma_j' + \sigma_j'} \]

Comparison of FPMM-BLUP and MM-BLUP

To Compare, Focus on...

...THIS Sample Space

Bigger Sample (n=3) Population (N=4)
n=3, What is Lily's Latent value?

- Use n=3 subject effects for MM
  1 possible sample set

- 8 sample points

- 8x(6 permutations)=48 sample points

- \( N \choose n = \binom{4}{3} = 4 \) Combinations

- Select one sequence

192 Sample Points
Select one sequence, Observe Sample Point

FPMM-Average MSE of Predictor over Permutations

Ave MSE

Summary MSE Results

Conclusions
Conclusions

Evaluate Performance Conditional on the Sample

Model Based

MM-BLUP
\[ \hat{\beta}_j = \hat{\beta} + k_j (Y^*_j - \hat{\mu}) \]

Design Based

FPMM-BLUP
\[ \hat{\beta}_j = \hat{\beta} + k_j (Y^*_j - \hat{\mu}) \]

Conclusions

Conceptual “Priors”

Model Based

MM-BLUP
\[ \hat{\beta}_j = \hat{\beta} + k_j (Y^*_j - \hat{\mu}) \]

To Evaluate Performance of BLUP Estimators:
- For Mixed Model: Condition on \( P=y \)
  i.e. MM Latent Values match subject Latent Values
- For the FPMM: Condition on the sample set

MSE for BLUPs not evaluated Correctly
- Extends to WLS estimate of mean
- MM-BLUP not always best

Conclusions

Evaluate Performance Conditional on the Sample

Model Based

MM-BLUP
\[ \hat{\beta}_j = \hat{\beta} + k_j (Y^*_j - \hat{\mu}) \]

Thanks

Any thoughts? Next steps? Questions?