Introduction to Mixed Models

• Linear Models
• Fixed and Random Effects
• Fixed Models
  – Hypothesis testing
  – Estimation
• Mixed Models
  – Hypothesis testing
  – Prediction
Linear Models

• Needed to correct for unbalancedness in data
  – Different sires in different herds
  – Account for accuracy of information
Genetic level confounded with herds

Problem:
The contemporaries of some animals may have higher genetic mean than of others

Example

<table>
<thead>
<tr>
<th>sire</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5: link sire</th>
</tr>
</thead>
<tbody>
<tr>
<td>herd 1</td>
<td>325</td>
<td>275</td>
<td>-</td>
<td>-</td>
<td>325</td>
</tr>
<tr>
<td>herd 2</td>
<td>-</td>
<td>-</td>
<td>325</td>
<td>275</td>
<td>375</td>
</tr>
</tbody>
</table>

_progeny means from 4 sires in 2 herds_
Genetic level confounded with herds

Problem 3:
The contemporaries of some animals may have higher genetic mean than of others

Example

<table>
<thead>
<tr>
<th>sire</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5: link sire</th>
</tr>
</thead>
<tbody>
<tr>
<td>herd 1</td>
<td>325</td>
<td>275</td>
<td>-</td>
<td>-</td>
<td>325</td>
</tr>
<tr>
<td>herd 2</td>
<td>-</td>
<td>-</td>
<td>325</td>
<td>275</td>
<td>375</td>
</tr>
</tbody>
</table>
Genetic level confounded with herds

Conclusion

Need links between herds (reference sires)

Need a simultaneous evaluation of all herd and sire effects

The argument for BLUP sire evaluation.....1970ties
Another example

<table>
<thead>
<tr>
<th>Cow</th>
<th>Breed</th>
<th>Feeding regime</th>
<th>Weight (kg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Angus</td>
<td>intensive</td>
<td>494</td>
</tr>
<tr>
<td>2</td>
<td>Angus</td>
<td>intensive</td>
<td>556</td>
</tr>
<tr>
<td>3</td>
<td>Angus</td>
<td>extensive</td>
<td>542</td>
</tr>
<tr>
<td>4</td>
<td>Hereford</td>
<td>extensive</td>
<td>473</td>
</tr>
<tr>
<td>5</td>
<td>Hereford</td>
<td>intensive</td>
<td>632</td>
</tr>
<tr>
<td>6</td>
<td>Hereford</td>
<td>extensive</td>
<td>544</td>
</tr>
</tbody>
</table>
## Order data

<table>
<thead>
<tr>
<th></th>
<th>Intensive</th>
<th>Extensive</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angus</td>
<td>494, 556</td>
<td>542</td>
<td>531</td>
</tr>
<tr>
<td>Hereford</td>
<td>632</td>
<td>473, 544</td>
<td>550</td>
</tr>
<tr>
<td>Mean</td>
<td>561</td>
<td>520</td>
<td></td>
</tr>
</tbody>
</table>
Using a Linear Model

\[
\begin{bmatrix} X & y \end{bmatrix} =
\begin{bmatrix}
1 & 1 & 1 & 494 \\
1 & 1 & 1 & 556 \\
1 & 1 & -1 & 542 \\
1 & -1 & -1 & 473 \\
1 & -1 & 1 & 632 \\
1 & -1 & -1 & 544
\end{bmatrix}
\]

\[
y = X \ b + e
\]

\[
= m + \text{breed} + \text{feed}
\]

class variables

\[
b = \text{inv}(X'X)X'y =
\]

\[
540.1667 \\
-18.3750 \quad \text{Angus} = \text{Hereford} \\
26.6250 \quad \text{Intensive} = \text{Extensive}
\]
Using a Linear Model

\[
\begin{bmatrix}
X & y
\end{bmatrix} =
\begin{bmatrix}
1 & 18 & 494 \\
1 & 21 & 556 \\
1 & 19 & 542 \\
1 & 17 & 473 \\
1 & 23 & 632 \\
1 & 19 & 544 \\
\end{bmatrix}
\]

\[y = X \ b + e\]

\[= m + \text{age}\]

\text{continuous variables}

\[b = \text{inv}(X'X)X'y = \]

59 \hspace{1em} \text{intercept: weight at } 0 \text{ months}

24 \hspace{1em} \text{slope: weight change per month}
A linear regression model
Using a Linear Model

\[
X = \\
\begin{array}{cccc}
1 & 1 & 1 & 18 \\
1 & 1 & 1 & 21 \\
1 & 1 & -1 & 19 \\
1 & -1 & -1 & 17 \\
1 & -1 & 1 & 23 \\
1 & -1 & -1 & 19 \\
\end{array}
\]

\[
b = \text{inv}(X'X)X'y = \\
\begin{array}{c}
-11.3522 \\
-0.6981 \\
-12.2642 \\
28.2830 \\
\end{array}
\]

\[
y = Xb + e \\
= m + \text{breed} + \text{feed} + \text{age}
\]

class variables

and

continuous variables

weight at age = 0
breed effect
feeding effect
age effect
Fixed and Random Effects

Fixed Effects
- Defined classes, comprise all the possible levels of interest
- Number of levels relatively small and confined to this number after repeated sampling.
- E.g. sex, age, breed, contemporary group

Random Effects
- Levels that are considered to be drawn from an infinite large population of levels
- E.g. animal effects
The linear model

- Equation: \( y = Xb + Zu + e \)
- Expectations and Variance Structures
- Assumptions and restrictions
  - Residuals IID?
  - Random sampling of u?
Expectations and Variance Structures

\[
E \begin{pmatrix} y \\ u \\ e \end{pmatrix} = \begin{pmatrix} Xb \\ 0 \\ 0 \end{pmatrix}
\]

\[
\text{Var} \begin{pmatrix} u \\ e \end{pmatrix} = \begin{pmatrix} G & 0 \\ 0 & R \end{pmatrix}
\]

\[
\text{var}(y) = V = ZGZ' + R
\]
Estimating Fixed effects

\[ b = (X'X)^{-1} X'y \]  
Ordinary Least Squares

\[ V = I \sigma^2_e \]

\[ b = (X'V^{-1} X)^{-1} X'V^{-1} y \]  
Weighted Least Squares

\[ V = D \sigma^2_e \]

\[ b = (X'V^{-1} X)^{-1} X'V^{-1} y \]  
Generalized Least Squares

\[ V = V \]
A simple example of variance structure

animal  obs’n
1       9
1       11
2       10

$Z = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$

$Z'Z = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$

$ZZ' = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

$G = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \sigma^2_{anm}$

$\text{var}(y) = ZGZ' + R = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \sigma^2_{anm} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \sigma^2_e$
Estimability

example:

<table>
<thead>
<tr>
<th>Year of Birth</th>
<th>Pedigree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990:</td>
<td>Animal No.: 1</td>
</tr>
<tr>
<td></td>
<td>Sex: ♂</td>
</tr>
<tr>
<td></td>
<td>Weight: 354</td>
</tr>
<tr>
<td>1991:</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>5</td>
</tr>
<tr>
<td>1992:</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>7</td>
</tr>
</tbody>
</table>

(Animal 7 is unrelated to the others.)
\[
Y = X b + e
\]

\[
\begin{pmatrix}
354 \\
251 \\
327 \\
328 \\
301 \\
270 \\
330
\end{pmatrix}
= \begin{pmatrix}
1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 \\
1 & 0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
b_{\text{mean}} \\
b_{1990} \\
b_{1991} \\
b_{1992} \\
e_1 \\
e_2 \\
e_3 \\
e_4 \\
e_5 \\
e_6 \\
e_7
\end{pmatrix}
\]

solutions:
\[
\hat{b} = X' X^{-1} X' Y
\]

\[
X' X = \begin{pmatrix}
7 & 2 & 4 & 1 \\
2 & 2 & 0 & 0 \\
4 & 0 & 4 & 0 \\
1 & 0 & 0 & 1
\end{pmatrix}
\quad \text{and} \quad
X' y = \begin{pmatrix}
2161 \\
605 \\
1226 \\
330
\end{pmatrix}
\]

X is dependent
X’X can not be inverted
Can only estimate 3 parameters from 3 means

Need restriction to solution
\[
\begin{pmatrix}
1 & 1 & 0 \\
1 & 1 & 0 \\
1 & 0 & 1 \\
1 & 0 & 1 \\
1 & 0 & 1 \\
1 & 0 & 0 \\
\end{pmatrix}
\]

e.g. put effect of 1992 to zero

\[
\hat{b} = (X'X)^{-1} X'Y
\]

\[
\hat{b} = \begin{pmatrix}
7 & 2 & 4 \\
2 & 2 & 0 \\
4 & 0 & 4
\end{pmatrix}^{-1}
\begin{pmatrix}
2161 \\
605 \\
1226
\end{pmatrix}
= \begin{pmatrix}
330 \\
-27.5 \\
-23.5
\end{pmatrix}
\]

There are more solutions possible

<table>
<thead>
<tr>
<th>General mean zero</th>
<th>First year zero</th>
<th>Last year zero</th>
<th>Sum of years to zero</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu = 0 )</td>
<td>( \mu = 302.5 )</td>
<td>( \mu = 330 )</td>
<td>( \mu = 313 )</td>
</tr>
<tr>
<td>1990 = 320.5</td>
<td>1990 = 0</td>
<td>1990 = -27.5</td>
<td>1990 = -10.5</td>
</tr>
</tbody>
</table>

estimable functions are unchanged

- expected value of an observation
- difference between years
fitting mean and year effect

\[
\hat{b} = \begin{pmatrix}
330 \\
-27.5 \\
-23.5
\end{pmatrix}
\begin{array}{c}
4.0 \\
\text{the mean of 1992} \\
\text{the effect of year 1990 (relative to 1992)} \\
\text{the effect of year 1991 (relative to 1992)}
\end{array}
\]

fitting mean, year effect and sex

\[
X \hat{b} \quad \text{meaning}
\]

\[
\begin{pmatrix}
1 & 1 & 0 & 1 \\
1 & 1 & 0 & 0 \\
1 & 0 & 1 & 1 \\
1 & 0 & 1 & 0 \\
1 & 0 & 1 & 1 \\
1 & 0 & 1 & 0 \\
1 & 0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
285.7 \\
-5.3 \\
-1.3 \\
44.3
\end{pmatrix} \rightarrow 4.0
\]

the mean of females in 1992
the effect of year 1990 (relative to 1992)
the effect of year 1991 (relative to 1992)
the effect of males (relative to females)

Note: year 1992 appears not so good after all!
Conclusion

• Linear models are a powerful, and relatively simple way to correct for different fixed effects in unbalanced designs
## Connectedness and Confounding

<table>
<thead>
<tr>
<th>Year</th>
<th>sex</th>
<th>Male</th>
<th>Steer</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1991</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>1992</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Connectedness and Confounding

1990 Male 316
1990 Female 314
1990 Male 312
1990 Male 324
1991 Female 311
1991 Male 312
1991 Female 293
1991 Female 304

model statement: weight ~ mu con(sex) con(year)
con(year)  1  2.06  2.06  [DF F_i F_a]
con(sex)  1  4.36  1.19  [DF F_i F_a]

model statement: weight ~ mu con(year) con(sex)
6 con(sex)  1  1.19  1.19  [DF F_i F_a ]
5 con(year)  1  5.23  2.06  [DF F_i F_a ]
Exmp4.dat
15 109 287
17 116 298
18 119 306
18 116 303
19 117 302
19 119 312
20 121 316
21 122 324

weight ~ mu height age
   1 age   1   3.03   3.03   [DF F_inc F_all]
   2 height   1   70.50   1.67   [DF F_inc F_all]

weight ~ mu age height
   2 height   1   1.67   1.67   [DF F_inc F_all]
   1 age   1   71.87   3.03   [DF F_inc F_all]
Hypothesis Testing

• $H'b = c$
• Or: $H'b-c = 0$

$$F = \frac{s / r(H')}{SSE / (N - r(X))}$$

where $s = (H'b-c)'(H'CH^{-1})(H'b-c)$

and $C = (X'V^{-1}X)^{-1}$
Accuracy of estimates

- $V(K'b) = K'(X'V^{-1}X)^{-1}K \cdot se^2$
Mixed model

\[ y = Xb + Zu + e \]

\[
\begin{bmatrix}
X' R^{-1} X & X' R^{-1} Z \\
Z' R^{-1} X & Z' R^{-1} Z + G^{-1}
\end{bmatrix}
\begin{bmatrix}
b \\
u
\end{bmatrix} =
\begin{bmatrix}
X' R^{-1} y \\
Z' R^{-1} y
\end{bmatrix}
\]

simple version

\[
\begin{bmatrix}
X' X & X' Z \\
Z' X & Z' Z + \lambda A^{-1}
\end{bmatrix}
\begin{bmatrix}
b \\
u
\end{bmatrix} =
\begin{bmatrix}
X' y \\
Z' y
\end{bmatrix}
\]

\[
\text{var}(u) = A \sigma_a^2
\]

\[
\text{var}(e) = I \sigma_e^2
\]
In the MMM we estimate

\[ \text{BLUE}(b) = \beta \]
\[ = (X' V^{-1}X)^{-1} X' V^{-1}y \]

is a GLS estimate

\[ \text{BLUP}(u) = \hat{u} \]
\[ = (Z' R^{-1} Z + G^{-1})^{-1} Z' R^{-1}(y - X\beta) \]
Accuracy of Estimates

\[ \text{Var}(b) = C_{XX} \]

\[ V(\hat{u}) = G - C_{ZZ} \]

And \[ V(u - \hat{u}) = C_{ZZ} \]
Prediction Error variance

prob(TBV)

SEP

EBV
Hypothesis Testing in Mixed Models

- Not well defined
- Can
  - Ignore random effects
  - Treat them as fixed
  - Estimate them