Mixed Models

- 1: Introduction to matrix algebra and calculus (W)
- 2: The General Linear Model (W)
- 3: Overview and Derivation of the mixed model (M)
- 4: Application: BLUP breeding values (M)
- 5: Application: Genomic selection (M)
- 6: Application: QTL/association mapping (W)
- 7: Application: G x E (W)
- 8: Application: BLUP maternal genetic (W)
- 9: Application: Associative effects (M)
- 10: Summary and wrap-up (W and M)

Overview And Introduction to Mixed Models

Linear Models
OLS and ML Estimators

Searle, S.R. 1971 Linear Models, Wiley

Schaefer, L.R., Linear Models and Computer Strategies in Animal Breeding

Lynch and Walsh Chapter 8
Linear vs non-linear

Linear
2nd order Polynomial

\[ Y_i = b_0 + b_1 X_{1i} + b_2 X_{1i}^2 + b_3 X_{2i} + b_4 X_{4i}^2 + b_5 (X_{1i}, X_{2i}) \cdot \varepsilon_i \]

Non-linear

\[ Y_i = b_0 e^{-b_1 X_i} \varepsilon_i \]

Log-linear

\[ \ln(Y_i) = \ln(b_0) - b_1 X_i + \ln(\varepsilon_i) \]

Why Linear

Taylor Expansion

\[ Y = f(X) \]

\[ Y \approx f(a) + \frac{f'(a)(X-a)}{1!} + \frac{f''(a)(X-a)^2}{2!} + \ldots + \frac{f^n(a)(X-a)^n}{n!} \]

\[ Y = e^{-X} \quad Y' = -e^{-X} \quad Y'' = e^{-X} \]

At \( a = 0 \)

\[ Y \approx 1 - X + \frac{X^2}{2!} - \frac{X^3}{3!} + \frac{X^4}{4!} + \ldots \]
Lower Order Terms Are more Important than higher

Works for other values of $a$ but not as exact, example $a=1$

\[ Y = e^{-X} \]
\[ Y \approx 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \frac{x^4}{4!} + \ldots \]

$x=1$

\[ Y \approx 1 \]

\[ Y = e^{-1} = .904837 \]

\[ Y \approx 1 - x = 1 - .1 = .9 \]

\[ Y \approx 1 - x + \frac{x^2}{2!} = 1 - .1 + \frac{.1^2}{2!} = .905 \]

\[ Y \approx 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} = 1 - .1 + \frac{.1^2}{2!} - \frac{.1^3}{3!} = .904833 \]

Generality

- Any underlying unknown function can be approximated by a polynomial equation (linear Model)
  - Lower order terms are more important than higher order
  - Model does not have any basis in biological function
  - Even highly non-linear systems can be approximated by a linear model with only lower order terms
  - Purely Descriptive
  - Allows tests of hypothesis related to treatment effects
  - Allows limited prediction (expansion is around a point)
Linear Model

- Can be used to approximate highly non-additive genetic systems, including dominance and epistasis
- Predictive ability is fairly good, even if underlying mode of gene action is non-additive
- Linear Models Extensively Used in Animal Breeding

One Random effect Linear Model

\[ Y_j = b_0 X_{0j} + b_1 X_{1j} + b_2 X_{1j}^2 + b_3 X_{3j} + b_4 X_{4j}^2 + \varepsilon_j \]
Matrix Notation

\[ Y_1 = b_0 X_{01} + b_1 X_{11} + b_2 X_{11}^2 + b_3 X_{31} + b_4 X_{41}^2 + \varepsilon_1 \]
\[ Y_2 = b_0 X_{02} + b_1 X_{12} + b_2 X_{12}^2 + b_3 X_{32} + b_4 X_{42}^2 + \varepsilon_2 \]
\[ \vdots \]
\[ Y_n = b_0 X_{0n} + b_1 X_{1n} + b_2 X_{1n}^2 + b_3 X_{3n} + b_4 X_{4n}^2 + \varepsilon_n \]

\[
\begin{bmatrix}
Y_1 \\
Y_2 \\
\vdots \\
Y_n
\end{bmatrix} = 
\begin{bmatrix}
X_{01} & X_{11} & X_{11}^2 & X_{21} & X_{21}^2 \\
X_{02} & X_{12} & X_{12}^2 & X_{22} & X_{22}^2 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
X_{0n} & X_{1n} & X_{1n}^2 & X_{2n} & X_{2n}^2
\end{bmatrix}
\begin{bmatrix}
b_0 \\
b_1 \\
b_2 \\
b_3 \\
b_4
\end{bmatrix} + 
\begin{bmatrix}
\varepsilon_1 \\
\varepsilon_2 \\
\vdots \\
\varepsilon_n
\end{bmatrix}
\]

\[ Y = XB + \varepsilon \]

Estimation

\[ Y = XB + \varepsilon \]

Ordinary Least Squares

- Independent variables (\(X\))
  - fixed
  - measured without error

- Residuals
  - Random
  - Independently and Identically Distributed (IID) with Mean 0 and variance \(\sigma^2\)
Independently and Identically Distributed with Mean 0 and variance $\sigma^2$

$$V(\varepsilon) = E\left[\varepsilon - E(\varepsilon)\right]^2$$

$$V(\varepsilon) = \begin{bmatrix}
\sigma^2 & 0 & 0 & 0 \\
0 & \sigma^2 & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \sigma^2
\end{bmatrix}$$

The error distribution from which each observation is sampled is the same

No Environmental Correlations

When would these assumptions be violated?

Ordinary Least Squares Estimator

$$\sum_{j=1}^{k} \varepsilon_j = \mathbf{e}' \varepsilon = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k]$$

$$E(Y_j) = \sum_{i=0}^{k} b_i X_{ij}$$

$$\varepsilon_j = Y_j - E(Y_j) = Y_j - \sum_{i=0}^{k} b_i X_{ij}$$

$$\varepsilon' \varepsilon = \sum_{j=1}^{n} \varepsilon_j^2 = \sum_{j=1}^{n} \left(Y_j - \sum_{i=0}^{k} b_i X_{ij}\right)^2$$

Find solutions such that the sum of the residuals squared is minimum
Least Square Estimators

\[ e'e = \sum_{j=1}^{n} e_j^2 = \sum_{j=1}^{n} \left( Y_j - \sum_{i=0}^{m} b_i X_{ij} \right)^2 \]

\[ \frac{\partial (e'e)}{\partial b_i} = 2 \sum_{j=1}^{n} \left( Y_j - \sum_{i=0}^{m} b_i X_{ij} \right) \left[ -X_{ij} \right] \]

Set to 0 for each i and solve system

Normal Equations

\[
\begin{bmatrix}
\sum x_{ij}^2 & \sum x_{ij}x_{i,j} & \cdots & \sum x_{ij}y_{ij} \\
\sum x_{ij}x_{i,j} & \sum x_{ij}^2 & \cdots & \sum x_{ij}x_{i,j} \\
\vdots & \vdots & \ddots & \vdots \\
\sum x_{ij}y_{ij} & \sum x_{ij}x_{i,j} & \cdots & \sum x_{ij}^2 \\
\end{bmatrix}
\begin{bmatrix}
\hat{b}_0 \\
\hat{b}_1 \\
\vdots \\
\hat{b}_k \\
\end{bmatrix}
= 
\begin{bmatrix}
\sum x_{ij}y_{ij} \\
\sum x_{ij}x_{i,j} \\
\vdots \\
\sum x_{ij}y_{ij} \\
\end{bmatrix}
\]

\[ X'XB = X'Y \]

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

\[ V(\hat{B}) = \sigma_e^2(X'X)^{-1} \]
Prediction

\[ \hat{Y} = X \hat{B} \]
\[ \hat{B} = \left( X'X \right)^{-1} \left( X'Y \right) \]
\[ V(\hat{Y}) = V(X\hat{B}) \]
\[ V(\hat{B}) = \sigma^2 \left( X'X \right)^{-1} \]
\[ V(\hat{Y}) = XV(\hat{B})X' \]
\[ V(\hat{Y}) = X(X'X)^{-1}X' \sigma^2 \]

Example Factor Affecting Fatty Acid
From Gill, J. Design and Analysis of experiments

<table>
<thead>
<tr>
<th>Fatty Acid</th>
<th>Amount over Weight (Kg)</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>6</td>
<td>28</td>
</tr>
<tr>
<td>20</td>
<td>12</td>
<td>40</td>
</tr>
<tr>
<td>17</td>
<td>10</td>
<td>32</td>
</tr>
<tr>
<td>12</td>
<td>8</td>
<td>36</td>
</tr>
<tr>
<td>11</td>
<td>9</td>
<td>34</td>
</tr>
</tbody>
</table>
OLS by IML

- *data from Gill, Design and Analysis of experiments;*
- *Demonstrate OLS estimation via matrix methods;*

```
proc iml;
start main;
y={ 10,
    20,
    17,
    12,
    11};
x={ 1 6 28,
     1 12 40,
     1 10 32,
     1 8 36,
     1 9 34};
b=inv(x`*x)*x`*y;
Yhat=x*b;
e=y-x*b;
sse=e`*e;
df=1/2;
ms=sse#df;
vb=inv(x`*x)#ms;
Vyhat=x*inv(x`*x)*x`#ms;
print b vb y yhat e sse ms vyhat;
finish main;
run;quit;
```

BY GLM

- **data** one;
- input fatty_acid over_wt age;
- cards;
  - 10 6 28
  - 20 12 40
  - 11 10 32
  - 12 8 36
  - 11 9 34
- proc glm;
- model fatty_acid=over_wt age /
  solution;
- run;
- quit;
- Compare results from IML to GLM
Generalized Least Squares (GLS)

- Ordinary Least Squares
  - Independent variables
    - fixed
    - measured without error
  - Residuals
    - Random
    - Independently and Identically Distributed (IID) with Mean 0 and variance $\sigma^2$

- Generalized Least Squares
  - Independent variables
    - fixed
    - measured without error
  - Residuals
    - Random
  \[ V(\varepsilon) = V \]

GLS

Minimize \[ (y - Xb)' V^{-1} (y - Xb) \]

Weighting by the inverse of the variance

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

If

\[ V = I \sigma^2 \]

\[ \hat{b} = (X' X)^{-1} (X' y) \]
Maximum Likelihood (ML) Solution to Same Problem

- Generalized Least Squares
  - Independent variables
    - fixed
    - measured without error
  - Residuals
    - Random
  \[ V(\varepsilon) = V \]

- Maximum Likelihood
  - Independent variables
    - fixed
    - measured without error
  - Residuals
    - Random
  \[ V(\varepsilon) = V \]
  \[ \varepsilon \approx N(0, V) \]

ML

\[ L = \frac{1}{(2\pi)^{N/2} \lVert V \rVert_2^{1/2}} e^{-\frac{1}{2}(y-Xb)^\top V^{-1} (y-Xb)} \]

Maximize w.r.t \( b \)

\[ \frac{\partial (\ln L)}{\partial b} = 0 \]

\[ \ln L = \ln(C) - \frac{1}{2}(y-Xb)^\top V^{-1} (y-Xb) \]
\[
\frac{\partial (\ln L)}{\partial b} = -\frac{1}{2} (y - Xb)'V^{-1}(-X) - \frac{1}{2} (-X)'V^{-1}(y - Xb)
\]

\[
\frac{\partial (\ln L)}{\partial b} = (y - Xb)'V^{-1}X
\]

\[
(y - Xb)'V^{-1}X = 0
\]

\[
(y' - (Xb)'V^{-1}X = 0'
\]

\[
(y' - b'X)'V^{-1}X = 0
\]

\[
b'X'V^{-1}X = y'V^{-1}X
\]

Same as GLS

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

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

**Variance of b**

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

Note if \( V = I\sigma^2_e \)

\[
V(b) = \sigma^2_e(X'X)^{-1}
\]