

Overview And Introduction to Mixed Models

- References
 - 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

$$Y_i = b_0 + b_1 X_{1i} + b_2 X_{1i}^2 + b_3 X_{3i} + b_4 X_{4i}^2 + b_5 (X_{1i} X_{2i}) + \varepsilon_i$$

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: Life is Non-Linear

Taylor Expansion

$$Y = f(X)$$

$$Y \approx f(a) + \frac{f'(a)(x-a)}{1!} + \frac{f''(a)(x-a)^2}{2!} + \dots + \frac{f^n(a)(x-a)^n}{n!}$$

$$Y = e^{-X} \quad Y' = -e^{-X} \quad Y'' = e^{-X}$$

At a=0

$$Y \approx 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \frac{x^4}{4!} + \dots$$

Lower Order Terms Are more Important than higher

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

$$Y = e^{-X} \quad Y \approx 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \frac{x^4}{4!} + \dots$$

X=.1

$$Y \approx 1$$

$$Y = e^{-.1} = .904837$$

$$Y \approx 1 - x = 1 - .1 = .9$$

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

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

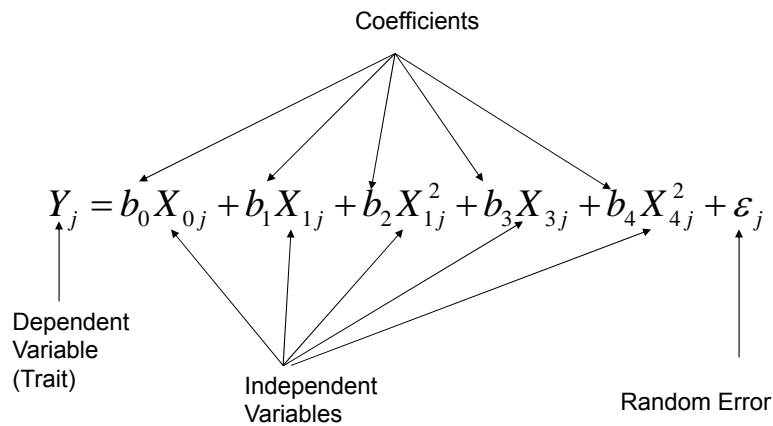
Generality

- 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



Matrix Notation

$$Y_1 = b_0 X_{01} + b_1 X_{11} + b_2 X_{11}^2 + b_3 X_{31} + b_4 X_{41}^2 + \epsilon_1$$

$$Y_2 = b_0 X_{02} + b_1 X_{12} + b_2 X_{12}^2 + b_3 X_{32} + b_4 X_{42}^2 + \epsilon_2$$

⋮

$$Y_n = b_0 X_{0n} + b_1 X_{1n} + b_2 X_{1n}^2 + b_3 X_{3n} + b_4 X_{4n}^2 + \epsilon_n$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} X_{01} & X_{11} & X_{11}^2 & X_{31} & X_{41}^2 \\ X_{02} & X_{12} & X_{12}^2 & X_{32} & X_{42}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{0n} & X_{1n} & X_{1n}^2 & X_{3n} & X_{4n}^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

$$\mathbf{Y} = \mathbf{XB} + \boldsymbol{\epsilon}$$

Estimation

$$\mathbf{Y} = \mathbf{XB} + \boldsymbol{\varepsilon}$$

Ordinary Least Squares

- Independent variables (X)
 - fixed
 - measured without error
- Residuals
 - Random
 - Independently and Identically Distributed (IID) with Mean 0 and variance σ^2

Independently and Identically Distributed with Mean 0 and variance σ^2

$$V(\boldsymbol{\varepsilon}) = E[\boldsymbol{\varepsilon} - E(\boldsymbol{\varepsilon})]^2$$

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

$$V(\boldsymbol{\varepsilon}) = \mathbf{I}\sigma_e^2$$

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

$$\sum_{j=1}^n \varepsilon_j^2 = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = [\varepsilon_1 \quad \varepsilon_2 \quad \cdots \quad \varepsilon_n] \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad \varepsilon_j = Y_j - E(Y_j)$$

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

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

Find Solutions such that the sum of the residuals squared is minimum

Least Square Estimators

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

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

Set=0 for each i and solve system

Normal Equations

$$\begin{bmatrix} \sum x_{0j}^2 & \sum x_{0j}x_{1j} & \cdots & \sum x_{0j}x_{kj} \\ \sum x_{0j}x_{1j} & \sum x_{1j}^2 & \cdots & \sum x_{1j}x_{kj} \\ \vdots & \vdots & \ddots & \vdots \\ \sum x_{0j}x_{kj} & \sum x_{1j}x_{kj} & \cdots & \sum x_{kj}^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{bmatrix} = \begin{bmatrix} \sum x_{0j}y_j \\ \sum x_{1j}y_j \\ \vdots \\ \sum x_{kj}y_j \end{bmatrix}$$

$$\mathbf{X}'\mathbf{X}\mathbf{B} = \mathbf{X}'\mathbf{Y}$$

$$\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y})$$

$$V(\hat{\mathbf{B}}) = \sigma_e^2(\mathbf{X}'\mathbf{X})^{-1}$$

Prediction

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\mathbf{B}} \qquad \hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y})$$

$$V(\hat{\mathbf{Y}}) = V(\mathbf{X}\hat{\mathbf{B}}) \qquad V(\hat{\mathbf{B}}) = \sigma_e^2(\mathbf{X}'\mathbf{X})^{-1}$$

$$V(\hat{\mathbf{Y}}) = \mathbf{X}V(\hat{\mathbf{B}})\mathbf{X}'$$

$$V(\hat{\mathbf{Y}}) = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma_e^2$$

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

Fatty Acid	Amount over Weight (Kg)	Age
10	6	28
20	12	40
17	10	32
12	8	36
11	9	34

OLS by IML

```

• *data from Gill, Design and Analysis of experiments;
• *Demonstrates 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;
rms=sse#df;
vb=inv(x`*x)#rms;
Vyhat=x*inv(x`*x)*x`#rms;
print b vb y yhat e sse rms vyhat;
finish main;
run;quit;

```


BY GLM

- **data** one;
 - input fatty_acid over_wt age;
 - cards;
 - 10 6 28
 - 20 12 40
 - 17 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 σ^2
 - **Generalized Least Squares**
 - Independent variables
 - fixed
 - measured without error
 - Residuals
 - Random
- $$V(\boldsymbol{\varepsilon}) = \mathbf{V}$$

GLS

Minimize
weighted SS

$$(\mathbf{y} - \mathbf{X}\mathbf{b})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\mathbf{b})$$

Weighting by the inverse of the variance

$$\hat{\mathbf{b}} = (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} (\mathbf{X}' \mathbf{V}^{-1} \mathbf{y})$$

if

$$\mathbf{V} = \mathbf{I} \sigma_e^2$$

$$\hat{\mathbf{b}} = (\mathbf{X}' \mathbf{X})^{-1} (\mathbf{X}' \mathbf{y})$$

Maximum Likelihood (ML) Solution to Same Problem

- Generalized Least Squares

- Independent variables

- fixed
- measured without error

- Residuals

- Random

$$V(\boldsymbol{\varepsilon}) = \mathbf{V}$$

- Maximum Likelihood

- Independent variables

- fixed
- measured without error

- Residuals

- Random

$$V(\boldsymbol{\varepsilon}) = \mathbf{V}$$

$$\boldsymbol{\varepsilon} \approx N(\mathbf{0}, \mathbf{V})$$

ML

$$L = \frac{1}{(2\pi)^{\frac{N}{2}} |\mathbf{V}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{y}-\mathbf{Xb})' \mathbf{V}^{-1}(\mathbf{y}-\mathbf{Xb})}$$

Maximize w.r.t \mathbf{b} $\frac{\partial(\ln L)}{\partial \mathbf{b}} = 0$

$$\ln L = \ln(C) - \frac{1}{2}(\mathbf{y} - \mathbf{Xb})' \mathbf{V}^{-1}(\mathbf{y} - \mathbf{Xb})$$

$$\frac{\partial(\ln L)}{\partial \mathbf{b}} = -\frac{1}{2}(\mathbf{y} - \mathbf{Xb})' \mathbf{V}^{-1}(-\mathbf{X}) - \frac{1}{2}(-\mathbf{X})' \mathbf{V}^{-1}(\mathbf{y} - \mathbf{Xb})$$

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

$$(\mathbf{y} - \mathbf{Xb})' \mathbf{V}^{-1} \mathbf{X} = 0$$

$$(\mathbf{y}' - (\mathbf{Xb})') \mathbf{V}^{-1} \mathbf{X} = 0'$$

$$(\mathbf{y}' - \mathbf{b}' \mathbf{X}') \mathbf{V}^{-1} \mathbf{X} = 0$$

$$\mathbf{b}' \mathbf{X}' \mathbf{V}^{-1} \mathbf{X} = \mathbf{y}' \mathbf{V}^{-1} \mathbf{X}$$

Same as GLS

$$\hat{\mathbf{b}}' = (\mathbf{y}' \mathbf{V}^{-1} \mathbf{X}) (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1}$$

$$\hat{\mathbf{b}} = (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} (\mathbf{X}' \mathbf{V}^{-1} \mathbf{y})$$

Variance of \mathbf{b}

$$V(\mathbf{b}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$$

Note if $\mathbf{V} = \mathbf{I}\sigma_e^2$

$$V(\mathbf{b}) = \sigma_e^2(\mathbf{X}'\mathbf{X})^{-1}$$

BLUP Best Linear Unbiased Prediction-Estimation

- References
- Searle, S.R. 1971 Linear Models, Wiley
- Schaefer, L.R., Linear Models and Computer Strategies in Animal Breeding
- Lynch and Walsh Chapter 26

Software

Ignacy Misztal UGA

- [Overview](#)
 - <http://nce.ads.uga.edu/~ignacy/newprograms.html>
- [General Documents](#)
 - <http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=lupf90.pdf>
 - <http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=emlf90.pdf>
- **Programs**
 - Binaries (UNIX, Windows, Max)
 - <http://nce.ads.uga.edu/html/projects/programs/>

OLS Independently and Identically Distributed Errors with Mean 0 and variance σ^2

$$V(\boldsymbol{\varepsilon}) = E[\boldsymbol{\varepsilon} - E(\boldsymbol{\varepsilon})]^2$$

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

$$V(\boldsymbol{\varepsilon}) = \mathbf{I}\sigma_e^2$$

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

For Some Traits Mean and Variance Correlated

Individuals Reared in Same Pen or Cage Cause these to be nonzero

Residuals independent

Solutions

- GLS
 - Fixes problem with changing variances and correlations in the data
- What about fixed effects?
 - How does one correct for
 - Environmental trend without a control
 - Herd effects
 - Year effects
 - Hatch effects
 - Confounding

Confounding of data

- Herd effects
 - Balanced design no problem
 - Require sample of every family in every herd
 - Old solution was within herd deviations
 - What if better herds have better genetics
- Fixed effects must be adjusted for genetic differences
- Random effects must be adjusted for fixed effects

Mixed Model

Simultaneous Adjustment of Fixed and Random effects

- Separates Independent variable into those that are

- Fixed **Xb**
- Random **Zu**

X=value of each fixed effect
b=linear regression coefficients
Z=incidence matrix of random effect, usually a 1 corresponding to each animal
u=estimate of random effects (breeding value)

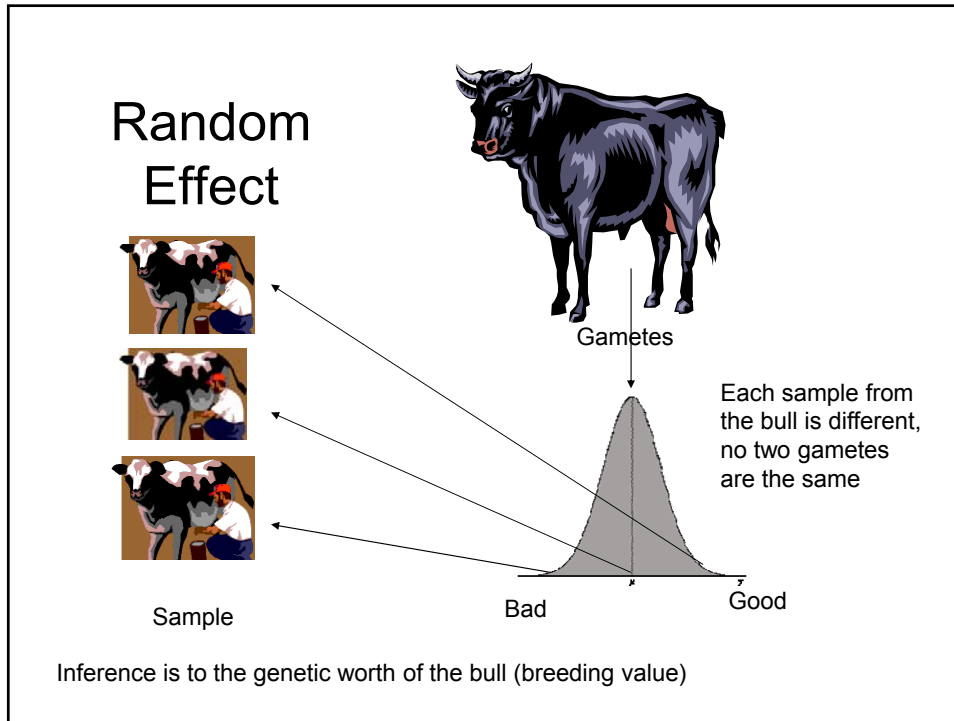
$$Y = Xb + Zu + e$$

In simple case $y_i = b_0 + g_i + e_i$

More importantly model's the variance structure

Fixed and Random Effects

- Fixed Effect
 - Inference Space only to those levels
 - Age, Hatch, Location, Parity, and Sex effects
- Random Effect
 - Effect Sampled From a Distribution of Effects
 - Inference Space To The Population From Which The Random Effect Was Sampled
 - If a new sample of observations were made (a new experiment), and the levels were completely different between the two samples, then the factors is usually random



Variances In Mixed Models

$$\begin{aligned}
 \mathbf{Y} &= \mathbf{Xb} + \mathbf{Zu} + \mathbf{e} & V(\mathbf{b}) &= \mathbf{0} \\
 & & V(\mathbf{u}) &= E(\mathbf{uu}') = \mathbf{G} \\
 & & V(\mathbf{e}) &= E(\mathbf{ee}') = \mathbf{R} \\
 V(\mathbf{Y}) &= V(\mathbf{Xb} + \mathbf{Zu} + \mathbf{e}) = \mathbf{ZGZ}' + \mathbf{R}
 \end{aligned}$$

Estimate the breeding values " \mathbf{u} " and fixed effects simultaneously

Old concept was to first adjust for the fixed effects, output the residuals and estimate the random effects

Resulted in Biased Estimates of Both Fixed and Random Effects

ML Derivation of Solutions

Joint density of \mathbf{y} and \mathbf{u} $f(\mathbf{y}, \mathbf{u}) = g(\mathbf{y}/\mathbf{u})h(\mathbf{u})$

$$g(\mathbf{y}/\mathbf{u}) = g(\mathbf{e})$$

$$g(\mathbf{e}) = \frac{1}{(2\pi)^{\frac{1}{2}N} V(\mathbf{e})^{\frac{1}{2}}} e^{-\frac{1}{2}\mathbf{e}'V(\mathbf{e})^{-1}\mathbf{e}} \quad h(\mathbf{u}) = \frac{1}{(2\pi)^{\frac{1}{2}N} V(\mathbf{u})^{\frac{1}{2}}} e^{-\frac{1}{2}\mathbf{u}'V(\mathbf{u})^{-1}\mathbf{u}}$$

$$f(\mathbf{y}, \mathbf{u}) = \frac{1}{(2\pi)^{\frac{1}{2}N} v(\mathbf{e})^{\frac{1}{2}}} e^{-\frac{1}{2}\mathbf{e}'\mathbf{R}^{-1}\mathbf{e}} \frac{1}{(2\pi)^{\frac{1}{2}N} v(\mathbf{u})^{\frac{1}{2}}} e^{-\frac{1}{2}\mathbf{u}'\mathbf{G}^{-1}\mathbf{u}}$$

$$f(\mathbf{y}, \mathbf{u}) = c_1 e^{-\frac{1}{2}\mathbf{e}'\mathbf{R}^{-1}\mathbf{e}} c_2 e^{-\frac{1}{2}\mathbf{u}'\mathbf{G}^{-1}\mathbf{u}}$$

$$f(\mathbf{y}, \mathbf{u}) = c e^{-\frac{1}{2}\mathbf{e}'\mathbf{R}^{-1}\mathbf{e}} e^{-\frac{1}{2}\mathbf{u}'\mathbf{G}^{-1}\mathbf{u}}$$

$$f(\mathbf{y}, \mathbf{u}) = L = c e^{-\frac{1}{2}\mathbf{e}'\mathbf{R}^{-1}\mathbf{e}} e^{-\frac{1}{2}\mathbf{u}'\mathbf{G}^{-1}\mathbf{u}}$$

Maximize w.r.t \mathbf{b} and \mathbf{u}

$$\ln(L) = \ln(c) - \frac{1}{2}\mathbf{e}'\mathbf{R}^{-1}\mathbf{e} - \frac{1}{2}\mathbf{u}'\mathbf{G}^{-1}\mathbf{u}$$

$$\mathbf{e} = \mathbf{Y} - \mathbf{X}\mathbf{b} - \mathbf{Z}\mathbf{u}$$

$$\ln(L) = \ln(c) - \frac{1}{2}(\mathbf{Y} - \mathbf{X}\mathbf{b} - \mathbf{Z}\mathbf{u})'\mathbf{R}^{-1}(\mathbf{Y} - \mathbf{X}\mathbf{b} - \mathbf{Z}\mathbf{u}) - \frac{1}{2}\mathbf{u}'\mathbf{G}^{-1}\mathbf{u}$$

SIMPLIFY FIRST THEN TAKE DERIVATIVES

$$\begin{aligned}
 & (\mathbf{Y} - \mathbf{Xb} - \mathbf{Zu})' \mathbf{R}^{-1} (\mathbf{Y} - \mathbf{Xb} - \mathbf{Zu}) + \mathbf{u}' \mathbf{G}^{-1} \mathbf{u} \\
 & = [\mathbf{Y}' - (\mathbf{Xb})' - (\mathbf{Zu})'] \mathbf{R}^{-1} (\mathbf{Y} - \mathbf{Xb} - \mathbf{Zu}) + \mathbf{u}' \mathbf{G}^{-1} \mathbf{u} \\
 & = \mathbf{Y}' \mathbf{R}^{-1} \mathbf{Y} - \mathbf{Y}' \mathbf{R}^{-1} \mathbf{Xb} - \mathbf{Y}' \mathbf{R}^{-1} \mathbf{Zu} \\
 & \quad - (\mathbf{Xb})' \mathbf{R}^{-1} \mathbf{Y} + (\mathbf{Xb})' \mathbf{R}^{-1} \mathbf{Xb} + (\mathbf{Xb})' \mathbf{R}^{-1} \mathbf{Zu} \\
 & \quad - (\mathbf{Zu})' \mathbf{R}^{-1} \mathbf{Y} + (\mathbf{Zu})' \mathbf{R}^{-1} \mathbf{Xb} + (\mathbf{Zu})' \mathbf{R}^{-1} \mathbf{Zu} + \mathbf{u}' \mathbf{G}^{-1} \mathbf{u}
 \end{aligned}$$

$$\frac{\partial(\ln L)}{\partial \mathbf{b}} = 0 \quad \begin{aligned} & -\mathbf{Y}' \mathbf{R}^{-1} \mathbf{X} - \mathbf{X}' \mathbf{R}^{-1} \mathbf{Y} + (\mathbf{Xb})' \mathbf{R}^{-1} \mathbf{X} + \\ & \mathbf{X}' \mathbf{R}^{-1} \mathbf{Xb} + \mathbf{X}' \mathbf{R}^{-1} \mathbf{Zu} + (\mathbf{Zu})' \mathbf{R}^{-1} \mathbf{X} = 0 \end{aligned}$$

$$\begin{aligned}
 & -2\mathbf{X}' \mathbf{R}^{-1} \mathbf{Y} + 2\mathbf{X}' \mathbf{R}^{-1} \mathbf{Xb} + 2\mathbf{X}' \mathbf{R}^{-1} \mathbf{Zu} = 0 \\
 & \mathbf{X}' \mathbf{R}^{-1} \mathbf{Xb} + \mathbf{X}' \mathbf{R}^{-1} \mathbf{Zu} = \mathbf{X}' \mathbf{R}^{-1} \mathbf{Y}
 \end{aligned}$$

Take Derivative w.r.t \mathbf{u}

$$\begin{aligned}
 & (\mathbf{Y} - \mathbf{Xb} - \mathbf{Zu})' \mathbf{R}^{-1} (\mathbf{Y} - \mathbf{Xb} - \mathbf{Zu}) + \mathbf{u}' \mathbf{G}^{-1} \mathbf{u} \\
 & = \mathbf{Y}' \mathbf{R}^{-1} \mathbf{Y} - \mathbf{Y}' \mathbf{R}^{-1} \mathbf{Xb} - \mathbf{Y}' \mathbf{R}^{-1} \mathbf{Zu} \\
 & \quad - (\mathbf{Xb})' \mathbf{R}^{-1} \mathbf{Y} + (\mathbf{Xb})' \mathbf{R}^{-1} \mathbf{Xb} + (\mathbf{Xb})' \mathbf{R}^{-1} \mathbf{Zu} \\
 & \quad - (\mathbf{Zu})' \mathbf{R}^{-1} \mathbf{Y} + (\mathbf{Zu})' \mathbf{R}^{-1} \mathbf{Xb} + (\mathbf{Zu})' \mathbf{R}^{-1} \mathbf{Zu} + \mathbf{u}' \mathbf{G}^{-1} \mathbf{u}
 \end{aligned}$$

$$\frac{\partial(\ln L)}{\partial \mathbf{u}} = 0 \quad \begin{aligned} & -\mathbf{Y}' \mathbf{R}^{-1} \mathbf{Z} + (\mathbf{Xb})' \mathbf{R}^{-1} \mathbf{Z} - (\mathbf{Zu})' \mathbf{R}^{-1} \mathbf{Y} + (\mathbf{Zu})' \mathbf{R}^{-1} \mathbf{Xb} \\ & + (\mathbf{Zu})' \mathbf{R}^{-1} \mathbf{Zu} + (\mathbf{Zu})' \mathbf{R}^{-1} \mathbf{Z} + 2\mathbf{G}^{-1} \mathbf{u} = 0 \end{aligned}$$

$$\begin{aligned}
 & -2\mathbf{Y}' \mathbf{R}^{-1} \mathbf{Z} + 2\mathbf{Z}' \mathbf{R}^{-1} \mathbf{Xb} + 2\mathbf{Z}' \mathbf{R}^{-1} \mathbf{Zu} + 2\mathbf{G}^{-1} \mathbf{u} = 0 \\
 & \mathbf{Z}' \mathbf{R}^{-1} \mathbf{Xb} + \mathbf{Z}' \mathbf{R}^{-1} \mathbf{Zu} + \mathbf{G}^{-1} \mathbf{u} = \mathbf{Z}' \mathbf{R}^{-1} \mathbf{Y}
 \end{aligned}$$

Mixed Model Equations

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\mathbf{b} + \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\mathbf{u} = \mathbf{X}'\mathbf{R}^{-1}\mathbf{Y}$$

$$\mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}\mathbf{b} + \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}\mathbf{u} + \mathbf{G}^{-1}\mathbf{u} = \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Y}$$

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{Y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Y} \end{bmatrix}$$

Simplifications If $\mathbf{R} = \mathbf{I}\sigma_e^2$

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \sigma_e^2\mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{Y} \\ \mathbf{Z}'\mathbf{Y} \end{bmatrix}$$

Alternative derivations are possible that do not require Normal Dist'n Assumptions, resulting in these same solutions and are therefore also Best Linear Unbiased Predictors (BLUP)

BLUP Breeding Values

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \sigma_e^2\mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{Y} \\ \mathbf{Z}'\mathbf{Y} \end{bmatrix}$$

With Diploid Organisms and Assuming Additivity $\mathbf{G} = \mathbf{A}\sigma_a^2$ $\mathbf{G}^{-1} = \frac{1}{\sigma_a^2}\mathbf{A}^{-1}$

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \frac{\sigma_e^2}{\sigma_a^2}\mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{Y} \\ \mathbf{Z}'\mathbf{Y} \end{bmatrix}$$

Only Estimate of Ratio is Needed

Only inverse is needed

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \frac{\sigma_e^2}{\sigma_a^2} \mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{Y} \\ \mathbf{Z}'\mathbf{Y} \end{bmatrix}$$

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \quad \mathbf{X}'\mathbf{Z} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

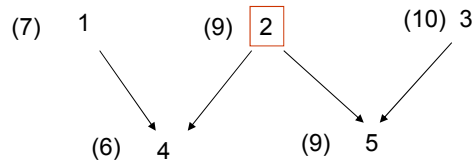
$\mathbf{X}'\mathbf{X} = 5$
 $\mathbf{X}'\mathbf{Z} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \end{bmatrix}$

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \frac{\sigma_e^2}{\sigma_a^2} \mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{Y} \\ \mathbf{Z}'\mathbf{Y} \end{bmatrix}$$

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

$$\mathbf{Z}'\mathbf{X} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

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



$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 1 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 1 & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 & 1 & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{4} & 1 \end{bmatrix}$$

Assume heritability=.5

$$\frac{\sigma_e^2}{\sigma_a^2} = 1 \quad \mathbf{Z}'\mathbf{Z} + \frac{\sigma_e^2}{\sigma_a^2} \mathbf{A}^{-1} = \begin{bmatrix} \frac{5}{2} & \frac{1}{2} & 0 & -1 & 0 \\ \frac{1}{2} & 3 & \frac{1}{2} & -1 & -1 \\ 0 & \frac{1}{2} & \frac{5}{2} & 0 & -1 \\ -1 & -1 & 0 & 3 & 0 \\ 0 & -1 & -1 & 0 & 3 \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \frac{\sigma_e^2}{\sigma_a^2} \mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{Y} \\ \mathbf{Z}'\mathbf{Y} \end{bmatrix}$$

$$\mathbf{X}'\mathbf{Y} = [1 \quad 1 \quad 1 \quad 1 \quad 1] \begin{bmatrix} 7 \\ 9 \\ 10 \\ 6 \\ 9 \end{bmatrix}$$

$$\mathbf{X}'\mathbf{Y} = [41]$$

$$\mathbf{Y}'\mathbf{Z} = \begin{bmatrix} 7 \\ 9 \\ 10 \\ 6 \\ 9 \end{bmatrix}$$

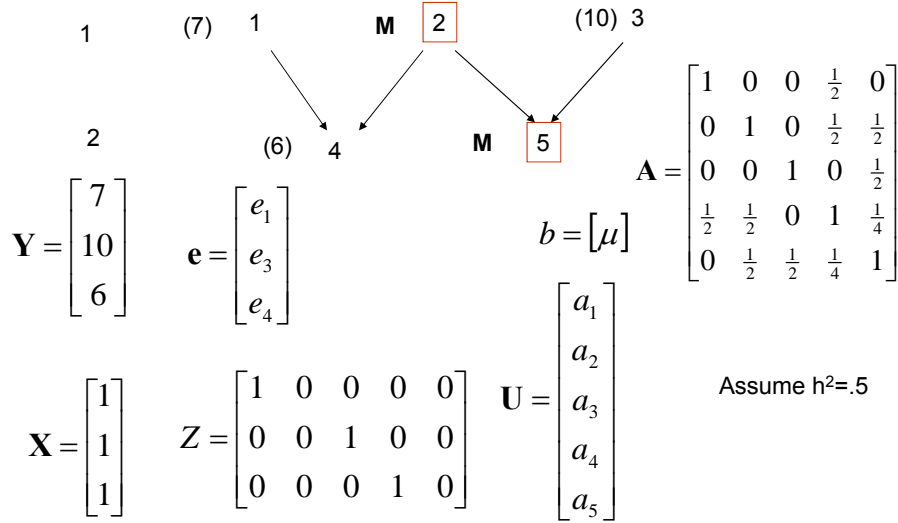
MME

$$\begin{bmatrix} 5 & 1 & 1 & 1 & 1 & 1 \\ 1 & \frac{5}{2} & \frac{1}{2} & 0 & -1 & 0 \\ 1 & \frac{1}{2} & 3 & \frac{1}{2} & -1 & -1 \\ 1 & 0 & \frac{1}{2} & \frac{5}{2} & 0 & -1 \\ 1 & -1 & -1 & 0 & 3 & 0 \\ 1 & 0 & -1 & -1 & 0 & 3 \end{bmatrix} \begin{bmatrix} \mu \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix} = \begin{bmatrix} 41 \\ 7 \\ 9 \\ 10 \\ 6 \\ 9 \end{bmatrix}$$

Y		Z		X		X	
7	1	0	0	0	0	1	
9	0	1	0	0	0	1	
10	0	0	1	0	0	1	
6	0	0	0	1	0	1	
9	0	0	0	0	1	1	
XZ=							1 1 1 1 1
1							sig(a)= 1
1							sig(e)= 0.5
							lam= 0.5
							A
ZX=	1	0	0	0.5	0	1.5 0.5 0 -1 0	
1	0	1	0	0.5	0.5	0.5 2 0.5 -1 -1	
1	0	0	1	0	0.5	0 0.5 1.5 0 -1	
1	0.5	0.5	0	1	0.25	-1 -1 0 2 0	
XY=	41	0	0.5	0.5	0.25	0 -1 -1 0 2	
ZY=	7	1.75	0.25	0	-0.5	0 0.75 0.25 0 -0.5 0	
9	0.25	2	0.25	-0.5	-0.5	0.25 1 0.25 -0.5 -0.5	
10	0	0.25	1.75	0	-0.5	0 0.25 0.75 0 -0.5	
6	-0.5	-0.5	0	2	0	-0.5 -0.5 0 1 0	
9	0	-0.5	-0.5	0	2	0 -0.5 -0.5 0 1	
Z'Z+INV(A)							INV(A)*Lam
5	1	1	1	1	1	41	
1	1.75	0.25	0	-0.5	0	u1 7	
1	0.25	2	0.25	-0.5	-0.5	u2 9	
1	0	0.25	1.75	0	-0.5	u3 10	
1	-0.5	-0.5	0	2	0	u4 6	
1	0	-0.5	-0.5	0	2	u5 9	
LHS						RHS	
b	0.918367	-0.65306	-0.69388	-0.65306	-0.79592	-0.79592	41
u1	-0.65306	1.083203	0.44898	0.467818	0.709576	0.55573	7
u2	-0.69388	0.44898	1.102041	0.44898	0.734694	0.734694	9
u3	-0.65306	0.467818	0.44898	1.083203	0.55573	0.709576	10
u4	-0.79592	0.709576	0.734694	0.55573	1.259027	0.720565	6
u5	-0.79592	0.55573	0.734694	0.709576	0.720565	1.259027	9
INV(LHS)						RHS	
mean + bv							
b	8.367347						
u1	-1.21507						7.152276
u2	0.122449						8.489796
u3	1.092622						9.459969
u4	-1.45683						6.910518
u5	0.620094						8.987441

Missing Values (Sex Limited Traits)

Generation



Y				Z			X		X	
7		1	0	0	0	0	1		1	1
10		0	0	1	0	0	1			
6		0	0	0	1	0	1		XX=	3
XZ=	1	0	1	1	0				XY=	23
	1			1	0	0	sig(a)=	1		
	0			0	0	0	sig(e)=	1	1.5	0.5
	1		Z=	0	1	0	lam=	1	0.5	2
	1			0	0	1			0	0.5
	0			0	0	0			-1	-1
ZX=									0	1.5
									-1	0
									0	2
									0	-1
									-1	0
									0	2
ZY=	7								1.5	0.5
	0			1	0	0			0.5	2
	10			0	1	0			0	0.5
	6			0	0	1			-1	-1
	0			0.5	0.5	0			0	1.5
				0	0.5	0.5			-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0
									0	2
									0	-1
									-1	0

Variance of the Estimates

$$\begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \frac{\sigma_e^2}{\sigma_a^2} \mathbf{A}^{-1} \end{bmatrix}^{-1}$$

$$V(\hat{\mathbf{b}}) = \mathbf{C}_{11} \sigma_e^2$$

$$V(\hat{\mathbf{u}} - \mathbf{u}) = \mathbf{C}_{22} \sigma_e^2 \quad \text{Prediction Error Variance}$$

$$V(\hat{\mathbf{u}}) = \mathbf{A} \sigma_a^2 + \mathbf{C}_{22} \sigma_e^2 \quad \begin{array}{l} \text{Prediction Error Variance} \\ \text{Including Drift Variance} \end{array}$$

Kennedy and Sorensen *Quantitative Genetics*

Run "BLUP by proc iml.sas"

PEV

1.12236	0.29509	0.32030	0.65827	0.39093
0.29509	1.14758	0.29509	0.68854	0.68854
0.32030	0.29509	1.12236	0.39093	0.65827
0.65827	0.68854	0.39093	1.2686	0.60026
0.39093	0.68854	0.65827	0.60026	1.2686

EV

2.86014	0.29509	0.32030	1.52716	0.39093
0.29509	2.88536	0.29509	1.5574	1.5574
0.32030	0.29509	2.86014	0.39093	1.52716
1.52716	1.5574	0.39093	3.00643	1.03471
0.39093	1.5574	1.52716	1.03471	3.00643

Extensions of Model

- Inclusion of Dominance and Epistasis
 - Dominance
 - Dominance effects are the result of interaction of alleles within a locus
 - Dominance relationship matrix needed
 - Reflects the probability that individuals have the same pair of alleles in common at a locus
 - Epistasis
 - Epistatic genetic effects are the result of interactions between alleles at different loci
 - Epistatic relationship matrix needed
 - Reflects the probability that individuals have the same pair of alleles in common at different loci (4 possible pairings of 2 alleles at 2 loci)
 - Useful in crossbreeding programs but generally not useful in pure breeding programs
 - An individual does not pass on dominance or epistatic effects (without inbreeding or cloning), which are a function of both parents
 - Exception is Additive x Additive epistasis is a function of 2 alleles at different loci in the same gamete, but dissipates with recombination and/or segregation

Estimation of Variances Using all Data in a Pedigree

- REML
 - EM-REML iterative process whereby
 - A value is assumed for additive variance
 - Estimates of breeding values found
 - Additive variance $V(A)$ is estimated as variance of breeding values $V(A) = (u'A^{-1}u + \text{stuff})/n$
 - The new value of $V(A)$ is substituted into the MME
 - Estimates of breeding values (u) are found
 - The process repeated until convergence
 - DF-REML work by trial and error finding a value of $V(A)$ that maximize the likelihood

Estimation of Effects and Parameters via Iteration and MCMC

Distributions	Estimates
$b_i \mathbf{b}_i, \mathbf{a}, \sigma_a^2, \sigma_\varepsilon^2, y \sim N\left(\hat{b}_i, \frac{\sigma_\varepsilon^2}{LHS_{ii}}\right)$	$\hat{S}_a^2 = (\mathbf{a}' \mathbf{A}^{-1} \mathbf{a} + v_a S_a^2) / \hat{v}_a$
$a_i \mathbf{b}, \mathbf{a}_i, \sigma_a^2, \sigma_\varepsilon^2, y \sim N\left(\hat{a}_i, \frac{\sigma_\varepsilon^2}{LHS_{ii}}\right)$	$\hat{v}_a = q + v_a$
$\sigma_a^2 \mathbf{b}, \mathbf{a}, \sigma_\varepsilon^2, y \sim \hat{v}_a \hat{S}_a^2 \chi_{\hat{v}_a}^{-2}$	$\hat{S}_\varepsilon^2 = (\boldsymbol{\varepsilon}' \boldsymbol{\varepsilon} + v_\varepsilon S_\varepsilon^2) / \hat{v}_\varepsilon$
$\sigma_\varepsilon^2 \mathbf{b}, \mathbf{a}, \sigma_a^2, y \sim \hat{v}_\varepsilon \hat{S}_\varepsilon^2 \chi_{\hat{v}_\varepsilon}^{-2}$	$\hat{v}_\varepsilon = N + v_\varepsilon$
	$\boldsymbol{\varepsilon} = \mathbf{Y} - \mathbf{X}\mathbf{B} - \mathbf{Z}\mathbf{a}$

1. Solutions to MME are found using iterative approach (Gauss-Seidel)
2. With each iteration a random amount is added to each solution based on the expected distribution
3. After processing all equations in the MME, new variances are computed and a random amount is added to each solution based on the expected distribution
4. After a burn in period, and many 1000 iterations, the average value of each parameter, with the empirical standard error is the best estimate of the effects and variances

LHS					RHS				
5	1	1	1	1	1	b0	238.2		
1	3	0.5	-0.5	-1	0	u1	38.5		
1	0.5	3.071429	-1	0.571429	-1.14286	u2	48.9		
1	-0.5	-1	3.5	-1	0	u3	64.3		
1	-1	0.571429	-1	3.571429	-1.14286	u4	50.5		
1	0	-1.14286	0	-1.14286	3.285714	u5	36		
					B	RND=random normal deviate (0,1)			
						1.953281			
inv(A)									
2	0.5	-0.5	-1	0	0	i=N+ve CHI= Random chi-square variate (idf)			
0.5	2.071429	-1	0.571429	-1.14286	0	105	102.7981		
-0.5	-1	2.5	-1	0	0	l=q+va			
-1	0.571429	-1	2.571429	-1.14286	0	105	99.71213		
0	-1.14286	0	-1.14286	2.285714	0				
					B (guess)	Hyperparameters			
					Prior Variance Guess va=' 100 S2e=' 10 ve=' 100 S2a=' 10 k=' S2e/S2a 1				
					a' -0.40127 1.38057 7.609528 37.85838 10.41943 Aa' a' -41.7754 -0.40127 4.775032 1.38057 -20.0145 7.609528 79.02284 37.85838 -21.0287 10.41943				
					2643.624 = 'a'inv(A)a a' -0.40127 1.38057 7.609528 37.85838 10.41943 Aa' a' -41.7754 -0.40127 4.775032 1.38057 -20.0145 7.609528 79.02284 37.85838 -21.0287 10.41943				
					Residuals' Square b0 hat 47.64 b0 new 50.40236 u1 hat -3.96745 u1 new -0.40127 -11.5011 132.2751 u2 hat -0.42382 u2 new 1.38057 -2.88293 8.311268 u3 hat 4.30788 u3 new 7.609528 6.288115 39.54039 u4 hat 34.58992 u4 new 37.85838 -37.7607 1425.873 u5 hat 7.011821 u5 new 10.41943 -24.8218 616.1211				
					Sum 2222.21 Va new= 36.54144 Va new=' (a'inv(A)a+vaS2a)/CHI (e'e+veS2e)/CHI				

Slide 53

WMM1 S_a is a prior guess about $\text{sig}(a)$
 V_a is the degree of belief in that prior

S_e is prior guess about $\text{sig}(e)$
 V_e is degrees of belief in that prior

q is number of random effects
 N is the number of phenotypes

William Muir, 5/20/2009

Genomic Selection

References

- Goddard (2008) Genomic selection: prediction of accuracy and maximization of long term response *Genetica* DOI 10.1007/s10709-008-9308-0
- NejatiJavaremi, A., C. Smith, and J. P. Gibson, 1997 Effect of total allelic relationship on accuracy of evaluation and response to selection. *Journal Of Animal Science* **75**: 1738-1745.
- VanRaden, P. M., 2008 Efficient Methods to Compute Genomic Predictions. *Journal Of Dairy Science* **91**: 4414-4423
- Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard, 2001 Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**: 1819-1829

Genomic Relationship Matrix

- Assumes
 - Dense markers evenly spaced across the genome
 - Assumes markers are in LD with QTL affecting trait(s) of interest
 - Alike in State (AIS) alleles were at one time a result of a single mutation, thus IBD when traced back in evolutionary time
 - Each marker accounts for an equal proportion of genetic variance (infinitesimal model)
 - Genetic Effects are Normally Distributed

Compute (AIS) relationship matrix (G)

$$TA_k = 2 \frac{\sum_{i=1}^2 \sum_{j=1}^2 I_{ij}}{4}$$

$$G = \sigma_{A^*}^2 G^*$$

$$\sigma_{A^*}^2$$

TA_k=total allelic relationship at kth locus
 TA_k=2x coefficient of relationship
 (Malecot. 1948)

Is the additive genetic variance associated with the markers for the trait

$$\sigma_{A^*}^2 < \sigma_A^2$$

$$G_{xy}^* = \frac{\sum_{k=1}^L TA_k}{L} = \frac{2 \sum_{k=1}^L \left(\frac{\sum_{i=1}^2 \sum_{j=1}^2 I_{ij}}{4} \right)}{L}$$

Note: with low marker density the markers may not capture any genetic variance

Individual	LOCUS										Pedigree							
	A		B		C		D		E									
1	2	2	1	1	1	2	1	1	2	2								
2	1	2	1	2	2	2	1	2	1	1								
3	1	2	1	1	1	2	1	2	1	2								
4	2	2	1	1	2	2	1	1	2	1								
5	2	1	1	2	2	2	1	1	2	1								
6	2	2	1	1	2	2	1	1	2	1								
Individuals (XY)											Total	relationship=axy						
x=1	2	2	1	1	1	2	1	1	2	2	9	1.8						
y=1	2	2	1	1	1	2	1	1	2	2								
sum	4		4		2		4		4									
shared alleles	2		2		1		2		2									
x=1	2	2	1	1	1	2	1	1	2	2	4	0.8						
y=2	2	2	1	1	1	2	1	1	2	2								
sum	2		2		2		2		0									
shared alleles	1		1		1		1		0									
AIS											G=GRM		IBD		PEDIGREE		A	
1	1.8	0.8	1.2	1.6	1.2	1.6	1	0	0.5	0.5	0.5	0.5	0.5	0.5				
2	0.8	1.4	1	1.2	1.2	1.2	0	1	0.5	0.5	0.5	0.5	0.5	0.5				
3	1.2	1.2	1.2	1.2	1	1.2	0.5	0.5	1	0.5	0.5	0.5	0.5	0.5				
4	1.6	1.2	1.2	1.8	1.4	1.8	0.5	0.5	0.5	1	0.5	0.5	0.5	0.5				
5	1.2	1.2	1	1.4	1.4	1.4	0.5	0.5	0.5	0.5	1	0.5	0.5	0.5				
6	1.6	1.2	1.2	1.8	1.4	1.8	0.5	0.5	0.5	0.5	0.5	0.5	1	1				
Parents assumed not related (False)											Parents assumed non inbred (false)		Full sibs assumed = relationship (false)					

G* Computed Directly from M

Individual	LOCUS										code	
	A		B		C		D		E			
1	2	2	1	1	1	2	1	1	2	2	22=2	1
2	1	2	1	2	2	2	1	2	1	1	12=1	0
3	1	2	1	1	1	2	1	2	1	2	11=0	-1
4	2	2	1	1	2	2	1	1	2	1		
5	2	1	1	2	2	2	1	1	2	1		
6	2	2	1	1	2	2	1	1	2	1		
	M N individuals x p markers					M' p markers x N individuals						
1	1	-1	0	-1	1	1	0	0	1	0	1	
2	0	0	1	0	-1	-1	0	-1	-1	0	-1	
3	0	-1	0	0	0	0	1	0	1	1	1	
4	1	-1	1	-1	0	-1	0	0	-1	-1	-1	
5	0	0	1	-1	0	1	-1	0	0	0	0	
6	1	-1	1	-1	0							
	0.8	-0.2	0.2	0.6	0.2	0.6	1.8	0.8	1.2	1.6	1.2	1.6
	-0.2	0.4	0	0.2	0.2	0.2	0.8	1.4	1	1.2	1.2	1.2
	0.2	0	0.2	0.2	0	0.2	1.2	1	1.2	1.2	1	1.2
	0.6	0.2	0.2	0.8	0.4	0.8	1.6	1.2	1.2	1.8	1.4	1.8
	0.2	0.2	0	0.4	0.4	0.4	1.2	1.2	1	1.4	1.4	1.4
	0.6	0.2	0.2	0.8	0.4	0.8	1.6	1.2	1.2	1.8	1.4	1.8
	MM'/5 +constant					=	G*					
	dimension nxn											

Coding Genomic Relationship Matrix

Genotype	012	101	centered
AA	0	-a _i	-2p _i a _i
Aa	a _i	0	(1-2p _i)a _i
aa	2a _i	a _i	2(1-p _i)a _i
Mean	2p _i a _i	a _i (1-2p _i)	0

Does it make a difference?

GBLUP, NO

ssGBLUP, Yes G matrix needs to be scaled the same as A matrix

Legarra, A., I. Aguilar, and I. Misztal, 2009 A relationship matrix including full pedigree and genomic information. *Journal Of Dairy Science* **92**: 4656-4663.

Chen, C., I. Misztal, I. Aguilar, A. Legarra, and W. Muir, 2011 Effect of different genomic relationship matrices on accuracy and scale. *Journal Of Animal Science* **89**: 2673-2679.

Mixed Model Equations

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{Y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Y} \end{bmatrix}$$

Simplifications If $\mathbf{R} = \mathbf{I}\sigma_e^2$

$$\mathbf{G} = \sigma_{A^*}^2 \mathbf{M}\mathbf{M}' / L$$

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \frac{\sigma_e^2}{\sigma_{A^*}^2} (\mathbf{M}\mathbf{M}' / L)^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{Y} \\ \mathbf{Z}'\mathbf{Y} \end{bmatrix}$$

Note one can multiply both sides of the second equation by \mathbf{G} to avoid an inverse

Dimension of Mixed Model Equations

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \frac{\sigma_e^2}{\sigma_{A^*}^2} (\mathbf{M}\mathbf{M}' / L)^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{Y} \\ \mathbf{Z}'\mathbf{Y} \end{bmatrix}$$

\mathbf{M} (n individuals x p markers)

$\mathbf{M}(n,p)\mathbf{M}'(p,n)$

$\mathbf{M}\mathbf{M}'(n,n)$

n effects to estimate

Equivalent Model Estimation of Marker effects

$$\begin{bmatrix} \mathbf{X}'_{1,N} \mathbf{X}_{N,1} & \mathbf{X}_{1,N}' \mathbf{M}_{N,p} \\ \mathbf{M}'_{p,N} \mathbf{X}_{N,1} & \mathbf{M}'_{p,N} \mathbf{M}_{N,p} + \frac{\sigma_e^2}{\sigma_g^2} \mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_{1,1} \\ \mathbf{g}_{p,1} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'_{1,N} \mathbf{Y}_{N,1} \\ \mathbf{M}'_{p,N} \mathbf{Y}_{N,1} \end{bmatrix}$$

Assumption depends on method

- 1) (GBLUP, ssGBLUP) Genetic variance associated with each marker is equal $\sigma_s^2 = \begin{pmatrix} \sigma_{A_s} \\ L \end{pmatrix}$
- 2) (Bayes A) sampled from a t distribution
- 3) (Bayes B and Bayes C π) from a mixture of distributions (null and t)

$$\hat{u}_i = GEBV_i = \mathbf{M}\mathbf{g} = \sum_j M_{ij} \hat{\delta}_j$$

M'						M						
1	0	0	1	0	1	1	-1	0	-1	1		
-1	0	-1	-1	0	-1	0	0	1	0	-1		
0	1	0	1	1	1	0	-1	0	0	0		
-1	0	0	-1	-1	-1	1	-1	1	-1	0		
1	-1	0	0	0	0	0	0	1	-1	0		
						1	-1	1	-1	0		
M'M						M'X		M'Y		Y X		
3	-3	2	-3	1		3		24		7	1	
-3	4	-2	3	-1		-4		-34		9	1	
2	-2	4	-3	-1		4		35		10	1	
-3	3	-3	4	-1		-4		-33		6	1	
1	-1	-1	-1	2		0		-2		9	1	
I						V(A)=		V(E)=		X'X		
1	0	0	0	0		5				6		
0	1	0	0	0		20						
0	0	1	0	0								
0	0	0	1	0								
0	0	0	0	1								
LHS						RHS		X'Y		52		
6	3	-4	4	-4	0	52						
3	23	-3	2	-3	1	24						
-4	-3	24	-2	3	-1	-34						
4	2	-2	24	-3	-1	35						
-4	-3	3	-3	24	-1	-33						
0	1	-1	-1	-1	22	-2						
inverse(LHS)						RHS		8.762384		-0.08491		
B	0.238268	-0.02055	0.030921	-0.03165	0.029414	52						
g1	-0.02055	0.046795	0.002074	-0.00012	0.002068	24						
g2	0.030921	0.002074	0.047116	-0.00139	-0.00057	-34						
g3	-0.03165	-0.00012	-0.00139	0.047031	0.00085	35						
g4	0.029414	0.002068	-0.00057	0.00085	0.047091	-33						
g5	0.002238	-0.00194	0.001958	0.002119	0.002059	-2						
M						g		-0.25929		0.094488		
u1	1	-1	0	-1	1	1						
u2	0	0	1	0	-1	0.021935						
u3	0	-1	0	0	0	0.012177						
u4	1	-1	1	-1	0	0.070134						
u5	0	0	1	-1	0	-0.08231						
u6	1	-1	1	-1	0							
same as before						-0.02194		-0.1648		-0.05796		
same as before						-0.1648		-0.1648		-0.1648		

Single Step ssGBLUP

- Merge G matrix into regular A matrix
 - Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta *et al.* 2010 Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal Of Dairy Science* **93**: 743-752.
 - Corrects for multi-trait selection bias
 - Vitezica, Z. G., I. Aguilar, I. Misztal, and A. Legarra, 2011 Bias in genomic predictions for populations under selection. *Genetics Research* **93**: 357-366.
 - Uses all information
 - Phenotypes of animals without genotypes
 - Chen, C.Y., I. Misztal, I. Aguilar, S. Tsuruta, T. H. E. Meuwissen, S. E. Aggrey, T. Wang and W. M. Muir. *Genome Wide Marker Assisted selection Combining All Pedigree* 2011. *Phenotypic Information with Genotypic data in One Step: An Example Using Broiler Chickens*. *J. Anim Sci.* 2011. 89:23-28. doi:10.2527/jas.2010-3071
- Software
 - <http://nce.ads.uga.edu/~ignacy/genomic-blupf90/>

ssGBLUP

GBLUP only animals genotyped included in analysis

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{Y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Y} \end{bmatrix}$$

ssGBLUP all animals with phenotypes included

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{H}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{Y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Y} \end{bmatrix}$$

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}$$

GWAS

GBLUP

$$\begin{bmatrix} \mathbf{X}'_{1,N} & \mathbf{X}_{N,1} & \mathbf{X}_{1,N}'\mathbf{M}_{N,p} \\ \mathbf{M}'_{p,N} & \mathbf{X}'_{N,1} & \mathbf{M}'_{p,N}\mathbf{M}_{N,p} + \frac{\sigma_e^2}{\sigma_g^2}\mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_{1,1} \\ \mathbf{g}_{p,1} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'_{1,N} & \mathbf{Y}_{N,1} \\ \mathbf{M}'_{p,N} & \mathbf{Y}_{N,1} \end{bmatrix}$$

SNP effects using only genotyped individuals

ssGBLUP

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{H}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{Y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Y} \end{bmatrix}$$

$\mathbf{g} = \mathbf{Z}(\mathbf{Z}\mathbf{Z}')^{-1}\mathbf{u}$ SNP effects using all information

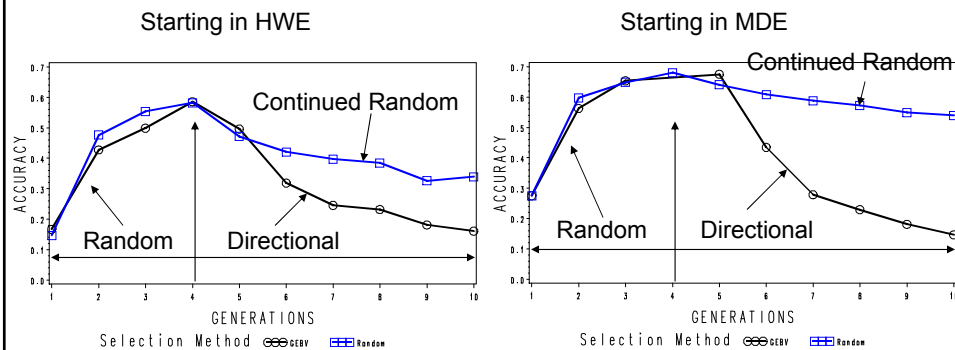
H. Wang, I. Misztal, I. Aguilar, A. Legarra, and W. M. Muir, 2012. Genome-wide association mapping including phenotypes from relatives without genotypes. *Genetics Research* (In press)

Issues

Genomic Selection and GWAS

- Admixture (Walsh Lectures)
 - Major problem
 - False Positives
 - Spurious Correlations
 - Correlation does not mean Causation
 - Partial Solution
 - Use of Igenstrat to correct for structure
 - Use of Structure to correct for structure
 - Scaling of G to combine all populations and cross in common relationship matrix
- Pedigree errors
 - More costly in terms of accuracy with G than A matrix
- Cost
 - Use dense SNP genotyping all breeders (parents)
 - Low Density on all candidates
 - Reducing the number of markers down to those that are most predictive
 - Going from 60,000 SNP to 384 SNP for genotyping
 - GWAS SNP selection
- Selection (See next slide)
 - Allele frequencies change
 - Older data becomes a liability

Effect of Random vs. Directional Selection on Accuracy



$h^2=.1$ $N=256$, $N_e=32$, 100/100 Marker/QTL loci distributed on 100cM.
 (average over 60 replicates, SEM=.02).

Appendix

SAS IML BLUP Programs

For Examples 1 and 2

```

proc iml;
start main;

y={ 7,
    9,
    10,
    6,
    9};

X={1,
    1,
    1,
    1,
    1};

A={1 0 0.5 0,
    0 1 0.5 .5,
    0 0 1 0 .5,
    .5 .5 0 1 .25,
    0 .5 .5 .25 1};

lam=1;

Z={1 0 0 0 0,
    0 1 0 0 0,
    0 0 1 0 0,
    0 0 0 1 0,
    0 0 0 0 1};

LHS=((X`*X)||((X`*Z)))/((Z`*X)||((Z`*
Z+INV(A)#LAM));

RHS=(X`*Y)/(Z`*Y);
C=INV(HS);

BU=C*RHS;
print C BU;

finish main;
run;
quit;

```

Estimates

BU

8.3018868
-0.960813
0.0754717
0.8853411
-1.062409
0.5529753

$b = [\hat{\mu}]$

$U = \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \end{bmatrix}$

```

proc iml;
start main;

y={ 7,
    10,
    6};

X={1,
    1,
    1};

A={1 0 0.5 0,
    0 1 0.5 .5,
    0 0 1 0 .5,
    .5 .5 0 1 .25,
    0.5 .5 .25 1};

lam=1;

Z={1 0 0 0 0,
    0 0 1 0 0,
    0 0 0 0 1};

LHS=((X`*X)||X`*Z)/((Z`*X)||Z`*
Z+INV(A)#LAM);

RHS=(X`*Y)/(Z`*Y);
C=INV(LHS);

BU=C*RHS;
print C BU;

finish main;
run;
quit;

```

Estimates

$$\begin{array}{r}
 7.84 \\
 -0.64 \\
 -0.43 \\
 1.07 \\
 -0.97 \\
 0.32
 \end{array}
 \left. \begin{array}{l} \\ \\ \\ \\ \\ \\ \end{array} \right\} \mathbf{U} = \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \end{bmatrix}$$

$b = [\hat{\mu}]$

```

LHS=(X`X)||(X`Z)||((Z`X)||((Z`Z+INV(A)#LAM)));
RHS=(X`Y)/(Z`Y);
C=INV(LHS);
BU=C*RHS;
RMSE=(Y`Y-BU`RHS)#(1/3);

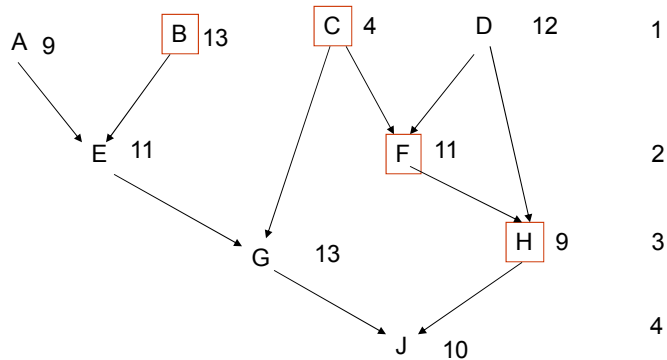
print C BU RMSE;

CSUB=C[2:6,2:6];           Sub matrix with rows 2 through 6 and
PEV=CSUB#RMSE;           Columns 2 through 6 of C
SIGMA_A=RMSE*(1/LAM);
PV=A*SIGMA_A+PEV;

PRINT PEV PV;
finish main;
run;
quit;

```

Problem 1



Find the best estimate of the genetic worth of each animal.
Assume a heritability of .5

Answer Problem 1

```

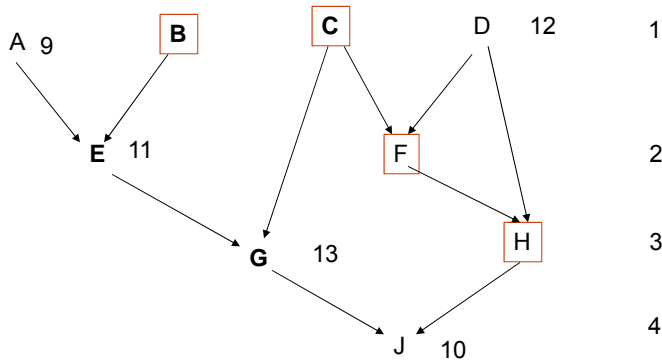
proc iml;
start main;
A={1 0 0 0 0.5 0 0.25 0 0.125,
0 1 0 0 0.5 0 0.25 0 0.125,
0 0 1 0 0 0.5 0.5 0.25 0.375,
0 0 0 1 0 0.5 0 0.75 0.375,
0.5 0.5 0 0 1 0 0.5 0 0.25,
0 0 0.5 0.5 0 1 0.25 0.75 0.5,
0.25 0.25 0.5 0 0.5 0.25 1 0.125 0.5625,
0 0 0.25 0.75 0 0.75 0.125 1.25 0.6875,
0.125 0.125 0.375 0.375 0.25 0.5 0.5625 0.6875 1.0625};
AINV=INV(A);
lam=1;
X={1,
1,
1,
1,
1,
1,
1,
1,
1,
1};
Z={1 0 0 0 0 0 0 0 0,
0 1 0 0 0 0 0 0 0,
0 0 1 0 0 0 0 0 0,
0 0 0 1 0 0 0 0 0,
0 0 0 0 1 0 0 0 0,
0 0 0 0 0 1 0 0 0,
0 0 0 0 0 0 1 0 0,
0 0 0 0 0 0 0 1 0,
0 0 0 0 0 0 0 0 1};
LHS=((X`*X)|| (X`*Z)) / ((Z`*X)|| (Z`*Z+AINV#LAM));
RHS=(X`*Y)/(Z`*Y);
C=INV(LHS);
BU=C*RHS;

```

Answer

10.07
4.779
4.779
-0.31
1.689
BU= -2.28
0.905
1.145
-0.31
0.564
-0.19
0.105

Problem 2: Sex Limited Trait



Estimate breeding values for the males.
Assume a heritability of .5

Answer Problem 2

```

proc iml;
start main;
A={1 0 0 0 0.5 0 0.25 0 0.125,
  0 1 0 0 0.5 0 0.25 0 0.125,
  0 0 1 0 0 0.5 0.5 0.25 0.375,
  0 0 0 1 0 0.5 0 0.75 0.375,
  12, 0.5 0.5 0 0 1 0 0.5 0 0.25,
  11, 0 0 0.5 0.5 0 1 0.25 0.75 0.5,
  13, 0.25 0.25 0.5 0 0.5 0.25 1 0.125 0.5625,
  10};
X={1,
  1,
  1,
  1,
  1};
AINV=INV(A);
lam=1;
Z={1 0 0 0 0 0 0 0 0,
  0 0 1 0 0 0 0 0,
  0 0 0 1 0 0 0 0,
  0 0 0 0 0 1 0 0,
  0 0 0 0 0 0 0 1};
LHS=((X`*X)||X`*Z)/((Z`*X)||Z`*Z+AINV#LAM);
RHS=(X`*Y)/(Z`*Y);
C=INV(LHS);
BU=C*RHS;

```

Answer

```

11.03
-0.89
0.247
0.338
0.307
BU= -0.075
0.206
0.587
0.023
-0.102

```