Lecture 6

Estimation of Basic Genetic Parameters

Guilherme J. M. Rosa
University of Wisconsin-Madison

Introduction to Quantitative Genetics
SISG, Seattle
20 - 22 July 2020

Heritability

Narrow vs. broad sense

Narrow sense: $h^2 = V_A/V_P$

Slope of midparent - offspring regression
(sexual reproduction)

Broad sense: $H^2 = V_G/V_P$

Slope of a parent - cloned offspring regression
(asexual reproduction)

When one refers to heritability, the default is narrow-sense, h^2

h^2 is the measure of (easily) usable genetic variation under sexual reproduction

Why h^2 instead of h?

Blame Sewall Wright, who used h to denote the correlation between phenotype and breeding value. Hence, h^2 is the total fraction of phenotypic variance due to breeding values

$$r(A,P) = \frac{\sigma(A,P)}{\sigma_A \sigma_P} = \frac{\sigma_A^2}{\sigma_A \sigma_P} = \frac{\sigma_A}{\sigma_P} = h$$

Heritabilities are functions of populations

Heritability values only make sense in the context of the population for which it was measured

Heritability measures the *standing genetic variation* of a population

A zero heritability DOES NOT imply that the trait is not genetically determined

Heritabilities are functions of the distribution of environmental values (i.e., the *universe* of E values)

Decreasing V_P increases h^2.

Heritability values measured in one environment (or distribution of environments) may not be valid under another

Measures of heritability for lab-reared individuals may be very different from heritability in nature

Heritability and the Prediction of Breeding Values

If P denotes an individual's phenotype, then best linear predictor of their breeding value A is

$$A = \frac{\sigma(P,A)}{\sigma_P^2}(P - \mu_P) + e = h^2(P - \mu_P) + e$$

The residual variance is also a function of h^2:

$$\sigma_e^2 = (1 - h^2)\sigma_P^2$$

The larger the heritability, the tighter the distribution of true breeding values around the value $h^2(P - \mu_P)$ predicted by an individual's phenotype.

Heritability and Population Divergence

Heritability is a completely unreliable predictor of long-term response

Measuring heritability values in two populations that show a difference in their means provides no information on whether the underlying difference is genetic

Sample Heritabilities

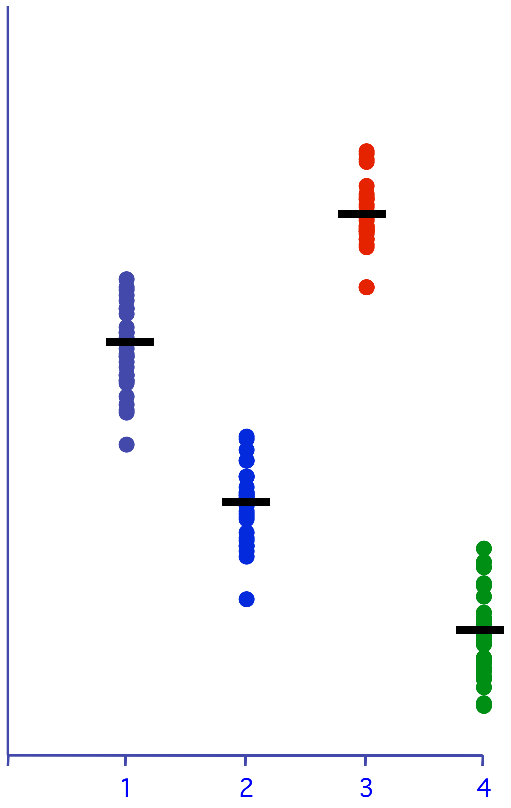
		h^2
People	Height	0.80
	Serum IG	0.45
Pigs	Back-fat	0.70
	Weight gain	0.30
	Litter size	0.05
Fruit Flies	Abdominal Bristles	0.50
	Body size	0.40
	Ovary size	0.30
	Egg production	0.20

Traits more closely associated with fitness tend to have lower heritabilities

ANOVA: Analysis of Variance

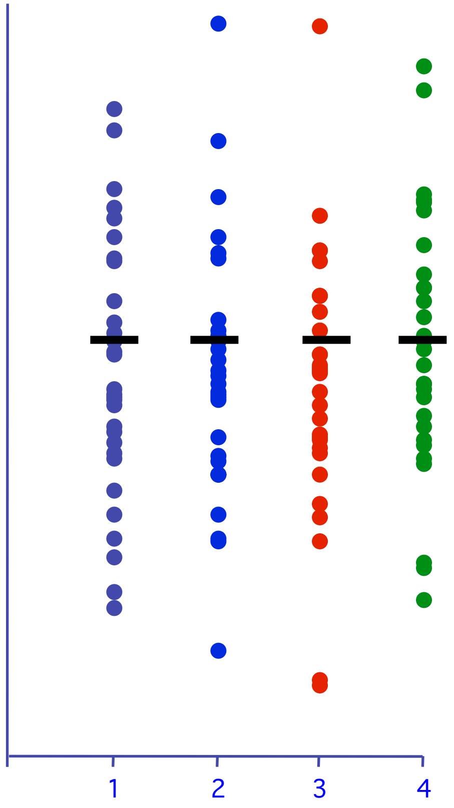
- Partitioning of trait variance into within- and among-group components
- Two key ANOVA identities
  - Total variance = between-group variance + within-group variance
    - Var(T) = Var(B) + Var(W)
  - Variance(between groups) = covariance (within groups)
  - Intraclass correlation, t = Var(B)/Var(T)
- The more similar individuals are within a group (higher within-group covariance), the larger their between-group differences (variance in the group means)

Situation 1



$$\left.\begin{array}{l}\mathrm{Var}(B) = 2.5\\ \mathrm{Var}(W) = 0.2\\ \mathrm{Var}(T) = 2.7\end{array}\right\} t = 2.5/2.7 = 0.93$$

Situation 2



$$\left.\begin{array}{l}\mathrm{Var}(B) = 0\\ \mathrm{Var}(W) = 2.7\\ \mathrm{Var}(T) = 2.7\end{array}\right\} t = 0$$

Phenotypic Resemblance Between Relatives

Relatives	Covariance	Regression (b) or correlation (t)
Offspring and one parent	$\frac{1}{2}V_A$	$b = \frac{1}{2}\frac{V_A}{V_P}$
Offspring and mid-parent	$\frac{1}{2}V_A$	$b = \frac{V_A}{V_P}$
Half sibs	$\frac{1}{4}V_A$	$t = \frac{1}{4}\frac{V_A}{V_P}$
Full sibs	$\frac{1}{2}V_A + \frac{1}{4}V_D + V_{E_c}$	$t = \frac{\frac{1}{2}V_A + \frac{1}{4}V_D + V_{E_c}}{V_P}$

Why cov(within) = variance(among)?

- Let z_{ij} denote the jth member of group i.
  - Here $z_{ij} = u + g_i + e_{ij}$
  - g_i is the group effect
  - e_{ij} the residual error
- Covariance within a group $Cov(z_{ij}, z_{ik})$

  $= Cov(u + g_i + e_{ij}, u + g_i + e_{ik})$

  $= Cov(g_i, g_i)$ as all other terms are uncorrelated

  - $Cov(g_i, g_i) = Var(g)$ is the among-group variance

PAUSE

- ⇨ h^2 and H^2
- ⇨ ANOVA
- ⇨ Variance(between groups) = covariance (within groups)
- ⇨ Intraclass correlation

Next PAUSE, slide 29

Estimation: One-way ANOVA

Simple (balanced) full-sib design: N full-sib families, each with n offspring: One-way ANOVA model

Trait value in sib j from family i

Common mean

Deviation of sib j from the family mean

$$z_{ij} = m + f_i + w_{ij}$$

Effect for family i; deviation of mean of i from the common mean

Covariance between members of the same group equals the variance among (between) groups

$$Cov(Full\ Sibs) = \sigma(z_{ij}, z_{ik})$$

$$= \sigma[(\mu + f_i + w_{ij}), (\mu + f_i + w_{ik})]$$

$$= \sigma(f_i, f_i) + \sigma(f_i, w_{ik}) + \sigma(w_{ij}, f_i) + \sigma(w_{ij}, w_{ik})$$

$$= \sigma_f^2$$

Hence, the variance among family effects equals the covariance between full sibs

$$\sigma_f^2 = \sigma_A^2 / 2 + \sigma_D^2 / 4 + \sigma_{Ec}^2$$

The within-family variance $\sigma^2_w = \sigma^2_P - \sigma^2_f$,

$$\begin{aligned}
\sigma^2_{w(FS)} &= \sigma^2_P - (\sigma^2_A / 2 + \sigma^2_D / 4 + \sigma^2_{Ec}) \\
&= \sigma^2_A + \sigma^2_D + \sigma^2_E - (\sigma^2_A / 2 + \sigma^2_D / 4 + \sigma^2_{Ec}) \\
&= (1/2)\sigma^2_A + (3/4)\sigma^2_D + \sigma^2_E - \sigma^2_{Ec}
\end{aligned}$$

One-way ANOVA: N families with n sibs, T = Nn

Factor	Degrees of freedom, df	Sum of squares (SS)	Mean squares (MS)	E[MS]
Among family	N-1	$SS_f = n\sum_{i=1}^{N}(\bar{z}_i - \bar{z})^2$	$SS_f/(N-1)$	$\sigma^2_w + n\ \sigma^2_f$
Within family	T-N	$SS_w = \sum_{i=1}^{N}\sum_{j=1}^{n}(z_{ij} - \bar{z}_i)^2$	$SS_w/(T-N)$	σ^2_w

Estimating the variance components:

$$Var(f) = \frac{MS_f - MS_w}{n}$$

$$Var(w) = MS_w$$

$$Var(z) = Var(f) + Var(w)$$

Since $\sigma_f^2 = \sigma_A^2 / 2 + \sigma_D^2 / 4 + \sigma_{Ec}^2$

2Var(f) is an upper bound for the additive variance

Assigning standard errors (= square root of Var)

Fun fact: Under normality, the (large-sample) variance for a mean-square is given by

$$\sigma^2(MS_x) \cong \frac{2(MS_x)^2}{df_x + 2}$$

$$Var[Var(w(FS))] = Var(MS_w) \cong \frac{2(MS_w)^2}{T - N + 2}$$

$$Var[Var(f)] = Var\left[\frac{MS_f - MS_w}{n}\right]$$

$$\cong \frac{2}{n^2}\left[\frac{(MS_f)^2}{N+1} + \frac{(MS_w)^2}{T - N + 2}\right]$$

Estimating heritability

$$t_{FS} = \frac{Var(f)}{Var(z)} = \frac{1}{2}h^2 + \frac{\sigma_D^2/4 + \sigma_{Ec}^2}{\sigma_z^2}$$

Hence, $h^2 \leq 2\,t_{FS}$

An approximate large-sample standard error for h^2 is given by

$$SE(h^2) \cong 2(1 - t_{FS})[1 + (n-1)t_{FS}]\sqrt{2/[Nn(n-1)]}$$

Worked Example

10 full-sib families, each with 5 offspring are measured

Factor	df	SS	MS	EMS
Among-families	9	$SS_f = 405$	45	$\sigma^2_w + 5\,\sigma^2_f$
Within-families	40	$SS_w = 800$	20	σ^2_w

$$Var(f) = \frac{MS_f - MS_w}{n} = \frac{45-20}{5} = 5 \longrightarrow V_A < 10$$

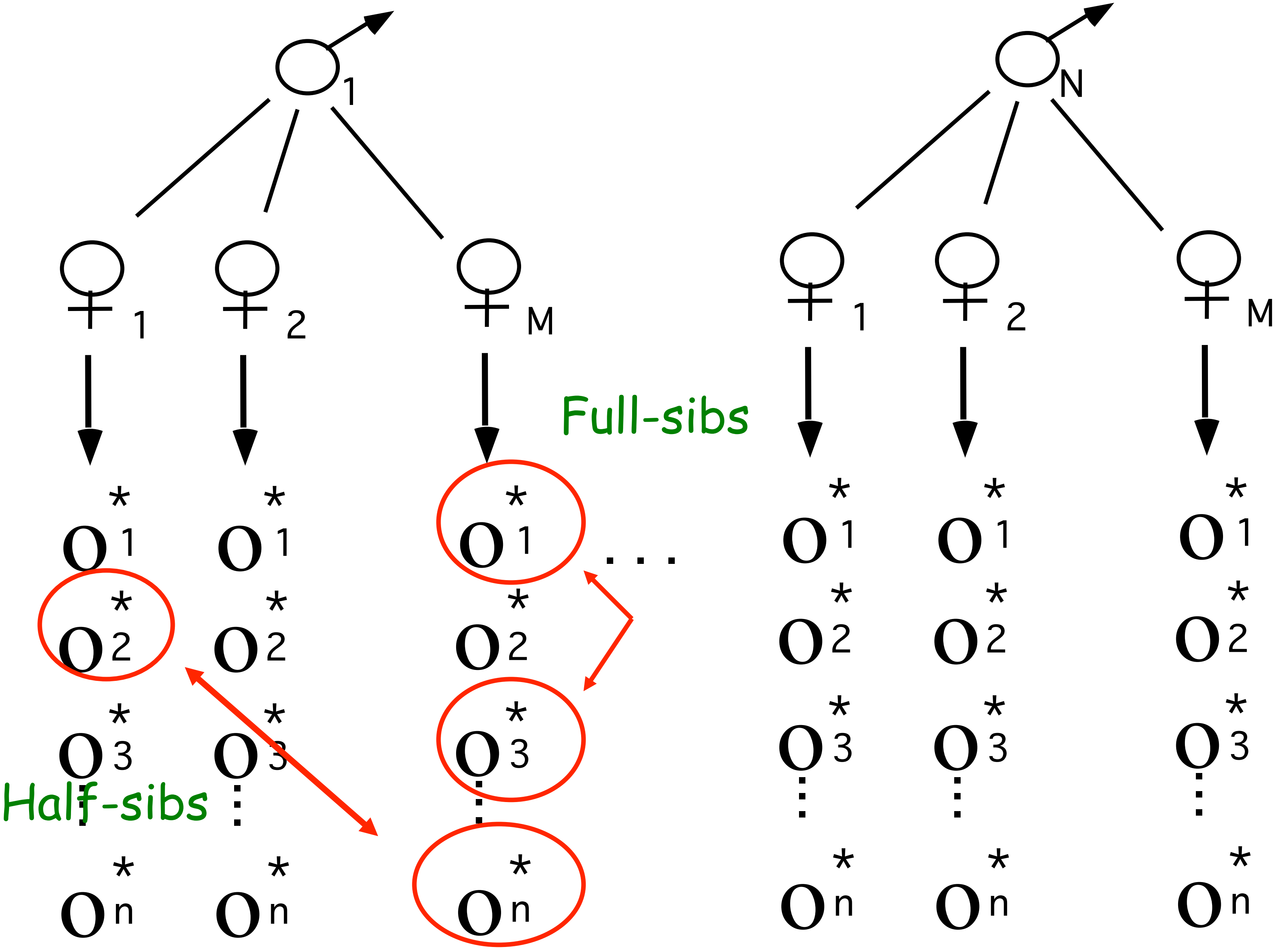
$$Var(w) = MS_w = 20$$

$h^2 < 2\,(5/25) = 0.4$

$$Var(z) = Var(f) + Var(w) = 25$$

$$SE(h^2) \cong 2(1-0.4)[1+(5-1)0.4]\sqrt{2/[50(5-1)]} = 0.312$$

Full sib-half sib design: Nested ANOVA



Estimation: Nested ANOVA

Balanced full-sib / half-sib design: N males (sires) are crossed to M dams each of which has n offspring:

Nested ANOVA model

Value of the kth offspring from the jth dam for sire i

Overall mean

Effect of dam j of sire i; deviation of mean of dam j from sire and overall mean

$$z_{ijk} = m + s_i + d_{ij} + w_{ijk}$$

Effect of sire i; deviation of mean of i's family from overall mean

Within-family deviation of kth offspring from the mean of the ij-th family

Nested ANOVA Model

$$z_{ijk} = m + s_i + d_{ij} + w_{ijk}$$

σ^2_s = between-sire variance = variance in sire family means

σ^2_d = variance among dams within sires = variance of dam means for the same sire

σ^2_w = within-family variance

$\sigma^2_T = \sigma^2_s + \sigma^2_d + \sigma^2_w$

Nested ANOVA: N sires crossed to M dams, each with n sibs, T = NMn

Factor	df	SS	MS	E[MS]
Sires	N-1	SS_s	$SS_s/(N-1)$	$\sigma_w^2 + n\sigma_d^2 + Mn\sigma_s^2$
Dams(Sires)	N(M-1)	SS_d	$SS_d/[N(M-1)]$	$\sigma_w^2 + n\sigma_d^2$
Sibs(Dams)	T-NM	SS_w	$SS_w/(T-NM)$	σ_w^2

where: $SS_s = Mn\sum_{i=1}^{N}(\overline{z}_i - \overline{z})^2$

$SS_d = n\sum_{i=1}^{N}\sum_{j=1}^{M}(\overline{z}_{ij} - \overline{z}_i)^2$ and $SS_w = n\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{k=1}^{n}(z_{ijk} - \overline{z}_{ij})^2$

Estimation of sire, dam, and family variances:

$$Var(s) = \frac{MS_s - MS_d}{Mn}$$

$$Var(d) = \frac{MS_d - MS_w}{n}$$

$$Var(e) = MS_w$$

Translating these into the desired variance components:

- Var(Total) = Var(between FS families) + Var(within FS)

$$\rightarrow \sigma_w^2 = \sigma_z^2 - Cov(FS)$$

- Var(Sires) = Cov(Paternal half-sibs)

$$\sigma_d^2 = \sigma_z^2 - \sigma_s^2 - \sigma_w^2 = \sigma(FS) - \sigma(PHS)$$

Summarizing:

$$\sigma_s^2 = \sigma(PHS)$$

$$\sigma_w^2 = \sigma_z^2 - \sigma(FS)$$

$$\sigma_d^2 = \sigma_z^2 - \sigma_s^2 - \sigma_w^2$$
$$= \sigma(FS) - \sigma(PHS)$$

Expressing these in terms of the genetic and environmental variances:

$$\sigma_s^2 \cong \frac{\sigma_A^2}{4}$$

$$\sigma_d^2 \cong \frac{\sigma_A^2}{4} + \frac{\sigma_D^2}{4} + \sigma_{Ec}^2$$

$$\sigma_w^2 \cong \frac{\sigma_A^2}{2} + \frac{3\sigma_D^2}{4} + \sigma_{Es}^2$$

Intraclass correlations and estimating heritability

$$t_{PHS} = \frac{Cov(PHS)}{Var(z)} = \frac{Var(s)}{Var(z)} \rightarrow 4t_{PHS} = h^2$$

$$t_{FS} = \frac{Cov(FS)}{Var(z)} = \frac{Var(s) + Var(d)}{Var(z)} \rightarrow h^2 \leq 2t_{FS}$$

Note that $4\mathrm{t}_{\mathrm{PHS}} = 2\mathrm{t}_{\mathrm{FS}}$ implies no dominance or shared family environmental effects

Worked Example: N = 10 sires, M = 3 dams, n = 10 sibs/dam

Factor	df	SS	MS	E[MS]
Sires	9	4,230	470	$\sigma_w^2 + 10\sigma_d^2 + 30\sigma_s^2$
Dams(Sires)	20	3,400	170	$\sigma_w^2 + 10\sigma_d^2$
Within Dams	270	5,400	20	σ_w^2

$$\sigma_w^2 = MS_w = 20$$

$$\sigma_d^2 = \frac{MS_d - MS_w}{n} = \frac{170 - 20}{10} = 15$$

$$\sigma_s^2 = \frac{MS_s - MS_d}{Nn} = \frac{470 - 170}{30} = 10$$

$$\sigma_P^2 = \sigma_s^2 + \sigma_d^2 + \sigma_w^2 = 45$$

$$\sigma_d^2 = 15 = (1/4)\sigma_A^2 + (1/4)\sigma_D^2 + \sigma_{Ec}^2$$

$$= 10 + (1/4)\sigma_D^2 + \sigma_{Ec}^2$$

$$\boxed{\sigma_A^2 = 4\sigma_s^2 = 40}$$

$$\boxed{h^2 = \frac{\sigma_A^2}{\sigma_P^2} = \frac{40}{45} = 0.89}$$

$$\boxed{\sigma_D^2 + 4\sigma_{Ec}^2 = 20}$$

PAUSE

- ⇨ One-way ANOVA
- ⇨ Nested ANOVA
- ⇨ ANOVA variance components and the desired genetic (functional) components

Next PAUSE, slide 39

Parent-offspring Regression

Single parent - offspring regression

$$z_{o_i} = \mu + b_{o|p}(z_{p_i} - \mu) + e_i$$

The expected slope of this regression is:

$$E(b_{o|p}) = \frac{\sigma(z_o, z_p)}{\sigma^2(z_p)} \cong \frac{(\sigma_A^2/2) + \sigma(E_o, E_p)}{\sigma_z^2} = \frac{h^2}{2} + \frac{\sigma(E_o, E_p)}{\sigma_z^2}$$

Residual error variance (spread around expected values)

$$\sigma_e^2 = \left(1 - \frac{h^2}{2}\right)\sigma_z^2$$

The expected slope of this regression is:

$$E(b_{o|p}) = \frac{\sigma(z_o, z_p)}{\sigma^2(z_p)} \cong \frac{(\sigma_A^2/2) + \sigma(E_o, E_p)}{\sigma_z^2} = \frac{h^2}{2} + \boxed{\frac{\sigma(E_o, E_p)}{\sigma_z^2}}$$

Shared environmental values

To avoid this term, typically regressions are male-offspring, as female-offspring more likely to share environmental values

Midparent-offspring regression:

$$z_{o_i} = \mu + b_{o|MP}\left(\frac{z_{m_i} + z_{f_i}}{2} - \mu\right) + e_i$$

$$b_{o|MP} = \frac{Cov[z_o,(z_m + z_f)/2]}{Var[(z_m + z_f)/2]}$$

$$= \frac{[Cov(z_o,z_m) + Cov(z_o,z_f)]/2}{[Var(z) + Var(z)]/4}$$

$$= \frac{2Cov(z_o,z_p)}{Var(z)} = 2b_{o|p}$$

The expected slope of this regression is h^2

Residual error variance (spread around expected values)

$$\sigma_e^2 = \left(1 - \frac{h^2}{2}\right)\sigma_z^2$$

Standard Errors

Single parent-offspring regression, N parents, each with n offspring

Square regression slope

$$Var(b_{o|p}) \cong \frac{n(t - b^2_{p|p}) + (1 - t)}{Nn}$$

Total number of offspring

Sib correlation $t = \begin{cases} t_{HS} = h^2/4 & \text{for half-sibs} \\ t_{FS} = h^2/2 + \dfrac{\sigma^2_D + \sigma^2_{Ec}}{\sigma^2_z} & \text{for full-sibs} \end{cases}$

$$Var(h^2) = Var(2b_{o|p}) = 4Var(b_{o|p})$$

Midparent-offspring regression,
N sets of parents, each with n offspring

$$Var(h^2) = Var(b_{o|MP}) \cong \frac{2[n(t_{FS} - b^2_{o|MP}/2) + (1 - t_{FS})]}{Nn}$$

- Midparent-offspring variance half that of single parent-offspring variance

$$Var(h^2) = Var(2b_{o|p}) = 4Var(b_{o|p})$$

Estimating Heritability in Natural Populations

Often, sibs are reared in a laboratory environment, making parent-offspring regressions and sib ANOVA problematic for estimating heritability

Let b′ be the slope of the regression of the values of lab-raised offspring regressed in the trait values of their parents in the wild

A lower bound can be placed of heritability using parents from nature and their lab-reared offspring,

$$h^2_{min} = (b'_{o|MP})^2 \frac{Var_n(z)}{Var_l(A)}$$

Trait variance in nature

Additive variance in lab

Why is this a lower bound?

Covariance between breeding value in nature and BV in lab

$$(b'_{o|MP})^2 \frac{Var_n(z)}{Var_l(A)} = \left[\frac{Cov_{l,n}(A)}{Var_n(z)}\right]^2 \frac{Var_n(z)}{Var_l(A)} = \gamma^2 h_n^2$$

where $\gamma = \dfrac{Cov_{l,n}(A)}{\sqrt{Var_n(A)Var_l(A)}}$

is the additive genetic covariance between environments and hence $\Upsilon^2 \leq 1$

Defining H^2 for Plant Populations

Plant breeders often do not measure individual plants (especially with pure lines), but instead measure a plot or a block of individuals. This can result in inconsistent measures of H^2 even for otherwise identical populations

Genotype i

Environment j

Effect of plot k for genotype i in environment j

$$z_{ijkl} = G_i + E_j + GE_{ij} + p_{ijk} + e_{ijkl}$$

Interaction between genotype i and environment j

Deviations of individual plants within plots

$$z_{ijkl} = G_i + E_j + GE_{ij} + p_{ijk} + e_{ijkl}$$

$$\sigma^2(z_i) = \sigma^2_G + \sigma^2_E + \frac{\sigma^2_{GE}}{e} + \frac{\sigma^2_p}{er} + \frac{\sigma^2_e}{ern}$$

e = number of environments
r = (replicates) number of plots/environment
n = number of individuals per plot

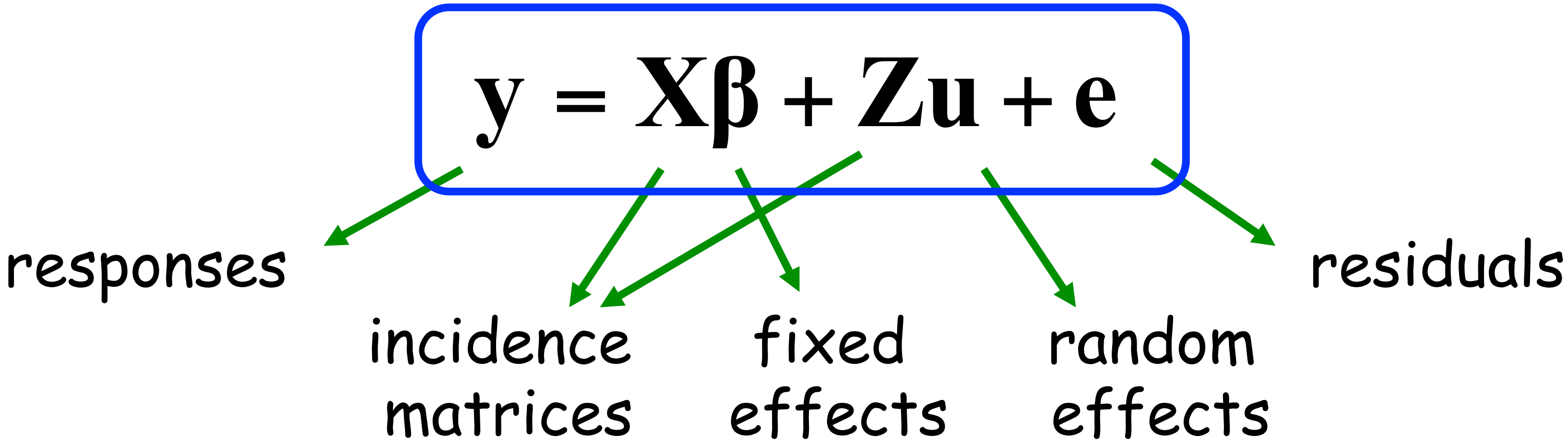
Hence, V_P, and hence H^2, depends on our choice of e, r, and n

PAUSE

- ⇨ Regression approach
- ⇨ Parent-offspring variations
- ⇨ Standard errors

Next PAUSE, slide 50

Linear Mixed Effects Model



$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

responses

incidence matrices

fixed effects

random effects

residuals

$$\begin{bmatrix} \mathbf{u} \\ \mathbf{e} \end{bmatrix} \sim \mathrm{MVN}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} \end{bmatrix} \right)$$

Estimation of Fixed Effects

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

with $\boldsymbol{\varepsilon} = \mathbf{Zu} + \mathbf{e}$, such that $\mathrm{Var}[\boldsymbol{\varepsilon}] = \mathbf{ZGZ}^{\mathrm{T}} + \boldsymbol{\Sigma}$

➔ MLE of β :

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{y} \sim \mathrm{MVN}(\boldsymbol{\beta}, (\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X})^{-1})$$

where $\mathbf{V} = \mathbf{ZGZ}^{\mathrm{T}} + \boldsymbol{\Sigma}$

Prediction of Random Effects

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{u} \end{bmatrix} \sim \mathrm{MVN}\left(\begin{bmatrix} \mathbf{X\beta} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{V} & \mathbf{ZG} \\ \mathbf{GZ}^{\mathrm{T}} & \mathbf{G} \end{bmatrix} \right)$$

$$\begin{aligned} \mathrm{E}[\mathbf{u} \mid \mathbf{y}] &= \mathrm{E}[\mathbf{u}] + \mathrm{Cov}[\mathbf{u}, \mathbf{y}^{\mathrm{T}}]\mathrm{Var}^{-1}[\mathbf{y}](\mathbf{y} - \mathrm{E}[\mathbf{y}]) \\ &= \mathbf{GZ}^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X\beta}) = \mathbf{GZ}^{\mathrm{T}}(\mathbf{ZGZ}^{\mathrm{T}} + \mathbf{\Sigma})^{-1}(\mathbf{y} - \mathbf{X\beta}) \end{aligned}$$

Replacing β by its estimate:

$$\boxed{\hat{\mathbf{u}} = \mathbf{GZ}^{\mathrm{T}}(\mathbf{ZGZ}^{\mathrm{T}} + \mathbf{\Sigma})^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}$$

Mixed Model Equations

$$\begin{bmatrix} \mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X} & \mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{Z} \\ \mathbf{Z}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X} & \mathbf{Z}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{Z}+\mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{y} \\ \mathbf{Z}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{y} \end{bmatrix}$$

BLUP and BLUE:

$$\begin{cases} \hat{\mathbf{u}} = (\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{Z}+\mathbf{G}^{-1})^{-1}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}) \\ \hat{\boldsymbol{\beta}} = \{\mathbf{X}^{\mathrm{T}}[\boldsymbol{\Sigma}^{-1}-\boldsymbol{\Sigma}^{-1}\mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{Z}+\mathbf{G}^{-1})^{-1}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}]\mathbf{X}\}^{-1} \\ \qquad \times \mathbf{X}^{\mathrm{T}}[\boldsymbol{\Sigma}^{-1}-\boldsymbol{\Sigma}^{-1}\mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{Z}+\mathbf{G}^{-1})^{-1}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}]\mathbf{y} \end{cases}$$

Estimation of Variance Components

BLUE and BLUP require knowledge of **G** and Σ

These matrices, however, are rarely known and must be estimated

Variance and covariance components estimation:

- *Analysis of Variance (ANOVA)*
- Maximum Likelihood
- Restricted Maximum Likelihood (REML)
- Bayesian Inference

Mixed Models in Animal and Plant Breeding

Many statistical methods for analysis of genetic data are specific (or more appropriate) for phenotypic measurements obtained from planned experimental designs and with balanced data sets

While such situations may be possible within laboratory or greenhouse experimental settings, data from natural populations and agricultural species are generally highly unbalanced and fragmented by numerous kinds of relationships

Animal Model

The mixed model methodology allows efficient estimation of genetic parameters (such as variance components and heritability) and breeding values while accommodating extended pedigrees, unequal family sizes, overlapping generations, sex-limited traits, assortative mating, and natural or artificial selection

To illustrate such application of mixed models in breeding programs, we consider here the so-called Animal Model in situations with a single trait and a single observation (including missing values) per individual

Animal Model

The animal model can be described as:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \mathbf{e}$$

$\mathbf{y}$ is an $(n \times 1)$ vector of observations (phenotypic scores)

β is a $(p \times 1)$ vector of fixed effects (e.g. herd-year-season effects)

$\mathbf{u} \sim N(\mathbf{0}, \boldsymbol{G})$ is a $(q \times 1)$ vector of breeding values (relative to all individuals with record or in the pedigree file, such that q is in general bigger than n)

$\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}_n\sigma_e^2)$ represents residual effects, where σ_e^2 is the residual variance

The Matrix *A*

The matrix ***G*** describing the covariances among the random effects (here the breeding values) follows from standard results for the covariances between relatives

It is seen that the additive genetic covariance between two relatives i and i' is given by $2\theta_{ii'}\sigma_a^2$, where $\theta_{ii'}$ is the coefficient of coancestry between individuals i and i', and σ_a^2 is the additive genetic variance in the base population

Hence, under the animal model, $\mathbf{G} = \mathbf{A}\sigma_a^2$, where ***A*** is the additive genetic (or numerator) relationship matrix, having elements given by $a_{ii'} = 2\theta_{ii'}$

The Matrix *A*

For each animal i in the pedigree (i = 1, 2,…,n), going from older to younger animals, compute a_{ii} and a_{ij} (j = 1, 2,…,i-1) as follows:

If both parents (s and d) of animal i are known:

$$a_{ij} = a_{ji} = (a_{js} + a_{jd})/2 \text{ and } a_{ii} = 1 + a_{sd}/2$$

If only one parent (e.g. d) of animal i is known:

$$a_{ij} = a_{ji} = a_{jd}/2 \text{ and } a_{ii} = 1$$

If parents unknown:

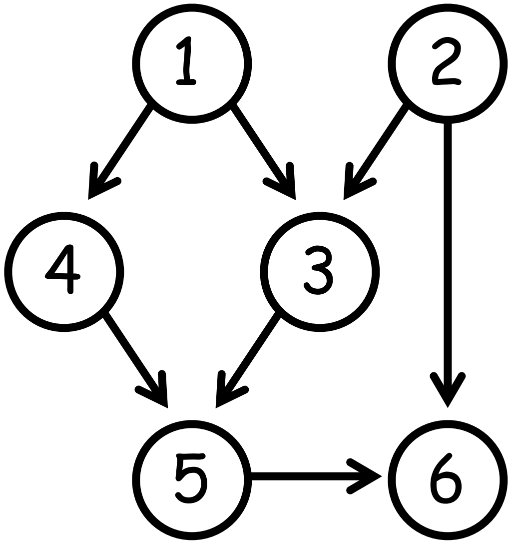
$$a_{ij} = a_{ji} = 0 \text{ and } a_{ii} = 1$$

PAUSE

- Mixed Model approach
- BLUE, BLUP, and REML
- Examples to follow

Next PAUSE, slide 64 (end of lecture)

Example



Animal	Sire	Dam
1	-	-
2	-	-
3	1	2
4	1	-
5	4	3
6	5	2

$$A = \begin{bmatrix} 1 & 0 & .5 & .5 & .5 & .25 \\ 0 & 1 & .5 & 0 & .25 & .625 \\ .5 & .5 & 1 & .25 & .625 & .563 \\ .5 & 0 & .25 & 1 & .625 & .313 \\ .5 & .25 & .625 & .625 & 1.125 & .688 \\ .25 & .625 & .563 & .313 & .688 & 1.125 \end{bmatrix}$$

Animal Model

In general, in animal/plant breeding interest is on prediction of breeding values (for selection of superior individuals), and on estimation of variance components and functions thereof, such as heritability

The fixed effects are, in some sense, nuisance factors with no central interest in terms of inferences, but which need to be taken into account (i.e., they need to be corrected for when inferring breeding values)

Animal Model

Since under the animal model $\mathbf{G}^{-1} = \mathbf{A}^{-1}\sigma_a^{-2}$ and $\mathbf{R}^{-1} = \mathbf{I}_n\sigma_e^{-2}$, the mixed model equations can be expressed as:

$$\begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z} + \lambda\mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{Z}^T\mathbf{y} \end{bmatrix}$$

where $\lambda = \frac{\sigma_e^2}{\sigma_a^2} = \frac{1-h^2}{h^2}$, such that:

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z} + \lambda\mathbf{A}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{Z}^T\mathbf{y} \end{bmatrix}$$

Conditional on the variance components ratio λ, the BLUP of the breeding values are given then by:

$$\hat{\mathbf{u}} = (\mathbf{Z}^{\mathrm{T}}\mathbf{Z} + \lambda\mathbf{A}^{-1})^{-1}\mathbf{Z}^{\mathrm{T}}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

These are generally referred to as Estimated Breeding Values (EBV)

Alternatively, some breeders associations express their results as Predicted Transmitting Abilities (PTA) (or Estimated Transmitting Abilities (ETA) or Expected Progeny Difference (EPD)), which are equal to half the EBV, representing the portion of an animal's breeding values that is passed to its offspring

The amount of information contained in an animal's genetic evaluation depends on the availability of its own record, as well as how many (and how close) relatives it has with phenotypic information

As a measure of amount of information in livestock genetic evaluations, EBVs are typically reported with its associated accuracies

Accuracy of predictions is defined as the correlation between true and estimated breeding values, i.e., $r_i = \rho(\hat{u}_i, u_i)$

Instead of accuracy, some livestock species genetic evaluations use reliability, which is the squared correlation of accuracy (r_i^2)

Prediction Accuracy

The calculation of $\rho(\hat{u}_i, u_i)$ requires the diagonal elements of the inverse of the MME coefficient matrix, represented as:

$$\mathbf{C} = \begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z} + \lambda\mathbf{A}^{-1} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{C}^{\beta\beta} & \mathbf{C}^{\beta u} \\ \mathbf{C}^{u\beta} & \mathbf{C}^{uu} \end{bmatrix}$$

It is shown that the prediction error variance of EBV $\hat{u}_i$ is given by:

$$\mathrm{PEV} = \mathrm{Var}(\hat{u}_i - u_i) = c_i^{uu}\sigma_e^2$$

where c_i^{uu} is the i-th diagonal element of $\mathbf{C}^{uu}$, relative to animal i.

Prediction Accuracy

The PEV can be interpreted as the fraction of additive genetic variance not accounted for by the prediction

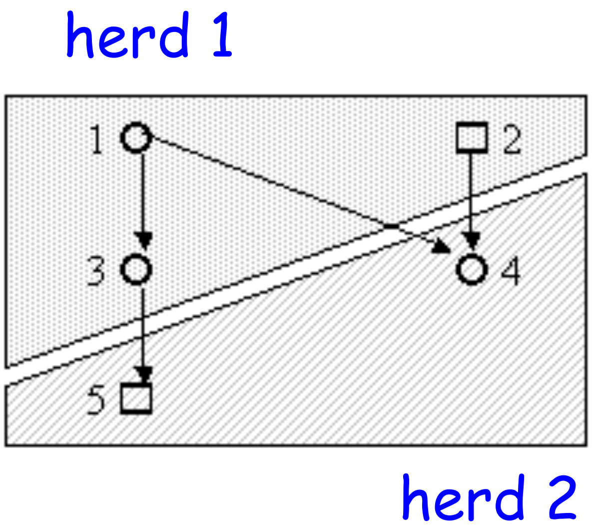
Therefore, PEV can be expressed also as:

$$\mathrm{PEV} = (1 - r_i^2)\sigma_a^2$$

such that $c_i^{uu}\sigma_e^2 = (1 - r_i^2)\sigma_a^2$, from which the reliability is obtained as:

$$r_i^2 = 1 - c_i^{uu}\sigma_e^2 / \sigma_a^2 = 1 - \lambda c_i^{uu}$$

Animal Model



Animal	Sire	Dam	Herd	Observation
1	–	–	h1	310
2	–	–	h1	–
3	–	1	h1	270
4	2	1	h2	350
5	–	3	h2	–

$$\underset{\mathbf{y}}{\begin{bmatrix}310\\270\\350\end{bmatrix}} = \underbrace{\begin{bmatrix}1&0\\1&0\\0&1\end{bmatrix}\begin{bmatrix}h_1\\h_2\end{bmatrix}}_{\mathbf{X}\,\beta} + \underbrace{\begin{bmatrix}1&0&0&0&0\\0&0&1&0&0\\0&0&0&1&0\end{bmatrix}\begin{bmatrix}u_1\\u_2\\u_3\\u_4\\u_5\end{bmatrix}}_{\mathbf{Z}\,\mathbf{u}} + \underset{\mathbf{e}}{\begin{bmatrix}e_1\\e_3\\e_4\end{bmatrix}}$$

Animal Model

Breeding values: $\mathbf{u} \sim \mathrm{N}(\mathbf{0}, \mathbf{A}\sigma_u^2)$, with

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0.5 & 0.5 & 0.25 \\ 0 & 1 & 0 & 0.5 & 0 \\ 0.5 & 0 & 1 & 0.25 & 0.5 \\ 0.5 & 0.5 & 0.25 & 1 & 0.125 \\ 0.25 & 0 & 0.5 & 0.125 & 1 \end{bmatrix}$$

Animal Model

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z} + \lambda\mathbf{A}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{Z}^T\mathbf{y} \end{bmatrix}$$

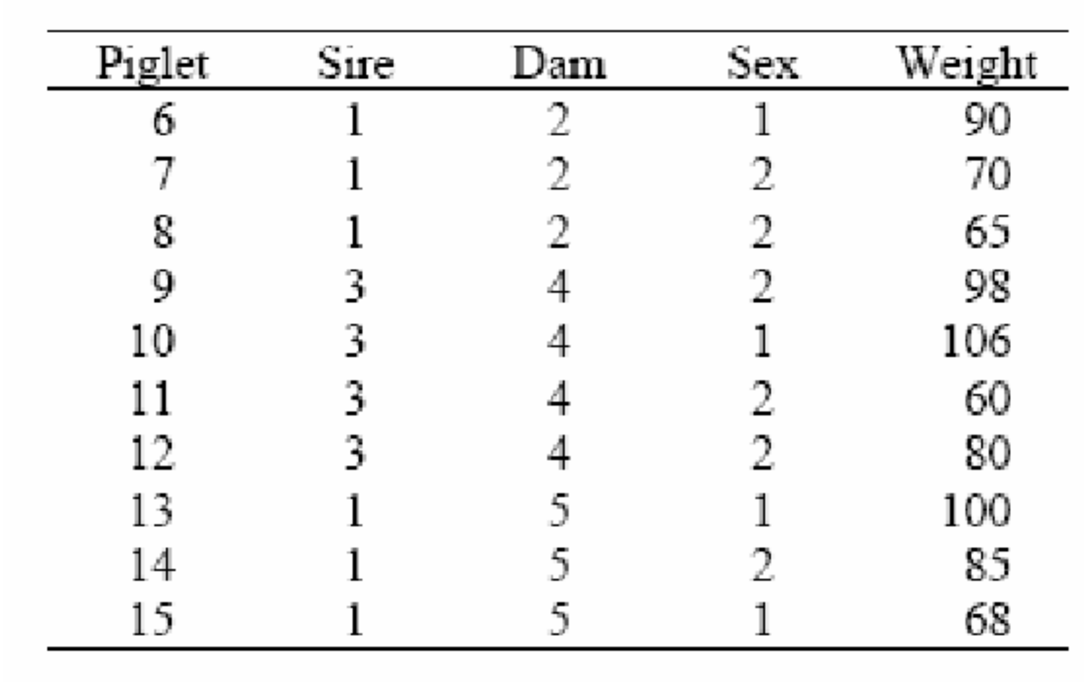
$$\lambda = \frac{\sigma_e^2}{\sigma_u^2} = \frac{1-h^2}{h^2}$$

$$h^2 = \frac{1}{3} \rightarrow \alpha = 2 \quad \Rightarrow \quad \begin{cases} \hat{h}_1 = 290 \\ \hat{h}_2 = 348 \\ \hat{u}_1 = 4.0 \\ \hat{u}_2 = 0.0 \\ \hat{u}_3 = -4.0 \\ \hat{u}_4 = 2.0 \\ \hat{u}_5 = -2.0 \end{cases}$$

Animal Model

The animal model can be extended to model multiple (correlated) traits, multiple random effects (such as maternal effects and common environmental effects), repeated records (e.g. test day models), and so on

Example (Mrode 1996, pp74-76): Weaning weight (kg) of piglets, progeny of three sows mated to two boars:



Piglet	Sire	Dam	Sex	Weight
6	1	2	1	90
7	1	2	2	70
8	1	2	2	65
9	3	4	2	98
10	3	4	1	106
11	3	4	2	60
12	3	4	2	80
13	1	5	1	100
14	1	5	2	85
15	1	5	1	68

A linear model with the (fixed) effect of sex, and the (random) effects of common environment (related to each litter) and breeding values can be expressed as **X**:

$$\underbrace{\mathbf{y}}_{\text{Weight}} = \underbrace{\mathbf{X}\boldsymbol{\beta}}_{\text{Sex}} + \underbrace{\mathbf{Z}\mathbf{u}}_{\text{Breeding values}} + \underbrace{\mathbf{W}\mathbf{c}}_{\text{Common environment}} + \underbrace{\mathbf{e}}_{\text{Residual}}$$

Assuming that $\sigma_u^2 = 20$, $\sigma_c^2 = 15$ and $\sigma_e^2 = 65$, the MME are as follows:

$$\begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} & \mathbf{X}^T\mathbf{W} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z} + \mathbf{A}^{-1}\lambda_1 & \mathbf{Z}^T\mathbf{W} \\ \mathbf{W}^T\mathbf{X} & \mathbf{W}^T\mathbf{Z} & \mathbf{W}^T\mathbf{W} + \mathbf{I}\lambda_2 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \\ \hat{\mathbf{c}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{Z}^T\mathbf{y} \\ \mathbf{W}^T\mathbf{y} \end{bmatrix}$$

where $\lambda_1 = \frac{\sigma_e^2}{\sigma_u^2} = 3.25$ and $\lambda_2 = \frac{\sigma_e^2}{\sigma_c^2} = 4.\dot{3}$

The BLUEs and BLUPs (inverting the numerator relationship matrix) are:

Effects	Solutions
Sex	
1	91.493
2	75.764
Animals	
1	-1.441
2	-1.175
3	1.441
4	1.441
5	-0.266
6	-1.098
7	-1.667
8	-2.334
9	3.925
10	2.895
11	-1.141
12	1.525
13	0.448
14	0.545
15	-3.819
Environ.	
2	-1.762
4	2.161
5	-0.399