Lecture 7
Estimation of Basic Genetic Parameters

Guilherme J. M. Rosa
University of Wisconsin-Madison

Introduction to Quantitative Genetics
SISG, Seattle
16 - 18 July 2018
Heritability

Narrow vs. broad sense

Narrow sense: \( h^2 = \frac{V_A}{V_P} \)
- Slope of midparent - offspring regression
  (sexual reproduction)

Broad sense: \( H^2 = \frac{V_G}{V_P} \)
- 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 content 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

<table>
<thead>
<tr>
<th>Species</th>
<th>Trait</th>
<th>$h^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>People</td>
<td>Height</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>Serum IG</td>
<td>0.45</td>
</tr>
<tr>
<td>Pigs</td>
<td>Back-fat</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>Weight gain</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>Litter size</td>
<td>0.05</td>
</tr>
<tr>
<td>Fruit Flies</td>
<td>Abdominal Bristles</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>Body size</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>Ovary size</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>Egg production</td>
<td>0.20</td>
</tr>
</tbody>
</table>

Traits more closely associated with fitness tend to have lower heritabilities.
Expected Value and Variance

Expected Value (Mean)

Notation: $E[X] = \mu_X$

- Discrete random variable, finite case:

$$E[X] = \sum_{i=1}^{k} x_i p_i \text{, where } p_i = \text{Pr}[X = x_i] \text{ (weighted average)}$$

If $p_1 = p_2 = \ldots = p_k = 1/k$ then:

$$E[X] = \frac{1}{k} \sum_{i=1}^{k} x_i \text{ (simple average)}$$
Expected Value

• Discrete random variable, countable case:

\[ E[X] = \sum_{i=1}^{\infty} x_i p_i \quad \text{and} \quad E[g(X)] = \sum_{i=1}^{\infty} g(x_i)p_i \]

• Continuous random variable:

\[ E[X] = \int_{-\infty}^{\infty} xf(x)dx \quad \text{and} \quad E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx \]

where \( f(x) \): probability density function

---

Expected Value

• Properties:

Constant \( c \):

\[ E[c] = c \]

\[ E[cX] = cE[X] \]

\[ E[X + Y] = E[X] + E[Y] \]

\[ E[X | Y = y] = \sum x \Pr(X = x | Y = y) \]

\[ E[X] = E[E[X | Y]] \]
Variance

**Notation:** $\text{Var}[X] = \sigma_X^2$

- **Definition:** expected value of the square deviation from the mean, i.e. $\text{Var}[X] = E[(X - \mu)^2]$

$$\text{Var}[X] = E[(X - E[X])^2]$$
$$= E[X^2] - 2XE[X] + (E[X])^2$$
$$= E[X^2] - 2E[X]E[X] + (E[X])^2$$
$$= E[X^2] - (E[X])^2$$
$$= E[X^2] - \mu^2$$

**Variance**

- **Discrete random variable:**

$$\text{Var}[X] = \sum_{i=1}^{\infty} (x_i - \mu)^2 p_i = \sum_{i=1}^{\infty} x_i^2 p_i - \mu^2$$

- **Continuous random variable:**

$$\text{Var}[X] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) \, dx = \int_{-\infty}^{\infty} x^2 f(x) \, dx - \mu^2$$
**Variance**

- **Properties:**

  **Constant c:** $\text{Var}[c] = 0$
  $\text{Var}[c + X] = \text{Var}[X]$
  $\text{Var}[cX] = c^2 \text{Var}[X]$

  $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2\text{Cov}[X, Y]$
  $\text{Var}[X - Y] = \text{Var}[X] + \text{Var}[Y] - 2\text{Cov}[X, Y]$
  $\text{Var}[X] = \mathbb{E}_Y[\text{Var}[X | Y]] + \text{Var}_Y[\mathbb{E}[X | Y]]$

**Covariance**

**Notation:** $\text{Cov}[X, Y] = \sigma_{X,Y}$

$\text{Cov}[X, Y] = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)]$

$= \mathbb{E}[XY] - \mu_X \mu_Y$

**Correlation**

**Notation:** $\text{Corr}[X, Y] = \rho_{X,Y}$

$\rho_{X,Y} = \frac{\text{Cov}[X, Y]}{\sigma_X \sigma_Y}$
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
    - \( \text{Var}(T) = \text{Var}(B) + \text{Var}(W) \)
  - Variance (between groups) = covariance (within groups)
  - Intraclass correlation, \( t = \frac{\text{Var}(B)}{\text{Var}(T)} \)

- 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**

\[
\begin{align*}
\text{Var}(B) &= 2.5 \\
\text{Var}(W) &= 0.2 \\
\text{Var}(T) &= 2.7 \\
t &= \frac{2.5}{2.7} = 0.93
\end{align*}
\]

**Situation 2**

\[
\begin{align*}
\text{Var}(B) &= 0 \\
\text{Var}(W) &= 2.7 \\
\text{Var}(T) &= 2.7 \\
t &= 0
\end{align*}
\]
Phenotypic Resemblance Between Relatives

<table>
<thead>
<tr>
<th>Relatives</th>
<th>Covariance</th>
<th>Regression (b) or correlation (t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Offspring and one parent</td>
<td>( \frac{1}{2} V_A )</td>
<td>( b = \frac{1}{2} \frac{V_A}{V_p} )</td>
</tr>
<tr>
<td>Offspring and mid-parent</td>
<td>( \frac{1}{2} V_A )</td>
<td>( b = \frac{V_A}{V_p} )</td>
</tr>
<tr>
<td>Half sibs</td>
<td>( \frac{1}{4} V_A )</td>
<td>( t = \frac{1}{4} \frac{V_A}{V_p} )</td>
</tr>
<tr>
<td>Full sibs</td>
<td>( \frac{1}{2} V_A + \frac{1}{4} V_D + V_E_c )</td>
<td>( t = \frac{\frac{1}{2} V_A + \frac{1}{4} V_D + V_E_c}{V_p} )</td>
</tr>
</tbody>
</table>

Why \( \text{cov(within)} = \text{variance(among)} \)?

- 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 \( \text{Cov}(z_{ij}, z_{ik}) \)
  = \( \text{Cov}(u + g_i + e_{ij}, u + g_i + e_{ik}) \)
  = \( \text{Cov}(g_i, g_i) \) as all other terms are uncorrelated
  - \( \text{Cov}(g_i, g_i) = \text{Var}(g) \) is the among-group variance
**Estimation: One-way ANOVA**

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

\[
z_{ij} = m + f_i + w_{ij}
\]

- Trait value in sib j from family i
- Common mean
- Deviation of sib j from the family mean
- 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**

\[
\text{Cov(Full Sibs)} = \sigma(z_{ij}, z_{ik}) = \sigma[(\mu + f_i + w_{ij}),(\mu + f_i + w_{ik})]
\]

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

\[
= \sigma_f^2
\]

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

\[
\sigma_f^2 = \frac{\sigma_A^2}{2} + \frac{\sigma_D^2}{4} + \sigma_{Ec}^2
\]
The within-family variance $\sigma_w^2 = \sigma_p^2 - \sigma_f^2$,

$$
\sigma_{w(FS)}^2 = \sigma_p^2 - (\sigma_A^2 / 2 + \sigma_D^2 / 4 + \sigma_{Ec}^2)
= \sigma_A^2 + \sigma_D^2 + \sigma_E^2 - (\sigma_A^2 / 2 + \sigma_D^2 / 4 + \sigma_{Ec}^2)
= (1 / 2)\sigma_A^2 + (3 / 4)\sigma_D^2 + \sigma_E^2 - \sigma_{Ec}^2
$$

---

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

<table>
<thead>
<tr>
<th>Factor</th>
<th>Degrees of freedom, df</th>
<th>Sum of squares (SS)</th>
<th>Mean squares (MS)</th>
<th>E[MS]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among family</td>
<td>N-1</td>
<td>$SS_f = n \sum_{i=1}^{N} (z_i - \bar{z})^2$</td>
<td>$SS_f/(N-1)$</td>
<td>$\sigma_w^2 + n \sigma_f^2$</td>
</tr>
<tr>
<td>Within family</td>
<td>T-N</td>
<td>$SS_w = \sum_{i=1}^{N} \sum_{j=1}^{n} (z_{ij} - \bar{z}_i)^2$</td>
<td>$SS_w/(T-N)$</td>
<td>$\sigma_w^2$</td>
</tr>
</tbody>
</table>
Appendix: Calculating $E(\text{MS})$

**Model:**
\[ z_{ij} = m + f_i + w_{ij} \quad \text{with} \]
\[ \begin{align*}
  m & \text{ fixed } \rightarrow E[m] = m, \ E[m^2] = m^2, \ \text{Var}[m] = 0 \\
  f_i & \sim \text{N}(0, \sigma_f^2) \rightarrow E[f_i] = 0, \ E[f_i^2] = \sigma_f^2 \\
  w_{ij} & \sim \text{N}(0, \sigma_w^2) \rightarrow E[w_{ij}] = 0, \ E[w_{ij}^2] = \sigma_w^2 \\
  \text{Cov}[f_i, f_j] = \text{Cov}[f_i, w_{ij}] = \text{Cov}[w_{ij}, f_{i'}, f_{j'}] = 0
\end{align*} \]

**Sum of Squares:**
\[ SS_t = n \sum_{i=1}^{N} (Z_i - \bar{Z})^2 = \frac{1}{n} \sum_{i=1}^{N} Z_i^2 - \frac{1}{T} \sum_{i=1}^{N} Z_i \]
\[ SS_w = \sum_{i=1}^{N} \sum_{j=1}^{n} (Z_{ij} - \bar{Z}_i)^2 = \sum_{i=1}^{N} \sum_{j=1}^{n} Z_{ij}^2 - \frac{1}{n} \sum_{i=1}^{N} Z_i^2 \]

**Key Expectations:**
\[ \begin{align*}
  E \left[ \sum_{i=1}^{N} \sum_{j=1}^{n} Z_{ij}^2 \right] &= \sum_{i=1}^{N} \sum_{j=1}^{n} E[Z_{ij}]^2 = \sum_{i=1}^{N} \sum_{j=1}^{n} E[m + f_i + w_{ij}]^2 \\
  &= \sum_{i=1}^{N} \sum_{j=1}^{n} E[m^2 + f_i^2 + w_{ij}^2 + 2mf_i + 2mw_{ij} + 2f_iw_{ij}] \\
  &= \sum_{i=1}^{N} \sum_{j=1}^{n} \left( m^2 + E[f_i^2] + E[w_{ij}^2] + 2mE[f_i] + 2mE[w_{ij}] + 2E[f_i]E[w_{ij}] \right) \\
  &= \sum_{i=1}^{N} \sum_{j=1}^{n} \left( m^2 + \sigma_f^2 + \sigma_w^2 \right) \\
  &= Tm^2 + T\sigma_f^2 + T\sigma_w^2
\end{align*} \]
\[
E \left( \frac{1}{T} z_i^2 \right) = \frac{1}{T} E \left[ \left( \sum_{i=1}^{N} \sum_{j=1}^{n} z_{ij} \right)^2 \right] = \frac{1}{T} E \left[ \left( \sum_{i=1}^{N} \sum_{j=1}^{n} (m + f_i + w_{ij}) \right)^2 \right]
\]

\[
= \frac{1}{T} E \left[ \left( Tm + n \sum_{i=1}^{N} f_i + \sum_{i=1}^{N} \sum_{j=1}^{n} w_{ij} \right)^2 \right]
\]

\[
= \frac{1}{T} E \left[ T^2 m^2 + n^2 \left( \sum_{i=1}^{N} f_i \right)^2 + \left( \sum_{i=1}^{N} \sum_{j=1}^{n} w_{ij} \right)^2 + DPs \right]
\]

\[
= \frac{1}{T} (T^2 m^2 + n^2 N \sigma_f^2 + T \sigma_w^2 + 0)
\]

\[
= Tm^2 + n \sigma_f^2 + \sigma_w^2
\]

\[
E \left[ \frac{1}{n} \sum_{i=1}^{N} z_i^2 \right] = \frac{1}{n} \sum_{i=1}^{N} E[z_i^2] = \frac{1}{n} \sum_{i=1}^{N} E \left[ \left( \sum_{j=1}^{n} z_{ij} \right)^2 \right]
\]

\[
= \frac{1}{n} \sum_{i=1}^{N} E \left[ \left( nm + n f_i + \sum_{j=1}^{n} z_{ij} \right)^2 \right]
\]

\[
= \frac{1}{n} \sum_{i=1}^{N} E \left[ \left( n^2 m^2 + n^2 f_i^2 + \left( \sum_{j=1}^{n} z_{ij} \right)^2 + DPs \right) \right]
\]

\[
= \frac{1}{n} \sum_{i=1}^{N} (n^2 m^2 + n^2 \sigma_f^2 + n \sigma_w^2 + 0)
\]

\[
= Tm^2 + T \sigma_f^2 + N \sigma_w^2
\]
Expected MS

\[
E[MS_i] = \frac{1}{N-1} E[SS_i] = \frac{1}{N-1} E \left[ \frac{1}{n} \sum_{j=1}^{N} z_{ij}^2 - \frac{1}{T} z_{ij}^2 \right]
\]

\[= \frac{1}{N-1} \left[ (Tm^2 + T\sigma_t^2 + N\sigma_w^2) - (Tm^2 + n\sigma_i^2 + \sigma_w^2) \right]\]

\[= \frac{1}{N-1} \left[ n(N-1)\sigma_i^2 + (N-1)\sigma_w^2 \right] = n\sigma_i^2 + \sigma_w^2
\]

\[
E[MS_w] = \frac{1}{T-N} E[SS_w] = \frac{1}{T-1} E \left[ \sum_{i=1}^{N} \sum_{j=1}^{n} z_{ij}^2 - \frac{1}{n} \sum_{i=1}^{N} z_{i}^2 \right]
\]

\[= \frac{1}{T-N} E \left[ (Tm^2 + T\sigma_t^2 + T\sigma_w^2) - (Tm^2 + T\sigma_i^2 + N\sigma_w^2) \right]
\]

\[= \frac{1}{T-N} (T-N)\sigma_w^2 = \sigma_w^2
\]

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

\[ \sigma^2(MS_x) = \frac{2(MS_x)^2}{df_x + 2} \]

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

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

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

Estimating heritability

\[ t_{FS} = \frac{\text{Var}(f)}{\text{Var}(z)} = \frac{1}{2} h^2 + \frac{\sigma_D^2}{4 + \sigma_{Ec}^2} \frac{\sigma_z^2}{\sigma_w^2} \]

Hence, \( h^2 \leq 2 \, t_{FS} \)

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

\[ SE(h^2) \approx 2(1 - t_{FS}) [1 + (n - 1)t_{FS}] \frac{\sqrt{2}}{[Nn(n-1)]} \]
**Worked Example**

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

<table>
<thead>
<tr>
<th>Factor</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>EMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among-families</td>
<td>9</td>
<td>405</td>
<td>45</td>
<td>$\sigma^2_w + 5 \sigma^2_f$</td>
</tr>
<tr>
<td>Within-families</td>
<td>40</td>
<td>800</td>
<td>20</td>
<td>$\sigma^2_w$</td>
</tr>
</tbody>
</table>

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

$Var(w) = MS_w = 20$

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

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

---

**Full sib-half sib design: Nested ANOVA**

[Diagram of full-sib and half-sib design with nested ANOVA structure.]
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

\[ z_{ijk} = m + s_i + d_{ij} + w_{ijk} \]

- Value of the \( k \)th offspring from the \( j \)th dam for sire \( i \)
- Overall mean
- Effect of sire \( i \): deviation of mean of \( i \)'s family from overall mean
- Effect of dam \( j \) of sire \( i \): deviation of mean of dam \( j \) from sire and overall mean
- Within-family deviation of \( k \)th offspring from the mean of the \( ij \)-th family

\( \sigma^2_s \) = between-sire variance = variance in sire family means

\( \sigma^2_d \) = variance among dams within sires = variance of dam means for the same sire

\( \sigma^2_w \) = 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, \( T = NMn \)

<table>
<thead>
<tr>
<th>Factor</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>E[MS]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sires</td>
<td>N-1</td>
<td>SS(_s)</td>
<td>SS(_s)/(N-1)</td>
<td>( \sigma_w^2 + n\sigma_d^2 + Mn\sigma_s^2 )</td>
</tr>
<tr>
<td>Dams(Sires)</td>
<td>N(M-1)</td>
<td>SS(_d)</td>
<td>SS(_d)/[N(M-1)]</td>
<td>( \sigma_w^2 + n\sigma_d^2 )</td>
</tr>
<tr>
<td>Sibs(Dams)</td>
<td>T-NM</td>
<td>SS(_w)</td>
<td>SS(_w)/(T-NM)</td>
<td>( \sigma_w^2 )</td>
</tr>
</tbody>
</table>

where:

\[
SS_s = Mn \sum_{i=1}^{N} (\bar{z}_i - \bar{z})^2
\]

\[
SS_d = n \sum_{i=1}^{N} \sum_{j=1}^{M} (\bar{z}_{ij} - \bar{z}_i)^2
\]

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

---

### 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:

- \( \text{Var(Total)} = \text{Var(between FS families)} + \text{Var(within FS)} \)
  \[
  \rightarrow \sigma_w^2 = \sigma_s^2 - \text{Cov}(FS)
  \]

- \( \text{Var(Sires)} = \text{Cov(Paternal half-sibs)} \)
  \[
  \sigma_d^2 = \sigma_s^2 - \sigma_w^2 = \sigma(\text{FS}) - \sigma(\text{PHS})
  \]
Summarizing:

\[
\begin{align*}
\sigma_s^2 &= \sigma(PHS) \\
\sigma_d^2 &= \sigma_z^2 - \sigma_s^2 - \sigma_w^2 \\
\sigma_w^2 &= \sigma_z^2 - \sigma(FS) \\
\end{align*}
\]

Expressing these in terms of the genetic and environmental variances:

\[
\begin{align*}
\sigma_s^2 &\equiv \frac{\sigma_A^2}{4} \\
\sigma_d^2 &\equiv \frac{\sigma_A^2}{4} + \frac{\sigma_D^2}{4} + \sigma_{Ec}^2 \\
\sigma_w^2 &\equiv \frac{\sigma_A^2}{2} + \frac{3\sigma_D^2}{4} + \sigma_{Es}^2 \\
\end{align*}
\]

Intraclass correlations and estimating heritability

\[
\begin{align*}
t_{PHS} &= \frac{\text{Cov}(PHS)}{\text{Var}(z)} = \frac{\text{Var}(s)}{\text{Var}(z)} \rightarrow 4t_{PHS} = h^2 \\
t_{FS} &= \frac{\text{Cov}(FS)}{\text{Var}(z)} = \frac{\text{Var}(s) + \text{Var}(d)}{\text{Var}(z)} \rightarrow h^2 \leq 2t_{FS} \\
\end{align*}
\]

Note that $4t_{PHS} = 2t_{FS}$ implies no dominance or shared family environmental effects
**Worked Example: N = 10 sires, M = 3 dams, n = 10 sibs/dam**

<table>
<thead>
<tr>
<th>Factor</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>E[MS]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sires</td>
<td>9</td>
<td>4,230</td>
<td>470</td>
<td>(\sigma_w^2 + 10\sigma_d^2 + 30\sigma_s^2)</td>
</tr>
<tr>
<td>Dams(Sires)</td>
<td>20</td>
<td>3,400</td>
<td>170</td>
<td>(\sigma_w^2 + 10\sigma_d^2)</td>
</tr>
<tr>
<td>Within Dams</td>
<td>270</td>
<td>5,400</td>
<td>20</td>
<td>(\sigma_w^2)</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
\sigma_w^2 &= MS_w = 20 \\
\sigma_d^2 &= \frac{MS_s - MS_{we}}{n} = \frac{170 - 20}{10} = 15 \\
\sigma_s^2 &= \frac{MS_d - MS_{sd}}{Nn} = \frac{470 - 170}{30} = 10 \\
\sigma_P^2 &= \sigma_s^2 + \sigma_d^2 + \sigma_w^2 = 45 \\
\sigma_d^2 &= 15 = (1/4)\sigma_A^2 + (1/4)\sigma_D^2 + \sigma_{Ec}^2 \\
&= 10 + (1/4)\sigma_D^2 + \sigma_{Ec}^2 \\
\sigma_A^2 &= 4\sigma_s^2 = 40 \\
\sigma_P^2 &= \frac{\sigma_d^2}{\sigma_P^2} = \frac{40}{45} = 0.89 \\
\sigma_D^2 + 4\sigma_{Ec}^2 &= 20
\end{align*}
\]

**Beetle Example**

Messina and Fry (2003): 24 males each mated to 4 or 5 dams (different for each sire), and 5 female progeny from each dam were measured for two traits, mass eclosion and lifetime fecundity.

**ANOVA for fecundity**

<table>
<thead>
<tr>
<th>Factor</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sires</td>
<td>23</td>
<td>33,983</td>
<td>1,477.5</td>
</tr>
<tr>
<td>Dams(Sires)</td>
<td>86</td>
<td>64,441</td>
<td>749.3</td>
</tr>
<tr>
<td>Sibs(Dams)</td>
<td>439</td>
<td>77,924</td>
<td>177.5</td>
</tr>
</tbody>
</table>
Beetle Example

Expected Mean Squares (EMS)

Sires: \( \sigma_R^2 + n\sigma_D^2 + nq\sigma_S^2 \)
Dams (Sires): \( \sigma_R^2 + n\sigma_D^2 \)
Sibs (Dams): \( \sigma_R^2 \)

Approximately \( n = 5 \) progeny by mating, and an average of \( q = 4.58 \) dams per sire, so:

\[
\begin{align*}
\sigma_R^2 &= 177.5 \\
\sigma_D^2 &= (749.3 - 177.5)/5 = 114.36 \\
\sigma_S^2 &= (1,477.5 - 749.3)/22.9 = 31.80
\end{align*}
\]

• Note: ANOVA method works only with balanced or slightly unbalanced data sets; otherwise ML or REML should be preferred

Estimation of genetic (causal) parameters:

\[
\begin{align*}
\sigma_S^2 &= V_A/4 \\
\sigma_D^2 &= V_A/4 + V_D/4 + V_{Ec} \\
\sigma_R^2 &= V_A/2 + 3V_D/4 + V_{Es}
\end{align*}
\]

For simplicity, assuming \( V_D = 0 \), the following estimates are obtained for the causal components:

\[
\begin{align*}
V_A &= 4\sigma_S^2 = 127.2 \\
V_{Ec} &= \sigma_D^2 - \sigma_S^2 = 82.56 \\
V_{Es} &= \sigma_R^2 - 2\sigma_S^2 = 113.9
\end{align*}
\]

Heritability: \( h^2 = V_A/(\sigma_R^2 + \sigma_D^2 + \sigma_S^2) = 0.393 \)
Parent-offspring Regression

Single parent - offspring regression

\[ z_{o_i} = \mu + b_{olp} (z_{p_i} - \mu) + e_i \]

The expected slope of this regression is:

\[ E(b_{olp}) = \frac{\sigma(z_o, z_p)}{\sigma^2(z_p)} \equiv \frac{(\sigma_A^2 / 2) + \sigma(E_o, E_p)}{\sigma_z^2} = \frac{h^2}{2} + \frac{\sigma(E_o, E_p)}{\sigma_z^2} \]

Residual error variance (spread around expected values)

\[ \sigma^2_e = \left( 1 - \frac{h^2}{2} \right) \sigma_z^2 \]

The expected slope of this regression is:

\[ E(b_{olp}) = \frac{\sigma(z_o, z_p)}{\sigma^2(z_p)} \equiv \frac{(\sigma_A^2 / 2) + \sigma(E_o, E_p)}{\sigma_z^2} = \frac{h^2}{2} + \frac{\sigma(E_o, E_p)}{\sigma_z^2} \]

Shared environmental values

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

\[ z_{oj} = \mu + b_{oMP} \left( \frac{z_{mj} + z_{fj}}{2} - \mu \right) + e_i \]

\[ b_{oMP} = \frac{\text{Cov}[z_o, (z_m + z_f) / 2]}{\text{Var}[(z_m + z_f) / 2]} \]

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

\[ = \frac{2\text{Cov}(z_o, z_p)}{\text{Var}(z)} = 2b_{op} \]

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

\[ \text{Var}(b_{op}) = \frac{n(t - b_{pl}^2) + (1 - t)}{Nn} \]

\( t = \begin{cases} t_{HS} = \frac{h^2}{4} & \text{for half-sibs} \\ t_{FS} = \frac{h^2}{2} + \frac{\sigma_D^2 + \sigma_E^2}{\sigma_z^2} & \text{for full-sibs} \end{cases} \)

\[ \text{Var}(h^2) = \text{Var}(2b_{op}) = 4\text{Var}(b_{op}) \]
Midparent-offspring regression, 
N sets of parents, each with n offspring

\[ Var(h^2) = Var(b_{olMP}) = \frac{2[n(t_{FS} - b_{olMP}^2 / 2) + (1 - t_{FS})]}{Nn} \]

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

\[ Var(h^2) = Var(2b_{olp}) = 4Var(b_{olp}) \]

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,

\[ h_{\text{min}}^2 = (b'_{olMP})^2 \frac{Var_n(z)}{Var_l(A)} \]

- Trait variance in nature
- Additive variance in lab
Why is this a lower bound?

\[
(b_{ol,MP}')^2 \frac{\text{Var}_n(z)}{\text{Var}_i(A)} = \left[ \frac{\text{Cov}_{i,n}(A)}{\text{Var}_n(z)} \right]^2 \frac{\text{Var}_n(z)}{\text{Var}_i(A)} = \gamma^2 h_n^2
\]

where \( \gamma = \frac{\text{Cov}_{i,n}(A)}{\sqrt{\text{Var}_n(A) \text{Var}_i(A)}} \)

is the additive genetic covariance between environments and hence \( \gamma^2 \leq 1 \)

---

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.

\[
z_{ijkl} = G_i + E_j + G_E_{ij} + p_{ijk} + e_{ijkl}
\]

- Genotype \( i \)
- Environment \( j \)
- Effect of plot \( k \) for genotype \( i \) in environment \( j \)
- Interaction between genotype \( i \) and environment \( j \)
- Deviations of individual plants within plots
$z_{ijkl} = G_i + E_j + GE_{ij} + p_{ijk} + e_{ijkl}$

$\sigma^2(z_i) = \sigma_G^2 + \sigma_E^2 + \frac{\sigma_{GE}^2}{e} + \frac{\sigma_p^2}{er} + \frac{\sigma_e^2}{ern}$

\[\begin{aligned}
e &= \text{number of environments} \\
r &= \text{(replicates) number of plots/environment} \\
n &= \text{number of individuals per plot}
\end{aligned}\]

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