

Lecture 7

Estimation of Basic Genetic Parameters

Guilherme J. M. Rosa

University of Wisconsin-Madison

Introduction to Quantitative Genetics

SISG, Seattle

17 - 19 July 2023

Estimation of Basic Genetic Parameters



Notation

Expected Value (Mean), Variance, and Covariance:

$$\left\{ \begin{array}{l} \text{Expected value: } E(X) = \mu_X \\ \text{Variance: } \text{Var}(X) = V_X = \sigma_X^2 \\ \text{Covariance: } \text{Cov}(X, Y) = \sigma_{X,Y} \end{array} \right.$$

Example:

- Variable W , normally distributed with mean μ_W and variance σ_W^2 , i.e. $W \sim N(\mu_W, \sigma_W^2)$

Heritability

Narrow vs. broad sense

Narrow sense: $h^2 = V_A/V_P = \sigma_A^2/\sigma_P^2$

Slope of midparent - offspring regression
(sexual reproduction)

Broad sense: $H^2 = V_G/V_P = \sigma_G^2/\sigma_P^2$

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

$$r_{A,P} = \frac{\sigma_{A,P}}{\sqrt{\sigma_A^2 \sigma_P^2}} = \frac{\sigma_A^2}{\sigma_A \sigma_P} = \frac{\sigma_A}{\sigma_P} = h$$

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

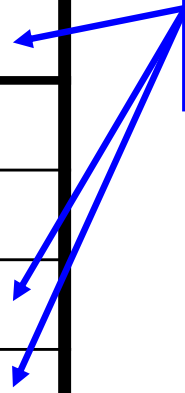
Heritability is a completely unreliable predictor of long-term response

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





Expected Value and Variance



Expected Value (Mean)

Notation: $E[X] = \mu_X$

- Discrete random variable, finite case:

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

If $p_1 = p_2 = \dots = p_k = 1/k$ 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:

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

where $f(x)$: probability density function

Expected Value

- Properties:

Constant c : $E[c] = c$

$$E[cX] = cE[X]$$

$$E[X + Y] = E[X] + E[Y]$$

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

$$E[X] = E[E[X | Y]]$$

Variance

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

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

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

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 : $\text{Var}[c] = 0$

$$\text{Var}[c + X] = \text{Var}[X]$$

$$\text{Var}[cX] = c^2 \text{Var}[X]$$

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2\text{Cov}[X, Y]$$

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

$$\text{Var}[X] = E_Y[\text{Var}[X | Y]] + \text{Var}_Y[E[X | Y]]$$

Covariance

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

$$\begin{aligned}\text{Cov}[X, Y] &= E[(X - \mu_X)(Y - \mu_Y)] \\ &= E[XY] - \mu_X \mu_Y\end{aligned}$$

Correlation

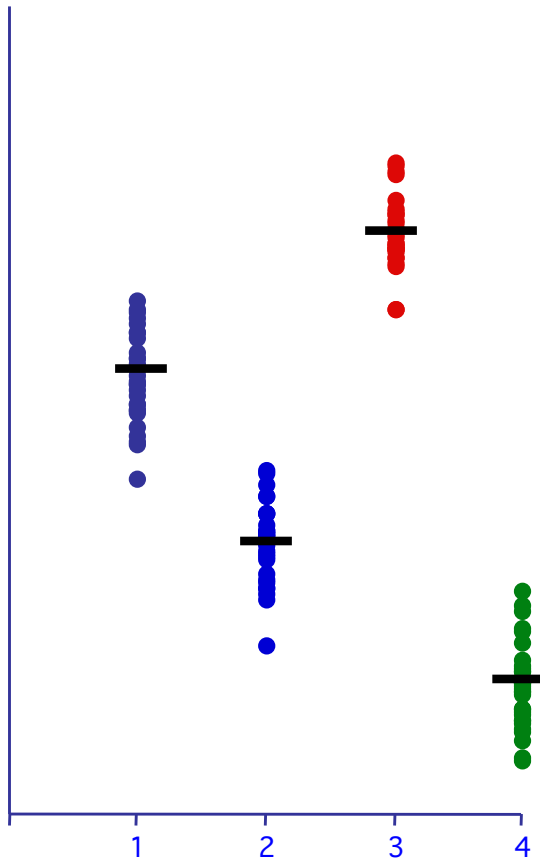
Notation: $\text{Corr}[X, Y] = \rho_{X,Y}$

$$\rho_{X,Y} = \frac{\text{Cov}[X, Y]}{\sigma_X \sigma_Y}$$

ANOVA: Analysis of Variance

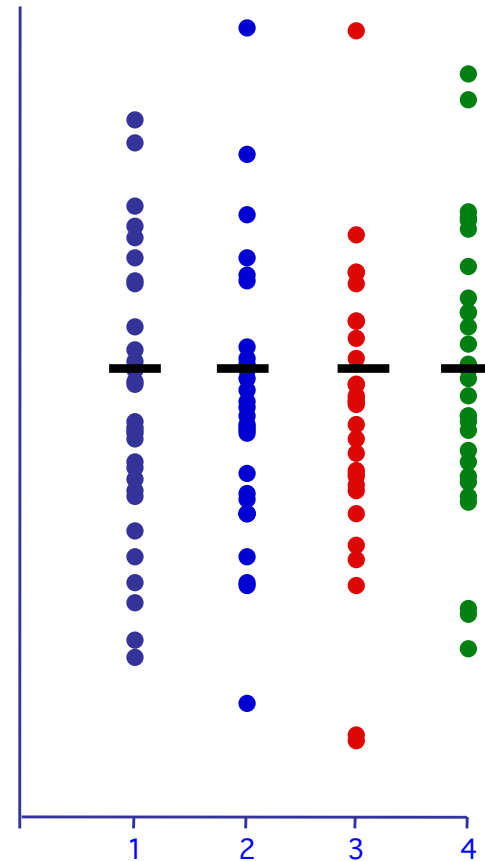
- Partitioning of trait variance into within- and among-group components
- Two key ANOVA identities
 - Total variance = between-group variance + within-group variance
 - $\text{Var}(T) = \text{Var}(B) + \text{Var}(W)$
 - Variance(between groups) = covariance (within groups)
 - Intraclass correlation, $t = \text{Var}(B)/\text{Var}(T)$
- 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



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

Situation 2



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

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

- Let z_{ij} denote the j th 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

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

The diagram illustrates the decomposition of a trait value z_{ij} into three components: a common mean m , a family effect f_i , and a within-family deviation w_{ij} . The equation $z_{ij} = m + f_i + w_{ij}$ is centered. Three arrows point from descriptive text to the terms: 'Common mean' points to m , 'Effect for family i; deviation of mean of i from the common mean' points to f_i , and 'Deviation of sib j from the family mean' points to w_{ij} . Additionally, an arrow points from 'Trait value in sib j from family i' to the entire equation.

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

$$\begin{aligned} \text{Cov}(\text{Full Sibs}) &= \sigma(z_{ij}, z_{ik}) \\ &= \sigma[(\mu + f_i + w_{ij}), (\mu + f_i + w_{ik})] \\ &= \sigma(f_i, f_i) + \sigma(f_i, w_{ik}) + \sigma(w_{ij}, f_i) + \sigma(w_{ij}, w_{ik}) \\ &= \sigma_f^2 \end{aligned}$$

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 $\sigma_w^2 = \sigma_p^2 - \sigma_f^2$,

$$\begin{aligned}\sigma_{w(FS)}^2 &= \sigma_P^2 - (\sigma_A^2 / 2 + \sigma_D^2 / 4 + \sigma_{Ec}^2) \\ &= \sigma_A^2 + \sigma_D^2 + \sigma_E^2 - (\sigma_A^2 / 2 + \sigma_D^2 / 4 + \sigma_{Ec}^2) \\ &= (1 / 2)\sigma_A^2 + (3 / 4)\sigma_D^2 + \sigma_E^2 - \sigma_{Ec}^2\end{aligned}$$

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	$SS_f = n \sum_{i=1}^N (\bar{z}_i - \bar{z})^2$	$SS_f / (N-1)$	$\sigma_w^2 + n \sigma_f^2$
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_w^2

Appendix: Calculating E(MS)

Model:

$$z_{ij} = m + f_i + w_{ij} \quad \text{with} \quad \left\{ \begin{array}{l} m \text{ fixed} \rightarrow E[m] = m, E[m^2] = m^2, \text{Var}[m] = 0 \\ f_i \stackrel{\text{iid}}{\sim} N(0, \sigma_f^2) \rightarrow E[f_i] = 0, E[f_i^2] = \text{Var}[f_i] = \sigma_f^2 \\ w_{ij} \stackrel{\text{iid}}{\sim} N(0, \sigma_w^2) \rightarrow E[w_{ij}] = 0, E[w_{ij}^2] = \text{Var}[w_{ij}] = \sigma_w^2 \\ \text{Cov}[f_i, f_{i'}] = \text{Cov}[f_i, w_{ij}] = \text{Cov}[w_{ij}, w_{i'j'}] = 0 \end{array} \right.$$

Sum of Squares:

$$SS_f = n \sum_{i=1}^N (\bar{z}_i - \bar{z})^2 = \frac{1}{n} \sum_{i=1}^N z_{i\cdot}^2 - \frac{1}{T} z_{\cdot\cdot}^2$$

$$SS_w = \sum_{i=1}^N \sum_{j=1}^n (z_{ij} - \bar{z}_i)^2 = \sum_{i=1}^N \sum_{j=1}^n z_{ij}^2 - \frac{1}{n} \sum_{i=1}^N z_{i\cdot}^2$$

$$z_{i\cdot} = \sum_{j=1}^n z_{ij}$$

$$z_{\cdot\cdot} = \sum_{i=1}^N \sum_{j=1}^n z_{ij}$$

Key Expectations:

$$\mathbb{E} \left[\sum_{i=1}^N \sum_{j=1}^n z_{ij}^2 \right], \quad \mathbb{E} \left[\frac{1}{T} z_{..}^2 \right], \quad \text{and} \quad \mathbb{E} \left[\frac{1}{n} \sum_{i=1}^N z_{i\cdot}^2 \right]$$

$$\begin{aligned} \mathbb{E} \left[\sum_{i=1}^N \sum_{j=1}^n z_{ij}^2 \right] &= \sum_{i=1}^N \sum_{j=1}^n \mathbb{E} [z_{ij}^2] = \sum_{i=1}^N \sum_{j=1}^n \mathbb{E} [m + f_i + w_{ij}]^2 \\ &= \sum_{i=1}^N \sum_{j=1}^n \mathbb{E} [m^2 + f_i^2 + w_{ij}^2 + 2mf_i + 2mw_{ij} + 2f_i w_{ij}] \\ &= \sum_{i=1}^N \sum_{j=1}^n (m^2 + \mathbb{E}[f_i^2] + \mathbb{E}[w_{ij}^2] + 2m\mathbb{E}[f_i] + 2m\mathbb{E}[w_{ij}] + 2\mathbb{E}[f_i]\mathbb{E}[w_{ij}]) \\ &= \sum_{i=1}^N \sum_{j=1}^n (m^2 + \sigma_t^2 + \sigma_w^2) \\ &= Tm^2 + T\sigma_t^2 + T\sigma_w^2 \end{aligned}$$

$$\begin{aligned}
\mathbf{E}\left[\frac{1}{T}z_{\bullet\bullet}^2\right] &= \frac{1}{T}\mathbf{E}\left[\left(\sum_{i=1}^N\sum_{j=1}^nz_{ij}\right)^2\right] = \frac{1}{T}\mathbf{E}\left[\left(\sum_{i=1}^N\sum_{j=1}^n(m+f_i+w_{ij})\right)^2\right] \\
&= \frac{1}{T}\mathbf{E}\left[\left(Tm+n\sum_{i=1}^Nf_i+\sum_{i=1}^N\sum_{j=1}^nw_{ij}\right)^2\right] \\
&= \frac{1}{T}\mathbf{E}\left[T^2m^2+n^2\left(\sum_{i=1}^Nf_i\right)^2+\left(\sum_{i=1}^N\sum_{j=1}^nw_{ij}\right)^2+DPs\right] \\
&= \frac{1}{T}(T^2m^2+n^2N\sigma_f^2+T\sigma_w^2+0) \\
&= Tm^2+n\sigma_f^2+\sigma_w^2
\end{aligned}$$

$$\begin{aligned}
\mathbb{E} \left[\frac{1}{n} \sum_{i=1}^N z_{i\cdot}^2 \right] &= \frac{1}{n} \sum_{i=1}^N \mathbb{E}[z_{i\cdot}^2] = \frac{1}{n} \sum_{i=1}^N \mathbb{E} \left[\left(\sum_{j=1}^n z_{ij} \right)^2 \right] \\
&= \frac{1}{n} \sum_{i=1}^N \mathbb{E} \left[\left(nm + nf_i + \sum_{j=1}^n z_{ij} \right)^2 \right] \\
&= \frac{1}{n} \sum_{i=1}^N \mathbb{E} \left[\left(n^2 m^2 + n^2 f_i^2 + \left(\sum_{j=1}^n z_{ij} \right)^2 + \text{DPs} \right) \right] \\
&= \frac{1}{n} \sum_{i=1}^N (n^2 m^2 + n^2 \sigma_f^2 + n \sigma_w^2 + 0) \\
&= Tm^2 + T\sigma_f^2 + N\sigma_w^2
\end{aligned}$$

Expected MS

$$\begin{aligned} E[MS_f] &= \frac{1}{N-1} E[SS_f] = \frac{1}{N-1} E \left[\frac{1}{n} \sum_{i=1}^N z_{i\cdot}^2 - \frac{1}{T} z_{\cdot\cdot}^2 \right] \\ &= \frac{1}{N-1} \left[(Tm^2 + T\sigma_f^2 + N\sigma_w^2) - (Tm^2 + n\sigma_f^2 + \sigma_w^2) \right] \\ &= \frac{1}{N-1} \left[n(N-1)\sigma_f^2 + (N-1)\sigma_w^2 \right] = n\sigma_f^2 + \sigma_w^2 \end{aligned}$$

$$\begin{aligned} E[MS_w] &= \frac{1}{T-N} E[SS_w] = \frac{1}{T-1} E \left[\sum_{i=1}^N \sum_{j=1}^n z_{ij}^2 - \frac{1}{n} \sum_{i=1}^N z_{i\cdot}^2 \right] \\ &= \frac{1}{T-N} E \left[(Tm^2 + T\sigma_t^2 + T\sigma_w^2) - (Tm^2 + T\sigma_f^2 + N\sigma_w^2) \right] \\ &= \frac{1}{T-N} (T-N)\sigma_w^2 = \sigma_w^2 \end{aligned}$$

Estimating the variance components:

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

$$\text{Var}(w) = MS_w$$

$$\text{Var}(z) = \text{Var}(f) + \text{Var}(w)$$

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

$2\text{Var}(f)$ 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	$\sigma_w^2 + 5 \sigma_f^2$
Within-families	40	$SS_w = 800$	20	σ_w^2

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

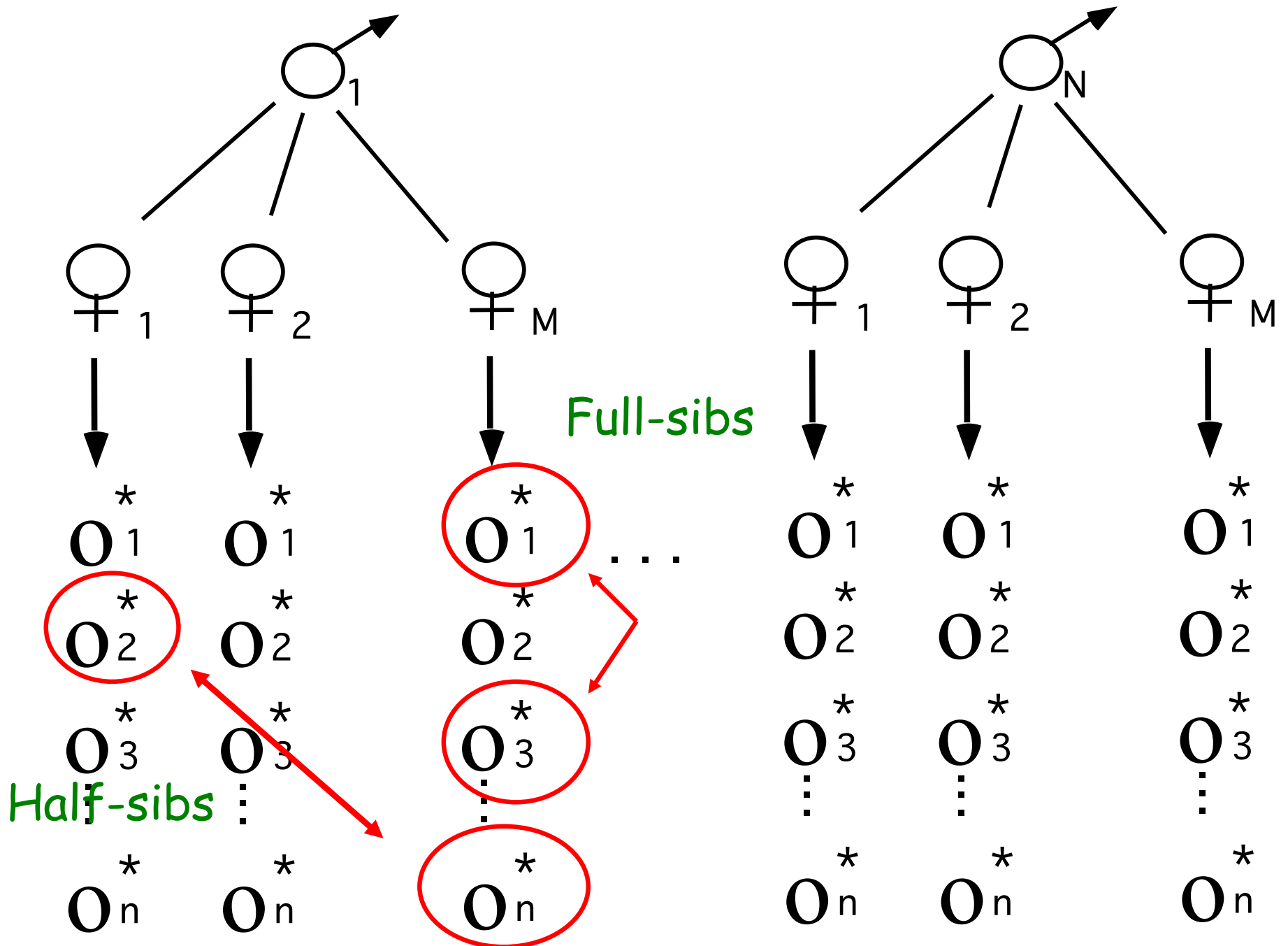
$$Var(w) = MS_w = 20$$

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

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

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

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 k-th offspring from the j-th 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: $SS_s = Mn \sum_{i=1}^N (\bar{z}_i - \bar{z})^2$

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

Estimation of sire, dam, and family variances:

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

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

$$\text{Var}(e) = MS_w$$

Translating these into the desired variance components:

- $\text{Var}(\text{Total}) = \text{Var}(\text{between FS families}) + \text{Var}(\text{within FS})$

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

- $\text{Var}(\text{Sires}) = \text{Cov}(\text{Paternal half-sibs})$

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

Summarizing:

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

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

$$\begin{aligned}\sigma_d^2 &= \sigma_z^2 - \sigma_s^2 - \sigma_w^2 \\ &= \sigma(FS) - \sigma(PHS)\end{aligned}$$

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 $4t_{PHS} = 2t_{FS}$ 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

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

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

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

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

$$\begin{aligned} \sigma_d^2 = 15 &= (1/4)\sigma_A^2 + (1/4)\sigma_D^2 + \sigma_{Ec}^2 \\ &= 10 + (1/4)\sigma_D^2 + \sigma_{Ec}^2 \end{aligned}$$

$$\sigma_A^2 = 4\sigma_s^2 = 40$$

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

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

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

Factor	df	SS	MS
Sires	23	33,983	1,477.5
Dams(Sires)	86	64,441	749.3
Sibs(Dams)	439	77,924	177.5



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):	σ_R^2

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

$$\sigma_R^2 = 177.5$$

$$\sigma_D^2 = (749.3 - 177.5)/5 = 114.36$$

$$\sigma_S^2 = (1,477.5 - 749.3)/22.9 = 31.80$$

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

Beetle Example

Estimation of genetic (causal) parameters:

$$\sigma_S^2 = V_A/4$$

$$\sigma_D^2 = V_A/4 + V_D/4 + V_{Ec}$$

$$\sigma_R^2 = V_A/2 + 3V_D/4 + V_{Es}$$

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

$$V_A = 4\sigma_S^2 = 127.2$$

$$V_{Ec} = \sigma_D^2 - \sigma_S^2 = 82.56$$

$$V_{Es} = \sigma_R^2 - 2\sigma_S^2 = 113.9$$

Heritability: $h^2 = V_A/(\sigma_R^2 + \sigma_D^2 + \sigma_S^2) = 0.393$

Parent-offspring Regression

Single parent - offspring regression

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

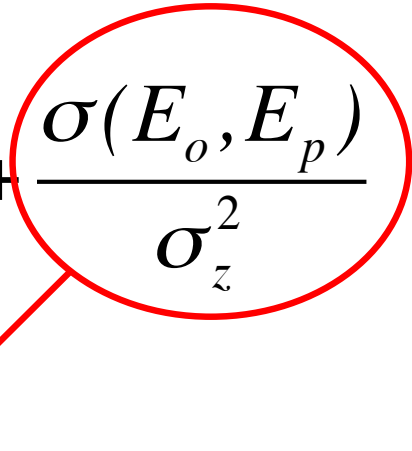
The expected slope of this regression is:

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

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:

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


Shared environmental values

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

Midparent-offspring
regression:

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

$$\begin{aligned} b_{olMP} &= \frac{Cov[z_o, (z_m + z_f) / 2]}{Var[(z_m + z_f) / 2]} \\ &= \frac{[Cov(z_o, z_m) + Cov(z_o, z_f)] / 2}{[Var(z) + Var(z)] / 4} \\ &= \frac{2Cov(z_o, z_p)}{Var(z)} = 2b_{olp} \end{aligned}$$

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

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

Square regression slope

Total number of offspring

Sib correlation

$$t = \begin{cases} t_{HS} = h^2 / 4 & \text{for half-sibs} \\ t_{FS} = h^2 / 2 + \frac{\sigma_D^2 + \sigma_{Ec}^2}{\sigma_z^2} & \text{for full-sibs} \end{cases}$$

$$\text{Var}(h^2) = \text{Var}(2b_{olp}) = 4\text{Var}(b_{olp})$$

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

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

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

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

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,


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

Trait variance in nature

Additive variance in lab

Why is this a lower bound?

Covariance between
breeding value in nature
and BV in lab



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

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

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

The diagram shows the following equation with labels and arrows:

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

Labels and arrows:

- Genotype i (green text, arrow pointing to G_i)
- Environment j (green text, arrow pointing to E_j)
- Effect of plot k for genotype i in environment j (green text, arrow pointing to p_{ijk})
- Interaction between genotype i and environment j (green text, arrow pointing to GE_{ij})
- Deviations of individual plants within plots (green text, arrow pointing to e_{ijkl})

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

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

$\left\{ \begin{array}{l} e = \text{number of environments} \\ r = (\text{replicates}) \text{ number of plots/environment} \\ n = \text{number of individuals per plot} \end{array} \right.$

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

