Basic Designs for Estimation of Genetic Parameters

Bruce Walsh lecture notes
Uppsala EQG 2012 course
version 28 Jan 2012

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 h² instead of h?

Blame Sewall Wright, who used h to denote the correlation between phenotype and breeding value. Hence, h² 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:

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

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
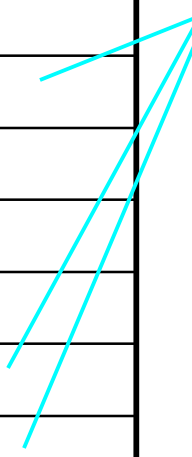
Heritability is a *completely unreliable predictor* *of long-term response*

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

People		h^s
	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	0.3
	Egg production	0.20

Traits more closely associated with fitness tend to have lower heritabilities

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

$$z_{ij} = \mu + f_i + w_{ij}$$

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

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

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 $\sigma^2_w = \sigma^2_P - \sigma^2_f$,

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

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

Factor	Degrees of freedom, df	Sums of Squares (SS)	Mean sum of 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:

$$\mathrm{Var}(f) = \frac{\mathrm{MS}_f - \mathrm{MS}_w}{n}$$

$$\mathrm{Var}(w) = \mathrm{MS}_w$$

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

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

$$\sigma^2(\mathrm{MS}_x) \simeq \frac{2\,(\mathrm{MS}_x)^2}{\mathrm{df}_x + 2}$$

$$\mathrm{Var}\big[\,\mathrm{Var}(w(FS))\,\big] = \mathrm{Var}(\mathrm{MS}_w) \simeq \frac{2(\mathrm{MS}_w)^2}{T - N + 2}$$

$$\mathrm{Var}\big[\,\mathrm{Var}(f)\,\big] = \mathrm{Var}\left[\frac{\mathrm{MS}_f - \mathrm{MS}_w}{n}\right]$$

$$\simeq \frac{2}{n^2}\left(\frac{(\mathrm{MS}_f)^2}{N+1} + \frac{(\mathrm{MS}_w)^2}{T - N + 2}\right)$$

Estimating heritability

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

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

An approximate large-sample standard error
for h² is given by

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

Worked example

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

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

$$\mathrm{Var}(f) = \frac{\mathrm{MS}_f - \mathrm{MS}_w}{n} = \frac{45 - 20}{5} = 5 \longrightarrow V_A < 10$$

$$\mathrm{Var}(w) = \mathrm{MS}_w = 20$$

$$\mathrm{Var}(z) = \mathrm{Var}(f) + \mathrm{Var}(w) = 25$$
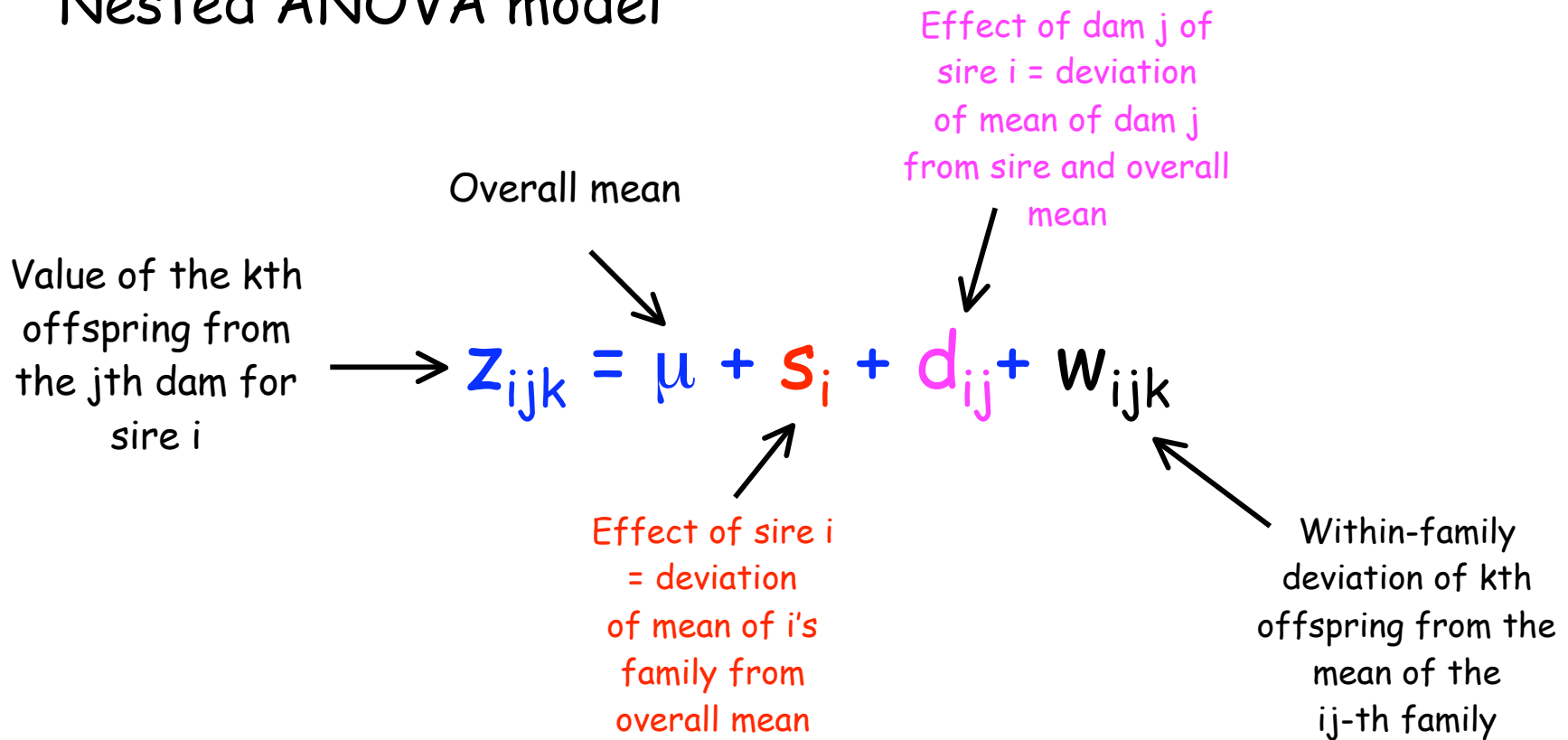
$h^2 < 2\,(5/25) = 0.4$

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

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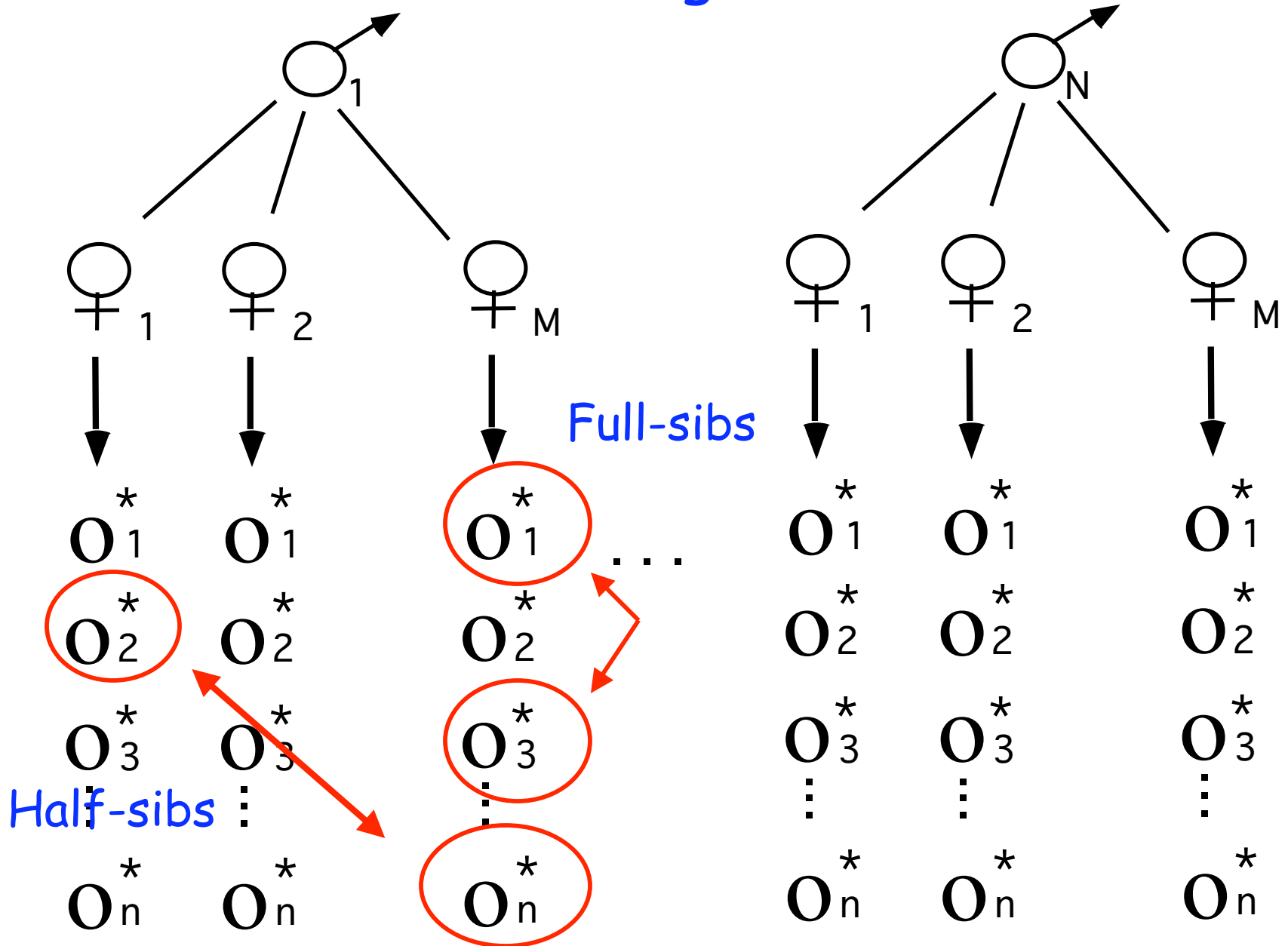
Estimation: Nested ANOVA

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

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

Overall mean

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

$$z_{ijk} = \mu + s_i + d_{ij} + w_{ijk}$$

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:

$$z_{ijk} = \mu + s_i + d_{ij} + w_{ijk}$$

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	EMS
Sires	N-1	$SS_s = Mn \sum_{i=1}^{N} \sum_{j=1}^{M_i} (\bar{z}_i - \bar{z})^2$	$SS_s/(N-1)$	$\sigma_w^2 + n\sigma_d^2 + Mn\sigma_s^2$
Dams(Sires)	N(M-1)	$SS_d = n \sum_{i=1}^{N} \sum_{j=1}^{M} (\bar{z}_{ij} - \bar{z}_i)^2$	$SS_d/(N[M-1])$	$\sigma_w^2 + n\sigma_d^2$
Sibs(Dams)	T-NM	$SS_w = \sum_{i=1}^{N} \sum_{j=1}^{M} \sum_{k=1}^{n} (z_{ijk} - \bar{z}_{ij})^2$	$SS_w/(T-NM)$	σ_w^2

Estimation of sire, dam, and family variances:

$$\mathrm{Var}(s) = \frac{\mathrm{MS}_s - \mathrm{MS}_d}{Mn}$$

$$\mathrm{Var}(d) = \frac{\mathrm{MS}_d - \mathrm{MS}_w}{n}$$

$$\mathrm{Var}(e) = \mathrm{MS}_w$$

Translating these into the desired variance components

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

$$\longrightarrow \quad \sigma^2_w = \sigma^2_z - \mathrm{Cov(FS)}$$

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

$$\sigma^2_d = \sigma^2_z - \sigma^2_s - \sigma^2_w$$
$$= \sigma(\mathrm{FS}) - \sigma(\mathrm{PHS})$$

Summarizing,

$$\sigma_s^2 = \sigma(\text{PHS})$$

$$\sigma_d^2 = \sigma_z^2 - \sigma_s^2 - \sigma_w^2$$

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

$$= \sigma(\text{FS}) - \sigma(\text{PHS})$$

Expressing these in terms of the genetic and environmental variances,

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

$$\sigma_d^2 \simeq \frac{\sigma_A^2}{4} + \frac{\sigma_D^2}{4} + \sigma_{E_c}^2$$

$$\sigma_w^2 \simeq \frac{\sigma_A^2}{2} + \frac{3\sigma_D^2}{4} + \sigma_{E_s}^2$$

Intraclass correlations and estimating heritability

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

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

Note that $4t_{\text{PHS}} = 2t_{\text{FS}}$ implies no dominance or shared family environmental effects

Worked Example:

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

Factor	Df	SS	MS	EMS
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

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

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

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

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

$$\sigma_d^2 = 15 = (1/4)\sigma_A^2 + (1/4)\sigma_D^2 + \sigma_{E_c}^2$$
$$= 10 + (1/4)\sigma_D^2 + \sigma_{E_c}^2$$

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

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

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

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:

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

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:

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

Shared environmental values

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

Midparent - offspring regression

$$z_{oi} = \mu + b_{o|MP} \left(\frac{z_{mi} + z_{fi}}{2} - \mu \right) + e_i$$

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

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

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

The expected slope of this regression is h²

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

Squared regression slope

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

Total number of offspring

Sib correlation

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

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

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

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

- 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,

$$h^2_{min} = (b'_{o|MP})^2 \frac{\mathrm{Var}_n(z)}{\mathrm{Var}_l(A)}$$

←—Trait variance in nature

Additive variance in lab

Why is this a lower bound?
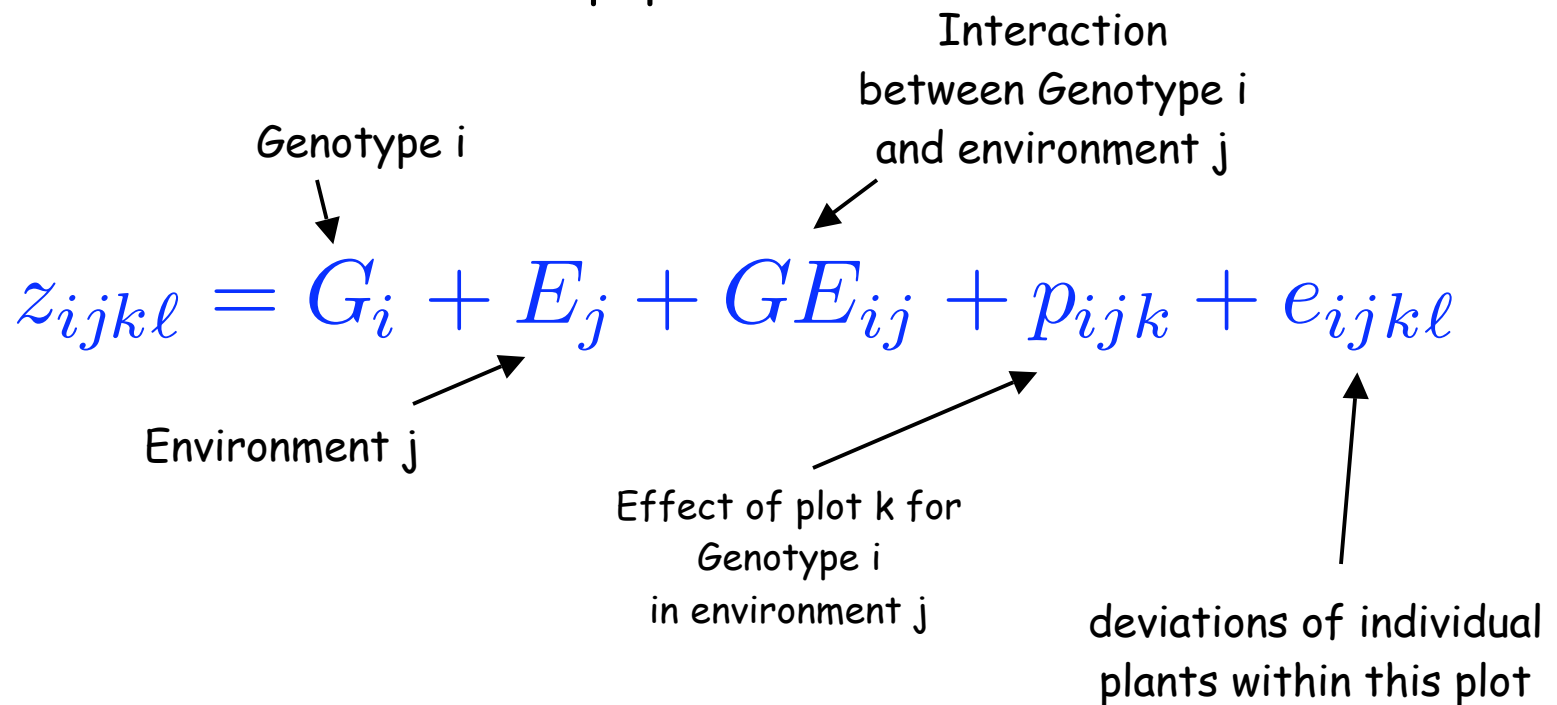
Covariance between
breeding value in nature
and BV in lab

$$(b'_{o|MP})^2 \frac{\text{Var}_n(z)}{\text{Var}_l(A)} = \left[\frac{\text{Cov}_{l,n}(A)}{\text{Var}_n(z)}\right]^2 \frac{\text{Var}_n(z)}{\text{Var}_l(A)} = \boxed{\gamma^2 h_n^2}$$

where
$$\gamma = \frac{\text{Cov}_{l,n}(A)}{\sqrt{\text{Var}_n(A)\text{Var}_l(A)}}$$

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

Defining H² 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 H² even for otherwise identical populations

Genotype i

Interaction
between Genotype i
and environment j

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

Environment j

Effect of plot k for
Genotype i
in environment j

deviations of individual
plants within this plot

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

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

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