Lecture 4

Heritability

Bruce Walsh. July 2005. Asian Institute on Statistical Genetics

Narrow -vs. Broad-sense Heritability

The reason for our focus, indeed obsession, on the heritability is that it determines the degree of resemblance between parents and offspring, which in turn determines the response to selection. In particular, the slope of a midparent-offspring regression is just $h^2 = V_A/V_P$. The fact that the regression involves midparents implies sexual reproduction. In many plant breeding settings, the parent-offspring regression involves offspring that are asexual clones of the single parent. In this case, the parent-offspring regression has slope given by the **broad-sense heritability**, $H^2 = V_G/V_P$.

When we refer to heritability (without making use of either h^2 or H^2), we are by default referring to the **narrow-sense heritability** h^2. Use of the broad-sense heritability H^2 is generally restricted to discussions of clones (such as identical twins or asexual propagates of an individual). While H^2 also gives the total fraction of variation in a trait due to differences in genotypic values, for sexually reproducing species only variation in breeding values is (easily) converted into selection response. Hence, h^2 rather than H^2 is a better measure for sexual species of the fraction of (easily) usable genetic variance.

Why h^2 instead of h?

Students often ask why we use h^2 rather than h to refer to heritability. You can blame Sewall Wright for this, as he used h to denote the correlation between breeding and phenotypic values within an individual. Recalling that the square of the correlation is the total fraction of variation accounted by a variable leads to the universal use of h^2. Indeed, whenever we speak of heritability, we always refer to h^2 and never to h. To see that h is indeed the correlation between breeding and phenotypic values within an individual, note from the definition of a correlation that

$$r_{Ap} = \frac{\sigma(A, P)}{\sigma_A \, \sigma_P} = \frac{\sigma_A^2}{\sigma_A \, \sigma_P} = \frac{\sigma_A}{\sigma_P} = h \tag{4.1}$$

Heritabilities are Functions of a Population

As heritability (in this case, either narrow h^2 or broad H^2 sense) is a function of both the genetic and environmental variances, it is strictly a property of a particular population. Different populations, even if closely related, can have very different heritabilities. Since heritability is a measure of the *standing genetic variation* of a population (either breeding values for h^2 or genotypic values for H^2), a zero heritability DOES NOT mean that a trait is not genetically determined. For example, an inbred line may show consist features that are clearly the result of genetic differences relative to other lines. However, since there is no variation *within* this hypothetical inbred population, both h^2 and H^2 are zero.

Increasing the Heritability

Note that both h^2 and H^2 decrease as the phenotype variance σ_P^2 increases. Hence, reducing the environmental variance (for example, by more careful measurements of a trait or by using a more uniform environment) generally increases the heritability. One note of caution, however. The heritability is not only a genetic property of a population, but also of the distribution (or *universe*) of

Lecture 4, pg. 1

environmental values that the population experiences. Thus, a heritability measured in a laboratory population may be rather different from the same population measured in a natural setting due to a wider range of environments. This is not a serious problem for breeders and experimentalists, *provided* that genotype-environment interactions are small. As the universe of environments change, when significant G x E is present, this can change the genotypic values, and hence can change the genetic variances. This issue is of special concern to plant breeders, where even slightly different growing regions may have subtle, but consistent, differences in their distribution of environmental values.

Heritability and the Prediction of Breeding Values

As mentioned, h^2 is the proportion of the total variance attributable to differences in breeding values. Further, h^2 is the slope of the regression predicting breeding value given an individual's phenotypic value, as

$$A = \frac{\sigma(P, A)}{\sigma_P^2} (P - \mu_p) + e = h^2(P - \mu_p) + e \qquad (4.2a)$$

This follows from the definition of a regression slope and the fact that the regression must pass through the mean of both A and P (0 and μ_p, respectively). The error e in predicting breeding value A from phenotypic value P has mean zero and variance

$$\sigma_e^2 = (1 - h^2)\sigma_A^2 \qquad (4.2b)$$

Hence, the larger the heritability, the tighter the distribution of true breeding values around the value $h^2(P - \mu_p)$ predicted by an individuals' phenotype.

Since heritability is a function of genetic variances, as allele frequencies change (for example, by selection and/or drift), the heritabilities also change. The slope of the parent-offspring regression changes during the course of selection, and as a result, our prediction of the response to selection using some initial estimate of heritability from an unselected population is good for only a few generations.

Heritability Values for Real Traits

Typically heritability values range from 0.1 to 0.6, although higher and lower values are certainly found. One general trend is that traits more closely related to fitness tend to have lower heritabilities.

People	h^2
Height	0.65
Serum IG	0.45
Pigs	
Back-fat thickness	0.70
Daily 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

Heritability Values and Population Divergence

While the heritability of a population provides a measure of its genetic potential to response to a generation of selection, the magnitude of h^2 only provides information on the potential over a few generations. As allele frequencies change, so does heritability. A population showing a high h^2 value may have heritability erode to zero very quickly, while another population with a much smaller h^2 value may actually have heritability increase during selection as rare alleles become more frequent. Hence, *heritability is a completely unreliable predictor for long-term response*, although it is generally a good to excellent predictor of short-term response.

Likewise, measuring heritability values in two populations that show a difference in their means provides no information on whether the underlying difference is genetic — h^2 is only a measure of the *current variation* in each population, it provides no information on the past history of either population. Thus, high estimated h^2 values in two divergent populations does not imply that the divergence is genetic (it could be strictly environmental). Likewise, low estimates of h^2 does not imply that an observed difference between two populations is environmental — both population could have exhausted genetic variation during selection for their divergence. In short, variances within populations and means between populations are not comparable.

Estimation: One-way ANOVA and the simple full-sib design

We now turn to common designs for estimating heritability, starting the collections of sibs. Perhaps the simplest sib design is to examine N full-sib families, each with n offspring. The traditional approach to analyzing such data is the **one-way analysis of variance**, based on the linear model

$$z_{ij} = \mu + f_i + w_{ij} \tag{4.3}$$

where z_{ij} is the phenotype of the jth offspring of the ith family, f_i is the effect of the ith family and w_{ij} is the residual error resulting from segregation, dominance, and environmental contributions. We further assume that the w_{ij} are uncorrelated with each other and have common variance σ_w^2, the **within-family variance** (we will also use w_{FS} and w_{HS} to distinguish between the within-family variance for full- and half-sibs respectively, but for now it is clear that we are simply dealing with a full-sib family). The variance among family effects (the **between-family**, or **among family, variance**) is denoted by σ_f^2.

A basic assumption of linear models underlying ANOVA is that the random factors are uncorrelated with each other. This leads to a key feature:

- *The analysis of variance partitions the total phenotypic variance into the sum of the variances from each of the contributing factors.*

For example, for the full-sib model, the critical assumption is that the residual within-family deviations are uncorrelated with the family effects, i.e., $\sigma(f_i, w_{ij}) = 0$. Thus, the total phenotypic variance equals the variance due to sires plus the residual variance,

$$\sigma_z^2 = \sigma_f^2 + \sigma_{w(FS)}^2 \tag{4.4}$$

The second ANOVA relationship that proves to be very useful is that

- *The phenotypic covariance between members of the same group equals the variance among groups.*

To see this for full-sibs, note that members of the same group (full sibs) share family effects, but have independent residual deviations, so

$$
\begin{aligned}
\text{Cov(Full Sibs)} &= \sigma(\, z_{ij}, z_{ik}\,) \\
&= \sigma[\,(\mu + f_i + w_{ij}), (\mu + f_i + w_{ik})\,] \\
&= \sigma(\, f_i, f_i\,) + \sigma(\, f_i, w_{ik}\,) + \sigma(\, w_{ij}, f_i\,) + \sigma(\, w_{ij}, w_{ik}\,) \\
&= \sigma_f^2
\end{aligned}
\tag{4.5}
$$

The identity Cov(within) = Var(between) allow us to relate an estimated variance component (e.g., the between-family variance σ_f^2) with the casual underlying variance components (e.g., σ_A^2) that are our real interest. For example, the variance among family effects equals the covariance between full sibs,

$$\sigma_f^2 = \sigma_A^2/2 + \sigma_D^2/4 + \sigma_{Ec}^2 \tag{4.6a}$$

where E_c is the common (or shared) family environmental effects (such as shared maternal effects) Likewise, since $\sigma_P^2 = \sigma_f^2 + \sigma_{w(FS)}^2$, the within-group variance $\sigma_{w(FS)}^2$ (i.e., the variance of full-sib values about their family mean) is

$$
\begin{aligned}
\sigma_{w(FS)}^2 &= \sigma_P^2 - \sigma_f^2 \\
&= \sigma_P^2 - (\,\sigma_A^2/2 + \sigma_D^2/4 + \sigma_{Ec}^2\,) \\
&= \sigma_A^2 + \sigma_D^2 + \sigma_E^2 - (\,\sigma_A^2/2 + \sigma_D^2/4 + \sigma_{Ec}^2\,) \\
&= (1/2)\sigma_A^2 + (3/4)\sigma_D^2 + \sigma_E^2 - \sigma_{Ec}^2
\end{aligned} \tag{4.6b}
$$

The ANOVA table for a balanced full-sib design becomes:

Table 4.1.   ANOVA for a balanced full-sib design for N families each with n sibs.

Factor	df	SS	MS	E(MS)
Among-families	$N-1$	$\mathrm{SS}_f = n \sum\limits_{i=1}^{N} (\bar{z}_i - \bar{z})^2$	$\mathrm{MS}_f = \mathrm{SS}_f/(N-1)$	$\sigma_{w(FS)}^2 + n\sigma_f^2$
Within-families	$T-N$	$\mathrm{SS}_w = \sum\limits_{i=1}^{N}\sum\limits_{j=1}^{n} (z_{ij} - \bar{z}_i)^2$	$\mathrm{MS}_w = \mathrm{SS}_w/(T-N)$	$\sigma_{w(FS)}^2$

Note: The total sample size is $T = Nn$. Degrees of freedom are denoted by df, observed sums of squares by SS, and expected mean squares by E(MS).

Estimating Variances and Variance Components

Unbiased estimators of σ_f^2, $\sigma_{w(FS)}^2$, and σ_z^2 follow from the expected mean squares

$$\mathrm{Var}(f) = \frac{\mathrm{MS}_f - \mathrm{MS}_w}{n} \tag{4.7a}$$

$$\mathrm{Var}(w) = \mathrm{MS}_w \tag{4.7b}$$

$$\mathrm{Var}(z) = \mathrm{Var}(f) + \mathrm{Var}(w) \tag{4.7c}$$

Recalling Equation 4.6,

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

so that $2\sigma_f^2$ provides an *upper bound* on σ_A^2.

Standard errors for the variance estimators given by Equation 4.7a-4.7c follow (under the assumptions of normality and balanced design) since the observed mean squares extracted from an analysis of variance are distributed independently with expected sampling variance

$$\sigma^2(\mathrm{MS}_x) \simeq \frac{2(\mathrm{MS}_x)^2}{\mathrm{df}_x + 2} \tag{4.8}$$

Since Equations 4.7a–4.7c are linear functions of the observed mean squares, the rules for obtaining variances and covariances of linear functions (Lecture 3) can be used in conjunction with Equation 4.8 to obtain the large-sample approximations

$$\text{Var}[\,\text{Var}(w(FS))\,] = \text{Var}(\text{MS}_w) \simeq \frac{2(\text{MS}_w)^2}{T - N + 2} \tag{4.9a}$$

$$\text{Var}[\,\text{Var}(f)\,] = \text{Var}\left[\frac{\text{MS}_f - \text{MS}_w}{n}\right] \simeq \frac{2}{n^2}\left(\frac{(\text{MS}_f)^2}{N + 1} + \frac{(\text{MS}_w)^2}{T - N + 2}\right) \tag{4.9b}$$

Estimating heritability

Since the intraclass correlation for full-sibs is given by

$$t_{\text{FS}} = \frac{\text{Var}(f)}{\text{Var}(z)} = \frac{1}{2}h^2 + \frac{\sigma_D^2/4 + \sigma_{E_c}^2}{\sigma_z^2}, \tag{4.10a}$$

an upper bound for the estimate of heritability is given by

$$h^2 \simeq 2t_{\text{FS}} \tag{4.10b}$$

This has a (large-sample) standard error of

$$\text{SE}(h^2) \simeq 2(1 - t_{\text{FS}})[1 + (n - 1)t_{\text{FS}}]\sqrt{2/[Nn(n - 1)]} \tag{4.10c}$$

Worked Example of a Full-sib Design

Table 4.2. Suppose $N = 10$ full-sib families each with $n = 5$ offspring are measured.

Factor	df	SS	MS	$E(\text{MS})$
Among-families	9	$\text{SS}_f = 405$	45	$\sigma_w^2 + 5\sigma_f^2$
Within-families	40	$\text{SS}_w = 800$	20	σ_w^2

$$\text{Var}(f) = \frac{\text{MS}_f - \text{MS}_w}{n} = \frac{45 - 20}{5} = 5, \quad \text{Var}(w) = \text{MS}_w = 20, \quad \text{Var}(z) = \text{Var}(f) + \text{Var}(w) = 25$$

Hence, an upper bound for the additive variance is $Var(A) = 2Var(f) = 10$. Likewise, the estimated heritability (assuming dominance and shared environmental effects can be ignored) is

$$2t_{FS} = 2\frac{5}{25} = 0.4$$

with

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

illustrating the (usually) large standard errors on heritability estimates.

Estimation: The Nested Full-sib, Half-sib Analysis

The simple full-sib design suffers in that one cannot obtain a clean estimate of σ_A^2. A more efficient design is the **nested full-sib, half-sib design**, wherein each male (or **sire**) is mated to several unrelated females (or **dams**), generating a series of full-sib families nested within half-sibs.
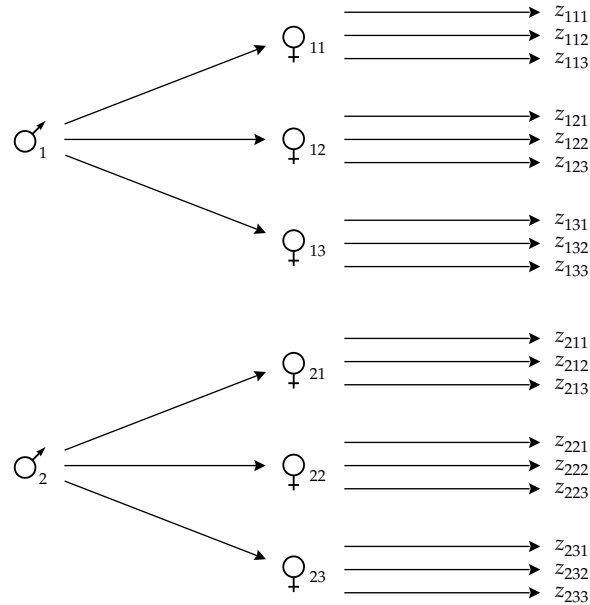

Figure 4.1 A nested full-sib, half-sib mating design. Each male is mated to several unique (unrelated) females, from each of which several offspring are assayed.

The linear model for this **nested design** is

$$z_{ijk} = \mu + s_i + d_{ij} + w_{ijk} \tag{4.11a}$$

where z_{ijk} is the phenotype of the kth offspring from the family of the ith sire and jth dam, s_i is the effect of the ith sire, d_{ij} is the effect of the jth dam mated to the ith sire, and w_{ijk} is the residual deviation (the within-full-sib family deviations). As usual, under the assumption that individuals are random members of the same population, the s_i, d_{ij}, and w_{ijk} are defined to be independent random variables with expectations (mean values) equal to zero. Because of indepedence, the total phenotypic variance is the sum of individual variances,

$$\sigma_z^2 = \sigma_s^2 + \sigma_d^2 + \sigma_w^2 \tag{4.11b}$$

where σ_s^2 is the variance among sires, σ_d^2 the variance among dams within sires, and σ_w^2 the variance within full-sib families.

To relate the observable components of variance (Equation 4.11b) to covariances between relatives, first note that the total phenotypic variance can be partitioned into two components, the variance within- and among- full-sib families. Since the variance among groups is equivalent to the covariance of members within groups, the variance among full-sib families equals the phenotypic covariance of full sibs, $\sigma(\text{FS})$. Thus, the variance within full-sib families (the residual variance in the model) is simply

$$\sigma_w^2 = \sigma_z^2 - \sigma(\text{FS}) \tag{4.12a}$$

Similarly, the variance among sires is equivalent to the covariance of individuals with the same father but different mothers, i.e., the covariance of paternal half sibs,

$$\sigma_s^2 = \sigma(\text{PHS}) \tag{4.12b}$$

Since the three components of variance must sum to the phenotypic variance σ_z^2, the dam variance is found to be

$$
\begin{aligned}
\sigma_d^2 &= \sigma_z^2 - \sigma_s^2 - \sigma_w^2 \\
&= \sigma(\text{FS}) - \sigma(\text{PHS})
\end{aligned}
\tag{4.12c}
$$

Recalling covariances $\sigma(\text{PHS})$ and $\sigma(\text{FS})$ among half- and full-sibs gives

$$\sigma_s^2 = \frac{\sigma_A^2}{4} \tag{4.13a}$$

$$\sigma_d^2 = \frac{\sigma_A^2}{4} + \frac{\sigma_D^2}{4} + \sigma_{E_c}^2 \tag{4.13b}$$

$$\sigma_w^2 = \frac{\sigma_A^2}{2} + \frac{3\sigma_D^2}{4} + \sigma_{E_s}^2 \tag{4.13c}$$

where $\sigma_{E_c}^2$ is the component of variance due to common family environmental effects, and $\sigma_{E_s}^2$ the remaining environmental variation. An obvious problem with this set of equations is that they are overdetermined — there are four causal sources of variance $(\sigma_A^2, \sigma_D^2, \sigma_{E_c}^2, \sigma_{E_c}^2)$ but only three observable variance components $(\sigma_s^2, \sigma_d^2, \sigma_w^2)$. We will deal with this in the worked example below.

The variance-component estimators are given by,

$$\text{Var}(s) = \frac{\text{MS}_s - \text{MS}_d}{Mn} \tag{4.14a}$$

$$\text{Var}(d) = \frac{\text{MS}_d - \text{MS}_w}{n} \tag{4.14b}$$

$$\text{Var}(e) = \text{MS}_w \tag{4.14c}$$

while the intraclass correlations for paternal half sibs and full sibs are

$$t_{\text{PHS}} = \frac{\text{Cov}(\text{PHS})}{\text{Var}(z)} = \frac{\text{Var}(s)}{\text{Var}(z)} \tag{4.15a}$$

$$t_{\text{FS}} = \frac{\text{Cov}(\text{FS})}{\text{Var}(z)} = \frac{\text{Var}(s) + \text{Var}(d)}{\text{Var}(z)} \tag{4.15b}$$

$4t_{\text{PHS}}$ provides the best estimate of h^2 since it is not inflated by dominance and/or common environmental effects. If, however, $\text{Var}(s)$ and $\text{Var}(d)$ are found to be approximately equal, then dominance and maternal effects can be ruled out as significant causal sources of covariance.

Table 4.3. Summary of a (balanced) nested analysis of variance involving N sires, M dams per sire and and n offspring per dam. $T = MNn$ is the total number of sibs in the design.

Factor	df	Sums of Squares	MS	E(MS)
Sires	$N-1$	$Mn\sum_{i=1}^{N}\sum_{j=1}^{M_i}(\bar{z}_i - \bar{z})^2$	SS_s/df_s	$\sigma_w^2 + n\sigma_d^2 + Mn\sigma_s^2$
Dams (sires)	$N(M-1)$	$n\sum_{i=1}^{N}\sum_{j=1}^{M}(\bar{z}_{ij} - \bar{z}_i)^2$	SS_d/df_d	$\sigma_w^2 + n\sigma_d^2$
Sibs (dams)	$T-NM$	$\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{k=1}^{n}(z_{ijk} - \bar{z}_{ij})^2$	SS_w/df_e	σ_w^2

Worked Example of a Nested Design

Table 4.4. Suppose $N = 10$ sires are each crossed to $M = 3$ dams and $n = 10$ offspring are measured in each full-sib family, with resulting ANOVA table

Factor	df	SS	MS	E(MS)
Sires	9	$\text{SS}_s = 4,230$	470	$\sigma_w^2 + 10\sigma_d^2 + 30\sigma_s^2$
Dams (sires)	20	$\text{SS}_d = 3,400$	170	$\sigma_w^2 + 10\sigma_d^2$
Within Dams	270	$\text{SS}_w = 5,400$	20	σ_w^2

$$\sigma_w^2 = MS_w = 20$$
$$\sigma_d^2 = \frac{MS_d - MS_w}{n} = \frac{170 - 20}{10} = 15$$
$$\sigma_s^2 = \frac{MS_s - MS_d}{Nn} = \frac{470 - 170}{30} = 10$$
$$\sigma_P^2 = \sigma_s^2 + \sigma_d^2 + \sigma_w^2 = 45$$

Hence

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

and

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

Likewise, since

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

we are left with the estimate of the linear combination

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

Hence, if $\sigma_D^2 = 0$, then $\sigma_{E_c}^2 = 5$, while if $\sigma_{E_c}^2 = 0$, then $\sigma_D^2 = 20$. These represent the extreme values of these two variance components consistent with the ANOVA.

Estimation: Parent-Offspring Regressions

In some sense the simplest design is the parent-offspring regression, the regression of offspring phenotype z_o on the phenotypic value of one of its parents, z_p. Here the linear model is

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

The second expression follows from the first since the regression passes through the mean of both variables (offspring and parental phenotypes).

The expected regression slope $b_{o|p}$ is

$$E(b_{o|p}) = \frac{\sigma(z_o, z_p)}{\sigma^2(z_p)} \simeq \frac{(\sigma_A^2/2) + \sigma(E_o, E_p)}{\sigma_z^2} \tag{4.16}$$

For males, it is generally expected that the covariance between parent and offspring environmental values is zero and the regression slope is $h^2/2$. This is not necessarily the case for females, as one can imagine how a larger female could better provision her offspring, leading to larger offspring, creating a positive environmental covariance. For this reason, single parent-offspring regressions usually involve fathers, although if the regression slopes for father-offspring and mother-offspring are the same, we can rule out shared mother-offspring environmental values. Thus a simple (possibly biased) estimate of $h^2 = \sigma_A^2/\sigma_z^2$ is twice the (single) parent-offspring regression, $2b_{o|p}$.

Greater precision is possible when both parents can be measured, in which case one can regress offspring phenotypes on the mean phenotypes of their parents (also known as the **midparent values**). The linear model is now

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

where z_{mi} and z_{fi} refer to the phenotypes of mothers and fathers. The slope $b_{o|MP}$ is a direct estimate of the heritability. To see this, note that

$$
\begin{aligned}
b_{o|MP} &= \frac{\text{Cov}[z_o, (z_m + z_f)/2]}{\text{Var}[(z_m + z_f)/2]} \\
&= \frac{[\text{Cov}(z_o, z_m) + \text{Cov}(z_o, z_f)]/2}{[\text{Var}(z) + \text{Var}(z)]/4} \\
&= \frac{2\text{Cov}(z_o, z_p)}{\text{Var}(z)} = 2b_{o|p} \tag{4.17}
\end{aligned}
$$

What happens when multiple (n) offspring are measured in each family? The expected phenotypic covariance of a parent i and the average of its $j = 1, \cdots, n$ offspring may be written $\sigma[(\sum_{j=1}^{n} z_{oij}/n), z_p]$. Since all n of the covariance terms have the same expected value, this reduces to $n\sigma(z_o, z_p)/n = \sigma(z_o, z_p)$, the same as the expectation for single offspring. Thus, provided family sizes are equal, the interpretation of a single parent-offspring regression is the same whether individual offspring data or the progeny means are used in the analysis.

The sampling variance of the regression of a single parent on its (n) offspring is approximately

$$\text{Var}(b_{o|p}) \simeq \frac{n(t - b_{o|p}^2) + (1 - t)}{Nn} \tag{4.18a}$$

where N is the number of parent-offspring pairs and t is the covariance between sibs. Since sibs in a single-parent regression can potentially be either full- or half-sibs,

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

Since h^2 is estimated as $2b_{o|p}$,

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

Likewise, the sampling variance for a midparent-offspring regression with N parent-offspring pairs and n offspring per set of parents is approximately

$$\text{Var}(b_{o|MP}) \simeq \frac{2[n(t_{FS} - b_{o|MP}^2/2) + (1 - t_{FS})])}{Nn} \qquad (4.18b)$$

Since h^2 is estimated by $b_{o|MP}$, $\text{Var}(h^2)$ is given by Equation 4.18b.

Estimating Heritability in Natural Populations from Lab-reared Offspring

A lower bound, h_{min}^2, to the heritability in the field can be estimated by regressing the phenotypes of lab-reared progeny on their field-reared parents. Let the regression coefficient involving wild midparents and lab-reared offspring be $b'_{o|MP}$, the phenotypic variance of the natural population be $\text{Var}_n(z)$, and the additive genetic variance in the laboratory environment (obtained either from the covariance of lab-reared sibs or of lab-reared parents and offspring) be $\text{Var}_l(A)$. Then,

$$h_{min}^2 = (b'_{o|MP})^2 \frac{\text{Var}_n(z)}{\text{Var}_l(A)} = \left[\frac{\text{Cov}_{l,n}(A)}{\text{Var}_n(z)}\right]^2 \frac{\text{Var}_n(z)}{\text{Var}_l(A)} \qquad (4.19a)$$

where $\text{Cov}_{l,n}(A)$ is the additive genetic covariance between the trait as expressed in the wild and in the lab. To see that this provides a lower bound, define

$$\gamma = \frac{\text{Cov}_{l,n}(A)}{\sqrt{\text{Var}_n(A)\text{Var}_l(A)}} \qquad (4.19b)$$

to be the additive genetic correlation between environments. The expected value of h_{min}^2 is then $\gamma^2 h_n^2$, which is necessarily $\leq h_n^2$ (as $\gamma^2 \leq 1$), the heritability in nature. h_{min}^2 is an unbiased estimate of h_n^2 only if the genetic correlation across environments is equal to one, i.e., there is no $G \times E$.

Estimating Variances Under General Pedigree Structures

All of the above designs are special cases of local analysis on just parts of a fuller pedigree that an investigator might have in hand. A general solution is offered by the method of **Restricted Maximum Likelihood**, or **REML** to estimate variances, and the associated method of **best linear unbiased predictor**, or **BLUP** to estimate breeding values. These approaches allow us to simultaneously exploit all the information in a complex pedigree as well as estimate other fixed factors such as effects of sex, age, etc. These methods, while fairly straightforward, require a fair amount of background information to be presented. In a sentence, they use a general linear model with a variance-covariance matrix specified by the matrix of known relationships. The interested reader is referred to Lynch and Walsh for a full treatment.

Defining H^2 for in Plant Populations

Our last point about heritability deals with how plant breeders define the board-sense heritability H^2. Our key point is that identical populations may have different H^2 values, depending on the unit of analysis chosen by the investigator.

In plant breeding, pure lines are often used, and instead of measuring individuals directly for trait values (such as yield), one often measures a block or plot of individuals as the sampling

unit. Suppose our design is to measure each line in r plots, each consisting of n individuals, over e environments. The resulting linear model for the ℓth individual of i genotype in plot k in environment j

$$z_{ijk\ell} = G_i + E_j + G \times E_{ij} + p_{ijk} + e_{ijk\ell} \tag{4.20}$$

where p_{ijk} is the plot effect for the kth replicate of the plot for genotype i in environment j, and $e_{ijk\ell}$ the residual value for the ℓth individual. If we simply take $z_i = z_{i\dots}$ (the average value of genotype i over all plots, environments, and individuals) as our measurement then σ_G^2 is unchanged, but the phenotypic variances of the z_i becomes

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

here σ_p^2 is the between-plot environmental variance and σ_e^2 the within-plot individual variance. The problem with defining H^2 in a consistent fashion is that different investigators may chose different values of r, e and n to measure the trait, and hence (even with identical variance components) get different values for the phenotypic variance, and hence for H^2.

Heritability Problems

1. Consider a simple full-sib design. Suppose $\sigma_A^2 = 30$, $\sigma_D^2 = 10$, $\sigma_{Ec}^2 = 5$ and $\sigma_E^2 - \sigma_{Ec}^2 = 5$.

    a: What are σ_f^2? σ_w^2?

    b: Now consider a nested full-sib design. What are the sire (σ_s^2) and dam (σ_d^2) variances?

2. Recalling Equation 4.9b, what is the variance of our estimate of σ_A^2 for the worked full-sib problem?

3. Create your own ANOVA! Consider a strictly half-sib analysis, wherein each of N sires are mated to n dams, each of which leaves exactly one offspring (an example of this is beef or dairy cattle). Under this model, the ijth observation is the jth offspring from sire i and is the sum of a sire effect s_i and a within (half-sib) family deviation from the sire effect.

    a: What is the linear model for this design?

    b: Express the sire σ_s^2 and within-family $\sigma_{w(HS)}^2$ variance in terms of the genetic and environmental variance components.

    c: What would the resulting ANOVA table look like? (i.e., what are the required sums of squares, the associated degrees of freedom, the mean squares, and the expected value of the mean squares expressed in terms of the genetic and environmental variance components).

Solution to Heritability Problems

1:   a: $\sigma_f^2 = \sigma_A^2/2 + \sigma_D^2/4 + \sigma_{E_c}^2 = 30/2 + 10/4 + 5 = 22.5$ $\sigma_w^2 = \sigma_z^2 - \sigma_f^2$. Here $\sigma_z^2 = 30 + 10 + 5 + 5 = 50$, giving $\sigma_w^2 = 50 - 22.5 = 27.5$

   b: The within-sire variance $\sigma_s^2 = \sigma_A^2/4 = 30/4 = 7.5$, while (Equation 4.12a) $\sigma_d^2 = Cov(FS) - \sigma_s^2 = \sigma_f^2 - \sigma_s^2 = 22.5 - 7.5 = 15$.

2:

$$\text{Var}[\,\text{Var}(A)\,] = \text{Var}[\,2\text{Var}(f)\,] \simeq 2^2 \frac{2}{n^2} \left(\frac{(\text{MS}_f)^2}{N+1} + \frac{(\text{MS}_w)^2}{T-N+2} \right)$$

$$= 4 \cdot \frac{2}{5^2} \left(\frac{(45)^2}{11} + \frac{(20)^2}{42} \right) = 61.96$$

Giving an standard error ($\sqrt{Var}$) of 7.87.

3:

   a: $z_{ij} = \mu + s_i + w_{ij}$

   b: $\sigma_s^2 = \text{Cov(half-sibs)} = \sigma_A^2/4, \quad \sigma_w^2 = \sigma_z^2 - \sigma_s^2 = (3/4)\sigma_A^2 + \sigma_D^2 + \sigma_E^2$

   c:

Factor	df	SS	MS	E(MS)
Between-sires	$N-1$	$\text{SS}_s = n \sum\limits_{i=1}^{N} (\overline{z}_i - \overline{z})^2$	$\text{SS}_s/(N-1)$	$\sigma_w^2 + n\sigma_s^2$
Within-sire families	$T-N$	$\text{SS}_w = \sum\limits_{i=1}^{N} \sum\limits_{j=1}^{n} (z_{ij} - \overline{z}_i)^2$	$\text{SS}_w/(T-N)$	σ_w^2

where $T = Nn$