

# Lecture 4

## Basic Designs for Estimation of Genetic Parameters

Bruce Walsh. Aug 2003. Nordic Summer Course

### 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  $h^2 = V_A/V_P$ .

#### 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 for 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

$$r_{Ap} = \frac{\sigma(A, P)}{\sigma_A \sigma_P} = \frac{\sigma(A, A + D + E)}{\sigma_A \sigma_P} = \frac{\sigma(A, A)}{\sigma_A \sigma_P} = \frac{\sigma_A^2}{\sigma_A \sigma_P} = \frac{\sigma_A}{\sigma_P} = h$$

#### Heritabilities are Functions of a Population

As heritability is a function of both the genetic and environmental variances, it is strictly a property of a 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, a zero heritability DOES NOT mean that a trait is not genetically determined. For example, an inbred line may show consistent features that are clearly the result of genetic differences relative to other lines. However, since there is no variation *within* this hypothetical inbred population,  $h^2$  is zero.

#### Increasing the Heritability

Note that  $h^2$  decreases as the phenotype variance  $\sigma_P^2$  increases. Hence, if one can reduce the environmental variance (for example, by more careful measurements of a trait or by using a more uniform environment) one 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 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 any appropriate genetic variance. 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

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

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  $\mu_p$ , respectively). The error  $e$  in predicting breeding value  $A$  from phenotypic value  $P$  has mean zero and variance

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

Hence, 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.

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 and Population Divergence

While the heritability of a population provides a measure of its genetic potential to respond 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 favorable 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 populations 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. Perhaps the simplest sib-based 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

$$z_{ij} = \mu + f_i + w_{ij} \tag{4.1}$$

where  $z_{ij}$  is the phenotype of the  $j$ th offspring of the  $i$ th family,  $f_i$  is the effect of the  $i$ th 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  $\sigma_w^2$ , 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  $\sigma_f^2$ .

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 differences among family means plus the residual variance of individual family members about their mean,

$$\sigma_z^2 = \sigma_f^2 + \sigma_{w(FS)}^2 \quad (4.2)$$

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

$$\begin{aligned} \text{Cov(Full Sibs)} &= \sigma(z_{ij}, z_{ik}) \\ &= \sigma[(\mu + f_i + w_{ij}), (\mu + f_i + w_{ik})] \\ &= \sigma(f_i, f_i) + \sigma(f_i, w_{ik}) + \sigma(w_{ij}, f_i) + \sigma(w_{ij}, w_{ik}) \\ &= \sigma_f^2 \end{aligned} \quad (4.3)$$

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

$$\sigma_f^2 = \sigma_A^2/2 + \sigma_D^2/4 + \sigma_{E_c}^2 \quad (4.4a)$$

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

$$\begin{aligned} \sigma_{w(FS)}^2 &= \sigma_P^2 - (\sigma_A^2/2 + \sigma_D^2/4 + \sigma_{E_c}^2) \\ &= \sigma_A^2 + \sigma_D^2 + \sigma_E^2 - (\sigma_A^2/2 + \sigma_D^2/4 + \sigma_{E_c}^2) \\ &= (1/2)\sigma_A^2 + (3/4)\sigma_D^2 + \sigma_E^2 - \sigma_{E_c}^2 \end{aligned} \quad (4.4b)$$

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	$E(\text{MS})$
Among-families	$N - 1$	$SS_f = n \sum_{i=1}^N (\bar{z}_i - \bar{z})^2$	$SS_f / (N - 1)$	$\sigma_{w(FS)}^2 + n\sigma_f^2$
Within-families	$T - N$	$SS_w = \sum_{i=1}^N \sum_{j=1}^n (z_{ij} - \bar{z}_i)^2$	$SS_w / (T - N)$	$\sigma_{w(FS)}^2$

*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  $E(\text{MS})$ .

### Estimating Variances and Variance Components

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

$$\text{Var}(f) = \frac{\text{MS}_f - \text{MS}_w}{n} \quad (4.5a)$$

$$\text{Var}(w) = \text{MS}_w \quad (4.5b)$$

$$\text{Var}(z) = \text{Var}(f) + \text{Var}(w) \quad (4.5c)$$

Recalling Equation 4.4,

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

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

Standard errors for the variance estimators given by Equation 4.5a-4.5c 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

$$\sigma^2(\text{MS}_x) \simeq \frac{2(\text{MS}_x)^2}{\text{df}_x + 2} \quad (4.6)$$

Since Equations 4.5a–4.5c 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 4.6 to obtain the large-sample approximations

$$\text{Var}[\text{Var}(w(FS))] = \text{Var}(\text{MS}_w) \simeq \frac{2(\text{MS}_w)^2}{T - N + 2} \quad (4.7a)$$

$$\text{Var}[\text{Var}(f)] = \text{Var}\left[\frac{\text{MS}_f - \text{MS}_w}{n}\right] \simeq \frac{2}{n^2} \left( \frac{(\text{MS}_f)^2}{N + 1} + \frac{(\text{MS}_w)^2}{T - N + 2} \right) \quad (4.7b)$$

### Estimating heritability

Since the intraclass correlation for full-sibs is given by

$$t_{\text{FS}} = \frac{\text{Var}(f)}{\text{Var}(z)} = \frac{1}{2}h^2 + \frac{\sigma_D^2/4 + \sigma_{E_c}^2}{\sigma_z^2}, \quad (4.8a)$$

an upper bound for the estimate of heritability is given by

$$h^2 \simeq 2t_{\text{FS}} \quad (4.8b)$$

This has a (large-sample) standard error of

$$\text{SE}(h^2) \simeq 2(1 - t_{\text{FS}})[1 + (n - 1)t_{\text{FS}}] \sqrt{2/[Nn(n - 1)]} \quad (4.8c)$$

### 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	$\sigma_w^2$

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

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

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

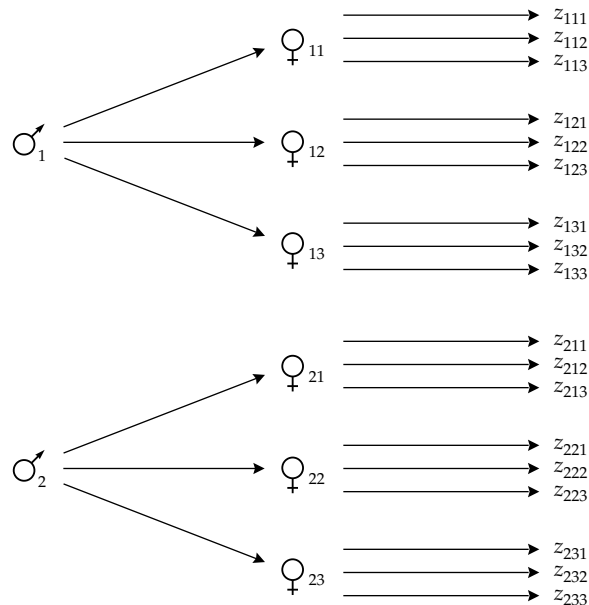
with

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

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  $\sigma_A^2$ . 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 18.3** 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

$$z_{ijk} = \mu + s_i + d_{ij} + w_{ijk} \quad (4.9a)$$

where  $z_{ijk}$  is the phenotype of the  $k$ th offspring from the family of the  $i$ th sire and  $j$ th dam,  $s_i$  is the effect of the  $i$ th sire,  $d_{ij}$  is the effect of the  $j$ th dam mated to the  $i$ th 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 equal to zero. It then follows that the total phenotypic variance is

$$\sigma_z^2 = \sigma_s^2 + \sigma_d^2 + \sigma_w^2 \quad (4.9b)$$

where  $\sigma_s^2$  is the variance among sires,  $\sigma_d^2$  the variance among dams within sires, and  $\sigma_w^2$  the variance within full-sib families.

To relate the observable components of variance 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

$$\sigma_w^2 = \sigma_z^2 - \sigma(\text{FS}) \quad (4.10a)$$

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,

$$\sigma_s^2 = \sigma(\text{PHS}) \quad (4.10b)$$

Since the three components of variance must sum to  $\sigma_z^2$ , the dam variance is found to be

$$\begin{aligned} \sigma_d^2 &= \sigma_z^2 - \sigma_s^2 - \sigma_w^2 \\ &= \sigma(\text{FS}) - \sigma(\text{PHS}) \end{aligned} \quad (4.10c)$$

Recalling the genetic/environmental interpretations of  $\sigma(\text{PHS})$  and  $\sigma(\text{FS})$  gives

$$\sigma_s^2 \simeq \frac{\sigma_A^2}{4} \quad (4.11a)$$

$$\sigma_d^2 \simeq \frac{\sigma_A^2}{4} + \frac{\sigma_D^2}{4} + \sigma_{E_c}^2 \quad (4.11b)$$

$$\sigma_w^2 \simeq \frac{\sigma_A^2}{2} + \frac{3\sigma_D^2}{4} + \sigma_{E_s}^2 \quad (4.11c)$$

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 ( $\sigma_A^2, \sigma_D^2, \sigma_{E_c}^2, \sigma_{E_s}^2$ ) 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,

$$\text{Var}(s) = \frac{\text{MS}_s - \text{MS}_d}{Mn} \quad (4.12a)$$

$$\text{Var}(d) = \frac{\text{MS}_d - \text{MS}_w}{n} \quad (4.12b)$$

$$\text{Var}(e) = \text{MS}_w \quad (4.12c)$$

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

$$t_{\text{PHS}} = \frac{\text{Cov}(\text{PHS})}{\text{Var}(z)} = \frac{\text{Var}(s)}{\text{Var}(z)} \quad (4.13a)$$

$$t_{\text{FS}} = \frac{\text{Cov}(\text{FS})}{\text{Var}(z)} = \frac{\text{Var}(s) + \text{Var}(d)}{\text{Var}(z)} \quad (4.13b)$$

$4t_{\text{PHS}}$  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  $n$  offspring per dam.  $T = MNn$  is the total number of sibs in the design.

Factor	df	Sums of Squares	MS	$E(\text{MS})$
Sires	$N - 1$	$Mn \sum_{i=1}^N \sum_{j=1}^{M_i} (\bar{z}_i - \bar{z})^2$	$\text{SS}_s/\text{df}_s$	$\sigma_w^2 + n\sigma_d^2 + Mn\sigma_s^2$
Dams (sires)	$N(M - 1)$	$n \sum_{i=1}^N \sum_{j=1}^M (\bar{z}_{ij} - \bar{z}_i)^2$	$\text{SS}_d/\text{df}_d$	$\sigma_w^2 + n\sigma_d^2$
Sibs (dams)	$T - NM$	$\sum_{i=1}^N \sum_{j=1}^M \sum_{k=1}^n (z_{ijk} - \bar{z}_{ij})^2$	$\text{SS}_w/\text{df}_e$	$\sigma_w^2$

### 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	$E(\text{MS})$
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	$\sigma_w^2$

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

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  $z_o$  given the phenotypic value of one of its parents,  $z_p$ . Here the linear model is

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

The alternative formulation follows since the regression passes through the mean of both variables (offspring and parental phenotypes).

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

$$E(b_{o|p}) = \frac{\sigma(z_o, z_p)}{\sigma^2(z_p)} \simeq \frac{(\sigma_A^2/2) + \sigma(E_o, E_p)}{\sigma_z^2} \quad (4.14)$$

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 the 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

$$\begin{aligned} b_{o|MP} &= \frac{\text{Cov}[z_o, (z_m + z_f)/2]}{\text{Var}[(z_m + z_f)/2]} \\ &= \frac{[\text{Cov}(z_o, z_m) + \text{Cov}(z_o, z_f)]/2}{[\text{Var}(z) + \text{Var}(z)]/4} \\ &= \frac{2\text{Cov}(z_o, z_p)}{\text{Var}(z)} = 2b_{o|p} \end{aligned} \quad (4.15)$$

What happens when multiple ( $n$ ) offspring are measured in each family? The expected phenotypic covariance of a parent  $i$  and the average of its  $j = 1, \dots, n$  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

$$\text{Var}(b_{o|p}) \simeq \frac{n(t - b_{o|p}^2) + (1 - t)}{Nn} \quad (4.16a)$$

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,

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

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

$$\text{Var}(b_{o|MP}) \simeq \frac{2[n(t_{FS} - b_{o|MP}^2/2) + (1 - t_{FS})]}{Nn} \quad (4.16b)$$

Since  $h^2$  is estimated by  $b_{o|MP}$ ,  $\text{Var}(h^2)$  is given by Equation 4.16b.

## Lecture 4 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 is  $\sigma_f^2$ ?  $\sigma_w^2$ ?
  - b: Now consider a nested full-sib design. What are the sire ( $\sigma_s^2$ ) and dam ( $\sigma_d^2$ ) variances?
2. Recalling Equation 4.7b, what is the variance of our estimate of  $\sigma_A^2$  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  $ij$ th observation is the  $j$ th 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  $\sigma_s^2$  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 Chapter 4 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 4.10a)  $\sigma_d^2 = Cov(FS) - \sigma_s^2 = \sigma_f^2 - \sigma_s^2 = 22.5 - 7.5 = 15$ .

2:

$$\begin{aligned} \text{Var}[\text{Var}(A)] &= \text{Var}[2\text{Var}(f)] \simeq 2^2 \frac{2}{n^2} \left( \frac{(\text{MS}_f)^2}{N+1} + \frac{(\text{MS}_w)^2}{T-N+2} \right) \\ &= 4 \cdot \frac{2}{5^2} \left( \frac{(45)^2}{11} + \frac{(20)^2}{42} \right) = 61.96 \end{aligned}$$

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

3:

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

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

c:

Factor	df	SS	MS	E(MS)
Between-sires	$N - 1$	$SS_s = n \sum_{i=1}^N (\bar{z}_i - \bar{z})^2$	$SS_s / (N - 1)$	$\sigma_w^2 + n\sigma_s^2$
Within-sire families	$T - N$	$SS_w = \sum_{i=1}^N \sum_{j=1}^n (z_{ij} - \bar{z}_i)^2$	$SS_w / (T - N)$	$\sigma_w^2$

where  $T = Nn$