

RESEMBLANCE AMONG RELATIVES

Maternal Genotype	Frequency	Offspring Genotype		
		A_1A_1	A_1A_2	A_2A_2
A_1A_1	f_{11}	p	q	
A_1A_2	f_{12}	$\frac{p}{2}$	$\frac{1}{2}$	$\frac{q}{2}$
A_2A_2	f_{22}		p	q

Note: $f_{11} = p^2$
 $f_{12} = 2pq$
 $f_{22} = q^2$

Note also that the probabilities in the above table are appropriate *only* if the progeny are from half-sib families. If the progeny are from full-sib families, we must specify the frequency of each of the nine possible matings (keeping track of the genotype of both mother and father) and the offspring that each will produce.¹

COVARIANCE OF TWO RANDOM VARIABLES

Let p_{xy} be the probability that random variable X takes the value x and random variable Y takes the value y . Then the covariance between X and Y is:

$$Cov(X, Y) = \sum p_{xy}(x - \mu_x)(y - \mu_y) \quad ,$$

where μ_x is the mean of X and μ_y is the mean of Y .

COVARIANCE BETWEEN HALF-SIBLINGS

Now we can calculate the covariance between half-siblings. To do so we imagine selecting huge number of half-sibs pairs at random. Now the phenotype of the first half-sib in the pair is a random variable (call it S_1), as is the second (call it S_2). The mean of S_1 is just the mean phenotype in *all* the progeny taken together, \bar{x} . Similarly, the mean of S_2 is just \bar{x} . Now with one locus, two alleles we have three possible phenotypes: x_{11} (corresponding to the genotype A_1A_1), x_{12} (corresponding to the genotype A_1A_2), and x_{22} (corresponding to the genotype A_2A_2). So all we need to do to calculate the covariance between half-sibs is to write down all possible pairs of phenotypes and the frequency with which they will occur in our sample based on the table above and the frequency of maternal genotypes. It's actually a bit easier to keep track of it all if we write down the frequency of each maternal genotype and the frequency with which each possible phenotypic combination will occur in

¹ To check your understanding of all of this, you might want to try to produce the appropriate table.

her progeny.

$$\begin{aligned}
Cov(S_1, S_2) &= f_{11}[p^2(x_{11} - \bar{x})^2 + 2pq(x_{11} - \bar{x})(x_{12} - \bar{x}) + q^2(x_{12} - \bar{x})^2] \\
&\quad + f_{12}[\frac{1}{4}p^2(x_{11} - \bar{x})^2 + \frac{1}{2}p(x_{11} - \bar{x})(x_{12} - \bar{x}) + \frac{1}{2}pq(x_{11} - \bar{x})(x_{22} - \bar{x}) \\
&\quad + \frac{1}{4}(x_{12} - \bar{x})^2 + \frac{1}{2}q(x_{12} - \bar{x})(x_{22} - \bar{x}) + \frac{1}{4}q^2(x_{22} - \bar{x})^2] \\
&\quad + f_{22}[p^2(x_{12} - \bar{x})^2 + 2pq(x_{12} - \bar{x})(x_{22} - \bar{x}) + q^2(x_{22} - \bar{x})^2] \\
&= f_{11}[p(x_{11} - \bar{x}) + q(x_{12} - \bar{x})]^2 \\
&\quad + f_{12}[\frac{1}{2}p(x_{11} - \bar{x}) + \frac{1}{2}q(x_{12} - \bar{x}) + \frac{1}{2}p(x_{12} - \bar{x}) + \frac{1}{2}q(x_{22} - \bar{x})]^2 \\
&\quad + f_{22}[p(x_{12} - \bar{x}) + q(x_{22} - \bar{x})]^2 \\
&= p^2[p(x_{11} - \bar{x}) + q(x_{12} - \bar{x})]^2 \\
&\quad + 2pq[\frac{1}{2}p(x_{11} - \bar{x}) + \frac{1}{2}q(x_{12} - \bar{x}) + \frac{1}{2}p(x_{12} - \bar{x}) + \frac{1}{2}q(x_{22} - \bar{x})]^2 \\
&\quad + q^2[p(x_{12} - \bar{x}) + q(x_{22} - \bar{x})]^2 \tag{1} \\
&= p^2[px_{11} + qx_{12} - \bar{x}]^2 \\
&\quad + 2pq \left[\frac{1}{2}(px_{11} + qx_{12} - \bar{x}) + \frac{1}{2}(px_{12} + qx_{22} - \bar{x}) \right]^2 \\
&\quad + q^2[px_{12} + qx_{22} - \bar{x}]^2 \\
&= p^2 \left[\alpha_1 - \frac{\bar{x}}{2} \right]^2 + 2pq \left[\frac{1}{2}(\alpha_1 - \frac{\bar{x}}{2}) + \frac{1}{2}(\alpha_2 - \frac{\bar{x}}{2}) \right]^2 + q^2 \left[\alpha_2 - \frac{\bar{x}}{2} \right]^2 \\
&= p^2 \left[\frac{1}{2}(2\alpha_1 - \bar{x}) \right]^2 + 2pq \left[\frac{1}{2}(\alpha_1 + \alpha_2 - \bar{x}) \right]^2 + q^2 \left[\frac{1}{2}(2\alpha_2 - \bar{x}) \right]^2 \\
&= \left(\frac{1}{4} \right) \{ p^2(2\alpha_1 - \bar{x})^2 + 2pq[(\alpha_1 + \alpha_2 - \bar{x})]^2 + q^2[2\alpha_2 - \bar{x}]^2 \} \\
&= \left(\frac{1}{4} \right) V_a
\end{aligned}$$

A NUMERICAL EXAMPLE

Genotype	A_1A_1	A_1A_2	A_2A_2
Phenotype	0.0	0.8	2.0
$p = 0.4$	$\implies \bar{x} = 1.104$		
	$V_a = 0.5192$		
	$V_d = 0.0092$		

Maternal Genotype	Frequency	Offspring Genotype		
		A_1A_1	A_1A_2	A_2A_2
A_1A_1	0.16	0.4	0.6	—
A_1A_2	0.48	0.2	0.5	0.3
A_2A_2	0.36	—	0.4	0.6

$$\begin{aligned}
Cov(S, S) &= [(0.4)^2(0.16) + (0.2)^2(0.48)](0 - 1.104)^2 \\
&\quad + [(0.6)^2(0.16) + (0.5)^2(0.48) + (0.4)^2(0.36)](0.8 - 1.104)^2 \\
&\quad + [(0.3)^2(0.48) + (0.6)^2(0.36)](2 - 1.104)^2 \\
&\quad + 2[(0.4)(0.6)(0.16) + (0.2)(0.5)(0.48)](0 - 1.104)(0.8 - 1.104) \\
&\quad + 2(0.2)(0.3)(0.48)(0 - 1.104)(2 - 1.104) \\
&\quad + 2[(0.5)(0.3)(0.48) + (0.4)(0.6)(0.36)](0.8 - 1.104)(2.0 - 1.104) \\
&= 0.1298 \\
&= \left(\frac{1}{4}\right) 0.5192
\end{aligned}$$

A SUMMARY OF GENETIC COVARIANCES AMONG RELATIVES

MZ twins (Cov_{MZ}):	$V_a + V_d$
Parent-offspring (Cov_{PO}):	$\left(\frac{1}{2}\right) V_a$
Full sibs (Cov_{FS}):	$\left(\frac{1}{2}\right) V_a + \left(\frac{1}{4}\right) V_d$
Half sibs (Cov_{HS}):	$\left(\frac{1}{4}\right) V_a$

ESTIMATION OF HERITABILITY

Galton introduced the term *regression* to describe the inheritance of height in humans. He noted that there is a tendency for adult offspring of tall parents to be tall and of short parents to be short, but he also noted that offspring tended to be less extreme than the parents. He described this as a “regression to mediocrity,” and statisticians have adopted the term to describe the process of describing the functional relationship between two variables.

Regression analysis

Measure the parents. Regress the offspring phenotype on: (1) the phenotype of one parent or (2) the mean of the parental phenotypes. In either case, the covariance between the parental phenotype and the offspring genotype is $\left(\frac{1}{2}\right) V_a$. Now the regression coefficient between one parent and offspring, $b_{P \rightarrow O}$, is

$$\begin{aligned}
b_{P \rightarrow O} &= \frac{Cov_{PO}}{Var(P)} \\
&= \frac{\left(\frac{1}{2}\right) V_a}{V_p} \\
&= \left(\frac{1}{2}\right) h_N^2
\end{aligned}$$

In short, the slope of the regression line is equal to one-half the narrow sense heritability. In the regression of offspring on mid-parent value,

$$\begin{aligned}
 \text{Var}(MP) &= \text{Var}\left(\frac{M + F}{2}\right) \\
 &= \frac{1}{4}\text{Var}(M + F) \\
 &= \frac{1}{4}(\text{Var}(M) + \text{Var}(F)) \quad . \\
 &= \frac{1}{4}(2V_p) \\
 &= \frac{1}{2}V_p
 \end{aligned}$$

Thus, $b_{MP \rightarrow O} = h_N^2$. In short, the slope of the regression line is equal to the narrow sense heritability.

Sib analysis

Mate a number of males (sires) with a number of females (dams). Each sire is mated to more than one dam, but each dam mates only with one sire. Do an analysis of variance on the phenotype in the progeny, treating sire and dam as main effects. The result is:

Source	d.f.	Mean square	Composition of mean square
Among sires	$s - 1$	MS_S	$= \sigma_W^2 + k\sigma_D^2 + dk\sigma_s^2$
Among dams (within sires)	$s(d - 1)$	MS_D	$= \sigma_W^2 + k\sigma_D^2$
Within progenies	$sd(k - 1)$	MS_W	$= \sigma_W^2$

s = number of sires

d = number of dams per sire

k = number of offspring per dam

Now we need some way to relate the variance components (σ_W^2 , σ_D^2 , and σ_s^2) to the components of the genetic variance. How do we do that? Well,

$$V_p = \sigma_T^2 = \sigma_S^2 + \sigma_D^2 + \sigma_W^2 \quad .$$

σ_S^2 estimates the variance among the means of the half-sib families fathered by each of the different sires or, equivalently, the covariance among half-sibs.²

$$\begin{aligned}
 \sigma_S^2 &= \text{Cov}_{HS} \\
 &= \left(\frac{1}{4}\right)V_a \quad .
 \end{aligned}$$

² To see why consider this is so, consider the following: The mean genotypic value of half-sib families with an A_1A_1 mother is $px_{11} + qx_{12}$; with an A_1A_2 mother, $px_{11}/2 + qx_{12}/2 + px_{12}/2 + qx_{22}/2$; with an A_2A_2 mother, $px_{12} + qx_{22}$. The equation for the variance of these means is identical to the third equation in (1), which is an expression for the covariance among half-sibs.

Now consider the within progeny component of the variance, σ_W^2 . In general, it can be shown that *any* among group variance component is equal to the covariance among the members within the groups.³ Thus, a within group component of the variance is equal to the total variance minus the covariance within groups. In this case,

$$\begin{aligned}\sigma_W^2 &= V_p - Cov_{FS} \\ &= V_a + V_d + V_e - \left[\left(\frac{1}{2} \right) V_a + \left(\frac{1}{4} \right) V_d \right] \\ &= \left(\frac{1}{2} \right) V_a + \left(\frac{3}{4} \right) V_d + V_e\end{aligned}$$

There remains only σ_D^2 . Now $\sigma_W^2 = V_p - Cov_{FS}$, $\sigma_S^2 = Cov_{HS}$, and $\sigma_T^2 = V_p$. Thus,

$$\begin{aligned}\sigma_D^2 &= \sigma_T^2 - \sigma_S^2 - \sigma_W^2 \\ &= V_p - Cov_{HS} - (V_p - Cov_{FS}) \\ &= Cov_{FS} - Cov_{HS} \\ &= \left[\left(\frac{1}{2} \right) V_a + \left(\frac{1}{4} \right) V_d \right] - \left(\frac{1}{4} \right) V_a \\ &= \left(\frac{1}{4} \right) V_a + \left(\frac{1}{4} \right) V_d\end{aligned}$$

In summary,

$$\begin{aligned}V_a &= 4\sigma_S^2 \\ V_d &= 4(\sigma_D^2 - \sigma_S^2) \\ V_e &= \sigma_W^2 - 3\sigma_D^2 + \sigma_S^2\end{aligned}$$

³ With $x_{ij} = a_i + \epsilon_{ij}$, where a_i is the mean group effect and ϵ_{ij} is random effect on individual j in group i (with mean 0), $Cov(x_{ij}, x_{ik}) = E(a_i + \epsilon_{ij} - \mu)(a_i + \epsilon_{ik} - \mu) = E((a_i - \mu)^2 + a_i(\epsilon_{ij} + \epsilon_{ik}) + \epsilon_{ij}\epsilon_{ik}) = Var(A)$.

AN EXAMPLE: BODY WEIGHT IN FEMALE MICE⁴

The analysis involves 719 offspring from 74 sires and 192 dams, each with one litter. The offspring were spread over 4 generations, and the analysis is performed as a nested ANOVA with the genetic analysis nested *within* generations. An additional complication is that the design was unbalanced, i.e., unequal numbers of progeny were measured in each sibship. As a result the degrees of freedom don't work out to be quite as simple as what I showed you.⁵

Source	d.f.	Mean square	Composition of mean square
Among sires	70	17.10	$= \sigma_W^2 + k'\sigma_D^2 + dk'\sigma_s^2$
Among dams (within sires)	118	10.79	$= \sigma_W^2 + k\sigma_D^2$
Within progenies	527	2.19	$= \sigma_W^2$

$d = 2.33$
 $k = 3.48$
 $k' = 4.16$

Using the expressions for the composition of the mean square we obtain

$$\begin{aligned}
 \sigma_W^2 &= MS_W \\
 &= 2.19 \\
 \sigma_D^2 &= \left(\frac{1}{k}\right) (MS_D - \sigma_W^2) \\
 &= 2.47 \\
 \sigma_s^2 &= \left(\frac{1}{dk'}\right) (MS_S - \sigma_W^2 - k'\sigma_D^2) \\
 &= 0.48
 \end{aligned}$$

Thus,

$$\begin{aligned}
 V_p &= 5.14 \\
 V_a &= 1.92 \\
 V_d + V_e &= 3.22 \\
 V_d &= (0.00—1.64) \\
 V_e &= (1.58—3.22)
 \end{aligned}$$

Why didn't I give a definite number for V_d after my big spiel above about how we can get it? Two reasons. First, if you plug these values into the formula above for V_d you get $V_d = 7.96, V_e = -4.74$, which is clearly impossible since V_d has to be less than V_p and V_e has to be greater than zero, since it's a variance. Second, the experimental design actually confounds two sources of resemblance among full siblings: (1) genetic covariance and (2)

⁴ Falconer & Mackay, p. 169–170

⁵ What did you expect from real data?

environmental covariance. Since the full-sib families were all raised by the same mother in the same pen, we don't know to what extent their resemblance is due to a common natal environment.⁶ If we assume $V_d = 0$, we can estimate the amount of variance accounted for by exposure to a common natal environment, $V_{Ec} = 1.99$, and by environmental variation within sibships, $V_{Ew} = 1.23$.⁷ Similarly, if we assume $V_{Ew} = 0$, then $V_d = 1.64$ and $V_{Ec} = 1.58$. In any case, we can estimate the narrow sense heritability as

$$h_N^2 = \left(\frac{1.92}{5.14} \right) .$$
$$= 0.37$$

⁶ Notice that this doesn't affect our analysis of half-sib families, i.e., the progeny of different sires, since each father was bred with several females

⁷ See Falconer for details.