Resemblance among relatives

Introduction

Just as individuals may differ from one another in phenotype because they have different genotypes, because they developed in different environments, or both, relatives may resemble one another more than they resemble other members of the population because they have similar genotypes, because they developed in similar environments, or both. In an experimental situation, we typically try to randomize individuals across environments. If we are successful, then any tendency for relatives to resemble one another more than non-relatives must be due to similarities in their genotypes.

Using this insight, we can develop a statistical technique that allows us to determine how much of the variance among individuals in phenotype is a result of genetic variance and how much is due to environmental variance. *Remember*, we can only ask about how much of the variability is due to genetic differences, and we can only do so *in a particular environment* and *with a particular set of genotypes*, and we can only do it when we *randomize genotypes across environments*.

An outline of the approach

The basic approach to the analysis is either to use a linear regression of offspring phenotype on parental phenotype, which as we'll see estimates h_n^2, or to use a nested analysis of variance. One of the most complete designs is a full-sib, half-sib design in which each male sires offspring from several dams but each dam mates with only one sire.

The offspring of a single dam are full-sibs (they are nested within dams). Differences among the offspring of dams indicates that there are differences in maternal "genotype" in the trait being measured.[1]

The offspring of different dams mated to a single sire are half-sibs. Differences among the offspring of sires indicates that there are differences in paternal "genotype" in the trait

---

[1] Assuming that we've randomized siblings across environments. If we haven't, siblings may resemble one another because of similarities in the environment they experienced, too.

Maternal genotype	Frequency	Offspring genotype A_1A_1	A_1A_2	A_2A_2
A_1A_1	p^2	p	q	0
A_1A_2	$2pq$	$\frac{p}{2}$	$\frac{1}{2}$	$\frac{q}{2}$
A_2A_2	q^2	0	p	q

Table 1: Half-sib family structure in a population with genotypes in Hardy-Weinberg proportions.

being measured.[2]

As we'll see, this design has the advantage that it allows both additive and dominance components of the genetic variance to be estimated. It has the additional advantage that we don't have to assume that the distribution of environments in the offspring generation is the same as it was in the parental generation. To use the regression approach to estimate heritability, we have to assume that the distribution of environmental effects is the same in parental and offspring generations.

The gory details

OK, so I've given you the basic idea. Where does it come from, and how does it work? Funny you should ask. The whole approach is based on calculations of the degree to which different relatives resemble one another. For these purposes we're going to continue our focus on phenotypes influenced by one locus with two alleles, and we'll do the calculations in detail only for half sib families. We start with something that may look vaguely familiar.[3] Take a look at Table 1.

Note that the probabilities in Table 1 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.[4]

---

[2]You'll see the reason for the quotes around genotype in this paragraph and the last a little later. It's a little more complex than what I've suggested.

[3]Remember our mother-offspring combinations with *Zoarces viviparus*?

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

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:

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

where μ_x is the mean of X and μ_y is the mean of Y. The covariance between two random variables is a measure of how much they vary together — covary. If the covariance is large and positive, they tend to vary in the same way. Positive deviations from the mean in one are associated with positive deviations from the mean in the other, and negative deviations are similarly associated. If the covariance is large and negative, they tend to vary in opposite ways. Positive deviations from the mean in one variable are associated with *negative* deviations in the other, and vice versa. If the covariance is small, it means there isn't a strong tendency for deviations from the mean in one variable to be associated with deviations in the other.

Covariance between half-siblings

Here's how we can calculate the covariance between half-siblings: First, imagine selecting huge number of half-sibs pairs at random. The phenotype of the first half-sib in the pair is a random variable (call it S_1), as is the phenotype of 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 $\bar{x}$.[5] 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 of randomly chosen half-sibs based on the frequencies in Table 1 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 her progeny.

$$
\begin{aligned}
\text{Cov}(S_1, S_2) \;=\; & p^2[p^2(x_{11} - \bar{x})^2 + 2pq(x_{11} - \bar{x})(x_{12} - \bar{x}) + q^2(x_{12} - \bar{x})^2] \\
& + 2pq[\frac{1}{4}p^2(x_{11} - \bar{x})^2 + \frac{1}{2}p(x_{11} - \bar{x})(x_{12} - \bar{x}) + \frac{1}{2}pq(x_{11} - \bar{x})(x_{22} - \bar{x})
\end{aligned}
$$

---

[5]The reasoning here gets a little tricky, since the mean of different half-sib families may be different. Think about it this way. We picked this particular half-sib family at random from among all half-sib families in the population. It takes a bit of algebra to show it, but the mean phenotype of a randomly chosen half-sib family is $\bar{x}$, meaning that $\bar{x}$ is the mean phenotype for both S_1 and S_2. They're part of the same family, so they share the same family mean.

$$+\frac{1}{4}(x_{12}-\bar{x})^2 + \frac{1}{2}q(x_{12}-\bar{x})(x_{22}-\bar{x}) + \frac{1}{4}q^2(x_{22}-\bar{x})^2]$$
$$+q^2[p^2(x_{12}-\bar{x})^2 + 2pq(x_{12}-\bar{x})(x_{22}-\bar{x}) + q^2(x_{22}-\bar{x})^2]$$
$$= \quad p^2[p(x_{11}-\bar{x}) + q(x_{12}-\bar{x})]^2$$
$$+2pq[\frac{1}{2}p(x_{11}-\bar{x}) + \frac{1}{2}q(x_{12}-\bar{x}) + \frac{1}{2}p(x_{12}-\bar{x}) + \frac{1}{2}q(x_{22}-\bar{x})]^2$$
$$+q^2[p(x_{12}-\bar{x}) + q(x_{22}-\bar{x})]^2$$
$$= \quad p^2[px_{11} + qx_{12} - \bar{x}]^2$$
$$+2pq\left[\frac{1}{2}(px_{11} + qx_{12} - \bar{x}) + \frac{1}{2}(px_{12} + qx_{22} - \bar{x})\right]^2$$
$$+q^2[px_{12} + qx_{22} - \bar{x}]^2$$
$$= \quad p^2\left[\alpha_1 - \frac{\bar{x}}{2}\right]^2 + 2pq\left[\frac{1}{2}(\alpha_1 - \frac{\bar{x}}{2}) + \frac{1}{2}(\alpha_2 - \frac{\bar{x}}{2})\right]^2 + q^2\left[\alpha_2 - \frac{\bar{x}}{2}\right]^2$$
$$= \quad p^2\left[\frac{1}{2}(2\alpha_1 - \bar{x})\right]^2 + 2pq\left[\frac{1}{2}(\alpha_1 + \alpha_2 - \bar{x})\right]^2 + q^2\left[\frac{1}{2}(2\alpha_2 - \bar{x})\right]^2$$
$$= \quad \left(\frac{1}{4}\right)\left[p^2(2\alpha_1 - \bar{x})^2 + 2pq[(\alpha_1 + \alpha_2 - \bar{x})]^2 + q^2(2\alpha_2 - \bar{x})^2\right]$$
$$= \quad \left(\frac{1}{4}\right)V_a$$

A numerical example

Now we'll return to an example we saw earlier (Table 2). This set of genotypes and phenotypes may look familiar. It is the same one we encountered earlier when we calculated additive and dominance components of variance. Let's assume that $p = 0.4$. We know from our earlier calculations that

$$\begin{aligned}
\bar{x} &= 54.4 \\
V_a &= 1505.28 \\
V_d &= 207.36 \quad .
\end{aligned}$$

We can also calculate the numerical version of Table 1, which you'll find in Table 3.

So now we can follow the same approach we did before and calculate the numerical value of the covariance between half-sibs in this example:

$$\begin{aligned}
\text{Cov}(S_1, S_2) &= \quad [(0.4)^2(0.16) + (0.2)^2(0.48)](100 - 54.4)^2 \\
&\quad +[(0.6)^2(0.16) + (0.5)^2(0.48) + (0.4)^2(0.36)](80 - 54.4)^2
\end{aligned}$$

4

Genotype	A_1A_1	A_1A_2	A_2A_2
Phenotype	100	80	0

Table 2: An example of a non-additive relationship between genotypes and phenotypes.

Maternal genotype	Frequency	Offspring genotype		
		A_1A_1	A_1A_2	A_2A_2
A_1A_1	0.16	0.4	0.6	0.0
A_1A_2	0.48	0.2	0.5	0.3
A_2A_2	0.36	0.0	0.4	0.6

Table 3: Mother-offspring combinations (half-sib) when the frequency of A_1 is 0.4.
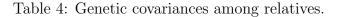
$$
\begin{aligned}
&+[(0.3)^2(0.48) + (0.6)^2(0.36)](0 - 54.4)^2 \\
&+2[(0.4)(0.6)(0.16) + (0.2)(0.5)(0.48)](100 - 54.4)(80 - 54.4) \\
&+2(0.2)(0.3)(0.48)(100 - 54.4)(0 - 54.4) \\
&+2[(0.5)(0.3)(0.48) + (0.4)(0.6)(0.36)](80 - 54.4)(0 - 54.4) \\
=\quad& 376.32 \\
=\quad& \left(\frac{1}{4}\right)1505.28 \quad.
\end{aligned}
$$

Covariances among relatives

Well, if we can do this sort of calculation for half-sibs, you can probably guess that it's also possible to do it for other relatives. I won't go through all of the calculations, but the results for common forms of relationship are summarized in Table 4

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

[1]One parent or mid-parent.

Table 4: Genetic covariances among relatives.

Estimating 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.[6] He described this as a "regression to mediocrity," and statisticians adopted the term to describe a standard technique for 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, $b_{P \to O}$, is

$$
\begin{aligned}
b_{P \to O} &= \frac{\text{Cov}_{PO}}{\text{Var}(P)} \\
&= \frac{\left(\frac{1}{2}\right) V_a}{V_p} \\
&= \left(\frac{1}{2}\right) h_N^2 \quad .
\end{aligned}
$$

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,

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

---

[6]It's worth noting that Galton is often "credited" with establishing the field of eugenics. He was a proponent of encouraging the "best" people to marry one another to "improve" the human race. In 2020, University College London renamed two lecture theaters and a building that bore the names of Francis Galton and Karl Pearson (`https://www.theguardian.com/education/2020/jun/19/ucl-renames-three-facilities-that-honoured-prominent-eugenicists`).

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	σ_W^2

s = number of sires
d = number of dams per sire
k = number of offspring per dam

Table 5: Analysis of variance table for a full-sib analysis of quantitative genetic variation.

Thus, $b_{MP\to O} = \frac{1}{2}V_a / \frac{1}{2}V_p = h_N^2$. 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 shown in Table 5.

Now we need some way to relate the variance components (σ_W^2, σ_D^2, and σ_S^2) to V_a, V_d, and V_e.[7] 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.[8]

$$\sigma_S^2 = \mathrm{Cov}_{HS}$$
$$= \left(\frac{1}{4}\right) V_a \quad .$$

---

[7]σ_W^2, σ_D^2, and σ_S^2 are often referred to as the *observational* components of variance, because they are estimated from observations we make on phenotypic variation. V_a, V_d, and V_e are often referred to as the *causal* components of variance, because they represent the genetic and environmental influences on trait expression.

[8]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 among these means is identical to the equation 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.[9] Thus, a within group component of the variance is equal to the total variance minus the covariance within groups. In this case,

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

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,

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

So if we rearrange these equations, we can express the genetic components of the phenotypic variance, the *causal* components of variance, as simple functions of the *observational* components of variance:

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

Furthermore, the narrow-sense heritability is given by

$$h_N^2 = \frac{4\sigma_s^2}{\sigma_S^2 + \sigma_D^2 + \sigma_W^2} \quad .$$

---

[9]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), $Cov(x_{ij}, x_{ik}) = E(a_i + \epsilon_{ij} - \mu)(a_i + \epsilon_{ik} - \mu) = E((a_i - \mu^2) + a_i(\epsilon_{ij} + \epsilon_{ik}) + \epsilon_{ij}\epsilon_{ik}) = Var(A)$.

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	σ_W^2

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

Table 6: Quantitative genetic analysis of the inheritance of body weight in female mice (from Falconer and Mackay, pp. 169–170.)

An example: body weight in female mice

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.[10] The results are summarized in Table 6.

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

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

Thus,

$$
\begin{aligned}
V_p &= 5.14 \\
V_a &= 1.92
\end{aligned}
$$

---

[10]What did you expect from real data? This example is extracted from Falconer and Mackay, pp. 169–170. See the book for details.

$$V_d + V_e = 3.22$$
$$V_d = (0.00\text{---}1.64)$$
$$V_e = (1.58\text{---}3.22)$$

Why didn't I give a definite number for V_d after my big spiel above about how we can estimate it from a full-sib crossing design? Two reasons. First, if you plug the estimates for σ_D^2 and σ_S^2 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. It's a variance. Second, the experimental design confounds two sources of resemblance among full siblings: (1) genetic covariance and (2) environmental covariance. The full-sib families were all raised by the same mother in the same pen. Hence, we don't know to what extent their resemblance is due to a common natal environment.[11] 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, $V_{Ew} = 1.23$.[12] 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

$$\begin{aligned} h_N^2 &= \left(\frac{1.92}{5.14}\right) \\ &= 0.37 \quad . \end{aligned}$$

Creative Commons License

---

[11]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

[12]See Falconer for details.