The genetic components of variance

A summary of notation and properties

Genotype	A_1A_1	A_1A_2	A_2A_2
Frequency	p^2	$2pq$	q^2
Genotypic value	x_{11}	x_{12}	x_{22}
Additive genotypic value	$2\alpha_1$	$\alpha_1 + \alpha_2$	$2\alpha_2$

Least squares estimates for α_1 and α_2

The objective is to find values for α_1 and α_2 that minimize the following function:

$$f = p^2[x_{11} - 2\alpha_1]^2 + 2pq[x_{12} - (\alpha_1 + \alpha_2)]^2 + q^2[x_{22} - 2\alpha_2]^2 \quad . \tag{1}$$

To do this we take the partial derivative of f with respect to both α_1 and α_2, set the resulting pair of equations equal to zero, and solve for α_1 and α_2.[1]

$$\begin{aligned} \frac{\partial f}{\partial \alpha_1} &= p^2\{2[x_{11} - 2\alpha_1][-2]\} + 2pq\{2[x_{12} - (\alpha_1 + \alpha_2)][-1]\} \\ &= -4p^2[x_{11} - 2\alpha_1] - 4pq[x_{12} - (\alpha_1 + \alpha_2)] \\ \frac{\partial f}{\partial \alpha_2} &= q^2\{2[x_{22} - 2\alpha_2][-2]\} + 2pq\{2[x_{12} - (\alpha_1 + \alpha_2)][-1]\} \\ &= -4q^2[x_{22} - 2\alpha_2] - 4pq[x_{12} - (\alpha_1 + \alpha_2)] \end{aligned} \tag{2}$$

Thus, $\frac{\partial f}{\partial \alpha_1} = \frac{\partial f}{\partial \alpha_2} = 0$ if and only if

$$\begin{aligned} p^2(x_{11} - 2\alpha_1) + pq(x_{12} - \alpha_1 - \alpha_2) &= 0 \\ q^2(x_{22} - 2\alpha_2) + pq(x_{12} - \alpha_1 - \alpha_2) &= 0 \end{aligned} \tag{3}$$

Adding the equations in (3) we obtain (after a little bit of rearrangement)

$$[p^2x_{11} + 2pqx_{12} + q^2x_{22}] - [p^2(2\alpha_1) + 2pq(\alpha_1 + \alpha_2) + q^2(2\alpha_2)] = 0 \quad . \tag{4}$$

Now the first term in square brackets is just the mean phenotype in the population, $\bar{x}$. Thus, we can rewrite (4) as:

$$\begin{aligned} \bar{x} &= 2p^2\alpha_1 + 2pq(\alpha_1 + \alpha_2) + 2q^2\alpha_2 \\ &= 2p\alpha_1(p + q) + 2q\alpha_2(p + q) \\ &= 2(p\alpha_1 + q\alpha_2) \end{aligned} \quad . \tag{5}$$

Now divide the first equation in (3) by p and the second by q.

$$\begin{aligned} p(x_{11} - 2\alpha_1) + q(x_{12} - \alpha_1 - \alpha_2) &= 0 \\ q(x_{22} - 2\alpha_2) + p(x_{12} - \alpha_1 - \alpha_2) &= 0 \end{aligned} \tag{6}$$

[1] We won't bother with proving that the resulting estimates produce the minimum possible value of f. Just take my word for it.

Thus,

$$\begin{aligned} px_{11} + qx_{12} &= 2p\alpha_1 + q\alpha_1 + q\alpha_2 \\ &= \alpha_1(p+q) + p\alpha_1 + q\alpha_2 \\ &= \alpha_1 + p\alpha_1 + q\alpha_2 \\ &= \alpha_1 + \bar{x}/2 \end{aligned} \tag{7}$$

$$\alpha_1 = px_{11} + qx_{12} - \bar{x}/2 \tag{8}$$

Similarly,

$$\begin{aligned} px_{12} + qx_{22} &= 2q\alpha_2 + p\alpha_1 + p\alpha_2 \\ &= \alpha_2(p+q) + p\alpha_1 + q\alpha_2 \\ &= \alpha_2 + p\alpha_1 + q\alpha_2 \\ &= \alpha_2 + \bar{x}/2 \end{aligned} \tag{9}$$

$$\alpha_2 = px_{12} + qx_{22} - \bar{x}/2 \tag{10}$$

Components of the genetic variance

$$\begin{aligned} V_g &= p^2[x_{11} - \bar{x}]^2 + 2pq[x_{12} - \bar{x}]^2 + q^2[x_{22} - \bar{x}]^2 \\ &= p^2[x_{11} - 2\alpha_1 + 2\alpha_1 - \bar{x}]^2 + 2pq[x_{12} - (\alpha_1 + \alpha_2) + (\alpha_1 + \alpha_2) - \bar{x}]^2 \\ &\quad + q^2[x_{22} - 2\alpha_2 + 2\alpha_2 - \bar{x}]^2 \\ &= p^2[x_{11} - 2\alpha_1]^2 + 2pq[x_{12} - (\alpha_1 + \alpha_2)]^2 + q^2[x_{22} - 2\alpha_2]^2 \\ &\quad + p^2[2\alpha_1 - \bar{x}]^2 + 2pq[(\alpha_1 + \alpha_2) - \bar{x}]^2 + q^2[2\alpha_2 - \bar{x}]^2 \\ &\quad + p^2[2(x_{11} - 2\alpha_1)(2\alpha_1 - \bar{x})] + 2pq[2(x_{12} - \{\alpha_1 + \alpha_2\})(\{\alpha_1 + \alpha_2\} - \bar{x})] \\ &\quad + q^2[2(x_{22} - 2\alpha_2)(2\alpha_2 - \bar{x})] \end{aligned} \tag{11}$$

There are two terms in (11) that have a biological (or at least a quantitative genetic) interpretation. The term on the first line of the last equation is the average squared deviation between the genotypic value and the additive genotypic value. It will be zero only if the effects of the alleles can be decomposed into strictly additive components, i.e., only if the heterozygote is exactly intermediate. Thus, it is a measure of how much variation is due to non-additivity (dominance) of allelic effects. In short, the dominance genetic variance, V_d, is

$$V_d = p^2[x_{11} - 2\alpha_1]^2 + 2pq[x_{12} - (\alpha_1 + \alpha_2)]^2 + q^2[x_{22} - 2\alpha_2]^2 \quad . \tag{12}$$

Similarly, the term on the second line of the last equation in (11) is the average squared deviation between the additive genotypic value and the mean genotypic value in the population. Thus, it is a measure of how much variation is due to differences between genotypes in their additive genotype. In short, the additive genetic variance, V_a, is

$$V_a = p^2[2\alpha_1 - \bar{x}]^2 + 2pq[(\alpha_1 + \alpha_2) - \bar{x}]^2 + q^2[2\alpha_2 - \bar{x}]^2 \quad . \tag{13}$$

What about the terms on the third and fourth lines of the last equation in (11)? Well, they can be rearranged as follows:

$$
\begin{aligned}
p^2&[2(x_{11}-2\alpha_1)(2\alpha_1-\bar{x})] + 2pq[2(x_{12}-\{\alpha_1+\alpha_2\})(\{\alpha_1+\alpha_2\}-\bar{x})] \\
&\quad + q^2[2(x_{22}-2\alpha_2)(2\alpha_2-\bar{x})] \\
&= 2p^2(x_{11}-2\alpha_1)(2\alpha_1-\bar{x}) + 4pq[x_{12}-(\alpha_1+\alpha_2)][(\alpha_1+\alpha_2)-\bar{x})] \\
&\quad + 2q^2(x_{22}-2\alpha_2)(2\alpha_2-\bar{x}) \\
&= 4p^2(x_{11}-2\alpha_1)[\alpha_1-(p\alpha_1+q\alpha_2)] \\
&\quad + 4pq[x_{12}-(\alpha_1+\alpha_2)][(\alpha_1+\alpha_2)-2(p\alpha_1+q\alpha_2)] \\
&\quad + 4q^2(x_{22}-2\alpha_2)[\alpha_2-(p\alpha_1+q\alpha_2)] \\
&= 4p[\alpha_1-(p\alpha_1+q\alpha_2)][p(x_{11}-2\alpha_1)+q(x_{12}-\{\alpha_1+\alpha_2\})] \\
&\quad + 4q[\alpha_2-(p\alpha_1+q\alpha_2)][p(x_{11}-2\alpha_1)p+q(x_{12}-\{\alpha_1+\alpha_2\})] \\
&= 0
\end{aligned}
\tag{14}
$$

Where we have used the identities $\bar{x} = 2(p\alpha_1 + q\alpha_2)$ [see (5)] and

$$
\begin{aligned}
p(x_{11}-2\alpha_1) + q(x_{12}-\alpha_1-\alpha_2) &= 0 \\
q(x_{22}-2\alpha_2) + p(x_{12}-\alpha_1-\alpha_2) &= 0
\end{aligned}
$$

[see (6)]. In short, we have now shown that the total genotypic variance in the population, V_g, can be subdivided into two components — the additive genetic variance, V_a, and the dominance genetic variance, V_d. Specifically,

$$
V_g = V_a + V_d \quad ,
$$

where V_g is given by the first line of (11), V_a by (13), and V_d by (12).

An alternative expression for V_a

$$
\begin{aligned}
V_a &= p^2(2\alpha_1)^2 + 2pq(\alpha_1+\alpha_2)^2 + q^2(2\alpha_2)^2 - 4(p\alpha_1+q\alpha_2)^2 \\
&= 4p^2\alpha_1^2 + 2pq(\alpha_1+\alpha_2)^2 + 4q^2\alpha_2^2 - 4(p^2\alpha_1^2 + 2pq\alpha_1\alpha_2 + q^2\alpha_2^2) \\
&= 2pq[(\alpha_1+\alpha_2)^2 - 4\alpha_1\alpha_2] \\
&= 2pq[(\alpha_1^2 + 2\alpha_1\alpha_2 + \alpha_2^2) - 4\alpha_1\alpha_2] \\
&= 2pq[\alpha_1^2 - 2\alpha_1\alpha_2 + \alpha_2^2] \\
&= 2pq[\alpha_1 - \alpha_2]^2 \\
&= 2pq\alpha^2
\end{aligned}
\tag{15}
$$

Calculating the genetic variance with known genotypes

$$
\left.
\begin{array}{cccc}
\text{Genotype} & A_1A_1 & A_1A_2 & A_2A_2 \\
\text{Phenotype} & 0 & 1 & 2
\end{array}
\right\} \Longrightarrow \text{additive phenotype}
$$

For $p = 0.4$

$$\begin{aligned}
\bar{x} &= (0.4)^2(0) + 2(0.4)(0.6)(1) + (0.6)^2(2) \\
&= 1.20 \\
\alpha_1 &= (0.4)(0) + (0.6)(1) - (1.20)/2 \\
&= 0.0 \\
\alpha_2 &= (0.4)(1) + (0.6)(2) - (1.20)/2 \\
&= 1.0 \\
V_g &= (0.4)^2(0 - 1.20)^2 + 2(0.4)(0.6)(1 - 1.20)^2 + (0.6)^2(2 - 1.20)^2 \\
&= 0.48 \\
V_a &= (0.4)^2[2(0.0) - 1.20]^2 + 2(0.4)(0.6)[(0.0 + 1.0) - 1.20]^2 + (0.6)^2[2(1.0) - 1.20]^2 \\
&= 0.48 \\
V_d &= (0.4)^2[0 - 2(0.0)]^2 + 2(0.4)(0.6)[1 - (0.0 + 1.0)]^2 + (0.6)^2[2 - 2(1.0)]^2 \\
&= 0.00 \quad .
\end{aligned}$$

For $p = 0.2$, $\bar{x} = 1.60$, $V_g = V_a = 0.32$, $V_d = 0.00$. You should verify for yourself that $\alpha_1 = 0$ and $\alpha_2 = 1$ for $p = 0.2$. If you are ambitious, you could try to prove that $\alpha_1 = 0$ and $\alpha_2 = 1$ for *any* allele frequency.

$$\left.\begin{array}{lccc}
\text{Genotype} & A_1A_1 & A_1A_2 & A_2A_2 \\
\text{Phenotype} & 0 & 0.8 & 2
\end{array}\right\} \Longrightarrow A_1 \text{ partially dominant}$$

For $p = 0.4$

$$\begin{aligned}
\bar{x} =& (0.4)^2(0) + 2(0.4)(0.6)(0.8) + (0.6)^2(2) \\
=& 1.104 \\
\alpha_1 =& (0.4)(0) + (0.6)(0.8) - (1.104)/2 \\
=& -0.072 \\
\alpha_2 =& (0.4)(0.8) + (0.6)(2) - (1.104)/2 \\
=& 0.968 \\
V_g =& \ (0.4)^2(0 - 1.104)^2 + 2(0.4)(0.6)(1 - 1.104)^2 + (0.6)^2(2 - 1.104)^2 \\
=& \ 0.5284 \\
V_a =& \ (0.4)^2[2(-0.072) - 1.104]^2 + 2(0.4)(0.6)[(-0.072 + 0.968) - 1.104]^2 \\
& + (0.6)^2[2(0.968) - 1.104]^2 \\
=& \ 0.5192 \\
V_d =& \ (0.4)^2[0 - 2(-0.072)]^2 + 2(0.4)(0.6)[0.8 - (-0.072 + 0.968)]^2 \\
& + (0.6)^2[2 - 2(0.968)]^2 \\
=& \ 0.0092 \quad .
\end{aligned}$$

To test your understanding, it would probably be useful to calculate $\bar{x}$, α_1, α_2, V_g, V_a, and V_d for one or two other allele frequencies, say $p = 0.2$ and $p = 0.8$. Is it still true that α_1 and α_2 are independent of allele frequencies? If you are *really* ambitious you could try to prove that α_1 and α_2 are independent of allele frequencies if and only if $x_{12} = (x_{11} + x_{12})/2$, i.e., when heterozygotes are exactly intermediatef.