Additive Genetic Variance

DANIELLE POSTHUMA

Volume 1, pp. 18–22

in

Encyclopedia of Statistics in Behavioral Science

ISBN-13: 978-0-470-86080-9
ISBN-10: 0-470-86080-4

Editors

Brian S. Everitt & David C. Howell

Additive Genetic Variance

The starting point for gene finding is the observation of population variation in a certain trait. This 'observed', or phenotypic, variation may be attributed to genetic and environmental causes. Although environmental causes of phenotypic variation should not be ignored and are highly interesting, in the following section we will focus on the biometric model underlying genetic causes of variation, specifically *additive* genetic causes of variation.

Within a population, one, two, or many different alleles may exist for a gene (*see* **Allelic Association**). Uniallelic systems will not contribute to population variation. For simplicity, we assume in this treatment one gene with two possible alleles, alleles A1 and A2. By convention, allele A1 has frequency p, while allele A2 has frequency q, and $p + q = 1$. With two alleles, there are three possible **genotypes**: A1A1, A1A2, and A2A2, with corresponding genotypic frequencies p^2, $2pq$, and q^2 (assuming random mating, equal viability of alleles, no selection, no migration and no mutation, see [3]). The genotypic effect on a phenotypic trait (i.e., *the genotypic value*) of genotype A1A1, is by convention called 'a' and the effect of genotype A2A2 ' $-a$'. The effect of the heterozygous genotype A1A2 is called 'd'. If the genotypic value of the heterozygote lies exactly at the midpoint of the genotypic values of the two homozygotes ($d = 0$), there is said to be no genetic dominance. If allele A1 is completely dominant over allele A2, effect d equals effect a. If d is larger than a, there is overdominance. If d is unequal to zero and the two alleles produce three discernable phenotypes of the trait, d is unequal to a. This model is also known as the *classical biometrical model* [3, 6] (see Figure 1 for a worked example).

The genotypic contribution of a gene to the population mean of a trait (i.e., the mean effect of a gene, or μ) is the sum of the products of the frequencies and the genotypic values of the different genotypes:

$$\begin{aligned}\text{Mean effect} &= (ap^2) + (2pqd) + (-aq^2)\\ &= a(p - q) + 2pqd. \qquad (1)\end{aligned}$$

This mean effect of a gene consists of two components: the contribution of the homozygotes $[a(p - q)]$ and the contribution of the heterozygotes $[2pqd]$. If there is no dominance, that is d equals zero, there is no contribution of the heterozygotes and the mean is a simple function of the allele frequencies. If d equals a, which is defined as complete dominance, the population mean becomes a function of the square of the allele frequencies; substituting d for a gives $a(p - q) + 2pqa$, which simplifies to $a(1 - 2q^2)$.

Complex traits such as height or weight are not very likely influenced by a single gene, but are assumed to be influenced by many genes. Assuming only additive and independent effects of all of these genes, the expectation for the population mean (μ) is the sum of the mean effects of all the separate genes, and can formally be expressed as $\mu = \sum a(p - q) + 2\sum dpq$ (see also Figure 2).

Average Effects and Breeding Values

Let us consider a relatively simple trait that seems to be mainly determined by genetics, for example eye color. As can be widely observed, when a brown-eyed parent mates with a blue-eyed parent, their offspring will not be either brown eyed or blue eyed, but may also have green eyes. At present, three genes are known to be involved in human eye color. Two of these genes lie on chromosome 15: the EYCL2 and EYCL3 genes (also known as the *BEY* 1 and *BEY* 2 gene respectively) and one gene lies on chromosome 19; the EYCL1 gene (or GEY gene) [1, 2]. For simplicity, we ignore one gene (BEY1), and assume that only GEY and BEY2 determine eye color. The BEY2 gene has two alleles: a blue allele and a brown allele. The brown allele is completely dominant over the blue allele. The GEY gene also has two alleles: a green allele and a blue allele. The green allele is dominant over the blue allele of GEY but also over the blue allele of BEY2. The brown allele of BEY2 is dominant over the green allele of *GEY*. Let us assume that the brown-eyed parent has genotype *brown–blue* for the BEY2 gene and *green–blue* for the GEY gene, and that the blue-eyed parent has genotype *blue–blue* for both the *BEY2* gene and the GEY gene. Their children can be (a) brown eyed: *brown–blue* for the *BEY2* gene and either *blue–blue* or *green–blue* for the GEY gene; (b) green eyed: *blue–blue* for the BEY2 gene and *green–blue* for the GEY gene; (c) blue eyed: *blue–blue* for the BEY2 gene and *blue–blue* for the GEY gene. The possibility of having green-eyed children from a brown-eyed

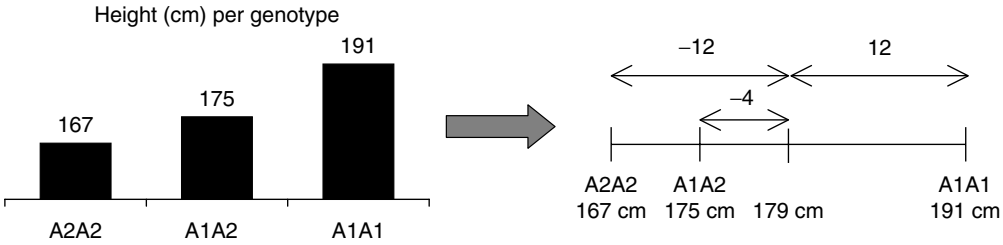

Figure 1 Worked example of genotypic effects, average effects, breeding values, and genetic variation. Assume body height is determined by a single gene with two alleles A1 and A2, and frequencies $p = 0.6$, $q = 0.4$. Body height differs per genotype: A2A2 carriers are 167 cm tall, A1A2 carriers are 175 cm tall, and A1A1 carriers are 191 cm tall. Half the difference between the heights of the two homozygotes is a, which is 12 cm. The midpoint of the two homozygotes is 179 cm, which is also the intercept of body height within the population, that is, subtracting 179 from the three genotypic means scales the midpoint to zero. The deviation of the heterozygote from the midpoint $(d) = -4$ cm. The mean effect of this gene to the population mean is thus $12(0.6 - 0.4) + 2 * 0.6 * 0.4 * -4 = 0.48$ cm. To calculate the average effect of allele A1 (α_1) c, we sum the product of the conditional frequencies and genotypic values of the two possible genotypes, including the A1 allele. The two genotypes are A1A1 and A1A2, with genotypic values 12 and – 4. Given one A1 allele, the frequency of A1A1 is 0.6 and of A1A2 is 0.4. Thus, $12 * 0.6 - 4 * 0.4 = 5.6$. We need to subtract the mean effect of this gene (0.48) from 5.12 to get the average effect of the A1 allele (α_1): $5.6 - 0.48 = 5.12$. Similarly, the average effect of the A2 allele (α_2) can be shown to equal -7.68. The breeding value of A1A1 carriers is the sum of the average effects of the two A1 alleles, which is $5.12 + 5.12 = 10.24$. Similarly, for A1A2 carriers this is $5.12 - 7.68 = 2.56$ and for A2A2 carriers this is $-7.68 - 7.68 = -15.36$. The genetic variance (V_G) related to this gene is 82.33, where V_A is 78.64 and V_D is 3.69

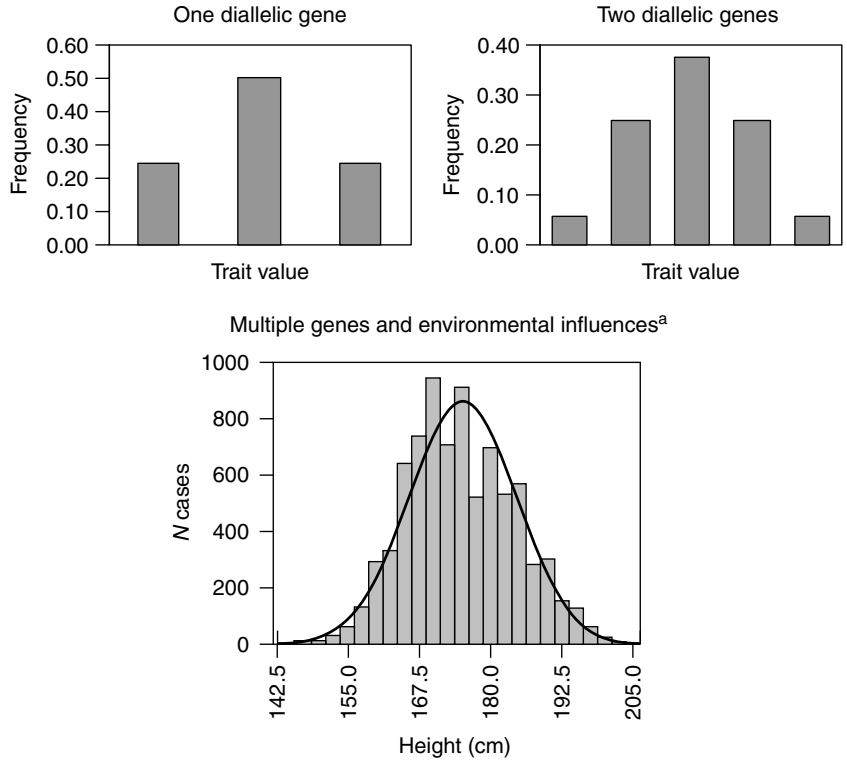

Figure 2 The combined discrete effects of many single genes result in continuous variation in the population. [a]Based on 8087 adult subjects from the Dutch Twin Registry (`http://www.tweelingenregister.org`)

parent and a blue-eyed parent is of course a consequence of the fact that parents transmit alleles to their offspring and not their genotypes. Therefore, parents cannot directly transmit their genotypic values a, d, and $-a$ to their offspring. To quantify the transmission of genetic effects from parents to offspring, and ultimately to decompose the observed variance in the offspring generation into genetic and environmental components, the concepts *average effect* and *breeding value* have been introduced [3].

Average effects are a function of genotypic values and allele frequencies within a population. The average effect of an allele is defined as '.. *the mean deviation from the population mean of individuals which received that allele from one parent, the allele received from the other parent having come at random from the population*' [3]. To calculate the average effects denoted by α_1 and α_2 of alleles A1 and A2 respectively, we need to determine the frequency of the A1 (or A2) alleles in the genotypes of the offspring coming from a single parent. Again, we assume a single locus system with two alleles. If there is random mating between gametes carrying the A1 allele and gametes from the population, the frequency with which the A1 gamete unites with another gamete containing A1 (producing an A1A1 genotype in the offspring) equals p, and the frequency with which the gamete containing the A1 gamete unites with a gamete carrying A2 (producing an A1A2 genotype in the offspring) is q. The genotypic value of the genotype A1A1 in the offspring is a and the genotypic value of A1A2 in the offspring is d, as defined earlier. The mean value of the genotypes that can be produced by a gamete carrying the A1 allele equals the sum of the products of the frequency and the genotypic value. Or, in other terms, it is $pa + qd$. The average genetic effect of allele A1 (α_1) equals the deviation of the mean value of all possible genotypes that can be produced by gametes carrying the A1 allele from the population mean. The population mean has been derived earlier as $a(p - q) + 2pqd$ (1). The average effect of allele A1 is thus: $\alpha_1 = pa + qd - [a(p - q) + 2pqd] = q[a + d(q - p)]$. Similarly, the average effect of the A2 allele is $\alpha_2 = pd - qa - [a(p - q) + 2pqd] = -p[a + d(q - p)]$. $\alpha_1 - \alpha_2$ is known as α or the average effect of gene substitution. If there is no dominance, $\alpha_1 = qa$ and $\alpha_2 = -pa$, and the average effect of gene substitution α thus equals the genotypic value a ($\alpha = \alpha_1 - \alpha_2 = qa + pa = (q + p)a = a$).

The *breeding value* of an individual equals the sum of the average effects of gene substitution of an individual's alleles, and is therefore directly related to the mean genetic value of its offspring. Thus, the breeding value for an individual with genotype A1A1 is $2\alpha_1$ (or $2q\alpha$), for individuals with genotype A1A2 it is $\alpha_1 + \alpha_2$ (or $(q - p)\alpha$), and for individuals with genotype A2A2 it is $2\alpha_2$ (or $-2p\alpha$).

The breeding value is usually referred to as the *additive effect* of an allele (note that it includes both the values a and d), and differences between the genotypic effects (in terms of a, d, and $-a$, for genotypes A1A1, A1A2, A2A2 respectively) and the breeding values ($2q\alpha$, $(q - p)\alpha$, $-2p\alpha$, for genotypes A1A1, A1A2, A2A2 respectively) reflect the presence of dominance. Obviously, breeding values are of utmost importance to animal and crop breeders in determining which crossing will produce offspring with the highest milk yield, the fastest race horse, or the largest tomatoes.

Genetic Variance

Although until now we have ignored environmental effects, quantitative geneticists assume that populationwise the phenotype (P) is a function of both genetic (G) and environmental effects (E): P = G + E, where E refers to the environmental deviations, which have an expected average value of zero. By excluding the term GxE, we assume no interaction between the genetic effects and the environmental effects (*see* **Gene-Environment Interaction**). If we also assume there is no covariance between G and E, the variance of the phenotype is given by $V_P = V_G + V_E$, where V_G represents the variance of the genotypic values of all contributing loci including both additive and nonadditive components, and V_E represents the variance of the environmental deviations. Statistically, the total genetic variance (V_G) can be obtained by applying the standard formula for the variance: $\sigma^2 = \sum f_i(x_i - \mu)^2$, where f_i denotes the frequency of genotype i, x_i denotes the corresponding genotypic mean of that genotype, and μ denotes the population mean, as calculated in (1). Thus, $V_G = p^2[a - (a(p - q) + 2pqd)]^2 + 2pq[d - (a(p - q) + 2pqd)]^2 + q^2[-a - (a(p - q) + 2pqd)]^2$. This can be simplified to $V_G = p^2[2q(a - dp)]^2 + 2pq[a(q - p) + d(1 - 2pq)]^2 + q^2[-2p(a + dq)]^2$, and further simplified to $V_G = 2pq[a + d(q - p)]^2 + (2pqd)^2 = V_A + V_D$ [3].

If the phenotypic value of the heterozygous genotype lies midway between A1A1 and A2A2, the total genetic variance simplifies to $2pqa^2$. If d is not equal to zero, the 'additive' genetic variance component contains the effect of d. Even if $a = 0$, V_A is usually greater than zero (except when $p = q$). Thus, although V_A represents the variance due to the additive influences, it is not only a function of p, q, and a but also of d. Formally, V_A represents the variance of the breeding values, when these are expressed in terms of deviations from the population mean. The consequences are that, except in the rare situation in which all contributing loci are diallelic with $p = q$ and $a = 0$, V_A is usually greater than zero. Models that decompose the phenotypic variance into components of V_D, without including V_A, are therefore biologically implausible. When more than one locus is involved and it is assumed that the effects of these loci are uncorrelated and there is no interaction (i.e., no **epistasis**), the V_G's of each individual locus may be summed to obtain the total genetic variances of all loci that influence a trait [4, 5].

In most human quantitative genetic models, the observed variance of a trait is not modeled directly as a function of p, q, a, d, and environmental deviations (as all of these are usually unknown), but instead is modeled by comparing the observed resemblance between pairs of differential, known genetic relatedness, such as monozygotic and dizygotic twin pairs (*see* **ACE Model**). Ultimately, p, q, a, d, and environmental deviations are the parameters that quantitative geneticists hope to 'quantify'.

Acknowledgments

The author wishes to thank Eco de Geus and Dorret Boomsma for reading draft versions of this chapter.

References

[1] Eiberg, H. & Mohr, J. (1987). Major genes of eye color and hair color linked to LU and SE, *Clinical Genetics* **31**(3), 186–191.
[2] Eiberg, H. & Mohr, J. (1996). Assignment of genes coding for brown eye colour (BEY2) and brown hair colour (HCL3) on chromosome 15q, *European Journal of Human Genetics* **4**(4), 237–241.
[3] Falconer, D.S. & Mackay, T.F.C. (1996). Introduction to Quantitative Genetics, Longan Group Ltd, Fourth Edition.
[4] Fisher, R.A. (1918). The correlation between relatives on the supposition of Mendelian inheritance, *Transactions of the Royal Society of Edinburgh: Earth Sciences* **52**, 399–433.
[5] Mather, K. (1949). *Biometrical Genetics*, Methuen, London.
[6] Mather, K. & Jinks, J.L. (1982). *Biometrical Genetics*, Chapman & Hall, New York.

DANIELLE POSTHUMA