## Quantitative Genetics: II - Advanced Topics:

In this section, mathematical models are developed for the computation of different types of genetic variance. Several substantive points about genetic variance components and their effect on the analysis of behavioral data are also made. The reader uninterested in the mathematics can read the two text boxes to gain the substantive conclusions.

Quantitative geneticists partition total genetic variance into three types—*additive, dominance*, and *epistatic* variance. Additive genetic variance measures the extent to which phenotypic individual differences are predictable from the additive effects of allelic substitutions. Dominance genetic variance is variance associated with dominant gene action—the fact that the genetic value for a heterozygote is not exactly the average of the genetic value of the two homozygotes. Epistatic genetic variance is the variance associated with the statistical interaction among loci—gene by gene interaction as it is often called.

Additive and dominance variance may be illustrated by examining a single locus, designated here as M locus, with two alleles $M_1$ and $M_2$. The additive effect of allele $M_2$ is the average change in genotypic values seen by substituting an $M_2$ allele for an $M_1$ allele. To find this effect, simply construct a new variable, called $X_1$ here, that equals the number of $M_2$ alleles for the individual's genotype. For genotype $M_1M_1$, $X_1 = 0$; for $M_1M_2$, $X_1 = 1$; and for $M_2M_2$, $X_1 = 2$. To account for dominance, construct another new variable, $X_2$, with values of 0, 1, and 0 for, respectively, genotypes $M_1M_1$, $M_1M_2$, and $M_2M_2$. Table 18A.1 provides hypothetical data set up with the new variables[1]. It is assumed that the phenotype is scaled so that the population mean is 100 and the population standard deviation is 15.

[Insert Table 18A.1 here]

---

[1] This coding system can be used to account for any number of alleles at a locus. For example, to model additive effects of allele $M_3$, construct another variable giving the number of $M_3$ alleles in a genotype. There would then be three dominance variables, one for the heterozygote $M_1M_2$, a second for the

If we had actual data on individuals we would calculate the additive and dominance effects and their variance components by performing two regression. In the first, we would regress the phenotypic score, noted as $Y$ herein, on $X_1$. The squared multiple correlation from this regression equals the *additive heritability* ($h^2_A$), the proportion of phenotypic variance associated with additive gene action at this locus. For the data in Table 18A.1, $R^2$ = .137 = $h^2_A$, so 13.7% of phenotypic variance is predicted from additive gene effects at this locus. The regression line for this equation will be the strait line of best fit that through the three genotypic values for the three genotypes. It is illustrated in Figure 18A.1.

The second regression model would be of the form

$$Y = b_0 + b_1 X_1 + b_2 X_2 \quad .$$

The intercept, $b_0$, will equal the genotypic value for $M_1M_1$. The regression coefficient $b_1$ equals the average effect of substituting allele $M_2$ for $M_1$ in any genotype. Finally, the coefficient $b_2$ equals the genotypic value of the heterozygote less the average of the genotypic values of the two homozygotes. This measures dominant gene action. For the present example, $b_0 = 94$, $b_1 = 7$, and $b_2 = -5$. The value of coefficient $b_1$ informs us that on average one $M_2$ allele increases phenotypic values by 7 units. Because the value of $b_2$ is not 0, we can conclude that there is some degree of dominance. Because $b_2 = -5$, we can conclude that there is partial dominance for allele $M_1$ so that the genotypic value of the heterozygote is moved 5 units away from the midpoint of the two homozygotes and toward genotype $M_1M_1$.

The multiple correlation from this model equals the additive heritability plus the dominance heritability ($h^2_A + h^2_D$). For the present example, $R^2$ = .162. Dominance heritability can be found by subtracting the $R^2$ from the first regression model from this value: $h^2_D = .162 - .137 = .025$ .

---

heterozygote $M_1M_3$, and the third for the heterozygote $M_2M_3$. For each of the three dominance variables, the

The regression coefficients can also be used to calculate additive and dominance heritability. Let $p_1$ denote the frequency of allele $M_1$ and $p_2$, the frequency of $M_2$; and let $V$ denote the variance of the phenotype (which would be 225 for this example). Then,

$$h^2_{\mathrm{A}} = \frac{2p_1p_2[b_1 + (p_1 - p_2)b_2]^2}{V},$$

and

$$h^2_{\mathrm{D}} = \frac{(2p_1p_2b_2)^2}{V}.$$

You should verify that these equations give the same heritability estimates and the regression procedure.

To examine epistasis, consider the N locus with alleles $N_1$ and $N_2$. Just as we created two new variables for the M locus, we could also create two new variables to model the additive effect and the dominance effect at this locus. Call these variables $X_3$ and $X_4$. For genotypes $N_1N_1$, $N_1N_2$,and $N_2N_2$, the respective values for $X_3$ will be 0, 1, and 2; the respective values for $X_4$ would be 0, 1, and 0. The coding for the additive and the dominance effects at both the M and the N loci are given in Table 18A.2.

[Insert Table 18A.2 here]

In regression, an interaction between two predictor variables is modeled by creating a new variable that is the product of the two predictor variables and entering this variable. Genetic epistasis is modeled in the same way. Multiplying the additive variable for the *M* locus ($X_1$) by the additive variable for the *N* locus ($X_3$) gives a new variable ($X_5$ in Table 18A.2) that geneticists call *additive by additive epistasis*. The variance associated with this is termed *additive by additive epistatic variance*.

There are two different ways to model the interaction between an additive effect at one locus and a dominance effect at a second locus. First, we could multiply the additive variable for the *M* locus by the dominance variable for the *N* locus. This new variable is

---

appropriate heterozygote would have a value of 1 and all other genotypes would be assigned a value of 0.

given as $X_6$ in Table 18A.2. The second way is to multiply the dominance variable for *M* by the additive variable for *N*, giving variable $X_7$ in Table 18A.2. Together variables $X_6$ and $X_7$ model *additive by dominance epistasis* and the variance associated with these two variables is called *additive by dominance epistatic variance*.

The final interactive term is the product of the two dominance variables for the *M* and *N* loci. It is given as $X_8$ in Table 18A.2. This is *dominance by dominance epistasis* and it associated variance is *dominance by dominance epistatic variance*.

Estimation of epistatic variance components proceeds in the hierarchical manner described previously for additive and dominance variance at a single locus. The regression models and their associated variance components are listed in Table 18.3. To calculate the proportion of phenotypic variance associated with any single effect, one simply takes the $R^2$ for the model and subtracts from it the $R^2$ of the model above it in Table 18A.3.

[Insert Table 18A.3 here]

For a numerical example consider the genotypic values presented in Table 18A.4 measured on a scale such as IQ with a population mean of 100 and a population standard deviation of 15. In calculating these numbers, the frequency of allele $M_2$ was set to .60 and the frequency of $N_2$ was .70. Variables $X_1$ through $X_8$ were constructed as in Table 18A.2,

[Insert Table 18A.4 here]

and the regression models in Table 18A.2were fitted to the data. The results of the regression models and the heritability components are given in Table 18A.5. In this table, the heritability due to dominance variance ($h^2_D$) equals the $R^2$ for model 2 less the $R^2$ for model 1 or .20881 - .18321 = .0256. The heritability due to additive by additive epistasis ($h^2_{AA}$) equals the $R^2$ for model 3 less the $R^2$ for model 2 or .23854 - .20881 = .0297, and so on.

[Insert Table 18A.5 here]

Notice how the hierarchical decomposition of genetic variance tends to extract large amounts for the additive variance and progressively smaller amounts for the dominance and epistatic variance.

## TEXT BOX: Gene Action and Genetic Variance Components.

Gene action is required for a variance component. For example, without some degree of dominant gene action, there can be no dominance variance. However, genetic variance components are a function of *both* gene action and genotypic frequencies. Consequently, one can have strong gene action, but if the genotypic frequencies are just right, then the variance component associated with it can be very small. To illustrate this consider the two equations given in the text to compute the additive and dominance variance at a single locus from the regression parameters,

$$h^2_A = 2p_1p_2[b_1 + (p_1 - p_2)b_2]^2 \ ,$$

and

$$h^2_D = (2p_1p_2b_2)^2 \ .$$

Let us assume that allele $M_2$ shows complete dominance to allele $M_1$. In this case, $b_2 = b_1$. Let us substitute $b_1$ for $b_2$ in the above equations and derive the ratio of additive to dominance variance,

$$\frac{h^2_A}{h^2_D} = \frac{2p_1p_2[b_1 + (p_1 - p_2)b_1]^2}{(2p_1p_2b_1)^2}$$

which reduces to

$$\frac{h^2_A}{h^2_D} = 2\frac{p_1}{p_2} \ .$$

Even though we have modeled a completely dominant gene, this equation tells us that the ratio of additive to dominance variance depends *only* on the allele frequencies! When allele frequencies are even (i.e., $p_1 = p_2$) then the ratio is 2 and we will have twice as much additive variance as dominance variance. As the frequency of the recessive allele increases, $p_1$ becomes larger and larger and there is more and more additive variance relative

to dominance variance. This tells us that rare dominant alleles have large additive variance relative to their dominance variance.

As $p_1$ becomes smaller and smaller relative to $p_2$, the ratio will get less than 1 and approach 0. Thus, rare recessive alleles have large dominance variance relative to their additive variance.

## TEXT BOX: Hierarchical Decomposition of Genetic Variance Components

Return to the single locus with two alleles. In performing the regression to calculate additive genetic variance, we fitted a regression line that minimizes the squared differences between the observed genotypic values and those predicted by the regression line. This procedure is deliberately geared to maximize additive genetic variance (the variance associated with the regression line) and minimize residual genetic variance (the dominance genetic variance).

The hierarchical decomposition of genetic variance components follows this maximization/minimization algorithm. In the polygenic case, after additive variance has been extracted the procedure will try to maximize dominance genetic variance and minimize the residual genetic variance (epistatic variance). After additive and dominance variances have been extracted, the regression will maximize the additive by additive epistatic variance and minimize the other epistatic variance components.

As a consequence, additive genetic variance tends to be the largest component with continually smaller and smaller components following. There is no mathematical guarantee that this will *always* happen, but the pattern is almost always expressed in biologically plausible models of gene action. The major exception to this rule occurs when phenotypes are due to rare recessive alleles at a single locus.

Any human behavioral trait is probably influenced by several different genes. It is unlikely that nonadditive gene action (dominance and epistasis) are completely absent at all loci that contribute to behavior. Would anyone care to bet that the genotypic value of the heterozygote lies exactly at the midpoint of the two homozygotes for every single locus operating in the central nervous system? But the hierarchical decomposition of variance for polygenic traits is likely to generate considerable additive variance with relatively small

dominance and epistatic variance. The net result is that the typical assumption used in fitting genetic models to human behavioral data—that all genetic variance is additive—will give the wrong answer but will not give a substantively misleading answer.

To illustrate, let us examine what twin correlations would look like using the variance components given in Table 18A.5. The identical twin correlation would be

$$R_{mz} = h_A^2 + h_D^2 + h_{AA}^2 + h_{AD}^2 + h_{DD}^2 = .243 \ ,$$

and the fraternal twin correlation would be

$$R_{dz} = \frac{1}{2}h_A^2 + \frac{1}{4}h_D^2 + \frac{1}{4}h_{AA}^2 + \frac{1}{8}h_{AD}^2 + \frac{1}{16}h_{DD}^2 = .106 \ .$$

(See Kempthorne (**NEED REF HERE**) for the coefficients for nonadditive effects for relatives other than MZ twins.) If we estimated heritability with the traditional formula that assumes no additive genetic variance, we would have

$$\hat{h}^2 = 2(R_{mz} - R_{dz}) = 2(.243 - .106) = .274 \ .$$

Although the estimate of .274 is not correct, it is not very different from the total heritability of .243.

Table 18A.1. Hypothetical data for estimating genetic variance components at a single locus.

| | | | Numerical Codes: | |
|---|---|---|---|---|
| Genotype | Frequency | Genotypic Value | Additive = $X_1$ | Dominance = $X_2$ |
| $M_1M_1$ | .16 | 94 | 0 | 0 |
| $M_1M_2$ | .48 | 96 | 1 | 1 |
| $M_2M_2$ | .36 | 108 | 2 | 0 |

Table 18A.2. An example of numerical coding to calculate genetic variance components for two loci with two alleles at each locus.

| | | **Contrast Codes:** | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | **Interactive** | | | |
| | | ***M* Locus:** | | ***N* Locus:** | | A*A | A*D | D*A | D*D |
| **Genotypes:** | | Add. | Dom. | Add. | Dom. | $X_1*X_3$ | $X_1*X_4$ | $X_2*X_3$ | $X_2*X_4$ |
| | | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | $X_7$ | $X_8$ |
| $M_1M_1N_1N_1$ | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $M_1M_1N_1N_2$ | | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| $M_1M_1N_2N_2$ | | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 |
| $M_1M_2N_1N_1$ | | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| $M_1M_2N_1N_2$ | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| $M_1M_2N_2N_2$ | | 1 | 1 | 2 | 0 | 2 | 0 | 2 | 0 |
| $M_2M_2N_1N_1$ | | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $M_2M_2N_1N_2$ | | 2 | 0 | 1 | 1 | 2 | 2 | 0 | 0 |
| $M_2M_2N_2N_2$ | | 2 | 0 | 2 | 0 | 4 | 0 | 0 | 0 |

Table 18A.3. Regression models for estimating heritability components.

| | Model: | Heritability Component ($R^2$): |
|---|---|---|
| 1 | $Y = X_1 + X_3$ | Additive = $h^2_A$ |
| 2 | $Y = X_1 + X_2 + X_3 + X_4$ | Additive + Dominance = $h^2_A + h^2_D$ |
| 3 | $Y = X_1 + X_2 + X_3 + X_4 + X_5$ | Additive + Dominance + A*A Epistasis = $h^2_A + h^2_D + h^2_{AA}$ |
| 4 | $Y = X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7$ | Additive + Dominance + A*A Epistasis + A*D Epistasis = $h^2_A + h^2_D + h^2_{AA} + h^2_{AD}$ |
| 5 | $Y = X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8$ | Additive + Dominance + A*A Epistasis + A*D Epistasis + D*D Epistasis = $h^2_A + h^2_D + h^2_{AA} + h^2_{AD} + h^2_{DD}$ |

Table 18A.4. Genotypic values for two genotypes.

| | $N_1N_1$ | $N_1N_2$ | $N_2N_2$ | Mean |
|---|---|---|---|---|
| $M_2M_2$ | 93.00 | 103.78 | 114.37 | 108.00 |
| $M_1M_2$ | 93.00 | 95.00 | 97.41 | 96.00 |
| $M_1M_1$ | 93.00 | 94.00 | 94.18 | 94.00 |
| Mean | 93.00 | 98.00 | 103.00 | 100.00 |

Table 18A.5. Components of heritability for the data in Table 18A.4.

| | Model: | $R^2$ | Heritability Component |
|---|---|---|---|
| 1 | $Y = X_1 + X_2$ | .18321 | $h^2_A$ = .1832 |
| 2 | $Y = X_1 + X_2 + X_3 + X_4$ | .20881 | $h^2_D$ = .0256 |
| 3 | $Y = X_1 + X_2 + X_3 + X_4 + X_5$ | .23854 | $h^2_{AA}$ = .0297 |
| 4 | $Y = X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7$ | .24309 | $h^2_{AD}$ = .0046 |
| 5 | $Y = X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8$ | .24314 | $h^2_{DD}$ = .0001 |

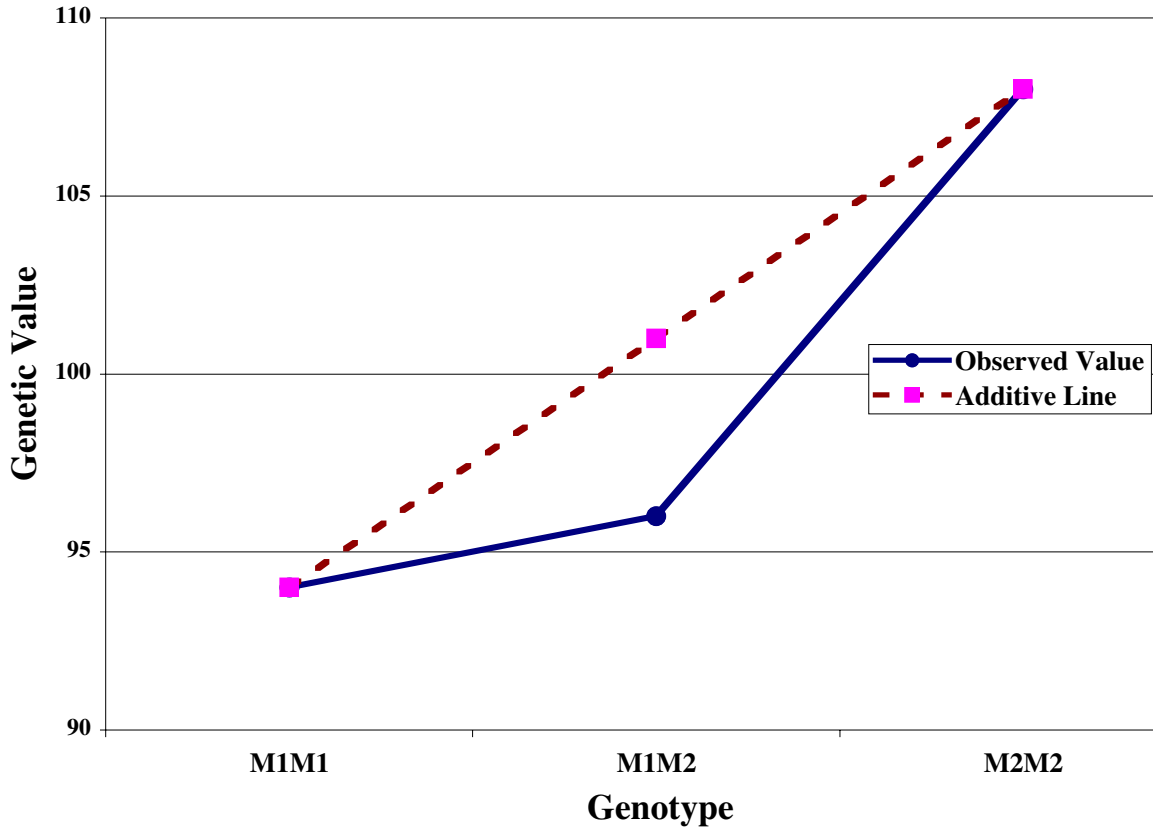


Figure 18A.1. Observed genetic values and the linear regression line of best fit.

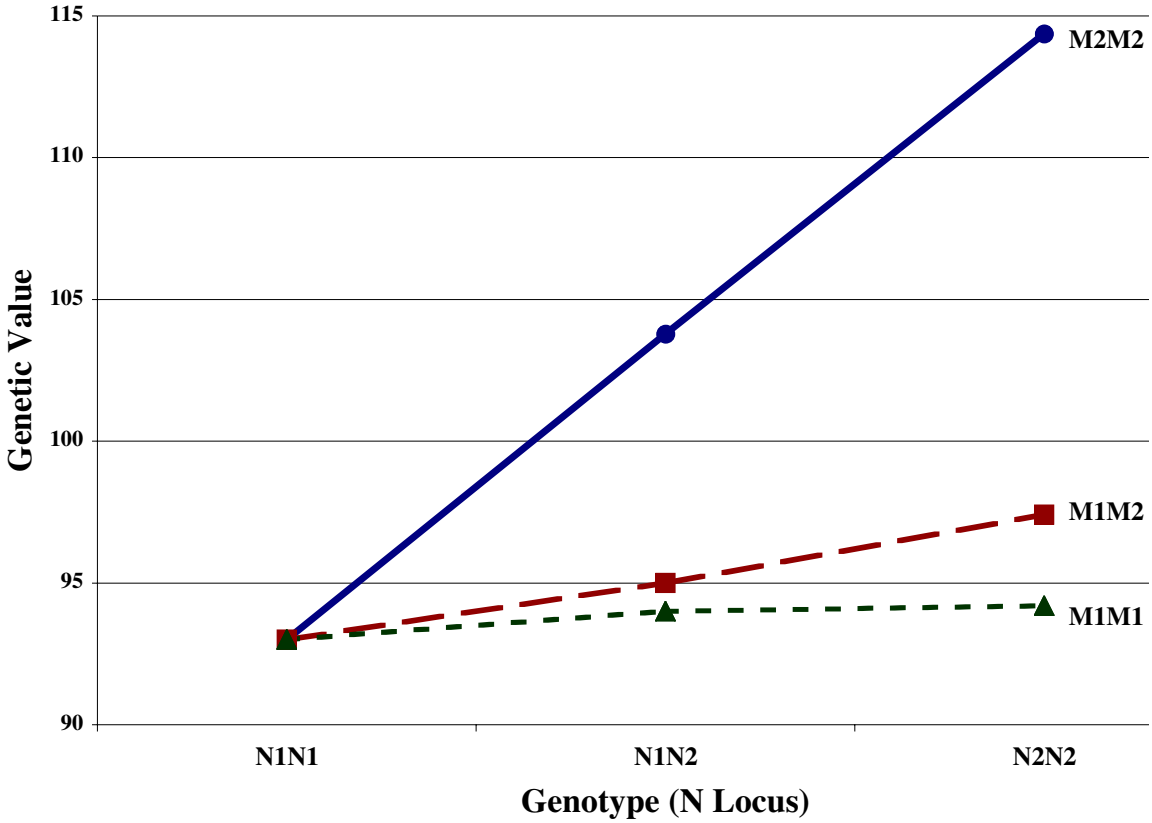


Figure 18A.2. Plot of the genetic values for two loci. Epistasis or gene-gene interaction is evident because the three lines are not parallel.