### POLYGENIC OR QUANTITATIVE INHERITANCE: NONINBRED POPULATIONS

**Graphic Outline of Chapter 8**

<table>
<thead>
<tr>
<th>No epistasis</th>
<th>Epistasis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two or more loci</td>
<td>Two loci</td>
</tr>
<tr>
<td>Multiple alleles</td>
<td>Multiple alleles</td>
</tr>
</tbody>
</table>

#### Arbitrary 'male' and 'female' allelic freq. (cross between two pop.)

<table>
<thead>
<tr>
<th>Multiple alleles</th>
<th>Two alleles</th>
<th>Multiple alleles</th>
<th>Two alleles</th>
<th>Multiple alleles</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Section</strong> 8.1.1</td>
<td><strong>Section</strong> 8.1.3</td>
<td><strong>Section</strong> 8.1.6</td>
<td><strong>Section</strong> 8.3.1</td>
<td><strong>Section</strong> 8.3.5</td>
</tr>
<tr>
<td><strong>page</strong> 8.4</td>
<td><strong>page</strong> 8.20</td>
<td><strong>(1) linkage eq.</strong></td>
<td><strong>(1) linkage eq.</strong></td>
<td><strong>(1) linkage eq.</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>(2) linkage diseq.</strong></td>
<td><strong>(2) linkage diseq.</strong></td>
<td><strong>(2) linkage diseq.</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>page</strong> 8.57</td>
<td><strong>page</strong> 8.76</td>
<td><strong>page</strong> 8.127</td>
</tr>
</tbody>
</table>

#### Equal male and female allelic freq. (Hardy-Weinberg)

<table>
<thead>
<tr>
<th>Multiple alleles</th>
<th>Two alleles</th>
<th>Multiple alleles</th>
<th>Two alleles</th>
<th>Multiple alleles</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Section</strong> 8.1.2</td>
<td><strong>Section</strong> 8.1.4</td>
<td><strong>Section</strong> 8.1.5</td>
<td><strong>Section</strong> 8.1.7</td>
<td><strong>Section</strong> 8.3.3</td>
</tr>
<tr>
<td><strong>page</strong> 8.19</td>
<td><strong>page</strong> 8.28</td>
<td><strong>p₁=p₂</strong> or <strong>p₁=p₂</strong></td>
<td><strong>p₁=p₂</strong> or <strong>p₁=p₂</strong>, no dominance</td>
<td><strong>Page</strong> 8.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>Page</strong> 8.39</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>Page</strong> 8.39</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>Page</strong> 8.64</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>Page</strong> 8.96</td>
</tr>
</tbody>
</table>

**Section** 8.3.6 | **(1) linkage eq.** | **(1) linkage eq.** | **Page** 8.119 |
| | **(2) linkage diseq.** | **(2) linkage diseq.** | | **Page** 8.135 |
# CHAPTER 8

## POLYGENIC OR QUANTITATIVE INHERITANCE: NONINBRED POPULATIONS

<table>
<thead>
<tr>
<th>Table of Contents</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.1 Means, allelic effects, and variances</td>
<td>8.4</td>
</tr>
<tr>
<td>8.1.1 One locus, multiple alleles, arbitrary &quot;male&quot; and &quot;female&quot; allelic frequencies (cross between two populations): General</td>
<td>8.4</td>
</tr>
<tr>
<td>8.1.2 One locus, multiple alleles, equal male and female allelic frequencies (Hardy-Weinberg)</td>
<td>8.19</td>
</tr>
<tr>
<td>8.1.3 One locus, two alleles, arbitrary &quot;male&quot; and &quot;female&quot; allelic frequencies (cross between two populations): General</td>
<td>8.20</td>
</tr>
<tr>
<td>8.1.4 One locus, two alleles, equal male and female allelic frequencies (Hardy-Weinberg) with arbitrary or equal allelic frequencies within sexual arrays and with dominance</td>
<td>8.28</td>
</tr>
<tr>
<td>1. Factorial approach</td>
<td>8.28</td>
</tr>
<tr>
<td>2. Regression approach</td>
<td>8.30</td>
</tr>
<tr>
<td>8.1.5 One locus, two alleles, equal male and female allelic frequencies (Hardy-Weinberg) with arbitrary or equal allelic frequencies within sexual arrays and with no dominance</td>
<td>8.39</td>
</tr>
<tr>
<td>8.1.5.1 Change of population mean</td>
<td>8.46</td>
</tr>
<tr>
<td>8.1.5.2 Changes in additive, dominance, and genotypic variances</td>
<td>8.48</td>
</tr>
<tr>
<td>1. No dominance (d = 0)</td>
<td>8.48</td>
</tr>
<tr>
<td>2. Complete dominance (d = a)</td>
<td>8.50</td>
</tr>
<tr>
<td>3. Partial dominance (0 &lt; d &lt; a)</td>
<td>8.54</td>
</tr>
<tr>
<td>4. &quot;Pure&quot; overdominance (a = -a = 0, d &gt; 0)</td>
<td>8.56</td>
</tr>
<tr>
<td>5. Overdominance (a &gt; -a, d &gt; a)</td>
<td>8.56</td>
</tr>
</tbody>
</table>
8.1.6 Two or more loci, multiple alleles, arbitrary "male" and "female" allelic frequencies (cross between two populations), no epistasis: General ........................................ 8.57
  1. Linkage equilibrium ........................................ 8.57
  2. Linkage disequilibrium ..................................... 8.63

8.1.7 Two or more loci, multiple alleles, equal male and female allelic frequencies (Hardy-Weinberg), no epistasis .................................................. 8.64
  1. Linkage equilibrium ........................................ 8.64
  2. Linkage disequilibrium ..................................... 8.66

8.2 Relating formulas herein to those presented by other authors .............. 8.66

8.2.1 Scaling equivalences ........................................ 8.66
  1. To convert any equation in the A, D, U, or W scale into that of the G scale 8.67
  2. To convert any equation in the G scale into that of another scale .................. 8.71

8.2.2 Separation of additive variance into Mather’s genetic parameters ........... 8.74

8.3 Gene effects and variances with epistasis ..................................... 8.77

8.3.1 Two loci, multiple alleles, arbitrary "male" and "female" allelic frequencies (cross between two populations), epistasis: General ........................................ 8.77
  1. Linkage equilibrium ........................................ 8.77
  2. Linkage disequilibrium ..................................... 8.96

8.3.2 Two loci, multiple alleles, equal male and female allelic frequencies at each locus (Hardy-Weinberg), epistasis ............................................. 8.96
  1. Linkage equilibrium ........................................ 8.96
  2. Linkage disequilibrium ..................................... 8.98

8.3.3 Two loci, two alleles, equal male and female allelic frequencies at each locus (Hardy-Weinberg), linkage equilibrium, epistasis. Explicit expressions for epistatic variances ........................................ 8.98

8.3.4 Two loci, two alleles, equal male and female allelic frequencies at each locus (Hardy-Weinberg), linkage disequilibrium, epistasis. Effect of linkage disequilibrium ........................................ 8.119
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.3.5</td>
<td>More than two loci, multiple alleles, arbitrary &quot;male&quot; and &quot;female&quot; allelic frequencies (cross between two populations), epistasis. General</td>
<td>8.127</td>
</tr>
<tr>
<td>1.</td>
<td>Linkage equilibrium</td>
<td>8.127</td>
</tr>
<tr>
<td>2.</td>
<td>Linkage disequilibrium</td>
<td>8.134</td>
</tr>
<tr>
<td>8.3.6</td>
<td>More than two loci, multiple alleles, equal male and female frequencies at each locus (Hardy-Weinberg), epistasis</td>
<td>8.135</td>
</tr>
<tr>
<td>1.</td>
<td>Linkage equilibrium</td>
<td>8.135</td>
</tr>
<tr>
<td>2.</td>
<td>Linkage disequilibrium</td>
<td>8.137</td>
</tr>
<tr>
<td>8.4</td>
<td>Covariances of characters</td>
<td>8.137</td>
</tr>
<tr>
<td>8.4.1</td>
<td>Covariance of two characters</td>
<td>8.138</td>
</tr>
<tr>
<td>8.4.1.1</td>
<td>One locus, multiple alleles, arbitrary &quot;male&quot; and &quot;female&quot; allelic frequencies (cross between two populations): General</td>
<td>8.138</td>
</tr>
<tr>
<td>8.4.1.2</td>
<td>One locus, multiple alleles, equal male and female allelic frequencies (Hardy-Weinberg)</td>
<td>8.139</td>
</tr>
<tr>
<td>8.4.1.3</td>
<td>Two or more loci, multiple alleles, arbitrary &quot;male&quot; and &quot;female&quot; allelic frequencies (cross between two populations), linkage equilibrium, no epistasis. General</td>
<td>8.140</td>
</tr>
<tr>
<td>8.4.1.4</td>
<td>Two or more loci, multiple alleles, equal male and female allelic frequencies (Hardy-Weinberg), linkage equilibrium, no epistasis</td>
<td>8.142</td>
</tr>
<tr>
<td>8.4.2</td>
<td>Variance of an index: An application</td>
<td>8.142</td>
</tr>
<tr>
<td>8.4.2.1</td>
<td>One locus, multiple alleles, arbitrary &quot;male&quot; and &quot;female&quot; allelic frequencies (cross between two populations): General</td>
<td>8.142</td>
</tr>
<tr>
<td>8.4.2.2</td>
<td>One locus, multiple alleles, equal male and female allelic frequencies (Hardy-Weinberg)</td>
<td>8.143</td>
</tr>
<tr>
<td>8.4.2.3</td>
<td>Two or more loci, multiple alleles, arbitrary &quot;male&quot; and &quot;female&quot; allelic frequencies (cross between two populations), linkage equilibrium, no epistasis: General</td>
<td>8.143</td>
</tr>
<tr>
<td>8.4.2.4</td>
<td>Two or more loci, multiple alleles, equal &quot;male&quot; and &quot;female&quot; allelic frequencies (Hardy-Weinberg), linkage equilibrium, no epistasis</td>
<td>8.144</td>
</tr>
</tbody>
</table>
CHAPTER 8

POLYGENIC OR QUANTITATIVE INHERITANCE: NONINBRED POPULATIONS

Polygenic inheritance is the study of the inheritance of multifactorial traits -- characters which are determined by genes at many loci. Three kinds of multifactorial traits may be distinguished: (1) continuous traits, (2) meristic traits, and (3) threshold traits (Hartl and Clark, 1989, p. 431; Falconer, 1989, p. 106). Continuous traits, often called quantitative traits, are ones for which there is a continuum of phenotypes, e.g., height, weight, grain yield, milk yield, and growth rate. We will most commonly consider this kind of multifactorial trait. Meristic traits are ones for which the phenotype is expressed in discrete classes, e.g., number of offspring in a litter, number of flowers in an inflorescence, and number of bristles on a fruit fly. Even though the phenotypic values are discrete, there is visualized an underlying scale of quantitative effects with multiple thresholds. Threshold traits are ones that are either present or absent in any one individual; there are only two phenotypic classes. For example, an individual manifests a disease or does not, or an individual of a mammalian species, which commonly bears only one offspring at a time, gives birth to twins (or multiple births) or does not. Quantitative genetics is the study of genetic principles underlying the inheritance of multifactorial traits.

One of the main reasons for our interest in the study of quantitative genetics is to predict the outcome of selection and to determine the best method of performing the selection.
To study the genetics of polygenic traits, we must first subdivide the phenotypic value into that part due to genotype and that to environment. Thus, we have the model

\[ P = G + E \]  \hspace{1cm} (8.1)

where \( P \) = phenotypic value of an individual,
\( G \) = genotypic value of an individual,
\( E \) = environmental effect.

The phenotypic value \( P \) of an individual is that value of any character actually observed, but in practice, it may be subject to an adjustment -- the same one as that of the environmental effect as discussed below. The phenotypic value may be the number of bristles on the ventral side of the first two abdominal segments of a Drosophila individual reared in a culture bottle in a controlled environmental room, number of grams of grain from an individual plant in a block or replication on some experimental farm in a given year, number of offspring in a litter of pigs from an individual sow in a particular breed on a particular farm in a given year, or 300-day milk yield from an individual cow in the herd of some dairyman in a given year. The genotypic value \( G \) is conceptually the mean phenotypic value of a large number of individuals of the same genotype, observed under all possible uncontrollable environmental conditions. It is conceptual or hypothetical in that in most cases it is not possible to observe more than one individual of any given genotype. More than one individual of the same genotype is possible only when one has a pure line, as is common in a self-fertilizing species, or an \( F_1 \) cross between two fully inbred lines, or clonally propagated individuals, or other replicated individuals such as identical twins. It is the genotypic value that we shall be initially concerned with in this chapter. The environmental effect \( E \) represents the departure of the observed phenotypic value of an individual of a given genotype from its genotypic value. Since all individuals cannot be in the
same place at any given time, individuals within a population experience different environmental conditions. The environmental effect is due to the composite of different environmental conditions that an individual is exposed to from the time of fertilization to the time when it is measured. In most situations, the environmental effect includes only those environmental effects which are uncontrollable by the experimenter. For example, in plants the environmental effect does not include that due to classifiable environmental effects such as blocks, locations, or years. In the simplest situation, the environmental effect represents the effect of environmental variation from plant to plant of the same genotype within a single block of a randomized complete block design. Likewise, in animals the environmental effect usually has had classifiable effects such as hatch, herd (or management effect), season, year, sex, and age removed. We will have more to say about the whole matter in Section 8.

We now turn our attention to the genotypic value of an individual. With sexual reproduction, an individual does not transmit its genotype intact to its offspring. Instead it transmits only a single random gene at any given locus to its offspring. That single gene is combined with another single random gene at that same locus, transmitted by the other parent to the offspring. In each generation the genotypes are recombined or reconstituted anew. Thus, Fisher (1918) in his classical paper reasoned that since individuals pass on their genes, not their genotype, an average effect for each of the two genes carried by the individual must be defined such that their sum plus the population mean will best predict that individual's genotypic value. To define the average effect of an allele leads to the whole idea of genotypic structure of the population -- the array of genotypes in the population and its corresponding array of genotypic frequencies. To define the average effect of an allele, we must make certain assumptions about the frequencies of occurrence of genotypes in the population.
We will initially define the genotypic structure at one locus in its broadest sense (Section 8.1.1), and we will obtain other specific cases of interest by imposing certain restrictions (Sections 8.1.2 to 8.1.5). Then we will do similarly for two loci with no epistasis (Sections 8.1.6 and 8.1.7). Then in later sections we will introduce epistasis.

8.1. Means, allelic effects, and variances. Assumptions:

8.1.1. One locus, multiple alleles, arbitrary "male" and "female" allelic frequencies (cross between two populations): General. Initially a single autosomal locus is considered. Mating is assumed to be at random (see Section 3.2), so all genes are randomly associated or are independent, i.e., the frequency of every genotype is determined solely by the frequencies of the alleles which it carries, namely, \( p_{ij} = p_i^m f_j \). This condition exists in a population in Hardy-Weinberg equilibrium (\( p_i^m = p_i^f = p_i \) for \( i = 1, \ldots, m \), so \( p_{ij} = p_i p_j \) (2.102) (3.26)), or in a cross between two populations, population \( m \) (not necessarily males only) and population \( f \) (not necessarily females only) (\( p_i^m \neq p_i^f \) for \( i = 1, \ldots, m \), so \( p_{ij} = p_i^m p_j^f \)). In this latter case, \( m \) is the total number of different alleles in the two populations considered together. One or more alleles may be absent in the \( m \) population, but present in the \( f \) population. In this case, \( p_i^m \) equals zero for those alleles that are absent in the \( m \) population. In addition, in the \( f \) population one or more different alleles may also be absent for which their frequencies \( p_i^f \) are zero (see second paragraph, Section 3.4.4).

The model to be presented will be used primarily as a definitional system of effects and variances, but it does lend itself for analysis also. Normally all the genotypic values and frequencies are unknown, but we start out assuming that we know all of them and then ask what we can estimate in an analysis.
Let the value of the genotype of an individual be

$$G_{ij} = \mu + \alpha_i^m + \alpha_j^f + \delta_{ij}$$  \hspace{1cm} (8.2)

where $G_{ij}$ = genotypic value, or the true average measurement of an individual, attributable to genotype $A_iA_j$ over the population of all possible uncontrollable environments,

$\mu$ = mean of population,

$\alpha_i^m$ = additive or average effect of allele $A_i$, $i = 1, \ldots, m$, received from its male parent (commonly we will equate superscript $m$ to male, but more generally, $m$ simply denotes the $m$ parent being either a male or female individual from the population designated $m$),

$\alpha_j^f$ = additive or average effect of allele $A_j$, $j = 1, \ldots, m$, received from its female parent,

$\delta_{ij}$ = dominance effect or interaction effect between alleles $A_i$ and $A_j$ in the sense that there is lack of complete additivity based on the sum of the values $\alpha_i^m$ and $\alpha_j^f$.

The genotypic value may be designated as that for either a male or a female individual, $\delta$ or $\varphi$, if the sexes differ in their expression. If the sexes do not differ, no distinction needs to be made. It is the latter situation that we assume throughout these notes. If, however, sexual expressions are different, then every thing that is defined herein can be regarded as that for either the male or female subpopulation. In reality, two subpopulations exist within the one. Considering the right side of this basic linear additive model (8.2), its rationale is that the genotypic value or true value of an individual with genotype $A_iA_j$ is determined by the population mean ($\mu$) and its two genes. Hence, a term exists to denote the "average" effect of each of the two alleles, namely, $\alpha_i^m$ and $\alpha_j^f$. These two terms plus the population mean $\mu$ best predict that individual's
genotypic value in a least-squares sense. Finally, since we know that from elementary genetics gene effects do not combine in an additive sense but show dominance or interaction, we must include a term \( \delta \) to account for such a phenomenon. The \( \delta_{ij} \) effect is an intra-locus interaction or the interaction between the two genes carried by the individual at that locus.

Next we desire to give the least-squares definitions of \( \mu, \alpha_i^m, \) and \( \alpha_j^f, \) \( i, j = 1, \ldots, m, \) in terms of the genotypic values \( G_{ij}, \) when uniting alleles are independent, i.e.,

\[
p_{ij} = p_i^m p_j^f \quad \text{for } i, j = 1, \ldots, m \tag{8.3}
\]

Fisher (1918) first set forth the definitional system given below in that it gives the rate at which a population mean changes under selection (Section 8.1.5.1). It has also led to a natural formulation for the covariance between relatives (Chapter 9). From (8.2) we have

\[
G_{ij} = \mu + \alpha_i^m + \alpha_j^f + \delta_{ij}
\]

Then weighting the square of both sides by the genotypic frequency, and summing over both subscripts, we have

\[
Q = \sum_{i=1}^{m} \sum_{j=1}^{m} p_i^m p_j^f (G_{ij} - \mu - \alpha_i^m - \alpha_j^f)^2 = \sum_{i=1}^{m} \sum_{j=1}^{m} p_i^m p_j^f \delta_{ij}^2 \tag{8.5}
\]

What we want to do next is to find the values of \( \mu, \alpha_i^m, \) and \( \alpha_j^f, \) \( i, j = 1, \ldots, m, \) which will minimize \( Q \) or the dominance variance -- the weighted squares of the dominance deviations. To do so, we partially differentiate \( Q \) with respect to each of the unknowns to obtain the normal equations, namely (see Box 8.1),

\[
\mu: \quad \mu = \sum_{i} p_i^m \alpha_i^m + \sum_{j} p_j^f \alpha_j^f = \sum_{i} p_i^m p_j^f G_{ij} \tag{8.6a}
\]

\[
\alpha_i^m: \quad p_i^m \mu + p_i^m \alpha_i^m + \sum_{j} p_j^f \alpha_j^f = p_i^m \sum_{j} p_j^f G_{ij} \quad i = 1, \ldots, m \tag{8.6b}
\]

\[ \alpha^f: \quad f_{j\mu} + \sum_i p_{ij}^m = p_{ij}^m \alpha_i^m + p_{ij}^f f - p_{ij}^m G_{ij}^m \quad j = 1, \ldots, m \] (8.6c)

Box 8.1

Derivation of (8.6)

1. Normal equation for \( \mu \). Differentiating (8.5)

\[ Q = \sum_{i=1}^{m} \sum_{j=1}^{m} p_{ij}^m (G_{ij} - \mu - \alpha_i^m - \alpha_j^f)^2 = \sum_{i=1}^{m} \sum_{j=1}^{m} p_{ij}^m \delta_{ij} \]

with respect to \( \mu \), and setting the derivative equal to zero, we obtain

\[ \frac{\partial Q}{\partial \mu} = -2 \sum_{i=1}^{m} \sum_{j=1}^{m} p_{ij}^m (G_{ij} - \mu - \alpha_i^m - \alpha_j^f) = 0 \] (1)

Then dividing both sides of the equation by -2, and moving the summations inside the parentheses, we obtain

\[ \sum_j p_{ij}^m f - \sum_i \sum_j p_{ij}^m \alpha_i^m - \sum_i \sum_j p_{ij}^m \alpha_j^f = 0 \]

By summing over all "free" subscripts associated with each term, we have

\[ \sum_j p_{ij}^m f (G_{ij} - (\sum_i p_{ij}^m \alpha_i^m - \sum_j p_{ij}^m \alpha_j^f) = 0 \]

Free subscripts are any subscripts not present on \( G, \mu, \alpha, \) or \( \delta \) as the case may be. There are no free subscripts on \( G_{ij} \) or \( \delta_{ij} \), but there are two \((i \text{ and } j)\) for the \( \mu \) term, one \((j)\) for the \( \alpha_i^m \) term, and one \((i)\) for the \( \alpha_j^f \) term. It is random mating that permits us to write the above expression, so that free subscripts exist. Then using (2.4), namely,

\[ \sum_{i=1}^{m} p_{ij}^m = \sum_{j=1}^{m} p_{ij}^m = 1 \] (2)

and moving the "observed" or \( G \) terms to the right side, we have

\[ \mu + \sum_i p_{ij}^m \alpha_i^m + \sum_j p_{ij}^m \alpha_j^f = \sum_i \sum_j p_{ij}^m G_{ij}^f \] (3)

which is the normal equation for \( \mu \).
2. Normal equations for $\alpha_i^m$. Differentiating (8.5) or (1) again with respect to each $\alpha_i^m$, and setting each derivative equal to zero, we have for each $\alpha_i^m$

$$\frac{\partial Q}{\partial \alpha_i^m} = -2 \sum_{j=1}^{m} p_i^m f(G_{ij} - \mu - \alpha_i^m - \alpha_j^f) = 0 \quad \text{for } i = 1, \ldots, m$$

Dividing by -2, moving the summations inside the parentheses, and summing over "free" subscripts, we have

$$p_i^m \sum_j f_{G_{ij}} - p_i^m (\Sigma_j f_{p_j}) \mu - p_i^m (\Sigma_j f_{p_j} \alpha_i^m) - p_i^m (\Sigma_j f_{p_j} \alpha_j^f) = 0$$

Then substituting (2) and moving the $G$ term to the right-hand side, we obtain

$$p_i^{m+f} + p_i^{m} \alpha_i^m + p_j^{m} \Sigma f_{p_j} \alpha_j^f - p_i^m (\Sigma_j f_{p_j} \alpha_i^m) - p_i^m (\Sigma_j f_{p_j} \alpha_j^f) = 0 \quad \text{for } i = 1, \ldots, m \quad (4)$$

which are the normal equations for $\alpha_i^m$'s. (Note that it is important not to divide both sides by $p_i^m$ (see 8.19).)

3. Normal equations for $\alpha_j^f$. Differentiating (8.5) or (1) again with respect to $\alpha_j^f$, and setting each derivative equal to zero

$$\frac{\partial Q}{\partial \alpha_j^f} = -2 \sum_{i=1}^{m} p_i^m f(G_{ij} - \mu - \alpha_i^m - \alpha_j^f) = 0 \quad \text{for } j = 1, \ldots, m$$

Summing over the "free" subscripts, we have

$$p_j^f \sum_i p_i^{m} G_{ij} - p_j^f (\Sigma_i p_i^{m}) \mu - p_j^f (\Sigma_i p_i^{m} \alpha_i^m) - p_j^f (\Sigma_i p_i^{m} \alpha_j^f) = 0$$

Then substituting (2) and moving the $G$ term to the right side, we obtain

$$p_j^{f} \mu + p_j^f \Sigma_i p_i^{m} \alpha_i^m + p_j^f \Sigma_i p_i^{m} \alpha_j^f - p_j^f (\Sigma_i p_i^{m} \alpha_i^m) - p_j^f (\Sigma_i p_i^{m} \alpha_j^f) = 0 \quad \text{for } j = 1, \ldots, m \quad (5)$$

which are the normal equations for the $\alpha_j^f$'s.

We now desire to solve for the parameters or effects in the model. If the m normal equations for $\alpha_i^m$ in (8.6b) are added, we obtain the same equation as for $\mu$
(8.6a) which shows that the \( m \) \( \alpha^m \)-equations are not independent of the normal equation for \( \mu \). In order to solve these equations we must introduce another equation or auxiliary relationship. A reasonable one is

\[
\sum_{i=1}^{m} \Sigma p_i \alpha_i = 0
\]  

(8.7)

Similarly, the \( m \) equations for \( \alpha^f_j \) (8.6c) are not independent of the normal equation for \( \mu \), so we make the restriction

\[
\sum_{j=1}^{m} \Sigma p_j \alpha_j = 0
\]  

(8.8)

The main argument for the reasonableness of the restrictions (8.7) (8.8) is that with them the expectation of \( G \) becomes \( \mu \) (see (8.9) below). With these restrictions we may solve for the effects in the model. We substitute (8.7) and (8.8) in (8.6a) to obtain

\[
\mu: \quad \mu = \sum_{i=1}^{m} \sum_{j=1}^{m} p_i p_j G_{ij} = E(G_{ij}) = G..
\]  

(8.9)

Notice that another symbol for \( \mu \) is \( G_.. \), where each dot signifies summation over the subscript which the dot has replaced. This common statistical convention will be followed throughout.

Again making the substitution (8.8) in (8.6b), we obtain

\[
P_i \mu + \sum_{i=1}^{m} \sum_{j=1}^{m} P_i p_j G_{ij}
\]

\[
\sum_{i=1}^{m} \sum_{j=1}^{m} p_i p_j G_{ij} - P_i \mu
\]

\[
= P_i (\sum_{j=1}^{m} p_j G_{ij} - \mu)
\]

\[
\alpha^m: \quad \alpha_i^m = \sum_{j=1}^{m} p_j G_{ij} - \mu = G_i - G.. \quad \text{for } i = 1, \ldots, m
\]  

(8.10)
Note that $\alpha^m_1$ is the difference between a weighted mean of all genotypic values for those genotypes which carry the allele $A_1$ from the male parents and the overall (cross) population mean. That weighted mean is the average or expected value of genotypes or individuals receiving $A_1$ from any parent in the $m$ population and the other gene at random from the $f$ population, or it is the mean of all offspring of the $A_1A_1$ homozygote. In that weighted mean, the genotypic values are weighted according to the respective frequency of occurrence of every allele from the female parents or $f$ population, because these genotypic frequencies are proportional to the allelic frequencies of the $f$ population with random mating. As such that weighted mean is a property of the $f$ population. Any $\alpha^m_1$ or average effect of an allele is a property of the corresponding allele $A_1$, but it is also a property of the particular population (the $f$ population, in this case) in that the magnitude of $\alpha^m_1$ obviously depends upon the allelic frequencies in that $f$ population. If the allelic frequencies in that $f$ population change, $\alpha^m_1$ will also change. Although $\alpha^m_1$ is labeled with a superscript $m$, because that allele came from the $m$ population, it is not really a property of the $m$ population, but a property of the $f$ population. From an operational viewpoint, this average effect of an allele ($\alpha^m_1$) is the deviation of the mean of all offspring of a homozygote for that allele in the $f$ population from the mean of the crossed population. The $\alpha^m_1$ differ from the corresponding ones for the $f$ population per se by a constant -- the difference between the mean of the cross population and the mean of the $f$ population.

Similarly from (8.6c) we obtain

$$\alpha^f_j = \frac{\Sigma}{i=1}^m \alpha^m_i c_{ij} - \mu = G_j - G.. \quad \text{for } j = 1, \ldots, m \quad (8.11)$$

These means and effects (8.9) (8.10) (8.11) may be arrived at easiest from a factorial table shown in Table 8.1, and are analogous to estimation in a two-
factor experiment. However, definitions are involved here and not estimation in that \( G_{ij} \), \( G_i \), \( G_j \), and \( G_{..} \) are true means. These effects are defined for an infinitely large population. Note that

\[
G_{i..} = \sum_{j} \frac{m_f}{p_i} G_{ij} = \frac{m}{p_i} \sum_{j} \frac{f}{p_j} G_{ij} = \sum_{j} \frac{f}{p_j} G_{ij}
\]  

(8.12)

Table 8.1. Marginal means and effects of alleles at a single locus, when gametes unite at random.

<table>
<thead>
<tr>
<th>Female gametes and frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Male</strong></td>
</tr>
<tr>
<td><strong>Gamete</strong></td>
</tr>
<tr>
<td>( A_1 )</td>
</tr>
<tr>
<td>( A_2 )</td>
</tr>
<tr>
<td>( A_m )</td>
</tr>
</tbody>
</table>

Mean

\[
G_1 = \mu + \alpha_1^f
\]

\[
G_2 = \mu + \alpha_2^f
\]

\[
= \mu + \alpha_1^f
\]

Frequency

\[
\frac{f}{P_1} \quad \frac{f}{P_2} \quad \frac{f}{P_m} \quad 1
\]

To obtain the \( \delta_{ij} \) or dominance effects, we substitute (8.9), (8.10) and (8.11) into (8.4), namely,
\[ \delta_{ij} = G_{ij} - \mu - \alpha_i^m - \alpha_j^f \]
\[ = G_{ij} - G_{..} - (G_{i..} - G_{..}) - (G_{j..} - G_{..}) \]
\[ = G_{ij} - G_{..} - G_{i..} + G_{..} - G_{j..} + G_{..} \]
\[ = G_{ij} - G_{i..} + G_{j..} \] for \( i, j = 1, \ldots, m \) \hspace{1cm} (8.13)

(This is analogous to the residual effect in a randomized complete block design (see Steel and Torrie, 1980, Section 9.4).) Because \( G_{i..} \neq G_{..} \), we have, in general, that

\[ \delta_{ij} \neq \delta_{ji} \] \hspace{1cm} (8.14)

We note that (see Box 8.2)

\[ \sum_{j=1}^{m} p_j^f \delta_{ij} = 0 \] \hspace{1cm} for \( i = 1, \ldots, m \) \hspace{1cm} (8.15a)

\[ \sum_{i=1}^{m} p_i^m \delta_{ij} = 0 \] \hspace{1cm} for \( j = 1, \ldots, m \) \hspace{1cm} (8.15b)

Since the sum of the weighted dominance deviations equals zero for every row or column, it is also true that the sum of all dominance deviations weighted by their respective frequencies equals zero, namely,

\[ \sum_{i=1}^{m} \sum_{j=1}^{m} p_i^m p_j^f \delta_{ij} = 0 \] \hspace{1cm} (8.16)

---

**Box 8.2**

**Derivation of (8.15)**

We substitute (8.4) in (8.15) to obtain

\[ \sum_{j=1}^{m} p_j^f \delta_{ij} = \sum_{j=1}^{m} p_j^f (G_{ij} - \mu - \alpha_i^m - \alpha_j^f) \] \hspace{1cm} for \( i = 1, \ldots, m \)

\[ = \sum_{j=1}^{m} p_j^f G_{ij} - \sum_{j=1}^{m} p_j^f \mu - \sum_{j=1}^{m} p_j^f \alpha_i^m - \sum_{j=1}^{m} p_j^f \alpha_j^f \] \hspace{1cm} (sub. (2), Box 8.1)

\[ = \left( \sum_{j=1}^{m} p_j^f G_{ij} - \mu \right) - \alpha_i^m - \sum_{j=1}^{m} p_j^f \alpha_j^f \] \hspace{1cm} (sub. (8.10) and (8.8))
\[
\begin{align*}
- \alpha_1^m & - \alpha_1^m - 0 \\
- 0 & \text{ for } i = 1, \ldots, m \\
\end{align*}
\]

Similarly, we can show
\[
\sum_{i=1}^{m} p_i^m \delta_{ij} = 0 \quad \text{for } j = 1, 2, \ldots, m
\]

which is (8.15).

This weighted least-squares analysis is completely analogous to the case of proportional subclass numbers in the analysis of variance (see Scheffe, 1959, Section 4.4, particularly "Case of proportional frequencies"; Steel and Torrie, 1980, Section 18.3).

**Remark.** It might also be noted here that if alleles are not independent \( p_{ij} \neq p_i^m p_j^m \), i.e., random mating has not occurred, least-squares values can still be found. The least squares value for \( \mu \) is the value which minimizes \( \sum p_{ij}(G_{ij} - \mu)^2 \). After finding \( \mu \), the least-squares values of \( \alpha^m \)'s are the values which minimize \( \sum p_{ij}(G_{ij} - \mu - \alpha_1^m)^2 \). Then after finding \( \mu \) and \( \alpha_1^m \)'s, the least-squares values of \( \alpha^f \)'s are the values which minimize \( \sum p_{ij}(G_{ij} - \mu - \alpha_1^m - \alpha_1^f)^2 \), and the \( \delta \)'s are the remainders or residuals. Alternatively, the \( \alpha^f \)'s may be fitted before the \( \alpha^m \)'s in which case the \( \alpha^f \)'s will be different from those obtained after fitting the \( \alpha^m \)'s first. In this case of sequential fitting, this procedure will lead to an additive set of variances in that the variance of the parts sums to the total variance. However, the variance of either \( \alpha^m \) or \( \alpha^f \) will be different depending upon the order of fitting in the same way as the values of \( \alpha^m \) and \( \alpha^f \) themselves were different depending upon order. This is not a very satisfactory situation and will not be considered henceforth.

* * * * * * *
The variance of each term in the model (8.2) may be written down by definition of a variance (see (2.91) and (2.92)). First, by the restriction in (8.7), the true mean of the random variable \( \alpha^m \) is

\[
\mu_{\alpha^m} = E(\alpha^m) = \sum_{i=1}^{m} p_i \alpha_i^m = 0 \tag{8.17}
\]

Then

\[
\sigma_{\alpha^m}^2 = E[(\alpha^m - \mu_{\alpha^m})^2] = E[(\alpha^m - 0)^2] = E[(\alpha_i^m)^2] = E(\alpha^m)^2
\]

\[
= \sum_{i=1}^{m} p_i (\alpha_i^m)^2 \tag{8.18}
\]

This is a very important expression. It follows directly from the definition of the variance of any variable (2.91) (2.92) in that it is the sum of the squares of the deviations, each weighted by its relative frequency in the population. Many other expressions exist for this quantity, such as (8.41a) which is expressed in terms of the average effect of an allelic substitution in the population, but its definitional meaning is often lost if the equivalence is not recognized. It is noted that the variance \( \sigma_{\alpha^m}^2 \) in (8.18) is determined by the allelic frequencies in the \( m \) population, but that each \( \alpha_i^m \) is a function of the allelic frequencies of the \( f \) population (8.10).

Alternatively, by utilizing a property of the general linear hypothesis (multiple linear regression) the variance component of \( G \) due to the \( \alpha^m \) term in the model is equal to the sum of the products of each \( \alpha^m \) effect and the corresponding element on the right-hand side of the normal equation. (This is analogous to obtaining the sum of squares due to regression for the set of \( \alpha^m \) in a block orthogonal model, i.e., \( SS(\alpha^m) = \sum_{i} b_i \sum(X_i - \bar{X}_i)Y \) where the \( b_i \)'s are the \( \alpha_i^m \) effects and \( \sum(X_i - \bar{X}_i)Y \) are the corresponding elements on the right-hand side of the normal equations.) That is, by using the right-hand side of (8.6b) we obtain
\[ \sigma^2_{\alpha_m} = \sum_{i=1}^{m} \alpha_i^m (p_i^m \sum_{j=1}^{m} f_{G_{ij}}^j) \]

\[ \text{subtract and add} \]
\[ = \sum_{i=1}^{m} p_i^m \alpha_i^m \sum_{j=1}^{m} f_{G_{ij}}^j - \sum_{i=1}^{m} p_i^m \alpha_i^m \mu + \sum_{i=1}^{m} p_i^m \alpha_i^m \mu \quad \text{(sub. (8.7))} \]
\[ = \sum_{i=1}^{m} p_i^m (\alpha_i^m)^2 \]
\[ = \sum_{i=1}^{m} p_i^m (\alpha_i^m)^2 \]
\[ (8.19) \]

which is the same as (8.18) written by definition.

Similar to that for \( \alpha_j^m \) (8.18) (8.19), the variance of \( \alpha_j^f \) is

\[ \sigma^2_{\alpha_j^f} = \sum_{j=1}^{m} p_j^f (\alpha_j)^2 \]

\[ (8.20) \]

and by definition (2.91) (2.92) the variance of \( \delta \) is

\[ \sigma^2_{\delta} = \sum_{i=1}^{m} \sum_{j=1}^{m} p_i^m p_j^f \delta_{ij}^2 \]

\[ (8.21) \]

The total variance among genotypes or the total genotypic variance for any given locus is the sum of the variances of the three random variables in the model (8.2), since the three variables are uncorrelated (see Box 8.3)

\[ \sigma^2_G = \sum_{i=1}^{m} \sum_{j=1}^{m} p_i^m p_j^f (G_{ij} - \mu)^2 \]

\[ \text{(sub. (8.2))} \]
\[ = \sum_{i=1}^{m} \sum_{j=1}^{m} p_i^m p_j^f (\alpha_i^m + \alpha_j^f + \delta_{ij})^2 \]
\[ = \sum_{i=1}^{m} \sum_{j=1}^{m} p_i^m p_j^f [(\alpha_i^m)^2 + (\alpha_j^f)^2 + \delta_{ij}^2 + 2\alpha_i^m \alpha_j^f + 2\alpha_i^m \delta_{ij} + 2 \alpha_j^f \delta_{ij}] \]
\[ = \sum_{i=1}^{m} (\alpha_i^m)^2 + \sum_{j=1}^{m} p_j^f (\alpha_j)^2 + \sum_{i=1}^{m} \sum_{j=1}^{m} p_i^m p_j^f \delta_{ij}^2 + 0 + 0 + 0 \]

\[ \text{(sub. (8.18) (8.20) (8.21))} \]
\[ = \sigma^2_m + \sigma^2_f + \sigma^2_\delta \]
\[ = \sigma^2_{Am} + \sigma^2_{Af} + \sigma^2_{Dmf} \]

(8.22)

where \( \sigma^2_G \) = genotypic variance for the particular locus,
\( \sigma^2_{Am} = \sigma^2_{\alpha^m} \) = additive variance due to \( \alpha^m \) for the particular locus for the male population,
\( \sigma^2_{Af} = \sigma^2_{\alpha^f} \) = additive variance due to \( \alpha^f \) for the particular locus for the female population,
\( \sigma^2_{Dmf} = \sigma^2_\delta \) = dominance variance for the particular locus in the cross population.

This is an application of the theorem for the variance of a linear function of uncorrelated random variables (Kempthorne, 1969, Section 12.5 and 12.7; Steel and Torrie, 1980, Section 5.10). An understanding of this theorem and its application is imperative, because we shall be applying it repeatedly in our study of statistical genetics. All of these variances are, by definition, weighted means of the squares of the corresponding effects and are properties of a given population in that their magnitude depends upon allelic frequencies (8.18) (8.20) (8.21) (also see discussion following (8.10)).

**Box 8.3**

Proof that all covariances or expectations between effects in one-locus model (8.2) are zero (see (8.22))

Considering each of the covariances in (8.22) between effects in the model (8.2), we have by definition (2.95) (2.96):

1. \( \text{Cov}(\alpha^m_i, \alpha^f_j) = E[(\alpha^m_i)(\alpha^f_j)] = \sum_{i=1}^{m} \sum_{j=1}^{m} p_{\alpha^m_i \alpha^f_j}(\alpha^m_i)(\alpha^f_j) = \sum_{i=1}^{m} \sum_{j=1}^{m} p_{ij}(\alpha^m_i)(\alpha^f_j) \)

where \( p_{\alpha^m_i \alpha^f_j} = p_{ij} = p_{G_{ij}} \) = joint frequency or frequency of occurrence of effects \( \alpha^m_i \) and \( \alpha^f_j \) together, which is the frequency of \( G_{ij} \).
With random mating $p_{ij}^m = p_{ij}^f = p_{Gij} = p_{ij}^m$,

$$\text{Cov}(\alpha_i^m, \alpha_j^m) = \Sigma_{i=1}^{m} \Sigma_{j=1}^{m} p_{ij}^m (\alpha_i^m)(\alpha_j^m) = \Sigma_{i=1}^{m} \Sigma_{j=1}^{m} p_{ij}^m \mu \sigma_{ij}^m = (0)(0) = 0 \text{ (sub. (8.7) and (8.8))}$$

2. \text{Cov}(\alpha_i^m, \delta_{ij}) = \mathbb{E}[(\alpha_i^m)(\delta_{ij})] = \Sigma_{i=1}^{m} \Sigma_{j=1}^{m} p_{ij}^m (\alpha_i^m)(\delta_{ij}) = \Sigma_{i=1}^{m} \Sigma_{j=1}^{m} p_{ij}^m \mu \delta_{ij}$

where $p_{ij}^m = p_{Gij}$ = joint frequency of occurrence of effects $\alpha_i^m$ and $\delta_{ij}$ together, which is the frequency of $G_{ij}$.

With random mating $p_{ij}^m = p_{ij}^f = p_{ij}^m$,

$$\text{Cov}(\alpha_i^m, \delta_{ij}) = \Sigma_{i=1}^{m} \Sigma_{j=1}^{m} p_{ij}^m (\alpha_i^m)(\delta_{ij}) = \Sigma_{i=1}^{m} \Sigma_{j=1}^{m} p_{ij}^m \mu \delta_{ij} = (0)(0) = 0 \text{ (2)}$$

3. That the remaining $\text{Cov}(\alpha_j^f, \delta_{ij})$ equals zero can be shown in a manner similar to that for $\text{Cov}(\alpha_i^m, \delta_{ij})$ in (2).

One might ask what kinds of phenomena would give nonzero covariances or correlations between the variables or terms in the model (8.2). The answer is:

Any phenomena which would bring about an association between any two terms in the model, in either a positive or negative sense, more frequently than expected based on random events. For example, considering the $\alpha^m$'s and the $\alpha^f$'s, one can depict the following with random mating for any two effects, say 1 and 2,
second variable
\( (\alpha^f) \)

\[
\begin{array}{c}
m^f_{P2P1} \times \\
\alpha_1^m \times m^f_{P1P1}
\end{array}
\]

(8.23)

first variable \((\alpha^m)\)

\[
\begin{array}{c}
m^f_{P2P2} \times \\
\alpha_2^m \times m^f_{P1P2}
\end{array}
\]

A general phenomenon which can bring about departures from random mating is assortative mating (Section 3.1). For example, positive phenotypic assortative mating would tend to increase the frequency \(m^f_{P1P1}\) of \(A_1A_1\) homozygotes and \(m^f_{P1P2}\) of \(A_2A_2\) which brings about an association of like \(\alpha\) effects or a positive covariance. When the covariances are nonzero, the total genotypic variance is (Kempthorne 1969, Sections 12.5 and 12.7; Steel and Torrie, 1980, Section 5.10)

\[
\sigma_G^2 = \sigma_{\alpha^m}^2 + \sigma_{\alpha^f}^2 + \sigma_\delta^2 + 2 \text{Cov}(\alpha^m, \alpha^f) + 2 \text{Cov}(\alpha^m, \delta) + 2 \text{Cov}(\alpha^f, \delta) \\
= \sigma_{Am}^2 + \sigma_{Af}^2 + \sigma_{Dmf}^2 + 2(\text{sum of covariances})
\]

(8.24)

It is emphasized that the variances are defined only in terms of a random-mated population structure for the cross between two populations. In practice, many estimates of so-called additive variance are made from populations which are not strictly random-mating ones. In such cases, the estimates are biased either upward or downward in that they absorb the covariances.
8.1.2. **One locus, multiple alleles, equal male and female allelic frequencies (Hardy-Weinberg).** So far the development has been general in that a cross between two populations has been supposed. The arrays of allelic frequencies in the two parental populations were presumed to be different, i.e., \( p_i^m \neq p_i^f \) for at least two or more values of \( i \). However, if one supposes that the \( m \) and \( f \) allelic arrays are the same, then the above development is applicable for a single population in Hardy-Weinberg equilibrium. Also, if the cross population itself is random-mated one generation, i.e., random-mate the individuals of the cross itself, then the male (\( m \)) and female (\( f \)) allelic frequencies of uniting gametes from males and females will be the same (see discussion following (3.7)). In either case, \( p_i^m = p_i^f = p_i \) for \( i = 1, \ldots, m \); the \( m \) and \( f \) superscripts can be dropped throughout in the previous section. Thus, from (8.10) and (8.11) we have

\[
\alpha_i^m = \alpha_i^f = \alpha_i = G_i - \mu = \Sigma p_j G_{ij} \mu \quad (8.25)
\]

and since \( G_i = G_i \), (8.14) becomes

\[
\delta_{ij} = \delta_{ji} \quad \text{for } i, j = 1, \ldots, m \quad (8.26)
\]

It may be noted here that each \( \alpha_i \) is one-half of the breeding value of the corresponding homozygote (8.67).

From (8.18) and (8.20), we have

\[
\sigma^2_\alpha^m = \sigma^2_\alpha^f = \sigma^2_\alpha^i = \Sigma p_i \alpha_i^2 \quad (8.27)
\]

and

\[
\sigma^2_\delta = \Sigma \Sigma p_i p_j \delta_{ij} \quad (8.28)
\]

The genotypic variance (8.22) for the particular locus becomes

\[
\sigma^2_G = 2\sigma^2_\alpha + \sigma^2_\delta
\]

\[
= \sigma^2_A + \sigma^2_D \quad (8.29)
\]
where $\sigma^2_A = 2\sigma^2_\alpha$ = additive variance for the particular locus,

$\sigma^2_D = \sigma^2_\delta$ = dominance variance for the particular locus.

All effects (8.25) (8.26) and their variances (8.27) (8.28) are descriptive of the particular population, and as such are dependent on allelic frequencies as before.

8.1.3. One locus, two alleles, arbitrary "male" and "female" allelic frequencies (cross between two populations): General. This section supposes the same conditions as those in Section 8.1.1, namely, a cross population, except only two alleles are now assumed, i.e., $m = 2$. All formulas presented in that section are applicable by letting $m = 2$. What we want to do here is to illustrate the general case for two alleles and to present some additional formulas that are peculiar to the two-allele case.

With two alleles, Table 8.1 becomes ($G_{12} = G_{21}$)

<table>
<thead>
<tr>
<th></th>
<th>$A_1$</th>
<th>$A_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f$</td>
<td>$p_1^m f_{G_{11}} + p_2^m f_{G_{12}}$</td>
<td>$p_1^m f_{G_{11}} + p_2^m f_{G_{12}}$</td>
</tr>
<tr>
<td>$p_1^m$</td>
<td>$p_1^m p_1^m f_{G_{11}} + p_2^m f_{G_{12}}$</td>
<td>$p_1^m p_1^m f_{G_{11}} + p_2^m f_{G_{12}}$</td>
</tr>
<tr>
<td>$f$</td>
<td>$p_1 f_{G_{11}} + p_2 f_{G_{12}}$</td>
<td>$p_1 f_{G_{11}} + p_2 f_{G_{12}}$</td>
</tr>
<tr>
<td>$p_2^m$</td>
<td>$p_2 f_{G_{11}} + p_2 f_{G_{12}}$</td>
<td>$p_2 f_{G_{11}} + p_2 f_{G_{12}}$</td>
</tr>
</tbody>
</table>

Freq. $p_1^m f_{G_{11}} + p_2^m f_{G_{12}}$

Mean $p_1 f_{G_{11}} + p_2 f_{G_{12}}$

With two alleles, Table 8.1 becomes ($G_{12} = G_{21}$)

<table>
<thead>
<tr>
<th></th>
<th>$A_1$</th>
<th>$A_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f$</td>
<td>$p_1^m f_{G_{11}} + p_2^m f_{G_{12}}$</td>
<td>$p_1^m f_{G_{11}} + p_2^m f_{G_{12}}$</td>
</tr>
<tr>
<td>$p_1^m$</td>
<td>$p_1^m p_1^m f_{G_{11}} + p_2^m f_{G_{12}}$</td>
<td>$p_1^m p_1^m f_{G_{11}} + p_2^m f_{G_{12}}$</td>
</tr>
<tr>
<td>$f$</td>
<td>$p_1 f_{G_{11}} + p_2 f_{G_{12}}$</td>
<td>$p_1 f_{G_{11}} + p_2 f_{G_{12}}$</td>
</tr>
<tr>
<td>$p_2^m$</td>
<td>$p_2 f_{G_{11}} + p_2 f_{G_{12}}$</td>
<td>$p_2 f_{G_{11}} + p_2 f_{G_{12}}$</td>
</tr>
</tbody>
</table>

Freq. $p_1^m f_{G_{11}} + p_2^m f_{G_{12}}$

Mean $p_1 f_{G_{11}} + p_2 f_{G_{12}}$

With two alleles, Table 8.1 becomes ($G_{12} = G_{21}$)

<table>
<thead>
<tr>
<th></th>
<th>$A_1$</th>
<th>$A_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f$</td>
<td>$p_1^m f_{G_{11}} + p_2^m f_{G_{12}}$</td>
<td>$p_1^m f_{G_{11}} + p_2^m f_{G_{12}}$</td>
</tr>
<tr>
<td>$p_1^m$</td>
<td>$p_1^m p_1^m f_{G_{11}} + p_2^m f_{G_{12}}$</td>
<td>$p_1^m p_1^m f_{G_{11}} + p_2^m f_{G_{12}}$</td>
</tr>
<tr>
<td>$f$</td>
<td>$p_1 f_{G_{11}} + p_2 f_{G_{12}}$</td>
<td>$p_1 f_{G_{11}} + p_2 f_{G_{12}}$</td>
</tr>
<tr>
<td>$p_2^m$</td>
<td>$p_2 f_{G_{11}} + p_2 f_{G_{12}}$</td>
<td>$p_2 f_{G_{11}} + p_2 f_{G_{12}}$</td>
</tr>
</tbody>
</table>

Freq. $p_1^m f_{G_{11}} + p_2^m f_{G_{12}}$

Mean $p_1 f_{G_{11}} + p_2 f_{G_{12}}$
With only two alleles, coded genotypic values are commonly used. On some occasions, we will also use coded values, particularly those of Falconer (1989, p. 112), i.e.,

\[
\begin{array}{c|c|c}
A_2A_2 & \quad & A_1A_2 \\
-\alpha & \quad & \quad \quad d \\
\end{array}
\]

where \( a = G_{11} - MH \),
\[
d = G_{12} - MH = G_{12} - \frac{G_{11} + G_{22}}{2} = \frac{1}{2} (2G_{12} - G_{11} - G_{22}),
\]

\( 2G_{12} - G_{11} - G_{22} = 2d = \text{dominance contrast (positive form)}, \)
\[
-a = G_{22} - MH,
\]
\[
MH = \frac{G_{11} + G_{22}}{2} = \text{mid-homozygote value}.
\]

We follow Falconer (1989, p. 112) and adopt the convention that \( A_1 \) is the allele that increases the genotypic value. Although we will generally use the positive form of the dominance contrast, there are occasions when the negative form is more natural and will be used, namely, \( G_{11} - 2G_{12} + G_{22} = -2d \). In conjunction with such coded values, we will also generally use, but not invariably, the \( p \) and \( q \) notation (see (3.33) (3.34)) for allelic frequencies in the \( m \) population (not necessarily the male population (8.2)), namely,

\[
\begin{align*}
p_1^m &= p, & p_2^m &= q \\
\end{align*}
\]

For the \( f \) population (not necessarily the female population), we will generally use the \( r \) and \( s \) notation, namely,

\[
\begin{align*}
p_1^f &= r, & p_2^f &= s \\
\end{align*}
\]

The mean of the cross between the two populations is
\[
M = p_1^m p_1^a + (p_1^m p_2^f + p_2^m p_1^f) d + p_2^m p_2^f (-\alpha) \quad \text{(sub. (8.32) (8.33))}
\]
\[
= pra + (ps + qr) d - qsa
\]
\[
= (pr - qs) a + (ps + qr) d \quad \text{(8.34)}
\]
where \( M \) = mean of (cross) population in terms of coded genotypic values.

Then the average effects of the alleles for the \( m \) (or male) population in terms of uncoded values are (see Box 8.4) (sub. (8.32))

\[
\alpha_m^1 = -\frac{p_2^m}{p_1^m} \alpha_m^2 = p_2^m[p_1^m(G_{11} - G_{21}) + p_2^m(G_{12} - G_{22})] = p_2^m \alpha^M = q \alpha^M \quad (8.35a)
\]

\[
\alpha_m^2 = -\frac{p_1^m}{p_2^m} \alpha_m^1 = -p_1^m[p_1^m(G_{11} - G_{21}) + p_2^m(G_{12} - G_{22})] = -p_1^m \alpha^M = -q \alpha^M \quad (8.35b)
\]

where \( \alpha^F = p_1^f(G_{11} - G_{21}) + p_2^f(G_{12} - G_{22}) \) = average effect of an allelic substitution in the \( f \) population.

Note that the quantity \( \alpha^F = p_1^f(G_{11} - G_{21}) + p_2^f(G_{12} - G_{22}) \), called the average effect of an allelic substitution in the \( f \) population, is the weighted mean of two contrasts or differences: (1) difference between the genotypic values for homozygote \( A_1A_1 \) and the heterozygote \( A_1A_2 \), \( G_{11} - G_{21} \), i.e., having substituted allele \( A_1 \) for the \( A_2 \) allele in the heterozygote \( A_2A_1 \) and observing the change, and (2) difference between heterozygote \( A_1A_2 \) and the homozygote \( A_2A_2 \), \( G_{12} - G_{22} \), i.e., having substituted allele \( A_1 \) for one of the \( A_2 \) genes in the homozygote \( A_2A_2 \) and observing the change. The weight \( p_1^f \) and \( p_2^f \) are the frequencies with which these two contrasts occur. Because the weights are the frequencies in the \( f \) population, we symbolize \( \alpha^F \) with an \( F \) superscript. We use a capital letter to avoid confusion with \( \alpha^f \) in the model (8.2) when the subscript \( j \) is dropped for various reasons. Another way to describe this effect is to visualize changing, at random, those \( A_2 \) alleles, contributed by the \( m \) population, to \( A_1 \) alleles. With random mating, the \( A_2 \) alleles from the \( m \) population occur with \( A_1 \) alleles with frequency \( p_1^f \) and with \( A_2 \) alleles with frequency \( p_2^f \). When \( A_2 \) alleles in \( A_2A_1 \) are changed to \( A_1 \), giving an \( A_1A_1 \) genotype, a response of \( G_{11} - G_{21} \) is observed. When \( A_2 \) alleles in \( A_2A_2 \) are changed to \( A_1 \), giving an \( A_1A_2 \) genotype, a response of \( G_{12} - G_{22} \) is observed. Since these responses occur randomly, they occur with
relative frequencies of \( p^f_1 \) and \( p^f_2 \). Thus, we have \( \alpha^F = p^f_1(G_{11} - G_{21}) + p^f_2(G_{12} - G_{22}) \). This average effect of an allelic substitution in the \( f \) population is the difference between the two marginal means, \( G_{1.} - G_{2.} \), or the difference between \( \alpha^m_1 \) and \( \alpha^m_2 \), i.e., (8.30),

\[
\alpha^F = \alpha^m_1 - \alpha^m_2 = (G_{1.} - G_{2.}) = G_{1.} - G_{2.}.
\]

The average effect of an allelic substitution (8.35) in terms of coded genotypic values is

\[
\alpha^F = p^f_1(G_{11} - G_{12}) + p^f_2(G_{12} - G_{22}) \quad \text{(sub. (8.31)(8.33))}
\]

\[
= r(a - d) + s[d - (-a)]
\]

\[
= (r + s)a + (s - r)d
\]

\[
= a + (s - r)d \quad (8.37)
\]

The average effects of the alleles for the \( f \) (or female) population are

(derived or written in an analogous manner to (8.35) (sub. (8.33))

\[
\alpha^f_1 = -\frac{p^f_2}{p^f_1} \alpha^f_2 = p^f_2[p^m_1(G_{11} - G_{12}) + p^m_2(G_{21} - G_{22})] = p^f_2 \alpha^M = s\alpha^M \quad (8.38a)
\]

\[
\alpha^F_2 = -\frac{p^f_1}{p^f_2} \alpha^f_1 = -p^f_1[p^m_1(G_{11} - G_{12}) + p^m_2(G_{21} - G_{22})] = -p^f_1 \alpha^M = -r\alpha^M \quad (8.38b)
\]

where \( \alpha^M = p^m_1(G_{11} - G_{12}) + p^m_2(G_{21} - G_{22}) \)

\[
= a + (q - p)d \quad \text{average effect of an allelic substitution in the \( m \) population (derivation analogous to (8.37)).}
\]

The quantity, \( \alpha^M = p^m_1(G_{11} - G_{12}) + p^m_2(G_{21} - G_{22}) \) or average effect of an allelic substitution in the \( m \) population, involves the same two contrasts as that for the \( f \) population (8.35), but the weights are allelic frequencies in the \( m \) population.

It is also the difference between the other two marginal means, \( G_{1.} - G_{2.} \), or the difference between \( \alpha^f_1 \) and \( \alpha^f_2 \), i.e.,

\[
\alpha^M = \alpha^f_1 - \alpha^f_2 = (G_{1.} - G_{2.}) = G_{1.} - G_{2.}.
\]

(8.39)
With regard to dominance deviations, we have (see Box 8.4) (sub. (8.31) (8.32) (8.33))

\[
\begin{align*}
\delta_{11} &= -m_{f}^{m}p_{2}^{2}(2G_{12} - G_{11} - G_{22}) = -2p_{2}^{m}p_{2}^{d} = -2qsd \\
\delta_{12} &= -m_{f}^{m}p_{2}^{1}(2G_{12} - G_{11} - G_{22}) = 2p_{2}^{m}p_{1}^{d} = 2qrd \\
\delta_{21} &= -m_{f}^{m}p_{1}^{2}(2G_{12} - G_{11} - G_{22}) = 2p_{1}^{m}p_{2}^{d} = 2psd \\
\delta_{22} &= -m_{f}^{m}p_{1}^{1}(2G_{12} - G_{11} - G_{22}) = -2p_{1}^{m}p_{1}^{d} = -2prd
\end{align*}
\] (8.40)

The variances of \( \alpha^{m} \) (8.18) and \( \alpha^{f} \) (8.20) (8.22) in terms of contrasts for the particular locus are (see Box 8.4)

\[
\begin{align*}
\sigma_{\alpha^{m}}^{2} &= \sigma_{\alpha^{f}}^{2} = m_{f}^{m}p_{2}^{2}[p_{1}^{1}(G_{11} - G_{12}) + p_{2}^{1}(G_{12} - G_{22})]^{2} = p_{2}^{m}p_{2}^{d} = pq(\alpha^{F})^{2} \quad (8.41a) \\
\sigma_{\alpha^{f}}^{2} &= \sigma_{\alpha^{m}}^{2} = f_{f}^{m}m_{m}^{f}p_{1}^{2}[p_{1}^{1}(G_{11} - G_{12}) + p_{2}^{1}(G_{12} - G_{22})]^{2} = p_{1}^{m}p_{1}^{d} = rs(\alpha^{M})^{2} \quad (8.41b) \\
\sigma_{\alpha^{m}+\alpha^{f}}^{2} &= \sigma_{\alpha^{m^{+}}+\alpha^{f^{+}}}^{2} = m_{f}^{m}p_{2}^{2}[p_{1}^{1}(G_{11} - G_{12}) + p_{2}^{1}(G_{12} - G_{22})]^{2} \\
&\quad + f_{f}^{m}m_{m}^{f}p_{1}^{2}[p_{1}^{1}(G_{11} - G_{12}) + p_{2}^{1}(G_{12} - G_{22})]^{2} \\
&\quad - pq(\alpha^{F})^{2} + rs(\alpha^{M})^{2} \quad (8.41c)
\end{align*}
\]

The dominance variance or the variance of \( \delta \) in terms of both the dominance contrast and \( d \) is (see Box 8.4)

\[
\sigma_{Dmf}^{2} = \sigma_{\delta}^{2} = m_{f}^{m}m_{f}^{f}p_{1}^{2}p_{2}^{2}(2G_{12} - G_{11} - G_{22})^{2} = 4pqrsd^{2} \quad (8.42)
\]

**Box 8.4**

Equation (8.35): From (8.10) and (8.30)

\[
\alpha_{1}^{m} = G_{1} - G_{m} = (p_{1}^{m}G_{11} + p_{2}^{m}G_{12}) - [p_{1}^{m}p_{1}^{1}G_{11} + p_{2}^{m}p_{2}^{1}G_{12} + p_{2}^{m}p_{2}^{1}G_{21} + p_{2}^{m}p_{2}^{2}G_{22}]
\]

\[
= p_{1}(1 - p_{1}^{m})G_{11} + p_{2}(1 - p_{2}^{m})G_{12} - p_{2}^{m}p_{2}^{1}G_{21} - p_{2}^{m}p_{2}^{2}G_{22}
\]

\[
= f_{m}^{m}p_{1}^{1}G_{11} - f_{m}^{m}p_{2}^{1}G_{12} + f_{m}^{m}m_{m}^{f}p_{2}^{2}(G_{11} - G_{22})
\]

\[
= p_{2}^{m}p_{2}^{2}[(G_{11} - G_{21}) + p_{2}^{m}G_{12} - G_{22})]
\] (1)
Then, to obtain $a_2^m$ we substitute (1) in (8.7)

$$\begin{align*}
  p_1 a_1^m + p_2 a_2^m &= 0 \\
  a_2^m &= -\frac{p_1^m}{p_2^m} a_1^m = -\frac{p_1^m}{p_2^m} p_2^f [p_1^f (G_{11} - G_{21}) + p_2^f (G_{12} - G_{22})] \\
  &= -p_1^m [p_1^f (G_{11} - G_{21}) + p_2^f (G_{12} - G_{22})] \\
\end{align*}$$

From (2) we have

$$\begin{align*}
  a_1^m &= -\frac{p_2^m}{p_1^m} a_2^m = p_2^m [p_1^f (G_{11} - G_{21}) + p_2^f (G_{12} - G_{22})] \\
\end{align*}$$

Equations (1) (2) and (3) complete the derivation of (8.35).

**Equation (8.40):** From (8.13) and (8.30)

$$\begin{align*}
  \delta_{11} &= G_{11} - G_{11} - G_{11} + G_{11} \\
  &= G_{11} - (p_1^f G_{11} + p_2^f G_{12}) - (p_1^m G_{11} + p_2^m G_{12}) + [p_1^m G_{11} + (p_1^f + p_2^f) G_{12} + p_2^m G_{22}] \\
  &= G_{11} - p_1^f G_{11} - p_2^f G_{12} + p_1^m G_{11} - p_2^m G_{12} + [p_1^m G_{11} + (p_1^f + p_2^f) G_{12} + p_2^m G_{22}] \\
  &= (1 - p_1^f) p_1^m G_{11} - (p_2^f + p_1^m p_1^f) G_{11} - (p_2^m + p_1^f p_1^m) G_{12} + p_2^m G_{22} \\
  &= (1 - p_1^m) (1 - p_1^f) G_{11} - [p_2^f (1 - p_1^m) + p_2^m (1 - p_1^f)] G_{12} + p_2^m G_{22} \\
  &= p_2^m G_{11} - (p_2^m + p_2^f) G_{12} + p_2^m G_{22} \\
  &= p_2^m (G_{11} - 2G_{12} + G_{22}) \\
  &= -p_2^m (2G_{12} - G_{11} - G_{22}) \\
\end{align*}$$

Then, solving for $\delta_{12}$ in (8.15a) and substituting (4), we obtain

$$\begin{align*}
  f_{p_1 \delta_{11}} + f_{p_2 \delta_{12}} &= 0 \\
  \delta_{12} &= -\frac{f_{p_1 \delta_{11}}}{p_2} = -\frac{f_{p_1}}{p_2} [p_2^m (2G_{12} - G_{11} - G_{22})] \\
  &= p_2^m (2G_{12} - G_{11} - G_{22}) \\
\end{align*}$$
Similarly, solving for $\delta_{21}$ in (8.15b) and substituting (4), we obtain

$$p_1^{m} \delta_{11} + p_2^{m} \delta_{21} = 0$$

$$\delta_{21} = -\frac{p_1^{m}}{p_2} \delta_{11} = -\frac{p_1^{m}}{p_2} [-p_2 p_2^{f}(2G_{12} - G_{11} - G_{22})]$$

$$= p_1^{m} p_2^{f}(2G_{12} - G_{11} - G_{22})$$

(6)

Finally, using (8.15a) again and substituting (6), we obtain

$$p_1^{f} \delta_{21} + p_2^{f} \delta_{22} = 0$$

$$\delta_{22} = -\frac{p_1^{f}}{p_2} \delta_{21} = -\frac{p_1^{f}}{p_2} [-p_1 p_2^{f}(2G_{12} - G_{11} - G_{22})]$$

$$= -p_1^{m} p_2^{f}(2G_{12} - G_{11} - G_{22})$$

(7)

Thus, (4) (5) (6) and (7) give (8.40).

Equation (8.41): We substitute (1) and (2) in (8.18) to obtain

$$\sigma_{\alpha}^{2} = \sum_{i=1}^{2} p_1^{m}(\alpha_i)$$

$$= p_1^{m}(\alpha_1)^2 + p_2^{m}(\alpha_2)^2$$

$$+ p_1^{m}(p_2^{f}-p_1^{f})[p_1^{f}(G_{11} - G_{12}) + p_2^{f}(G_{12} - G_{22})]^2$$

(8)

Similarly, from (8.20) we obtain

$$\sigma_{\delta}^{2} = p_1^{f} p_2^{f}[p_1^{m}(G_{11} - G_{12}) + p_2^{m}(G_{12} - G_{22})]$$

(9)

Thus, (8) and (9) give (8.41).

Equation (8.42): From (8.21)

$$\sigma_{\delta}^{2} = \sum_{i=1}^{2} \sum_{j=1}^{2} p_1^{m} p_j^{f} \delta_{ij}^{2}$$

$$= \frac{m_{f}^{2}}{p_1^{P_{j}^{d}} 11} + \frac{m_{f}^{2}}{p_2^{P_{j}^{d}} 12} + \frac{m_{f}^{2}}{p_2^{P_{j}^{d}} 21} + \frac{m_{f}^{2}}{p_2^{P_{j}^{d}} 22}$$

(sub. (4)(5)(6) and (7))
\[
= p_{1}p_{2}\left[p_{1}(2G_{12} - G_{11} - G_{22}) + p_{2}(2G_{12} - G_{11} - G_{22})\right]^{2} + p_{1}p_{2}\left[p_{2}p_{1}(2G_{12} - G_{11} - G_{22})\right]^{2} \\
+ p_{2}p_{1}\left[p_{1}p_{2}(2G_{12} - G_{11} - G_{22})\right]^{2} + p_{2}p_{1}\left[-p_{1}p_{2}(2G_{12} - G_{11} - G_{22})\right]^{2} \\
= p_{1}p_{2}p_{1}p_{2}(2G_{12} - G_{11} - G_{22})^{2} + p_{1}p_{2}p_{1}p_{2}(2G_{12} - G_{11} - G_{22})^{2} \\
= p_{1}p_{2}p_{1}p_{2}(p_{1} + p_{2})(2G_{12} - G_{11} - G_{22})^{2} \\
= p_{1}p_{2}p_{1}p_{2}(2G_{12} - G_{11} - G_{22})^{2} \\
\text{(since } p_{1} + p_{2} = 1) \\
\text{(10)}
\]
which is (8.42).

That the level of dominance does affect the additive effects, \( \alpha_m \)'s and \( \alpha_f \)'s, and their variances can be seen by rewriting the variance of \( \sigma^2_m \) (8.41) (see Box 8.5)

\[
\sigma^2_m = p_{1}p_{2}\left[G_{11} - G_{22}\right]^{2} + (p_{2} - p_{1})\left[2G_{12} - G_{11} - G_{22}\right]^{2} \\
\text{(8.43)}
\]
which is expressed as a function of the difference between the two homozygotes and a measure of dominance, the dominance contrast, \( 2G_{12} - G_{11} - G_{22} \) (8.31). If there is dominance, it is only when the allelic frequencies are \( \frac{1}{2} \), \( p_{1} = p_{2} \), that the additive variance \( \sigma^2_m \) is not affected by the level of dominance. In other words, the variance is a function of only the difference between the two homozygotes, when there is either (1) the two alleles are equally frequent, \( p_{1} = p_{2} = \frac{1}{2} \) or (2) no dominance, \( 2G_{12} - G_{11} - G_{22} = 0 \), (see (8.72) below).

**Box 8.5**

**Derivation of (8.43)**

From (8.41)

\[
\sigma^2_m = p_{1}p_{2}p_{1}(G_{11} - G_{12}) + p_{2}(G_{12} - G_{22})^{2} \\
= p_{1}p_{2}\left[2p_{1}G_{11} - 2p_{1}G_{12} + 2p_{2}G_{12} - 2p_{2}G_{22}\right]^{2} \\
= p_{1}p_{2}\left[-2p_{1}G_{11} + p_{2}G_{11} - 2p_{1}G_{12} + 2p_{2}G_{12} - p_{1}G_{22} - 2p_{2}G_{22}\right]^{2}
\]
add and subtract

\[ -p_1 p_2 \left( \frac{1}{2} \left( (p_1 + p_2) G_{11} + (p_1 - p_2) G_{11} - (p_1 - p_2)^2 G_{11} + (p_1 - p_2) G_{22} \right) \right) \]

\[ -p_1 p_2 \left( \frac{1}{2} \left[ G_{11} - G_{22} + (p_1 - p_2) (G_{11} - 2G_{12} + G_{22}) \right] \right) \]

\[ = p_1 p_2 \left[ \frac{G_{11} - G_{22}}{2} + (p_2 - p_1) \frac{2G_{12} - G_{11} - G_{22}}{2} \right]^2 \]

which is (8.43).

8.1.4. **One locus, two alleles, equal male and female allelic frequencies** (Hardy-Weinberg) with arbitrary or equal allelic frequencies within sexual arrays and with dominance. 1. **Factorial approach.** This section differs only from the previous one in that we assume that the allelic frequencies in the m (male) and f (female) arrays are equal, i.e., we assume Hardy-Weinberg proportions and not a cross between two populations. Thus, we can drop all m and f superscripts in all equations in the previous section, i.e., \( p_1^m = p_1^f = p_1 \) and \( p_2^m = p_2^f = p_2 \). This section differs only from Section 8.1.2, which discussed the Hardy-Weinberg population, in that we assume only two alleles here instead of multiple alleles. The situation discussed in this section is the one commonly discussed by other authors (e.g., Falconer, 1989).

The population mean in terms of coded values (8.31) is (see (8.34) where \( p = r, q = s \))

\[ M = (pr - qs)a + (ps + qr)d \]

\[ = (p^2 - q^2)a + (pq + qp)d \]

\[ = (p - q)a + 2pqd \]  \hspace{1cm} (8.44)

Then from either (8.35) or (8.38), the additive effects are \( (p_1^m - p_1^f - p_1, p_2^m = p_2^f = p_2) \) or \( p = r = p_1, q = s = p_2 \).
\[ \alpha_1 = \frac{p_2}{p_1} \alpha_2 = p_2[p_1(G_{11} - G_{12}) + p_2(G_{12} - G_{22})] = p_2\alpha = q\alpha \quad (8.45a) \]

\[ \alpha_2 = \frac{p_1}{p_2} \alpha_1 = p_1[p_1(G_{11} - G_{12}) + p_2(G_{12} - G_{22})] = -p_1\alpha = -p\alpha \quad (8.45b) \]

where \( \alpha = p_1(G_{11} - G_{12}) + p_2(G_{12} - G_{22}) \) is the average effect of an allelic substitution

\[ = a + (q - p)d \quad \text{(see (8.38))} \]

The quantity, \( \alpha = p_1(G_{11} - G_{12}) + p_2(G_{12} - G_{22}) \), is called the average effect of an allelic substitution (Falconer, 1989, p. 116), and is the difference between the two marginal means \( G_{1.} - G_{2.} = G_{1} - G_{2} \), or the difference between \( \alpha_1 \) and \( \alpha_2 \), i.e.,

\[ \alpha = \alpha_1 - \alpha_2 = (G_{1.} - G_{2.}) - (G_{2.} - G_{1.}) = G_{1.} - G_{2.} \quad (8.46) \]

It is completely analogous to the average effects of allelic substitutions described for the f population in (8.35) and (8.36) and for the m population in (8.38) and (8.39).

From (8.40), the dominance effects are

\[ \delta_{11} = -p_2^2(2G_{12} - G_{11} - G_{22}) = -2p_2^2d = -2q^2d \quad (8.47a) \]

\[ \delta_{12} = \delta_{21} = p_1p_2(2G_{12} - G_{11} - G_{22}) = 2p_1p_2d = 2pqd \quad (8.47b) \]

\[ \delta_{22} = -p_1^2(2G_{12} - G_{11} - G_{22}) = -2p_1^2d = -2p^2d \quad (8.47c) \]

and their relations to each other are

\[ \frac{p_1}{p_2} \delta_{11} = \frac{p_2}{p_1} \delta_{22} = \delta_{12} = \delta_{21} = p_1p_2(2G_{12} - G_{11} - G_{22}) = 2p_1p_2d = 2pqd \quad (8.48) \]

Multiplying both sides of (8.48) by \(-p_1p_2\) gives

\[ \frac{p_1}{p_2} \delta_{11} = \frac{1}{2}(2p_1p_2\delta_{12}) = \frac{2}{p_2^{2}} \delta_{22} \quad (8.49) \]

which means that the dominance deviation for the one homozygote weighted by its frequency in the population is equal to that of the other homozygote, and, of
course, each is equal to one-half of the negative of the sum of the two weighted deviations for the heterozygote. This is a consequence of \( p_1 \delta_{11} + p_2 \delta_{12} = p_1^2 \delta_{11} + p_1 p_2 \delta_{12} = 0 \), \( p_1 \delta_{21} + p_2 \delta_{22} = p_2 p_1 \delta_{21} + p_2^2 \delta_{22} = 0 \) (8.15a), and \( \delta_{12} = \delta_{21} \) (8.26). It is also, in part, a consequence of (8.16), the sum of \( p_1^2 \delta_{11} + p_1 p_2 \delta_{12} + p_2^2 \delta_{22} = 0 \), namely,

\[
\begin{align*}
2p_1^2 \delta_{11} + 2p_1 p_2 \delta_{12} + 2p_2^2 \delta_{22} &= 0 \\
p_1^2 \delta_{11} + p_2^2 \delta_{22} &= -2p_1 p_2 \delta_{12}
\end{align*}
\]  
(8.50)

From (8.41) the additive variance for the particular locus is, substituting \( p_1^m = f \), \( p_1 = p \), \( p_2^m = f \), \( p_2 = q \), and \( \alpha \) (see (8.45)),

\[
\sigma_A^2 = 2\sigma_\alpha^2 = 2p_1 p_2 [p_1 (G_{11} - G_{12}) + p_2 (G_{12} - G_{22})]^2 = 2pq[a + (q-p)d]^2 = 2pq^2 \]  
(8.51)

From (8.42) the dominance variance for the particular locus is

\[
\sigma_D^2 = \sigma_\delta^2 = p_1 p_2 (2G_{12} - G_{11} - G_{22})^2 = 4p^2 q^2 d^2
\]  
(8.52)

In an analogous way to that discussed above (8.43), the level of dominance does affect the additive effects, \( \alpha \)'s, and thereby the additive variance, i.e.,

\[
\sigma_A^2 = 2\sigma_\alpha^2 = 2p_1 p_2 \left[ \frac{G_{11} - G_{22}}{2} + (p_2 - p_1) \right]^2
\]  
(8.53)

If there is dominance, it is only when the allelic frequencies are \( 1/2 \), \( p_1 = p_2 \), that the additive variance is not affected by the level of dominance, as stated above.

2. Regression approach. The above definitions of effects have been from a two-factor, factorial point of view, where both factors and their levels, the male and female allelic arrays, are one and the same. For two alleles, it is also convenient to view this least-squares fitting process from a regression point of view. It is from this viewpoint that one can best understand the usefulness of the additive variance and its role in predicting genetic gain from selection. Let the \( X \) variable be the number of \( A_1 \) alleles in a genotype. Its mean is the mean number of \( A_1 \) alleles per individual in the population, or twice the allelic
frequency of $A_1$ in the population. (Some writers have let the $X$ variable be the frequency of the $A_1$ allele in the genotype, i.e. 0, 1/2, and 1, or one-half of the number of $A_1$ alleles in a genotype. See Empig, Gardner, and Compton, 1972.) The $Y$ variable is the genotypic value of the genotype. Thus, we have the following

\[
\begin{align*}
G_{11} & \quad 2p_1p_2 \\
G_{12} & \quad x_2 \\
G_{22} & \quad P_2 - P_1
\end{align*}
\]

Number of $A_1$ alleles in genotype
or
mean number of $A_1$ alleles per individual in population

We desire to obtain a least-squares regression line for these three points. A weighted regression procedure must be used, because each genotype is not equally frequent in the population (see Steel and Torrie, 1980, Section 10.13). To contrast this procedure with the usual regression procedure of equal weights, we first briefly review the latter. The usual regression model is written as

\[
Y_i = \mu + \beta(X_i - \overline{X}) + \epsilon_i \quad i = 1, \ldots, n
\]

and estimates of $\mu$ and $\beta$ which minimize

\[
Q = \Sigma \epsilon_i^2 = \Sigma [Y_i - \mu - \beta(X_i - \overline{X})]^2
\]

are (Steel and Torrie, 1980, p. 242)
\[
\hat{\beta}_1 = \frac{\Sigma (X_1 - \bar{X})(Y_1 - \bar{Y})}{\Sigma (X_1 - \bar{X})^2} = \frac{\Sigma X_1 Y_1}{\Sigma X_1^2} - \frac{(\Sigma X_1)(\Sigma Y_1)}{n} \frac{n}{\Sigma X_1^2 - (\Sigma X_1)^2}
\]

\[
\hat{\mu} = \frac{\Sigma Y_1}{n} = \bar{Y}
\]

(8.57)

In weighted regression, we have the same model (8.55), but desire estimates of \(\mu\) and \(\beta\) which minimize

\[
Q = \Sigma w_i \varepsilon_i^2 = \Sigma w_i [Y_i - \mu - \beta(X_i - \bar{X})]^2
\]

where \(w_i\) = relative weight or frequency of ith Y value.

The least-squares estimates are

\[
\hat{\beta} = \frac{\Sigma w_i (X_i - \bar{X})(Y_i - \bar{Y})}{\Sigma w_i(X_i - \bar{X})^2} = \frac{\Sigma w_i X_i Y_i}{\Sigma w_i (X_i - \bar{X})^2}
\]

\[
= \frac{\Sigma w_i X_i Y_i}{\Sigma w_i (X_i - \bar{X})^2} - \frac{(\Sigma w_i X_i)(\Sigma w_i Y_i)\Sigma w_i}{\Sigma w_i X_i^2 - (\Sigma w_i X_i)^2}
\]

(8.59)

where \(\bar{X} = \frac{\Sigma w_i X_i}{\Sigma w_i}\),

\[
\hat{\mu} = \bar{Y} = \frac{\Sigma w_i Y_i}{\Sigma w_i}
\]

The estimates in (8.59) are exactly the ones that we want to use in fitting a regression line in (8.54), where the weights are the genotypic frequencies, i.e.,

\[
w_1 = p_1^2, \quad w_2 = 2p_1p_2, \quad w_3 = p_2^2
\]

(8.60)

So from (8.59)

\[
\bar{X} = \frac{\Sigma w_i X_i}{\Sigma w_i} = \frac{p_1^2(2) + 2p_1p_2(1) + p_2^2(0)}{p_1^2 + 2p_1p_2 + p_2^2} = 2p_1^2 + 2p_1p_2 - 2p_1(p_1 + p_2) = 2p_1
\]

(8.61)
\[ \bar{Y} = \frac{p_1^2 G_{11} + 2p_1p_2 G_{12} + p_2^2 G_{22}}{p_1^2 + 2p_1p_2 + p_2^2} = p_1^2 G_{11} + 2p_1p_2 G_{12} + p_2^2 G_{22} = \mu \]

To calculate \( \hat{\beta} \), it is convenient to write (8.60) and (8.61) in table form as follows:

<table>
<thead>
<tr>
<th>i</th>
<th>Genotype</th>
<th>( w_i )</th>
<th>Number of ( A_1 ) alleles</th>
<th>Genotypic value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( A_1A_1 )</td>
<td>( p_1^2 )</td>
<td>2</td>
<td>( G_{11} = G_2 )</td>
</tr>
<tr>
<td>2</td>
<td>( A_1A_2 )</td>
<td>( 2p_1p_2 )</td>
<td>1</td>
<td>( G_{12} = G_1 )</td>
</tr>
<tr>
<td>3</td>
<td>( A_2A_2 )</td>
<td>( p_2^2 )</td>
<td>0</td>
<td>( G_{22} = G_0 )</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td>2( p_1 )</td>
<td>( G_{..} )</td>
</tr>
</tbody>
</table>

Substituting (8.62) in (8.59), we obtain (see Box 8.6)

\[ \hat{\beta} = \beta = p_1(p_{11} - G_{12}) + p_2(G_{12} - G_{22}) = \alpha \]  

(8.63)

which means that the slope of the least-squares regression line is equal to the average effect of an allelic substitution \( \alpha \) (8.45). We omit the hat on \( \mu \) in (8.61) and that on \( \alpha \) in (8.63), because we are not sampling. We are using the least squares procedure to define population parameters and not to obtain their estimates or statistics.

**Box 8.6**

Derivation of (8.63)

From (8.59) we choose the form

\[ \hat{\beta} = \frac{\Sigma w_i (X_i - \bar{X})Y_i}{\Sigma w_i (X_i - \bar{X})^2} \]

and substitute values from (8.62) to obtain (let \( p_1 = p \), \( p_2 = q \))

\[ \hat{\beta} = \frac{p^2 (2 - 2p)G_{11} + 2pq(1 - 2p)G_{12} + q^2(0 - 2p)G_{22}}{p^2(2 - 2p)^2 + 2pq(1 - 2p)^2 + q^2(0 - 2p)^2} \]
\[
\begin{align*}
\frac{2p^2qG_{11} + 2pq(q - p)G_{12} - 2pq^2G_{22}}{4p^2(1 - p)^2 + 2pq(q - p)^2 + 4pq^2} \\
= \frac{2pq[pG_{11} + (q - p)G_{12} - qG_{22}]}{4p^2q^2 + 2pq(q - p)^2 + 4pq^2} \\
= \frac{2pq(pG_{11} - pG_{12} + qG_{12} - qG_{22})}{2pq(2pq + q^2 - 2pq + p^2 + 2pq)} \\
= \frac{p(G_{11} - G_{12}) + q(G_{12} - G_{22})}{p^2 + 2pq + q^2} \\
= p(G_{11} - G_{12}) + q(G_{12} - G_{22})
\end{align*}
\]

which is (8.63).

Substituting \( \alpha \) for \( \beta \) (8.63) in the general statistical model (8.55), we have

\[
Y_i = \mu + \alpha (X_i - \bar{X}) + \epsilon_i \quad i = 1, 2, 3 \quad (8.64)
\]

The best fitted genotypic values are

\[
\hat{Y}_i = \hat{G}_i = \mu + \alpha (X_i - \bar{X}) \quad i = 1, 2, 3 \quad (8.65)
\]

which for each \( i \) or genotype becomes ((8.62), \( \bar{X} = 2p_1 \) (8.61))

\[
\begin{align*}
\hat{G}_2 = \hat{G}_{11} &= \mu + \alpha (2 - 2p_1) \\
&= \mu + 2(1 - p_1)\alpha \\
&= \mu + 2p_2\alpha \quad \text{(sub. (8.45a))} \\
&= \mu + 2\alpha_1 \\
&= \mu + \alpha_1 + \alpha_1 \quad (8.66a)
\end{align*}
\]

\[
\begin{align*}
\hat{G}_1 = \hat{G}_{12} &= \mu + \alpha (1 - 2p_1) \\
&= \mu + \alpha (p_1 + p_2 - 2p_1) \\
&= \mu + (p_2 - p_1)\alpha \\
&= \mu + p_2\alpha - p_1\alpha \quad \text{(sub. (8.45)} \\
&= \mu + \alpha_1 + \alpha_2 \quad (8.66b)
\end{align*}
\]
\[ \hat{G}_0 = \hat{G}_{22} = \mu + a(0 - 2p_1) \]
\[ = \mu - 2p_1a \quad \text{(sub. (8.45b))} \]
\[ = \mu + 2a_2 \]
\[ = \mu + a_2 + a_2 \quad (8.66c) \]

where \( \hat{G}_{ij} \) = fitted genotypic value or occasionally breeding or additive or genic value.

Thus, these fitted values (8.66) are equal to the sum of the first three terms in the model (8.2) or are equal to the predicted cell values in the factorial approach (8.30). Although the mean is occasionally included in the definition of the breeding or additive or genic value, more commonly they are defined as the fitted genotypic value minus the population mean or simply as the sum of the two additive terms in the model (8.66), namely,

\[ A_2 = A_{11} = \hat{G}_{11} - \mu = a_1 + a_2 \]
\[ A_1 = A_{12} = \hat{G}_{12} - \mu = a_1 + a_2 \]
\[ A_0 = A_{22} = \hat{G}_{22} - \mu = a_2 + a_2 \quad (8.67) \]

The breeding value of a parental individual is twice the deviation of its potential offspring and that that deviation is doubled to adjust for the fact that only half of the offspring values are derived from the genes of the parent being evaluated. The deviations from the regression line are equal to the dominance effects as defined in (8.47). The regression model (8.55) can then be written completely in genetic notation as

\[ G_{ij} = \mu + a(X_i - \bar{X}) + \delta_{ij} \quad (8.68) \]

The additive variance is that variation due to regression of genotypic value on the number of \( A_1 \) alleles in the genotype. It is the expectation of the squares of the breeding values, i.e., the mean of the squares of the breeding values, each weighted by their frequencies, namely,
\[ \sigma_A^2 = \text{E}(\alpha^m + \alpha^f)^2 = p_1^2(2p_2 \alpha)^2 + 2p_1 p_2 [(p_2 - p_1) \alpha]^2 + p_2^2(-2p_1 \alpha)^2 \]
\[ = 2p_1 p_2 (2p_1 p_2 + p_2 - 2p_1 p_2 + p_1^2 + 2p_1 p_2) \alpha^2 \]
\[ = 2p_1 p_2 \alpha^2 \] ~ (8.69)

which is the same as (8.51). The expectation of the square of the sum, \( \alpha^m + \alpha^f \), is equal to the sum of the expectation of the square of each individual term, because \( \text{E}(2\alpha^m \alpha^f) = 0 \), namely,
\[ \sigma_A^2 = \text{E}(\alpha^m + \alpha^f)^2 = \text{E}[(\alpha^m)^2 + 2\alpha^m \alpha^f + (\alpha^f)^2] \]
\[ = \text{E}(\alpha^m)^2 + \text{E}(2\alpha^m \alpha^f) + \text{E}(\alpha^f)^2 \] ~ (see Box 8.3)
\[ = \text{E}(\alpha^m)^2 + 0 + \text{E}(\alpha^f)^2 \]
\[ = \sigma_{\alpha}^2 + \sigma_{\alpha}^2 \]
\[ = 2\sigma_{\alpha}^2 \] ~ (8.70)

With independence between the male and female parts of the breeding value, one may simply square each part, neglecting the cross product, and weight the squares by their frequencies to obtain the same answer as in (8.70), namely,
\[ p_1^2(2\alpha_1^2) + 2p_1 p_2 (\alpha_1^2 + \alpha_2^2) + p_2^2(2\alpha_2^2) = 2p_1 (p_1 \alpha_1^2 + p_2 \alpha_2^2) + 2p_2 (p_1 \alpha_1^2 + p_2 \alpha_2^2) \]
\[ = 2p_1 \alpha_1^2 + 2p_2 \alpha_2^2 \]
\[ = 2\sigma_{\alpha}(p_1 + p_2) \]
\[ = 2\sigma_{\alpha}^2 \] ~ (8.70A)

The dominance variance is the residual variation or the variance of the departures from the regression line. It is the expectation of the squares of the dominance deviations (8.47), i.e.,
\[ \sigma_D^2 = \text{E}(\delta^2) = p_1^2(-2p_2d)^2 + 2p_1 p_2 (2p_1 p_2 d)^2 + p_2^2(-2p_1 d)^2 \]
\[ = (2p_1 p_2 d)^2 \] ~ (8.71)

The reader is referred to Kempthorne (1969, Sections 15.1 to 15.5) for a complete development of the regression approach with two alleles. Other less mathematical presentations are given by Li (1976), Chapter 3, Sections 1 and 2;
Falconer (1989), Chapters 7 and 8; and Turner and Young (1969), Chapter 3, Section 3.1.

Alternatively, one may use orthogonal polynomial coefficients (see Steel and Torrie, Sections 15.7 and 19.5, for unequal frequencies or unequal numbers of replicates; Anderson and Bancroft, 1952, Chapter 16; Kendall and Stuart, 1967, pp. 356-359) to obtain the linear regression coefficient, $\beta$ or $\alpha$ (8.63), and the quadratic regression coefficient, say $\alpha_D$, and then to obtain the additive and dominance variances therefrom. The linear and quadratic orthogonal polynomial coefficients, $w_{ij}$, are (Cockerham, 1954)

\[
\begin{align*}
\text{Linear (additive), } w_{1j} & : \quad \begin{array}{ccc}
A_1A_1 & A_1A_2 & A_2A_2 \\
2p_2 & p_2 - p_1 & -2p_1 \\
\end{array} \\
\text{Quadratic (dominance), } w_{2j} & : \quad \begin{array}{ccc}
1/p_1^2 & -2/2p_1p_2 & 1/p_2^2 \\
\end{array}
\end{align*}
\]

(8.71Aa) (8.71Ab)

Because the genotypic frequencies are unequal, the weighted sum of the orthogonal polynomial coefficients must equal zero, i.e.,

\[
\sum_j w_{1j} = p_1(2p_2) + 2p_1p_2(p_2 - p_1) + p_2(-2p_1) = 0
\]

(8.71Ba)

\[
\sum_j w_{2j} = p_1(1/p_1^2) + 2p_1p_2(-2/2p_1p_2) + p_2(1/p_2^2) = 0
\]

(8.71Bb)

The weights, $w_1$, are the corresponding genotypic frequencies (8.60). The linear orthogonal polynomial coefficients are the deviations of the values of the $X$ variable (2, 1, and 0) from their mean $2p_1$ (see 8.54). The derivation of the quadratic coefficients require more sophisticated procedures given in the above references. The linear and quadratic polynomial coefficients are also orthogonal, i.e.,

\[
\sum_j w_{1j}w_{2j} = p_1(2p_2)(1/p_1^2) + 2p_1p_2(p_2 - p_1)(-2/2p_1p_2) + p_2(-2p_1)(1/p_2^2) = 0
\]

(8.71C)

Then from least squares, multiple regression theory, any regression coefficient is equal to the weighted sum of cross products of the linear or
quadratic orthogonal polynomial coefficients times the genotypic values divided by the weighted sum of squares of the orthogonal polynomial coefficients (see Anderson and Bancroft, 1952, p. 211, Draper and Smith, 1981, Chapter 2) (note that the genotypic values are numbered in the reverse order, \(G_11 = G_1, G_{12} = G_2, G_{22} = G_3\), i.e.,

Linear (additive)(8.63):

\[
\alpha = \frac{\sum w_j w_{1j} G_j}{\sum w_j w_{1j}^2} = \frac{p_1^2(2p_2)G_{11} + 2p_1p_2(p_2 - p_1)G_{12} + p_2^2(-2p_1)G_{22}}{p_1^2(2p_2)^2 + 2p_1p_2(p_2 - p_1)^2 + p_2^2(-2p_1)^2}
\]

\[
= \frac{2p_1p_2[p_1G_{11} - p_1G_{12} + p_2G_{12} - p_2G_{22}]}{2p_1p_2(2p_1p_2 + p_2^2 - 2p_1p_2 + p_1^2 + 2p_1p_2)}
\]

\[
= p_1(G_{11} - G_{12}) + p_2(G_{12} - G_{22}) \quad (8.71Da)
\]

Quadratic (dominance):

\[
\alpha_D = \frac{\sum w_j w_{2j} G_j}{\sum w_j w_{2j}^2} = \frac{p_1^2(1/p_1)^2G_{11} + 2p_1p_2(-2/2p_1p_2)G_{12} + p_2^2(1/p_2)^2G_{22}}{p_1^2(1/p_1)^2 + 2p_1p_2(-2/2p_1p_2)^2 + p_2^2(1/p_2)^2}
\]

\[
= \frac{G_{11} - 2G_{12} + G_{22}}{1 + \frac{2}{p_1^2} + \frac{1}{p_2^2}}
\]

\[
= \frac{G_{11} - 2G_{12} + G_{22}}{p_2^2 + 2p_1p_2 + p_1}
\]

\[
= \frac{2p_1^2G_{11} - 2G_{12} + G_{22}}{p_1p_2} \quad (8.71Db)
\]

That these are the correct regression coefficients can be easily verified in that

\[
G_j = u + \alpha w_{1j} + \alpha_D w_{2j} \quad (8.71E)
\]

The additive and dominance variances are obtained by multiplying the corresponding regression coefficient by the weighted sum of cross products of the orthogonal polynomial coefficients by the genotypic values, i.e.,
\[ \sigma^2_A = \alpha \sum_j w_j W_{1j} C_j - \alpha(2p_1 p_2 \alpha) = 2p_1 p_2 \alpha^2 \]  \hspace{1cm} \text{(see (8.69))} \hspace{1cm} \text{(8.71Fa)}

\[ \sigma^2_D = \sigma_D^2 \sum_j w_j W_{2j} G_j = p_1^2 p_2^2 (C_{11} - 2G_{12} + C_{22})^2 = 4p_1^2 p_2^2 d^2 \]  \hspace{1cm} \text{(see (8.71))} \hspace{1cm} \text{(8.71Fb)}

8.1.5. **One locus, two alleles, equal male and female allelic frequencies** (Hardy-Weinberg) with arbitrary or equal allelic frequencies within sexual arrays and with no dominance. The dominance effects, \( \delta \)'s, and the dominance variance, \( \sigma^2_{\delta} = \sigma^2_D \), are measures of the failure of alleles to act additively. When alleles are additive in their effects, i.e., show no dominance, the difference between genotypic values for genotypes \( A_1^- \) and \( A_2^- \), where the blank denotes the presence of either \( A_1 \) or \( A_2 \) in both genotypes, is the same regardless of whether the \( A_1 \) or \( A_2 \) allele occupy the blank position, i.e., \( G_{11} - G_{21} = G_{12} - G_{22} \). With no dominance, the dominance variance, \( \sigma^2_{\delta} = \sigma^2_D \), equals zero, and the additive variance from (8.51) is

\[ \sigma^2_A = 2\sigma^2_\alpha = 2p_1 p_2 [p_1(G_{11} - G_{12}) + p_2(G_{12} - G_{22})]^2 \]  \hspace{1cm} \text{(sub. } G_{11} - G_{12} = G_{12} - G_{22} \text{)}

\[ = 2p_1 p_2 [(G_{11} - G_{12})(p_1 + p_2)]^2 \]

\[ = 2p_1 p_2 [(G_{11} - G_{22})/2]^2 \]  \hspace{1cm} \text{(note } a = G_{11} - G_{12} = G_{12} - G_{22} = (G_{11} - G_{22})/2 \text{)}

\[ = 2pqa^2 \]  \hspace{1cm} \text{(8.72)}

is a function of the comparison of homozygotes only. Equation (8.72) can also be obtained from (8.53), when the dominance contrast is zero (no dominance).

In the regression approach with no dominance, the regression line passes through the three genotypic values exactly, and hence the slope of the line (\( a = \alpha \)) is one-half of the difference between the two homozygotes as can be seen in (8.72) (compare \( 2pqa^2 \) (8.72) with \( 2pqa^2 \) (8.51)).

**Example 8.1.** With two alleles at one locus in Hardy-Weinberg proportions and with partial dominance, we assume the genotypic values and allelic frequencies to be
\[ G_{11} = 100, \ G_{12} = 95, \ G_{22} = 40; \quad p_1 = 0.6, \ p_2 = 0.4 \]

then

\[
\begin{array}{cc|cc}
A_1 & A_2 & & \\
 & & & \\
0.6 & 0.4 & & \\
A_1 & A_1A_1 & A_1A_2 & \\
G_{11} = 100 & G_{12} = 95 & & \\
0.36 & 0.24 & & \\
A_2 & A_2A_1 & A_2A_2 & \\
G_{21} = 95 & G_{22} = 40 & & \\
0.24 & 0.16 & & \\
\end{array}
\]

\[ G_.1 = 98 \quad G_.2 = 73 \quad G_. = \frac{.36(100) + .24(95)}{.36 + .24} = .6(100) + .4(95) = 98 \]
\[ G_.2 = \frac{.24(95) + .16(40)}{.24 + .16} = .6(95) + .4(40) = 73 \]

From (8.10) the average effects of the alleles are (could also use (8.45))

\[ \alpha_1 = p_1G_{11} + p_2G_{12} - G_. = G_.1 - G_. = 98 - 88 = 10 \]
\[ \alpha_2 = p_1G_{21} + p_2G_{22} - G_. = G_.2 - G_. = 73 - 88 = -15 \]

Check: From (8.7) or (8.8), \( p_1\alpha_1 + p_2\alpha_2 = .6(10) + .4(-15) = 6.0 - 6.0 = 0. \)

From (8.46) the average effect of an allelic substitution is the difference between the two row or column means

\[ \alpha = G_.1 - G_.2 = G_.1 - G_.2 = \alpha_1 - \alpha_2 = 10 - (-15) = 25 \]

The fitted cell means based on the effects of the particular row and column in which the cell occurs are (same as (8.66))

\[ \hat{G}_{11} = \mu + \alpha_1 + \alpha_1 = 88 + 10 + 10 = 108 \]
\[ \hat{G}_{12} = \mu + \alpha_1 + \alpha_2 = 88 + 10 + (-15) = 83 \]
\[ \hat{G}_{22} = \mu + \alpha_2 + \alpha_2 = 88 + (-15) + (-15) = 58 \]

The breeding or additive genetic values, measured from the population mean, are the above fitted cell means minus the population mean. Alternatively, the breeding values can be calculated by adding the average effects of the genes
\[ A_{11} = \alpha_1 + \alpha_1 = 2(10) = 20 \]
\[ A_{12} = \alpha_1 + \alpha_2 = 10 + (-15) = -5 \]
\[ A_{22} = \alpha_2 + \alpha_2 = 2(-15) = -30 \]

From (8.13) the dominance effects are (could also use (8.47))

\[ \delta_{11} = G_{11} - G_1 - G_1 + G_1 = 100 - 98 - 98 + 88 = -8 \]
\[ \delta_{12} = G_{12} - G_1 - G_2 + G_1 = 95 - 98 - 73 + 88 = 12 \]
\[ \delta_{21} = G_{21} - G_2 - G_1 + G_1 = 95 - 73 - 98 + 88 = 12 \]
\[ \delta_{22} = G_{22} - G_2 - G_2 + G_1 = 40 - 73 - 73 + 88 = 18 \]

Check: From (8.15)

\[ p_1 \delta_{11} + p_2 \delta_{12} = .6(-8) + .4(12) = -4.8 + 4.8 = 0 \]
\[ p_1 \delta_{21} + p_2 \delta_{22} = .6(12) + .4(-18) = 7.2 - 7.2 = 0 \]
\[ p_1 \delta_{11} + p_2 \delta_{21} = .6(-8) + .4(12) = -4.8 + 4.8 = 0 \]
\[ p_1 \delta_{12} + p_2 \delta_{22} = .6(12) + .4(-18) = 7.2 - 7.2 = 0 \]

Also note that from (8.49)

\[ p_1^2 \delta_{11} = p_2^2 \delta_{22} = \frac{1}{2} (2p_1p_2 \delta_{12}) \]

\[ (.6)^2(-8) = (.4)^2(-18) = \frac{1}{2} 2 (.6)(.4)12 \]

\[ -2.88 = -2.88 \]

Then the additive variance may be calculated by definition of a variance from the additive effects (8.18) (8.20)

\[ \sigma_A^2 = 2\sigma_\alpha^2 = 2 \sum_{i=1}^{2} p_i \alpha_i^2 = 2[.6(+10)^2 + .4(-15)^2] = 2(60 + 90) = 2(150) = 300 \]

or from the genotypic values themselves (8.51), using the average effect of an allelic substitution,

\[ \sigma_A^2 = 2\sigma_\alpha^2 = 2p_1p_2[p_1(G_{11} - G_{12}) + p_2(G_{12} - G_{22})]^2 \]
\[ = 2(.4)(.6)[.6(100 - 95) + .4(95 - 40)]^2 = .48(3 + 22)^2 = 300 \]
Similarly, the dominance variance may be calculated by definition from the dominance effects (8.21)

\[ \sigma_D^2 = \sigma_{\delta}^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} p_i p_j \delta_{ij}^2 = p_{11}^2 + 2p_{12}^2 + p_{22}^2 \]

\[ = .36(-8)^2 + 2(.24)(12)^2 + .16(-18)^2 = 144 \]

or from the genotypic values themselves (8.52), using the dominance contrast,

\[ \sigma_D^2 = \sigma_{\delta}^2 = p_{11}^2 p_{22}^2 (2G_{12} - G_{11} - G_{22})^2 \]

\[ = (.4)^2(.6)^2[2(95) - 100 - 40]^2 = .16(.36)(50)^2 = 144 \]

The total genotypic variance equals the sum of the additive and dominance variances (8.29)

\[ \sigma_G^2 = \sigma_A^2 + \sigma_D^2 = 300 + 144 = 444 \]

The total genotypic variance may also be calculated from (8.22) as a check

\[ \sigma_G^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} p_i p_j (G_{ij} - \mu)^2 \]

\[ = .36(100 - 88)^2 + .48(95 - 88)^2 + .16(40 - 88)^2 \]

\[ = .36(12)^2 + .48(7)^2 + .16(-48) = 444 \]

This example is also presented from a regression point of view in Figure 8.1. The genotypic values are given for three scales of measurement: (1) scale 1 is the uncoded G scale where \( G_{11} = MP + a = 100, G_{12} = MP + d = 95, G_{22} = MP + (-a) = 40, \mu = G.. = MP + M = 88 \) (8.31) (8.44), (2) scale 2 is the coded one where \( a = 30, d = 25, \) and \(-a = -30,\) the mid-homozygote value \( MH = 70 \) (8.31), and the population mean \( M = (p - q)a + 2pqd = 18 \) (8.44) on the coded scale, and (3) scale 3 is simply one that measures the genotypic value on either scale 1 or 2 from its corresponding mean, \( \mu \) or \( M. \) The breeding values are also given for all three scales, although they are usually measured from the population mean -- scale 3. The breeding values are symbolized \( A_{ij} = \alpha_i + \alpha_j \) in the case for scale 3. Scales
Figure 8.1. Graphic representation of genotypic values (closed circles) and breeding values (open circles), and average effects of alleles and the average effect of an allelic substitution, $a = 30$, $d = 25$, $p = 0.6$, $q_{A1} = p = 0.6$, $q_{A2} = q = 0.4$.
1 and 2 really give the fitted genotypic values for the uncoded and coded scales, respectively. It is entirely coincidental that \( d = \alpha = 25 \).

The additive variance is the expectation of the square of the breeding values (scale 3)

\[
\sigma_A^2 = E(A_{ij}^2) = E(\hat{G}_{ij} - \mu)^2 = \sum_{i,j} p_i p_j (\hat{G}_{ij} - \mu)^2 = .36(20)^2 + .48(-5)^2 + .16(-30)^2 = 300
\]

and the dominance variance is

\[
\sigma_D^2 = E(\delta_{ij}^2) = \sum_{i,j} p_i p_j \delta_{ij}^2 = .36(-8) + .48(12)^2 + .16(-18)^2 = 144
\]

The average effect of each of the alleles, and the breeding value of each of the genotypes are identified and labeled. Note that the average effect of an allelic substitution is the slope of the regression line.

**Example 8.2** With two alleles at one locus in Hardy-Weinberg equilibrium and with no dominance, we assume the genotypic values and allelic frequencies to be

\( G_{11} = 60, \ G_{12} = 40, \ G_{22} = 20; \quad p_1 = 0.7, \ p_2 = 0.3 \)

then

\[
\begin{array}{cccc}
A_1 & A_2 \\
0.7 & 0.3 \\
\hline
A_1 & A_1 A_1 & A_1 A_2 \\
G_{11} = 60 & G_{12} = 40 & \quad G_{1.} = .7(60) + .3(40) = 54 \\
0.7 & .49 & .21 \\
A_2 & A_2 A_1 & A_2 A_2 \\
G_{21} = 40 & G_{22} = 20 & \quad G_{2.} = .7(40) + .3(20) = 34 \\
0.3 & .21 & .09 \\
\hline
G_{1.} = 54 & G_{2.} = 34 & \quad G_{..} = .49(60) + .42(40) + .09(20) = 48
\end{array}
\]
From (8.10) (or also (8.45)) the additive effects are

\[ \alpha_1 = (p_1 G_{11} + p_2 G_{12}) - G.. - G_1 - G_1 = 54 - 48 = 6 \]

\[ \alpha_2 = (p_1 G_{21} + p_2 G_{22}) - G.. - G_2 - G_2 = 34 - 48 = -14 \]

Check: From (8.7) or (8.8), \( p_1 \alpha_1 + p_2 \alpha_2 = .7(6) + .3(-14) = 4.2 - 4.2 = 0 \).

From (8.13) (or also (8.47)) the dominance effects are

\[ \delta_{11} = G_{11} - G_1 - G_1 + G.. = 60 - 54 - 54 + 48 = 0 \]

\[ \delta_{12} = G_{12} - G_1 - G_2 + G.. = 40 - 54 - 34 + 48 = 0 \]

\[ \delta_{21} = G_{21} - G_2 - G_1 + G.. = 40 - 34 - 54 + 48 = 0 \]

\[ \delta_{22} = G_{22} - G_2 - G_2 + G.. = 20 - 34 - 34 + 48 = 0 \]

Then the additive variance may be calculated by definition from the additive effects (8.18) (8.20)

\[ \sigma_A^2 = 2\sigma_a^2 - 2 \sum_{i=1}^{2} p_i \alpha_i^2 = 2\{.7(6)^2 + .3(-14)^2\} = 2(25.2 + 58.8) = 168 \]

or from the genotypic values themselves (8.51)

\[ \sigma_A^2 = 2\sigma_a^2 - 2p_1p_2[p_1(G_{11} - G_{12}) + p_2(G_{12} - G_{22})]^2 \]

\[ = 2(.3)(.7)[.7(60 - 40) + .3(40 - 20)]^2 = .42(14 + 6)^2 = 168 \]

There is no dominance variance, so the total genotypic variance equals the additive variance

\[ \sigma_G^2 = \sigma_A^2 = 168 \]

Alternatively, the total genotypic variance may also be calculated from (8.22) as a check, namely,

\[ \sigma_G^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} p_ip_j(G_{ij} - \mu)^2 \]

\[ = .49(60 - 48)^2 + .42(40 - 48)^2 + .09(20 - 48)^2 \]

\[ = .49(12)^2 + .42(-8)^2 + .09(-28)^2 = 168 \]

* * * * * * *
8.1.5.1. **Change of population mean.** Although we do not discuss selection here (see Chapter 12), all of the above is defined with selection in mind. Let us think what happens under selection if the \( A_1 \) allele were favored. First, with respect to the \( X \) axis, the mean number \( 2p \) of \( A_1 \) alleles per individual in the population would increase, because the frequency of the \( A_1 \) allele would increase. Second, with respect to the \( Y \) axis, the mean of the population would also increase. The question is: Would the new mean lie exactly on the line? To investigate this question, we first recall that the original population mean is (8.44) (coded values will be used throughout in this and the following section)

\[
M_0 = p^2a + 2pqd + q^2(-a) = (p - q)a + 2pqd \tag{8.73}
\]

By selection for one generation the frequency of \( A_1 \) allele is changed \( \Delta p \), so the mean of the selected population after random mating is

\[
M_1 = (p + \Delta p)^2a + 2(p + \Delta p)(q - \Delta p)d + (q - \Delta p)^2(-a)
\]

\[
= [p^2 + 2p\Delta p + (\Delta p)^2]a + 2[pq - p\Delta p + q\Delta p - (\Delta p)^2]d
\]

\[
+ [q^2 - 2q\Delta p + (\Delta p)^2](-a)
\]

\[
= M_0
\]

\[
= p^2a + 2pqd + q^2(-a) + [2p\Delta p + (\Delta p)^2 + 2q\Delta p - (\Delta p)^2]a
\]

\[
+ 2\Delta p[(q - p) + \Delta p]d
\]

\[
= M_0 + 2\Delta p(p + q)a + 2\Delta p[(q - p) - \Delta p]d
\]

\[
= M_0 + 2\Delta p[a + (q - p)d] - 2(\Delta p)^2d \quad \text{(sub. (8.45))}
\]

\[
= M_0 + 2\Delta p a - 2(\Delta p)^2d \tag{8.74}
\]

Thus, from (8.74) we note that if we neglect the second-order term of \( \Delta p \), the new mean would lie on the line \( \Delta p \) is the change on the \( X \) axis and \( a \) is the slope of the line. This would be true for either a positive \( \Delta p \) or a negative \( \Delta p \). Hence, the least-squares regression line as defined gives a first-order approximation to the new mean under selection. In fact, if \( d \) were zero, the new population mean would lie exactly on the line.
However, if we consider the second-order term, the term itself will be positive for either positive or negative $\Delta_p$ whenever $d$ itself is positive. Hence for positive $d$, the population mean always lies below the line, regardless of the direction of selection. For positive $\Delta_p$, we have the following as shown in part (a) for one generation of selection and that in part (b) for several successive generations.

Likewise, for negative $\Delta_p$
The curve of population means is one that passes through \((2p, M_0)\), lies below the original straight line for the entire range of \(p\) from 0 to 1, and also lies below each newly constructed regression line for each generation of selection for the entire range of \(p\) greater than 0 and less than 1. Whenever \(d\) is positive, the fitted regression line is always above "-a" and "a", and below \(d\). That the regression line lies above "-a" and "a", and that the population mean must converge on "a" as \(p\) approaches 1 for positive \(\Delta_p\) or converge on "-a" as \(p\) approaches 0 for negative \(\Delta_p\) accounts for the locus of population means always being below the regression line. If \(d\) is negative, the reverse of the above situation occurs.

8.1.5.2. \textbf{Changes in additive, dominance, and genotypic variances.} Since the additive, dominance, and genotypic variances depend upon allelic frequencies, it is worthwhile to compare their relative magnitude for different allelic frequencies. We do this in varying degrees for five cases: (1) no dominance \((d = 0)\), (2) complete dominance \((d = a)\), (3) partial dominance \((0 < d < a)\), (4) "pure" overdominance \((a = -a = 0, d > 0)\) (see Falconer, 1989, p. 130), and (5) overdominance \((a > -a, d > a)\).

1. \textbf{No dominance} \((d = 0)\). The genotypic variance equals the additive variance, and is (8.51) (8.72)

\[
\sigma^2_G = \sigma^2_A = 2pq\alpha = 2pq\alpha^2 = 2p(1 - p)a^2 = 2a^2(p - p^2) \tag{8.77}
\]

It reaches a maximum when \(p = q = 1/2\), as can be easily found by differentiating \(\sigma^2_G\) (8.77) with respect to \(p\), namely,

\[
\frac{d\sigma^2_G}{dp} = 2a^2(1 - 2p) = 0
\]

\[
2p = 1
\]

\[
p = 1/2 \tag{8.78}
\]
Thus, the maximum genotypic variance is obtained by substituting (8.78) in (8.77) and equals one-half of \( a \) squared, namely,

\[
\sigma_G^2 = 2\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)a^2 = \frac{1}{2} a^2
\]  

(8.79)

The genotypic variance is plotted in Figure 8.2(a).

![Graphs illustrating genotypic variance](image)

Figure 8.2. Relative magnitude of additive, dominance, and genotypic variances for (a) no dominance \((d = 0)\), (b) complete dominance \((d = a)\), and (c) "pure" overdominance \((a = -a = 0, d > 0)\).
2. **Complete dominance** \((d = a)\). The additive variance can be obtained by substituting \(a\) for \(d\) in (8.51), giving

\[
\sigma_A^2 = 2pq[a + (q - p)d]^2 \quad \text{(sub. } a = d) \\
= 2pq[a + (q - p)a]^2 \\
= 2pqa^2[1 + q - p]^2 \\
= 2pqa^2[p + q + q - p]^2 \\
= 2pqa^2(2q)^2 \\
= 8pq^3a^2
\]  

(8.80)

Similarly, substituting \(a\) for \(d\) in (8.52) gives

\[
\sigma_D^2 = 4p^2q^2d^2 \\
= 4p^2q^2a^2
\]  

(8.81)

Thus, from (8.80) and (8.81)

\[
\sigma_C^2 = \sigma_A^2 + \sigma_D^2 = 8pq^3a^2 + 4p^2q^2a^2 \\
= 4pq^2a^2(2q + p) \\
= 4pq^2a^2(q + 1)
\]  

(8.82)

The maximum value of each of the functions for the additive variance (8.80), dominance variance (8.81), and genotypic variance (8.82) occurs for the following allelic frequency (see Box 8.7)

- **Additive variance:** \(p = 0.25\)
- **Dominance variance:** \(p = 0.50\)  
- **Genotypic variance:** \(p = 1 - \frac{1}{2} \sqrt{2} = 0.293\)  

(8.83)

For that of the genotypic variance, the maximum occurs when the sum of the frequencies of the \(A_1A_1\) and \(A_1A_2\) genotypes is equal to the frequency of the \(A_2A_2\) genotype, i.e.,
\[ p^2 + 2pq = q^2 \]
\[ = (1 - p)^2 \quad \text{(sub. (8.83))} \]
\[ = [1 - (1 - \frac{1}{2} \sqrt{2})]^2 \]
\[ = \left( \frac{1}{2} \sqrt{2} \right)^2 \]
\[ = \frac{2}{4} \]
\[ = \frac{1}{2} \quad \text{(8.84)} \]

**Box 8.7**

**Derivation of (8.83)**

1. To find \( p \) when the additive variance is a maximum, we must differentiate (8.80) with respect to \( p \)

\[ \sigma^2_A = 8pq^3a^2 = 8a^2 p(1 - p)^3 \]  

(1)

Thus, using the differentiation formula, \( d(uv) = u \frac{dv}{dx} + v \frac{du}{dx} \) for a product of two factors, \( u = p \) and \( v = (1 - p)^3 \), we have

\[ \frac{d\sigma^2_A}{dp} = 8a^2[p(3)(1 - p)^2(-1) + (1 - p)^3 (1)] \]
\[ = 8a^2[-3p(1 - 2p + p^2) + (1 - 3p + 3p^2 - p^3)] \]
\[ = 8a^2(-3p + 6p^2 - 3p^3 + 1 - 3p + 3p^2 - p^3) \]
\[ = 8a^2(1 - 6p + 9p^2 - 4p^3) \]
\[ = 8a^2(1 - p)(1 - 5p + 4p^2) \]
\[ = 8a^2(1 - p)(1 - p)(1 - 4p) \]  

(2)

(The above actually involves more algebra than if we would have first expressed \( 8a^2 p(1 - p)^3 \) as a polynomial \( 8a^2(p - 3p^2 + 3p^3 - p^4) \) and differentiated it.)

Set the first derivative (2) equal to zero, and solve for \( p \) by setting each factor equal to zero, namely,
\[(1 - p)(1 - p)(1 - 4p) = 0\]
\[1 - p = 0; \quad 1 - 4p = 0\]
\[p = 1 \quad 4p = 1\]
\[p = 1/4\] (3)

Thus, the additive variance is a maximum at \(p = 1/4\).

2. For the dominance variance (8.81), we have
\[
\sigma_D^2 = 4p^2 q^2 a^2 = 4a^2 p^2 (1 - p)^2 = 4a^2 (p^2 - 2p^3 + p^4)
\] (4)
\[
\frac{d\sigma_D^2}{dp} = 4a^2 (2p - 6p^2 + 4p^3)
\]
\[= 8a^2 p (1 - 3p + 2p^2)
\]
\[= 8a^2 p (1 - p) (1 - 2p)
\] (5)

and setting (5) equal to zero, and solving for \(p\)
\[p = 0; \quad 1 - p = 0; \quad 1 - 2p = 0\]
\[p = 1 \quad 2p = 1\]
\[p = 1/2\] (6)

Thus, the dominance variance is a maximum at \(p = 1/2\).

3. For the genotypic variance (8.82), we have
\[
\sigma_G^2 = 4pq^2 a^2 (q + 1) = 4a^2 p (1 - p)^2 (2 - p)
\]
\[= 4a^2 (p - 2p^2 + p^3) (2 - p)
\]
\[= 4a^2 (2p - 5p^2 + 4p^3 - p^4)
\] (7)
\[
\frac{d\sigma_G^2}{dp} = 4a^2 (2 - 10p + 12p^2 - 4p^3)
\]
\[= 8a^2 (1 - 5p + 6p^2 - 2p^3)
\]
\[= 8a^2 (1 - p) (1 - 4p + 2p^2)
\] (8)

and setting (8) equal to zero, and solving for \(p\), we obtain
\[ 1 - p = 0; \quad 1 - 4p + 2p^2 = 0 \]
\[ p = 1 \]
\[ p = \frac{-(-4) \pm \sqrt{(-4)^2 - 4(2)(1)}}{2(2)} \]
\[ = 1 \pm \sqrt{(1/2)} \]

Thus, the genotypic variance is a maximum for \( p = 1 - \sqrt{(1/2)} = 0.293. \)

Each of the three variances is plotted against \( p \), the frequency of the \( A_1 \) or the "favorable" allele, in Fig. 8.2(b). We notice that as \( p \) increases beyond 0.67, \( \sigma_D^2 > \sigma_A^2 \). Hence, in a "nearly plateaued" population undergoing selection for \( A_1 \), one would expect \( \sigma_D^2 \) to be much greater than \( \sigma_A^2 \). This ratio of the dominance variance to the additive variance is given by ((8.81) (8.80))

\[ \frac{\sigma_D^2}{\sigma_A^2} = \frac{4p^2q^2a^2}{8pq^2a^2} = \frac{p}{2q} = \frac{p}{2(1 - p)} \quad (8.85) \]

and is plotted in Fig. 8.3(a). The additive variance decreases whereas the dominance variance increases relative to the genotypic variance as \( p \) increases.

This is shown by plotting ((8.80) (8.81) (8.82))

\[ \frac{\sigma_A^2}{\sigma_G^2} = \frac{8pq^2a^2}{4pq^2a^2(q + 1)} = \frac{2(1 - p)}{2 - p} \quad (8.86) \]

and

\[ \frac{\sigma_D^2}{\sigma_G^2} = \frac{4p^2q^2a^2}{4pq^2a^2(q + 1)} = \frac{p}{2q} = \frac{p}{2(1 - p)} \quad (8.87) \]

in Fig. 8.3(b) and (c).
Figure 8.3. Values of the three possible ratios involving additive, dominance, and genotypic variances for complete dominance.

3. Partial dominance \((0 < d < a)\). Before the reader reads this section, the writer recommends reading Section 8.2.1. To investigate various properties for partial dominance, it is convenient to use the U scale of Comstock and Robinson (8.101). In their scale, a is the proportion of \(d\) to \(a\) in a scale of Falconer (8.106). From Example 8.4, the additive variance is

\[ \sigma_A^2 = 2p(1 - p)[1 + (1 - 2p)a]^2u^2 \]  

(8.87A)

and by substituting \(au\) for \(d\) (8.101) in (8.52), we obtain

\[ \sigma_D^2 = 4p^2(1 - p)^2a^2u^2 \]  

(8.87B)

Thus, from (8.80) and (8.81).

\[ \sigma_G^2 = \sigma_A^2 + \sigma_D^2 = 2p(1 - p)[1 + (1 - 2p)a]^2u^2 + 4p^2(1 - p)^2a^2u^2 \]

\[ = 2u^2p(1 - p)[2a^2p^2 - 2a(a + 2)p + (a + 1)^2] \]  

(8.87C)

The first derivative of \(\sigma_A^2\) (8.87A) with respect to \(p\) is
\[
\frac{d\sigma^2_A}{dp} = 2u^2[-16a^2p^3 + 12a(2a + 1)p^2 - 2(5a + 1)(a + 1)p + (a + 1)^2] \tag{8.87D}
\]

I have not taken the time to formalize the properties of partial dominance any further, but from plotting the ratios of \(\sigma_D^2/\sigma_A^2\) for \(a = 0.1, 0.2 \ldots, 0.9\) with respect to \(p\), it is apparent that all curves start at zero for \(p = 0\) and increase to different maxima and then decrease to zero for \(p = 1\). For the maximum value of \(\sigma_D^2/\sigma_A^2\), \(p\) increases, as \(a\) increases. In addition, the maximum value of the ratio increases, as \(a\) increases. All maximum values are less than one for \(a \leq 0.8\), but for \(a = 0.9\), the maximum value exceeds 2.1 for about \(p = 0.95\), and then the ratio rapidly declines to zero for \(p = 1\). As \(a\) approaches one, the ratio \(\sigma_D^2/\sigma_A^2\) goes to infinity (see Fig. 8.3a), instead of returning to zero.

For the ratio of \(\sigma_A^2/\sigma_C^2\), all curves for different values of \(a\) start at one for \(p = 0\), decline to different minima and then return to one for \(p = 1\). For the minimum, \(p\) increases, as \(a\) increases. The minimum value decreases, as \(a\) increases. For \(a = 0.9\), the minimum value is about 0.32 for \(p = 0.95\), and then the curve rapidly returns to one. As \(a\) approaches one, the ratio \(\sigma_A^2/\sigma_C^2\) goes to zero (see Fig. 8.3b), instead of returning to one.

The curves for the ratio \(\sigma_D^2/\sigma_C^2\) are somewhat similar to those for \(\sigma_D^2/\sigma_A^2\). All curves start at zero for \(p = 0\) and return to zero for \(p = 1\). For the maximum, \(p\) increases, as \(a\) increases. For \(a = 0.9\); the maximum is about 0.68 and then the curve rapidly returns to zero for \(p = 1\). Again as \(a\) approaches one, the ratio \(\sigma_D^2/\sigma_C^2\) goes to one (see Fig. 8.3c), instead of returning to zero.
4. "Pure" overdominance \((a = -a = 0, d > 0)\). The additive variance can be obtained by setting "a" equal to zero in (8.51), giving
\[
\sigma^2_A = 2pq[a + (q - p)d]^2
= 2pq[(q - p)d]^2
\]
(8.88)
The dominance variance from (8.52) is
\[
\sigma^2_D = (2pqd)^2
\]
(8.89)
The genotypic variance is
\[
\sigma^2_G = \sigma^2_A + \sigma^2_D = 2pq[(q - p)d]^2 + (2pqd)^2
= 2pq[q^2 + 2pq + 2pqd]^2
= 2pq[q^2 + p^2]d^2
\]
(8.90)
The maximum values for the additive variance are for allelic frequencies 0.149 and 0.851, and the maximum values for the dominance and genotypic variances are for allelic frequency of one-half (derivation is similar to that in Box 8.7).

Each of the three variances is plotted in Figure 8.2(c). Note that when the two homozygotes have equal genotypic values and are equally frequent (equal allelic frequencies), the additive variance becomes zero, and the dominance variance equals the genotypic variance. This should be intuitive from a consideration of the regression of genotypic values on the number of \(A_1\) alleles (8.54).

5. Overdominance \((a > -a, d > a)\). Again this section will not be developed thoroughly. For the case of overdominance, in general, there is an intermediate frequency \(p\) for which the additive variance equals zero. This occurs whenever the average effect of an allelic substitution (8.45) (8.63) in the expression for the additive variance (8.51) equals zero, i.e.,
\[
P_1(G_{11} - G_{12}) + P_2(G_{12} - G_{22}) = 0
\]
\[
P_1(G_{11} - G_{12}) = -P_2(G_{12} - G_{22})
\]
(8.90A)
For example, when \( G_{11} = 20 \), \( G_{12} = 30 \), and \( G_{22} = 15 \), and

\[
p_1(20 - 30) = -p_2(30 - 15) \\
-10p_1 = -15p_2 \\
= -15(1 - p_1) \\
25p_1 = 15 \\
p_1 = 15/25 = 0.60
\]  
(8.90B)

then the additive variance equals zero.

8.1.6. **Two or more loci, multiple alleles, arbitrary "male" and "female" allelic frequencies (cross between two populations), no epistasis:** General. 1.

**Linkage equilibrium.** To be general, we assume the same conditions as we did in Section 8.1.1, namely, multiple alleles, and a cross between two populations, \( m \) and \( f \), such that random union of alleles at any locus exists. In this section we assume many segregating loci whose genes at different loci do not interact or show no epistasis, i.e., loci are said to be additive. In that case, the model for a genotypic value can be obtained by summing over the \( n \) loci. First, we rewrite the single-locus model (8.2) for the \( k \)-th locus as

\[
G_{k_1 k_j} = \mu_k + \alpha_{k_1}^m + \alpha_{k_j}^f + \delta_{k_1 k_j} = \mu_k + y_k
\]  
(8.91)

where \( y_k = G_{k_1 k_j} - \mu_k = \alpha_{k_1}^m + \alpha_{k_j}^f + \delta_{k_1 k_j} = \) total genotypic or locus effect at \( k \)-th locus.

Then we obtain the **locus form** of the model for the genotypic value for \( n \) loci as follows

\[
G = \sum_{k=1}^{n} G_{k_1 k_j} = \sum_{k=1}^{n} (\mu_k + y_k) = \sum_{k=1}^{n} \mu_k + \sum_{k=1}^{n} y_k = \mu + \sum_{k=1}^{n} y_k
\]

\[
= \sum_{k=1}^{n} \mu_k + \sum_{k=1}^{n} (\alpha_{k_1}^m + \alpha_{k_j}^f + \delta_{k_1 k_j})
\]
\[ = \mu + (\alpha_{i1}^m + \alpha_{j1}^f + \delta_{11}^11) + (\alpha_{i2}^m + \alpha_{2j}^f + \delta_{2i}^22) \\
+ \ldots + (\alpha_{i1}^m + \alpha_{nj}^f + \delta_{n1}^n n_j) \]  
(8.92)

where \( G_{1121 \ldots k_1 k_2 \ldots n_1 n_j} \) is the genotypic value for genotype \( A_{11} A_{1j} A_{2i} A_{2j} \ldots A_{ki} A_{kj} \ldots A_{n1} A_{nj} \),

- \( A_{1i} \) = ith allele at locus 1 from m population, \( i = 1, 2, \ldots, m_1 \),
- \( A_{1j} \) = jth allele at locus 1 from f population, \( j = 1, 2, \ldots, m_1 \),

- \( \alpha_{i1}^m \) = additive effect of allele \( A_{1i} \) from m population, \( i = 1, \ldots, m_1 \),
- \( \alpha_{lj}^f \) = additive effect of allele \( A_{lj} \) from f population, \( j = 1, \ldots, m_1 \),
- \( \delta_{11}^11 \) = dominance effect or interaction between allele \( A_{11} \) and \( A_{1j} \),

Moving the summation sign ahead of each term in (8.92), we have the gametic form of the model

\[ G_{\prod k_i k_j}^{n} = \mu + \sum_{k=1}^{n} \alpha_{k_i}^m + \sum_{k=1}^{n} \alpha_{k_j}^f + \sum_{k=1}^{n} \delta_{k_i k_j} = \mu + g_{s}^m + g_{c}^f + (gg)_{st}^{mf} \]  
(8.93)

where \( \prod k_i k_j \) = description of a genotype for the genes present at the n loci,

- \( g_{s}^m = \sum_{k=1}^{n} \alpha_{k_i}^m \) = effect of sth male gamete over n loci from m population, \( s = 1, \ldots, M = \prod m_i \) (2.12),
\[ g_t^f = \sum_{k=1}^{n} \alpha_{k}^{f} \] = effect of \( n \) loci from \( f \) population, \( t = 1, \ldots, M = \Pi_{i} m_{i} \),

\[ (g g)_{st}^{m f} = \sum_{k=1}^{n} \delta_{k_{1}k_{j}} \] = interaction between the genes carried by the \( s \)th male gamete and those carried by the \( t \)th female gamete.

(Note: Although the phrase "gametic model" has not been used previously, the one-locus model (8.2), \( G_{ij} = \mu + \alpha_{i}^{m} + \alpha_{j}^{f} + \delta_{ij} \), is a gametic model in its simplest form. The additive effect \( \alpha_{i}^{m} \) is the effect of the male gamete, the additive effect \( \alpha_{j}^{f} \) is the effect of the female gamete, and the dominance effect \( \delta_{ij} \) is the interaction effect between the two gametes.)

To define the effects in (8.92) and their variances, certain assumptions about both the intralocus and interlocus genotypic structures must be made. With respect to the intralocus structure, we make the same assumption of random union of alleles between the \( m \) and \( f \) populations that we did in the above Section 8.1.1. This ensures that the frequency of every genotype at any given locus is equal to the product of the frequencies of the two alleles which the genotype possesses at that locus (8.3).

Next we consider the genotypic values at that locus. We desire to have the same differences between genotypic values, as defined in Table 8.1, so the intralocus additive and dominance variances are those defined for a single locus. To achieve this the value of every genotype at any given locus must be averaged over the same genotypic background, i.e., the genotypic combinations at all other loci must be the same for every genotype at the locus being considered. To define the genotypic values in that manner, the interlocus genotypic structure of the population must be considered. What this implies theoretically is that each of the two parental populations themselves must be in pairwise linkage equilibrium, i.e., every allele at a locus in each population is randomly associated frequency-
wise with the same set of genes at all other loci (see (2.105) to (2.107), Table 2.3, Sections 3.4 to 3.7). This is achieved by having linkage equilibria between genes at every pair of loci (3.60). Three-locus linkage disequilibria (3.112), for example, may exist in the parental populations, but all pair-wise linkage equilibria involving the three loci -- specifically the one of interest with each of the other two loci -- can and must exist (3.118) (3.129). It makes no difference how the genes at those \( n - 1 \) other loci (other than by pairs) are associated with respect to each other, or how pairs of genes at any two of those other loci are associated with the alleles at the locus under consideration. The linkage equilibria at all pairs of loci and random mating between the two parental populations ensure that the covariances between any two terms at different loci in (8.92) are uncorrelated (Box 8.8). In summary, random mating ensures that the variance at each locus is the sum of the variances of the three terms for that locus as given in (8.22) (Box 8.3), and pairwise linkage equilibrium in both parental populations (3.60) and random mating between the two parental populations ensure that the total genotypic variance for \( n \) loci is equal to the sum of the genotypic variances at each locus, because all interlocus effects are uncorrelated. The loci may or may not be linked. The total genotypic variance is simply the sum of the variances of all terms in the model, because every pair of terms is uncorrelated. This is another application of the variance of a linear function of uncorrelated random variables (see (8.22)). It might be pointed out that we are not concerned about the linkage disequilibrium of the gametic output of the cross population itself (3.102). Thus, substituting (8.22)

\[
\sigma_G^2 = \sum_{k=1}^{n} \sigma_{G_k}^2 = \sum_{k=1}^{n} \left( \sigma_{\alpha_k}^2 + \sigma_{\delta_k}^2 \right) + 2 \sum_{k<k'} \Sigma \Sigma \text{Cov}(G_k, G_{k'})
\]

\[
\sigma_{G'}^2 = \sum_{k=1}^{n} \sigma_{\alpha_k}^2 + \sum_{k=1}^{n} \sigma_{\delta_k}^2 + 0
\]
\[ \sigma^2_m + \sigma^2_f + \sigma^2_\delta. \]
\[ = \sigma^2_{Am} + \sigma^2_{Af} + \sigma^2_{Dmf} \]  
(8.94)

where \( \sigma^2_G = \sigma^2_G \) = total genotypic variance for \( n \) loci,
\( \sigma^2_{Am} = \sigma^2_{Af} \) = total haplotypic (that due to gametes) or additive variance for \( n \) loci from the \( m \) population,
\( \sigma^2_{Af} \) = total haplotypic or additive variance for \( n \) loci from the \( f \) population,
\( \sigma^2_\delta \) = total dominance variance for \( n \) loci in the crossed population.

Box 8.8

Prove that all covariances or expectations between effects
at any two loci in the \( n \)-locus model (8.92) (8.93) are zero (see (8.94)).

In a manner similar to Box 8.3, we consider the possible covariances between loci in the model (8.92) (8.93) by taking any pair of loci, say 1 and 2, and evaluating the expectation of each of the nine possible terms from the following product

\[ \text{Cov}(G_1, G_2) = \text{E}(\alpha^m_{11} + \alpha^f_{1j} + \delta^m_{1j})(\alpha^m_{2i} + \alpha^f_{2j} + \delta^m_{2j}) \]  
(1)

There are four \( \alpha \) by \( \alpha \) terms, four \( \alpha \) by \( \delta \) terms, and one \( \delta \) by \( \delta \) term (see (8.95)). We consider one in each of the three categories.

First, consider the covariance between \( \alpha^m_{11} \) and \( \alpha^m_{21} \). By definition (2.95) (2.96), we have

\[ \text{Cov}(\alpha^m_{11}, \alpha^m_{21}) = \text{E}
\[ \left( \alpha^m_{11}, \alpha^m_{21} \right) = \sum \sum p_{1121} \left( \alpha^m_{11} \right)(\alpha^m_{21}) = \sum \sum p_{1121} \left( \alpha^m_{11} \right)(\alpha^m_{21}) \]

where \( p_{1121}, p_{1121} \) = joint frequency of the \( \alpha^m_{11} \) and \( \alpha^m_{21} \) effects occurring together.
Only when linkage equilibrium exists can one substitute the product of the corresponding allelic frequencies for that joint frequency, whereby a zero-covariance is obtained, namely,

\[
\text{Cov}(\alpha_{11}^m, \alpha_{21}^m) = \sum_i \sum_i' p_{1i} p_{2i}' \alpha_{11}^m \alpha_{21}^m = (\sum_i p_{1i} \alpha_{1i}^m)(\sum_i p_{2i} \alpha_{2i}^m) = 0(0) \quad (\text{sub. (8.7) and (8.8)})
\]

Note that if one considered another covariance, \(\text{Cov}(\alpha_{11}^m, \alpha_{21}^f)\), one effect from the \(m\) population and the other from the \(f\) population, the assumption of random mating instead of linkage equilibrium, would be required.

Similarly, the covariance between any \(\alpha\) and \(\delta\) terms is similar to

\[
\text{Cov}(\alpha_{11}^m, \delta_{21}^f) = E \left[ (\alpha_{11}^m)(\delta_{21}^f) \right] = \sum_i \sum_j p_{1i}^m p_{1i}^2 \delta_{1i}^m \delta_{21}^f = \sum_i \sum_j p_{1i}^m p_{1i}^2 \delta_{2i}^m \delta_{21}^f
\]

Making the assumptions of linkage equilibrium and random mating, we write

\[
\text{Cov}(\alpha_{11}^m, \delta_{21}^f) = \sum_i \sum_j p_{1i}^m p_{2i}^f \delta_{1i}^m \delta_{21}^f = (\sum_i p_{1i} \alpha_{1i}^m)(\sum_j p_{2j} \delta_{2j}^f) = 0(0) - 0
\]

Finally, the covariance between the two \(\delta\) terms is

\[
\text{Cov}(\delta_{11}^f, \delta_{21}^f) = E \left[ (\delta_{11}^f)(\delta_{21}^f) \right] = \sum_i \sum_j \sum_{i'} \sum_{j'} p_{1i} p_{1i} p_{2j} \delta_{1i} \delta_{2j} = (\sum_i \sum_j \sum_{i'} \sum_{j'} p_{1i} p_{1i} p_{2j} \delta_{1i} \delta_{2j}) = 0(0) - 0
\]

Again, making the same assumptions of linkage equilibrium and random mating, we write
\[
\text{Cov}(\delta_{11j}, \delta_{21j}) = \sum_{ij} \sum_{i'j'} p_i p_j p_{i'} p_{j'} \left( \delta_{11j} \delta_{21j} \right) \\
= (\sum_{ij} p_i p_j \delta_{11j}) (\sum_{i'j'} p_{i'} p_{j'} \delta_{21j}) \quad \text{(sub. (8.16))} \\
= (0)(0) = 0 \\
\text{(4)}
\]

This proves that all nine covariances in (1) equal zero.

2. **Linkage disequilibrium.** When linkage disequilibria exist, covariances between pairs of loci exist (see (8.94) or Box 8.8). The covariance between any two loci, say, \( k = 1 \) and \( k' = 2 \), consists of nine separate possible covariances, namely,

\[
\text{Cov}(G_k, G_{k'}) = \text{Cov}(G_{111j}, G_{212j}) = E(G_{111j} - \mu)(G_{212j} - \mu) \\
= E(\alpha_{11}^m + \alpha_{1j}^f + \delta_{11j}^m)(\alpha_{21}^m + \alpha_{2j}^f + \delta_{21j}^m) \\
= E(\alpha_{11}^m \alpha_{21}^m) + E(\alpha_{11}^f \alpha_{2j}^m) + E(\alpha_{1j}^f \alpha_{21}^m) + E(\alpha_{1j}^f \alpha_{2j}^f) + E(\alpha_{1j}^f \delta_{21j}^m) + E(\alpha_{1j}^f \delta_{21j}^m) \\
+ E(\delta_{11j}^m \alpha_{21}^m) + E(\delta_{11j}^m \alpha_{2j}^f) + E(\delta_{11j}^m \delta_{21j}^m) \\
\quad \text{(8.95)}
\]

Weir, Cockerham, and Reynolds (1980) have shown that all covariances in (8.95) equal zero, except \( E(\alpha_{11}^m \alpha_{21}^m) \), \( E(\alpha_{11}^f \alpha_{21}^f) \), and \( E(\delta_{11j}^m \delta_{21j}^m) \). The total genotypic variance in this case, expressed in the locus form of the model (8.92), is

\[
\sigma_G^2 = \sum_{k=1}^{n} \sigma_{G_k}^2 + 2 \sum_{k<k'} \text{Cov}(G_k, G_{k'}) \\
\quad \text{(sub. (8.94))} \\
= \sum_{k=1}^{n} \left( \sigma_m^2 \alpha_k^m + \sigma_f^2 \alpha_k^f + \sigma_k^2 \delta_k^m \right) + 2 \sum_{k<k'} \text{Cov}(G_k, G_{k'}) \\
= \sigma_m^2 + \sigma_f^2 + \sigma_{Df}^2 + 2 \sum_{k<k'} \text{Cov}(G_k, G_{k'}) \\
\quad \text{(8.96)}
\]

where \( \text{Cov}(G_k, G_{k'}) = \text{Cov}(\alpha_k^m, \alpha_{k'}^m) + \text{Cov}(\alpha_k^f, \alpha_{k'}^f) + \text{Cov}(\delta_k^m, \delta_{k'}^m) \).
In the gametic form of the model (8.93), the total genotypic variance from (8.96) is

\[
\sigma^2_G = \sum_{k=1}^{n} \sigma^2_{a_k} + 2 \sum \sum \text{Cov}(a^m_k, a^m_{k'}) + \sum_{k=1}^{n} \sigma^2_f + 2 \sum \sum \text{Cov}(a^f_k, a^f_{k'}) + \sum_{k=1}^{n} \sigma^2_{\delta_k} + 2 \sum \sum \text{Cov}(\delta_k, \delta_{k'})
\]

(8.97)

The sum of the \(a^m\) variances and covariances represents the total gametic variance for the \(m\) population in the cross. It is the variance of the \(M = \Pi m_i\) (2.12) marginal means for the \(m\) population. It is no longer composed of only the additive variance for the \(m\) population. It may be biased upward or downward depending upon the net effect of the sum of all covariances, which in turn depend upon the nature of the linkage disequilibrium. Similarly, the \(a^f\) variances and covariances represent the total gametic variance for the \(f\) population.

8.1.7. Two or more loci, multiple alleles, equal male and female allelic frequencies (Hardy-Weinberg), no epistasis. 1. Linkage equilibrium. The development in the previous section has been general in that a cross population was accommodated. In this section we suppose that all allelic frequencies in the two populations are the same, i.e., \(p^m_i = p^f_i\), \(i = 1, \ldots, m\), for every locus, so that Hardy-Weinberg proportions exist. (This section bears the same relation to the previous Section 8.1.6 as Section 8.1.2 does to Section 8.1.1.) Then, as above, the genes at different loci are assumed to be associated independently in the population, i.e., no linkage disequilibrium exists in the parental population [see (2.105) to (2.107), Table 2.3, Sections 3.4 to 3.7] (the loci may or may not be linked). Note that if the population being considered herein is that resulting from random mating of a crossed population itself, the conditions that must exist for the gametic output of the crossed population or for that first generation of
random mating to be in linkage equilibrium are summarized in (3.103) as deduced from (3.102), i.e., both parental populations must have been in linkage equilibrium (or, if not, the sum of the linkage disequilibrium values in the two populations for all gametes must equal zero) and at least one of the following conditions must exist: (1) independence between loci \( \rho_0 = \rho_1 = 1/2 \), or (2) equal allelic frequencies in the two parental populations for either one of the two loci. This latter condition of equal male and female allelic frequencies clearly exists for every locus in the gametic output of the crossed population itself (see (3.7) ff and (8.151) ff), so linkage equilibrium would exist, if the two parental populations were both in linkage equilibrium. In any case, if linkage equilibrium exists, the total genotypic variance is the sum over loci of the genotypic variances at the individual loci. Thus, substituting (8.22) (8.29)

\[
\sigma_G^2 = \sum_{k=1}^{n} \sigma_{G_k}^2 = \sum_{k=1}^{n} \left( \sigma_{m_k}^2 + \sigma_{f_k}^2 + \sigma_{h_k}^2 \right) \\
= \sum_{k=1}^{n} 2\sigma_{a_k}^2 + \sum_{k=1}^{n} \sigma_{h_k}^2 \\
= \sum_{k=1}^{n} \sigma_{A_k}^2 + \sum_{k=1}^{n} \sigma_{D_k}^2 \\
= \sigma_A^2 + \sigma_D^2
\]

(8.98)

where

\[
\sigma_A^2 = \sum_{k=1}^{n} \sigma_{A_k}^2 \\
\sigma_D^2 = \sum_{k=1}^{n} \sigma_{D_k}^2
\]

The additive variances at individual loci are summed together to obtain the total additive variance \( \sigma_A^2 \), and likewise for the total dominance variance, \( \sigma_D^2 \). (This is simply another application of the theorem for the variance of linear functions of uncorrelated random variables.)
2. **Linkage disequilibrium.** If linkage disequilibrium exists, the same expressions as those in (8.95), (8.96), and (8.97) exist, except we no longer need the m and f superscripts, except possibly for convenience, to distinguish between the effects for the m (male) and f (female) parental populations. From (8.96) the genotypic variance, expressed in the locus form of the model, becomes

\[
\sigma^2_G = \sum_{k=1}^{n} \sigma^2_{G_k} + 2\sum_{k}^{n} \sum_{k' < k} \text{Cov}(G_k, G_{k'}) \quad (\text{sub. (8.29))}
\]

\[
= \sum_{k=1}^{n} (2\sigma^2_D + \sigma^2_\delta) + 2\sum_{k}^{n} \sum_{k' < k} \text{Cov}(G_k, G_{k'})
\]

\[
= \sigma^2_A + \sigma^2_D + 2\sum_{k < k'} \text{Cov}(G_k, G_{k'}) \quad (8.99)
\]

where \(\text{Cov}(G_k, G_{k'}) = 2\text{Cov}(\alpha_k, \alpha_{k'}) + \text{Cov}(\delta_k, \delta_{k'}).\)

From (8.97), the total genotypic variance for the gametic form of the model becomes

\[
\sigma^2_G = 2 \left[ \sum_{k=1}^{n} \sigma^2_{\alpha_k} + 2\sum_{k}^{n} \sum_{k' < k} \text{Cov}(\alpha_k, \alpha_{k'}) \right] + \sum_{k=1}^{n} \sigma^2_{\delta_k} + 2\sum_{k}^{n} \sum_{k' < k} \text{Cov}(\delta_k, \delta_{k'})
\]

\[
= \sigma^2_A + 4\sum_{k < k'} \text{Cov}(\alpha_k, \alpha_{k'}) + \sigma^2_D + 2\sum_{k < k'} \text{Cov}(\delta_k, \delta_{k'}) \quad (8.100)
\]

(For what this covariance is equal to in terms of the linkage disequilibrium with only two alleles present, see (8.173).)

8.2. **Relating formula herein to those presented by other authors.**

Assumptions:

8.2.1. **Scaling equivalences.** In the above development, uncoded genotypic values denoted by \(G\)'s have been generally used; this particularly follows
Cockerham (he used the symbol Y many years ago). However, many authors have used coded values (8.31) of various kinds. Any set of genotypic values or parameters, one for each different genotype, can always be reparameterized to be one less than the number of genotypes. For two alleles at a locus we can define each genotypic value about a mid-homozygote value which is the mean of the two homozygotes. In this case the number of parameters can be reduced by one in that the genotype for one homozygote is the negative of the other homozygote. Another example is Wright's scale where one genotypic value is assigned the constant value one. One needs to be able to transfer readily from the values or scale of one author to that of another in order to compare results. To be considered are the a and d notation of Fisher (1918) (used by Falconer (1989)), the d and h notation of Mather and Jinks (1982), the u and a notation of Comstock and Robinson (1952), and the s and h notation of Wright (1931). While Wright's notation is for relative survival values, it may be compared to the others in terms of dominance interpretations. Care must be exercised herein to avoid confusing the a of Comstock and Robinson and that of Fisher and Falconer, and likewise the h of Mather and that of Wright. These scales of genotypic measure are summarized below:

<table>
<thead>
<tr>
<th>Scale value</th>
<th>Parameterization of genotype</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>G_{11} G_{12} G_{22}</td>
<td>Cockerham (1954) (G = Y) (8.101)</td>
</tr>
<tr>
<td>A</td>
<td>a d -a</td>
<td>Fisher (1918), Falconer (1989)</td>
</tr>
<tr>
<td>D</td>
<td>d h -d</td>
<td>Mather and Jinks (1982)</td>
</tr>
<tr>
<td>U</td>
<td>u au -u</td>
<td>Comstock and Robinson (1952)</td>
</tr>
<tr>
<td>W</td>
<td>l 1 - hs 1 - s</td>
<td>Wright (1931)</td>
</tr>
</tbody>
</table>
1. To convert any equation in the A, D, U, or W scale into that of the G scale. To do so we need to express the genotypic values of those scales in terms of the G scale.

   a. General expression relating the genotypic values of all scales except the W scale to the G scale. The general expression which relates the genotypic values used by Fisher (Falconer) (A scale), Mather (D scale), and Comstock and Robinson (U scale) to Cockerham's genotypic values (G scale) is

   \[ A_{ij} = D_{ij} = U_{ij} = G_{ij} - \frac{G_{11} + G_{22}}{2} \]  \hspace{1cm} (8.102)

   Hence, to find the equivalence between either a, d, or u and \( G_{11} \) for the \( A_1A_1 \) genotype, we substitute \( G_{11} \) for \( G_{ij} \) in (8.102) to obtain

   \[ a = d = u = G_{11} - \frac{G_{11} + G_{22}}{2} = \frac{2G_{11} - G_{11} - G_{22}}{2} = \frac{G_{11} - G_{22}}{2} \]  \hspace{1cm} (8.103)

   In like manner, for the \( A_2A_2 \) genotype we obtain \(-a\), \(-d\), and \(-u\), which is the negative of (8.103), so

   \[ -a = -d = -u = -\frac{G_{11} - G_{22}}{2} = \frac{G_{22} - G_{11}}{2} \]  \hspace{1cm} (8.104)

   Similarly, by substituting \( G_{12} \) in (8.102), we find the corresponding expression for \( d, h, \) and \( au \) to be

   \[ d = h = au = G_{12} - \frac{G_{11} + G_{22}}{2} = \frac{2G_{12} - G_{11} - G_{22}}{2} = \frac{G_{11} - 2G_{12} + 2G_{22}}{2} \]  \hspace{1cm} (8.105)

   This is one-half of the positive form (or the negative of the negative form) of the dominance contrast (8.31).

   The degree of dominance \( a \) in the notation of Comstock and Robinson may be obtained by substituting (8.105) and (8.103) in the expression below

   \[ a = \frac{d}{d} = \frac{h}{u} = \frac{2G_{12} - G_{11} - G_{22}}{2G_{11} - G_{22}} = \frac{2G_{12} - G_{11} - G_{22}}{G_{11} - G_{22}} \]  \hspace{1cm} (8.106)
b. General expression relating parameters of the $W$ scale to the $G$ and $U$ scales. The general expression which relates the notation used by Wright to that commonly used herein is

$$W_{ij} = G_{ij}/G_{11}$$  (8.107)

All genotypes are expressed as a ratio to $G_{11}$ in the denominator. Hence, for the $A_1A_1$ genotype

$$W_{11} = 1 - \frac{G_{11}}{G_{11}}$$  (8.108)

for the $A_2A_2$ genotype

$$W_{22} = 1 - s = \frac{G_{22}}{G_{11}}$$  (i.e., $W_{22}$ equals a fraction or proportion of $W_{11}$ which equals one (8.108))

and for the $A_1A_2$ genotype

$$W_{12} = 1 - hs = \frac{G_{12}}{G_{11}}$$  (8.110)

For the selection coefficient, $s$, we first express it in terms of $W_{ij}$'s which is shown to be (use (8.108) (8.109))

$$s = 1 - (1 - s) = W_{11} - W_{22}$$  (8.111)

and then by substituting (8.107) (or (8.108) and (8.109)) in (8.111), we have

$$s = \frac{G_{11}}{G_{11}} - \frac{G_{22}}{G_{11}} = \frac{G_{11} - G_{22}}{G_{11}}$$  (8.112)

The measure of degree of dominance $h$ in the Wright scale is expressed in the $G$ scale as follows. Using (8.107) we obtain

$$1 - hs = W_{12} = \frac{G_{12}}{G_{11}}$$

$$hs = 1 - \frac{G_{12}}{G_{11}} = \frac{G_{11} - G_{12}}{G_{11}}$$
(sub. (8.112))
\[ h = \frac{G_{11} - G_{12}}{G_{11} - G_{12}} = \frac{G_{11} - G_{12}}{G_{11}} \cdot \frac{1}{\frac{G_{12}}{G_{11}} - \frac{G_{22}}{G_{11}}} = \frac{G_{11} - G_{12}}{G_{11} - G_{22}} \]  
(8.113)

Wright's measure of degree of dominance \( h \) as a function of Comstock and Robinson's measure \( a \) is (see Box 8.9)

\[ h = \frac{1 - a}{2} \]  
(8.114)

**Box 8.9**

Derivation of (8.114)

We manipulate (8.113) to incorporate the measure \( a \) (8.106), obtaining

\[ h = \frac{G_{11} - G_{12}}{G_{11} - G_{22}} = \frac{2(G_{11} - G_{12})}{2(G_{11} - G_{22})} = \frac{2G_{11} - 2G_{12}}{2(G_{11} - G_{22})} = \frac{2G_{11} - G_{22} + G_{22} - 2G_{12}}{2(G_{11} - G_{22})} \]  
(sub and add

\[ h = \frac{G_{11} - G_{22} - 2G_{12} + G_{11} + G_{22}}{2(G_{11} - G_{22})} = \frac{1}{2} \cdot \frac{2G_{12} - G_{11} - G_{22}}{2(G_{11} - G_{22})} = \frac{1}{2} \cdot \frac{1}{2} a = \frac{1 - a}{2} \]

which is (8.114).

c. **General expression relating means of all scales to the \( G \) scale.** The mean of any scale is related to the mean of the \( G \) scale in the same way as any particular genotype is related to that of the \( G \) scale. Hence, for the \( A, D, \) and \( U \) scales, the mean of any of those scales is related to the mean of the \( G \) scale (\( \bar{G} = G_{..} = \mu \)) by (8.102), so

\[ \bar{A} = \bar{D} = \bar{U} = \bar{G} - \frac{G_{11} + G_{22}}{2} \]  
(8.115)

Similarly for the \( W \) scale, we have from (8.107)

\[ \bar{W} = \frac{\bar{G}}{G_{11}} \]  
(8.116)
Summary of equivalences, which are used to express formulas given by others into the notation used by Cockerham and that used herein, is presented below.

<table>
<thead>
<tr>
<th>Scale</th>
<th>$A_1A_1$</th>
<th>$A_1A_2$</th>
<th>$A_2A_2$</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>$a = \frac{G_{11} - G_{22}}{2}$</td>
<td>$d = \frac{2G_{12} - G_{11} - G_{22}}{2}$</td>
<td>$-a$</td>
<td>$\bar{A}$</td>
</tr>
<tr>
<td>D</td>
<td>$d = \frac{G_{11} - G_{22}}{2}$</td>
<td>$-d$</td>
<td>$\bar{D}$</td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>$u = \frac{G_{11} - G_{22}}{2}$</td>
<td>$-u$</td>
<td>$\bar{U}$</td>
<td></td>
</tr>
</tbody>
</table>

\[
a = \frac{2G_{12} - G_{11} - G_{22}}{G_{11} - G_{22}}
\]

\[
W = \frac{G_{11}}{G_{11}} (8.108)
\]

\[
1 - hs = \frac{G_{12}}{G_{11}} (8.110)
\]

\[
1 - s = \frac{G_{22}}{G_{11}} (8.109)
\]

\[
\bar{W} = \frac{G}{G_{11}} (8.116)
\]

\[
s = \frac{G_{11} - G_{22}}{G_{11}} (8.112)
\]

\[
h = \frac{G_{11} - G_{12}}{G_{11} - G_{22}} (8.113)
\]

2. To convert any equation in the $G$ scale into that of another scale. If one desires to reparameterize or reexpress any of the results contained in these notes into other scales or parameters, one may solve the general expression (8.102) or (8.107) for $G_{ij}$, and then substitute the result for $G_{ij}$ in any equation given in these notes. Such expressions for either the Fisher (Falconer), Mather, or Comstock and Robinson scale are

\[
G_{ij} = \begin{cases} 
A_{ij} + \frac{G_{11} + G_{22}}{2} & \text{(Fisher; Falconer)} \\
D_{ij} + \frac{G_{11} + G_{22}}{2} & \text{(Mather; Mather and Jinks)} \\
U_{ij} + \frac{G_{11} + G_{22}}{2} & \text{(Comstock and Robinson)}
\end{cases}
\]  

and that for the Wright scale is
\[ G_{ij} = W_{ij} G_{11} \]  

(8.119)

In the literature other symbols denoting genotypic values are occasionally used. In most cases they can be equated to one of the scales presented herein. For example, Kempthorne (1969, Section 15.1 and elsewhere) uses \( d, h, \) and \( r \) as uncoded, genotypic values, and \( i, j, \) and \( k \) as their deviations from the mean. In either case, they may be equated to the \( G \) scale as follows

\[
\begin{align*}
G_{11} &= d \\
G_{12} &= h \\
G_{22} &= r
\end{align*}
\]

Thus, in summary, any results for one scale can be translated into that of another scale by the substitution of the proper equivalences.

**Example 8.3.** We desire to reparameterize the additive variance (8.51) into parameters used by Fisher (or Falconer) (also given in the same equation (8.51) where it was obtained in a roundabout way). Thus, substituting (8.118), \( p_1 = p, \) and \( p_2 = q \) in (8.51)

\[
\begin{align*}
\sigma^2_A &= 2 \sigma^2_x = 2 p_1 p_2 [p_1 (G_{11} - G_{12}) + p_2 (G_{12} - G_{22})]^2 \\
&= 2pq \left[ p \left[ a + \frac{G_{11} + G_{22}}{2} \right] - \left[ d + \frac{G_{11} + G_{22}}{2} \right] \right]^2 + q \left[ d + \frac{G_{11} + G_{22}}{2} - \left[ -a + \frac{G_{11} + G_{22}}{2} \right] \right]^2 \\
&= 2pq[p(a - d) + q(d + a)]^2 \\
&= 2pq(pa - pd + qd + qa)^2 \\
&= 2pq[a + (q - p)d]^2 \quad \text{(sub. (8.45))} \\
&= 2pqa^2
\end{align*}
\]

which agrees with that also given in (8.51) and by Falconer (1989, equation 8.3b).

**Example 8.4.** To reparameterize the same additive variance (8.51) into \( a \) and \( u \) notation (Comstock and Robinson), we proceed as follows (\( p_1 = p, p_2 = 1 - p \))
\[ \sigma_A^2 = 2\sigma_\alpha^2 = 2p_1p_2[p_1(G_{11} - G_{12}) + p_2(G_{12} - G_{22})]^2 \quad \text{(sub. (8.118))} \]

\[ = 2p(1 - p)[p\left[u + \frac{G_{11} + G_{22}}{2} - \left[u + \frac{G_{11} + G_{22}}{2}\right]\right] + (1 - p)[au + \frac{G_{11} + G_{22}}{2} - \left[-u + \frac{G_{11} + G_{22}}{2}\right]]]^2 \]

\[ = 2p(1 - p)[p(u - au) + (1 - p)(au + u)]^2 \]

\[ = 2p(1 - p)[p(1 - a) + (1 - p)(a + 1)]^2u^2 \]

\[ = 2p(1 - p)[p - pa + a + 1 - pa - p]^2u^2 \]

\[ = 2p(1 - p)[1 - pa + a - pa]^2u^2 \]

\[ = 2p(1 - p)[1 - (p - 1 + p)a]^2u^2 \]

\[ = 2p(1 - p)[1 - (2p - 1)a]^2u^2 \text{ or } 2p(1 - p)[1 + (1 - 2p)a]^2u^2 \]

which is the expression given by Comstock and Robinson (1948, equation (3)).

** * * * * * * *

How the value of the heterozygote relates to that of the homozygotes describes the degree of dominance. For the various scales the descriptions are:

<table>
<thead>
<tr>
<th>Positive over-</th>
<th>Complete dominance</th>
<th>Partial dominance</th>
<th>No dominance</th>
<th>Partial dominance*</th>
<th>Complete dominance*</th>
<th>Negative over-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale dominance</td>
<td>G_{12} &gt; G_{11} &gt; G_{12} &gt; \frac{(G_{11} + G_{22})}{2} &gt; G_{12} &gt; G_{22} &gt; G_{12}</td>
<td>Partial dominance</td>
<td>No dominance</td>
<td>Partial dominance*</td>
<td>Complete dominance*</td>
<td>Negative over-</td>
</tr>
<tr>
<td>G</td>
<td>G_{12} &gt; G_{11} &gt; G_{12} &gt; \frac{(G_{11} + G_{22})}{2} &gt; G_{12} &gt; G_{22} &gt; G_{12}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>d &gt; a &gt; d &gt; 0 &gt; d &gt; -a &gt; d (8.121)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>h &gt; d &gt; h &gt; 0 &gt; h &gt; -d &gt; h</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>a &gt; 1 &gt; a &gt; 0 &gt; a &gt; -1 &gt; a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W</td>
<td>h &lt; 0 &lt; h &lt; \frac{1}{2} &lt; h &lt; 1 &lt; h</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* This is dominance for the allele that decreases the value.
** This is overdominance in the decreasing direction.
The genotypic values of the heterozygote, $G_{12}$, d, and h (Mather), or the measures of the degree of dominance, $a$ and h (Wright), for complete dominance, no dominance, and complete dominance of the allele in the decreasing direction are given in the correspondingly designated columns. For all other situations the possible ranges of the heterozygote or measures of the degree of dominance are indicated. Overdominance in the decreasing direction is sometimes called underdominance, meaning that the heterozygote is less than both homozygotes, but the term negative overdominance is preferred.

Dominance for more than two alleles is not simple to describe. For example, in Comstock and Robinson's notation, there would be an $a_{ij}$ corresponding to every ijth heterozygote for the measure of degree of dominance relative to the i and jth homozygotes, and each allele would have multiple dominance statements, $a_{12}$, $a_{13}$, and so on. If all a's are zero, there is no dominance whatsoever.

Although s and h (Wright) could be used as a metric scale, its primary use has been in terms of relative survival values, and was included with the others to compare the descriptions of dominance.

8.2.2. Separation of additive variance into Mather's genetic parameters.

The Mather school (Hayman, 1960; Mather and Jinks, 1971, p. 251; Mather and Jinks, 1977, p. 50) has defined various quadratic (or higher) genetic parameters that compose the additive variance. To relate these to the additive variance, consider an expansion of the additive variance (8.51) in the u and a notation (see Example 8.4) into its four parts, namely,

$$2\sigma^2 = 2p(1 - p)[1 - (2p - 1)a]^2u^2$$

$$= 2p(1 - p)[1 - 2(2p - 1)a + (2p - 1)^2a^2]u^2$$

$$= 2p(1 - p)[1 - 2(2p - 1)a + (4p^2 - 4p + 1)a^2]u^2$$
\[ \sigma_A^2 = \sum_{i=1}^{n} \left[ \Sigma p_i(1 - p_i)u_i^2 - 4 \Sigma p_i(1 - p_i)(2p_i - 1)u_i^2 + 2p(1 - p)4p(p - 1)a^2u^2 \right] \]

and summing (8.122) over all loci, which are assumed to be in linkage equilibrium, we obtain (Hayman, 1960)

\[ \sigma_A^2 = \sum_{i=1}^{n} \left[ 2 \Sigma p_i(1 - p_i)u_i^2 - 4 \Sigma p_i(1 - p_i)(2p_i - 1)a_iu_i^2 + 2 \Sigma p_i(1 - p_i)a_i^2u_i^2 - 8 \Sigma p_i(1 - p_i)^2a_i^2u_i^2 \right] \]

\[ = D - F + \frac{H_1}{2} - \frac{H_2}{2} = \frac{1}{2} (D - F + H_1 - H_2) \]

(8.123)

where

\[ D = 4 \Sigma p_i(1 - p_i)u_i^2, \]

\[ F = 8 \Sigma p_i(1 - p_i)(2p_i - 1)a_iu_i^2, \]

\[ H_1 = 4 \Sigma p_i(1 - p_i)a_i^2u_i^2, \]

\[ H_2 = 16 \Sigma p_i(1 - p_i)^2a_i^2u_i^2. \]

Equation (8.123) relates particularly to (8.53) in that the first term in (8.123) equals the square of the first term in the binomial in (8.53), after summing over loci, i.e.,

\[ 2 \Sigma p_i(1 - p_i)u_i^2 = 2 \Sigma p_ip_iG_{i11} - G_{i22} \]

(sub. (8.103))

\[ = 2 \Sigma p_i(1 - p_i)u_i^2 \]

(8.124)

When \( p_1 = p_2 = 1/2 \), the second term in (8.123) equals zero, because \( (2p_i - 1) = 0 \), and the last two terms cancel for every i-th locus, i.e.,
\[ 2p_I(1 - p_I)a_{i1}^2 + 8p_I^2(1 - p_I)a_{i1}^2 = 2 \left( \frac{1}{2} \right) a_{i1}^2 8 \left( \frac{1}{2} \right) a_{i1}^2 = \frac{1}{2} a_{i1}^2 + \frac{1}{2} a_{i1}^2 = 0 \] (8.125)

so (8.123) becomes (compare (8.72))

\[ \sigma_A^2 = D = \frac{1}{2} \sum_{i=1}^{n} u_{i1}^2 \] (8.126)

In this case of equal allelic frequencies of one-half at all loci, twice the additive variance equals the component D, which is then equal to \( \sum u_{i1}^2 \) -- simply the sum of the squares of all deviations of the homozygotes from the mid-homozygote value.

In the Mather school, interest centers on these four genetic components, D, F, \( H_1 \), and \( H_2 \) (8.123). Component D is a function of \( u^2 \), the square of the amount either homozygote deviates from the mid-homozygote value (8.103). Each square is weighted by the allelic frequencies at that locus. Loci with intermediate allelic frequencies would contribute the most, other things being equal. \( H_1 \) and \( H_2 \) are functions of \( (au)^2 \), the square of the amount the heterozygote deviates from the mid-homozygote value -- again each is weighted differently by the allelic frequencies at that locus. Since \( H_1 \) has the same coefficient of 4 as D, the ratio \( \sqrt{H_1/D} = \bar{a} \) is a measure of the average degree of dominance (weighted by \( p_I(1 - p_I) \)). Also the ratio \( H_2/4H_1 = p(1 - p) \) measures the average value of \( p(1 - p) \) over all loci. In the case of allelic frequencies of one-half at all loci (\( p_1 = p_2 = 1/2 \)), component D which is equal to \( \sum u_{i1}^2 \) -- simply the sum of the squares of all deviations of the homozygotes from the mid-homozygote value. Likewise, \( H_1 = H_2 = \Sigma(a_{i1}u_{i1})^2 \) -- simply the sum of the squares of all deviations of the heterozygotes from the mid-homozygote value, and F = 0 (see Mather and Jinks, 1982, p. 257).
8.3. Gene effects and variances with epistasis. Assumptions:

8.3.1. Two loci, multiple alleles, arbitrary "male" and "female" allelic frequencies (cross between two populations), epistasis: General. 1. Linkage equilibrium. Previously the genotypic variance for any number of loci was described under the assumptions of random mating (either a cross obtained by random mating between two populations in Section 8.1.6 or a population in Hardy-Weinberg equilibrium in Section 8.1.7), linkage equilibrium, and no interaction between loci or no epistasis, i.e., additivity between loci. We will now relax the assumption of no interaction or no epistasis between loci.

Interaction between genes at different loci is accommodated by utilizing a complete factorial model (Anderson and Kempthorne, 1954; Cockerham, 1954; Kempthorne, 1954, 1955, 1969, Section 19.3; Bulmer, 1980, Chapter 4; Turner and Young, 1969, Section 3.2). Its utilization is natural, because the idea of factorial experiments in statistics, which is due to Fisher, arose from his knowledge of the simultaneous inheritance of Mendelian factors (see Bulmer, 1980, p. 46). To examine interlocus interactions or epistasis, we initially consider, in some detail, only two loci, after which its extension to n loci will be clear (see Section 8.3.5). In the absence of interaction between loci, the model (8.92) for any two loci, say 1 and 2, redesignated a and b (or later as A and B, depending upon the suitability of the context), is

\[
G_{i1j2k2} = G_{aiajbkbl} = \mu + y_{ai}y_{aj} + y_{bkl}y_{bkl} = y_{ai}y_{aj} + y_{bkl}y_{bkl} - u + (\alpha_{ai}^m + \alpha_{aj}^f + \delta_{aiaj}^m) + (\alpha_{bk}^m + \alpha_{bkl}^f + \delta_{bklbkl}^m)
\]  

(8.127)

This model (8.127) presently has four main-effect terms, \(\alpha_{ai}^m, \alpha_{aj}^f, \alpha_{bk}^m\), and \(\alpha_{bkl}^f\), one for each gene, \(A_i, A_j, B_k\), and \(B_l\), that the genotype possesses and only two interaction terms, \(\delta_{aiaj}^m\) and \(\delta_{bklbkl}^m\) (see (8.2) for definition of dominance effect). All possible interaction effects between loci may be written down by considering
all combinations of each of the three genic effects at locus a with each of the
three genic effects at locus b. Thus, the complete factorial model of genic
effects for two loci is

\[
G_{a_i a_j b_k b_\ell} = \mu + \left[ \alpha_{a_i}^m + \alpha_{a_j}^m + \delta_{a_i a_j}^m \right] + \left[ \alpha_{b_k}^f + \alpha_{b_\ell}^f + \delta_{b_k b_\ell}^f \right]
\]

\[
+ \left[ (\alpha \alpha)^m m f_{a_i b_k} + (\alpha \alpha)^m f m_{a_i b_\ell} + (\alpha \alpha)^f m m_{b_k b_\ell} \right]
\]

\[
+ (\alpha \delta)^m m f_{a_i b_k b_\ell} + (\alpha \delta)^f m f_{a_i b_k b_\ell} + (\delta \alpha)^m m f_{a_i a_j b_k} + (\delta \alpha)^m f f_{a_i a_j b_\ell} + (\delta \delta)^m f f_{a_i a_j b_k b_\ell}
\]

where \( G_{a_i a_j b_k b_\ell} \) = genotypic value for individual genotype \( A_i A_j B_k B_\ell \); \( i, j = 1, \ldots, m_a; k, \ell = 1, \ldots, m_b; \)

\( \mu \) = overall mean,

\( \alpha_{a_i}^m \) = additive or average effect of \( i \)th allele \( (A_i) \) at locus a
received from male parent (or more generally the \( m \) parental population (8.2)),

\( \alpha_{a_j}^f \) = additive or average effect of \( j \)th allele \( (A_j) \) at locus a
received from female parent (or more generally the \( f \) parental population),

\( \delta_{a_i a_j}^m \) = dominance effect due to interaction between \( i \)th and \( j \)th alleles
\( (A_i \) and \( A_j \) at locus a,

\( \alpha_{b_k}^m \) = additive or average effect of allele \( B_k \) at locus b received from
male parent,

\( \alpha_{b_\ell}^f \) = additive or average effect of allele \( B_\ell \) at locus b received from
female parent,

\( \delta_{b_k b_\ell}^m \) = dominance effect due to interaction between alleles \( B_k \) and \( B_\ell \),

\( (\alpha \alpha)^m m_{a_i b_k} \) = additive-by-additive effect or interlocus interaction effect
between \( i \)th allele \( (A_i) \) at locus a from the male parent and \( k \)th
allele \( (B_k) \) at locus b from the male parent,
\((a\delta)^m_{a_1b_k} = \) additive-by-dominance effect or interlocus interaction effect between ith allele \((A_i)\) at locus a from the male parent and the kth and \(\ell\)th alleles \((B_k\) and \(B_{\ell}\)) at locus b,

\((b\delta)^m_{a_1a_jb_k} = \) dominance-by-dominance effect or interlocus interaction effect between ith and jth alleles \((A_i\) and \(A_j)\) at locus a and the kth and \(\ell\)th alleles \((B_k\) and \(B_{\ell}\)) at locus b.

All other analogous terms have similar meanings to those given. Note, for example, that there are four interlocus additive-by-additive terms in the model. Since an individual carries two genes at each locus, there are four possible pairs, i.e.,

\[
\begin{align*}
\begin{array}{c}
a^m \\
\downarrow \\
af \\
\uparrow \\
b^m \\
\end{array}
\end{align*}
\]

(8.129)

and an additive-by-additive term exists for each such pair. If the individual genotype is homozygous at both loci, then all four terms for the individual will be equal. If one locus is heterozygous and the other homozygous, then two terms will be equal, and the other two terms equal. When both loci are heterozygous, all four terms will generally be different. Cockerham (1954) partitioned the epistatic variance and coined the terms additive-by-additive, additive-by-dominance, dominance-by-additive, and dominance-by-dominance.

This genic factorial model for two loci has 15 terms plus the mean and is analogous to a four-factor experiment in statistics as shown in Table 8.2. The model has four
Table 8.2. Analogy between genic factorial model for two loci and a four-factor experiment.

<table>
<thead>
<tr>
<th>Number of factors involved</th>
<th>Kind of term</th>
<th>Number of terms</th>
<th>Term in genic factorial model</th>
<th>Term in model of factorial experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Main effects</td>
<td>$4 = \binom{4}{1}$</td>
<td>$\alpha_i^m \alpha_j^f$</td>
<td>$A_i$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\alpha_i^m \alpha_k^f$</td>
<td>$B_j$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\delta \alpha_i^m \alpha_j^f$</td>
<td>$C_k$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$(\delta \alpha_i^m \alpha_j^f \alpha_k^f \alpha_l^f)$</td>
<td>$D_{\ell}$</td>
</tr>
<tr>
<td></td>
<td>Two-factor or two-gene interaction effects</td>
<td>$6 = \binom{4}{2}$</td>
<td>{ $(\alpha \alpha_i^m \alpha_j^f \alpha_k^f \alpha_l^f)$, $(\alpha \alpha_i^m \alpha_j^f \alpha_k^f \alpha_l^f)$ }</td>
<td>{ ( (A\alpha_i^m \alpha_j^f \alpha_k^f \alpha_l^f) ), $(AC)_{ik}$ }</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>{ $(\alpha \alpha_i^m \alpha_j^f \alpha_k^f \alpha_l^f)$, $(\alpha \alpha_i^m \alpha_j^f \alpha_k^f \alpha_l^f)$ }</td>
<td>{ $(AD)_{il}$ }</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$(\alpha \alpha_i^m \alpha_j^f \alpha_k^f \alpha_l^f)$</td>
<td>$(BC)_{jk}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$(\alpha \alpha_i^m \alpha_j^f \alpha_k^f \alpha_l^f)$</td>
<td>$(BD)_{jl}$</td>
</tr>
<tr>
<td></td>
<td>Three-factor or three-gene interaction effects</td>
<td>$4 = \binom{4}{3}$</td>
<td>{ $(\alpha \alpha_i^m \alpha_j^f \alpha_k^f \alpha_l^f)$, $(\alpha \alpha_i^m \alpha_j^f \alpha_k^f \alpha_l^f)$ }</td>
<td>{ $(ACD)_{ik\ell}$ }</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$(\alpha \alpha_i^m \alpha_j^f \alpha_k^f \alpha_l^f)$</td>
<td>$(BCD)_{jk\ell}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$(\alpha \alpha_i^m \alpha_j^f \alpha_k^f \alpha_l^f)$</td>
<td>$(ABC)_{ijk}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$(\alpha \alpha_i^m \alpha_j^f \alpha_k^f \alpha_l^f)$</td>
<td>$(ABD)_{ij\ell}$</td>
</tr>
<tr>
<td></td>
<td>Four-factor or four-gene interaction effects</td>
<td>$1 = \binom{4}{4}$</td>
<td>$(\alpha \alpha_i^m \alpha_j^f \alpha_k^f \alpha_l^f)$</td>
<td>$(ABCD)_{ijk\ell}$</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td>15</td>
</tr>
</tbody>
</table>
terms for main effects, one for each gene; six terms for two-gene interactions of which two are intralocus interactions, the dominance effects, and four are interlocus interactions, the additive-by-additive effects; four terms for three-gene interactions, the additive-by-dominance effects; and finally one term for the four-gene interaction, the dominance-by-dominance effects. The additive-by-additive effect is the interaction between two genes, one at each locus. The additive-by-dominance effect is the interaction between three genes, one at one locus and two at another locus. Dominance-by-dominance is the interaction between four genes, two at one locus and two at another. The effects are due to the interaction between the genes and not to the interaction between the effects of the genes, i.e., the additive-by-additive effect is not the interaction between the additive effects of the two genes, as too many writers often carelessly state.

The genic effects in the model (8.128) may be grouped to form a locus

factorial model or a model of locus effects (8.92), i.e.,

$$c_{a_i a_j b_k b_\ell}^m f m f = \mu + y_{a_i a_j}^m f + y_{b_k b_\ell}^m f + (y y)_{a_i a_j b_k b_\ell}^m f m f \quad (8.130)$$

where

$$y_{a_i a_j}^m f = \alpha_{a_i}^m f + \alpha_{a_j}^m f + \delta_{a_i a_j}^m f$$

is average effect of locus a due to alleles $A_i$ and $A_j$,

$$y_{b_k b_\ell}^m f = \alpha_{b_k}^m f + \alpha_{b_\ell}^m f + \delta_{b_k b_\ell}^m f$$

is average effect of locus b due to alleles $B_k$ and $B_\ell$,

$$(y y)_{a_i a_j b_k b_\ell}^m f m f = (\alpha \alpha)_{a_i b_k}^m f m + (\alpha \alpha)_{a_j b_\ell}^m f m + (\alpha \alpha)_{a_i b_k}^f f + (\alpha \alpha)_{a_j b_\ell}^f f$$

+ $$(\alpha \delta)_{a_i b_k}^m f m + (\alpha \delta)_{a_j b_\ell}^m f m + (\delta \alpha)_{a_i a_j b_k}^m f m + (\delta \alpha)_{a_i a_j b_\ell}^m f m + (\delta \delta)_{a_i a_j b_k b_\ell}^m f m f$$

is interlocus interaction effect between genotype $A_i A_j$ at locus a and genotype $B_k B_\ell$ at locus b; residual effect of genotypic value not accounted for by effect $y_{a_i a_j}^m f$ and $y_{b_k b_\ell}^m f$; the total epistatic effect between loci a and b.
Another grouping of genic effects in model (8.128) may be made to form a 
gametic factorial model or a model of gametic effects (8.93), i.e.,

\[ G_{\text{m m f f}}^{a_1a_jb_kb_{\text{L}}} = G_{\text{m m f f}}^{a_1b_kb_jb_{\text{L}}} = \mu + g_{\text{a_i b_k}} + g_{\text{a_j b_L}} + (gg)_{\text{a_i b_k a_j b_L}} \]

(8.131)

where \[ a_i b_k = \alpha_{\text{a_i}} + \alpha_{\text{b_k}} + (\alpha \alpha)_{\text{a_i b_k}} \] = average effect of gamete from male parent with \( i \)th and \( k \)th alleles \((A_i \text{ and } B_k)\) at loci \( a \) and \( b \),

\[ g_{\text{a_j b_L}} = \alpha_{\text{a_j}} + \alpha_{\text{b_L}} + (\alpha \alpha)_{\text{a_j b_L}} \] = average effect of gamete from female parent with \( j \)th and \( \text{L} \)th alleles \((A_j \text{ and } B_{\text{L}})\) at loci \( a \) and \( b \),

\[ (gg)_{\text{a_i b_k a_j b_L}} = \delta_{\text{a_i a_j}} + \delta_{\text{b_k b_L}} + (\alpha \alpha)_{\text{a_i a_j b_k b_L}} + (\alpha \delta)_{\text{a_i b_k a_j b_L}} + (a \delta)_{\text{a_i a_j b_k b_L}} + (\delta \delta)_{\text{a_i a_j b_k b_L}} \]

= interaction effect between gamete \( A_i B_k \) from male parent and gamete \( A_j B_{\text{L}} \) from female parent.

The 15 effects in the model (8.128) have been classified in two ways

(Cockerham, 1966): (1) being part of either the locus \( a \) or \( b \) effect or the interlocus interaction effect (8.130), and (2) being part of either the male or female gametic effect or the intergametic interaction effect (8.131). This is set forth in Figure 8.4.
Figure 8.4 Joint classification of effects in the genic factorial model according to locus and interlocus effects, and according to gametic and intergametic effects.

By columns they are grouped according to each locus effect, $y_{a_i a_j}^m f$ and $y_{b_k b_{k'}}^m f$, and the interaction effect between the loci, $(yy)_{a_i a_j b_k b_{k'}}^m f$. By rows they are grouped according to the effects of uniting gametes, $e_{a_i b_k}^m m$ and $e_{a_j b_{k'}}^m f f$, and the interaction between the gametes, $(gg)_{a_i b_k a_j b_{k'}}^m m f f$. This manner of classifying the effects illustrates the connection between locus effects and gametic effects. All
epistatic effects, \((a\alpha)'s = additive-by-additive, (a\delta)'s = additive-by-dominance, (\delta\delta)'s = dominance-by-dominance,\) represent interactions among nonallelic genes, and thus constitute interactions between loci. Thus, all epistatic effects lie in the last column. Each of the gametic effects contains an additive-by-additive effect, \((a\alpha)\). Interaction among the gametes includes dominance effects and the remaining epistatic effects.

The epistatic effects involving dominance have been written to correspond to the interactions of effects between loci, as is usually done, i.e., written in locus or \(\delta\) notation. However, to correspond to interactions of gametic effects the five kinds of epistatic effects involving dominance could be expressed as follows

\[
\begin{align*}
(a\delta)^{m\ m\ f}_{a_1b_kb_\ell} &= [(a\alpha)a]^{m\ m\ f}_{a_1b_kb_\ell}, \\
(a\delta)^{f\ m\ f}_{a_1b_kb_\ell} &= [a(a\alpha)]^{m\ f\ f}_{b_kbJa_jb_\ell}, \\
(\delta a)^{m\ f\ m}_{a_1a_jb_k} &= [(a\alpha)a]^{m\ m\ f}_{a_1b_kb_j}, \\
(\delta a)^{m\ f\ f}_{a_1a_jb_\ell} &= [a(a\alpha)]^{m\ f\ f}_{a_1a_jb_\ell}, \\
(\delta \delta)^{m\ f\ m\ f}_{a_1a_jb_kb_\ell} &= [(a\alpha)(a\alpha)]^{m\ m\ f\ f}_{a_1b_kbJa_jb_\ell}.
\end{align*}
\] (8.132)

In all terms, \(\delta\) has been rewritten as \(a\alpha\), signifying the interaction between the two alleles at the locus. This is customary in a typical factorial experiment (Table 8.2). Then the \(a'\)s within a gamete are grouped by the use of parentheses.

While the \(\delta\) terminology will generally be used, it is in the context of interactions of gametic effects that differences in male and female allelic frequencies and covariances of relatives with linkages can be accommodated most easily (see Section 9.5).

We now desire to consider the least-squares definitions of the effects in the two-locus epistatic model (8.128). Under random mating and linkage equilibrium, the frequency of every genotypic value \(G_{ijkl}\) can be written as a product of four gene frequencies, two at one locus and two at the other,
\[ p_{ijk\ell} = (p_{a_i}^m f p_{a_j}^m f) (p_{b_k}^f p_{b_\ell}^f) = (p_{a_i}^m p_{b_k}^f) (p_{a_j}^m p_{b_\ell}^f) = (p_{a_i}^m p_{b_k}^f) (p_{a_j}^m p_{b_\ell}^f) \] (8.133)

where \( p_{ijk\ell} \) = frequency of genotypic value \( c_{a_i a_j b_k b_\ell}^m f m f \) in crossed population,

\( p_{a_i}^m \) = frequency of \( i \)th allele at locus \( a \) in \( m \) (male) population,

\( p_{a_j}^f \) = frequency of \( j \)th allele at locus \( a \) in \( f \) (female) population,

\( p_{b_k}^m \) = frequency of \( k \)th allele at locus \( b \) in \( m \) (male) population,

\( p_{b_\ell}^f \) = frequency of \( \ell \)th allele at locus \( b \) in \( f \) (female) population.

Any two gene frequencies at the same locus or at different loci, one of which has an \( m \) superscript and the other an \( f \) superscript, can be written as a product due to random mating. The product between the frequency of a particular male gamete with a particular female gamete can be written as a product because of random mating (see Section 3.2). Any two gene frequencies at different loci with the same superscript, either \( m \) or \( f \), can be written as a product because of linkage equilibrium. For linkage equilibrium to be achieved, the only condition is that both parental populations must be in linkage equilibrium, as was pointed out in Section 8.1.6.

The values for the effects in the model are the usual least-squares values obtained by minimizing \( \sum \sum \sum \sum p_{a_i}^m f p_{a_j}^m f p_{b_k}^f p_{b_\ell}^f [(\delta \delta)^m f m f]_{a_i a_j b_k b_\ell}^2 \). (This is identical in nature to those derived earlier in a simpler context, Box 8.1, (8.9), (8.10), (8.11), and (8.13).) In the expressions, which are to follow for the least-squares values only, the following abbreviated notation will be used for simplicity. We drop the \( m \) and \( f \) superscripts, and the \( a \) and \( b \) subscripts, letting
i imply the male at the a locus, j imply the female at the a locus, k imply the male at the b locus, and l imply the female at the b locus.

\[ G_{ijkl} = \sum_{m} \sum_{f} G_{ijkl} \]

\[ p_{a_i} = p_{i}, \quad p_{a_j} = p_{j}, \quad p_{b_k} = p_{k}, \quad p_{b_l} = p_{l} \]

\[ a = a_i, \quad a_f = a_j, \quad a_{b_k} = \alpha_k, \quad a_{b_l} = \alpha_l \]

\[ \delta_{m} = \delta_{ij}, \quad \delta_{k} = \delta_{kl} \]

\[ (\alpha \alpha)_{ijkl} = (\alpha \alpha)_{ik}, \quad (\alpha \alpha)_{ijkl} = (\alpha \alpha)_{il}, \quad (\alpha \alpha)_{ijkl} = (\alpha \alpha)_{jk}, \quad (\alpha \alpha)_{ijkl} = (\alpha \alpha)_{jl} \]

\[ (\alpha \delta)_{ijkl} = (\alpha \delta)_{ikl}, \quad (\alpha \delta)_{ijkl} = (\alpha \delta)_{jkl}, \quad (\alpha \delta)_{ijkl} = (\alpha \delta)_{ijk}, \quad (\delta \delta)_{ijkl} = (\delta \delta)_{ijk} \]

(This abbreviated notation remains clear as long as the letter subscript is present. Upon substituting numbers for i, j, k and l, e.g., i = j = k = l = 1 in the p's, their distinction would be lost, and the notation would obviously be inadequate, so one would need to resort to the full notation.) In addition, the range of summation is omitted, but it always is

\[ i, j = 1, 2, \ldots, m_a; \quad k, l = 1, 2, \ldots, m_b \]

The following definitions of effects or parameters are identical to that for a factorial experiment with unequal frequencies of cell values (Scheffe, 1959, Section 4.4, particularly "Case of proportional frequencies"; Steel and Torrie, 1980, Section 18.3; Bancroft, 1968, Section 3.4, for three-way proportional subclass frequencies).

The least-squares values for the mean in the genic factorial model (8.128) is

\[ \mu = \sum_{i} \sum_{j} \sum_{k} \sum_{l} p_{i} p_{j} p_{k} p_{l} G_{ijkl} = G \ldots \]
This was obtained by making the restrictions that the weighted mean of every
effect in the genic factorial model (8.128), except the last one, equals zero over
each subscript which the effect possesses (analogous to (8.7) (8.8)), namely,

$$\sum_{i} \alpha_i - \sum_{j} \alpha_j = \sum_{k} \alpha_k = \sum_{\ell} \alpha_{\ell} = 0$$  \hspace{1cm} (8.137)

$$\sum_{j} \delta_{ij} = 0 \text{ for every } i, \quad \sum_{i} \delta_{ij} = 0 \text{ for every } j$$  \hspace{1cm} (8.138a)

$$\sum_{\ell} \delta_{k\ell} = 0 \text{ for every } k, \quad \sum_{k} \delta_{k\ell} = 0 \text{ for every } \ell$$  \hspace{1cm} (8.138b)

$$\sum_{k} (\alpha\alpha)_{ik} = 0 \text{ for every } i, \quad \sum_{i} (\alpha\alpha)_{ik} = 0 \text{ for every } k$$  \hspace{1cm} (8.139a)

$$\sum_{\ell} (\alpha\alpha)_{i\ell} = 0 \text{ for every } i, \quad \sum_{i} (\alpha\alpha)_{i\ell} = 0 \text{ for every } \ell$$  \hspace{1cm} (8.139b)

$$\sum_{k} (\alpha\alpha)_{jk} = 0 \text{ for every } j, \quad \sum_{j} (\alpha\alpha)_{jk} = 0 \text{ for every } k$$  \hspace{1cm} (8.139c)

$$\sum_{\ell} (\alpha\alpha)_{j\ell} = 0 \text{ for every } j, \quad \sum_{j} (\alpha\alpha)_{j\ell} = 0 \text{ for every } \ell$$  \hspace{1cm} (8.139d)

$$\sum_{\ell} (\alpha\delta)_{ik\ell} = 0 \text{ for every combination of } i \text{ and } k$$  \hspace{1cm} (8.140a)

$$\sum_{k} (\alpha\delta)_{ik\ell} = 0 \text{ for every combination of } i \text{ and } \ell$$  \hspace{1cm} (8.140b)

$$\sum_{i} (\alpha\delta)_{ik\ell} = 0 \text{ for every combination of } k \text{ and } \ell$$  \hspace{1cm} (8.140c)

$$\sum_{\ell} (\alpha\delta)_{jk\ell} = 0 \text{ for every combination of } j \text{ and } k$$  \hspace{1cm} (8.140d)

$$\sum_{k} (\alpha\delta)_{jk\ell} = 0 \text{ for every combination of } j \text{ and } \ell$$  \hspace{1cm} (8.140e)

$$\sum_{j} (\alpha\delta)_{jk\ell} = 0 \text{ for every combination of } k \text{ and } \ell$$  \hspace{1cm} (8.140f)

$$\sum_{k} (\delta\alpha)_{ijk} = 0 \text{ for every combination of } i \text{ and } j$$  \hspace{1cm} (8.140g)

$$\sum_{j} (\delta\alpha)_{ijk} = 0 \text{ for every combination of } i \text{ and } k$$  \hspace{1cm} (8.140h)

$$\sum_{i} (\delta\alpha)_{ijk} = 0 \text{ for every combination of } j \text{ and } k$$  \hspace{1cm} (8.140i)

$$\sum_{\ell} (\delta\alpha)_{ij\ell} = 0 \text{ for every combination of } i \text{ and } j$$  \hspace{1cm} (8.140j)

$$\sum_{j} (\delta\alpha)_{ij\ell} = 0 \text{ for every combination of } i \text{ and } \ell$$  \hspace{1cm} (8.140k)

$$\sum_{i} (\delta\alpha)_{ij\ell} = 0 \text{ for every combination of } j \text{ and } \ell$$  \hspace{1cm} (8.140l)

Similar conditions also exist for the last effect in the model as a consequence of
the above, (8.137) to (8.140), namely,
\[ \Sigma p_{k}(\delta \delta)_{ijkl} = 0 \text{ for every combination of } i, j, \text{ and } k \]  
\[ \Sigma p_{i}(\delta \delta)_{ijkl} = 0 \text{ for every combination of } i, j, \text{ and } l \]  
\[ \Sigma p_{j}(\delta \delta)_{ijkl} = 0 \text{ for every combination of } i, k, \text{ and } l \]  
\[ \Sigma p_{l}(\delta \delta)_{ijkl} = 0 \text{ for every combination of } j, k, \text{ and } l \]  

We also note that

\[ \Sigma p_{i} = \Sigma p_{j} = \Sigma p_{k} = \Sigma p_{l} = 1 \]  
\[ \Sigma \Sigma p_{i}p_{j} = \Sigma \Sigma p_{i}p_{k} = \Sigma \Sigma p_{i}p_{l} = \Sigma \Sigma p_{j}p_{k} = \Sigma \Sigma p_{j}p_{l} = \Sigma \Sigma p_{k}p_{l} = 1 \]  
\[ \Sigma \Sigma \Sigma p_{i}p_{j}p_{k} = \Sigma \Sigma \Sigma p_{i}p_{j}p_{l} = \Sigma \Sigma \Sigma p_{i}p_{k}p_{l} = \Sigma \Sigma \Sigma p_{j}p_{k}p_{l} = 1 \]  

The least-squares definitions of the other effects in the model \((8.128)\) are

as follows, and are expressed in terms of \(G's\) by substituting expressions in terms

of \(G's\) for the lower-order effects:

\[ \alpha_{i} = \alpha_{i} = \Sigma \Sigma \Sigma p_{j}p_{k}p_{l} G_{ijkl} - \mu \]  
\[ = G_{i...} - G_{...} \text{ for every } i \]  
\[ \text{(sub. (8.136))} \]  
\[ \alpha_{j} = \Sigma \Sigma \Sigma p_{i}p_{k}p_{l} G_{ijkl} - \mu \]  
\[ = G_{j...} - G_{...} \text{ for every } j \]  
\[ \alpha_{k} = \Sigma \Sigma \Sigma p_{i}p_{j}p_{l} G_{ijkl} - \mu \]  
\[ = G_{...k} - G_{...} \text{ for every } k \]  
\[ \alpha_{l} = \Sigma \Sigma \Sigma p_{i}p_{j}p_{k} G_{ijkl} - \mu \]  
\[ = G_{...l} - G_{...} \text{ for every } l \]  
\[ \delta_{ij} = \delta_{ij} = \Sigma \Sigma p_{k}p_{l} G_{ijkl} - \mu - \alpha_{i} - \alpha_{j} \]  
\[ = G_{ij...} - [G_{...} + (G_{i...} - G_{...}) + (G_{j...} - G_{...})] \]  
\[ = G_{ij...} - G_{i...} - G_{j...} + G_{...} \text{ for every combination of } i \text{ and } j \]  
\[ \delta_{kl} = \delta_{kl} = \Sigma \Sigma p_{i}p_{j} G_{ijkl} - \mu - \alpha_{k} - \alpha_{l} \]  
\[ = G_{...kl} - G_{...k} - G_{...l} + G_{...} \text{ for every combination of } k \text{ and } l \]
\[(\alpha\alpha)_{ik} = \sum_j \sum_l p_j p_l G_{ijkl} - \mu - \alpha_i - \alpha_k \]
\[= G_{i\cdot k\cdot} - G_{i\cdot \cdot k\cdot} + G_{\cdot \cdot \cdot \cdot} \quad (8.145a) \]
for every combination of \(i\) and \(k\)

\[(\alpha\alpha)_{il} = \sum_j \sum k G_{ijkl} - \mu - \alpha_i - \alpha_l \]
\[= G_{i\cdot k\cdot l} - G_{i\cdot \cdot \cdot l} + G_{\cdot \cdot \cdot l\cdot} + G_{\cdot \cdot \cdot \cdot} \quad (8.145b) \]
for every combination of \(i\) and \(l\)

\[(\alpha\alpha)_{jk} = (\alpha\alpha)_{jk} = \quad \text{similar to (8.145a) and (8.145b)} \quad (8.145c)\]

\[(\alpha\alpha)_{jl} = (\alpha\alpha)_{jl} = \quad \text{similar to (8.145a) and (8.145b)} \quad (8.145d)\]

\[(\alpha\delta)_{ikl} = \sum_j G_{ijkl} - \mu - \alpha_i - \alpha_k - \alpha_l - (\alpha\alpha)_{ik} - (\alpha\alpha)_{il} - \delta_{kl} \]
\[= G_{i\cdot k\cdot l} - G_{i\cdot \cdot k\cdot} + G_{i\cdot \cdot \cdot \cdot} + G_{\cdot \cdot \cdot \cdot} \quad (8.146a) \]
for every combination of \(i, k\), and \(l\)

\[(\alpha\delta)_{jkl} = \sum_i G_{ijkl} - \mu - \alpha_j - \alpha_k - \alpha_l - (\alpha\alpha)_{jk} - (\alpha\alpha)_{jl} - \delta_{kl} \]
\[= G_{j\cdot k\cdot l} - G_{j\cdot \cdot k\cdot} + G_{j\cdot \cdot \cdot \cdot} + G_{\cdot \cdot \cdot \cdot} \quad (8.146b) \]
for every combination of \(j, k\), and \(l\)

\[(\delta\alpha)_{ijk} = (\delta\alpha)_{ijk} = \quad \text{similar to (8.146a) and (8.146b)} \quad (8.147a)\]

\[(\delta\alpha)_{ijl} = (\delta\alpha)_{ijl} = \quad \text{similar to (8.146a) and (8.146b)} \quad (8.147b)\]

\[(\delta\delta)_{ijkl} = G_{ijkl} - \mu - \alpha_i - \alpha_j - \alpha_k - \alpha_l \]
\[= \delta_{ij} - (\alpha\alpha)_{ik} - (\alpha\alpha)_{il} - (\alpha\alpha)_{jk} - (\alpha\alpha)_{jl} - \delta_{kl} \]
\[= (\alpha\delta)_{ikl} - (\alpha\delta)_{jkl} - (\delta\alpha)_{ijk} - (\delta\alpha)_{ijl} \]
\[= G_{ijk\cdot} - G_{ij\cdot k\cdot} - G_{i\cdot kl\cdot} - G_{j\cdot kl\cdot} + G_{\cdot \cdot kl\cdot} + G_{\cdot \cdot \cdot k\cdot} + G_{\cdot \cdot \cdot \cdot} \]
\[+ G_{i\cdot j\cdot \cdot\cdot} + G_{i\cdot \cdot \cdot \cdot} + G_{\cdot \cdot \cdot \cdot} + G_{\cdot \cdot \cdot \cdot} \quad (8.148) \]
for every combination of \(i, j, k\), and \(l\)

The following pattern is observed in all of the above definitions. For all parameters the \(G\)'s are weighted and summed over all subscripts not possessed by the effect. Then the main effects (\(\alpha\)'s) are adjusted for only the overall mean. The two-gene or two-factor interaction effects are adjusted for the overall mean
and the two corresponding main effects. The three-gene or three-factor interaction effects are adjusted for the overall mean, the three corresponding main effects, and the three corresponding two-gene effects. The four-gene or four-factor interaction effects are adjusted for the overall mean, the four corresponding main effects, the six corresponding two-gene effects and the four corresponding three-gene effects. The last equation (8.148) given for the (δδ) effect is really a consequence of the least-squares fitting process of the constant μ being present in the model and of the restrictions of all other effects, as noted above (8.141).

We note that these single-locus effects (α's and δ's) ((8.143) (8.144)) are completely analogous to those defined for the single-locus model (8.10) (8.11) (8.13). What has been done is that the genotypic values entered in the cells in a table like Table 8.1 have already been averaged over all the genotypic combinations at the other locus for the particular genotype of that cell at that locus. This averaging must have already occurred, because with four subscripts we are initially dealing with a four-way table.

Turning to the four (αα) terms in the model, each is a two-factor interaction effect, and hence is the residual in a two-way table similar to the dominance effect. For example, for (αα)_{ik} = (αα)^{m}_{a_i b_k} = G_{i \cdot k} - μ - α_i - α_k = G_{i \cdot k} - (μ + α^m_{a_i} + α^m_{b_k}) (8.145a), it is the residual in an m_a by m_b, two-way table similar to Table 8.1, but in this case the levels of one factor are the alleles at the a locus and the levels of the other factor are the alleles at the b locus. For the intragametic (αα)_{ik}^{mm} term, the allelic frequencies for both loci are those in the male population, and the frequency of every intragametic effect can be written as a product of the allelic frequencies because of linkage equilibrium. The (αα)^{mm}_{a_i b_k} effects are the residuals in Table 8.3. The marginal effects, α_i = α^m_{a_i}, are the same as those row effects in Table 8.1
Table 8.3. A two-way table in which the residuals are the intragametic (male by male) additive-by-additive effects. The allelic frequencies at both loci are those of the male population. We let the number of alleles at the A and B loci be \( m_a = u \) and \( m_b = v \), respectively.

<table>
<thead>
<tr>
<th>( B_1 )</th>
<th>( B_2 )</th>
<th>( \ldots )</th>
<th>( B_v )</th>
<th>Mean and frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p_{b1}^m )</td>
<td>( p_{b2}^m )</td>
<td>( \ldots )</td>
<td>( p_{bv}^m )</td>
<td>( G_1 \ldots = \mu + \alpha_{a1}^m )</td>
</tr>
<tr>
<td>( A_1 )</td>
<td>( G_{1.1} )</td>
<td>( G_{1.2} )</td>
<td>( \ldots )</td>
<td>( G_{1.v} )</td>
</tr>
<tr>
<td>( p_{a1}^m )</td>
<td>( p_{a1}^m p_{b1}^m )</td>
<td>( \ldots )</td>
<td>( \ldots )</td>
<td>( \ldots )</td>
</tr>
<tr>
<td>( A_2 )</td>
<td>( G_{2.1} )</td>
<td>( G_{2.2} )</td>
<td>( \ldots )</td>
<td>( G_{2.v} )</td>
</tr>
<tr>
<td>( p_{a2}^m )</td>
<td>( p_{a2}^m p_{b1}^m )</td>
<td>( \ldots )</td>
<td>( \ldots )</td>
<td>( \ldots )</td>
</tr>
<tr>
<td>( \ldots )</td>
<td>( \ldots )</td>
<td>( \ldots )</td>
<td>( \ldots )</td>
<td>( \ldots )</td>
</tr>
<tr>
<td>( A_u )</td>
<td>( G_{u.1} )</td>
<td>( G_{u.2} )</td>
<td>( \ldots )</td>
<td>( G_{u.v} )</td>
</tr>
<tr>
<td>( p_{au}^m )</td>
<td>( p_{au}^m p_{b1}^m )</td>
<td>( \ldots )</td>
<td>( \ldots )</td>
<td>( \ldots )</td>
</tr>
<tr>
<td>Mean</td>
<td>( G_{..1} )</td>
<td>( G_{..2} )</td>
<td>( \ldots )</td>
<td>( G_{..v} )</td>
</tr>
<tr>
<td>Frequency</td>
<td>( p_{b1}^m )</td>
<td>( p_{b2}^m )</td>
<td>( \ldots )</td>
<td>( p_{bv}^m )</td>
</tr>
</tbody>
</table>

representing the a locus. Similarly, the marginal effects of \( \alpha_k = \alpha_{bk}^m \) are the same as the row effects in another table similar to Table 8.1, but representing the b locus. Both sets represent the rows of two different tables, because we are letting the rows represent the male population. For the remaining three terms, \( (\alpha\alpha)^mF \), \( (\alpha\alpha)^mF \) and \( (\alpha\alpha)^fF \), each would be defined from a different table. For example, \( (\alpha\alpha)^mF \) would be defined by replacing the allelic frequencies for locus b.
in Table 8.3 by those from the female population. In this case, the frequency of every intergametic effect can be written as a product of the allelic frequencies because of random mating. Random mating is not strictly a single-locus phenomenon.

In a similar manner, the additive-by-dominance and dominance-by-additive effects are the residuals in different three-way tables, and likewise the dominance-by-dominance effects are the residuals in the four-way table.

With the effects defined as above under random mating and linkage equilibrium, the variance of each term in the model may be written down by definition of a variance (2.91) (2.92) as follows:

\[
\begin{align*}
\text{Locus a:} & \quad \sigma^2_a = \sum a_{i}^m a_{i}^m (\sigma_a^2), \quad \sigma^2_f = \sum a_{j}^m a_{j}^m (\sigma_f^2) \\
& \quad \sigma^2_{mf} = \sum a_{i}^m a_{j}^m (\sigma_{mf}^2) \\
\text{Locus b:} & \quad \sigma^2_b = \sum b_{k}^m b_{k}^m (\sigma_b^2), \quad \sigma^2_f = \sum b_{l}^m b_{l}^m (\sigma_f^2) \\
& \quad \sigma^2_{mf} = \sum b_{k}^m b_{l}^m (\sigma_{mf}^2)
\end{align*}
\]

(8.149a)

(8.149b)

Epistatic effects between loci a and b:
\[
\begin{align*}
\sigma^2_{(\alpha\alpha)_{mm}} &= \sum_{ab} \sum_{i} \sum_{k} p_{a}^{m} p_{b}^{m} (\alpha_{\alpha})^{m m}_{a_{i} b_{k}} \\
\sigma^2_{(\alpha\alpha)_{mf}} &= \sum_{ab} \sum_{i} \sum_{\ell} p_{a}^{m} p_{b}^{f} (\alpha_{\alpha})^{m f}_{a_{i} b_{\ell}} \\
\sigma^2_{(\alpha\alpha)_{fm}} &= \sum_{ab} \sum_{j} \sum_{k} p_{a}^{f} p_{b}^{m} (\alpha_{\alpha})^{f m}_{a_{j} b_{k}} \\
\sigma^2_{(\alpha\alpha)_{ff}} &= \sum_{ab} \sum_{j} \sum_{\ell} p_{a}^{f} p_{b}^{f} (\alpha_{\alpha})^{f f}_{a_{j} b_{\ell}} \\
\sum_{\alpha\delta}^{2}_{mmf} &= \sum_{abb} \sum_{i} \sum_{k} \sum_{\ell} p_{a}^{m} p_{b}^{m} p_{b}^{f} (\alpha_{\delta})^{m m f}_{a_{i} b_{k} b_{\ell}} \\
\sum_{\alpha\delta}^{2}_{fmf} &= \sum_{abb} \sum_{j} \sum_{k} \sum_{\ell} p_{a}^{f} p_{b}^{m} p_{b}^{f} (\alpha_{\delta})^{f m f}_{a_{j} b_{k} b_{\ell}} \\
\sum_{\delta\alpha}^{2}_{mfm} &= \sum_{aab} \sum_{i} \sum_{j} \sum_{k} p_{a}^{m} p_{a}^{m} p_{b}^{f} (\delta_{\alpha})^{m f m}_{a_{i} a_{j} b_{k}} \\
\sum_{\delta\alpha}^{2}_{mff} &= \sum_{aab} \sum_{i} \sum_{j} \sum_{\ell} p_{a}^{m} p_{a}^{f} p_{b}^{f} (\delta_{\alpha})^{m f f}_{a_{i} a_{j} b_{\ell}} \\
\sum_{\delta\delta}^{2}_{mff} &= \sum_{aabb} \sum_{i} \sum_{j} \sum_{k} \sum_{\ell} p_{a}^{m} p_{a}^{f} p_{b}^{f} p_{b}^{f} (\delta_{\delta})^{m f f}_{a_{i} a_{j} b_{k} b_{\ell}}
\end{align*}
\]

(8.149c)

(8.149d)

(8.149e)

Hence, since all parameters or effects in model (8.128) are uncorrelated or have zero covariances under random mating and linkage equilibrium (see Box 8.10), then applying the theorem for the variance of a linear combination of uncorrelated variables, the total variance among genotypes is

\[
\sigma^2_G = \sigma^2_{mff} = \sigma^2_{m} + \sigma^2_{f} + \sigma^2_{mf} + \sigma^2_{m} + \sigma^2_{f} + \sigma^2_{mf} + \sigma^2_{m} + \sigma^2_{f} + \sigma^2_{ff}
\]

\[
+ \sigma^2_{(\alpha\alpha)_{mm}} + \sigma^2_{(\alpha\alpha)_{mf}} + \sigma^2_{(\alpha\alpha)_{fm}} + \sigma^2_{(\alpha\alpha)_{ff}}
\]

\]
\[ + \sigma^2_{(\alpha \delta)}_{abb} \sigma^2_{(\alpha \delta)}_{aabb} + \sigma^2_{(\delta \alpha)}_{aab} + \sigma^2_{(\delta \alpha)}_{aabb} + \sigma^2_{(\delta \delta)}_{aabb} \]

\[ (8.150) \]

where \( \sigma^2_{ya} = \sigma^2_{m} + \sigma^2_{f} + \sigma^2_{mf} = \text{genotypic variance due to locus } a \) (8.149a),

\( \sigma^2_{yb} = \sigma^2_{m} + \sigma^2_{f} + \sigma^2_{mf} = \text{genotypic variance due to locus } b \) (8.149b),

\( \sigma^2_{1} = \sigma^2_{yyab} = \sigma^2_{(aa)}_{ab} + \ldots + \sigma^2_{(\delta \delta)}_{aabb} = \text{variance due to interactions between loci } a \text{ and } b, \text{ or due to epistasis } (8.149c) \)

(8.149d) (8.149e).

The variances, \( \sigma^2_{ya}, \sigma^2_{yb}, \text{ and } \sigma^2_{yyab} \), are the variances of the locus factorial model (8.130). Similarly, a different grouping of genic effects with terms having only the m superscript, only f, and both m and f, gives the variances, \( \sigma^2_{gm}, \sigma^2_{gf}, \text{ and } \sigma^2_{gmgf} \), respectively, of the gametic factorial model (8.131).

**Box 8.10**

Example that covariance between any two terms in genic factorial model (8.128) is zero (see (8.150))

In a manner similar to that in Boxes 8.3 and 8.8, we consider, as an example, the covariance between \( \delta^m_{a_i a_j} \) and \( (aa)^m_{a_i b_k} = (aa)_{ik} \) (see (8.133) (8.134) for abbreviated notation)

\[ \text{Cov}[(\delta_{ij}, (aa))_{ik}] = \sum \sum P_{\delta_{ij} (aa)_{ik}} (\delta_{ij})[(aa)_{ik}] = \sum \sum P_{ijk} (\delta_{ij})[(aa)_{ik}] \]

where \( P_{\delta_{ij} (aa)_{ik}} = P_{ijk} \) = joint frequency of \( \delta_{ij} \) and \( (aa)_{ik} \) effects occurring together, which is the occurrence of \( G_{ijk} \).

Under random mating and linkage equilibrium,

\[ P_{\delta_{ij} (aa)_{ik}} = P_{ijk} = P_i P_j P_k \]
\begin{align*}
\text{Cov}[\delta_{ij}, (\alpha\alpha)_{ik}] = & \sum_i \sum_j \sum_k p_i p_j p_k (\delta_{ij}) (\alpha\alpha)_{ik} \\
= & -(0)(0) = 0 \quad \text{(sub. (8.138a) (also see (8.16) (8.139a))}
\end{align*}

Limiting ourselves to two loci, we may rewrite (8.150)

\[
\sigma^2_G = \sigma^2_{G, \text{mfm}} = \sigma^2_{\text{Am}} + \sigma^2_{\text{Af}} + \sigma^2_{\text{Dmf}} + \sigma^2_{\text{AmAm}} + \sigma^2_{\text{AmAf}} + \sigma^2_{\text{AfAf}} \\
+ \sigma^2_{\text{AmD}} + \sigma^2_{\text{AfD}} + \sigma^2_{\text{DDmf}}
\]

(8.151)

where

\[
\sigma^2_{\text{Am}} = \sigma^2_{\text{Am}}, \quad \sigma^2_{\text{Af}} = \sigma^2_{\text{Af}}, \quad \sigma^2_{\text{Dmf}} = \sigma^2_{\text{Dmf}}, \\
\sigma^2_{\text{AmAm}} = \sigma^2_{(\alpha\alpha)_{am}}, \quad \sigma^2_{\text{AmAf}} = \sigma^2_{(\alpha\alpha)_{af}}, \\
\sigma^2_{\text{AfAf}} = \sigma^2_{(\alpha\alpha)_{af}}, \quad \sigma^2_{\text{AmD}} = \sigma^2_{(\alpha\delta)_{md}}, \\
\sigma^2_{\text{AfD}} = \sigma^2_{(\alpha\delta)_{df}}, \quad \sigma^2_{\text{DDmf}} = \sigma^2_{(\delta\delta)_{mf}}.
\]

It should be pointed out here that these variances apply only to the crossed population itself, i.e., to the \(F_1\) cross population between the two populations (Griffing, 1962). The variances do not apply to the cross population subsequently derived by random mating one or any number of generations thereafter. This is so
because if the allelic frequencies in the two parental populations are not equal
\( p^m_i \neq p^f_i \) for at least one \( i \), the allelic frequencies in the crossed
population are \( p_i = \frac{p^m_i + p^f_i}{2} \) and would produce equal male and female allelic
frequencies at each locus. Hence, all effects in the random-mated population are
different from those in either parental population, and their variances are
changed.

2. **Linkage disequilibrium.** Nyquist -- discuss linkage disequilibrium here.

8.3.2. **Two loci, multiple alleles, equal male and female allelic frequencies
at each locus (Hardy-Weinberg), epistasis.** 1. **Linkage equilibrium.** The
development in the previous section was general and included a cross between two
populations. The allelic frequencies, \( p^m \) and \( p^f \), were assumed to be different in
the two sexual arrays for both loci \( a \) and \( b \). In this section we restrict
ourselves to a population in Hardy-Weinberg equilibrium, i.e.,
\[
\begin{align*}
\frac{p^m_a}{p^m_i} &= \frac{p^f_a}{p^f_i} = p^a_i & \text{for } i = 1, \ldots, m_a, \\
\frac{p^m_b}{p^m_k} &= \frac{p^f_b}{p^f_k} = p^b_k & \text{for } k = 1, \ldots, m_b,
\end{align*}
\]  
(8.152)

Again by dropping the \( m \) and \( f \) superscripts throughout in the previous section, we
obtain
\[
\begin{align*}
\frac{\sigma^2_{a}}{\sigma^2_{a}} = \frac{\sigma^2_{f}}{\sigma^2_{a}} = \frac{\sigma^2_{a}}{\sigma^2_{a}} &= \frac{\sigma^2_{A}}{4} \\
\sigma^2_{\delta_b} &= \sigma^2_{D_a} \\
\sigma^2_{a} = \sigma^2_{A_b} &= \frac{\sigma^2_{D_b}}{4} \\
\sigma^2_{\delta_a} &= \sigma^2_{D_a} \\
\sigma^2_{(aa)_{ab}} = \sigma^2_{(aa)_{ab}} &= \sigma^2_{(aa)_{ab}} = \sigma^2_{(aa)_{ab}} = \frac{\sigma^2_{A_{ab}}}{4} \\
\end{align*}
\]  
(8.153a)
(8.153b)
(8.153c)
(8.153d)
(8.153e)
\[ \sigma_{(\alpha \delta)}^2_{abb} = \sigma_{(\alpha \delta)}^2_{fmm} - \sigma_{(\alpha \delta)}^2_{mfa} = \sigma_{(\alpha \delta)}^2_{ss} = \sigma_{AD}^2_{ab} - \frac{\sigma_{AD}^2_{ab}}{2} \]  
(8.153f)

\[ \sigma_{(\delta \alpha)}^2_{aab} = \sigma_{(\delta \alpha)}^2_{mfm} - \sigma_{(\delta \alpha)}^2_{mfm} = \sigma_{(\delta \alpha)}^2_{ss} = \sigma_{DA}^2_{ab} - \frac{\sigma_{DA}^2_{ab}}{2} \]  
(8.153g)

\[ \sigma_{\delta \delta}^2_{ab} = \sigma_{DD}^2_{ab} \]  
(8.153h)

Hence, (8.150) becomes

\[ \sigma_G^2 = \sigma_G^2_{a^i b^j k^i b^j} = 2\sigma_{\alpha}^2_a + \sigma_{\delta}^2_a + 2\sigma_{\alpha}^2_b + \sigma_{\delta}^2_b \]

\[ = 4\sigma_{(\alpha \alpha)}^2_{ab} + 2\sigma_{(\alpha \delta)}^2_{ab} + 2\sigma_{(\delta \alpha)}^2_{ab} + \sigma_{(\delta \delta)}^2_{ab} \]  
(8.154)

Further, we may rewrite (8.154)

\[ \sigma_G^2 = \sigma_G^2_{a^i b^j k^i b^j} = \sigma_{A_a}^2 + \sigma_{D_a}^2 + \sigma_{A_b}^2 + \sigma_{D_b}^2 \]

\[ + \sigma_{AA}^2_{ab} + \sigma_{AD}^2_{ab} + \sigma_{DA}^2_{ab} + \sigma_{DD}^2_{ab} \]  
(8.155)

where

\[ \sigma_{A_a}^2 = 2\sigma_{\alpha}^2_a = \text{additive variance for locus a}, \]

\[ \sigma_{A_b}^2 = 2\sigma_{\alpha}^2_b = \text{additive variance for locus b}, \]

\[ \sigma_{D_a}^2 = \sigma_{\delta}^2_a = \text{dominance variance for locus a}, \]

\[ \sigma_{D_b}^2 = \sigma_{\delta}^2_b = \text{dominance variance for locus b}, \]

\[ \sigma_{AA}^2_{ab} = 4\sigma_{(\alpha \alpha)}^2_{ab} = \text{additive-by-additive variance for loci a and b}, \]

\[ \sigma_{AD}^2_{ab} = 2\sigma_{(\alpha \delta)}^2_{ab} = \text{additive-by-dominance variance for loci a and b}, \]

\[ \sigma_{DA}^2_{ab} = 2\sigma_{(\delta \alpha)}^2_{ab} = \text{dominance-by-additive variance for loci a and b}, \]

\[ \sigma_{DD}^2_{ab} = \sigma_{(\delta \delta)}^2_{ab} = \text{dominance-by-dominance variance for loci a and b}. \]
Limiting ourselves to two loci, we may rewrite (8.155)

\[
\sigma_G^2 = \sigma_A^2 + \sigma_D^2 + \sigma_{AA}^2 + \sigma_{AD}^2 + \sigma_{DD}^2
\]  

(8.156)

where \( \sigma_A^2 = \sigma_{A_a}^2 + \sigma_{A_b}^2 \) = total additive variance for loci a and b,

\( \sigma_D^2 = \sigma_{D_a}^2 + \sigma_{D_b}^2 \) = total dominance variance for loci a and b,

\( \sigma_{AA}^2 \) = additive-by-additive variance for loci a and b,

\( \sigma_{AD}^2 = \sigma_{AD_{ab}}^2 + \sigma_{DA_{ab}}^2 \) = total additive-by-dominance variance for loci a and b.

\( \sigma_{DD}^2 \) = dominance-by-dominance variance for loci a and b.

2. **Linkage disequilibrium.** Nyquist -- discuss linkage disequilibrium here, as I did in Section 8.1.6 or 8.1.7.

8.3.3. **Two loci, two alleles, equal male and female allelic frequencies at each locus (Hardy-Weinberg), linkage equilibrium, epistasis.** Explicit expressions for epistatic variances. In the previous section, we considered multiple alleles. When we go from multiple alleles to only two alleles at each locus, as we do here, we can enumerate all genotypes at one locus as the levels of one factor, and similarly for the other locus. We have the following factorial table:
### B locus

<table>
<thead>
<tr>
<th>A locus</th>
<th>A1A1</th>
<th>A1A2</th>
<th>A2A2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>G111</td>
<td>G112</td>
<td>G112</td>
</tr>
<tr>
<td></td>
<td>G22</td>
<td>G21</td>
<td>G20</td>
</tr>
<tr>
<td></td>
<td>G121</td>
<td>G122</td>
<td>G122</td>
</tr>
<tr>
<td></td>
<td>G12</td>
<td>G11</td>
<td>G10</td>
</tr>
<tr>
<td></td>
<td>G211</td>
<td>G212</td>
<td>G222</td>
</tr>
<tr>
<td></td>
<td>G02</td>
<td>G01</td>
<td>G00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{A1}^2$</td>
<td>$G_{11..}$</td>
</tr>
<tr>
<td>$-G_{22}$</td>
<td>$-G_2$</td>
</tr>
<tr>
<td>$2p_{A1}p_{A2}$</td>
<td>$G_{12..}$</td>
</tr>
<tr>
<td>$-G_{12}$</td>
<td>$-G_1$</td>
</tr>
<tr>
<td>$p_{A2}^2$</td>
<td>$G_{22..}$</td>
</tr>
<tr>
<td>$-G_{02}$</td>
<td>$-G_0$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{B1}^2$</td>
<td>$G_{..11}$</td>
</tr>
<tr>
<td>$2p_{B1}p_{B2}$</td>
<td>$G_{..12}$</td>
</tr>
<tr>
<td>$p_{B2}^2$</td>
<td>$G_{..22}$</td>
</tr>
<tr>
<td>$-G_{..22}$</td>
<td>$-G_{..}$</td>
</tr>
</tbody>
</table>

In a random-mating population the marginal frequencies for each locus represent Hardy-Weinberg proportions. With linkage equilibrium, the genotypes at different loci are independent, so their joint frequencies are products of the marginal frequencies, e.g., the frequency of $G_{1111} = G_{22}$ is $p_{A1}^2 p_{B1}^2$. Note the abbreviated notation for the $G$'s, changing from four subscripts (or factors) with two levels each to two subscripts (or factors) with three levels each. The 2, 1, 0 notation at the A locus, for example, corresponds to the number of $A_1$ alleles present in the genotype (same as in Figure 8.1). This is convenient when there are only two alleles at each locus.

The total genotypic variance among the $G$'s in the above table (8.157) can be partitioned first into the components of the locus model (see model (8.130)), i.e., the marginal variance due to the A locus, the marginal variance due to the B locus, and an interaction or epistatic variance due to the residual effects in the table. The epistatic variance can be calculated in the same manner that the dominance variance (8.21) was calculated from Table 8.1, or likewise that each
component of the additive-by-additive variance (8.149c) was calculated from Table 8.3. If the effects of loci A and B are additive (no epistasis) and the genotypes at the different loci are independent, as we have assumed (hence, all covariances between the effects of the two loci are zero), the total variation in the table (8.157) is obtained from marginal means and their marginal frequencies, and is the sum of the marginal variances only (8.94). If the effects of the two loci are not additive, this is not true. There is an additional variance, namely, the interlocus variance component, the epistatic variance $\sigma_{yy}^2 = \sigma_I^2$ (8.150). When an interaction exists in one or more of the following two-by-two subtables in (8.157), there is nonadditivity between loci, or a nonzero epistatic variance exists.

$\begin{array}{cccc}
B_1B_1 & B_1B_2 & B_2B_1 & B_2B_2 \\
A_1A_1 & G_{22} & G_{21} & A_1A_2 & G_{12} & G_{11} & A_1A_2 & G_{11} & G_{10} \\
A_1A_2 & G_{12} & G_{11} & A_1A_2 & G_{11} & G_{10} & \end{array}$

We have already learned that dominance, an interaction between alleles, is present when $G_{11} - G_{12} \neq G_{12} - G_{22}$ exists (see sentences above (8.72) and (8.30)), namely, a difference in the simple effects of a $2 \times 2$ table. Similarly, when an interaction exists in any of the $2 \times 2$ subtables (8.158), an epistatic interaction exists. For example, in the first $2 \times 2$ table, if $G_{22} - G_{21} \neq G_{12} - G_{11}$, then an interaction between loci A and B occurs. In terms of genotypes we note that $G_{22} - G_{21}$ is the difference associated with genotypes $B_1B_1$ and $B_1B_2$ in the presence of $A_1A_1$, whereas $G_{12} - G_{11}$ is the difference associated with genotypes $B_1B_1$ and $B_1B_2$ in the presence of $A_1A_2$. Hence, if these two differences associated with $B_1B_1$ and $B_1B_2$ are different in the presence of $A_1A_1$ and $A_1A_2$, then an interaction exists in the $2 \times 2$ table.
It is worth pointing out how likely the presence of epistasis is, at least, from some theoretical point of view. For example, for no interactions (no dominance, no epistasis) to exist in the 3 × 3 table in (8.157) only three values need to be specified to determine all other values in the table (only two if the mean or a reference point is given). Suppose the values 6, 5, and 4 are given in the unshaded cells in the diagram below. Then all other values in the diagram are determined under the assumption of no interaction. From some point of view this appears rather unlikely or unreal.

\[
\begin{array}{ccc}
B_1B_1 & B_1B_2 & B_2B_2 \\
A_1A_1 & 6 & 4 & 2 \\
A_1A_2 & 5 & 3 & 1 \\
A_2A_2 & 4 & 2 & 0 \\
\end{array}
\]

(8.159)

As another example, consider a model with dominance but no epistasis. In this case, only five values need to be specified.

\[
\begin{array}{ccc}
B_1B_1 & B_1B_2 & B_2B_2 \\
A_1A_1 & 4 & 3 & 1 \\
A_1A_2 & 4 & 3 & 1 \\
A_2A_2 & 3 & 2 & 0 \\
\end{array}
\]

(8.160)

Note that the two differences in each of the four 2 × 2 tables are equal in both of the above 3 × 3 tables, so no epistasis exists. Even with dominance, many cells are fixed after values are assigned to a few cells.

After partitioning according to the locus factorial model into \( \sigma^2_{ya} \), \( \sigma^2_{yb} \), and \( \sigma^2_{yyab} \), further partitioning of the locus components, as discussed above (8.155), can be closely allied to the partitioning of the variation in a 3 × 3 factorial experiment in statistics, where both factors are quantitative and a regression or response surface analysis is performed (Steel and Torrie, 1980, Section 15.7; Snedecor and Cochran, 1967, Section 12.8). The additive and
dominance variances for each locus can be calculated from the expressions given in (8.51) and (8.52), respectively, using the mean values, \( G_{i,j} \)'s for the A locus and \( G_{k,l} \)'s for the B locus (8.157). The additive variance is that portion of the variance at a given locus accounted for by a least squares regression line (the linear component) (8.69) (8.70), and the dominance variance is that portion about the regression line (the quadratic component) (8.71). The additive-by-additive epistatic variance corresponds to the A-linear-by-B-linear component, additive-by-dominance to A-linear-by-B-quadratic component, dominance-by-additive to A-quadratic-by-B-linear, and dominance-by-dominance to the A-quadratic-by-B-quadratic components in the partitioning of an interaction in a 3 \( \times \) 3, A \( \times \) B factorial experiment. Computing formulas for all partitions of the epistatic variance can be expressed in terms of the interactions in the four 2 \( \times \) 2 subtables (8.158) (Cockerham, 1954). The interaction in any 2 \( \times \) 2 table is measured by the difference between its two simple effects or the main diagonal minus the other diagonal (Steel and Torrie, 1980, p. 340). Hence, for each 2 \( \times \) 2 table, we define (note that these are defined, for convenience in subsequent expressions, somewhat differently from those in Cockerham, 1954)
\[
\begin{align*}
\epsilon_{11} &= G_{1111} - G_{1112} - G_{1211} + G_{1212} = G_{22} - G_{21} - G_{12} + G_{11} \\
\epsilon_{12} &= G_{1112} - G_{1122} - G_{1212} + G_{1222} = G_{21} - G_{20} - G_{11} + G_{10} \\
\epsilon_{21} &= G_{1211} - G_{1212} - G_{2211} + G_{2212} = G_{12} - G_{11} - G_{02} + G_{01} \\
\epsilon_{22} &= G_{1212} - G_{1222} - G_{2212} + G_{2222} = G_{11} - G_{10} - G_{01} + G_{00}
\end{align*}
\]
Note that the double heterozygote is involved in all \( \epsilon \)'s. For no interlocus interaction or no epistatic effects, we have
\[
\epsilon_{11} = \epsilon_{12} = \epsilon_{21} = \epsilon_{22} = 0
\]
The epistatic variances are (these are not derived herein, but the orthogonal polynomial coefficients for each epistatic source can be easily obtained by multiplying the corresponding orthogonal polynomial coefficient for the linear
(additive) or quadratic (dominance) at the separate loci (8.71A) together (a Box
should be developed), see Cockerham (1954) and (8.161))

\[
\sigma_{AA}^2_{ab} = 4\sigma_{(aa)}^2_{ab} = 4p_{A1}p_{A2}p_{B1}p_{B2}(p_{A1}p_{B1}e_{11} + p_{A1}p_{B2}e_{12} + p_{A2}p_{B1}e_{21} + p_{A2}p_{B2}e_{22})^2
\]

(8.163a)

\[
\sigma_{AD}^2_{ab} = 2\sigma_{(a\delta)}^2_{ab} = 2p_{A1}p_{A2}p_{B1}^2p_{B2}^2[p_{A1}(e_{11} - e_{12}) + p_{A2}(e_{21} - e_{22})]^2
\]

(8.163b)

\[
\sigma_{DA}^2_{ab} = 2\sigma_{(\delta a)}^2_{ab} = 2p_{A1}^2p_{A2}p_{B1}p_{B2}^2[p_{B1}(e_{11} - e_{21}) + p_{B2}(e_{12} - e_{22})]^2
\]

(8.163c)

\[
\sigma_{DD}^2_{ab} = \sigma_{\delta\delta}^2_{ab} = 2p_{A1}p_{A2}p_{B1}^2p_{B2}^2(e_{11} - e_{12} - e_{21} + e_{22})^2
\]

(8.163d)

where

\[
\sum_{i}^{2} \sum_{j}^{2} p_{Ai}p_{Bj}e_{ij} = \text{additive-by-additive contrast},
\]

\[
p_{A1}(e_{11} - e_{12}) + p_{A2}(e_{21} - e_{22}) = \text{additive-by-dominance contrast},
\]

\[
p_{B1}(e_{11} - e_{21}) + p_{B2}(e_{12} - e_{22}) = \text{dominance-by-additive contrast},
\]

\[
e_{11} - e_{12} - e_{21} + e_{22} = \text{dominance-by-dominance contrast}.
\]

The additive-by-additive variance is a function of the weighted mean of all e's.

The additive-by-dominance variance is a weighted mean of two differences of the

e's, namely,

<table>
<thead>
<tr>
<th>e_{11}</th>
<th>e_{12}</th>
</tr>
</thead>
<tbody>
<tr>
<td>p_{A1}</td>
<td>e_{11} - e_{12}</td>
</tr>
<tr>
<td>e_{21}</td>
<td>e_{22}</td>
</tr>
</tbody>
</table>

(8.164)

The dominance-by-additive variance is a weighted mean of two differences of the
e's obtained in the other direction, namely,

<table>
<thead>
<tr>
<th>e_{11}</th>
<th>e_{12}</th>
</tr>
</thead>
<tbody>
<tr>
<td>e_{21}</td>
<td>e_{22}</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>p_{B1}</th>
<th>p_{B2}</th>
</tr>
</thead>
<tbody>
<tr>
<td>e_{11} - e_{21}</td>
<td>e_{12} - e_{22}</td>
</tr>
</tbody>
</table>

(8.165)
The dominance-by-dominance variance is another function of the \( e \)'s, namely, the sum of the main diagonal minus the other diagonal, \( (e_{11} + e_{22}) - (e_{12} + e_{21}) = e_{11} - e_{12} - e_{21} + e_{22} \).

The coefficients 4, 2, 2, and 1 in (8.163) represent the number of variances of the same type of effects in the genic factorial model that have been summed together (8.153) (8.154). For example, in the model (8.128) there are four \((aa)\)'s. In a random-mating population in linkage equilibrium, each effect of the same type has the same variance (8.153).

Crow and Kimura (1970, pp. 127 to 129) also give some formulas with arbitrary allelic frequencies for calculating \( \sigma_{AA}^2 \), \( \sigma_{AD}^2 \), and \( \sigma_{DD}^2 \).

Any specific numerical model with two equally frequent alleles with only certain variance components can be generated, as follows. For example, a model with additive and dominance variances at both loci and additive (at the A locus)-by-dominance (at the B locus) only can be obtained as shown in the following table:

<table>
<thead>
<tr>
<th></th>
<th>( B_1B_1 )</th>
<th>( B_1B_2 )</th>
<th>( B_2B_2 )</th>
<th>A locus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1/4</td>
<td>1/2</td>
<td>1/4</td>
<td>add. dom.</td>
</tr>
<tr>
<td>( A_1A_1 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/4</td>
<td>3 + 3</td>
<td>3 + 2</td>
<td>3 + 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+0 + 0</td>
<td>+0 + 1</td>
<td>+0 + 0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+3(0)</td>
<td>+3(1)</td>
<td>+3(0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>= 6</td>
<td>= 9</td>
<td>= 4</td>
<td></td>
</tr>
<tr>
<td>( A_1A_2 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/2</td>
<td>2 + 3</td>
<td>2 + 2</td>
<td>2 + 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+1 + 0</td>
<td>+1 + 1</td>
<td>+1 + 0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+3(0)</td>
<td>+2(1)</td>
<td>+2(0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>= 6</td>
<td>= 8</td>
<td>= 4</td>
<td></td>
</tr>
<tr>
<td>( A_2A_2 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/4</td>
<td>1 + 3</td>
<td>1 + 2</td>
<td>1 + 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+0 + 0</td>
<td>+0 + 1</td>
<td>+0 + 0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+1(0)</td>
<td>+1(1)</td>
<td>+1(0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>= 4</td>
<td>= 5</td>
<td>= 2</td>
<td></td>
</tr>
</tbody>
</table>

B locus

<table>
<thead>
<tr>
<th></th>
<th>add.</th>
<th>dom.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

(8.165A)
To generate such a table we assigned a linear (additive) sequence (1, 2, and 3) and a quadratic (dominance) sequence (0, 1, 0) (which gave complete dominance) to both loci. Then since we desired only the additive (at the A locus)-by-dominance (at the B locus), we multiplied the linear (additive) value at the A locus by the quadratic (dominance) coefficient at the B locus, and added all values in the cells together to obtain the genotypic values. That one has only the $\sigma^2_{ADab}$ epistatic variance component can be established easily by calculating the $e'$s, and verifying that the only nonzero comparison is the additive-by-dominance comparison.

$$\begin{array}{c|c}
e_{11} & e_{12} \\
-1 & +1 \\
\hline
e_{21} & e_{22} \\
-1 & +1 \\
\end{array}$$

$$e_{11} - e_{12} = -2$$

$$e_{21} - e_{22} = -2$$

$$e_{1j} - e_{2j} = 0$$

When more loci are encompassed, the G's in (8.157) and (8.161) are mean values over all other loci. Additive effects, $\alpha$'s, are average effects of alleles, averaged over all other alleles and nonalleles in the population, and are dependent on their frequencies in the population. Two-locus epistatic effects are average effects, averaged over other loci, and so on.

**Example 8.5.** Assume two loci in a population, two alleles at each locus, equal allelic frequencies, $p_{A1} = p_{A2} = 1/2$, $p_{B1} = p_{B2} = 1/2$, linkage equilibrium, and complementary gene action (a 9:7 $F_2$ genetic ratio) with genotypic values as given below (no assumption about the recombination or linkage value between the two loci need to be made with linkage equilibrium):
First the total genotypic variance will be partitioned according to the locus factorial model (8.130) into its component parts: variance due to locus a, variance due to locus b, and the interaction between loci (the epistatic variance) (8.150). This partitioning is analogous to that of a proportional subclass two-factor experiment. The variances are as follows:
Total genotypic variance: By definition (2.91) (2.92)

\[
\sigma_G^2 = \sum \sum p_{ij} G_{ij}^2 - (G..)^2 = \sum \sum p_i \cdot p_j G_{ij}^2 - (G..)^2 \\
= \left( \frac{1}{16} + \frac{1}{8} + \frac{1}{4} \right)(4)^2 + \left( \frac{1}{16} + \frac{1}{8} + \frac{1}{16} + \frac{1}{8} + \frac{1}{16} \right)(2)^2 - (3 1/8)^2 \\
= \frac{9}{16} (16) + \frac{7}{16} (4) - \left( \frac{25}{8} \right)^2 = \frac{63}{64} \tag{2}
\]

Locus A variance: Locus A is symbolized a.

\[
\sigma_a^2 = \sigma_y^2 = \sum p_i^2 \cdot G_i - (G..)^2 = \left( \frac{1}{4} + \frac{1}{2} \right)(3 1/2)^2 + \frac{1}{4} (2)^2 - (3 1/8)^2 = \frac{27}{64} \tag{3}
\]

Locus B variance: Locus B is symbolized b.

\[
\sigma_b^2 = \sigma_y^2 = \sigma_G^2 = \frac{27}{64} \quad \text{(same as locus A)} \tag{4}
\]

Variance for interaction between loci A and B: Epistatic variance. Initially this will be calculated by definition to emphasize the fact that the epistatic variance is the variance of the residual effects, or the interaction variance in the above two-way table. The fitted or predicted ijth cell mean based on marginal information only is equal to the grand mean plus the deviation of the mean of the ith row from the grand mean plus the deviation of the mean of the jth column from the grand mean, i.e., \( \hat{G}_{ij} = G.. + (G_{i..} - G..) + (G_{.j} - G..) = G_{i..} + G_{.j} - G.. \), and are shown below
\[
\begin{array}{c|c|c|c}
A_1 A_1 & B_1 B_1 & B_1 B_2 & B_2 B_2 \\
1/4 & 1/4 & 1/2 & 1/4 \\
G_{22} & G_{21} & G_{20} \\
4 & 4 & 2 \\
1/16 & 1/8 & 1/16 \\
\hat{G}_{22} = 3 \ 7/8 & \hat{G}_{21} = 3 \ 7/8 & \hat{G}_{20} = 2 \ 3/8 \\
\hline
A_1 A_2 & G_{12} & G_{11} & G_{10} \\
1/2 & 4 & 4 & 2 \\
1/8 & 1/4 & 1/8 \\
\hat{G}_{12} = 3 \ 7/8 & \hat{G}_{11} = 3 \ 7/8 & \hat{G}_{10} = 2 \ 3/8 \\
\hline
A_2 A_2 & G_{02} & G_{01} & G_{00} \\
1/4 & 2 & 2 & \frac{9}{8} \\
1/16 & 1/8 & 1/16 \\
\hat{G}_{02} = 2 \ 3/8 & \hat{G}_{01} = 2 \ 3/8 & \hat{G}_{00} = 7/8 \\
\end{array}
\]

Mean: \( G_{..} = 3 \ \frac{1}{2} \quad G_{.1} = 3 \ \frac{1}{2} \quad G_{.0} = 2 \)

Deviation from grand mean:
\[
\begin{align*}
\frac{3}{8} & \quad \frac{3}{8} & \quad -\frac{9}{8} & \quad G_{..} = 3 \ \frac{1}{8} \\
\end{align*}
\]

Hence,
\[
\sigma_{(yy)}^{2}_{ab} = \left( \frac{1}{16} + \frac{1}{8} + \frac{1}{8} + \frac{1}{4} \right) (4 - 3 \ \frac{7}{8})^2 + \left( \frac{1}{16} + \frac{1}{8} + \frac{1}{16} + \frac{1}{8} \right) (2 - 2 \ \frac{3}{8})^2 + \frac{1}{16} \left( 2 - \frac{7}{8} \right)^2 \\
= \frac{9}{64}
\]

which is equal to that obtained more easily by the total genotypic variance \( (2) \) minus the individual locus components, \( (3) \) and \( (4) \).

\[
\sigma_{(yy)}^{2}_{ab} = \sigma_G^2 - \sigma_{Ga}^2 - \sigma_{Gb}^2 = \frac{63}{64} - \frac{27}{64} - \frac{27}{64} = \frac{9}{64}
\]

We will now partition each of the above variances into their component variances. Each of the locus variances will be partitioned into additive and
dominance variances. Then the interlocus or epistatic variance will be partitioned into additive-by-additive, additive-by-dominance, dominance-by-additive, and dominance-by-dominance variances. Each of the components will be calculated by using the definition formula to impart a fundamental understanding of their nature, and will also be verified by the computing formula. However, in practice the definition formula is too laborious. We will then show that the sum of the components adds to their respective totals.

**Locus A: Additive variance.** To use the definition formula for the additive variance \( (8.18) \) \( (8.20) \) \( (8.29) \), we must first compute \( a_{a_1} \) and \( a_{a_2} \), which are the row or column effects in the following table, namely,

<table>
<thead>
<tr>
<th></th>
<th>( A_1 )</th>
<th>( A_2 )</th>
<th>( \alpha_{a_1} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( G_{2} - G_{11} )</td>
<td>3 1/2</td>
<td>3 1/2</td>
<td>1/2</td>
</tr>
<tr>
<td>( 3 1/2 )</td>
<td>1/4</td>
<td>1/4</td>
<td>1/2</td>
</tr>
<tr>
<td>( G_{1} - G_{12} )</td>
<td>3 1/2</td>
<td>2</td>
<td>1/2</td>
</tr>
<tr>
<td>( 3 1/2 )</td>
<td>1/4</td>
<td>1/4</td>
<td>1/2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Mean</th>
<th>Deviation from grand mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/2</td>
<td>3 1/2</td>
<td>+3/8</td>
</tr>
<tr>
<td>3 1/2</td>
<td>2</td>
<td>-3/8</td>
</tr>
</tbody>
</table>

\[ G_{...} = 3 1/8 \]

So

\[
\sigma^2_{a_a} = 2\sigma^2_{a_a} - 2 \sum A_1 \left( \alpha_{a_1} \right)^2 = 2[(1/2)(3/8)^2 + (1/2)(-3/8)^2] = 9/32 \] \( (9) \)

Using computing formula \( (8.51) \)

\[
\sigma^2_{a_a} = 2\sigma^2_{a_a} - 2p_{A_1} p_{A_2} \left[ p_{A_1} \left( G_{11} - G_{11} \right) + p_{A_2} \left( G_{12} - G_{12} \right) \right]^2
= 2(1/2)(1/2) [(1/2)(3 1/2 - 3 1/2) + (1/2)(3 1/2 - 2)]^2 = 9/32 \] \( (10) \)
Locus A: Dominance variance. To calculate the dominance variance by definition (8.21), we compute the expected or fitted value for each cell $\hat{G}_{ij..} = G_{..} + \alpha_{ai} + \alpha_{aj}$

\[
\begin{array}{ccc}
A_1 & A_2 & a_{ai} \\
1/2 & 1/2 & \\

\begin{array}{ccc}
A_1 & A_1 & G_{11..} = 3 7/8 \\
3 1/2 & 3 1/2 & +3/8 \\
1/4 & 1/4 & \\

\begin{array}{ccc}
A_2 & A_2 & G_{12..} = 3 1/8 \\
3 1/2 & 2 & -3/8 \\
1/4 & 1/4 & \\

\begin{array}{ccc}
A_1 & A_2 & \hat{G}_{12..} = 3 1/8 \\
1/2 & 1/2 & \\

\begin{array}{ccc}
A_1 & A_1 & \hat{G}_{21..} = 3 1/8 \\
1/2 & 1/2 & \\

\begin{array}{ccc}
A_2 & A_2 & \hat{G}_{22..} = 2 3/8 \\
1/2 & 1/2 & \\
\end{array}
\end{array}
\end{array}
\end{array}
\end{array}
\]

Then the dominance variance

\[
\sigma^2_{Da} = \sigma^2_{\delta a} = \sum_i \sum_j p_{A_1} p_{A_2} (G_{ij..} - \hat{G}_{ij..})^2
\]

\[
= (1/4)(3 1/2 - 3 7/8)^2 + (1/4 + 1/4)(3 1/2 - 3 1/8)^2 + (1/4)(2 - 2 3/8)^2
\]

\[
= 9/64
\]

Using the computing formula (8.52)

\[
\sigma^2_{Da} = \sigma^2_{\delta a} = p_{A_1}^2 p_{A_2}^2 (2G_{12} - G_{11} - G_{22})^2 = (1/2)^2 (1/2)^2 [2(3 1/2 - 3 1/2 - 2)]^2
\]

\[
= 9/64
\]

Check: Variance for locus A equals additive variance plus dominance variance

(8.22) (8.29). Thus, adding (10) and (13)

\[
\sigma^2_G = \sigma^2_y = 2\sigma^2_a + \sigma^2_{\delta a} = 9/32 + 9/64 = 27/64
\]

which is (3).
Locus B: Additive and dominance variances. Since the marginal means and frequencies for the genotypes at the B locus are the same as for the A locus, the additive and dominance variances at the B locus are

\[ 2\sigma_{a_b}^2 = 2\sigma_{a_a}^2 = 9/32 \]

\[ \sigma_{\delta_b}^2 = \sigma_{\delta_a}^2 = 9/64 \]

(15)

Additive-by-additive variance. Let us consider the first additive-by-additive term \((a_a^m a_b^m)_{ik}\). It is a two-factor interaction term, and, hence, is the residual effect in the following two-way table:

<table>
<thead>
<tr>
<th></th>
<th>B₁</th>
<th>B₂</th>
<th>Freq.</th>
<th>G₁...</th>
<th>G₁... - G...</th>
</tr>
</thead>
<tbody>
<tr>
<td>A₁</td>
<td>A₁-B₁⁻</td>
<td>A₁-B₂⁻</td>
<td>1/2</td>
<td>3 1/2</td>
<td>α₁ = 3/8</td>
</tr>
<tr>
<td>P_{A₁}</td>
<td>1/4</td>
<td>1/4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G₁.₁⁻</td>
<td>4</td>
<td>G₁.₂⁻</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G₁.₁⁻ = 3</td>
<td>G₁.₂⁻ = 3</td>
<td>3 1/2</td>
<td>3 1/2</td>
<td>α₁ = 3/8</td>
</tr>
<tr>
<td>A₂</td>
<td>A₂-B₁⁻</td>
<td>A₂-B₂⁻</td>
<td>1/2</td>
<td>2 3/4</td>
<td>α₂ = 3/8</td>
</tr>
<tr>
<td>P_{A₂}</td>
<td>1/4</td>
<td>1/4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G₂.₁⁻</td>
<td>3</td>
<td>G₂.₂⁻</td>
<td>2 1/2</td>
<td>2 3/4</td>
<td>α₂ = 3/8</td>
</tr>
<tr>
<td></td>
<td>G₂.₁⁻ = 3</td>
<td>G₂.₂⁻ = 2 3/4</td>
<td>2 3/4</td>
<td>2 3/4</td>
<td>α₂ = 3/8</td>
</tr>
</tbody>
</table>

(16)

Frequency

\[ 1/2 \quad 1/2 \]

\[ 3 1/2 \quad 2 3/4 \]

\[ G._... = 3 1/8 \]

\[ G._... - G._... = 3 \frac{3}{8} \]

\[ \alpha_{b₁} = 3 \frac{3}{8} \]

\[ \alpha_{b₂} = 3 \frac{3}{8} \]
The genotypic value for $A_iB_k$ is obtained by averaging over the genotypic values represented by all the possible combinations of genes corresponding to the blanks, holding $i$ and $k$ constant. The genotypic value of $A_iB_k$ is the weighted average of the genotypic values of the following genotypes, $A_iA_1B_kB_1$, $A_iA_1B_kB_2$, $A_iA_2B_kB_1$, and $A_iA_2B_kB_2$. Thus,

$$G_{i,k} = p_{A_1B_1}G_{i1k1} + p_{A_1B_2}G_{i1k2} + p_{A_2B_1}G_{i2k1} + p_{A_2B_2}G_{i2k2} \tag{17}$$

For the possible combinations of $i$ and $k$, we have

$$G_{1,1} = \frac{1}{4}G_{1111} + \frac{1}{4}G_{1112} + \frac{1}{4}G_{1211} + \frac{1}{4}G_{1212} = \frac{1}{4}(4 + 4 + 4 + 4) = 4$$
$$G_{1,2} = \frac{1}{4}G_{1121} + \frac{1}{4}G_{1122} + \frac{1}{4}G_{1221} + \frac{1}{4}G_{1222} = \frac{1}{4}(4 + 2 + 4 + 2) = 3 \tag{18}$$
$$G_{2,1} = \frac{1}{4}G_{2111} + \frac{1}{4}G_{2112} + \frac{1}{4}G_{2211} + \frac{1}{4}G_{2212} = \frac{1}{4}(4 + 4 + 2 + 2) = 3$$
$$G_{2,2} = \frac{1}{4}G_{2121} + \frac{1}{4}G_{2122} + \frac{1}{4}G_{2221} + \frac{1}{4}G_{2222} = \frac{1}{4}(4 + 2 + 2 + 2) = 2\frac{1}{2}$$

The deviations of the marginal means from the grand mean for the alleles at the A locus in (16) are the same as those marginal deviations for the two-way table for the A locus (8), and similarly for the B locus. The fitted cell means are

$$\hat{G}_{i,k} = \mu + \alpha_{a_i}^m + \alpha_{b_k}^m \tag{19}$$

For example,

$$\hat{G}_{1,1} = 3\frac{1}{8} + \frac{3}{8} + \frac{3}{8} = 3\frac{7}{8} \tag{20}$$

The variance of the first additive-by-additive term is

$$\sigma_{a_i}^m\sigma_{b_k}^m = \frac{1}{4}(4 - 3\frac{7}{8})^2 + \frac{1}{4}(3 - 3\frac{1}{8})^2 + \frac{1}{4}(3 - 3\frac{7}{8})^2 + \frac{1}{4}(2\frac{1}{2} - 2\frac{3}{8})^2$$
$$= \frac{1}{4}\left(\frac{1}{8}\right)^2 + \frac{1}{4}\left(-\frac{1}{8}\right)^2 + \frac{1}{4}\left(-\frac{1}{8}\right)^2 + \frac{1}{4}\left(\frac{1}{8}\right)^2 = \frac{1}{64} \tag{21}$$

There are three other additive-by-additive terms in (8.128). For each term, a corresponding two-way table would be set up, and the variance of the residuals would be calculated. The values obtained from all such tables would be the same as in (21), because male and female allelic frequencies for both the A and B loci are the same. Hence,
\[
\sigma_{AA}^2 = \sigma_{(aa)_{ab}}^{mm} + \sigma_{(aa)_{ab}}^{mf} + \sigma_{(aa)_{ab}}^{fm} + \sigma_{(aa)_{ab}}^{ff} = 4 \left(\frac{1}{64}\right) = \frac{1}{16}
\]  

(22)

Using the computing formula for the additive-by-additive variance (8.163a), we first compute the \(e_{ij}\)'s (8.161), namely,

\[
\begin{align*}
    e_{11} &= 4 - 4 - 4 + 4 = 0 \\
    e_{12} &= 4 - 2 - 4 + 2 = 0 \\
    e_{21} &= 4 - 4 - 2 + 2 = 0 \\
    e_{22} &= 4 - 2 - 2 + 2 = 2
\end{align*}
\]  

(23)

Then the additive-by-additive variance is

\[
\sigma_{AA_{ab}}^2 = 4p_A p_A p_B p_B (p_A p_B e_{11} + p_A p_B e_{12} + p_A p_B e_{21} + p_A p_B e_{22})^2 \\
= 4(1/2)(1/2)(1/2)(1/2)[(1/2)(1/2)0 + (1/2)(1/2)0 + (1/2)(1/2)0 \\
+ (1/2)(1/2)2]^2 = 1/16
\]  

(24)

**Additive-by-dominance variance.** Turning to the additive-by-dominance term, \(\alpha_{a_b}^m\), it is a three-factor interaction term and is the residual in the following three-way table:
The genotypic value $G_{i\cdot k\ell}$ for every cell is the average over the two different genotypes represented by the dot, weighted by their frequency. For example,

$$G_{1\cdot 11} = \frac{1}{2} G_{1111} + \frac{1}{2} G_{1211} = \frac{1}{2} (4 + 4) = 4$$

Next, the fitted genotypic value is calculated for each cell. It is composed of all lower-order terms -- three main and three two-factor interaction effects, namely,
\[
\hat{G}_{1.11} = \mu + \alpha_{a_1} + \alpha_{b_1} = \frac{1}{8} + 3 \cdot \frac{1}{8} + 3 \cdot \frac{1}{8} + 3 \cdot \frac{1}{8} + 1 \cdot \frac{1}{8} + 1 \cdot \frac{1}{8} + (-3) \cdot \frac{-3}{8} = 4 \cdot \frac{1}{8}
\]
\[
\hat{G}_{1.21} = \mu + \alpha_{a_1} + \alpha_{b_2} = \frac{1}{8} + 3 \cdot \frac{1}{8} + (-3) \cdot \frac{1}{8} + 3 \cdot \frac{1}{8} + (-1) \cdot \frac{1}{8} + 1 \cdot \frac{1}{8} + 3 \cdot \frac{1}{8} = 3 \cdot \frac{7}{8}
\]
\[
\hat{G}_{2.22} = \mu + \alpha_{a_2} + \alpha_{b_2} = \frac{1}{8} + (-3) \cdot \frac{1}{8} + (-3) \cdot \frac{1}{8} + (-3) \cdot \frac{1}{8} + 1 \cdot \frac{1}{8} + 1 \cdot \frac{1}{8} + (-3) \cdot \frac{1}{8} = 1 \cdot \frac{7}{8}
\]

The variance of that additive-by-dominance term in the model is
\[
\sigma_{\alpha a \delta b}^2 = \frac{1}{8} \left( 4 - 4 \cdot \frac{1}{8} \right)^2 + \frac{1}{8} \left( 4 - 3 \cdot \frac{7}{8} \right)^2 + \ldots + \frac{1}{8} \left( 2 - 1 \cdot \frac{7}{8} \right)^2
\]
\[
= \frac{1}{8} \left[ (-\frac{1}{8})^2 + (\frac{1}{8})^2 + \ldots + (\frac{1}{8})^2 \right] = \frac{1}{64}
\]  

There is another additive-by-dominance term in (8.128), and setting up the proper three-way table we obtain the same variance as in (29). Thus, the total additive-by-dominance variance with order considered is
\[
\sigma_{AD}^2 = \sigma_{\alpha a \delta b}^2 + \sigma_{\alpha f \delta b}^2 = \frac{1}{64} + \frac{1}{64} = \frac{1}{32}
\]

Using the computing formula (8.163b) and the above e's (23), we obtain
\[
\sigma_{ADab}^2 = 2p_{A_1}p_{A_2}^2p_{B_1}^2p_{B_2}^2 [p_{A_1}^2 (e_{11} - e_{12}) + p_{A_2}^2 (e_{21} - e_{22})]^2
\]
\[
= 2(1/2)(1/2)(1/2)^2(1/2)^2[(1/2)(0 - 0) + (1/2)(2 - 0)]^2 = 1/32
\]

**Dominance-by-additive variance.** Similarly, for the two dominance-by-additive terms in (8.128), the appropriate three-factor table for each term leads to the same variance as in (29). Thus, the total dominance-by-additive variance with order considered is
\[
\sigma_{DA}^2 = \sigma_{\delta a \alpha b}^2 + \sigma_{\delta f a \alpha b}^2 = \frac{1}{64} + \frac{1}{64} = \frac{1}{32}
\]
Using the computing formula (8.163c), we obtain
\[
\sigma_{DAab}^2 = 2p_{A_1}^2 p_{A_2}^2 p_{B_1}^2 p_{B_2}^2 (e_{11} - e_{21})^2 + p_{B_1}^2 (e_{12} - e_{22})^2
\]
\[= 2(1/2)^2 (1/2)^2 (1/2) [(1/2)(0-0) + (1/2)(2-0)]^2 = 1/32
\] (33)

The total additive-by-dominance without regard to order is
\[
\sigma_{AD}^2 = \sigma_{AD}^2 + \sigma_{DA}^2 = \frac{1}{32} + \frac{1}{32} = \frac{1}{16}
\] (34)

**Dominance-by-dominance variance.** Finally, the four-factor interaction of the dominance-by-dominance effect is the residual in a four-way table. There are \(2^4 = 16\) cells, and the fitted genotypic value for any cell is

\[
\hat{G}_{ijkl} = \mu + \alpha_{a_i} + \alpha_{b_j} + \alpha_{a_i} + \alpha_{b_j}
\]

**six two-factor interaction effects**

\[
\delta_{a_i} + \delta_{b_j} + (\alpha\alpha)_{a_i} m_{a_i} + (\alpha\alpha)_{b_j} m_{b_j} + (\alpha\alpha)_{a_i} m_{a_i} + (\alpha\alpha)_{b_j} m_{b_j}
\]

**four three-factor interaction effects**

\[
(\alpha\delta)_{a_i} m_{a_i} m_{b_j} + (\alpha\delta)_{b_j} m_{b_j} m_{a_j} + (\delta\alpha)_{a_i} m_{a_i} m_{b_j} + (\delta\alpha)_{b_j} m_{b_j} m_{a_j}
\] (35)

The fitted value for two of the sixteen cells, for example, from (35) is

\[
\hat{G}_{1111} = 3 \frac{1}{8} + 3 \frac{3}{8} + 3 \frac{3}{8} + 3 \frac{3}{8} + 3 \frac{3}{8} + (-3) + 1 \frac{1}{8} + 1 \frac{1}{8} + 1 \frac{1}{8}
\]
\[+ (-\frac{1}{8}) + (-\frac{1}{8}) + (-\frac{1}{8}) + (-\frac{1}{8}) = 3 \frac{7}{8}
\] (36)

\[
\hat{G}_{1221} = 3 \frac{1}{8} + 3 \frac{3}{8} + (-\frac{3}{8}) + (-\frac{3}{8}) + 3 \frac{3}{8} + 3 \frac{3}{8} + 3 \frac{3}{8} + (-\frac{1}{8}) + 1 \frac{1}{8} + 1 \frac{1}{8} + (-\frac{1}{8})
\]
\[+ 1 \frac{1}{8} + (-\frac{1}{8}) + (-\frac{1}{8}) + 1 \frac{1}{8} = 3 \frac{7}{8}
\]

All sixteen fitted values are

\[
\hat{G}_{1111} = 3 \frac{7}{8} \quad \hat{G}_{1112} = 4 \frac{1}{8} \quad \hat{G}_{1121} = 4 \frac{1}{8} \quad \hat{G}_{1122} = 1 \frac{7}{8}
\]

\[
\hat{G}_{1211} = 4 \frac{1}{8} \quad \hat{G}_{1212} = 3 \frac{7}{8} \quad \hat{G}_{1221} = 3 \frac{7}{8} \quad \hat{G}_{1222} = 2 \frac{1}{8}
\] (37)
\[
\hat{G}_{2111} = 4 \frac{1}{8} \quad \hat{G}_{2112} = 3 \frac{7}{8} \quad \hat{G}_{2121} = 3 \frac{7}{8} \quad \hat{G}_{2122} = 2 \frac{1}{8} \\
\hat{G}_{2211} = 1 \frac{7}{8} \quad \hat{G}_{2212} = 2 \frac{1}{8} \quad \hat{G}_{2221} = 2 \frac{1}{8} \quad \hat{G}_{2222} = 1 \frac{7}{8}
\]

Some genotypic values are 4, and others are only 2 (1). In either case, every cell deviation \((G_{ijkl} - \hat{G}_{ijkl})\) equals a positive or negative one-eighth. It is noted that the sum of the deviations over any two cells for which three subscripts are common equals zero \((8.141)\). The dominance-by-dominance variance is the sum of the squares of every deviation, each weighted by its frequency. Thus,

\[
\sigma_{DD}^2 = \delta_{ab}^2 = 16 \left( \frac{1}{16} \right) \left( \pm \frac{1}{8} \right)^2 = \frac{1}{64} \quad (38)
\]

Using the computing formula \((8.163d)\), we obtain

\[
\sigma_{DD,ab}^2 = p_{A1}^2 p_{A2}^2 p_{B1}^2 p_{B2}^2 (e_{11} - e_{12} - e_{21} + e_{22})^2 \\
= (1/2)^2 (1/2)^2 (1/2)^2 (1/2)^2 (0 - 0 - 0 + 2)^2 = 1/64 \quad (39)
\]

Check: Note that the sum of the four partitions of the epistatic variance from \((22), (30), (32)\) and \((38)\) does equal the epistatic variance calculated in \((6)\) and \((7)\)

\[
\sigma_{IY,ab}^2 = \sigma_{AA}^2 + \sigma_{AD}^2 + \sigma_{DA}^2 + \sigma_{DD}^2 = \frac{1}{16} + \frac{1}{32} + \frac{1}{32} + \frac{1}{64} = \frac{9}{64} \quad (40)
\]

Note that both the additive-by-additive and the total additive-by-dominance variances constitute 6.35% of the total genotypic variance, and the dominance-by-dominance variance 1.59%.

Check: Likewise the sum of all components \((9), (12), (15), (22), (30), (32), (38)\) equals the genotypic variance, i.e.,

\[
\sigma_G^2 = \sigma_{A_a}^2 + \sigma_{D_a}^2 + \sigma_{A_b}^2 + \sigma_{D_b}^2 + \sigma_{AA,ab}^2 + \sigma_{AD,ab}^2 + \sigma_{DA,ab}^2 + \sigma_{DD,ab}^2 \\
= 9/32 + 9/64 + 9/32 + 9/64 + 1/16 + 1/32 + 1/32 + 1/64 \\
= 63/64 \quad (41)
\]

which is the same as \((2)\).
**Example 8.6.** Consider the same situation as in Example 8.5 except with \( p_{A_1} = 0.6, p_{A_2} = 0.4, p_{B_1} = 0.7 \) and \( p_{B_2} = 0.3 \). Only computational formulas will be used.

<table>
<thead>
<tr>
<th>B locus</th>
<th>( B_{1B_1} )</th>
<th>( B_{1B_2} )</th>
<th>( B_{2B_2} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_{1A_1} )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>.1764</td>
<td>.1512</td>
<td>.0324</td>
<td></td>
</tr>
<tr>
<td>( A_{1A_2} )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>.2352</td>
<td>.2016</td>
<td>.0432</td>
<td></td>
</tr>
<tr>
<td>( A_{2A_2} )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>.0784</td>
<td>.0672</td>
<td>.0144</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 2 p_{A_1} )</td>
<td>( p_2 )</td>
</tr>
<tr>
<td>( = .36 )</td>
<td>( = 3.82 )</td>
</tr>
</tbody>
</table>

The variances are as follows (equation used is listed on the left):

\[
\sigma_G^2 = (0.1764 + 0.1512 + 0.2352 + 0.2016)(4)^2 \]
\[
+ (0.0324 + 0.0432 + 0.0784 + 0.0672 + 0.0144)(2)^2 \]
\[
- [(0.1764 + \ldots + 0.2016)4 + (0.0324 + \ldots + 0.0144)2]^2 \]
\[
= 0.720 \quad 370 \quad 56 \]

\[
\sigma_{A_a}^2 = 2(0.6)(0.4)[(0.6)(3.82 - 3.82) + 0.4(3.82 - 2)]^2 = 0.254 \quad 392 \quad 32 \]

\[
\sigma_{D_a}^2 = (0.6)^2(0.4)^2[2(3.82) - 3.82 - 2]^2 = 0.190 \quad 794 \quad 24 \]

\[
\sigma_{A_b}^2 = 2(0.7)(0.3)[(0.7)(3.68 - 3.68) + (0.3)(3.68 - 2)]^2 = 0.106 \quad 686 \quad 72 \]
(8.52) \[ \sigma_{DB}^2 = (0.7)^2(0.3)^2[2(3.68) - 3.68 - 2]^2 = 0.124\ 467\ 84 \]

(8.163a) \[ \sigma_{AA_{ab}}^2 = 4(0.6)(0.4)(0.7)(0.3)[0 + 0 + 0 + (0.4)(0.3)]^2 = 0.011\ 612\ 16 \]

(8.163b) \[ \sigma_{AD_{ab}}^2 = 2(0.6)(0.4)(0.7)^2(0.3)^2[0 + (0.4)(0 - 2)]^2 = 0.013\ 547\ 52 \]

(8.163c) \[ \sigma_{DA_{ab}}^2 = 2(0.6)^2(0.4)^2(0.7)(0.3)[0 + 0.3(0 - 2)]^2 = 0.008\ 709\ 12 \]

(8.163d) \[ \sigma_{DD_{ab}}^2 = (0.6)^2(0.4)^2(0.7)(0.3)^2[0 - 0 - 0 + 2]^2 = 0.010\ 160\ 64 \]

Check:

Variance Value Proportion

| \sigma_{A_a}^2 | 0.254 392 32 | 0.3531 |
| \sigma_{D_a}^2 | 0.190 794 24 | 0.2649 | 0.5012 \( \alpha \sigma_{A_a}^2 \)
| \sigma_{A_b}^2 | 0.106 686 72 | 0.315 262 08 | \( \sigma_{D_b}^2 \) = 0.1481 | 0.4376 \( \alpha \sigma_{D_a}^2 \)
| \sigma_{D_b}^2 | 0.124 467 84 | 0.1728 |
| \sigma_{AA_{ab}}^2 | 0.011 612 16 | \( \sigma_{A_a}^2 \) = 0.0161 \( \alpha \sigma_{AA_{ab}}^2 \)
| \sigma_{AD_{ab}}^2 | 0.013 547 52 | 0.022 256 64 | \( \sigma_{AD_{ab}}^2 \) = 0.0188 | 0.0309 \( \alpha \sigma_{AD_{ab}}^2 \)
| \sigma_{DA_{ab}}^2 | 0.008 709 12 | 0.0121 |
| \sigma_{DD_{ab}}^2 | 0.010 160 64 | \( \sigma_{DD_{ab}}^2 \) = 0.0141 | 1.0000 |
| \sigma_{G}^2 | 0.720 370 56 | 0.0611 \( \alpha \sigma_{I}^2 \)

8.3.4. Two loci, two alleles, equal male and female allelic frequencies at each locus (Hardy-Weinberg), linkage disequilibrium, epistasis. Effect of linkage disequilibrium. To illustrate the effects of linkage disequilibrium on the means and variances, consider the same model as in the previous section, but with gametic frequencies \( p_{A_1B_k} = p_{A_1}p_{B_k} + A_{1B_k} \) (3.63), i.e., we now consider the model with two loci, two alleles per locus, random mating, linkage disequilibrium, and epistasis. The gametic frequencies can be expressed as \( p_{A_1B_k} = p_{A_1}p_{B_k} + (-1)^{i+k}A \)
in that the linkage disequilibrium for each gamete is (3.65)

\[
\begin{align*}
P_{A_1B_1} - P_{A_1B_1} A_{11} - \Delta \\
P_{A_1B_2} - P_{A_1B_2} A_{12} - \Delta \\
P_{A_2B_1} - P_{A_2B_1} A_{21} - \Delta \\
P_{A_2B_2} - P_{A_2B_2} A_{22} - \Delta
\end{align*}
\]  \hspace{1cm} (8.166)

The gametic frequencies are assumed to be equal in the two sexes, so no
distinction between male and female gametes needs to be made, i.e., \( p_{11}^{mm} = p_{11}^{ff} \)
etc. The same subscript identification of the G's in (8.157) is maintained in the
following factorial arrangement of the genotype in terms of uniting gametes.

**Female gametes**

<table>
<thead>
<tr>
<th>( A_1B_1 )</th>
<th>( A_1B_2 )</th>
<th>( A_2B_1 )</th>
<th>( A_2B_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( ff ) ( P_{11} )</td>
<td>( ff ) ( P_{12} )</td>
<td>( ff ) ( P_{21} )</td>
<td>( ff ) ( P_{22} )</td>
</tr>
<tr>
<td>( G_{1111} ) ( P_{11}^{mm} )</td>
<td>( G_{1112} ) ( P_{11}^{mm} )</td>
<td>( G_{1211} ) ( P_{11}^{mm} )</td>
<td>( G_{1212} ) ( P_{11}^{mm} )</td>
</tr>
<tr>
<td>( G_{1121} ) ( P_{11}^{mm} )</td>
<td>( G_{1122} ) ( P_{11}^{mm} )</td>
<td>( G_{1221} ) ( P_{11}^{mm} )</td>
<td>( G_{1222} ) ( P_{11}^{mm} )</td>
</tr>
<tr>
<td>( mm ) ( P_{11} )</td>
<td>( mm ) ( P_{11} )</td>
<td>( mm ) ( P_{11} )</td>
<td>( mm ) ( P_{11} )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( A_1B_2 )</th>
<th>( mm ) ( P_{12} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( G_{1121} ) ( P_{12}^{mm} )</td>
<td>( G_{1122} ) ( P_{12}^{mm} )</td>
</tr>
<tr>
<td>( G_{1221} ) ( P_{12}^{mm} )</td>
<td>( G_{1222} ) ( P_{12}^{mm} )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( A_2B_1 )</th>
<th>( mm ) ( P_{21} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( G_{2111} ) ( P_{21}^{mm} )</td>
<td>( G_{2112} ) ( P_{21}^{mm} )</td>
</tr>
<tr>
<td>( G_{2211} ) ( P_{21}^{mm} )</td>
<td>( G_{2212} ) ( P_{21}^{mm} )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( A_2B_2 )</th>
<th>( mm ) ( P_{22} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( G_{2121} ) ( P_{22}^{mm} )</td>
<td>( G_{2122} ) ( P_{22}^{mm} )</td>
</tr>
<tr>
<td>( G_{2221} ) ( P_{22}^{mm} )</td>
<td>( G_{2222} ) ( P_{22}^{mm} )</td>
</tr>
</tbody>
</table>

| Frequency | \( P_{11} \) | \( P_{12} \) | \( P_{21} \) | \( P_{22} \) |
|----------------|----------------|----------------|----------------|
| \( =p_{A_1B_1} + \Delta \) | \( =p_{A_1B_1} \) + \( \Delta \) | \( =p_{A_1B_2} \) - \( \Delta \) | \( =p_{A_2B_2} \) + \( \Delta \) |

\[
\begin{align*}
\text{Mean} & \quad G_{1\cdot1} \quad G_{1\cdot2} \quad G_{2\cdot1} \quad G_{2\cdot2} \\
\text{Frequency} & \quad P_{11} \quad P_{12} \quad P_{21} \quad P_{22}
\end{align*}
\]  \hspace{1cm} (8.167)

(\text{each column mean equals corresponding row mean})

\[
\begin{align*}
\text{Mean} & \quad G_{1\cdot1} \quad G_{1\cdot2} \quad G_{2\cdot1} \quad G_{2\cdot2} \\
\text{Frequency} & \quad P_{11} \quad P_{12} \quad P_{21} \quad P_{22}
\end{align*}
\]  \hspace{1cm} (8.171)
Under random mating the frequency of each genotype, \( p_{ijkl} \), is the product of the frequency of its gametes, i.e.,

\[
p_{ijkl} = p_{ik}p_{jk} = p_{ik}p_{jk} = [p_{A_i}p_{B_k} + (-1)^{i+k}\Delta] [p_{A_j}p_{B_l} + (-1)^{j+l}\Delta]
\]  
(8.168)

First, we want to look at the marginal means and the overall mean. Each row (or column) mean is a weighted mean of the genotypic values in that row, where the weights are the frequencies of the gametes (analogous to (8.12), Table 8.1, (8.30)). Hence, the genotypic mean of the row (or column) associated with gamete \( A_1B_1 \) is

\[
\hat{G}_{1.1} = (p_{A_1}p_{B_1} + \Delta)G_{1111} + (p_{A_1}p_{B_2} - \Delta)G_{1112} + (p_{A_2}p_{B_1} - \Delta)G_{1211} + (p_{A_2}p_{B_2} + \Delta)G_{1212}
\]

\[
= p_{A_1}p_{B_1}G_{1111} + p_{A_1}p_{B_2}G_{1112} + p_{A_2}p_{B_1}G_{1211} + p_{A_2}p_{B_2}G_{1212} + \Delta(G_{1111} - G_{1112} - G_{1211} + G_{1212}) \quad \text{(sub. (8.161))}
\]

\[
= \hat{G}_{1.1} + \Delta e_{11}
\]  
(8.169)

where \( \hat{G}_{1.1} \) = mean genotypic value of all genotypes that carry the male gamete \( A_1B_1 \); the gametic genotypic mean or value,

\( \hat{G}_{1.1} \) = mean genotypic value associated with male gamete \( A_1B_1 \) for a random mating population in linkage equilibrium,

\( e_{11} \) = interaction contrast (8.161).

Or, in general, for the \( ik \)th row, we have

\[
\hat{G}_{i.1} = \hat{G}_{i.1} + (-1)^{i+k}\Delta e_{ik} = \hat{G}_{i.1} + \Delta e_{ik}
\]  
(8.170)

Notice that the gametic genotypic mean deviates from that under random mating and linkage equilibrium by the product of its gametic linkage disequilibrium value and the corresponding \( e \) value. Averaging over all row means, we obtain the overall mean (see Box 8.11)

\[
\hat{G}_{..} = \hat{G}_{..} + 2\Delta \sum \sum p_{A_i}p_{B_k} e_{ik} + \Delta^2(e_{11} - e_{12} - e_{21} + e_{22})
\]  
(8.171)
where \( \hat{G} \ldots \) = overall mean of population in linkage equilibrium and random mating,

\[
\Sigma \Sigma p_{A_i} p_{B_k} e_{ik} = \text{additive-by-additive contrast (8.163a)},
\]

\[
e_{11} - e_{12} - e_{21} + e_{22} = \text{dominance-by-dominance contrast (8.163d)}.
\]

Box 8.11

Derivation of (8.171)

Weighting each marginal mean by its frequency, we have

\[
\hat{G} \ldots = (p_{A_1} p_{B_1} + \Delta)(\hat{\Delta}_1 + \Delta e_{11}) + (p_{A_1} p_{B_2} - \Delta)(\hat{\Delta}_2 - \Delta e_{12})
\]

\[
+ (p_{A_2} p_{B_1} - \Delta)(\hat{\Delta}_1 - \Delta e_{21}) + (p_{A_2} p_{B_2} + \Delta)(\hat{\Delta}_2 + \Delta e_{22})
\]

\[
= p_{A_1} p_{B_1} \hat{G}_{11}^1 + p_{A_1} p_{B_2} \hat{G}_{11}^2 + p_{A_2} p_{B_1} \hat{G}_{21}^1 + p_{A_2} p_{B_2} \hat{G}_{21}^2
\]

\[
+ p_{A_1} p_{B_1} \Delta e_{11} + p_{A_1} p_{B_2} \Delta e_{12} + p_{A_2} p_{B_1} \Delta e_{21} + p_{A_2} p_{B_2} \Delta e_{22}
\]

\[
+ \Delta (\hat{G}_1 - \hat{G}_2 - \hat{G}_1 - \hat{G}_2)
\]

\[
+ \Delta^2 (e_{11} - e_{12} - e_{21} + e_{22})
\]

\[
= \hat{G} \ldots + \Delta \Sigma \Sigma p_{A_i} p_{B_k} e_{ik}
\]

\[
+ \Delta^2 (e_{11} - e_{12} - e_{21} + e_{22})
\]

(see 8.161)
\[ + \Delta [p_{A_1}p_{B_1} e_{11} + p_{A_1}p_{B_2} e_{12} + p_{A_2}p_{B_1} e_{21} + p_{A_2}p_{B_2} e_{22}] \\
+ \Delta^2 (e_{11} - e_{12} - e_{21} + e_{22}) \\
= \hat{G} + 2 \Delta \sum \sum p_{A_i} p_{B_k} e_{ik} + \Delta^2 (e_{11} - e_{12} - e_{21} + e_{22}) \]

which is (8.171).

Thus, from (8.170) and (8.171) we see that the marginal (haplotypic or "gametic") means and the overall mean are affected only if both linkage disequilibrium and epistasis (one or more nonzero $e_{ik}$) are present. This is an illustration of a general result: Linkage disequilibrium affects the "gametic" and overall means of a population only in conjunction with epistatic effects. (Relate this to the mean and variance of a linear function, where the mean is not affected when variables are correlated. Why is the mean affected?)

Secondly, we look at the marginal or gametic genetic variance, which is the variance among the gametic means. The same above conditions for the population mean do not hold for gametic genotype variance, however. Linkage disequilibrium always affects the marginal (haplotypic or "gametic") variance, whether epistasis is present or not. With linkage disequilibrium and epistasis either the male or female gametic genotypic variance is

\[ \sigma^2_g = \sum \sum p_{A_i} p_{B_k} (G_{i \cdot k} - G_{\ldots})^2 \]

\[ = [p_{A_i} p_{B_k} + (-1)^{i+k} \Delta] (G_{i \cdot k} - G_{\ldots})^2 \]

(8.172)

and is a function of linkage disequilibrium ($\Delta$) and epistasis ($e_{ik}$'s) in that $(G_{i \cdot k} - G_{\ldots})$ contains $e_{ik}$'s (see (8.170) and (8.171)). Both linkage disequilibrium and epistasis affect the gametic variance.

If one assumes only linkage disequilibrium and no epistasis, (8.172) becomes (see Box 8.12)
\[ \sigma^2_g = \sigma^2_a + \sigma^2_b + \frac{2\Delta}{\pi A_1 P_{A_2} P_{B_1} P_{B_2}} \sigma_\alpha \sigma_\beta + \sigma^2_\alpha + \sigma^2_\beta + 2 \text{Cov}(\alpha, \beta) \quad (8.173) \]

where \[ \frac{\Delta}{\pi A_1 P_{A_2} P_{B_1} P_{B_2}} \sigma_\alpha \sigma_\beta = \text{covariance between the additive effects, } \alpha \text{ and } \beta, \]
at the A and B loci due to linkage disequilibrium,

\[ \sigma_\alpha = \text{square root of the variance of } \sigma^2_a \quad (8.149a), \]

\[ \sigma_\beta = \text{square root of the variance of } \sigma^2_\beta \quad (8.149b). \]

Even though (8.149a) and (8.149b) are definitions given in the section with epistasis, the definitions are the same as those for Section 8.1.7. The covariance between the additive effects, given in (8.173), is an explicit function in terms of the linkage disequilibrium itself, and is that covariance noted in (8.100) to be part of the gametic variance.

For the case when no linkage disequilibrium exists, we have the following. If epistasis (additive-by-additive) is present, the gametic variance decomposes exactly into the components for gene effects in the model (8.131), namely,

\[ \sigma^2_g = \sigma^2_a + \sigma^2_b + \sigma^2_{(aa)_{ab}} \quad (8.174) \]

If epistasis is absent, the same exists as above (8.174), except no variance term for the additive-by-additive effect exists.

---

**Box 8.12**

**Derivation of (8.173)**

To derive (8.173), we must determine what the gametic variance (8.172) is under the assumption of linkage disequilibrium, but with no epistasis, and random mating. First, we must express \( G_{i \cdot k} \cdot - G_{\cdot \cdot} \) in terms of the genic factorial model (8.128). We note that with no epistasis all \( e_{ik} = 0 \), so (8.170) becomes \( G_{i \cdot k} = \hat{G}_{i \cdot k} \), and (8.171) becomes \( G_{\cdot \cdot} = \hat{G}_{\cdot \cdot} \) which are the values expected under linkage equilibrium and random mating. Hence with linkage equilibrium,
random mating, and no epistasis, we then show formally that \( G_{i \cdot \cdot} = \sum_{j=1}^{2} p_{A_j} p_{B_j} G_{i j k l} - G_{\cdot \cdot \cdot} \)

\[
G_{i \cdot \cdot} = G_{i \cdot \cdot} - G_{\cdot \cdot \cdot} = \sum_{j=1}^{2} \sum_{l=1}^{2} p_{A_j} p_{B_l} G_{i j k l} - G_{\cdot \cdot \cdot}
\]

\[
=- \sum_{j=1}^{2} \sum_{l=1}^{2} p_{A_j} p_{B_l} \left( \mu + \alpha_{a_i}^m + \alpha_{a_j}^f + \delta_{a_i} + \alpha_{b_k}^m + \alpha_{b_l}^f + \delta_{b_k l} \right) - G_{\cdot \cdot \cdot}
\]

\[
= \left( \sum_{j} p_{A_j} \right) \left( \sum_{l} p_{B_l} \right) \mu + \alpha_{a_i}^m \left( \sum_{j} p_{A_j} \right) \left( \sum_{l} p_{B_l} \right) + \alpha_{b_k}^m \left( \sum_{l} p_{B_l} \right) \left( \sum_{j} p_{A_j} \right) + \left( \sum_{j} p_{A_j} \right) \left( \sum_{l} p_{B_l} \right) \alpha_{a_j}^f \left( \sum_{j} p_{A_j} \right) \left( \sum_{l} p_{B_l} \right)
\]

\[
= \alpha_{a_i}^m + \alpha_{b_k}^m
\]

Then, substituting (1) in (8.172), we have

\[
\sigma^2_g = \sum_{i k} \left[ p_{A_i} p_{B_k} + (-1)^{i+k} \right] \left[ (\alpha_{a_i}^m)^2 + (\alpha_{b_k}^m)^2 \right]
\]

\[
= \sum_{i k} \left[ p_{A_i} p_{B_k} + (-1)^{i+k} \right] \left[ (\alpha_{a_i}^m)^2 + (\alpha_{b_k}^m)^2 + 2(\alpha_{a_i}^m)(\alpha_{b_k}^m) \right]
\]

\[
= \sum_{i k} \left[ p_{A_i} \left( \alpha_{a_i}^m \right)^2 \left( \sum_{k} p_{B_k} \right) + \Delta \sum_{i} \left( \alpha_{a_i}^m \right)^2 \left( \sum_{k} (-1)^{i+k} \right) \right]
\]

\[
+ \sum_{i k} \left[ p_{B_k} \left( \alpha_{b_k}^m \right)^2 \left( \sum_{i} p_{A_i} \right) + \Delta \sum_{k} \left( \alpha_{b_k}^m \right)^2 \left( \sum_{i} (-1)^{i+k} \right) \right]
\]

\[
+ 2 \left( \sum_{i} p_{A_i} \alpha_{a_i}^m \right) \left( \sum_{k} p_{B_k} \alpha_{b_k}^m \right) + 2 \Delta \sum_{i} \sum_{k} (-1)^{i+k} (\alpha_{a_i}^m)(\alpha_{b_k}^m)
\]

(2)
We observe that for every $i$ (and similarly, for every $k$)

$$
\begin{align*}
\sum_{k=1}^{2} (-1)^{i+k} &= 1 - 1 = -1 + 1 = 0 \\
&\quad \text{for } i = 1, 2
\end{align*}
$$

(3)

Then substituting (8.7), (3), $\Sigma p_{A_i} = \Sigma p_{B_k} = 1$, and (8.18) in (2), we have

$$
\begin{align*}
\sigma_g &= \sigma_a^2 + \sigma_b^2 + 2\Delta \sum_{i=1}^{2} \sum_{k=1}^{2} (-1)^{i+k} \sigma_m \left( \sigma_m \right) \\
&= \sigma_a^2 + \sigma_b^2 + 2\Delta \left[ \sigma_m \sigma_m - \sigma_m \sigma_m - \sigma_m \sigma_m + \sigma_m \sigma_m \right] \\
&= \sigma_a^2 + \sigma_b^2 + 2\Delta \left[ \sigma_m \sigma_m - \sigma_m \sigma_m - \sigma_m \sigma_m + \sigma_m \sigma_m \right] \\
&= \sigma_a^2 + \sigma_b^2 + 2\Delta \left[ \sigma_m - \sigma_m \right] \left( \sigma_m - \sigma_m \right)
\end{align*}
$$

(4)

Using (8.51), we may write for locus A

$$
2\sigma_a^2 = 2p_{A_1}p_{A_2}\sigma_A^2 \quad \text{(sub. (8.46))}
$$

$$
= 2p_{A_1}p_{A_2} (\sigma_{a_1} - \sigma_{a_2})^2
$$

$$
(\sigma_{a_1} - \sigma_{a_2}) = \frac{\sigma_a}{\sqrt{p_{A_1}p_{A_2}}}
$$

(5)

Similarly, for locus B

$$
(\sigma_{b_1} - \sigma_{b_2}) = \frac{\sigma_b}{\sqrt{p_{B_1}p_{B_2}}}
$$

(6)

Substituting (5) and (6) in (4), we have

$$
\sigma_g^2 = \sigma_a^2 + \sigma_b^2 + 2\Delta \frac{\sigma_a}{\sqrt{p_{A_1}p_{A_2}p_{B_1}p_{B_2}}} \frac{\sigma_b}{\sqrt{p_{A_1}p_{A_2}p_{B_1}p_{B_2}}}
$$

which is (8.173).
8.3.5. More than two loci, multiple alleles, arbitrary "male" and "female" allelic frequencies (cross between two populations), epistasis. General. 1. Linkage equilibrium. Extension of the genic factorial model (8.128) to n loci with all possible interactions is straightforward but tedious. This section and the next one can be omitted, at least, upon first reading, if desired. The justification for its inclusion here is that it does deal with the real world of more than two loci, and thereby fully accounts for the genotypic value. Proceeding in a manner analogous to (8.92), but using the model with epistasis (8.128), we have

\[ G_{i_1, j_1, i_2, j_2, \ldots, i_n, j_n} = \prod_{k=1}^{n} k_{i_k} k_{j_k} = \mu + \sum_{k=1}^{n} (\alpha_{i_k} + \beta_{j_k}) + \delta_{i_k j_k} \]

\[ N \binom{n}{2} \text{ AA terms} \]

\[ + \sum_{k=1}^{n} \sum_{k' < k} \left[ \alpha_{i_k} \alpha_{i_{k'}} + \alpha_{i_k} \beta_{j_{k'}} + \beta_{j_k} \alpha_{i_{k'}} + \beta_{j_k} \beta_{j_{k'}} \right] \]

\[ n(n - 1) \binom{n}{2} \text{ AD terms} \]

\[ + \sum_{k=1}^{n} \sum_{k' = k}^{n} \left[ \alpha_{i_k} \beta_{j_k} \alpha_{i_{k'}} + \alpha_{i_k} \beta_{j_k} \beta_{j_{k'}} + \beta_{j_k} \alpha_{i_{k'}} + \beta_{j_k} \beta_{j_{k'}} \right] \]

\[ \sum_{k=1}^{n} \sum_{k' = k}^{n} \left[ \delta_{i_k j_k} \delta_{i_{k'} j_{k'}} \right] \]

All possible epistatic effects between two loci
\[ \binom{n}{4} 2^4 \text{ AAAA terms} \]
\[ + \sum \sum \sum \sum \left[ 2^4 = 16 \text{ terms of (aaa\alpha) type} \right] \]
\[ k \ k' \ k'' \ k''' \]
\[ 1 = k < k' < k'' < k''' \]

\[ \binom{n}{3} (n - 3) 2^3 \text{ AAAD terms} \]
\[ + \sum \sum \sum \sum \left[ 2^3 = 8 \text{ terms of (aaa\delta) type} \right] \]
\[ k \ k' \ k'' \ k''' \]
\[ 1 = k < k' < k'' \]
\[ k''' \not\approx k, k', k'' \]

\[ \binom{n}{2} (n - 2) 2^2 \text{ AADD terms} \]
\[ + \sum \sum \sum \sum \left[ 2^2 = 4 \text{ terms of (aa\delta\delta) type} \right] \]
\[ k \ k' \ k'' \ k''' \]
\[ 1 = k < k' \ 1 = k'' < k''' \]
\[ k''' \not\approx k, k' \]
\[ k'' \not\approx k, k' \]

\[ n \binom{n}{3} 2 \text{ ADDD terms} \]

\[ + \sum \sum \sum \sum \left[ 2 \text{ terms of (a\delta\delta\delta) type} \right] \]
\[ k = 1 \ k' \ k'' \ k''' \]
\[ 1 = k < k'' < k''' \]
\[ k = k', k'', k''' \]

\[ \binom{n}{4} \text{ DDDD terms} \]

\[ + \sum \sum \sum \sum \left[ \text{ one term of (\delta\delta\delta\delta) type} \right] \]
\[ k \ k' \ k'' \ k''' \]
\[ 1 = k < k' < k'' \]

\[ + \ldots \quad (8.175) \]

* All possible epistatic effects between three loci

All terms involving the same number of loci can be grouped into types. For example, for one locus, there are additive (A or \( \alpha \)) and dominance (D or \( \delta \)) terms; for two loci, additive-by-additive (AA or \( \alpha \alpha \)), additive-by-dominance (AD or \( \alpha \delta \) and \( \delta \alpha \)), and dominance-by-dominance (DD or \( \delta \delta \)), and so on. Thus, when \( q \) loci are involved in the effects, there are \( q + 1 \) types. Table 8.4 summarizes the general expressions for the number of terms of each type of effect in the genic factorial model of an individual. When \( q \), the number of loci involved in an epistatic
interaction effect, equals \( n \), the genotypic model would be complete, i.e., the genotype value would be completely accounted for by the terms on the right-hand side.

Table 8.4. General expressions for the number of terms of each type of effect in the genic factorial model of \( n \) loci.

<table>
<thead>
<tr>
<th>Type of effect</th>
<th>Number of terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A ) ( n(2) )</td>
<td>( = \binom{n}{2} = n(2+1) = n(2+1)^1 = n(3) )</td>
</tr>
<tr>
<td>( D ) ( n )</td>
<td>( = \binom{n}{1} )</td>
</tr>
<tr>
<td>( AA ) ( \binom{n}{2}^4 )</td>
<td>( = \binom{n}{2}^2 )</td>
</tr>
<tr>
<td>( AD ) ( n(n-1)2 )</td>
<td>( = \binom{n}{2}(2)^2 = \binom{n}{2}(4+4+1) = \binom{n}{2}(2+1)^2 = \binom{n}{2}^3 )</td>
</tr>
<tr>
<td>( DD ) ( \binom{n}{2} )</td>
<td>( = \binom{n}{2} )</td>
</tr>
<tr>
<td>( AAA ) ( \binom{n}{3}^2^3 )</td>
<td>( = \binom{n}{3}^2^3 )</td>
</tr>
<tr>
<td>( AAD ) ( \binom{n}{2}(n-2)2^2 )</td>
<td>( = \binom{n}{3}(3)^2^2 = \binom{n}{3}(8+12+6+1) = \binom{n}{3}(2+1)^3 = \binom{n}{3}^3 )</td>
</tr>
<tr>
<td>( ADD ) ( \binom{n}{1}(n-1)2 )</td>
<td>( = \binom{n}{3}(3)^2 )</td>
</tr>
<tr>
<td>( DDD ) ( \binom{n}{3} )</td>
<td>( = \binom{n}{3} )</td>
</tr>
<tr>
<td>( AAAAA ) ( \binom{n}{4}2^4 )</td>
<td>( = \binom{n}{4}2^4 )</td>
</tr>
<tr>
<td>( AAAD ) ( \binom{n}{3}(n-3)2^3 )</td>
<td>( = \binom{n}{4}(4)2^3 )</td>
</tr>
<tr>
<td>( AADD ) ( \binom{n}{2}(n-2)2^2 )</td>
<td>( = \binom{n}{4}(6)2^2 = \binom{n}{4}(16+32+24+8+1) )</td>
</tr>
<tr>
<td>( ADDD ) ( \binom{n}{1}(n-1)2 )</td>
<td>( = \binom{n}{4}(4)2 = \binom{n}{4}(2+1)^4 = \binom{n}{4}^3 )</td>
</tr>
<tr>
<td>( DDDD ) ( \binom{n}{4} )</td>
<td>( = \binom{n}{4} )</td>
</tr>
</tbody>
</table>
\[
\begin{align*}
A^q & \quad \binom{n}{q}2^q = \binom{n}{q}2^q \\
A^q_{-1D} & \quad \binom{n}{q-1}(n-(q-1))2^{q-1} = \binom{n}{q-1}2^{q-1} \\
A^q_{-2D^2} & \quad \binom{n}{q-2}(n-(q-2))2^{q-2} = \binom{n}{q-2}2^{q-2} \\
A^q_{-3D^3} & \quad \binom{n}{q-3}(n-(q-3))2^{q-3} = \binom{n}{q-3}2^{q-3} \\
& \quad \vdots \\
A^q_{-sD^s} & \quad \binom{n}{q-s}(n-(q-s))2^{q-s} = \binom{n}{q-s}2^{q-s} = \binom{n}{q}(2+1)^s = \binom{n}{q}3^q \\
& \quad \vdots \\
A^q_{D^q-3} & \quad \binom{n}{q-3}2^3 = \binom{n}{q}2^3 \\
A^q_{D^q-2} & \quad \binom{n}{q-2}2^2 = \binom{n}{q}2^2 \\
A^q_{D^q-1} & \quad \binom{n}{q}2 = \binom{n}{q}2 \\
A^q_{D^q} & \quad \binom{n}{q} = \binom{n}{q} \\
& \quad \vdots \\
A^n & \quad \binom{n}{2^n} = \binom{n}{2^n} \\
A^n_{-1D} & \quad \binom{n}{2^{n-1}} = \binom{n}{2^{n-1}} \\
A^n_{-2D^2} & \quad \binom{n}{2^{n-2}} = \binom{n}{2^{n-2}} = \binom{n}{2^{n-2}}(2+1)^n = 3^n \\
& \quad \vdots \\
\end{align*}
\]
\[ A^{n-1} \binom{n}{1}^2 = \binom{n}{n}^2 \]
\[ D^n \binom{n}{0}2^0 = 1 = \binom{n}{n} = 1 \]

Total \[ 2^{2n} - 1 \]

* Letter A denotes additive; D denotes dominance. The number of A and D letters in the series or their superscript denotes the number of additive and dominance words in the effect, e.g., AAD = A^2D means "additive-by-additive-by-dominance".

\( q = \) number of loci involved in epistatic interaction effect,
\( s = \) number of loci at which both genes at locus (dominance) are involved in interaction.

Table 8.5 gives numerical values for the number of terms of each type of effect for different values of \( n \). It is seen that as the number of loci increases, the number of terms increases very rapidly. Even grouping the variances of individual terms according to their type, the number of variances or parameters in the model still remains large. This is one of the disadvantages of this multi-locus, genic factorial model which, in many ways, has several nice properties and appears to be general. What is done in practice is to estimate various sets of lower-order variances separately, and to assume higher-order variances to be negligible. In some cases, these higher-order variances as a group can be tested for their presence.
Table 8.5. Example of total number of terms of each type of effect in the genic factorial model for n loci, n = 1, 2, 3 and 4.

<table>
<thead>
<tr>
<th>Type of effect</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>A²</td>
<td>4</td>
<td>12</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>AD</td>
<td>4</td>
<td>12</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>D²</td>
<td>1</td>
<td>3</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>A³</td>
<td>8</td>
<td>32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A²D</td>
<td>12</td>
<td>48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A D²</td>
<td>6</td>
<td>24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D³</td>
<td>1</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A⁴</td>
<td></td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A³D</td>
<td></td>
<td>32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A²D²</td>
<td></td>
<td>24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A D³</td>
<td></td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D⁴</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3</td>
<td>15</td>
<td>63</td>
<td>255</td>
</tr>
</tbody>
</table>

Similarly, extension of the locus factorial model (8.130) to more loci is also possible. There are all possible two-locus interactions (yy)'s, all possible three-locus interactions (yyy)'s, and so on, with a corresponding decomposition into interaction effects for combinations of genes.

Likewise the gametic factorial model may be extended. The gametic effects would include all additive types of effects, α's, (αα)'s, (ααα)'s, and so on, corresponding to the genes in the gametes, with the remaining effects being assigned to the interactions between gametes.
Under random mating between two populations and linkage equilibrium, the effects in the model (8.175) can be defined in a manner similar to that in (8.136), (8.143) to (8.148), and their variances defined similar to that in (8.149). Then the total variance among the genotypes for the model (8.175) may also be written down in a manner analogous to that in (8.150) and (8.151).

2. **Linkage disequilibrium.** If linkage disequilibrium exists, but yet mating is at random, then the three composite effects in the following gametic model are independent

\[
G_{st}^{k_i k_j} = G_{st} = \mu + g_s^m + g_t^f + (gg)_{st}
\]

where \( G_{st} \) = genotypic value obtained from the union of gametes \( s \) and \( t \) \( \sum_{k=1}^{n} A_{k_i}^m \) and \( \sum_{k=1}^{n} A_{k_j}^f \)

\[
g_s^m = \sum_{k=1}^{n} \alpha_{k_i}^m + \sum_{k=1}^{n} \sum_{k'}^n (\alpha\alpha)_{k_i k'}^m + \sum_{k=1}^{n} \sum_{k'}^n \sum_{k''}^n (\alpha\alpha)_{k_i k' k''}^m m m m + \ldots
\]

\[
= \text{total effect of gamete } s \text{ from male parent,}
\]

\[
g_t^f = \sum_{k=1}^{n} \alpha_{k_j}^f + \sum_{k=1}^{n} \sum_{k'}^n (\alpha\alpha)_{k_j k'}^f + \sum_{k=1}^{n} \sum_{k'}^n \sum_{k''}^n (\alpha\alpha)_{k_j k' k''}^f + \ldots
\]

\[
= \text{total effect of gamete } t \text{ from female parent,}
\]

\[
(gg)_{st}^{k_i k_j} = \sum_{k=1}^{n} \delta_{k_i k_j} + \sum_{k=1}^{n} \sum_{k'}^n (\alpha\alpha)_{k_i k' k_j}^m + \sum_{k=1}^{n} \sum_{k'}^n (\alpha\alpha)_{k_i k_j k'} + \sum_{k=1}^{n} \sum_{k'}^n \sum_{k''}^n (\alpha\alpha)_{k_i k' k''} + \ldots
\]

\[
= \text{total effect of interactions between gametes } s \text{ and } t.
\]

Then
\[ \sigma_G^2 = \sigma_m^2 + \sigma_f^2 + \sigma_{(\text{gg})mf}^2 \]  
\hfill (8.177)

but \( \sigma_m^2 \), for example, must now include all possible covariances between the effects in the male gamete due to the linkage disequilibrium, i.e.,
\[ \sigma_m^2 = \sigma_A^2 + \sigma_{AA}^2 + \sigma_{AAA}^2 + \ldots + \text{all possible covariances between effects} \]  
\hfill (8.178)

With respect to the variance \( \sigma_{gg}^2 \) of the interaction effect between gametes, I am presently unclear what covariances between terms exist.

8.3.6. More than two loci, multiple alleles, equal male and female frequencies at each locus (Hardy-Weinberg), epistasis. 1. Linkage equilibrium.

When random mating occurs and all allelic frequencies in the male population are equal to the allelic frequencies in the female population,

\[ p_{1i}^m = p_{1i}^f = p_{1i} \quad \text{for } i = 1, \ldots, m_1 \]
\[ p_{2i}^m = p_{2i}^f = p_{2i} \quad \text{for } i = 1, \ldots, m_2 \]
\[ \vdots \]
\[ p_{ni}^m = p_{ni}^f = p_{ni} \quad \text{for } i = 1, \ldots, m_n \]  
\hfill (8.179)

the following variances for terms in the model (8.175) are equal

\[ \sigma_k^2 = \sigma_k^2 \quad \text{for } k = 1, \ldots, n \]  
\hfill (8.180a)

\[ \sigma_{(aa)}^2_{kk'} = \sigma_{(aa)}^2_{kk'} = \sigma_{(aa)}^2_{kk'} = \sigma_{(aa)}^2_{kk'} \]  
\hfill (8.180b)

for every combination of \( k, k' = 1, \ldots, n; k < k' \)

\[ \sigma_{(\alpha\delta)}^2_{kk'k'} = \sigma_{(\alpha\delta)}^2_{kk'k'} = \sigma_{(\alpha\delta)}^2_{kk'k'} \]  
\hfill (8.180c)

for every combination of \( k, k' = 1, \ldots, n; k \neq k' \)
\[ \sigma^2_{(\delta\alpha)}_{kkk'} = \sigma^2_{(\delta\alpha)}_{kkk}, \quad \text{for every combination of } k, k' = 1, \ldots, n; \ k \neq k' \]

\[ \sigma^2_{(\alpha\alpha)}_{mm} = \sigma^2_{(\alpha\alpha)}_{kk} = \sigma^2_{(\alpha\alpha)}_{ff} = \sigma^2_{(\alpha\alpha)}_{kk}, \quad \text{for every combination of } k, k' = 1, \ldots, n; \ k < k' < k'\]

\[ \sigma^2_{(\alpha\delta)}_{mm} = \sigma^2_{(\alpha\delta)}_{kk}, \quad \sigma^2_{(\alpha\delta)}_{ff} = \sigma^2_{(\alpha\delta)}_{kk}, \quad \text{for every combination of } k, k', k'' = 1, 2, \ldots, n; \ k < k', k'' \neq k, k'' \neq k' \]

\[ \sigma^2_{(\alpha\delta)}_{mm} = \sigma^2_{(\alpha\delta)}_{kk}, \quad \text{for every combination of } k, k', k'' = 1, 2, \ldots, n; \ k' < k'', k' \neq k, k'' \neq k, \text{ etc.} \]

Hence, the variance among genotypes for the model (8.175) is

\[ \sigma^2_G = \sigma^2_A + \sigma^2_D + \sigma^2_{AA} + \sigma^2_{AD} + \sigma^2_{DD} + \sigma^2_{AAA} + \sigma^2_{AAD} + \sigma^2_{ADD} + \sigma^2_{DDD} + \ldots \]

where

\[ \sigma^2_A = \sum_{k=1}^{n} \sigma^2_{a_k} = \text{total additive variance}, \]

\[ \sigma^2_D = \sum_{k=1}^{n} \sigma^2_{\delta_k} = \text{total dominance variance}, \]

\[ \sigma^2_{AA} = 4 \sum_{k=1}^{n} \sum_{k'}^{n} \sigma^2_{(\alpha\alpha)}_{kk}, \quad \text{total additive-by-additive variance}, \]

\[ \sigma^2_{AD} = 2 \sum_{k} \sum_{k'} \sigma^2_{(\alpha\delta)}_{kk'} = \text{total additive-by-dominance variance}, \]

\[ \sigma^2_{DD} = \sum_{k} \sum_{k'} \sigma^2_{\delta_{kk'}} = \text{total dominance-by-dominance variance}, \]
\[
\sigma_{\text{AAA}}^2 = 8 \sum_{k} \sum_{k'} \sum_{k''} \sigma^2_{(aa\alpha)kk'k''} = \text{total additive-by-additive-by-additive variance},
\]
\[
\sigma_{\text{AAD}}^2 = 4 \sum_{k} \sum_{k'} \sum_{k''} \sigma^2_{(aa\delta)kk'k''} = \text{total additive-by-additive-by-dominance variance},
\]
\[
\sigma_{\text{ADD}}^2 = 2 \sum_{k} \sum_{k'} \sum_{k''} \sigma^2_{(a\delta\delta)kk'k''} = \text{total additive-by-dominance-by-dominance variance},
\]
\[
\sigma_{\text{DDD}}^2 = \sum_{k} \sum_{k'} \sum_{k''} \sigma^2_{(\delta\delta\delta)kk'k''} = \text{total dominance-by-dominance-by-dominance variance},
\]

etc.

Each variance term in (8.181) represents the appropriate sum of variances of effects of the same type in (8.175) and takes into account the equality of certain variances (8.180).

2. **Linkage disequilibrium.** Nyquist discuss.

8.4. **Covariances of characters.** Assumptions:

The individual is the basic unit in a population. Although it may be abstracted in terms of genes, uniting gametes, a single characteristic, and so on, as we have done, it is the whole individual that must be reckoned with. Each individual is a composite of many characters. We will first discuss the covariance between two characters in various genetical contexts, and then discuss an application of its use in the variance of an index.
8.4.1. Covariance of two characters.

8.4.1.1. One locus, multiple alleles, arbitrary "male" and "female" allelic frequencies (cross between two populations): General. To reckon with more than one character requires a measure of relationship between characters. For this purpose consider the covariance between two characters, character 1 and character 2. Each character has the same model (8.2) but we distinguish the effects for each character by the subscripts 1 and 2, namely,

\[ G_{ij,1} = \mu_1 + \alpha_{i,1}^m + \alpha_{j,1}^f + \delta_{ij,1} \]
\[ G_{ij,2} = \mu_2 + \alpha_{i,2}^m + \alpha_{j,2}^f + \delta_{ij,2} \]

(8.182)

In a population in which genes are associated at random (a population in Hardy-Weinberg equilibrium or the cross from random mating between two populations (see p. 8.4)), the covariance between \( G_1 \) and \( G_2 \) for the particular locus, by definition, is

\[
\text{Cov}(G_1, G_2) = \sum_{i=1}^{m} \sum_{j=1}^{m} p_i p_j (G_{ij,1} - \mu_1)(G_{ij,2} - \mu_2) \quad \text{(sub. (8.182))}
\]

\[
= \sum_{i=1}^{m} \sum_{j=1}^{m} p_i p_j (\alpha_{i,1}^m + \alpha_{j,1}^f + \delta_{ij,1})(\alpha_{i,2}^m + \alpha_{j,2}^f + \delta_{ij,2}) \quad (8.183)
\]

Taking the frequency of each genotype times each possible product between the effects of the two characters, we have

\[
\text{Cov}(G_1, G_2) = \sum_{i=1}^{m} \sum_{j=1}^{m} p_i p_j (\alpha_{i,1}^m)(\alpha_{i,2}^m) + \sum_{i=1}^{m} \sum_{j=1}^{m} p_i p_j (\alpha_{j,1}^f)(\alpha_{j,2}^f) + \sum_{i=1}^{m} \sum_{j=1}^{m} p_i p_j (\delta_{ij,1})(\delta_{ij,2})
\]

+ \sum_{i=1}^{m} \sum_{j=1}^{m} p_i p_j (\alpha_{i,1}^m)(\alpha_{j,2}^m) + \sum_{i=1}^{m} \sum_{j=1}^{m} p_i p_j (\alpha_{i,1}^m)(\delta_{ij,2}) + 4 \text{ more terms}

\[
= \sum_{i=1}^{m} p_i^m (\alpha_{i,1}^m)(\alpha_{i,2}^m)(\alpha_{j,1}^f)(\alpha_{j,2}^f) + \sum_{j=1}^{m} p_j^m (\alpha_{j,1}^f)(\alpha_{j,2}^f)(\delta_{ij,1})(\delta_{ij,2})
\]

+ \sum_{i=1}^{m} p_i^m (\delta_{ij,1})(\delta_{ij,2}) + 4 \text{ more terms (8.184)}
Substituting (2.4), (8.7), (8.8), and (8.15) in (8.184), we have for the particular locus

\[
\text{Cov}(G_1, G_2) = \sum_{i=1}^{m} p_i^m (\alpha_1^m, \alpha_2^m) + \sum_{j=1}^{m} p_j^f (\alpha_1^f, \alpha_2^f) + \sum_{i=1}^{m} \sum_{j=1}^{m} p_i^m p_j^f (\delta_{ij}^m, \delta_{ij}^f)
\]

\[
= \text{Cov}(\alpha_1^m, \alpha_2^m) + \text{Cov}(\alpha_1^f, \alpha_2^f) + \text{Cov}(\delta_1^m, \delta_2^f)
\]

\[
= C_{\alpha_1^m, \alpha_2^m} + C_{\alpha_1^f, \alpha_2^f} + C_{\delta_1^m, \delta_2^f}
\]

where \(C_{\alpha_1^m, \alpha_2^m}\) = covariance between \(\alpha^m\) for character 1 and \(\alpha^m\) for character 2,

\(C_{\alpha_1^f, \alpha_2^f}\) = covariance between \(\alpha^f\) for character 1 and \(\alpha^f\) for character 2,

\(C_{\delta_1^m, \delta_2^f}\) = covariance between \(\delta\) for character 1 and \(\delta\) for character 2.

The result is analogous to genetic variances (8.22), except in terms of covariances. The covariance between the genotypic values for two characters, 1 and 2, is the algebraic sum of the three component covariances involving the three effects in the model. We are concerned with the same set of genes and their effects upon two characters. These covariances are due to pleiotropy.

If we do not have random mating, positive assortative mating will increase the covariance between two characters. A positive covariance between two characters would increase and likewise a negative covariance would increase, approaching zero. Negative assortative mating will decrease the covariance between two characters. A positive covariance would decrease toward zero and a negative covariance would also decrease (become more negative).

8.4.1.2. One locus, multiple alleles, equal male and female allelic frequencies (Hardy-Weinberg). If we assume one population and only one locus in Hardy-Weinberg equilibrium, then \(p_i^m = p_i^f = p_i\) for all alleles, \(i = 1, \ldots, m\).
Then \( C_{m}^{\alpha} = C_{f}^{\alpha} = C_{\alpha} \) so (8.185) for the particular locus becomes
\[
C_{1,2} = C_{m}^{\alpha} + C_{f}^{\alpha} + C_{\delta}^{\alpha}
\]
\[
= 2 C_{\alpha} + C_{\delta}^{\alpha}
\]
\[
= C_{A_{1,2}} + C_{D_{1,2}}
\]
where \( C_{A_{1,2}} \) = total covariance between characters 1 and 2 due to additive effects of alleles for the particular locus.

8.4.1.3. Two or more loci, multiple alleles, arbitrary "male" and "female" allelic frequencies (cross between two populations), linkage equilibrium, no epistasis. General. For several loci, the covariance between characters 1 and 2 is simply the sum over loci of each of the three component covariances in (8.185) when the \( n \) loci are in linkage equilibrium and do not interact (no epistasis), namely, (see (8.92) and (8.93) for notation)

\[
\text{Cov}(G_{1}, G_{2}) = \sum_{m} \sum_{n} \sum_{m_{1}} \sum_{n_{1}} \sum_{m_{2}} \sum_{n_{2}} \ldots \sum_{m_{n}} \sum_{n_{n}} p_{m_{1}} p_{n_{1}} p_{m_{2}} p_{n_{2}} \ldots p_{m_{n}} p_{n_{n}}
\]

\[
(\sum_{k=1}^{n} k_{i,j,1} k_{j,i,1} - \mu_{1})(\sum_{k=1}^{n} k_{i,j,2} k_{j,i,2} - \mu_{2})
\]

\[
= \sum_{k=1}^{n} (\sum_{i=1}^{m_{k}} k_{i,j,1} + \sum_{k=1}^{n} k_{k_{i},1}) + \sum_{j=1}^{m_{k}} k_{i,j,1} + \sum_{k=1}^{n} k_{k_{i},1} + \sum_{\delta_{k_{i},k_{j},1}} k_{k_{i},k_{j},1}
\]

\[
= \sum_{k=1}^{n} (\sum_{i=1}^{m_{k}} k_{i,j,2} + \sum_{k=1}^{n} k_{k_{i},2}) + \sum_{j=1}^{m_{k}} k_{i,j,2} + \sum_{k=1}^{n} k_{k_{i},2} + \sum_{\delta_{k_{i},k_{j},2}} k_{k_{i},k_{j},2}
\]

(8.187)

where \( \prod_{k=1}^{n} p_{k_{i}} p_{k_{j}} \) = frequency of genotype \( \prod_{k=1}^{n} A_{k_{i}} A_{k_{j}} \).

Then proceeding in a manner similar to (8.185), we obtain all possible products between the \( 3n \) effects for character 1 and the \( 3n \) effects for character 2, and multiply each by the frequency of genotype \( \prod_{k=1}^{n} A_{k_{i}} A_{k_{j}} \). For the same reasons as in
(8.185), all terms are zero except the 3n terms involving corresponding effects for characters 1 and 2, so

\[
\text{Cov}(G_1, G_2) = \sum_{k=1}^{n} \left[ \sum_{i=1}^{m} \frac{p_{k_i} (a_{k_i,1}) (a_{k_i,2})}{k_{i=1}} \right] + \sum_{k=1}^{n} \left[ \sum_{j=1}^{m} \frac{p_{k_j} (a_{k_j,1}) (a_{k_j,2})}{k_{j=1}} \right]
\]

\[
+ \sum_{k=1}^{n} \sum_{i=1}^{m} \frac{p_{k_i} p_{k_i} (\delta_{k_i,1}) (\delta_{k_i,2})}{k_{i=1} k_{i=1}}
\]

(8.188)

Then making substitutions similar to that in (8.185), we have

\[
\text{Cov}(G_1, G_2) = \sum_{k=1}^{n} \text{Cov}(a_{k,1}, a_{k,2}) + \sum_{k=1}^{n} \text{Cov}(a_{k,1}, a_{k,2}) + \sum_{k=1}^{n} \text{Cov}(\delta_{k,1}, \delta_{k,2})
\]

(8.189)

where \(\text{Cov}(a_{k,1}, a_{k,2}) = \sum_{k=1}^{m} p_{k_i} (a_{k_i,1}) (a_{k_i,2})\), and so on.

Resymbolizing (8.189) further, we write

\[
\text{Cov}(G_1, G_2) = C_{m} + C_{f} + C_{d}
\]

(8.190)

where \(C_{m} = \sum_{k=1}^{n} C_{m,1,2} a_{k,1,2}\) = total covariance between additive effects \(a_{m}\) for characters 1 and 2 summed over loci due to the genes from the male parents, or from the \(m\) population,

\(C_{f} = \sum_{k=1}^{n} C_{f,1,2} a_{k,1,2}\) = total covariance between additive effects \(a_{f}\) for character 1 and 2 summed over loci due to the genes from the female parents, or from the \(f\) population,

\(C_{d} = \sum_{k=1}^{n} C_{d,1,2} \delta_{k,1,2}\) = total covariance between dominance effects \(\delta\) for characters 1 and 2 summed over loci.

When nonallelic genes are independent, any genetic covariance of two characters must stem from pleiotropic effects of genes, at whatever physiological stage these effects occur. It should be pointed out that a zero covariance does not necessarily imply no pleiotropic effect of the alleles, because some alleles
may give positive product effects and others negative product effects such that they exactly cancel each other. If the genes at different loci are not independent, i.e., nonallelic genes are not associated at random, namely, linkage disequilibrium exists, then the covariance of nonallelic genes would have to be taken into account in Cov(G₁, G₂) as it would also in $\sigma_G^2$ for each characteristic (see (8.99) (8.100)).

8.4.1.4. Two or more loci, multiple alleles, equal male and female allelic frequencies (Hardy-Weinberg), linkage equilibrium, no epistasis. With more than one locus, equal male and female allelic frequencies, and linkage equilibrium,

$$C_{A_1,2}^m = C_{A_1,2}^f = \frac{1}{2} C_{A_1,2}$$

so (8.190) becomes

$$C_{G_1,2} = C_{A_1,2}^m + C_{A_1,2}^f + C_{D_1,2}$$

$$= C_{A_1,2}^m + C_{D_1,2}$$

(8.191)

where $C_{A_1,2} = C_{A_1,2}^m + C_{A_1,2}^f$

Alternatively, (8.191) could be obtained by summing (8.186) over the n loci.

8.4.2. Variance of an index: An application. Covariances between characters have many applications. We will discuss one application in the variance of an index used in selection.

8.4.2.1. One locus, multiple alleles, arbitrary "male" and "female" allelic frequencies (cross between two populations): General. When considering composites of characteristics or (selection) indexes, the variance of linear combinations can be accommodated by using standard procedures

$$I = b_1G_1 + b_2G_2$$

(8.192)

where I = index for a linear function of characters 1 and 2.

Then substituting (8.182) in (8.192), we have for the particular locus
\[ I = b_1(\mu_1 + \alpha_{1,1}^m + \alpha_{1,1}^f + \delta_{1,1}) + b_2(\mu_2 + \alpha_{2,1}^m + \alpha_{2,1}^f + \delta_{2,1}) \]
\[ = (b_1\mu_1 + b_2\mu_2) + (b_1\alpha_{1,1}^m + b_2\alpha_{2,1}^m) + (b_1\alpha_{1,1}^f + b_2\alpha_{2,1}^f) + (b_1\delta_{1,1} + b_2\delta_{2,1}) \]
\[ = \mu_I + \alpha_I^m + \alpha_I^f + \delta_I \]  
(8.193)

where \( \mu_I = b_1\mu_1 + b_2\mu_2 \) = index value of \( \mu \)'s,
\( \alpha_I^m = b_1\alpha_{1,1}^m + b_2\alpha_{2,1}^m \) = index value of \( \alpha^m \)'s,
\( \alpha_I^f = b_1\alpha_{1,1}^f + b_2\alpha_{2,1}^f \) = index value of \( \alpha^f \)'s,
\( \delta_I = b_1\delta_{1,1} + b_2\delta_{2,1} \) = index value of \( \delta \)'s.

Then from the theorem for the variance of a linear combination for uncorrelated variables, assuming random union of male and female gametes, we have from (8.193)

\[ \sigma_I^2 = \sigma_I^m + \sigma_I^f + \sigma_I^\delta \]  
(8.194)

where
\[ \sigma_I^m = b_1^2 \sigma_{1,1}^m + b_2^2 \sigma_{2,1}^m + 2b_1b_2 \text{Cov}(\alpha_{1,1}^m, \alpha_{2,1}^m), \]
\[ \sigma_I^f = b_1^2 \sigma_{1,1}^f + b_2^2 \sigma_{2,1}^f + 2b_1b_2 \text{Cov}(\alpha_{1,1}^f, \alpha_{2,1}^f), \]
\[ \sigma_I^\delta = b_1^2 \sigma_{1,1}^\delta + b_2^2 \sigma_{2,1}^\delta + 2b_1b_2 \text{Cov}(\delta_{1,1}, \delta_{2,1}). \]

8.4.2.2. One locus, multiple alleles, equal male and female allelic frequencies (Hardy-Weinberg). If \( p_i^m = p_i^f \) for \( i = 1, \ldots, m \), \( \sigma_I^m = \sigma_I^f = \sigma_I^\delta \), so (8.194) becomes

\[ \sigma_I^2 = \sigma_I^m + \sigma_I^f + \sigma_I^\delta \]
(8.195)

8.4.2.3. Two or more loci, multiple alleles, arbitrary "male" and "female" allelic frequencies (cross between two populations), linkage equilibrium, no epistasis: General. Summing over loci which are in linkage equilibrium and do not interact (no epistasis), (8.194) becomes
\[ \begin{align*}
\sigma^2_G &= \sum_{k=1}^{n} \sigma^2_{I_k} = \sum_{k=1}^{n} \sigma^2_{\alpha_{I_k}} + \sum_{k=1}^{n} \sigma^2_{\delta_{I_k}} \\
&= \sigma^2_{A_i} + \sigma^2_{D_i} \\
\text{where } \sigma^2_{A_i} &= b_1^2 \sum_{k=1}^{n} \sigma^2_{\alpha_{I_k}} + b_2^2 \sum_{k=1}^{n} \sigma^2_{\alpha_{I_k}} + 2b_1b_2 \sum_{k=1}^{n} \text{Cov}(\sigma^2_{\alpha_{I_k}}, \sigma^2_{\alpha_{I_k}}) \\
&= b_1^2 \sigma^2_{A_m,1} + b_2^2 \sigma^2_{A_m,2} + 2b_1b_2 \sigma^2_{A_m,1,2} \\
\sigma^2_{D_i} &= b_1^2 \sum_{k=1}^{n} \sigma^2_{\delta_{I_k}} + b_2^2 \sum_{k=1}^{n} \sigma^2_{\delta_{I_k}} + 2b_1b_2 \sum_{k=1}^{n} \text{Cov}(\sigma^2_{\delta_{I_k}}, \sigma^2_{\delta_{I_k}}) \\
&= b_1^2 \sigma^2_{D_f,1} + b_2^2 \sigma^2_{D_f,2} + 2b_1b_2 \sigma^2_{D_f,1,2} \\
\end{align*} \]

(8.196)

8.4.2.4. Two or more loci, multiple alleles, equal "male" and "female" allelic frequencies (Hardy-Weinberg), linkage equilibrium, no epistasis. With equal male and female allelic frequencies, \( \sigma^2_{A_i} = \sigma^2_{F_i} = \frac{1}{2} \sigma^2_{A_i} \) so (8.196) becomes

\[ \begin{align*}
\sigma^2_G &= \sum_{k=1}^{n} \sigma^2_{I_k} = \sum_{k=1}^{n} \sigma^2_{\alpha_{I_k}} + \sum_{k=1}^{n} \sigma^2_{\delta_{I_k}} \\
&= \sigma^2_{A_i} + \sigma^2_{D_i} \\
\text{where } \sigma^2_{A_i} &= \sigma^2_{A_i} + \sigma^2_{D_i}. \\
\end{align*} \]

(8.197)

Alternatively, (8.197) could be obtained by summing (8.195) over the \( n \) loci.

The extension to any number of characters in the index is straightforward and will lead to the same simple notation as in (8.197). However, the decomposition of \( \sigma^2_{A_i} \), for example, into its parts as in (8.196) indicates the complexity of the situation.
References

Anderson, R. L., and T. A. Bancroft

Anderson, V. L., and O. Kempthorne

Bancroft, T. A.

Bulmer, M. G.

Cockerham, C. C.

Cockerham, C. C.
1966 Notes on quantitative genetics, Section 2. North Carolina State University, Raleigh, North Carolina.

Comstock, R. E., and H. F. Robinson

Comstock, R. E., and H. F. Robinson

Crow, J. F., and M. Kimura

Draper, N. R., and H. Smith

Empig, L. T., C. O. Gardner, and W. A. Compton
1972 Theoretical gains for different population improvement procedures. University of Nebraska, Agricultural Experiment Station, MP28, revised.

Falconer, D. S.
Fisher, R. A.

Griffing, B.

Hartl, D. L., and A. G. Clark

Hayman, B. I.

Kempthorne, O.

Kempthorne, O.

Kempthorne, O.

Kendall, M. G., and A. Stuart

Li, C. C.

Mather, K.

Mather, K., and J. L. Jinks

Mather, K., and J. L. Jinks

Scheffe, H.

Snedecor, G., and W. G. Cochran
Steel, R. G. D., and J. H. Torrie

Turner, H. N., and S. S. S. Young

Weir, B. S., C. C. Cockerham, and J. Reynolds

Wright, S.