1953

A model for the study of quantitative inheritance

Virgil L. Anderson
Iowa State College

Follow this and additional works at: https://lib.dr.iastate.edu/rtd
Part of the Statistics and Probability Commons

Recommended Citation
https://lib.dr.iastate.edu/rtd/12878

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.
NOTE TO USERS

This reproduction is the best copy available.

UMI
A MODEL FOR THE STUDY OF
QUANTITATIVE INHERITANCE

by

Virgil L. Anderson

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major Subject: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State College

1953
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II. REVIEW OF LITERATURE</td>
<td>7</td>
</tr>
<tr>
<td>III. THE FACTORIAL MODEL</td>
<td>17</td>
</tr>
<tr>
<td>A. General Remarks on Models</td>
<td>17</td>
</tr>
<tr>
<td>B. The Factorial Model</td>
<td>22</td>
</tr>
<tr>
<td>C. Discussion of the Factorial Model</td>
<td>28</td>
</tr>
<tr>
<td>D. The Scope of the Body of the Thesis</td>
<td>32</td>
</tr>
<tr>
<td>IV. THE APPLICATION OF THE FACTORIAL MODEL TO SELF-FERTILIZATION</td>
<td>33</td>
</tr>
<tr>
<td>V. TWO PARENTS</td>
<td>65</td>
</tr>
<tr>
<td>VI. THREE PARENTS</td>
<td>82</td>
</tr>
<tr>
<td>VII. SUMMARY AND CONCLUSIONS</td>
<td>104</td>
</tr>
<tr>
<td>VIII. BIBLIOGRAPHY</td>
<td>108</td>
</tr>
<tr>
<td>IX. ACKNOWLEDGMENTS</td>
<td>110</td>
</tr>
</tbody>
</table>
I. INTRODUCTION

Methods for examining and studying the inheritance of quantitative characters are extremely important to plant and animal breeders. Breeders are continuously striving for a better understanding of the complex inheritance of quantitative character. These characters are controlled by a large number of hereditary units, called genes, that may act independently or may interact with one another. In general, the effect of an individual gene is relatively small compared to the joint effect of the entire complex of genes affecting the character involved. Individual genes probably do not act independently, therefore it is quite advantageous to study the whole group of genes at once. Since most quantitative characters are measurable and are determined by systems of genes which must be handled "en masse", statistical techniques are necessary for an evaluation of the inheritance involved.

Before proceeding to a discussion of these statistical techniques, a review of basic genetic terminology should be useful. The genes are arranged along threadlike bodies, called chromosomes and particular genes are generally located at particular points along the chromosome called loci. At each locus two or more genes are possible and these are called alleles. In the simple situation only two alleles are possible at a given locus on a chromosome pair.
This is the only case we will discuss in this thesis. Frequently, however, there are more than two alleles possible at a given locus. These alleles are called multiple alleles.

Only the diploid condition, that is the condition which exists when chromosomes appear in pairs, will be considered in this dissertation. The basis for statistical investigations in genetics is Mendelian segregation. This is a random process by which an offspring receives with equal probability one of the two genes possessed by the parent at a locus. In the process of reproduction the parent passes on to the offspring one chromosome from each of the pairs of chromosomes. The whole array of a single set of chromosomes is called a gamete. The result of the uniting of two gametes is a zygote.

The possible genotypes, that is the genetic classes of zygotes, in the case of two alleles and one locus are AA, Aa and aa. The genotype which has different alleles, Aa, is called heterozygous, while the other two genotypes which have like alleles, AA and aa, are called homozygous. If we denote the expression that the genotypes exhibit by the genotype symbol, the concepts of dominance are as follows:

(a) Complete dominance if \( Aa = AA \)
(b) Overdominance if \( Aa > AA \)
(c) No dominance if \( Aa = \frac{1}{2} (aa + AA) \)
(d) Positive partial dominance if \( \frac{1}{2} (aa + AA) < Aa < AA \)
(e) Negative partial dominance if \( aa < Aa < \frac{1}{2} (aa + AA) \)
When several loci affecting the same character are on the same chromosome these loci are said to be linked. The double heterozygote, Aa Bb, is said to be in the coupling phase if A and B are on one chromosome and a and b are on the other. It is said to be in the repulsion phase if A and b are on one chromosome and a and B are on the other. In general, for any number of loci affecting a given character the possible combinations of genes will not be passed on to the offspring with equal frequency and the inequality of the frequencies in the case of two loci is measured by the recombination value.

The observed value of a given character of an individual is called the phenotypic value of that character for that individual. In the simple situations we can visualize the phenotypic value as being made up of two parts, the genotypic value and the environmental value. The genotypic value is the average of the phenotypic value over a large number of repetitions in a specified population of environments. The environmental portion is assumed to be additive with the genotypic value.

That part of the genotypic value contributed by a given locus is called the locus effect. One portion of the locus effect is the basic gene effect which is defined to be

\[ \frac{1}{2} (AA - aa). \]

The other portion of the locus effect is the dominance effect which is defined to be

\[ \frac{1}{2} (2Aa - AA - aa). \]
If several loci affect the same character but the genotypic value of these loci in combination is not equal to the sum of the basic gene effects plus the dominance effects, there is said to be an interaction between loci called epistasis. If it is assumed that there is no epistasis then the genotypic value is made up of a set of loci effects only.

The roles of these three loci contributions, namely basic gene effect, dominance effect and epistatic effect in quantitative inheritance has been a subject of controversy for several years.

The methods by which these matters are investigated depend upon the observation of various types of crosses of genetic material. The simple types are as follows. When a plant is fertilized with pollen from the same plant (that is the plant in monoecious), the process is called self fertilization. When two parents with homozygous loci for the character being considered are crossed, the whole group of progenies is called an $F_1$ population. If $F_1$ individuals from the same parents are crossed, the group of progenies is called an $F_2$ population. If $F_2$ individuals from the same $F_1$'s are self fertilized the group of progenies is called an $F_3$ population. Similarly for any number of generations, $m$, of self fertilization if $F_{m-1}$ individuals from the same $F_{m-2}$'s are self fertilized the group of progenies is called an $F_m$ population.

The basis of the analysis of inheritance of quantitative characters is the postulation of a model which states that the
phenotypic expression of an individual (or population) is equal to a genetic effect which is a function of the genetic structure of the individual plus an environmental contribution which is independent of the genetic effect. A gene model is then a function of the genetic structure. For example, two gene models are the following:

(a) \( AA = a, Aa = d \) and \( aa = -a \)
where:
\( a = \) basic gene effect
\( d = \) dominance effect.

(b) \( AAbb = a(1/e), Aabb = \frac{a}{2}(1-e)(1+h), aabb = 0 \)
where:
\( a = \) basic gene effect
\( h = \) dominance effect
\( e = \) epistatic effect.

These gene models which have been used will be described in the next section.

The aim of the present thesis is to examine a general model which includes all possible interaction contributions, in connection with self fertilization, crossing of two parents, three parents and four parents. In this way it will be possible to examine the way in which epistatic effects enter into the means of various populations. It will then be possible to examine what are known as scaling tests, which are tests for the existence of epistasis, and determine the exact role of epistatic effects in these tests. It
will also be possible, with some further assumptions about the type of interactions involved, to estimate the epistatic contribution to certain genotypic values and means of genotypic values for populations.
The following gene model of a single gene pair, that is a single locus, showing dominance was devised by Fisher (7) in 1918 and later, in 1932, with different symbols was used by Fisher, Immer and Tedin (8):

\[
\begin{array}{c}
\text{aa} & \text{d} & \text{AA} \\
\text{a} & \text{0} & \text{a}
\end{array}
\]

where: \( d \) represents the dominance relations.

If \( -a < d < a \), there is incomplete or partial dominance.

If \( d = a \), there is complete dominance.

If \( d > a \), there is overdominance or superdominance.

This model is one of the first gene models used to estimate genetic parameters. It has the very desirable property of symmetry, in which aa is the negative of AA and the mean of the two parents, assuming aa and AA are the two parents, is zero. The model is simple because it deals with only one locus and has only the locus effect involved. Fisher, et al, state that if heritable variance observable among any group of organisms is the sum of the variance due to the individual loci, the model is easily extended so that any number of loci may be considered affecting the character being studied.
Since the model is extremely simple and quantitative inheritance is so complex, it is not surprising that some restrictive assumptions must be made to use the model. Fisher, et al. (8) assumed that epistasis, the interaction among loci, was negligible. Fisher (7) stated that a deviation from additive effects similar to dominance may occur between loci. This he called epistacy. The term epistasis was originally used by Bateson (2) to describe the effects of genes which covered up the effects of other genes. Fisher (7) represented the epistatic deviations for a dihybrid in the $i, j$th genotype as $e_{ij}$, where $(i = 1, 2$ or $3)$ and $(j = 1, 2$ or $3)$; and used the term dual epistacy for the interaction which exists between two loci. The assumption that the effect of the interaction of loci is negligible simplifies the model considerably, but it may affect the results in many genetical investigations.

Rasmusson (16) explained the interaction of loci by the hypothesis that "the effect of each factor on the genotype is dependent upon all the other factors present, the visible effect of a certain factor being smaller the greater the number of factors acting in the same direction." The term "factor" here is equivalent to the term "locus".

East (6) stated that it was not necessary to accept Rasmusson's hypothesis exactly as presented. He suggested a model that involves more than two alleles at the same locus, e.g., $A_1, A_2, A_3, A_4, a$. East presented the following relationships only:
where \( \alpha > \beta > \gamma > \delta \). East stated that \( \alpha \) is virtually equivalent to \( A_1 \). He stated that the physiological efficiency of \( A_1 A_4 \) may be supposed to be greater than \( A_1 A_2 \). He made no statement concerning the representation of \( A_2 A_2, A_3 A_3, A_4 A_4 \) and the other combinations of \( A_2, A_3, A_4 \) and \( a \).

Powers (14) obtained results from studying inheritance of habit of growth that supported Rasmusson's hypothesis. In a later study of the interaction of the factors affecting weight of seed per plant, number of spikes per plant, height of plant and length of awn, Powers (15) found the reverse of that expected from Rasmusson's hypothesis, that is the effect of certain factors was not smaller the greater the number acting in a certain direction. Powers made the further statement that the nature of the interaction of genes affecting the quantitative characters was sufficiently variable to render any hypothesis of doubtful value as a means of prediction.

Mather (13), in chapter three, discusses scales of measurement of data. For a scale to be adequate for analytical purposes it must satisfy the following two criteria:
(a) The genic effects must on the average be additive, and
(b) The contribution due to nonheritable agents must be
independent of the genotype.

Mather states that nonadditiveness due to genic interaction may be
allowed by including a special term in the analysis; but the large
variation of possible interactions prevents an easy interpretation
of the term and its use for predicting purposes is not clear.

Three quantities used to test the first criterion of the adequacy
of the scale are the following:

\[ A = \bar{x}_1 - \bar{F}_1 - \bar{F}_1, \]
\[ B = \bar{x}_2 - \bar{F}_2 - \bar{F}_1, \]
\[ C = \bar{x}_2 - \bar{x}_1 - \bar{F}_1 - \bar{F}_2 \]

where \( \bar{F}_1, \bar{F}_2, \bar{F}_1 \) and \( \bar{F}_2 \) are means of parental, \( F_1 \) and \( F_2 \) populations.

The variances of these quantities, denoted by \( V \), are

\[ V_A = 4V_{F_1} + V_{F_1} + V_{F_1}, \]
\[ V_B = 4V_{F_2} + V_{F_2} + V_{F_1}, \]
\[ V_C = 16V_{F_2} + 4V_{F_1} + V_{F_1} + V_{F_2} \]

where \( V_{F_1}, V_{F_2}, V_{F_1} \) and \( V_{F_2} \) are the variances on the respective
populations. These tests are of no value if there is differential
viability or fertility in the segregating generations. For the
first criterion of an adequate scale to be met \( A, B \) and \( C \) must each
be equal to zero within the limits of sampling error. The second
criterion is usually tested by a test of homogeneity of variances within the true breeding parents and their $F_1$ in the experiment.

Griffing (9) considered the following gene model using one locus only:

\[
\begin{array}{cccccc}
\text{aa} & d & h & Aa & d & Aa \\
0 & d & - & d & 2d
\end{array}
\]

The dominance deviation is $hd$, so that $h$ may be designated as follows:

- $h = 0$, no dominance
- $-1 < h < 0$, partial negative dominance
- $0 < h < 1$, partial positive dominance
- $h = -1$, complete negative dominance
- $h = 1$, complete positive dominance
- $h > 1$ or $h < -1$, overdominance

He used a model in which both epistasis and dominance are present. He obtained the following expressions for the genotypic values (we have denoted the genotypic value by the genotype):

- $aabb = 0$
- $Aabb$ or $aaBb = \frac{c}{2}(1-e)(1+h)$
- $AAbb$ or $aaBB = c(1-e)$
- $AaBb$ or $aABb = \frac{c}{2}(2-e+2h+eh^2)$
- $AABB$ or $AaBB = \frac{c}{2}(3-e+h+eh)$
- $AAAB = 2a$

where the epistatic parameter, $e$, may take on positive or negative
values. Griffing used a negative sign before e when there was a homozygous recessive locus present in the genotype.

The basic gene effect of AA or BB is represented by $\alpha$, $h$ is the dominance effect at both loci and $e$ is the epistatic effect.

Griffing's model was constructed by setting

$$aabb = 0$$
$$AABB = 2\alpha,$$

and building up the intermediate genotypes of the dihybrid by following certain rules for using $h$ and $e$. The $h$ in the one locus model is the same as the $h$ in the two loci model and $2h$ in the one locus model is equivalent to $\alpha$ in the two loci model.

The assumptions used by Griffing are:

(a) The basic gene effects of all genes involved are equal.
(b) The dominance effect at each locus is constant.
(c) Epistasis is considered as the interaction between loci, which is analogous to dominance considered as interaction between alleles at the same locus.

Six genotypes can be measured with Griffing's model; however, each of three of these genotypes is actually a mixture of two fundamental genotypes. For example, the second genotype given above can be $Aabb$, $aaBb$ or a mixture of the two. This model therefore does not give a complete specification and includes an assumption which is somewhat of the nature of symmetry. The need for or the desirability of making certain symmetry assumptions may be
questioned. Griffing's two loci model is, however, an approach to the problem of determining how important epistasis is in quantitative inheritance.

Charles and Smith (3) used the following gene models:

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Additive Effects</th>
<th>Multiplicative Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a_1a_2a_2)</td>
<td>(A_0)</td>
<td>(A_0)</td>
</tr>
<tr>
<td>(a_1a_2a_2)</td>
<td>(A_0 + \alpha_1 - \delta_1)</td>
<td>(A_0 (1+k_1))</td>
</tr>
<tr>
<td>(a_1a_2a_2)</td>
<td>(A_0 + 2 \alpha_1)</td>
<td>(A_0 (1+k_1)^2)</td>
</tr>
<tr>
<td>(a_1a_2a_2)</td>
<td>(A_0 + \alpha_2 - \delta_2)</td>
<td>(A_0 (1+k_2))</td>
</tr>
<tr>
<td>(a_1a_2a_2)</td>
<td>(A_0 + \alpha_1 + \alpha_2 - \delta_2)</td>
<td>(A_0 (1+k_1) (1+k_2))</td>
</tr>
<tr>
<td>(a_1a_2a_2)</td>
<td>(A_0 + 2 \alpha_1 + \alpha_2 - \delta_2)</td>
<td>(A_0 (1+k_1)^2 (1+k_2))</td>
</tr>
<tr>
<td>(\ldots)</td>
<td>(\ldots)</td>
<td>(\ldots)</td>
</tr>
<tr>
<td>(\ldots)</td>
<td>(\ldots)</td>
<td>(\ldots)</td>
</tr>
<tr>
<td>(a_1a_2a_2)</td>
<td>(A_0 + 2 \alpha_1 + 2 \alpha_2 + \ldots + 2 \alpha_n)</td>
<td>(A_0 (1+k_1)^2 (1+k_2)^2 \ldots (1+k_n)^2)</td>
</tr>
</tbody>
</table>

In the additive scheme, dominance at the \(i^{th}\) locus is shown by

\[
(a_{1a_1} - a_{1a_1}) = (\alpha_i - \delta_i) < (a_{1a_1} - a_{1a_1}) = \alpha_i + \delta_i
\]

where \(\alpha_i = \frac{1}{2} (a_{1a_1} - a_{1a_1})\) and \(\delta_i = \text{deviation of } a_{1a_1} \text{ from true intermediacy.}\n
In order to use their models on means, Charles and Smith assumed

(a) There were no linkages present

(b) For those genes which are not common to the two parents, the representations of the parents are
(i) the average size of the small parent strain =
\[ a_1 a_2 a_2 \cdots a_n a_n = \bar{v}_0 = A_0 \]

(ii) the average size of the large parent strain =
\[ a_1 a_1 a_1 a_2 \cdots a_1 a_1 = \bar{v}_0 = A_0 + 2 \sum_{i=1}^{n} \alpha_i \]

where (\hat{\cdot}) is the symbol for "is an estimate of".

The average size of the \( F_1 \) progeny =
\[ a_1 a_1 a_2 a_2 \cdots a_1 a_1 = \bar{v}_1 = A_0 + \frac{1}{2} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \delta_i \]

From the results of the two parents and \( F_1 \) described above Charles and Smith (3) found that
\[ \frac{1}{4}(\bar{v}_0 + 2\bar{v}_1 + \bar{v}_0) = A_0 + \frac{1}{2} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \delta_i \]

and they showed that the average size of the \( F_2 \),
\[ \bar{v}_2 = A_0 + \frac{1}{2} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \delta_i \]

Since the two observable values estimate the same unknown, the estimates should be equal within the limits of sampling error if the genes have arithmetic effects, that is
\[ \bar{v}_2 = \frac{1}{4}(\bar{v}_0 + 2\bar{v}_1 + \bar{v}_0) \]

Similarly
\[ \frac{1}{2}(\bar{v}_0 + \bar{v}_1) = A_0 + \frac{1}{2} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \delta_i \]

and the average size of the backcross to the small parent strain,
\[ \bar{v}_r = A_0 + \frac{1}{2} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \delta_i \]

Hence
\[ \bar{v}_r = \frac{1}{2}(\bar{v}_0 + \bar{v}_1) \]
Also, the backcross to the large parent strain
\[ \overline{V}_R = \frac{1}{2} (\overline{V}_0 + \overline{V}_1). \]

When tomato data were used, the results using the multiplicative model were superior to those using the additive model. One explanation may be that a multiplicative model will account for some of the interaction of loci whereas the additive model has no parameter for epistasis.

Comstock and Robinson (4) estimated the average degree of dominance for a character which is influenced by \( n \) loci under the following assumptions:

(a) There was no epistasis present.

(b) Either there were no linkages among the \( n \) loci, or if there were linkages, the distribution of the genotypes with respect to linked loci was at equilibrium.

(c) The gene frequency of the dominant allele was constant for all \( n \) pairs of genes.

Their model was similar to Fisher's, et al (3) except that the frequency of the dominant allele was considered.

They considered the following populations:

(a) Random mating in which each female is mated to any one of the males in the population.

(b) Random mating in which each female may be mated to a group of males in the population, e.g. multiflowered plants.

From each of these populations the authors obtained the variance due to additive genetic effects and the variance due to dominance effects
by using certain variance components of males and of females. They then let the gene frequency be .5 and obtained an estimate of the average degree of dominance. They tested the average degree of dominance by using an approximate \( F \) test in which certain mean squares were added together to estimate the expected value of the mean square desired. The three types of dominance for which they tested were overdominance, partial dominance, and complete dominance. They stated that, if epistasis is present, there is an upward bias to the estimate of the average degree of dominance, because epistatic variance will be present in the mean squares that estimate the variance due to females in population (a) above, and the variance due to the interaction of males and females in population (b).

They further asserted that if assumption (b) is not satisfied, there is no bias in the estimate of the average degree of dominance if the linkages are in the coupling phase. If the linkages are in the repulsion phase, overdominance may seem to be present when, in fact, there is only partial dominance in all the linked genes. They maintained that linkage in the repulsion phase may be detected by overdominance in early generations of a cross of two genetically divergent materials, but that the estimate of the average degree of dominance is smaller in later generations. However, there may be linkage effects in the repulsion phases which persist in later generations, causing the average degree of dominance to appear high when in fact it is low.
III. THE FACTORIAL MODEL

A. General Remarks on Models

A few of the desirable features of a gene model are:

(a) Additivity of the parameters, including epistatic parameters, for ease of handling.
(b) Parameters which have genetical meaning.
(c) Applicability to a genotypic value for any number of loci.
(d) Symmetry with respect to the expressions for the homosygous loci because the concepts of complete dominance and recessiveness at individual loci are useful only if the genotypes corresponding to these loci can be identified and examined and because such symmetry will undoubtedly lead to simplification in calculations without loss of generality.
(e) Adaptability of the model with respect to increasing assumptions, in the sense that increased assumptions result in the dropping of some terms of the model.

An example of a model which displays the difficulty which may exist when we try to extend a particular model to any number of loci, say p, and include epistatic parameters follows. The model constructed by Griffing (9) for two loci gives a basis for a generalised model for p loci. Griffing built up the expressions for the intermediate genotypes between the genotypes aabb and AABB to which he
gave values 0 and 2a respectively. Griffing's technique of setting a value of 2a on AABB assumed that there was no epistasis for that genotype. As a consequence of this assumption it is difficult to define interaction of loci in his model. We can show where $a$, his epistatic parameter, enters in the expression of the individual genotypic value but that is all. In our model for the two loci case we allow for epistasis in AABB and we are able to define interaction of loci more clearly.

In the following list of genotypic values we will denote the genotypic values by the genotypes:

\[
\begin{align*}
abbb &= (0) (0) (1 + 12e) + (0) (0) (1 + 12e) \\
Aabb &= a_1 (1 + h_1)(1 + 2e_1) + (0) (0)(1 + 2e_1) \\
AAbb &= a_1 (2) (1 + e_1) + (0) (0)(1 + 2e_1) \\
saBb &= (0) (0)(1 + e_2) + a_2 (1 + h_2)(1 + e_2) \\
AaBb &= a_1 (1 + h_1)(1 + e_{12}) + a_2 (1 + h_2)(1 + e_{12}) \\
AAbb &= a_1 (2)(1 + e_{12}) + a_2 (1 + h_2)(1 + e_{12}) \\
saBB &= (0) (0)(1 + e_2) + a_2 (2)(1 + e_2) \\
AaBB &= a_1 (1 + h_1)(1 + e_{12}) + a_2 (2)(1 + e_{12}) \\
AAaB &= a_1 (2)(1 + e_{12}) + a_2 (2)(1 + e_{12})
\end{align*}
\]

where the epistatic parameters enter and vary as follows:

\[
\begin{align*}
12e &= \text{epistatic parameter when both loci are homozygous recessive} \\
1e_2 &= \text{epistatic parameter when locus 1 is homozygous recessive and locus 2 has at least one dominant allele}
\end{align*}
\]
\( \epsilon_{1} = \) epistatic parameter when locus 2 is homozygous recessive and locus 1 has at least one dominant allele.

\( \epsilon_{12} = \) epistatic parameter when both loci have at least one dominant allele.

The basis for this model is the introduction of epistasis or interaction as in the following way:

\[
\begin{align*}
x_0y_0 &= 0 \\
x_1y_0 &= x_1 \\
x_0y_1 &= y_1 \\
x_1y_1 &= (x_1+y_1)(1+e) 
\end{align*}
\]

where \( x_0, x_1, y_0, y_1 \) are levels of 2 factors and \( e \) is a proportional deviation of the yield of \( x_1y_1 \) from what would be expected from \( x_0y_0, x_0y_1, x_1y_0 \) with additivity. For every genotypic value, if there is interaction of loci the product of the sum of the loci parameters and the epistatic parameter makes up the increment which is added to the sum of the loci effects.

In Griffing's two loci model there were only three parameters, one for basic gene effect, one for dominance effect and one for epistatic effect. If we let

\[
\begin{align*}
\alpha_1 &= \alpha_2 = \frac{\alpha}{2}, \\
h_1 &= h_2 = h, \\
12^e &= 1^2 = 2^e = -e, \\
\text{and } \quad \epsilon_{21} &= e, 
\end{align*}
\]
we may compare the two models. Again we will denote the genotypic values by the genotypes in the following:

<table>
<thead>
<tr>
<th>Genotypic Value</th>
<th>New Model</th>
<th>Griffing's Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>aabb</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Aabb or aABb</td>
<td>(\frac{a}{2}(1+e)(1-e))</td>
<td>(\frac{a}{2}(1+h)(1-e))</td>
</tr>
<tr>
<td>AAbb or aAABB</td>
<td>(a(1-e))</td>
<td>(a(1-e))</td>
</tr>
<tr>
<td>AaBB</td>
<td>(a(1+h)(1+e))</td>
<td>(\frac{a}{2}(3-e+2h+eh))</td>
</tr>
<tr>
<td>AADD or AaBB</td>
<td>(\frac{a}{2}(3+h)(1+e))</td>
<td>(\frac{a}{2}(3-e+2h+eh))</td>
</tr>
<tr>
<td>AABB</td>
<td>(2a(1+e))</td>
<td>(2a)</td>
</tr>
</tbody>
</table>

The same three pairs of genotypic values have equal expressions in both models. The first three expressions in the table are identical but the last three are different due to Griffing's definition of AABB as \(2a\).

Returning to the new model with no restrictions there are 8 parameters to be estimated, \(a_1\), \(a_2\), \(h_1\), \(h_2\), \(12^e\), \(1e_2\), \(2^e_1\) and \(e_{12}\) and nine equations to estimate the parameters. It is unfortunate that we cannot obtain estimates of the basic gene effects and epistatic effects separately for this simple case. Estimates of dominance can be obtained for this case but modifications or additional assumptions similar to those of Griffing or Charles and Smith (3) must be made before the model can be utilized.

This model lends itself to extension to the \(n\) loci case because with \(n\) loci we may expect to have a sum of \(n\) parts. Each part is

(a) a product of the basic gene effect and one plus the dominant effect of that locus
(b) the product in (a) is then multiplied by various epistatic effects to form the interaction of loci increment and the products in (a) are maintained in an additive state by adding one to every epistatic effect as follows:

\[
\text{Genotypic value for an } n \text{ loci character} = \sum_{i=1}^{n} \alpha_i (1+h_i) \prod_{j \neq i} \left(1+e_{ij}\right) \prod_{k \neq i,j} \left(1+e_{ijk}\right)...
\]

where

(a) \( \alpha_i \) = the basic gene effect of the \( i^{th} \) locus.
(b) \( h_i \) = \( \{1 \) when both alleles are dominant
= \(-1 \) when both alleles are recessive
= the dominance effect of the \( i^{th} \) locus.
(c) \( e_{ij} \) = the epistatic effect of the \( i^{th} \) locus with the \( j^{th} \) locus.
\( e_{ij} = \begin{cases} i^j & \text{if both loci are homozygous recessive.} \\ i^j & \text{if locus } i \text{ is homozygous recessive and locus } j \text{ has at least one dominant allele.} \\ j^i & \text{if locus } j \text{ is homozygous recessive and locus } i \text{ has at least one dominant allele.} \\ e_{ij} & \text{if both loci have at least one dominant allele.} \end{cases} \)
(d) \( e_{ijk} \) = the epistatic effect of the \( i^{th} \) locus with the \( j^{th} \) locus with the \( k^{th} \) locus, for all possible cases similar to those described in detail in (c) above.
(e) the epistatic effects may be represented for all possible combinations of loci up to \( n \) loci.
This model is limited in applicability because it is so complex. We have described the difficulties for the two loci case and the difficulty of handling the model increases as the number of loci increases. The main difficulty appears to be that loci effects enter additively while dominance and epistatic effects enter multiplicatively. This results in the aforementioned difficulties of estimation of the parameters or of any relationships among the parameters. It is interesting to note that there are \( (3^N - 1) \) parameters for any \( N \), the number of loci. Since the utility of this model is so limited we shall develop a model which incorporates all the interactions of all possible loci. This model will be called the factorial model. As stated earlier, the assumption is made in this model that there are only two alleles at each locus.

B. The Factorial Model

The factorial model will be developed intuitively from the case of a dihybrid. Let us represent the nine possible gene combinations of two loci with relative frequencies in the following manner:
locus a

<table>
<thead>
<tr>
<th></th>
<th>aa</th>
<th>Aa</th>
<th>AA</th>
</tr>
</thead>
<tbody>
<tr>
<td>or</td>
<td>a_0</td>
<td>a_1</td>
<td>a_2</td>
</tr>
<tr>
<td>bb</td>
<td>b_0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>locus b</td>
<td>Bb</td>
<td>b_1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>BB</td>
<td>b_2</td>
<td>1</td>
</tr>
</tbody>
</table>

These relative frequencies are the expected relative frequencies arising from the selfing of the double heterozygotes in the absence of linkage: for example, the selfing of AaBb or a_1b_1 gives AaBB (or a_1b_2) in \( \frac{2}{16} \) of the cases.

Genotypic values will be denoted also by the symbol a_1b_j so that a_2b_1 denotes both the genotype AABb and the genotypic value of AABb individuals. Whether the genotype or genotypic value is meant will always be clear from the context. Let \( \mu \) equal the mean of the genotypic values in the table, that is

\[
\mu = \frac{(a_0^2 + 2a_1 + a_2)(b_0 + 2b_1 + b_2)}{16}
\]

where the expression is to be expanded and the genotypic values inserted.

Now let

\[
A_0 = \frac{(a_0^2b_0 + 2a_0b_1 + a_0b_2)}{4} - \mu
\]

\[
A_1 = \frac{(2a_1b_0 + 2a_1b_1 + 2a_1b_2)}{8} - \mu
\]
B₀, B₁ and B₂ can be expressed similarly. At this point it is interesting to note that

\[ A₀ + 2A₁ + A₂ = B₀ + 2B₁ + B₂ = 0 \]

The interaction deviation is

\[ aₙbₖ - μ - A₁ - B₁ \]

and is therefore a deviation of \( aₙbₖ \) from an additive model. For example, the deviation for \( a₀b₀ \) is

\[ a₀b₀ - μ - A₀ - B₀ \]  \[= \frac{1}{16}(3a₀ - 2a₁ - a₂)(3b₀ - 2b₁ - b₂) \]

and for \( a₂b₁ \) it is

\[ a₂b₁ - μ - A₂ - B₂ \]  \[= \frac{1}{16}(-a₀ - 2a₁ + 3a₂)(-b₀ + 2b₁ - b₂) \]

In general the interaction deviations can be written down formally by the multiplication of certain symbols. Note that formally

\[
A₀ = a₀ \frac{(b₀ + 2b₁ + b₂)}{4} - μ
\]

\[ = \left[ a₀ - \frac{1}{4}(a₀ + 2a₁ + a₂) \right] \left[ \frac{(b₀ + 2b₁ + b₂)}{4} \right] \]

\[= \frac{1}{16}(3a₀ - 2a₁ - a₂)(b₀ + 2b₁ + b₂). \]

Likewise

\[ A₁ = \frac{1}{16}(-a₀ + 2a₁ - a₂)(b₀ + 2b₁ + b₂) \]

and

\[ A₂ = \frac{1}{16}(-a₀ - 2a₁ + 3a₂)(b₀ + 2b₁ + b₂). \]

Similarly \( B₀ = \frac{1}{16}(a₀ + 2a₁ + a₂)(3b₀ - 2b₁ - b₂) \)

\[ B₁ = \frac{1}{16}(a₀ + 2a₁ + a₂)(-b₀ + 2b₁ - b₂) \]

\[ B₂ = \frac{1}{16}(a₀ + 2a₁ + a₂)(-b₀ - 2b₁ + 3b₂). \]
We will now denote

\[
\frac{1}{4}(3a_0-2a_1-a_2) \text{ by } A_0, \quad \frac{1}{4}(3b_0-2b_1-b_2) \text{ by } B_0, \\
\frac{1}{4}(-a_0+2a_1-a_2) \text{ by } A_1, \quad \frac{1}{4}(-b_0+2b_1-b_2) \text{ by } B_1, \\
\frac{1}{4}(-a_0-2a_1+3a_2) \text{ by } A_2 \text{ and } \frac{1}{4}(-b_0-2b_1+3b_2) \text{ by } B_2.
\]

Then the interaction deviation for \(a_0b_0\) can be denoted by \(A_0B_0\) since this deviation is equal to

\[
\left[\frac{1}{4}(3a_0-2a_1-a_2)\right] \left[\frac{1}{4}(3b_0-2b_1-b_2)\right] = \frac{1}{16}(3a_0-2a_1-a_2)(3b_0-2b_1-b_2)
\]

as given previously. The same formulation holds for the other interaction deviations so that we have

\[
a_{1j} = \mu + A_1 + B_j + A_1B_j
\]

where the symbol \(A_0\), for example, means \(\frac{1}{4}(3a_0-2a_1-a_2)\) evaluated over all possibilities for unmentioned loci, that is over the \(b\) locus, \(B_0\) means \(\frac{1}{4}(3b_0-2b_1-b_2)\) evaluated over unmentioned loci, that is over the \(a\) locus, and \(A_0B_0\) means \(\frac{1}{16}(3a_0-2a_1-a_2)(3b_0-2b_1-b_2)\) evaluated over all unmentioned loci, in this case none.

The model for the two loci case, given here, corresponds to the \(2^2\) factorial model for two factors \(a\) and \(b\), say, in which

\[np = \mu + N + P + NP.\]

(See Kempthorne (11, p. 235)).

In general the model is given by the following equation:

\[
a_{1j}c_k = \mu + A_1 + B_j + A_1B_j + C + A_1C + B_jC + A_1B_jC_k + \ldots
\]

where:
(a) \( a_i b_j c_k \), \( a_i b_j c_k \) = genotypic value for a character with any number of loci

(b) \( \mu \) = overall mean

(c) \( A_0 = \frac{1}{4}(3a_0 - 2a_1 - a_2) \) evaluated over all the possible combinations for all other loci,

(d) \( A_1 = \frac{1}{4}(-a_0 + 2a_1 - a_2) \) evaluated over all the possible combinations for all other loci,

(e) \( A_2 = \frac{1}{4}(-a_0 - 2a_1 + 3a_2) \) evaluated over all the possible combinations for all other loci,

(f) Similarly for \( B_j (j = 0, 1 \text{ or } 2), C_k (k = 0, 1 \text{ or } 2) \) and so on

(g) \( A_i B_j = \) product of \( A_i \) and \( B_j \) (for all \( i \) and \( j \)), evaluated over all other loci

(h) Similarly for \( A_i C_k, B_j C_k \) (for all \( i, j \) and \( k \)) and so on, evaluated over all unmentioned loci

(i) \( A_i B_j C_k = \) product of \( A_i B_j \) and \( C_k \) (for all \( i, j \) and \( k \)), evaluated over all unmentioned loci

(j) Similarly, all other combinations are formed by products.

In order to validate the accuracy of the model we must show that the equation above is an identity. That is, we must prove that

\[
a_i b_j c_k = \mu + A_i + B_j + C_k + A_i B_j + C_k + A_i C_k + B_j C_k + A_i B_j C_k + \ldots
\]

We know that

\[
\mu = \frac{(a_0 + 2a_1 + a_2)}{4} \frac{(b_0 + 2b_1 + a_2)}{4} \frac{(c_0 + 2c_1 + c_2)}{4} \ldots
\]

but let us make the notation:
\[ \bar{a} = \frac{(a_0 + 2a_1 + a_2)}{4} \]
\[ \bar{b} = \frac{(b_0 + 2b_1 + b_2)}{4} \]
\[ \bar{c} = \frac{(c_0 + 2c_1 + c_2)}{4} \]

and so on.

It follows that

\[ a_1 = \bar{a} + A_1 \]
\[ b_j = \bar{b} + B_j \]
\[ c_k = \bar{c} + C_k \]

so we have to prove

\[ (\bar{a} + A_1) (\bar{b} + B_j) (\bar{c} + C_k) \ldots \]

\[ = ABC \ldots + A_1BC \ldots + AB_jC \ldots + A_1B_jC \ldots \]
\[ + ABC_k \ldots + A_1BC_k \ldots + AB_jC_k \ldots + A_1B_jC_k \ldots \]

It is clear that the right-hand side of this equation is the expansion of the left-hand side, for we may first remove the factor \((\bar{a} + A_1)\) from the right-hand side, then the factor \((\bar{b} + B_j)\) and so on to exhaust completely the right-hand side. Hence it follows that

\[ a_1 b_j c_k \ldots = \mu + A_1 + B_j + A_1 B_j + C_k + A_1 C_k + B_j C_k + A_1 B_j C_k \ldots \]
C. Discussion of the Factorial Model

For the general case of n loci, the factorial model is given by the equation,

$$a_{ijk} = \mu + A_i + B_j + C_k + A_iB_j + A_iC_k + B_jC_k + A_iB_jC_k + \ldots$$

which was developed previously. This model has all the desirable features mentioned earlier. Thus we may note

(a) Additivity of all the parameters is obvious because all possible interactions as well as the effects of the individual loci, are displayed in an additive manner.

(b) We can show that the parameters have genetical meaning by the following:

(i) The basic gene effect of any locus r, say, is

$$\frac{1}{2}(R_2 - R_0),$$

where $R_2$ is the effect of RR

$R_0$ is the effect of rr.

(ii) The dominance effect at locus r is

$$\frac{1}{2}(2R_1 - R_2 - R_0)$$

where $R_1$ is the effect of Rr

and $R_2$ and $R_0$ are denoted as in (i) above.

Because of the relation $R_0 + 2R_1 + R_2 = 0$, the dominance effect is given by $2R_1$. Hence complete lack of dominance, that is the heterozygote being exactly half way between the homozygotes, is indicated by $R_1$ equal to zero. Deviations in the dominance effect from the case of no dominance
are indicated by values for $R_1$ unequal to zero.
Partial dominance is indicated by $R_1$ lying between $R_0$ and $R_2$ but not exactly half way between them.
Overdominance is indicated by
\[
\frac{(4R_1)}{(R_2-R_0)} \quad \text{being greater than one.}
\]

(c) We have shown that the model applies for any number of loci in its development.

(d) The symmetry is best seen when we look at the original definitions of $A_0$ and $A_2$. It has been given for example that
\[
A_0 = \frac{(a_0 b_1 + 2a_1 b_0)}{4} - \mu
\]
and
\[
A_2 = \frac{(a_2 b_0 + 2a_0 b_1 + a_1 b_2)}{4} - \mu.
\]
If we replace $a_0$ by $a_2$ in the expression for $A_0$ we get $A_2$, and if we replace $a_2$ by $a_0$ in the expression for $A_2$ we get $A_0$. Hence the factorial model is symmetric.

(e) We can see that the model is adaptable to increasing assumptions with respect to epistasis by noting that when any combination of loci is assumed not to interact we merely drop the corresponding interaction terms from the model and the model retains, without further computations,
its additive structure. Examples given below explain the
feature, adaptability of the model with respect to assump-
tions, in more detail.

A few examples which show the desirable features of the
factorial model follow:

(a) If there is no epistasis, that is no interaction, then the
model becomes

\[ a_{ik} b_{jk} = \mu + \alpha_{i} + \beta_{j} + \gamma_{k} + \ldots, \]

that is the model is additive in the loci effects.

(b) If all the loci have the same effect and interactions, we
may then write the model as

\[ a_{ik} b_{jk} = \mu + \pi_{0} E_{0} + \pi_{1} E_{1} + \pi_{2} E_{2}, \]

where \( E_{0} \) is the effect of \( aa \)
\( E_{1} \) is the effect of \( Aa \)
\( E_{2} \) is the effect of \( AA \)
at each and every locus.

(c) If there is no dominance at say the \( r \) locus then \( R_{r} \) equals
zero so that \( R_{0} \) equals minus \( R_{2} \) and if this held for all
loci we would have

\[ a_{ik} b_{jk} = \mu + \sum_{(2)} R - \sum_{(0)} R, \]

where:

\[ \sum_{(2)} R = \text{summation of basic gene effects over loci} \]

for which the subscript is 2.
\[ \sum_{(6)} R = \text{summation of basic gene effects over loci for which the subscript is 0.} \]

Alternatively, let \( r \) be the basic gene effect for locus \( R \), or

\[ r = \frac{1}{2}(R_2 - R_0). \]

It follows that

\[ a_{i,j^k} = \mu + \sum_{(2)} r - \sum_{(6)} r \]

where:

\[ \sum_{(2)} r = \text{sum of basic gene effects for the RR loci.} \]

\[ \sum_{(6)} r = \text{sum of basic gene effects for the rr loci.} \]
D. The Scope of the Body of the Thesis

The factorial model which has been developed will be applied to certain breeding situations. In the next section the model will be applied to an $F_1$ population and to succeeding populations obtained by repeated self-fertilization. This will result in a sequence of expected means in terms of the parameter in the model and these formulas will be applied to some data.

In the succeeding section the model will be applied to two homozygous populations, the $F_1$ population derivable from these, the two backcross populations, the $F_2$ population and some other populations derivable from the two parental populations. In this section these results will be applied to the consideration of scaling tests given by Mather (13), the aim of which is to determine if epistasis is present.

The next section will deal with the case of three inbred parental populations and some of the possible populations which can be obtained from these by crossing.
IV. THE APPLICATION OF THE FACTORIAL MODEL TO
SELF-FERTILIZATION

We consider the case in which only one heterozygous individual
is used as a parent and is self-fertilized for any number of genera-
tions, m. This method of breeding is most restrictive because the
genetic material must be plants, and many precautions must be taken
to insure self-fertilization of the plants used. There are
advantages in studying self-fertilized material, however, because
the regular pattern of breeding allows a general mathematical
treatment of some genetical problems. The problems that will be
dealt with here are:

(a) Reasons for the inbreeding depression, which is quite
    apparent when cross fertilized genetic material is self-
    fertilized;

(b) Estimation of genetic parameters, especially epistatic
    parameters; and

(c) Prediction of the results of certain crosses, using the
    estimates obtained in (b).

It is of special interest to deal with quantitative inheritance
problems by the most general method possible, because of the com-
plexity of this type of inheritance and because of the unknown
number of loci involved. A general mathematical scheme in which n
loci and m generations of inbreeding are considered will be used to
investigate the three genetical problems outlined above.

In any one heterozygous individual there are two sets of loci, one set consisting of all the homozygous loci and the other set consisting of all the heterozygous loci. Therefore the factorial representation of the genotypic value must contain

(a) The mean effect, \( \mu \),
(b) The loci effects of the homozygous set,
(c) The loci effects of the heterozygous set,
(d) The interactions among loci within the homozygous set,
(e) The interactions among loci within the heterozygous set, and
(f) The interactions among loci between the homozygous set and the heterozygous set.

Either of the two sets of loci may be empty. If the heterozygous set is empty, as is usually assumed for an inbred line, the factorial representation does not contain (c), (e) or (f); hence, the problems are trivial, because homozygotes breed true, for example the \( F_1 \)'s are the same as the parent. If the homozygous set is empty, the factorial representation does not contain (b), (d) or (f).

The problem of dealing with both a homozygous set and a heterozygous set of loci may be more clearly presented by the use of the following example:

The genotype = aaBbCCDDd.
The genotypic value is \( a_0 b_1 c_2 d_1 \).

The homozygous set is \( aaCC \).

The heterozygous set is \( BbDd \).

The factorial representation of the genotypic value contains

the following effects:

(a) \( \mu \)

(b) \( A_0 \cdot C_2 \)

(c) \( B_1 \cdot D_1 \)

(d) \( A_0 \cdot C_2 \)

(e) \( B_1 \cdot D_1 \)

(f) \( A_0 \cdot B_1, A_0 \cdot D_1, C_2 \cdot B_1, C_2 \cdot D_1, A_0 \cdot B_1 \cdot C_2, A_0 \cdot B_1 \cdot D_1, A_0 \cdot C_2 \cdot D_1, B_1 \cdot C_2 \cdot D_1, A_0 \cdot B_1 \cdot C_2 \cdot D_1 \).

The homozygous set, effects (b) and (d), will breed true, so that all its effects will remain constant under self-fertilization; hence we can incorporate effects (a), (b) and (d) into one group, called the \( K \) group. Thus, the problem resolves into studying (c), (e) and (f), while carrying \( K \) along in the manipulations. The genotypic value can be expressed as

\[
\begin{align*}
\text{a}_0 \text{b}_1 \text{c}_2 \text{d}_1 &= \mu + A_0 + B_1 + A_0 B_1 + C_2 + A_0 C_2 + B_1 C_2 + A_0 B_1 C_2 + A_0 D_1 + B_1 D_1 + \\
&+ A_0 B_1 D_1 + A_0 C_2 D_1 + B_1 C_2 D_1 + A_0 B_1 C_2 D_1 \\
&= \mu + (A_0 + C_2 + A_0 C_2) + B_1 + (A_0 + C_2 + A_0 C_2) + D_1 + D_1 (A_0 + C_2 + A_0 C_2) \\
&+ B_1 D_1 + B_1 D_1 (A_0 + C_2 + A_0 C_2).
\end{align*}
\]

Let \( A_0 + C_2 + A_0 C_2 = K \) and \( \mu + A_0 + C_2 + A_0 C_2 = k_1 \).
Then it follows that

\[ a_0b_1c_2d_1 = K_1 + E_1 + D_1 + KD_1 + F_1 + D_1 + KB_1 D_1. \]

Hence the genotypic value is broken up into a constant \((K_1)\), loci effects of variable loci \((E_1 + D_1)\), interactions of fixed loci and one variable locus \((KD_1)\), interaction of variable loci \((F_1 + D_1)\) and interactions of the variable loci with the fixed loci \((KB_1 D_1)\).

Using this breakdown as a basis of analysis we could write

\[ a_0b_1c_2d_1 = K_1 + \beta_1 + \beta_2 \]

where

\[ \beta_1 = E_1 + D_1 + K(E_1 + D_1) = (E_1 + D_1)(1+K) \]
\[ \beta_2 = E_1 D_1 (1+K). \]

The parameter \(\beta_1\) measures the effect due to loci effects of variable loci plus the effect of these loci effects interacting with fixed loci, and it is entirely justifiable to include these interaction effects with the main effects. The parameter \(\beta_2\) measures the interaction of two variable loci plus the interaction of the two variable loci with the fixed loci.

In general suppose the fixed loci are \(f_1,f_2,\ldots,f_x\) and the variable loci are \(g_1,g_2,\ldots,g_z\), where the two groups of loci are mutually exclusive, then the model for the genotypic value is

\[ a_0b_1c_2d_1 = K_1 + \beta_1 + \beta_2 + \ldots + \beta_z \]

where

\[ K = \mu + K, \]
\[ K = \sum_{i=1}^{x} f_i + \sum_{i,j}^{x} f_i f_j + \sum_{i,j,k}^{x} f_i f_j f_k + \ldots. \]
$$\sum_{i=1}^{k} F_i = \text{the sum of all fixed loci effects},$$

$$\sum_{i \neq j}^{k} F_i F_j + \sum_{i \neq j \neq k}^{k} F_i F_j F_k + \ldots = \text{the sum of all the possible interactions among fixed loci},$$

$$\beta_1 = \frac{s}{p=1} G_p + K(\sum_{p=1}^{s} G_p),$$

$$\sum_{p=1}^{s} G_p = \text{the sum of all variable loci effects},$$

$$\beta_2 = \sum_{p \neq q} G_p G_q + K(\sum_{p \neq q} G_p G_q),$$

$$\sum_{p \neq q} G_p G_q = \text{the sum of all the two factor interactions of possible pairs of variable loci},$$

$$\vdots$$

$$\beta_s = (G_1 G_2 \ldots G_s) + K(G_1 G_2 \ldots G_s),$$

$$(G_1 G_2 \ldots G_s) = \text{the s factor interaction of all s variable loci.}$$

Given the initial heterozygote, we shall obtain expressions for the $F_2$, $F_3$ and so on, in terms of the factorial model. To do this it is necessary to determine the frequency array of genotypes in say the $m$th generation and then combine these genotypic frequencies with the genotypic values given by the model. We shall show how the genotypic array in the $m$th generation can be obtained in a simple manner for an arbitrary initial array of genotypes. This is not essential when the initial array consists simply of the complete
heterozygote but it is preferable to obtain the general form for use in more complex problems.

We have to find the relation

$$p_{m,n} = G(m,n) p_{0,n}$$

where: $p_{0,n}$ = the initial frequency array of genotypes arranged in a column matrix,

$G(m,n)$ = the operation by which we get from $p_{0,n}$ to $p_{m,n}$, and

$p_{m,n}$ = the frequency array of the genotypes in the $m$th generation in a column matrix.

It is known that if

$$p_{1,n} = G(1,n) p_{0,n} = G_n p_{0,n}$$

where $G(1,n) = G_n$ is a $3^{n} \times 3^{n}$ matrix, then

$$G(m,n) = G_n^m.$$ 

The matrix $G_n$ is known as the generation matrix. We therefore have to find $G_n$ and it is first necessary to specify the order in which the frequencies of the genotypes are arranged in the $3^n \times 1$ column matrix. The order which will be followed is the lexicographical order exemplified by

000, 001, 002, 010, 011, 012, 020, 021, 022 and so on.

Denote the ordered set of genotypes for the $(r-1)$ loci by $a$.

Then the three classes are

$\alpha 0,$

$\alpha 1,$
and α2.

Suppose that the generation matrix for (r-1) loci is \( G_{r-1} \), and consider the selfing of the above three classes. Because the loci are segregating independently, the segregation of any class, say \( αi \), is the product of the segregations of the class \( α \) and the genotype \( i \). The class α0 does not segregate at the \( r \)th locus.

If \( p_0 \) denotes the column array of genotypic frequencies and \( (α0) \) denotes the column of genotypes, the initial genotypic array is \( p_0 (α0) \), where \( (α0) \) denotes the row of genotypes or the transpose of \( (α0) \). The genotypic array resulting from selfing is

\[
G_{r-1} p_0 (α0)'.
\]

The class α0 does not give rise to any individuals in α1 or α2.

The class α1 gives \( \frac{1}{4} \) 0's, \( \frac{1}{2} \) 1's and \( \frac{1}{4} \) 2's at the \( r \)th locus. If \( q_0 \) denotes the column array of the genotypic frequencies and \( (α1) \) the column array of genotypes, the initial genotypic array is \( q_0 (α1) \). On selfing, this array gives

\[
G_{r-1} q_0 (α1)'.
\]

for the first (r-1) loci. Hence, in all, the class α1 gives

\[
\frac{1}{4} G_{r-1} q_0 (α0)', \quad \frac{1}{2} G_{r-1} q_0 (α1)', \quad \text{and} \quad \frac{1}{4} G_{r-1} q_0 (α2)'.
\]

As in class α0, class α2 does not segregate at the \( r \)th locus. If \( t_0 \) denotes the column array of genotypic frequencies and \( (α2) \) denotes the column of genotypes, the initial genotypic array is \( t_0 (α2) \). The genotypic array resulting from selfing is

\[
G_{r-1} t_0 (α2)'.
\]
The class a2 does not give rise to any individuals in a0 or a1.

Hence the genotypic array after one generation of selfing is, say, $p_1(a0)' + q_1(a1)' + t_1(a2)'$, where $p_1(a0) + q_1(a1) + t_1(a2) = \frac{1}{4} G_{r-1} p_0(a0)' + \frac{1}{2} G_{r-1} q_0(a0)' + \frac{1}{2} G_{r-1} q_0(a1)' + \frac{1}{4} G_{r-1} q_0(a2)' + G_{r-1} t_0(a2)'.

This is an identity in the $(ai)'$, $i = 0,1,2$, hence

$$\begin{bmatrix}
p_1 \\
qu_1 \\
t_1
\end{bmatrix} = \begin{bmatrix}
\frac{1}{4} G_{r-1} & \frac{1}{4} G_{r-1} & 0 \\
0 & \frac{1}{2} G_{r-1} & 0 \\
0 & \frac{1}{4} G_{r-1} & G_{r-1}
\end{bmatrix} \begin{bmatrix}
p_0 \\
qu_0 \\
t_0
\end{bmatrix}.$$  

It follows that the generation matrix for $r$ loci is defined by

$$G_r = \begin{bmatrix}
\frac{1}{4} G_{r-1} & \frac{1}{4} G_{r-1} & 0 \\
0 & \frac{1}{2} G_{r-1} & 0 \\
0 & \frac{1}{4} G_{r-1} & G_{r-1}
\end{bmatrix}.$$  

Since $r$ may run from 2 to $n$, we may express the generation matrix of the $n$ loci case as

$$G_n = \begin{bmatrix}
\frac{1}{4} G_{n-1} & \frac{1}{4} G_{n-1} & 0 \\
0 & \frac{1}{2} G_{n-1} & 0 \\
0 & \frac{1}{4} G_{n-1} & G_{n-1}
\end{bmatrix}.$$  

We can summarize our result by stating that the column matrix of genotypic frequencies in generation $m$ for $n$ loci, $p_m, q_m, t_m$, is given by
The process of forming powers of matrices directly is tedious and it is not possible to write down a general formula. The way out of the difficulty is to develop $G_n^m$ in terms of matrices of its characteristic roots and characteristic vectors.

We know that

$$G_r V_r = V_r A_r$$

where:  
$V_r$ = the matrix of characteristic column vectors of $G_r$  
$A_r$ = the matrix of characteristic roots of $G_r$, which is diagonal.

Then

$$G_r = G_r V_r V_r^{-1} = V_r A_r V_r^{-1}.$$  

When $r = n-1$, we have $G_{n-1} = V_{n-1} A_{n-1} V_{n-1}^{-1}$ and if we substitute this for $G_{n-1}$ in the matrix, $G_n$, we may write

$$G_n = \begin{bmatrix}
V_{n-1} A_{n-1} V_{n-1}^{-1} & \frac{1}{4} V_{n-1} A_{n-1} V_{n-1}^{-1} & 0 \\
0 & \frac{1}{2} V_{n-1} A_{n-1} V_{n-1}^{-1} & 0 \\
0 & \frac{1}{4} V_{n-1} A_{n-1} V_{n-1}^{-1} & V_{n-1} A_{n-1} V_{n-1}^{-1}
\end{bmatrix}.$$  

We shall now find the $n^{th}$ power of the generation matrix when $n$ loci are involved. The generation matrix for selfing with one locus is
Following the algebraic procedures given by Bartlett (1), page 18, we may write the matrix of characteristic column vectors as

\[
G_1 = \begin{bmatrix}
1 & 1 & 1 \\
0 & 1/2 & 0 \\
0 & 1/4 & 1 \\
\end{bmatrix}
\]

The matrix of characteristic roots for one generation of selfing is

\[
\Lambda_1 = \begin{bmatrix}
1/2 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
\end{bmatrix}
\]

of which the \(m\)th power is

\[
\Lambda_1^m = \begin{bmatrix}
1/2^m & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
\end{bmatrix}
\]

Since \(G_1 = V_1 \Lambda_1 V_1^{-1}\),

\[
c_{1}^{m} = V_1 \Lambda_1^m V_1^{-1}
= \begin{bmatrix}
1 & (1/2 - 1/2^m) & 0 \\
0 & 1/2^m & 0 \\
0 & (1/2 - 1/2^m) & 1 \\
\end{bmatrix}
\]

We found
and it can be verified that the matrix of characteristic roots of $G_r$ is

$$\Lambda_r = \begin{bmatrix}
\frac{1}{2} \Lambda_{r-1} & 0 & 0 \\
0 & \Lambda_{r-1} & 0 \\
0 & 0 & \Lambda_{r-1}
\end{bmatrix}, \text{ so that}
$$

$$\Lambda^m_r = \begin{bmatrix}
\frac{1}{2^m} \Lambda^m_{r-1} & 0 & 0 \\
0 & \Lambda^m_{r-1} & 0 \\
0 & 0 & \Lambda^m_{r-1}
\end{bmatrix}.$$ 

Using the known relationship

$$G_{r-1} V_{r-1} = V_{r-1} \Lambda_{r-1}$$

and the matrix $\Lambda_r$, we find that $V_r$ is given by

$$V_r = \begin{bmatrix}
V_{r-1} & V_{r-1} & V_{r-1} \\
-2V_{r-1} & 0 & 0 \\
V_{r-1} & V_{r-1} & -V_{r-1}
\end{bmatrix}.$$ 

Since $G_r V_r = V_r \Lambda_r$, ...
After some algebraic manipulations we obtain $G^n_r$, the $n$th power of the generation matrix when $n$ loci are involved, as

$$G^n_r = v_r \wedge^n_r = 
\begin{bmatrix}
\frac{1}{2^m_r} v_{r-1} \wedge^{m-n} r-1 & v_{r-1} \wedge^{m-n} r-1 & v_{r-1} \wedge^{m-n} r-1 \\
0 & 0 & 0 \\
\frac{1}{2^m_r} v_{r-1} \wedge^{m-n} r-1 & v_{r-1} \wedge^{m-n} r-1 & -v_{r-1} \wedge^{m-n} r-1
\end{bmatrix}.
$$

We therefore have a recursion relationship on $G^n_r$. Thus

$$G^n_r = v_n \wedge^{m-n} n-1 v^{-1}_{n-1} =
\begin{bmatrix}
v_{n-1} \wedge^{m} n-1 v^{-1}_{n-1} & (\frac{1}{2} - \frac{1}{2^{m+1}}) v_{n-1} \wedge^{m} n-1 v^{-1}_{n-1} & 0 \\
0 & \frac{1}{2^m} v_{n-1} \wedge^{m} n-1 v^{-1}_{n-1} & 0 \\
0 & (\frac{1}{2} - \frac{1}{2^{m+1}}) v_{n-1} \wedge^{m} n-1 v^{-1}_{n-1} & v_{n-1} \wedge^{m} n-1 v^{-1}_{n-1}
\end{bmatrix}.
$$

with $G^n_0$ equal to unity. Hence we can write down the matrix $G^n_r$ by which we find the genotypic array resulting from $n$ generations of selfing of an arbitrary genotypic array.
We are concerned here only with the case when the initial
genotypic array consists of the completely heterozygous individual.
Either by the use of $g_n^m$ above or directly it can be seen that the
frequency of a genotype in the $n$th generation depends only on the
number of heterozygous loci. In fact the frequency of any geno­
type with $j$ heterozygous loci is

$$\left[ \frac{1}{2} \left( 1 - \frac{1}{2^m} \right) \right]^{n-j} \left[ \frac{1}{2^m} \right]^j.$$

The next problem is to combine the frequencies of the possible
genotypes with the coefficients of these parameters of the model in
such a way that we obtain the coefficients of the loci effects and
interactions for the mean genotypic value of the population. An
example will illustrate the problem and method of attack. Consider
the case of two loci. The genotypic values may be expressed as

- $a_0^2b_0 = \mu + A_0 + B_0 + A_0B_0$
- $a_0b_1 = \mu + A_0 + B_1 + A_0B_1$
- $a_0b_2 = \mu + A_0 + B_2 + A_0B_2$
- $a_1b_0 = \mu + A_1 + B_0 + A_1B_0$
- $a_1b_1 = \mu + A_1 + B_1 + A_1B_1$
- $a_1b_2 = \mu + A_1 + B_2 + A_1B_2$
- $a_2b_0 = \mu + A_2 + B_0 + A_2B_0$
- $a_2b_1 = \mu + A_2 + B_1 + A_2B_1$
- and $a_2b_2 = \mu + A_2 + B_2 + A_2B_2$
Among the nine genotypes there are the following three classes:

(a) No heterozygous locus, \( a_0 b_0 + a_0 b_2 + a_2 b_0 + a_2 b_2 \),

(b) One heterozygous locus, \( a_1 b_1 + a_1 b_0 + a_2 b_1 + a_1 b_2 \) and

(c) Two heterozygous loci, \( a_1 b_1 \).

This is a classification of genotypes according to the presence or absence of heterozygous loci. The general expression for the frequency of any genotype with \( m \) loci, \( j \) of which are heterozygous, in a population arising by selfing the completely heterozygous individual for \( n \) generations, is

\[
\left[ \frac{1}{2}(1 - \frac{1}{2^m}) \right]^{n-j} \left( \frac{1}{2^m} \right)^j.
\]

The genotypic frequencies and the total of the genotypic values for the three classes of genotypes are

<table>
<thead>
<tr>
<th>Class (h)</th>
<th>Number of Genotypes</th>
<th>Genotype frequency ((f_h))</th>
<th>Total of genotypic values per class ((y_h))</th>
</tr>
</thead>
</table>
| (a)       | 4                   | \[
\left[ \frac{1}{2}(1 - \frac{1}{2^m}) \right]^2
\]
|           |                     | \(4a_0^4 + 4a_1^4 + 4a_2^4 + 4a_1^2a_2^2\) |
| (b)       | 4                   | \[
\left[ \frac{1}{2}(1 - \frac{1}{2^m}) \right]^2 \left( \frac{1}{2^m} \right)
\]
|           |                     | \(4a_1^4 + 4a_2^4 + 4a_1^2a_2^2\) |
| (c)       | 1                   | \[
\left( \frac{1}{2^m} \right)^2
\]
|           |                     | \(4a_1^4 + 4a_2^4 + 4a_1^2a_2^2\) |

We note that the total frequency

\(4f_1 + 4f_2 + f_3 = 1\).

The expressions for \((y_h)\) have been simplified by the use of the identities
\[ A_0 + 2A_1 + A_2 = B_0 + 2B_1 + B_2 = A_1B_0 + 2A_1B_1 + A_1B_2 = A_0B_0 + 2A_1B_1 + A_2B_2 = 0, \]

and

\[ \sum_{i=0}^{2} \sum_{j=0}^{2} y_{ij} f_{ij} = \sum_{i=1}^{3} y_{ih} f_{ih}. \]

It follows that

\[ \sum_{i=1}^{3} y_{ih} f_{ih} = \mu + \left( \frac{1}{2^{m-1}} - 1 \right) \left[ A_1 + B_1 \right] + \left[ \frac{1}{2^{m-1}} - 1 \right] ^2 \left[ A_1B_1 \right] \]

where we could call \( \mu = X_1, A_1 + B_1 = \beta_1 \) and \( A_1B_1 = \beta_2 \).

In exactly the same way we find that if we self-fertilize a dihybrid which contains one heterozygous locus for \( m \) generations we obtain

\[ \sum_{i=1}^{3} y_{ih} f_{ih} = X_1 + \left( \frac{1}{2^{m-1}} - 1 \right) \beta_1 \]

where \( X_1 = \mu \) plus the locus effect of the homozygous locus

\( \beta_1 = \) the locus effect of the variable locus plus the interaction of the variable locus and the fixed locus.

If we self-fertilize an individual homozygous at both its loci, there is no segregation so we obtain

\[ \sum_{i=1}^{3} y_{ih} f_{ih} = X_1 \]

where \( X_1 = \mu + A_0 + B_0 + A_0B_0 \).

If we self-fertilize a dihybrid for one generation, that is \( m = 1 \),

\[ \sum_{i=1}^{3} y_{ih} f_{ih} = X_1 \]
no matter whether the parental genotype has none, one or two heterozygous loci. This indicates that the genotypic value $F_1$, the mean of the genotypic values after one generation of self-fertilization, provides an estimate of the fixed effects. If we self-fertilize a dihybrid indefinitely, that is let $m$ tend to infinity,

$$
\sum_{h=1}^{3} y_{n,n} = x_1 - \beta_1 + \beta_2
$$

where $x_1$, $\beta_1$ and $\beta_2$ are defined for the specific cases as before.

We will next derive the coefficients of the loci effects and interactions for the mean genotypic value of the population which comes from a completely heterozygous $n$-loci genotype which is self-fertilized for $m$ generations. Symmetry as regards the loci in any $F_m$ population makes the approach which follows possible. For example, in considering the $p$-factor interactions we need only consider the $p$-factor interactions arising from one set of $p$ loci. All the other $p$-factor interactions will have the same coefficient because of this symmetry.

First we shall obtain the coefficient of the loci effects. We shall take the case in which the first $j$ loci are heterozygous and the other $(n-j)$ loci are homozygous. There are $2^{n-j}$ different genotypes which have the first $j$ loci heterozygous and the other $(n-j)$ loci homozygous. Hence there will be $2^{n-j}$ loci effects which come from the heterozygous loci, for example $A_1$, but only
loci effects which come from the homozygous loci, for example $L_0$ and $L_2$. The reason there are only half as many loci effects from the homozygous loci as from the heterozygous loci is that either $L_0$ or $L_2$ can occur in one genotype but not both and each occurs with equal frequency. The sum of the loci effects for this ordered arrangement is

\[
\left[ \frac{2^{n-j}}{2} \right] = 2^{n-j-1}
\]

\[2^{n-j}(A_1+B_1+\ldots+J_1) + 2^{n-j-1}(L_0+L_2+\ldots+N_0+N_2) =
\]

\[2^{n-j}(A_1+B_1+\ldots+J_1) + 2^{n-j-1}(-2L_1\ldots-2N_1) =
\]

\[2^{n-j}(A_1+B_1+\ldots+J_1) - 2^{n-j}(H_1-A_1-B_1-\ldots-J_1) =
\]

\[2^{n-j-2}(A_1+B_1+\ldots+J_1) - 2^{n-j}M_1,
\]

where $M_1 = (A_1+B_1+\ldots+H_1)$.

There are \( \binom{n}{j} \) possible ways in which \( j \) of the \( n \) loci may be heterozygous and as a result the expression for the loci effects is

\[2^{n-j} \left[ 2^{(n-1)}(A_1+B_1+\ldots+H_1) - \binom{n}{j}M_1 \right]
\]

because any one contribution, say $D_1$, occurs for \( \binom{n-1}{j-1} \) genotypes. However,

\[M_1 = A_1+B_1+\ldots+H_1\], by definition,

so the expression above reduces to

\[2^{n-j} \left( \frac{2^j}{\binom{n}{j}} - 1 \right) \binom{n}{j} M_1, \quad \text{for} \ 0 \leq j \leq n,
\]

which may be written \( \binom{n}{j} M_1 \).
We found that after $m$ generations of selfing a completely heterozygous individual with $n$ loci, the frequency of any genotype with $j$ heterozygous loci is

$$\left[ \frac{1}{2}(1 - \frac{1}{2^n}) \right]^{n-j} \left( \frac{1}{2^n} \right)^j, \text{ for } 0 \leq j \leq n.$$ 

Hence multiplying this frequency by the coefficient of $H_1$ and summing over $j$ from zero to $n$, we obtain

$$\frac{1}{n} \sum_{j=0}^{n} \binom{n}{j} \left(1 - \frac{1}{2^n}\right)^{n-j} \left(\frac{1}{2^n}\right)^j = \frac{2}{n} \sum_{j=0}^{n} \binom{n}{j} (1 - \frac{1}{2^n})^{n-j} \left(\frac{1}{2^n}\right)^j.$$ 

However,

$$\frac{1}{n} \sum_{j=0}^{n} \binom{n}{j} (1 - \frac{1}{2^n})^{n-j} \left(\frac{1}{2^n}\right)^j = \frac{n}{2^n}.$$ 

Hence the coefficient of the loci effects for any number of loci using the factorial model is

$$\left[ \frac{2}{n} \left(\frac{1}{2^n} - 1\right) \right] = \left[ \frac{1}{2^{m-1}} - 1 \right].$$

We note that this coefficient is the same as that obtained for the dihybrid loci effects, $\beta_1$, and does not depend upon the number of loci in the genotype.

The contribution of the $p$-factor interactions to any $F_m$ mean is equal to

$$\left(\frac{1}{2^n-1} - 1\right)^P \sum_{\text{all } p\text{-factor interactions symbols of type } A_1 B_1 C_1 \ldots P_1}.$$
This can be shown generally in the following way. In view of the symmetry with respect to loci, we need only consider the p-factor interactions arising from one set of the p loci. We consider the interactions arising from the first p loci for which the letters are say a, b, c, ..., p. The total genotypic array in the mth generation is

\[
\left[ \frac{1}{2} (1-q)aa + qAa + \frac{1}{2} (1-q)AA \right] \left[ \frac{1}{2} (1-q)bb + qBb + \frac{1}{2} (1-q)BB \right] \ldots \\
\ldots \left[ \frac{1}{2} (1-q)pp + qPp + \frac{1}{2} (1-q)PP \right] \ldots
\]

and so on, where q is equal to \( \frac{1}{2^m} \).

The total frequency over all phases for loci other than the a, b, c, ..., p loci is unity. To evaluate the mean genotypic value we expand the above product and then insert the genotypic value as given by the model for each genotype. We consider only the p-factor interaction contributions from loci a, b, c, ..., p. The mean contribution is obtained by expanding

\[
\left[ \frac{1}{2} (1-q)A_0 + qA_1 + \frac{1}{2} (1-q)A_2 \right] \left[ \frac{1}{2} (1-q)B_0 + qB_1 + \frac{1}{2} (1-q)B_2 \right] \ldots \\
\ldots \left[ \frac{1}{2} (1-q)P_0 + qP_1 + \frac{1}{2} (1-q)P_2 \right].
\]

From the method of construction of the model if \( X, \ Y, \ Z \ldots \) and so on are symbols such as \( E_1, \ C_2 \) and so on. We know that

\[
A_0 XYZ\ldots + 2A_1 XYZ\ldots + A_2 XYZ\ldots = 0.
\]

Hence \((A_0 + A_2) XYZ\ldots = -2A_1 XYZ\ldots \).

We may therefore substitute for \((A_0 + A_2)\) in the first factor by
means of this expression, likewise in the other factors to get
\[
\sum_{\text{first } p \text{ loci}} \left[ q A_1 + \left( \frac{1-q}{2} \right) (A_0 + A_2) \right] =
\]
\[
\sum_{\text{first } p \text{ loci}} \left[ q A_1 - (1-q) A_1 \right] =
\]
\[
\sum_{\text{first } p \text{ loci}} (2q-1) A_1 =
\]
\[
(2q-1)^p A_1 B_1 C_1 \cdots P_1 .
\]

where \( q \) is the frequency of a heterozygous allele after
\( m \) generations of selfing or is \( \frac{1}{2^m} \).

Substituting \( \frac{1}{2^m} \) for \( q \) in the last expression above we obtain
\[
\left( \frac{1}{2^{m-1}} - 1 \right)^p A_1 B_1 C_1 \cdots P_1 .
\]

We now sum over all possible \( p \) loci and let
\[
\beta_p = \sum_{\text{all sets of } p \text{ loci}} A_1 B_1 C_1 \cdots P_1 .
\]

We find that the \( p \)-factor interaction contribution is
\[
\left( \frac{1}{2^{m-1}} - 1 \right)^p \beta_p .
\]

This proof also holds for the interaction of the fixed loci with
any \( p \)-factor interaction of the variable loci.

It follows that the mean genotypic value of the population
arising from selfing a completely heterozygous individual for \( m \)
generations can be summarized as:
\[
\overline{F}_m = \mu + \sum_{p=1}^{n} \left( \frac{1}{2^{m-1}} - 1 \right) \beta_p
\]

where \(\overline{F}_m\) = the mean genotypic value of the population obtained after \(m\) generations of self-fertilization of any given genotype,

\(\mu\) = the overall mean, plus the loci effects of the fixed loci, and

\(\beta_p\) = the \(p\)-factor interaction of the variable loci plus the interaction of the \(p\)-factor interactions with the fixed loci.

Hence

\[
\overline{F}_0 = \mu + \sum_{p=1}^{n} \beta_p , \quad \text{where } \overline{F}_0 \text{ is the original heterozygous parent,}
\]

\[
\overline{F}_1 = \mu
\]

\[
\overline{F}_2 = \mu + \sum_{p=1}^{n} \left( -\frac{1}{2} \right)^p \beta_p
\]

\[
\overline{F}_3 = \mu + \sum_{p=1}^{n} \left( -\frac{3}{4} \right)^p \beta_p
\]

\[
\vdots
\]

\[
\overline{F}_\infty = \mu + \sum_{p=1}^{n} (-1)^p \beta_p .
\]

The algebraic value of the coefficient of the loci effects of the variable loci and the interaction of each variable locus with
the fixed loci \( (\beta_1) \) decreases as the number of generations of self-fertilization increases. The coefficient of the two-factor interactions of the variable loci and the interaction of every possible combination of two of the variable loci with the fixed loci \( (\beta_2) \) decreases to the \( F_2 \) and increases slowly as the number of generations of self-fertilization increases from \( F_2 \). The coefficient of \( \beta_3 \) acts as the coefficient of \( \beta_1 \); however, the algebraic value of the coefficient of \( \beta_3 \) is much smaller than the coefficient of \( \beta_1 \) beyond the \( F_1 \).

The \( F_1 \) corresponds to the mean of the usual \( F_2 \) population. Fisher, Immer and Tedin (8) found the variance of the \( F_2 \) to be \( \left( \frac{\mu^2}{4} + \frac{h^2}{4} \right) \). We will obtain a corresponding formula for the factorial model. When there is no differential fertilization or viability, the \( F_2 \) for a segregating locus, \( a \), will be

\[
\begin{align*}
\frac{1}{4} \text{ AA } & ; \frac{1}{2} \text{ Aa } ; \frac{1}{4} \text{ aa },
\end{align*}
\]

and we may set up the following:

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Genotype</th>
<th>Factorial representation of the genotypic value</th>
<th>Deviation from mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \frac{1}{4} )</td>
<td>aa</td>
<td>( \mu + A_0 )</td>
<td>( A_0 )</td>
</tr>
<tr>
<td>( \frac{1}{2} )</td>
<td>Aa</td>
<td>( \mu + A_1 )</td>
<td>( A_1 )</td>
</tr>
<tr>
<td>( \frac{1}{4} )</td>
<td>AA</td>
<td>( \mu + A_2 )</td>
<td>( A_2 )</td>
</tr>
</tbody>
</table>

It follows that the variance of the \( F_2 \) is

\[
\frac{1}{4}(A_0^2 + 2A_1^2 + A_2^2).
\]
This corresponds to the expression obtained by Fisher, et al. (8).

Fisher's d and h are given by

\[ 2d = A_2 - A_0 \]

\[ 2h = 2A_1 - A_0 - A_2 = 4A_1 \]

or

\[ h = 2A_1 = -(A_0 + A_2) . \]

The quantity

\[ \frac{1}{4} \left\{ A_0^2 + 2A_1^2 + A_2^2 \right\} \]

is equal to

\[ \frac{1}{4} \left\{ (A_0 - A_2)^2 + 2A_1^2 + 2A_0A_2 \right\} \]

\[ = \frac{1}{4} \left\{ (A_0 - A_2)^2 + 2A_1^2 \right\} + \frac{1}{2} \left[ (A_0 + A_2)^2 - (A_0 - A_2)^2 \right] \]

\[ = \frac{1}{4} \left\{ 4d^2 + h^2 + \frac{h^2}{2} - 2d^2 \right\} \]

\[ = \frac{1}{4} \left( 2d^2 + h^2 \right) . \]

For \( n \) loci the procedure for finding the variance of the \( F_2 \) is the same. The mean of the \( F_2 \) is always \( \mu \) and the variance is obtained by squaring each deviation of the genotypic value from \( \mu \), multiplying by the frequency and summing. When \( n \) is greater than one, interaction terms will enter, but the cross products of all the loci effects and all the interactions sum to zero because

\[ A_0 \cdot XYZ... + 2A_1 \cdot XYZ... + A_2 \cdot XYZ... = \]

\[ (A_0 + 2A_1 + A_2)(XYZ...) = 0 \]

where the dot (.) denotes multiplication, for example, of the number \( A_0 \) by the number \( XYZ... \). It follows that we have

\[ V(F_2) = V_1 + V_2 + \ldots + V_n \]
where

\[ V_1 = \text{variance due to loci effects which is the sum of expressions like the one obtained in the one locus case,} \]

\[ V_2 = \text{variance due to interactions of two loci, which is the weighted sum of squares of terms like } A_i B_j, \]

\[ V_3 = \text{variance due to interactions of these loci which is the weighted sum of squares of terms like } A_i B_j C_k \]

and so on.

We will not discuss variances in the advanced generations beyond \( F_2 \) because the object of this thesis is more towards examination of scaling tests rather than towards estimation of degrees of dominance. Difficulties arise in the case of generations beyond the \( F_2 \) because loci effects and interactions enter multiplicatively.

For example, the \( F_3 \) variance for the case of two loci is

\[
\frac{3}{8}(A_0^2 + A_1^2 + B_0^2 + B_1^2) - \frac{3}{8}(A_0 A_1 B_0 + A_0 A_1 B_1 + A_1 B_0 + A_1 B_1 + A_0 B_0 + A_0 B_1 + A_1 B_0 + A_1 B_1) +
\]

\[
\frac{3}{64}(A_0 B_0)^2 + (A_1 B_0)^2 + (A_0 B_1)^2 + (A_1 B_1)^2.
\]

To examine explanations for inbreeding depression, we shall suppose that interactions involving three loci or more are trivial.

This corresponds to Fisher's (7) statement

\[ \ldots \text{it is very improbable that any statistical effect, of a nature other than that which we are considering is actually produced by more complex somatic connections,} \]

prior to which he had explained dual epistacy. Then we have
There is no particular reason for $\beta_1$ to be negative or positive. If $\beta_1$ is negative and we assume the interactions are negligible, we should get no inbreeding depression. In fact we would expect a slight increase rather than a decrease in the mean genotypic value with inbreeding. If $\beta_2$ were negative and large enough to influence the genotypic value, the inbreeding depression could be explained by epistasis. Hence an explanation of inbreeding depression must, it appears, give attention to epistasis as well as to loci effects.

It should perhaps be restated that the development given here is based on the complete absence of selection, including of course the absence of lethals, and on equal viability of all genotypes. The extent to which these contribute to inbreeding depression cannot be ascertained from the type of data we are considering.

If the assumption is made that $\beta_3$ and higher-factor interactions are negligible the parameters $X_1$, $\beta_1$ and $\beta_2$ can be estimated from the first three generations.

The first three equations will be
It follows that estimates are

\[ \hat{x}_1 = \bar{F}_1 \]

\[ \hat{\beta}_1 = \frac{1}{3} \bar{F}_0 + \bar{F}_1 - \frac{h}{3} \bar{F}_2 \]

\[ \hat{\beta}_2 = \frac{2}{3} \bar{F}_0 - 2\bar{F}_1 + \frac{4}{3} \bar{F}_2 \]
Table 1
Means of Selfed Generations

<table>
<thead>
<tr>
<th>Generation</th>
<th>Number of Observations</th>
<th>Length</th>
<th>Width</th>
<th>Shape = (Length/Width)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>arithmetic</td>
</tr>
<tr>
<td>F₀</td>
<td>300</td>
<td>51.3</td>
<td>25.1</td>
<td>2.04</td>
</tr>
<tr>
<td>F₁</td>
<td>900</td>
<td>48.9</td>
<td>25.0</td>
<td>2.10</td>
</tr>
<tr>
<td>F₂</td>
<td>898</td>
<td>51.5</td>
<td>23.4</td>
<td>2.42</td>
</tr>
<tr>
<td>F₃</td>
<td>897</td>
<td>50.1</td>
<td>23.2</td>
<td>2.40</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Generation</th>
<th>Green Fruit Weight (arithmetic)</th>
<th>Green Fruit Weight (logarithm)</th>
<th>Dry Fruit Weight (arithmetic)</th>
<th>Dry Fruit Weight (logarithm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F₀</td>
<td>59.6</td>
<td>1.771</td>
<td>10.1</td>
<td>1.000</td>
</tr>
<tr>
<td>F₁</td>
<td>61.7</td>
<td>1.758</td>
<td>9.6</td>
<td>0.963</td>
</tr>
<tr>
<td>F₂</td>
<td>56.3</td>
<td>1.694</td>
<td>9.1</td>
<td>0.930</td>
</tr>
<tr>
<td>F₃</td>
<td>53.7</td>
<td>1.673</td>
<td>8.8</td>
<td>0.915</td>
</tr>
</tbody>
</table>
The variance of $\hat{\beta}_2$ is $\frac{4}{9} \sigma_F^2 + \frac{16}{9} \sigma_{F_2}^2$ if the yields of $F_0$, $F_1$ and $F_2$ are independent. We may obtain an estimate of the standard error of $\hat{\beta}_2$ for each character by using the mean square of $F_0$, $F_1$ and $F_2$ appropriately.

The estimates of $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$ obtained from the populations usually called $F_1$, $F_2$ and $F_3$ when we assume $\beta_2$ is not zero and the estimated standard error of $\hat{\beta}_2$ (s.e. $\hat{\beta}_2$) are given in Table 2.

The following conclusions may be drawn:

(a) For length and shape of fruit, $\hat{\beta}_1$ is negative and $\hat{\beta}_2$ is positive. From our knowledge of the model we should expect no apparent depression. The data substantiate this view. For width and fruit weight $\hat{\beta}_1$ is positive and $\hat{\beta}_2$ is negative, except for dry fruit weight. We expect a decided depression, and the data uphold this view also.

(b) If we assume the sampling distribution of $\hat{\beta}_2$ to be normal and all tests to be independent, we would reject at the 5 per cent level the hypothesis that $\beta_2$ is zero in all cases except for dry fruit weight.

A better statistical procedure for analysis of these data is to obtain the least squares estimates of $\beta_1$, $\beta_1$ and $\beta_2$ from all four generations and obtain the standard errors of these estimates.
Table 2

Estimates of Loci and Epistatic Effects

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Length (arithmetic)</th>
<th>Width</th>
<th>Shape = length/width</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Arithmetic</td>
</tr>
<tr>
<td>$\hat{\xi}_1$</td>
<td>48.9</td>
<td>25.0</td>
<td>2.10</td>
</tr>
<tr>
<td>$\hat{\beta}_1$</td>
<td>-2.667</td>
<td>2.167</td>
<td>-0.4467</td>
</tr>
<tr>
<td>$\hat{\beta}_2$</td>
<td>5.067</td>
<td>-2.067</td>
<td>0.3867</td>
</tr>
<tr>
<td>s.e. $\hat{\beta}_2$</td>
<td>±0.131</td>
<td>±0.045</td>
<td>±0.085</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Green Fruit Weight (arithmetic)</th>
<th>Green Fruit Weight (logarithm)</th>
<th>Dry Fruit Weight (arithmetic)</th>
<th>Dry Fruit Weight (logarithm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\xi}_2$</td>
<td>61.7</td>
<td>1.758</td>
<td>9.60</td>
<td>0.963</td>
</tr>
<tr>
<td>$\hat{\beta}_1$</td>
<td>6.5</td>
<td>0.089</td>
<td>0.833</td>
<td>0.056</td>
</tr>
<tr>
<td>$\hat{\beta}_2$</td>
<td>-8.6</td>
<td>-0.077</td>
<td>-0.333</td>
<td>0.021</td>
</tr>
<tr>
<td>s.e. $\hat{\beta}_2$</td>
<td>±2.11</td>
<td>±0.015</td>
<td>±0.239</td>
<td>±0.011</td>
</tr>
</tbody>
</table>
Following the notation and method given by Kempthorne (11), chapter five, we will use the matrix notation

\[ y = X \beta + e \]

where

\[
\begin{bmatrix}
  \mathbf{F}_0 \\
  \mathbf{F}_1 \\
  \mathbf{F}_2 \\
  \mathbf{F}_3
\end{bmatrix},
\begin{bmatrix}
  1 & 1 & 1 \\
  1 & 0 & 0 \\
  1 & \frac{1}{2} & \frac{1}{4} \\
  1 & \frac{3}{4} & \frac{9}{16}
\end{bmatrix},
\begin{bmatrix}
  \beta_1 \\
  \beta_2
\end{bmatrix},
\begin{bmatrix}
  e_0 \\
  e_1 \\
  e_2 \\
  e_3
\end{bmatrix}
\]

and \( e_i (i = 0, 1, 2, 3) \) is the nonheritable portion of the \( i \)th phenotypic value.

The estimates are

\[
\begin{bmatrix}
  \hat{\beta}_1 \\
  \hat{\beta}_2
\end{bmatrix} = S^{-1}X'y
\]

where

\[
S = \begin{bmatrix}
  1 & 1 & 1 & 1 \\
  1 & 0 & \frac{1}{2} & \frac{3}{4} \\
  1 & 0 & \frac{1}{4} & \frac{9}{16} \\
  1 & \frac{3}{4} & \frac{9}{16}
\end{bmatrix}
\]

\[
\begin{bmatrix}
  1 & 1 & 1 \\
  1 & 0 & 0 \\
  1 & \frac{1}{2} & \frac{1}{4} \\
  1 & \frac{3}{4} & \frac{9}{16}
\end{bmatrix}
\]

\[
\begin{bmatrix}
  4 & -\frac{1}{4} & \frac{29}{16} \\
  -\frac{1}{4} & 4 & \frac{29}{16} \\
  \frac{29}{16} & \frac{29}{16} & 4
\end{bmatrix}
\]

The inverse of \( S \) is

\[
S^{-1} = \begin{bmatrix}
  .84193548 & .42795699 & -1.24731183 \\
  .42795699 & .81863799 & -.83154122 \\
  -1.24731183 & -.83154122 & 2.63799263
\end{bmatrix}
\]
so the estimates are given by

\[
\begin{bmatrix}
\hat{X}_1 \\
\hat{\beta}_1 \\
\hat{\beta}_2
\end{bmatrix} = \begin{bmatrix}
-0.02258064 \\
0.141505376 \\
-0.55913978
\end{bmatrix} F_0 + \begin{bmatrix}
-0.04193548 \\
0.42795699 \\
-1.24731183
\end{bmatrix} F_1 + \begin{bmatrix}
0.31635483 \\
-0.18924731 \\
-0.17204301
\end{bmatrix} F_2 + \begin{bmatrix}
-0.1505376 \\
-0.65376344 \\
0.86021505
\end{bmatrix} F_3.
\]

The estimates of the variances of \( \hat{X}_1 \), \( \hat{\beta}_1 \) and \( \hat{\beta}_2 \) are

\[
V(\hat{X}_1) = (0.02258064)^2 V(F_0) + (-0.04193548)^2 V(F_1) + (0.31635483)^2 V(F_2) + (-0.1505376)^2 V(F_3).
\]

\[
V(\hat{\beta}_1) = (0.141505376)^2 V(F_0) + (0.42795699)^2 V(F_1) + (-0.18924731)^2 V(F_2) + (-0.65376344)^2 V(F_3),
\]

\[
V(\hat{\beta}_2) = (0.55913978)^2 V(F_0) + (-1.24731183)^2 V(F_1) + (-0.17204301)^2 V(F_2) + (0.86021505)^2 V(F_3).
\]

The results are summarized in Table 3.

This full analysis of the available data provides the general conclusions already stated from the analysis of the first three generations.
Table 3

Estimates of Parameters and Their Standard Errors

<table>
<thead>
<tr>
<th>Character</th>
<th>$\hat{x}_1$ (s.e. $\hat{x}_1$)</th>
<th>$\hat{\beta}_1$ (s.e. $\hat{\beta}_1$)</th>
<th>$\hat{\beta}_2$ (s.e. $\hat{\beta}_2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length of fruit</td>
<td>49.571 $\pm$ .0516</td>
<td>-.2804 $\pm$ .0798</td>
<td>1.9269 $\pm$ .1189</td>
</tr>
<tr>
<td>Width of fruit</td>
<td>24.826 $\pm$ .0176</td>
<td>1.5211 $\pm$ .0278</td>
<td>-1.2172 $\pm$ .0412</td>
</tr>
<tr>
<td>Shape (arithmetic)</td>
<td>2.146 $\pm$ .0336</td>
<td>-.2816 $\pm$ .0519</td>
<td>.1695 $\pm$ .0773</td>
</tr>
<tr>
<td>Shape (logarithmic)</td>
<td>0.288 $\pm$ .0062</td>
<td>-.0268 $\pm$ .0096</td>
<td>.0463 $\pm$ .0143</td>
</tr>
<tr>
<td>Green Fruit Weight (arithmetic)</td>
<td>61.403 $\pm$ .3270</td>
<td>5.3804 $\pm$ 1.2974</td>
<td>-7.1269 $\pm$ 1.9256</td>
</tr>
<tr>
<td>Green Fruit Weight (logarithmic)</td>
<td>1.754 $\pm$ .0058</td>
<td>.0731 $\pm$ .0093</td>
<td>-.0548 $\pm$ .0137</td>
</tr>
<tr>
<td>Dry Fruit Weight (arithmetic)</td>
<td>9.600 $\pm$ .0922</td>
<td>.8252 $\pm$ .1486</td>
<td>-.3226 $\pm$ .2187</td>
</tr>
<tr>
<td>Dry Fruit Weight (logarithmic)</td>
<td>0.962 $\pm$ .0042</td>
<td>.0530 $\pm$ .0070</td>
<td>-.0149 $\pm$ .0102</td>
</tr>
</tbody>
</table>
V. TWO PARENTS

We shall consider as parents only inbred lines which are homozygous, and the phase of any given locus in the parents will be denoted by either R or D, depending on whether the number of dominant alleles at that locus is zero or two respectively. The treatment used for the present situation utilizes the fact that the loci fall into groups of loci within which the loci have identical phase relationships to the parents. There are, in fact, four loci groups as follows:

<table>
<thead>
<tr>
<th>Loci Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent</td>
</tr>
<tr>
<td>P₁</td>
</tr>
<tr>
<td>P₂</td>
</tr>
</tbody>
</table>

For loci groups one and four the loci are in the same homozygous phase for both parents. Hence, on crossing, these loci will always be fixed. The loci which vary in crossing are from loci groups two and three. It should be noted that the number of loci in any loci group may be extremely low or even zero for a given set of parents.

The populations which will be discussed in the two parent case are obtained by crossing or selfing after certain crosses are made from the two parents. In the crossing of two populations random mating is assumed. When we cross two parents the fixed
loci groups do not change, but loci groups two and three will contain all heterozygous loci, that is, have phase denoted by $H$.

Hence the $F_1$, which is $(F_1)(F_2)$, where the brackets around $F_1$ and $F_2$ denote crossing, has the following phase structure:

<table>
<thead>
<tr>
<th>Loci Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$F_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R$</td>
</tr>
<tr>
<td>$H$</td>
</tr>
<tr>
<td>$H$</td>
</tr>
<tr>
<td>$D$</td>
</tr>
</tbody>
</table>

Loci groups one and four will be fixed in any succeeding matings. Also, loci group three is a reflection of loci group two, in the sense that phase $D$ is a reflection of phase $R$, while phase $H$ is a reflection of itself. In this sense, group four is a reflection of group one, but this consideration is unnecessary because group four is fixed. Hence we may specify completely any genotype by displaying only loci group two for that genotype.

If we cross $F_1$ by itself we obtain the $F_2$ which is specified by the phase frequencies of any single locus of group two, which are as follows:

<table>
<thead>
<tr>
<th>Loci Group Two</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_2$</td>
</tr>
<tr>
<td>$\frac{1}{4} R$</td>
</tr>
<tr>
<td>$\frac{1}{2} H$</td>
</tr>
<tr>
<td>$\frac{1}{4} D$</td>
</tr>
</tbody>
</table>

where the coefficients represent the proportion of cases in which loci of the group have the specified genetic makeup, for example, $\frac{1}{2} H$ indicates that one half of the loci of loci group two are in the $H$ phase for the $F_2$ population.
The following list includes the representation of the parents, crosses made by random mating and selfed populations obtained by crossing and then selfing. (The letter $S$ will be used to denote selfed populations, for example, $(F_2)S = F_3$):

<table>
<thead>
<tr>
<th>Population</th>
<th>Loci Group Two</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_1$</td>
<td>D</td>
</tr>
<tr>
<td>$F_2$</td>
<td>R</td>
</tr>
<tr>
<td>$(F_1)(F_2) = F_1$</td>
<td>$\frac{1}{4}R, \frac{1}{2}H, \frac{1}{4}D$</td>
</tr>
<tr>
<td>$(F_1)(F_1) = F_2$</td>
<td>$\frac{1}{2}R, \frac{1}{4}D$</td>
</tr>
<tr>
<td>$(F_2)(F_1)$</td>
<td>$\frac{1}{2}R, \frac{1}{2}H$</td>
</tr>
<tr>
<td>$(F_1)[(F_1)(F_1)]$</td>
<td>$\frac{1}{4}R, \frac{3}{4}H$</td>
</tr>
<tr>
<td>$(F_2)[(F_1)(F_1)]$</td>
<td>$\frac{3}{4}R, \frac{1}{4}H$</td>
</tr>
<tr>
<td>$(F_1)[(F_2)(F_1)]$</td>
<td>$\frac{1}{8}R, \frac{1}{2}H, \frac{3}{8}D$</td>
</tr>
<tr>
<td>$(F_1)[(F_2)(F_1)]$</td>
<td>$\frac{3}{8}R, \frac{1}{2}H, \frac{1}{8}D$</td>
</tr>
<tr>
<td>$(F_1)S = F_2$</td>
<td>$\frac{1}{4}R, \frac{1}{2}H, \frac{1}{4}D$</td>
</tr>
<tr>
<td>$(F_2)S = F_3$</td>
<td>$\frac{3}{8}R, \frac{1}{4}H, \frac{3}{8}D$</td>
</tr>
<tr>
<td>$(F_3)S = F_4$</td>
<td>$\frac{7}{16}R, \frac{1}{8}H, \frac{7}{16}D$</td>
</tr>
<tr>
<td>$[(F_1)(F_1)]S$</td>
<td>$\frac{1}{8}R, \frac{1}{4}H, \frac{5}{8}D$</td>
</tr>
<tr>
<td>$[(F_2)(F_1)]S$</td>
<td>$\frac{5}{8}R, \frac{1}{4}H, \frac{1}{8}D$</td>
</tr>
</tbody>
</table>
We shall now obtain the mean genotypic value of a population represented by

\[ pR + qH + rD. \]

Apart from fixed loci the genotypic array is

\[ \sum_{\text{loci of group two}} (pR + qH + rD) \] \[ \sum_{\text{loci of group three}} (pR + qH + pD). \]

Consider in order the loci effects, two factor interactions and so on which occur when this array is expanded and the factorial model is inserted in place of the genotype.

The total effect contribution is

\[ \sum_{\text{fixed loci}} \sum_{\text{loci of loci}} \sum_{\text{loci of group two}} \sum_{\text{loci of group three}} \]

The two factor interactions are of three types: (a) interactions between variable loci, (b) interactions of a variable locus with a fixed locus and (c) interactions of two fixed loci. As regards (c), we shall get a certain fixed quantity. As regards (b), suppose \( a_0 \) refers to a fixed locus then we get

\[ \sum_{\text{loci of group two}} (p a_0 A_0 + q a_0 A_1 + r a_0 A_2) \]

with similar expressions for all loci of group two and group three.

As regards (a) we have three types (i) interaction of two loci in group two, (ii) interaction of two loci in group three, (iii) interaction of one locus in group two with one locus in group three.

When we combine these terms together with the loci effects
and ignore interactions of more than two loci we obtain the
expression

\[ X_2 + p \left( \frac{\sum A_0}{2} + \frac{\sum B_2}{3} \right) + q \left( \frac{\sum A_1}{2} + \frac{\sum B_1}{3} \right) + r \left[ \frac{\sum A_2}{2} + \frac{\sum B_0}{3} \right] 
+ p^2 \left( \frac{\sum A_1 A_1}{2} + \frac{\sum B_2 B_1}{3} + \frac{\sum A_1 B_2}{2,3} \right) 
+ q^2 \left( \frac{\sum A_2 A_1}{2} + \frac{\sum B_1 B_1}{3} + \frac{\sum A_2 B_1}{2,3} \right) 
+ r^2 \left( \frac{\sum A_0 A_1}{2} + \frac{\sum B_0 B_1}{3} + \frac{\sum A_0 B_0}{2,3} \right) 
+ pq \left( \frac{2 \sum A_0 A_1}{2} + \frac{2 \sum B_2 B_1}{3} + \frac{\sum A_0 B_1}{2,3} + \frac{\sum A_2 B_1}{2,3} \right) 
+ pr \left( \frac{2 \sum A_0 A_2}{2} + \frac{2 \sum B_0 B_2}{3} + \frac{\sum A_0 B_2}{2,3} + \frac{\sum A_2 B_2}{2,3} \right) 
+ qr \left( \frac{2 \sum A_1 A_2}{2} + \frac{2 \sum B_0 B_1}{3} + \frac{\sum A_1 B_0}{2,3} + \frac{\sum A_2 B_1}{2,3} \right) \]

when for example,

\[ \sum A_0 \quad = \quad \text{the sum of the loci effects for loci in the } R \text{ phase} \]

in group two

\[ \sum A_0 A_1 \quad = \quad \text{the sum of interactions contributions from two loci} \]

in group two, both of which are in the R phase

\[ \sum A_0 A_2 \quad = \quad \text{the sum of interaction contributions from two loci} \]

in group two, one of which is in the R phase and

the other of which is in the D phase

\[ \sum A_0 B_1 \quad = \quad \text{the sum of interaction contributions arising from} \]

two loci, one of which is in group two and is in the
R phase, and the other of which is in group three
and is in the H phase,
and where all effects and interactions are defined with respect to
the variable loci only.

There are ten parameters in this expression as follows:

(a) $X_2$

(b) three loci effects parameters, $(\frac{\sum A_0}{2} + \frac{\sum B_2}{3})$,
\[ (\frac{\sum A_1}{2} + \frac{\sum B_1}{3}) \text{ and } \left[ \frac{\sum A_2}{2} + \frac{\sum B_0}{3} \right] \]

and (c) six epistatic parameters, which are the sum of interactions with common coefficients $p^2$, $q^2$, $r^2$, $pq$, $pr$, and $qr$.

However we can reduce the number of parameters to six when all the restrictions of the type

$$ (A_0 + 2A_1 + A_2) = (B_0 + 2B_1 + B_2) = 0, $$

are used.

We will substitute $(-2A_1 - A_2)$ for $A_0$ and $(-2B_1 - B_2)$ for $B_2$

wherever they occur. The resulting expression is

$$ X_2 + (q-2p) \left( \frac{\sum A_1}{2} + \frac{\sum B_1}{3} \right) + (r-p) \left( \frac{\sum A_2}{2} + \frac{\sum B_0}{3} \right) + $$

$$ (q-2p)^2 \left( \frac{\sum A_1 A_1'}{2} + \frac{\sum B_1 B_1'}{3} + \frac{\sum A_1 B_1}{2} \right) + $$

$$ (r-p)^2 \left( \frac{\sum A_2 A_2'}{2} + \frac{\sum B_0 B_0'}{3} + \frac{\sum A_2 B_0}{2} \right) + $$

$$ (q-2p)(r-p) \left( 2 \frac{\sum A_1 A_2}{2} + 2 \frac{\sum B_1 B_1'}{3} + \frac{\sum A_1 B_0}{2} + \frac{\sum A_2 B_1}{2} \right). $$
Hence we have six parameters as follows:

(a) \( k_2 \)

(b) two loci effects parameters

1. \( \left( \frac{\sum A_1}{2} + \frac{\sum B_1}{3} \right) \)

2. \( \left( \frac{\sum A_2}{2} + \frac{\sum B_0}{3} \right) \)

and (c) three epistatic parameters

1. \( \left( \frac{\sum A_1 A_1}{2} + \frac{\sum B_1 B_1}{3} + \frac{\sum A_1 B_1}{2, 3} \right) \)

2. \( \left( \frac{\sum A_2 A_2}{2} + \frac{\sum B_0 B_0}{3} + \frac{\sum A_2 B_0}{2, 3} \right) \)

3. \( \left( 2 \frac{\sum A_1 A_2}{2} + 2 \frac{\sum B_1 B_0}{3} + \frac{\sum A_1 B_0}{2, 3} + \frac{\sum A_2 B_1}{2, 3} \right) \).

To simplify the notation we will let

\[ B = \left( \frac{\sum A_1}{2} + \frac{\sum B_1}{3} \right), \]

\[ F = \left( \frac{\sum A_2}{2} + \frac{\sum B_0}{3} \right), \]

\[ G = \left( \frac{\sum A_1 A_1}{2} + \frac{\sum B_1 B_1}{3} + \frac{\sum A_1 B_1}{2, 3} \right), \]

\[ L = \left( 2 \frac{\sum A_1 A_2}{2} + 2 \frac{\sum B_1 B_0}{3} + \frac{\sum A_1 B_0}{2, 3} + \frac{\sum A_2 B_1}{2, 3} \right), \]

and \( M = \left( \frac{\sum A_2 A_2}{2} + \frac{\sum B_0 B_0}{3} + \frac{\sum A_2 B_0}{2, 3} \right). \)

Due to the construction of the model

\[ \left( \frac{\sum A_0}{2} + \frac{\sum B_2}{3} \right) = -(2E + F), \]
This specifies all loci effects and possible interactions of two loci. Hence we may describe the six parameters as,

1. $K_2$ which is the contribution due to the overall mean plus the loci effects and interactions of the fixed loci
2. $E$ and $F$ which are the contributions due to the variable loci effects plus the interaction of these loci effects with the fixed loci, and
3. $G$, $L$, and $M$ which are the contributions due to the interaction of all possible pairs of variable loci plus the interaction of these interactions with the fixed loci.

It should be emphasized that due to the construction of the model the coefficient of $G$ is the square of the coefficient of $E$, the coefficient of $L$ is equal to the product of the coefficient of $E$ and of the coefficient of $F$, and the coefficient of $M$ is the square of the coefficient of $F$. Hence after we obtain the coefficients of $E$ and $F$, we can write down the coefficients of $G$, $L$ and $M$ immediately.

Using these definitions of the loci effects and interactions and the frequencies of the various phases of the loci in group two we can write the factorial representation of the means of the genotypic values of the populations listed previously as follows:
In many cases experimenters will grow the parents, $F_1$, $F_2$ and backcrosses but no other crosses, due to lack of space and time.

In these cases we can estimate the parameters $K_2, E, F, G, L$ and $M$ by
solving the following matrix equation obtained from the first six equations above:

\[
\begin{bmatrix}
1 & 0 & 1 & 0 & 0 & 1 \\
1 & -2 & -1 & 4 & 2 & 1 \\
1 & 1 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\
1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4}
\end{bmatrix}
\begin{bmatrix}
X_2 \\
E \\
F \\
G \\
L \\
M
\end{bmatrix}
=
\begin{bmatrix}
F_1 \\
F_2 \\
F_1 \\
F_2 \\
(F_1)(F_1) \\
(F_2)(F_1)
\end{bmatrix}
\]

Then it follows that the estimates of the six parameters expressed in matrix form are:

\[
\begin{bmatrix}
X_2 \\
E \\
F \\
G \\
L \\
M
\end{bmatrix}
=
\begin{bmatrix}
0 & 0 & 0 & 1 & 0 & 0 \\
\frac{1}{4} & \frac{1}{4} & \frac{1}{2} & -2 & 1 & 1 \\
\frac{1}{4} & \frac{1}{4} & \frac{1}{2} & 2 & 0 & -2 \\
\frac{1}{4} & \frac{1}{4} & \frac{1}{2} & 1 & -1 & -1 \\
-1 & 0 & -1 & -2 & 3 & 1 \\
\frac{3}{4} & \frac{1}{4} & \frac{1}{2} & -3 & 0 & 2
\end{bmatrix}
\begin{bmatrix}
F_1 \\
F_2 \\
F_1 \\
F_2 \\
(F_1)(F_1) \\
(F_2)(F_1)
\end{bmatrix}
\]
The variances of these estimates are

\[
\begin{pmatrix}
\sigma_{e_1}^2 \\
\sigma_{e_2}^2 \\
\sigma_{e_3}^2 \\
\sigma_{e_4}^2 \\
\sigma_{e_5}^2
\end{pmatrix}
= 
\begin{pmatrix}
0 & 0 & 0 & 1 & 0 \\
\frac{1}{16} & \frac{1}{16} & \frac{1}{4} & 4 & 1 \\
\frac{1}{16} & \frac{1}{16} & \frac{1}{4} & 4 & 0 \\
\frac{1}{16} & \frac{1}{16} & \frac{1}{4} & 1 & 1 \\
\frac{1}{16} & \frac{1}{16} & \frac{1}{4} & 9 & 0
\end{pmatrix}
\begin{pmatrix}
\sigma_{e_1}^2 \\
\sigma_{e_2}^2 \\
\sigma_{e_3}^2 \\
\sigma_{e_4}^2 \\
\sigma_{e_5}^2
\end{pmatrix}
\]

if the yields of the \(P_1\)'s, \(P_2\)'s, \(F_1\)'s, \(F_2\)'s, \((P_1)(F_1)\)'s and \((P_2)(F_1)\)'s are independent.

G. H. Stringfield (17) has furnished the means on three characters in corn, number of days to mid-silk, ear node height in inches and yield of corn per acre in bushels. From these data we have taken all possible pairs of two parents and obtained Tables 4, 5 and 6. The variance of the parents and crosses were not obtainable so that standard errors of the estimates could not be obtained.
Table 4
The Means of the Parents and Crosses and the Estimate of
the Six Parameters for the Character, Days to Mid silk

<table>
<thead>
<tr>
<th></th>
<th>28x51A</th>
<th>Ryz28</th>
<th>28x40B</th>
<th>Ryz51A</th>
<th>40Bx51A</th>
<th>Ryz40B</th>
</tr>
</thead>
<tbody>
<tr>
<td>P_1</td>
<td>86.3</td>
<td>90.0</td>
<td>86.3</td>
<td>90.0</td>
<td>81.0</td>
<td>90.0</td>
</tr>
<tr>
<td>P_2</td>
<td>79.3</td>
<td>86.3</td>
<td>81.0</td>
<td>79.3</td>
<td>79.3</td>
<td>81.0</td>
</tr>
<tr>
<td>F_1</td>
<td>77.5</td>
<td>81.3</td>
<td>80.0</td>
<td>76.3</td>
<td>74.8</td>
<td>80.0</td>
</tr>
<tr>
<td>P_1F_1</td>
<td>82.5</td>
<td>82.5</td>
<td>81.0</td>
<td>82.0</td>
<td>78.5</td>
<td>83.5</td>
</tr>
<tr>
<td>P_2F_1</td>
<td>78.0</td>
<td>83.0</td>
<td>79.0</td>
<td>76.5</td>
<td>76.5</td>
<td>82.0</td>
</tr>
<tr>
<td>F_2</td>
<td>80.0</td>
<td>84.0</td>
<td>80.8</td>
<td>80.0</td>
<td>77.8</td>
<td>82.0</td>
</tr>
<tr>
<td>E</td>
<td>-2.15</td>
<td>-5.92</td>
<td>-3.42</td>
<td>-5.68</td>
<td>-3.28</td>
<td>-1.25</td>
</tr>
<tr>
<td>F</td>
<td>6.65</td>
<td>5.42</td>
<td>5.42</td>
<td>11.18</td>
<td>5.28</td>
<td>2.75</td>
</tr>
<tr>
<td>G</td>
<td>-0.35</td>
<td>3.22</td>
<td>2.62</td>
<td>1.98</td>
<td>0.28</td>
<td>-0.75</td>
</tr>
<tr>
<td>L</td>
<td>1.70</td>
<td>-8.80</td>
<td>-5.90</td>
<td>-3.80</td>
<td>0.60</td>
<td>-4.50</td>
</tr>
<tr>
<td>M</td>
<td>-0.35</td>
<td>0.58</td>
<td>0.08</td>
<td>-1.18</td>
<td>-2.08</td>
<td>5.25</td>
</tr>
</tbody>
</table>
Table 5
The Means of the Parents and Crosses and the Estimates of the Six Parameters for the Character, Ear Node Height

<table>
<thead>
<tr>
<th></th>
<th>51Ax28</th>
<th>Hyx28</th>
<th>40Bx28</th>
<th>Hyx51A</th>
<th>51Ax40b</th>
<th>Hyx40B</th>
</tr>
</thead>
<tbody>
<tr>
<td>P₁</td>
<td>24.9</td>
<td>31.0</td>
<td>24.5</td>
<td>31.0</td>
<td>24.9</td>
<td>31.0</td>
</tr>
<tr>
<td>P₂</td>
<td>21.5</td>
<td>21.5</td>
<td>21.5</td>
<td>24.9</td>
<td>24.5</td>
<td>24.5</td>
</tr>
<tr>
<td>F₁</td>
<td>37.5</td>
<td>43.0</td>
<td>34.2</td>
<td>46.5</td>
<td>32.4</td>
<td>43.9</td>
</tr>
<tr>
<td>P₁F₁</td>
<td>31.2</td>
<td>39.9</td>
<td>30.1</td>
<td>42.4</td>
<td>31.2</td>
<td>40.3</td>
</tr>
<tr>
<td>P₂F₁</td>
<td>31.5</td>
<td>33.1</td>
<td>26.5</td>
<td>37.8</td>
<td>28.8</td>
<td>34.1</td>
</tr>
<tr>
<td>F₂</td>
<td>30.1</td>
<td>36.6</td>
<td>27.2</td>
<td>37.1</td>
<td>28.6</td>
<td>37.4</td>
</tr>
<tr>
<td>K 2</td>
<td>30.10</td>
<td>36.60</td>
<td>27.20</td>
<td>37.10</td>
<td>28.60</td>
<td>37.40</td>
</tr>
<tr>
<td>E</td>
<td>9.65</td>
<td>8.18</td>
<td>7.80</td>
<td>15.28</td>
<td>6.65</td>
<td>7.68</td>
</tr>
<tr>
<td>F</td>
<td>-9.65</td>
<td>-1.38</td>
<td>-4.20</td>
<td>-10.68</td>
<td>-4.25</td>
<td>-1.48</td>
</tr>
<tr>
<td>G</td>
<td>-2.25</td>
<td>-1.78</td>
<td>-0.80</td>
<td>-5.88</td>
<td>-2.85</td>
<td>-1.18</td>
</tr>
<tr>
<td>L</td>
<td>2.50</td>
<td>5.60</td>
<td>3.70</td>
<td>13.30</td>
<td>7.90</td>
<td>5.30</td>
</tr>
<tr>
<td>H</td>
<td>4.45</td>
<td>4.22</td>
<td>1.50</td>
<td>4.58</td>
<td>0.55</td>
<td>4.92</td>
</tr>
</tbody>
</table>
Table 6
The Means of the Parents and Crosses and the Estimates of the Six Parameters for the Character, Yield per Acre

<table>
<thead>
<tr>
<th></th>
<th>51Ax2B</th>
<th>28xHy</th>
<th>40Bx2B</th>
<th>51AxHy</th>
<th>51Ax40B</th>
<th>40BxHy</th>
</tr>
</thead>
<tbody>
<tr>
<td>P&lt;sub&gt;1&lt;/sub&gt;</td>
<td>53.3</td>
<td>39.5</td>
<td>46.0</td>
<td>53.3</td>
<td>53.3</td>
<td>46.0</td>
</tr>
<tr>
<td>P&lt;sub&gt;2&lt;/sub&gt;</td>
<td>39.5</td>
<td>27.2</td>
<td>39.5</td>
<td>27.2</td>
<td>46.0</td>
<td>27.2</td>
</tr>
<tr>
<td>F&lt;sub&gt;1&lt;/sub&gt;</td>
<td>86.0</td>
<td>85.2</td>
<td>86.8</td>
<td>88.1</td>
<td>84.6</td>
<td>78.6</td>
</tr>
<tr>
<td>P&lt;sub&gt;1&lt;/sub&gt;F&lt;sub&gt;1&lt;/sub&gt;</td>
<td>74.3</td>
<td>70.7</td>
<td>65.9</td>
<td>69.4</td>
<td>74.6</td>
<td>69.9</td>
</tr>
<tr>
<td>P&lt;sub&gt;2&lt;/sub&gt;F&lt;sub&gt;1&lt;/sub&gt;</td>
<td>66.5</td>
<td>66.8</td>
<td>71.6</td>
<td>60.5</td>
<td>65.9</td>
<td>58.7</td>
</tr>
<tr>
<td>F&lt;sub&gt;2&lt;/sub&gt;</td>
<td>62.9</td>
<td>64.8</td>
<td>61.7</td>
<td>60.6</td>
<td>61.8</td>
<td>61.5</td>
</tr>
<tr>
<td>E&lt;sub&gt;2&lt;/sub&gt;</td>
<td>62.90</td>
<td>64.80</td>
<td>61.70</td>
<td>60.60</td>
<td>61.80</td>
<td>61.50</td>
</tr>
<tr>
<td>E</td>
<td>34.78</td>
<td>33.82</td>
<td>36.12</td>
<td>32.60</td>
<td>34.35</td>
<td>26.60</td>
</tr>
<tr>
<td>F</td>
<td>-26.98</td>
<td>-29.92</td>
<td>-41.82</td>
<td>-23.70</td>
<td>-25.65</td>
<td>-15.40</td>
</tr>
<tr>
<td>L</td>
<td>24.30</td>
<td>24.60</td>
<td>13.10</td>
<td>6.10</td>
<td>28.20</td>
<td>20.80</td>
</tr>
<tr>
<td>M</td>
<td>17.38</td>
<td>4.62</td>
<td>26.12</td>
<td>16.40</td>
<td>17.15</td>
<td>11.90</td>
</tr>
</tbody>
</table>
It is unfortunate that we do not have standard errors by which to assess the magnitudes of the estimates of the parameters. However we can make a comparison of magnitudes between loci effects and epistatic effects, and in the case of all characteristics some of the epistatic estimates are of the same magnitude as or larger than the loci effects. Hence we conclude that the epistatic contributions are important in the particular material under investigation.

If the experimenter is willing to make the assumption that epistasis is negligible and he has only the parents and F₁ populations, he can estimate \( k₂ \) and the loci parameters in the reduced factorial model from the equations:

\[
P₁ = k₂ + P
\]
\[
P₂ = k₂ - 2E - F
\]
and \( P₁ = k₂ + E \).

Hence the estimates of the parameters are

\[
\hat{k}₂ = \frac{1}{4} (P₁ + P₂ + 2F₁)
\]
\[
\hat{E} = \frac{1}{4} (P₁ - P₂ + 2F₁)
\]
and \( \hat{F} = \frac{1}{4} (3P₁ - P₂ + 2F₁) \).

The variances of these estimates are

\[
\sigma_{k₂}^2 = \frac{1}{16} (\sigma_{P₁}^2 + \sigma_{P₂}^2 + 4 \sigma_{F₁}^2)
\]
\[
\sigma_{E}^2 = \frac{1}{16} (\sigma_{P₁}^2 + \sigma_{P₂}^2 + 4 \sigma_{F₁}^2)
\]
and $c_F^2 = \frac{1}{16}(9c_{P_1}^2 + c_{P_2}^2 + 4c_{F_1}^2)$

if the yields of the $P_1$'s, $P_2$'s and $F_1$'s are independent. The quantity $E$ clearly measures the potence, as described by Mather (13). These expressions lead to the following equations for predicting the two backcross and $F_2$ means:

$\hat{F}_2 = \frac{1}{4}(P_1 + P_2 + 2F_1)$,

$(P_1)(F_1) = \frac{1}{2}(P_1 + F_1)$

and $(P_2)(F_1) = \frac{1}{2}(P_2 + F_1)$,

which are the standard expressions given, for example, by Mather (13, p. 42). We may note here the flexibility of the model which is such that the deletion of parameters corresponding to genetic assumptions results in the simple known expression.

It is of interest to investigate the scaling tests given by Mather (13). The test, using $A$, $B$ and $C$, given by Mather tests the additiveness of the genic effects regardless of the presence of dominance or linkages. This seems to imply, and it is certainly accepted by many, that this is a test for absence of epistasis. If the interaction of three or more loci is known to be zero then this is a test of the two-factor interactions being zero. However, if the interactions of three loci only are not zero, this is not a test for the absence of epistasis. The reason for this may be summarized by expressing $A$, $B$ and $C$ in terms of the factorial model as follows:
where $\alpha$, $\beta$, $\gamma$ and $\delta$ are the three factor interaction parameters. Hence if $A = B = C = 0$, it does not necessarily follow that the parameters are zero.

We have not discussed in this study the variances of populations although it may be expected that such an examination would yield further information on the general problem of the importance of epistasis.
VI. THREE PARENTS

The number of distinct loci groups for this case is eight and we may represent the three parents as follows:

<table>
<thead>
<tr>
<th>Parent</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>P₁</td>
<td>R</td>
<td>D</td>
<td>R</td>
<td>D</td>
<td>R</td>
<td>D</td>
<td>R</td>
<td>D</td>
</tr>
<tr>
<td>P₂</td>
<td>R</td>
<td>R</td>
<td>D</td>
<td>D</td>
<td>R</td>
<td>R</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>P₃</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
</tbody>
</table>

Using the same technique we did for the previous section, we may reduce the number of loci groups to three and still specify the individual genotypes completely. In this case, loci group one and eight constitute the fixed loci and group two is the reflection of group seven, group three is the reflection of group six and group four is the reflection of group five. Hence we need only write loci groups two, three and four to specify the following populations:
<table>
<thead>
<tr>
<th>Population</th>
<th>Loci Group 2</th>
<th>Loci Group 3</th>
<th>Loci Group 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>(P₁)</td>
<td>D</td>
<td>R</td>
<td>D</td>
</tr>
<tr>
<td>(P₂)</td>
<td>R</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>(P₃)</td>
<td>R</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td>(P₁) (P₂) = (P₁₁₂)</td>
<td>H</td>
<td>H</td>
<td>D</td>
</tr>
<tr>
<td>(P₁) (P₃) = (P₁₁₃)</td>
<td>H</td>
<td>R</td>
<td>H</td>
</tr>
<tr>
<td>(P₂) (P₃) = (P₁²₃)</td>
<td>R</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>(P₁₁₂) (P₁₁₂)</td>
<td>γ</td>
<td>γ</td>
<td>D</td>
</tr>
<tr>
<td>(P₁₁₃) (P₁₁₃)</td>
<td>γ</td>
<td>R</td>
<td>γ</td>
</tr>
<tr>
<td>(P₁²₃) (P₁²₃)</td>
<td>R</td>
<td>γ</td>
<td>γ</td>
</tr>
<tr>
<td>(P₁₁₂) (P₁₁₃)</td>
<td>γ</td>
<td>ρ</td>
<td>δ</td>
</tr>
<tr>
<td>(P₁₁₂) (P₁²₃)</td>
<td>ρ</td>
<td>γ</td>
<td>δ</td>
</tr>
<tr>
<td>(P₁₁₃) (P₁²₃)</td>
<td>ρ</td>
<td>ρ</td>
<td>γ</td>
</tr>
<tr>
<td>(P₁)(P₁₁₂)</td>
<td>ϵ</td>
<td>ρ</td>
<td>D</td>
</tr>
<tr>
<td>(P₁)(P₁₁₃)</td>
<td>ϵ</td>
<td>R</td>
<td>δ</td>
</tr>
<tr>
<td>(P₂)(P₁₁₂)</td>
<td>ρ</td>
<td>δ</td>
<td>D</td>
</tr>
<tr>
<td>(P₂)(P₁²₃)</td>
<td>R</td>
<td>δ</td>
<td>δ</td>
</tr>
<tr>
<td>(P₃)(P₁₁₃)</td>
<td>ρ</td>
<td>R</td>
<td>ρ</td>
</tr>
<tr>
<td>(P₃)(P₁²₃)</td>
<td>R</td>
<td>ρ</td>
<td>ρ</td>
</tr>
</tbody>
</table>
where:

\[
\rho = \begin{pmatrix}
\frac{1}{2} & R \\
\frac{1}{2} & H
\end{pmatrix},
\]

\[
\delta = \begin{pmatrix}
\frac{1}{2} & H \\
\frac{1}{2} & D
\end{pmatrix},
\]

\[
\gamma = \begin{pmatrix}
\frac{1}{2} & R \\
\frac{1}{2} & H
\end{pmatrix}.
\]

and the coefficients of the phases have the same meaning as they did for the two parent case.

Next we shall obtain the genotypic value of a population whose genotypic array for the variable loci is

\[\begin{array}{cccc}
\text{loci of} & \text{loci of} & \text{loci of} & \text{loci of} \\
\text{group 2} & \text{group 3} & \text{group 4} & \\
(pR+qH+rD) & (sR+tH+uD) & (vR+wH+yD) & \\
\end{array}\]

\[\times\begin{array}{cccc}
\text{loci of} & \text{loci of} & \text{loci of} & \\
\text{group 5} & \text{group 6} & \text{group 7} & \\
(pR+qH+pD) & (sR+tH+sR) & (vR+wH+yD) & \\
\end{array}\]

\[\text{Population} \quad \begin{array}{ccc}
\text{Loci Group} & 2 & 3 & 4 \\
(p_1) (p_1^{23}) & H & \rho & \delta \\
(p_2) (p_1^{13}) & \rho & H & \delta \\
(p_3) (p_1^{12}) & \rho & \rho & H \\
\end{array}\]
If we place restrictions of the type $A_0 + 2A_1 + A_2 = 0$ on all effects, the total effect contribution is

$$\sum_{\text{fixed effects}} + \sum_{2} \left[ (q-p)A_1 + (r-p)A_2 \right] + \sum_{3} \left[ (t-2s)C_1 + (u-s)C_2 \right]$$

$$+ \sum_{4} \left[ (w-2v)E_1 + (y-2v)E_2 \right] + \sum_{5} \left[ (x-v)F_0 + (v-2r)F_1 \right]$$

$$+ \sum_{6} \left[ (u-s)D_0 + (t-2s)D_1 \right] + \sum_{7} \left[ (v-p)B_0 + (q-2p)B_1 \right].$$

In addition to the three types of two factor interactions for all reflection pairs of loci groups given in the two parent case, there are the two factor interactions between loci of the non-reflecting pairs. The coefficients of the two factor interaction terms involving say loci $x$ and $y$ are obtained by expanding the product of the factor for locus $x$ and the factor for locus $y$ in the genotypic array given above. The same result may be obtained by expanding formally the product of the contribution from locus $x$ and the contribution from locus $y$ in the total effect contribution. The possible coefficients are easily written down, there being six for the loci effects and twenty-one for the two-factor interactions. Because of the particular populations we are considering here, the number of distinct coefficients of variable effects reduces to fifteen in all, six for the loci effects, and nine for the interactions. There are sixteen parameters in all, including the contribution for the fixed loci, $E_3$. 
After considerable collection of terms it is found that the expression for the mean genotypic value is

\[ k_0 + (q-2p) \left( \sum_{j=1}^{3} A_j + \sum_{j=1}^{2} B_j \right) + (r-p) \left( \sum_{j=1}^{3} A_j + \sum_{j=1}^{2} B_j \right) + (t-2a) \left( \sum_{j=1}^{3} C_j + \sum_{j=1}^{6} D_j \right) \]

\[ + (u-s) \left( \sum_{j=1}^{3} E_j + \sum_{j=1}^{6} F_j \right) + (w-2v) \left( \sum_{j=1}^{3} G_j + \sum_{j=1}^{6} H_j \right) + (y-v) \left( \sum_{j=1}^{3} I_j + \sum_{j=1}^{6} J_j \right) \]

\[ + (q-2p)^2 \left( \sum_{j=1}^{3} A_j A_j + \sum_{j=1}^{6} B_j B_j + \sum_{j=1}^{3} C_j C_j + \sum_{j=1}^{6} D_j D_j \right) \]

\[ + (r-p)^2 \left( \sum_{j=1}^{3} A_j A_j + \sum_{j=1}^{6} B_j B_j + \sum_{j=1}^{3} C_j C_j + \sum_{j=1}^{6} D_j D_j \right) \]

\[ + (t-2a)^2 \left( \sum_{j=1}^{3} C_j C_j + \sum_{j=1}^{6} D_j D_j + \sum_{j=1}^{3} E_j E_j + \sum_{j=1}^{6} F_j F_j \right) \]

\[ + (u-s)^2 \left( \sum_{j=1}^{3} C_j C_j + \sum_{j=1}^{6} D_j D_j + \sum_{j=1}^{3} E_j E_j + \sum_{j=1}^{6} F_j F_j \right) \]

\[ + (w-2v)^2 \left( \sum_{j=1}^{3} G_j G_j + \sum_{j=1}^{6} H_j H_j + \sum_{j=1}^{3} I_j I_j + \sum_{j=1}^{6} J_j J_j \right) \]

\[ + (y-v)^2 \left( \sum_{j=1}^{3} G_j G_j + \sum_{j=1}^{6} H_j H_j + \sum_{j=1}^{3} I_j I_j + \sum_{j=1}^{6} J_j J_j \right) \]

\[ + (q-2p)(r-p) \left( \sum_{j=1}^{3} A_j A_j + \sum_{j=1}^{6} B_j B_j + \sum_{j=1}^{3} C_j C_j + \sum_{j=1}^{6} D_j D_j \right) + \text{interactions of pairs of loci between non-reflecting groups} \]

\[ + (t-2a)(u-s) \left( \sum_{j=1}^{3} C_j C_j + \sum_{j=1}^{6} D_j D_j + \sum_{j=1}^{3} E_j E_j + \sum_{j=1}^{6} F_j F_j \right) + \text{interactions of pairs of loci between non-reflecting groups} \]
\[(w-2v)(y-v)\left(\frac{2}{5}E_{1}E_{2} + \frac{2}{5}F_{1}F_{0} + \frac{4}{5}E_{1}F_{0} + \frac{4}{5}E_{2}F_{1} + \text{interactions of pairs of loci between non-reflecting groups}\right),\]

where \((q-2p), (r-p), (t-2s), (u-s), (w-2v), (y-v)\)

may be \(-2, -1, -\frac{1}{2}, 0, \frac{1}{2} \text{ or } 1\) only.

To simplify the notation we will let

\[
E' = \frac{\sum A_{1} + \sum B_{1}}{2}, \quad F' = \frac{\sum A_{2} + \sum B_{0}}{2},
\]

\[
G' = \frac{\sum A_{1}A_{1} + \sum E_{1}B_{1} + \sum A_{1}B_{1} + \text{interactions of pairs of loci}}{2, 7}, \quad \text{between non-reflecting groups}
\]

\[
L' = \left(\frac{2}{7}A_{1}A_{1} + \frac{2}{7}B_{1}B_{0} + \sum A_{1}B_{0} + \sum A_{2}B_{1} + \text{interactions of pairs of loci between non-reflecting groups}\right)
\]

\[
M' = \left(\frac{2}{7}A_{2}A_{2} + \sum B_{0}B_{1} + \sum A_{2}B_{0} + \text{interactions of pairs of loci between non-reflecting groups}\right)
\]

\[
E'' = \frac{\sum C_{1} + \sum D_{1}}{3}, \quad F'' = \frac{\sum C_{2} + \sum D_{0}}{3}
\]

\[
G'' = \frac{\sum C_{1}C_{1} + \sum D_{1}D_{1} + \sum C_{1}D_{1} + \text{interactions of pairs of loci}}{3, 6}, \quad \text{between non-reflecting groups}
\]

\[
L'' = \left(\frac{2}{3}C_{1}C_{1} + \frac{2}{3}D_{1}D_{0} + \sum C_{1}D_{0} + \sum C_{2}D_{1} + \text{interactions of pairs of loci between non-reflecting groups}\right)
\]

\[
M'' = \left(\frac{2}{3}C_{2}C_{2} + \sum D_{0}D_{0} + \sum C_{2}D_{0} + \text{interactions of pairs of loci between non-reflecting groups}\right)
\]
The interaction of pairs of loci between non-reflecting groups may be represented as in the two parent case, for example

\[ \sum \frac{A_0 + \sum B_2}{2} = -(2E' + F'). \]

Hence we have specified all loci effects and all possible interactions of two loci. We may describe the sixteen parameters as follows:

- \( K \) is the contribution due to the overall mean plus the loci effects and interactions of the fixed loci.
- \( E', F', E'', F'' \) and \( F''' \) are the contributions due to the variable loci effects plus the interaction of these effects with the fixed loci and
- \( G', L', M', G'', L'', M'' \) and \( M''' \) are the contributions due to the interactions of all possible pairs of variable loci plus the interaction of these interactions with the fixed loci.
Due to the construction of the model, the same relationships among the coefficients of the loci effects parameters and the coefficients of the interaction parameters within each set of reflecting group parameters exists here as in the two parent case. Therefore, knowing the coefficients of

(a) $E'$ and $F'$ we know the coefficients of $G'$, $L'$ and $M'$,
(b) $E''$ and $F''$ we know the coefficients of $G''$, $L''$ and $M''$, and
(c) $E'''$ and $F'''$ we know the coefficients of $G'''$, $L'''$ and $M'''$.

Using these definitions of the loci effects and interactions and the frequencies of the various phases of the loci in groups two, three and four we can write the factorial representation of the means of the genotypic values of the populations listed previously in a matrix equation as follows:
\[
\begin{bmatrix}
0 & 1 & -2 & -1 & 4 & 2 & 1 & 0 & 1 & 0 & 0 & 1 \\
2 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\
2 & 1 & -2 & -1 & 4 & 2 & 1 & -2 & -1 & 4 & 2 & 1 \\
0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & -2 & -1 & 4 & 2 & 1 & 1 & 0 & 1 & 0 & 0 \\
2 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & -2 & -1 & 4 & 2 & 1 & 0 & 0 & 0 & 0 & 0 \\
2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -1/2 & -1/2 & 1/4 & 1/4 & 1/4 & 1/2 & 1/2 & 1/4 & 1/4 & 1/4 \\
4 & 1/4 & 1/4 & 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 & 1/4 & 1/4 & 1/4 \\
4 & 1/4 & 1/4 & -1/2 & -1/2 & 1/4 & 1/4 & 1/4 & 0 & 0 & 0 & 0 & 0 \\
4 & 1/4 & 1/4 & -1/2 & -1/2 & 1/4 & 1/4 & 1/4 & 0 & 1 & 0 & 0 & 1 \\
4 & 1/4 & 1/4 & -2 & -1 & 4 & 2 & 1 & 1/2 & 1/2 & 1/4 & 1/4 & 1/4 \\
4 & 1/4 & 1/4 & 1/2 & 1/2 & 1/4 & 1/4 & 1/4 & 0 & 1 & 0 & 0 & 1 \\
2 & 1 & 1/2 & 1/2 & 1/4 & 1/4 & 1/4 & 1/2 & 1/2 & 1/4 & 1/4 & 1/4 \\
4 & 1/4 & 1/4 & -2 & -1 & 4 & 2 & 1 & -1/2 & -1/2 & 1/4 & 1/4 & 1/4 \\
2 & 1 & -1/2 & -1/2 & 1/4 & 1/4 & 1/4 & -1/2 & -1/2 & 1/4 & 1/4 & 1/4 \\
0 & 0 & -1/2 & -1/2 & 1/4 & 1/4 & 1/4 & 1/2 & 1/2 & 1/4 & 1/4 & 1/4 \\
4 & 1/4 & 1/4 & 1 & 0 & 1 & 0 & 0 & 1/2 & 1/2 & 1/4 & 1/4 & 1/4 \\
4 & 1/4 & 1/4 & -1/2 & -1/2 & 1/4 & 1/4 & 1/4 & 1 & 0 & 1 & 0 & 0
\end{bmatrix}
\]
On inspection of this matrix, we find that the coefficients of $K_3$, $F'$, $F''$ and $F'''$ are linearly related in the following manner:

Coefficient of $K_3 = \text{coefficient of } F''' - \text{coefficient of } F'' - \text{coefficient of } F'$.

Further, if we let

$W = F'' + F'''$,

$X = F' + F'''$,

and $Y = K_3 + F'''$,

we find that there are no more linear relationships among the parameters. The coefficient of $W$ actually equals the coefficient of $F''$, the coefficient of $X$ equals that of $F'$ and the coefficient of $Y$ equals that of $K_3$. From these new parameters we can set up the following matrix equation:
\[
\begin{bmatrix}
    p_1 & \vdots & p_n \\
    c_1 & \vdots & c_m \\
\end{bmatrix}
= 
\begin{bmatrix}
    1 & 1 & -1 & 0 & -2 & 0 & 0 & 0 & 1 & 4 & 2 \\
    1 & -1 & 1 & -2 & 0 & 0 & 4 & 2 & 1 & 0 & 0 \\
    1 & -1 & -1 & -2 & -2 & -2 & 4 & 2 & 1 & 4 & 2 \\
    1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\
    1 & 0 & -1 & 1 & -2 & 1 & 1 & 0 & 0 & 4 & 2 \\
    1 & -1 & 0 & -2 & 1 & 1 & 4 & 2 & 1 & 1 & 0 \\
    1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
    1 & 0 & -1 & 0 & -2 & 0 & 0 & 0 & 0 & 4 & 2 \\
    1 & -1 & 0 & -2 & 0 & 0 & 1 & 2 & 1 & 1 & 0 \\
    1 & 0 & -1/2 & 0 & -1/2 & 1/2 & 0 & 0 & 0 & 1/4 & 1/4 \\
    1 & -1/2 & 0 & -1/2 & 0 & 1/2 & 1/4 & 1/4 & 1/4 & 0 & 0 \\
    1 & -1/2 & -1/2 & -1/2 & -1/2 & 0 & 1/4 & 1/4 & 1/4 & 1/4 & 1/4 \\
    1 & 1/2 & -1/2 & 1/2 & -1/2 & 0 & 1/4 & 1/4 & 1/4 & 1/4 & 1/4 \\
    1 & 1/2 & -1 & 1/2 & -2 & 1/2 & 1/4 & 1/4 & 1/4 & 1/4 & 2 \\
    -1/2 & 1/2 & -1/2 & 1/2 & 0 & 1/4 & 1/4 & 1/4 & 1/4 & 1/4 \\
    1 & -1 & 1/2 & -2 & 1/2 & 1/2 & 1/4 & 2 & 1 & 1/4 & 1/4 \\
    1 & -1/2 & -1 & -1/2 & -2 & -1/2 & 1/4 & 1/4 & 1/4 & 4 & 2 \\
    1 & -1 & -1/2 & -2 & -1/2 & -1/2 & 4 & 2 & 1 & 1/4 & 1/4 \\
    1 & 0 & -1/2 & 1 & -1/2 & 1/2 & 1 & 0 & 0 & 1/4 & 1/4 \\
    1 & -1/2 & 0 & -1/2 & 1 & 1/2 & 1/4 & 1/4 & 1/4 & 1/4 & 1/4 \\
    1 & -1/2 & -1/2 & -1/2 & 1 & 1/4 & 1/4 & 1/4 & 1/4 & 1/4 & 1/4 \\
\end{bmatrix}
\]
We can estimate the fifteen parameters from the parent and cross means,

\[ P_1, P_2, P_3, F_{12}, F_{13}, F_{23}, (F_{12})^2, (F_{13})^2, (F_{23})^2, \]

\[ (F_{12})^3, (F_{13})^3, (F_{23})^3, (F_{12})(F_{13}), (F_{12})(F_{23}), \]

\[ (F_{13})(F_{23}), (F_{12})(F_{13})(F_{23}), (P_1)(F_{12}), (P_2)(F_{12}) \text{ and} \]

\[ (P_3)(F_{13}). \]

If there are more than these fifteen parents and crosses available, we may estimate the parameters by the method of least squares. Using these fifteen parent and cross means we obtain the estimates of the parameters in a matrix equation as follows:
$$
\begin{bmatrix}
I & \hat{I} \\
\hat{I} & \hat{I}
\end{bmatrix}
= 
\begin{bmatrix}
0 & 0 & 1/4 & 1/4 & -1/4 & -1/4 & -1 & 1 & 1 & 0 & 2 & 0 \\
1/4 & 0 & 1/4 & 0 & -1/2 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\
0 & 1/4 & 1/4 & 0 & 0 & -1/2 & -2 & 2 & 0 & 0 & 2 & 0 \\
-1/4 & 0 & 0 & 1/4 & 1/4 & -1/4 & -1 & 1 & 0 & -1 & 0 & 0 \\
0 & -1/4 & 0 & 1/4 & -1/4 & 1/4 & 1 & -1 & 0 & -1 & 0 & 0 \\
0 & 0 & -1/4 & -1/4 & 1/4 & 1/4 & 1 & 0 & -1 & 0 & -1 & 0 \\
1/4 & 0 & 0 & 1/4 & 1/4 & -1/4 & 1 & -1 & 0 & 1/2 & -1/2 & 1/2 \\
-1 & 0 & 0 & -1/2 & -1/2 & 1/2 & -2 & 2 & -1 & -3/2 & 3/2 & -1/2 \\
3/4 & 0 & 0 & 1/4 & 1/4 & -1/4 & -1 & 1 & 1 & -1 & -1 & 0 \\
0 & 1/4 & 0 & 1/4 & -1/4 & 1/4 & -1 & 1 & 0 & 1/2 & 1/2 & -1/2 \\
0 & -1 & 0 & -1/2 & 1/2 & -1/2 & 2 & -2 & -1 & -3/2 & -1/2 & 3/2 \\
0 & 3/4 & 0 & 1/4 & -1/4 & 1/4 & 1 & -1 & 1 & -1 & 0 & -1 \\
0 & 0 & 1/4 & -1/4 & 1/4 & 1/4 & -1 & 0 & 1 & 1/2 & 1/2 & -1/2 \\
0 & 0 & 0 & 1/2 & -1/2 & -1/2 & 2 & 1 & -2 & -3/2 & -1/2 & 1/2 \\
0 & 0 & 0 & -1/4 & -1/4 & 1/4 & 1/4 & 1 & -1 & -1 & 1 & -2 & 0
\end{bmatrix}
$$
\[
\begin{array}{cccccccccc}
\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & 0 & -1 & 1 & 1 & 0 & 2 & 0 & 0 & 0 & -2 \\
0 & -\frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & -2 \\
0 & 0 & -\frac{1}{2} & -2 & 2 & 0 & 0 & 2 & 0 & 0 & 0 & -2 \\
1/4 & 1/4 & -1/4 & -1 & 1 & 0 & -1 & 0 & 0 & 1 & 0 \\
1/4 & -1/4 & 1/4 & 1 & -1 & 0 & -1 & 0 & 0 & 0 & 1 \\
-1/4 & 1/4 & 1/4 & 1 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 1 \\
1/4 & 1/4 & -1/4 & 1 & -1 & 0 & 1/2 & -1/2 & 1/2 & -1 & 0 & 0 \\
-1/2 & -1/2 & 1/2 & -2 & 2 & -1 & -3/2 & 3/2 & -1/2 & 3 & 0 & 0 \\
1/4 & 1/4 & -1/4 & -1 & 1 & 1 & -1 & -1 & 0 & 0 & 0 & 0 \\
1/4 & -1/4 & 1/4 & -1 & 1 & 0 & 1/2 & 1/2 & -1/2 & 0 & -1 & 0 \\
-1/2 & 1/2 & -1/2 & 2 & -2 & -1 & -3/2 & -1/2 & 3/2 & 0 & 3 & 0 \\
1/4 & -1/4 & 1/4 & 1 & -1 & 1 & -1 & 0 & -1 & 0 & 0 & 0 \\
-1/4 & 1/4 & 1/4 & -1 & 0 & 1 & 1/2 & 1/2 & -1/2 & 0 & 0 & -1 \\
1/2 & -1/2 & -1/2 & 2 & 1 & -2 & -3/2 & -1/2 & 1/2 & 0 & 0 & 1 \\
-1/4 & 1/4 & 1/4 & 1 & -1 & -1 & 1 & -2 & 0 & 0 & 0 & 2
\end{array}
\]
The variances of these estimates are obvious.

Using the data from Stringfield (17), we construct Tables 7, 8, 9, 10, 11 and 12.

Table 7

The Means of the Parents and Crosses for the Character,

Days to Midsilk

<table>
<thead>
<tr>
<th>Parent or Cross Mean</th>
<th>$H_y \times 28 \times 51A$</th>
<th>$H_y \times 28 \times 40B$</th>
<th>$H_y \times 40B \times 51A$</th>
<th>$28 \times 40B \times 51A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_1$</td>
<td>90.0</td>
<td>90.0</td>
<td>86.3</td>
<td>81.0</td>
</tr>
<tr>
<td>$P_2$</td>
<td>86.3</td>
<td>86.3</td>
<td>81.0</td>
<td>81.0</td>
</tr>
<tr>
<td>$P_3$</td>
<td>79.3</td>
<td>81.0</td>
<td>79.3</td>
<td>79.3</td>
</tr>
<tr>
<td>$P_{1\times 12}$</td>
<td>81.3</td>
<td>81.3</td>
<td>81.3</td>
<td>80.0</td>
</tr>
<tr>
<td>$P_{1\times 13}$</td>
<td>76.3</td>
<td>81.3</td>
<td>76.3</td>
<td>77.5</td>
</tr>
<tr>
<td>$P_{1\times 23}$</td>
<td>77.5</td>
<td>80.0</td>
<td>74.8</td>
<td>74.8</td>
</tr>
<tr>
<td>$(P_{1\times 12})(P_{1\times 13})$</td>
<td>81.0</td>
<td>79.5</td>
<td>79.5</td>
<td>78.5</td>
</tr>
<tr>
<td>$(P_{1\times 12})(P_{1\times 23})$</td>
<td>80.5</td>
<td>80.5</td>
<td>78.0</td>
<td>77.5</td>
</tr>
<tr>
<td>$(P_{1\times 13})(P_{1\times 23})$</td>
<td>77.5</td>
<td>79.0</td>
<td>78.0</td>
<td>78.0</td>
</tr>
<tr>
<td>$(P_{1\times 12})(P_{1\times 12})$</td>
<td>84.0</td>
<td>84.0</td>
<td>82.0</td>
<td>80.8</td>
</tr>
<tr>
<td>$(P_{1\times 13})(P_{1\times 13})$</td>
<td>80.0</td>
<td>82.0</td>
<td>80.0</td>
<td>80.0</td>
</tr>
<tr>
<td>$(P_{1\times 23})(P_{1\times 23})$</td>
<td>80.0</td>
<td>80.8</td>
<td>77.8</td>
<td>77.8</td>
</tr>
<tr>
<td>$(P_{1\times 12})(P_{1\times 12})$</td>
<td>82.5</td>
<td>82.5</td>
<td>83.5</td>
<td>81.0</td>
</tr>
<tr>
<td>$(P_{2\times 12})(P_{1\times 12})$</td>
<td>83.0</td>
<td>83.0</td>
<td>82.0</td>
<td>79.0</td>
</tr>
<tr>
<td>$(P_{3\times 13})(P_{1\times 13})$</td>
<td>76.5</td>
<td>82.0</td>
<td>76.5</td>
<td>78.0</td>
</tr>
</tbody>
</table>
Table 5
The Estimates of the Fifteen Parameters for the Character, Days to Midsilk

<table>
<thead>
<tr>
<th>Estimate of Parameter</th>
<th>Hy \times 28 x 51A</th>
<th>Hy \times 28 x 40B</th>
<th>Hy \times 40B x 51A</th>
<th>28 x 40B x 51A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>85.700</td>
<td>80.250</td>
<td>85.875</td>
<td>82.750</td>
</tr>
<tr>
<td>X</td>
<td>11.175</td>
<td>2.100</td>
<td>11.175</td>
<td>6.650</td>
</tr>
<tr>
<td>W</td>
<td>8.650</td>
<td>3.825</td>
<td>6.675</td>
<td>4.675</td>
</tr>
<tr>
<td>E'</td>
<td>-4.475</td>
<td>-2.350</td>
<td>-1.800</td>
<td>-1.700</td>
</tr>
<tr>
<td>E''</td>
<td>-1.450</td>
<td>-3.575</td>
<td>-1.200</td>
<td>-1.725</td>
</tr>
<tr>
<td>E'''</td>
<td>-1.700</td>
<td>0.250</td>
<td>4.375</td>
<td>-3.250</td>
</tr>
<tr>
<td>G'</td>
<td>2.525</td>
<td>1.050</td>
<td>1.100</td>
<td>1.550</td>
</tr>
<tr>
<td>L'</td>
<td>-7.050</td>
<td>-4.200</td>
<td>-3.800</td>
<td>-4.750</td>
</tr>
<tr>
<td>M'</td>
<td>0.525</td>
<td>2.150</td>
<td>2.700</td>
<td>1.600</td>
</tr>
<tr>
<td>G''</td>
<td>-0.700</td>
<td>2.175</td>
<td>-1.200</td>
<td>1.075</td>
</tr>
<tr>
<td>L''</td>
<td>-1.050</td>
<td>-4.100</td>
<td>3.800</td>
<td>-3.150</td>
</tr>
<tr>
<td>M''</td>
<td>-0.650</td>
<td>-2.075</td>
<td>-4.00</td>
<td>-1.475</td>
</tr>
<tr>
<td>G''''</td>
<td>-0.050</td>
<td>-0.350</td>
<td>1.375</td>
<td>0.900</td>
</tr>
<tr>
<td>L'''</td>
<td>-1.750</td>
<td>-3.100</td>
<td>-1.500</td>
<td>-1.950</td>
</tr>
<tr>
<td>M'''</td>
<td>-1.700</td>
<td>3.750</td>
<td>-3.875</td>
<td>-1.950</td>
</tr>
</tbody>
</table>
Table 9
The Means of the Parents and Crosses for the Character, Ear Node Height

<table>
<thead>
<tr>
<th>Parent or Cross</th>
<th>Hy x 51A x 28</th>
<th>Hy x 40B x 28</th>
<th>Hy x 51A x 40B</th>
<th>51A x 40B x 28</th>
</tr>
</thead>
<tbody>
<tr>
<td>F₁</td>
<td>31.0</td>
<td>31.0</td>
<td>31.0</td>
<td>24.9</td>
</tr>
<tr>
<td>P₂</td>
<td>24.9</td>
<td>24.5</td>
<td>24.9</td>
<td>24.5</td>
</tr>
<tr>
<td>P₃</td>
<td>21.5</td>
<td>21.5</td>
<td>24.5</td>
<td>21.5</td>
</tr>
<tr>
<td>F₁₁₂</td>
<td>46.5</td>
<td>43.9</td>
<td>46.5</td>
<td>32.4</td>
</tr>
<tr>
<td>F₁¹³</td>
<td>43.0</td>
<td>43.0</td>
<td>43.9</td>
<td>37.5</td>
</tr>
<tr>
<td>F₁²³</td>
<td>37.5</td>
<td>34.2</td>
<td>32.4</td>
<td>34.2</td>
</tr>
<tr>
<td>F₁¹₂ F₁⁻¹³</td>
<td>44.2</td>
<td>40.7</td>
<td>43.5</td>
<td>33.2</td>
</tr>
<tr>
<td>F₁¹₂ F₁⁻²³</td>
<td>41.6</td>
<td>37.9</td>
<td>40.8</td>
<td>33.7</td>
</tr>
<tr>
<td>F₁¹³ F₁⁻²³</td>
<td>38.8</td>
<td>36.7</td>
<td>37.7</td>
<td>32.4</td>
</tr>
<tr>
<td>F₁¹₂ F₁⁻¹²</td>
<td>37.1</td>
<td>37.4</td>
<td>37.1</td>
<td>28.6</td>
</tr>
<tr>
<td>F₁¹³ F₁⁻¹³</td>
<td>36.6</td>
<td>36.6</td>
<td>37.4</td>
<td>30.1</td>
</tr>
<tr>
<td>F₁²³ F₁⁻²³</td>
<td>30.1</td>
<td>27.2</td>
<td>28.6</td>
<td>27.2</td>
</tr>
<tr>
<td>P₁ F₁⁻¹²</td>
<td>42.4</td>
<td>40.3</td>
<td>42.4</td>
<td>31.2</td>
</tr>
<tr>
<td>P₂ F₁⁻¹²</td>
<td>37.8</td>
<td>34.1</td>
<td>37.8</td>
<td>28.6</td>
</tr>
<tr>
<td>P₃ F₁⁻¹³</td>
<td>33.1</td>
<td>33.1</td>
<td>34.1</td>
<td>31.5</td>
</tr>
</tbody>
</table>
Table 10
The Estimates of the Fifteen Parameters for the Character, Ear Node Height

<table>
<thead>
<tr>
<th>Estimate of Parameter</th>
<th>$\hat{Y}$</th>
<th>$\hat{X}$</th>
<th>$\hat{W}$</th>
<th>$\hat{B}'$</th>
<th>$\hat{E}'$</th>
<th>$\hat{E}'$</th>
<th>$\hat{E}''$</th>
<th>$\hat{G}'$</th>
<th>$\hat{L}'$</th>
<th>$\hat{M}'$</th>
<th>$\hat{G}''$</th>
<th>$\hat{L}''$</th>
<th>$\hat{M}''$</th>
<th>$\hat{G}'''$</th>
<th>$\hat{L}'''$</th>
<th>$\hat{M}'''$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{Y}$</td>
<td>40.075</td>
<td>37.950</td>
<td>40.275</td>
<td>25.550</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{X}$</td>
<td>-1.375</td>
<td>-1.375</td>
<td>-1.475</td>
<td>-9.950</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{W}$</td>
<td>-5.350</td>
<td>4.200</td>
<td>-2.650</td>
<td>-7.600</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{B}'$</td>
<td>7.950</td>
<td>5.525</td>
<td>9.350</td>
<td>5.700</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{E}'$</td>
<td>7.325</td>
<td>2.150</td>
<td>5.925</td>
<td>0.950</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{E}''$</td>
<td>5.025</td>
<td>3.450</td>
<td>3.825</td>
<td>6.750</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{G}'$</td>
<td>-3.750</td>
<td>-2.575</td>
<td>-3.300</td>
<td>-3.650</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{L}'$</td>
<td>10.400</td>
<td>6.450</td>
<td>10.250</td>
<td>7.900</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{M}'$</td>
<td>-1.250</td>
<td>-3.675</td>
<td>-1.750</td>
<td>1.700</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{G}''$</td>
<td>-2.125</td>
<td>1.400</td>
<td>-2.575</td>
<td>0.750</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{L}''$</td>
<td>5.600</td>
<td>4.450</td>
<td>7.250</td>
<td>-3.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{M}''$</td>
<td>3.125</td>
<td>2.050</td>
<td>2.125</td>
<td>1.850</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{G}'''$</td>
<td>-2.825</td>
<td>0.000</td>
<td>-3.375</td>
<td>-1.450</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{L}'''$</td>
<td>9.600</td>
<td>1.550</td>
<td>11.550</td>
<td>3.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{M}'''$</td>
<td>-2.975</td>
<td>-0.550</td>
<td>-3.175</td>
<td>3.050</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 11

The Means of the Parents and Crosses for the Character, Yield

<table>
<thead>
<tr>
<th>Parent or Cross</th>
<th>51A x 28 x Hy</th>
<th>40B x 28 x Hy</th>
<th>51A x 40B x Hy</th>
<th>51A x 40B x 28</th>
</tr>
</thead>
<tbody>
<tr>
<td>P₁</td>
<td>53.3</td>
<td>46.0</td>
<td>53.3</td>
<td>53.3</td>
</tr>
<tr>
<td>P₂</td>
<td>39.5</td>
<td>39.5</td>
<td>46.0</td>
<td>46.0</td>
</tr>
<tr>
<td>P₃</td>
<td>27.2</td>
<td>27.2</td>
<td>27.2</td>
<td>39.5</td>
</tr>
<tr>
<td>P₁₁²</td>
<td>86.0</td>
<td>86.6</td>
<td>84.6</td>
<td>84.6</td>
</tr>
<tr>
<td>P₁₁³</td>
<td>88.1</td>
<td>78.6</td>
<td>88.1</td>
<td>86.0</td>
</tr>
<tr>
<td>P₁²³</td>
<td>85.2</td>
<td>85.2</td>
<td>78.6</td>
<td>86.8</td>
</tr>
<tr>
<td>(P₁₁²) (P₁₁³)</td>
<td>76.9</td>
<td>76.2</td>
<td>77.4</td>
<td>75.6</td>
</tr>
<tr>
<td>(P₁¹²) (P₁²³)</td>
<td>81.4</td>
<td>82.0</td>
<td>70.0</td>
<td>69.7</td>
</tr>
<tr>
<td>(P₁¹³) (P₁²³)</td>
<td>66.9</td>
<td>74.8</td>
<td>74.8</td>
<td>76.9</td>
</tr>
<tr>
<td>(P₁¹²) (P₁¹²)</td>
<td>62.9</td>
<td>61.7</td>
<td>61.8</td>
<td>61.8</td>
</tr>
<tr>
<td>(P₁¹³) (P₁¹³)</td>
<td>60.6</td>
<td>61.5</td>
<td>60.6</td>
<td>62.9</td>
</tr>
<tr>
<td>(P₁²³) (P₁²³)</td>
<td>64.8</td>
<td>64.8</td>
<td>61.5</td>
<td>61.7</td>
</tr>
<tr>
<td>(P₁) (P₁¹²)</td>
<td>74.3</td>
<td>65.9</td>
<td>74.6</td>
<td>74.6</td>
</tr>
<tr>
<td>(P₂) (P₁¹²)</td>
<td>66.5</td>
<td>71.6</td>
<td>65.9</td>
<td>65.9</td>
</tr>
<tr>
<td>(P₃) (P₁¹³)</td>
<td>60.5</td>
<td>58.7</td>
<td>60.5</td>
<td>66.5</td>
</tr>
</tbody>
</table>
Table 12

The Estimates of the Fifteen Parameters for the Character, Yield

<table>
<thead>
<tr>
<th>Estimate of Parameter</th>
<th>51A x 28 x 40B</th>
<th>x 40B x 28 x 51A</th>
<th>x 51A x 40B</th>
<th>x 51A x 28</th>
</tr>
</thead>
<tbody>
<tr>
<td>^ I</td>
<td>56.575</td>
<td>73.750</td>
<td>53.875</td>
<td>51.625</td>
</tr>
<tr>
<td>^ X</td>
<td>-23.725</td>
<td>-15.400</td>
<td>-23.725</td>
<td>-27.000</td>
</tr>
<tr>
<td>^ W</td>
<td>-16.725</td>
<td>-8.725</td>
<td>-35.600</td>
<td>-41.025</td>
</tr>
<tr>
<td>^ E</td>
<td>24.800</td>
<td>18.550</td>
<td>15.600</td>
<td>14.525</td>
</tr>
<tr>
<td>^ E II</td>
<td>10.000</td>
<td>17.575</td>
<td>18.775</td>
<td>19.850</td>
</tr>
<tr>
<td>^ G</td>
<td>-9.700</td>
<td>-7.650</td>
<td>1.000</td>
<td>-4.125</td>
</tr>
<tr>
<td>^ L</td>
<td>31.400</td>
<td>15.700</td>
<td>1.300</td>
<td>10.700</td>
</tr>
<tr>
<td>^ M</td>
<td>10.100</td>
<td>11.950</td>
<td>8.500</td>
<td>7.225</td>
</tr>
<tr>
<td>^ G II</td>
<td>-2.000</td>
<td>-3.375</td>
<td>-12.575</td>
<td>-7.450</td>
</tr>
<tr>
<td>^ L II</td>
<td>15.100</td>
<td>16.100</td>
<td>23.400</td>
<td>12.300</td>
</tr>
<tr>
<td>^ M II</td>
<td>-14.900</td>
<td>-4.525</td>
<td>12.175</td>
<td>15.150</td>
</tr>
<tr>
<td>^ G III</td>
<td>12.525</td>
<td>4.850</td>
<td>-5.325</td>
<td>-1.775</td>
</tr>
<tr>
<td>^ L III</td>
<td>26.000</td>
<td>14.100</td>
<td>2.400</td>
<td>-3.800</td>
</tr>
<tr>
<td>^ M III</td>
<td>6.325</td>
<td>-12.050</td>
<td>7.925</td>
<td>10.175</td>
</tr>
</tbody>
</table>
The parameters which were estimated in Tables 8, 10 and 12 are defined as follows:

Y is the overall average contribution,

X, W, E', E'' and E''' are the loci effects parameters
and G', L', M', G'', L'', M'' and G'''' are the two-factor interaction or the epistatic parameters.

The importance of epistasis could be more clearly stated if the standard errors of the estimates were available. However, we can compare the absolute values of the estimates of the epistatic parameters with those of the loci effects for these data. For all characters there are some estimates of epistatic parameters which are of the same magnitude as those of the loci effects. In many cases the estimates of epistatic parameters are larger than the estimates of loci effects. Therefore, we conclude that the contributions due to epistasis are of major importance in these data.

If there is reason to believe that there is no epistasis present in the character being investigated, estimates of E', E'', E''', X, Y and Z may be obtained from the parent and F1 means only. The estimates of these parameters, displayed in a matrix equation, are as follows:
The variances of the estimates are obvious.

Using $\hat{E}'$, $\hat{E}''$, $\hat{E}'''$, $\hat{W}$, $\hat{X}$, $\hat{Y}$ we may estimate the three way crosses as

\[
\begin{align*}
(P_1)(F_1^{23}) &= \frac{1}{2} \left[ F_1^{12} + F_1^{13} \right] \\
(P_2)(F_1^{13}) &= \frac{1}{2} \left[ F_1^{12} + F_1^{23} \right] \\
(P_3)(F_1^{12}) &= \frac{1}{2} \left[ F_1^{13} + F_1^{23} \right]
\end{align*}
\]

These estimates agree with the method of prediction given by \textit{Dorotator} and \textit{Johnson} (5). Here, again as in the two parent case, we exemplify the flexibility of the model in that when we delete the epistatic parameters for three parents we obtain well-known results. If all three of these estimates give satisfactory results, this constitutes some evidence that epistasis is not important for that character because these formulae are derived on the basis of there being no epistasis.

It is interesting to note that the prediction equations for the three way crosses, allowing that epistasis is present and using the estimates described previously, are
\begin{align*}
&\left(\hat{p}_1\right)\left(\hat{p}_1^{23}\right) = \frac{1}{2} \hat{p}_1^{12} + \frac{1}{2} \hat{p}_1^{13} - \frac{1}{2} \hat{p}_1^{23} + \left(\hat{p}_1^{12}\right)\left(\hat{p}_1^{13}\right) - \\
&\quad \frac{1}{8}\left(\hat{p}_1^{12}\right)\left(\hat{p}_1^{12}\right) - \frac{1}{2}\left(\hat{p}_1^{13}\right)\left(\hat{p}_1^{13}\right) + \frac{1}{8}\left(\hat{p}_1^{23}\right)\left(\hat{p}_1^{23}\right) \\
&\left(\hat{p}_2\right)\left(\hat{p}_1^{13}\right) = \frac{1}{2} \hat{p}_1^{12} + \frac{1}{2} \hat{p}_1^{23} - \frac{1}{2} \hat{p}_1^{13} + \left(\hat{p}_1^{12}\right)\left(\hat{p}_1^{23}\right) - \frac{1}{2}\left(\hat{p}_1^{12}\right)\left(\hat{p}_1^{12}\right) + \\
&\quad \frac{1}{8}\left(\hat{p}_1^{13}\right)\left(\hat{p}_1^{13}\right) - \frac{1}{2}\left(\hat{p}_1^{23}\right)\left(\hat{p}_1^{23}\right) \\
&\left(\hat{p}_3\right)\left(\hat{p}_1^{12}\right) = \frac{1}{2} \hat{p}_1^{13} + \frac{1}{2} \hat{p}_1^{23} - \frac{1}{2} \hat{p}_1^{12} + \left(\hat{p}_1^{13}\right)\left(\hat{p}_1^{23}\right) + \frac{1}{2}\left(\hat{p}_1^{12}\right)\left(\hat{p}_1^{12}\right) + \\
&\quad \frac{1}{8}\left(\hat{p}_1^{13}\right)\left(\hat{p}_1^{13}\right) - \frac{1}{2}\left(\hat{p}_1^{23}\right)\left(\hat{p}_1^{23}\right).
\end{align*}
VII. SUMMARY AND CONCLUSIONS

The aim of the present study was the development of a model which could be applied towards the interpretation and understanding of data on quantitative inheritance.

Previous work in the area was reviewed and discussed in the light of characteristics which such a model should possess. The desirable characteristics were taken to be:

(a) additivity of the parameters,
(b) parameters which have genetic interpretation,
(c) applicability to a genotypic value for any number of loci,
(d) symmetry with respect to the homozygous phases at each locus
and (e) flexibility of the model with respect to increasing assumptions.

The model which was developed is related to the factorial model used in the design of experiments. The factorial model used is based on comparisons in the population of all possible genotypes in which heterozygotes are given twice the weight of homozygotes. As a result the model for the genotypic value of an individual represented by \(a_i b_j c_k\ldots\), that is, an individual in which the \(a\) locus is in the \(i\)th phase, the \(b\) locus is in the \(j\)th phase and so on, is

\[
\mu + A_i + B_j + A_i B_j + C_k + A_i C_k + B_j C_k + A_i B_j C_k + \ldots
\]
where symbols such as $A_i B_j$ while formally, in a sense, a product of $A_i$ and $B_j$ are in fact definite quantities which are not the products of the quantities represented by $A_i$ and $B_j$, for example. The terms in the model have explicit meaning which is described in the body of the thesis. It suffices here to state that $\mu$ is an overall effect, symbols involving one letter such as $A_i$ represent the average effect of a locus when in the $i^{th}$ phase, symbols such as $A_i B_j$ represent the two-factor interaction contribution arising as a deviation from the overall effect plus the loci effects and so on. Symbols involving $n$ loci represent the $n$-factor interaction arising from the deviation of a genotypic value for a genotype involving $n$ loci, from the value which would be predicted on the basis of effects and interactions involving a subset of the $n$ loci. Thus the model represents the genotypic value as the sum of a constant effect, of loci effects and contributions from interaction of loci. The latter constitute epistatic contributions. This model possesses all of the desirable characteristics described above.

In all the work described in this thesis it was assumed that linkages and lethal genes were absent and that viability was constant for all genotypes.

The factorial model was applied to the consideration of means of populations arising successively by selfing from a single individual. The general formula for the mean genotypic value of
the population produced by $m$ generations of selfing is

$$F_m = X_1 + \sum \left( \frac{1}{2^m - 1} - 1 \right) \beta_p$$

where $\beta_p$ is the contribution of the $p$-factor interactions and $X_1$ is the overall effect of the loci which are homozygous in the original individual.

The model was applied to data, Khambanonda (12), and it was concluded that in his material the epistatic contributions may be an important cause of inbreeding depression. Loci effects and epistatic parameters along with their standard errors were estimated from these data and it was found in almost all cases that the epistatic contribution differed significantly, at the 5% level, from zero.

The model was applied to some of the populations which arise by crossing of two inbred lines and subsequent crossing and selfing. From the observed population means certain functions of the loci effects and epistatic effects can be estimated. These results were applied to data given by Stringfield (17) and it was found for the characters considered, assuming low order interactions only, that the epistatic components were an important part of the observed mean genotypic values. The flexibility of the model was shown by assuming no epistasis was present, deleting the interaction parameters and obtaining the usual estimates of the $F_2$ and backcross means. The model was used to show that the scaling tests described
by Mather (13) are not necessarily tests for the absence of epistasis.

The difficulty of keeping an accurate count of the various parameters increases with increasing number of parents. However populations arising from three homozygous parents were considered extensively under the assumption that interactions between three or more loci could be ignored. Using this framework it was shown that the epistatic components were not trivial relative to loci effects, for the material examined by Stringfield (17).

The model was also applied to simple populations arising from four homozygous parents and the formula of Doxtator and Johnson (5) for the prediction of double crosses was shown to hold in the absence of epistasis. The examination of this situation in the presence of epistasis was very tedious and none of the material on this case is included in the thesis.

One of the problems which could be investigated is the derivation of the variances of the useful n-loci populations. A procedure has been indicated of obtaining the variance of the F_2 of a two-loci population; however, the problem will undoubtedly be much more difficult for other generations. Another problem, the solution to which might prove quite interesting, is the derivation of a scaling test which would test for the absence of interactions of three or more loci.
VIII. BIBLIOGRAPHY


17. Stringfield, G. H. Senior Agronomist, Agricultural Experiment Station, Wooster, Ohio. (Data obtained by personal communication).
IX. ACKNOWLEDGMENTS

The author wishes to express his deepest appreciation to Professor Oscar Kempthorne for suggesting the problem and guiding the research as well as for invaluable assistance in the final construction of the thesis.

The author also wishes to thank the Iowa Experiment Station for financial support during his period of study at Iowa State College.