R. E. COMSTOCK
and
H. F. ROBINSON
North Carolina State College

Chapter 30

Estimation of Average Dominance of Genes

This discussion will center around three experimental procedures used at the North Carolina Experiment Station for investigating the degree of dominance involved in the action of genes that affect quantitative characters of economic plants. The objective is twofold: (1) to outline and, in so far as possible, evaluate these methods; and (2) by example, to point up the role of statistics in genetical research.

Basic criteria for the usefulness of a projected experiment are: (1) Will data obtained provide a logical basis for inference relative to the research objective? (2) Will the random variability in the experiment be of an order that will permit satisfactory certainty of conclusions? The latter has obvious statistical overtones, but statistics is not always deeply involved in the former. In genetic work, random variability in the experimental material is generated in part by the genetic mechanism, and can therefore be used as a basis for inference in genetic problems. Hence statistics plays an inescapable role in both aspects of the evaluation of many genetic experiments.

Examination of any proposed basis for inference must obviously center on the premises involved and the validity of deductions predicated on those premises. We will turn first, therefore, to description of the experiments and the logical basis for the estimates they are designed to provide.

THE EXPERIMENTAL DESIGNS

The designs of each of these experiments have two aspects: (1) the genetic background and (2) the field arrangement of the material on which data are collected.

Experiment I

The experimental material is produced from matings among plants of the F_2 generation of a cross of two inbred lines. Each plant used as a pollen parent is mated with n seed parents, no seed parent being involved in more than one mating. Thus, if sm pollen parents are used there will be smn seed parents used in smn matings. The progenies of these smn matings comprise the experimental material. All parent plants are chosen from the F_2 population at random.

Pollen parent and seed parent plants will for brevity be referred to in what follows as males and females, respectively. A group of n progenies having the same male parent will be referred to as a male group.

The field arrangement of the material is based on division of the sm male groups into s sets each of which contains mn progenies in m male groups. Each set of progenies comprises the material for a distinct unit of the experiment and is planted in a randomized block arrangement having mn entries and r replications. Thus the total field arrangement is composed of s independent units, each unit being devoted to a different set of progenies. Data on characters of interest are collected on k plants per plot.

Experiment II

This is a modification of Experiment I that can be used when dealing with multi-flowered plants. The foundation stock is again the F_2 generation of a cross of inbred lines. In this case, however, a set of mn progenies is produced by making all of the mn possible matings of m males and n females chosen at random from the F_2 population. With annual plants this can be done (and the progenies kept distinct) only if more than one pistillate flower per plant is available. It could not be done, for example, with single-eared corn.

The field arrangement is as described for Experiment I, the sets arising from the mating plan being maintained intact in the units of the field structure of the experiment.

Experiment III

The experimental material is produced from backcross matings of F_2 plants to the two inbred lines from which the F_2 was derived; the F_2 plants are used as pollen parents. A set of progenies is made up of the $2n$ progenies obtained from backcrossing n random F_2 plants to each of the parent inbreds. The number of inbred plants used in production of each backcross progeny is important only with respect to insuring sufficient seed.

The total experimental material consists of s sets of n pairs of progenies. The members of each pair have the same F_2 (male) parent but different inbred parents. The two inbred parents are, of course, the same for all pairs of progenies.

The field arrangement is analogous to that for Experiments I and II. The

unit in this case is a randomized block arrangement ($2n$ entries and r replications) of r plots of each of the progenies of a set.

ANALYSIS OF DATA

The appropriate variance analyses for the data of the three experiments are outlined in Tables 30.1 to 30.3. The expected value (the value that

TABLE 30.1

VARIANCE ANALYSIS (EXPERIMENT I)

Source of Variance	d.f.	m.s.	Expectation of m.s.*
Sets	$s-1$	.....	.....
Replications in sets	$s(r-1)$	.....	.....
Males in sets	$s(m-1)$	M_{11}	$\sigma^2+r\sigma_f^2+rn\sigma_m^2$
Females in males in sets	$sm(n-1)$	M_{12}	$\sigma^2+r\sigma_f^2$
Remainder among plots	$s(mn-1)(r-1)$	M_{13}	σ^2

* σ^2 is the error variance among plots of the same progeny (due in part to soil variation among plots in the same block and in part to variation among plants of the same progeny).

σ_f^2 is progeny variance arising from genetic differences among female parents.

σ_m^2 is progeny variance arising from genetic differences among male parents.

TABLE 30.2

VARIANCE ANALYSIS (EXPERIMENT II)

Source of Variance	d.f.	m.s.	Expectation of m.s.*
Sets	$s-1$	.....	.....
Replications in sets	$s(r-1)$	.....	.....
Males in sets	$s(m-1)$	M_{21}	$\sigma^2+r\sigma_{fm}^2+rn\sigma_m^2$
Females in sets	$s(n-1)$	M_{22}	$\sigma^2+r\sigma_{fm}^2+rm\sigma_f^2$
Males×females in sets	$s(m-1)(n-1)$	M_{23}	$\sigma^2+r\sigma_{fm}^2$
Remainder among plots	$s(mn-1)(r-1)$	M_{24}	σ^2

* σ_{fm}^2 is progeny variance arising from interaction of genotypes of male and female parents. Other symbols are defined in Table 30.1.

would be approached as a limit if the amount of data were made infinitely large) is listed for each mean square to be used in interpretations.

In order to specify the significance of components of the total variance of which estimates can be used for inferences about dominance, some additional symbolism must first be established. Consider the three genotypes possible at a locus where there is segregation between two alleles. Let the difference in effect of the two homozygous genotypes on a measured character be $2u$ and the deviation of the effect of the heterozygous genotype from the mean effect of the homozygous genotypes be au. Note that u and au have the same significance as d and h, respectively, in the symbolism used by Fisher *et al.* (1932). Also, they have the same significance as d and k in the symbolism

employed earlier in the Heterosis Conference. The symbols u and au are used here for consistency with usage in articles by Comstock and Robinson (1948) and Comstock *et al.* (1949). Let the number of segregating gene pairs that affect a particular character be symbolized as N, and a numerical subscript to u or a specify the locus to which the symbolized quantity is relevant. Thus $2u_3$ is the difference in effect of the two homozygous genotypes of the third locus and a_5u_5 is the dominance deviation for the fifth locus.

Now granting validity of several assumptions (to be listed and discussed later) σ_m^2, σ_f^2, σ_{mf}^2, and σ_{ml}^2 have genetic meaning as set out in Table 30.4.

TABLE 30.3

VARIANCE ANALYSIS (EXPERIMENT III)

Source of Variance	d.f.	m.s.	Expectation of m.s.*
Sets	$s-1$	.....	.....
Replications in sets	$s(r-1)$	.....	.....
Inbred line in sets	s	.....	.....
F_2 parent in sets	$s(n-1)$	M_{31}	$\sigma^2+2r\sigma_m^2$
F_2 parent×line in sets	$s(n-1)$	M_{32}	$\sigma^2+r\sigma_{ml}^2$
Remainder among plots	$s(2n-1)(r-1)$	M_{33}	σ^2

* σ_m^2 is progeny variance arising from genetic differences among F_2 (male) parents.

σ_{ml}^2 is progeny variance arising from interaction of genotypes of F_2 and inbred parents.

THE ESTIMATE OF AVERAGE DOMINANCE

The magnitude of a measures the degree of dominance in the action of any one pair of genes, being related to qualitative classification of dominance as follows:

Class of Dominance	Numerical Value of a
No dominance	$a=0$
Partial dominance	$0<a<1.0$
Complete dominance	$a=1.0$
Overdominance	$a>1.0$

However, a problem arises concerning the best way to represent the average dominance for all loci with a single number. An obvious possibility is the unweighted mean of a's for all gene pairs. On the other hand, it can be argued that a mean in which individual a's are weighted relative to the importance of loci would be more useful. This in turn raises the question of how the relative importance of loci should be measured. However, the matter will not be pursued further, since the experiments under consideration offer no choice of measure to be estimated.

The estimate that can be made is of

$$\bar{a}^2 = \frac{\Sigma a^2u^2}{\Sigma u^2} \quad \text{or} \quad \bar{a} = \sqrt{\frac{\Sigma a^2u^2}{\Sigma u^2}}$$

$\bar{a}^2$ is a weighted average of all a^2s, weighting being relative to the square of u (one of the possible measures of the importance of loci). $\bar{a}^2$ and $\bar{a}$ can exceed unity only if one or more individual a's are larger than one, but values of $\bar{a}^2$ in excess of one do not exclude the possibility of partial dominance at numerous

TABLE 30.4

GENETIC NATURE OF COMPONENTS OF PROGENY VARIANCE*

Component	Experiment		
	I	*II*	*III*
σ_m^2	$\frac{1}{8}\Sigma u^2$	$\frac{1}{8}\Sigma u^2$	$\frac{1}{8}\Sigma u^2$
σ_f^2	$\frac{1}{8}\Sigma u^2+\frac{1}{16}\Sigma a^2u^2$	$\frac{1}{8}\Sigma u^2$	
σ_{mf}^2		$\frac{1}{16}\Sigma a^2u^2$	
σ_{ml}^2			$\frac{1}{4}\Sigma a^2u^2$

* Summation is in all cases over loci, i.e.,
$\Sigma u^2 = (u_1^2+u_2^2+ \ldots u_N^2)$ and
$\Sigma a^2u^2 = (a_1^2u_1^2 + a_2^2u_2^2 + \ldots a_N^2u_N)$

loci. On the other hand, $\bar{a}^2$ will not be less than one unless dominance is less than complete at one or more loci, but values less than one do not insure absence of overdominance at all loci.

Experiment I

In accordance with the mean square expectations of Table 30.1 we can estimate

$$\sigma_m^2 \quad \text{as} \quad (M_{11} - M_{12}) / rn$$

and

$$\sigma_f^2 \quad \text{as} \quad (M_{12} - M_{13}) / r$$

and from Table 30.4 we see that in this experiment

$$\sigma_m^2 = \tfrac{1}{8}\Sigma u^2$$

and

$$\sigma_f^2 - \sigma_m^2 = \tfrac{1}{16}\Sigma a^2u^2$$

Hence,

$$\frac{2\,[\,(n+1)\,M_{12} - M_{11} - nM_{13}]}{M_{11} - M_{12}} \text{ is an estimate of } \frac{\Sigma a^2u^2}{\Sigma u^2} = \bar{a}^2\,.$$

Experiment II

Note first from Table 30.4 that in this experiment $\sigma_m^2 = \sigma_f^2$. If the experiment is designed with $m = n$ (this will be assumed in what follows since it is a rational procedure where possible) this means that the expectation of M_{21}

and M_{22} (see Table 30.2) are equal and hence that the two mean squares may be pooled.[1] Let the pooled mean square be symbolized by M_{20}. Then

$$(M_{20} - M_{23}) / rn \text{ estimates } \sigma_f^2 = \sigma_m^2$$

and

$$(M_{23} - M_{24}) / r \text{ estimates } \sigma_{fm}^2$$

In this experiment (see Table 30.4)

$$\sigma_f^2 = \sigma_m^2 = \tfrac{1}{8}\Sigma u^2$$

and

$$\sigma_{fm}^2 = \tfrac{1}{16}\Sigma a^2 u^2$$

It follows that

$$\frac{2n\,(M_{23} - M_{24})}{M_{20} - M_{23}} \text{ estimates } \frac{\Sigma a^2 u^2}{\Sigma u^2} = \bar{a}^2 .$$

Experiment III

Following Tables 30.3 and 30.4 we see that

$$(M_{31} - M_{33}) / 2r \text{ estimates } \sigma_m^2 = \tfrac{1}{8}\Sigma u^2$$

$$(M_{32} - M_{33}) / r \text{ estimates } \sigma_{ml}^2 = \tfrac{1}{4}\Sigma a^2 u^2$$

so that

$$\frac{M_{32} - M_{33}}{M_{31} - M_{33}} \text{ estimates } \bar{a}^2 .$$

ASSUMPTIONS

Evaluation of procedures described above should obviously begin with examination of assumptions underlying derivations of mean square expectations listed in Tables 30.1 to 30.3 and genetic interpretations placed on variance components in Table 30.4. Premises involved in the derivation of mean square expectations were as follows:

1. Random choice of individuals mated for production of experimental progenies.
2. Random distribution of genotypes relative to variations in environment.
3. No non-genic maternal effects.

The first of these can be assured easily in the conduct of the experiment. The second is equally easy to assure in so far as environmental variations within the experiment are concerned. On the other hand, the environments encountered in an experiment conducted within the confines of a single year

1. By taking an unweighted mean since degrees of freedom will also be equal when $m = n$.

and location do not constitute a random sample of the environments that occur within the wider limits of time and space for which we would like experimental findings to apply. The consequence of this is that, if interaction of genotype with environment is a source of variation, each mean square arising from variation among progenies will contain some variance from such interaction. Thus, to have been rigorously correct, the expectations of all mean squares between progenies should have included terms recognizing contributions from this source. Separate estimation of the genetic and interaction components of mean squares between progenies could not be effected with data collected in a single year and location. If the ratio of these two sorts of variance is constant for the several mean squares, and there is no obvious reason why it should vary, the presence of interaction variance does not bias the estimates of $\bar{a}^2$ since numerator and denominator are affected proportionately. Nevertheless this constitutes a possible weakness of the methods but one which, if important, can be corrected by replication of all progenies over years and locations.

There are many characters and organisms for which it appears safe to assume maternal affects are absent or of no consequence. This assumption must be viewed with some suspicion when dealing with seedling characters of plants or any character for which there is any hint that cytoplasmic inheritance may be operating, and it is definitely not tenable for pre-maturity characters of mammals. Maternal effects do not contribute to the pertinent mean squares in the variance analysis of Experiment III and only to M_{22} in that of Experiment II. Thus these two experiments are useful in the presence of maternal effects, though if II is used Σu^2 must be estimated from M_{21} instead of jointly from M_{21} and M_{22}.

Assumptions involved in deriving the genetic interpretations of variance components are as follows:

1. Regular diploid behavior at meiosis.
2. Population gene frequencies of one-half at all loci where there is segregation (not necessary for Experiment III).
3. No multiple allelism.
4. No correlation of genotypes at separate loci. This implies no linkage among genes affecting the character studied or that, if linkages exist, the distribution of genotypes is at equilibrium with respect to coupling and repulsion phases.
5. No epistasis, *i.e.*, the effect of variation in genotype at any single locus is not modified by genes at other loci.

In accord with the first of these, usefulness of the procedures described is limited to studies with diploids or amphidiploids in which multivalent meiotic associations are entirely absent or are absent in meiotic divisions giving rise to fertile gametes.

Save for deviations due to natural selection, the second assumption is assured by the fact that the population used is an F_2 of a cross of homozygous lines. Moreover, natural selection strong enough to have more than a trivial effect on gene frequencies can only occur if the development of a moderately large proportion of F_2 plants is so slow or aberrant as to prevent their effective use as parents. Thus a good stand of usable plants constitutes insurance that this assumption is satisfied.

Number three is also assured by the origin of the populations. Multiple alleles in an F_2 of homozygous lines can result only from mutation, and in the light of present knowledge of mutation rates would be expected very infrequently.

On the other hand, complete validity of the fourth assumption is improbable. Present day geneticists are in general agreement that quantitative characters, and particularly physiologically complex ones such as yield, are influenced by many genes. If that is so, there may well be linkages among some of the genes affecting any single character. Furthermore, specific linkage relationships in an F_1 of homozygous lines must be in either the coupling or repulsion phase, and equilibrium between the phases cannot occur in the F_2. In fact the approach to equilibrium in later generations is rather slow unless linkage is very loose (see Wright, 1933).

The effect of linkage is to cause upward bias in estimates of $\bar{a}$. Thus Comstock and Robinson (1948) in discussion of Experiments I and II and Robinson *et al.* (1949) in discussing results obtained using Experiment I indicated that values of $\bar{a}$ larger than one can result either from true overdominance or from repulsion linkage of genes that are completely or partially dominant to their alleles. The same conclusion can be inferred from Mather (1949).

The situation can be summarized in another manner by stating that values of $\bar{a}$ in excess of unity do not distinguish true overdominance in the action of alleles from what Mangelsdorf has termed pseudo-overdominance or overdominance at the gamete level. However, in defense of the procedures under discussion, it must be emphasized that knowing one or the other of these two phenomena is at work is an advance over being uncertain as to whether either is operative. On the other hand there is good reason to attempt to distinguish which is responsible if estimates of $\bar{a}$ by the methods described are much greater than one. One source of such supplementary information is an extension of Experiment III to be considered briefly in the next section.

The assumption of no epistasis is no more realistic than that of no linkages. It has been pointed out (Comstock and Robinson, 1948) that epistasis probably causes upward bias in the estimates of $\bar{a}$, but that the amount of bias may not be large. Subsequent investigation of several simple epistatic models with respect to expected values of estimates of $\bar{a}$ from Experiments I and II have turned up nothing to change that point of view. It must be emphasized

that the matter has not been considered exhaustively, and the possibility remains that in some materials epistasis would be responsible for serious overestimation of $\bar{a}$ by the methods being discussed.

The authors' knowledge of the situation may be summarized briefly as follows. It appears possible that with complete dominance the rule, $\bar{a} = 1.0$, epistasis might bias estimates upwards by as much as .10 to .25. This cannot be considered serious against the background of an actual estimate of 1.6 as reported for grain yield in corn by Robinson *et al.* (1949). On the other hand, genetic models can be specified in which the consequences of epistasis would be serious, but to date no such models have been discovered that seem likely to have reality in nature.

Much investigation of the epistasis problem remains to be done. Theoretical studies of a variety of epistatic models are needed as a basis for understanding (1) how and to what extent inferences based on expectations derived from non-epistatic models may be in error, and (2) how epistasis may be measured and characterized experimentally. Equally important are experimental investigations of the role of epistasis in inheritance of quantitative characters of various organisms. The problem in this connection is one of knowing how to obtain critical information. The most familiar approach is that of studying the regression of character measurements on levels of homozygosity as represented at the extremes by inbred lines and F_1's and at intermediate levels by F_2's and various sorts of backcrosses. While this approach has admitted shortcomings, it has not been exploited to the limit of its usefulness. Other possibilities are suggested by Mather (1949).

EVALUATION OF THE ROLE OF LINKAGE

It was pointed out above that repulsion linkages are a source of upward bias in estimates of $\bar{a}$. In fact if a moderately high number of genes is postulated, one finds on careful examination that estimates in excess of one seem inevitable unless dominance at the locus level is considerably less than complete. From the point of view of breeding methods it then becomes important to distinguish between true overdominance and pseudo-overdominance. Particularly is this true if the latter is to any important degree a consequence of linkages that are loose enough to allow their effects to be dissipated by recombination in a few generations of random breeding, as opposed to the rather durable associations that appear to be postulated by Anderson (1949).

If the assumption that frequencies of genes at all segregating loci are one-half were tenable for generations beyond the F_2, any of the three experiments would provide a basis for obtaining information on the role of linkage. The procedure would be to compare estimates obtained as described with others obtained when parents were taken from an advanced generation (produced under random mating, not with inbreeding) rather than from the F_2. In fact one might systematically repeat the experiment using each successive genera-

tion as it became available. Then if loose linkages were of much importance in the first estimate, one would anticipate a downward trend in the estimates of $\bar{a}$ as more and more advanced generations were employed. Natural selection too weak to have much effect on results when the F_2 is used could be the source of significant changes in gene frequencies over a period of generations. Hence the effects of recombination and of shifting gene frequencies would be confounded in trends observed using either Experiment I or II.

Fortunately, Experiment III does not depend on any assumption about gene frequencies. Letting q symbolize the population frequency of any gene and $1 - q$ that of its allele, the genetic interpretation of σ^2_m and σ^2_{ml} can be expressed more generally than in Table 30.4 as $\frac{1}{2}\Sigma q(1 - q)u^2$ and $\Sigma q(1 - q)a^2u^2$, respectively. One possible weakness of the proposal is apparent. If shifts in gene frequency are variable by loci the weighting of individual a's in $\bar{a}^2$ is shifted slightly since it is now relative to $q(1 - q)u^2$ rather than u^2. However, barring shifts greater than .2 which are unlikely unless a gene has a very important effect, shifts in weights will be of minor magnitude since $q(1 - q)$ varies only between .21 and .25 as q varies from .3 to .7. Furthermore, shifts in weight are not a source of bias unless degree of dominance (size of a) is correlated with importance of the gene. While this weakness should not be overlooked, it appears of minor consequence. A partial check could be made by accumulating seed of each generation for a yield comparison of the successive generations. If major gene frequency trends have occurred at important loci they should be evidenced in higher yields by the later generations.

The suggested extension of Experiment III is intrinsically the same sort of technic as Mather (1949) has outlined for investigating linkage effects on genetic variances.

DERIVATION OF GENETIC INTERPRETATIONS OF COMPONENTS OF VARIANCE BETWEEN PROGENIES

The genetic constitution of σ^2_m and σ^2_{ml} of Experiment III will be derived as examples. Derivations for components of the other two experiments are given elsewhere (Comstock and Robinson, 1948). Initial assumptions will include only the following: regular diploid behavior at meiosis, no multiple allelism, no epistasis. Restrictions are not being placed on gene frequencies or linkage. To that extent the derivations to be given below are more general than those cited above which assumed absence of linkage and gene frequencies all equal to one-half.

The population sampled in Experiment III is outlined in Table 30.5. It consists of an infinity of pairs of backcross progencies, one pair for each variable parent that might be chosen from the F_2 or a later generation from crossing the two homozygous lines. Expected genetic values of each progeny are indicated symbolically. Because all progenies must be of finite size, there will

be a sampling deviation between the actual and expected values of the progenies. The variation among expected values is the variation due to genetic differences among parents, and hence that to be considered in evaluating σ_m^2 and σ_{ml}^2. σ_m^2 is the progeny variance due to genetic differences among the variable parents, *i.e.*, the variance of the pair means indicated in the next to the last column of Table 30.5. σ_{ml}^2, the progeny variance from interaction of genotypes of the variable and homozygous parents, is one-half the variance of the pair differences indicated in the final column.[2]

TABLE 30.5

POPULATION SAMPLED IN EXPERIMENT III

Variable Parent*	Homozygous Parent		Mean	Difference
	Line A	Line B		
1	X_{a1}	X_{b1}	$\bar{X}_1$	D_1
2	X_{a2}	X_{b2}	$\bar{X}_2$	D_2
3	X_{a3}	X_{b3}	$\bar{X}_3$	D_3
.				
.				
.				
s	X_{as}	X_{bs}	$\bar{X}_s$	D_s

* The one chosen from F_2 or later generation of the cross between lines *A* and *B*.

s symbolizes an infinitely large number.

X's are expected genetic values of progenies, subscripts indicate parentage of individual progenies.

$\bar{X}_i$ = mean of X_{ai} and X_{bi} [where the subscript i identifies the variable parent, e.g., $\bar{X}_1 = (X_{a1} + X_{b1})/2$].

$D_i = X_{ai} - X_{bi}$.

Now note that a pair mean or difference is the sum of contributions from individual loci. Let

x_{ij} be the contribution of the jth locus to the ith pair mean, and

d_{ij} be the contribution of the jth locus to the ith pair difference.

Then

$$\bar{X}_i = x_{i1} + x_{i2} + \ldots x_{iN}$$
$$D_i = d_{i1} + d_{i2} + \ldots d_{iN}$$

where N is the number of contributing loci. Then the variances of pair means and differences must be as follows:[3]

$$\sigma_{\bar{X}}^2 \left(= \sigma_m^2 \right) = \sum_j \sigma_{xj}^2 + 2 \sum_{j,k} \sigma_{xjk} \qquad (1)$$

$$\sigma_D^2 \left(= 2\sigma_{ml}^2 \right) = \sum_j \sigma_{dj}^2 + 2 \sum_{j,k} \sigma_{djk} \qquad (2)$$

2. It is well known and easily verified that in the analysis of variance of any $2 \times s$ table, the interaction variance is one-half that of the pair differences.

3. Since the variance of the sum of any number of variables is the sum of the variances of those variables plus twice the sum of all covariances among them.

where σ^2_{xj} is the variance of contributions from the jth locus to pair means,

σ_{xjk} is the covariance of contributions from the jth and kth loci to pair means,

σ^2_{dj} is the variance of contributions from the jth locus to pair differences,

σ^2_{djk} is the covariance of contributions from the jth and kth loci to pair differences, and $\sum_{j,k}$ indicates summation over all pairs of loci.

From equations (1) and (2) it is apparent that general expressions for σ^2_m and σ^2_{ml} can be written as soon as we know (i) the variance of contributions of any single locus to pair means and differences, and (ii) the covariance of contributions of any two loci to pair means and differences.

With respect to two loci, there will be ten types of variable parent (when classification is by types of gametes giving rise to the parent plant). Table 30.6 lists these, together with their frequencies in the population and the

TABLE 30.6

FREQUENCIES OF VARIABLE PARENT TYPES AND CONTRIBUTIONS* OF INDIVIDUAL LOCI TO EXPECTED GENETIC VALUES OF PROGENIES

Variable Parent	Freq.†	Homozygous Line				Mean		Difference	
		B_1b_2/B_1b_2		b_1B_2/b_1B_2					
		1st locus	2d locus	1st locus	2d locus	x_1	x_2	d_1	d_2
B_1B_2/B_1B_2	p^2	u_1	a_2u_2	a_1u_1	u_2	$(u_1+a_1u_1)/2$	$(u_2+a_2u_2)/2$	$u_1-a_1u_1$	$a_2u_2-u_2$
B_1B_2/B_1b_2	$2pr$	u‡	$(au-u)/2$	au	$(u+au)/2$	$(u+au)/2$	$au/2$	$u-au$	$-u$
B_1b_2/B_1b_2	r^2	u	$-u$	au	au	$(u+au)/2$	$(au-u)/2$	$u-au$	$-u-au$
B_1B_2/b_1B_2	$2ps$	$(u+au)/2$	au	$(au-u)/2$	u	$au/2$	$(u+au)/2$	u	$au-u$
B_1B_2/b_1b_2	$2pt$	$(u+au)/2$	$(au-u)/2$	$(au-u)/2$	$(u+au)/2$	$au/2$	$au/2$	u	$-u$
B_1b_2/b_1B_2	$2rs$	$(u+au)/2$	$(au-u)/2$	$(au-u)/2$	$(u+au)/2$	$au/2$	$au/2$	u	$-u$
B_1b_2/b_1b_2	$2rt$	$(u+au)/2$	$-u$	$(au-u)/2$	au	$au/2$	$(au-u)/2$	u	$-u-au$
b_1B_2/b_1B_2	s^2	au	au	$-u$	u	$(au-u)/2$	$(u+au)/2$	$u+au$	$au-u$
b_1B_2/b_1b_2	$2st$	au	$(au-u)/2$	$-u$	$(u+au)/2$	$(au-u)/2$	$au/2$	$u+au$	$-u$
b_1b_2/b_1b_2	t^2	au	$-u$	$-u$	au	$(au-u)/2$	$(au-u)/2$	$u+au$	$-u-au$

* Coded by subtraction of $u_1 + z_1$ (or $u_2 + z_2$) where z is the contribution from the locus when homozygous for the b allele.

† On the basis that frequencies in which B_1B_2, B_1b_2, b_1B_2, and b_1b_2 gametes are produced in the generation preceding that used for variable parents are p, r, s, and t, respectively. $p + r + s + t = 1.0$.

‡ For ease of printing, subscripts to a and u are omitted in rows beyond the first. However, the subscript used in the first row of each column applies throughout the column.

contributions of the two loci to the expected genetic values of progenies and to pair means and differences. As is evident from genotypes indicated for the homozygous lines, the initial linkage phase assumed is repulsion. The required variances and covariances can be worked out directly from information in the table. For example, the variance of contributions from the 1st locus to pair means is

$$(p^2 + 2pr + r^2)(u_1 + a_1u_1)^2/4 + (2ps + 2pt + 2rs + 2rt)\,a_1^2u_1^2/4 + (s^2 + 2st + t^2)(a_1u_1 - u_1)^2/4 - (\Sigma x_1)^2$$

and the covariance of contributions to pair means from the 1st and 2d loci is

$$p^2(u_1+a_1u_1)(u_2+a_2u_2)/4+2pr(u_1+a_1u_1)(a_2u_2)/4+\ldots$$
$$\ldots+t^2(a_1u_1-u_1)(a_2u_2-u_2)/4-(\Sigma x_1)(\Sigma x_2).$$

The algebraic reductions are tedious, particularly for the covariances, and will not be written out. However, the final expressions, for both the repulsion and coupling phase, are listed in Table 30.7.

TABLE 30.7

VARIANCES AND COVARIANCES OF SINGLE LOCUS CONTRIBUTIONS TO PAIR MEANS AND DIFFERENCES

Item	Initial Linkage Phase	
	Coupling	Repulsion
σ^2_{x1}.....	$\frac{1}{2}(p+r)(s+t)u_1^2$	$\frac{1}{2}(p+r)(s+t)u_1^2$
σ^2_{x2}.....	$\frac{1}{2}(p+s)(r+t)u_2^2$	$\frac{1}{2}(p+s)(r+t)u_2^2$
σ_{x12}....	$\frac{1}{2}(pt-rs)u_1u_2$	$\frac{1}{2}(pt-rs)u_1u_2$
σ^2_{d1}.....	$2(p+r)(s+t)a_1^2u_1^2$	$2(p+r)(s+t)a_1^2u_1^2$
σ_{d2}.....	$2(p+s)(r+t)a_2^2u_2^2$	$2(p+s)(r+t)a_2^2u_2^2$
σ_{d12}....	$2(pt-rs)a_1u_1a_2u_2$	$2(rs-pt)a_1u_1a_2u_2$

Note now that if the frequencies of B_1 and B_2 (in the population from which the variable parents are taken) are symbolized as q_1 and q_2, then

$$p+r=q_1 \qquad s+t=1-q_1$$
$$p+s=q_2 \qquad r+t=1-q_2$$

and

$$\sigma^2_{x1}=\tfrac{1}{2}q_1(1-q_1)u_1^2$$
$$\sigma^2_{x2}=\tfrac{1}{2}q_2(1-q_2)u_2^2$$

In general

$$\sigma^2_{xj}=\tfrac{1}{2}q_j(1-q_j)u_j^2$$

and

$$\sigma^2_{dj}=2q_j(1-q_j)a_j^2u_j^2$$

Substituting in equations (1) and (2), we have

$$\sigma^2_m=\sigma^2_{\underline{X}}=\frac{1}{2}\sum_j q_j(1-q_j)u_j^2+\sum_{j,k}(pt-rs)_{jk}u_ju_k \tag{3}$$

and

$$\sigma^2_{ml}=\tfrac{1}{2}\sigma^2_D=\sum_j q_j(1-q_j)a_j^2u_j^2+2\sum_{j,k}^{c}(pt-rs)_{jk}a_ju_ja_ku_k$$
$$+2\sum_{j,k}^{r}(rs-pt)_{jk}a_ju_ja_ku_k \tag{4}$$

where $\sum_{j,k}^{c}$ indicates summation over all pairs of loci for which the initial linkage phase was coupling and $\sum_{j,k}^{r}$, summation over pairs for which the initial phase was repulsion.

When the associations between alleles at two loci are at equilibrium with respect to coupling and repulsion phases, either because the loci are not linked or because there has been sufficient opportunity for recombination

$$p = q_j q_k \qquad r = q_j (1 - q_k)$$

$$s = (1 - q_j)\, q_k \qquad t = (1 - q_j)(1 - q_k)$$

and $(pt - rs) = 0$. Thus assuming no linkages (3) and (4) reduce to

$$\sigma^2_m = \tfrac{1}{2}\Sigma q\,(1 - q)\,u^2$$

$$\sigma^2_{ml} = \sigma\, q\,(1 - q)\,a^2u^2$$

as indicated in the preceding section. If, in addition, gene frequencies at all segregating loci are assumed to be one-half, σ^2_m and σ^2_{ml} reduce to the values assigned them in Table 30.4.

If there are linkages and equilibrium has not been reached, $(pt - rs)$ will be negative if the initial phase was repulsion, positive if the initial phase was coupling. Thus covariances from repulsion and coupling linkages will tend to cancel in σ^2_m. In fact if one assumes that enough loci are involved so that the number of linked pairs must be high and that there is no reason why the closer linkages should be predominantly in one phase, one is tempted to conclude that the sum of covariance will not be very important in σ^2_m.

On the other hand the covariance term is always positive in σ^2_{ml}, being a function of $(pt - rs)$ for coupling and of $(rs - pt)$ for repulsion.[4] Thus presence of any linkage, regardless of whether the two phases are equally frequent, will cause σ^2_{ml} to be greater than $\Sigma q(1 - q)a^2u^2$, except in the improbable event that a for either or both members of pairs of linked loci is zero. And unless all linkages were in the coupling phase (in which case the ratio of σ^2_m to $\frac{1}{2}\Sigma q(1 - q)u^2$ would be the same as of σ^2_{ml} to $\Sigma q(1 - q)a^2u^2$ and hence the ratio of σ^2_m to σ^2_{ml} unaffected by the linkages) $\sigma^2_{ml}/2\sigma^2_m$ would overestimate $\Sigma q(1 - q)a^2u^2/\Sigma q(1 - q)u^2$ so long as equilibrium in linkage associations had not been attained through recombination. However, as stated in the preceding section, the linkage bias becomes progressively smaller as equilibrium is approached.

For purposes of illustration, consider application of the formulae in a simple hypothetical situation. Assume that Experiment III is applied as first described, with variable parents taken from an F_2, and that the quanti-

4. This assumes generality of dominance of the more favorable allele—that a will almost always be positive.

tative character to be studied is affected by seven pairs of genes that are distributed as follows in the parent lines:

$$\text{Line } A\text{—}B_1B_1\,b_2b_2\,b_3b_3\,b_4b_4B_5B_5B_6B_6\,b_7b_7$$

$$\text{Line } B\text{—}b_1b_1B_2B_2B_3B_3B_4B_4\,b_5b_5\,b_6b_6B_7B_7$$

with u's and a's having the following values:

	Locus						
	1	2	3	4	5	6	7
u.....	1	2	1	2	1	2	1
a.....	6	.6	.8	.8	.8	1.0	.8

Note that less than complete dominance has been assumed for every locus. Gene frequencies should be one-half in an F_2, so $\frac{1}{2}\Sigma q(1-q)u^2$ becomes $\frac{1}{8}\Sigma u^2$ and $\Sigma q(1-q)a^2u^2$ becomes $\frac{1}{4}\Sigma a^2u^2$. Substituting numerical values of u and a listed above, we obtain

$$\tfrac{1}{2}\Sigma q\,(1-q)\,u^2 = 2.0$$

and

$$\Sigma q\,(1-q)\,a^2u^2 = 2.57$$

Now assume the following recombination values for pairs of loci:

Pair	Recombination Value (v)
1 and 2	.3
3 and 4	.2
5 and 6	.1
5 and 7	.2
6 and 7	.1
All others	.5

Thus the seven loci fall in three groups that are either on three separate chromosomes or, if on the same chromosome, far enough apart to allow free recombination. In an F_2 the values of p, r, s, and t will depend on v, the recombination value, and the original linkage phase as follows:

	p	r	s	t
Coupling.....	$(1-v)/2$	$v/2$	$v/2$	$(1-v)/2$
Repulsion....	$v/2$	$(1-v)/2$	$(1-v)/2$	$v/2$

Hence $(pt - rs)$ takes the following values:

	v			
	.1	.2	.3	.5
Coupling.....	.20	.15	.10	.0
Repulsion.....	−.20	−.15	−.10	.0

Substituting these and the numerical values of the a's and u's, we find

Locus Pair	Linkage Phase	$(pt-rs)u_iu_k$	$(pt-rs)a_iu_ia_ku_k$	$(rs-pt)a_iu_ia_ku_k$
1 and 2	Repulsion	−.2		.0720
3 and 4	Coupling	.3	.1920	
5 and 6	Coupling	.4	.3200	
5 and 7	Repulsion	−.15		.0960
6 and 7	Repulsion	−.40		.3200
		−.05	.5120	.4880

With these three sums and the values found above for $\frac{1}{2}\Sigma q(1-q)u^2$ and $\Sigma q(1-q)a^2u^2$ we compute

$$\sigma^2_m = 2.00 + (-.05) = 1.95$$

and

$$\sigma^2_{ml} = 2.57 + 2(.512) + 2(.488) = 4.57$$

Thus, while

$$\bar{a}^2 = \frac{\Sigma a^2u^2}{\Sigma u^2} = \frac{2.57}{2(2.0)} = .64,$$

the experiment would estimate

$$\frac{\sigma^2_{ml}}{2\sigma^2_m} = \frac{4.57}{2(1.95)} = 1.17.$$

Put differently, the estimate of $\bar{a}^2$ provided by Experiment III would in this case have positive bias in the amount $1.17 - .64 = .53$.

The foregoing example is given only to clarify the meaning of the formulae, not to suggest the amount of bias that may actually be present in practice. The actual bias with any specific material would depend on the amount of linkage and the relative prevalence of coupling and repulsion phases. However, the bias can only be positive and may range from a negligible to a large amount depending on the prevalence of repulsion linkage. While such bias detracts from the described estimate as a criterion of average dominance at the locus level, it is worth emphasizing that it represents a pseudo-overdominance effect which if persistent (due to closeness of linkages responsible) has much the same significance for short-run breeding practice as true overdominance. If the bias declines fairly rapidly as opportunity is provided for recombination, Experiment III offers a means of measuring that decline and thereby gaining an idea of the extent to which apparent dominance stems from linkage relationships that are loose enough to allow a near approach to equilibrium of linkage phases within a moderate number of generations.

AMOUNT OF DATA REQUIRED

An exhaustive consideration of this problem would require more space than can be devoted to it here. Detailed discussion will therefore be limited to one specific question. Let P symbolize the probability of an estimate of $\bar{a}$

that is significantly[5] greater than one. The question to be considered is as follows: Assuming a particular value (>1.0) for $\bar{a}$ how much data is required if P is to be one-half? Procedure and the argument involved will be given in detail for Experiment III; only comparative results will be indicated for the other two.

If the values of σ_m^2 and σ_{ml}^2 listed in Table 30.4 are substituted into the expectations of M_{31} and M_{32} of Table 30.3, we have

$$E(M_{31}) = \sigma^2 + \frac{r}{4}\Sigma u^2$$

$$E(M_{32}) = \sigma^2 + \frac{r}{4}\Sigma a^2u^2$$

Note that when $\Sigma u^2 = \Sigma a^2u^2$, *i.e.*, when $\bar{a} = 1.0$, the two expectations are equal. But if $\bar{a} > 1.0$, which means $\Sigma a^2u^2 > \Sigma u^2$; then $E(M_{32}) > E(M_{31})$. Also, the estimate of $\bar{a}$ will exceed one only where $M_{32} > M_{31}$. It follows that a one-tailed test of the hypothesis that $E(M_{32}) - E(M_{31} \leqslant 0$ is also a test of the hypothesis that $\bar{a} \leqslant 0$. Since both mean squares are functions of random variables (fixed effects do not contribute to either of them) the variance ratio test, the F test, is applicable and P is equivalent to the probability that the test ratio, M_{32}/M_{31}, will exceed F_α, where α is the probability level of the test.

Let $E(M_{32})/E(M_{31}) = \phi$. If $\phi = 1.0$, M_{32}/M_{31} will be distributed in samples in the same manner as F, otherwise it will be distributed as ϕF, *i.e.*, M_{32}/M_{31} for any probability point in its distribution will be exactly ϕ times the value of F for the same point in the F distribution. Thus the probability of a sample value of M_{32}/M_{31} equal or greater than F_α is the same as that of a sample value of F equal or greater than F_α/ϕ. When degrees of freedom are equal for the two mean squares, as will always be true in Experiment III, the 50 per cent point of the F distribution is 1.0. Hence P will be one-half when the amount of data is that for which F_α (the lowest value of M_{32}/M_{31} to be considered significantly different from one) is equal to ϕ.

We now must know the magnitude of ϕ when $\bar{a}$ is not unity.

$$\phi = \frac{E(M_{32})}{E(M_{31})} = \frac{4\sigma^2 + r\Sigma a^2u^2}{4\sigma^2 + r\Sigma u^2}$$

It varies with r, the number of replications in the experiment; with the ratio of Σa^2u^2 to Σu^2 which is $\bar{a}^2$; and with the ratio of σ^2 to Σu^2. Let $c = \sigma^2/\Sigma u^2$. Then

$$\phi = \frac{4c + r\bar{a}^2}{4c + r}$$

Number of replications is subject to the will of the experimenter, but c and $\bar{a}$

5. In the statistical sense, that the probability of the observed or a larger estimate as a consequence of random sampling is small.

are not. The logical procedure is to compute ϕ for various combinations of values of r, c, and $\bar{a}$. This is tedious but very useful if the three items are varied over rational ranges. A set of values for ϕ is presented in Table 30.8. Choice of rational values for $\bar{a}$ presented no difficulty since, in this connection, we are not so much concerned with its actual value as with the smallest for which sufficient data to make $P = .50$ are not beyond the reach of the experimenter.

TABLE 30.8

VALUE OF ϕ FOR $r = 2$ AND VARYING VALUES OF c AND $\bar{a}$

Expt.	$\bar{a}$	c				
		.25	.50	1.00	2.00	4.00
III	1.2	1.29	1.22	1.15	1.09	1.05
	1.4	1.64	1.48	1.32	1.19	1.11
	1.6	2.04	1.78	1.52	1.31	1.17
	2.0	3.00	2.50	2.00	1.60	1.33
II	1.4	1.27	1.17	1.09	1.05	1.03
	1.6	1.44	1.27	1.15	1.08	1.04
	2.0	1.80	1.50	1.29	1.15	1.08
I	1.4	1.13	1.10	1.06	1.04	1.02
	1.6	1.21	1.15	1.10	1.06	1.03
	2.0	1.38	1.28	1.18	1.11	1.06

Appropriate values for c will vary with the experimental material. The range listed in the table was chosen for application to work with grain yield of corn. σ^2 is plot error variance which, judging from experience, will usually be between 50 and 160 when yield is measured in bushels per acre.[6] This corresponds to a range of about 10 to 18 per cent for the coefficient of variation if mean bushel yield is 70. Σu^2 is twice the additive genetic variance in the F_2 population used. Robinson *et al.* (1949) worked with three F_2 populations and reported .0056 as an estimate of the average amount of additive genetic variance where yield was measured as pounds per plant. Converted to bushels per acre this figure becomes 78.4. More recent work at the North Carolina Experiment Station has yielded estimates of the same order of magnitude. From these results it appears that additive genetic variance will in many cases be between 20 and 100 and hence that Σu^2 will be between 40 and 200. The extreme values for c, if σ^2 and Σu^2 are within ranges suggested above,[7] are $50/200 = .25$ and $160/40 = 4.0$.

6. In work at the North Carolina station it has been quite close to 50.

7. Note that the suggested range for σ^2 is off-center upwards and that for Σu^2 is off-center downwards with respect to estimates from North Carolina data. This was done deliberately in an effort to be on the safe side. Efficiency of the experiment suffers from large σ^2 or small Σu^2.

All values of ϕ listed in the table are for $r = 2$. However, the effect of multiplying r by any constant is the same as dividing c by the same constant. Hence, ϕ for $c = 1$ and $r = 8$ is the same as for $c = .25$ and $r = 2$; ϕ for $c = 4$ and $r = 4$ is the same as for $c = 2$ and $r = 2$; etc.

Table 30.9 lists the approximate degrees of freedom required for M_{31} and M_{32} if $F_{.05}$ is to equal ϕ so that P will be .50. As an example to clarify the significance of this table, assume that $c = 1.0$, $\bar{a} = 1.4$, and $r = 2$. Then if the data provide 142 degrees of freedom for both M_{31} and M_{32}, the probabili-

TABLE 30.9

APPROXIMATE DEGREES OF FREEDOM* REQUIRED TO MAKE $P = .50$ IN EXPERIMENT III

$\bar{a}$	c				
	.25	.50	1.00	2.00	4.00
1.2	168	275	555	1460	4550
1.4	45	72	142	360	995
1.6	23	34	63	150	450
2.0	10	14	24	50	134

* Obtained assuming normal distribution of Fisher's z and employing the facts that $\sigma_z^2 = \frac{1}{2}(1/f_1 + 1/f_2)$ (where f_1 and f_2 are degrees of freedom for the two mean squares) and $F = e^{2z}$.

ty of the estimate of $\bar{a}$ being significantly greater (at the 5 per cent point) than one is one-half. Degrees of freedom can be related to amount of data as follows. Suppose that n, the number of progeny pairs per set, is 8. Then degrees of freedom will be 7/8 the number of progeny pairs, and assuming two replications, $r = 2$, degrees of freedom will be 7/32 the number of plots in the experiment. The 142 degrees of freedom indicated in the specific instance singled out above would require data on a total of about 650 plots.

An obvious question is whether increasing replications is as effective as increasing the number of progeny pairs. Consider the case where $c = 4.0$ and $\bar{a} = 1.6$. Degrees of freedom required are 450 when $r = 2$. But remembering that multiplying r by a constant has the same effect on ϕ as division of c by the same constant, we see that with four replications degrees of freedom required would be only 150. Thus with two replications a total of about 2056 plots would be required, whereas with four replications only about 1370 would be needed. The same is not true for the entire area of the table. Careful inspection will show that when c is 1.00 or less, doubling the number of progeny pairs is more effective than increasing replications from two to four. But when c is 2.0 or greater, the opposite is true.

Also pertinent are (1) the effect on P of increasing data above amounts indicated in Table 30.9, and (2) the probability of an estimate of $\bar{a}$ that is less

than one even though the true value exceeds one. P becomes about .75 if the data are doubled, between .85 and .9 if the data are tripled, and about .95 if the data are quadrupled. With the degrees of freedom indicated, the probability of an estimate less than 1.0 for $\bar{a}$ is in all cases close to .05, and that of an estimate significantly less than one is much smaller. This is an important point since it means a very small chance of erroneously concluding that $\bar{a}$ is less than one if its real value is greater than one by any very important amount.

The general point to note is that the amounts of data indicated in Table 30.9 are moderate for any combination of $c \leqslant 1.0$ and $\bar{a} \geqslant 1.4$. In addition, it is not prohibitive when both $\bar{a}$ and c are (within the ranges considered) either large or small. Actually, as indicated by earlier references, estimates of c for corn yield from data collected to date at the North Carolina Experiment Station have been somewhat less than .50.

An exact F test of the hypothesis that $\bar{a} \leqslant 1.0$ is not provided in the variance analysis of either Experiment I or II. In both instances there is a function (R) of three mean squares that provides an approximate F test. They are given below. Remember for Experiment II that we are assuming $m = n$

Experiment	R
I	$R_1 = (2n+3)M_{12}/(3M_{11}+2nM_{13})$
II	$R_2 = (2n+1)M_{23}/(M_{20}+2nM_{24})$

and using M_{20} to symbolize the mean of M_{21} and M_{22}. As was true for the test ratio of Experiment III, the expectations of numerator and denominator are equal in both of these ratios when $\bar{a} = 1.0$, but when $\bar{a} > 1.0$ the expectation of the numerator exceeds that of the denominator. Also, the estimate of $\bar{a}$ is greater than one only when the test ratio is greater than one. Values of ϕ for Experiments I and II in Table 30.8 are the ratios of expectations of numerator and denominator in these test ratios. As suggested by relative sizes of ϕ for the three experiments, more data are required in Experiment II than in III, and still more are required in I. However, the degrees of freedom supplied are greater relative to numbers of plots used than in III so differences in data required cannot be judged properly in terms of the ϕ's.

The data requirement cannot be determined as accurately as for Experiment III, primarily because degrees of freedom that should properly be assigned to the denominators of the test ratios cannot be known exactly though they can be approximated by the method of Satterthwaite (1946). For the same reason, determination of the approximate data requirement is more time-consuming. Attention will therefore be confined to the three situations indicated below. Degrees of freedom for Experiment I refer to the mean square, M_{12}, and for Experiment II to M_{23}. In both cases, n was assumed to be 4.0. Thus in II, progenies per set would be 16 as was assumed for Experiment III. This would make degrees of freedom for M_{23} be 9/32 of the number

of plots, if $r = 2$. If male groups per progeny set are 4.0, in Experiment I, as in the work of Robinson *et al.*, there would also be 16 progenies per set and degrees of freedom for M_{12} would be 12/32 of the number of plots.

Experiment III is obviously the most powerful and I the least powerful of the three. In the three cases examined, the plot requirement for I is from ten to twelve times that of III. Experiment II is intermediate, requiring from two to four times the data needed in III. It may be of interest that in the

c	$\bar{a}$	Degrees of Freedom Required			Plots Required if r=2		
		Expt. I	II	III	I	II	III
.50	1.4	1525	315	72	4066	1120	329
1.00	2.0	440	120	24	1173	426	110
.25	1.6	480	60	23	1280	213	105

work reported by Robinson *et al.* (1949) in which Experiment I was used in studying corn yield there were about 500 degrees of freedom for M_{12}. The estimate of $\bar{a}$ was 1.64 and, by the approximate F test, was just significant at the 5 per cent point.

Before leaving the subject, it should be noted that the problem of data required has been dealt with under the original assumptions. If what have been called estimates of $\bar{a}$ are biased upward by linkage or epistasis, their expected values are larger than $\bar{a}$, and the foregoing has relevance to the expected values of the estimates rather than to $\bar{a}$ itself. To exemplify, suppose that $\bar{a}$ were 1.2, but as a result of bias from epistasis and linkage the expected value of the Experiment III estimate were 1.2. Then assuming $c = .25$ and $r = 2$, the probability of the estimate being significantly above 1.0 would be .50 if the data furnished 168 degrees of freedom (Table 30.9), the same number required if $\bar{a}$ were 1.0 and the estimate unbiased. Thus, we see that the probability of an estimate significantly greater than one is a function of the expected value of the estimate rather than of $\bar{a}$ when the two are not equal. The corollary, that an estimate (obtained as described) significantly greater than one is not final proof of overdominance at the locus level, has been indicated in preceding sections.

CONCLUDING REMARKS

To attempt a general discussion of what has been presented appears unwise. It would almost certainly lead to some unnecessary repetition and could do more to confuse than to clarify. However, certain comments seem in order.

With regard to the experiments themselves, III appears definitely the most useful (1) because it is the most powerful, and (2) because it can be employed to learn something about the effect of linkage on the estimate of $\bar{a}$.

It should not be necessary to comment on the role of statistics in the devising and evaluation of schemes for investigating the inheritance of quantitative characters. If the importance of statistics in this area of research has not been adequately demonstrated by the foregoing, general statements could hardly be expected to be convincing. The point to be emphasized is that *more* theoretical investigation of experimental technics in quantitative inheritance is badly needed. For example, insofar as the three experiments considered here are concerned, more information is needed on the biases resulting from various sorts of epistasis. It is possible that such biases are greatest in Experiment III and would detract from its apparent superiority. It is also possible that the biases from epistasis differ between the experiments and that the differences vary with type of epistasis. In that event, comparison of results from two or more of the experiments could conceivably contribute to our knowledge of epistasis.

Investigation of the power of a variety of technics used in quantitative genetic research also would be fruitful. The intent is not to imply that there are no such procedures for which the power is known within satisfactory limits, but only to point out that there are some for which this is not the case. For example, mention has been made of the use of parent, F_1, F_2, and backcross means for investigating epistasis, but to the authors' knowledge there is nothing in the literature concerning amount of data required to insure that the chances of erroneous conclusions from such a study would be small.

Equally important is continued search for useful technics and procedures. It is entirely possible that approaches may thereby be discovered which are more efficient than any presently known. As a case in point, at the time the work described by Robinson *et al.* (1949) was planned we had not thought of the procedure designated here as Experiment III which, so far as we know, has not been previously described as a technic for investigation of dominance. Judging from findings of the preceding section, the same amount of work using the latter procedure would have provided considerably more precise estimates.

While attention herein has been devoted to estimation of average level of dominance, the experiments described provide other information as well. The data collected can be used also for estimation of additive genetic variance, variance due to dominance deviations, and the genetic and phenotypic covariances and correlations of pairs of characters.

LITERATURE

No attempt has been made to cite all of the various publications that in one way or another were stimulatory to the above discussion, since a careful attempt to assign credit where due would have made the manuscript considerably longer. Most interested readers will be familiar with relevant litera-

ture, but examples will be given here of papers that might have been cited.

The utilization of genetic variance component estimates is illustrated by numerous publications, for example, Baker *et al.* (1943). The composition (in terms of additive genetic variance and variance due to dominance deviations) of the estimable genetic variance components in the sort of population on which Experiment I is based is known generally and is indicated by Lush *et al.* (1948).

An experiment very similar to II but not designed with as specific information about dominance as its objective has been reported by Hazel and Lamoreux (1947).

The general pattern for genetic interpretation of variance components arising from Mendelian segregation was set in such papers as those by Fisher (1918), Fisher *et al.* (1932), and Wright (1935).

Other procedures for estimation of dominance have been described by Fisher *et al.* (1932), Mather (1949), and Hull (in this volume).