# THE COMPONENTS OF GENETIC VARIANCE IN POPULATIONS OF BIPARENTAL PROGENIES AND THEIR USE IN ESTIMATING THE AVERAGE DEGREE OF DOMINANCE*

R. E. COMSTOCK AND H. F. ROBINSON†

THE phenotypic expression of a character can be considered the sum of a genetic effect and a deviation attributable to environment and interaction between the genotype and environment involved. Symbolically,

$$p = y + e$$

where $p$ is the phenotype; $y$, the genetic effect; and $e$, the deviation of $p$ from $y$. Then, if genotypes are randomly distributed relative to variations in environment, phenotypic variance is

(1) $$\sigma_p^2 = \sigma_y^2 + \sigma_e^2$$

where $\sigma_y^2$ is variance of genetic effects, and $\sigma_e^2$ is the portion of $\sigma_p^2$ resulting from variation in environment.

Wright [9] defined three components of $\sigma_y^2$ as follows:

1. Additive genetic variance
2. Variance due to dominance deviations from the additive scheme
3. Variance due to epistatic deviations from the additive scheme.

These components will be symbolized in what follows as $\sigma_v^2$, $\sigma_d^2$, and $\sigma_i^2$, respectively. Assuming that either of two allelic genes ($B$ and $b$) may occupy a given locus, the genotype of a diploid organism with respect to that locus may be $BB$, $Bb$, or $bb$. In the absence of dominance the effect of the heterozygous genotype is the average of the effects of the other two genotypes, i.e.

$$y_{Bb} = (y_{BB} + y_{bb})/2 \quad \text{and} \quad y_{BB} - y_{Bb} = y_{Bb} - y_{bb}\,.$$

If $$y_{BB} - y_{Bb} \neq y_{Bb} - y_{bb}\,,$$

there is dominance and there will be variance due to dominance deviations.

*Contribution from Institute of Statistics, of The University of North Carolina, Raleigh, and the North Carolina Agricultural Experiment Station. Journal paper No. 297.

†Professor and Associate professor, respectively.





The term, epistasis, as used by Wright and others following his usage, covers all types of interactions among non-allelic genes. If a character is conditioned by genes at $n$ loci and there is no epistasis,

$$y = y_1 + y_2 + \cdots y_n$$

where $y_1$ is the effect of the genotype at the first locus, $y_2$ is the effect of the genotype at the second locus, etc. Then $\sigma_i^2 = 0$, i.e., there is no epistasis. If, in addition, there are no linkages among the $n$ loci or if there are linkages but the distribution of genotypes is at equilibrium, then

$$\sigma_v^2 = \sum_1^n \sigma_{v_i}^2\,, \qquad \sigma_d^2 = \sum_1^n \sigma_{d_i}^2\,, \qquad \text{and} \qquad \sigma_y^2 = \sum_1^n \sigma_{y_i}^2 = \sigma_v^2 + \sigma_d^2$$

where $\sigma_{v_i}^2$ is the additive genetic variance and $\sigma_{d_i}^2$, the variance due to dominance deviations resulting from segregation at the $i$th locus. This is the genetic model (no epistasis and either no linkage or the equilibrium distribution of genotypes with respect to linked loci) to be considered herein.

Though Wright defined them in slightly different terms, the additive genetic variance for the $i$th locus may be defined as the portion of the variance of genetic effects explained by regression on the number of $B$ genes in the genotype (or, alternatively, the number of $b$ genes) and the variance due to dominance deviations as the variance of deviations of the genetic effects from that regression. Let $x$ be the number of $B$'s in the genotype, $q$ be the frequency of $B$ in the population, and $(1 - q)$ the frequency of $b$ in the population. Then, for a population under random mating, the distribution of genotypes for a single locus will be as tabulated below.

| Genotype | Frequency | $x$ | $y$ | $y'$ |
|---|---|---|---|---|
| $BB$ | $q^2$ | 2 | $z + 2u$ | $u$ |
| $Bb$ | $2q(1 - q)$ | 1 | $z + u + au$ | $au$ |
| $bb$ | $(1 - q)^2$ | 0 | $z$ | $-u$ |

The symbols, $u$ and $a$ have the significance of $d$ and $h/d$, respectively, in the notation of Fisher *et al.* [3]. The $y'$ values are coded $y$ values obtained by subtracting $(z + u)$ from each. Note that $a$ is a measure of dominance; it equals zero when dominance is absent and increases in magnitude as $y_{Bb}$ deviates from the midpoint between $y_{BB}$ and $y_{bb}$.

Working from the above table we find that

$$\sigma^2_{y_i} = 2q(1-q)[1 + 2(1-2q)a + (1 - 2q + 2q^2)a^2]u^2 \tag{2}$$

$$\mathrm{Cov}_{xy_i} = 2q(1-q)[1 + (1-2q)a]u$$

$$\sigma^2_x = 2q(1-q)$$

$$\sigma^2_{g_i} = \frac{[\mathrm{Cov}_{xy_i}]^2}{\sigma^2_x} = 2q(1-q)[1 + 2(1-2q)a + (1 - 4q + 4q^2)a^2]u^2 \tag{3}$$

and finally

$$\sigma^2_{d_i} = \sigma^2_{y_i} - \sigma^2_{g_i} = 4q^2(1-q)^2a^2u^2 \tag{4}$$

Expressions for $\sigma^2_{g_i}$ and $\sigma^2_{d_i}$ are equivalent to those derived by Wright [9].

*Estimation of Average a Based on the Composition of Variance in Populations of Biparental Progenies*

Let $mn$ females be taken from a population produced by random mating and let them be mated, $n$ to each of $m$ males taken from the same

TABLE 1

ANALYSIS OF VARIANCE OF A POPULATION OF BIPARENTAL PROGENIES

| *Source of variance* | *d.f.* | *m.s.* | *Expectation of m.s.* |
|---|---|---|---|
| Males | $m - 1$ | $M_1$ | $\sigma^2 + k\sigma^2_p + rk\sigma^2_f + nrk\sigma^2_m$ |
| Females in males | $m(n-1)$ | $M_2$ | $\sigma^2 + k\sigma^2_p + rk\sigma^2_f$ |
| Plots in females in males | $mn(r-1)$ | $M_3$ | $\sigma^2 + k\sigma^2_p$ |
| Within plots | $mnr(k-1)$ | $M_4$ | $\sigma^2$ |
| Total | $mnrk - 1$ | | |

KEY TO TABLE 1

$\sigma^2$ is the sum of the intra "plot" environmental variance and the genetic variance among individuals of the same progeny.
$\sigma^2_p$ is the variance of "plot" effects.
$\sigma^2_f$ is the variance of female effects, and
$\sigma^2_m$ is the variance of male effects



population. Assume random choice of all individuals involved in the $mn$ matings. Let $rk$ offspring from each mating be grown, $k$ in each of $r$ "plots" assigned at random. The variance among the individuals of the $mn$ matings can be partitioned as indicated in Table 1. The expectations of the mean squares (See Crump [2]) are also indicated.

The four variance components can all be estimated from appropriate mean squares. For example,

$$(M_1 - M_2)/rkn \sim \sigma^2_m .$$

Under the hypothesis of random mating the expected frequencies of various types of matings, relative to a single pair of allelic genes, will be as listed below. Expected mean $y'$ values of progeny are listed for each mating.

| Mating | | Frequency | mean $y'$ |
|---|---|---|---|
| Male | Female | | |
| *BB* | *BB* | $q^4$ | $u$ |
| | *Bb* | $2q^3(1-q)$ | $(u + au)/2$ |
| | *bb* | $q^2(1-q)^2$ | $au$ |
| *Bb* | *BB* | $2q^3(1-q)$ | $(u + au)/2$ |
| | *Bb* | $4q^2(1-q)^2$ | $au/2$ |
| | *bb* | $2q(1-q)^3$ | $(au - u)/2$ |
| *bb* | *BB* | $q^2(1-q)^2$ | $au$ |
| | *Bb* | $2q(1-q)^3$ | $(au - u)/2$ |
| | *bb* | $(1-q)^4$ | $-u$ |

From this the frequencies and expected mean $y'$ values of progeny for the three types of males are found to be as follows:

| Male | Frequency | mean $y'$ |
|---|---|---|
| *BB* | $q^2$ | $qu + (1-q)au$ |
| *Bb* | $2q(1-q)$ | $\frac{(2q-1)}{2}u + \frac{au}{2}$ |
| *bb* | $(1-q)^2$ | $qau - (1-q)u$ |

The grand mean of $y'$ for all progenies is

$$(2q - 1)u + 2q(1 - q)au$$

The expected genetic variance (from the pair of alleles under consideration) among means of the progenies of different males is

$$q^2[qu + (1 - q)au]^2 + 2q(1 - q)\left[\frac{(2q - 1)}{2}u + \frac{au}{2}\right]^2$$

$$+ (1 - q)^2[qau - (1 - q)u]^2 - [(2q - 1)u + 2q(1 - q)au]^2$$

$$= \frac{q(1 - q)}{2}[1 + 2(1 - 2q)a + (1 - 4q + 4q^2)a^2]u^2$$

which by reference to equation (3) is seen to be equal to $\sigma^2_{\sigma i}/4$. Since the genetic model assumed (a) no epistasis and (b) the equilibrium state relative to distribution of linked genes, i.e., no correlation among effects of genotypes at different loci, the total genetic variance among male progeny means is

$$\frac{1}{4}\sum_1^n \sigma^2_{\sigma i} = \sigma^2_\sigma/4 \tag{5}$$

The expected genetic variance of means of progenies from different females but the same male is

$$q^4u + 2q^3(1 - q)\left(\frac{u + au}{2}\right)^2 \cdots (1 - q)^4(-u)^2$$

$$- q^2[qu + (1 - q)au]^2 - 2q(1 - q)\left[\frac{2q - 1}{2}u + \frac{au}{2}\right]^2$$

$$- (1 - q)^2[qau - (1 - q)u]^2$$

$$= \frac{q(1 - q)}{2}[1 + 2(1 - 2q)a + (1 - 2q + 2q^2)a^2]u^2$$

which by reference to equation (2) is seen to be equal to $\sigma^2_{\sigma i}/4$ or $\sigma^2_{\sigma i}/4 + \sigma^2_{di}/4$. Thus the total genetic variance among progenies of females within males is

$$\frac{1}{4}\sum_1^n \sigma^2_{\sigma i} + \frac{1}{4}\sum_1^n \sigma^2_{di} = \frac{1}{4}(\sigma^2_\sigma + \sigma^2_d) \tag{6}$$



If there is no maternal influence and if the progenies have been assigned to blocks at random, $\sigma^2_m$ and $\sigma^2_f$ will contain only the genetic variance among male progeny means and the genetic variance among the means of progenies of females within males, respectively. Then from (5) and (6)

$$\sigma^2_m = \sigma^2_\sigma/4 \tag{7}$$

and

$$\sigma^2_f = \tfrac{1}{4}\sigma^2_\sigma + \tfrac{1}{4}\sigma^2_d \tag{8}$$

The sum of $\sigma^2_m$ and $\sigma^2_f$ is the total of the additive genetic variance and variance due to dominance deviations among full sib families, $\frac{1}{2}\sigma^2_\sigma + \frac{1}{4}\sigma^2_d$, as reported by Wright [9].

Provided that $q$ has the same value for all gene pairs, estimates of $\sigma^2_d$ and $\sigma^2_\sigma$ furnish information on the magnitude of $a$. If the population of progenies described above were in the $F_3$ generation of a cross between two isogenic lines, it could be assumed that $q = .5$ for all loci at which there was segregation. From (3) and (4) when $q = .5$

$$\sigma^2_{\sigma i} = \tfrac{1}{2}u_i^2 \qquad \text{and} \qquad \sigma^2_{di} = \tfrac{1}{4}a_i^2u_i^2$$

Then

$$\sigma^2_\sigma = \tfrac{1}{2}\sum u_i^2, \qquad \sigma^2_d = \tfrac{1}{4}\sum a_i^2u_i^2,$$

and

$$\left(\frac{2\sigma^2_d}{\sigma^2_\sigma}\right)^{1/2} = \left(\frac{a_1^2u_1^2 + a_2^2u_2^2 + \cdots a_n^2u_n^2}{u_1^2 + u_2^2 + \cdots u_n^2}\right)^{1/2} = (\bar{a}^2)^{1/2} \tag{9}$$

where $\bar{a}^2$ is a weighted mean of the $a^2$'s for all loci, the individual $a^2$'s being weighted relative to the $u^2$'s for the corresponding loci. In what follows $(\bar{a}^2)^{1/2}$ will be symbolized as "$a$".

From (7) and (8), we have

$$\left(\frac{2(\sigma^2_f - \sigma^2_m)}{\sigma^2_m}\right)^{1/2} = \left(\frac{2\sigma^2_d}{\sigma^2_\sigma}\right)^{1/2} = \text{"}a\text{"} \tag{10}$$

Thus given data on the sort of population described above, estimates of $\sigma^2_f$ and $\sigma^2_m$ can be made and these used to estimate "$a$".

While "$a$" $> \bar{a}$ unless all $a$'s are equal, the bias in "$a$" as an estimate

of $\bar{a}$ cannot be very great unless the $a$'s vary a great deal[1] and "$a$" cannot exceed unity unless one or more of the $a$'s are larger than one. Therefore if the estimate of "$a$" is significantly greater than one, it can be concluded that one or more $a$ is greater than one, i.e., that there is overdominance of genes at one or more locus.

A slightly different sort of population might be set up if the organism is one in which a female parent can produce progenies by more than one male. Suppose that from each of $n$ females, progenies are obtained by each of $m$ male parents. (This is possible, for example, with multiflowered plants.) The result would be a set of $nm$ progenies, one from the mating of each of the $m$ males with each of the $n$ females. Assume $s$ such sets and that $k$ members of each progeny are grown in each of $r$ "plots". The form for the analysis of variance data so obtained is shown in Table 2.

TABLE 2

ANALYSIS OF VARIANCE FOR DATA OBTAINED FROM PROGENIES PRODUCED BY MATING EACH OF A SERIES OF MALES TO EACH OF A SERIES OF FEMALES

| *Source of variance* | *d.f.* | *m.s.* | *Expectation of mean square* |
|---|---|---|---|
| Sets of progenies | $s - 1$ | | |
| Males in sets | $s(m - 1)$ | $M_0$ | $\sigma^2 + k\sigma_p^2 + rk\sigma_{fm}^2 + rkn\sigma_m^2$ |
| Females in sets | $s(n - 1)$ | $M_1$ | $\sigma^2 + k\sigma_p^2 + rk\sigma_{fm}^2 + rkm\sigma_f^2$ |
| Males × females in sets | $s(m - 1)(n - 1)$ | $M_2$ | $\sigma^2 + k\sigma_p^2 + rk\sigma_{fm}^2$ |
| Plots in progenies | $smn(r - 1)$ | $M_3$ | $\sigma^2 + k\sigma_p^2$ |
| Within plots | $smnr(k - 1)$ | $M_4$ | $\sigma^2$ |
| Total | $smnrk - 1$ | | |

$\sigma^2_{fm}$ is the variance of effects due to interactions among males and females
Other symbols have the same significance as in Table 1.

[1]The bias would become very large if $a$ were positive for some loci and negative for others. Clearly it is the average absolute magnitude of $a$ that is being estimated since $a^2$ is positive whether $a$ is positive or negative.



Again assuming random choice of parents from the population available and random allotment of progenies to "plots", $\sigma_m^2$, $\sigma_f^2$, and $\sigma_{fm}^2$ will be entirely genetic in origin. Proceeding as before it can be shown that in this case

$$\sigma_m^2 = \sigma_f^2 = \sigma_a^2/4$$

and

$$\sigma_{fm}^2 = \sigma_d^2/4$$

Thus from (9) we have

$$\left(\frac{2\sigma_{fm}^2}{\sigma_m^2}\right)^{1/2} = \left(\frac{2\sigma_{fm}^2}{\sigma_f^2}\right)^{1/2} = \left(\frac{2\sigma_d^2}{\sigma_a^2}\right)^{1/2} = \text{"}a\text{"} \tag{11}$$

Given the same total number of progenies, $\sigma_a^2$ and $\sigma_d^2$, and hence "$a$", will be estimated more precisely by data from this sort of population than by data from the first type of population described. The most obvious advantage is in the fact that in this case the estimate of $\sigma_d^2$ is a function of only $M_2$ and $M_3$, both of which will have relatively many degrees of freedom, whereas in the other case it is a function of $M_1$, $M_2$, and $M_3$.

*Numerical Example*

The data to be used is taken from a study with corn reported by Robinson *et al.* [7]. Parent plants were selected at random from the $F_2$ generation of crosses between long inbred (essentially isogenic) lines. Individual plants used as pollen parents (males) were each mated to four seed-producing plants (females). Data on progenies produced in one of the crosses will suffice to exemplify the estimation procedure.

The analysis of variance in Table 3 is on yield of grain of 192 progenies produced from matings of 48 males with 192 females. The 192 progenies are comprised of 48 male groups; the four progenies of a group had a common male parent but each progeny had a different female parent. The field lay-out was in blocks containing 32 plots each. The material was divided into 12 sets of 4 male groups each and each such set of 16 progenies was assigned to a block and replicated twice within the block. A new randomization of the 16 progenies was made within each of the two replications. Each plot consisted of two 10 plant rows. Yield was measured in pounds of grain produced by 10 guarded plants (plants having another plant on each side in the row).

TABLE 3
ANALYSIS OF VARIANCE OF GRAIN YIELD OF 192 BIPARENTAL PROGENIES

| *Source of Variance* | *d.f.* | *m.s.* | *m.s. expectations (see Table 1)* |
|---|---|---|---|
| Blocks | 11 | .153 | |
| Replications in blocks | 12 | .063 | |
| Males in blocks | 36 | .167 $= M_1$ | $\sigma^2 + 10\sigma_p^2 + 20\sigma_f^2 + 80\sigma_m^2$ |
| Females in males in blocks | 144 | .069 $= M_2$ | $\sigma^2 + 10\sigma_p^2 + 20\sigma_f^2$ |
| Males × Replications in blocks | 36 | .031* $= M_3$ | $\sigma^2 + 10\sigma_p^2$ |
| Females in males × Replications in blocks | 144 | | |
| Within plots | 207 | .0153† | $\sigma^2$ |

*The two interaction mean squares were pooled for a single progeny x replication mean square. This is rather common procedure. The underlying assumption is that intra-block genotype x replication interaction was unimportant in magnitude. The point was supported by the fact that the Males x Replication mean square was actually the smaller.

†This estimate was obtained from data on only a portion of the individual plants. Individual plant data were taken only in about every 12th plot.

From the mean squares in Table 3 variance components are estimated as follows:

| *Variance Component* | *Estimate* |
|---|---|
| $\sigma_m^2$ | $(M_1 - M_2)/80 = .001225$ |
| $\sigma_f^2$ | $(M_2 - M_3)/20 = .0019$ |
| $\sigma_p^2$ | $(M_3 - M_4)/10 = .00157$ |
| $\sigma^2$ | $M_4 = .0153$ |



Then using the estimates of $\sigma_m^2$ and $\sigma_f^2$ in accordance with equation (10) we obtain

$$\left(\frac{2(.0019 - .001225)}{.001225}\right)^{1/2} = 1.05$$

as an estimate of "$a$".

The above data also furnish the material for estimating heritability. Heritability is the additive genetic variance as a fraction of the phenotypic variance. Hence, since $\sigma_m^2 = \frac{1}{4}\sigma_g^2$,

$$\frac{4(.001225)}{.001225 + .0019 + .00157 + .0153} = .245$$

is an estimate of the heritability of the variance among individual plants. Numerous estimates of the heritability of animal characteristics have been made in this way (for an example see Hetzer *et al.* [4]).[2]

*Significant Test for the Deviation of the Estimate of "a" from any Hypothetical Value*

An estimate of "$a$" will have its greatest value only if the probability that its deviation from critical hypothetical values is of chance origin can be established. If it is significantly greater than zero, a degree of dominance in the action of genes conditioning the character in question is indicated. If it is significantly greater than one, over-dominance is indicated.

An approximate $F$ test[3] can be made in the following manner. Considering the variance analysis for the first type of population discussed, the expected value of $M_2$ assuming a specific value of "$a$" can be estimated as a linear function of $M_1$ and $M_3$. From equation (10),

$$\sigma_m^2 = \frac{2\sigma_f^2}{2 + (\text{“}a\text{”})^2}$$

Then $M_2'$, the estimate of the expected value of $M_2$, is

$$(12) \qquad \left[\frac{2 + (\text{“}a\text{”})^2}{2 + (\text{“}a\text{”})^2 + 2n}\right]M_1 + \left[\frac{2n}{2 + (\text{“}a\text{”})^2 + 2n}\right]M_3$$

[2]Estimates made in this manner from data obtained at one location in a single year will overevaluate genetic improvement possible through selection if there are important interactions of genotype with the variations in environment that occur from year to year and between locations within the area in which an improved strain or variety is to be used. Such interactions are known to be important for many plants and estimates of their magnitudes are needed.

[3]Suggested by W. G. Cochran of the Institute of Statistics of The University of North Carolina, Raleigh. He states that the test outlined is reasonably good but that work directed toward finding a more precise one is in progress.

in which "$a$" is given the hypothetical value against which the estimated value is to be tested. The test is based on a comparison of $M_2'$ with the observed $M_2$. Three types of cases should be distinguished.

1. The object is to test whether the estimate of "$a$" is significantly *larger* than a specified hypothetical value. For example, in testing for over-dominance we wish to know if our estimate of "$a$" is significantly larger than 1.0. Since the expected value of $M_2$ increases as "$a$" becomes larger, $F$ will in this case be computed as $M_2/M_2'$.
2. The object is to test whether the estimate of "$a$" is significantly *smaller* than a specified hypothetical value. For example, if we want to establish that there are loci at which there is either no dominance or only partial dominance we are concerned with whether our estimate of "$a$" is significantly smaller than 1.0. Because smaller values of "$a$" mean smaller expected values of $M_2$, $F$ must in this case be computed as $M_2'/M_2$.
3. The object is to test whether the estimate of "$a$" deviates significantly (in either direction) from a specified hypothetical value. In this case $F$ is taken as $M_2'/M_2$ if $M_2' > M_2$ and as $M_2/M_2'$ if $M_2 > M_2'$.

In cases 1 and 2 the probability of $F$ is taken from the standard $F$ table in the usual way. In case 3, a two-tailed $F$ test is involved and hence the probability from the $F$ table must be doubled. Degrees of freedom are assigned $M_2'$ in the manner described by Satterthwaite [8]. However since quantities such as $M_2'$ are not distributed precisely like ordinary mean squares, the test is clearly not an exact one.

It may be noted that an alternative procedure involving comparison of $M_1$ with an estimate, $M_1'$, of its expected value based on $M_2$ and $M_3$ could be used. However, the expression for $M_1'$ would involve the *difference* between $M_2$ multiplied by a constant and $M_3$ multiplied by another constant, and it is known from theory that the distribution of a linear function of mean squares deviates further from that of a single mean square when the function involves differences than when it does not.

Using the data of table 3 in testing the divergence of the estimate, 1.05, of "$a$" from zero we find

$$M_2' = \left(\frac{2+0}{2+0+8}\right).167 + \left(\frac{8}{2+0+8}\right).031 = .0582$$

Degrees of freedom assigned $E(M_2)$ are

$$\frac{[c_1M_1 + c_3M_3]^2}{\frac{c_1^2M_1^2}{f_1} + \frac{c_3^2M_3^2}{f_3}} = 98$$



where $c_1$ and $c_3$ are the coefficients of $M_1$ and $M_3$ in (11) and $f_1$ and $f_3$ are the degrees of freedom of $M_1$ and $M_3$.

$F = 1.19$ which is non-significant for 144 and 98 degrees of freedom.

In the case of the analysis of variance in table 2 the test of "$a$" would be an $F$ test of $M_2$ against its expected value based on $M_0$, $M_1$, and $M_3$. If $m$ and $n$ were equal $M_2'$ would be as follows:

$$\left[\frac{(\text{“}a\text{”})^2}{(\text{“}a\text{”})^2 + 2n}\right]\left[\frac{M_0 + M_1}{2}\right] + \left[\frac{2n}{(\text{“}a\text{”})^2 + 2n}\right]M_3$$

When the deviation of "$a$" from zero is being tested this reduces to $M_3$ and the test becomes the exact test of $F = M_2/M_3$.

It should be noted that a rather large amount of data is required for a reasonably good estimate of "$a$". Robinson *et al.* [7] in the study referred to above had 146 and 518 degrees of freedom for $M_1$ and $M_2$, respectively, yet an estimate of 1.64 had a probability of about .05 assuming the true value of "$a$" was 1.0. If it had been possible to utilize the alternative type of population described and this had been done using sets of 16 progenies produced by mating each of 4 females with each of 4 males, the probability of "$a$" = 1.64 being greater than 1.0 would have been below .01 assuming that all variance estimates had turned out the same.

## DISCUSSION

The foregoing has assumed (a) no epistasis, and (b) equilibrium with respect to segregation of linked genes. In many, if not most, instances neither assumption will be strictly valid. It appears to the authors that the presence of genetic variance due to epistatic deviations will cause upward bias in the estimate of "$a$" because this variance is distributed among the mean squares in somewhat the same manner as that due to dominance deviations. The probable size of such bias is not clear at this time. It is known that the epistatic variance arising from some types of non-allelic gene interactions is small relative to $\sigma_d^2$ but that in certain genetic situations it can be large, Lush [6]. Further, in the case of at least some types of epistasis the fraction of the epistatic variance contained in $\sigma_f^2$ (or $\sigma_{fm}^2$ in the case of the second type of population discussed) is considerably smaller than that for $\sigma_d^2$. Further work is needed in evaluation of bias possible as a consequence of epistasis and in devising techniques for determining when a serious amount of epistatic variance is involved.

Coupling phase linkages would not be a source of bias in estimates of "$a$". On the other hand, the net effect of genes tightly linked in repulsion could be the same as for over-dominance in the action of independently segregating genes even though none of the linked genes were

individually more than partially dominant to their alleles. As a result it would appear that estimates of "$a$" > 1.0 might well be obtained from data collected in an early generation following a cross of genetically divergent material even though there were no over-dominance in the action of the individual gene pairs. In that event somewhat smaller estimates would be expected in later generations, though it is possible that this linkage effect would always be present to a degree. However, whether apparent over-dominance is an attribute of the individual gene pairs or a consequence of linkage it poses problems relative to methods for genetic improvement, Hull [5], and Comstock *et al.* [1]. One difference will be that, if what is measured as overdominance results largely from linkage, information on its magnitude must be interpreted in the light of the history of the material from which that information was obtained and applied with consideration to the history of the material on which one is attempting to work genetic improvement.

SUMMARY

Two sorts of populations of biparental progenies from which data can be used for the estimation of the degree of dominance in the action of genes have been described. The mathematical basis for the estimation methods have been presented in detail together with some discussion of limiting assumptions involved. A numerical example of the arithmetic procedures is given and an approximate test of significance of the estimate of dominance is outlined.

REFERENCES

[1] Comstock, R. E., H. F. Robinson, and P. H. Harvey. A Breeding Procedure Designed to Make Maximum Use of Both General and Specific Combining Ability. (In Press).
[2] Crump, S. Lee. The Estimation of Variance Components in Analysis of Variance. *Biometrics Bulletin, 2:* 7–11, 1946.
[3] Fisher, R. A., F. R. Immer and O. Tedin. The Genetical Interpretation of Statistics of the Third Degree in the Study of Quantitative Inheritance. *Genetics, 17:* 107–124, 1932.
[4] Hetzer, H. O., G. E. Dickerson, and J. H. Zeller. Heritability of Type in Poland China Swine Evaluated by Scoring. *Jour. Anim. Sci.: 3,* 390–398, 1944.
[5] Hull, Fred H. Recurrent Selection for Specific Combining Ability in Corn. *Jour. Am. Soc. Agron. 37:* 134–145, 1945.
[6] Lush, J. L. Animal Breeding Plans. The Collegiate Press, Inc., Ames, Iowa. 2nd Edition (see pp. 76–79) 1943.
[7] Robinson, H. F., R. E. Comstock, and P. H. Harvey. Estimates of Heritability and the Degree of Dominance in Corn. (In press).
[8] Satterthwaite, F. E. An Approximate Distribution of Estimates of Variance Components. *Biometrics Bulletin, 2:* 110–114, 1946.
[9] Wright, Sewall. The Analysis of Variance and the Correlations between Relatives with Respect to Deviations from an Optimum. *Jour. Genetics, 30:* 243–256, 1935.

# QUERIES

**61** QUERY: A problem was recently referred to me for criticism and it now seems that a further opinion is needed. I would appreciate your apparisal of the problem. Here is the problem as presented to me.

In a cattle feeding experiment, twelve rations were tried on 12 lots of animals in each of two replications. There were two animals per plot. The twelve rations consisted of all combinations of four winter rations and three summer rations. Table 1 is the analysis of variance presented to me.

TABLE 1
ANALYSIS OF VARIANCE SUMMARY

| Source of Variation | D/F | MS | F |
|---|---|---|---|
| Total | 47 | | |
| Btw. winter rations | 3 | 9,388.06 | 1.02 |
| Btw. summer rations | 2 | 3,308.77 | 2.88 |
| Btw. reps | 1 | 28,226.97 | 1.64 |
| W X S | 6 | 9,528.99 | 4.85* |
| W X Reps | 3 | 1,862.38 | 24.81* |
| S X Reps | 2 | 52,581.43 | 1.14 |
| W X S X Reps | 6 | 46,198.99 | 4.00** |
| Error | 24 | 11,558.92 | |

My opinion is that the experiment as set up has 24 plots (only 2 reps are accounted for in the summary above) and the analysis of variance summary should be:

| | |
|---|---|
| Total | 23 |
| Winter | 3 |
| Summer | 2 |
| Reps | 1 |
| W X S | 6 |
| W X R | 3 |
| S X R | 2 |
| W X S X R | 6 |

It seems to me that if all 48 animals were considered separate plots, then there are 4 reps and the summary of analysis of variance would be: