THEORETICAL AND EXPERIMENTAL STUDY OF SELF FERTILIZED POPULATIONS*

THEODORE W. HORNER
Agricultural Experiment Station, Ames, Iowa, U.S.A.
AND
CHARLES R. WEBER
Field Crops Research Branch, A.R.S., U. S. Department of Agriculture

1. INTRODUCTION

Plant breeders frequently classify the material of a generation into groups according to the ancestry of the group. Replication of the groups permits the variance component due to that type of grouping to be estimated. Groups of one generation may be related to groups of another generation through common ancestry and hence covariance can be estimated. Such estimated quantities are functions of the underlying genetic mechanism and in particular of genotypic variances and covariances. They thus provide a basis for inference about the genetic mechanism.

Genotypic variances and covariances have been defined and derived under various assumptions by numerous authors; notably Fisher [1932], Mather [1949], Horner [1955], Gates [1954] and Kempthorne [1956]. The definitions have been based on infinite populations. These may be divided into sub-populations whose genotypic composition and relative frequency are known from genetic theory. The actual magnitude of a genotypic variance or covariance depends on the value assigned to each genotype. This depends on the physiology of the expression of the genotype and is usually a function, simple or complicated, of the genes of the genotype.

The relationship of sample covariances (or sample variance components) to genotypic variances and covariances is not always obvious. The expected values of the sample covariances may be a function of more than one genotypic covariance. The relationship is often further obscured by the fact that actual plant populations are finite and a defined infinite subpopulation may be represented by only a few plants. The problem of relationship becomes particularly acute in populations which reproduce by self fertilization. These give rise to populations

---

*Journal paper No. 281, Iowa Agr. Exp. Sta. Project Nos. 1285 and 1179.

Journal Paper No. J-3014 of the Iowa Agricultural Experiment Station, Ames, Iowa. Project Nos. 1179 and 1285.

having a hierarchical structure and hence to a whole set of genotypic variances and covariances.

The objective in the present paper is to describe a formal procedure for obtaining the relationship of sample covariances and variance components to genotypic variances and covariances. This procedure is described with reference to an experiment conducted at the Iowa Agricultural Experiment Station on a soybean cross using data from the F_2 through the F_7 generation.

2. DESCRIPTION OF SOYBEAN EXPERIMENT

Homozygous soybean varieties Adams and Hawkeye were crossed in 1947. Their F_1 was produced in 1948, their F_2 in 1949, etc. For the purposes of this study, soybeans are nearly completely self fertilized. The experiment has the structure indicated in Table 1.

TABLE 1

STRUCTURE OF THE DATA

Year	Population	Structure							
1949	F_2	A_1		A_2	$\cdots$	A_i	$\cdots$	A_{94}	
1950	F_3	B_1		B_2		B_i		B_{94}	
1951	F_4	C_1	C_1^*	C_2	C_2^*	C_i	C_i^*	C_{94}	C_{94}^*
1952	F_5	D_1	D_1^*	D_2	D_2^*	D_i	D_i^*	D_{94}	D_{94}^*
1953	F_6	E_1	E_1^*	E_2	E_2^*	E_i	E_i^*	E_{94}	E_{94}^*
1954	F_7	F_1	F_1^*	F_2	F_2^*	F_i	F_i^*	F_{94}	F_{94}^*

Unselected progeny: *. Selected progeny: no star

Ninety four random F_2 plants, indicated by A's in Table 1, were selected. The 94 progenies in the F_3 generation, indicated by B's, were grown in a simple lattice design together with parents and bulk populations. Two plants were selected at random from each F_3 progeny. The 188 resulting F_4 progenies, which are indicated by C's in Table 1, were again grown in a simple lattice design. To obtain the F_5 generation one of the two progenies in the F_4 generation tracing to each particular F_2 individual was selected at random. Two random plants of each selected progeny produced F_5 progenies. The cycle was then repeated. Selected progenies are indicated in Table 1 by the absence of a star.

The letters B through F in addition to being used to represent progenies will also be used to represent adjusted phenotypic progeny means, and the letter A will also be used to represent a measured value of an F_2 plant.

TABLE 2

SAMPLE COVARIANCES AND VARIANCE COMPONENTS AND THEIR EXPECTED VALUES

Sample covariance			Expected value
A	with	B	$\mathrm{Cov}\,(2; 2, 3)$
A	"	$\frac{1}{2}(C + C^*)$	$\mathrm{Cov}\,(2; 2, 4)$
A	"	$\frac{1}{2}(D + D^*)$	$\mathrm{Cov}\,(2; 2, 5)$
A	"	$\frac{1}{2}(E + E^*)$	$\mathrm{Cov}\,(2; 2, 6)$
A	"	$\frac{1}{2}(F + F^*)$	$\mathrm{Cov}\,(2; 2, 7)$
B	"	$\frac{1}{2}(C + C^*)$	$\mathrm{Cov}\,(2; 3, 4) + a\,\mathrm{Cov}\,(3; 3, 4) - a\,\mathrm{Cov}\,(2; 3, 4)$††
B	"	$\frac{1}{2}(D + D^*)$	$\mathrm{Cov}\,(2; 3, 5) + a\,\mathrm{Cov}\,(3; 3, 5) - a\,\mathrm{Cov}\,(2; 3, 5)$
B	"	$\frac{1}{2}(E + E^*)$	$\mathrm{Cov}\,(2; 3, 6) + a\,\mathrm{Cov}\,(3; 3, 6) - a\,\mathrm{Cov}\,(2; 3, 6)$
B	"	$\frac{1}{2}(F + F^*)$	$\mathrm{Cov}\,(2; 3, 7) + a\,\mathrm{Cov}\,(3; 3, 7) - a\,\mathrm{Cov}\,(2; 3, 7)$
C	"	$\frac{1}{2}(D + D^*)$	$\mathrm{Cov}\,(3; 4, 5) + b\,\mathrm{Cov}\,(4; 4, 5) - b\,\mathrm{Cov}\,(3; 4, 5)$
C	"	$\frac{1}{2}(E + E^*)$	$\mathrm{Cov}\,(3; 4, 6) + b\,\mathrm{Cov}\,(4; 4, 6) - b\,\mathrm{Cov}\,(3; 4, 6)$
C	"	$\frac{1}{2}(F + F^*)$	$\mathrm{Cov}\,(3; 4, 7) + b\,\mathrm{Cov}\,(4; 4, 7) - b\,\mathrm{Cov}\,(3; 4, 7)$
D	"	$\frac{1}{2}(E + E^*)$	$\mathrm{Cov}\,(4; 5, 6) + c\,\mathrm{Cov}\,(5; 5, 6) - c\,\mathrm{Cov}\,(4; 5, 6)$
D	"	$\frac{1}{2}(F + F^*)$	$\mathrm{Cov}\,(4; 5, 7) + c\,\mathrm{Cov}\,(5; 5, 7) - c\,\mathrm{Cov}\,(4; 5, 7)$
E	"	$\frac{1}{2}(F + F^*)$	$\mathrm{Cov}\,(5; 6, 7) + d\,\mathrm{Cov}\,(6; 6, 7) - d\,\mathrm{Cov}\,(5; 6, 7)$
C^*	"	$\frac{1}{2}(D + D^*)$	$\mathrm{Cov}\,(2; 4, 5)$
C^*	"	$\frac{1}{2}(E + E^*)$	$\mathrm{Cov}\,(2; 4, 6)$
C^*	"	$\frac{1}{2}(F + F^*)$	$\mathrm{Cov}\,(2; 4, 7)$
D^*	"	$\frac{1}{2}(E + E^*)$	$\mathrm{Cov}\,(3; 5, 6)$
D^*	"	$\frac{1}{2}(F + F^*)$	$\mathrm{Cov}\,(3; 5, 7)$
E^*	"	$\frac{1}{2}(F + F^*)$	$\mathrm{Cov}\,(4; 6, 7)$

Year	Population	Sample variance component	Expected value
1950	F_3	Among A groups	$\mathrm{Cov}\,(2; 3, 3)$†
1951	F_4	Among B groups	$\mathrm{Cov}\,(2; 4, 4)$
1951	F_4	Within B groups	$\mathrm{Cov}\,(3; 4, 4) - \mathrm{Cov}\,(2; 4, 4)$†
1952	F_5	Among C groups	$\mathrm{Cov}\,(3; 5, 5)$
1952	F_5	Within C groups	$\mathrm{Cov}\,(4; 5, 5) - \mathrm{Cov}\,(3; 5, 5)$†
1953	F_6	Among D groups	$\mathrm{Cov}\,(4; 6, 6)$
1953	F_6	Within D groups	$\mathrm{Cov}\,(5; 6, 6) - \mathrm{Cov}\,(4; 6, 6)$†
1954	F_7	Among E groups	$\mathrm{Cov}\,(5; 7, 7)$
1954	F_7	Within E groups	$\mathrm{Cov}\,(6; 7, 7) - \mathrm{Cov}\,(5; 7, 7)$†

†Strictly speaking year × genotype interaction components should also be included in these expected values.

††a, b, c and d are the reciprocals of the harmonic means of the numbers of plants in B, C, D, and E respectively. The harmonic means are around 100 and hence the terms involving these coefficients are negligible.

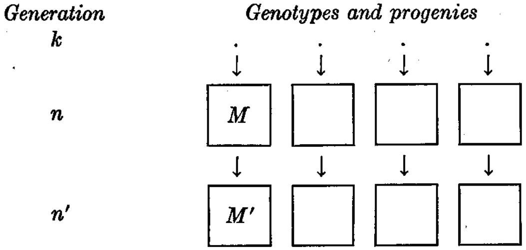
In this experiment there were 21 sample covariances and 9 variance components which were unconfounded with environmental effects and which could be given a genetic interpretation. These are listed in Table 2 along with their expected values.

3. INTERPRETATION OF GENOTYPIC VARIANCES AND COVARIANCES

The meaning of the symbolization Cov $(2; 2, 3)$ or in general Cov $(k; n, n')$ is indicated by Figure 1. Cov $(k; n, n')$ is the genotypic covariance of progenies in the nth generation from particular genotypes in the kth generation with progenies in the n'th generation from those same genotypes of the kth generation.

FIGURE 1

INTERPRETATION OF Cov $(k; n, n')$



M and M' are genotypic progeny means of progenies in the n and n' generations which are descended from genotypes in the *kth* generation.

$$\text{Cov}\ (k; n, n') = \text{Cov}\ (M, M')$$

All genotypic variances and covariances of populations having the hierarchical structure described can be expressed as linear functions of the Cov $(k; n, n')$. With this notation, the genotypic variance among F_2 plants is Cov $(2; 2, 2)$, and the covariance of F_2 plants with their progenies in the F_3 is Cov $(2; 2, 3)$. The variance of F_3 progeny means is Cov $(2; 3, 3)$ and the average variance within these progenies is [Cov $(3; 3, 3)$ − Cov $(2; 3, 3)$].

The interpretation of the Cov $(k; n, n')$ is a logical consequence of the genetic assumptions. As an example of interpretation, consider Cov $(3; 4, 5)$; that is, the genotypic covariance of progenies in the F_4 and F_5 generations which trace to particular genotypes in the F_3 generation. Letting p_t represent the frequency of the tth genotype in the F_3 and letting $y_{t.}$ and $y_{t..}$ represent the respective F_4 and F_5 progeny means, then Cov $(3; 4, 5) = \sum_t p_t(y_{t.} - y_{..})(y_{t..} - y_{...})$ where $y_{..}$ and $y_{...}$ are the genotypic means of the F_4 and F_5 generations. In this notation the generation with which a symbol is associated is indicated by the number of subscripts of the symbol. Each successive generation has one more subscript. Taking the case of a single locus with two alleles, let the values of the three genotypes be $z + 2u$, $z + u + au$ and z. From genetic theory it is known that the frequencies of the three genotypes in the F_3 are 3/8, 2/8 and 3/8. The genotypic progeny means in the F_4 generation are $z + 2u$, $z + u + 1/2\ au$ and z; and in the F_5 generation, $z + 2u$, $z + u + 1/4\ au$ and z. Thus, Cov (3; 4, 5) $=$ $3/4u^2 + 3/128a^2u^2 = (3/2)\sigma_A^2 + (3/32)\sigma_D^2$, where σ_A^2 and σ_D^2 are the additive genetic and dominance variances of the F_2 generation.

4. DERIVATION OF THE COVARIANCE OF C WITH $\frac{1}{2}(D + D^*)$

4.1 *Notation and model for an adjusted progeny mean:*

The covariance of C with $\frac{1}{2}(D + D^*)$ is derived in this section to illustrate the derivation of covariances and variance components. For this purpose C_i will be represented by

$$x_{i.} = \frac{\sum_j x_{ij}}{n_i}$$

where x_{ij} is the value of the jth plant in C_i and n_i is the number of plants in C_i. For D_i we will use

$$x_{ij'.} = \frac{\sum_k x_{ij'k}}{n_{ij'}}$$

where $x_{ij'k}$ is the value of the kth plant in D_i. The prime on the subscript j denotes the fact that the D_i progeny is from a particular plant in C_i, the j' plant say. The j' plant is one of the two plants which was selected at random in C_i. The other will be indicated by j''. For D_i^* we will use

$$x_{ij''.} = \frac{\sum_k x_{ij''k}}{n_{ij''}}.$$

The expected value of the sample covariance is

$$E\{\widehat{\text{Cov}}\ [C, \tfrac{1}{2}(D + D^*)]\}$$

$$= \frac{1}{2} E\left\{\frac{\sum_i (x_{i.} - x_{..})(x_{ij'.} + x_{ij''.} - x_{.j'.} - x_{.j''.})}{n - 1}\right\}$$

$$i = 1, 2, \cdots, n$$

which can be evaluated most easily by making use of the equivalent expression

$$E\{\widehat{\text{Cov}}\ [C, \tfrac{1}{2}(D + D^*)]\}$$

$$= \frac{1}{2n(n-1)} \sum_i \sum_{i'} E\{(x_{i.} - x_{i'.})[\tfrac{1}{2}(x_{ij'.} + x_{ij''.} - x_{i'j'.} - x_{i'j''.})]\}$$

$$= \frac{1}{2n(n-1)} \sum_{\substack{i \\ i' \neq i}} \sum_{i'} E\{PP'\}, \quad \text{say}$$

which thus reduces the problem to that of finding the expected covariance associated with individual degrees of freedom.

The model for $x_{i.}$ is $x_{i.} = \mu_1 + z_{i.} + \epsilon_i$ where μ_1 and ϵ have the usual meanings and $z_{i.}$ is the mean of the genotypic effects in the C_i progeny assumed to be random. The model for $x_{ij'.}$ is similar except that μ_2 replaces μ_1 and $z_{ij'.}$ replaces $z_{i.}$. Strictly speaking the models should have an additional term for year × genotype interaction. For present purposes, this term is not considered. Only certain of the sample variance components would have their expectations changed under the more complex model.

4.2 *Genotypic covariance of observed progeny means*:

Assuming that the environmental effects of one year are uncorrelated with those of another year, then EPP' reduces to

$$EPP' = \tfrac{1}{2}\{E(z_{i.}z_{ij'.}) + E(z_{i.}z_{ij''.}) + E(z_{i'.}z_{i'j'.}) + E(z_{i'.}z_{i'j''.}) - E(z_{i.}z_{i'j'.}) - E(z_{i.}z_{i'j''.}) - E(z_{i'.}z_{ij'.}) - E(z_{i'.}z_{ij''.})\}.$$

By the procedure to be described subsequently it can be shown that the four terms entering with a plus sign are equal while the other four are zero and thus $EPP' = 2E(z_{i.}z_{ij'.}\ )$. This genotypic covariance of observed progeny means can be written as

$$E(z_{i.}z_{ij'.}) = E\left\{\frac{z_{i1} + \cdots z_{ij'} + \cdots z_{in_i}}{n_i}\right\}\left\{\frac{z_{ij'1} + \cdots z_{ij'n_{ij'}}}{n_{ij'}}\right\}$$

which reduces to

$$\frac{n_i - 1}{n_i} E(z_{ij}z_{ij'k}) + \frac{1}{n_i} E(z_{ij'}z_{ij'k}).$$

$E(z_{ij}z_{ij'k})$ is the genotypic covariance of a plant in C_i with a plant in the progeny of a sib. $E(z_{ij'}z_{ij'k})$ is the genotypic covariance of a plant in C_i with a plant in its own progeny.

4.3 *Model for genotypic effect of an observed plant:*

The genotypic covariances of observed plants, $E(z_{ij}z_{ij'k})$ and $E(z_{ij'}z_{ij'k})$, can be related to population genotypic covariances. To do this let δ^{ij}_{tu} be a random variable that assumes the value *one* when the genotypic effect of the jth plant in C_i corresponds to the genotypic effect of the uth genotype in the progeny of the tth genotype in the F_3 generation and zero otherwise. Likewise let δ^{ijk}_{tuv} be a random variable which assumes the value *one* if the kth plant in the progeny from the jth plant in the progeny of C_i corresponds to the genotypic effect of the vth genotype in the progeny from the uth genotype in the progeny from the tth genotype of the F_3 generation and zero otherwise.

The properties of the δ's are known from genetic theory. For example $E(\delta^{ij}_{tu}) = p_t p_{tu}$ where p_t is the probability of the tth genotype in the F_3 generation and p_{tu} is the probability of the uth genotype in the progeny of the tth genotype given the tth genotype. Similarly $E(z^{ijk}_{tuv}) = p_t p_{tu} p_{tuv}$ and

$$E(\delta^{ij}_{tu}\delta^{i'j'k}_{t'u'v}) = \begin{cases} p_t p_{tu} p_{t'} p_{t'u'} p_{t'u'v} & \text{for } i' \neq i \\ p_t p_{tu} p_{tu'} p_{tu'v} & \text{for } i' = i,\ t' = t \text{ and } j' \neq j \\ p_t p_{tu} p_{tuv} & \text{for } i' = i,\ t' = t,\ j' = j, \text{ and } u' = u \\ 0 & \text{otherwise} \end{cases}$$

The genotypic effects of observed plants are now related to population genotypic effects by the equations

$$z_{ij} = \sum_t \sum_u \delta^{ij}_{tu}(y_{tu} - y_{..})$$

$$z_{ij'k} = \sum_t \sum_u \sum_v \delta^{ij'k}_{tuv}(y_{tuv} - y_{...})$$

where y_{tu} is the genotypic value of the uth genotype in the progeny of the tth genotype, $y_{..}$ is the genotypic mean of the F_3 generation and y_{tuv} and $y_{...}$ have similar meanings. It follows that

$$E(z_{ij}z_{ij'k}) = E\{\sum_t \sum_u \delta^{ij}_{tu}(y_{tu} - y_{..})\}\{\sum_t \sum_u \sum_v \delta^{ij'k}_{tuv}(y_{tuv} - y_{...})\}.$$

After simplifying the right hand side by making use of the properties of the δ's, we have

$$E(z_{ij}z_{ij'k}) = \sum_t p_t(y_{t.} - y_{..})(y_{t..} - y_{...})$$

$$= \text{Cov }(3; 4, 5).$$

In a similar manner it can be shown that

$$E(z_{ij'}z_{ij'k}) = \text{Cov }(4; 4, 5).$$

4.4 *Expected Value of* $\widehat{\text{Cov}}\,[C, (D + D^*)/2]$:

This expected value is then

$$\frac{1}{2n(n-1)} \sum_{i} \sum_{\substack{i' \\ i' \neq i}} 2\left\{\left(\frac{n_i - 1}{n_i}\right) \text{Cov}\,(3; 4, 5) + \frac{1}{n_i}\,[\text{Cov}\,(4; 4, 5)]\right\}.$$

However, Cov (4; 4, 5) represents the total genotypic covariance of the F_4 and F_5 generations. Hence Cov (4; 4, 5) contains Cov (3; 4, 5) plus an additional covariance that might be represented as Cov (4; 4, 5) − Cov (3; 4, 5). Making this substitution and simplifying, we have

$$E\{\widehat{\text{Cov}}\,[C, \tfrac{1}{2}(D + D^*)]\}$$
$$= \text{Cov}\,(3; 4, 5) + b[\text{Cov}\,(4; 4, 5) - \text{Cov}\,(3; 4, 5)]$$

where b is the reciprocal of the harmonic mean of the n_i, which is the number of plants in C_i.

5. NUMERICAL EXAMPLE

The derivations above have shown that the expected values of the sample covariances are linear functions of the Cov $(k; n, n')$; it remains to be indicated how actual genotypic models are fitted to the data. Accordingly, let Y represent a sample covariance. Then

$$Y = \text{Cov}\,(k; n, n') + \epsilon$$

where ϵ is an error.

For instance, if an additive model with dominance with additive × additive interaction, two alleles per locus, no linkage and a gene frequency of one-half is assumed, then it can be shown that

$$\text{Cov}\,(k; n, n') = \left(\frac{2^{k-1} - 1}{2^{k-2}}\right)\sigma_A^2 + \left(\frac{2^{k-1} - 1}{2^{n+n'-4}}\right)\sigma_D^2 + \left(\frac{2^{k-1} - 1}{2^{k-2}}\right)^2 \sigma_{AA}^2 .$$

If one makes other assumptions then Cov $(k; n, n')$ will be different. Assuming the above model, however, and defining

$$\beta_1 = \sigma_A^2, \qquad \beta_2 = \sigma_D^2, \qquad \beta_3 = \sigma_{AA}^2,$$

$$X_1 = \frac{2^{k-1} - 1}{2^{k-2}}, \qquad X_2 = \frac{2^{k-1} - 1}{2^{n+n'-4}}, \quad \text{and} \quad X_3 = \left(\frac{2^{k-1} - 1}{2^{k-2}}\right)^2,$$

the model for the sample covariance becomes

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

and least squares procedures can be utilized to give estimates of β_1, β_2 and β_3. Values of the X's for the thirty sample covariances are shown in Table 3 along with the sample covariances for the character maturity. This character maturity was the number of days after August 31 that 95 to 100% of the pods in a plot were ripe as measured by eye.

TABLE 3

OBSERVED AND ESTIMATED VALUES OF COVARIANCES

				Sample covariance	Estimated value		
Covariance	X_1	X_2	X_3	Y	$\hat{Y}_1$	$\hat{Y}_2$	$\hat{Y}_3$
Cov (2; 2, 3)	1.00	.50	1.00	9.7	10.9	11.6	11.5
Cov (2; 2, 4)	1.00	.25	1.00	7.1	10.9	11.2	11.2
Cov (2; 2, 5)	1.00	.13	1.00	7.3	10.9	11.0	11.1
Cov (2; 2, 6)	1.00	.06	1.00	6.0	10.9	10.9	11.0
Cov (2; 2, 7)	1.00	.03	1.00	7.6	10.9	10.9	11.0
Cov (2; 3, 4)	1.00	.13	1.00	13.2	10.9	11.0	11.1
Cov (2; 3, 5)	1.00	.06	1.00	14.2	10.9	10.9	11.0
Cov (2; 3, 6)	1.00	.03	1.00	13.1	10.9	10.9	11.0
Cov (2; 3, 7)	1.00	.02	1.00	13.8	10.9	10.9	11.0
Cov (3; 4, 5)	1.50	.09	2.25	16.6	16.4	16.4	16.5
Cov (3; 4, 6)	1.50	.05	2.25	13.9	16.4	16.3	16.4
Cov (3; 4, 7)	1.50	.02	2.25	15.3	16.4	16.3	16.4
Cov (4; 5, 6)	1.75	.05	3.06	18.0	19.1	19.1	19.1
Cov (4; 5, 7)	1.75	.03	3.06	20.2	19.1	19.0	19.1
Cov (5; 6, 7)	1.88	.03	3.52	19.3	20.5	20.4	19.3
Cov (2; 4, 5)	1.00	.03	1.00	11.8	10.9	10.9	11.0
Cov (2; 4, 6)	1.00	.02	1.00	9.7	10.9	10.9	11.0
Cov (2; 4, 7)	1.00	.01	1.00	11.4	10.9	10.9	11.0
Cov (3; 5, 6)	1.50	.02	2.25	17.0	16.4	16.3	16.4
Cov (3; 5, 7)	1.50	.01	2.25	18.2	16.4	16.3	16.4
Cov (4; 6, 7)	1.75	.01	3.06	16.9	19.1	19.0	19.1
Cov (2; 3, 3)	1.00	.25	1.00	19.8	10.9	11.2	11.2
Cov (2; 4, 4)	1.00	.06	1.00	10.9	10.9	10.9	11.0
Cov (3; 4, 4) − Cov (2; 4, 4)	.50	.13	1.25	5.5	5.5	5.6	5.5
Cov (3; 5, 5)	1.50	.05	2.25	20.0	16.4	16.3	16.4
Cov (4; 5, 5) − Cov (3; 5, 5)	.25	.06	.81	3.8	2.7	2.8	2.7
Cov (4; 6, 6)	1.75	.03	3.06	17.1	19.1	19.0	19.1
Cov (5; 6, 6) − Cov (4; 6, 6)	.13	.03	.45	3.0	1.4	1.4	1.4
Cov (5; 7, 7)	1.88	.01	3.52	21.8	20.5	20.4	20.4
Cov (6; 7, 7) − Cov (5; 6, 6)	.06	.02	.24	2.4	.7	.7	.7
				R^2:	.9623	.9625	.9627

$\hat{Y}_1 = 10.92X_1$	Additive Model
$\hat{Y}_2 = 10.85X_1 + 1.43X_2$	Additive Model with dominance
$\hat{Y}_3 = 11.06X_1 + 1.07X_2 - .10X_3$	Additive Model with dominance and additive × additive interaction

The three $\hat{Y}$ columns shown in Table 3 represent the population covariances as estimated from linear regression, associated with a completely additive model, an additive model with dominance, and an additive model with dominance and additive × additive interaction.

A detailed interpretation of this set of data will not be given here. It might be noted, however, that the completely additive model fits this set of data as well as the two more complex models, explaining about 96% of the variation in the dependent variable.

Confidence limits for the Cov $(k; n, n')$ and prediction limits for the sample covariances in the case of the completely additive model are shown in Table 4. In interpreting the data one must bear in mind that certain of the estimates of the variance components may be biased upward because of the ignoring of genotype × year interaction. The estimates of the 21 sample covariances and the 9 sample variance components are not completely independent. Linkage may be an important factor in the early generations and natural selection progressively more important in the latter generations. The F_2 data were individual plant values, while the data of other generations were based on plot values.

TABLE 4

CONFIDENCE AND PREDICTION LIMITS FOR THE COMPLETELY ADDITIVE MODEL

Coefficient of σ_A^2, X_1	Confidence limits for population covariances l_1	l_2	Prediction limits for sample covariances l_1	l_2
.0625	.63	.73	−4.9	6.3
.125	1.26	1.47	−4.3	7.0
.25	2.53	2.94	−2.9	8.4
.5	5.05	5.87	− .2	11.1
1.0	10.1	11.7	5.2	16.6
1.5	15.2	17.6	10.6	22.1
1.75	17.7	20.5	13.3	24.9
1.875	18.9	22.0	14.6	26.3

SUMMARY

This paper is concerned with the derivation of the expected values of sample covariances and variance components in terms of genotypic variances and covariances for populations produced from crossing two homozygous lines and subsequent self-fertilization. Such genotypic variances and covariances have been defined by authors including

Fisher, Mather, Horner and Kempthorne. Use is made of random variables such as δ_t^i, where $\delta_t^i = 1$ if the genotype of the plant labeled i in an experiment corresponds to the tth genotype in a list of possible genotypes and zero otherwise. Application was made to maturity data of a soybean experiment. In this experiment 21 covariances and 9 variance components, unconfounded with environmental effects, were estimated. A completely additive model fitted this set of data as well as did two more complex models, explaining 96% of the variation among the sample covariances and variance components.

REFERENCES

Fisher, R. A., Immer, F. R., and Teden, O. [1932]. The genetical interpretation of statistics of the third degree in the study of quantitative inheritance. *Genetics, 17:* 107–124.

Gates, Charles, E. [1954]. The constitution of genetic variances and covariances in self fertilized crops assuming linkage. Ph.D. Dissertation. North Carolina State College, Raleigh, N. C.

Horner, T. W., Comstock, R. E. and Robinson, H. F. [1955]. Non-allelic gene interaction and the interpretation of quantitative genetic data. *Tech. Bull. No. 118.* North Carolina Agricultural Experiment Station, Raleigh, N. C.

Kempthorne, O. [1956]. On the covariances between relatives under selfing with general epistacy. *Proc. Roy. Soc. London Ser. B.* In press.

Mather, K. [1949]. *Biometrical genetics*. Dover Publications, Inc., New York.