1952

A Statistical Model for Estimating Components of Genetic Variance in Bulk Yield Tests of Self-Pollinated Small Grains

J.E. Grafius

Technical Bulletin No. 9 May 1952

A Statistical Model For Estimating Components of Genetic Variance in Bulk Yield Tests of Self-Pollinated Small Grains

AGRONOMY
DEPARTMENT

AGRICULTURAL EXPERIMENT STATION
South Dakota State College of Agriculture and Mechanic Arts
Brookings, South Dakota

A Statistical Model For Estimating Components of Genetic Variance in Bulk Yield Tests of Self-Pollinated Small Grains

AGRONOMY
DEPARTMENT

AGRICULTURAL EXPERIMENT STATION
South Dakota State College of Agriculture and Mechanic Arts
Brookings, South Dakota

ACKNOWLEDGMENT

The author wishes to thank Professors H. B. MacDougal and O. E. Walder of South Dakota State College for their assistance. Credit also is due Dr. R. E. Comstock of the North Carolina State College for his critical review of this manuscript and for his many helpful suggestions.

A STATISTICAL MODEL FOR ESTIMATING THE COMPONENTS OF GENETIC VARIANCE IN BULK YIELD TESTS OF SELF-POLLINATED SMALL GRAINS

J. E. GRAFIUS [1]

Comstock and Robinson (1)[2] have given a statistical model for the isolation of the components of genetic variance in populations of biparental progenies. Their model has made possible the construction of one for self-pollinated progenies to be applied to the method of small grain breeding (3) (4), wherein large numbers of bulked F_2 and F_3 progenies are tested for yielding ability. Inasmuch as the genetic variance of high yielding progenies will be due to both additive and non-additive genetic effects, it is important that these effects be separated.

In self-pollinated small grains, the final objective of the breeding program is an isogenic line. Hence the additive, or heritable genetic variance is highly important, while the non-additive genetic variance is, in general, not usable.

Limiting Assumptions and Definitions

The genetic models to be proposed assume that the genes at the separate loci are randomly distributed in the population of homozygous lines from which the parents were drawn at random.

It is also assumed that there is no epistasis and (for the purposes of summation) that genes 1 to n have equal effect. In the absence of bias caused by epistasis, linkage will not interfere with the conclusions that can be drawn from the model for the analysis of variance. This analysis is based on cross means, and in the absence of epistasis the cross mean will be the result of the average effect of the segregating loci irrespective of the linkage relationships.

The additive genetic variance at the ath locus has been defined (1) (6) as that part of the variance of the genetic effects attributable to regression on the number of A genes in the genotype. Conversely, the variance of the genetic deviations from regression represents the

[1] Agronomist, South Dakota Agricultural Experiment Station.

[2] Figures in parentheses refer to "Literature Cited," p. 13.

variance due to deviations from the additive scheme caused by dominance and epistasis. The additive genetic variance may, in turn, be divided into an unfixable and a fixable portion. It is this latter part of the additive genetic variance that may be retained in an isogenic line.[3]

Estimations of Components of Genetic Variance Within Crosses Between Self-Pollinated Varieties

After the genetic model presented by Comstock and Robinson, let y be the mean effects of the genotype, x be the number of A's in the genotype, u be the frequency of A, and $v = (1-u)$ be the frequency of a in the population. Then for an F_2, the genotypes for a single locus will have the distribution shown in Table 1.

TABLE 1. THE DISTRIBUTION OF GENOTYPES AND THEIR MEAN EFFECTS FOR AN F_2 POPULATION

Genotype	Genotypic frequency	Frequency of A x	Mean effects y
AA	u_a^2	2	d_a
Aa	$2u_a v_a$	1	h_a
$a\,a$	v_a^2	0	$-d_a$

Before proceeding to calculate the F_2 variance it is necessary to define certain symbols. For a single segregating locus let

$\sigma^2_{y_a}$ = the total genetic variance

$\sigma^2_{d_a}$ = the additive genetic variance

$\sigma^2_{h_a}$ = the non-additive genetic variance

$Cov_{x_a y_a}$ = the covariance of x_a and y_a

and $\sigma^2_{x_a}$ = the variance of x_a.

Upon summing the variance from a series of such segregating loci

$$\sigma^2_y = \sum_1^n \sigma^2_{y_a}, \qquad \sigma^2_d = \sum_1^n \sigma^2_{d_a}, \qquad \sigma^2_h = \sum_1^n \sigma^2_{h_a}, \qquad \sigma^2_x = \sum_1^n \sigma^2_{x_a},$$

$$\text{and } Cov_{xy} = \sum_1^n Cov_{x_a y_a}$$

[3] For example, the additive genetic variance for the ath locus in an F_2 may be written $2\Sigma u_a v_a \left[d_a + h_a (v_a - u_a) \right]^2$. The term containing d_a^2 is fixable, but the terms containing h_a are not fixable.

Working from Table 1 it was found that

$$\sigma^2_{y_a}(F_2) = d_a^2(u_a^2+v_a^2) + 2u_a v_a h_a^2 - \left[(u_a - v_a)\, d_a + 2u_a v_a h_a\right]^2$$

$$= d_a^2(2u_a v_a) + 4u_a v_a d_a h_a (v_a - u_a) + 2u_a v_a h_a^2 (1 - 2u_a v_a)$$

$$= 2u_a v_a \left[d_a + h_a(v_a - u_a)\right]^2 + 4u_a^2 v_a^2 h_a^2$$

$$\sigma^2_y(F_2) = 2\Sigma u_a v_a \left[d_a + h_a(v_a - u_a)\right]^2 + 4\Sigma u_a^2 v_a^2 h_a^2 \qquad [1]$$

Equation [1] gives the total F_2 genetic variance. By definition, the additive genetic variance at the ath locus equals

$$\frac{\left[Cov_{x_a y_a}(F_2)\right]^2}{\sigma^2_{x_a}}.$$

This may be obtained as follows

$$Cov_{x_a y_a}(F_2) = 2u_a^2 d_a + 2u_a v_a h_a - \left[(2u_a^2 + 2u_a v_a)(u_a^2 d_a + 2u_a v_a h_a - v_a^2 d_a)\right]$$

$$= 2u_a v_a \left[d_a + h_a(1 - 2u_a)\right]$$

$$= 2u_a v_a \left[d_a + h_a(v_a - u_a)\right]$$

$$\sigma^2_{x_a} = u_a^2(2)^2 + 2u_a v_a (1)^2 - \left[2u_a^2 + 2u_a v_a\right]^2 = 2u_a v_a, \quad \sigma^2_x = 2\Sigma u_a v_a$$

$$\sigma^2_{d_a}(F_2) = \frac{\left[Cov_{x_a y_a}(F_2)\right]^2}{\sigma^2_{x_a}} = \left\{2u_a v_a \left[d_a + h_a(v_a - u_a)\right]\right\}^2$$

$$\sigma^2_d(F_2) = 2\Sigma u_a v_a \left[d_a + h_a(v_a - u_a)\right]^2 \qquad [2]$$

The non-additive genetic variance in [1] may be obtained by subtraction

$$\sigma^2_{h_a}(F_2) = \sigma^2_{y_a}(F_2) - \sigma^2_{d_a}(F_2) = 4u_a^2 v_a^2 h_a^2$$

$$\sigma^2_h(F_2) = 4\Sigma u_a^2 v_a^2 h_a^2 \qquad [3]$$

Similarly it was found that for the variance of the F_2 family means that

$$\sigma_y^2(\bar{F}_3) = 2\Sigma u_a v_a \left[d_a + \frac{h_a}{2}(v_a - u_a) \right]^2 + \Sigma u_a^2 v_a^2 h_a^2 \qquad [4]$$

$$\sigma_d^2(\bar{F}_3) = 2\Sigma u_a v_a \left[d_a + \frac{h_a}{2}(v_a - u_a) \right]^2 \qquad [5]$$

$$\sigma_h^2(\bar{F}_3) = \Sigma u_a^2 v_a^2 h_a^2 \qquad [6]$$

and that for the variance of the F_4 family means from unselected F_3 families that

$$\sigma_y^2(\bar{F}_4) = 2\Sigma u_a v_a \left[d_a + \frac{h_a}{4}(v_a - u_a) \right]^2 + \frac{\Sigma u_a^2 v_a^2 h_a^2}{4} \qquad [7]$$

$$\sigma_d^2(\bar{F}_4) = 2\Sigma u_a v_a \left[d_a + \frac{h_a}{4}(v_a - u_a) \right]^2 \qquad [8]$$

and that

$$\sigma_h^2(\bar{F}_4) = \frac{\Sigma u_a^2 v_a^2 h_a^2}{4} \qquad [9]$$

In a cross of two homozygous lines where $u = v = \frac{1}{2}$, equations [1-9] will reduce to terms of d^2 and/or h^2. For example, equation [1] will reduce to $\frac{1}{2}\Sigma d_a^2 + \frac{1}{4}\Sigma h_a^2$

which is the expression obtained by Fisher, Tedin and Immer (2) for the total F_2 genetic variance. The reduced equation [7] equals

$$\frac{1}{2}\Sigma d_a^2 + \frac{1}{64}\Sigma h_a^2.$$

This can be verified by following the method given by Mather (5).

If $u \neq v \neq \frac{1}{2}$ and $h \neq 0$, the contributions made by d and h are not completely separable. Under these conditions σ_d^2 will include some of the effects of h and σ_h^2 will be correspondingly less than the summed effects of all the squared h deviations (5).

Estimation of the Composition of Variance between Self-Fertilized Bulked Progenies from Different Crosses

Let j homozygous lines be used as female parents to be crossed with each of k males. The bulk progeny from each cross will be grown

in r replications. The variance can then be partitioned as shown in Table 2.

TABLE 2. ANALYSIS OF VARIANCE BETWEEN SELF-FERTILIZED BULKED PROGENIES FROM CROSSES OF HOMOZYGOUS LINES

Source of variance	d.f.	m.s.	Expectations of m.s.
Replication	$r-1$		
Crosses	$jk-1$		
Between progeny of different ♂ ♂	$k-1$	M_1	$\sigma^2 + r\sigma^2_{m\times f} + rj\sigma^2_m$
Between progeny of different ♀ ♀	$j-1$	M_2	$\sigma^2 + r\sigma^2_{m\times f} + rk\sigma^2_f$
Interaction	$(j-1)(k-1)$	M_3	$\sigma^2 + r\sigma^2_{m\times f}$
Error	$(r-1)(jk-1)$	M_4	σ^2
Total	$jkr-1$		

KEY TO TABLE 2

σ^2 = environmental variance

σ^2_m = the variance of male effects

σ^2_f = the variance of female effects

$\sigma^2_{m\times f}$ = the variance due to the interaction of male and female effects

The four components of variance can be estimated from the appropriate mean squares. To illustrate,

$$(M_1 - M_3) / rj = \sigma^2_m .$$

The genetic model for Table 2 will be as follows:

Starting with homozygous parents, the frequency of gene $A-a$ will also equal the frequency of the genotype AA and aa. Let u equal the frequency of the AA genotype and $v = (1-u)$ the frequency of the aa genotype in the hypothetical parental population. Then, on the average,[4] u of the males and females will be AA and v will be aa. An AA female may have F_1 progeny of two types, AA

[4] Where the males and females are picked at random.

and Aa with a frequency of u^2 and uv respectively. An aa female may have F_1 progeny of two types Aa and aa with frequencies of uv and v^2 respectively. The same reasoning applies to the male parents.

TABLE 3. MEAN EXPRESSIONS FOR F_1 PROGENIES FROM CROSSES BETWEEN HOMOZYGOUS LINES

Parental genotype			Means of progeny from female parent
Female	Male		
	AA u	aa v	
	Progeny means		
AA u	$\bar{y}_a = d_a$	h_a	$u_a d_a + v_a h_a$
aa v	h_a	$-d_a$	$u_a h_a - v_a d_a$
Means of progeny from male parents	$u_a d_a + v_a h_a$	$u_a h_a - v_a d_a$	Grand mean $= d_a(u_a - v_a) + 2u_a v_a h_a$

The total genetic variance will be

$$\sigma^2_{y_a} = d_a^2 (u_a^2 + v_a^2) + 2u_a v_a h_a^2 - \left[(u_a - v_a) d_a + 2u_a v_a h_a \right]^2$$

$$= 2u_a v_a \left[d_a + h_a (v_a - u_a) \right]^2 + 4u_a^2 v_a^2 h_a^2$$

which upon summation for n such genes will be seen to equal equation [1].[5]

The variance for between progeny of different females is

$$\sigma^2_{f_a} = u_a (u_a d_a + v_a h_a)^2 + v_a (u_a h_a - v_a d_a)^2 - \left[(u_a - v_a) d_a + 2u_a v_a h_a \right]^2$$

$$= d_a^2 (u_a^3 + v_a^3 - u_a^2 - v_a^2 + 2u_a v_a) + 2 u_a v_a d_a h_a (v_a - u_a) + u_a v_a h_a^2 (1 - 4u_a v_a)$$

[5] The equivalence of σ^2_y, σ^2_m, σ^2_f, and $\sigma^2_{m \times f}$ to equations [1-9] for within cross variance is algebraic only. Under experimental conditions the estimates of σ^2_y, σ^2_m, σ^2_f, and $\sigma^2_{m \times f}$ will differ from the within cross estimates whenever: (a) $u \neq v \neq \frac{1}{2}$; (b) linkage influences the within cross variance; (c) where epistasis is present.

$$= d_a^2\left[u_a^3+1-3u_a+3u_a^2-u_a^3-u_a^2-v_a^2+2u_av_a\right]+2u_av_ad_ah_a(v_a-u_a)$$
$$+u_av_ah_a^2(1-4u_av_a)$$

$$= d_a^2\left[1-3u_a+3u_a^2-u_a^2-(1-2u_a+u_a^2)+2u_av_a\right]+2u_av_ad_ah_a(v_a-u_a)$$
$$+u_av_ah_a^2(v_a-u_a)^2$$

$$= d_a^2(-u_av_a+2u_av_a)+2u_av_ad_ah_a(v_a-u_a)+u_av_ah_a^2(v_a-u_a)^2$$

$$= u_av_a\left[d_a+h_a(v_a-u_a)\right]^2$$

which upon summation for n such genes will be seen to equal $\frac{1}{2}\sigma_d^2(F_2)$ in equation [2].

Similarly upon summation, the variance for between progeny of different males turns out to be

$$\sigma_m^2 = \Sigma u_av_a\left[d_a+h_a(v_a-u_a)\right]^2 \quad \text{and} \quad \sigma_f^2 = \sigma_m^2$$

Then the variance for the interaction will be

$$\sigma_{m_a\times f_a}^2 = \sigma_{y_a}^2-(\sigma_{m_a}^2+\sigma_{f_a}^2) = 4u_a^2v_a^2h_a^2$$

which when summed for a series of such genes will be found to equal $\sigma_h^2(F_2)$ in equation [3].

The model for the mean expressions for the F_1 may be used for the F_2 and F_3. Only one change is necessary. The mean of the heterozygous classes becomes $h/2$ and $h/4$ for the F_2 and F_3 respectively.

In the F_2

$\sigma_m^2 = \sigma_f^2 = \Sigma u_av_a\left[d_a+\frac{h_a}{2}(v_a-u_a)\right]^2 = \frac{1}{2}\sigma_d^2(\bar{F}_3)$ in equation [5] and

$\sigma_{m\times f}^2 = \Sigma u_a^2v_a^2h_a^2 = \sigma_h^2(\bar{F}_3)$ in equation [6].

In the F_3

$\sigma_m^2 = \sigma_f^2 = \Sigma u_av_a\left[d_a+\frac{h_a}{4}(v_a-u_a)\right]^2 = \frac{1}{2}\sigma_d^2(\bar{F}_4)$ in equation [8] and

$\sigma_{m\times f}^2 = \frac{\Sigma u_a^2v_a^2h_a^2}{4} = \sigma_h^2(\bar{F}_4)$ in equation [9].

It is thus apparent that algebraically in the F_1

$$\sigma_m^2 = \sigma_f^2 = \frac{1}{2}\sigma_d^2(F_2)$$

$$\sigma^2_{m\times f} = \sigma^2_h\,(F_2)$$

in the F_2

$$\sigma^2_m = \sigma^2_f = \frac{1}{2}\sigma^2_d\,(\bar{F}_3)$$

$$\sigma^2_{m\times f} = \frac{1}{4}\sigma^2_h\,(F_2)$$

and in the F_3

$$\sigma^2_m = \sigma^2_f = \frac{1}{2}\sigma^2_d\,(\bar{F}_4)$$

$$\sigma^2_{m\times f} = \frac{1}{16}\sigma^2_h\,(F_2)$$

The estimate of heritability of yield for any generation equals the additive genetic variance divided by the total variance.

$$= \frac{\sigma^2_m + \sigma^2_f}{\sigma^2_m + \sigma^2_f + \sigma^2_{m\times f} + \sigma^2}$$

Discussion

In the proposed genetic models it is assumed that the genes at the separate loci are randomly distributed in a population of homozygous lines from which the parents were drawn at random. It has also been assumed that epistatic interactions are absent. No assumptions can be made concerning the size of u relative to v as the effects of selection, natural or otherwise, are unavoidable which would have a tendency to make $u > v$.

If $h \neq o$ and $u > v$ the fixable portion of the additive genetic variance will have a negative bias which will, however, diminish as homozygosity is approached. Where $h = o$, σ^2_m and σ^2_f estimate $\Sigma u_a v_a d^2_a$ regardless of the values of u. If $\sigma^2_m = \sigma^2_f$ is expanded, the way in which $\Sigma u_a v_a d^2_a$ is biased when $u \neq v$ and $h \neq o$ can be seen.

$$F_1\sigma^2_m = \Sigma u_a v_a d^2_a + 2\Sigma u_a v_a d_a h_a\,(v_a - u_a) + \Sigma u_a v_a h^2_a\,(v_a - u_a)^2$$

$$F_2\sigma^2_m = \Sigma u_a v_a d^2_a + \Sigma u_a v_a d_a h_a\,(v_a - u_a) + \Sigma u_a v_a \frac{h^2_a}{4}\,(v_a - u_a)^2$$

$$F_3\sigma^2_m = \Sigma u_a v_a d^2_a + \Sigma u_a v_a d_a \frac{h_a}{2}\,(v_a - u_a) + \Sigma u_a v_a \frac{h^2_a}{16}\,(v_a - u_a)^2$$

For an arithmetic example, assume that $h = d$ (complete dom-

inance). Then the three terms of σ_m^2 take the following values as u changes for one locus. Table 4.[6]

TABLE 4. INFLUENCE OF CHANGES IN u ON σ_m^2 WHERE $h = d$, FOR ONE LOCUS

Generation	u	$u_a v_a d_a^2$	$2u_a v_a d_a h_a (v_a - u_a)$	$u_a v_a h_a^2 (v_a - u_a)^2$	Total bias
F_1	0.5	$0.25d_a^2$	0	0	0
	0.6	$0.24d_a^2$	$-0.096d_a^2$	$0.0096d_a^2$	$-0.0864d_a^2$
	0.7	$0.21d_a^2$	$-0.168d_a^2$	$0.0336d_a^2$	$-0.1344d_a^2$
	0.8	$0.16d_a^2$	$-0.192d_a^2$	$0.0576d_a^2$	$-0.1344d_a^2$
	0.9	$0.09d_a^2$	$-0.144d_a^2$	$0.0576d_a^2$	$-0.0864d_a^2$
F_2	0.5	$0.25d_a^2$	0	0	0
	0.6	$0.24d_a^2$	$-0.048d_a^2$	$0.0024d_a^2$	$-0.0456d_a^2$
	0.7	$0.21d_a^2$	$-0.084d_a^2$	$0.0084d_a^2$	$-0.0756d_a^2$
	0.8	$0.16d_a^2$	$-0.096d_a^2$	$0.0144d_a^2$	$-0.0816d_a^2$
	0.9	$0.09d_a^2$	$-0.072d_a^2$	$0.0144d_a^2$	$-0.0576d_a^2$
F_3	0.5	$0.25d_a^2$	0	0	0
	0.6	$0.24d_a^2$	$-0.024d_a^2$	$0.0006d_a^2$	$-0.0234d_a^2$
	0.7	$0.21d_a^2$	$-0.042d_a^2$	$0.0021d_a^2$	$-0.0399d_a^2$
	0.8	$0.16d_a^2$	$-0.048d_a^2$	$0.0036d_a^2$	$-0.0444d_a^2$
	0.9	$0.09d_a^2$	$-0.036d_a^2$	$0.0036d_a^2$	$-0.0324d_a^2$

Values for other than $h = d$ may be readily obtained by multiplying the numbers in the fourth and fifth column by h/d and $(h/d)^2$ respectively.

For a single gene pair in the F_1 the effects of $u \neq v$ will be increased to a maximum bias due to h_a where $h_a = \dfrac{-d_a}{(v_a - u_a)}$.

This result can be obtained by setting the total bias equal to y and differentiating with respect to h_a.

[6] The author is indebted to Dr. R. E. Comstock for suggesting this table.

$$y = 2u_a v_a d_a h_a (v_a - u_a) + u_a v_a h_a^2 (v_a - u_a)^2$$

$$\frac{dy}{dh} = 2u_a v_a d_a (v_a - u_a) + 2u_a v_a h_a (v_a - u_a)^2$$

$$h_a = \frac{-d_a}{v_a - u_a}$$

For example, if $d_a = 1$ and $u_a = 0.7$, a maximum bias due to h_a in the F_1 would be obtained when $h_a = 1/0.4 = 2.5$. Substituting back in the original equation, we see that a value of $h_a = 2.5$ (where $d_a = 1$) will reduce σ_m^2 to zero. Values of h_a this high would imply extreme over-dominance. Similar maximum values may be calculated for the F_2 and F_3. In addition, estimates of $\sigma_m^2 = \sigma_f^2$ may be biased where epistasis is present. This bias may be minimized by transformation of the data to a more suitable scale such as logarithms (5). Also, the bias due to epistasis will be less as the crosses become more homozygous.

Summary

A model for the estimation of the components of genetic variance between self-fertilized bulked progenies from crosses of isogenic lines has been presented. The mathematical basis for the method of estimation has been presented in detail. The limiting assumptions have been briefly discussed.

LITERATURE CITED

1. COMSTOCK, R. E. and ROBINSON, H. F. The components of genetic variance in populations of biparental progenies and their use in estimating the average degree of dominance. Biometrics, **4**:254-266. 1948.

2. FISHER, R. A., IMMER, F. R., and TEDIN, O. The genetical interpretation of statistics of the third degree in the study of quantitative inheritance. Genetics, **17**:107-124. 1932.

3. HARLAN, H. V., MARTINI, M. L., and STEVENS, HARLAND. A study of methods in barley breeding. U.S.D.A. Tech. Bull. No. 720. 1940.

4. HARRINGTON, J. B. Yielding capacity of wheat crosses as indicated by bulk hybrid tests. Can. Jour. Res., C, **18**:578-584. 1940.

5. MATHER, K. Biometrical Genetics. 158 pp. Dover Publications, Inc. 1949.

6. WRIGHT, SEWALL. The analysis of variance and the correlations between relatives with respect to deviations from an optimum. Jour. Genetics, **30**:243-256. 1935.