Theor Appl Genet (1985) 70:318–321

Sampling bias in estimating Design II variance components with S_1 families*

L. P. Jones [1] and W. A. Compton [2]

[1] Department of Agriculture, P.O. Box 4041, G.P.O. Melbourne, Victoria 3001, Australia
[2] Professor of Agronomy, University of Nebraska, Lincoln, Nebraska, USA

Received November 30, 1984; Accepted January 8, 1985
Communicated by A. L. Kahler

Summary. The use of several S_1 individuals to represent an S_0 individual permits the use of a Design II mating scheme for plants with only one pistillate flower per plant. Estimates of additive (V_A) and dominance (V_D) variance from this mating scheme will be biased upwards, when a small number (10) of individuals of each S_1 line are used. This bias can be computed, and the additive and dominance estimates can be corrected. Of particular interest is the observation that the additive genetic variance contributes to bias in estimates of V_D. When S_0 plants are non inbred and their selfed-progeny (S_1 lines) are used to represent them in developing families for use in the Design II, $\hat{V}_A = V_A(1+1/2m_1)$ and $\hat{V}_D = V_D\left(1+\frac{5}{4m_2}\right)+\left(\frac{1}{m_2}-\frac{1}{m_1}\right)V_A$ where m_1 is the number of individuals used to represent an S_1 line in developing half sib-families and m_2 is the number of individuals used to represent the S_1 line in making up full sib-families. For example, in a 3×3 Design II, with about 10 individuals used to represent each S_1 line in each cross, $m_2 = 10$ and $m_1 = 30$. When $m_1 = m_2 = 1$, $\hat{V}_A = (1+1/2)\,V_A = (1+F)\,V_A$ and $\hat{V}_D = V_D + \frac{5}{4} V_D = (1+F)^2 V_D$.

Key words: Quantitative genetics – Design II bias – Dominance variance

Introduction

Comstock and Robinson (1948) proposed two mating schemes to estimate additive and dominance variance in populations. These two mating schemes are commonly referred to as Design I and Design II. In Design I, a number of males are each mated to a set of females, a female being used in only one mating. The Design II is similar to a diallel mating scheme, except that each of a set of males is mated to each of a different set of females.

The Design II provides more precise estimates of dominance variance than the Design I (Comstock and Robinson 1952) and would be preferred where it can be used. Cockerham (1956) suggests a modification of the Design II that could be used with plants that have only one pistillate flower per plant (e.g., single-eared corn *Zea mays* L). Selfed progeny (S_1) of a parent plant (S_0) is used to represent the parent plant. The gametes from the selfed progeny are the same as those of the parent plant except as affected by the opportunity for recombination of linked genes. Cockerham notes that several plants from the selfed progeny should be used in each cross to ensure a representative sample. Sample sizes and methods of sampling were discussed by Hammond and Gardner (1974).

If F_0 is the inbreeding of the S_0 individual, then $F_1 = \frac{1+F_0}{2}$ is the inbreeding of the S_1 family derived by selfing the S_0 plant. If an infinite number of individuals are used to represent the S_1 family then F_0 is the appropriate inbreeding coefficient to use in computing the genetic variances from the covariances of relatives. On the other hand, if a single S_1 individual is used to make all crosses, the appropriate inbreeding level to use is F_1 (0.5 if $F_0 = 0$). Both of these alternatives are clear. However, when a small (say 2–10) number of individuals are used to represent each S_1 family, the appropriate coefficients for the additive and dominance variances have not been clear. This

---

* Joint contribution from Department of Agronomy, University of Nebraska 68583, and the S. S. Cameron Laboratory, Werribee, Victoria 3030, Australia. Published as paper No. 7395, Journal Series

Table 1[a]. Analysis of variance of progeny from a diallel crossing system

Source	Degrees of freedom	Mean squares	Expectations of mean squares	
			Components	Covariances
Between male parents (A)	n_1-1	M_{14}	$\sigma^2 + k\,\sigma^2_{f\times m} + k\,n_2\,\sigma^2_m$	$(\sigma^2_T - Cov) + k\,(Cov - Cov_1 - Cov_2) + k\,n_2\,Cov_2$
Between female parents (B)	n_2-1	M_{13}	$\sigma^2 + k\,\sigma^2_{f\times m} + k\,n_1\,\sigma^2_f$	$(\sigma^2_T - Cov) + k\,(Cov - Cov_1 - Cov_2) + k\,n_1\,Cov_1$
A × B	$(n_1-1)\,(n_2-1)$	M_{12}	$\sigma^2 + k\,\sigma^2_{f\times m}$	$(\sigma^2_T - Cov) + k\,(Cov - Cov_1 - Cov_2)$
Within crosses	$n_1\,n_2\,(k-1)$	M_{11}	σ^2	$(\sigma^2_T - Cov)$

The symbols have the following meaning:

n_1 = number of male parents
n_2 = number of female parents
k = number of progeny for each mating or cross
σ^2 = variance among individuals from the same mating
$\sigma^2_{f\times m}$ = progeny variance arising from interaction of the contributions of male and female parents
σ^2_f = progeny variance arising from differences among female parents
σ^2_m = progeny variance arising from differences among male parents
σ^2_T = total variance or covariance between an individual and itself
Cov = covariance between full sibs – plants with both parents common
Cov_1 = covariance between maternal half sibs – plants with only the female parent in common
Cov_2 = covariance between paternal half sibs – plants with only the male parent in common

[a] from Cockerham (1956)

paper is intended to clarify that point and also to show how to remove the biases that can occur.

Model

The analysis of variance from a Design II is given by Cockerham (1956) and is reproduced in Table 1. Relationships between the variance components and covariance among relatives are discussed by Cockerham, who also gives the expectations of the covariances in terms of the genetic parameters in the base population. Thus,

$$Cov_2 = \left(\frac{1+F_m}{4}\right) V_A + \left(\frac{1+F_m}{4}\right)^2 V_{AA} + \begin{array}{l}\text{higher order}\\ \text{epistatic terms.}\end{array} \qquad (1)$$

$$Cov_1 = \left(\frac{1+F_f}{4}\right) V_A + \left(\frac{1+F_f}{4}\right)^2 V_{AA} + \begin{array}{l}\text{higher order}\\ \text{epistatic terms.}\end{array} \qquad (2)$$

$$\begin{aligned} Cov = {} & \left(\frac{2+F_m+F_f}{4}\right) V_A + \left(\frac{1+F_m}{2}\right)\left(\frac{1+F_f}{2}\right) V_D \\ & + \left(\frac{2+F_m+F_f}{4}\right)^2 V_{AA} \\ & + \left(\frac{2+F_m+F_f}{2}\right)\left(\frac{1+F_m}{2}\right)\left(\frac{1+F_f}{2}\right) V_{AD} \\ & + \text{higher order epistatic terms}\,, \qquad (3) \end{aligned}$$

where V_A is the additive genetic variance, V_D is the dominance genetic variance; and all other terms are part of the epistatic variance, V_{AA} is the additive × additive, V_{AD} is the additive × dominance, and so on; and F_m, F_f are the inbreeding coefficients of the paternal and maternal plants, respectively.

The covariances will be increased when selfed progeny of a parent are used to represent the parent in making the crosses. To determine the bias of the estimates, six assumptions are made to simplify the calculations:

1) Because the selfed (S_1) progeny of any of the S_0 parents can be used as either male or female, it is possible to use a square diallel, i.e. $n_1 = n_2 = n$.

2) All plants are equally inbred, i.e. $F_m = F_f = F_i$.

3) The number (k) of plants measured from a cross is larger than the number (m_2) of S_1 plants used to represent an S_0 in a particular cross.

4) The number (m_2) of S_1 plants used to make a cross is the same for each S_0 plant and each S_0 plant is represented over all crosses by $m_1\,S_1$ plants, and each S_1 plant is used to make only one cross, $m_1 = n\,m_2$.

5) There is no limit to the number of generations of inbreeding of S_0 parents but all S_1 plants are equally inbred with inbreeding coefficient F_j.

6) Maternal effects are unimportant. When $F_m = F_f = F_i$ equations (1), (2) and (3) reduce to:

$$Cov_1 = Cov_2 = \left(\frac{1+F_i}{4}\right) V_A + \left(\frac{1+F_i}{4}\right)^2 V_{AA} + \ldots \qquad (4)$$

$$\begin{aligned} Cov = {} & \left(\frac{1+F_i}{2}\right) V_A + \left(\frac{1+F_i}{2}\right)^2 V_D + \left(\frac{1+F_i}{2}\right)^2 V_{AA} \\ & + \left(\frac{1+F_i}{2}\right)^3 V_{AD} + \left(\frac{1+F_i}{2}\right)^4 V_{DD} + \ldots\,. \qquad (5) \end{aligned}$$

When the epistatic variance is absent, the additive genetic variance can be estimated from (4) and the dominance variance from

$$\mathrm{Cov}-\mathrm{Cov}_1-\mathrm{Cov}_2=\mathrm{Cov}-2\,\mathrm{Cov}_1=\left(\frac{1+F_i}{2}\right)^2 V_D. \qquad (6)$$

Solution and particular case examples

The bias introduced by using S_j's (individual members of S_i lines) to estimate V_A and V_D is determined by considering covariances. The contributions of epistatic components to estimates of V_A and V_D are included in the Appendix.

We assume there are m_1 individuals of a particular S_i line. S_j is one of these individuals. When this S_i line is crossed to several other S_i lines, members of the resulting progeny are half-siblings with the S_i line in question as the common parent. Among the half-siblings, $\frac{1}{m_1}$ have a single S_j plant as their common parent and the remainder are from different S_j plants in the S_i line. The covariance between half-sibs with a common S_j would be $\left(\frac{1+F_j}{4}\right)V_A$ and its probability of appearing in a sample of two individuals from the half-sib family would be $\frac{1}{m_1}$. The covariance for the other types of half-siblings in the family would be $\left(\frac{1+F_i}{4}\right)V_A$ with a probability of $\frac{m_1-1}{m_1}$. Thus, the covariance of half sibs (Cov_1) would be

$$\begin{aligned}\mathrm{Cov}_1 &= 1/4\,V_A\left\{\left(\frac{m_1-1}{m_1}\right)(1+F_i)+\frac{1}{m_1}(1+F_j)\right\}\\ &= 1/4\,V_A\left\{(1+F_i)+\frac{(F_j-F_i)}{m_1}\right\}. \qquad (7)\end{aligned}$$

In the case of full-siblings, there are m_2 members of each S_i line so the probability that two full-siblings have the same S_j parent is

$$\begin{aligned}\frac{1}{m_2}\text{ and }\widehat{\mathrm{Cov}} &= \frac{1}{m_2}\left\{\left(\frac{1+F_j}{2}\right)V_A+\left(\frac{1+F_j}{2}\right)^2V_D\right\}\\ &\quad+\left(1-\frac{1}{m_2}\right)\left\{\left(\frac{1+F_i}{2}\right)V_A+\left(\frac{1+F_i}{2}\right)^2V_D\right\}\\ &=\left\{\left(\frac{1+F_i}{2}\right)V_A+\left(\frac{1+F_i}{2}\right)^2V_D\right\}\\ &\quad+\frac{1}{m_2}\left\{\left(\frac{F_j-F_i}{2}\right)V_A\right.\\ &\quad\left.+\frac{V_D}{4}(F_j-F_i)(2+F_i+F_j)\right\}. \qquad (8)\end{aligned}$$

Thus, $\widehat{\mathrm{Cov}}-\widehat{\mathrm{Cov}}_1-\widehat{\mathrm{Cov}}_2=\widehat{\mathrm{Cov}}-2\,\widehat{\mathrm{Cov}}_1$ (because $\widehat{\mathrm{Cov}}_1=\widehat{\mathrm{Cov}}_2$) =

$$\begin{aligned}&\left\{\left(\frac{F_j-F_i}{2}\right)\left(\frac{1}{m_2}-\frac{1}{m_1}\right)V_A\right\}+1/4\,(1+F_i)^2V_D\\ &\quad+\frac{1}{4m_2}(F_j-F_i)(2+F_i+F_j)V_D. \qquad (9)\end{aligned}$$

The estimate of additive genetic variance ($\hat{V}_A$) can be found by substituting $\widehat{\mathrm{Cov}}_1$ for Cov_1 in equation (4), and then $V_A=\frac{4}{1+F_i}\widehat{\mathrm{Cov}}_1$ and using equation (7)

$$\hat{V}_A=V_A\left\{1+\left(\frac{1}{1+F_i}\right)\left(\frac{F_j-F_i}{m_1}\right)\right\}\text{ and the part in brackets is bias}. \qquad (10)$$

The estimate of dominance variance ($\hat{V}_D$) can be found from equation (6) using $\widehat{\mathrm{Cov}}$, $\widehat{\mathrm{Cov}}_1$, and $\widehat{\mathrm{Cov}}_2$ instead of Cov, Cov_1, and Cov_2. We let $\widehat{\mathrm{Cov}}_1=\widehat{\mathrm{Cov}}_2$ so that $\hat{V}_D=\frac{4}{(1+F_i)^2}(\widehat{\mathrm{Cov}}-2\,\widehat{\mathrm{Cov}}_1)$, and using (6) and (9)

$$\begin{aligned}\hat{V}_D&=\frac{4}{(1+F_i)^2}\left\{\left(\frac{F_j-F_i}{2}\right)\left(\frac{1}{m_2}-\frac{1}{m_1}\right)\right\}V_A\\ &\quad+1/4\,(1+F_i)^2V_D+\frac{1}{4m_2}(F_j-F_i)(2+F_i+F_j)V_D\\ &=V_D+2\frac{(F_j-F_i)}{(1+F_i)^2}\left(\frac{1}{m_2}-\frac{1}{m_1}\right)V_A\\ &\quad+\frac{1}{m_2}\frac{(F_j-F_i)}{(1+F_i)^2}(2+F_i+F_j)V_D. \qquad (11)\end{aligned}$$

In the case of S_0 plants which are non-inbred, $F_i=0$, $F_j=\frac{1}{2}$, and equations (10) and (11) simplify to

$$\hat{V}_A=V_A\left(1+\frac{1}{2m_1}\right) \qquad (12)$$

and

$$\hat{V}_D=V_D+\left(\frac{1}{m_2}-\frac{1}{m_1}\right)V_A+\frac{5}{4m_2}V_D. \qquad (13)$$

Thus, it is clear that estimates of V_A and V_D are biased upwards with small samples of S_1 plants used to represent the S_0 individuals from which the S_1 lines are derived by self-pollination. It is interesting to note that $\hat{V}_D$ includes part of the additive genetic variance.

If S_0 plants were inbred ($F_i \neq 0$) the effect of small numbers of S_1 plants representing S_0 individuals would be reduced and less bias would occur.

If a single individual was used to represent each S_1 line from each S_0 plant ($m_1=m_2=1$), then from (12) $\hat{V}_A=(1+\frac{1}{2})V_A=(1+F)V_A$, and from (13) $\hat{V}_D=V_D+\frac{5}{4}V_D=(1+F)^2V_D$. Thus, bias due to V_A disappears in this case. The same result would be obtained if an infinite number of individuals were used to represent each S_1 when S_0 individuals were inbred to the 0.5 level (or when S_2 lines were used to represent S_1 individuals which represented S_0 individuals).

If m_1 and m_2 vary between families, arguments of King and Henderson (1954) can be used to obtain average values of m_1 and m_2. There are at least two cases that need to be considered: 1) When some S_1 plants are used in more than one mating per full-sib family, $\frac{1}{m_2}$ should be rapladed by $\frac{\sum n_i^2}{(\sum n_i)^2}$, where n_i is the number of times each S_1 plant is used. This is the probability that two full-siblings come from mating the same S_1 individual. If we let P_1 be the proportion of a half-sib family contributed by any S_1 plant, then $\frac{1}{m_1}$ can be replaced by $\frac{\sum P_i^2}{(\sum P_i)^2}$; (2) when m_1 and m_2 vary between S_0 individuals in the same set of the Design II or among sets, then m_1 and m_2 are replaced by the harmonic means of their values.

Discussion

The utilization of several S_1 plants to represent an S_0 plant in matings allows the Design II system to compute variance components for species such as single-eared corn where the Design II could not be used normally. Comstock and Robinson (1952) noted that this design gives a more precise estimate of dominance genetic variance than Design I.

Unfortunately, there will be a bias in estimates of additive and dominance variance. This can be reduced by using larger numbers of S_1 plants to represent each S_0. For a constant relationship between m_1 and m_2, the coefficients of V_A and V_D in their overestimate will be inversely proportional to the number of S_1 plants used to represent each S_0. The bias decreases fairly rapidly until m_2 is about 10 and then the reduction slows considerably. It would seem reasonable to have at least 10 S_1's represent each S_0. If a record of the number of plants used to represent each S_0 is kept, a weighted estimate of m_1 and m_2 can be obtained and the bias estimated. This would enable a more accurate estimate of V_A and V_D. The estimate of V_D from this method will not be completely independent of V_A, as is normally the case with a Design II. However, the increased precision in estimating V_D relative to a Design I will compensate for this deficiency. The degree of dominance will be overestimated, when the biases in estimates of V_A and V_D are not considered. The bias in the coefficients of V_A and V_D will be less when the initial parents are partially inbred. Thus, use of an S_4 to represent an S_3 will make less difference than use of an S_1 for an S_0. However, the coefficients of higher terms in the epistatic variance will be greater where S_0 plants are inbred.

Conclusion

The use of S_1 plants to represent S_0 parents in a Design II mating system will lead to overestimates of additive and dominance genetic variance.

Recording the number of S_1 plants used for each S_0 enables the estimation of these biases and hence their correction. When this is done the Design II becomes a valuable tool for estimating components of variance.

Appendix

The coefficients of epistatic components to estimates of additive and dominance genetic variance can be calculated by extension of the same arguments, as in the body of this paper, to the epistatic components given by Cockerham (1956). The arguments are as follows:

(a) The coefficient (C_{AA}) of V_{AA} (additive × additive variance) in $\hat{V}_A$ (estimate of additive genetic variance) is given by

$$C_{AA} = \left(\frac{1+F_i}{4}\right) + \left(\frac{1}{16\,m_1}\right)\frac{(F_j - F_i)\,(2+F_i+F_j)}{(1+F_i)}.$$

The first term is the coefficient of V_{AA} in a Design II without sampling bias and the second term is the extra contribution from using S_1's to represent an S_0.

(b) The coefficient (C'_{AA}) of V_{AA} in $\hat{V}_D$ (estimate of dominance genetic variance) is given by

$$C'_{AA} = 1/2 + \frac{1}{(1+F_i)^2}(F_j - F_i) \cdot (2+F_i+F_j)\left(\frac{1}{m_2} - \frac{1}{2\,m_1}\right).$$

In this case, the first term represents the coefficient of V_{AA} in the estimate of V_D from a Design II not containing sampling bias while the second term represents the extra contribution from using S_1's to represent an S_0.

(c) Coefficient (C'_{AD}) of V_{AD} in $\hat{V}_D$.

$$C'_{AD} = \frac{1+F_i}{2} + \frac{1}{2\,m_2(1+F_i)^2}\{(1+F_j)^3 - (1+F_i)^3\}.$$

(d) Coefficient (C'_{DD}) of V_{DD} in estimate of V_D

$$C'_{DD} = \left(\frac{1+F_i}{2}\right)^2 + \frac{1}{4\,m_2(1+F_i)^2}\{(1+F_j)^4 - (1+F_i)^4\}.$$

References

Cockerham CC (1956) Analysis of quantitative gene action. In: Genetics in plant breeding. Brookhaven Symp Biol, vol 9, pp 53–68

Comstock RE, Robinson HF (1948) The components of genetic variance in populations. Biometrics 4:254–266

Comstock RE, Robinson HF (1952) Estimation of average dominance of genes. In: Gowen JW (ed) Heterosis. Iowa State Press, Ames, Iowa, pp 494–516

Hammond JJ, Gardner CO (1974) Effect of genetic sampling technique on variation within populations derived by crossing, selfing, or random mating other populations. Crop Sci 14:63–66

King SC, Henderson CR (1954) Variance components analysis in heritability studies. Poult Sci 33:147–154