Theor Appl Genet (1988) 76:385—392

Parametric and jackknife confidence interval estimators for two-factor mating design genetic variance ratios*

S. J. Knapp[1] **and W. C. Bridges**[2]

[1] Department of Crop Science, Oregon State University, Corvallis, OR 97331, USA
[2] Experimental Statistics Unit, Clemson University, Clemson, SC 29634, USA

Received December 26, 1986; Accepted August 19, 1987
Communicated by A. R. Hallauer

Summary. Confidence interval estimators have not been defined for dominance to additive genetic variance (ϱ) and average degree of dominance (δ) for the nested, factorial, and backcross mating designs. The objective of this paper was to describe interval estimators for these parameters. Approximate F random variables were defined for expected mean square (EMS) ratios for linear models with one environmental effect. Approximate $1-\alpha$ parametric interval estimators were defined for ϱ and δ using these random variables. Random variables defined for linear models with no environmental effects are not approximately distributed as F random variables because common EMS are involved in the numerators and denominators of the EMS ratios. Delete-one jackknife (jackknife) interval estimators were defined for ϱ and δ for linear models with zero or one environmental effect(s); ln transformed analysis of variance point estimates were used in pseudovalue estimators.

Key words: Random linear models – Analysis of variance – Jackknifing

Introduction

Inferences about variances or variance ratios are frequent in quantitative genetics; however, methods (hypothesis tests and interval estimators) needed for statistical inferences about variances or variance ratios are often lacking. Interval estimators and hypothesis tests have generally been developed more slowly than point estimators. This reflects, in many cases anyway, the intractability or inadequacy of parametric methods.

The focus of our paper is statistical inference methods for genetic variance ratios. Comstock and Robinson (1948, 1952) described the variance ratios we investigated. These are nested (I), factorial (II), and backcross (III) mating design dominance to additive genetic variance ($\varrho = \sigma_d^2/\sigma_a^2$) and average degree of dominance [$\delta = (2\sigma_d^2/\sigma_a^2)^{1/2}$] variance ratios. The assumption of equal allele frequencies ($p_1 = p_2$) was used to define ϱ and δ (Comstock and Robinson 1948, 1952). Variances or confidence intervals have not been defined for these parameters.

Parametric methods have been used to define interval estimators for random linear model parameters similar to ϱ and δ for balanced data. A parametric interval estimator may be defined for a parameter if it is a function of constants and a random variable with a known distribution. A probability statement is defined for the random variable and subsequently transformed to obtain an interval estimator for the parameter.

Graybill (1976) defined parametric interval estimators for some variance ratios. Bogyo and Becker (1963), Knapp et al. (1985), Knapp (1986), and Knapp and Bridges (1987) used parametric methods to define interval estimators for various expected mean square ratios. The common feature of these parameters is that each is a function of an expected mean square (EMS) ratio and constants. There are, nevertheless, numerous parameters that cannot be expressed as functions of an EMS ratio and constants; expected selection response is an example.

A limitation of parametric interval estimators for variances or mean square ratios is that they are not robust (Box 1953; Scheffè 1959); realized coverage probabilities may be greater or less than stated coverage probabilities when data are non-normal. Arvesen (1969) and Miller (1974a) recommend against parametric interval estimators for variances or mean square ratios because of

---

* Oregon Agricultural Experiment Station Technical Paper No. 8067

the robustness problem. They instead recommend jackknife interval estimators. Arvesen (1969), Arvesen and Schmitz (1970), and Miller (1968, 1974a) have demonstrated that realized and stated coverage probabilities are approximately equal for jackknife interval estimators of these parameters for several random effect distributions, i.e., they are robust.

Our objectives were to define parametric and jackknife interval estimators for ϱ and δ for the nested, factorial, and backcross mating designs and zero or one environmental effects(s) experiment designs. Parametric interval estimators were defined for ϱ and δ for linear models with one environmental effect. Jackknife interval estimators were defined for ϱ and δ for linear models with zero or one environmental effect(s).

Confidence interval estimators

Nested mating design

The analysis of variance for the nested mating design and replications-in-incomplete blocks experiment design in one environment is given in Table 1. The relationships between males and females nested in males variances, genetic variances, and EMS are

$$\sigma_m^2 = [E(M_m) - E(M_f)]/rf = COV(HS) = 1/4\,\sigma_a^2$$

and

$$\sigma_f^2 = [E(M_f) - E(M_e)]/r = COV(FS) - COV(HS) = 1/4\,\sigma_a^2 + 1/4\,\sigma_d^2,$$

respectively, where σ_a^2 is additive genetic variance, σ_d^2 is dominance genetic variance, COV (HS) is the covariance among half-sib progenies, and COV (FS) is the covariance among full-sib progenies (Comstock and Robinson 1952). Additive and dominance genetic variances are $\sigma_a^2 = 4\,\sigma_m^2$ and $\sigma_d^2 = 4(\sigma_f^2 - \sigma_m^2)$, respectively. Epistatic variances were omitted for simplicity in these expressions and others in this paper.

The males to females nested in males variance ratio expressed in terms of EMS is

$$\begin{aligned}\sigma_m^2/\sigma_f^2 &= \{[E(M_m) - E(M_e)]/[E(M_f) - E(M_e)] - 1\}/f \\ &= [E(M'_n)/E(M'_d) - 1]/f \qquad (1) \\ &= \{[E(M_m) - E(M_f)]/[E(M_f) - E(M_e)]\}/f \\ &= [E(M''_n)/E(M''_d)]/f \qquad (2)\end{aligned}$$

σ_m^2/σ_f^2 was expressed in terms of two EMS ratios $E(M'_n)/E(M'_d)$ and $[E(M''_n)/E(M''_d)]$. There are no apparent advantages to using (1) versus (2), therefore, additional results are presented using (1) only.

Dominance to additive genetic variance is

$$\sigma_d^2/\sigma_a^2 = \{[E(M'_n)/E(M'_d) - 1]/f\}^{-1} - 1. \qquad (3)$$

$R = [M'_n/E(M'_n)]/[M'_d/E(M'_d)]$ is the random variable required for defining a parametric interval for (3). M'_n and M'_d in (3) are mean square differences, therefore, approximate degrees of freedom for M'_n and M'_d (df'_n and df'_d, respectively) may be estimated using standard methods (Satterthwaite 1946). The random variables $U'_n = df'_n M'_n/E(M'_n)$ and $U'_d = df'_d M'_d/E(M'_d)$ may be approximately distributed chi-square (Graybill 1976). U_n and U_d are not, however, always approximately distributed as chi-square random variables because they involve mean square differences (Gaylor and Hopper 1969).

An additional problem is that U'_n and U'_d are not approximately distributed as independent chi-square random variables. M_e and $E(M_e)$ are involved in U'_n and U'_d, therefore, U'_n and U'_d are not independent and R is not approximately distributed as an *F* random variable (Graybill 1976). We did not define parametric interval estimators for (3) for these reasons.

Average degree of dominance expressed in terms of EMS and constants is

$$(2\,\sigma_d^2/\sigma_a^2)^{1/2} = [2(\{[E(M'_n)/E(M'_d) - 1]/f\}^{-1} - 1)]^{1/2}. \qquad (4)$$

Equation (4) uses the same EMS ratio used to define (3), therefore, parametric interval estimators were not defined for (4) either.

Parametric interval estimators were not defined in the aforementioned cases because necessary distributional assumptions about U'_n and U'_d were not met. Delete-one jackknifing, on the other hand, does not require distributional assumptions about U'_n and U'_d (Miller 1968) and may be used to define interval estimators for (3) and (4).

It is sufficient to define an interval estimator for $E(M'_n)/E(M'_d)$ in order to define interval estimators for (3) and (4), therefore, $E(M'_n)/E(M'_d)$ was used in the description of jackknifing. The parameters we investigated differ slightly from the parameters investigated by Miller (1968) and Arvesen and Schmitz (1970) in that they are functions of ratios of EMS differences. Arvesen and Schmitz (1970) and Miller (1968) demonstrated that it is necessary to use ln transformed analysis of variance (ANOVA) mean square difference or ratio estimates in pseudovalue estimates to obtain accurate jackknife interval estimates. The distributions of mean square differences and ratios are skewed, thus, requiring a 'variance stabilizing' transformation (Arvesen and Schmitz 1970; Miller 1968). We used ln transformed ANOVA estimates, $\ln(M'_n/M'_d)$, to define delete-one jackknife interval estimators for $E(M'_n)/E(M'_d)$ for this reason.

Delete-one jackknifing involves dividing the original data set into subsets of size one (all replications of one treatment in one-factor ANOVA), deleting a subset, calculating point estimates from the remaining data, and iterating the process until each size one subset has been deleted once. This process has been described for the one-factor experiment design (Miller 1974a).

Table 1. Analysis of variance for the nested mating design and replications-in-incomplete blocks experiment design in one environment

Source of variation	Degrees of freedom[a]	Mean square	Expected mean square[b]
Incomplete blocks (B)	$b-1$		
Replications: B	$b(r-1)$		
Males: B (M:B)	$df_m = b(m-1)$	M_m	$\sigma_e^2 + r\sigma_f^2 + rf\sigma_m^2$
Females: M:B	$df_f = bm(f-1)$	M_f	$\sigma_e^2 + r\sigma_f^2$
Residual	$df_e = b(r-1)(mf-1)$	M_e	σ_e^2

[a] r is the number of replications, m is the number of male and f the number of female parents used to produce nested mating design progeny for one incomplete block, and b is the number of incomplete blocks

[b] σ_e^2 is the residual, σ_f^2 is the females nested in males, and σ_m^2 is the males variance

Miller (1974a) subsetted data into sets containing all replications of a factor level for the one-factor experiment design. This produces a subset number equal to the factor level number. Miller (1974a) considered this to be the least arbitrary subsetting strategy.

The method used to resample the original data set is important in jackknifing. The optimum method is not clear for experiment designs more complex than the one-factor experiment design. Data resampling decisions are affected by the parameter of interest and mating and experiment designs. Data resampling alternatives for complex analyses, using the Table 1 analysis as an example, are considered below.

The number of pseudovalues in a particular delete-one jackknifing problem is equal to the number of data subsets and determines the degrees of freedom for *t*-percentage points used in intervals (Miller 1974a). Data subsetting strategies resulting in balanced resampled data may give pseudovalue numbers different than strategies resulting in unbalanced resampled data. Consider an experiment where nested mating design progenies are grown in a completely randomized experiment design. Suppose $m = 50$ and $f = 3$ where m is the number of male parents and f is the number of female parents. Pseudovalue number is greater when data are subdivided into subsets consisting of all replications of one full-sib family ($mf = 150$ pseudovalues) instead of all replications of one half-sib family ($m = 50$ pseudovalues); however, unbalanced data results when a subset consisting of r replications of one full-sib family is deleted while balanced data results when a subset consisting of r replications of one half-sib family is deleted.

Experiment design also complicates data resampling. Consider the Table 1 incomplete blocks experiment design. Suppose that $m = 5$, $f = 3$, and $b = 10$ where b is the number of incomplete blocks. This gives $m = 5$ half-sib families and $(m \times f) = 15$ full-sib families per incomplete block. A balanced data resampling strategy, where data subsets consist of all replications of one half-sib family from each incomplete block, results in $m = 5$ pseudovalues. Compare this to a strategy where subsets consist of all replications of one full-sib family from one incomplete block. This strategy results in $mfb = 150$ pseudovalues and is analogous in an important way to the resampling strategy used by Miller (1974a) for the one-factor experiment design. They are analogous because pseudovalue number is equal to full-sib family number in our example while pseudovalue number was equal to factor level number in the one-factor linear model (Miller 1974a). This is important because simulations verifying the accuracy of jackknife interval estimators were based on equivalence between pseudovalue and factor level numbers (Arvesen and Schmitz 1970; Miller 1968).

The problem of developing a data sampling (or resampling) plan for complex experiment designs is not unique to delete-one jackknifing. Felsenstein (1985) described similar problems in applying bootstrapping to phylogeny interval estimation. Bootstrap and jackknife variance estimation theory has naturally been developed using simple models (Efron 1979; Efron and Gong 1983; Wu 1986). The extension of bootstrapping and jackknifing to complex data structures is not always clear. The challenge we anticipate is in developing accurate interval estimators for complex data structures. This is not to say that the problem of jackknife or bootstrap variance estimation is trivial but that it is less complex than interval estimation and hypothesis testing problems.

The problem of alternative resampling strategies has not been examined because previous research has largely been limited to one-factor linear model parameters. It is easy to achieve equivalence between pseudovalue and factor level numbers in the one-factor and other simple experiment designs without resorting to a data resampling plan yielding unbalanced data.

The data resampling strategy used directly affects pseudovalue number (g) and, consequently, the value of $t_{\alpha:g-1}$ used in jackknife interval estimators. Jackknife interval estimate width decreases as $t_{\alpha:g-1}$ decreases and $t_{\alpha:g-1}$ decreases as g increases. Interval estimate accuracy depends on how resampling is done. The rationale for the data subsetting strategy we suggest, where pseudovalue and full-sib family or factor level numbers are equal, is this equivalence.

The difficulty introduced by the proposed subsetting strategy is that resampled data is unbalanced. We suggest using fitting constants estimation (Searle 1971) in this case and assume that jackknife variance and interval estimators are accurate for general unbalanced linear model problems; accuracy has been verified for one-factor linear model problems (Arvesen and Layard 1971; Hinkley 1977; Miller 1974b). The accuracy assumption seems reasonable in view of the minor unbalance associated with the proposed data resampling plan.

Suppose that jackknifing is performed using the proposed data resampling plan. There are, for the nested mating design, m half-sib families, m f full-sib families, and r replications in each incomplete block (Table 1). Data are divided into $\mathrm{m\,f\,b} = \mathrm{g}$ subsets containing r replications of one full-sib family from one incomplete block. The ANOVA estimate of $\ln[\mathrm{E}(\mathrm{M}'_n)/\mathrm{E}(\mathrm{M}'_d)]$, estimated from the entire data set, is denoted $\mathrm{J_g}$. The fitting constants (Searle 1971) estimate of $\ln[\mathrm{E}(\mathrm{M}'_n)/\mathrm{E}(\mathrm{M}'_d)]$, estimated from data remaining after the deletion of the ith subset, is denoted J^i_{g-1}. Pseudovalue estimates are equal to $\mathrm{J_i} = \mathrm{g\,J_g} - (\mathrm{g}-1)\,\mathrm{J}^i_{g-1}$. They are calculated for each resampled data set. This process is repeated until each subset has been deleted once, thus, producing g pseudovalue estimates ($\mathrm{J_i}$). The pseudovalue mean (jackknife point estimate) for $\ln[\mathrm{E}(\mathrm{M}'_n)/\mathrm{E}(\mathrm{M}'_d)]$ is $\Sigma\,\mathrm{J_i}/\mathrm{g} = \mathrm{J}$ (Arvesen and Schmitz 1970; Miller 1974a).

The variance of $\mathrm{J_i}$ is $\mathrm{S}^2_J = [\Sigma(\mathrm{J_i}-\mathrm{J})^2]/(\mathrm{g}-1)$, therefore,

$$t = \{\mathrm{J} - \ln[\mathrm{E}(\mathrm{M}'_n)/\mathrm{E}(\mathrm{M}'_d)]\}/(\mathrm{S_J}/\mathrm{g}^{1/2})$$

is an approximate t random variable with $\mathrm{g}-1$ degrees of freedom (Arvesen and Schmitz 1970). An approximate $1-\alpha$ interval estimator for $\ln[\mathrm{E}(\mathrm{M}'_n)/\mathrm{E}(\mathrm{M}'_d)]$ is

$$\mathrm{J} - t_{\alpha:g-1}\,\mathrm{S_J}/\mathrm{g}^{1/2},\ \mathrm{J} + t_{\alpha:g-1}\,\mathrm{S_J}/\mathrm{g}^{1/2} = \mathrm{J} - t\,\mathrm{S}^*_J,\ \mathrm{J} + t\,\mathrm{S}^*_J \quad (5)$$

where $t = t_{\alpha:g-1}$ is a percentage point from the t-distribution with $\mathrm{g}-1$ degrees of freedom and $\mathrm{S}^*_J = \mathrm{S_J}/\mathrm{g}^{1/2}$ is the standard error of a pseudovalue mean (Arvesen and Schmitz 1970; Miller 1974a).

An approximate $1-\alpha$ jackknife interval estimator for (3) (dominance to additive genetic variance) from (5) is

$$[(\mathrm{e}^{(\mathrm{J}+t\mathrm{S}^*_j)}-1)/\mathrm{f}]^{-1}-1,\quad [(\mathrm{e}^{(\mathrm{J}-t\mathrm{S}^*_j)}-1)/\mathrm{f}]^{-1}-1 \qquad (6)$$

where e is the base of the natural logarithm. An approximate $1-\alpha$ jackknife interval estimator for (4) (average degree of dominance) is

$$(2\{[(\mathrm{e}^{(\mathrm{J}+t\mathrm{S}^*_j)}-1)/\mathrm{f}]^{-1}-1\})^{1/2},$$
$$(2\{[(\mathrm{e}^{(\mathrm{J}-t\mathrm{S}^*_j)}-1)/\mathrm{f}]^{-1}-1\})^{1/2}. \qquad (7)$$

The usefulness of jackknifing is apparent from interval estimators (6) and (7). First, jackknifing provides a solution to the interval estimator problem where no obvious or adequate parametric solution was found. Second, the estimation process is conceptually simple.

We investigated, in addition to the above model (Table 1), a one environmental effect linear model (Table 2). Males and females nested in males variances are

$$\sigma^2_m = [\mathrm{E}(\mathrm{M_m}) + \mathrm{E}(\mathrm{M_{fs}}) - \mathrm{E}(\mathrm{M_f}) - \mathrm{E}(\mathrm{M_{ms}})]/\mathrm{r\,f\,s}$$

and

$$\sigma^2_f = [\mathrm{E}(\mathrm{M_f}) - \mathrm{E}(\mathrm{M_{fs}})]/\mathrm{r\,s},$$

respectively. The males to females variance ratio (σ^2_m/σ^2_f) expressed in terms of EMS and constants is

$$\begin{aligned}&\{[\mathrm{E}(\mathrm{M_m}) + \mathrm{E}(\mathrm{M_{fs}}) - \mathrm{E}(\mathrm{M_f}) - \mathrm{E}(\mathrm{M_{ms}})]/\mathrm{r\,f\,s}\}/\\ &\qquad\{[\mathrm{E}(\mathrm{M_f}) - \mathrm{E}(\mathrm{M_{fs}})]/\mathrm{r\,s}\}\\ &\qquad= \{[\mathrm{E}(\mathrm{M_m}) - \mathrm{E}(\mathrm{M_{ms}})]/[\mathrm{E}(\mathrm{M_f}) - \mathrm{E}(\mathrm{M_{fs}})] - 1\}/\mathrm{f}\\ &\qquad= [\mathrm{E}(\mathrm{M}^*_n)/\mathrm{E}(\mathrm{M}^*_d) - 1]/\mathrm{f},\end{aligned}$$

Table 2. Analysis of variance for the nested mating design and replications-in-incomplete blocks experiment design in more than one location

Source of variation	Degrees of freedom[a]	Mean squares	Expected mean squares[b]
Incomplete blocks (B)	$b-1$		
Locations (L)	$s-1$		
L × B	$(b-1(s-1)$		
Replications: L:B	$b\,s(r-1)$		
Males (M:B)	$df_m = b(m-1)$	M_m	$\sigma^2_e + r\,\sigma^2_{fs} + r\,s\,\sigma^2_f + r\,f\,\sigma^2_{ms} + r\,f\,s\,\sigma^2_m$
Females: M(F:M:B)	$df_f = b\,m(f-1)$	M_f	$\sigma^2_e + r\,\sigma^2_{fs} + r\,s\,\sigma^2_f$
M: B × L	$df_{ms} = b(s-1)(m-1)$	M_{ms}	$\sigma^2_e + r\,\sigma^2_{fs} + r\,f\,\sigma^2_{ms}$
F:M:B × L	$df_{fs} = b(s-1)\,m(f-1)$	M_{fs}	$\sigma^2_e + r\,\sigma^2_{fs}$
Residual	$b\,s(r-1)(m\,f-1)$		

[a] s is the number of locations, and b is the number of incomplete blocks, r is the number of replications, and m is the number of male and f the number of female parents used to produce nested mating design progenies for one incomplete block

[b] σ^2_m is the male parent, σ^2_f is the female parent nested in male parent, σ^2_{ms} is the male parent × location interaction, σ^2_{fs} is the female parent nested in male parent × location interaction, and σ^2_e is the residual variance

thus, ϱ is

$$(\sigma_f^2 - \sigma_m^2)/\sigma_m^2 = \sigma_f^2/\sigma_m^2 - 1 = \{[E(M_n^*)/E(M_d^*) - 1]/f\}^{-1} - 1 \tag{8}$$

where $E(M_n^*)$ and $E(M_d^*)$ are linear functions of EMS. Denote the respective approximate degrees of freedom for M_n^* and M_d^* (mean square differences corresponding to $E(M_n^*)$ and $E(M_d^*)$) df_n^* and df_d^*, respectively; where df_n^* and df_d^* are estimated with standard methods (Satterthwaite 1946).

Assume that the random variable $S = [M_n^*/E(M_n^*)]/[M_d^*/E(M_d^*)]$ is approximately distributed as an *F* random variable then

$$P(F_{1-\alpha/2:df_n^*, df_d^*} \leqq S \leqq F_{\alpha/2:df_n^*, df_d^*}) \cong 1 - \alpha$$

and an approximate $1-\alpha$ interval estimator for (8) (dominance to additive genetic variance) is

$$[(F_{\alpha/2:df_n^*, df_d^*} M_n^*/M_d^* - 1)/f]^{-1} - 1,$$
$$[(F_{1-\alpha/2:df_n^*, df_d^*} M_n^*/M_d^* - 1)/f]^{-1} - 1. \tag{9}$$

Average degree of dominance is

$$[2(\{[E(M_n^*)/E(M_d^*) - 1]/f\}^{-1} - 1)]^{1/2}, \tag{10}$$

thus, an approximate $1-\alpha$ parametric interval estimator for (10) is

$$(2\{[(F_{\alpha/2:df_n^*, df_d^*} M_n^*/M_d^* - 1)/f]^{-1} - 1\})^{1/2},$$
$$(2\{[(F_{1-\alpha/2:df_n^*, df_d^*} M_n^*/M_d^* - 1)/f]^{-1} - 1\})^{1/2}. \tag{11}$$

Random variable S involves different mean squares and EMS in the numerator and denominator. It was assumed, for the purpose of defining interval estimators (9) and (11), that $df_n^* M_n^*/E(M_n^*)$ and $df_d^* M_d^*/E(M_d^*)$ are approximately distributed as independent chi-square random variables and that relevant random effects are normally distributed. The parametric interval estimators are, however, non-robust (Box 1953; Scheffè 1959) and differences between chi-square random variables may not be distributed as chi-square random variables (Gaylor and Hopper 1969). Realized coverage probabilities for interval estimators (9) and (11) may be different, perhaps greatly different, from stated coverage probabilities when either of these assumptions are not met (Arvesen 1969; Arvesen and Schmitz 1970; Miller 1974a). Thus, jackknifing is justified in this case even though simple parametric interval estimators are available for (9) and (11).

Jackknifing for the one environmental effect linear model proceeds exactly as for the no environmental effect linear model except that it is suggested that data resampling consist of deleting all replications and environments of one full-sib family from one incomplete block. The main difference in this case is that $\ln(M_n^*/M_d^*)$ is jackknifed instead of $\ln(M'_n/M'_d)$; thus, pseudovalue means and variances have interpretations identical to those in (6) and (7) but for a different EMS ratio.

An approximate $1-\alpha$ jackknife interval estimator for (8), jackknifing $\ln(M_n^*/M_d^*)$, is

$$[(e^{(J+tS_J^*)} - 1)/f]^{-1} - 1, \quad [(e^{(J-tS_J^*)} - 1)/f]^{-1} - 1 \tag{12}$$

and for (10) is

$$\{2[(e^{(J+tS_J^*)} - 1)/f]^{-1} - 1\}^{1/2},$$
$$\{2[(e^{(J-tS_J^*)} - 1)/f]^{-1} - 1\}^{1/2}. \tag{13}$$

Factorial mating design

Results are given first for the zero environmental effects linear model (Table 3) and second for the one environmental effect linear model (Table 4) for the factorial mating design. The males variance was used in all expressions, however, the females or males and females pooled variances may also be used.

Dominance to additive genetic variance expressed in terms of EMS is

$$\begin{aligned}\sigma_d^2/\sigma_a^2 &= \sigma_{mf}^2/\sigma_m^2 \\ &= (\{[E(M_m) - E(M_e)]/[E(M_{mf}) - E(M_e)] - 1\}/f)^{-1} \\ &= (\{[E(M_f) - E(M_{mf})]/[E(M_{mf}) - E(M_e)]\}/f)^{-1} \\ &= \{[E(M'_n)/E(M'_d)]/f\}^{-1}.\end{aligned} \tag{14}$$

Table 3. Analysis of variance for the factorial mating design and replications-in-incomplete blocks experiment design in one environment

Source of variation	Degrees of freedom[a]	Mean square	Expected mean square[b]
Incomplete blocks (B)			
Replications/B			
Males/B	$df_m = b(m-1)$	M_m	$\sigma_e^2 + r\sigma_f^2 + rf\sigma_m^2$
Females/B	$df_f = b(f-1)$	M_f	$\sigma_e^2 + r\sigma_f^2 + rm\sigma_f^2$
(Males × females)/B	$df_{mf} = b(m-1)(f-1)$	M_{mf}	$\sigma_e^2 + r\sigma_f^2$
Residual	$df_e = b(r-1)(mf-1)$	M_e	σ_e^2

[a] b is the number of incomplete blocks, r is the number of replications, m is the number of males, and f is the number of females

[b] σ_e^2 is the residual, σ_{mf}^2 is the males × females nested in incomplete blocks, σ_f^2 is the females nested in incomplete blocks, and σ_m^2 is the males nested in incomplete blocks variance

Table 4. Analysis of variance for the factorial mating design and replications-in-incomplete blocks experiment design in more than one year

Source of variance	Degrees of freedom[a]	Mean square	Expected mean squares[b]
Incomplete blocks (B)	$b-1$		
Years (Y)	$y-1$		
Y × B	$(b-1)(y-1)$		
Replications: Y:B	$yb(r-1)$		
Males: B(M:B)	$df_m = b(m-1)$	M_m	$\sigma_e^2 + r\sigma_{mfy}^2 + rf\sigma_{my}^2 + ry\sigma_{mf}^2 + ryf\sigma_m^2$
Females: B(F:B)	$df_f = b(f-1)$	M_f	$\sigma_e^2 + r\sigma_{mfy}^2 + rm\sigma_{fy}^2 + ry\sigma_{mf}^2 + rym\sigma_f^2$
(M × F):B	$df_{mf} = b(m-1)(f-1)$	M_{mf}	$\sigma_e^2 + r\sigma_{mfy}^2 + ry\sigma_{mf}^2$
M: B × Y	$df_{my} = b(m-1)(y-1)$	M_{my}	$\sigma_e^2 + r\sigma_{mfy}^2 + rf\sigma_{my}^2$
F: B × Y	$df_{fy} = b(f-1)(y-1)$	M_{fy}	$\sigma_e^2 + r\sigma_{mfy}^2 + rm\sigma_{fy}^2$
(M × F): B × Y	$df_{mfy} = b(m-1)(f-1)(y-1)$	M_{mfy}	$\sigma_e^2 + r\sigma_{mfy}^2$
Residual	$yb(r-1)(mf-1)$		

[a] r is the number of replications nested in incomplete blocks, b is the number of incomplete blocks, m is the number of male and f the number of female parents used to produce factorial mating design progenies for one incomplete block, and y is the number of years

[b] σ_m^2 is the male and σ_f^2 the female parent, σ_{mf}^2 is the male × female parent interaction, σ_{my}^2 is the male parent × year and σ_{fy}^2 the female parent × year interaction, σ_{mfy}^2 is the male parent × female parent × year interaction, and σ_e^2 is the residual variance

A parametric interval estimator was not defined for (14) because $E(M_{mf})$ is involved in the numerator and denominator. Jackknife interval estimators may, nevertheless, be defined for (14) based on jackknifing $\ln M'_n/M'_d$). The jackknife point estimate for (14) is $(e^J/f)^{-1}$. An approximate $1-\alpha$ jackknife interval estimator for (14) is

$$[(e^{(J+t_{\alpha:g-1}S_J/g^{1/2})}/f]^{-1},\ [(e^{(J-t_{\alpha:g-1}S_J/g^{1/2})}/f]^{-1}$$
$$=[(e^U/f)^{-1},\quad (e^L/f)^{-1}]$$

where J is the pseudovalue mean and jackknife point estimate and S_J is the pseudovalue variance for $\ln[E(M'_n)/E(M'_d)]$.

The number of pseudovalues used in the calculation of J and S_J depends, as before, on how the original data are subsetted. Balanced data results from eliminating one subset containing all replications of one half-sib family from each incomplete block. If $m=4$, $f=4$, and $b=10$, where m is the number of males, f is the number of females, and b is the number of incomplete blocks, then the balanced data resampling plan produces $m+f=8$ pseudovalues. This is obviously inadequate and inefficient. A data resampling plan that results in subsets consisting of all replications of one full-sib family from one incomplete block gives $mfb=160$ pseudovalues. We suggest, analogous to the nested mating design, using this resampling plan because it produces mfb pseudovalues, is efficient, and results in equivalence between pseudovalue and full-sib family numbers. Fitting constants variance estimation (Searle 1971) may be used to analyze the resampled data.

Average degree of dominance expressed in terms of EMS is

$$(2\{[E(M'_n)/E(M'_d)]/f\}^{-1})^{1/2} \tag{15}$$

where $E(M'_n)$, $E(M'_d)$, and f were defined in (14). A jackknife point estimator for (15) is $[2(e^J/f)^{-1}]^{1/2}$. An approximate $1-\alpha$ jackknife interval estimator for (15) is

$$[2(e^U/f)^{-1}]^{1/2},\quad [2(e^L/f)^{-1}]^{1/2}.$$

We investigated a one environmental effect linear model (Table 4) in addition to the zero environmental effects linear model (Table 3). The males variance was used in all expressions, however, the females or males and females pooled variance may be used.

Dominance to additive genetic variance expressed in terms of EMS is

$$\sigma_d^2/\sigma_a^2 = \sigma_{mf}^2/\sigma_m^2$$
$$=(\{[E(M_m)-E(M_{my})]/[E(M_{mf})-E(M_{mfy})]-1\}/f)^{-1}$$
$$=\{[E(M''_n)/E(M''_d)-1]/f\}^{-1}. \tag{16}$$

$E(M''_n)$ and $E(M''_d)$ do not involve common EMS, therefore, $df''_n M''_n/E(M''_n)$ and $df''_d M''_d/E(M''_d)$ may be approximately distributed as independent chi-square random variables and $T=[M''_n/E(M''_n)]/[M''_d/E(M''_d)]$ may be approximately distributed as an F random variable. An approximate parametric interval estimator for (17) based on T is

$$[(F_{\alpha/2:df''_n,df''_d}M''_n/M''_d-1)/f]^{-1},$$
$$[(F_{1-\alpha/2:df''_n,df''_d}M''_n/M''_d-1)/f]^{-1}.$$

Average degree of dominance expressed in terms of EMS is

$$(2\{[E(M''_n)/E(M''_d)-1]/f\}^{-1})^{1/2}. \tag{17}$$

An approximate $1-\alpha$ parametric interval estimator for (17) is

$$\{2[(F_{\alpha/2:\mathrm{df}''_n,\mathrm{df}''_d}\mathrm{M}''_n/\mathrm{M}''_d-1)/\mathrm{f}]^{-1}\}^{1/2},$$
$$\{2[(F_{1-\alpha/2:\mathrm{df}''_n,\mathrm{df}''_d}\mathrm{M}''_n/\mathrm{M}''_d-1)/\mathrm{f}]^{-1}\}^{1/2}.$$

Jackknife point and interval estimators for (15) and (17) were defined as in previous examples. Data are subsetted so that each subset contains every replication of one full-sib family from one incomplete block and all environments. Resampled data are unbalanced, therefore, fitting constants estimation may be used to obtain point estimates. The jackknifed statistic is $\ln(\mathrm{M}''_n/\mathrm{M}''_d)$. The jackknife point estimate for (15) is $[(\mathrm{e}^{\mathrm{J}}-1)/\mathrm{f}]^{-1}$ where J is the pseudovalue mean. An approximate $1-\alpha$ jackknife interval estimator for (15) is

$$[(\mathrm{e}^{\mathrm{L}}-1)/\mathrm{f}]^{-1},\quad [(\mathrm{e}^{\mathrm{U}}-1)/\mathrm{f}]^{-1}.$$

Jackknife point and interval estimators for (17) are $\{2[(\mathrm{e}^{\mathrm{J}}-1)/\mathrm{f}]^{-1}\}^{1/2}$ and

$$\{2[(\mathrm{e}^{\mathrm{L}}-1)/\mathrm{f}]^{-1}\}^{1/2},\quad \{2[(\mathrm{e}^{\mathrm{U}}-1)/\mathrm{f}]^{-1}\}^{1/2},$$

respectively.

Backcross mating design

The analysis for backcross mating design (mating design III) progenies in a replications-in-incomplete blocks experiment design in one environment is given in Table 5. Dominance to additive genetic variance expressed in terms of EMS is

$$\begin{aligned}\sigma_d^2/\sigma_a^2 &= \sigma_{mp}^2/4\,\sigma_m^2\\ &= \{2[\mathrm{E}(\mathrm{M}_m)-\mathrm{E}(\mathrm{M}_e)]/[\mathrm{E}(\mathrm{M}_{mp})-\mathrm{E}(\mathrm{M}_e)]\}^{-1}\\ &= [2\,\mathrm{E}(\mathrm{M}'_n)/\mathrm{E}(\mathrm{M}'_d)]^{-1}. \qquad (18)\end{aligned}$$

Parametric intervals were not defined for (18) because $\mathrm{E}(\mathrm{M}_e)$ is involved in numerator and denominator EMS differences. An approximate $1-\alpha$ jackknife interval estimator was, however, defined for (18) consistent with the approach used in previous sections. The jackknife point estimate for (18) is $(2\,\mathrm{e}^{\mathrm{J}})^{-1}$. An approximate $1-\alpha$ jackknife interval estimator for (18), based on jackknifing $\ln(\mathrm{M}'_n/\mathrm{M}'_d)$, is

$$\begin{aligned}&[2\,\mathrm{e}^{(\mathrm{J}+t_{\alpha:g-1}\mathrm{S_J}/g^{1/2})}]^{-1},\quad [2\,\mathrm{e}^{(\mathrm{J}-t_{\alpha:g-1}\mathrm{S_J}/g^{1/2})}]^{-1}\\ &\quad=(2\,\mathrm{e}^{\mathrm{U}})^{-1},\quad (2\,\mathrm{e}^{\mathrm{L}})^{-1}\end{aligned}$$

where J is the pseudovalue mean or jackknife point estimate and $\mathrm{S_J}$ is the pseudovalue variance for $\ln(\mathrm{M}'_n/\mathrm{M}'_d)$.

Average degree of dominance expressed in terms of EMS is

$$\{[\mathrm{E}(\mathrm{M}'_d)/\mathrm{E}(\mathrm{M}'_n)]^{-1}\}^{1/2}. \qquad (19)$$

The jackknife point estimate for (19) is $[(\mathrm{e}^{\mathrm{J}})^{-1}]^{1/2}$. An approximate $1-\alpha$ jackknife interval estimator for (19) is

$$[(\mathrm{e}^{\mathrm{U}})^{-1}]^{1/2},\quad [(\mathrm{e}^{\mathrm{L}})^{-1}]^{1/2}.$$

We investigated a one environmental effect linear model for the backcross mating design (Table 6) in addition to the zero environmental effects linear model (Table 5). Dominance to additive genetic variance expressed in terms of EMS is

$$\begin{aligned}\sigma_d^2/\sigma_a^2 &= \sigma_{mp}^2/4\,\sigma_m^2\\ &= \{2[\mathrm{E}(\mathrm{M}_m)-\mathrm{E}(\mathrm{M}_{ml})]/[\mathrm{E}(\mathrm{M}_{mp})-\mathrm{E}(\mathrm{M}_e)]\}^{-1}\\ &= [2\,\mathrm{E}(\mathrm{M}''_n)/\mathrm{E}(\mathrm{M}''_d)]^{-1}. \qquad (20)\end{aligned}$$

$\mathrm{E}(\mathrm{M}''_n)$ and $\mathrm{E}(\mathrm{M}''_d)$ do not involve common EMS, therefore, $\mathrm{df}''_n\,\mathrm{M}''_n/\mathrm{E}(\mathrm{M}''_n)$ and $\mathrm{df}''_d\,\mathrm{M}''_d/\mathrm{E}(\mathrm{M}''_d)$ may be approximately distributed as independent chi-square random variables and $\mathrm{U}=[\mathrm{M}''_n/\mathrm{E}(\mathrm{M}''_n)]/\mathrm{M}''_d/\mathrm{E}(\mathrm{M}''_d)]$ may be approximately distributed as an *F* random variable. An approximate parametric $1-\alpha$ interval estimator for (20) based on U is

$$(2\,F_{\alpha/2:\mathrm{df}''_n,\mathrm{df}''_d}\mathrm{M}''_n/\mathrm{M}''_d)^{-1},\quad (2\,F_{1-\alpha/2:\mathrm{df}''_n,\mathrm{df}''_d}\mathrm{M}''_n/\mathrm{M}''_d)^{-1}.$$

Average degree of dominance expressed in terms of EMS is

$$(2\,\sigma_d^2/\sigma_a^2)^{1/2}=(\sigma_{mp}^2/2\,\sigma_m^2)^{1/2}=\{[\mathrm{E}(\mathrm{M}''_n)/\mathrm{E}(\mathrm{M}''_d)]^{-1}\}^{1/2}. \qquad (21)$$

Table 5. Analysis of variance for the backcross mating design and replications-in-incomplete blocks experiment design in one environment

Source of variation	Degrees of freedom[a]	Mean square	Expected mean square[b]
Incomplete blocks (B)	b − 1		
Replications: B	b(r − 1)		
Inbred Lines (P:B)	b(p − 1)		
Males (M:B)	b(m − 1)	M_m	$\sigma_e^2+\mathrm{pr}\,\sigma_m^2$
(M × P):B	b(m − 1)(p − 1)	M_{mp}	$\sigma_e^2+\mathrm{r}\,\sigma_{mp}^2$
Residual	b(r − 1)(pm − 1)	M_e	σ_e^2

[a] b is the number of incomplete blocks, r is the number of replications, m is the number of males used to produce backcross progenies, and p = 2 is the number of inbred lines

[b] σ_m^2 is the male parent, σ_{mp}^2 is the male parent × inbred line parent interaction, and σ_e^2 is the residual variance

Table 6. Analysis of variance for the backcross mating design and replications-in-incomplete blocks experiment design in more than one location

Source of variation	Degrees of freedom[a]	Mean square	Expected mean square[b]
Incomplete blocks (B)	$b-1$		
Locations (L)	$s-1$		
L × B	$(b-1)(s-1)$		
Replications: L:B	$bs(r-1)$		
Inbred Lines: B(P:B)	$b(p-1)$		
L × P:B	$(s-1)(p-1)$		
Males: (M:B)	$b(m-1)$	M_m	$\sigma_e^2 + r\sigma_{mpl}^2 + rp\sigma_{ml}^2 + rps\sigma_m^2$
(M × P):B	$b(m-1)(p-1)$	M_{mp}	$\sigma_e^2 + r\sigma_{mpl}^2 + rs\sigma_{mp}^2$
L × M:B	$b(m-1)(s-1)$	M_{ml}	$\sigma_e^2 + r\sigma_{mpl}^2 + rp\sigma_{ml}^2$
(L × M × P):B	$b(m-1)(s-1)(p-1)$	M_{mpl}	$\sigma_e^2 + r\sigma_{mpl}^2$
Residual	$bs(r-1)(pm-1)$		

[a] b is the number of incomplete blocks, s is the number of locations, r is the number of replications, m is the number of male parents used to produce backcross lines, and $p = 2$ is the number of inbred lines

[b] σ_m^2 is the male parent, σ_{mp}^2 is the male parent × inbred line parent interaction, σ_{ml}^2 is the male parent × location, σ_{mpl}^2 is the male parent × inbred line parent × location interaction, and σ_e^2 is the residual variance

An approximate $1-\alpha$ parametric interval estimator for (21) is

$$[(F_{\alpha/2:\mathrm{df}''_n,\mathrm{df}''_d}\,M''_n/M''_d)^{-1}]^{1/2},$$
$$[(F_{1-\alpha/2:\mathrm{df}''_n,\mathrm{df}''_d}\,M''_n/M''_d)^{-1}]^{1/2}.$$

Jackknife point and interval estimators for (21), based on jackknifing $\ln(M''_n/M''_d)$, are $[(e^J)^{-1}]^{1/2}$ and

$$[(e^U)^{-1}]^{1/2}, \quad [(e^L)^{-1}]^{1/2},$$

respectively, where J is the pseudovalue mean or jackknife point estimate and S_J is the pseudovalue variance for $\ln(M''_n/M''_d)$.

References

Arvesen JN (1969) Jackknifing U-statistics. Ann Math Statist 40: 2076–2100

Arvesen JN, Layard MWJ (1971) Asymptotically robust tests in unbalanced variance component models. Tech Rep No. 263, Dept Statist, Purdue Univ

Arvesen JN, Schmitz TH (1970) Robust procedures for variance component problems using the jackknife. Biometrics 26: 677–686

Bogyo TP, Becker WA (1963) Exact confidence intervals for genetic heritability estimated from paternal half-sib correlations. Biometrics 19:494–496

Box GEP (1953) Non-normality and tests on variances. Biometrika 40:318–335

Comstock RE, Robinson HF (1948) The components of genetic variance in populations of biparental progenies and their use in estimating the average degree of dominance. Biometrics 4:254–266

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

Efron B (1979) Bootstrap methods: Another look at the jackknife. Ann Statist 7:1–26

Efron B, Gong G (1983) A leisurely look at the bootstrap, the jackknife and cross-validation. Amer Statist 37:36–48

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Gaylor DW, Hopper FN (1969) Estimating the degrees of freedom for linear combinations of mean squares by Satterthwaite's formula. Technometrics 11:691–706

Graybill FA (1976) Theory and application of the linear model. Duxbury Press, Boston

Hinkley DV (1977) Jackknifing in unbalanced situations. Technometrics 19:285–292

Knapp SJ (1986) Confidence intervals for heritability for two-factor mating design single environment linear models. Theor Appl Genet 72:587–591

Knapp SJ, Bridges WC (1987) Confidence interval estimators for heritability for several mating and experiment designs. Theor Appl Genet 73:759–763

Knapp SJ, Stroup WW, Ross WM (1985) Exact confidence intervals for heritability on a progeny mean basis. Crop Sci 25: 192–194

Miller RG (1968) Jackknifing variances. Ann Math Statist 39: 567–582

Miller RG (1974a) The jackknife – a review. Biometrika 61: 1–15

Miller RG (1974b) An unbalanced jackknife. An Statist 2: 880–891

Satterthwaite FE (1946) An approximate distribution of estimates of variance components. Biom Bull 2: 110–114

Scheffè H (1959) The analysis of variance. Wiley, New York

Searle SR (1971) Linear models. Wiley, New York

Wu CFJ (1986) Jackknife, bootstrap and other resampling methods in regression analysis. Ann Statist 14:1261–1295