Theor Appl Genet (1987) 73:759–763

Confidence interval estimators for heritability for several mating and experiment designs*

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

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

Received November 1, 1986; Accepted November 18, 1986
Communicated by A. R. Hallauer

Summary. Confidence interval estimators have not been described for several heritability (H) estimators relevant to recurrent family selection. Previously described H interval estimators do not apply to one-factor mating designs in split-plot in time experiment designs in one or more locations, one-factor mating designs for several experiment designs in two or more locations and years, and two-factor mating designs for several experiment designs in two or more locations or years. Our objective was to derive H interval estimators for these cases. H reduced to a function of constants and a single expected mean square ratio in every case; $H = 1 - E(M')/E(M'')$ where $E(M')$ is a linear function of expected mean squares and $E(M'')$ is a single expected mean square. It was shown that $F' = [M''/E(M'')]/[M'/E(M')]$ has an approximate *F*-distribution with df″ and df′ degrees of freedom, respectively, where M′ and M″ are mean squares corresponding to $E(M')$ and $E(M'')$, respectively. H is a function of F′, therefore, we used F′ to define an approximate $(1-\alpha)$ interval estimator for H.

Key words: Random linear models – Recurrent selection – Forages – Maize

Introduction

Confidence interval estimators for heritability (H) estimators relevant to family selection have been derived for several one- and two-factor mating and experiment designs (Knapp et al. 1985; Knapp 1986). H interval estimators for the one-factor mating design do not apply to split-plot in time experiment designs in one or more locations or other experiment designs in locations and years (Knapp et al. 1985). H interval estimators for two-factor mating designs do not apply to experiment designs in more than one location or year (Knapp 1986). The objective of this research was to derive H interval estimators for these cases.

General results

Heritability for the mating and experiment designs we considered reduced to functions of constants and expected mean square ratios if data were balanced. The general form of H is

$$H = 1 - E(M')/E(M''), \tag{1}$$

where $E(M'')$ is a single expected mean square and $E(M')$ is a linear function of expected mean squares that does not involve $E(M'')$. $E(M')$ is defined as follows:

$$E(M') = \Sigma\, k_i\, E(M_i),$$

where the k_i are known constants and $E(M_i)$ are expected mean squares for $i = 1, 2, \ldots, n$. Probability distribution results used to derive H interval estimators are given below.

Let df_i, M_i and $E(M_i)$ denote degrees of freedom, mean squares, and expected mean squares, respectively, for the effects in the balanced random linear models used to describe the mating and experiment designs we examined. M′ and M″ denote functions of M_i and are estimates of $E(M')$ and $E(M'')$, respectively.

If model effects are normally distributed then the random variable $U_i = df_i M_i / E(M_i)$ has a chi-square

---

* Oregon Agricultural Experiment Station Technical Paper No. 7923

(χ^2) distribution with df_i degrees of freedom (Graybill 1976). M'' involves only one M_i, therefore, the random variable $U^{2''} = df'' M''/E(M'')$ has a χ^2-distribution with df'' degrees of freedom. The distribution of the random variable $U^{2'} = df' M'/E(M')$ is not clear because M' is a linear function of M_i's (Graybill 1976).

Approximate degrees of freedom for M' are obtained from

$$df' = (M')^2 [\Sigma (k_i M_i)^2/df_i]^{-1},$$

where df_i are degrees of freedom for M_i (Satterthwaite 1946). $U^{2'}$ has an approximate χ^2-distribution for certain conditions (Welch 1956; Gaylor and Hopper 1969). If all k_i in M' and $E(M')$ are positive then $U^{2'}$ is approximately distributed χ^2. If some k_i are negative then $U^{2'}$ may not be distributed χ^2 (Gaylor and Hopper 1969). The results of Gaylor and Hopper (1974) relevant to the distribution of $U^{2'}$ with negative k_i are discussed below.

The χ^2 distribution may be a poor approximation of the distribution of $U^{2'}$ because M' may be negative. Gaylor and Hopper (1969) defined groups of mean squares with positive and negative k_i such that

$$M' = M'_P - M'_N,$$

where M'_P and M'_N are the sums of M_i with positive and negative k_i, respectively. Degrees of freedom for M'_P and M'_N are df'_P and df'_N, respectively, and are estimated from Satterthwaite's approximation (1946). The probability of a negative M' is

$$\begin{aligned} P[M'<0] &= P[M'_P/M'_N < 1] \\ &= P[F_{\alpha:\, df'_P,\, df'_N} < E(M'_N)/E(M'_P)], \end{aligned}$$

where $E(M'_P)$ and $E(M'_N)$ are the expected values for M'_P and M'_N, respectively, and $F_{\alpha:\, df'_P,\, df'_N}$ is a value from the F-distribution with df'_P numerator and df'_N denominator degrees of freedom (Searle 1971). It is, of course, desirable for $P(M' < 0)$ to be extremely small.

Gaylor and Hopper (1969) indicated that $U^{2'} = df' M'/E(M')$ is approximately distributed χ^2 if $P(M' < 0)$ is small. We expect $U^{2'}$ to be approximately distributed χ^2 for the mating and experiment designs studied because M' is a sum of several variances, therefore, $P(M' < 0)$ is, on the average, very small. The expected value for M' in (4), for example, is

$$\begin{aligned} E(M') &= E(M_{bf}) + E(M_{fy}) - E(M_e) \\ &= \sigma_e^2 + y\sigma_{bf}^2 + b\sigma_{fy}^2, \end{aligned}$$

where terms are defined in Table 1. $E(M')$ is a sum of at least three variances for the Table 2, 3, and 4 examples also.

If the random variables $U_i = df_i M_i/E(M_i)$ and $U_j = df_j M_j/E(M_j)$ are independent and have χ^2 distributions then the random variable

$$F = (U_i/df_i)/(U_j/df_j) = [M_i/E(M_i)]/[M_j/E(M_j)]$$

has an F-distribution with df_i and df_j degrees of freedom (Graybill 1976). The random variables $U^{2''}$ and $U^{2'}$, described above, are independent and have χ^2 and approximate χ^2 distributions, respectively; therefore, the random variable

$$F' = (U^{2''}/df'')/(U^{2'}/df') = [M''/E(M'')]/[M'/E(M')]$$

has an approximate F-distribution with df'' and df' degrees of freedom (Searle 1971).

Given that F' has an approximate F-distribution then we obtain the following probability statement:

$$\begin{aligned} P\{F_{1-\alpha/2:\, df'', df'} &< [M''/E(M'')]/[M'/E(M')] \\ &< F_{\alpha/2:\, df'', df'}\} \cong 1 - \alpha, \end{aligned} \qquad (2)$$

where $F_{1-\alpha/2:\, df'', df'}$ and $F_{\alpha/2:\, df'', df'}$ denote $(1-\alpha/2)$th and $(\alpha/2)$th quantiles of an F-distribution, respectively,

Table 1. Analysis of variance for a balanced completely random linear model for family evaluation in a perennial crop with measurements repeated over years. The experiment design is a split-plot in time with complete blocks (whole plots) arranged in a randomized complete blocks experiment design

Source of variation	Degrees of freedom[a]	Mean square	Expected mean square[b]
Blocks (B)	$b-1$		
Years (Y)	$y-1$		
Y × B	$(y-1)(b-1)$		
Families (F)	$df_f = f-1$	M_f	$\sigma_e^2 + y\sigma_{bf}^2 + b\sigma_{fy}^2 + yb\sigma_f^2$
B × F	$df_{bf} = (b-1)(f-1)$	M_{bf}	$\sigma_e^2 + y\sigma_{bf}^2$
F × Y	$df_{fy} = (f-1)(y-1)$	M_{fy}	$\sigma_e^2 + b\sigma_{fy}^2$
F × Y × B	$df_e = (f-1)(y-1)(b-1)$	M_e	σ_e^2

[a] f is the number of families, b is the number of complete blocks, and y is the number of incomplete blocks or years

[b] σ_f^2 is the family, σ_{fy}^2 is the family × year, σ_{bf}^2 is the block × family, and σ_e^2 is the residual variance

Table 2. Analysis of variance for a balanced completely random linear model for family evaluation in different locations in a perennial crop with measurements repeated over years. The experiment design is a split-plot in time with complete blocks (whole plots) arranged in a randomized complete blocks experiment design

Source of variation	Degrees of freedom[a]	Mean square	Expected mean square[b]
Years (Y)	$y-1$		
Locations (S)	$s-1$		
Blocks:S (B:S)	$s(b-1)$		
Y × S	$(y-1)(s-1)$		
Y × B:S	$(y-1)(b-1)s$		
F × B:S	$(f-1)(b-1)s$		
Families (F)	$df_f = (f-1)$	M_f	$\sigma_e^2 + y\sigma_{fb:s}^2 + rf\sigma_{ms}^2 + by\sigma_{fs}^2 + bs\sigma_{fy}^2 + bsy\sigma_f^2$
F × S	$df_{fs} = (f-1)(s-1)$	M_{fs}	$\sigma_e^2 + y\sigma_{fb:s}^2 + b\sigma_{fys}^2 + by\sigma_{fs}^2$
F × Y	$df_{fy} = (f-1)(y-1)$	M_{fy}	$\sigma_e^2 + b\sigma_{fys}^2 + bs\sigma_{fy}^2$
F × Y × S	$df_{fsy} = (f-1)(y-1)(s-1)$	M_{fys}	$\sigma_e^2 + b\sigma_{fys}^2$
Y × F × B:S	$(y-1)(f-1)(b-1)s$		

[a] f is the number of families, s is the number of locations, and y is the number of incomplete blocks or years

[b] b is the number of complete blocks, σ_f^2 is the family, σ_{fs}^2 is the family × location, σ_{fy}^2 is the family × year, σ_{fsy}^2 is the family × year × location, $\sigma_{fb:s}^2$ is the family × blocks, and σ_e^2 is the incomplete blocks residual variance

Table 3. Analysis of variance for a balanced completely random linear model for the nested mating 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	$sb(r-1)$		
Males:B (M:B)	$df_m = b(m-1)$	M_m	$\sigma_e^2 + r\sigma_{fs}^2 + rs\sigma_f^2 + rf\sigma_{ms}^2 + rfs\sigma_m^2$
Females:M:B (F:M:B)	$df_f = bm(f-1)$	M_f	$\sigma_e^2 + r\sigma_{fs}^2 + rs\sigma_f^2$
M:B × L	$df_{ms} = (s-1)\,b(m-1)$	M_{ms}	$\sigma_e^2 + r\sigma_{fs}^2 + rf\sigma_{ms}^2$
F:M:B × L	$df_{fs} = (s-1)\,bm(f-1)$	M_{fs}	$\sigma_e^2 + r\sigma_{fs}^2$
Residual	$sb(r-1)(mf-1)$		

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

[b] r is the number of replications, σ_m^2 is the male parent, σ_f^2 is the female parent nested in male parent, σ_{ms}^2 is the male parent × lcoation interaction, σ_{fs}^2 is the female parent nested in male parent × location interaction, and σ_e^2 is the residual variance

with df″ numerator and df′ denominator degrees of freedom. An equation algebraically equivalent to (2) is

$$P\{1 - F_{\alpha/2:\,df'',\,df'}\,M'/M'' \leq 1 - E(M')/E(M'') \leq 1 - F_{1-\alpha/2:\,df'',\,df'}\,M'/M''\} \cong 1 - \alpha\,.$$

An approximate $(1-\alpha)$ interval estimator for $H = 1 - E(M')/E(M'')$ is

$$1 - F_{\alpha/2:\,df'',\,df'}\,M'/M''\,, \quad 1 - F_{1-\alpha/2:\,df'',\,df'}\,M'/M''\,. \qquad (3)$$

In the remaining sections of this paper we demonstrate that various H have the form in (1) and that approximate $1-\alpha$ interval estimators for these are obtained by substitution in (3).

One-factor mating design

Split-plot in time experiment designs are used for half-sib family evaluation in perennial crops with measure-

Table 4. Analysis of variance for a balanced completely random linear model for the factorial mating 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) and Females: B (F: B) Pooled	$df_c = b(m+f-2)$	M_c	$\sigma_e^2 + r\sigma_{mfy}^2 + rc\sigma_{py}^2 + ry\sigma_{mf}^2 + ryc\sigma_c^2$
M: B	$df_m = b(m-1)$	M_m	$\sigma_e^2 + r\sigma_{mfy}^2 + rf\sigma_{my}^2 + ry\sigma_{mf}^2 + ryf\sigma_m^2$
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 and F: B × Y Pooled	$df_{cy} = b(m-1)(y-1) + b(f-1)(y-1)$	M_{cy}	$\sigma_e^2 + r\sigma_{mfy}^2 + rc\sigma_{cy}^2$
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] b is the number of incomplete blocks, m is the number of male and f the number of female parents used to procedure factorial mating design progenies for one incomplete block, y is the number of years, and $c = (m+f)/2$

[b] σ_m^2 is the male and σ_f^2 the female parent, $\sigma_c^2 = (\sigma_m^2 + \sigma_f^2)/2$ is the male and female parent pooled, 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, $\sigma_{cy}^2 = (\sigma_{my}^2 + \sigma_{fy}^2)/2$ is the male and female parent × year pooled interaction, σ_{mfy}^2 is the male parent × female parent × year interaction, and σ_e^2 is the residual variance

ments repeated over years (Table 1). Heritability on a half-sib family mean basis for the Table 1 analysis is

$$\begin{aligned} H &= \sigma_f^2/\sigma_{\bar{p}}^2 = \sigma_f^2/(\sigma_f^2 + \sigma_{fy}^2/y + \sigma_{bf}^2/b + \sigma_e^2/yb) \\ &= \{[E(M_f) - E(M_{bf}) - E(M_{fy}) + E(M_e)]/yb\}/[M_f/yb] \\ &= 1 - [E(M_{bf}) + E(M_{fy}) - E(M_e)]/E(M_f)\,; \qquad (4) \end{aligned}$$

terms are defined in Table 1. If we define $E(M') = [(E(M_{bf}) + E(M_{fy}) - E(M_e)]$ and $E(M'') = E(M_f)$ then H in (4) has the general form in (1) that we described. The approximate $1-\alpha$ H interval estimator obtained by substitution of $M_{bf} + M_{fy} - M_{bfy}$ for M' and M_f for M'' in (3) is

$$1 - F_{\alpha/2:\,df''\,df'}(M_{bf} + M_{fy} - M_{bfy})/M_f\,,$$

$$1 - F_{1-\alpha/2:\,df''\,df'}(M_{bf} + M_{fy} - M_{bfy})/M_f\,.$$

The analysis of variance shown in Table 2 follows that of Table 1 except that the experiment was conducted over locations. For this experiment H is

$$\begin{aligned} H &= \sigma_f^2/\sigma_{\bar{p}}^2 = \sigma_f^2/(\sigma_f^2 + \sigma_{fy}^2/y + \sigma_{fs}^2/s + \sigma_{fys}^2/ys \\ &\qquad + \sigma_b^2/sy + \sigma_e^2/bsy) \\ &= 1 - [E(M_{fs}) + E(M_{fy}) - E(M_{fsy})]/E(M_f)\,. \qquad (5) \end{aligned}$$

If we define $E(M') = E(M_{fs}) + E(M_{fy}) - E(M_{fys})$ and $E(M'') = E(M_f)$ then H in (5) is equivalent to (1). An approximate $1-\alpha$ interval estimator for (5) is obtained by substitution of $M_{fs} + M_{fy} + M_{fys}$ for M' and M_f for M'' in (3).

H for several other experiment designs, e.g., randomized complete blocks and replications-in-incomplete blocks, in locations and years are similar to (5). Nguyen and Sleper (1983) give tall fescue half-sib family data for a randomized complete blocks experiment design in locations and years. We obtained an interval estimate for heritability on a half-sib family mean basis from their reproductive herbage yield data to illustrate the method. Mean squares were reconstructed from variance data. We obtained $M' = 59{,}475$ and $M'' = 116{,}179$. This gave $H = 1 - 0.512 = 0.488$ which differs from their estimate (0.47) because of rounding error. There were 14 degrees of freedom for every mean square. Approximate degrees of freedom for M' were $df' = 21.021 \cong 21$. Pertinent F-values were obtained by using interpolation equations partly given in Laubscher (1965) and detailed in Knapp (1986). The lower limit for a $1 - \alpha \cong 0.95$ interval estimate for H was

$$1 - F_{0.025:\,14,\,21}\,M'/M'' = 1 - 2.561\,(0.512) = -0.311$$

and the upper limit was

$$1 - F_{0.975:\,14,\,21}\,M'/M'' = 1 - 0.354\,(0.512) = 0.819\,.$$

Two-factor mating designs

Knapp (1986) described interval estimators for heritability on a half-sib family mean basis for the nested and factorial mating designs in one environment. We examined these estimators for experiments in more than one environment.

The analysis of variance for the nested mating design is given in Table 3. Heritability on a half-sib family mean basis is

$$\begin{aligned} H &= \{[E(M_m) - E(M_f) - E(M_{ms}) + E(M_{fs})]/rfs\}/(M_m/rfs) \\ &= 1 - [E(M_f) + E(M_{ms}) - E(M_{fs})]/E(M_m). \end{aligned} \quad (6)$$

If we define $E(M') = E(M_f) + E(M_{ms}) - E(M_{fs})$ and $E(M'') = E(M_m)$ then H in (6) has the general form in (1). Substitution of $M_f + M_{ms} - M_{fs}$ for M' and M_m for M'' in (3) gives an interval estimator for this example.

The analysis of variance for the factorial mating design is given in Table 4. Heritability on a half-sib family mean basis is

$$\begin{aligned} H &= \{[E(M_c) - E(M_{mf}) - E(M_{cy}) + E(M_{mfy})]/ryc\}/[E(M_c)/ryc] \\ &= 1 - [E(M_{mf}) + E(M_{cy}) - E(M_{mfy})]/E(M_c). \end{aligned} \quad (7)$$

If we define $E(M') = E(M_{mf}) + E(M_{cy}) - E(M_{mfy})$ and $E(M'') = E(M_c)$ then H in (7) has the general form in (1). Substitution of $M_{mf} + M_{cy} - M_{mfy}$ for M' and M_c for M'' in (3) gives interval estimators for this example.

The results for the two-factor mating designs in replications-in-incomplete blocks experiment designs also apply to other experiment designs, e.g., completely randomized and randomized complete blocks experiment designs.

References

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, Mass

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

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

Laubscher NF (1965) Interpolation in F-tables. Am Stat 19:39–40

Nguyen HT, Sleper DA (1983) Genetic variability of seed yield and reproductive characters in tall fescue. Crop Sci 23:621–626

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

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

Welch BL (1956) On linear combinations of several variances. J Am Stat Assoc 51:132–148