Theoretical and Applied Genetics 40, 357—359 (1970)
© by Springer-Verlag 1970

Sampling Variance of the Correlation Coefficients Estimated from Analyses of Variance and Covariance[1]

M. GROSSMAN[2]

Agricultural Experiment Station, Purdue University, Lafayette, Indiana (USA)

Summary. A generalized sampling variance of correlation coefficients is derived for phenotypic, genetic and environmental correlations estimated from nested analyses of variance and covariance for the equal number case. A numerical example is presented to estimate the sampling variance for the genetic correlation coefficient based on the relationship among full sibs using unequal subclass numbers.

Introduction

In a breeding program, the genetic correlation between traits is usually of interest, and a minimum variance estimate is important for precise prediction on a short term basis. From a practical viewpoint, it is also critical that the variance not be biased downward and thus underestimated. Robertson (1959) has derived the sampling variance of the genetic correlation coefficient between two traits with equal heritability, based on the intraclass correlation and estimated from analyses of variance and covariance. Using the analyses of variance and covariance in a different development, Tallis (1959) derived the sampling variance of the genetic correlation with different heritabilities in the between-and-within sire case. This special case was extended by Scheinberg (1966) to include the phenotypic and environmental correlations as well. In 1959, Mode and Robinson estimated the variance of the genetic correlation based on the sire component (assuming only additive components) and of the genotypic correlation based on the dam component (involving additive and dominance components) from a nested analysis. They also obtained the variance of the phenotypic correlation.

It is the purpose of this paper to extend the development of Mode and Robinson (1959) in order to obtain estimates of the sampling variance of the phenotypic, genetic and environmental correlations estimated by appropriate combinations of the sire, dam and progeny components from nested analyses of variance and covariance with equal numbers.

Estimation Procedure

A commonly employed mating scheme in quantitative genetic studies is the hierarchial or nested experiment in which a random sample of progeny is observed from each dam, a random sample of dams is mated to each sire chosen also at random. This type of experiment requires a dam to be mated to only one sire. The population is assumed to be under random selection with an inbreeding coefficient of zero and no assortative mating.

The assumed random effects model for the covariables to be analyzed is:

$$X_{tijk} = \mu_t + S_{ti} + D_{tij} + E_{tijk}$$

where X_{tijk} is the t^{th} trait measured on the k^{th} progeny of the j^{th} dam by the i^{th} sire; μ_t is the true mean, S_{ti} the effect of the i^{th} sire, D_{tij} the effect of the j^{th} dam within the i^{th} sire and E_{tijk} the residual effect; and where $t =$ trait 1 or 2, $i = 1, \ldots, s$ sires, $j = 1, \ldots, d$ dams/sire, $k = 1, \ldots, n$ progeny/dam, nd is the number of progeny/sire and $snd = N$ the total number of progeny. The following assumptions are necessary:

(1) All effects have mean zero and are mutually uncorrelated for fixed t.

(2) $E(S_{qi}\,S_{ri}) = S_{qr}$; $E(D_{qij}\,D_{rij}) = D_{qr}$; $E(E_{qijk}\,E_{rijk}) = E'_{qr}$ where q and r denote traits 1 or 2. If $q = r$ the expectations are interpreted as variance components and if $q \neq r$ as covariance components.

(3) There is no interaction between sires and dams. It should be noted that if interaction is indeed present it is completely confounded with dam effects.

The model for analyses of variance and covariance of individual values is presented in Table 1.

Estimates of the variance or covariance components are:

$$\hat{E}'_{qr} = W_{qr}$$
$$\hat{D}_{qr} = [V_{qr} - W_{qr}]/n$$
$$\hat{S}_{qr} = [U_{qr} - V_{qr}]/nd\,.$$

---

[1] Journal Paper No. 3472 of the Purdue University Agricultural Experiment Station. This research was supported, in part, by NIH Biometry Training Grant, GM-00024.

[2] Present address: Department of Dairy Science, 215 Animal Sciences Laboratory, University of Illinois, Urbana, Illinois 61801, U.S.A.

Table 1. *Analyses of Variance and Covariance*

Source	d.f.	MS/MCP*	E(MS)/E(MCP)
Sires (S)	$s - 1 = u$	U_{qr}	$E'_{qr} + n\,D_{qr} + ndS_{qr}$
Dams (D)/S	$s\,(d - 1) = v$	V_{qr}	$E'_{qr} + n\,D_{qr}$
Progeny/D/S	$sd\,(n-1) = w$	W_{qr}	E'_{qr}

* q = r denotes mean square
q ≠ r denotes mean cross product

When giving genetic interpretations to these components, it shall be assumed throughout this development that (a) the genetic effects are additive, (b) the environmental correlation between members of families is negligible and (c) the variance due to dominance deviations is zero. The latter assumption is implied in assumption (3) for the model.

Thus from Dickerson (1960) for variance components, and Grossman and Gall (1968) for covariance components:

$$S_{qr} = D_{qr} = G_{qr}/4$$
$$E'_{qr} = (G_{qr}/2) + E_{qr}$$

where G_{qr} and E_{qr} are the genetic and environmental variances or covariances, respectively. The phenotypic variance or covariance is:

$$P_{qr} = G_{qr} + E_{qr} = S_{qr} + D_{qr} + E'_{qr}\,.$$

The values of P_{qr}, G_{qr} and E_{qr} can be estimated, assuming only additive genetic components, as follows:

$$\begin{aligned}\hat{P}_{qr} &= \hat{S}_{qr} + \hat{D}_{qr} + \hat{E}'_{qr} = [U_{qr} + (d-1)\,V_{qr} \\ &\quad + d\,(n-1)\,W_{qr}]/nd \qquad (1)\\ \hat{G}_{qr} &= 4\,\hat{S}_{qr} = 4\,[U_{qr} - V_{qr}]/nd \qquad (2)\\ &= 4\,\hat{D}_{qr} = 4\,[dV_{qr} - dW_{qr}]/nd \qquad (3)\\ &= 2\,(\hat{S}_{qr} + \hat{D}_{qr}) = 2\,[U_{qr} + (d-1)\,V_{qr} \\ &\quad - dW_{qr}]/nd \qquad (4)\\ \hat{E}_{qr} &= \hat{E}'_{qr} - 2\,\hat{S}_{qr} = 2\,[-U_{qr} + V_{qr} \\ &\quad + (nd)\,W_{qr}/2]/nd \qquad (5)\\ &= \hat{E}'_{qr} - 2\,\hat{D}_{qr} = 2\,[-\,dV_{qr} + d((n/2)+1)\,W_{qr}] \\ &\quad /nd \qquad (6)\\ &= \hat{E}'_{qr} - \hat{S}_{qr} - \hat{D}_{qr} = [-U_{qr} - (d-1)\,V_{qr} \\ &\quad + d\,(n+1)\,W_{qr}]/nd \qquad (7)\\ &= \hat{E}'_{qr} + \hat{S}_{qr} - 3\,\hat{D}_{qr} = [U_{qr} - (3\,d+1)\,V_{qr} \\ &\quad + d\,(n+3)\,W_{qr}]/nd \qquad (8)\\ &= \hat{E}'_{qr} + \hat{D}_{qr} - 3\,\hat{S}_{qr} = [-\,3\,U_{qr} + (d+3)\,V \\ &\quad + d\,(n-1)\,W_{qr}]/nd\,. \qquad (9)\end{aligned}$$

Using the notation of Scheinberg (1966), the correlation coefficients between traits 1 and 2 are estimated from analyses of variance and covariance by:

$$\hat{r}_\theta = \hat{\theta}_{12}/(\hat{\theta}_{11}\,\hat{\theta}_{22})^{1/2}$$

where r_θ is the phenotypic, genetic or environmental correlation coefficient depending on whether θ represents P, G or E; θ_{12} is the covariance between traits 1 and 2; θ_{11} and θ_{22} are the variances of traits 1 and 2, respectively.

The variance of the correlation coefficient is given by Kendall and Stuart (1963, p. 235) as:

$$\begin{aligned}Var(\hat{r}_\theta) = r_\theta^2\,\{&[Var(\hat{\theta}_{12})/\theta_{12}^2] + [Var(\hat{\theta}_{11})/4\,\theta_{11}^2] \\ &+ [Var(\hat{\theta}_{22})/4\,\theta_{22}] \\ &- [Cov\,(\hat{\theta}_{11}, \hat{\theta}_{12})/\theta_{11}\,\theta_{12}] \\ &- [Cov\,(\hat{\theta}_{12}, \hat{\theta}_{22})/\theta_{12}\,\theta_{22}] \\ &+ [Cov\,(\hat{\theta}_{11}, \hat{\theta}_{22})/2\,\theta_{11}\,\theta_{22}]\} \qquad (10)\end{aligned}$$

In evaluating the variances and covariances of the θ's in terms of U_{qr}, V_{qr} and W_{qr} the following relationships presented by Tallis (1959) are necessary:

$$Var(M_{qr}) \sim (M_{qq}\,M_{rr} + M_{qr}^2)/\text{d. f.}$$
$$Cov(M_{qr}, M_{st}) \sim (M_{qs}\,M_{rt} + M_{qt}\,M_{rs})/\text{d. f.}$$

where M is U, V or W and d. f. refers to the appropriate degrees of freedom u, v or w, respectively.

It follows that:

$$\begin{aligned}\text{Est. } Var(\hat{\theta}_{qr}) \sim f^2\,\{&[a^2\,(U_{qq}\,U_{rr} + U_{qr}^2)/u] \\ &+ [b^2\,(V_{qq}\,V_{rr} + V_{qr}^2)/v] \\ &+ [c^2\,(W_{qq}\,W_{rr} + W_{qr}^2)/w]\}/(nd)^2 \qquad (11)\\ \text{Est. } Cov(\hat{\theta}_{qr}, \hat{\theta}_{st}) \sim f^2\,\{&[a^2\,(U_{qs}\,U_{rt} + U_{qt}\,U_{rs})/u] \\ &+ [b^2\,(V_{qs}\,V_{rt} + V_{qt}\,V_{rs})/v] \\ &+ [c^2\,(W_{qs}\,W_{rt} + W_{qt}\,W_{rs})/w]\}/(n\,d)^2 \qquad (12)\end{aligned}$$

where a, b, c and f are defined in Table 2 depending on the correlation and method used to estimate it.

Substituting equations (11) and (12) into (10), the estimated sampling variance of the correlation coefficient between traits 1 and 2 is:

$$\begin{aligned}Est.\ Var(\hat{r}_\theta) = f^2\,\hat{r}_\theta^2\,[(&[\{a^2\,(U_{11}\,U_{22} + U_{12}^2)/u\} \\ &+ \{b^2\,(V_{11}\,V_{22} + V_{12}^2)/v\} \\ &+ \{c^2\,(W_{11}\,W_{22} + W_{12}^2)/w\}]/\hat{\theta}_{12}^2) \\ &+ ([\{a^2\,U_{11}^2/u\} + \{b^2\,V_{11}^2/v\} \\ &+ \{c^2\,W_{11}^2/w\}]/2\,\hat{\theta}_{11}^2) \\ &+ ([\{a^2\,U_{22}^2/u\} + \{b^2\,V_{22}^2/v\} \\ &+ \{c^2\,W_{22}^2/w\}]/2\,\hat{\theta}_{22}^2) \qquad (13)\\ &- ([\{a^2\,U_{11}\,U_{12}/u\} + \{b^2\,V_{11}\,V_{12}/v\} \\ &+ \{c^2\,W_{11}\,W_{12}/w\}]/\hat{\theta}_{11}\,\hat{\theta}_{12}) \\ &- ([\{a^2\,U_{12}\,U_{22}/u\} + \{b^2\,V_{12}\,V_{22}/v\} \\ &+ \{c^2\,W_{12}\,W_{22}/w\}]/\hat{\theta}_{12}\,\hat{\theta}_{22}) \\ &+ ([\{a^2\,U_{11}\,U_{22}/u\} + \{b^2\,V_{11}\,V_{22}/v\} \\ &+ \{c^2\,W_{11}\,W_{22}/w\}]/\hat{\theta}_{11}\,\hat{\theta}_{22})]/(n\,d)^2\,.\end{aligned}$$

For the case of unequal subclass numbers, it should be noted that the mean squares are not distributed as $\chi^2\,\sigma^2$/d.f. but as weighted sums of different χ^2's and the sampling variances of their estimates are very complex (Anderson and Bancroft, 1952). In addition,

Table 2. *Coefficients Used in Variance of Correlation Coefficient*

Equation number used to estimate correlation	Coefficients a	b	c	f
Phenotypic				
(1)	1	$(d-1)$	$d(n-1)$	1
Genetic				
(2) Sire component	1	1	0	4
(3) Dam component	0	d	d	4
(4) Sire plus dam components	1	$(d-1)$	d	2
Environmental				
(5)	1	1	$n\,d/2$	2
(6)	0	d	$d((n/2)+1)$	2
(7)	1	$(d-1)$	$d(n+1)$	1
(8)	1	$(3\,d+1)$	$d(n+3)$	1
(9)	3	$(d+3)$	$d(n-1)$	1

Table 3. *Analyses of Variance and Covariance*

Source	d.f.	MS(8)	MCP(8,12)	MS(12)
Sires (*S*)	31	23,410	35,592	66,634
Dams (*D*)/*S*	18	19,134	25,187	37,381
Progeny/*D*/*S*	175	9,349	13,241	23,613
Total	224			

Kendall and Stuart (1963, p. 236) do not recommend the use of the standard error developed here for tests of significance since the sampling distribution of r is unknown and tends to normality slowly.

Numerical Example

A numerical example is presented to illustrate the use of analyses of variance and covariance with unequal subclass numbers in estimating the variance of the genetic correlation coefficient. Body weight data was collected at the Purdue University Poultry Center Farm on Rhode Island Red males at 8 and 12 weeks of age. The experiment was in a nested arrangement with between one and seven progeny per dam and one to three dams per sire. The average number of progeny per sire was 6.9, the average number of dams per sire was 1.5 and the average number of progeny per dam was 4.6. These values should be substituted for nd, d and n, respectively. The average numbers may be obtained by substituting $1/2\,(k_1+k_2)$ for n and k_3 for nd. The method of calculating k_1, k_2 and k_3 is contained in Grossman and Gall (1968).

The method of calculation presented in this example uses the full sib (sire plus dam) components, equation (4). Therefore, the values for a, b, c and f are 1.0, 0.5, 1.5 and 2.0, respectively (see Table 2). The appropriate values of U, V, W and u, v, w are found in Table 3. From equation 4, the genetic variances and covariance can be estimated as:

$$\hat{G}_{8,8} = 2\,[23{,}410 + 0.5\,(19{,}134) - 1.5\,(9{,}349)]/6.9 = 5{,}494.$$

$$\hat{G}_{12,12} = 2\,[66{,}634 + 0.5\,(37{,}381) - 1.5\,(23{,}613)]/6.9 = 14{,}465.$$

$$\hat{G}_{8,12} = 2\,[35{,}592 + 0.5\,(25{,}187) - 1.5\,(13{,}241)]/6.9 = 8{,}210.$$

From these estimates, the genetic correlation coefficient can be estimated as:

$$\hat{r}_G = 8{,}210/(5{,}494 \cdot 14{,}465)^{1/2} = .92\,.$$

The above information can be combined to estimate the variance of the genetic correlation coefficient:

$$\begin{aligned} Est.\ Var(\hat{r}_G) = {} & 4\,(.92)^2\,[([\{(23{,}410 \cdot 66{,}634 \\ & + 35{,}592^2)/31\} \\ & + \{.25\,(19{,}134 \cdot 37{,}381 + 25{,}187^2)/ \\ & \quad 18\} \\ & + \{2.25\,(9{,}349 \cdot 23{,}613 \\ & + 13{,}241^2)/175\}]/8{,}210^2) + \cdots \\ & + ([\{(23{,}410 \cdot 66{,}634)/31\} \\ & + \{0.25\,(19{,}134 \cdot 37{,}381)/18\} \\ & + \{2.25\,(9{,}349 \cdot 23{,}613)/175\}]/ \\ & 5{,}494 \cdot 14{,}465)]/6.9^2 \\ & = 0.12\,. \end{aligned}$$

Therefore, the standard error of the genetic correlation coefficient is 0.35.

Acknowledgements

The author wishes to thank Dr. B. B. Bohren for his encouragement and critical comments as well as to extend appreciation to Miss Peggy Fahl and Miss Rosalyn Buhr for their typographical assistance.

References

1. Anderson, R. L., Bancroft, T. A.: Statistical Theory in Research. New York: McGraw-Hill Book Company Inc. 1952. — 2. Dickerson, G. E.: Techniques for research in quantitative animal genetics. In: Techniques and Procedures in Animal Production Research. Monograph of American Society of Animal Production, 1960. — 3. Grossman, M., Gall, G. A. E.: Covariance analysis with unequal subclass numbers: Component-estimation in quantitative genetics. Biometrics **24**, 49—59 (1968). — 4. Kendall, M. G., Stuart, A.: The Advanced Theory of Statistics, Vol. 1, 2nd Ed. New York: Hafner Publishing Company 1963. — 5. Mode, C. G., Robinson, H. F.: Pleiotropism and the genetic variance and covariance. Biometrics **15**, 518—537 (1959). — 6. Robertson, A.: The sampling variance of the genetic correlation coefficient. Biometrics **15**, 469—485 (1959). — 7. Scheinberg, E.: The sampling variance of the correlation coefficients estimated in genetic experiments. Biometrics **22**, 187 to 191 (1966). — 8. Tallis, G. M.: Sampling errors of genetic correlation coefficients calculated from the analyses of variance and covariance. Aust. Jour. of Stat. **1**, 35—43 (1959).

Received June 1, 1970
Communicated by H. Abplanalp

M. Grossman
Agricultural Experiment Station
Purdue University
Lafayette, Indiana 47907 (USA)