Theoretical and Applied Genetics 42, 97–100 (1972)

The Sampling Variance of the Correlation Coefficients estimated from Two-fold Nested and Offspring-Parent Regression Analyses

K. HAMMOND and **F. W. NICHOLAS**

Department of Animal Husbandry, University of Sydney (Australia)

Summary. A formula is presented for the large-sample variance of phenotypic, genetic and environmental correlation coefficients, estimated from two-fold nested genetic analyses of balanced or unbalanced offspring data. Where parents of these offspring are scored, offspring on parent regression estimates of genetic parameters may be obtained for each parental level of the nested model. The large-sample variance of the offspring-parent genetic correlation coefficients, computed from either level or by using mid-parent data, is given. The first formula is a correction and generalisation of an expression given by Grossman (1970), while the second formula is an extension of a relationship derived by Reeve (1955).

Introduction

Both sib and offspring on parent regression estimates of genetic parameters may be obtained from completely nested data provided the traits studied are measured in both parents and offspring. During the development of a computer programme which was designed to perform both these analyses, it became apparent that the existing methods for computing sampling variances of the correlation coefficients obtained from each analysis, *e. g.* Mode and Robinson (1959) and Reeve (1955), were insufficiently general to be applied.

Sib Analysis

The appropriate generalisation for obtaining the sampling variance of correlation coefficients estimated from sib analysis was given by Abe (1969) for the genetic correlation coefficients only, and by Grossman (1970) for all correlation coefficients. However, the formulae given in these two papers do not agree when used to estimate the sampling variance of the genetic correlation coefficients. The final formula, (13), of Grossman is incorrect. In addition, the analyses were not extended to unbalanced data. Grossman's reason for not providing this extension is valid, because the variance of the genetic correlation coefficient will be underestimated with unbalanced data (Tallis, 1959). Rohde and Tallis (1969), using matrix methods, derived the exact moments of estimates of components of covariance. In an empirical study, and using the same single-fold model that Tallis examined, Shook and Barr (1968) showed that "the tendency for the formula [of Tallis] to underestimate the sampling variance in samples with unequal subclass size decreases as genetic correlation increases. The estimated sampling variance is not seriously in error, but should be considered conservative when subclass size is not constant". Their subclass sizes followed an exponential distribution of 5, 10, 20, 40 and 80 progeny per sire. In practice this degree of inequality probably would be rare and the other major sources of bias indicated by them would become all-important. These are low heritabilities, small subclass size and small total sample size. This also would apply to estimates of the variance of the phenotypic and environmental correlations, and to all estimates computed using the two-fold model. In practice data are rarely completely balanced — even the data of Grossman's example are unbalanced. Thus the generalisation of his analysis, to cater for unbalanced data, while maintaining a simple formula, is warranted.

In providing a corrected and general formula for the sampling variance of the correlation coefficients derived from a sib analysis, we apply throughout the model, assumptions, notation and method of Grossman except where otherwise stated.

The standard analysis of variance and covariance model is given in Table 1. With unequal numbers of progeny per sire and/or per dam the coefficients, k_1, k_2, and k_3, of the components of variance (covariance) must be computed using formulae of the form given by Reeve (1955), equation (34) formula (2).

Table 1. *Analyses of Variance and Covariance*†

Source	d.f.	MS or Cov*	E (MS) or E (Cov)
Sires (S)	$s - 1 = u$	U_{qr}	$E'_{qr} + k_2 D_{qr} + k_3 S_{qr}$
Dams (D)/S	$d. - s = v$	V_{qr}	$E'_{qr} + k_1 D_{qr}$
Progeny/D/S	$n.. - d. = w$	W_{qr}	E'_{qr}

* q and r denote two characters, and for mean square $q = r$.

† where, following Grossman, for the tth trait

$$d. = \sum_{i=1}^{s} d_i \text{ and } n.. = \sum_{i=1}^{s} \sum_{j=1}^{d_i} n_{ij}$$

Estimates of each of the three variance (covariance) components are then obtained from Table 1 as follows:

$$\hat{E}'_{qr} = W_{qr}$$

$$\hat{D}_{qr} = [V_{qr} - W_{qr}]/k_1$$

$$\hat{S}_{qr} = [U_{qr} - (k_2/k_1)\, V_{qr} - ((k_1 - k_2)/k_1)\, W_{qr}]/k_3\,.$$

The estimates of the phenotypic (P_{qr}), and of genetic (G_{qr}) and environmental (E_{qr}) variances (covariances, $q \neq r$) derived by utilizing the sire, dam and sire plus dam components and designated G^s_{qr}, E^s_{qr}; G^d_{qr}, E^d_{qr}; G^{sd}_{qr}, E^{sd}_{qr}, respectively become

$$\begin{aligned}\hat{P}_{qr} &= \hat{S}_{qr} + \hat{D}_{qr} + \hat{E}'_{qr}\\ &= [k_1\, U_{qr} + (k_3 - k_2)\, V_{qr}\\ &\quad + (k_2 - k_1 + k_3\,(k_1 - 1))\, W_{qr}]/k_1\, k_3 \qquad (1)\end{aligned}$$

$$\begin{aligned}\hat{G}^s_{qr} &= 4\, \hat{S}_{qr}\\ &= 4\, [U_{qr} - (k_2/k_1)\, V_{qr} + ((k_2 - k_1)/k_1)\, W_{qr}]/k_3 \qquad (2)\end{aligned}$$

$$\begin{aligned}\hat{G}^d_{qr} &= 4\, \hat{D}_{qr}\\ &= 4\, [V_{qr} - W_{qr}]/k_1 \qquad (3)\end{aligned}$$

$$\begin{aligned}\hat{G}^{sd}_{qr} &= 2\, (\hat{S}_{qr} + \hat{D}_{qr})\\ &= 2\, [k_1\, U_{qr} + (k_3 - k_2)\, V_{qr}\\ &\quad + (k_2 - k_1 - k_3)\, W_{qr}]/k_1\, k_3 \qquad (4)\end{aligned}$$

$$\begin{aligned}\hat{E}^s_{qr} &= \hat{E}'_{qr} - 2\, \hat{S}_{qr}\\ &= [-\, 2\, U_{qr} + 2\, (k_2/k_1)\, V_{qr}\\ &\quad + (((k_1 - k_2)/k_1) + k_3)\, W_{qr}]/k_3 \qquad (5)\end{aligned}$$

$$\begin{aligned}\hat{E}^d_{qr} &= \hat{E}'_{qr} - 2\, \hat{D}_{qr}\\ &= [-2\, V_{qr} + (k_1 + 2)\, W_{qr}]/k_1 \qquad (6)\end{aligned}$$

$$\begin{aligned}\hat{E}^{sd}_{qr} &= \hat{E}'_{qr} - \hat{S}_{qr} - \hat{D}_{qr}\\ &= [-\, k_1\, U_{qr} + (k_2 - k_3)\, V_{qr}\\ &\quad + (k_1 - k_2 + k_3\,(k_1 + 1))\, W_{qr}]/k_1\, k_3\,. \qquad (7)\end{aligned}$$

When the data are completely balanced, $k_1 = k_2 = n$, the number of progeny per dam and $k_3 = nd$, the number of progeny per sire.

Equations (1) to (7) take the general form

$$\hat{\theta}_{qr} = f[a\, U_{qr} + b\, V_{qr} + c\, W_{qr}]/l\,, \qquad (8)$$

where relevant values for f, l, a, b and c are given in Table 2. The estimated correlation coefficients, for characters 1 and 2 for example, are then

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

Following Grossman, the estimated variances and covariances of the θ's may then be represented approximately as

$$\begin{aligned}\text{Est. Var }(\hat{\theta}_{qr}) &\sim f^2[(a^2\, (U_{qq}\, U_{rr} + U^2_{qr})/u')\\ &\quad + (b^2\, (V_{qq}\, V_{rr} + V^2_{qr})/v')\\ &\quad + (c^2\, (W_{qq}\, W_{rr} + W^2_{qr})/w')]/l^2 \qquad (9)\\ \text{Est. Cov }(\hat{\theta}_{qr}, \hat{\theta}_{st}) &\sim f^2[(a^2(U_{qs}\, U_{rt} + U_{qt}\, U_{rs})/u')\\ &\quad + (b^2\, (V_{qs}\, V_{rt} + V_{qt}\, V_{rs})/v')\\ &\quad + (c^2\, (W_{qs}\, W_{rt} + W_{qt}\, W_{rs})/w)]/l^2\,.\end{aligned}$$

In practive the mean squares and products will always be sample estimates, so 2 is added to each of the degrees of freedom in (9) to correct the variance of these estimates for bias, *e. g.* $u' = u + 2$.

By substituting values from (9) in equation (10)* of Grossman the estimated variance of the correlation coefficient is given by

$$\begin{aligned}\text{Est. Var }(\hat{r}_\theta) = 2\, f^2\, \hat{r}^2_\theta\, [&(((a^2(U_{11}\, U_{22} + U^2_{12})/u')\\ &+ (b^2(V_{11}\, V_{22} + V^2_{12})/v') + (c^2(W_{11}\, W_{22}\\ &+ W^2_{12})/w'))/2\, \hat{\theta}^2_{12})\\ &+ (((a^2\, U^2_{11}/u') + (b^2\, V^2_{11}/v')\\ &+ (c^2\, W^2_{11}/w'))/4\, \hat{\theta}^2_{11})\\ &+ (((a^2\, U^2_{22}/u') + (b^2\, V^2_{22}/v')\\ &+ (c^2\, W^2_{22}/w'))/4\, \hat{\theta}^2_{22})\\ &- (((a^2\, U_{11}\, U_{12}/u') + (b^2\, V_{11}\, V_{12}/v')\\ &+ (c^2\, W_{11}\, W_{12}/w'))/\hat{\theta}_{11}\, \hat{\theta}_{12})\\ &- (((a^2\, U_{12}\, U_{22}/u') + (b^2\, V_{12}\, V_{22}/v')\\ &+ (c^2\, W_{12}\, W_{22}/w'))/\hat{\theta}_{12}\, \hat{\theta}_{22})\\ &+ (((a^2\, U^2_{12}/u') + (b^2\, V^2_{12}/v')\\ &+ (c^2\, W^2_{12}/w'))/2\, \hat{\theta}_{11}\, \hat{\theta}_{22})]/l^2\,. \qquad (10)\end{aligned}$$

Table 2. *Values for the Coefficients of Equations (8), (9) and (10)*

Equation number used to estimate variance (covariance)	Coefficients f	l	a	b	c
Phenotypic					
(1)	1	$k_1\, k_3$	k_1	$(k_3 - k_2)$	$[k_2 - k_1 + k_3\,(k_1 - 1)]$
Genetic					
(2)	4	k_3	1	$-\, k_2/k_1$	$(k_2 - k_1)/k_1$
(3)	4	k_1	0	1	$-\, 1$
(4)	2	$k_1\, k_3$	k_1	$(k_3 - k_2)$	$(k_2 - k_1 - k_3)$
Environmental					
(5)	1	k_3	$-\, 2$	$2\, k_2/k_1$	$[((k_1 - k_2)/k_1) + k_3]$
(6)	1	k_1	0	$-\, 2$	$(k_1 + 2)$
(7)	1	$k_1\, k_3$	$-\, k_1$	$(k_2 - k_3)$	$[k_1 - k_2 + k_3\,(k_1 + 1)]$

---

* The denominator on the second line of equation (10) of Grossman should read $4\, \theta^2_{22}$.

By substituting into equation (10) the appropriate values from the example given by Grossman, the standard error of the calculated genetic correlation coefficient is 0.04. Since he gives no estimate for k_2, his estimates of a, b and c have been used.

Offspring-Parent Regression Analysis

When both parent and offspring measurements are taken on two or more traits, and the data assume the model of Table 1, sire- and dam-offspring estimates of the genetic correlations between pairs of these traits can be obtained. Analysis is carried out by using this same model on both the parent and progeny data and then by substituting the appropriate covariances into the formula(e) for the offspring-parent (O—P) genetic correlation — equation (1) of Reeve (1955). In data such as these, each sire will provide families of full-sibs in a population of half-sibs.

Reeve derived the formula for estimating the variance of the O—P genetic correlation for the two specific cases of mid-parent with full-sib progeny and common parent with half-sib progeny. By modifying the basic formula — his equation (13) — he extended its usefulness, but it is still not sufficiently general to enable the estimation of separate variances for the sire- and dam-offspring genetic correlations computed assuming the model of Table 1. Reeve's formula (13) is now re-derived to obtain an expression for the sampling variance of the O—P genetic correlation coefficient applicable to all methods of estimation which assume random mating. The notation used by Reeve is followed where practicable.

In the general model of Table 1 we have k_1 progeny per dam and k_3 progeny per sire on average. Now an individual progeny will have $(k_1 - 1)$ full sibs in a total sire family of $(k_3 - 1)$ relatives. The proportion of full-sibs in a sire family is then $(k_1 - 1)/(k_3 - 1)$, and the proportion of half-sibs in the same family is $\big(1 - ((k_1 - 1)/(k_3 - 1))\big)$. The intraclass correlations among full-sibs and among half-sibs, for character 1, are $h_1^2/2$ and $h_1^2/4$ respectively, so the intraclass correlation among sibs for character 1 is given by

$$\begin{aligned} r_{11} &= (h_1^2\,(k_1 - 1)/2\,(k_3 - 1)) + (h_1^2\,(k_3 - k_1)/4\,(k_3 - 1)) \\ &= h_1^2\,(k_3 + k_1 - 2)/4\,(k_3 - 1) \end{aligned}$$

where h_1^2 is the heritability of differences among individuals; and similarly with r_{22} for character 2.

Following Reeve we let a and x be the parent phenotypes and b and y the corresponding progeny (or progeny mean) phenotypes for characters 1 and 2 respectively, and the correlation coefficients, [ ], of his equation (12) now become

$$\left.\begin{aligned} [a\,b] &= h_1^2/(2\,B\,p/k_3)^{1/2} \\ [x\,y] &= h_2^2/(2\,Y\,p/k_3)^{1/2} \\ [x\,b] &= h_1\,h_2\,r_G/(2\,B\,p/k_3)^{1/2} \\ [a\,y] &= h_1\,h_2\,r_G/(2\,Y\,p/k_3)^{1/2} \\ [a\,x] &= r_P \\ [b\,y] &= (r_P + h_1\,h_2\,r_G\,(k_3 + k_1 - 2)/4)/(B\,Y)^{1/2}, \end{aligned}\right\} \qquad (11)$$

where

$$B = 1 + h_1^2\,(k_3 + k_1 - 2)/4$$

$$Y = 1 + h_2^2\,(k_3 + k_1 - 2)/4\,.$$

r_G and r_P are the genetic and phenotypic correlations between characters 1 and 2, and p is equal to either 1 or 2 for estimates involving mid-parent values or those involving measurement on one parent, respectively. All possible proportions of half- and full-sibs are covered, including, Reeve's two original cases of full-sibs only ($k_1 = k_3$) and half-sibs only ($k_1 = 1$). Substituting (11) in Reeve's (11) and simplifying gives the variance of the O—P genetic correlations analogous to Reeve's (13), as

$$\begin{aligned} \text{Est. Var}\,(r_G) = [\,((1 - r_G^2)^2/2) &+ (p\,(k_3 + k_1)\,(1 - r_G^2) \\ &\times ((1/D) - (r_P\,r_G/C))/4\,k_3) \\ &+ (2\,p\,((r_G/D) - (r_P/C))^2/k_3) \\ &+ ((1 - r_G^2)\,((p\,(1 - r_P^2)/C^2) \\ &- (p/2\,D) + (p\,r_P\,r_G/2\,C))/k_3)]/d\,, \end{aligned} \qquad (12)$$

where d is one less than the number of degrees of freedom of the relevant level in the model used for estimation, $C = h_1\,h_2$ and $1/D = ((1/h_1^2) + (1/h_2^2))/2$. r_P, r_G, h_1^2 and h_2^2 are the best available estimates of their respective population parameters. As suggested by Reeve, for a given population, these need not necessarily come from the data currently under study. Furthermore, equation (12) applies equally to estimates of r_G calculated from either a geometric or an arithmetic average of the reciprocal covariances of offspring on parent, *i. e.* from either equation (1 a) or (1 b) of Reeve (1955).

Discussion

Considerable caution is needed in using estimates of the standard error of the correlations obtained from equations (10) or (12) for tests of significance. Empirical studies of the sampling variances of genetic correlations have been made by Van Vleck and Henderson (1961) for the O—P estimate and Shook and Barr (1968) for the sib estimate, while Brown (1969) has considered the distributions of the O—P estimate of the genetic correlation coefficient and its variance. Guidelines to the accuracy of estimates of genetic correlations and their variances may be obtained from these studies. However, to our knowledge no studies of the distributions of the other correlation coefficients estimated from sib analysis, nor of their variances, have been attempted.

A computer programme has been written (for an IBM 7040 computer and adapted to a CDC 6600) to provide a general quantitative genetic analysis of nested data, of which the final 3 levels are arranged as in Table 1. Equation (8) is solved for the seven θ's described here and these are then used to obtain all estimates of sib heritabilities and correlation coefficients. Offspring-parent regression analyses can be done on request by the method suggested in this paper, and the respective heritabilities and genetic correlation coefficients are computed. Standard errors are calculated for all estimates. A partitioning of the phenotypic variance is then performed. A copy of the programme is available from Associate Professor J. S. F. Barker, Department of Animal Husbandry, University of Sydney.

Acknowledgements

The encouragement and guidance of Associate Professor J. S. F. Barker throughout this study have been invaluable. Further, we wish to thank Dr. L. P. Jones, for his helpful suggestions on the extension of Reeve's formula, and Mr. J. W. James and Dr. B. D. H. Latter for their constructive criticism of the manuscript. This work was carried out while each author held a Commonwealth Postgraduate Research Studentship.

References

1. Abe, T.: On the sampling variances of the genetic correlation estimates from analysis of variance and covariance. Jap. Poult. Sci. **6**, 209—214 (1969). — 2. Brown, G. H.: An empirical study of the distribution of the sample genetic correlation coefficient. Biometrics **22**, 63—72 (1969). — 3. Grossman, M.: Sampling variance of the correlation coefficients estimated from analyses of variance and covariance. Theoret. Appl. Genet. **40**, 357—359 (1970). — 4. Mode, C. G., Robinson, H. F.: Pleiotropism and the genetic variance and covariance. Biometrics **15**, 518—537 (1959). — 5. Reeve, E. C. R.: The variance of the genetic correlation coefficient. Biometrics **11**, 357—374 (1955). — 6. Rohde, C. A., Tallis, G. M.: Exact first- and second-order moments of estimates of components of covariance. Biometrika **56**, 517—525 (1969). — 7. Shook, G. E., Barr, G. R.: Empirical sampling variances of genetic correlations computed from sire components of variance and covariance. Paper at 63rd Annual Meet., American Dairy Science Assoc. (Mimeograph) (1968). — 8. Tallis, G. M.: Sampling errors of genetic correlation coefficients calculated from the analyses of variance and covariance. Aust. J. Statist. **1**, 35—43 (1959). — 9. Van Vleck, L. D., Henderson, C. R., Empirical sampling estimates of genetic correlations. Biometrics **17**, 359—371 (1961).

Received June 17, 1971

Communicated by J. S. F. Barker

K. Hammond and
F. W. Nicholas
Department of Animal Husbandry
University of Sydney
Sydney, N.S.W. 2006 (Australia)