J. Ind. Soc. Agril. Statist.
60(2), 2006 : 126–130

Estimation of Variance Components when Errors are Correlated by Autoregressive of Order One

N. Okendro Singh, V.K. Bhatia[1] and A.K. Paul[1]
National Research Centre on Coldwater Fisheries, Nainital District, Uttaranchal
(Received : March, 2006)

SUMMARY

In the present investigation, expressions for estimating variance components of one-way classification random model are developed in case of correlated errors which follows autoregressive of order one i.e. AR(1). These expressions are further used to see the influence of correlated errors on the estimate of heritability by half-sib method. The expected mean sums of square due to error and due to sire are overestimated and underestimated respectively when errors are negatively correlated. The former increases and the later decreases as the degree of correlation increases. When the correlation is positive, just reverse results are obtained for both the expected mean sums of squares. The heritability values obtained by neglecting the correlation present in errors are underestimated when they are negatively correlated. In contrast, heritability values are overestimated if the correlation is positive. These results are found to be consistent for all the levels of heritability. Also, heritability increases from zero to nearly four as the autoregressive coefficient increases from minus unity to approximately unity irrespective of the heritability values. Thus, when the coefficient of AR(1) is positive and is very high, heritability value goes far beyond its actual upper limit (i.e. unity) but if it is negatively correlated, it will never cross the actual minimum limit of heritability even for maximum value of auto regression coefficient, and simply it matches to the minimum limit of heritability in such a situation.

Key words: Variance components, Autoregressive coefficient, Correlated errors, Half-sib, Heritability.

1. INTRODUCTION

The information on genetic components of variances of the important characters is the prime interest of plant and animal breeders. Thus, the estimation of various genetic variances and inferring about their inheritance, based on estimates of different genetic parameters is very important from plant and animal breeding programme point of view. The development of methods of estimating variance components was initiated in the early 20th century. Fisher (1925) made a major contribution to variance component models through initiating the concept of the analysis of the variance method of estimation. Jackson (1939) dealt with a mixed model for the first time in the literature of variance component estimation. Cochran (1939) initiated for unbalanced data. Henderson (1953) gave a method of how to use unbalanced data for estimating the variance components in a difficult problem. All these theories are available for estimating the variance components for those models which are having uncorrelated errors but not for correlated error structure. So, it is required to develop the theory for estimating variance components in case of correlated error structure so as to meet the requirements of practical situations. Keeping in view of these points, in the present investigation expressions for estimating variance components when errors are having autoregressive of order one i.e., AR(1) structure has been developed. Further, the influence of correlation on the inheritance of characters is also presented by taking an example of half-sib method.

[1] *Indian Agricultural Statistics Research Institute, New Delhi-12*

ONE-WAY CLASSIFICATION MODEL

The random model for the one-way classification is

$$y_{ij} = \mu + \alpha_i + e_{ij}$$

where y_{ij} is the j^{th} observation in i^{th} class, μ is the general mean, α_i is the effect on the y-variable of its being observed on an observational unit that is in the i^{th} class, and e_{ij} is a residual error. The number of classes in the data shall be denoted by a, and the number of observations in the i^{th} class by n. Thus i = 1, 2, ...,a and j = 1, 2, ..., n.

Under the normality assumptions

$$E(e_{ij}) = 0$$
$$E(\alpha_i) = 0$$
$$E(e_{ij}^2) = \sigma_e^2$$
$$E(\alpha_i^2) = \sigma_\alpha^2$$
$$Cov(\alpha_i, \alpha_{i'}) = 0 \forall i \neq i'$$
$$Cov(\alpha_i, e_{i'j}) = 0 \forall i \text{ and } i'$$
$$Cov(e_{ij}, e_{i'j'}) = 0 \text{ except for } i = i' \text{ and } j = j'$$

The two sums of squares that are the basis of the analysis of variance of balanced data from a one-way classification are

$$SSA = \sum_{i=1}^{a} n(\bar{y}_{i.} - \bar{y}_{..})^2$$

and

$$SSE = \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{i.})^2$$

The ANOVA method of estimation is based on deriving the expected values of SSA and SSE. Searle *et al.* (1992) had given the expressions for estimating variance components of the above model as follows

$$E(MSA) = \frac{E(SSA)}{a-1} = n\sigma_\alpha^2 + \sigma_e^2$$

and

$$E(MSE) = \frac{E(SSE)}{a(n-1)} = \sigma_e^2$$

In case of the one-way classification model with errors having AR(1) structure i.e.

$$y_{ij} = \mu + \alpha_i + e_{ij}$$
$$e_{ij} = \rho e_{i(j-1)} + \eta_{ij}$$

where $|\rho| < 1$, $Var(\eta_{i1}) = \frac{1}{1-\rho^2}$

and $\eta_{ij} \approx IIDN(0,1) \text{ for } j > 1$

Assumptions

$$E(e_{ij}) = 0$$
$$E(\alpha_i) = 0$$
$$E(e_{ij}^2) = \sigma_e^2$$
$$E(\alpha_i^2) = \sigma_\alpha^2$$
$$Cov(\alpha_i, \alpha_{i'}) = 0 \forall i \neq i'$$
$$Cov(\alpha_i, e_{i'j}) = 0 \ \forall \ i \text{ and } i'$$
$$Cov(e_{ij}, e_{i'j}) = 0 \ \forall \ i \text{ and } i'$$
$$Cov(e_{ij}, e_{ij'}) = \rho^{|j-j'|}\sigma_e^2 \ \forall \ j \neq j'$$
$$Cov(e_{ij}, e_{i'j'}) = 0 \ \forall \ i \neq i' \text{ and } j \neq j'$$

In the similar fashion, the expressions for estimating the variance components are developed for the above model as follows (see Appendix-I for derivation).

$$E(MSA) = n\sigma_\alpha^2 + \sigma_e^2\left[1 - \left(\frac{2}{n}\right)\rho - \rho^2 + \left(\frac{2}{n}\right)\rho^{n+1}\right] \Big/ (1-\rho)^2$$

and

$$E(MSE) = \frac{\sigma_e^2\left[1 - 2\left(\frac{n+1}{n}\right)\rho + \left(\frac{n+1}{n-1}\right)\rho^2 - \left(\frac{2}{n(n-1)}\right)\rho^{n+1}\right]}{(1-\rho)^2} \quad (1)$$

INFLUENCE OF CORRELATED ERRORS ON THE ESTIMATE OF HERITABILITY

The half-sib analysis model can be written as follows

$$y_{ij} = \mu + s_i + e_{ij}; \ i = 1,2,...,s; \ j = 1,2,...,p$$

where

y_{ij} is the observed value on the progeny of the j^{th} dam mated to the i^{th} sire

μ is the general mean

s_i is the effect due to i^{th} sire

e_{ij} is the random effect associated with j^{th} member of the i^{th} sire

$$E(s_i) = 0; \ E(e_{ij}) = 0$$
$$E(s_i^2) = \sigma_s^2; \ E(e_{ij}^2) = \sigma_e^2$$

The value of heritability by half-sib method is given as follows

$$h^2 = \frac{4\sigma_s^2}{\sigma_s^2 + \sigma_e^2} \quad (2)$$

Suppose that sires are independent but within sire progenies are correlated. Further, assume that the correlated errors follow AR(1) i.e.

$$e_{ij} = \rho e_{i(j-1)} + \eta_{ij}$$

where $|\rho| < 1$, $\text{Var}(\eta_{i1}) = \frac{1}{1-\rho^2}$

and $\eta_{ij} = \text{IIDN}(0,1)$ for $j > 1$

In case of correlated errors, estimation of heritability is not as simple as given by equation (2). The influence of correlation is studied by using equation (1) for different values of autoregressive coefficient, ρ and is discussed below.

RESULTS AND DISCUSSION

The expressions for estimation of variance components for one-way classification random model are developed when errors are correlated by autoregressive of order one. The developed expressions of variance components are further used to see the influence of correlated errors on the estimate of heritability and is demonstrated by considering the half-sib method. Three different levels of heritability namely low (0.1), medium (0.25) and high (0.5) with their respective optimum structure (i.e., number of progenies per sire, $n = 4/h^2$) are used to plot the expected mean sums of squares given by equation (1) for different values of autoregressive coefficient, ρ. The variance of error (σ_e^2) is taken to be unity and the corresponding variance due to sire (σ_s^2) is obtained by using the relation for half-sib, $h^2 = 4\sigma_s^2/(\sigma_s^2 + \sigma_e^2)$. The expected mean sums of squares thus obtained are further used to get the corresponding heritability values by neglecting the correlation present in the error term for different values of autoregressive coefficient, ρ and are shown in Table 1. The expected mean sums of squares due to error are overestimated when the correlation is negative and they increase as degree of correlation increases. But these expected mean sums of squares are underestimated if errors are positively correlated and they decrease with increase in degree of correlation and it approaches to zero as ρ tends to unity. On the other hand, just reverse results are obtained for estimating the mean sums of squares due to sire i.e., these expected mean sums of squares are underestimated when ρ is negative and they are overestimated if the correlation is positive. As ρ tends to unity, expected mean sums of squares due to sire approaches to its maximum value. The maximum and minimum values of expected mean sums of squares due to error and due to sire respectively coincide at a point when ρ is negative and unity. These results are found to be consistent for all the levels of heritability. Further, the heritability values obtained by neglecting the

Table 1. Expected mean sums of squares when errors are correlated by AR(1) and corresponding heritability values after neglecting the correlation by half-sib method for different levels of heritability and autoregressive coefficient

ρ	$h^2 = 0.1, n = 40, \sigma_e^2 = 1$			$h^2 = 0.25, n = 16, \sigma_e^2 = 1$			$h^2 = 0.5, n = 8, \sigma_e^2 = 1$		
	E (MSE)	E (MSA)	h^2	E (MSE)	E (MSA)	h^2	E (MSE)	E (MSA)	h^2
-1.000	1.026	1.026	0.000	1.067	1.067	0.000	1.143	1.143	0.000
-0.900	1.024	1.091	0.006	1.061	1.145	0.020	1.130	1.231	0.044
-0.800	1.022	1.149	0.012	1.057	1.208	0.035	1.120	1.305	0.081
-0.700	1.021	1.214	0.019	1.053	1.273	0.052	1.109	1.376	0.117
-0.600	1.019	1.287	0.026	1.048	1.346	0.070	1.099	1.450	0.154
-0.500	1.017	1.370	0.034	1.043	1.428	0.090	1.087	1.532	0.194
-0.400	1.014	1.464	0.044	1.036	1.521	0.114	1.074	1.622	0.240
-0.300	1.012	1.573	0.055	1.029	1.627	0.140	1.060	1.726	0.291
-0.200	1.008	1.699	0.067	1.021	1.751	0.171	1.043	1.844	0.351
-0.100	1.005	1.848	0.082	1.011	1.895	0.207	1.023	1.982	0.419
0.000	1.000	2.026	0.100	1.000	2.067	0.250	1.000	2.143	0.500
0.100	0.994	2.242	0.122	0.986	2.273	0.302	0.973	2.334	0.596
0.200	0.988	2.510	0.148	0.969	2.528	0.365	0.940	2.565	0.711
0.300	0.979	2.852	0.183	0.948	2.847	0.445	0.899	2.847	0.852
0.400	0.967	3.303	0.228	0.920	3.261	0.549	0.849	3.199	1.028
0.500	0.951	3.926	0.290	0.883	3.817	0.688	0.785	3.645	1.251
0.600	0.928	4.838	0.381	0.831	4.598	0.883	0.703	4.221	1.539
0.700	0.890	6.303	0.528	0.753	5.764	1.174	0.595	4.977	1.917
0.800	0.821	9.026	0.800	0.629	7.637	1.643	0.452	5.982	2.419
0.900	0.652	15.592	1.457	0.411	10.901	2.459	0.259	7.328	3.093
0.9999	0.001	40.972	3.995	0.001	17.058	3.998	0.000	9.141	3.999
1.0001	-0.001	41.079	4.005	-0.001	17.075	4.002	0.000	9.145	4.001

correlation present in errors are underestimated when errors are negatively correlated. In contrast, heritability values are overestimated if the correlation is positive. The same trend follows for all the levels of heritability. Also, heritability increases from zero to nearly four as the autoregressive coefficient increases from minus unity to approximately unity. Thus, when the coefficient of AR(1) is positive and is very high, heritability value goes far beyond its actual upper limit (i.e., unity) but if it is negatively correlated, it will never cross the actual minimum limit of heritability even for maximum value of autoregressive coefficient, and simply it matches to the minimum limit of heritability in such a situation.

REFERENCES

Cochran, W.G. (1939). The use of analysis of variance in enumeration by sampling. *J. Amer. Stat. Assoc.*, **34**, 492-510.

Fisher, R.A. (1925). *Statistical Methods for Research Workers*. 1st ed., Oliver & Boyd, Edinburgh, London.

Henderson, C.R. (1953). Estimation of variance and covariance components. *Biometrics*, **9**, 226-252.

Jackson, R.W.B. (1939). Reliability of method tests. *Brit. J. Psychol.*, **29**, 267-287.

Searle, S.R., Casella, G. and McCulloch, C.E. (1992). *Variance Components*. John Wiley, New York.

APPENDIX-I

Random model for the one-way classification is

$$y_{ij} = \mu + \alpha_i + e_{ij};\ i = 1,2,\ldots,a \text{ and } j=1,2,\ldots,n$$

$$e_{ij} = \rho e_{i(j-1)} + \eta_{ij}$$

where $|\rho| < 1, \mathrm{Var}(\eta_{i1}) = \dfrac{1}{1-\rho^2}$

and $\eta_{ij} \approx \mathrm{IIDN}(0,1)$ for $j>1$

Then

$$\bar{y}_{i.} = \mu + \alpha_i + \bar{e}_{i.} \text{ for } \bar{e}_{i.} = \frac{1}{n}\sum_{j=1}^{n} e_{ij}; \qquad \bar{y}_{..} = \mu + \bar{\alpha}_{.} + \bar{e}_{..}$$

where $\bar{\alpha}_{.} = \dfrac{1}{a}\sum_{i=1}^{a}\alpha_i$ and $\bar{e}_{..} = \dfrac{1}{a}\sum_{i=1}^{a}\bar{e}_{i.}$

Define

$$SSA = \sum_{i=1}^{a} n(\bar{y}_{i.} - \bar{y}_{..})^2 \text{ and } SSE = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2$$

Now

$$E(SSA) = E\left[n\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2\right]$$

$$= E\left[n\sum_{i=1}^{a}\{(\alpha_i - \bar{\alpha}_{.}) + (\bar{e}_{i.} - \bar{e}_{..})\}^2\right]$$

$$= \left[n\sum_{i=1}^{a}\{E(\alpha_i - \bar{\alpha}_{.})^2 + E(\bar{e}_{i.} - \bar{e}_{..})^2\}\right]$$

Since covariance term vanishes.

$$= n\sum_{i=1}^{a}[\mathrm{Var}(\alpha_i - \bar{\alpha}_{.}) + \mathrm{Var}(\bar{e}_{i.} - \bar{e}_{..})]$$

$$= n\sum_{i=1}^{a}\left(\sigma_\alpha^2 + \frac{1}{a}\sigma_\alpha^2 - \frac{2}{a}\sigma_\alpha^2\right)$$

$$+ n\sum_{i=1}^{a}\left[\frac{\sigma_e^2(n - 2\rho - n\rho^2 + 2\rho^{n+1}}{n^2(1-\rho)^2} + \frac{\sigma_e^2(n - 2\rho - n\rho^2 + 2\rho^{n+1})}{an^2(1-\rho)^2} - 2\frac{\sigma_e^2(n - 2\rho - n\rho^2 + 2\rho^{n+1})}{an^2(1-\rho)^2}\right]$$

$$= n(a-1)\sigma_\alpha^2 + \frac{(a-1)}{n}\sigma_e^2\left[n - 2\rho - n\rho^2 + 2\rho^{n+1}\right]/(1-\rho)^2$$

Therefore

$$E(MSA) = \frac{E(SSA)}{(a-1)}$$

$$= n\sigma_\alpha^2 + \frac{\sigma_e^2\left[1 - \left(\frac{2}{n}\right)\rho - \rho^2 + \left(\frac{2}{n}\right)\rho^{n+1}\right]}{(1-\rho)^2}$$

$$E(SSE) = E\left[\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2\right]$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{n} E(e_{ij} - \bar{e}_{i.})^2$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{n} \mathrm{Var}(e_{ij} - \bar{e}_{i.})$$

$$= \sum_{i=1}^{a} \sum_{j=1}^{n} \left[\text{Var}(e_{ij}) + \text{Var}(\bar{e}_{i.}) - 2\text{Cov}(e_{ij} - \bar{e}_{i.})\right]$$

$$= an\sigma_e^2 - \frac{a}{n} \frac{\left[n - 2\rho - n\rho^2 + 2\rho^{n+1}\right]}{(1-\rho)^2} \sigma_e^2$$

$$= a\left[n - \frac{(n - 2\rho - n\rho^2 + 2\rho^{n+1})}{n(1-\rho)^2}\right]\sigma_e^2$$

$$= a\left[\frac{n(n-1) - 2\rho(n^2-1) + \rho^2 n(n+1) - 2\rho^{n+1}}{n(1-\rho)^2}\right]\sigma_e^2$$

Therefore

$$E(MSE) = \frac{E(SSE)}{a(n-1)}$$

$$= \sigma_e^2\left[1 - 2\left(\frac{n+1}{n}\right)\rho + \left(\frac{n+1}{n-1}\right)\rho^2 - \left(\frac{2}{n(n-1)}\rho^{n+1}\right)\right] \Big/ (1-\rho)^2$$

where

$$\text{Var}(\bar{e}_{i.}) = \text{Var}\left(\frac{1}{n}\sum_{j=1}^{n} e_{ij}\right)$$

$$= \frac{1}{n^2}\text{Var}(e_{i1} + e_{i2} + \ldots + e_{in})$$

$$= \frac{1}{n^2}\text{Cov}\{(e_{i1} + e_{i2} + \ldots + e_{in}),\ (e_{i1} + e_{i2} + \ldots + e_{in})\}$$

$$= \frac{1}{n^2}\sigma_e^2\{(1 + \rho + \rho^2 + \ldots + \rho^{n-1}) + (\rho + 1 + \rho + \rho^2 + \ldots + \rho^{n-2}) + \ldots + (\rho^{n-1} + \rho^{n-2} + \ldots + \rho^2 + \rho + 1)\}$$

$$= \frac{1}{n^2}\sigma_e^2\left[n - 2\rho - n\rho^2 + 2\rho^{n+1}\right] \Big/ (1-\rho)^2$$

Similarly

$$\text{Var}(\bar{e}_{..}) = \text{Var}\left(\frac{1}{a}\sum_{i=1}^{a} \bar{e}_{i.}\right)$$

$$= \frac{1}{a^2}\sum_{i=1}^{a} \text{Var}\left(\frac{1}{n}\sum_{j=1}^{n} e_{ij}\right)$$

$$= \frac{1}{a^2 n^2}\sum_{i=1}^{a} \text{Var}(e_{i1} + e_{i2} + \ldots + e_{in})$$

$$= \frac{1}{an^2}\sigma_e^2\left[n - 2\rho - n\rho^2 + 2\rho^{n+1}\right] \Big/ (1-\rho)^2$$

$$\text{Cov}(\bar{e}_{i.}, \bar{e}_{..}) = \text{Cov}\left(\bar{e}_{i.}, \frac{1}{a}\sum_{i=1}^{a} \bar{e}_{i.}\right)$$

$= \frac{1}{a}\text{Var}(\bar{e}_{i.})$ Since i's are independent and hence

$$\text{Cov}(\bar{e}_{i.}, \bar{e}_{i'.}) = 0 \quad \forall\ i = i'$$

$$= \frac{1}{a}\text{Var}\left(\frac{1}{n}\sum_{j=1}^{n} e_{ij}\right)$$

$$= \frac{1}{an^2}\text{Var}(e_{i1} + e_{i2} + \ldots + e_{in})$$

$$= \frac{1}{an^2}\text{Cov}\{(e_{i1} + e_{i2} + \ldots + e_{in}),\ (e_{i1} + e_{i2} + \ldots + e_{in})\}$$

$$= \frac{1}{an^2}\sigma_e^2\{(1 + \rho + \rho^2 + \ldots + \rho^{n-1}) + (\rho + 1 + \rho + \rho^2 + \ldots + \rho^{n-2}) + \ldots + (\rho^{n-1} + \rho^{n-2} + \ldots + \rho^2 + \rho + 1)\}$$

$$= \frac{1}{an^2}\sigma_e^2\left[n - 2\rho - n\rho^2 + 2\rho^{n+1}\right] \Big/ (1-\rho)^2$$

$$\text{Cov}(e_{ij}, \bar{e}_{i.}) = \text{Cov}\left\{e_{ij}, \left(\frac{1}{n}\sum_{j=1}^{n} e_{ij}\right)\right\}$$

$$= \frac{1}{n}\text{Cov}\{e_{ij}, (e_{i1} + e_{i2} + \ldots + e_{in})\}$$

$$= \frac{1}{n}\{\text{Cov}(e_{ij}, e_{i1}) + \text{Cov}(e_{ij}, e_{i2}) + \ldots + \text{Cov}(e_{ij}, e_{in})\}$$

$$= \frac{1}{n}\sigma_e^2\left\{\rho^{|j-1|} + \rho^{|j-2|} + \ldots + \rho^{|j-n|}\right\}$$

Therefore

$$\sum_{i=1}^{a}\sum_{j=1}^{n} \text{Cov}(e_{ij}, \bar{e}_{i.}) = \frac{1}{n}\sigma_e^2 \sum_{i=1}^{a}\sum_{j=1}^{n}\left\{\rho^{|j-1|} + \rho^{|j-2|} + \ldots + \rho^{|j-n|}\right\}$$

$$= \frac{a}{n}\sigma_e^2 \sum_{j=1}^{n}\left\{\rho^{|j-1|} + \rho^{|j-2|} + \ldots + \rho^{|j-n|}\right\}$$

$$= \frac{a}{n}\sigma_e^2\left[n - 2\rho - n\rho^2 + 2\rho^{n+1}\right] \Big/ (1-\rho)^2$$