DATA UNBALANCE AND THE ESTIMATION OF HERITABILITY USING PATERNAL HALF-SIB TECHNIQUES

JOHN M. RENDEL,* G.A. WICKHAM and A.L. RAE

Department of Animal Science,

Massey University,

Palmerston North, NZ.

* Present address: Livestock Improvement Corporation Ltd, Private Bag 3016,

Hamilton, NZ.

INTRODUCTION

In devising efficient breeding schemes, heritability estimates are required. In recent years, not only has the number of these estimates grown rapidly, but also the number of methods used to estimate them. It is usually necessary to assess the likely accuracy of several estimates in deciding the value to use.

This paper uses simulated data to compare three methods of variance component estimation, data set size and level of unbalance on the accuracy of heritabilities in estimating population values. It was undertaken as part of a larger study to ascertain the accuracy of the standard error in estimating the variance about a population heritability.

MATERIALS AND METHODS

Data sets were constructed using a random number generator. The data were based on a model that included a general mean (18), a random effect to represent sires ($\sigma^2 = 0.6783$) and a random effect to represent individuals within sire groups in a paternal half sib analysis ($\sigma^2 = 11.0106$). These values were those obtained for weaning weight in a flock analysed by Rendel (1985) and resulted in a heritability of 0.2321.

Sire and error variances were estimated for each of the 100 replicates of each data set by 3 methods:

Henderson's method 1 (HM) (Henderson 1953);

Maximum likelihood (ML) (Hartley and Rao 1967);

Restricted maximum likelihood (REML) (Patterson and Thompson 1974, K Meyer's programme).

The stopping criterion for ML was a difference of 4×10^{-8} between successive likelihoods; for REML a change of 0.005% in the sire variance between iterations.

The mean heritability estimate for each set of 100 replicates was calculated and deviations of these from the population value (0.2321) were calculated to indicate any bias. Mean squared errors (MSE) for each set were derived from the deviations of the individual estimates from the population value.

Initially sets were constructed which ranged from 20 to 200 sires with mean numbers of progeny per sire of 20 to 100. In some cases the number of progeny per sire was variable, the standard deviations being 0 (SD0), 2 (SD2) and 7 (SD7). Without departing from a normal distribution it was not possible to achieve a higher standard deviation when the mean number of progeny was low.

The unbalancedness of the data was estimated by the parameter γ, which Ahrens and Pincus (1981) derived as:

$$\gamma = \frac{\mathrm{m}}{\bar{\mathrm{n}} \sum_{j=1}^{m} \frac{1}{\mathrm{n}_j}}$$

where m = number of sires
n_j = number of observations on the j^{th} sire (j=1,2,...,m)
$\bar{\mathrm{n}}$ = mean number of observations per sire.

Theoretical values of γ range from near 0 (extreme unbalance) to 1 (balanced). Estimates for the data sets constructed indicated that these were not nearly as unbalanced as many sets of sheep data used for heritability estimation. Hence further sets were constructed with 100 sires and a mean of 100 progeny per sire and standard deviations of 15, 25 and 29. The γ's achieved (0.98, 0.93 and 0.91 respectively) again did not approach commercial flock values.

Finally, data sets were constructed based on numbers of sires and progeny in 6 flocks representative of the 31 commercial sheep flocks studied by Rendel (1985). The distribution of progeny per sire was very non-normal in 5 of these flocks. The γ's are indicated in Table 1.

Table 1. The number of sires (Sires), mean number (Mean) and standard deviation (SD) of progeny per sire and estimate of unbalance (γ) for flocks A to F.

	A	B	C	D	E	F
Sires	105	44	84	60	60	87
Mean	63.61	81.59	81.51	99.03	97.61	101.75
SD	27.99	37.99	45.08	67.08	61.53	51.25
γ	0.861	0.674	0.524	0.402	0.391	0.389

RESULTS AND DISCUSSION

The mean heritability estimates were close to the population value for all sets and all methods. There were small downward biases of as much as -0.03 in most of the data sets with 20 or 50 sires. In only one (<1%) of the data sets with 100 sires or more did the mean estimate and the population value differ by more than 0.01. In flock A it was not possible to get the REML estimates as the solutions to the equations went outside the parameter space during iteration. This was a deficiency of the algorithm used not of REML.

The differences obtained are similar to those reported by Olausson and Rönningen (1975) for heritabilities of 0.1 and 0.5, Rothschild et al. (1979) for a heritability of 0.3, and Rönningen (1972) for heritabilities of 0.1 and 0.3.

The mean squared errors (MSE) of the data types SD0, SD2 and SD7 decreased with both increasing numbers of sires and number of progeny per sire (Table 2). There appears to be a distinct levelling off of the MSE at 100 sires with 20 progeny per sire, and at 50 sires for the remaining progeny per sire groups. There was little difference in MSE due to methods of variance component estimation, the largest being with data sets of 20 sires with 20 progeny per sire.

Table 2. The mean squared errors (MSE) of the heritability estimates estimated by HM (H), REML (R) and ML (M) for data types SD0, SD2 and SD7 (Number of sires on the vertical and mean number of progeny per sire on the horizontal).

		SD0				SD2				SD7			
		20	50	70	100	20	50	70	100	20	50	70	100
	H	17.49	6.45	6.62	6.06	16.40	9.72	6.51	7.57				
20	R	17.14	6.45	6.62	6.06	16.05	9.70	6.47	7.52				
	M	15.20	6.68	6.55	5.69	15.36	9.15	6.56	7.18				
	H	5.20	3.89	2.41	2.06	9.85	4.04	2.77	2.47	6.47	3.22	2.74	2.05
50	R	5.20	3.89	2.41	2.06	9.76	4.03	2.76	2.47	6.15	3.23	2.69	2.10
	M	5.46	3.81	2.38	1.90	9.65	4.06	2.82	2.45	6.26	3.38	2.64	2.08
	H	2.84	2.19	1.88	1.15	2.99	1.53	1.46	1.02	3.26	1.75	2.06	1.54
100	R	2.84	2.19	1.88	1.15	2.95	1.54	1.47	1.02	3.34	1.70	1.98	1.54
	M	2.86	2.09	1.84	1.13	2.96	1.54	1.48	1.02	3.39	1.61	1.95	1.57
	H	1.78	1.50	1.18	0.80	2.73	1.28	0.88	0.96	2.71	1.41	1.39	0.74
150	R	1.78	1.50	1.18	0.80	2.73	1.28	0.88	0.96	2.74	1.41	1.32	0.72
	M	1.79	1.48	1.16	0.78	2.71	1.30	0.87	0.95	2.79	1.41	1.30	0.73
	H	2.31	1.00	1.06	0.87	1.43	0.69	0.67	0.61	1.43	1.20	0.37	0.65
200	R	2.31	1.00	1.06	0.87	1.43	0.69	0.67	0.61	1.41	1.21	0.38	0.65
	M	2.27	0.98	1.05[1]	0.77[2]	1.45	0.69	0.66	0.62	1.40	1.23	0.37	0.65[1]

[1] = 98 replicates [2] = 87 replicates

There was no consistent effect of unbalance on the MSE for flocks A to F (Table 3). The MSE of the heritabilities estimated from variance components by HM were larger than either REML or ML. This may indicate that HM has a larger error in estimating the heritability, or it may have been a reflection on the algorithm used to solve the equations needed to obtain the ML and REML estimates. The stopping criteria may have been too large and the point of maximum likelihood had not been reached. The MSE were similar to data types SD0, SD2 and SD7 for similar numbers of sires and progeny per sire.

Rothschild et al. (1979) reported similar MSE's of heritabilities estimated by HM and ML for balanced data. With unbalanced data the MSE's were slightly larger for heritabilities estimated using HM than ML. Lin and M^{C}Allister (1984) reported MSE's, from unbalanced data, were larger for ML than for HM or REML.

Table 3. The mean squared error (MSE) of heritability estimates estimated by HM (H), REML (R) and ML (M) for flocks A to F.

	H	R	M
A	1.68	.	1.56
B	3.62	3.09	3.23
C	2.75	1.89	1.88
E	4.89	3.75	3.88
F	1.80	1.50	1.55

CONCLUSIONS

The heritabilities estimated using HM, REML and ML were unaffected by the level of data unbalance. The heritability MSE's of the REML and ML estimates were lower than those from HM for the distributions based on the 6 flocks. This is probably a reflection on the algorithms used to solve the equations for REML and ML, especially as very little difference would be expected between the accuracy of the 3 methods in a 1-way model.

REFERENCES

AHRENS, H. and PINCUS, R. (1981). Biometrical Journal 23: 227.

HARTLEY, H.O. and RAO, J.N.K. (1967). Biometrika 54: 93.

HENDERSON, C.R. (1953). Biometrics 9: 226.

LIN, C.Y. and M^{c}ALLISTER, A.J. (1984). Journal of Dairy Science 67: 2389.

OLAUSSON, A. and RÖNNINGEN, K. (1975). Acta Agriculturae Scandinavica 25: 201.

PATTERSON, H.D. and THOMPSON, R. (1974). Proceedings of the 8th International Biometrics Conference: 197.

RENDEL, J.M. (1985). M.Ag.Sci. Thesis, Massey University, NZ.

RÖNNINGEN, K. (1972). Acta Agriculturae Scandinavica 22: 90.

ROTHSCHILD, M.F., HENDERSON, C.R. and QUAAS, R.L. (1979). Journal of Dairy Science 62: 996.