Journal of Genetics Volume 61 Number 3 December 1974

A SCALING WHICH REMOVES BIAS OF ABEYWARDENA'S ESTIMATOR OF REPEATABILITY[1]

J. J. Rutledge

Department of Animal Sciences,

University of Vermont, Burlington - 05401

INTRODUCTION

Abeywardena (1972) proposed that the proportion of the variance accounted for by a principal component (usually the first) of the correlation matrix of observed scores on individuals was a "far superior" estimator of repeatability to those currently in use. It is the purpose of this note to show that Abeywardena's estimator is highly sensitive to the number of scores per individual (p), and that bias may be large for small p. A scaling of his estimator which removes bias will be suggested.

DISCUSSION

Repeatability (ρ) is defined as the correlation among scores of the same individual (Lush, 1945). Kempthorne (1969) and Turner and Young (1969) give particularly good discussions of repeatability, its estimation and uses. An assumption in repeatability estimation is that the correlation among scores is homogeneous over all pairs of scores. Using the definition and assumption, one can write the population correlation matrix required for a principal components analysis as:

$$\Sigma = \begin{bmatrix} 1 & \rho & . & . & . & \rho \\ \rho & 1 & . & . & . & \rho \\ . & . & . & . & . & . \\ \rho & \rho & . & . & . & 1 \end{bmatrix}_{p\mathrm{x}p} \qquad \text{Eq. 1.}$$

---

[1]University of Vermont Agricultural Experiment Station Journal Article No. 331.

The component structure of a matrix such as Σ is discussed by Morrison (1967). The first characteristic vector of Σ is:

$$CV_1 = \left[p^{-\frac{1}{2}}, p^{-\frac{1}{2}}, p^{-\frac{1}{2}}, \ldots, p^{-\frac{1}{2}} \right]_{1xp} \qquad \text{Eq. 2.}$$

The proportion of the variance, say λ, accounted for by a vector such as CV_1 ("same sign and nearly equal weights to all elements") is claimed to be a "truer measure of repeatability" by Abeywardena. Following Morrison (1967), this proportion is:

$$\lambda = (1 + (p-1)\rho)/p. \qquad \text{Eq. 3.}$$

Note that λ is a function of p, and

$$\lim_{p \to \infty} \lambda = \rho. \qquad \text{Eq. 4.}$$

Equation 4 indicates that $\lambda = \rho$ only for infinite p. A scaling of λ which equals ρ for all p is:

$$\theta = (\lambda-1/p)(p/(p-1)). \qquad \text{Eq. 5.}$$

With Σ, which is a model for the dependence structure of a repeatability data set, theoretical values of λ and θ were calculated for several values of p and ρ. These are shown in Table 1. That λ is biased is clear, especially for small p and small ρ; θ is unbiased.

The preceding analyses are exact algebraic solutions based on Σ. The properties of λ and θ as estimators of ρ were further examined from a stochastic viewpoint using data from Abeywardena's simulation study. Each of the 45 possible 2x2 correlation matrices was extracted from the 10x10 correlation matrix of Abeywardena's situation 1. A principal components analysis was calculated for each 2x2, and $\hat{\lambda}$ and $\hat{\theta}$ were calculated. The true value of repeatability in the population from which Abeywardena's sample was drawn was 0.67. The mean $\hat{\lambda}$ from the 45 analyses was 0.86, with range 0.70 to 0.93. Thus, $\hat{\lambda}$ consistently overestimated ρ. The mean $\hat{\theta}$ from the analyses was 0.69, with range 0.40 to 0.86. It is concluded that Abeywardena's estimator of repeatability is inappropriate, especially for small p. However, a scaling of his estimator which removes bias can be used.

SUMMARY

Abeywardena's (J. Genetics, 61, 27-51, 1972) estimator of repeatability was shown to be biased, using both the theoretical model for the dependence structure of a repeatability data set and analyses of his simulated data. A scaling involving the number of observations per individual corrected the bias.

Table 1. Values of Abeywardena's estimator of repeatability (λ) and the unbiased estimator (θ) for various numbers of scores per individual (p) at several levels of repeatability (ρ).

	ρ							
	.2		.4		.6		.8	
p	λ	θ	λ	θ	λ	θ	λ	θ
2	.60	.20	.70	.40	.80	.60	.90	.80
4	.40	.20	.55	.40	.70	.60	.85	.80
6	.33	.20	.50	.40	.67	.60	.83	.80
8	.30	.20	.48	.40	.65	.60	.83	.80
10	.28	.20	.46	.40	.64	.60	.82	.80
100	.21	.20	.41	.40	.60	.60	.80	.80
1000	.20	.20	.40	.40	.60	.60	.80	.80

LITERATURE CITED

Abeywardena, V. (1972). An application of principal component analysis in genetics. J. Genet. 61, 27.

Kempthorne, Oscar (1969). An Introduction to Genetic Statistics. The Iowa State University Press, Ames, Iowa.

Lush, Jay L. (1945). Animal Breeding Plans. The Iowa State College Press, Ames, Iowa.

Morrison, Donald F. (1967). Multivatiate Statistical Methods. McGraw-Hill Book Co., New York.

Turner, Helen Newton and Young, Sydney S. Y. (1969). Quantitative Genetics in Sheep Breeding. Cornell University Press, Ithaca, New York.