267

JA 258

Indian J. Genet., **45** (2) : 288 - 291 (1985)

THE ADEQUACY OF GENETIC MODEL BASED ON COMPONENTS OF MEAN WHEN SOME GENERATION VARIANCES ARE ZERO

M. SINGH

International Crops Research Institute for the Semi-Arid Tropics, Patancheru 502 324

(Accepted : May 24, 1985)

ABSTRACT

The unified theory of least squares of Rao has been demonstrated to test the adequacy of genetic model from plant breeding data when variances of some generations are zero (and thus the weights are infinite).

Index words : Generalised inverse, genetic parameters.

It is sometimes required in plant breeding experiments to estimate the parameters m, [d], [h] following the joint scaling test of Cavalli (1952), which is explained in Mather and Jinks (1971, Ch 4, pp 73). For the estimation of these parameters using the above scaling test, one raises four or more types of generations and makes use of weighted least squares taking weights as the reciprocal of variances of each generation mean and tests the sum of weighted squared differences between observed and expected generation means against a chi-square statistic.

It the variances of each generation are non-zero, then weights can be easily found by inverting the variances. Sometimes in plant breeding experiments, observations from some generations may be constant yielding zero as estimates of variances. The usual weights for these will be infinite, so normal scaling tests cannot be applied.

In order to tackle this problem, the unified theory of least squares reported by Rao (1973a) has been proposed in this note and illustrated with the help of the data from a plant breeding trial conducted at ICRISAT centre, India.

RAO'S UNIFIED THEORY OF LEAST SQUARES

The above situation can be considered as if an n-component vector(generation means), $\bar{Y}$ is being observed with possibly a singular variance-covariance

*ICRISAT Journal Article No. 258

matrix, $\sigma^2\mathbf{G}$ $(n \times n)$ while the interest lies in estimating the parameters of the m-component vector β, of which the expected value of the random vector is a linear combination.

We shall reproduce here, for the sake of completeness the results from Rao (1973a,b) which are required for our purpose.

A situation often met is when **G** is non-singular, for which the weighted or generalised least squares estimates, calculated by Aitken's (1935) method, is an optimal choice. In the situations when **G** is singular Aitken's procedure cannot be applied. However, in order to deal with this case Rao's procedure (1973a) provides the estimation of β which uses the following steps. These steps consist of finding a matrix **M** such that the estimate $\hat{\beta}$ of β is a stationary value of

$$S = (Y - \mathbf{X}\beta)' \mathbf{M} (Y - \mathbf{X}\beta)$$

where **X** is the $n \times m$ matrix of known constants in the model of the expected value $\bar{Y}$, $E(\bar{Y}) = \mathbf{X}\beta$. One choice of **M** proposed in Rao (1973 a) is

$\mathbf{M} = (\mathbf{G} + \mathbf{X}\,\mathbf{U}\,\mathbf{X}')^-$ for any g–inverse (generalised inverse)

(where a g–inverse $\mathbf{A}^-$ of a matrix **A** is given by $\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$) such that Rank $(\mathbf{G}:\mathbf{X}) = \text{Rank}\ (\mathbf{G} + \mathbf{X}\,\mathbf{U}\,\mathbf{X}')$. Also one simple choice of **U** is $\mathbf{U} = k^2\,\mathbf{I}$, $k \neq 0$ where **I** is the identity matrix of the order n. k can be zero when the column space of **X** is contained in the column space of **G**. Further define $\mathbf{T} = \mathbf{G} + \mathbf{X}\mathbf{U}\mathbf{X}'$.
We thus find an estimate of β as a solution of $\mathbf{X}'\,\mathbf{T}^-\mathbf{X}\,\beta = \mathbf{X}'\,\mathbf{T}^- Y$

given by $\hat{\beta} = (\mathbf{X}'\,\mathbf{T}^-\mathbf{X})^-\,\mathbf{X}'\,\mathbf{T}^-\,Y$

and that of $\hat{\sigma}^2 = (Y - \mathbf{X}\,\beta)'\,\mathbf{T}^-\,(Y - \mathbf{X}\,\beta)/f$

where $f = R\,(\mathbf{G} : \mathbf{X}) - R\,(\mathbf{X}) = R(\mathbf{T}) - R(\mathbf{X})$ and $R(\mathbf{A}) =$ rank of **A** (we take k to be equal to 1 in our computation). The variance of a linear combination $p'\beta$ of β is given by $V\ (p'\beta) = \sigma^2 p'\ [(\mathbf{X}'\,\mathbf{T}^-\,\mathbf{X})^- - \mathbf{U}]p$. For further details refer to Rao (1973a,b).

A GENSTAT computer program was prepared to execute the result for $\hat{\beta}$ and $Q = (Y - \mathbf{X}\beta)'\,\mathbf{T}^-\,(Y - \mathbf{X}\beta)$ to measure the appropriateness of the fit. Q was tested against χ_3^2. This program when run with data in the book by Mather & Jinks (1971, pp. 74; Table 13) gave the same results as reported by them.

We now illustrate the above procedure with the data from a plant breeding trial conducted at ICRISAT Centre, India.

The six generations used were raised at ICRISAT Centre, India from two groundnut parents, namely, Gangapuri (P_1) and EC 76 (P_2) and the second rust score (RS-II) recorded.

Using the inverse sine transformation on RS-II the following means and variances of the six generations were observed. We are following the notations of Mather & Jinks (1971) to denote generation names.

Generation	P_1	P_2	F_1	F_2	B_1	B_2
No. of observations	20	15	20	226	194	92
Mean	9.000	4.133	9.000	7.831	8.860	6.728
Variance	0.000	0.409	0.000	3.660	0.410	4.331

We take here n=6, m=3.

$$Y=(\bar{P}_1, \bar{B}_1, \bar{F}_1, \bar{F}_2, \bar{B}_2, \bar{P}_2,)\ ;$$

$$\sigma^2 G=\text{diag}\ (0.0,0.002113,0.000,0.016195,\ 0.047076,0.027267);$$

$$X=\begin{Bmatrix} 1 & 1 & 0 \\ 1 & .5 & .5 \\ 1 & 0 & 1 \\ 1 & 0 & .5 \\ 1 & -.5 & .5 \\ 1 & -1 & 0 \end{Bmatrix} : \beta=(m,\ [d],\ [h]\ ).$$

The 'bar' ($\bar{p}$) over a generation name denotes the mean of the generation (p) and diag represents the diagonal matrix.

The following values were obtained.

(i) A particular set of estimates of m, [d], [h] parameters

$$\hat{m}=6.5931\quad \hat{[d]}=2.4069\quad \hat{[h]}=2.4069$$

(ii) Variances and covariances of $\hat{m}$, $\hat{[d]}$, $\hat{[h]}$:

$$V(\hat{m})=V(\hat{[d]})=V(\hat{[h]})=0.0055$$

$$\text{Cov}\ (\hat{m},\ \hat{[d]})=-.00545,\ \text{Cov}\ (\hat{m},\ \hat{[h]})=-.00545$$

$$\text{Cov}\ (\hat{[d]},\ \hat{[h]})=.00546$$

(iii) Generation means and their expected values;

Generation	Observed generation mean	Expected generation mean
P_1	9.000	9.000
B_1	8.860	9.000
F_1	9.000	9.000
F_2	7.830	7.797
B_2	6.728	6.593
P_2	4.133	4.186

(iv) Weighted sum of squares of deviations between observed and expected means of the six generations

$$Q=9.8340$$

The degrees of freedom of Q=R(T)−Rank (X)=6−3=3. A comparison with chi square value (P=.05)=7.815 shows that this test indicates a significant departure from the model.

It may be noted that generalised inverse of a matrix is not unique. Thus the values of estimates of the parameter would depend on the choice of the generalised inverse but the measure of the appropriateness of fit Q is valid. However, with the use of the generalised inverse function (INV) available in GENSTAT we did not get any difference in the computed values of the estimates of parameters and Q, at any rate, upto three places of decimals and for the values of k=.5,1,1.5,2,2.5,3,5 and 10 when applied on the above data.

ACKNOWLEDGEMENTS

Thanks are due to Dr. S.L. Dwivedi for pointing out the problem, Mr. B. Gilliver, Dr. D.G. Faris and a referee for some useful suggestions and Mr. C.P. Jaiswal for typing the manuscript.

REFERENCES

Aitken, A.C. (1935). On least squares and linear combinations of observations. *Proc. Roy Soc. Edin. A*, **55** : 42-48.

Cavalli, L.L. (1952). An analysis of linkage in quantitative inheritance. In : E.C.R. Rieve and C.H. Waddington (Eds). *Quantitative Inheritance* : 135-144, HMSO, London.

Mather, K and J.L. Jinks (1971). *Biometrical Genetics*, 2nd Ed. Chapman and Hall. London

Rao, C.R. (1973a). *Linear Statistical Inference and its Applications*. 2nd Ed. John Wiley. N.Y

Rao, C.R. (1973b). Unified theory of least squares. *Communications in Statistics*, **1**, 1-8