STATISTICS UNIT

Report No. 3/86

CONSULTANT'S REPORT TO ICRISAT

ON

VARIANCE COMPONENTS ESTIMATION IN AGRICULTURAL EXPERIMENTS WITH POSSIBLE APPLICATION TO ICRISAT DATA

Y.P. CHAUBEY

International Crops Research Institute for the Semi-Arid Tropics

Patancheru 502 324, Andhra Pradesh, India.

File name:YPChaubey on floppy disk
entered in rainbow

VARIANCE COMPONENTS ESTIMATION IN AGRICULTURAL EXPERIMENTS WITH POSSIBLE APPLICATION TO ICRISAT DATA

Y.P.CHAUBEY+
Visiting Scientist, Indian Statistical Institute, Calcutta.
On leave from Concordia University, Montreal, Canada.

1. INTRODUCTION

In agricultural experiments, especially in plant breeding, quite often interest lies in estimating different components of genetic variance either for the purpose of studying their relative magnitudes or for estimating certain functions of them e.g. heritability (see Kempthorne and Tandon 1953), average degree of dominance (see Comstock and Robinson 1948) etc. To be more specfic, let y denote the phenotypic value of certain trait, according to Wright's model it can be decomposed as

$$y = g + e \tag{1.1}$$

where g is the genotypic value and e reflects deviation of g from y. The variances are similarly decomposed, i.e.

$$\sigma_y^2 = \sigma_g^2 + \sigma_e^2 \tag{1.2}$$

if genotypes are randomly distributed relative to variations in environment. Genotypes may be randomized in a completely randomized design giving a one way lay out with a model as

(1.3) $$y_{ij} = \mu + a_i + e_{ij},\ j=1,\ldots,n_i,\ i=1,\ldots,t.$$

re y_{ij} is the j-th observation corresponding to i-th genotype, μ is the mean effect, a is the i-th genotype effect and e_{ij} is the random error. For the usual assumptions see Rao (1973). The variation in y is decomposed as

(1.4) $$\sigma_y^2 = \sigma_a^2 + \sigma_e^2$$

σ_a^2 and σ_e^2 being called the variance components. Variations in above models are also possible with the change in experimental design. We describe here, e.g. in the radomized block design, the model

(1.5) $$y_{ijk} = \mu + a_i + \beta_j + e_{ijk},\ i=1,\ldots,t;\ j=1,\ldots,r;\ k=1,\ldots,n_{ij}$$

where it is possible that $n_{ij} = 0$ or 1 and β_j are fixed effects. Such
)ls are conveniently put into the form of mixed linear model,

(1.6) $$Y = X\beta + U_1 \xi_1 + \ldots + U_p \xi_p + \varepsilon$$

X, $U_1, \ldots, U_p$ are known matrices, β is fixed unknown vector parameter and $\xi_1 \ldots \xi_p$, ε are unobservable random variables such

(1.7) $$E(\varepsilon)=0,\ E(\xi_i)=0,\ E(\xi_i \xi_j')=0,\ i \neq j$$

File name:YPChaubey on floppy disk
entered in rainbow

$$E(\varepsilon_i) = 0,\ E(\underline{\varepsilon\varepsilon}') = \sigma_0^2 I_n,\ E(\xi_i \xi_i') = \sigma_i^2 I_{n_i}$$

The unknown parameters $\sigma_0^2, \sigma_1^2, \ldots \sigma_p^2$ are called variance components. Such model have been used since the time of Fisher using mainly analysis of variance (ANOVA) tables, which was systematized by Henderson (1953). But it was only about 20 years later when Rao (1970, 1971a, 1971b, 1972) proposed a general method in a series of papers called MINQUE (Minimum Norm Quadratic Unbiased Estimator). This method also being subject to produce negative estimates like the ANOVA method has been modified and extended by Rao and Chaubey (1978), Chaubey (1980, 1982, 1983) and performances of various estimators have also been studies (see the review papers by Rao and Kleffe 1980, Rao, P.S.R.S 1977 and Chaubey 1984).

In what follows we describe here various methods for estimating the variance components. Some of these methods are available on SAS system, which are computed for given data. Other methods should be studied with special reference to such data.

METHODS ON

The different methods can be put into three cateogries. (1) ANOVA methods, (2) Maximum Likelihood method and Marginal or Restricted Maximum Likelihood method and (3) MINQUE method and its modifications. To give all the methods in detail will be very lengthy proposition. Instead, we describe here each method in short and provide references for its detailed account.

File name:YPChaubey on floppy disk
entered in rainbow

2.1 ANOVA Methods

These methods derive p+1 quadratic functions of y usually through mean squares from a corresponding analysis of variance table, $Y'A_iY$, i=1,2...p+1, whose expectations are linear functions of σ_i^2, i=0,...,p. To get estimates of σ_i^2 ,we solve

$$Y' A_i Y = E(Y' A_i Y) = a_{i0}\sigma_0^2 + \ldots + a_{ip}\sigma_p^2,\ i=1,\ldots p+1$$

For a detailed account of these methods, see Searle (1971)

2.2 Maximum Likelihood and Marginal Maximum Likelihood Methods

The approach of maximum likelihood was initiated by Hartley and Rao (1967) for such models, where $\xi_1\ \xi_2 \ldots \xi_p$ and ϵ are assumed normally distributed. Patterson and Thompson (1975) eliminated the fixed effects by invariance principle and considered the likelihood function generated by the least squares residuals. For a review and algorithms see Harville (1977).

2.3 MINQUE Method and its Modifications

In the MINQUE theory a natural estimator in terms of $\xi_1, \xi_2 \ldots \xi_p, \epsilon$ for a variance component is defined. But since $\xi_1 \xi_2 \ldots \xi_p\ \epsilon$ are unobservables, the working estimator (a quadratic form in Y) Y A Y is obtained so that it is "close" to the natural estimator. For a detailed account see Rao and Kleffe (1980). For three modifications, see Rao and Chaubey (1978). The methods are compared for one way model using simulation by Chaubey (1983). For the type of models we

may be concerned here, two modifications of MINQUE towards getting a non-negative estimate are provided by Chaubey (1980, 1983). One of them does not restrict the estimator to be unbiased and the other gives an estimator "closest" to MINQUE. Non-negative MINQUE was proposed by Pukelsheim (1977) which may not always exist.

This section provides variance component estimates using SAS. This system gives four estimators, Henderson's Type I, a special case of MINQUE, MIVQUE(0), ML, and REML. The drawback of this program is that it does not provide the variance-covariance matrix of estimators for type I and MIVQUE(0) which is of interest to investigators. It should also include simple modifications for comparision purposes. It should also be remarked that MINQUE methodology can incorporate the prior knowledge about σ_i^2/ σ_0^2 ; MIVQUE(0) method assumes these to be zero. One may start with these as initial prior guess and iterate until stable solutions are obtained. Kleffe (1985) has developed a general algorithm and corresponding software for computing MINQUE and some of its modifications.

For illustration we consider the data from two pearl millet trials conducted during 1983. 1 Trial 1 conducted at Bhawanisagar on 22 genotypes in 3 replicates and 2. Trial 2 conducted at ICRISAT Center at high fertility on 24 genotypes in 3 replicates. Since 21 genotypes are common to both the trials it has been decided first to

have a look on the basis of common genotypes. Table I summarises the data on Trial 1 and Table II summarises the data on Trial 2. Variables Y_1, Y_2, Y_3 represent days to bloom, plant height (cms) and Yield (kg/ha) respectively. GENO and Rep represent genotypes and blocks.

File name:YPChaubey on floppy disk
entered in rainbow

Table I

OBS	REP	GENO	Y_1	Y_2	Y_3
1	1	12	49	170	2164.0
2	1	17	44	180	4632.0
3	1	8	47	207	4477.0
4	1	3	46	183	3971.2
5	1	21	46	198	4774.0
6	1	4	45	202	4264.0
7	1	19	39	158	3322.6
8	1	16	47	202	4544.8
9	1	14	41	180	3830.4
10	1	7	45	190	3728.8
11	1	15	49	213	3517.8
12	1	10	48	245	4021.8
13	1	9	46	225	3480.0
14	1	18	44	170	2982.2
15	1	6	45	207	4020.0
16	1	11	50	213	4151.0
17	1	5	47	198	3891.2
18	1	1	44	182	4299.4
19	1	20	45	195	3832.0
20	1	13	45	195	3529.8
21	1	2	45	145	1850.0
22	2	4	47	180	2115.0

File name:YPChaubey on floppy disk
entered in rainbow

23	2	17	47	195	4297.8
24	2	13	48	220	3837.6
25	2	7	45	202	3689.3
26	2	19	41	150	3550.6
27	2	12	49	182	2496.0
28	2	9	4[illegible]	207	3328.8
29	2	8	48	208	4110.0
30	2	10	47	227	3608.7
31	2	16	49	227	3231.1
32	2	6	47	205	3250.0
33	2	5	48	203	2912.0
34	2	11	50	227	3952.4
35	2	15	51	217	3465.0
36	2	3	45	203	4487.2
37	2	21	47	182	1976.0
38	2	14	42	178	3655.6
39	2	18	44	182	3024.0
40	2	2	44	158	3837.6
41	2	1	44	145	3544.8
42	2	20	47	172	2832.0
43	3	6	47	188	2971.6
44	3	[illegible]6	48	187	2979.2
45	3	1	43	183	3863.7
46	3	10	48	206	3088.8
47	3	8	51	225	3556.4

File name:YPChaubey on floppy disk
entered in rainbow

48	3	14	40	162	2394.0
49	3	18	45	167	2100.0
50	3	11	51	225	3556.4
51	3	4	47	215	3252.8
52	3	12	48	192	2733.6
53	3	21	46	178	3427.6
54	3	2	44	163	1963.5
55	3	17	45	168	3085.6
56	3	7	46	162	1958.4
57	3	20	45	162	2114.8
58	3	9	45	197	3218.6
59	3	13	47	197	3515.0
60	3	19	40	165	2910.0
61	3	3	46	205	3474.8
62	3	15	48	223	4192.0
63	3	5	47	197	3888.0

File name:YPChaubey on floppy disk
entered in rainbow

Table II.

OBS	REP	GENO	Y_1	Y_2	Y_3
1	1	20	50	255	3191.15
2	1	16	51	261	2913.75
3	1	9	52	260	3105.18
4	1	11	56	278	3893.42
5	1	6	51	222	2541.01
6	1	17	56	241	2223.61
7	1	7	52	238	2871.01
8	1	15	54	250	2432.15
9	1	14	48	230	2432.15
10	1	12	52	240	2714.41
11	1	13	50	245	2638.61
12	1	1	51	230	3684.92
13	1	21	51	250	2976.95
14	1	3	49	255	2705.41
15	1	4	50	250	2861.61
16	1	5	50	270	2738.71
17	1	18	51	230	2484.35
18	1	8	52	260	2837.35
19	1	10	53	285	2933.35
20	1	[illegible]	49	252	2865.08
21	1	19	45	195	2833.48
22	2	3	49	242	2811.21
23	2	20	50	246	2418.10

File name:YPChaubey on floppy disk
entered in rainbow

24	2	12	52	270	2923.01
25	2	17	56	251	3009.02
26	2	13	50	242	2214.34
27	2	10	52	255	2938.81
28	2	16	51	240	2178.68
29	2	14	48	222	3679.22
30	2	18	52	231	1903.51
31	2	2	51	242	3140.92
32	2	5	51	267	3351.08
33	2	7	49	250	2888.18
34	2	11	57	275	2961.95
35	2	21	49	260	2993.55
36	2	19	44	188	2412.85
37	2	4	49	240	2725.35
38	2	1	50	235	3391.95
39	2	8	51	240	2725.35
40	2	15	53	262	3916.75
41	2	9	50	261	3185.32
42	2	6	50	260	2963.35
43	3	1	50	240	3434.65
44	3	12	52	279	2815.01
45	3	21	50	275	3368.68
46	3	10	52	285	3047.35
47	3	17	56	260	3444.75
48	3	18	52	220	2533.35

File name:YPChaubey on floppy disk
entered in rainbow

49	3	11	57	281	2735.01
50	3	4	49	256	2857.95
51	3	14	49	220	3608.20
52	3	15	55	255	3532.08
53	3	3	50	275	3280.02
54	3	13	49	249	2288.18
55	3	20	50	252	2629.81
56	3	16	51	265	2704.01
57	3	7	49	240	3061.35
58	3	9	50	260	2948.41
59	3	6	51	240	2580.68
60	3	5	51	252	2494.81
61	3	19	45	196	2982.91
62	3	2	51	245	3429.08
63	3	8	52	260	3019.22

File name:YPChaubey on floppy disk
entered in rainbow

The model of the following type is assumed ;

$$Y_{ij} = \mu + a_i + \beta_j + e_{ij};\ j = 1,2,3,\ i = 1,2,3\ldots,\ 21$$

where y_{ij} represent observation on i-th genotype and j-th block. Block effects β_j are assumed fixed. The variance component due to random effects a_i represent genotype variance component. In such a situation as this (of a balanced experiment) MIVQUE(0) and MINQUE are same and equal to ANOVA type estimators.

SAS program did not produce any REML estimates due to some inherent problem in the software. Available estimates are summarised below:

File name:YPChaubey on floppy disk
entered in rainbow

Table III. Variance component Estimates for Trial I

Character	Method	Estimate of variance component due to: Cultivar $\hat{\sigma}^2_g$	Error $\hat{\sigma}^2_e$	% ratio $\frac{\hat{\sigma}^2_g}{(\hat{\sigma}^2_g+\hat{\sigma}^2_e)} \times 10$
Y_1	ML	5.7498	.8428	87.22
	MINQUE	6.0373	.8849	87.22
Y_2	ML	333.4618	176.7226	65.36
	MINQUE	350.1349	185.5587	65.36
Y_3	ML	84183.1467	363002.2099	18.83
	MINQUE	88392.3041	381152.3204	18.83

File name:YPChaubey on floppy disk
entered in rainbow

Table IV. Variance Component Estimates for Trial II

Character	Method	Estimate of variance component due to Cultivar $\hat{\sigma}^2_g$	Estimate of variance component due to Error $\hat{\sigma}^2_e$	% ratio $\frac{\hat{\sigma}^2_g}{(\hat{\sigma}^2_g+\hat{\sigma}^2_e)} \times 100$
Y_1	ML	6.1376	.3144	95.13
	MINQUE	6.4444	.3302	95.13
Y_2	ML	298.0567	99.8027	74.92
	MINQUE	392.9595	104.7929	74.92
Y_3	ML	34845.2199	136426.4925	20.34
	MINQUE	36587.4809	143247.8172	20.34

From these tables one thing is brought out that the estimate of heritability is same whether ML is used or MINQUE is used for balanced data. But for unbalanced data this may not be true.

The difference in estimates of genotypic variances from different locations may be attributed to different locations, hence the following model on the combined data may be used;

$y_{ijk} = u + a_i + \beta_k + r_k + \delta_{ik} + \epsilon_{ij}$,

where k represents location. We consider two cases (A) location effects fixed, (B) location effects random. The results are summarized in Tables V and VI respectively.

File name:YPChaubey on floppy disk
entered in rainbow

Table V. Estimates of variance components on combined trials (fixed location case).

Character	Method	Estimate of variance component due to: Cultivar $\hat{\sigma}^2_g$	Loc x cultivar $\hat{\sigma}^2_{gL}$	Error $\hat{\sigma}^2_e$	$\frac{\hat{\sigma}^2_g}{(\hat{\sigma}^2_g+\hat{\sigma}^2_{gL}+\hat{\sigma}^2_e)}$ x 10(
Y_1	ML	4.2459	1.6978	.5786	65.10
	MINQUE	4.4582	1.7826	.6075	65.10
Y_2	ML	258.8859	56.8733	138.2627	57.02
	MINQUE	271.8302	59.7171	145.1758	57.02
Y_3	ML	36524.6101	24989.5733	249714.3512	11.16
	MINQUE	36250.8405	26239.0519	262200.0688	11.16

File name:YPChaubey on floppy disk
entered in rainbow

Table VI. Estimates of variance component on combined trials (Random location case).

Character	Method	Estimate of variance component due to Cultivar $\hat{\sigma}^2_g$	Loc $\hat{\sigma}^2_L$	Loc x Cult $\hat{\sigma}^2_{gL}$	Error $\hat{\sigma}^2_e$	$\frac{\hat{\sigma}^2_g}{\hat{\sigma}^2_g+\hat{\sigma}^2_L+\hat{\sigma}^2_{gL}+\hat{\sigma}^2_e}$ x100
Y_1	ML	4.4362	6.0784	1.7564	.6905	34.23
	MINQUE	4.4582	11.8561	1.7494	.7073	23.75
Y_2	ML	271.2192	817.8122	59.0996	147.1254	20.94
	MINQUE	271.8302	1615.7349	57.8712	150.7135	12.97
Y_3	ML	35408.8117	57938.1598	15647.9725	295367.9054	8.76
	MINQUE	36250.8405	118129.6437	12781.5573	302572.5528	7.72

Just to contrast balanced data and unbalanced data we consider 21st treatment missing in the 2nd location. ANOVA method does not give same estimates as MINQUE. The estimates for this data are summarised in Table VII and VIII respectively for the two cases described. The different methods may provide very different estimates; see estimates for Y_2 and Y_3 in Tables VI and VIII.

File name:YPChaubey on floppy disk
entered in rainbow

Table VII. Estimates of variance components on combined trials with missing data (fixed location case).

Character	Method	Estimate of variance component due to: Cultivar $\hat{\sigma}^2_g$	Loc x cultivar $\hat{\sigma}^2_{gL}$	Error $\hat{\sigma}^2_e$	$\frac{\hat{\sigma}^2_g}{(\hat{\sigma}^2_g+\hat{\sigma}^2_{gL}+\hat{\sigma}^2_e)}$ x 100
Y_1	ML	4.2629	1.7389	.5768	64.80
	MINQUE	4.7082	1.6755	.6064	67.36
	ANOVA	4.5374	1.8463	.6064	64.91
Y_2	ML	263.9908	51.9338	139.1033	58.02
	MINQUE	290.3462	46.2046	146.2365	60.14
	ANOVA	281.1910	55.3598	146.2365	58.24
Y_3	ML	34686.2649	25233.7543	255203.1547	11.01
	MINQUE	38334.3790	24871.9054	268289.7815	11.56
	ANOVA	34324.4834	28881.8011	268289.7815	10.35

File name:YPChaubey on floppy disk
entered in rainbow

Table VIII. Estimates of variance component on combined trials with missing data (Random location case).

Character	Method	Estimate of variance component due to cultivar $\hat{\sigma}^2_g$	Loc $\hat{\sigma}^2_L$	Loc x Cult $\hat{\sigma}^2_{gL}$	Error $\hat{\sigma}^2_e$	$\frac{\hat{\sigma}^2_g}{\hat{\sigma}^2_g+\hat{\sigma}^2_L+\hat{\sigma}^2_{gL}+\hat{\sigma}^2_e}$
Y_1	ML	4.4531	6.2117	1.8084	.6750	41.20
	MINQUE	4.6621	12.0773	1.6927	.6929	24.38
	ANOVA	4.5374	12.0713	1.8174	.6929	23.73
Y_2	ML	276.6470	793.8929	54.4806	146.8883	21.75
	MINQUE	299.0202	1580.1886	36.0892	150.5608	14.47
	ANOVA	281.1911	1579.3396	53.9183	150.5608	13.62
Y_3	ML	35562.1095	60244.2572	16645.7826	289491.0092	8.63
	MINQUE	38070.2939	1.791.6674	12239.5242	306979.1817	7.93
	ANOVA	34324.4834	122 93.2956	15985.3344	30697.1817	7.15

File name:YPChaubey on floppy disk
entered in rainbow

4. Suggestions for further investigations

Comparison of various estimators with reference to estimating non-linear functions of variance component pertinent to plant breeding data are of interest. This can be done using simulation. It is of further interest to bringout software for other estimators along with the estimates of their variances and covariances. This development will be useful to plant breeders. The general computer program developed by Kleffe and his associates at Akademie of Sciences, GDR is recommended as a very useful tool towards this methodology.

File name:YPChaubey on floppy disk
entered in rainbow

REFERENCES

Chaubey,Y.P. (1980). Minimum Norm Quadratic Estimators of variance components. Metrika 27, 255-262.

Chaubey,Y.P. (1982). Minimum Norm Invariant Quadratic Estimation of a covariance Matrix in a linear model. Biometrical Journal 24, 457-461.

Chaubey,Y.P. (1983). A Non negative Estimator of variance component closest to MINQUE. Sankhya, A 45, 201-211.

Chaubey,Y.P. (1984). On the comparison of some Non negative Estimators of variance components for Two Models Commun. Statist - Simul. Comput. B13(5), 619-633.

Comstock,R.E. and Robinson, H.F. (1948). The components of Genetic variance in populations of Biparental Progenies and their use in estimating Average degree of dominance Biometrics 4, 254-266.

Hartley, H.O. and Rao,J.N.K. (1967). Maximum Likelihood Estimation for Mixed Analysis of Variance Model Biometrika 54, 93-108

File name:YPChaubey on floppy disk
entered in rainbow

Harville, D.A. (1977). Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems J. Amer. Statist. Assoc., 72, 320-340.

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

Kempthorne,O. and Tandon,O.B. (1953). The estimation of Heritability by Regression of Offspring on Parent. Biometrics, 9, 90-100.

Kleffe, J (1985). Personal Communication.

Patterson, H.D. and Thompson,R. (1975). Maximum Likelihood Estimation of Components of Variance. In : Proc. of 8th International biometric Conference, pp. 197-207.

Pukelsheim, F. (1977). Linear Models and Convex Programs: Unbiased Non-negative Estimation in Variance components Models. Tech. Rep. 104, Stanford University.

Rao, C.R. (1970). Estimation of Heteroscedastic variances in Linear Models J. Amer. Stat. Assoc. 65, 161-172.

File name:YPChaubey on floppy disk
entered in rainbow

Rao, C.R. (1970a). Estimation of Variance and Covariance Components. J. Multivar. Anal. 1, 257-275.

Rao, C.R. (1971b). Minimum Variance Quadratic Unbiased Estimation of Variance Components. J. Multivar. Anal., 1, 445-456.

Rao, C.R. (1972). Estimation of Variance and Covariance Components in Linear Models: MINQUE Theory J. Amer Stat. Assoc., 67, 112-115.

Rao,C.R., and Kleffe, J. (1980). Estimation of variance components. In : Handbook of Statistics Vol. 1, Ed. P.R.Krishnaiah: North Holland Publishing Co., pp. 1-40.

Rao,P.S.R.S. (1977). theory of the MINQUE - A Review. Sankhya B, 201-210.

Rao,P.S.R.S. and Chaubey,Y.P. (1978). Three notifications of the Principle of MINQUE. Commun. Statist. A7, 767-778.