ESTIMATION OF EFFECTS AND THEIR STANDARD ERRORS IN A CROSSED MATING DESIGN

By

W. N. WICKREMASINGHE

SUMMARY

The estimators of standard errors of the estimated male, female and interaction effects of a Crossed Mating Design with m – males, f – females and r – replicates per each cross are, $\sqrt{(1/fr)(1-1/m)\sigma^2}$, $\sqrt{(1/mr)(1-1/f)\sigma^2}$ *and* $\sqrt{(1/r)(1-1/m-1/f+1/mf)\sigma^2}$ *respectively. A practical method using a FORTRAN program, to estimate the above effects along with their corresponding standard errors, is given. A numerical example is also given to illustrate the method. The program used to obtain the results in the numerical example is available with the RRISL computer.*

INTRODUCTION

A common problem in research is the investigation of effects of each of a number of treatments, relative to one or more response variables. A second important problem is to obtain appropriate standard errors for the estimates of these effects.

There is always the statistical theory associated with any kind of statistical problem, *e.g.* finding the best estimators of parameters of interest. However, in most practical cases experimentors are concerned with the use of such estimators in their research, and not with the underlying theory used to obtain the estimators. It is desirable that such estimators be of a simple and convenient form for use, but also that they have some generality such that they can be used in a variety of data situations.

A particular problem in the field of genetics is to find the male, female and their interaction effects along with the corresponding standard errors in a Crossed Mating Design (Comstock and Robinson, 1948) with fixed model (Model I : Eisenhart, 1947). This mating design is also known as Design II. In this design a number of males are each mated with each female in a set of females. This is possible in an organism in which a female parent can produce progenies by more than one male (mostly with plants).

The objective of this study is to investigate a method to compute estimates of standard errors of estimated male, female and interaction effects from a Design II experiment under fixed model assuming each cross gives equal number of offspring. This method utilizes a simple computer program written in FORTRAN. The statistical package called SAS could also be used to compute these estimates after a simple modification of procedures given in SAS. Because of the non-availability of SAS in Sri Lanka, that method will not be discussed in this paper.

Manual procedure for estimating these effects is tedious, time consuming and not very accurate. Especially, when there are a large number of males and females in the design, it is necessary to stick to a statistical package or a freshly written program. Since packages are not available with the RRISL computer at this time I have given more emphasis to write a new program using FORTRAN. This program can be changed occasionally to suit different data situations.

METHODS AND RESULTS

Crossed mating design (Design II)

The linear model for the above design is :

$$P_{ijk} = \mu + M_i + F_j + (MF)_{ij} + G_{ijk} \cdots\cdots\cdots\cdots (1)$$

Where P_{ijk} is the response for k^{th} replicate in ij^{th} combination. μ is the grand mean. M_i is the effect of the i^{th} male ; $i = 1, 2 \ldots\ldots m$. F_j is the effect of the j^{th} female ; $j = 1, 2, \ldots\ldots f$. $(MF)_{ij}$ is the effect of the ij^{th} MF interaction. G_{ijk} is the experimental error associated with the ij^{th} cross in the k^{th} replication. The general assumption on the error is that they are distributed normally and independently with mean zero and constant variance σ^2, *i.e.*, $E(\underline{G}\ \underline{G}') = \sigma^2 I_{mfr}$ where $\underline{G}$ is the vector of G_{ijk}'s. All the other effects are assumed to be fixed under this model.

Using normal equations for the above model and solving them (Kempthorne, 1979) after adjoining $(m + f + 1)$ linearly independent conditions on the solutions (Kempthorne, 1983) we get the following estimators for the effects :

$$\hat{M}_i = \bar{P}_{i..} - \bar{P}_{...},\ i = 1, 2 \ldots\ldots m$$

$$\hat{F}_j = \bar{P}_{.j.} - \bar{P}_{...},\ j = 1, 2 \ldots. f$$

$\widehat{(MF)}_{ij} = \bar{P}_{ij.} - \bar{P}_{i..} - \bar{P}_{.j.} + \bar{P}_{...}$, where the symbol " $\wedge$ " denotes the estimator of the corresponding effect. Also, $\bar{P}_{...} = \Sigma_{ijk} P_{ijk}/mfr$, $\bar{P}_{i..} = \Sigma_{jk} P_{ijk}/fr$, and so on.

Using the direct product notation (Graybill,1969) and the multi-part model version of the general linear model (Kempthorne, 1983), one can obtain the following estimators for the standard errors of the estimated male, female and interaction effects of a Design II experiment (Wickremasinghe, 1983).

These are $\sqrt{(1/fr)(1-1/m)\sigma^2}$, $\sqrt{(1/mr)(1-1/f)\sigma^2}$ and $\sqrt{(1/r)\ (1-1/m-1/f + 1/mf)\sigma^2}$ respectively for male, female and interaction effects with m-levels for males, f-levels for females and r - replicates per each cross.

Replacing σ^2 by $\hat{\sigma}^2$, where $\hat{\sigma}^2$ is the error mean square from the usual analysis of variance, we can get the relevant estimates of the effects under consideration.

Estimation of σ^2

The estimate of the error variance (σ^2) is denoted by ($\hat{\sigma}^2$.) Therefore, $\hat{\sigma}^2$ = (S. S. of error)/mf (r–1), if we assume a completely randomized design with r–replicates of each cross.

Sum of Squares (S.S.) of error $= \underline{P}' (I - P_x) \underline{P}$
where $P_x = (1/r)\, J_r \otimes I_{mf}$ (Wickremasinghe, 1983)
Using this, we have $\hat{\sigma}^2 = \underline{P}' (I-P_x)\underline{P}/mf(r-1)$
This is also equal to $(\Sigma_{ijk} P^2_{ijk} - (1/r)\Sigma_{ij} P^2_{ij.})/mf(r-1)$

("$\otimes$" is the notation for "direct product" of two matrices and $\underline{P}$ is the vector of responses given in the model (1) defined above).

Computational Method

The following is a general FORTRAN program which can be used to obtain required results in our problem. The particular program used in the numerical example is given in the appendix, and this program is available with the RRISL Computer :

```
     INTEGER  F, R

     DIMENSION  (will depend on the maximum number of levels for males and females
         and also replicates)

     COMMON

     ABSQ  = 0·0, SUM = 0·0, TSQ = 0·0

     ACCEPT R, M, F (Give specific levels for replicates, males and females)

     ACCEPT P (I, J, K) (feed data here)
     DO  15  I = 1, M
     DO  15  J = 1, F
     AB (I,  J) = 0·0
     DO 20 K  =  1, R

     AB  (I, J) = AB (I, J.) + P (I, J, K)

20 CONTINUE

     ABMEAN (I, J) = AB (I, J)/R
     ABSQ  = ABSQ + AB.(I. J) **2
```

```
15 CONTINUE
   DO 21 I = 1, M
   A (I) = 0·0
   DO 22 J = 1, F
   DO 22 K = 1 R
   A (I) = A (I) + P (I,J,K)
   SUM = SUM + P (I, J, K)
   TSQ = TSQ + P (I, J, K) **2
22 CONTINUE
   AMEAN (I) = A (I)/F *R)
21 CONTINUE
   DO 31 J = 1,F
   B (J) = 0·0
   DO 30 I = 1, M
   DO 30 K = 1, R
   B (J) = B (J) + P (I, J, K)
30 CONTINUE
   BMEAN (J) = B (J)/(M *R)
31 CONTINUE
   PBAR = SUM/(M*F*R)
   SIGMASQ = (TSQ — (ABSQ/R) ) / (M*F*(R — 1) )
   ASE = SQRT ( ( ( 1·0 — 1·0/M)*SIGMASQ)/(F*R) )
   BSE = SQRT ( ( (1·0 — 1·0/F)*SIGMASQ)/(M*R) )
   ABSE = SQRT ( ( (1·0—1·0/M—1·0/F + 1·0/(M*F) )*SIGMASQ)/R)
   WRITE (12) (Use this statement with relevant FORMAT to print results)
   STOP
   END
```

A numerical example

To illustrate the procedure given earlier to compute effects and standard errors of a Crossed Mating Design, a numerical example is given below. These data were provided by a graduate student in Horticulture of the University of Wisconsin. The data come from a real experiment on cucumber. The relevant entries are given in Table 1.

Table 1. *Entries of a crossed mating design corresponding to an experiment on cucumber*

		Male lines from 1928 population				
		K241	K301	K350	K311	K320
Female	K87	D43	D44	D45	D46	D47
Lines from	K64	D48	D49	D50	D51	D52
1925	K85	D53	D54	D55	D56	D57
Population	K43	D58	D59	D60	D61	D62
	K147	D63	D64	D65	D66	D67

In Table 1 there are twenty five crosses from D43 to D67 in order. There are 4 observations (replicates) for each cross taken on a measurement called "length". Let us call this variable "P". (This is the same P which appears in our model (1) defined earlier.) Raw data of this experiment are given in the appendix.

A print out of the results can be obtained as wished, by the FORMAT statement.

Table 2. *Estimated effects and standard errors of Males, Females and their interaction*

		Male lines					
		K241	K301	K350	K311	K320	
Female	K87	0·1295 (0·2673)	−0·0905 (0·2673)	0·2920 (0·2673)	−0·2880 (0·2673)	−0·0430 (0·2673)	0·7705 (0·1337)
lines	K64	−0·0705 (0·2673)	−0·0405 (0·2673)	−0·0330 (0·2673)	−0·4005 (0·2673)	0·5445 (0·2673)	−0·0420 (0·1337)
	K85	0·1570 (0·2673)	0·0120 (0·2673)	−0·1430 (0·2673)	−0·0105 (0·2673)	−0·0155 (0·2673)	−0·1995 (0·1337)
	K43	0·2620 (0·2673)	−0·3955 (0·2673)	0·1870 (0·2673)	0·4695 (0·2673)	−0·5230 (0·2673)	−1·062 (0·1337)
	K147	−0·4780 (0·2673)	0·5145 (0·2673)	−0·3030 (0·2673)	0·2295 (0·2673)	0·0370 (0·2673)	0·4530 (0·1337)
		1·0455 (0·1337)	−0·3970 (0·1337)	−0·7045 (0·1337)	−0·4245 (0·1337)	0·4805 (0·1337)	

(Standard errors are given within brackets)

In the above numerical example, the number of males is 5 (*i.e.* m = 5), that of females is 5 (*i.e.* f = 5) and the number of replicates per each cross is 4 (*i.e.* r = 4). The design is a completely randomized design. The actual computer program used to obtain the results in Table 2 is given in the appendix.

DISCUSSION

When statistical packages are readily available, researchers prefer to get their whole set of results of an experiment at one run of a computer program. Even when packages are not available, the availability of a simple program to do the task of a package is preferred by applied workers.

In this paper, I have presented such a method using a FORTRAN program, to obtain relevant estimates coming in a Design II experiment.

The same method as given in this paper can be used to obtain estimates with the same design but with any different crop. In short, this computational method remains the same with Design II experiments on rubber. It is only a matter of slightly adjusting the DIMENSION statement to suit different data situations, when required.

REFERENCE

ANON (1979). *SAS User's Guide*, SAS Institute, Inc. Cary, NC.

ANON (1981). *MP/FORTRAN IV Programmers' Reference*, Data General Corporation, Westboro, Mass.

COMSTOCK, R. E. AND ROBINSON, H. F. (1948). "The Components of Genetic Variance in Populations of Biparental Progenies and Their Use in Estimating the Average Degree of Dominance", *Biometrics*, **4** : 254 - 266.

EISENHART, C. (1947). "The Assumptions Underlying the Analysis of Variance", *Biometrics*, **3**, 1 - 21.

GRAYBILL, F. A. (1969). *Introduction to Matrices with Application in Statistics*, Wadsworth Publishing Company, Inc., Belmont, CA.

KEMPTHORNE, O. (1979). *Design and Analysis of Experiments*, Robert E. Krieger Publishing Company, Huntington, NY.

KEMPTHORNE, O. (1983). *Introduction to Linear Models*, a draft of material intended to be put into book form, Iowa State University, Ames, IA.

WICKREMASINGHE, N. (1983). *A Study on the Effects and Standard Errors in a Crossed Mating Design*, Creative Component for the M.Sc. Degree, Iowa State University, Ames, IA.

APPENDIX

A. Program used in the numerical example

```
      INTEGER F, R
      DIMENSION P (10, 10, 5), A (10), B (10), AB (10, 10), AMEAN (10), BMEAN (10),
  —   ABMEAN (10, 10), ACAP (10), BCAP (10), ABCAP (10, 10)
      COMMON/BK1/A, B, AMEAN, BMEAN, ACAP, BCAP/BK2/AB, ABMEAN,
  —   ABCAP
      ABSQ = 0·0
      SUM  = 0·0
      TSQ  = 0·0
      ACCEPT "TYPE NO. OF REPLICATES :", R
      ACCEPT "TYPE NO. OF LEVELS FOR MALES :", M
      ACCEPT "TYPE NO. OF LEVELS FOR FEMALES :", F

COMMENT : WE HAVE ASSUMED THAT THE MAX. FOR M & F IS 10
  —   AND THAT OF R IS 5. ON EXECUTION, ONE HAS TO TYPE THE
  —   REAL VALUES FOR R, M & F.
      DO 10 I = 1, M
      DO 10 J = 1, F
      TYPE "ENTER DATA FOR THE COMBINATION", I, J
      ACCEPT "DATA :", (P (I, J, K), K = 1, R)

COMMENT : WHEN EXECUTING THE PROGRAM, COMPUTER WILL FIRST
  —   ASK FOR
  —   DATA FOR THE (1,1) COMBINATION. WE ANSWER BY TYPING R VALUES
  —   EACH SEPARATED BY A COMMA OR A NEW LINE. THEN (1, 2) COMBI-
  —   NATION AND SO ON.

10    CONTINUE
      DO 15 I = 1, M
      DO 15 J = 1, F
      AB (I, J) = 0·0
      DO 20 K = 1, R
      AB (I, J) = AB (I, J) + P (I, J, K)

20    CONTINUE
      ABMEAN (I, J) = AB (I, J)/R
      ABSQ = ABSQ + AB (I, J) **2

15    CONTINUE
      DO 21 I = 1, M
      A (I) = 0·0
      DO 22 J  = 1, F
      DO 22 K = 1, R
      A (I) = A (I) + P (I,J,K.)
      SUM = SUM + P (I, J, K)
      TSQ  = TSQ + P (I, J, K) **2
```

```
 22 CONTINUE

    AMEAN (I) = A (I)/(F *R)

 21 CONTINUE

    DO 31 J = 1,F
    B (J) = 0·0
    DO 30 I  = 1, M
    DO 30 K = 1, R
    B (J) = B (J) + P (I, J, K)

 30 CONTINUE

    BMEAN (J) = B (J)/(M*R)

 31 CONTINUE

    PBAR = SUM/(M*F*R)
    SIGMASQ = (TSQ — (ABSQ/R) ) / (M*F*(R — 1) )
    ASE = SQRT ( ( ( 1·0 — 1·0/M)*SIGMASQ)/(F*R) )
    BSE = SQRT ( ( (1·0 — 1·0/F)*SIGMASQ)/(M*R) )
    ABSE = SQRT ( ( (1·0—1·0/M—1·0/F + 1·0/(M*F) )*SIGMASQ)/R)

    WRITE (12, 99)

 99 FORMAT (3X, "ESTIMATES OF MALE EFFECTS AND STD. ERRORS"/)

    DO 25 I = 1, M
    ACAP (I) = AMEAN (I) — PBAR
    WRITE (12,100) I, ACAP (I)

100 FORMAT (4X, I2, 4X, F9·4)

 25 CONTINUE

    WRITE (12, 110) ASE

110 FORMAT (3X, "STD. ERROR", 4X, F9·4)

    WRITE (12, 98)

 98 FORMAT (3X, "ESTIMATES OF FEMALE EFFECTS AND STD. ERRORS"/)

    DO 35 J = 1, F
    BCAP (J) = BMEAN (J) — PBAR
    WRITE (12, 120) J, BCAP (J)
```

```
120 FORMAT (4X, 12, 4X, F9·4)
 35 CONTINUE

    WRITE (12,130) BSE

130 FORMAT (3X, "STD. ERROR", 4X, F9·4)

    WRITE (12, 97)

 97 FORMAT (3X, "ESTIMATES OF INTERACTION EFFECTS AND STD.
                                                    ERRORS"/)

    DO 45 I = 1,M
    DO 45 J = 1, F

    ABCAP (I, J) = ABMEAN (I, J) — AMEAN (I) — BMEAN (J) + PBAR
    WRITE (12, 150) I, J, ABCAP (I, J)

150 FORMAT (1X, I2, 3X, I2, 4X, F9·4)

 45 CONTINUE
    WRITE (12, 160) ABSE

160 FORMAT (3X, "STD. ERROR", 4X, F9·4)

    STOP
    END.
```

B. Raw data of the numerical example

Table 3. "*Lengths*" *in cm. of a Design II experiment on cucumber*

		Male lines									
		K 241		K 301		K 350		K 311		K 320	
Female lines	K 87	13·05	15·05	12·10	13·40	13·30	11·55	11·05	12·35	12·35	13·10
		12·25	14·15	10·95	11·40	12·85	10·45	11·30	12·25	13·50	12·60
	K 64	13·40	11·90	11·90	10·40	11·25	11·60	11·35	10·70	12·60	12·95
		11·65	13·50	10·75	11·75	10·15	10·60	10·35	10·85	12·35	12·75
	K 85	12·60	13·05	10·80	11·30	10·75	10·30	11·45	11·20	12·20	11·65
		12·75	12·65	10·75	11·85	11·30	10·50	11·25	10·60	12·30	11·95
	K 43	11·55	13·25	9·95	10·10	10·85	9·60	11·30	10·50	11·80	9·80
		11·45	11·45	10·15	9·10	10·05	9·90	10·50	10·80	10·05	10·65
	K147	12·65	12·95	12·05	12·55	11·10	11·40	12·80	11·45	13·20	12·65
		12·15	13·05	12·30	12·10	11·35	10·65	11·80	11·70	12·70	12·05