Arunachalam, V.
Computer Programmes for Some Problems in Biometrical Genetics-II. Use of Canonical Variates in Deriving Group constellations.
Indian Journal of Genetics and Plant Breeding, 27 (1967): 70-79.

COMPUTER PROGRAMMES FOR SOME PROBLEMS IN BIOMETRICAL GENETICS—II. USE OF CANONICAL VARIATES IN DERIVING GROUP CONSTELLATIONS

V. Arunachalam

Division of Genetics, Indian Agricultural Research Institute, Delhi-12

(Accepted 16-xii-1966)

Canonical analysis or the principal component analysis helps in confirming the group constellations arrived at by the use of D^2–statistic. It also facilitates in many cases the representation of the populations in a two-dimensional chart. The size of the coefficients in the first two canonical vectors indicates the relative importance of the characters in the major and secondary axes of differentiation.

The programme presented in this paper (Appendix) computes the first four canonical roots and the first four canonical vectors by the iteration method described by C. R. Rao (1952). The utility of the programme is well illustrated by an example taken from the data collected at this Institute and published in *Sankhya* (1965). Since canonical analysis is supplementary to D^2 analysis, it will be useful to use this programme as a supplement to the one on D^2 published in part I.

Description of the programme :

Language.—Fortran II for an IBM 1620 Model II computer
Input medium.—Punch cards

Computational method :

(i) Starting from the matrix of y-values, i.e., the mean values of the uncorrelated linear combinations for all the characters and for all the varieties, the between product sum matrix for the characters is computed (Matrix A).

(ii) The fourth power of matrix A is computed to quicken the iteration process.

(iii) Assuming a trial vector (1, 1, —, 1) the programme proceeds to make as many iterations as needed until the difference between each element in any two successive vectors does not exceed 0·0009. The vector is standardized and designated as canonical vector 1 and the first canonical root is extracted simultaneously.

(iv) The residual matrix after eliminating vector 1 is found out and the process repeated to get the next vector and so on until the first four vectors and roots are obtained.

(v) The mean values of the first two vectors designated as Z_1 and Z_2 by C. R. Rao are computed for each variety.

(vi) Sum of all canonical roots and the percentage contribution of the first four roots to the total variation are computed.

Input Data.—The programme requires the following as input data.

(i) Title of the experiment as described in Part I of the paper.
(ii) The number of characters and the number of varieties punched in one card as described in Part I of the paper.
(iii) The mean values of the uncorrelated linear combinations (obtained as punched output from the programme in Part I).

In case the programme in Part I is not excecuted, the mean values of Y's may be punched with six quantities in each card, each quantity occupying 10 columns with four decimal digits. The decimal point need not be punched. Punching instructions are similar to Input (ii) described in Part I of the paper. If more than six characters say 10 are used, the remaining 4 quantities for a variety may be punched in another card. In this case, two cards are required for y-values for each variety.

Output.—The programme renders the following printed output in a neat form giving spaces and underlined sub-titles wherever necessary, as in the example presented here :

(i) Title of the experiment as fed in the input.
(ii) The between product sum matrix of y-values designated as V-matrix.
(iii) V-matrix to the power of 4.
(iv) Canonical vector 1.
(v) Residual matrix eliminating canonical vector 1 and so on until canonical vector 4 is obtained.
(vi) Mean values of the first two canonical variates.
(vii) λ_1, λ_2, λ_3, λ_4 and their percentage contributions; sum of the other canonical roots and sum of all canonical roots.

This programme will work up to 60 varieties and 10 characters without alteration. It can be easily modified for a larger number of varieties and characters.

REFERENCES

Murty, B. R., Mathur, J.B.L. and Arunachalam, V. (1965). Self-incompatibility and genetic divergence in *Brassica campestris* var. brown *sarson*, *Sankhya B*, **27:** 271–78.

Rao, C.R. (1952). *Advanced Statistical Methods in Biometrical Research*, John Wiley and Sons, New York.

APPENDIX

```
  ‡‡ JOB 5
  ‡‡ FORX53
C CANONICAL ANALYSIS-PROGRAMMED BY V. ARUNACHALAM, I.A.R.I.
      DIMENSION T(10), E(10), A(10,10), B(10,10), F(10), CV(10,2), FF(10,4), Y
     1 (60,10), Q(12), P(60,2), PC(4)
 3000 READ 300, (Q(I), I=1, 12)
  300 FORMAT (12A4)
      PRINT 301, (Q(I), I=1, 12)
  301 FORMAT (1X, 12A4/1X, 24(2H-*), 1H-//)
C THE ABOVE CAUSES THE PRINTING OF THE TITLE OF THE EXPERIMENT
      PRINT 302
  302 FORMAT (1X, 41HEVALUATION OF CANONICAL ROOTS AND VECTORS/1X, 14 (3H-*
     1)//)
C N—NO. OF CHARACTERS, IV—NO. OF VARIETIES
C Y(IV,N)-UNCORRELATED MEANS MATRIX OF DIMENSION IV X N
C T(N)-COLUMN TOTALS OF ((Y))
C A(N, N)-BETWEEN PRODUCT SUM MATRIX OF Y-VALUES
      READ 190, N, IV
  190 FORMAT (2I3)
      NO =N-1
      IVV=IV-1
      FIV=IV
      DO 111I=1, IV
  111 READ 1110, (Y (I, J), J=1, N)
 1110 FORMAT (6F10·4)
C STATEMENT 1110 TO BE CHANGED FOR MORE THAN SIX CHARACTERS
      DO 151 J=1, N
      T(J)=0.
      A(J,J)=0.
      DO 150 I=1, IV
      T(J)=T(J)+Y(I,J)
  150 A(J,J)=A(J,J)+Y(I,J)* Y(I,J)
      CFAC=T(J)* T(J)/FIV
  151A(J,J)=A(J,J)-CFAC
      DO 153 I=1, NO
      II=I+1
      DO 153 J=II, N
      A(I,J)=0.
      DO 152 K=1, IV
  152 A(I,J)=A(I,J)+Y(K,I) *Y(K, J)
```

```
      CF=T(I) *T(J)/FIV
  153 A(I,J)=A(I,J)-CF
      DO 154 I=1, NO
      II=I+1
      DO 154 J=II, N
  154 A(J, I)=A(I, J)
      CARE=0.
      DO 1540 I=1, N
 1540 CARE=CARE+A(I,I)
      PRINT 155
  155 FORMAT ( 1X,47HV-MATRIX-BETWEEN PRODUCT SUM MATRIX OF Y-VALUES/ 1X,
     147 (1H-)/)
      DO 157 I=1, N
  157 PRINT 156, (A(I, J), J=1,N)
  156 FORMAT (10F12·4)
      DO 161 MGR=1,2
      DO 158  I=1, N
      DO 158  J=I, N
      B(I, J)=0.
      DO 158 K=1, N
  158 B(I,J)=B (I,J) +A(I,K)  *A(K,J)
      DO 159 I=1, N
      DO 159  J=I, N
  159 A(I,J)=B(I,J)
      DO 160 I=1, NO
      II=I+1
      DO 160  J=II, N
  160 A(J,I)=A(I,J)
  161 CONTINUE
      PRINT 162
  162 FORMAT (/1X,23HV-MATRIX TO THE POWER 4/1X, 23 (1H-)/)
      DO 164 I=1,N
  164 PRINT 163, (A(I,J), J=1, N)
  163 FORMAT (6F20·3)
      EQUIVALENCE (I,E)
      MP=1
  195 DO 165 I=1, N
  165 E(I)=1.
      L=1
  166 DO 167 J=1, N
```

```
      F(J)=0.
      DO 167 I=1, N
  167 F(J)=F(J)+E(I)* A(I,J)
      DO 168 I=1, N
  168 E(I)=F(I)
      ZZ=E(1)
      DO 170 I=2, N
      IF(ZZ-E(I)) 169,170,170
  169 ZZ=E(I)
  170 CONTINUE
  173 DO 174 I=1, N
  174 E(I)=E(I)/ZZ
      DO 176 I=1, N
  176 CV(I,L)=E(I)
      L=L+1
      IF (L-2) 166, 166, 178
  178 L=L-2
      LL=L+1
      DO 179 I=1, N
      VTEST=CV(I,L)-CV(I,LL)
      IF (VTEST-·0009) 179, 179, 180
  179 CONTINUE
      GO TO 181
  180 DO 1800 I=1, N
 1800 CV(I,L)=CV(I,LL)
      L=L+1
      GO TO 166
  181 G=0.
      DO 182 I=1, N
  182 G=G+CV(I,LL) *CV(I,LL)
      G=SQRTF(G)
      DO 183 I=1, N
  183 CV(I,LL)=CV(I,LL)/G
      ZP=ZZ** ·25
      Q(MP)=ZP
      PRINT 185, MP
  185 FORMAT (//1X,17HCANONICAL VECTOR(,I1, 1H)/1X,19(1H-)/)
      PRINT 1850, (CV(I,LL),I=1, N)
 1850 FORMAT (1X,10(F8·4,2X)//)
      DO 196 I=1, N
```

```
196 FF(I,MP)=CV(I,LL)
    IF (MP-4) 193, 194, 194
193 DO 187 I=1, N
    DO 187 J=I, N
187 A(I,J)=A(I,J)-ZZ *CV(I,LL) *CV(J,LL)
    DO 188 I=1, NO
    II=I+1
    DO 188 J=II, N
188 A(J,I)=A(I,J)
    PRINT 189, MP
189 FORMAT(1X,45HRESIDUAL MATRIX ELIMINATING CANONICAL VECTOR(, I2, 1H)/
   C1X,48(1H-)/)
    DO 192 I=1, N
192 PRINT 191,(A(I,J), J=1,N)
191 FORMAT (6F20·3)
    MP=MP+1
    GO TO 195
194 DO 198 I=1, IV
    P(I,1)=0.
    P(I,2)=0.
    DO 197 J=1, N
    P(I,1)=P(I,1)+FF(J,1)* Y(I,J)
197 P(I,2)=P(I,2)+FF(J,2)* Y(I,J)
198 CONTINUE
    PRINT 199
199 FORMAT (//1X,47HMEAN VALUES OF THE FIRST TWO CANONICAL VARIATES/1X,
   C47(1H-)/)
    IXI=IV/4
    PRINT 202
202 FORMAT (12X, 4(3HVAR, 3X, 4HZ(1), 6X, 4HZ(2), 4X)/12X, 96(1H-)//)
    DO 204 IW=1, IXI
    IC=4* IW-3
    ID=IC+3
204 PRINT 203,(I,P(I,1),P(I,2),I=IC,ID)
203 FORMAT (12X, 4(I3,1X, F8·2, 2X, F8·2, 2X)
    IXX=IV-4* IXI
    IF(IXX) 207, 207, 205
250 IE=4* IXI+1
    PRINT 206, (I,P(I,1), P(I,2), I=IE, IV)
206 FORMAT (12X, 3(I3,1X, F8·2, 2X, F8·2, 2X)//)
```

```
  207 CAR=CARE-Q(1)-Q(2)-Q(3)-Q(4)
      PRINT 184, (MP, Q(MP), MP=1, 4), CAR, CARE
  184 FORMAT (3X,4(6HLAMDA(,I1,2H)=,F10.2,5X)//3X,29HSUM OF OTHER CANONIC
     1AL ROOTS,  F12.2//3X,27HSUM OF ALL CANONICAL ROOTS=, F12.2/)
      DO 208 I=1,4
      PC(I)=Q(I)* 100·/CARE
  208 PRINT 1840, I, PC(I)
 1840 FORMAT (3X, 31HCONTRIBUTION OF CANONICAL ROOT (, I1, 2H)=, F6·1,8HPER CE
     1NT/)
C  FORMAT AND DIMENSION STATEMENTS SHOULD BE ALTERED FOR PROCESSING MORE
C  THAN 10 CHARACTERS AND 60 VARIETIES
      PRINT 1841
 1841 FORMAT (10(/))
      GO TO 3000
      END
BRASSICA DATA-SANKHYA, B, 1966-CANONICAL ANALYSIS
   623
      15·1160   11·8473   -3·1924    5·5606   3·6497   6·3706
      15·5324   12·9287   -2·0690    4·7833   3·4052   7·1270
      18·5428   14·9390   -2·6803    5·7312   3·8314   6·9518
      16·3650   11·5694   -1·9772    5·6333   3·1882   5·9725
      19·0232   12·1486   -1·5315    7·7647   1·0820   4·5527
      19·1833   12·4032   -1·5027    7·2932   1·1418   4·0143
      24·1793   17·2472   -5·1362    5·8404   2·7597   8·2199
      17·4219   12·2763   -1·3841    5·7479   3·7544   6·6845
      23·1865   16·0625   -4·8066    5·7081   2·3286   8·1825
      19·2794   13·6159   -2·5012    7·2878   3·3529   5·6867
      16·5572   12·5057   -2·2889    6·4964   3·3316   5·7896
      14·9879   11·3939   -2·2701    5·8260   2·5898   5·8027
      15·2762   12·4381   -2·8655    5·1892   3·2524   6·9431
      13·1625   10·1437   -1·1102    5·3993   3·2529   6·7608
      15·6285   11·4997   -1·6675    6·2255   3·3480   5·3748
      18·9911   14·0445   -3·4776    6·2857   5·2858   5·4406
      17.7101   13·9770   -1·9918    6·9975   4·4867   5·7314
      18·4467   12·1895   -1·5123    7·4379   3·2613   7·1130
      16·4291   16·0547   -5·2971    5·8176   1·1825   6·5672
      17·9663   12·8025   -2·0771    6·4993   2·4416   5·7626
      17·8382   13·4058   -1·9417    9·9482   3·1315   4·7832
      19·8558   15·4322   -1·3504    7·7722   3·8823   4·8055
      22·8663   18·0509   -6·1709   10·4761   1·7620   5·2424
```

BRASSICA DATA—SANKHYA, B, 1966—CANONICAL ANALYSIS

EVALUATION OF CANONICAL ROOTS AND VECTORS

V-MATRIX—BETWEEN PRODUCT SUM MATRIX OF Y-VALUES

165·3509	97·9272	–47·6280	37·3482	–13·3917	6·7202
97·9272	89·4273	–50·3447	21·6137	–7·1615	9·8639
–47·6280	–50·3447	44·1011	–4·6897	8·8704	–12·5373
37·3482	21·6137	–4·6897	43·8503	–7·7153	–20·3847
–13·3917	–7·1615	8·8704	–7·7153	23·5991	3·9614
6·7202	9·8639	–12·5373	–20·3847	3·9614	25·8817

V—MATRIX TO THE POWER 4

2823475200·000	2003125600·000	–1102233600·000	694544370·000	–277034430·000	155475760·000
2003125600·000	1423224300·000	–784532360·000	490171750·000	–196259110·000	112805620·000
–1102233600·000	–784532360·000	433512220·000	–267564310·000	107798700·000	–64081514·000
694544370·000	490171750·000	–267564310·000	176784330·000	–69245698·000	33006606·000
–277034430·000	–196259110·000	107798700·000	–69245698·000	27686788·000	–14322162·000
155475760·000	112805620·000	–64081514·000	33006606·000	–14322162·000	13244854·000

CANONICAL VECTOR (1)

·7603	·5397	–·2971	·1867	–·0745	·0421

RESIDUAL MATRIX ELIMINATING CANONICAL VECTOR (1)

1393500·000	–144000·000	806800·000	1465200·000	–177080·000	–1064260·000
–144000·000	1192900·000	–1533270·000	–1814150·000	269540·000	1684870·000
806800·000	–1533270·000	2377210·000	3332990·000	–413920·000	–2896195·000
1465200·000	–1814150·000	3332990·000	6570010·000	–1251896·000	–5438286·000
–177080·000	269540·000	–413920·000	–1251896·000	525989·000	1035031·000
–1064260·000	1684870·000	–2896195·000	–5438286·000	1035031·000	4561627·000

CANONICAL VECTOR (2)

·1528	–·2169	·3739	·6744	–·1228	–·5654

RESIDUAL MATRIX ELIMINATING CANONICAL VECTOR (2)

1063751·700	324114·490	–224·820	9867·900	87969·680	155998·000
324114·490	528359·400	–387608·400	251856·100	–106727·550	–47422·100
–224·820	–387608·400	402100·500	–228784·400	234761·600	90259·800
9867·900	251856·100	–228784·400	146954·700	–82108·900	–52721·900
87969·680	–106727·550	234761·600	–82108·900	312943·690	54195·100
155998·000	–47422·100	90259·800	–52721·900	54195·100	45972·200

CANONICAL VECTOR (3)

·7609	·5393	–·2965	·1868	–·0742	·0420

RESIDUAL MATRIX ELIMINATING CANONICAL VECTOR (3)

313475·000	–207685·570	292163·150	–174414·570	161188·340	114506·150
–207685·570	151416·760	–180362·290	121235·730	–54829·790	–76831·743
292163·150	–180362·290	288154·920	–156968·270	206227·790	106429·450
–174414·570	121235·730	–156968·270	101691·370	–64124·981	–62913·101
161188·340	–54829·790	206227·790	–64124·981	305798·370	58244·243
114506·150	–76831·743	106429·450	–62913·101	58244·243	43677·615

CANONICAL VECTOR (4)

·5567	–·3482	·5429	–·3000	·3773	·2034

MEAN VALUES OF THE FIRST TWO CANONICAL VARIATES

VAR	Z(1)	Z(2)	VAR	Z(1)	Z(2)	VAR	Z(1)	Z(2)	VAR	Z(1)	Z(2)
1	19·87	–1·75	2	20·34	–2·42	3	24·03	–1·94	4	20·34	–·71
5	23·03	2·22	6	23·17	2·18	7	30·45	–3·01	8	21·35	–·88
9	28·96	–2·80	10	24·10	·34	11	21·22	–·34	12	19·36	–·70
13	20·19	–2·26	14	16·86	–1·18	15	19·72	·01	16	24·06	–·93
17	22·81	– ·14	18	22·50	·20	19	24·00	–2·88	20	22·46	·01
21	23·20	2·71	22	25·19	1·22	23	31·00	1·15			

LAMDA(1)= 264·32 LAMDA(2)= 61·30 LAMDA(3)= 33·73 LAMDA(4)=31·41

SUM OF OTHER CANONICAL ROOTS = 1·43
SUM OF ALL CANONICAL ROOTS= 392·21
CONTRIBUTION OF CANONICAL ROOT (1)= 67·3 PER CENT.
CONTRIBUTION OF CANONICAL ROOT (2)= 15·6 PER CENT.
CONTRIBUTION OF CANONICAL ROOT (3)= 8·6 PER CENT.
CONTRIBUTION OF CANONICAL ROOT (4)= 8·0 PER CENT.