Multiple discriminant analysis: A BASIC program for microcomputers

DANIEL COULOMBE

University of Ottawa, Ottawa, Ontario, Canada

Multiple discriminant analysis (MDA) is a multivariate data-analysis technique that can be used to examine the possibility that a linear combination (discriminant function) of P correlated variables can be used to discriminate K groups of subjects, to test the significance of difference between groups centroids, and/or to classify subjects into various groups along the discriminant function. The foundations, as well as the mechanics involved, are described in most textbooks on multivariate analysis (Lindeman, Merenda, & Gold, 1980; Marascuilo & Levin, 1983; Pedhazur, 1982; Tabachnick & Fidell, 1983). The purpose of the program presented here is to perform multiple discriminant analysis, including tests of significance and generation of the classification functions. All critical computations are performed in double precision to minimize rounding errors. The routine to obtain the eigenstructure of the asymmetrical matrix, using the deflation power method, was derived from Douglass's (1983) algorithm. The computational procedure used in the program development was described in Marascuilo and Levin (1983) and in Pedhazur (1982).

Input. The program requests the number of variables (NV), the number of groups (K), and the number of subjects within each group (n_j). Unequal numbers of subjects per cell are allowed. The raw data can then be entered either sequentially from the keyboard or from a disk file. In the latter case, the data file editor (described earlier, in Coulombe, 1983) can be used to generate a data file suitable as input to this program. The file should contain, first, the total number of scores ($NV \times \sum n_j$) and then the scores themselves, from $A_1S_1V_1$ to $A_KS_{n(K)}V_{nv}$, the last index moving fastest (A = group, S = subject, and V = variable).

Output. All outputs are directed to the line printer. If screen output is desired, the LPRINT statements should be replaced with PRINT statements. This can be performed easily with the line editor function (EDLIN) provided with MS-DOS. Table 1 is a sample printout obtained using data presented in Lindeman et al. (1980, p. 188). It includes a listing of: (1) the mean and standard deviation vectors observed for each group and variables; (2) the characteristic roots (eigenvalues), index of discriminatory power, Wilks' lambda, R squared, and test of significance (Bartlett's test); (3) the discriminant functions, expressed in both standardized and unstan-

Table 1
Sample Output From MDA Program

DESCRIPTIVE STATISTICS FOR GROUP 1

VARIABLE	MEAN	STD. DEV.
1	13.200	4.541
2	23.500	2.550
3	9.000	3.712

DESCRIPTIVE STATISTICS FOR GROUP 2

VARIABLE	MEAN	STD. DEV.
1	16.462	3.307
2	21.000	3.651
3	9.385	2.694

DESCRIPTIVE STATISTICS FOR GROUP 3

VARIABLE	MEAN	STD. DEV.
1	13.214	4.742
2	14.714	3.099
3	13.857	4.111

FUNCTION	1	2
Eigenvalues	1.92276	0.13490
Discr. Pwr	0.93444	0.06556
Lambda	0.30147	0.88114
R-Squared	0.69853	0.11886
Chi-Square	39.5693	4.1758
D. F.	6	2

Standardized Discriminant Function (weights):

	1	2
X 1	0.0913	0.9255
X 2	0.8994	-0.2984
X 3	-0.4274	-0.2333

Raw Score Discriminant Function (weights):

	1	2
X 1	0.0247	0.2227
X 2	0.2434	-0.0718
X 3	-0.1157	-0.0561
Intercept	-3.78	-1.19

Discriminant Function evaluated at Group Centroid

Group	1	2
1	1.223	-0.447
2	0.650	0.437
3	-1.477	-0.086

Structure Coefficients Matrix

	1	2
X 1	0.178	0.948
X 2	0.950	-0.168
X 3	-0.675	-0.199

Discriminant Scores for Group 1

SUBJECT	1	2
1	0.316	-1.277
2	1.409	0.791
3	1.728	-0.971
4	0.378	-0.425
5	0.572	-0.942
6	1.855	-2.281
7	1.422	1.788
8	1.237	0.414
9	1.509	-0.677
10	1.802	-0.894

The author's mailing address is: School of Psychology, University of Ottawa, 275 Nicholas, Ottawa, Ontario, Canada K1N 6N5.

Discriminant Scores for Group 2

SUBJECT	1	2
1	1.834	0.386
2	-0.748	-0.232
3	1.901	0.587
4	0.439	-0.163
5	-0.126	0.554
6	0.064	1.278
7	1.207	0.251
8	0.312	0.555
9	1.596	-0.193
10	1.390	0.508
11	-0.592	2.752
12	1.187	-1.214
13	-0.010	0.610

Discriminant Scores for Group 3

SUBJECT	1	2
1	-2.035	-1.783
2	-0.444	-0.634
3	-1.892	-0.387
4	-0.376	0.159
5	-0.868	0.953
6	-2.788	-0.673
7	-1.137	0.211
8	-0.650	0.068
9	-2.985	-1.864
10	-1.656	-0.334
11	-0.022	0.794
12	-0.650	0.068
13	-2.322	1.260
14	-2.858	0.958

Likelihood of Group Membership for GROUP 1

Subject	P(G 1)	P(G 2)	P(G 3)	Pred. Mem
1	0.360	0.522	0.117	2
2	0.558	0.432	0.009	1
3	0.605	0.391	0.004	1
4	0.374	0.523	0.103	2
5	0.414	0.518	0.067	2
6	0.623	0.374	0.003	1
7	0.560	0.431	0.009	1
8	0.531	0.454	0.014	1
9	0.573	0.419	0.007	1
10	0.615	0.381	0.004	1

Likelihood of Group Membership for GROUP 2

Subject	P(G 1)	P(G 2)	P(G 3)	Pred. Mem
1	0.620	0.377	0.003	1
2	0.106	0.283	0.611	3
3	0.629	0.368	0.003	1
4	0.387	0.523	0.090	2
5	0.254	0.474	0.272	2
6	0.302	0.505	0.194	2
7	0.527	0.458	0.015	1
8	0.360	0.522	0.118	2
9	0.586	0.408	0.006	1
10	0.555	0.435	0.010	1
11	0.139	0.338	0.523	3
12	0.523	0.461	0.016	1
13	0.283	0.495	0.222	2

Likelihood of Group Membership for GROUP 3

Subject	P(G 1)	P(G 2)	P(G 3)	Pred. Mem
1	0.005	0.029	0.966	3
2	0.173	0.388	0.438	3
3	0.008	0.039	0.954	3
4	0.190	0.409	0.401	2
5	0.085	0.242	0.674	3

6	0.001	0.006	0.993	3
7	0.048	0.160	0.791	3
8	0.126	0.318	0.556	3
9	0.000	0.004	0.996	3
10	0.014	0.062	0.924	3
11	0.280	0.493	0.227	2
12	0.126	0.318	0.556	3
13	0.002	0.016	0.982	3
14	0.001	0.005	0.994	3

dardized (raw score) form; (4) the discriminant functions evaluated at group centroids; (5) the coefficients of structure (correlations between each original variable and the discriminant functions); (6) the discriminant scores computed for each subject on each discriminant function; and, (7) the probability of belonging to each group determined for each subject, together with the most probable predicted group membership or classification.

Program Language and Requirements. The program was written in GW-BASIC on a Tandy 2000 with two disk drives, under MS-DOS. This language is directly compatible with that of the IBM PC and work-alikes. The program can be transported easily to other computers that use Microsoft BASIC. It occupies about 7K of memory, leaving approximately 55K for execution under GW-BASIC. Under these conditions, problems involving 20 groups, 15 variables, and 200 subjects per group can be analyzed. In order to increase the speed of execution, the program could be compiled with a BASIC compiler.

Availability. A listing of the program is available free from the author. A diskette for IBM PC or Tandy 2000 is available for \$15.00 (cost of diskette, handling, and shipping). Please send postal money order only.

REFERENCES

- COULOMBE, D. (1983). MULREG: A multiple regression analysis system for microcomputers. *Behavior Research Methods & Instrumentation*, 15, 608.
- DOUGLASS, B. P. (1983, August). Eigenvalues and eigenvectors. *80 Micro*, pp. 328-339.
- LINDEMAN, R. H., MERENDA, P. F., & GOLD, R. Z. (1980). *Introduction to bivariate and multivariate analysis*. Glenview, IL: Scott, Foresman.
- MARASCUILLO, L. A., & LEVIN, J. R. (1983). *Multivariate statistics in the social sciences: A researcher's guide*. Monterey, CA: Brooks/Cole.
- PEDHAZUR, E. J. (1982). *Multiple regression in behavioral research*. New York: Holt, Rinehart & Winston.
- TABACHNICK, B. G., & FIDELL, L. S. (1983). *Using multivariate statistics*. New York: Harper & Row.

(Revision accepted for publication February 12, 1985.)