OKHOWAT, Valiollah, 1933-
MULTIPLE DISCRIMINANT ANALYSIS APPLIED TO AMERICAN COLLEGE TEST SCORES FOR THREE GROUPS OF COLLEGE MAJORS IN FOUR OKLAHOMA STATE COLLEGES.

The University of Oklahoma, Ph.D., 1970
Psychology, general

University Microfilms, A XEROX Company, Ann Arbor, Michigan

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

MULTIPLE DISCRIMINANT ANALYSIS APPLIED TO AMERICAN COLLEGE TEST SCORES FOR THREE GROUPS OF COLLEGE MAJORS IN FOUR OKLAHOMA STATE COLLEGES

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

VALIOLLAH OKHOWAT

Norman, Oklahoma

1970

MULTIPLE DISCRIMINANT ANALYSIS APPLIED TO AMERICAN COLLEGE TEST SCORES FOR THREE GROUPS OF COLLEGE MAJORS IN FOUR OKLAHOMA STATE COLLEGES

APPROVED BY

O. J. Rupiper

Gerald Kowitz

Herbert R. Hengst

George A. Letchworth

DISSERTATION COMMITTEE

ACKNOWLEDGMENT

The writer gratefully acknowledges the unselfish, generous and invaluable assistance of Professor Omer John Rupiper, chairman of the doctoral committee whose continual support and supervision from the initial planning stages of the study through its completion, made this work possible.

The writer also wishes to extend appreciation and express especial gratitude to Professor Gerald T. Kowitz, whose guidance and consultation sharpened my thinking in design and formulation of the study.

Acknowledgment should also go to Professor Herbert R. Hengst and Assistant Professor George Letchworth for their generous encouragement and guidance in completion of this study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

MULTIPLE DISCRIMINANT ANALYSIS APPLIED TO AMERICAN COLLEGE TEST SCORES FOR THREE GROUPS OF COLLEGE MAJORS IN FOUR OKLAHOMA STATE COLLEGES

CHAPTER I

INTRODUCTION

Background of the Study

That the world in general and western society in particular are becoming increasingly test sophisticated is a commonly accepted fact. This sophistication with its accompanying advantages and disadvantages is largely the result of man's increasing knowledge in a competitive and rapidly growing technological society.

In the near future selection and classification of persons qualified for receiving college education will be of fundamental importance. Selection and classification will be based largely on scores obtained from college entrance examinations. These test scores will be used not only for prediction of success in a single criterion activity, but also in a number of different criterion activities such as in various jobs or in different courses taken in college.

Crawford (1933) indicated that the most disappointing factor in general prediction studies has proved to be college entrance examination grades in spite of the fact that these examinations were prepared and scored with great care. These studies revealed that scores on standardized tests failed to satisfactorily predict either students' subsequent college achievements generally, or their competence in specific subjects, particularly.

Upon improvement of the internal structure of the commonly used types of measuring instruments better discrimination among people in different professional groups or different fields of study was found. Eckhart (1936) found that individuals who received college degrees signifying different areas of specialization were differentiated in terms of variates which had not been influenced by specialized training at the college level. Tiedeman (1954) stated that because

> investigators utilized univariate tests of significance of differences among means of the groups or profile methods of comparison, meaningful statements as to the magnitude of obtained differences or their practical utility were difficult to make.

In utilizing Fisher's discriminant function technique, Selover (1942) was able to demonstrate that an appropriately weighted combination of different measures markedly differentiated students in various pairs of major fields of study. Tiedeman (1954) pointed out that students majoring in various scientific fields tended to have combinations of interests, abilities and achievement at time of college entrance which differentiated them from those students majoring in the non-scientific fields. A similar finding was reported by Adkins (1940) earlier.

Although much study has been given to the problem of selection and classification of college students with respect to multiple variates, a limited number of studies are available where multivariate statistical techniques have been used. Tatsuoka and Tiedeman (1954) in their extensive review of the theoretical developments of discriminant analysis indicate that it is virtually unused in educational and psychological research. With the "increased availability of electronic computers... multivariate analysis should play an even more prominent role in educational research than it presently does..." (Tatsuoka, 1969, p. 740). Multivariate discriminant analysis can be used for classifying an individual in one of several groups on the basis of multiple criterion variables simultaneously. It can provide a parsimonious description of group differences.

The American College Test, hereafter referred to as ACT, is widely used as an instrument in selecting and classifying college students. Studies utilizing the results of ACT in making comparisons among groups and schools are numerous as well as reviews of the literature (Sanders, 1969). However, only one study may be regarded as relevant here since the statistical technique of discriminant analysis was used. Stone (1965) investigated the possibility of predicting the drop out of students entering the field of agriculture during or immediately following their freshman year. He used entrance scores, ACT, and high-school rank taken from the records on file at Kansas State University. There were 161 ACT scores available for the subjects out of which 116 had high-school ranks in percentile form. From a two-group discriminate analysis of the variables he found an index of discrimination of R = .472 with an F ratio equal to 6.28, $p<.001$. In view of the obtained relative weights he found that Mathematics, Natural Science, and high-school rank were the most meaningful discriminator variables. By reducing the number of variables to two, little discriminating precision was lost. The function for the two variables, Natural Science and high-school rank was 0.0571X + 0.0117X. Stone found that by using these two variables he could predict with 70 percent accuracy of classification. He indicated that this prediction system was not sufficient for making admission decisions. The very sparseness of references to the use of multiple discriminant analysis with ACT scores in the literature shows a need for studying this problem.

Statement of Problem

There are several methods of analyzing data with respect to grouping and classification. In order to analyze more deeply the internal structure of the obtained data multiple-discriminate function was used to determine whether more homogeneous and appropriate groupings could be made.

Without external criterion, the obtained scores on ACT by three selected major fields of study in four state colleges will be separated by use of multiple-discriminate analysis to arrive at a classification according to the several categories.

Significance of Study

The study has particular significance for colleges that use the ACT scores for selection and classification purposes. It would enable them to place students more appropriately in the group that they are most like providing the new students meet the same criteria used in selecting the sample studied. The findings of the study could also reveal the efficacy of the use of ACT scores in the classification of students according to their major field of study. The results of the study should indicate whether according to their performance on the ACT the students are properly classified. If discriminations are not indicated, it must be concluded that in subsequent research other measuring instruments, statistical techniques, or general plan of investigation should be employed.

CHAPTER II

DESIGN OF THE STUDY

The Data

Since its inception in 1959, the American College Testing Program published a technical report which presented an extensive account of the ACT battery. Odd-even reliabilities ranged from R = .84 to .95, whereas test-retest reliabilities ranged from R = .78 to .87 when parallel forms were used with several hundred high-school students (ACT Technical Report, 1965 ed., p. 16). Data obtained from a random sample of colleges who participated in the 1962, 1963 and 1964 research services programs yielded median predictive validities of grade-point averages on the four subtests from R = .37 to .50. In view of these reliabilities and validities members of the American College Testing Program stressed the importance of considering all four subtests for making academic predictions.

The ACT battery is comprised of four subtests in the areas of English, Mathematics, Social Studies, and Natural Science with representative items of scholastic tasks designed to measure directly the abilities applicable to college work. These tasks involve the comprehension of reading passages and the solution of functional and practical problems requiring quantitative reasoning.

A description of the four subtests follow:

> English. This measures the student's understanding and use of the basic elements in correct and effective writing: punctuation, capitalization, usage, phraseology, style, and organization.
>
> Mathematics. This test measures the student's mathematical reasoning ability and emphasizes practical quantitative problems that are encountered in many college curricula. It also includes a sampling of mathematical techniques taught in the high school courses.
>
> Social Studies. This test measures the evaluative reasoning and problem-solving skills required in the social studies. It measures the student's comprehension of reading passages taken from typical social studies' materials. It also contains a few items that test his understanding of basic concepts, knowledge of sources of information, and knowledge of special study skills needed in college work in the social studies.
>
> Natural Sciences. This test measures critical reasoning and problem-solving skills required in the natural sciences. Emphasis is placed on the formulation and testing of hypotheses and the evaluation of reports of scientific experiments. (Sanders, 1969, pp. 38-39).

The study completed by Sanders (1969) showed relationships among ACT subtest scores and personality variables on the Edwards Personal Preference Schedule for 384 graduating seniors from four major state colleges in Oklahoma. Intercorrelations reported among ACT variables ranged from R = .42 to .66. These statistically significant, positive correlations indicated a common underlying cognitive ability as measured by the subtests.

Since these data were not analyzed by various major fields of study or by college Sanders (1969, p. 60) recommended further investigation of the data among the four colleges which represented four distinct geographical regions of Oklahoma by means of a more complex statistical design. In view of this recommendation permission was granted to use these ACT data for further analysis. The original data or raw scores on the ACT for this study are presented in Appendix A by sex, major field of study and by college.

Schmid (1950) illustrates two different methods; one method proceeds from data expressed in deviation score units whereas the other method implies that the data has been expressed in standard units. It has, algebraically, been shown that both procedures yield identical discriminant functions.

The 384 subjects in this investigation were graduating seniors in the four major state colleges in Oklahoma, who have taken the ACT as entering freshmen. The data were obtained from the students' record in the registrars' offices. The subjects were classified according to sex and their major area of study which was delimited to (1) Education (2) Business-political and persuasive fields included and (3) Scientific fields.

The four major state colleges in which the subjects were currently enrolled were as follows: Northeastern at Tahlequah, Central at Edmond, Southeastern at Durant and Southwestern at Weatherford, Oklahoma. For purpose of simplicity in reporting, these colleges will be referred to hereafter as college I, II, II and IV, respectively. Table I shows the sampling distribution of subjects by college, major field and sex. Because of the small number of females in several categories the sex variables were combined which resulted in twelve programs.

TABLE 1

SUBJECTS BY COLLEGE, MAJOR FIELD AND SEX

	College												
	I			II			III			IV			
Major Field	Educ	Bus	Sci	Educ	Bus	Sci	Educ	Bus	Sci	Educ	Bus	Sci	Total
Male	16	26	23	12	21	23	16	27	16	25	18	12	235
Female	22	6	6	22	13	4	17	8	8	21	12	10	149
Total	38	32	29	34	34	27	33	35	24	46	30	22	384

Procedures

The raw data were punched on Hollorith cards and processed at the Merrick Computer Center, The University of Oklahoma. A multivariate analysis was used in order to classify individual subjects in one of the twelve groups on the basis of multiple measurements, ACT scores. The computer program employed was a stepwise discriminant analysis by Sampson (1967). This program was selected because (1) it gave canonical variates or linear functions which represented the structural space of the four ACT variables and (2) it provided a plotted scattergram of the first two canonical variables for each subject. A copy of the computer program is presented in Appendix B.

CHAPTER III

ANALYSIS OF THE DATA

The primary purpose of the present study was to investigate the classification of students into twelve groups defined according to three major fields of study within four major state colleges in Oklahoma. The criterion measure used was performance on each of the four subtests of the ACT battery. Because of the relatively small number of cases evident in some of the categories when divided by sex, males and females were combined for purposes of this analysis.

Before presenting the analysis, a brief description of the method and procedures will be given. Since the program yielded a multiple discriminant analysis in a stepwise manner, at each step one variable was entered into the set of discriminating variables at a time. The variable with the largest F value, that which produced the highest multiple correlation, and that which gave the greatest decrease in ratio of within to total generalized variance was added. If the F value became too low or was not significant, the variable was not included. The program consisted of an output of canonical variates.

The canonical correlation differs from the classical multiple correlation in that weights are found to apply to several criterion measures to form a composite, whereas, in multiple correlation a group of weighted variables are combined to predict only one criterion estimate. The basic idea in canonical correlation is to find two linear combinations which have maximal correlation. These correlations represent the interdependence of the pairs of canonical variates which are used in the classification system. These variates are graphically represented in the scatterplot of the first two pairs in order to show the relationships. The scatterplot or graphical pictorialization is an important strategy in the analysis of the data and stimulation of insight.

A summary of the size of sample, mean and standard deviation for each variable for each group is presented in Table 2. The within groups covariance matrix given in Table 3 and within groups correlation matrix in Table 4. All correlations among the subtests for the sample size of 384 were significant at or beyond the .05 level of significance.

According to the program, at each step one variable is entered into the set of discriminating variables. For inclusion the F value would have to meet the .05 level of significance. The variable first entered is selected with respect to the one with the largest F value. In this case variable 4, Natural Science, carried the highest F value of 5.29 df 11/372. The following entries in order of F value magnitudes were variable 2, Mathematics, F = 4.59, variable 3, Social Studies, F = 4.07 and variable 1, English, F = 2.39, respectively. No variables were deleted since the F values did not become too low. It should be mentioned that the F or liklihood ratio of between group variance to within group variance developed by Rao (1952) indicated the probability of significant differences among variables. The ratio of variances indicated differences among means (Anderson and Fruchter, 1957).

After entry of each variable the program provided an output of the F matrix to test differences among means between each pair of groups, the function, constant and number of cases classified into groups. Since all variables were included in the analysis the functions and constants for each variable for each group is presented in Table 5.

TABLE 2

MEAN AND STANDARD DEVIATION OF THE FOUR ACT SUBTEST SCORES

BY MAJOR, FIELD AND COLLEGE

		Eng.		Math.		Soc. Stu.		Nat. Sci.	
		$\bar{X}$	S.D.	$\bar{X}$	S.D.	$\bar{X}$	S.D.	$\bar{X}$	S.D.
College I_n									
Educ	38	16.34	4.94	15.00	5.68	16.86	6.02	17.50	5.53
Bus	32	17.00	3.49	19.31	5.13	18.96	6.04	18.50	6.91
Sci	29	19.20	5.57	22.86	5.04	22.58	6.35	24.82	4.72
College II									
Educ	34	18.00	4.90	18.14	5.60	20.05	6.60	20.58	4.79
Bus	34	19.55	5.20	19.61	5.68	20.67	5.88	21.05	5.52
Sci	27	18.37	3.62	20.74	4.02	21.18	4.59	21.77	4.59
College III									
Educ	46	17.52	4.62	18.58	8.32	17.21	5.80	18.36	5.01
Bus	30	17.83	4.05	17.23	5.01	16.68	5.34	18.53	5.07
Sci	22	19.22	4.91	19.27	5.64	21.18	5.02	20.72	5.14
College IV									
Educ	33	16.81	4.01	16.90	5.05	16.72	4.63	18.84	4.84
Bus	35	17.57	4.75	19.02	6.16	19.05	5.89	20.31	5.74
Sci	24	20.75	3.41	22.41	4.91	20.16	5.07	23.12	4.32
Total	384								

TABLE 3

WITHIN GROUPS COVARIANCE MATRIX

	1 English	2 Math	3 Soc. Stu.	4 Nat. Sc.
1. English	20.65			
2. Math	8.84	33.42		
3. Soc. Stu.	13.75	11.64	32.46	
4. Nat. Sci.	10.34	11.65	18.73	27.39

TABLE 4

WITHIN GROUPS CORRELATION MATRIX

	1 English	2 Math	3 Soc. Stu.	4 Nat. Sci.
1. English	1.00			
2. Math	0.34	1.00		
3. Soc. Stu.	0.53	0.35	1.00	
4. Nat. Sci.	0.43	0.39	0.63	1.00

FUNCTIONS AND CONSTANTS BY VARIABLE AND GROUP

Group	Functions for ACT Variables				Constant
Coll/Major	Eng.	Math	Soc. Stu.	Nat. Sc.	
A I-Educ	0.51	0.18	0.04	0.34	-8.89
B I-Bus	0.47	0.31	0.10	0.29	-10.74
C I-Sci	0.47	0.35	0.06	0.54	-15.89
D II-Educ	0.51	0.24	0.08	0.40	-11.74
E II-Bus	0.59	0.27	0.07	0.38	-13.21
F II-Sci	0.48	0.32	0.10	0.41	-13.23
G III-Educ	0.56	0.29	-0.01	0.34	-10.67
H III-Bus	0.61	0.24	-0.06	0.39	-10.57
I III-Sci	0.56	0.26	0.11	0.35	-12.85
J IV-Educ	0.53	0.23	-0.03	0.41	-10.03
K IV-Bus	0.50	0.28	0.04	0.41	-11.58
L IV-Sci	0.65	0.35	-0.07	0.49	-15.73

After each case was evaluated in view of the functions and constants, the posterior probability coming from each group was determined as well as the square of the Mahalanobis distance (D^2). The D^2 statistic was applicable for determining the distances between all possible pairs of dependent variables.

The generalized distance functions as Anderson and Fruchter (1957) state, indicate the exact differences between paired variable and provide distances in terms of a common unit between a number of groups. It is useful in the classification of groups since they show which groups belong together and which ones are separated by value that cannot be attributable to chance. The D^2 was computed because the discriminant analysis did not provide a significant test for classifying groups. If the ratio of the determinant of the "within" matrix and determinant of "total" matrix was found to be significant, the D^2 statistic is applicable to determine the distance between all possible pairs of variable. With reference to the posterior probabilities a classification matrix was prepared. This matrix is presented in Table 6. Upon inspection of Table 6, it is evident that according to performance on the ACT, declared field of study usually advised on the basis of ACT scores and college location a relatively large number of cases were misclassified. Of the 384 actual subjects, 324 were misclassified. College I appears to have the best classification, while College II tends to have the highest rate of misclassification. In terms of major fields of study, it appears that members in science are more accurately classified at College I and IV.

In summary, table of the printout tabulated the variables entered and accompanied by its respective F value, number of variables included, and the U statistic which was used to test the equality of group means. The summary is presented in Table 7. Table 8 presents eigenvalues, cumulative proportion of total dispersion and coefficients of canonical variables. It should be noted that variable 4, Natural Science accounted for approximately 65 percent of the total dispersion. Variable 2, Mathematics, accounted for an additional 18 percent and variable 3, Social Studies and variable 1, English, accounted for the remaining 17 percent of total dispersion.

In order to arrive at a configuration of the groups graphically, it was necessary to obtain the canonical variate for purpose of plotting the direction cosines corresponding to the roots obtained, normalized and compounded into linear functions of the four test variables which are called canonical variates or dimensions. Each canonical variate is a linear compound of the four original variables. In other words, weights are attached to the variables in making up the canonical variate. The canonical correlations and coefficients for canonical variates are presented in Table 8.

TABLE 6

CLASSIFICATION MATRIX

	Number of Cases Classified Into Each Group												Number of Cases Misclassified
Group	A	B	C	D	E	F	G	H	I	J	K	L	
A I-Educ	14	3	2	1	1	0	1	4	3	6	0	3	24
B I-Bus	5	6	6	0	0	0	3	4	3	1	0	4	26
C I-Sci	1	2	14	1	0	0	0	2	2	0	1	6	15
D II-Educ	6	5	4	1	0	4	0	2	5	2	0	5	33
E II-Bus	5	6	7	1	1	2	1	2	3	0	1	5	33
F II-Sci	2	8	8	0	1	2	0	1	2	1	0	2	25
G III-Educ	7	7	2	2	2	2	5	4	2	7	0	6	41
H III-Bus	9	3	2	1	0	0	2	2	2	1	1	7	28
I III-Sci	3	4	4	0	2	1	0	3	1	1	0	3	21
J IV-Educ	5	7	2	2	0	0	3	5	1	3	1	4	30
K IV-Bus	5	4	7	3	1	0	3	2	3	3	0	4	35
L IV-Sci	1	0	6	0	0	0	2	2	1	0	1	11	13

TABLE 7

F-VALUES AND U-STATISTICS FOR VARIABLES ENTERED

Step No.	Variable Entered	F-Value to Enter	Number of Variables Included	U-Statistic
First	4	5.2940	1	0.8646
Second	2	2.2035	2	0.8116
Third	3	1.5409	3	0.7761
Fourth	1	1.2875	4	0.7474

TABLE 8

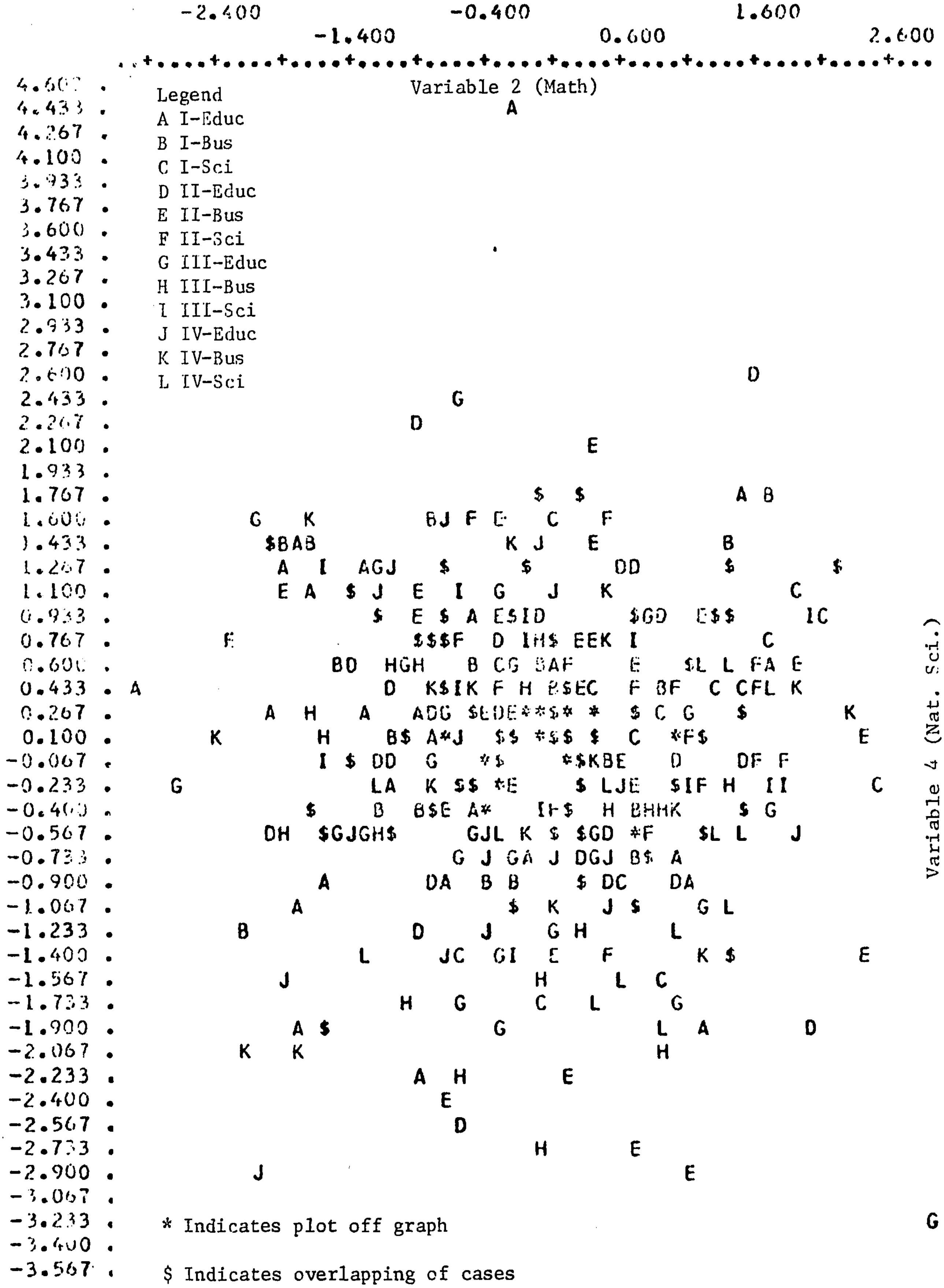
EIGENVALUES, CUMULATIVE PROPORTION OF TOTAL DISPERSION,

CANONICAL CORRELATIONS AND COEFFICIENT FOR CANONICAL VARIABLES

	Nat. Sci.	Math	Soc. Stu.	Eng.
Eigenvalues	0.2018	0.0579	0.0274	0.0244
Cum. Prop. of total dispersion	0.6479	0.8337	0.9216	1.0000
Canonical Correlation	0.4098	0.2339	0.1632	0.1544
Coefficients Original variable				
1	-0.0055	-0.1731	0.0238	0.2008
2	0.0892	-0.0266	0.1487	-0.0786
3	0.0312	0.2256	0.0354	0.0787
4	0.1072	-0.0878	-0.1930	-0.0889

The scatterplot is shown in Figure 1. The first canonical variate (variable 4, Natural Science) was plotted against the second variate (variable 2, Mathematics) using the respective mean ordinates for each group. The asterisk (*) represents the group means and the dollar sign ($) indicates overlap of cases. It is obvious from this figure that no significant grouping is evident. It appears that the findings here tend to corroborate the large number of misclassifications indicated earlier.

The results indicate the relatively high number of misclassifications and the lack of clusters merging from plotting each case with the pair of variables carrying most weight.



* Indicates plot off graph

$ Indicates overlapping of cases

Fig. 1. Scatterplot of first and second canonical variates

CHAPTER IV

SUMMARY, CONCLUSION AND RECOMMENDATIONS

The purpose of this study was to investigate the use of discriminant analysis with ACT subtest scores to determine the cogritive abilities associated with different college majors among four state colleges in Oklahoma, and to see if students majoring in different fields of study (Education, Business and Science) were properly classified as to their major field of study by college. The generalized multivariate hypothesis that the three groups of college majors in Education, Business and Science performed similarly on the subtests of the ACT among the four state colleges was tested.

The subjects employed in this study were comprised of 235 male and 149 female graduating seniors who were currently enrolled in four major colleges; Northeastern State College, Tahlequah, Oklahoma, Central State College, Edmond, Oklahoma, Southeastern State College, Durant, Oklahoma, and Southwestern State College, Weatherford, Oklahoma. Within the total number of 384 subjects randomly selected, 151 were majoring in Education, 131 were in field of Business and 102 were majoring in Science. All senior classes in the three major areas were listed and from this list the classes for testing were randomly selected (Sanders, 1969).

The scores of the subjects on the American College Test for English, Mathematics, Social Studies and Natural Science were subjected to discriminant analysis, a method in which the multiple data are combined so as to maximize the differences between each group. Review of the literature revealed that this method is the proper procedure to employ not only because it tests the hypothesis of no differences between groups, but, if there is a difference, it also provides information with respect to dimensions which account for the group variance (Tiedeman, 1951; Rulon, 1951 and McFadden, 1965).

ANALYSIS OF DATA

Employing a stepwise computer program written by Sampson (1967) for discriminant analysis, the raw data were punched and processed at the Merrick Computer Center, The University of Oklahoma. Since the structural space of the four ACT variables and a plotted scattergram of the first two canonical variates for each subject were desirable, this program was selected. Appendix B presents an outline of the program. In evaluating each case in view of the functions and constants the posterior probability and the square of the Mahalanobis distance from each group were determined.

Upon inspection of the classification matrix it was found that according to performance on the ACT, a large number of cases were misclassified. The Northeastern State College appeared to have the best classification, while the Central State College tended to have the highest misclassification. In terms of major field of study it appeared that students in Science were more accurately classified at Northeastern and Southwestern State Colleges.

An examination of the variables entered and respective F-Values, number of variables included in different steps of the computation and U-Statistic it was found that variable 4, Natural Science, accounted for approximately 65 percent of the total dispersion. Variable 2, Mathematics, for an additional 18 percent Social Studies and English were responsible for the remaining 17 percent of the total separation between groups.

In making up canonical variates weights were attached to the variables on the basis of the canonical correlation and coefficients. From the scatterplot it is obvious that a significant grouping was absent. This tended to corroborate the large number of misclassifications found in this study.

CONCLUSION AND RECOMMENDATIONS

The results indicated a relatively high number of misclassification and the lack of clusters emerging from plotting each case with the pairs of variables carrying the most weight. It was evident that performance on the ACT did not tend to differentiate between different groups.

Whereas the multivariate method for analyzing a large mass of data and the Generalized Distance Function (D^2) which provides distance seems to be useful for classification purposes, however, the findings herein may be due to the few variables employed and the large variety of groups used.

While this study did provide information about the limited discriminating power of the American College Test, analysis of the present investigation indicates the need for further research. Perhaps we should do either a cluster analysis or a straight forward discriminant analysis with major fields of study combined.

A replication of the study for further explanation of the nature of the functions is highly suggested. This investigation should provide additional information with respect to different subtests of the ACT. Further investigation is also recommended to determine whether lack of ability of the ACT to differentiate between groups were due to this possibility that all subjects were similar, or to questionable reliability or validity of the test, or combination of two or all three of the above.

REFERENCES

Adkins, Dorothy G. "The Relation of Primary Mental Abilities to Vocational Choice." American Council on Education Studies Series, Washington, D. C., The American Council on Education, 1940, 39-53.

American College Testing Program. Technical Report, 1965 Edition, Research Development Division of American College Testing Program and Science Research Associates, Inc., and Measurement of Research Center, Inc.

Anderson, Harry and Fruchter, Benjamin. "Statistical Procedures: Multivariate Analysis with the Generalized Distance Function and Canonical Variates, Research Guide No. 3, Psycometric Laboratory, Department of Educational Psychology, The University of Texas, 1957.

Crawford, Albert B. "Some Criticism of Current Practice in Educational Measurement." The Harvard Educational Review, 3, 1933, 67-81.

Eckhart, Ruth E. "The Significance of Curriculum Choice," Studies in Articulation of High School and College, Series II, Bulletin 8, University of Buffalo Studies, XIII, 1936, 311-335.

McFadden, Jack D. The Relationship of Values, Attitudes and Personality Characteristics of Student Teachers to Ratings by Their Supervisor. Unpublished Doctoral Dissertation, Northwestern University, Evanston, Illinois, 1965.

Rao, C. Radhakrishna. Advanced Statistical Methods in Biometric Research. New York: John Wiley and Sons, 1952, 390.

Sampson, Paul. "Class M-Multivariate Analysis, BMDO7M, Stepwise Discriminant Analysis, Health Science Computing Facility, University of California at Los Angeles, 1967.

Sanders, Robert G. The Relationship of Achievement and Personality Variables for Graduating Seniors Between Test Performance on the American College Test and the Edwards Personality Preference Schedule. Unpublished Doctoral Dissertation, University of Oklahoma, Norman, 1969.

Schmid, John, Jr., "A Comparison of Two Procedures for Calculating Discriminant Function Coefficients." Psychometrika, 15, 1950, 431-434.

Selover, Robert B. "A Study of the Sophomore Testing Program at the University of Minnesota, Part I, Journal of Applied Psychology, 26, 1942, 296-307.

Stone, Joice B. "Differential Prediction of Academic Success at Brigham Young University." Journal of Applied Psychology, 38, 1954, 109-110.

Stone, Le Roy A. "A Discriminant Analysis of Prediction of Dropouts for Freshman Year With Agricultural Students." The Journal of Educational Research, 59, 1965, No. 1, 37-38.

Tatsuoka, Maurice M. "Multivariate Analysis," Review of Educational Research, 39, 1969, 740.

Tatsuoka, Maurice M. and Tiedeman, David V. "Discriminant Analysis, Statistical Methodology in Educational Research, Review of Educational Research, 24, 1954, 402-416.

Tiedeman, David V. The Harvard Studies in Carrier Development, Working Paper No. 5, Paper read at Conference on Guidance, Harvard Teachers Association, Harvard University, 1954.

Tiedeman, David V. "The Utility of the Discriminant Function in Psychological and Guidance Investigation," Harvard Educational Review, 21, 1951, 71-80.

Tiedeman, David V. and Joseph G. "Prediction of College Field of Concentration," Harvard Educational Review, 24, 1954, 122-139.

Tiedeman, David V. and Sternberg, Jack J. "Information Appropriate for Curriculum Guidance." Harvard Educational Review, 22, 1952, 4, 267-274.

APPENDIX A

Original ACT Scores

NORTHEASTERN STATE COLLEGE, TAHLEQUAH - COLLEGE I

EDUCATION

MALE					FEMALE				
ACT					ACT				
	EN	MA	SS	NS		EN	MA	SS	NS
1	25	15	25	25	1	10	14	12	18
2	18	15	20	16	2	18	18	16	17
3	08	13	10	15	3	16	17	21	22
4	17	16	17	16	4	19	09	14	15
5	12	22	17	16	5	21	06	10	17
6	19	17	18	14	6	18	08	11	15
7	15	10	17	24	7	23	11	26	20
8	21	28	27	26	8	11	14	16	13
9	23	22	22	27	9	21	18	23	22
10	10	16	13	19	10	18	12	19	10
11	18	27	29	24	11	08	12	12	10
12	19	14	15	07	12	10	18	11	15
13	21	21	21	27	13	21	23	19	14
14	11	13	10	14	14	21	10	14	14
15	14	12	34	19	15	21	16	10	17
16	05	01	05	10	16	16	12	18	23
					17	22	26	17	26
					18	14	12	17	09
					19	15	12	15	26
					20	10	10	14	10
					21	17	13	15	16
					22	15	17	11	17

NORTHEASTERN STATE COLLEGE, TAHLEQUAH - COLLEGE I

BUSINESS

MALE					FEMALE				
ACT					ACT				
	EN	MA	SS	NS		EN	MA	SS	NS
1	11	17	12	09	1	21	24	21	22
2	25	26	22	20	2	21	18	24	20
3	13	08	17	12	3	19	13	16	13
4	22	22	26	24	4	10	10	14	10
5	14	27	21	27	5	16	19	13	18
6	18	25	30	32	6	16	11	07	10
7	20	25	28	26					
8	18	15	20	16					
9	16	19	13	18					
10	13	19	16	16					
11	20	13	34	37					
12	19	22	27	27					
13	17	21	15	16					
14	18	27	15	18					
15	20	18	15	11					
16	18	23	19	19					
17	10	15	11	15					
18	16	18	13	14					
19	15	25	19	22					
20	19	25	19	22					
21	13	15	17	17					
22	16	15	22	15					
23	17	20	14	19					
24	15	20	17	04					
25	15	20	17	04					
26	20	24	17	19					

NORTHEASTERN STATE COLLEGE, TAHLEQUAH - COLLEGE I

SCIENCE

MALE					FEMALE				
ACT					ACT				
	EN	MA	SS	NS		EN	MA	SS	NS
1	21	27	28	27	1	30	28	33	32
2	14	27	21	27	2	20	19	13	16
3	07	23	17	18	3	16	08	21	26
4	16	22	10	20	4	13	11	14	19
5	24	27	21	26	5	30	28	33	32
6	18	26	28	29	6	20	19	13	16
7	25	27	29	25					
8	18	21	28	29					
9	21	23	20	24					
10	23	23	25	22					
11	18	23	22	24					
12	05	18	13	20					
13	22	24	27	26					
14	19	21	19	26					
15	24	17	20	30					
16	19	17	23	27					
17	18	25	30	32					
18	25	26	21	20					
19	14	27	21	27					
20	21	27	28	27					
21	15	26	19	18					
22	20	25	28	26					
23	21	28	30	29					

CENTRAL STATE COLLEGE, EDMOND - COLLEGE II

EDUCATION

	MALE					FEMALE			
	ACT					ACT			
	EN	MA	SS	NS		EN	MA	SS	NS
1	22	21	21	27	1	22	07	22	19
2	21	26	28	24	2	27	26	31	20
3	10	09	21	19	3	24	17	27	26
4	22	20	20	22	4	22	20	24	22
5	17	23	17	22	5	20	20	22	12
6	16	09	15	18	6	15	18	25	22
7	05	09	03	14	7	12	12	14	14
8	14	21	18	16	8	19	19	20	17
9	17	18	26	25	9	21	18	28	24
10	14	24	31	27	10	22	30	19	30
11	16	15	22	22	11	12	20	14	14
12	20	23	19	24	12	23	24	26	28
					13	15	18	25	22
					14	20	21	23	27
					15	10	25	10	22
					16	20	18	24	20
					17	23	15	19	13
					18	25	12	17	18
					19	16	09	15	18
					20	13	15	04	22
					21	17	16	13	17
					22	20	19	19	13

CENTRAL STATE COLLEGE, EDMOND - COLLEGE II

BUSINESS

MALE					FEMALE				
ACT					ACT				
	EN	MA	SS	NS		EN	MA	SS	NS
1	14	09	16	11	1	21	10	10	15
2	18	18	22	17	2	22	22	13	21
3	11	21	10	14	3	17	09	16	08
4	23	18	21	19	4	18	22	13	20
5	19	15	22	29	5	25	18	27	25
6	13	22	20	14	6	19	20	22	21
7	18	17	24	24	7	31	25	19	26
8	18	27	19	25	8	19	19	20	18
9	15	16	06	20	9	19	14	22	15
10	16	22	24	20	10	19	15	21	14
11	20	21	26	27	11	32	25	19	22
12	23	28	29	32	12	24	17	28	26
13	17	12	19	22	13	32	26	28	26
14	14	22	14	20					
15	25	27	31	28					
16	11	22	23	20					
17	17	21	20	22					
18	22	21	29	27					
19	19	23	23	18					
20	15	18	25	22					
21	19	15	22	28					

CENTRAL STATE COLLEGE, EDMOND - COLLEGE II

SCIENCE

MALE					FEMALE				
ACT					ACT				
	EN	MA	SS	NS		EN	MA	SS	NS
1	19	23	19	24	1	25	21	24	18
2	18	22	20	18	2	20	21	16	24
3	17	24	19	20	3	23	21	26	27
4	21	16	20	23	4	20	16	19	16
5	23	21	25	28					
6	21	23	25	24					
7	20	26	22	25					
8	17	24	19	20					
9	12	20	18	20					
10	15	20	21	25					
11	17	18	23	19					
12	24	25	27	28					
13	20	27	26	26					
14	20	25	25	30					
15	16	14	06	13					
16	22	26	28	27					
17	16	14	23	18					
18	16	22	23	23					
19	17	16	19	14					
20	19	18	23	22					
21	12	19	16	15					
22	15	25	24	18					
23	11	13	16	23					

SOUTHEASTERN STATE COLLEGE, DURANT - COLLEGE III

EDUCATION

MALE					FEMALE				
ACT					ACT				
	EN	MA	SS	NS		EN	MA	SS	NS
1	22	12	19	21	1	16	16	11	11
2	15	61	06	16	2	18	13	18	19
3	19	25	10	12	3	22	11	23	15
4	21	23	17	27	4	24	16	22	26
5	14	15	16	15	5	11	05	07	10
6	12	16	18	16	6	10	14	17	23
7	12	15	18	16	7	23	16	20	17
8	11	17	22	15	8	23	21	17	17
9	24	20	31	28	9	18	23	15	10
10	10	13	07	13	10	20	26	20	15
11	07	14	05	12	11	18	10	21	19
12	22	19	22	22	12	19	25	10	12
13	18	14	21	23	13	23	18	25	23
14	19	21	11	18	14	25	17	19	16
15	19	27	22	28	15	19	25	10	12
16	21	22	25	26	16	17	15	18	17
17	16	19	22	18	17	14	03	19	14
18	19	24	18	20	18	27	27	24	24
19	15	18	19	19	19	12	17	15	21
20	10	17	21	21	20	16	13	16	21
21	19	24	25	22	21	21	18	17	23
22	17	22	11	17					
23	10	18	12	25					
24	14	15	18	12					
25	15	15	21	20					

SOUTHEASTERN STATE COLLEGE, DURANT - COLLEGE III

BUSINESS

MALE					FEMALE				
ACT					ACT				
	EN	MA	SS	NS		EN	MA	SS	NS
1	19	20	19	17	1	20	20	16	23
2	07	14	12	13	2	14	08	18	18
3	15	11	12	17	3	20	17	10	19
4	12	16	05	17	4	14	15	12	09
5	20	23	24	27	5	20	16	18	14
6	17	19	19	20	6	14	10	13	14
7	19	19	21	28	7	15	06	11	15
8	26	23	28	28	8	21	19	22	21
9	12	22	10	12	9	16	14	17	14
10	24	23	17	26	10	19	15	16	17
11	22	22	24	16	11	22	16	16	24
12	19	25	19	20	12	15	16	17	15
13	16	20	19	18					
14	21	09	17	19					
15	18	24	07	19					
16	16	15	11	10					
17	19	17	21	23					
18	23	23	23	23					

SOUTHEASTERN STATE COLLEGE, DURANT - COLLEGE III

SCIENCE

	MALE					FEMALE			
	ACT					ACT			
	EN	MA	SS	NS		EN	MA	SS	NS
1	10	17	21	21	1	23	21	17	17
2	07	14	12	13	2	28	24	27	25
3	21	32	29	25	3	23	16	20	17
4	21	25	20	21	4	19	06	17	17
5	25	26	28	29	5	14	10	18	13
6	17	20	12	19	6	18	14	19	26
7	18	19	22	18	7	24	21	31	28
8	22	25	25	29	8	20	19	20	14
9	19	18	18	21	9	24	24	25	25
10	16	19	20	14	10	20	15	20	20
11	14	19	19	19					
12	20	19	26	25					

SOUTHWESTERN STATE COLLEGE, WEATHERFORD - COLLEGE IV

EDUCATION

MALE					FEMALE				
ACT					ACT				
	EN	MA	SS	NS		EN	MA	SS	NS
1	15	16	11	21	1	18	10	08	13
2	13	22	21	17	2	21	17	19	23
3	11	16	14	21	3	23	21	21	23
4	08	14	13	14	4	24	06	07	15
5	14	27	16	16	5	17	18	19	24
6	19	29	22	29	6	15	09	12	17
7	15	15	19	13	7	18	17	17	18
8	16	15	19	17	8	11	16	17	16
9	19	19	21	25	9	14	19	20	13
10	21	20	19	17	10	21	20	20	22
11	13	16	21	23	11	16	18	12	13
12	18	10	19	22	12	20	14	24	25
13	22	21	19	23	13	15	07	13	21
14	17	25	18	18	14	18	17	20	15
15	17	18	15	18	15	16	16	11	10
16	08	18	08	20	16	20	18	13	16
					17	22	14	24	24

SOUTHWESTERN STATE COLLEGE, WEATHERFORD - COLLEGE IV

BUSINESS

	MALE					FEMALE			
	ACT					ACT			
	EN	MA	SS	NS		EN	MA	SS	NS
1	19	22	21	26	1	17	14	16	18
2	19	24	17	23	2	16	17	13	10
3	20	27	27	29	3	16	08	12	09
4	23	29	19	23	4	24	19	29	24
5	20	19	23	21	5	21	26	17	14
6	11	16	15	19	6	19	15	20	18
7	18	22	24	26	7	19	25	20	19
8	18	17	24	18	8	20	22	19	10
9	17	16	05	10					
10	19	14	25	27					
11	18	16	20	25					
12	06	19	12	15					
13	19	13	19	21					
14	24	26	24	22					
15	18	02	07	17					
16	20	29	27	31					
17	23	25	25	28					
18	05	10	12	13					
19	04	16	09	18					
20	12	19	21	24					
21	19	17	16	24					
22	15	14	18	24					
23	19	24	28	26					
24	19	23	17	18					
25	20	18	21	22					
26	19	28	23	16					
27	19	15	22	23					

SOUTHWESTERN STATE COLLEGE, WEATHERFORD - COLLEGE IV

SCIENCE

MALE					FEMALE				
ACT					ACT				
	EN	MA	SS	NS		EN	MA	SS	NS
1	22	23	28	27	1	26	25	28	24
2	20	18	19	17	2	24	27	29	27
3	20	23	20	21	3	26	27	22	26
4	16	25	23	25	4	21	09	19	21
5	20	23	14	22	5	25	27	25	26
6	17	25	19	27	6	15	15	08	14
7	26	28	26	25	7	20	19	18	23
8	26	21	20	27	8	18	19	18	22
9	21	18	17	27					
10	17	17	14	12					
11	19	29	25	26					
12	22	25	21	28					
13	21	27	22	22					
14	22	27	18	20					
15	17	23	16	27					
16	17	18	15	19					

APPENDIX B

Computer Program

BMD07M
STEPWISE DISCRIMINANT ANALYSIS

1. GENERAL DESCRIPTION

a. This program performs a multiple discriminant analysis in a stepwise manner. At each step one variable is entered into the set of discriminating variables. The variable entered is selected by the first of the following equivalent criteria:

(1) The variable with the largest F value (see computational procedure).
(2) The variable which when partialed on the previously entered variables has the highest multiple correlation with the groups.
(3) The variable which gives the greatest decrease in the ratio of within to total generalized variances.

A variable is deleted if its F value becomes too low. The program also computes canonical correlations and coefficients for canonical variables. It plots the first two canonical variables to give an optimal two-dimensional picture of the dispersion.

b. The output consists of:

(1) Group means and standard deviations
(2) Within groups covariance matrix
(3) Within groups correlation matrix
(4) At each step:

(a) Variables included and F to remove
(b) Variables not included and F to enter
(c) U statistic and approximate F statistic to test equality of group means
(d) Matrix of F statistics to test the equality of means between each pair of groups

(5) At certain specified steps and after the last step:

(a) Discriminant functions
(b) Classification matrix

(6) For each case:

(a) The posterior probability of coming from each group
(b) Square of the Mahalanobis distance from each group

(7) Summary table. For each step of the procedure the following is tabulated:

(a) Variable entered or removed
(b) F value to enter or remove
(c) Number of variables included
(d) U statistic

(8) Eigenvalues, canonical correlation, and coefficients of canonical variables
(9) Plot of the first canonical variable against the second

(10) Residuals and canonical coefficients (optional)

c. Limitations per problem:

(1) p, number of variables ($1 \leq p \leq 80$)
(2) t, total number of groups ($2 \leq t \leq 80$)
(3) j, number of Variable Format Card(s) ($1 \leq j \leq 16$)

d. Estimation of running time and output pages per problem:

Number of seconds = $.0006\, p^2(mp + 2n) + 60$ (for IBM 7094)

Number of pages = $.02n(m + 2k) + .01(pg^2 + p^2) + p + 10$

where p = number of variables
t = total number of groups
n = total number of cases
m = 1 if the canonical analysis is to be performed
0 otherwise
k = number of steps at which the cases are to be classified

2. ORDER OF CARDS IN JOB DECK

Cards indicated by letters enclosed in parentheses are optional. All other cards must be included in the order shown.

a. System Cards [Introduction, IV]

b. Problem Card

(c.) Covariance Weight Card(s)--COVAR

d. Sample-size Card(s)

e. Group Label Card(s)

f. F-type Variable Format Card(s) [Introduction, III-C]

(g.) F-type variable format for alternate output

(h.) Data Input Cards
(Place data input deck here
if data input is from cards.)

i. Subproblem Card

(j.) Control-Delete Card

Repeat g. and (h.) as specified on Problem Card

. . .

Repeat b. through (h.) as desired.

. . .

k. Finish Card [Introduction, III-D]

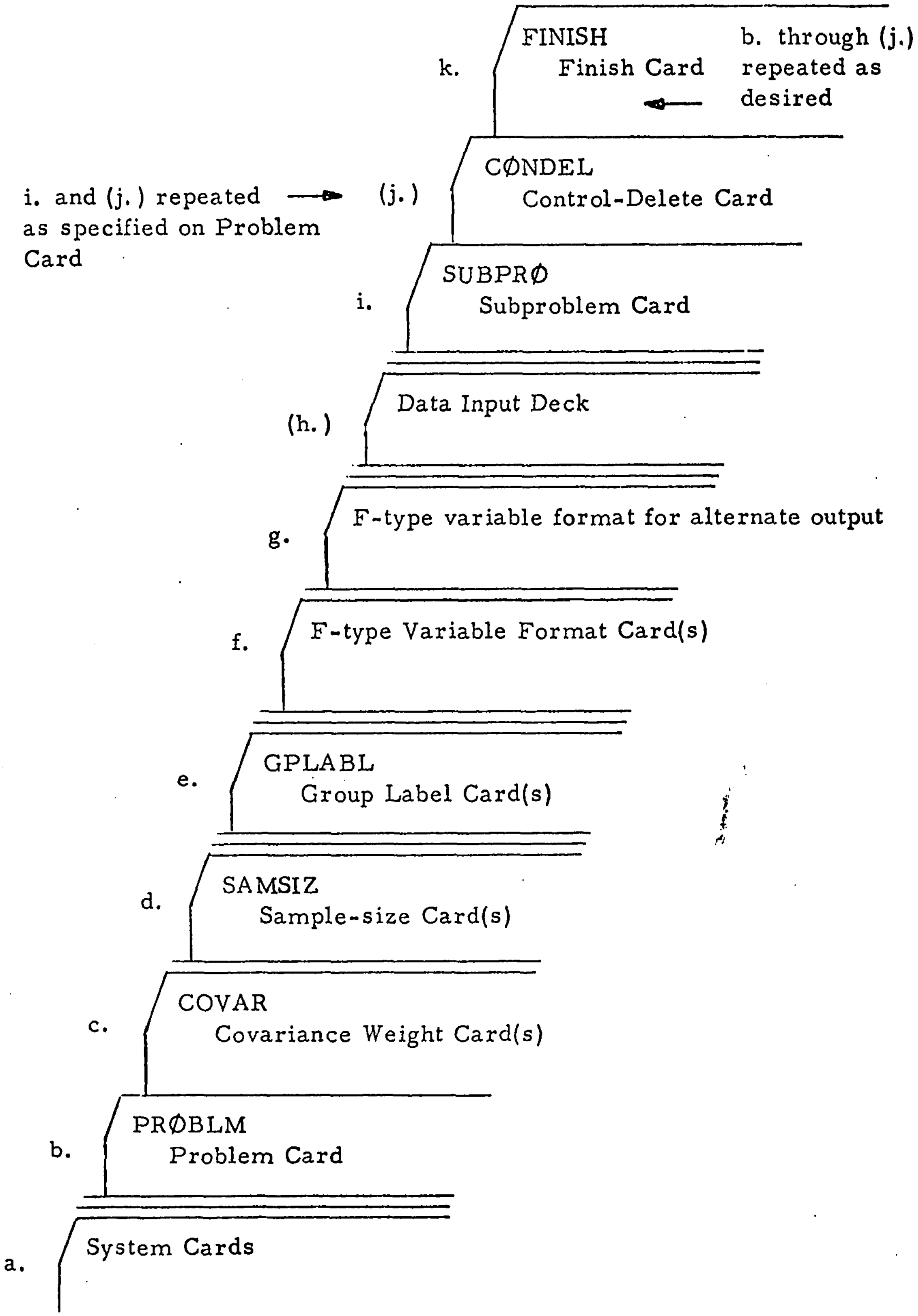
3. CARD PREPARATION (SPECIFIC FOR THIS PROBLEM)

Preparation of the cards listed below is specific for this program. All other cards listed in the preceding section are prepared according to instructions in the Introduction.

b. Problem Card (One Problem Card for each problem)

Col.	1-6	PRØBLM	(Mandatory)
Col.	7-12	Alphameric problem name	
Col.	13, 14	Number of variables	($1 \leq p \leq 80$)
Col.	15, 16	Number of groups	($2 \leq t \leq 80$)
Col.	17, 18	Number of Subproblem Cards	
Col.	19, 20	Number of Variable Format Cards	($1 \leq j \leq 16$)
Col.	21, 22	Tape number if data is from tape ($\neq$ logical 1, 2 or 6); otherwise leave blank	

Example of Job Deck Set Up:



k. FINISH
Finish Card

b. through (j.) repeated as desired

(j.) CØNDEL
Control-Delete Card

i. and (j.) repeated as specified on Problem Card

i. SUBPRØ
Subproblem Card

(h.) Data Input Deck

g. F-type variable format for alternate output

f. F-type Variable Format Card(s)

e. GPLABL
Group Label Card(s)

d. SAMSIZ
Sample-size Card(s)

c. COVAR
Covariance Weight Card(s)

b. PRØBLM
Problem Card

a. System Cards

Col. 23, 24 Number of groups to be plotted on each page if the canonical analysis is to be done. Leave blank otherwise.

Col. 25-27 YES if group means are to be printed.

Col. 28-30 YES if standard deviations are to be printed.

Col. 31-33 YES if within groups covariance matrix is to be printed.

Col. 34-36 YES if within groups correlation matrix is to be printed.

Col. 37-39 YES if weighting of covariance matrix is desired.

Col. 40 Tape number of optional output for canonical variables.

Col. 41, 42 NO if alternate input tape is not to be rewound.

Col. 43 Tape number of optional output for coefficients of canonical variables.

c. Covariance Weight Card(s)

Col. 1-5 COVAR (Mandatory)

Col. 6 Blank

Col. 7-12 Weight for first group (keypunch decimal)

Col. 13-18 Weight for second group (keypunch decimal)

• • •

Col. 67-72 Weight for eleventh group (keypunch decimal)

Additional cards may be used if needed.

d. Sample-size Card

Col. 1-6 SAMSIZ (Mandatory)

Col. 7-12 Number of cases in the first group.

Col. 13-18 Number of cases in the second group.

• • •

Col. 67-72 Number of cases in the eleventh group.

If required, additional cards of the same form are used until all the groups are specified.

If the number of cases for a group is preceded by a minus sign, that group is deleted from the computation of everything except the group means and standard deviations, classification, and plotting. This allows classification of new cases.

e. Group Label Card

Col. 1-6 GPLABL (Mandatory)

Col. 7-12 Alphameric name of the first group

Col. 13-18 Alphameric name of the second group

• • •

Col. 67-72 Alphameric name of the eleventh group

If required, additional cards of the same form are used until all the groups have been labeled. The first non-blank character of each group name is used for plotting.

i. Subproblem Card

Col. 1-6	SUBPRØ (Mandatory)
Col. 7-10	Maximum number of steps (if blank, 2p is assumed)
Col. 11-16	F value for inclusion (F to enter). Keypunch decimal. (if blank, .01 is assumed)
Col. 17-22	F value for deletion (F to remove). Keypunch decimal. (if blank, .005 is assumed)
Col. 23-28	Tolerance level. Keypunch decimal. (if blank, .0001 is assumed)
Col. 29-31	YES if a Control-Delete Card is present.
Col. 32-34	YES if the posterior probabilities are to be printed.
Col. 35, 36 Col. 37, 38 ... ... Col. 71, 72	A list of integers. When the number of variables in the set of discriminating variables is equal to one of these numbers, the discriminant functions are printed, evaluated for each case and a classification matrix is printed.

(j.) Control-Delete Card

Col. 1-6	CØNDEL (Mandatory)
Col. 7	Control value for the first variable
Col. 8	Control value for the second variable
...	
Col. 72	Control value for the sixty-sixth variable

If required, a second card of the same form is used to specify control values for the remaining variables. The control values specify the following:

0 or blank - Variable is not used for this subproblem
1 - Free variable
2 - Low level forced variable
.
.
9 - High level forced variable

If no Control-Delete Card is indicated on the Subproblem Card, a value of 1 will be assigned to each variable.

4. COMPUTATIONAL PROCEDURE

Notation: p = number of variables
g = number of groups used for the analysis. This excludes those with negative group size (see 3.d.)
t = total number of groups
n_m = number of cases in group m
n = total number of cases
x_{mki} = value of variable i for case k of group m

Assume for simplicity that the first g of the t groups are used for the analysis.

Step 1. The data are read and the following are formed:

$$\text{Means} \quad \bar{x}_i = \frac{1}{n} \sum_{m=1}^{g} \sum_{m=1}^{n_m} x_{mki} \qquad i = 1, 2, \ldots, p$$

$$\text{Group means} \quad \bar{x}_{mi} = \frac{1}{n_m} \sum_{k=1}^{n_m} x_{mki} \qquad \begin{array}{l} i = 1, 2, \ldots, p \\ m = 1, 2, \ldots, t \end{array}$$

Group standard deviations

$$s_{mi} = \sqrt{\frac{1}{n_m - 1} \sum_{k=1}^{n_m} (x_{mki} - \bar{x}_{mi})^2} \qquad \begin{array}{l} i = 1, 2, \ldots, p \\ m = 1, 2, \ldots, t \end{array}$$

Within and total cross-product matrices

$$W = \left\{ w_{ij} \right\} ; \quad w_{ij} = \sum_{m=1}^{g} \sum_{k=1}^{n_m} (x_{mki} - \bar{x}_{mi})(x_{mkj} - \bar{x}_{mj})$$

$$T = \left\{ t_{ij} \right\} ; \quad t_{ij} = \sum_{m=1}^{g} \sum_{k=1}^{n_m} (x_{mki} - \bar{x}_i)(x_{mkj} - \bar{x}_j)$$

$$i = 1, 2, \ldots, p$$
$$j = 1, 2, \ldots, p$$

Within groups covariance matrix

$$V = \left\{ v_{ij} \right\} ; \quad v_{ij} = \frac{1}{n-g} w_{ij} \qquad \begin{array}{l} i = 1, 2, \ldots, p \\ j = 1, 2, \ldots, p \end{array}$$

Within groups correlation matrix

$$R = \left\{ r_{ij} \right\} ; \quad r_{ij} = \frac{w_{ij}}{\sqrt{w_{ii} \, w_{jj}}} \qquad \begin{array}{l} i = 1, 2, \ldots, p \\ j = 1, 2, \ldots, p \end{array}$$

Step 2. At each step of the procedure the variables are divided into two disjoint sets; those included in the discriminant functions and those not included. Assume for simplicity that the first r are included.

Let $W = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix}$ and $T = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix}$

where W_{11} and T_{11} are $r \times r$.

$$\text{Let} \quad A = \begin{bmatrix} W_{11}^{-1} & W_{11}^{-1} W_{12} \\ W_{21} W_{11}^{-1} & W_{22} - W_{21} W_{11}^{-1} W_{12} \end{bmatrix} = \left\{ a_{ij} \right\}$$

$$\text{and} \quad B = \begin{bmatrix} T_{11}^{-1} & T_{11}^{-1} T_{12} \\ T_{21} T_{11}^{-1} & T_{22} - T_{21} T_{11}^{-1} T_{12} \end{bmatrix} = \left\{ b_{ij} \right\}$$

The coefficients and constant terms of the discriminant functions are computed:

$$c_{ki} = (n-g) \sum_{j=1}^{r} \bar{x}_{kj} \, a_{ij} \qquad \begin{array}{l} i = 1, 2, \ldots, r \\ k = 1, 2, \ldots, g \end{array}$$

$$c_{k0} = -\frac{1}{2} \sum_{i=1}^{r} c_{ki} \, \bar{x}_{ki} \qquad k = 1, 2, \ldots, g$$

The following statistics are computed:

a) F values for testing the differences between each pair of groups

$$F_{m\ell} = \frac{(n-g-r+1)\, n_m \, n_\ell}{r(n-g)(n_m + n_\ell)} \sum_{i=1}^{r} (c_{mi} - c_{\ell i})(\bar{x}_{mi} - \bar{x}_{\ell i}) \qquad \begin{array}{l} m = 1, 2, \ldots, g \\ \ell = 1, 2, \ldots, g \end{array}$$

with degrees of freedom r and n-g-r+1

b) F values for each variable

(1) If variable j has been entered

$$F_j = \frac{a_{jj} - b_{jj}}{b_{jj}} \; \frac{n-r-g+1}{g-1}$$

with degrees of freedom g-1 and n-r-g+1

(2) If variable j has not been entered

$$F_j = \frac{b_{jj} - a_{jj}}{a_{jj}} \; \frac{n-r-g}{g-1}$$

with degrees of freedom g-1 and n-g-r

Under the usual normality assumptions these are the likelihood ratio tests of the equality over all g groups of the conditional distribution of variable j given the (remaining) entered variables.

c) U statistic to test equality of group means

$$U = \mathrm{Det}\,(W_{11}) \,/\, \mathrm{Det}\,(T_{11})$$

with degrees of freedom (r, g-1, n-g)

d) Approximate F statistic to test equality of group means

$$F = \frac{1 - U^{1/s}}{U^{1/s}} \quad \frac{ms + 1 - rq/2}{rq}$$

$$\text{where } s = \sqrt{\frac{r^2 q^2 - 4}{r^2 + q^2 - 5}}, \text{ if } r^2 + q^2 \neq 5$$

$$s = 1, \text{ if } r^2 + q^2 = 5$$

$$m = n - \frac{r + q + 3}{2}$$

$$q = g - 1$$

its degrees of freedom are rq and ms + 1 - rq/2. If either r or q is 1 or 2 the approximation is exact.

Step 3. To move from one step to the next, one variable is added or removed from the discriminating set according to one of the following rules:

a) If there are one or more variables which are entered, have a control value of 1 and an F value less than 'F to remove ', the one with the smallest F will be deleted.

b) If no variable satisfies a) and there are one or more variables which have not been included, pass the tolerance test, and have a control value greater than 1, the one with the largest control value and the largest F among all those with the same control value will be included.

c) If no variable satisfies a) or b) and there are one or more variables not entered which pass the tolerance test, have a control value of 1 and an F value greater than 'F to enter ', the one with the largest F will be entered.

If no variable satisfies a), b) or c), the process is terminated.

The tolerance test is used to maintain accuracy and may be controlled by the tolerance level on the Subproblem Card. A value of 10^{-n} specifies that roughly not more than n significant digits will be lost due to round-off.

Step 4. When the number of variables entered is equal to one of the numbers indicated on the Subproblem Card and after the last step the following are computed for $\ell = 1, 2, \ldots, t$; $m = 1, 2, \ldots, g$; $k = 1, 2, \ldots, n_\ell$:

a) Value of the m^th discriminant function evaluated at case k of group ℓ

$$s_{\ell mk} = c_{mo} + \sum_{j=1}^{r} c_{mj}\, x_{mkj}$$

b) Posterior probability of case k in group ℓ having come from group m

$$p_{\ell mk} = \frac{\mathrm{Exp}(s_{\ell mk})}{\sum_{i=1}^{g} \mathrm{Exp}(s_{\ell il})}$$

c) Square of Mahalanobis distance of case k in group m from group ℓ

$$D^2_{\ell mk} = \sum_{i=1}^{r} \sum_{j=1}^{r} (x_{mki} - \bar{x}_{\ell i})\, a_{ij} (x_{mkj} - \bar{x}_{\ell j})$$

This may be used as a chi-square variable with r degrees of freedom for classification purposed.

Step 5. At this point let p denote the number of variables which are included after the last step and let W and T be their within and total sum of product matrices. Let $B = T - W$. The eigenvalue problem

$$Bu_i = \lambda_i W u_i \qquad i = 1, 2, \ldots, p$$

is solved to find coefficients, u_i, of canonical variables and the amount of dispersion λ_i explained by each canonical variable.

The vectors are normalized so that

$$u_i' W^{-1} u_j = \delta_{ij}$$

The canonical correlations of $\rho_1, \rho_2, \ldots, \rho_p$ relative to the groups are then computed

$$\rho_i = \sqrt{\lambda_i/(1 + \lambda_i)}$$

For each case the first two canonical variables y and z are computed and plotted on a scattergram

$$y_{mk} = \sum_{j=1}^{r} u_{j1}(x_{mkj} - \bar{x}_j)$$

$$z_{mk} = \sum_{j=1}^{r} u_{j2}(x_{mkj} - \bar{x}_j)$$

5. REFERENCES

Anderson, T. W., Introduction to Multivariate Statistical Analysis, Wiley, 1958.

Efroymsen, M. A., "Multiple Regression Analysis," Mathematical Methods for Digital Computers, Part V, (17). Edited by A. Ralston and H. S. Wilf, Wiley, 1960.

Rao, C. R., Advanced Statistical Methods in Biometric Research, Wiley, 1962.

***

This program was written by Paul Sampson, a member of the staff of Health Sciences Computing Facility, UCLA.