DOCUMENT RESUME

ED 341 730 TM 017 925

AUTHOR	Freidrich, Katherine R.
TITLE	Canonical Correlation Analysis: An Instructional Tool for All Parametric Statistical Procedures.
PUB DATE	Jan 92
NOTE	29p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Houston, TX, January-February 1992).
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Analysis of Covariance; Analysis of Variance; College Mathematics; Comparative Analysis; *Correlation; Heuristics; Higher Education; Instructional Effectiveness; Methods Courses; *Multivariate Analysis; *Statistics; *Teaching Methods
IDENTIFIERS	Linear Models; *Parametric Analysis

ABSTRACT

It is argued that, given the importance and the increased use of multivariate techniques such as factor analysis and canonical correlation, students need to be made aware of multivariate methods and the appropriate ways in which they can be applied. As a general linear model that subsumes all other parametric measures, canonical correlation analysis provides a natural framework for instruction involving all of the various parametric procedures (e.g., analysis of variance and analysis of covariance). Furthermore, when canonical correlation analysis is used as an instructional tool, students gain an understanding of how all parametric procedures are special cases of canonical correlation analysis, that all parametric procedures involve the application of weights to derive synthetic scores, and that all parametric procedures are correlational, thus yielding a measure of effect that is important to the interpretation of one's results. A small heuristic data set (18 examples) is used to demonstrate how canonical correlation analysis can be used as an instructional device in teaching both univariate and multivariate parametric methods. Fourteen tables present data from the example and illustrate the discussion through comparison of analytic techniques. A 30-item list of references and 1 appendix are included. (Author/SLD)

Canonical Correlation Analysis:

An Instructional Tool for All Parametric Statistical Procedures

Katherine R. Friedrich

Texas A&M University 77843-4225

Paper presented at the annual meeting of the Southwest Educational Research Association, Houston, January 1992.

ABSTRACT

It is argued that given the importance and the increased use of multivariate techniques such as factor analysis and canonical correlation, students need to be made aware of multivariate methods and the appropriate ways in which they can be applied. As a general linear model that subsumes all other parametric methods, canonical correlation analysis provides a natural framework for instruction involving all the various parametric procedures (e.g., ANOVA, ANCOVA). Furthermore, when canonical correlation analysis is used as an instructional tool, students gain an understanding of how all parametric procedures are special cases of canonical correlation analysis, that all parametric procedures involve the application of weights to derive synthetic scores, and that all parametric procedures are correlational, thus yielding a measure of effect important to the interpretation of one's results. A small heuristic data set is employed to demonstrate how canonical correlation analysis can be used as an instructional device in teaching both univariate and multivariate parametric methods.

Canonical Correlation Analysis:
An Instructional Tool for All Parametric Statistical Procedures

Traditionally, graduate level courses in research methodology have focused primarily on univariate procedures. As research by Willson (1982) indicates, until the 1970's textbooks in the field emphasized analysis of variance (ANOVA) methods. Following Cohen's (1968) seminal article on linear regression as a general linear model, textbooks, such as Kerlinger and Pedhazur's (1973) text, stressed the application of regression techniques. This also led to extensive application of regression analyses, as reported by Willson (1980) in a review of a decade of research. More recently, researchers have noted an increase in the application of multivariate techniques, such as factor analysis and canonical correlation analysis (Goodwin & Goodwin, 1985; Thompson, 1989a). While these techniques have existed for some time, this slow trend towards an increase in the use of multivariate procedures can be attributed to the incorporation of such methods into major statistical computer packages (Krus, Reynolds, & Krus, 1976), which removed associated problems of mathematical complexity involved in calculation by hand.

Many have stressed the importance of multivariate techniques and their advantages over traditional univariate methods (Campo, 1990; Fish, 1988; Kerlinger, 1986). The strongest argument in favor of multivariate techniques has been summarized by LaGaccia (1991) as follows:

> Researchers who believe that most outcomes are multiply caused, and that most interventions have multiple outcomes, simply must use multivariate analyses, or risk the seriously incorrect interpretations that can directly result from the failure

to use analytic methods that honor a view of reality presuming that reality is complex (p.153).

Given the importance and the increased use of such techniques, students need to be made aware of multivariate methods and the appropriate ways in which they can be applied. As Kerlinger (1986) states, "one cannot conceive of modern behavioral research without also recognizing the necessity for students of research to study these admittedly difficult yet indispensable approaches to research problems" (p. ix).

Multivariate methods do involve complex, restrictive mathematical manipulations. The mathematical training required to implement these procedures by hand, however, is typically beyond the scope of most graduate programs within the behavioral sciences. Therefore, if students are to acquire a basic conceptual understanding of multivariate methods, a heuristic framework that does not require an extensive mathematical background is essential. Such a framework is inherent in the multivariate procedure referred to as canonical correlation analysis.

Researchers have for some time recognized that canonical correlation analysis, not regression analysis, is the most general linear model that subsumes all other parametric procedures (Baggaley, 1981; Fornell, 1978; Knapp, 1978). As such, it provides a natural instructional tool for all parametric methods, both univariate and multivariate. Knapp (1978) and others (Campo, 1990; Thompson, 1985) have detailed how canonical correlation analysis will produce the same results as other parametric methods. The purpose of the present paper is to demonstrate in concrete, mathematically simple terms, how canonical correlation analysis can be employed as an instructional device for teaching research methodology. This discussion will include a brief review of the basics of canonical correlation analysis, its value as an instructional tool, and its advantages over other statistical procedures.

The Basics of Canonical Correlation Analysis

In general, canonical correlation analysis is a method for investigating the relationship between two sets of variables, a set of dependent variables and a set of independent variables, where each set contains two or more variables (Thompson, 1984). Simply put, it involves the calculation of a set of weights for each group of variables which, when applied, yields a linear composite, or synthetic score, for each set. These weights are derived such that the bivariate correlation between the pairs of composite scores is maximized. This bivariate correlation is the canonical correlation, R_C, which can be squared to obtain an estimate of the variance shared by the composite scores. If the first set of variables contains p variables and the second set has q variables, where q is less than or equal to p, then a total of q-1 linear combinations are possible, and that each set of composite scores will be perfectly uncorrelated with all previously derived composites (Cooley & Lohnes, 1971; Stevens, 1986; Thompson, 1984). For a more detailed discussion of canonical correlation analysis and its interpretation, the reader is referred to Thompson's (1984) treatment of the method.

Instructional Value of Canonical Correlation Analysis

Thompson (1984) describes to canonical correlation analysis using the framework of the very familiar bivariate technique as the resulting canonical correlation is the bivariate correlation coefficient. This approach to canonical correlation analysis is attractive instructionally "because most students feel comfortable working with bivariate correlation coefficients" (Thompson, 1987, p.3). Furthermore, it aids students in gaining "important insights regarding the relatedness of all parametric methods" (Campo, 1990, p.9). Campo (1990) discusses three such "insights" that contribute to the value of canonical correlation analysis as an instructional tool. First, since canonical correlation analysis subsumes all other parametric methods, all such methods can be considered special cases of canonical correlation analysis. As such, canonical correlation analysis can be applied to perform any parametric analysis. The obverse is not true, however; that is, canonical correlation analysis can not be performed using less sophisticated methods (Campo, 1990).

Use of canonical correlation analysis as a heuristic framework also enables the student to see how all parametric methods apply weights to create synthetic scores. Furthermore, it is these synthetic scores that is the focus of all analyses (Campo, 1990; Thompson, 1987).

Finally, a bivariate approach to canonical correlation analysis demonstrates that all parametric methods are correlational and, as such, yield a measure of effect size analogous to r^2. Thompson (1989b) emphasizes the importance of interpreting effect size estimates with all analyses in order to gain an understanding of the importance of one's results.

Canonical Correlation Analysis as a Heuristic Framework

Before examining how canonical correlation analysis yields the same results as other parametric procedures, a comment concerning some of the various statistics reported by the different methods is in order. Students are familiar with the F statistic reported by most univariate procedures, and in particular, in ANOVA techniques. They are not, however, as familiar with the various test statistics that are reported with many multivariate procedures. In canonical correlation analysis, a particular test statistic of interest, other than the canonical correlation R_c, is Wilk's lambda. As Glass and Stanley (1970) have pointed out, all test statistics, such as Z, t, chi-square and F, are related. Although the relationship between F and Wilk's Lambda is not a direct one, Rao (1952) has provided a formula for converting the lambda resulting from a canonical correlation analysis to a value whose distribution approximates the F distribution. When either p or q is less than or equal to two, which is the case when canonical correlation analysis is applied in place of many univariate procedures, this conversion is exact (Knapp, 1978) and simplifies to (Cooley & Lohnes, 1971; Thompson, 1985):

$$F = \frac{1 - \text{lambda}}{\text{lamda}} * \frac{\text{df error}}{\text{df effect}}$$

Presented in Table 1 is a small heuristic data set employed to demonstrate how canonical correlation analysis produces the same results as other univariate and multivariate methods. The data set contains two continuous dependent variables, Y and X, and two independent variables, a discrete variable, A, which might represent experimental groupings or categories, and a continuous variable, B. The discrete variable B' is a dichotomization of the variable B, created by collapsing B in a manner similar to how researchers often treat aptitude variables when investigating an aptitude-treatment interaction effect through an ANOVA design. Note that both sets contain two variables, the minimum requirement for performing a true canonical correlation analysis.

As many researchers have noted, in order to apply canonical correlation analysis in place of some parametric methods, specifically various

ANOVA techniques, some form of contrast coding must be employed (Campo, 1990; Knapp, 1978; Thompson, 1985). A review of the various methods of contrast coding is beyond the scope of this discussion; the interested reader is referred to Pedhazur (1982), who provides an excellent elaboration of such coding techniques. For the present discussion, included with the data set in Table 1 is the contrast coding for both of the independent variables A and B'. Appendix A contains the SAS program statements to perform all of the analyses discussed below. The statements for creating the contrast coding are included.

Insert Table 1 about here.

Table 2 compares the Pearson product moment correlation between X and B to the canonical correlation analysis between the same variables. As noted previously, R_c is the bivariate correlation coefficient between the two composite scores derived through canonical correlation analysis. Since multiplicative constants applied to variables have no impact whatsoever on correlations between the variables, the r between the variables is also the canonical R_c between the variables after weighting by the canonical coefficients to transform the observed variables into latent composite scores. In this case, each set contains only one variable, that is, $p=q=1$. Thus, the two sets of variables cannot be reduced any further and the canonical correlation R_c is equivalent to the bivariate correlation r. Note that while R_c can never be negative, the magnitude between R_c and r will always be the same.

Insert Table 2 about here.

Table 3 demonstrates how canonical correlation analysis and t-test analysis yield the same results. A t-test analysis was performed for the dependent variable Y with the independent variable B'. Recall that the squared value a t statistic with n degrees of freedom is equivalent to an F statistic with 1 and n degrees of freedom (Glass & Stanley, 1970). Thus, using Rao's conversion to calculate F, it is evident that the two procedures are equivalent.

Insert Table 3 about here.

The conventional ANOVA summary table for a 2 X 2 factorial analysis on the dependent variable Y and the factors A and B' is presented in Table 4. To conduct a two factor ANOVA through canonical correlation analysis, four separate canonical correlation analyses are required. The first analysis includes all of the contrast variables, A1, A2, B1, A1B2, A2B2, representing both factors and their interaction (i.e., all possible effects). Each of the remaining three analyses excludes the specific contrast variables associated with a specific effect. The resulting lambda's from these analyses are contained in table 5.

Insert Tables 4 and 5 about here.

As Thompson (1985) notes, Wilk's lambda is analogous to the sums of squares (SOS) within or error in conventional oneway ANOVA, that is, lambda = 1 - ($SOS_{Between}/SOS_{Total}$) or ($SOS_{Error}/SOS_{Total}$). Both are estimates of effect, but, whereas SOS gets larger as an effect increases, lambda gets smaller.

As a measure of effect, lamda estimates the effect of all those variables included in its calculation. The effect of a particular variable, as measured by lambda, can be determined by partitioning out the effect of the other variables from the overall lambda, which measures the effect of all the variables. The calculations yielding the individual lambda's for each effect are presented in Table 6. The individual lambda's may then be converted into F statistics by applying Rao's (1952) conversion as they are in Table 7. Comparison of the values in Tables 4 and 7 demonstrates that factorial ANOVA and CCA yield equivalent F statistics.

Insert Tables 6 and 7 about here.

It is also of note that 1 - lambda, where the lambda of interest is the multivariate lambda of the full model, is equal to the squared canonical correlation coefficient R_c for the full model (Thompson, 1988). For this example, 1 - 0.05755 = 0.94245, which is equal to R_c for the canonical correlation analysis and eta^2 for the factorial ANOVA, providing further evidence that canonical correlation does subsume ANOVA.

An explanation of how canonical correlation also subsumes factorial MANOVA follows directly from the example on factorial ANOVA. The results of a factorial MANOVA for the dependent variables X and Y with the variables A and B' and the corresponding canonical correlation analysis are presented in Tables 8 through 10. As Thompson (1985) notes, the calculation's are simplified since MANOVA results are reported in the form of lambda's.

Insert Tables 8, 9 and 10 about here.

Results of a multiple linear regression analysis of X and B on Y and the corresponding canonical correlati... analysis is presented in Table 11. Note that in multiple regression, the multiple R squared is the squared correlation coefficient between the predictor score, Y, and the composite score, Yhat. Thus, it follows logically that the multiple R squared of multiple regression is equivalent to the squared canonical coefficient resulting from canonical correlation analysis. Furthermore, although it is not so readily apparent, the regression beta weights are related to the function coefficients generated through canonical correlation analysis. Thompson and Borrello (1985) provide a detailed discussion of how the two sets of coefficients are equated through a variance adjustment applying either R_c or R. Table 12 demonstrates this relationship for the current example.

Insert Tables 11 and 12 about here.

The relationship between canonical correlation analysis and discriminant analysis has been previously demonstrated by Tatsuoka (1989) and others (Dunteman, 1984; Xitao, 1992). The objective of discriminant analysis is the prediction of group membership on the basis of some set of scores. For this example, Table 13 presents the results of a discriminant analysis with the variables X and Y to predict membership for the variable A. Also included in Table 13 is the results of a canonical correlation analysis for X and Y with the contrast variables A1 and A2 representing the levels of A.

Insert Table 13 about here.

Again, the results are equivalent, except for the function coefficients. Equivalence of the function coefficients can be demonstrated by setting the largest coefficient to one (Tatsuoka, 1989), as is shown in Table 13. Although this method clearly demonstrates that the two sets of coefficients have the same ratio, the relationship between them is not clear. A more explicit description of the relationship between the two sets of variables has been provided by Xitao (1992). Similar to the comparison between multiple regression and canonical correlation analysis, the relationship between the resulting two sets of function coefficients can be demonstrated through a variance adjustment involving the pooled within-group covariance matrix (Xitao, 1992). Basically, the relationship between the two sets of function coefficients is:

$$a_D = \frac{a_C}{\sqrt{a_C' \, s_{pooled} \, a_C}}$$

where a_D is the vector of function coefficients from the discriminant analysis, a_C is the vector of function coefficients from the canonical correlation analysis, and s_{pooled} is the pooled within-group covariance matrix of the original predictor variables (Xitao, 1992). For this example, correspondence between the two sets of function coefficients for the first function is presented in Table 14.

Insert Table 14 about here.

A concrete example of how canonical correlation analysis subsumes factor analysis, specifically the principal components method, has not been included as a part of this discussion. However, some general comments concerning this issue can be made. First, it should be noted that both canonical correlation analysis and principal components are variable reduction techniques (Stevens, 1986). Both methods reduce a set of variables into a set of synthetic scores which contains all or most of the variance of the original variables. Whereas in canonical correlation analysis, the objective is to derive a set of scores such that the correlation *between* the two sets of variables is maximized, in principal components the objective is to maximize the correlation *within* a single set of variables (Campo, 1990). Thus, principal components analysis may be thought of as the case where there is only one set of variables instead of two.

Advantages of Canonical Correlation Analysis

Believing that most of the phenomenon that is of interest in the behavioral sciences have multiple causes and outcomes, researchers typically measure several different but related variables. Having been trained to apply traditional univariate methods, researchers will often conduct multiple tests within a single study. There are several problems with multiple univariate tests, however, that can be avoided through the application of multivariate procedures, including canonical correlation analysis.

First, the use of multiple univariate tests disregards the variance that is shared between the multiple dependent and multiple independent variables that exists in reality (Thompson, 1984). Furthermore, ANOVA techniques discard even more variance by requiring that all independent variables be scaled at the nominal level of measurement. As a result, the reality the researcher wishes to generalize to is distorted. Canonical correlation analysis, however, allows for variables of any level of measurement and was designed to examine multiple variables simultaneously (Thompson, 1984). Thus, the reality the researcher has strived to represent by collecting multiple measures is preserved.

The second problem with multiple univariate tests concerns the probability of committing a Type I error. As the number of hypotheses within a study increases, the experimentwise error rate, that is, the probability that one or more Type I errors in a study as a whole has occurred, inflates (Thompson, 1988). This problem can be alleviated through the application of multivariate techniques such as canonical correlation analysis where fewer hypotheses or a single hypothesis is tested. And finally, as both Fish (1988) and Thompson (1986) have demonstrated, by employing several univariate tests one may fail to find statistically significant results that are present when a multivariate test is employed.

Summary

Given the multivariate nature of reality, it is imperative for students of research in the behavioral sciences to become familiar with the various multivariate statistical procedures that are readily available through the use of computers. Although such techniques are mathematically complicated, the use of canonical correlation analysis as a heuristic framework enables students to gain a deeper conceptual understanding of all parametric methods, both univariate and multivariate. Canonical correlation analysis (Thompson, 1991) employed as an instructional tool demonstrates (a) how all parametric procedures are special cases of canonical correlation analysis, (b) that all parametric procedures involve the application of weights to derive synthetic scores, and (c) that all parametric procedures are correlational, thus yielding a measure of effect important to the interpretation of one's results. Furthermore, the application of multivariate techniques such as canonical correlation analysis can overcome serious problems associated with the use of multiple univariate tests. This is not to imply that all analyses should be carried out through canonical correlation analysis, but that students should be made aware of the various procedures that are available such that they are able to apply the appropriate method in their own research and provide a more accurate representation of a complex reality.

References

Baggaley, A. R. (1981). Multivariate analysis: An introduction for consumers of behavioral research. Evaluation Review, 5, 123-131.

Campo, S. F. (1990, January). Canonical correlation as the most general parametric method: Implications for educational research. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin. (ERIC Document Reproduction Service No. ED 315 440)

Cohen, J. (1968). Multiple regression as a general data-analytic system. Psychological Bulletin, 70, 426-443.

Cooley, W. W., & Lohnes, P. R. (1971). Multivariate data analysis. New York: John Wiley & Sons.

Dunteman, G. H. (1984) Introduction to multivariate analysis. Beverly Hills: SAGE.

Fish, L. J. (1988). Why multivariate methods are usually vital. Measurement and Evaluation in Counseling Development, 21, 130-137.

Fornell, C. (Ed.) (1982). A second generation of multivariate analysis (Vol. 2). New York: Praeger.

Glass, G. V, & Stanley, J. C. (1970). Statistical methods in education and psychology. Englewood Cliff: Prentice-Hall.

Goodwin, L. D., & Goodwin, W. L. (1985). Statistical techniques in AERJ articles, 1979-1983: The preparation of graduate students to read the educational research literature. Educational Researcher, 14, 5-11.

Kerlinger, F. N. (1986). Foundations of behavioral research (3rd ed.). New York: Holt, Rinehart and Winston.

Kerlinger, F. N., & Pedhazur, E. J. (1973). Multiple regression in behavioral research. New York: Holt, Rinehart, & Winston.

Knapp, T. R. (1978). Canonical correlation analysis: A general parametric significance testing system. Psychological Bulletin, 85, 410-416.

Krus, D. J., Reynolds, T. S., & Krus, P. M. (1976). Rotation in canonical variate analysis. Educational and Psychological Measurement, 36, 725-730.

LaGaccia, S. S. (1991). Methodology choices in a cohort of education dissertations. In B. Thompson (Ed.). Advances in educational research: Substantive findings, methodological developments (Vol 1, pp. 149-158). Greenwich: JAI Press.

Pedhazur, E. J. (1982). Multiple regression in behavioral research (2nd ed.). New York: Holt, Rinehart, & Winston.

Rao, C. R. (1952). Advanced statistical methods in biometric research. New York: John Wiley and Sons.

Stevens, J. (1986). Applied multivariate statistics for the social sciences. Hillsdale: Lawrence Erlbaum.

Tatsuoka, M. M. (1989). Multivariate analysis: Techniques for educational and psychological research (2nd ed.). New York: Wiley.

Thompson, B. (1984). Canonical correlation analysis: Uses and interpretation. Newbury Park: SAGE.

Thompson, B. (1985, March). Heuristics for teaching multivariate general linear model techniques. Paper presented at the annual meeting of the American Educational Research Association, Chicago. (ERIC Document Reproduction Service No. ED 262 073)

Thompson, B. (1986, November). Two reasons why multivariate methods are usually vital. Paper presented at the annual meeting of the Mid-south Educational Research Association, New Orleans.

Thompson, B. (1987, April). Fundamentals of canonical correlation analysis: Basics and three common fallacies in interpretation. Paper presented at the annual meeting of the Society for Multivariate Experimental Psychology, New Orleans. (ERIC Document Reproduction Service No. ED 282 904)

Thompson, B. (1988, April). Canonical correlation analysis: An explanation with comments on correct practice. Paper presented at the annual meeting of the American Educational Research Association, New Orleans. (ERIC Document Reproduction Service No. ED 295 957)

Thompson, B. (1989a, August). Applications of multivariate statistics: A bibliography of canonical correlation analysis studies. Paper presented at the annual meeting of the American Psychological Association, New Orleans. (ERIC Document Reproduction Service No. ED 311 070)

Thompson, B. (1989b). Statistical significance, result importance, and result generalizability: Three noteworthy but somewhat different issues. Measurement and Evaluation in Counseling Development, 22, 2-6.

Thompson, B. (1991). A primer on the logic and use of canonical correlation analysis. Measurement and Evaluation in Counseling and Development, 24, 80-95.

Thompson, B., & Borrello, G. M. (1985) The importance of structure coefficients in regression research. Educational and Psychological Measurement, 45, 203-209.

Willson, V. L. (1980). Research techniques in AERJ articles: 1969 to 1975. Educational Researcher, 9, 5-10.

Willson, V. L. (1982, January). Misuses of regression approaches to ANOVA and ANCOVA. Paper presented at the annual meeting of the Southwest Educational Research Association, Dallas. (ERIC Document Reproduction Service No. ED 222 522)

Xitao, F. (1992, April). Conceptualizing canonical correlation analysis as an overarching statistical testing model. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Table 1

Hypothetical Data Set with Contrast Coding for Heuristic Demonstration

Y	X	A	B	B'	A1	A2	B1	A1B1	A2B1
5	28	1	1	1	1	-1	1	1	-1
3	27	1	2	1	1	-1	1	1	-1
4	26	1	4	1	1	-1	1	1	-1
6	25	1	6	2	1	-1	-1	-1	1
7	26	1	7	2	1	-1	-1	-1	1
8	24	1	9	2	1	-1	-1	-1	1
6	25	2	1	1	-1	-1	1	-1	-1
5	23	2	2	1	-1	-1	1	-1	-1
7	26	2	3	1	-1	-1	1	-1	-1
12	24	2	6	2	-1	-1	-1	1	1
11	23	2	7	2	-1	-1	-1	1	1
10	22	2	8	2	-1	-1	-1	1	1
10	22	3	2	1	0	2	1	0	2
9	21	3	4	1	0	2	1	0	2
8	22	3	4	1	0	2	1	0	2
13	21	3	6	2	0	2	-1	0	-2
14	20	3	7	2	0	2	-1	0	-2
15	20	3	8	2	0	2	-1	0	-2

Table 2

Pearson Product Moment Correlation through CCA (X with B)

CCA		Pearson Correlation	
Squared R_c	0.20206		
R_c	0.44951	r	-0.44951
lambda	0.79794		
F	4.052		
df	1/16		
p	0.0613	p	0.0613

Table 3

t-test Analysis through CCA [Y by B' (1,2)]

CCA		*t*-test Analysis	
Squared R_c	0.63661	Mean of group 1	6.33
R_c	0.40528	Sd	2.3452
lambda	0.59472	Mean of group 2	10.67
		Sd	3.1623
		t	-3.3020
F	10.9032	t^2	10.9032
df	1/16	df	16
p	0.0045	*p*	0.0045

Table 4

Factorial ANOVA [Y by A (1,3), B' (1.2)]

Source	SOS	df	MS	Fcalc
A	108.00	2	54.00	54.00
B'	84.50	1	84.50	84.50
AB'	4.00	2	2.00	2.00
Error	12.00	12	1.00	
Total	208.00	17		

$eta^2 = 196.5/208.5 = 0.94245$

Table 5

Canonical Analysis of Four Models

Model	Predictors of Y	Lambda
1	A1 A2 B1 A1B1 A2B1	0.05775
2	B1 A1B1 A2B1	0.57554
3	A1 A2 A1B1 A2B1	0.46283
4	A1 A2 B1	0.07674

Table 6

Conversion to ANOVA Lambda's

Source	Models	Calculation	Lambda
A	1/2	0.05755/0.57554	0.09999
B'	1/3	0.05755/0.46283	0.12434
AB'	1/4	0.05755/0.07674	0.74993

Table 7

Conversion of Lambda's to ANOVA F's

Source	[(1 - lambda)/lambda]*[df error/df effect]	= Fcalc
A	[(1 - 0.09999)/0.09999] * [12/2]	= 54.00
B'	[(1 - 0.12434)/0.12434] * [12/1]	= 84.50
AB'	[(1 - 0.74993)/0.74993] * [12/2]	= 2.00

$R_c^2 = 1 - \text{lambda} = 1 - 0.05775 = 0.94245 = \text{eta}^2$

Table 8

Factorial MANOVA [Y, X with A (1,3), B' (1,2)]

Source	Lambda	Fcalc	df	p
A	0.04133	21.56	4/22	0.0001
B'	0.08247	61.19	2/11	0.0001
AB'	0.74275	0.88	4/22	0.49

Table 9

Canonical Analysis of Four Models

Model	Predictors of Y and X	Lambda
1	A1 A2 B1 A1B1 A2B1	0.01802
2	B1 A1B1 A2B1	0.43614
3	A1 A2 A1B1 A2B1	0.21854
4	A1 A2 B1	0.02427

Table 10

Conversion to MANOVA Lambda's

Source	Models	Calculation	Lambda
A	1/2	0.01802/0.43614	0.04132
B'	1/3	0.01802/0.21854	0.08246
AB'	1/4	0.01802/0.02427	0.74248

Table 11

Multiple Regression through CCA [Y with X and B]

CCA		Regression Analysis	
Squared R_c	0.73638	Squared R	0.7364
R_c	0.85813	R	0.85813
lambda	0.26362		
F	20.95	F	20.95
df	2/15	df	2/15
p	0.0001	p	0.0001

Table 12

Function Coefficient and Beta Weight Conversions

Predictor	Function Coefficient	*	R_c (or R)	=	Beta Weight	/	R_c (or R)	=	Function Coefficient
X	-0.7852	*	0.85813	=	-0.67384	/	0.85813	=	-0.7852
B	0.3598	*	0.85813	=	0.38071	/	0.85813	=	0.3598

Table 13

Discriminant Analysis through CCA [A with X and Y]

CCA		Discriminant Analysis	
Squared R_C	0.75284		
R_C	0.86766	R_C	0.86766
lambda	0.24494		
F	7.1438	F	7.1438
df	4/28	df	4/28
p	0.0004	p	0.0004

Raw Function Coefficients

CCA: Function I	Discriminant: Function I
X 0.39696 ----> 0.39696/0.39696 = 1	X 0.75003 ----> 0.75003/0.75003 = 1
Y -0.0125 ----> -0.0125/0.39696 = -0.03148	Y -0.02361 ---->-0.02361/0.75003 = -0.03148

Table 14

Relationship Between Canonical and Discriminant Function Coefficients

$$s_{pooled} = \begin{bmatrix} 1.6556 & -1.8333 \\ -1.8333 & 6.70 \end{bmatrix}$$

$$\sqrt{a_C' \, s_{pooled} \, a_C} = [0.39696 \;\; -0.01250] \begin{bmatrix} 1.6556 & -1.8333 \\ -1.8333 & 6.70 \end{bmatrix} \begin{bmatrix} 0.39696 \\ -0.01250 \end{bmatrix} = 0.52956$$

$$\frac{a_C}{\sqrt{a_C' \, s_{pooled} \, a_C}} = \frac{1}{0.52956} \begin{bmatrix} 0.39696 \\ -0.0125 \end{bmatrix} = \begin{bmatrix} 0.75003 \\ -0.02361 \end{bmatrix} = a_D$$

Appendix A

```
DATA D1; INFILE CAN;
INPUT Y X A B;
IF A = 1 THEN A1 = 1;
 ELSE IF A = 2 THEN A1 = -1;
 ELSE IF A = 3 THEN A1 = 0;
IF A = 1 OR A = 2 THEN A2 = -1;
 ELSE IF A = 3 THEN A2 = 2;
IF B < 5 THEN BB = 1;
 ELSE IF B > 5 THEN BB = 2;
IF BB = 1 THEN B1 = 1;
 ELSE IF BB = 2 THEN B1 = -1;
A1B1 = A1*B1;
A2B1 = A2*B1;
PROC SORT;
 BY A B BB;
PROC PRINT;
  VAR Y X  A B BB A1 A2 B1 A1B1 A2B1;
 TITLE 'RAW DATA SET WITH CONTRAST CODING';
PROC CORR;
 VAR X B;
 TITLE 'CORRELATION OF PREDICTOR AND CRITERION VARIABLE';
PROC CANCORR SIMPLE CORR;
 VAR X;
 WITH B;
 TITLE 'CCA SUBSUMES PEARSON CORRELATION';
PROC TTEST;
 CLASS BB;
 VAR Y;
 TITLE 'T TEST FOR DEP VAR Y AND INDEP VAR B';
PROC CANCORR SIMPLE CORR;
 VAR Y;
 WITH B1;
 TITLE 'CCA SUBSUMES T TEST: INDEP VAR CONTRAST CODING';
PROC ANOVA;
 CLASS A BB;
 MODEL Y=A BB A*BB;
 TITLE 'ANOVA WITH DEP VAR Y AND INDEP VARS A AND B';
```

Note: BB refers to the contrast variable B' in the text as B' is not a valid variable name in SAS.

```
PROC CANCORR SIMPLE CORR;
 VAR Y;
  WITH A1 A2 B1 A1B1 A2B1;
 TITLE 'CCA SUBSUMES FACTORIAL ANOVA';
PROC CANCORR SIMPLE CORR;
 VAR Y;
  WITH B1 A1B1 A2B1;
 TITLE 'CCA SUBSUMES FACTORIAL ANOVA';
PROC CANCORR SIMPLE CORR;
 VAR Y;
  WITH A1 A2 A1B1 A2B1;
 TITLE 'CCA SUBSUMES FACTORIAL ANOVA';
PROC CANCORR SIMPLE CORR;
 VAR Y;
  WITH A1 A2 B1;
 TITLE 'CCA SUBSUMES FACTORIAL ANOVA';
PROC CANCORR SIMPLE CORR;
 VAR Y;
 WITH X B;
 TITLE 'CCA SUBSUMES MULTIPLE REGRESSION';
PROC ANOVA;
 CLASS A BB;
 MODEL Y X=A BB A*BB;
 MANOVA H=_ALL_/SUMMARY;
 TITLE 'FACTORIAL MANOVA';
PROC CANCORR SIMPLE CORR;
 VAR Y X;
  WITH A1 A2 B1 A1B1 A2B1;
 TITLE 'CCA SUBSUMES FACTORIAL MANOVA';
PROC CANCORR SIMPLE CORR;
 VAR Y X;
  WITH B1 A1B1 A2B1;
 TITLE 'CCA SUBSUMES FACTORIAL MANOVA';
PROC CANCORR SIMPLE CORR;
 VAR Y X;
  WITH A1 A2 A1B1 A2B1;
 TITLE 'CCA SUBSUMES FACTORIAL MANOVA';
PROC CANCORR SIMPLE CORR;
 VAR Y X;
 WITH A1 A2 B1;
 TITLE 'CCA SUBSUMES FACTORIAL MANOVA';
```

```
PROC REG;
 MODEL Y=X B/STB;
 TITLE 'MULTIPLE REGRESSION OF X AND B ON Y';
PROC DISCRIM SIMPLE WCOV WCORR PCOV PCORR;
 VAR X Y;
 CLASS A;
 TITLE 'DISCRIMINANT ANALYSIS';
PROC CANDISC ALL;
 VAR X Y;
 CLASS A;
PROC CANCORR SIMPLE CORR;
 VAR A1 A2;
 WITH X Y;
 TITLE 'CCA SUBSUMES DISCRIMINANT ANALYSIS';
```