DOCUMENT RESUME

ED 222 546 TM 820 698

AUTHOR	Thompson, Bruce
TITLE	Stepwise Canonical Correlation Analysis: A New Research Technique.
PUB DATE	Feb 82
NOTE	15p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Austin, TX, February 11-13, 1982).
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Correlation; Multiple Regression Analysis; *Multivariate Analysis; Predictor Variables; *Research Methodology; *Statistical Analysis
IDENTIFIERS	*Stepwise Canonical Correlation Analysis

ABSTRACT

Virtually all parametric statistical procedures have been shown to be special cases of canonical correlation analysis, which is a useful research methodology particularly when augmented by the calculation of canonical structure, index, and invariance coefficients. A logic for conducting stepwise canonical correlation analysis based upon evaluation of canonical communality coefficients is presented. The coefficients indicate how much of a variable's variance is reproducible from the canonical solution. Variables with the smallest communality coefficients may be deleted in a stepwise procedure as a direct analogue to stepwise backward multiple regression analysis. A heuristic demonstration of the technique involving two criterion variables and five predictor variables in a 7 by 7 correlation matrix is presented. The application provides more insight into the dynamics of social science phenomena, lessens the probability of Type II errors, and provides estimates of the generalizability of results. (Author/CM)

Stepwise Canonical Correlation Analysis:

A New Research Technique

Bruce Thompson

University of New Orleans 70148

Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, 1982.

Abstract

Virtually all parametric statistical procedures have been shown to be special cases of canonical correlation analysis, and canonical correlation analysis has been shown to be a useful research methodology, particulary when augmented by the calculation of canonical structure, index, and invariance coefficients. This paper presents a logic for conducting stepwise canonical correlation analysis, based upon evaluation of canonical communality coefficients. A heuristic demonstration of the technique is included.

Analysis of variance (ANOVA) techniques and their analogues (ANCOVA, MANOVA, and MANCOVA) have been among the most widely used research methods employed in the social sciences (Edgington, 1974; Willson, 1980). However, these techniques may not be fully appropriate for use in research involving multiple independent variables, if some of the independent variables are higher than intervally scaled. This frequently occurs because, as Kerlinger (1979, p. 119) notes, most non-manipulated variables, i.e., what Cronbach (1957) has termed aptitude variables, tend to be higher than nominally scaled. The use of ANOVA techniques in this situation can be unfortunate, because "when we reduce interval level of scale data to the nominal level of scale we are doing nothing less than thoughtlessly throwing away information which we previously went to some trouble to collect. If research is conducted for the purpose of acquiring knowledge, then is it consistent with our purpose to employ a method which 'throws away' information which might provide a more refined understanding of the phenomena which we are studying?" (Thompson, 1981, p. 8; see also Cohen, 1968).

This logic suggests that researchers should consider more frequent use of more general analytic techniques when one or more independent variables are higher than nominally scaled. When this is the case, and when the researcher is also investigating multiple dependent variables, canonical correlation analysis is an appropriate analytic technique

(Thompson, 1980a). The procedure has been usefully applied in previous studies (e.g., Thompson, 1980b; Thompson & Miller, 1981; Thompson & Pitts, 1981), and the technique is particularly helpful when the analysis is augmented by the calculation of indices such as canonical structure and index coefficients (Thompson & Frankiewicz, 1979) and canonical invariance coefficients (Thompson, 1982). Indeed, Knapp (1978, p. 410) has noted that "virtually all of the commonly encountered parametric tests of significance can be treated as special cases of canonical correlation analysis, which is the general procedure for investigating the relationships between two sets of variables."

This paper presents a method for implementing a new extension of the technique, stepwise canonical correlation analysis, which may make canonical correlation analysis an even more useful procedure. The procedure is a direct analogue of multiple regression analysis. A computer program which implements this new canonical technique is available from the author; the program was used to generate the results presented in this report's heuristic example.

A Stepwise Canonical Logic

A canonical structure coefficient (Cooley & Lohnes, 1971) represents the correlation between a variable and a canonical function. The square of a canonical structure coefficient indicates the proportion of variance which a variable linearly shares with a canonical function. A variable's canonical communality coefficient (Thompson, 1980a, p. 19) is equal to the sum of all the variable's squared canonical structure coefficients; the number of structure coefficients which a variable has is equal to the number of variables in the smaller of the two variable sets.

In effect, canonical communality coefficients indicate how much of a variable's variance is reproducable from the canonical solution. Variables with small canonical communality coefficients, i.e., coefficients close to zero, contribute little to a canonical solution. Thus, variables with the smallest communality coefficients may be deleted in a stepwise procedure as a direct analogue to stepwise backward multiple regression analysis.

Stepwise canonical correlation analysis will produce more parsimonious results and will conserve degrees of freedom for hypothesis testing. For example, the degrees of freedom for testing the canonical correlation associated with the first canonical function is equal to the number of variables in each variable set times each other. If both variable sets consist of five variables, the degrees of freedom for testing the statistical significance of the first canonical correlation would be 25 (five times five). After stepwise deletion of one variable, the degrees of freedom for the first function would then be 20 (four times five). Thus the conservation of degrees of freedom can be sizeable, and tends to reduce the likelihood of Type II errors occurring as a function of variable set sizes.

Heuristic Example

Pitts, as part of her dissertation research, performed a stepwise canonical correlation analysis employing this technique. The theoretical framework and substantive implications of her work will not be discussed here (see Pitts & Thompson, 1982), but her statistical analysis will be summarized in order to present a concrete example of implementation of the technique. Her analysis involved two criterion variables and five predictor variables. The seven-by-seven correlation matrix upon which the analysis was based is presented in Table 1.

INSERT TABLE 1 ABOUT HERE.

The two canonical functions extracted from the matrix are presented in Table 2. The squared canonical correlation, i.e., the eigenvalue, associated with the first function was .48 ($\chi^2 = 84.8$, $df = 10$, $p < .05$); the squared canonical correlation associated with the second function was .04 ($\chi^2 = 5.2$, $df = 4$, $p > .05$).

INSERT TABLE 2 ABOUT HERE.

Although in this case the results would not have been altered, Pitts (Note 1) has argued that generally only predictor variables should be considered for stepwise deletion. This position is reasonable since researchers are generally most interested in understanding criterion rather than predictor variables; this approach is also more comparable to stepwise regression analysis, since the criterion variable is never "removed" in a stepwise regression analysis. However, in some cases the two variable sets can not readily be characterized as "criterion" or "predictor," and in these cases all variables should be given equal consideration for deletion. Thus, since in this study the attentional style variable had the smallest canonical communality coefficient (.16), the variable was deleted at the first step.

Two canonical functions were extracted from the six-by-six correlation matrix produced from the Table 1 matrix by ignoring correlations involving the deleted variable. These results are presented in Table 3. The squared canonical correlation associated with the first function was .42 (χ^2 = 72.6, $\underline{df}$ = 8, $\underline{p}$ < .05); the squared canonical correlation associated with the second function was .04 (χ^2 = 4.6, $\underline{df}$ = 3, $\underline{p}$ > .05). Since the reflective-impulsive variable had the smallest canonical communality coefficient at this step (.19), the variable was deleted from the analysis at the end of the second step.

INSERT TABLE 3 ABOUT HERE.

Two canonical functions were then extracted from the five-by-five correlation matrix produced from the Table 1 matrix by ignoring the correlations involving the two deleted variables. These results are presented in Table 4. The squared canonical correlation associated with the first function was .39 ($\chi^2 = 65.8$, $df = 6$, $p < .05$); the squared canonical correlation associated with the second function was .04 ($\chi^2 = 4.7$, $df = 2$, $p > .05$). Since the canonical communality coefficients for the variables left in the equation at the end of this step were relatively homogeneous, stepwise deletion of variables was terminated at step three. Thus, the Table 4 results were the results interpreted by the researcher. Of course, only the first function was interpreted since the second function was not statistically significant.

INSERT TABLE 4 ABOUT HERE.

Conclusions

The logic recommended here represents an extension of conventional canonical correlation analysis. The extension makes clear that, as noted by Baggaley (1981), parametric statistical techniques, including stepwise multiple regression, are special cases of canonical correlation analysis. However, the logic should be of more than heuristic value. Its application will provide more insight into the dynamics of social science phenomena, lessen the probability of Type II errors, and provide estimates of the generalizability of results. As Thorndike (1978, p. 188) explains, "as the number of variables increases, the probable effect of these sources of [error] variation on the canonical correlations increases. Therefore, the fewer variables there are in a canonical analysis which yields a correlation of a given magnitude, the greater is the likelihood that that correlation is due to real, population-wide sources of covariation, rather than to sample-specific sources." Finally, it should be noted that a forward stepwise canonical analysis could also be couched on evaluation of canonical communality coefficients; at each step the variable which would then have the highest structure coefficient would be added to the equation.

Reference Notes

1. Pitts, M.C. Personal communication, December 2, 1981.

References

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

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

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

Cronbach, L.J. The two disciplines of scientific psychology. American Psychologist, 1957, 12, 671-684.

Edgington, E.S. A new tabulation of statistical procedures used in APA journals. American Psychologist, 1974, 29, 25-26.

Kerlinger, F.N. Behavioral research: A conceptual approach. New York: Holt, Rinehart, & Winston, 1979.

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

Pitts, M.C., & Thompson, B. Cognitive styles as mediating variables in inferential comprehension. Paper presented at the annual meeting of the American Educational Research Association, New York, 1982.

Thompson, B. Canonical correlation: Recent extensions for modelling educational processes. Paper presented at the annual meeting of the American Educational Research Association, Boston, 1980. (ERIC Document Reproduction

Service No. ED 199 269) (a)

Thompson, B. The instructional strategy decisions of teachers. Education, 1980, 101, 150-157. (b)

Thompson, B. The problem of OVAism. Paper presented at the annual meeting of the Mid-South Educational Research Association, Lexington, 1981. [Order document #03980 from National Auxillary Publication Service, P.O. Box 3513, Grand Central Station, New York, NY 10017]

Thompson, B. Comparison of two methods for computing canonical invariance coefficients. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, 1982.

Thompson, B., & Frankiewicz, R.G. CANON: A computer program which produces canonical structure and index coefficients. Educational and Psychological Measurement, 1979, 39, 219-222.

Thompson, B., & Miller, A.H. The utility of "Social Attitudes" theory. Journal of Experimental Education, 1981, 49, 157-160.

Thompson, B., & Pitts, M.C. Validity of teachers' perceptions of children's cognitive styles. Measurement and Evaluation in Guidance, 1981, 14, 90-95.

Thorndike, R.M. Correlational procedures for research. New York: Gardner Press, 1978.

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

Table 1

Correlation Matrix
(n = 127)

Variable		HARD	EASY	GEFT	FI	AS	RI	GL
Ability on Hard	(HARD)	---						
Ability on Easy	(EASY)	.66**	---					
GEFT	(GEFT)	.15*	.31**	---				
Field Independence	(FI)	.40**	.49**	.34**	---			
Attentional Style	(AS)	-.23**	-.22**	-.16*	.00	---		
Reflective Impulsive	(RI)	.24**	.28**	.14	.08	-.06	---	
Grade Level	(GL)	.38**	.38**	.29**	.06	-.03	.17*	---

*p < .05
**p < .01

Table 2

Canonical Solution for First Step
(n = 127)

Variable	Function I F	Function I S	Function I SSQ	Function II F	Function II S	Function II SSQ	h^2
Ability on Hard	.43	.87	.75	-1.26	-.50	.25	1.00
Ability on Easy	.66	.95	.89	1.16	.33	.11	1.00
GEFT	-.10	.39	.15	1.02	.83	.70	.84
Field Independence	.69	.71	.51	.00	.32	.10	.61
Attentional Style	-.34	-.36	.13	.31	.16	.03	.16
Reflective Impulsive	.26	.41	.17	.02	.06	.00	.18
Grade Level	.53	.60	.36	-.49	-.20	.04	.39

NOTE: "F" = canonical function coefficients; "S" = canonical structure coefficients; "SSQ" = squared canonical structure coefficients; "h^2" = canonical communality coefficients.

Table 3

Canonical Solution for Second Step
($\underline{n}$ = 127)

Variable	Function I F	Function I S	Function I SSQ	Function II F	Function II S	Function II SSQ	h^2
Ability on Hard	.39	.85	.72	-1.27	-.52	.28	1.00
Ability on Easy	.70	.96	.92	1.13	.29	.08	1.00
GEFT	-.02	.42	.18	1.03	.83	.69	.86
Field Independence	.71	.76	.58	-.07	.24	.06	.64
Reflective Impulsive	.29	.44	.19	-.03	.01	.00	.19
Grade Level	.54	.63	.40	-.58	-.29	.08	.50

NOTE: "F" = canonical function coefficients; "S" = canonical structure coefficients; "SSQ" = squared canonical structure coefficients; "h^2" = canonical communality coefficients.

Table 4

Final Canonical Solution
($\underline{n}$ = 127)

Variable	Function I F	Function I I	Function I S	Function I SSQ	Function II F	Function II I	Function II S	Function II SSQ	h^2
Ability on Hard	.38	.53	.85	.72	-1.27	-.10	-.53	.28	1.00
Ability on Easy	.70	.60	.96	.92	1.13	.06	.29	.08	1.00
GEFT	.00	.28	.44	.19	1.02	.16	.82	.68	.87
Field Independence	.76	.50	.79	.63	-.08	.05	.24	.06	.68
Grade Level	.61	.41	.66	.43	-.59	-.06	-.29	.09	.52

NOTE: "F" = canonical function coefficients; "I" = canonical index coefficients; "S" = canonical structure coefficients; "SSQ" = squared canonical structure coefficients; "h^2" = canonical communality coefficients.