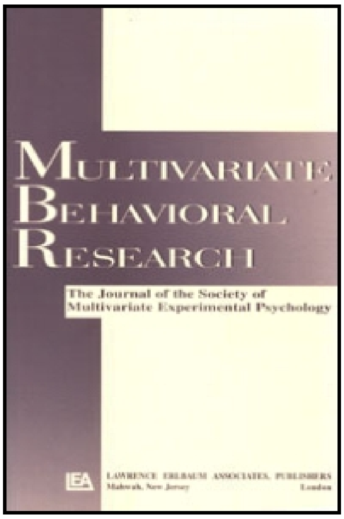


**Multivariate Behavioral Research**
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t775653673

# A Monte Carlo Study of the Stability of Canonical Correlations, Canonical Weights and Canonical Variate-Variable Correlations

Robert S. Barcikowski; James P. Stevens

Online Publication Date: 01 July 1975

# A MONTE CARLO STUDY OF THE STABILITY OF CANONICAL CORRELATIONS, CANONICAL WEIGHTS AND CANONICAL VARIATE-VARIABLE CORRELATIONS

ROBERT S. BARCIKOWSKI
Ohio University

and

JAMES P. STEVENS
University of Cincinnati



ABSTRACT

A Monte Carlo study was run to check the stability of canonical correlations, canonical weights, and canonical variate-variable correlations. Eight data matrices were selected from the literature for the canonical analyses, with the number of variables ranging from 7 to 41. The results showed that the canonical correlations are very stable upon replication. The results also indicated that there is no solid evidence for concluding that the components are superior to the coefficients, at least not in terms of being more reliable. However, the number of subjects per variable necessary to achieve reliability in detecting the most important variables, using components or coefficients, was quite large, ranging from 42/1 to 68/1.

Canonical correlation as a technique for determining the relationship between two sets of variables was brought to the attention of educational researchers by Cooley and Lohnes about ten years ago. Yet, there have been few studies reported in the literature that have used canonical correlation as a statistical tool.

It is important to realize that canonical correlation is a mathematical maximization procedure in which linear composites from each of two sets of variables are derived such that the correlation between each pair is maximized. Thus, the two sets of beta weights (for any pair of canonical variates) are optimal for that sample. For another sample different beta weights may be optimal. This difference in weights may be caused by considerable sample specific covariation, especially for relatively small sample size.

Two techniques that have been advocated for interpreting the canonical variates are (1) to examine the standardized coefficients of the variates, and (2) to examine the canonical variate-variable correlations (canonical components). For these approaches it is the variables with the largest coefficients or the largest correlations that one focuses on in interpetation.

Which of these techniques is better? Meredith (1964, p. 55) has commented, "If the variables within each set are moderately intercorrelated the possibility of interpreting the canonical variates by inspection of the appropriate regression weights is practically nil. However, the correlations between the canonical variates and the original measures can be very enlightening." Darlington, et al., (1973, p. 443) do not take quite such a strong position. They state, "The theoretical advantages of the two types of statistic have not been adequately explicated. A detailed analysis would probably show that the correlations are theoretically preferable in some situations and the weights in others." They go on to note that when the variables within a set are highly intercorrelated the researcher should emphasize the correlations, at least for small or medium sized samples, because they will have less sampling error.

However, the effect of sample specific covariation on the coefficients or components has not been examined (Thorndike and Weiss, 1973). That is, will the variables which have the largest coefficients or correlations for a given sample necessarily be the most important for another sample? Thorndike and Weiss (1973) carried out a study in which they recommended cross validation as a check on the extent of sample specific covariation. They suggested splitting the original sample randomly, and then performing a canonical correlation analysis on each of the samples. Then, apply the two sets of weights from each sample to the other sample to see if the relationship found is stable. They employed a similar procedure to check on the stability of the components. Basically their conclusions were that the components were consistent in cross validation, but that a relationship found between a pair of canonical variates in one sample may not hold up under cross validation.

In our study, through Monte Carlo techniques, the stability of the components and weights as well as the canonical correlations themselves was examined. Canonical correlation analyses were performed on sets of data having 7, 10, 12, 27, 31 and 41 variables. The various matrices examined displayed different types of within-set structure, i.e., the patterns of intercorrelations for the two sets of variables could be described as "weak" to "irregular" to "fairly strong."

## Theoretical Framework and Methods

The Monte Carlo technique employed in this study rests upon a procedure described by Huberty (1969) for generating data from a $p$-variate multivariate normal distribution. Essentially the procedure uses classical factor analysis to arrive at a population factor loading matrix **A**. The population correlation matrix **S** is then arrived at using the following equation: $\mathbf{S} = \mathbf{A}\ \mathbf{A}' + \mathbf{D}^2$, where **D** is the uniqueness diagonal matrix (Harman, 1967).

In this study **A** was developed using factor analysis procedures in data taken from actual studies, or was developed using regression techniques (Huberty, 1969). The regression techniques were used so that the resultant variables would have prescribed properties. For example, the communality of each of the variables could be arbitrarily set at .75. Then the reliability of each of the variables would be at least .75. In this case **D** would be a diagonal matrix with all elements equal to .50. The **A** matrix based on actual data was found using theBMDX72 program, such that the original correlation matrix was reproduced within rounding error.

After the population loading matrix **A** was developed using one of the preceding procedures, sample score matrices were generated and sample correlation matrices developed, using a technique similar to that suggested by Kaiser and Dickman (1962). Numbers were generated from a random normal (0,1) distribution using a subroutine called RANDNR. This technique enables the generation of elements for a sample ($m$ factor by $N$ people) matrix $\hat{\mathbf{F}}$, and a sample ($V$ variables by $N$ people) matrix $\hat{\mathbf{U}}$. A data matrix $\hat{\mathbf{X}}$ ($V$ variables by $N$ people) was then obtained using the following equation: $\hat{\mathbf{X}} = \mathbf{A}\ \hat{\mathbf{F}} + \mathbf{D}\ \hat{\mathbf{U}}$. From these data sample correlation matrices, $\hat{\mathbf{S}}$, were generated and the canonical correlation analyses were performed.

For each factor loading matrix the number of subjects varied from 200 to 3000, in increments of 200.[1] Each sample size was replicated 100 times. Canonical correlation analyses were performed on each replication, and the matrix of ranks for the standardized weights and canonical components were obtained. Also, for each set of 100 rankings Kendall's coefficient of concordance $W$ was calculated.

## Data Sources

There were eight data sources, i.e., eight correlation matrices from the literature were selected. We denote the two sets of varia-

[1]This procedure was varied slightly in some cases.

bles on which the canonical analysis will be performed as the right and left set. More specifically, the data sources are:

(1) 7 x 7 [five vars-right set, two vars-left set; Press (1972)]
(2) 10 x 10 [five vars-right set, five vars-left set; Huberty (1969)]
(3) 12 x 12 [six vars-right set, six vars-left set; Wechsler (7.5 yrs, 1947)]
(4) 12 x 12 [six vars-right set, six vars-left set; Wechsler (10.5 yrs, 1947)]
(5) 12 x 12 [six vars-right set, six vars-left set; Wechsler (13.5 yrs, 1947)]
(6) 27 x 27 [17 vars-right set, 10 vars-left set, Cooley (1965)]
(7) 31 x 31 [21 vars-right set, 10 vars-left set, Thorndike and Weiss (1973)]
(8) 41 x 41 [21 vars-right set, 20 vars-left set, Thorndike and Weiss (1973)]

Each of these matrices was treated as a *population* correlation matrix. Using the factor analytic procedure previously described, data sample correlation matrices were generated from each population matrix.

## Results

Five of the eight examples will be discussed individually, and then some general statements will be made. We examine first the Wechsler matrices for 7.5 and 10.5 years. For the 7.5 yrs matrix, the coefficients are quite superior for the right set and somewhat superior for the left set. The latter statement is in reference to the range from 100 to 600. On the other hand, for 10.5 yrs (where the within set correlational matrices are somewhat tighter than for 7.5 yrs) the components evidence clear superiority for the right sets for the two largest canonical correlations (Table 1). However, for the left set the superiority of the components over the coefficients vanishes at $N = 600$ for the largest canonical correlation. For the second largest canonical correlation the stability of the components and the coefficients is essentially the same.

We turn next to the Cooley data (27 x 27 matrix). Table 2 shows that the components are definitely superior for both the right and left sets for the largest canonical correlation. However, for the second largest canonical correlation the coefficients are

Table 1
Kendall's Coefficient of Concordance for 12 x 12 Matrices: 7.5 yrs—Largest Canonical Correlation, 10.5 yrs—Two Largest Canonical Correlations

| First Pair of Canonical Variates (7.5 yrs) | | (.6801)[a] 100 | Sample Size 200 | 400 | 600 | 1000 | 2000 | 3000 |
|---|---|---|---|---|---|---|---|---|
| R | Coefficients | 39 | 60 | 77 | 80 | 86 | 89 | 90 |
| | Components | 20 | 32 | 49 | 61 | 73 | 84 | 87 |
| L | Coefficients | 55 | 81 | 89 | 94 | 96 | 98 | 99 |
| | Components | 54 | 72 | 81 | 90 | 94 | 97 | 97 |
| **First Pair of Canonical Variates (10.5 yrs)** | | (.7329) 100 | Sample Size 200 | 400 | 600 | 1000 | 2200 | 3000 |
| R | Coefficients | 42 | 57 | 72 | 74 | 81 | 84 | 85 |
| | Components | 71 | 85 | 92 | 95 | 98 | 98 | 99 |
| L | Coefficients | 32 | 59 | 67 | 79 | 87 | 94 | 94 |
| | Components | 50 | 67 | 72 | 75 | 78 | 86 | 87 |
| **Second Pair of Canonical Variates (10.5 yrs)** | | (.3484) 100 | 200 | 400 | 600 | 1000 | 2200 | 3000 |
| R | Coefficients | 10 | 35 | 44 | 50 | 60 | 74 | 80 |
| | Components | 23 | 43 | 56 | 59 | 74 | 86 | 92 |
| L | Coefficients | 26 | 44 | 63 | 77 | 84 | 90 | 92 |
| | Components | 26 | 46 | 65 | 78 | 85 | 91 | 94 |

[a]Population value for the first canonical correlation.

definitely superior for the left set, while for the right set there is no real difference.

Tables 3 and 4 again show (for the 31 and 41 variable matrices), as the previous examples have suggested, that neither the components or the coefficients come out consistently as more reliable. The components are superior in three out of the four cases for the largest canonical correlation; however, in three out of the four cases for the second largest canonical correlation, the coefficients are somewhat superior. The other case is somewhat muddled.

Using components (rather than coefficients) can lead to the selection of different variables for interpretive purposes. This situation arose in our examination of the Wechsler and Cooley data. It

Table 2
Kendall's Coefficient of Concordance for 27 x 27 Matrix
and Population Values of Coefficients and Components for Largest Canonical Correlation.

| First Pair of Canonical Variates (.5090) | | Sample Size | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 400 | 600 | 800 | 1000 | 1400 | 1800 | 2400 |
| R | Coefficients | 16 | 23 | 26 | 37 | 43 | 54 | 58 |
| | Components | 46 | 64 | 60 | 65 | 77 | 85 | 83 |
| L | Coefficients | 32 | 40 | 42 | 41 | 50 | 57 | 58 |
| | Components | 45 | 60 | 54 | 59 | 69 | 76 | 72 |
| **Second Pair of Canonical Variates (.4472)** | | 400 | 600 | 800 | 1000 | 1400 | 1800 | 2400 |
| R | Coefficients | 19 | 27 | 30 | 30 | 40 | 51 | 54 |
| | Components | 18 | 29 | 28 | 34 | 49 | 59 | 56 |
| L | Coefficients | 34 | 49 | 51 | 55 | 66 | 73 | 74 |
| | Components | 25 | 35 | 30 | 30 | 43 | 48 | 46 |

Population Values for Coefficients and Components

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| R | Coefficients | −.015 | .150 | .057 | −.104 | .383 | .230 | .191 | .271 | −.069 | .246 | −.007 | .079 | −.113 | −.097 | −.313 | .015 | −.235 |
| | Components | .383 | .550 | .593 | .679 | .658 | .516 | .539 | .435 | −.011 | .528 | .338 | .319 | .111 | −.303 | −.367 | −.162 | .365 |
| L | Coefficients | .068 | .149 | −.099 | .096 | −.116 | .043 | .573 | .413 | .203 | −.119 | | | | | | | |
| | Components | .578 | .707 | .142 | .572 | .558 | .566 | .868 | .770 | .556 | .614 | | | | | | | |

Table 3
Kendall's Coefficient of Concordance for 31 x 31 Matrix
and Population Values of Coefficients and Components for Largest Canonical Correlation.

| First Pair of Canonical Variates | (.4968) | Sample Size | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 400 | 600 | 800 | 1000 | 1400 | 1800 | 2400 | 3000 |
| R | Coefficients | 14 | 22 | 24 | 33 | 42 | 47 | 55 | 63 |
| | Components | 50 | 67 | 72 | 82 | 83 | 89 | 90 | 93 |
| L | Coefficients | 26 | 35 | 36 | 47 | 48 | 62 | 64 | 71 |
| | Components | 31 | 50 | 53 | 61 | 70 | 76 | 78 | 84 |
| **Second Pair of Canonical Variates** | (.4268) | | | | | | | | |
| | | 400 | 600 | 800 | 1000 | 1400 | 1800 | 2400 | 3000 |
| R | Coefficients | 14 | 22 | 29 | 39 | 48 | 55 | 64 | 70 |
| | Components | 08 | 15 | 21 | 32 | 33 | 47 | 52 | 60 |
| L | Coefficients | 13 | 26 | 33 | 44 | 52 | 62 | 71 | 76 |
| | Components | 12 | 21 | 36 | 57 | 57 | 69 | 78 | 84 |

Population Values for Coefficients and Components

| | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| R | Coefficients | 119 | –282 | 267 | –758 | 114 | –046 | 295 | –017 | 015 | 178 | 151 | –393 | 117 | 127 | 328 | –013 | –328 | 302 | 088 | 044 | –335 |
| | Components | –544 | –561 | –486 | –666 | –403 | –616 | –327 | –019 | –415 | –069 | 304 | –396 | 100 | 577 | 736 | 304 | 555 | 718 | 356 | 395 | –169 |
| L | Coefficients | 482 | 098 | –331 | –151 | –840 | 359 | –341 | –429 | 089 | 212 | | | | | | | | | | | |
| | Components | 033 | 230 | 185 | 419 | 905 | 266 | 388 | 425 | 210 | 217 | | | | | | | | | | | |

Table 4
Kendall's Coefficient of Concordance for 41 x 41 Matrix
for the Three Largest Canonical Correlations.

| | | Sample Size | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **First Pair of Canonical Variates** (.6166) | | 400 | 600 | 800 | 1000 | 1500 | 2000 | 2500 | 3000 |
| R | Coefficients | 32 | 44 | 57 | 66 | 78 | 80 | 83 | 85 |
| | Components | 27 | 41 | 56 | 64 | 71 | 78 | 84 | 84 |
| L | Coefficients | 31 | 41 | 50 | 57 | 65 | 67 | 72 | 75 |
| | Components | 48 | 63 | 75 | 79 | 87 | 88 | 91 | 92 |
| **Second Pair of Canonical Variates** (.5191) | | 400 | 600 | 800 | 1000 | 1500 | 2000 | 2500 | 3000 |
| R | Coefficients | 15 | 19 | 29 | 31 | 42 | 52 | 58 | 66 |
| | Components | 08 | 12 | 16 | 15 | 31 | 36 | 47 | 54 |
| L | Coefficients | 11 | 22 | 32 | 34 | 48 | 57 | 62 | 71 |
| | Components | 07 | 14 | 22 | 22 | 36 | 42 | 48 | 57 |
| **Third Pair of Canonical Variates** (.4775) | | 400 | 600 | 800 | 1000 | 1500 | 2000 | 2500 | 3000 |
| R | Coefficients | 14 | 19 | 20 | 20 | 28 | 35 | 41 | 46 |
| | Components | 06 | 07 | 10 | 17 | 23 | 27 | 35 | 37 |
| L | Coefficients | 05 | 07 | 09 | 13 | 17 | 22 | 29 | 35 |
| | Components | 05 | 08 | 11 | 12 | 15 | 23 | 29 | 34 |

also occurs again for the 31 x 31 matrix. An examination of the population values for the coefficients and components for the right set (Table 3) shows that using coefficients would probadly lead to variables 4, 12, 15, 17 and 21 being selected. The components, however, would probably select variables 4, 6, 14, 15 and 18. These two sets have only two variables in common, i.e., variables 4 and 15. Furthermore, the variable that would be ranked first by the coefficients (variable 4) is different from the variable that would be ranked first by the components (variable 15).

Table 5 gives the frequency rank tables for the 31 x 31 matrix for the largest canonical correlation. Notice that for the case where $W > .50$, the most important variables are identified at least 70% of the time. This would be true if we were using three variables for interpretation purposes.

Table 5
Frequency Rank Tables for 31 x 31 Matrix
for Left Components for $N = 400$ and 800
(Largest Canonical Correlation)

$N = 400$, $W = .31$

| Variable | Rank 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | Frequency Ranked 1-4 | Population Rank |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 24 | 25 | 13 | 9 | 5 | 4 | 5 | 6 | 7 | 2 | — | – |
| 2 | 8 | 9 | 14 | 12 | 11 | 12 | 5 | 13 | 11 | 5 | — | – |
| 3 | 11 | 13 | 14 | 23 | 14 | 9 | 7 | 5 | 4 | 0 | — | – |
| 4 | 2 | 5 | 5 | 7 | 10 | 13 | 14 | 15 | 28 | 1 | 58 | 3 |
| 5 | 1 | 1 | 2 | 0 | 0 | 3 | 2 | 2 | 3 | 86 | 93 | 1 |
| 6 | 10 | 11 | 15 | 10 | 19 | 16 | 6 | 11 | 2 | 0 | — | – |
| 7 | 5 | 5 | 7 | 5 | 19 | 14 | 20 | 17 | 8 | 0 | 45 | 4 |
| 8 | 6 | 4 | 3 | 8 | 8 | 14 | 18 | 17 | 21 | 1 | 57 | 2 |
| 9 | 17 | 12 | 14 | 12 | 3 | 12 | 14 | 6 | 9 | 1 | — | – |
| 10 | 16 | 15 | 13 | 14 | 11 | 3 | 9 | 8 | 7 | 4 | — | – |

$N = 800$, $W = .53$

| Variable | Rank 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | Frequency Ranked 1-4 | Population Rank |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 36 | 18 | 11 | 17 | 5 | 4 | 1 | 4 | 4 | 0 | — | – |
| 2 | 11 | 9 | 20 | 12 | 12 | 19 | 2 | 3 | 11 | 1 | — | – |
| 3 | 18 | 29 | 21 | 13 | 7 | 4 | 3 | 4 | 1 | 0 | — | – |
| 4 | 0 | 2 | 3 | 1 | 6 | 11 | 19 | 27 | 31 | 0 | 77 | 3 |
| 5 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 97 | 98 | 1 |
| 6 | 3 | 7 | 12 | 21 | 29 | 11 | 7 | 5 | 5 | 0 | — | – |
| 7 | 2 | 0 | 2 | 6 | 12 | 13 | 31 | 22 | 12 | 0 | 65 | 4 |
| 8 | 1 | 1 | 3 | 3 | 5 | 12 | 18 | 29 | 28 | 0 | 75 | 2 |
| 9 | 18 | 17 | 13 | 9 | 11 | 11 | 9 | 3 | 8 | 1 | — | – |
| 10 | 9 | 17 | 15 | 18 | 13 | 15 | 10 | 2 | 0 | 1 | — | – |

We have been selective in terms of the tables that have been presented. The population correlation matrices for the first six examples, along with further tables for Kendall's $W$, and many frequency rank tables have been deposited with National Auxiliary Publications Service[2].

## Overview for the Eight Examples

From the various frequency rank tables available to us it was clear that a value of Kendall's $W > .50$ will generally do a good job of detecting those variables which are most important (from the examples, at least 70% of the time).

2. See NAPS document No. 02614 for 12 pages of supplementary material. Order from ASIS/NAPS c/o Microfiche Publications, 440 Park Avenue South, New York, NY 10016. Remit in advance for each NAPS accession number. Make checks payable to Microfiche Publications. Photocopies are \$5.00. Microfiche are \$1.50. Outside of the United States and Canada, postage is \$2.00 for a photocopy or \$ .50 for a fiche.

For the eight examples the number of subjects per variable necessary to obtain $W > .50$ for the right and left coefficients and components for the *two largest* canonical correlations was determined. The results were:

(1) 7 x 7, $N$ = 300, 43/1 ratio
(2) 10 x 10, $N$ = 500, 50/1 ratio
(3) 12 x 12 (7.5 yrs), $N$ = 500, 42/1 ratio
(4) 12 x 12 (10.5 yrs), $N$ = 600, 50/1 ratio
(5) 12 x 12 (13.5 yrs), $N$ = 800, 65/1 ratio
(6) 27 x 27, $N$ = 1800, 67/1 ratio
(7) 31 x 31, $N$ = 2000, 65/1 ratio
(8) 41 x 41, $N$ = 2800, 68/1 ratio

There are only two cases where $W < .50$. For (3), one of the $W$'s is .41, while for (6), one of the $W$'s is .48. A final note on these subject/variable ratios. The *minimum* value was set at .50. In most cases many of the $W$'s are substantially larger than .50.

A check of the eight examples was made to determine whether the components were more reliable than the coefficients, as measured by Kendall's $W$. For interpeting the largest two canonical correlations in six cases [7 x 7 and 12 x 12 (7.5 yrs) omitted], and the largest three canonical correlations in two cases (10 x 10 and 41 x 41), about 60% of the time the components were superior, while 40% of the time the coefficients were superior. In five cases there was essentially no difference in the $W$'s for the two approaches. Thus, if one were to hazard a generalization based on these eight examples, there is no basis for concluding that the components are a superior approach, at least not in terms of being more reliable.

## Stability of Canonical Correlations & Cross Validation

For all eight examples the canonical correlations are very stable under replication, i.e., where one is maximizing the correlation between each pair of variates for each replication (sample). Even for small sample sizes, such as 100 or 200, the variances were almost always less than .005. However, stability in this sense by no means implies stability under cross validation, i.e., where weights from one sample are applied to another sample to determine if the relationship found in the first sample will hold up. Now, as mentioned earlier, Thorndike and Weiss (1973, p. 133) concluded that if such a cross validation check held up then, "The canonical variates may be interpreted (via the canonical components) with confidence as having the degree of covariation found in the cross validation group, or the weights may be used for prediction."

There appears to be a problem with this approach. It is possible for such a cross validation check to show the relationship to be stable and yet to be interpreting the canonical variates in somewhat different ways. For example, the components found using sample (1) might show variables 1, 4, 6 and 10 to be the most important for a particular canonical variate; yet if the other sample components had been used, some other variables [including perhaps some of the same variables as found in sample (1)] might be used for interpretation purposes. In other words, depending on *which* sample is used for computing components, a somewhat different interpretation of the canonical variates might arise. The authors have several specific examples of where this has happened, so that it appears that it would not be unusual for it to happen in practice. What is needed is sufficient sample size to ensure that the relationship is stable *and* that the interpretation of the canonical variates is the same. However, what if one cannot obtain the large number of subjects per variable that our study suggests is necessary? The following seems like a reasonable method of proceeding. Do the cross validation to see if the relationship holds up. Then compute the components for *each* sample, and just use for interpretation those variables whose average components are highest.

It seems reasonable to assume that if the components are stable in the sense we have indicated ($W > .50$), then cross validation of the canonical correlations would hold up. However, the eight examples show that 42 to 68 subjects per variable are necessary to obtain this kind of stability. Of the two examples used by Thorndike and Weiss, one had about 13 subjects per variable and the other only six subjects per variable.



## Conclusions

Based on the eight examples, there is no solid evidence for concluding that the components are superior to the coefficients, at least not in terms of being reliable. For the examples considered the components tended to be more reliable more often for the largest canonical correlation, especially when the correlations among the variables within each set are fairly high. The advantage of the components, however, seemed to decrease for the second and third largest canonical correlations. The number of subjects per variable necessary to achieve reliability in determining the most important variables, using components or coefficients, was quite large, ranging from 42/1 to 68/1.

The canonical correlations are very stable upon replication. With respect to interpreting the canonical variates for small or medium-sized samples, one possible approach is to compute the components (coefficients) for each sample and for interpretation just use those variables whose average values are the largest.

## References

BMDX72, Factor analyses. In W. J. Dixon (Ed.), *Biomedical Computer Programs*. Berkeley: University of California Press, 1970, 90-103.

Cooley, W. W. Canonical correlation. Paper presented at the APA-Psychometric Society symposium on the applications of multivariate analysis, September, 1965.

Darlington, R. B., Weinberg, S. L. and Walberg, H. J. Canonical variate analysis and related techniques. *Review of Educational Research,* 1973, *43*, 433-454.

Harman, H. H. *Modern factor analysis*. Chicago: University of Chicago Press, 1967.

Huberty, C. J. An empirical comparison of selected classification rules in multiple discriminant analysis. Unpublished doctoral dissertation, University of Iowa, 1969.

Kaiser, K. F. and Dickman, K. Sample and population score matrices and sample correlation matrices from an arbitrary population correlation matrix. *Psychometrika*, 1962, *27*, 179-182.

Meredith, W. Canonical correlations with fallible data. *Psychometrika,* 1964, *29*, 55-65.

Press, S. J. *Applied multivariate analysis*. Holt, Rinehart and Winston: New York, 1972.

Thorndike, R. M. and Weiss, D. J. A study of the stability of canonical correlations and canonical components. *Educational and Psychological Measurement*, 1973, *33*, 123-134.

Wechsler, D. *Wechsler intelligence scale for children*. New York: The Psychological Corporation, 1947.