DOCUMENT RESUME

ED 120 234	TM 005 193

AUTHOR	Williams, John D.
TITLE	Canonical Analysis as a Generalized Regression Technique for Multivariate Analysis.
PUB DATE	Apr 76
NOTE	20p.: Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976); Some pages may reproduce poorly due to light print of the original
EDRS PRICE	MF-$0.83 HC-$1.67 Plus Postage
DESCRIPTORS	*Analysis of Variance; Hypothesis Testing; Matrices; *Multiple Regression Analysis; Predictor Variables; Research Design; *Statistical Analysis; *Tests of Significance
IDENTIFIERS	*Canonical Analysis; Multiple Linear Regression

ABSTRACT

The use of characteristic coding (dummy coding) is made in showing solutions to four multivariate problems using canonical analysis. The canonical variates can be themselves analyzed by the use of multiple linear regression. When the canonical variates are used as criteria in a multiple linear regression, the R2 values are equal to C, where 0 is the squared canonical correlation coefficient. Several different methods exist for testing multivariate hypotheses. Where the interest is in a two-way disproportionate multivariate analysis of variance, the trace criterion seems particularly applicable. Characteristic (dummy) coding has been used in multiple linear regression to analyze univariate analysis of variance problems; the same coding scheme can be extended to multiple criteria. While the resulting data are analyzed through canonical analysis, the design matrix conforms to the usual multiple linear regression design matrices. Thus, the utilization of multiple criteria can be pursued in a logical sequence without necessitating continuosly changing the entire terminology. (Author/RC)

Canonical Analysis as a Generalized Regression Technique for Multivariate Analysis

John D. Williams
The University of North Dakota

The use of characteristic coding (dummy coding) is made in showing solutions to four multivariate problems using canonical analysis. The canonical variates can be analyzed by the use of multiple linear regression. When the canonical variates are used as criteria in a multiple linear regression, the R^2 values are equal to θ, where θ is the squared canonical correlation coefficient. Several different methods exist for testing multivariate hypotheses. Where the interest is in a two-way disproportionate multivariate analysis of variance, the trace criterion ($\Sigma\theta_i$) seems particularly applicable.

Characteristic (dummy) coding has been used in multiple linear regression to analyze univariate analysis of variance problems; the same coding scheme can be extended to multiple criteria. While the resulting data are analyzed through canonical analysis, the design matrix conforms to the usual multiple linear regression design matrices. Thus, the utilization of multiple criteria can be pursued in a logical sequence without necessitating continuously changing the entire terminology.

In the present paper, four multivariate research designs are examined in a canonical analysis framework: a multivariate two-group situation, sometimes referred to as Hotelling's T^2 test; a multivariate multiple group situation; a multivariate two-way analysis; and a multivariate two-way analysis with disproportionate cell frequencies.

Tests of Significance in Canonical Analysis

In multivariate analysis, several different tests of significance are used. Typically, the multivariate analysis of variance has focused on solving the following equation for λ_i:

$$\left| S_{12}S_{22}^{-1}S_{21} - \lambda(S_{11} - S_{22}^{-1}S_{21}) \right| = 0,$$

whereas canonical analysis has focused on solving

$$\left| S_{12}S^{-1}_{22}S_{21} - \theta S_{11} \right| = 0$$

where $\theta = R^2_{c_i}$ and $\theta_i = \frac{\lambda_i}{1 + \lambda_i}$; also,

S_{11}, S_{12}, S_{21} and S_{22} are variance - covariance matrices.

Roy's (1957) largest root criterion tests the significance of the largest characteristic root. Hotelling's (1951) trace criterion tests the overall multivariate hypotheses for all dimensions simultaneously and is given by trace = $\Sigma\lambda_i$. Tables for testing either of these two tests have been given by Pillai (1960). A trace criterion using Pillai's tables wherein the trace is $\Sigma\theta_i$ is useful in testing the significance of the overall set of canonical correlations and is analogous to Hotelling's trace criterion.

Wilks Λ also provides a test of the overall hypothesis. All necessary tables for testing these hypotheses can be found in Timm (1975). Typically, in canonical analysis simultaneous tests of each characteristic root using Roy's approach is used. On the other hand, multivariate analysis of variance programs usually employ an overall test (either the trace criterion or Λ). Harris (1975) has argued that the use of the largest root criterion is more sensible in that if either the trace criterion or Λ shows significance but the largest root criterion does not, then the differences among the groups cannot be pinpointed by any single linear combination of variables. Harris would see the use of the overall hypotheses as being more useful in only those cases where λ_1 and λ_2 are close to the same value.

A Multivariate Two'Group Situation (Hotelling's T^2 Test)

The simplest multivariate analysis of variance situation is the multivariate analog to the usual t test; here, several criteria are observed for two groups and an overall test for group d'fferences can be made. As an example, suppose four criteria (Y_1, Y_2, Y_3 and Y_4) are observed for two groups as indicated in Table 1.

Table 1

Four Criteria Y_1, Y_2, Y_3, and Y_4

for the Multivariate Two-Group Situation

Group 1				Group 2			
Y_1	Y_2	Y_3	Y_4	Y_1	Y_2	Y_3	Y_4
20	17	17	25	15	26	13	16
22	19	16	28	19	25	15	15
24	14	18	23	23	21	17	17
26	16	17	17	24	17	22	18
28	18	16	29	25	19	16	22
30	20	15	32	26	14	8	23
32	22	14	35	27	22	14	24
34	16	16	42	28	20	18	20
36	9	18	38	30	17	21	18

To accomplish a canonical analysis with the data in Table 1, it is necessary to define a first-set and a second-set. For convenience, the criteria will constitute the first set and the predictors (group membership variables) will constitute the second set. Actually, only a single group membership variable is necessary:

$X_1 = 1$ if a member of Group One; 0 otherwise.

Table 2 contains the criteria and design matrix necessary to accomplish this analysis.

Table 2

Criteria and Design Matrix for a Two-Group Multivariate Analysis

Y_1	Y_2	Y_3	Y_4	X_1
20	17	17	25	1
22	19	16	28	1
24	14	18	23	1
26	16	17	17	1
28	18	16	29	1
30	20	15	32	1
32	22	14	35	1
34	16	16	42	1
36	9	18	38	1
15	26	13	16	0
19	25	15	15	0
23	21	17	17	0
24	17	22	18	0
25	19	16	22	0
26	14	8	23	0
27	22	14	24	0
28	20	18	20	0
30	17	21	18	0

Using canonical analysis to find the relationship between the left set and the right set, the following results are obtained:

θ = .60066; canonical $R_c = \sqrt{\theta}$ = .77502;

Wilk's Λ = .39934, with $p < .01$. Also, the trace = .60066; in every case, $p < .01$.

The coefficients for the first set are

	First root
Y_1	-.65403
Y_2	-.32522
Y_3	.24849
Y_4	1.26372.

The coefficient for the X_1 variable is of course 1.000.

Also,

$\bar{Y}_1$ = 26.0556, s_{Y_1} = 5.3189;

$\bar{Y}_2$ = 18.4444, s_{Y_2} = 4.0905;

$\overline{Y}_3 = 16.1667$, $s_{Y_3} = 3.0534$; and

$\overline{Y}_4 = 24.5556$, $s_{Y_4} = 7.9648$.

If a new variable, Y_5, is formed as

$$Y_5 = -.65403(Y_1 - \overline{Y}_1)/s_{Y_1} - .32522(Y_2 - \overline{Y}_2)/s_{Y_2} + .24849(Y_3 - \overline{Y}_3)/s_{Y_3}$$
$$+ 1.26372(Y_4 - \overline{Y}_4)/s_{Y_4},$$

then a regression equation can be formed as

$$Y_5 = b_0 + b_1X_1 + e_1. \qquad (1)$$

If a regression is completed with the formulation in equation 1, then

$R = .77502$,

$R^2 = .60066$, and

$1\text{-}R^2 = .39934$.

This information is identical to that found in the use of the canonical analysis; the relationship is, for the two-group situation:

$R = \sqrt{\theta} = R_c$;

$R^2 = \theta$ and $\Lambda = 1\text{-}R^2$.

The use of equation 1 shows that a composite variable, Y_5, is a linear composite of variables Y_1, Y_2, Y_3 and Y_4 such that the relationship with X_1 remains maximized.

Because there is only one group membership variable involved, an interesting reversal of the roles of the criteria and predictor can be made:

$$X_1 = b_0 + b_1Y_1 + b_2Y_2 + b_3Y_3 + b_4Y_4 + e_2. \qquad (2)$$

If equation 2 is utilized in a multiple regression framework,

$R = .77502$, $R^2 = .60066$, $1\text{-}R^2 = .39934$ and $F = 4.888$.

The first three results were obtained in the prior two analyses.

There are some differences between the two analyses, however. For the use of equation 2, $b_1 - b_4$ are different from the coefficients for the first set given previously; this, of course, was to be accepted. The beta coefficients also differ from the coefficients given earlier;

$\beta_1 = -.50688$,

$\beta_2 = -.25206$,

$\beta_3 = .19259$,

$\beta_4 = .97940$.

If some thought is given to it, this difference comes as no surprise either. In a canonical analysis, each canonical variate has a mean of zero and standard deviation of one. In a regression analysis, the beta coefficients are such that for every predictor variable, there is a mean of zero and standard deviation of one. The difference is that in canonical analysis, the new variate is created with mean zero and standard deviation of one.

Finally if a multivariate analysis of variance program is executed, $\Lambda = .39934$ and $F = 4.888$, results that were obtained earlier. Thus, if the interest is in comparing two groups on several criteria simultaneously, several different strategies allow equivalent solutions. In this special case, the execution of equation 2 (using the group membership variable as the criterion and the Y variables as predictors) is perhaps the easiest solution to employ. The use of canonical analysis and subsequent formation of a composite variable would also seem to be of some value.

A Multivariate Multiple Group Situation

If several groups are involved in the analysis with multiple criteria, then the usual one-way multivariate analysis of variance is often employed. As an example of such a situation, suppose three criteria are available for three groups of subjects. Such a situation is encountered in Table 3.

Table 3

Criteria and Design Matrix for Multivariate Analysis of Variance Through Regression

Y_1	Y_2	Y_3	X_1	X_2
17	23	1	1	0
22	28	2	1	0
14	22	3	1	0
18	27	4	1	0
29	25	5	1	0
22	32	6	0	1
24	34	8	0	1
26	36	10	0	1
28	42	12	0	1
25	31	14	0	1
26	23	15	0	0
29	32	16	0	0
32	29	17	0	0
35	42	18	0	0
33	23	19	0	0

Two group membership variables are used:

X_1 = 1 if a member of Group 1; 0 otherwise, and

X_2 = 1 if a member of Group 2; 0 otherwise.

Using the data in Table 3, a canonical analysis is performed with the Y scores (criteria) as the first set and the X variables (predictors) as the second set. Several useful items are typically available from a canonical analysis. Either the canonical roots or canonical correlations (or both) will be available. For the data in Table 3, $\theta_1 = .89286$, and $\theta_2 = .43602$. The first canonical correlation is $\sqrt{\theta_1} = .94491$, and the second canonical correlation is $\sqrt{\theta_2}$ = .66032. The weights for the Y side and X side are:

Y side weights	1	2	3
1	.09556	-.01807	-1.07322
2	-.79902	1.10739	.27948

X side weights	1	2
1	1.15457	.56235
2	.01725	1.00851

It is interesting to form variables to correspond to those suggested by the Y side and X side weights and investigate these transformed variables using ordinary multiple regression. To utilize the weights, it is first necessary to transform all the data in Table 1 into z scores. As $\overline{X}_1$ = .3333, $\overline{X}_2$ = .3333, $\overline{Y}_1$ = 25.3333, $\overline{Y}_2$ = 29.9333, $\overline{Y}_3$ = 10.0, s_{x_1} = .4880, s_{x_2} = .4880, s_{y_1} = 6.0198, s_{y_2} = 6.5407 and s_{y_3} = 6.2678, the transformation equations are:

$$Z_1 = \left[.09556(Y_1-25.3333)/6.0198\right] + \left[-.01807(Y_2-29.9333)/6.5407\right] + \left[-1.07322(Y_3-10.)/6.2678\right]$$

$$Z_2 = \left[-.79902(Y_1-25.3333)/6.0198\right] + \left[1.10739(Y_2-29.9333)/6.5407\right] + \left[.27938(Y_3-10.)/6.2678\right]$$

$$Z_3 = \left[1.15457(X_1-.3333)/.4880\right] + \left[.56235(X_2-.3333)/.4880\right]$$

$$Z_4 = \left[.01725(X_1-.3333)/.4880\right] + \left[1.00851(X_2-.3333)/.4880\right]$$

Using Z_1 as the criterion and X_1 and X_2 as predictors, R^2 = .89286, R = .94491, identically the same results as found for the first canonical root. Similarly, using Z_2 as the criterion and X_1 and X_2 as predictors, R^2 = .43602 and R = .66032. If Y_1, Y_2, and Y_3 are used as predictors of Z_3 and then Z_4, again the canonical correlations appear as multiple correlations.

Also the following correlations are of interest:

$r_{Z_1Z_2} = 0$ $\qquad$ $r_{Z_2Z_3} = 0$

$r_{Z_1Z_3} = .94491$ $\qquad$ $r_{Z_2Z_4} = .66032$

$r_{Z_1Z_4} = 0$ $\qquad$ $r_{Z_3Z_4} = 0$

If a traditional multivariate analysis of variance is performed, the test for H_2 (overall difference among all groups) yields $\Lambda = .06042$. While some canonical printouts (such as Cooley and Lohnes, 1971) include this value as part of the output, Λ can be found as $\Pi_i^k\ (1-\theta_i)$ where the θ_i are the canonical roots. For this particular data, $\Lambda = (1-.89286)(1-.43602) = .06042$. Because $\theta_i = R_i^2$, this result can be written as $\Lambda = \Pi_i^k(1-R_i^2)$. The trace criterion yields Tr = .89286 + .43602 = 1.2888.

Also, from the multivariate analysis printout, an F = 10.227 ($p < .01$) tests the overall group differences among the three groups.

The use of the canonical variates found through the use of the canonical vectors should present an attractive alternative to those researchers who wish to complete multiple comparisons after the rejection of the overall null hypothesis. One suggestion, made by Hummel and Sligo (1971) is to compare the groups on univariate tests on each variable after the rejection of the overall null hypothesis. An alternative is to use the first canonical variate for the criteria set and run an analysis among the groups with this (and subsequent) canonical variates from the criteria set. Scheffé's test would seem appropriate as a multiple comparison method.

A Two-Way Multivariate Situation

The two-way multivariate analysis of variance is not quite as available as the one-way multivariate analysis of variance, but several programs are available, including Cramer (1974), Finn (1974), Ondrack (1974), and Cooley and Lohnes (1971). Suppose three criteria are measured in a 2X3 design. The design matrix is given in Table 4.

Table 4

Criteria and Design Matrix For Two-Way
Analysis of Variance Through Regression

Y_1	Y_2	Y_3	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
13	18	16	1	0	0	0	0	1	1	0
14	19	17	1	0	0	0	0	1	1	0
15	18	22	1	0	0	0	0	1	1	0
16	15	27	1	0	0	0	0	1	1	0
14	19	20	1	0	0	0	0	1	1	0
17	18	23	1	0	0	0	0	1	1	0
15	9	13	0	1	0	0	0	1	0	1
17	13	21	0	1	0	0	0	1	0	1
16	16	12	0	1	0	0	0	1	0	1
17	18	22	0	1	0	0	0	1	0	1
14	13	17	0	1	0	0	0	1	0	1
15	11	18	0	1	0	0	0	1	0	1
16	17	27	0	0	1	0	0	1	0	0
16	14	21	0	0	1	0	0	1	0	0
17	16	23	0	0	1	0	0	1	0	0
18	15	24	0	0	1	0	0	1	0	0
20	13	22	0	0	1	0	0	1	0	0
22	9	18	0	0	1	0	0	1	0	0
17	15	16	0	0	0	1	0	0	1	0
18	17	15	0	0	0	1	0	0	1	0
19	19	14	0	0	0	1	0	0	1	0
20	21	14	0	0	0	1	0	0	1	0
21	23	18	0	0	0	1	0	0	1	0
22	25	19	0	0	0	1	0	0	1	0
13	17	22	0	0	0	0	1	0	0	1
15	13	22	0	0	0	0	1	0	0	1
13	16	21	0	0	0	0	1	0	0	1
12	12	17	0	0	0	0	1	0	0	1
11	11	19	0	0	0	0	1	0	0	1
11	17	18	0	0	0	0	1	0	0	1
17	15	9	0	0	0	0	0	0	0	0
18	17	8	0	0	0	0	0	0	0	0
20	15	7	0	0	0	0	0	0	0	0
19	17	6	0	0	0	0	0	0	0	0
17	19	5	0	0	0	0	0	0	0	0
16	22	4	0	0	0	0	0	0	0	0

where

X_1 = 1 for a member of Row 1 and Column 1, 0 otherwise;

X_2 = 1 for a member of Row 1 and Column 2, 0 otherwise;

X_3 = 1 for a member of Row 1 and Column 3, 0 otherwise;

X_4 = 1 for a member of Row 2 and Column 1, 0 otherwise;

X_5 = 1 for a member of Row 2 and Column 2, 0 otherwise;

X_6 = 1 for a member of Row 1, 0 otherwise;

X_7 = 1 for a member of Column 1, 0 otherwise and

X_8 = 1 for a member of Column 2, 0 otherwise.

An analysis of the data in Table 5 is necessarily complex. Four sets of canonical relationships are possible; the Y_1, Y_2 and Y_3 variables can be related to X_1 - X_5, then X_6 and X_7, then X_8 and finally X_6, X_7 and X_8. In the univariate situation these relationships correspond to the full model, the rows, the columns and the rows and columns as predictors. Table 5 contains the various canonical relationships from these four different sets of predictors.

Table 5

Eigenvalues, Canonical Correlations and Wilks Lambda For Two-Way Multivariate Analysis of Variance by Regression

Preditors: X_1, X_2, X_3, X_4, X_5 (Full Model)			
	$R^2_{c_i}$	Canonical R	Λ
First Root	.86535	.93024	.03594 **
Second Root	.58716	.76626	.26692 **
Third Root	.35345	.59452	.64655 **
Predictors: X_6 (Rows)			
First Root	.32400	.56921	.65600 **
Predictors: X_7, X_8 (Columns)			
First Root	.45924	.67767	.40033 **
Second Root	.25968	.50959	.74032
Predictors: X_6, X_7, X_8 (Rows and Columns)			
First Root	.54326	.73706	.25558 **
Second Root	.30626	.55341	.55957 **
Third Root	.19340	.43977	.80660 **

** significant at .01 level

Corresponding to each canonical root are the weights to create the canonical variables; they can be found (for the Y side only) in Table 6.

Table 6

Canonical Weights for Canonical Variates from Table 5

	Y-Weights		
Predictors: X_1 - X_5			
Variate	1	2	3
Z_1	-.44831	-.20691	.73004
Z_2	-.85876	-.05983	-.68227
Z_3	.40964	-1.02427	-.23006
Predictor: X_6			
Variate			
Z_4	.16443	-.45391	.85726
Predictors: X_7, X_8			
Z_5	.62240	.65303	.10464
Z_6	-.62870	.69649	.59305
Predictors: X_6, X_7, X_8			
Z_7	-.46410	-.62886	.32592
Z_8	.00110	-.58279	-.91247
Z_9	.94386	-.60035	.33548

If Z_1 - Z_9 are used as criteria in a multiple linear regression layout, then when X_1 - X_5 are used as predictors of Z_1, $R^2 = .86535 = \theta_1$; similar findings will occur with the Z variates that correspond to the X predictors used in the original canonical analysis. If the trace (sum of the squared canonical correlations) of the four models are found

$$\text{Tr (Full)} = .86535 + .58716 + .35345 = 1.80596$$

$$\text{Tr (Rows)} = .32400$$

$$\text{Tr (Columns)} = .44924 + .25968 = .71892$$

$$\text{Tr (Rows and Columns)} = .54326 + .30626 + .19340 = 1.04292$$

also,

$$\text{Tr (Rows)} + \text{Tr (Columns)} = \text{Tr (Rows \& Columns)}\ .$$

The interaction can be defined as

$$\text{Tr (Full)} - \text{Tr (Rows \& Columns)} = 1.80596 - 1.04292 = .76304.$$

The sum of the squared canonical correlations can be broken down into the separate R_c^2 values through the use of orthogonal coefficients, as there are an equal number of entries in each of the six cells. If five new variables are defined as follows

X_9 = 1 if a member of Row 1, -1 if a member of Row 2;

X_{10} = 1 if a member of Column 1, 0 if a member of Column 2, -1 if a member of Column 3;

X_{11} = 1 if a member of either Column 1 or Column 3, -2 if a member of Column 2;

$X_{12} = X_9 \cdot X_{10}$ and

$X_{13} = X_9 \cdot X_{11}$.

Using X_9 as the predictor of the three criteria, $R_c^2 = .32400$, the same result as found in Table 6 for rows. When X_{10} and X_{11} are used as predictors, $R^2_{c_1} = .45924$ and $R^2_{c_2} = .25968$, the same result as found in Table 5 for columns. If X_9, X_{10}, X_{11}, X_{12} and X_{13} are used as predictors, $R^2_{c_1} = .86535$, $R^2_{c_2} = .58716$ and $R^2_{c_3} = .35345$, the same results as found for the full model. If X_9, X_{10} and X_{11} were used as predictors, the results would duplicate those found by usinq X_6, X_7 and X_8 as predictors. If X_{12} and X_{13} are used as predictors, the following results are found:

$$R^2_{c_1} = .42305 \quad R_{c_1} = .65042 \quad \Lambda_1 = .38079, \quad p < .01; \quad \text{and}$$

$$R^2_{c_2} = .33999 \quad R_{c_2} = .58309 \quad \Lambda_2 = .66001, \quad p < .01.$$

Finding the interaction directly through the use of orthogonal polynomials appears to be limited to those cases in which the cell entries are either equal or proportional. The last problem to be discussed considers the multivariate disproportional case.

A Two-Way Disproportionate Multivariate Analysis

An analysis similar to the one employed for the two-way equal cell case shown in Table 4 can be considered. In fact, the same 36 "subjects" are reconsidered, after deliberately creating a disproportionate situation. The first 3 subjects are, for the disproportionate case, in cell 1 (Row 1 and Column 1); the next four subjects are in Cell 2 (Row 1 and Column 2); the next 10 scores are in Cell 3 (Row 1 and Column 3; the next 9 subjects are in Cell 4 (Row 2 and Column 1); the next 7 subjects are in Cell 5 (Row 2 and Column 2); finally, the last 3 subjects are in Cell 6 (Row 2 and Column 3). The number of entries in each cell for the 2X3 layout is as given in Table 7.

Table 7

Frequencies for 2X3 Multivariate Analysis with Disproportionate Cells

	Column 1	Column 2	Column 3
Row 1	3	4	10
Row 2	9	7	3

The design matrix is as before with X_1 - X_8 having the same meaning. The results of the canonical analysis are found in Table 8.

Table 8

Eigenvalues, Canonical Correlations and Wilks Lambda for Two-Way Disproportionate Cell Frequencies

Predictors: X_1, X_2, X_3, X_4, X_5 (Full Model)	$R^2_{c_i}$	Canonical R	Λ
First Root	.61960	.78715	.22565**
Second Root	.31954	.56528	.59320*
Third Root	.12824	.35811	.87176
Predictors: X_6 (Rows)			
First Root	.27262	.52213	.72738*
Predictors: X_7, X_8 (Columns)			
First Root	.21675	.46557	:76287
Second Root	.02601	.16127	.97399
Predictors: X_6, X_7, X_8 (Rows and Columns)			
First Root	.34802	.58993	.49760**
Second Root	.21674	.46555	.76322
Third Root	.02559	.15998	.97441

*significant at .05 level
**significant at .01 level

Interpretation of the data in Table 8 may be made, but the lack of consensus on interpreting univariate disproportionate situations will only be increased as the situation becomes multivariate. Many authors prefer the "fitting constants" solution (see Anderson & Bancroft, 1952, Overall and Spiegel, 1969, and Rao, 1965). Cohen (1968) describes a partioning solution called the hierarchical model. An unadjusted main effects solution is shown in Williams (1972). Searle (1971) and Appelbaum and Cramer (1974) prefer a multiple step decision making process that combines the fitting constants solution and the unadjusted main effects solution. The multivariate situation is complicated by the existence of several criteria for judging the significance of an experiment. The approach taken here is to describe both the fitting constants solution and the unadjusted main effects solution; those who prefer the decision rules given in Searle could easily employ them with the information given.

The Unadjusted Main Effects Solution

The unadjusted main effects solution follows in a manner very similar to the one presented in regard to the equal cell case. In fact, the data in Table 8 can be interpreted (except for the interaction) as being an unadjusted main effects solution. The interaction can be found as the difference between the trace of the full model and the trace of the rows and columns model:

$$(.61960 + .31954 + .12824) - (.34802 + .21674 + .02559) = .47703.$$

The only available method to test the interaction hypothesis is Pillai's trace criterion; $p < .05$.

A Multivariate Analog to the Fitting Constants Solution

Because of the disproportionality of the data, the direct calculation of the R_c^2 terms is precluded; the traces can be found in a manner similar to finding the trace for the interaction, however. The trace for rows (after removing the effect for columns) can be found as the trace for rows and columns minus the trace for columns:

$$(.34802 + .21674 + .02559) - (.21675 + .02601) = .34759,\ p <.01.$$

The trace for columns can be found as the trace for rows and columns minus the trace for rows:

$$(.34802 + .21674 + .02559) - (.27262) = .31773,\ p >.05.$$

The interaction is the same as given for the unadjusted main effects model.

Discussion

While Harris's argument for the use of the greatest characteristic root criterion as a measure for multivariate analysis is noteworthy, the trace criterion ($\Sigma\theta_i$) is particularly useful with disproportionate cell frequencies. If the focus of the greatest characteristic root and the corresponding canonical variates are made on the full model (cell model, or full rank model) then either criterion is applicable, and perhaps Harris's suggestion is appropriate. If the intent is on producing a two-way MANOVA with disproportionate cells and there is interest in the row, column and interactions effects, then Pillai's trace criterion is most appropriate. Even where there is interest in the usual effects, the most likely canonical variate to be of interest is the variate associate with the greatest characteristic root from the full model.

Four different multivariate applications have been shown herein. Other applications (multivariate trend analysis, multivariate analysis of covariance and other analogs to univariate designs) are possible through a canonical approach. Also, the univariate analyses that can be performed by multiple linear regression can be conceptualized as a canonical problem. That the canonical analyses and multiple regression analyses would yield quite similar results is not quite the same as saying that the analyses are identical for the univariate situation. As was shown in Hotelling's T^2 test, some differences in weighting coefficients occur. The overall results (R^2's) are identical, however.

References

Anderson, R.L. and Bancroft, T.A. *Statistical theory in research*. New York: McGraw-Hill, 1952.

Appelbaum, M.I. and Cramer, E.M. Some problems in the nonorthogonal analysis of variance. *Psychological Bulletin*, 1974, 81, 335-343.

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

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

Cramer, E.M. *Revised MANOVA program.* Chapel Hill, N.C.: The L.L. Thurstone Psychometric Laboratory, The University of North Carolina, 1974.

Finn, J.D. *A general model for multivariate analysis*. New York: Holt, Rinehart and Winston, 1974.

Harris, R.J. *A primer of multivariate statistics*. New York: Academic Press, 1975.

Hotelling, H. A generalized T-test and measure of multivariate dispersion. *Proceedings of the second Berkeley symposium on mathematics and statistics*, 23-41, 1951.

Hummel, T.J. and Sligo, J.R. Empirical comparison of univariate and multivariate analysis of variance procedures. *Psychological Bulletin*, 1971, 76, 49-57.

Ondrack, M. *Analysis of variance designs: A manual for the computer analysis of selected models*. Toronto: Institute for Behavioral Research, York University, 1974.

Overall, J.E. and Spiegel, D.K. Concerning least squares analysis of experimental data. *Psychological Bulletin*, 1969, 72, 311-322.

Pillai, K.C.S. *Statistical tables for tests of multivariate hypotheses*. Manila: Statistical Center, 1960.

Rao, C.R. *Linear statistical inference and its applications*. New York: Wiley, 1965.

Roy, S.N. Some aspects of multivariate analysis. New York: Wiley, 1957.

Searle, S.R. *Linear models*. New York: Wiley, 1971

Timm, N.H. *Multivariate analysis with applications in education and psychology*. Belmont, Calif: Wadsworth, 1975.

Williams, J.D. Two way fixed effects analysis of variance with disproportionate cell frequencies. *Multivariate Behavioral Research*, 1972, 7, 66-83.