CANONICAL VARIATES AND CORRELATIONS

by

CHI-YUNG CHIANG

B. A. National Taiwan University, 1958

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1965

Approved by

Gary F. Krause
Major Professor

TABLE OF CONTENTS

INTRODUCTION

The search for statistical relationships of one sort or another is a basic objective of scientific work. The method of simple correlation deals with pairs of quantitative variables such as the relationship between heights and the weights of human beings or the relationship between price of one commodity and the volume of its sales. Here only two variables are involved. Sometimes more than two variables will enter the problem. For example, one may be interested in the relation between college students' grade average and (1) their intelligence quotients and (2) the number of hours each studies per week. We may look at the relationship from two different standpoints. First, we may study the relation of a dependent variable and a whole group of independent variables. This is the method of multiple correlation. Thus, a multiple regression equation may be used to estimate or predict the value of the dependent variable on the basis of the several combined independent variables. Secondly, when several different variables are taken into a problem, we may also study the "net" relationship. Since two variables may correlate or fail to correlate largely because of the influence of some other factor upon them rather than their inherent relationship, this influence must be eliminated. The correlation independent of the other factors thus found is called a partial correlation.

Simple correlation represents the relationship between two variables ignoring the influence of other variables. Multiple correlation concerns the relationship between one dependent variable y and a set of independent variables $(x_1, x_2, \ldots, x_k)$. Partial correlation represents the relationship between variables x_i and x_j eliminating the influence of the other variables by holding other variables as constant.

Concepts of correlation and regression may be applied not only to ordinary one-dimensional variates but also to variates of two or more dimensions. The study of individual differences in mental and physical traits calls for a detailed study of the relations between sets of correlated variates. For example the scores on a number of mental tests may be compared with physical measurements on the same persons. The questions then arise of determining the number and nature of the independent relations of mind and body shown to exist by these data and of extracting, from the multiplicity of correlations in the system, suitable characterizations of these independent relations. In economics one may desire to find demand and supply functions. Since the consumptions of one commodity may be related as much to the prices of other commodities, it therefore seems appropriate that studies of demand and supply should be made by groups rather than by single commodities.

To find the relationship between sets of variables, the method of canonical correlations was first introduced into statistics by Hotelling (1936). Considerable additions have been made by various later writers particularly Bartlett (1947) who found the general canonical correlation distribution. Hsu (1941) found its limiting distribution.

The problem concerns relations between two sets of variates that are invariant under all internal linear transformations of each set separately. If one undertakes to find a linear function of the variates in each set which yields the highest possible correlation between the two, a set of linear equations in the coefficients and an interesting determinantal equation are obtained. These, at the same time, determine other linear functions in pairs. One member of a pair in each set will have zero correlations between all functions belonging to different pairs, but within the pairs, correlations are equal to roots of the determinantal equation, called canonical correlations. The linear functions, thus determined, are called the canonical variates and include the best prediction of one set in terms of the other.

Since Hotelling's paper makes use of rather complicated mathematics and does not spell out in detail the methods of numerical computation, few practical statisticans seem to know of the paper. It is the purpose of this report to illustrate the mathematical derivation of canonical variates and canonical correlations, computational procedures, the sampling distribution of canonical correlations and tests of various hypotheses.

DERIVATION AND COMPUTATIONAL PROCEDURE

(a) Derivation

The theory of canonical variates and canonical correlation was developed by Hotelling (1935, 1936). T. W. Anderson (1962), using matrix algebra, gave more modern expression of its derivation in detail.

Suppose the random vector X of p components,

$X' = (x_1\ x_2 \ldots\ x_{p_1}\ x_{p_1+1} \ldots\ x_{p_1+p_2}) \quad p_1+p_2 = p$

has $E(X) = 0$, with variance covariance matrix

$$\Sigma = (\sigma_{ij}),$$

where $\sigma_{ii} = Var(x_i)$ and $\sigma_{ij} = Cov(x_i,x_j)$.

Partition X into two subvectors of p_1 and p_2 (assume $p_1 \leq p_2$) components respectively. The variance covariance matrix is partitioned similarly into p_1 and p_2 rows and columns as follows

$$X' = (X^{(1)} \quad X^{(2)})' \quad \ldots\ldots\ldots\ldots \quad (1)$$

and

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}. \quad \ldots\ldots\ldots\ldots \quad (2)$$

Consider an arbitrary linear combination

$$U = \alpha' X^{(1)}$$

and

$$V = \gamma' X^{(2)}$$

The linear function that will maximize the correlation between U and V is wanted. Since the correlation of a multiple of U and a multiple of V is the same as the correlation of U and V, an arbitrary normalization of α and γ can be made. Therefore α and γ can be chosen such that U and V have unit variance.

Since $E(X) = 0$, therefore $E(U) = E(V) = 0$

$$Var(U) = E(U^2) = E(\alpha' X^{(1)} X^{(1)'} \alpha) = \alpha' \Sigma_{11} \alpha$$

and similarily

$$Var(V) = \gamma' \Sigma_{22} \gamma \ .$$

No generality is lost in requiring:

$$Var(U) = 1 \quad \dots\dots\dots\dots \quad (3)$$

$$Var(V) = 1 \quad \dots\dots\dots\dots \quad (4)$$

since a correlation is scale free.

Now the correlation between U and V is rewritten as

$$\rho\ (U,V) = E(UV) = E(\alpha' X^{(1)} X^{(2)'} \gamma) = \alpha' \Sigma_{12} \gamma \ . \quad \dots\dots\dots \quad (5)$$

Thus the algebraic problem is to find α and γ to maximize (5) subject to (3) and (4). Using Lagrange's method, let

$$\psi = \alpha' \Sigma_{12} \gamma - (\lambda/2)(\alpha' \Sigma_{11} \alpha - 1) - (\mu/2)(\gamma' \Sigma_{22} \gamma - 1) \ \dots\dots \ (6)$$

where λ and μ are Largange multipliers.

Normal equations are obtained by partial differentiation with respect to α and γ .

$$\Sigma_{12} \gamma - \lambda \Sigma_{11} \alpha = 0 \quad \dots\dots\dots\dots\dots\dots\dots\dots \quad (7)$$

$$\Sigma'_{12} \alpha - \mu \Sigma_{22} \gamma = 0 \quad \dots\dots\dots\dots\dots\dots\dots\dots \quad (8)$$

Their solution is expedited by multiplication of (7) on the left side by α' and (8) on the left side by γ' yielding

$$\alpha' \Sigma_{12} \gamma - \lambda \alpha' \Sigma_{11} \alpha = 0 \quad \dots\dots\dots\dots\dots\dots \quad (9)$$

$$\gamma' \Sigma_{21} \alpha - \mu \gamma' \Sigma_{22} \gamma = 0. \quad \dots\dots\dots\dots\dots\dots \quad (10)$$

Since, $\alpha' \Sigma_{11} \alpha = \gamma' \Sigma_{22} \gamma = 1$,

$$\lambda = \mu = \alpha' \Sigma_{12} \gamma \ ,$$

Thus (7) and (8) can be written as

$$-\lambda\Sigma_{11}\alpha + \Sigma_{12}\gamma = 0 \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (11)$$

$$\Sigma_{21}\alpha - \lambda\Sigma_{22}\gamma = 0 \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (12)$$

or

$$\begin{pmatrix} -\lambda\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\lambda\Sigma_{22} \end{pmatrix}\begin{pmatrix} \alpha \\ \gamma \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad \ldots\ldots\ldots\ldots\ldots \quad (13)$$

In order that there be a nontrivial solution, the determinant must be zero

$$\begin{vmatrix} -\lambda\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\lambda\Sigma_{22} \end{vmatrix} = 0. \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (14)$$

It can be shown that the determinant on the left is a polynomial of degree p. It therefore has p roots say,

$$\lambda_1 \geqq \lambda_2 \geqq \lambda_3 \geqq \ldots \geqq \lambda_p$$

and $V = \gamma' X^{(2)}$ when α and γ satisfy (13) for some value of λ. Thus the maximum correlation will be $\lambda = \lambda_1$.

Let a solution to (13) be

$$\alpha = \alpha^{(1)}, \qquad \gamma = \gamma^{(1)}$$

and let

$$U_1 = \alpha^{(1)\prime}X^{(1)}, \quad V_1 = \gamma^{(1)\prime}X^{(2)}$$

then U_1 and V_1 are normalized linear combinations of $X^{(1)}$ and $X^{(2)}$ respectively, with maximum correlation.

Now, consider a second linear combination of $X^{(1)}$ say, $U = \alpha' X^{(1)}$ and a second linear combination of $X^{(2)}$, $V = \gamma' X^{(2)}$, such that, of all linear combinations uncorrelated with U_1 and V_1 these have maximum correlation. This procedure is continued.

at the r^{th} step there are

$$U_1 = \alpha^{(1)\prime}X^{(1)}, \qquad V_1 = \gamma^{(1)\prime}X^{(2)}$$
$$U_2 = \alpha^{(2)\prime}X^{(1)}, \qquad V_2 = \gamma^{(2)\prime}X^{(2)}$$
$$\vdots \qquad\qquad \vdots$$
$$U_r = \alpha^{(r)\prime}X^{(1)}, \qquad V_r = \gamma^{(r)\prime}X^{(2)}$$

linear combinations with corresponding correlations

$$\lambda^{(1)} = \lambda_1, \lambda^{(2)}, \lambda^{(3)}, \dots, \lambda^{(r)}, \quad r \leq p_1.$$

Now, suppose we wish to evaluate U_{r+1}, V_{r+1}, then we must find a linear combination of $X^{(1)}$, $U = \alpha' X^{(1)}$ and a linear combination of $X^{(2)}$, $V = \gamma' X^{(2)}$ which is uncorrelated with all previous r sets of U and V but has maximum correlation. (α and γ now denote vectors distinct from the previous $\alpha^{(i)}$, $\gamma^{(i)}$, $i = 1, 2, \dots, r$).

The condition that U be uncorrelated with U_i is

$$E(UU_i) = E(\ \alpha' X^{(1)} X^{(1)\prime} \alpha^{(i)}\ ) = \alpha' \Sigma_{11} \alpha^{(i)} = 0 \quad \dots\dots\dots \ (15)$$

if $\lambda^{(i)} \neq 0$,

$$\Sigma_{11}\alpha^{(i)} = \frac{1}{\lambda^{(i)}} \Sigma_{12}\gamma^{(i)}.$$

Therefore

$$\alpha' \Sigma_{12} \gamma^{(i)} = 0.$$

This implies that

$$E(UV_i) = 0. \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \ (16)$$

If $\lambda^{(i)} = 0$, $\Sigma_{12}\gamma^{(i)} = 0$ and (16) also holds.

Similarly the condition that V be uncorrelated with V_i is

$$E(VV_i) = \gamma' \Sigma_{22} \gamma^{(i)} = 0$$

and this also implies that $E(VU_i) = 0$.

Now maximize $E(U_{r+1}V_{r+1})$, choosing α and γ to satisfy (3), (4), (15) and (17) for $i = 1, 2, \ldots, r$. Consider

$$\psi_{r+1} = \alpha'\Sigma_{12}\gamma - (\lambda/2)(\alpha'\Sigma_{11}\alpha - 1) - (\mu/2)(\gamma'\Sigma_{22}\gamma - 1) + \sum_{i=1}^{r} \nu_i\alpha'\Sigma_{11}\alpha^{(i)} + \sum_{i=1}^{r} \theta_i\gamma'\Sigma_{22}\gamma^{(i)} \quad \ldots\ldots\ldots \quad (19)$$

where $\lambda, \mu, \nu_1, \ldots, \nu_r, \theta_1, \ldots, \theta_r$ are Lagrange multipliers. Partial differentiation with respect to α and γ, gives

$$\Sigma_{12}\gamma - \lambda\Sigma_{11}\alpha + \sum_{i=1}^{r} \nu_i\Sigma_{11}\alpha^{(i)} = 0 \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (20)$$

$$\Sigma_{21}\alpha - \mu\Sigma_{22}\gamma + \sum_{i=1}^{r} \theta_i\Sigma_{22}\gamma^{(i)} = 0 \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (21)$$

Multiplication of (20) on the left by $\alpha^{(j)'}$ and (21) on the left by $\gamma^{(j)'}$ gives

$$\alpha^{(j)'}\Sigma_{12}\gamma - \lambda\alpha^{(j)'}\Sigma_{11}\alpha + \sum_{i=1}^{r} \nu_i\alpha^{(j)'}\Sigma_{11}\alpha^{(i)} = 0,$$

that is

$$\nu_i\alpha^{(j)'}\Sigma_{11}\alpha^{(j)} = 0. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (22)$$

Since $\quad \alpha^{(j)'}\Sigma_{11}\alpha^{(j)} = 1$

then $\quad \nu_j = 0.$

Similarily $\quad \theta_i = 0.$

Thus equation (20), (21) are simply (11) and (12). Therefore take the largest λ_i, say $\lambda^{(r+1)}$, such that there is a solution to (13) satisfying (3), (4), (15) and (17) for $i = 1, \ldots, r$. Let this solution be $\alpha^{(r+1)}, \gamma^{(r+1)}$ and let

$$U_{r+1} = \alpha^{(r+1)'} X^{(1)},$$
$$V_{r+1} = \gamma^{(r+1)'} X^{(2)}.$$

The procedure is continued step by step as long as successive solutions can be found which satisfy the conditions, namely, (13) for some λ_i, (3), (4), (15) and (17).

It can be shown that the number of steps for which this can be done is equal to the number of components in $X^{(1)}$, that is p_1.

The conditions on the λ's, α's and γ's can be summarized as follows:

Define $A = (\alpha^{(1)}\ \alpha^{(2)} \dots\ \alpha^{(p_1)})$,

$\Gamma_1 = (\gamma^{(1)}\ \gamma^{(2)} \dots\ \gamma^{(p_1)})$,

and

$$\Lambda = \begin{pmatrix} \lambda^{(1)} & 0 & 0 & \dots & 0 \\ 0 & \lambda^{(2)} & 0 & \dots & 0 \\ 0 & 0 & \lambda^{(3)} & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & \lambda^{(p_1)} \end{pmatrix} . \quad \dots\dots\dots \quad (23)$$

Then $A' \Sigma_{11} A = I$, ... (24)

$\Gamma_1' \Sigma_{22} \Gamma_1 = I$, ... (25)

$A' \Sigma_{12} \Gamma_1 = \Lambda$. .... (26)

Let $\Gamma_2 = (\gamma^{(p_1+1)}\ \gamma^{(p_1+2)} \dots\ \gamma^{(p_1+p_2)})$

be a $p_2 \times (p_2 - p_1)$ matrix satisfying

$\Gamma_2' \Sigma_{22} \Gamma_1 = 0$, ... (27)

$\Gamma_2' \Sigma_{22} \Gamma_2 = I$, ... (28)

and let $\Gamma = (\Gamma_1 \Gamma_2)$.

Since Γ_1 is a $p_2 \times p_1$ matrix and Γ_2 is a $p_2 \times (p_2 - p_1)$ matrix, therefore Γ is a square matrix of size $p_2 \times p_2$. The result

$$\Gamma_1' \Sigma_{22} \Gamma_1 = I$$

and

$$\Gamma_2' \Sigma_{22} \Gamma_2 = I$$

implies $\Gamma' \Sigma_{22} \Gamma = I$, that is, Γ is nonsingular.

Now, consider the determinant

$$\begin{vmatrix} A' & 0 \\ 0 & \Gamma_1' \\ 0 & \Gamma_2' \end{vmatrix} \begin{vmatrix} -\lambda\ \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\lambda\ \Sigma_{22} \end{vmatrix} \begin{vmatrix} A & 0 & 0 \\ 0 & \Gamma_1 & \Gamma_2 \end{vmatrix}$$

$$= \begin{vmatrix} -\lambda\ I & \Lambda & 0 \\ \Lambda & -\lambda\ I & 0 \\ 0 & 0 & -\lambda\ I \end{vmatrix}$$

$$= (-\lambda)^{p_2-p_1} \begin{vmatrix} -\lambda\ I & \Lambda \\ \Lambda & -\lambda\ I \end{vmatrix}$$

$$= (-\lambda)^{p_2-p_1} \left| -\lambda I \right| \left| -\lambda I - \Lambda(-\lambda I)^{-1}\Lambda \right|$$

$$= (-\lambda)^{p_2-p_1} \left| \lambda^2 I - \Lambda^2 \right|$$

$$= (-\lambda)^{p_2-p_1} \prod_{i=1}^{p_1} (\lambda^2 - \lambda^{(i)2}). \qquad \text{.................. (29)}$$

The above polynomial is

$$\begin{vmatrix} -\lambda\ \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\lambda\ \Sigma_{22} \end{vmatrix} \qquad \text{................... (30)}$$

except for a constant factor. Thus the roots of (14) are roots of

$$(-\lambda)^{p_2-p_1} \prod_{i=1}^{p_1} (\lambda^2 - \lambda^{(i)2}) = 0 \qquad \text{....... (31)}$$

where

$\lambda = \pm\lambda^{(i)}$ for $i = 1, 2, \ldots, p_1$

and (p_2-p_1) of the λ are zero.

Thus

$$(\lambda_1, \ldots, \lambda_p) = (\lambda^{(1)}, \ldots, \lambda^{(p_1)}, 0, \ldots, 0, -\lambda^{(p_1)}, \ldots -\lambda^{(1)}).$$

It can be shown that the set

$$\lambda^{(1)2},\ \lambda^{(2)2},\ \ldots\ .\ \lambda^{(p_1)}$$

is the set

$$\lambda_1^2,\ \lambda_2^2,\ \ldots\ ,\ \lambda_{p_1}^2\ .$$

Thus $\lambda^{(i)} = \lambda_i$ for $i = 1, 2, \ldots , p_1$.

Let

$$U = \begin{pmatrix} U_1 \\ \cdot \\ \cdot \\ \cdot \\ U_{p_1} \end{pmatrix} = A'X^{(1)}, \qquad \ldots\ldots\ldots\ldots\ldots\ldots \qquad (32)$$

$$V^{(1)} = \begin{pmatrix} V_1 \\ \cdot \\ \cdot \\ \cdot \\ V_{p_1} \end{pmatrix} = \Gamma_1' X^{(2)}, \qquad \ldots\ldots\ldots\ldots\ldots\ldots \qquad (33)$$

$$V^{(2)} = \begin{pmatrix} V_{p_1+1} \\ \cdot \\ \cdot \\ \cdot \\ V_{p_2} \end{pmatrix} = \Gamma_2' X^{(2)}. \qquad \ldots\ldots\ldots\ldots\ldots\ldots \qquad (34)$$

The components of U are one set of canonical variates and the components of $V = \begin{pmatrix} V^{(1)} \\ V^{(2)} \end{pmatrix}$ are the other set.

$$E\begin{pmatrix} U \\ V^{(1)} \\ V^{(2)} \end{pmatrix} (\ U' \quad V^{(1)'} \quad V^{(2)'}\ )$$

$$= E\begin{pmatrix} U\,U' & U\,V^{(1)'} & U\,V^{(2)'} \\ V^{(1)}U' & V^{(1)}V^{(1)'} & V^{(1)}V^{(2)'} \\ V^{(2)}U' & V^{(2)}V^{(1)'} & V^{(2)}V^{(2)'} \end{pmatrix}$$

$$= E\begin{pmatrix} A'X^{(1)}X^{(1)'}A & A'X^{(1)}X^{(2)'}\Gamma_1 & A'X^{(1)}X^{(2)'}\Gamma_2 \\ \Gamma_1'X^{(2)}X^{(1)'}A & \Gamma_1'X^{(2)}X^{(2)'}\Gamma_1 & \Gamma_1'X^{(1)}X^{(2)'}\Gamma_2 \\ \Gamma_2'X^{(2)}X^{(1)'}A & \Gamma_2'X^{(2)}X^{(2)'}\Gamma_1 & \Gamma_2'X^{(2)}X^{(2)'}\Gamma_2 \end{pmatrix}$$

$$= \begin{pmatrix} I & \Lambda & 0 \\ \Lambda & I & 0 \\ 0 & 0 & I \end{pmatrix}. \qquad \text{..........................} \qquad (35)$$

The previous discussion can be summarized in the following statements:

Definition: Let $X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}$ where $X^{(1)}$ has p_1 components and $X^{(2)}$ has $p_2(=p-p_1 \geq p_1)$ components. The r^{th} pair of canonical variates are the pair of linear combinations $U_r = \alpha^{(r)'}X^{(1)}$ and $V_r = \gamma^{(r)'}X^{(2)}$, each of unit variance and uncorrelated with the (r-1) pair of canonical variates and having maximum correlation. The correlation is the r^{th} canonical correlation.

Theorem: Let $X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}$ be a random vector with covariance matrix Σ then the r^{th} canonical correlation between $X^{(1)}$

and $X^{(2)}$ is the r^{th} largest root of (14). The coefficients of $\alpha^{(r)\prime}X^{(1)}$ and $\gamma^{(r)\prime}X^{(2)}$ define the r^{th} pair of canonical variates satisfy (13) for $\lambda = \lambda_r$ and (3) and (4).

A single matrix equation for α and γ can be derived. If (11) is multiplied by λ and (12) by Σ_{22}^{-1}, it becomes

$$-\lambda^2\Sigma_{11}\alpha + \lambda\Sigma_{12}\gamma = 0$$

$$\Sigma_{22}^{-1}\Sigma_{21}\alpha - \lambda\Sigma_{22}^{-1}\Sigma_{22}\gamma = 0$$

or

$$\lambda\Sigma_{12}\gamma = \lambda^2\Sigma_{11}\alpha \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (36)$$

$$\Sigma_{22}^{-1}\Sigma_{21}\alpha = \lambda\gamma . \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (37)$$

Thus

$$\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\alpha = \lambda\Sigma_{11}\alpha \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (38)$$

or

$$(\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - \lambda^2\Sigma_{11})\alpha = 0. \quad \ldots\ldots\ldots\ldots \quad (39)$$

The quantities $\lambda_1^2, \ldots , \lambda_{p_1}^2$ satisfy

$$| \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - \lambda^2\Sigma_{11} | = 0 \quad \ldots\ldots\ldots\ldots\ldots \quad (40)$$

and $\alpha^{(1)}, \alpha^{(2)}, \ldots , \alpha^{(p_1)}$ satisfy (38) for

$$\lambda^2 = \lambda_1^2, \lambda_2^2, \ldots , \lambda_{p_1}^2 \quad \text{respectively.}$$

Similary, if (11) is multiplied by Σ_{11}^{-1} and (12) by λ, it becomes

$$-\lambda\Sigma_{11}^{-1}\Sigma_{11}\alpha + \Sigma_{11}^{-1}\Sigma_{12}\gamma = 0$$

$$\lambda\Sigma_{21}\alpha - \lambda^2\Sigma_{22}\gamma = 0$$

or

$$\lambda\alpha = \Sigma_{11}^{-1}\Sigma_{12}\gamma$$

$$\lambda\Sigma_{21}\alpha = \lambda^2\Sigma_{22}\gamma .$$

Thus

$$\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\gamma = \lambda^2\Sigma_{22}\gamma \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (41)$$

or

$$(\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} - \lambda^2\Sigma_{22})\,\gamma = 0. \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (42)$$

The equantities $\lambda_1^2, \lambda_2^2, \ldots , \lambda_{p_2}^2$ satisfy

$$| \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} - \lambda^2\Sigma_{22}| = 0 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (43)$$

and $\gamma(1), \gamma(2), \ldots , \gamma(p_2)$ satisfy (41) for

$$\lambda^2 = \lambda_1^2, \lambda_2^2, \ldots , \lambda_{p_2}^2 \ .$$

(b) Computational procedure

From the previous discussion, equation (38), (39) or (40) usually will be used in computation. The computational procedure in terms of the population quantities is as follows:

(1) Obtain Σ_{11}^{-1} and Σ_{22}^{-1} .

(2) Obtain $B = \Sigma_{22}^{-1} \Sigma_{21}$.

(3) Obtain $R = \Sigma_{12} B$.

(4) Obtain $\Sigma_{11}^{-1} R = E$.

(5) Obtain largest characteristic root of E.

(6) Form $Q = E - \nu I$
(subtract ν from each diagonal element).

(7) Solve $Q\alpha = 0$, preferably by eliminating the last row of Q and putting the last column with opposite sign to right hand side. This sets $\alpha_{p_1} = 1$.

(8) Solve $\gamma = (\Sigma_{22}^{-1} \Sigma_{21})' \alpha$.

If p_1 is sufficiently small, the determinant Q can be expanded into a polynomial in ν and solved for ν. Then use (40) to find the vectors α.

If p_1 is too large, the direct method is hard to evaluate. An iterative method can be used. S. N. Roy and J. Roy (1958) suggested a very convenient iterative procedure for finding the characteristic roots of E of size $p_1 \times p_1$ as follows:

Let Y_0 be an arbitrary row vector with p_1 elements, not all zero, compute recursively the vector Z_0, Y_1, Z_1, Y_2, ... , as follows:

$$
\begin{array}{ll}
 & Z_0 = Y_0 / \| Y_0 \| \\
Y_1 = Z_0 E & Z_1 = Y_1 / \| Y_1 \| \\
Y_2 = Z_1 E & Z_2 = Y_2 / \| Y_2 \| \\
\vdots & \vdots \\
Y_{i+1} = Z_i E & Z_{i+1} = Y_{i+1} / \| Y_{i+1} \|
\end{array}
$$

where $\| Y_i \| = (Y_i Y_i')^{1/2}$.

It can be shown that the sequence of vectors Z_i will converge to a characteristic vector of E corresponding to the largest characteristic root of E and the root corresponding to Z can be computed from

$$\lambda = Z E Z'.$$

It is customary and convenient to start with the vector

$$Y = (S_1 \; S_2 \; \ldots \; S_{p_1})$$

where S_i is the sum of the elements in the i^{th} column of the matrix E. If a second characteristic root and / or vector is required, one first computes the residual matrix

$$E^* = E - \lambda Z' Z$$

and makes use of the fact that any characteristic root (or vector) of E^* is also a characteristic root (or vector) of E.

The rapidity of convergence of the iterative procedure can be increased by first raising the matrix E to some convenient power, say E^m, and making use of the fact that if $\lambda_1, \lambda_2, \ldots ,$ λ_n are the characteristic roots of E, then those of E^m are $\lambda_1^m, \lambda_2^m, \ldots , \lambda_n^m$. It is convenient to take m of the form m=2, and compute E, E^2, E^4, E^8, ... , etc. by successive squaring of matrices.

The procedure is continued to find as many characteristic roots and its corresponding vectors as desired.

Cooley and Lohnes (1962) gives two computer programs, one for computing the largest canonical correlation, the other for computing all.

SAMPLING DISTRIBUTION

The problem of determining the sampling distribution of the canonical correlation coefficient under general conditions is very complicated. However, under certain conditions the problem was solved approximately simultaneously by Fisher (1939), Girshick (1939), Hsu (1939), Mood (1951) (Mood's results were not published until twelve years after they were obtained) and Roy (1939), using different methods.

For samples from a normal population and the canonical correlation in the population assumed zero, Hotelling (1936) obtained the joint distribution of the sample canonical correlations for the special case in which there are only two variables in each set. The distribution of the canonical correlation r_1 and r_2 is of the form:

$$(n-2)(n-3)(r_1^2-r_2^2)(1-r_1^2)^{(n-5)/2}(1-r_2^2)^{(n-5)/2}\, dr_1 dr_2$$

where n is one less than the number in the sample for each variate.

Later, Grishick (1939) extended to the case in which there are two variables in one set and any number of variables in the other. The form is

$$\frac{(n-2)!}{(p_2-2)!(n-p_2-2)!}\,(r_1^2-r_2^2)(r_1 r_2)^{p_2-2}((1-r_1^2)(1-r_2^2))^{\frac{n-p_2-3}{2}}\,dr_1 dr_2.$$

At about the same time that Girshick published his paper, Fisher (1939) and Hsu (1939) gave the joint distribution of the canonical correlations for the most general case. The results due to Hsu can be summarized in the following statement:

Let (x_1, x_2, ... , x_p) and (x_{p+1}, x_{p+2}, ... , x_{p+q}) be two mutually independent sets of variables of which the first set is normally distributed. Let θ_1, θ_2, ... be the squares of the canonical correlations between the two sets, arranged in the decending order of magnitude. Then the joint distribution of the θ_i is given by

$$K \cdot \prod_{i\,j}^{p} (\theta_i - \theta_j) \prod_{i=1}^{p} \theta_i^{(q-p-1)/2} \prod_{i=1}^{p} (1-\theta_i)^{(n-p-q-2)/2} \prod_{i=1}^{p} d\theta_i, \quad \text{if } p \leq q,$$

where

$$K = (\pi)^{p/2} \prod_{i=1}^{p} \frac{\frac{1}{2}(n-i)}{\Gamma\frac{1}{2}(n-q-i)\Gamma\frac{1}{2}(p-i+1)\Gamma\frac{1}{2}(q-i+1)} ,$$

if $q < p$, the distribution is represented by the above formula with the letters p and q interchanged.

TESTING SIGNIFICANCE

The joint distribution of canonical correlation, given by Hsu, for any value of p and q makes further possible tests available. The percentage points of the distribution of the largest characteristic root have been computed and charts for it were prepared by Heck (see Roy, S. N. and Roy, J. (1958) appendix), which can be sued to test the significance of the largest canonical correlation. For large n, Bartlett (1941) has outlined procedures for testing the significance of canonical correlation. He defines

$$\Lambda = \prod_{i=1}^{p} (1-\lambda_i) \qquad p \leq q$$

and the Chi-square approximation for the distribution of Λ provides a test for the null hypotheses that the p variates are unrelated to the q variates, that is

$$\chi^2 = -(n-(p+q+1)/2) \log_e \Lambda$$

with pxq degrees of freedom.

If the null hypotheses can be rejected, the contribution of first root to Λ can be removed and the significance of the (p-1) roots can be tested using

$$\Lambda' = \prod_{i=1}^{p-1} (1-\lambda_i)$$

$$\chi^2 = -(n-(p+q+1)/2) \log_e \Lambda'$$

with (p-1)(q-1) degrees of freedom. In general, with r roots removed,

$$\Lambda' = \prod_{i=1}^{p-r} (1-\lambda_i)$$

and Chi-square is distributed with (p-r)(q-r) degrees of freedom.

EXAMPLE

The following example illustrates how to find canonical variates and canonical correlations using both the direct method and iterative method:

Let

x_1 = price of food index,

x_2 = price of other commodities indes,

x_3 = production durables index,

x_4 = production non-durables index,

x_5 = index agricultural production.

we wish to find the canonical correlation between production indices and price indices. The correlation matrix is partitioned as follows

$$R = \left(\begin{array}{cc|ccc} 1.000 & .914 & -.427 & .430 & .267 \\ .914 & 1.000 & -.203 & .584 & .378 \\ \hline -.427 & -.203 & 1.000 & .496 & .481 \\ .430 & .584 & .496 & 1.000 & .710 \\ .267 & .378 & .481 & .710 & 1.000 \end{array}\right) .$$

Here

$$R_{11} = \begin{pmatrix} 1.000 & .914 \\ .914 & 1.000 \end{pmatrix} ,$$

$$R_{22} = \begin{pmatrix} 1.000 & .496 & .481 \\ .496 & 1.000 & .710 \\ .481 & .710 & 1.000 \end{pmatrix} ,$$

$$R_{12} = \begin{pmatrix} -.427 & .430 & .267 \\ -.203 & .584 & .378 \end{pmatrix} .$$

(I) Direct method:

(1) Find Σ_{11}^{-1} and Σ_{22}^{-1}.

$$R_{11}^{-1} = \begin{pmatrix} 6.0752 & -5.5527 \\ -5.5527 & 6.0752 \end{pmatrix} ,$$

$$R_{22}^{-1} = \begin{pmatrix} 1.3879 & -0.4324 & -0.3606 \\ -0.4324 & 2.1512 & -1.3194 \\ -0.3606 & -1.3194 & 2.1102 \end{pmatrix} .$$

(2) Find $B = \Sigma_{22}^{-1} \Sigma_{21}$.

$$B = \begin{pmatrix} -.87485 & .75737 & .15006 \\ -.67057 & .84535 & .10036 \end{pmatrix} .$$

(3) Obtain $R = \Sigma_{12} B$.

$$R = \begin{pmatrix} .7393 & .6766 \\ .6766 & .6677 \end{pmatrix} .$$

(4) Obtain $E = \Sigma_{12}^{-1} R$.

$$E = \begin{pmatrix} .7344 & .4029 \\ .0054 & .2995 \end{pmatrix} .$$

(5) Find the largest characteristic root of E.

$$\begin{vmatrix} .7344 - \nu & .4029 \\ .0054 & .2995 - \nu \end{vmatrix} = 0$$

$$(.7344 - \nu)(.2995 - \nu) - (.0054)(.4029) = 0$$

$$\nu^2 - 1.0339\,\nu + .21778 = 0$$

$$\nu = .51695 \pm \sqrt{((.51695)^2 - .21778)}$$

$$= .51695 \pm .22239,$$

$$\nu_1 = .73934 = \lambda_1^2$$

$$\nu_2 = .29456 = \lambda_2^2 .$$

The canonical correlation is therefore

$$\lambda_1 = .860$$

$$\lambda_2 = .542$$

(6) Form $Q = E - \nu I$.

$$Q = \begin{pmatrix} .7344 - .73934 & .4029 \\ .0054 & .2995 - .73934 \end{pmatrix}$$

$$= \begin{pmatrix} -.00494 & .4029 \\ .0054 & -.43984 \end{pmatrix} .$$

(7) Solve $Q\alpha = 0$.

Set $\alpha_2 = 1$, i.e., [illegible]

the last column and

$$-.00494\,\alpha_1 = -.4029$$

$$\alpha_1 = 81.56$$

therefore

$$\alpha = \begin{pmatrix} 81.56 \\ 1.00 \end{pmatrix} .$$

(8) Solve $\gamma = (\Sigma_{22}^{-1}\,\Sigma_{21})'\alpha$.

$$= B' \alpha$$

$$= \begin{pmatrix} -.87485 & -.67057 \\ .75737 & .84534 \\ .15006 & .10033 \end{pmatrix} \begin{pmatrix} 81.56 \\ \\ 1.00 \end{pmatrix}$$

$$= \begin{pmatrix} -72.023 \\ 62.616 \\ 12.339 \end{pmatrix} .$$

Therefore the first pair of canonical variates is

$$U = \alpha' X^{(1)}$$

$$= (81.56 \quad 1.00) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= 81.56\, x_1 + x_2 ,$$

$$V = \gamma' X^{(2)}$$

$$= -72.023\, x_3 + 62.616\, x_4 + 12.339\, x_5 .$$

(II) Iterative method

Do step (1), (2), (3) and (4) as before, but for step (5), i. e., to find the characteristic roots of E, start with

$$Y_0 = (.7398 \quad .7024).$$

Then

$$\| Y_0 \| = (Y_0 Y_0')^{1/2} = 1.1861,$$

$$Z_0 = Y_0 / \| Y_0 \| = (.62372 \quad .59219),$$

$$Y_1 = Z_0 E = (.46126 \quad .59219) \begin{pmatrix} .7344 & .4029 \\ .0054 & .2995 \end{pmatrix}$$

$$= (.46126 \quad .42866),$$

$$Z_1 = Y_1 / \| Y_1 \| = (.73251 \quad .68074),$$

$$Y_2 = Z_1 E = (.54163 \quad .49901),$$

$$Z_2 = Y_2 / \| Y_2 \| = (.73544 \quad .67757),$$

$$Y_3 = Z_2 E = (.543766 \quad .49241),$$

$$Z_3 = Y_3 / \| Y_3 \| = (.73662 \quad .67630),$$

$$Y_4 = Z_3 E = (.54463 \quad .49935),$$

$$Z_4 = Y_4 / \| Y_4 \| = (.73709 \quad .67579),$$

$$Y_5 = Z_4 E = (.544968 \quad .499373),$$

$$Z_5 = Y_5 / \| Y_5 \| = (.73727 \quad .67559),$$

$$Y_6 = Z_5 E = (.54510 \quad .49939),$$

$$Z_6 = Y_6 / \| Y_6 \| = (.73735 \quad .67551),$$

$$Y_7 = Z_6 E = (.54516 \quad .49939),$$

$$Z_7 = Y_7 / \| Y_7 \| = (.73737 \quad .67547),$$

$$Y_8 = Z_7 E = (.54517 \quad .49939),$$
$$Z_8 = Y_8 / \| Y_8 \| = (.73739 \quad .67547),$$
$$Y_9 = Z_8 E = (.54519 \quad .49940),$$
$$Z_9 = Y_9 / \| Y_9 \| = (.73739 \quad .67545).$$

Since $Z_8 \cong Z_9$, Z_9 is the vector corresponding to the largest charăcteristic root.

$$\begin{aligned}
\nu_{max} &= Z\ E\ Z' \\
&= (.73739 \quad .67545) \begin{pmatrix} .7344 & .4029 \\ .0054 & .3995 \end{pmatrix} \\
&\qquad X \begin{pmatrix} .73739 \\ .67545 \end{pmatrix} \\
&= .73939, \\
\lambda_1 &= .860.
\end{aligned}$$

To find the second characteristic root, compute the residual matrix first:

$$\begin{aligned}
E^* &= E - \lambda\ Z'\ Z \\
&= \begin{pmatrix} .7344 & .4029 \\ .0054 & .2995 \end{pmatrix} - (.73937) \\
&\qquad \begin{pmatrix} .73737 \\ .67545 \end{pmatrix} (.73739 \quad .67545) \\
&= \begin{pmatrix} .7344 & .4029 \\ .0054 & .2995 \end{pmatrix} - \begin{pmatrix} .4020 & .3683 \\ .3683 & .3373 \end{pmatrix} \\
E^* &= \begin{pmatrix} .3324 & .0346 \\ -.3629 & -.0378 \end{pmatrix}.
\end{aligned}$$

Then, following the same procedure as before, let

$$Y_0 = (-.0305 \quad -.0032),$$

then,

$$\| Y_0 \| = (Y \; Y')^{1/2} = (.00094094)^{1/2} = .03067,$$

$$Z_0 = (-.99445 \quad -.10433),$$

$$Y_1 = (-.29269 \quad -.03046),$$

$$Z_1 = (-.99463 \quad -.10351),$$

$$Y_2 = (-.29305 \quad -.03050),$$

$$Z_2 = (-.99463 \quad -.10351).$$

Now, $Z_1 = Z_2$, the vector corresponding to the second characteristic root is found.

Then,

$$\nu_2 = Z E^* Z'$$

$$= (-.29305 \quad -.03050) \begin{pmatrix} .3324 & .0346 \\ -.3629 & -.0378 \end{pmatrix} \begin{pmatrix} -.29305 \\ -.03050 \end{pmatrix}$$

$$= .29463.$$

Therefore

$$\lambda_2 = .542.$$

The result is the same as found by direct method.

ACKNOWLEDGEMENT

The writer wishes to express his sincere appreciation to his major professor, Dr. Gary F. Krause, for suggesting this topic and for his advice and assistance during the preparation of this report.

REFERENCES

Anderson, T. W. (1962)

" An introduction to multivariate statistical analysis, "
288-305, John Wiley and Sons, Inc.

Bartlett, M. S. (1941)

" The statistical significance of canonical correlation, "
Biometrika, 32, 29-37.

Bartlett, M. S. (1947)

" The general canonical correlation distribution, "
Ann. Math. Stat., 18, 1-17.

Cooley, William W. and Lohnes, Paul R. (1962)

" Multivariate procedures for the behavioral sciences, "

Fisher, R. A. (1939)

" The sampling distribution of some statistics obtained from non-linear equations, "
Ann. Eugen., 9, 238-249.

Girshick, M. A. (1939)

" On the sampling theory of roots of determinantal equation,"
Ann. Math. Stat., 10, 203-224.

Hotelling, Harold (1935)

" The most predictable criterion, "
J. Educ. Psych., 26, 139-142.

Hotelling, Harold (1936)

" Relations between two sets of variates, "
Biometrika, 28, 321-377.

Hsu, P. L. (1939)

" On the distribution of the roots of certain determinantal equations, "

Ann. Eugen., 9, 250-258.

Hsu, P. L. (1941)

" On the limiting distribution of the canonical correlation, "

Biometrika, 32, 38-45.

Mood, A. M. (1951)

" On the distribution of the characteristic roots of normal second-moment matrices, "

Ann. Math. Stat., 22, 266-273.

Roy, S. N. (1939)

" P-statistics or some generalizations in analysis of variance appropriate to multivariate problems, "

Sankhyā, 4, 381-396.

Roy, S. N. and Roy, J. (1958)

" Analysis of variance with univariate or multivariate, fixed or mixed classical models,"

North Carolina University Institute of Stat. mimeograh series No. 208., II-17.

CANONICAL VARIATES AND CORRELATIONS

by

CHI-YUNG CHIANG

B. A. National Taiwan University, 1958

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

Kansas State University
Manhattan, Kansas

1965

Procedures for finding the correlation between two scalar variables, simple r, a scalar variable and an associated vector, multiple R, a scalar and a second scalar, holding a vector fixed, partial correlations, are discussed in standard texts. The procedure of canonical variables permits the definition of correlation between two vectors of arbitrary size, which is the most general case.

The basic theory of canonical variates and canonical correlation was set forth in 1936 by Hotelling. Considerable additions have been made by various later writers.

Consider p variables, x_i where $i = 1, 2, \ldots, p$, each has n observations, the variables are divided into two groups, p_1 of them contained in one group, the remaining p_2 variables in the other. The problem is to find two linear functions

$$U = \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_{p_1} x_{p_1} = \alpha' X^{(1)}$$

and

$$V = \gamma_{p_1+1} x_{p_1+1} + \gamma_{p_1+2} x_{p_1+2} + \ldots + \gamma_{p_1+p_2} x_{p_1+p_2} = \gamma' X^{(2)}$$

which have maximum correlation with each other, subject to the constraints that

$$Var(u) = Var(V) = 1.$$

The canonical correlation coefficient between U and V becomes

$$\lambda = \alpha' \Sigma_{12} \gamma .$$

To make it a maximum under the conditions that the variances are unity, a system of linear equations is obtained and a determinantal equation solved to determine the λ_i^2's, which are the square of the canonical correlation coefficients. The largest root of this equation when inserted into the system of linear equations, determines the scalar vectors α and γ. Thus the canonical variates U and V are obtained while U is most successful in predicting V and V the best predictor of U .

The sampling distribution of canonical correlations has been found by various authors, under the hypotheses that their population value is zero. Fisher and Hsu gave the joint distribution of the canonical correlations for the most general case.

The percentage points of the distribution of the largest characteristic root have been computed and the charts for it prepared by Heck. They can be used to test the significance of the largest canonical correlation. For large n, Bartlett defined

$$\Lambda = \prod_{i=1}^{p} (1 - \lambda_i)$$

and the Chi-square approximation for the distribution of Λ provides a test for the null hypotheses that the p variates are uncorrelated to the q variates.

The mathematical derivation and the computational procedures for finding the canonical correlations in detail are presented in this report. A numerical example, solved by using a direct method and an iterative method, illustrates how the canonical variates and canonical correlations can be found.