Canad. Math. Bull. Vol. **28** (1), 1985

# RESIDUAL CANONICAL CORRELATIONS

BY

ANANT M. KSHIRSAGAR AND R. P. GUPTA

ABSTRACT. Residual canonical correlations are defined and are derived in terms of canonical correlations. Some measures of residual association are also defined, in terms of the residual canonical correlations and some possible applications are suggested.

1. **Introduction**. The sample canonical correlations $r_1, r_2, \ldots, r_p$ between two vector variables $\boldsymbol{x}$ and $\boldsymbol{y}$ ($\boldsymbol{x}$ of $p$ components, $\boldsymbol{y}$ of $q$ components, with $p \leq q$ without loss of generality) are measures of association between them, and are fairly well known statistics. In many situations, however, one needs to eliminate $s$ ($s < p$) specified linear functions of $\boldsymbol{x}$ and obtain measures of the "residual" association between $\boldsymbol{x}$ and $\boldsymbol{y}$, that is left. These functions may either be known to the experimenter from prior information or might be under consideration for replacing the $p$ $x$-variables by a smaller set. One needs these measures in Discriminant Analysis, Contingency Table Analysis, Econometrics and many other fields. The reason for eliminating $s$ given functions may be that they are apriorily known to be irrelevant or it could also be that it is conjectured that these specified functions are the only ones that contribute to any meaningful association between $\boldsymbol{x}$ and $\boldsymbol{y}$. In the latter case then, the conjecture or hypothesis may be tested by the significance of the "residual" canonical correlations. In the former case, the residual canonical correlations are "refined" or "revised" measures when irrelevant functions are eliminated.

The idea of residual canonical correlations is originally due to Williams [5] (see also Kshirsagar [4]); but he introduced it only for the special case of $s = 1$. We consider this in a more general context and also develop some "overall" measures of residual association.

2. **Residual canonical correlations**. Let us denote by $C_{xx}$, $C_{xy}$, $C_{yy}$ the matrices of the corrected sum of squares and products (S.S. & S.P.) of observations on $\boldsymbol{x}$ and $\boldsymbol{y}$ and based on $n$ degrees of freedom (d.f.). This notation is due to Bartlett [1]. We shall further denote by $C_{xx \cdot y}$ the matrix

$$C_{xx} - C_{xy}C_{yy}^{-1}C_{yx}. \tag{2.1}$$

Received by the editors August 20, 1982 and, in final revised form, April 5, 1984.
AMS Subject Classification (1980): 62M20.
Key words: Canonical correlations, residual canonical correlations, measure of residual association.
This research was supported by Natural Science and Engineering Research Council of Canada. Grant No. A5290.

which is the matrix of the "residual" S.S. & S.P. in the regression of $\boldsymbol{x}$ **on** $\boldsymbol{y}$. Let $\boldsymbol{\xi} = L'\boldsymbol{x}$, where $L$ is $p \times s$, of rank $s$, denote the specified $s$ linear functions of $\boldsymbol{x}$, which are to be eliminated. Then we have the following Analysis of Dispersion table:

TABLE

| Source | d.f. | S.S. & S.P. matrix |
|---|---|---|
| Regression of $\boldsymbol{y}$ on $\boldsymbol{\xi}$ | $s$ | $C_{y\xi}C_{\xi\xi}^{-1}C_{\xi y}$ |
| Regression of $\boldsymbol{y}$ on $\boldsymbol{x}$, given $\boldsymbol{\xi}$ | $p - s$ | $C_{yx}C_{xx}^{-1}C_{xy} - C_{y\xi}C_{\xi\xi}^{-1}C_{\xi y} = C_{yy\cdot\xi} - C_{yy\cdot x}$ |
| Residual in the regression of $\boldsymbol{y}$ on $\boldsymbol{x}$ | $n - p$ | $C_{yy\cdot x}$ |
| Total | $n$ | $C_{yy}$ |

The usual unrestricted Canonical Correlations $r_i (i = 1, \ldots, p)$ between $\boldsymbol{x}$ and $\boldsymbol{y}$ are obtained from the roots of the determinantal equation,

$$|-r^2 C_{yy} + C_{yx}C_{xx}^{-1}C_{xy}| = 0, \tag{2.2}$$

or

$$|(1 - r^2)C_{yy} - C_{yy\cdot x}| = 0. \tag{2.3}$$

Correspondingly, we define the restricted canonical correlations $\phi_j (j = 1, \ldots, p - s)$ between $\boldsymbol{y}$ and $\boldsymbol{x}$, when $L'\boldsymbol{x}$ is eliminated, to be obtainable from the equation.

$$|-\phi^2 C_{yy\cdot\xi} + (C_{yy\cdot\xi} - C_{yy\cdot x})| = 0 \tag{2.4}$$

or

$$|(1 - \phi^2)C_{yy\cdot\xi} - C_{yy\cdot x}| = 0. \tag{2.5}$$

Williams [5], in the particular case $s = 1$, defined them alternatively as the principal semi-axes of the $(p - 1)$-dimensional ellipsoid obtained by the intersection of a $p$-dimensional ellipsoid and a hyperplane. However, in our case one would be dealing with a $(p - s)$-dimensional ellipsoid obtained by the intersection of a $p$-dimensional ellipsoid with a $(p - s)$ dimensional hyperplane.

To express the $\phi_j^2$ in terms of the $r_i^2$ and $\boldsymbol{\xi}$, we assume without loss of generality that the $x$'s and $y$'s are sample canonical variables, so that with the usual convention of scaling, we have

$$C_{xx} = nI_p, \qquad C_{yy} = nI_q$$

$$\frac{1}{n}C_{xy} = \begin{bmatrix} r_1 & & & 0 \ldots 0 \\ & r_2 & & \\ & & \ddots & \\ & & r_p & 0 \ldots 0 \end{bmatrix} \tag{2.6}$$

$$= [D\{r_i\}|0]$$

https://doi.org/10.4153/CMB-1985-003-2 Published online by Cambridge University Press

where $D\{a_i\}$ will denote, in future, a diagonal matrix with diagonal elements $a_1, a_2, \ldots, a_p$ and 0 will denote a null matrix of appropriate order. We also assume, without loss of generality, that the matrix $L$ satisfies

$$L'L = I_s, \tag{2.7}$$

so that the $s$ specified functions are orthonormal. Then

$$\frac{1}{n}C_{yy\cdot x} = I_q - \begin{bmatrix} D\{r_i^2\} & 0 \\ 0 & 0 \end{bmatrix} \tag{2.8}$$

and

$$\frac{1}{n}C_{yy\cdot \xi} = I_q - \begin{bmatrix} D\{r_i\}LL'D\{r_i\} & 0 \\ 0 & 0 \end{bmatrix} \tag{2.9}$$

Therefore, (2.5) reduces to

$$|(1 - \phi^2)(I_p - D\{r_1\}LL'D\{r_i\}) - I_p - D\{r_i^2\}| = 0. \tag{2.10}$$

This can also be expressed as

$$|D\{\phi^2 - r_i^2\} + (1 - \phi^2)D\{r_i\}LL'D\{r_i\}| = 0 \tag{2.11}$$

Simplifying further, (2.10) or (2.11) becomes,

$$|D\{\phi^2 - r_i^2\}|\left|I_p + (1 - \phi^2)D\left\{\frac{r_i^2}{\phi^2 - r_i^2}\right\}LL'D\{r_i\}\right| = 0. \tag{2.12}$$

We now use,

$$|I_p + PQ| = |I_s + QP|, \tag{2.13}$$

where $P$, $Q$ are matrices of order $p \times s$ and $s \times p$, respectively. This is

$$\left|I_s + (1 - \phi^2)L'D\left\{\frac{r_i^2}{\phi^2 - r_i^2}\right\}L\right|\prod_1^p (\phi^2 - r_i^2) = 0 \tag{2.14}$$

Using (2.7), this can also be written as

$$\left|L'D\left\{\frac{\phi^2(1 - r_i^2)}{\phi^2 - r_i^2}\right\}L'\right|\prod_1^p (\phi^2 - r_i^2) = 0. \tag{2.15}$$

The squared residual canonical correlations $\phi_j^2 (j = 1, \ldots, p - s)$ are the roots of the equations (2.15).

3. **Measures of Residual Association**. In the statistical literature,

$$\sum_1^p r_i^2, \qquad \sum_1^p \frac{r_i^2}{1 - r_i^2}, \qquad \prod_1^p (1 - r_i^2)$$

are some of the important measures of association between $\boldsymbol{x}$ and $\boldsymbol{y}$. Actually, the last one is a measure of lack of association. We, therefore, propose

https://doi.org/10.4153/CMB-1985-003-2 Published online by Cambridge University Press

$$S_1 = \sum_{i}^{p} r_i^2 - \sum_{1}^{p-s} \phi_j^2 \tag{3.1}$$

$$S_2 = \sum \frac{r_i^2}{1 - r_i^2} - \sum \frac{\phi_j^2}{1 - \phi_j^2} \tag{3.2}$$

$$S_3 = \Pi(1 - r_i^2)/\Pi(1 - \phi_j^2) \tag{3.3}$$

as measures of overall residual association (or lack of association) between $\boldsymbol{x}$ and $\boldsymbol{y}$, when $\xi$ is eliminated. We shall now express $S_1$, $S_2$, $S_3$ in terms of $r_i^2$ and $\boldsymbol{\xi}$.

From (2.10), we observe that $1 - \phi_j^2$ are the eigenvalues of the matrix

$$(I_p - D\{r_i\}LL'D\{r_i\})^{-1}(I_p - D\{r_i^2\}). \tag{3.4}$$

Hence, defining $\phi_j^2 = 0$ for $j > p - s$, and using tr for trace of a matrix,

$$\sum_{j=1}^{p} (1 - \phi_j^2) = \text{tr}\,[(I_p - D\{r_i\}LL'D\{r_i\})^{-1}(I_p - D\{r_i^2\})] \tag{3.5}$$

Again, observing (see for example, Graybill [2])

$$(I_p - D\{r_i\}LL'D\{r_i\})^{-1} = I_p + D\{r_i\}L(I_s - L'D\{r_i^2\}L)^{-1}L'D\{r_i\} \tag{3.6}$$

Substituting (3.6) into (3.5), one gets

$$\begin{aligned} S_1 = \Sigma r_i^2 - \Sigma \phi_j^2 &= \text{tr}\,[D\{r_i\}L(I_s - L'D\{r_i^2\}L)^{-1}L'D\{r_i\}(I_p - D\{r_i^2\})] \\ &= \text{tr}\,[L'D\{r_i\}(I_p - D\{r_i^2\}D\{r_i\}L)(I_s - L'D\{r_i^2\}L)^{-1}] \\ &= \text{tr}\,[L'D\{r_i^2\}(1 - r_i^2)\}L(L'D\{1 - r_i^2)\}L)^{-1}]. \end{aligned} \tag{3.7}$$

where tr $AB$ = tr $BA$ and $L'L = I_s$.

Similarly, from (3.4),

$$\prod_j (1 - \phi_j^2) = \frac{|I_p - D\{r_i^2\}|}{|I_p - D\{r_i\}LL'D\{r_i\}|}.$$

and, therefore, using (2.13) we get

$$S_3 = \frac{\Pi(1 - r_i^2)}{\Pi(1 - \phi_j^2)} = |L'D\{1 - r_i^2)\}L|. \tag{3.8}$$

To find $S_2$, we let

$$r_i^{*2} = \frac{r_i^2}{1 - r_i^2}, \qquad \phi_j^{*2} = \frac{\phi_j^2}{1 - \phi_j^2}. \tag{3.9}$$

Then from (2.10) or (2.11), we obtain, after some algebra,

$$|-\phi^{*2}I_p + D\{r_i^{*2}\} - D\{(1 + r_i^{*2})r_i\}LL'D\{r_i\}| = 0 \tag{3.10}$$

So, it follows that

$$\Sigma \phi_j^{*2} = \text{tr}\,[D\{r_i^{*2}\} + D\{1 - r_i^{*2})r_i\}LL'D\{r_i\}], \tag{3.11}$$

https://doi.org/10.4153/CMB-1985-003-2 Published online by Cambridge University Press

or that (using tr $AB$ = tr $BA$),

$$
\begin{aligned}
S_2 &= \Sigma r_i^{*2} - \Sigma \phi_j^{*2} \\
&= \text{tr}\,[D\{(1 + r_i^{*2})r_i\}LL'D\{r_i\}] \\
&= \text{tr}\,[L'D\{r_i\}D\{1 + r_i^{*2})r_i\}] \\
&= \text{tr}\,[L'D\{r_i^2(1 + r_i^{*2})\}L] \\
&= \text{tr}\,(L'D\{r_i^2/(1 - r_i^2)\}L).
\end{aligned} \tag{3.12}
$$

4. **Some applications**. If we consider a $(p + 1) \times (q + 1)$ contingency table with $n_{ij}$ as the entry in the $(i, j)$-th call, and if we denote by $x_i, y_j$ the dummy variables defined by

$$
x_i = \begin{cases} 1, & \text{if a member belongs to the } i\text{-th row class,} \\ 0, & \text{otherwise } (i = 1, 2, \ldots, p + 1) \end{cases}
$$

$$
y_j = \begin{cases} 1, & \text{if a member belongs to the } j\text{-th column class,} \\ 0, & \text{otherwise } (j = 1, 2, \ldots, q + 1), \end{cases}
$$

and carry out a canonical analysis between $\boldsymbol{x}$ and $\boldsymbol{y}$, it is well known that (see, for example Kendall & Stuart, [3])

$$
n \sum_{1}^{p} r_i^2 = \chi^2 \tag{4.1}
$$

Where $\chi^2$ is the usual chi-square statistic for testing the independence of the row and column attributes, and $n$ is the total frequency. Here $r_i^2 (i = 1, \ldots, p)$ are the canonical correlations between $\boldsymbol{x}$ and $\boldsymbol{y}$, excluding the root 1, which is always present in such a case, since $\Sigma x_i = \Sigma y_j = 1$. If we quantify the row categories, according to some specified scores $a_1, a_2, \ldots, a_{p+1}$ and wish to find out how much residual $\chi^2$ is left out, for testing the goodness of these conjectured scores, we will have to eliminate two specified functions, one corresponding to the irrelevent scores $(1, 1, \ldots, 1)$ leading to the irrelevant root, and the other corresponding to the hypothetical scores $(a_1, a_2, \ldots, a_p)$ and use our above theory of residual canonical correlations to find $\phi_j^2$ and employ

$$
nS_1 = n(\Sigma r_i^2 - \Sigma \phi_j^2)
$$

to test the significance of the residual association and if this is insignificant, the hypothesis of goodness of fit of the proposed scores is tenable. Bartlett [1] gives a numerical illustration of such a situation, where blood serological data are analysed and the goodness of fit of equidistant scores like $(1, 2, \ldots, p)$ is tested.

In discriminant analysis with several groups, one can employ this technique to test the goodness of fit of 2 or more hypothetical discriminants. Williams [6] has given an example, where the relationship between certain physical variables about lamb carcasses and two factors, grade and weight is considered. In the response surface which

https://doi.org/10.4153/CMB-1985-003-2 Published online by Cambridge University Press

expresses this relationship, he tests the hypothesis of the significance of only the main effects of these two factors. The main-effects can be represented by three specified functions and the hypothesis then can be tested by the significance or otherwise of a statistic like $S_1$.

ACKNOWLEDGEMENTS. We are grateful to the referee who made some very valuable suggestions to improve the paper.

## REFERENCES

1. M. S. Bartlett, *Goodness of fit of a hypothetical discriminant function in the case of several groups*. Ann of Eugen, **16** (1951), 199–214.

2. F. Graybill, *Matrices with Application in Statistics*. Wadsworth Publishing Co., Belmont, CA 2nd Edition (1983), 183–185.

3. M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, Vol. 2. Charles Griffin & Co., Ltd. London (1967), 568–572.

4. A. M. Kshirsagar, *A note on the derivation of some exact tests*. Biometrika **47** (1960), 480–482.

5. E. J. Williams, *Some exact tests in multivariate analysis*. Biometrika **39** (1952), 17–22.

6. E. J. Williams, *The analysis of association among many variables*. J. Roy. Statist. Soc. B, Vol. 20 (1967), 199–220.

DEPARTMENT OF BIOSTATISTICS
UNIVERSITY OF MICHIGAN
ANN ARBOR, MI 48109

DEPARTMENT OF MATHEMATICS, STATISTICS AND C.S.
DALHOUSIE UNIVERSITY
HALIFAX, N.S. B3H 4H8