Rothamsted Repository Download

A - Papers appearing in refereed journals

Gower, J. C. 1966. A Q-technique for the calculation of canonical variates. *Biometrika.* 53 (3-4), pp. 588-590.

The publisher's version can be accessed at:

- https://dx.doi.org/10.1093/biomet/53.3-4.588

The output can be accessed at: https://repository.rothamsted.ac.uk/item/8w1w1.

© Please contact library@rothamsted.ac.uk for copyright queries.

By integrating out variables, $\{U_i\}$ is a set of mutually independent standardized normal variables and so is $\{W_i\}$. Any square summable function $g(U_1, U_2, ..., U_p)$ can be approximated arbitrarily closely by the product set of orthonormal functions on the joint distribution of the U_i. This is true on the combined distribution of the U_i and W_j since both $g(u_1, u_2, ..., u_p)$ and the approximating series are constant for a given set $\{u_1, u_2, ..., u_p\}$.

Let us write the members of the product set on the distribution of the U's as P_i. We might take $P_0, P_1, P_2, ...$ to be, first a constant term, then the $u_k^{(1)}$, then the $u_k^{(2)}$, then the products $u_k^{(1)} u_{k'}^{(1)} ....$ The $Q_0, Q_1, Q_2, ...$ may be similarly defined on the distribution of the W's. We now have

$$\begin{aligned} \xi_1 &= \Sigma a_i P_i, \quad a_0 = 0, \quad \Sigma a_i^2 = 1, \\ \eta_1 &= \Sigma b_j Q_j, \quad b_0 = 0, \quad \Sigma b_j^2 = 1, \end{aligned} \tag{3.16}$$

and have to maximize $E(\xi_1 \eta_1)$. Note that

$$E\left\{ \prod_{i=1}^{p} u_i^{(j_i)} \prod_{i=1}^{q} w_i^{(k_i)} \right\} = \prod_{i=1}^{p} E\left\{ u_i^{(j_i)} w_i^{(k_i)} \right\} \prod_{i=p+1}^{q} E w_i^{(u)}, \tag{3.17}$$

since

$$\{U_1, W_1\}, \quad \{U_2, W_2\}, ..., \{U_p, W_p\}, \quad W_{p+1}, ..., W_q$$

are jointly independent.

But $E(u_i^{(j)} w_i^{(k)}) = \rho^j \delta_{jk}$ and so the only terms of (3.17) which do not vanish yield correlations of the form $\rho_1^{(j_1)} \rho_2^{(j_2)} ... \rho_p^{(j_p)}$. But of all such products ρ_1 is the greatest. If $\rho_k > \rho_1^2, \rho_2, \rho_3, ..., \rho_k$ are the next in order of magnitude.

By Lemma 1, the first pair of canonical variables is $U_1 \equiv u_1^{(1)}$, $W_1 \equiv w_1^{(1)}$, and the canonical correlation is ρ_1. Thus ξ_2 and η_2 must not contain any term in $u_1^{(1)}$ or $w_1^{(1)}$ respectively.

If $\rho_2 > \rho_1^2$ then the second pair of canonical variables is $U_2 \equiv u_2^{(1)}, W_2 \equiv w_2^{(1)}$. The process can be continued as long as $\rho_k > \rho_1^2$. If $\rho_p > \rho_1^2$, then there are p canonical variables which are linear in the U_i or W_i and hence in the X_i and Y_i respectively. In any case, at the $(p+1)$th or earlier choice of pair, it will be necessary to select $u_1^{(2)}$ and $w_1^{(2)}$ and the functions are no longer linear in the X_i and Y_i.

REFERENCES

HOTELLING, H. (1936). Relations between two sets of variables. *Biometrika* **28**, 321–77.

LANCASTER, H. O. (1957). Some properties of the bivariate normal distribution considered in the form of a contingency table. *Biometrika* **44**, 289–92.

LANCASTER, H. O. (1958). The structure of bivariate distributions. *Ann. Math. Statist.* **29**, 719–36; see also (1964), **35**, 1388.

MAUNG, K. (1942). Measurement of association in a contingency table with special reference to the pigmentation of hair and eye colours of Scottish school children. *Ann. Eugen., Lond.* **11**, 189–223.

SARMANOV, O. V. & ZAHAROV, V. K. (1960). Maximum coefficients of multiple correlation. *Dokl. Akad. Nauk. SSSR.* **130**, 269–71.

SCHWERDTFEGER, H. (1960). Direct proof of Lanczos's decomposition theorem. *Amer. Math. Mon.* **67**, 856–60.

A *Q*-technique for the calculation of canonical variates

BY J. C. GOWER

Rothamsted Experimental Station, Harpenden

SUMMARY

A *Q*-technique for evaluating canonical variates is described. It is shown to have computational, and statistical, advantages over the usual *R*-technique.

The canonical variates associated with k multivariate normal populations, based on v variates with common dispersion matrix Ω and means $\Gamma(k \times v)$, are linear combinations of the original variates. The coefficients ξ_i of the ith canonical variate are the elements of the latent vector associated with the ith largest latent root λ_i of the equation $|\Gamma'\Gamma - \lambda\Omega| = 0$, where the means are supposed measured from an

origin such that each column of Γ has zero sum. There are thus v sets of these coefficients and they can be taken as columns of a square matrix $\Xi = (\xi_1, \xi_2, ..., \xi_v)$. The latent root and vector relationship can be written

$$\Gamma'\Gamma\Xi = \Omega\Xi\Lambda_1, \tag{1}$$

where Λ_1 is the diagonal matrix of the latent roots.

The means Π of the canonical variates are given by

$$\Pi = \Gamma\Xi. \tag{2}$$

The rows of Π can be regarded as the co-ordinates of points representing the means, referred to orthogonal canonical variate axes, and the square of the ordinary Euclidean distance between two means γ_i and γ_j is $(\gamma_i - \gamma_j)\Xi\Xi'(\gamma_i - \gamma_j)'$. Thus if the vectors are scaled so that $\Xi\Xi' = \Omega^{-1}$, i.e. $\Xi'\Omega\Xi = I$, this distance becomes Mahalanobis's Δ^2_{ij}. The axes corresponding to the r largest latent roots maximize the total Δ^2 in r dimensions between all pairs of populations (Rao, 1952, p. 365) or, equivalently, the first r axes are the principal components of a set of k points, representing the means, whose $\frac{1}{2}k(k-1)$ squared Euclidean distances are the values of Δ^2 (Gower, 1966).

Alternatively, Π can be found directly as the matrix of latent vectors of the $k \times k$ symmetric matrix $\Theta = \Gamma\Omega^{-1}\Gamma'$ provided that the vectors are scaled to make $\Pi'\Pi = \Lambda_2$, the diagonal matrix of latent roots of Θ (Gower, 1966). Thus

$$\Theta\Pi = \Pi\Lambda_2. \tag{3}$$

The matrices Λ_1 and Λ_2 will have the same non-zero diagonal elements, there are min $(k-1, v)$ of these, and will have zeros appended as necessary so that Λ_1 is $v \times v$ and Λ_2 is $k \times k$. The matrix Ξ can be expressed in terms of Γ and Π by pre-multiplying both sides of (2) by Γ' and using (1) to give

$$\Gamma'\Pi = \Gamma'\Gamma\Xi = \Omega\Xi\Lambda_1.$$

Therefore
$$\Xi = \Omega^{-1}\Gamma'\Pi\Lambda_1^{-}, \tag{4}$$

where Λ_1^{-} is the generalized inverse of Λ_1 found by replacing all non-zero values by their reciprocals. In the special case when $\Gamma'\Gamma$ is non-singular, i.e. $v < k$,

$$\Xi = (\Gamma'\Gamma)^{-1}\Gamma'\Pi. \tag{5}$$

When calculating canonical variates, sample values must replace the population parameters. The pooled within population dispersion matrix W replaces Ω and the sample means G replace Γ. The usual method of computation is the R-technique derived from (1) and (2). It is then necessary to compute the vectors of $(G'G - \lambda W)x = 0$ which requires either the vectors of the $v \times v$ non-symmetric matrix $W^{-1}G'G$ or a special calculation to factorize W into components UU', where U is an upper triangular matrix (Ashton, Healy & Lipton, 1957). Both of these methods of computation are more complex than the Q-technique derived from (3) and (4) which needs only the roots and vectors of the $k \times k$ symmetric matrix $T = GW^{-1}G'$ and the standard matrix operations of inversion and multiplication. When $k < v$ the Q-technique has the additional advantage of requiring the vectors of a smaller matrix than the R-technique. The equation corresponding to (3) is

$$TP = PL_2, \tag{6}$$

whose solution gives P and L_2 and thus also L_1. The canonical variate coefficients X are then given by equation (7), which corresponds to (4)

$$X = W^{-1}G'PL_1^{-}, \tag{7}$$

or when $G'G$ is non-singular
$$X = (G'G)^{-1}G'P. \tag{8}$$

Other advantages of the Q-technique are that the formula $D^2_{ij} = t_{ii} + t_{jj} - 2t_{ij}$ gives a simple direct method of computing the generalized distances from the elements of T, and it is possible to take account of the bias in D^2 (Rao, 1952, p. 364). Thus a matrix E of D^2 values adjusted for bias by Rao's formula can be constructed. From this we form a new matrix F with elements $-\frac{1}{2}e_{ij}$ and finally define T_1 as a matrix with elements $t^{(1)}_{ij} = f_{ij} - f_{i.} - f_{.j} + f_{..}$ (Gower, 1966). This matrix T_1 can be analysed as T above.

Canonical variates are often used as discriminant functions derived by finding the linear functions that give stationary values to the ratio of between-population to within-population sums of squares. This requires the solution of

$$(B - \nu W)x_1 = 0; \tag{9}$$

here B is the usual sum of squares and products matrix between groups obtained by weighting each group mean by the sample size, whereas $G'G$ is the corresponding unweighted sum of squares and

products matrix. If canonical variates are derived from (9) to give a set of coefficients X_1 then scaling the vectors so that $X_1 X_1' = W^{-1}$ ensures that the squares of the distances between the means of the canonical variates are still equal to D_{ij}^2, but it is no longer generally true that the first r dimensions maximize the total D^2 in r dimensions. The degree of approximation depends on the difference in the sample sizes; when all samples are the same size the two methods give the same results. These observations are relevant when canonical variates are used mainly for descriptive purposes (see, for example, Ashton, Healy & Lipton, 1957), when it is reasonable to require a representation of the sample means which maximizes total D^2, using only a few dimensions.

The distribution functions of the diagonal elements of **B** are each proportional to χ^2, but this is not so for **G'G**, so that Bartlett's (1938) approximate test for the significance of canonical variates has an extra degree of approximation when **G'G** is used. Thus when **B** is used the D^2 interpretation becomes approximate but when **G'G** is used the significance test is less reliable.

REFERENCES

ASHTON, E. H., HEALY, M. J. R. & LIPTON, S. (1957). The descriptive use of discriminant functions in physical anthropology. *Proc. R. Soc.* B **146**, 552–72.

BARTLETT, M. S. (1938). Further aspects of the theory of multiple regression. *Proc. Camb. Phil. Soc.* **30**, 33–40.

GOWER, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **53**, 325–38.

RAO, C. R. (1952). *Advanced Statistical Methods in Biometric Research.* New York: John Wiley and Sons.

The non-null distribution of a statistic in principal components analysis

BY A. M. KSHIRSAGAR

Defence Science Laboratory, Delhi

SUMMARY

The non-null distribution is obtained of a test statistic for the null hypothesis that a matrix with a single non-isotropic principal component has that component in a prescribed direction.

1. INTRODUCTION

Suppose that one latent root of Σ, the dispersion matrix of p multinormal variables with zero means

$$\mathbf{x}' = [x_1, ..., x_p]$$

is σ_1^2 and the remaining roots $\sigma_2^2, ..., \sigma_p^2$ are all equal to $\sigma^2 < \sigma_1^2$. If the latent column vector corresponding to σ_1^2 is $\mathbf{h}_1$, the variable $\mathbf{h}_1' \mathbf{x}$ is called the single non-isotropic principal component of Σ, because the variance of any linear combination of the variables $\mathbf{x}$, orthogonal to $\mathbf{h}_1' \mathbf{x}$, is the same, namely σ^2. Practical cases, where one comes across such a situation, have already been described (Kshirsagar, 1961). Given that Σ admits such a single non-isotropic principal component and given a hypothetical direction vector $\mathbf{h}$, it is of interest to test whether the hypothetical function $\mathbf{h}'\mathbf{x}$ is the true non-isotropic principal component. In other words, one wishes to test the hypothesis

$$H_0 : \mathbf{h} \equiv \mathbf{h}_1. \tag{1·1}$$

The appropriate statistic for this is (Kshirsagar, 1961)

$$\chi_d^2 = \frac{\mathbf{h}' A^2 \mathbf{h}}{\mathbf{h}' A \mathbf{h}} - \mathbf{h}' A \mathbf{h}, \tag{1·2}$$

where

$$A = X'X,$$

and

$$X = [x_{ir}] \quad (i = 1, ..., p; r = 1, ..., n)$$