Journal of Classification 28:315-326 (2011)
DOI: 10.1007/s00357-011-9090-z

Between-Group Metrics

John C. Gower

The Open University, United Kingdom

Casper J. Albers

University of Groningen, The Netherlands

Abstract: In canonical analysis with more variables than samples, it is shown that, as well as the usual canonical means in the range-space of the within-groups dispersion matrix, canonical means may be defined in its null space. In the range space we have the usual Mahalanobis metric; in the null space explicit expressions are given and interpreted for a new metric.

Keywords: Between-group distances; Canonical analysis; Mahalanobis distance.

1. Introduction

In Canonical Variate Analysis measurements on each of p variables for n samples are distributed among k groups of sizes $n_1 + n_2 + \ldots + n_k = n$. These measurements are available in an $n \times p$ matrix $\mathbf{X}$, assumed column-centered, and therefore of rank at most $\min(n-1, p)$, with group-membership given in an $n \times k$ indicator matrix $\mathbf{G}$. Here, $g_{ij} = 1$ when the ith sample belongs to the jth group but otherwise $\mathbf{G}$ is zero. Thus $\mathbf{G1} = \mathbf{1}$ and $\mathbf{1}'\mathbf{G} = \mathbf{1}'\mathbf{N}$, where $\mathbf{N} = \mathrm{diag}(n_1, n_2, \ldots, n_k) = \mathbf{G}'\mathbf{G}$; we also write ${}_n\mathbf{H}_n = {}_n\mathbf{G}_k\mathbf{N}_k^{-1}\mathbf{G}'_n$ (idempotent). With this notation, the usual between and within-group orthogonal decomposition:

Authors' Addresses: J.C. Gower, Department of Mathematics and Statistics, The Open University, Walton Hall, Milton Keynes MK7 6AA, United Kingdom, e-mail: j.c.gower@open.ac,uk; C.J. Albers, Department of Psychometrics and Statistics, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands, e-mail: c.j.albers@rug.nl.

Published online 4 October 2011

$$ {}_n\mathbf{X}_p = {}_n\mathbf{G}_k\mathbf{N}_k^{-1}\mathbf{G}'_n\mathbf{X}_p + [\mathbf{I} - {}_n\mathbf{G}_k\mathbf{N}_k^{-1}\mathbf{G}'_n]_n\mathbf{X}_p = \mathbf{HX} + (\mathbf{I}-\mathbf{H})\mathbf{X} $$

has an associated analysis of variance

$$ \underbrace{{}_p\mathbf{X}'_n\mathbf{X}_p}_{\mathbf{T}} = \underbrace{{}_p\mathbf{X}'_n\mathbf{H}_n\mathbf{X}_p}_{\mathbf{B}} + \underbrace{{}_p\mathbf{X}'_n(\mathbf{I}-\mathbf{H})_n\mathbf{X}_p}_{\mathbf{W}}, $$

expressing that the Total sum-of-squares ($\mathbf{T}$) is the sum of the Between-Group sum-of-squares ($\mathbf{B}$) and the Within-Group sum-of-squares ($\mathbf{W}$). Note that the n rows of $\mathbf{HX}$ repeat the k different means $n_1, n_2, \ldots, n_k$ times; to get each mean only once, we require $\mathbf{N}^{-1}\mathbf{GX}$ which we write as $\bar{\mathbf{X}}$.

In classical canonical variate analysis, the spectral decomposition $\mathbf{W} = \mathbf{U}\boldsymbol{\Sigma}^2\mathbf{U}'$ underpins the transformation to canonical variables $\mathbf{XL}$ where $\mathbf{L} = \mathbf{U}\boldsymbol{\Sigma}^{-1}$. These define canonical means $\mathbf{HXL}$ with inner-products $(\mathbf{HXL})(\mathbf{HXL})' = \mathbf{HXW}^{-1}\mathbf{X}'\mathbf{H}$ that use the metric $\mathbf{LL}' = \mathbf{W}^{-1}$ to generate Mahalanobis distances between the canonical means; note that $\mathbf{L}'\mathbf{WL} = \mathbf{I}$. The rank of the canonical means is $k-1$ (or less) but they may be approximated in a smaller space, by using a conventional principal components analysis of $\mathbf{HXL}$. The two steps (i) define a metric, followed by (ii) a principal components analysis, are usually subsumed into a single two-sided eigenvalue calculation but the two-step process is better for understanding the following.

The above requires that $\mathbf{W}$ has full rank p. The case when $p > n$ is increasingly important where much of the interest is in overcoming computational difficulties, perhaps reducing the number of variables by identifying and rejecting those deemed irrelevant or by focussing on some form of functional multivariate analysis (see e.g. Krzanowski, Jonathan, McCarthy, and Thomas 1995; Mertens 1998). Here, we explore a novel structural property of canonical analysis that occurs when $p > n$. When $p > n$ then $\text{rank}\ \mathbf{W} = n-k$ and $\text{rank}\ \mathbf{T} = n-1$ and $\mathbf{W}$ does not have an ordinary inverse so the Mahalanobis metric is undefined. This need not be a major problem, because we may define the spectral decomposition

$$ \mathbf{W} = (\mathbf{U}_1, \mathbf{U}_0)\begin{pmatrix} \boldsymbol{\Sigma}^2 & \\ & \mathbf{0} \end{pmatrix}\begin{pmatrix} \mathbf{U}'_1 \\ \mathbf{U}'_0 \end{pmatrix}, $$

where $\boldsymbol{\Sigma}^2$ is a diagonal matrix of the non-zero eigenvalues of $\mathbf{W}$, $\mathbf{U}_1$ are the eigenvectors in the range space of $\mathbf{W}$ and $\mathbf{U}_0$ those in its null space. Then, we may define canonical means $\mathbf{HXL}$ where now $\mathbf{L} = \mathbf{U}_1\boldsymbol{\Sigma}^{-1}$ in the range space. No longer is $\mathbf{L}'\mathbf{WL} = \mathbf{I}$ but rather $\mathbf{L}'\mathbf{WL} = \mathbf{I}_{n-k}$. Then $(\mathbf{L}'\mathbf{WL})(\mathbf{L}'\mathbf{WL}) = \mathbf{I}_{n-k} = \mathbf{L}'\mathbf{WL}$. We may write this:

$$ \begin{pmatrix} \mathbf{L}' \\ \mathbf{U}'_0 \end{pmatrix} \mathbf{WLL}'\mathbf{W}\,(\mathbf{L}, \mathbf{U}_0) = \begin{pmatrix} \mathbf{L}' \\ \mathbf{U}'_0 \end{pmatrix} \mathbf{W}\,(\mathbf{L}, \mathbf{U}_0), $$

which, because $(\mathbf{L}, \mathbf{U}_0)$ is non-singular, gives $\mathbf{W}(\mathbf{LL}')\mathbf{W} = \mathbf{W}$ showing that the metric $\mathbf{LL}'$ is now a generalized inverse, rather than an inverse, of $\mathbf{W}$. With this minor change, we may proceed as before with a principal components analysis. The range space solution is a natural extension of classical canonical variate analysis and is reasonably well known (see e.g. Krzanowski et al. 1995; Rao and Yanai 1979). An interesting thing is that canonical means may also be defined in the null space and it is this solution that is the focus of interest in the following. This solution is less well known though it was noted by Krzanowski et al. (1995) and termed by them "zero variance discrimination". The null space solution follows from noting that the null vectors satisfy:

$$\mathbf{X}'(\mathbf{I} - \mathbf{H})\mathbf{X}\mathbf{U}_0 = 0\,,$$

and so

$$\mathbf{X}\mathbf{U}_0 = \mathbf{H}\mathbf{X}\mathbf{U}_0. \tag{1}$$

Note that the k different means are repeated $n_1, n_2, \ldots, n_k$ times in the n rows of both $\mathbf{XU}_0$ and, equivalently, $\mathbf{HXU}_0$. Being null vectors of $\mathbf{W}$, the canonical variables $\mathbf{XU}_0$ have zero variability within groups, but the corresponding canonical means $\mathbf{HXU}_0$ have non-zero sums-of-squares. Evidently, the computation of $\mathbf{HXU}_0$ is straightforward, as is any subsequent principal components analysis; see e.g. Albers and Gower (2010), where a practical application of the null space solution may be found. For a fuller understanding it is interesting to ask what functional form, analogous to Mahalanobis distance in the range space, is taken by the distance d_{ij} between the ith and jth canonical means in the null space of $\mathbf{W}$. This is our main objective below, but first we have to address a minor but troublesome technical matter.

The total dispersion $\mathbf{T} = \mathbf{X}'\mathbf{X}$ has rank $n-1$ so implying an extensive null space of rank $p-n+1$; this null space is also common to the null spaces of $\mathbf{B}$ and $\mathbf{W}$. This common null space is uninteresting; we are concerned only with the additional null spaces of $\mathbf{W}$ and $\mathbf{B}$ that are in the range space of $\mathbf{T}$, especially the intersection of the range space of $\mathbf{T}$ and the null space of $\mathbf{W}$ which normally has dimension/rank $k-1$. To simplify the following development we assume that the common null space has been eliminated by taking the spectral decomposition $\mathbf{T} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}'$ and redefining $\mathbf{X}$ as $\mathbf{XV}$. Throughout the following, we assume that $\mathbf{X}$ has been so redefined, now $\mathbf{X}'\mathbf{X} = \boldsymbol{\Lambda}$.

This initialization to give $\mathbf{X}$ with $n-1$ columns, eliminates the common null space from the dispersion matrices $\mathbf{T}$, $\mathbf{B}$ and $\mathbf{W}$. However, it does not remove null items from $\mathbf{X}$ itself. Indeed the vector $\mathbf{1}$, which eliminates the general mean, is one such null vector and is what gives rise to the rather extensive algebraic manipulations required in the following. Any null linear combinations among the rows of $\mathbf{X}$ would generate additional null vectors in the common null space. The position is complicated, because such linear combinations may be of two, not mutually exclusive, kinds (i) linear combinations within groups and (ii) linear combinations among the group means. Loss of rank within groups merely reduces the number of columns of the redefined $\mathbf{X}$ but to handle all variants that include (ii) is not trivial and would greatly extend this short paper. Therefore, apart from some passing concluding remarks, throughout the following, we assume that $\operatorname{rank} \bar{\mathbf{X}} = k-1$ and that $\operatorname{rank} \mathbf{X} = n-1$.

2. Derivation of d_{ij}^2

From here on we shall be working in the null space of $\mathbf{W}$ so we drop the suffix from $\mathbf{U}_0$. Starting from $\mathbf{XU} = \mathbf{HXU}$ for the null-vectors of $\mathbf{W}$, as in (1), we have that

$$\begin{aligned} {}_n\mathbf{X}_{n-1}\mathbf{U}_{k-1} &= \mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{XU} \\ &= {}_n\mathbf{G}_k\mathbf{A}_{k-1}\,, \end{aligned} \tag{2}$$

where ${}_k\mathbf{A}_{k-1} = \bar{\mathbf{X}}\mathbf{U}$ are the k group-mean coordinates given in repeated form in $\mathbf{XU}$.

Thus, the rows of $\mathbf{A} = (\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{XU}$ give coordinates that generate the distance d_{ij} between each pair of group-mean coordinates. To derive an explicit expression for d_{ij}^2, we do not require $\mathbf{A}$ itself, which has the usual rotational indeterminacy, but only $\mathbf{AA}'$. Then, $d_{ij}^2 = (\mathbf{AA}')_{ii} + (\mathbf{AA}')_{jj} - 2(\mathbf{AA}')_{ij}$.

From (2) we have

$$\begin{aligned} \mathbf{U}'\mathbf{U} &= ((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{GA})'((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{GA}) \\ &= \mathbf{A}'\mathbf{PA} \text{ (say)}, \end{aligned}$$

where $\mathbf{P} = \mathbf{G}'\mathbf{X}(\mathbf{\Lambda}^{-2})\mathbf{X}'\mathbf{G}$ with $\mathbf{X}'\mathbf{X} = \mathbf{\Lambda}$. So from the orthogonality of $\mathbf{U}$, our problem is to solve $\mathbf{A}'\mathbf{PA} = \mathbf{I}$ for $\mathbf{AA}'$. The solution would be trivial if $\mathbf{AA}'$ and $\mathbf{P}$ were not singular, but $\mathbf{1}'\mathbf{NA} = \mathbf{1}'\mathbf{XU} = \mathbf{0}$ and $\mathbf{1}'\mathbf{P} = \mathbf{1}'\mathbf{G}'\mathbf{X}(\mathbf{\Lambda}^{-2})\mathbf{X}'\mathbf{G} = \mathbf{1}'\mathbf{G}'\mathbf{X}(\mathbf{\Lambda}^{-2})\mathbf{X}'\mathbf{G} = \mathbf{0}$, showing that both $\mathbf{AA}'$ and $\mathbf{P}$ are singular (with rank $k-1$). To proceed, consider

$$\begin{pmatrix} \mathbf{A}'\mathbf{N} \\ \frac{1}{n}\mathbf{1}'\mathbf{N} \end{pmatrix} (\mathbf{Q} + \lambda\mathbf{11}')(\mathbf{NA}, \tfrac{1}{n}\mathbf{N1})\,, \tag{3}$$

where $\mathbf{1}'\mathbf{NQ} = \mathbf{1}'\mathbf{XPN}^{-1} = \mathbf{0}$ shows that $\mathbf{Q} = \mathbf{N}^{-1}\mathbf{PN}^{-1} = \bar{\mathbf{X}}\mathbf{\Lambda}^{-2}\bar{\mathbf{X}}'$. The introduction of λ may seem arbitrary but we shall show that it has no substantive effect. On expansion, (3) becomes

$\begin{pmatrix} \mathbf{I} & 0 \\ 0 & \lambda \end{pmatrix}$ giving:

$$\mathbf{Q} + \lambda \mathbf{1}\mathbf{1}' = \begin{pmatrix} \mathbf{A}'\mathbf{N} \\ \frac{1}{n}\mathbf{1}'\mathbf{N} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{I} & \\ & \lambda \end{pmatrix} (\mathbf{NA}, \tfrac{1}{n}\mathbf{N1})^{-1}$$

and

$$\left(\mathbf{Q} + \lambda \mathbf{1}\mathbf{1}'\right)^{-1} = (\mathbf{NA}, \tfrac{1}{n}\mathbf{N1}) \begin{pmatrix} \mathbf{I} & \\ & \frac{1}{\lambda} \end{pmatrix} \begin{pmatrix} \mathbf{A}'\mathbf{N} \\ \frac{1}{n}\mathbf{1}'\mathbf{N} \end{pmatrix}.$$

Thus,

$$\mathbf{N}^{-1}\left(\mathbf{Q} + \lambda \mathbf{1}\mathbf{1}'\right)^{-1}\mathbf{N}^{-1} = \mathbf{A}\mathbf{A}' + \mathbf{1}\mathbf{1}'/\lambda n^2. \tag{4}$$

From (4) we may calculate d_{ij}^2. The constant term $\mathbf{1}\mathbf{1}'/\lambda n^2$ has no effect on derived distances and we shall show that $(\mathbf{Q} + \lambda\mathbf{1}\mathbf{1}')^{-1}$ also is invariant to non-zero choices of λ. Thus (4) contains everything needed for finding $\mathbf{A}\mathbf{A}'$ but the evaluation of $(\mathbf{Q}+\lambda\mathbf{1}\mathbf{1}')^{-1}$ needs some care, because $\mathbf{1}'\mathbf{N}\mathbf{Q} = \mathbf{1}'\mathbf{X}\mathbf{P}\mathbf{N}^{-1} = \mathbf{0}$ shows that $\mathbf{Q}$ is singular and the equivalent of (A.4) is unavailable.

The technical details for inverting $\mathbf{Q} + \lambda\mathbf{1}\mathbf{1}'$ are rather extensive and have been exiled to Appendix B, where (B.7) gives our main result

$$d_{12}^2 = \frac{(n_1+n_2)^2}{n_1^2 n_2^2}\left[(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{\Lambda}^{-2}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) - (\mathbf{q}_1 - \mathbf{q}_2)'\mathbf{Q}_{12}^{-1}(\mathbf{q}_1 - \mathbf{q}_2)\right]^{-1}. \tag{5}$$

where $\mathbf{Q}_{12}$ is the matrix formed from $\mathbf{Q}$ by striking out its first two rows and columns and $\mathbf{q}_1$ and $\mathbf{q}_2$ are vectors of the final $k-2$ elements in the first two columns of $\mathbf{Q}$. We give the result for the first two groups but clearly it generalizes trivially to give the distance d_{ij} between groups i and j.

3. Interpretation

When $k = 2$ the vectors $\mathbf{q}_1$ and $\mathbf{q}_2$ of (5) do not exist and $\mathbf{C}$, as defined in Appendix B, becomes

$$\mathbf{C} = \begin{pmatrix} \mathbf{c}_{11} & \mathbf{c}_{12} \\ \mathbf{c}_{21} & \mathbf{c}_{22} \end{pmatrix},$$

and its inverse is immediately replacing (5) by

$$d_{12}^{-2} = \frac{n_1^2 n_2^2}{(n_1+n_2)^2}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{\Lambda}^{-2}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2). \tag{6}$$

We next examine the part of expression (5) that is enclosed in square brackets. We have

$$\bar{X} = \begin{pmatrix} \bar{\mathbf{x}}_1 \\ \bar{\mathbf{x}}_2 \\ \hline \bar{\mathbf{X}}_{12} \end{pmatrix} \quad \text{and} \quad \left.\begin{array}{rcl} (\mathbf{q}_1 - \mathbf{q}_2) & = & (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)\mathbf{\Lambda}^{-2}\bar{\mathbf{X}}_{12}' \\ \mathbf{Q}_{12} & = & \bar{\mathbf{X}}_{12}\mathbf{\Lambda}^{-2}\bar{\mathbf{X}}_{12}' \end{array}\right\}.$$

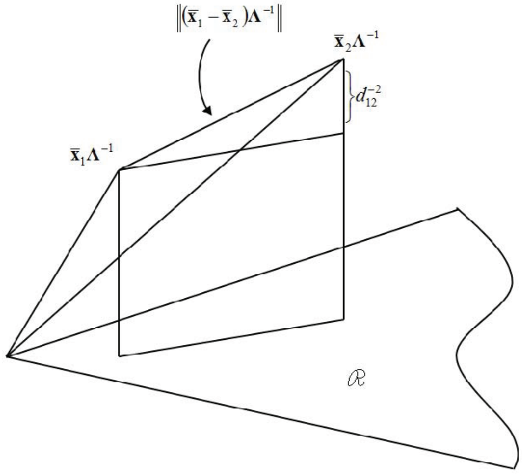

Figure 1. $\mathscr{R}$ is the space spanned by the means $\bar{\mathbf{x}}_3,\ldots,\bar{\mathbf{x}}_k$. Apart from the factor $n_1n_2/(n_1+n_2)$, the inverse of the distance between groups 1 and 2 in the intersection space is given by the illustrated projection onto the space orthogonal to $\mathscr{R}$.

Hence,

$$d_{12}^{-2} = \frac{n_1^2n_2^2}{(n_1+n_2)^2}\left[(\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2)'\mathbf{\Lambda}^{-2}(\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2)-(\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2)'\mathbf{\Lambda}^{-1}\mathbf{R}\mathbf{\Lambda}^{-1}(\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2)\right]$$

where $\mathbf{R}=\mathbf{\Lambda}^{-1}\bar{\mathbf{X}}_{12}'(\bar{\mathbf{X}}_{12}\mathbf{\Lambda}^{-2}\bar{\mathbf{X}}_{12}')^{-1}\bar{\mathbf{X}}_{12}\mathbf{\Lambda}^{-1}$ represents orthogonal projection in the metric $\mathbf{\Lambda}$ onto the space spanned by the $k-2$ rows of $\bar{\mathbf{X}}_{12}\mathbf{\Lambda}^{-1}$. Thus

$$d_{12}^{-2} = \frac{n_1^2n_2^2}{(n_1+n_2)^2}\left[(\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2)'\mathbf{\Lambda}^{-1}(\mathbf{I}-\mathbf{R})\mathbf{\Lambda}^{-1}(\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2)\right]. \qquad (7)$$

Thus, expressions (5), with its variants (6) and (7) are our main results.

In the more usual range space solution, inter-group distance is a Mahalonobis distance. In the intersection space solution discussed here, d_{12}^{-1}, especially in the special case (6), is seen to have some of the features of a Mahalonobis distance but with $\mathbf{W}$ replaced by $\mathbf{\Lambda}^2=\mathbf{T}^2$, the square of the total variation. Of course d_{12}^{-1} is not a Mahalonobis distance; formally it is not even a distance. Apart from the inverse relationship, we have $\mathbf{T}$ occuring in squared form. This is not surprising as a dimensional analysis shows: $\mathbf{R}$ is dimensionless, d_{12} has dimension L (for length) as do $\bar{\mathbf{x}}_1$ and $\bar{\mathbf{x}}_2$, while $\mathbf{\Lambda}$ has the dimensions of $\mathbf{T}=\mathbf{X}'\mathbf{X}$. Together, these show that the right and left side of (7) match correctly. The interpretation of (7), visualized in Figure 1, is that d_{12}^{-1} is a measure of how far the space of $\bar{\mathbf{x}}_1,\bar{\mathbf{x}}_2$ is from the space spanned by $\bar{\mathbf{x}}_3,\ldots,\bar{\mathbf{x}}_k$.

If $\bar{\mathbf{x}}_1$ and $\bar{\mathbf{x}}_2$ lie in $\mathscr{R}$ then (7) gives a null projection onto $\mathscr{R}^{\perp}$ but this case is excluded by our assumption that $\operatorname{rank} \bar{\mathbf{X}} = k - 1$. An indication of the corresponding results when $\operatorname{rank} \bar{\mathbf{X}} < k - 1$ is given by the collinearity case $k = 3$ and $\operatorname{rank} \bar{\mathbf{X}} = 1$. Then, without giving the detailed derivation, it can be shown that d_{12}^{-2} is proportional to $(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{\Lambda}^{-2}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$. That the projection term vanishes is consistent and plausible but we do not know whether it generalizes to other reduced rank cases.

4. Discussion

A referee remarked that the intersection space (i.e. the intersection of the range space of $\mathbf{T}$ with the null space of $\mathbf{W}$) is not so interesting for canonical analysis. Perhaps utility rather than interest was behind the remark; in a statistical context, something which is not useful might be regarded as uninteresting. Whatever was intended, it is true that the linked concepts of interest and utility deserve consideration.

We believe that the existence of the intersection space makes it interesting and its properties worthy of investigation. Such a study is an essential prerequisite for assessing utility. Our explicit formulae for d_{ij}^2 are interesting, with interesting interpretations. We apologize for the fairly heavy algebra used; the elegance of the results suggests that there ought to be a neater derivation. We have seen that the d_{ij}^{-2} (note the inverse) has some resemblance to Mahalanobis distance, using the total variation $\mathbf{T}$ rather than within-group variation $\mathbf{W}$. Unlike the Mahalonobis distance this quantity is not dimensionless, indicating the desirability of prior normalizations. As there is no within-group variation, there is complete separation between the groups. Is this separation meaningful? It might be thought that the distance depended on the vagaries of the number p of variables and that just by increasing p, different intersection space distances would be found. But, the size $k - 1$ of the intersection space remains unchanged and perhaps other properties of the intersection space remain invariant. Krzanowski et al. (1995) compared discrimination as effected by a variety of methods, mainly based on the range space of $\mathbf{W}$ but also using the intersection space. They found that the intersection space (Zero-variance Discrimination in their terminology) could behave well. Our own preliminary numerical results also suggest that distance in the intersection space is worth considering (Albers and Gower 2010). We suggest that performance depends on whether $\mathbf{W}$ contains variables with very small relative normalized variation. Such variables will be good discriminators and in the limit will become part of the intersection space. A variable with truly zero within group variation will be an ideal discriminator, provided the corresponding between group variation is not itself null.

To sum up, the intersection space is interesting and there is evidence that it could be useful. Research is now needed on invariance properties of the intersection space.

Appendix A
Basic Formulae

This appendix contains some basic results for convenience of reference. Most are well-known and, apart from (A.1) and A.2), no derivations are supplied.

Using $\mathbf{1}'\mathbf{NQ} = \mathbf{0}$ shows that

$$\begin{aligned} n_1 q_{11} + n_2 q_{12} + \mathbf{1}'\mathbf{M}\mathbf{q}_1 &= 0\,, \\ n_1 q_{12} + n_2 q_{22} + \mathbf{1}'\mathbf{M}\mathbf{q}_2 &= 0\,, \\ n_1 \mathbf{q}_1 + n_2 \mathbf{q}_2 + \mathbf{Q}_{12}\mathbf{M}\mathbf{1} &= 0\,, \end{aligned}$$

where $\mathbf{N} = \mathrm{diag}(n_1, n_2, \mathbf{M})$. It follows that

$$n_1^2 q_{11} + n_2^2 q_{22} + 2 n_1 n_2 q_{12} = -\mathbf{1}'\mathbf{M}(n_1\mathbf{q}_1 + n_2\mathbf{q}_2) = \mathbf{1}'\mathbf{M}\mathbf{Q}_{12}\mathbf{M}\,. \quad \text{(A.1)}$$

Similarly,

$$\begin{aligned} n_1\mathbf{c}_1 + n_2\mathbf{c}_2 &= n_1\mathbf{q}_1 + n_2\mathbf{q}_2 + \lambda(n_1+n_2)\mathbf{1} \\ &= -\mathbf{Q}_{12}\mathbf{M}\mathbf{1} + \lambda(n_1+n_2)\mathbf{1} \\ &= -\mathbf{C}_{12}\mathbf{M}\mathbf{1} + \lambda n\mathbf{1}. \end{aligned} \quad \text{(A.2)}$$

Furthermore, we need

$$\det\begin{pmatrix} \alpha & a' \\ a & \mathbf{A} \end{pmatrix} = \left(\alpha - a'\mathbf{A}^{-1}a\right)\det\mathbf{A}, \quad \text{(A.3)}$$

$$\det\mathbf{C}_{12} = \det(\mathbf{Q}_{12} + \lambda\mathbf{1}\mathbf{1}') = \det\mathbf{Q}_{12}\left(1 + \lambda\mathbf{1}'\mathbf{Q}_{12}^{-1}\mathbf{1}\right), \quad \text{(A.4)}$$

and

$$\mathbf{C}_{12}^{-1} = (\mathbf{Q}_{12} + \lambda\mathbf{1}\mathbf{1}')^{-1} = \mathbf{Q}_{12}^{-1} - \frac{\lambda\mathbf{Q}_{12}^{-1}\mathbf{1}\mathbf{1}'\mathbf{Q}_{12}^{-1}}{1 + \lambda\mathbf{1}'\mathbf{Q}_{12}^{-1}\mathbf{1}}. \quad \text{(A.5)}$$

Appendix B
The Inverse of $\mathbf{Q} + \lambda\mathbf{1}\mathbf{1}'$ and the Derivation of Explicit Formulae for d_{ij}^2

For simplicity, we derive d_{12}^2, the other values of d_{ij}^2 following by symmetry. Notation is established in the following partitioning of the matrix $\mathbf{C} = \mathbf{Q} + \lambda\mathbf{1}\mathbf{1}'$, written explicitly as

$$\mathbf{C} = \left(\begin{array}{c|c|c} c_{11} & c_{21} & \mathbf{c}_1' \\ \hline c_{12} & c_{22} & \mathbf{c}_2' \\ \hline & & \\ \mathbf{c}_1 & \mathbf{c}_2 & \mathbf{C}_{12} \\ & & \end{array}\right),$$

where

$$\begin{cases} \mathbf{C} &= \mathbf{Q} + \lambda\mathbf{1}\mathbf{1}', \\ \mathbf{C}_{12} &= \mathbf{Q}_{12} + \lambda\mathbf{1}\mathbf{1}', \\ \mathbf{c}_1 &= \mathbf{q}_1 + \lambda 1\,, \\ \mathbf{c}_2 &= \mathbf{q}_2 + \lambda 1\,, \\ (c_{11}, c_{12}, c_{22}) &= (q_{11}, q_{12}, q_{22}) + \lambda\,. \end{cases}$$

From (4) we obtain

$$\begin{aligned} \Delta d_{12}^2 = {} & \frac{1}{n_1^2} \det \begin{pmatrix} c_{22} & \mathbf{c}_2' \\ \mathbf{c}_2 & \mathbf{C}_{12} \end{pmatrix} + \frac{1}{n_2^2} \det \begin{pmatrix} c_{11} & \mathbf{c}_1' \\ \mathbf{c}_1 & \mathbf{C}_{12} \end{pmatrix} \\ & + \frac{2}{n_1 n_2} \det \begin{pmatrix} c_{12} & \mathbf{c}_2' \\ \mathbf{c}_1 & \mathbf{C}_{12} \end{pmatrix}, \end{aligned} \qquad \text{(B.1)}$$

where $\Delta = \det \mathbf{C}$, and the determinants are the cofactors of c_{11}, c_{22} and c_{12}. Using (A.3), (B.1) becomes

$$\begin{aligned} \Delta d_{12}^2 = {} & \det \mathbf{C}_{12} \left[\frac{1}{n_1^2} \left(c_{22} - \mathbf{c}_2' \mathbf{C}_{12}^{-1} \mathbf{c}_2 \right) + \frac{1}{n_2^2} \left(c_{11} - \mathbf{c}_1' \mathbf{C}_{12}^{-1} \mathbf{c}_1 \right) \right. \\ & \left. + \frac{2}{n_1 n_2} \left(c_{12} - \mathbf{c}_1' \mathbf{C}_{12}^{-1} \mathbf{c}_2 \right) \right], \end{aligned}$$

which simplifies to

$$\begin{aligned} n_1^2 n_2^2 \Delta d_{12}^2 = {} & \det \mathbf{C}_{12} \Big[n_1^2 c_{11} + n_2^2 c_{22} + 2 n_1 n_2 c_{12} \\ & - (n_1 \mathbf{c}_1 + n_2 \mathbf{c}_2)' \mathbf{C}_{12}^{-1} (n_1 \mathbf{c}_1 + n_2 \mathbf{c}_2) \Big]. \end{aligned} \qquad \text{(B.2)}$$

Using (A.1) and (A.2) we have

$$\begin{aligned} & (n_1 \mathbf{c}_1 + n_2 \mathbf{c}_2)' \mathbf{C}_{12}^{-1} (n_1 \mathbf{c}_1 + n_2 \mathbf{c}_2) \\ & \quad = (\mathbf{C}_{12} \mathbf{M} \mathbf{1} - \lambda n \mathbf{1})' \mathbf{C}_{12}^{-1} (\mathbf{C}_{12} \mathbf{M} \mathbf{1} - \lambda n \mathbf{1}) \\ & \quad = \mathbf{1}' \mathbf{M} \mathbf{C}_{12} \mathbf{M} \mathbf{1} - 2 \lambda n \mathbf{1}' \mathbf{M} \mathbf{1} + \lambda^2 n^2 \mathbf{1}' \mathbf{C}_{12}^{-1} \mathbf{1}. \end{aligned} \qquad \text{(B.3)}$$

Bringing everything together using (B.3), (A.4) and (A.5), and expanding in terms of q_{ij}, (B.2) becomes

$$\begin{aligned} n_1^2 n_2^2 \Delta d_{12}^2 &= (1+\lambda \mathbf{1}'\mathbf{Q}_{12}\mathbf{1}) \det \mathbf{Q}_{12} \Big[n_1^2 q_{11} + n_2^2 q_{22} + 2 n_1 n_2 q_{12} \\ &\quad + \lambda(n_1+n_2)^2 - \mathbf{1}'\mathbf{M}\mathbf{Q}_{12}\mathbf{M}\mathbf{1} - \lambda(\mathbf{1}'\mathbf{M}\mathbf{1})^2 + 2\lambda n \mathbf{1}'\mathbf{M}\mathbf{1} \\ &\quad - \lambda^2 n^2 \mathbf{1}' \left(\mathbf{Q}_{12}^{-1} - \frac{\lambda \mathbf{Q}_{12}^{-1}\mathbf{1}\mathbf{1}'\mathbf{Q}_{12}^{-1}}{1+\lambda \mathbf{1}'\mathbf{Q}_{12}^{-1}\mathbf{1}} \right) \mathbf{1} \Big], \end{aligned}$$

which, on using (A.1), simplifies to

$$\begin{aligned} n_1^2 n_2^2 \Delta d_{12}^2 &= \det \mathbf{Q}_{12} \Big[\lambda(n_1+n_2)^2 + \lambda(n-n_1-n_2)(n+n_1+n_2) - \\ &\quad \frac{\lambda n^2 \mathbf{1}'\mathbf{Q}_{12}^{-1}\mathbf{1}}{1+\lambda \mathbf{1}'\mathbf{Q}_{12}^{-1}\mathbf{1}} \Big] (1+\lambda \mathbf{1}'\mathbf{Q}_{12}^{-1}\mathbf{1}) \\ &= \det \mathbf{Q}_{12} \left[\lambda n^2 \left(1+\lambda \mathbf{1}'\mathbf{Q}_{12}^{-1}\mathbf{1}\right) - \lambda^2 n^2 \mathbf{1}'\mathbf{Q}_{12}^{-1}\mathbf{1} \right] \\ &= \lambda n^2 \det \mathbf{Q}_{12}. \end{aligned} \tag{B.4}$$

This simple result shows that d_{ij}^2 is proportional to $\det \mathbf{Q}_{ij}$.

To show that d_{ij}^2 in (B.4) is independent of λ requires an analysis of $\Delta = \det(\mathbf{Q} + \lambda \mathbf{1}\mathbf{1}')$. Let $\mathbf{R} = \mathbf{Q} + \mathbf{1}\mathbf{1}'$. Then $\mathbf{I} = \mathbf{R}^{-1}\mathbf{Q} + \mathbf{R}^{-1}\mathbf{1}\mathbf{1}'$ and so $\mathbf{1}' I \mathbf{N}\mathbf{1} = \mathbf{1}' \left(\mathbf{R}^{-1}\mathbf{Q} + \mathbf{R}^{-1}\mathbf{1}\mathbf{1}'\right) \mathbf{N}\mathbf{1}$ and because $\mathbf{Q}\mathbf{N}\mathbf{1} = \mathbf{0}$, we have that $\mathbf{1}'\mathbf{N}\mathbf{1} = \left(\mathbf{1}'\mathbf{R}^{-1}1\right)(\mathbf{1}'\mathbf{N}\mathbf{1})$, showing that

$$\mathbf{1}'\mathbf{R}^{-1}\mathbf{1} = 1.$$

Then

$$\begin{aligned} \Delta &= \det(\mathbf{R} + (\lambda - 1)\mathbf{1}\mathbf{1}') \\ &= \det \mathbf{R} \left(1 + (\lambda-1)\mathbf{1}'\mathbf{R}^{-1}\mathbf{1}\right) \\ &= \lambda \det \mathbf{R} \\ &= \lambda \det(\mathbf{Q} + \mathbf{1}\mathbf{1}'). \end{aligned}$$

That $\mathbf{R} \neq \mathbf{0}$ is guaranteed by our assumption that $\bar{\mathbf{X}} = k - 1$ made at the end of Section 1.

Thus, finally, (B.4) becomes

$$d_{12}^2 = \frac{1}{\det(\mathbf{Q} + \mathbf{1}\mathbf{1}')} \frac{n^2}{n_1^2 n_2^2} \det \mathbf{Q}_{12}, \tag{B.5}$$

showing that d_{ij}^2 depends only on the group sizes and $\det \mathbf{Q}_{12}$. Recall that $\mathbf{Q} = \bar{\mathbf{X}}\mathbf{\Lambda}^{-1}\bar{\mathbf{X}}'$ and that $\mathbf{Q}_{12}$ is obtained from $\mathbf{Q}$ by striking out its first two rows and columns.

Alternative expressions

The expression for Δ has many forms. Of special interest is that derived from writing

$$\Delta = \det \left(\begin{array}{cc|c} c_{11} & c_{12} & \mathbf{c}_1' \\ c_{12} & c_{22} & \mathbf{c}_2' \\ \hline \mathbf{c}_1 & \mathbf{c}_2 & \mathbf{C}_{12} \end{array} \right).$$

Multiplying the first row by n_1 and then adding $\mathbf{q}_i$ $(i = 2, \ldots, k)$ times the other rows, replaces the first row by $n\lambda\mathbf{1}$. This shows that λ may be subtracted from rows $2, \ldots, k$ to give

$$n_1\Delta = \lambda n \det \left(\begin{array}{cc|c} \mathbf{1} & \mathbf{1} & \mathbf{1}' \\ q_{12} & q_{22} & \mathbf{q}_2' \\ \hline \mathbf{q}_1 & \mathbf{q}_2 & \mathbf{Q}_{12} \end{array} \right),$$

whence similar operations on the columns give

$$\begin{aligned} n_1^2\Delta &= \lambda n^2 \det \left(\begin{array}{c|cc} \mathbf{1} & \mathbf{1} & \mathbf{1}' \\ 0 & q_{22} & q_2' \\ \hline 0 & q_2 & \mathbf{Q}_{12} \end{array} \right) \\ &= \lambda n^2 \left(q_{22} - \mathbf{q}_2'\mathbf{Q}_{12}^{-1}\mathbf{q}_2 \right) \det \mathbf{Q}_{12}. \end{aligned}$$

Similar expressions may be derived by annihilating the second row/column and the first row and second column to give

$$\left. \begin{aligned} n_2^2\Delta &= \lambda n^2 \left(q_{11} - \mathbf{q}_1'\mathbf{Q}_{12}^{-1}\mathbf{q}_1 \right) \det \mathbf{Q}_{12} \\ n_1^2\Delta &= \lambda n^2 \left(q_{22} - \mathbf{q}_2'\mathbf{Q}_{12}^{-1}\mathbf{q}_2 \right) \det \mathbf{Q}_{12} \\ -n_1 n_2\Delta &= \lambda n^2 \left(q_{12} - \mathbf{q}_1'\mathbf{Q}_{12}^{-1}\mathbf{q}_2 \right) \det \mathbf{Q}_{12} \end{aligned} \right\}. \tag{B.6}$$

Combining, gives the symmetric form

$$\begin{aligned} (n_1 + n_2)^2\Delta &= \\ & \lambda n^2 \left[(q_{11} + q_{22} - 2q_{12}) - (\mathbf{q}_1 - \mathbf{q}_2)'\mathbf{Q}_{12}^{-1}(\mathbf{q}_1 - \mathbf{q}_2) \right] \det \mathbf{Q}_{12}, \end{aligned}$$

which, on substitution into (B.5) gives

$$\begin{aligned} d_{12}^2 &= \\ & \frac{(n_1 + n_2)^2}{n_1^2 n_2^2} \left[(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{\Lambda}^{-2}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) - (\mathbf{q}_1 - \mathbf{q}_2)'\mathbf{Q}_{12}^{-1}(\mathbf{q}_1 - \mathbf{q}_2) \right]^{-1}. \end{aligned} \tag{B.7}$$

Other substitutions for Δ given by (B.6) give alternative, less symmetric, expressions for d_{ij}^2.

References

ALBERS, C.J., and GOWER, J.C. (2010). "Canonical Analysis: Ranks, Ratios and Fits", available at http://www.gmw.rug.nl/∼casper

KRZANOWSKI, W.J., JONATHAN, P., MCCARTHY, W.V., and THOMAS, M.R. (1995), "Discriminant Analysis with Singular Covariance Matrices: Methods and Applications to Spectroscopic Data", *Journal of the Royal Statistical Society, Series C (Applied Statistics), 44*, 101–115.

MERTENS, B.J.A. (1998), “Exact Principal Components Influence Measures Applied to the Analysis of Spectroscopic Data on Rice”, *Journal of the Royal Statistical Society, Series C (Applied Statistics), 47*, 527–542.

RAO, C.R.,and YANAI, H.(1979), “General Definition and Decomposition of Projectors and Some Applications to Statistical Problems”, *Journal of Statistical Planning and Inference, 3*, 1–17