Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 52 (2008) 1908–1927
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Sliced mean variance–covariance inverse regression

Simon J. Sheather[a,*], Joseph W. McKean[b], Kimberly Crimin[c]

[a] *Department of Statistics, Texas A&M University, College Station, TX 77843, USA*
[b] *Department of Statistics, Western Michigan University, Kalamazoo, MI 49008, USA*
[c] *Pfizer, Groton, CT 06340, USA*

Received 13 September 2006; received in revised form 4 June 2007; accepted 12 June 2007
Available online 16 June 2007

Abstract

Inverse regression methods have gained popularity over the last 10 years or so. More recently these methods have been applied to the classification problem. Sliced inverse regression (SIR) is equivalent to linear discriminant analysis and as such it detects mean differences between the classes. Sliced average variance estimation (SAVE) is designed to detect differences between the means, variances and covariances of the predictors across the classes. However, in SAVE each difference in variance across the groups takes up a dimension, and hence SAVE can have difficulty detecting mean differences and covariance differences. In this paper, we propose a new data analytic method, called sliced mean variance–covariance regression (SMVCIR) which can readily detect both first and second order differences between the classes even when many variance differences exist. Further, our procedure is based on an ordering of the dimensions based on their relative importance which is quite beneficial to interpretation. In particular, we demonstrate that useful data-analytic information about mean and covariance differences can be obtained from SMVCIR when SAVE finds many dimensions due to variance differences. Finally, the advantages of SMVCIR over SIR and SAVE are exemplified using a new data set from the enology literature.

Keywords: Classification; Dimension reduction; Discrimination; QR-decomposition; SAVE; SIR; Visualization

1. Introduction

Consider data sets which consist of k predictors taken on n objects, where each object can fall into one of g mutually exclusive groups or categories. Suppose n_i objects fall into category i, $i = 1, 2, \ldots, g$. Our goal is to obtain a visualization of the data set which displays differences in location and variance across the groups and identifies the important relationships among the predictors, using as few predictors as possible.

In an award winning article Cook and Yin (2001) studied inverse regression methods in this classification setting. In particular they studied the properties of sliced inverse regression (SIR; Li, 1991) and sliced average variance estimation (SAVE; Cook and Weisberg, 1991). Cook and Yin (2001) show that SAVE is more comprehensive that SIR identifying differences in variances and covariances as well as means across the groups.

SIR is actually equivalent to linear discriminant analysis (Kent, 1991). Linear discriminant analysis best separates the data based on differences in group means, assuming a constant covariance matrix. This method obtains a set

* Corresponding author.
E-mail addresses: sheather@stat.tamu.edu (S.J. Sheather), joe@stat.wmich.edu (J.W. McKean), kimberly.crimin@pfizer.com (K. Crimin).

0167-9473/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2007.06.007

of discriminant directions based on a space of k-dimensional vectors consisting of standardized differences in sample means across groups. The visualization is obtained by plotting the data along these directions. Discriminant coordinates were used by Fisher (1936). An informative discussion of them can be found in Gnanadesikan (1977).

Data sets, though, often differ in more than location. Besides location differences, predictors may be heteroscedastic or have different covariance structures across groups. SAVE is designed to detect differences between the means, variances and covariances of the predictors across the classes. Cook and Critchley (2000) have shown that SAVE is more comprehensive, generally having the ability to capture a larger part of the central subspace than SIR. However, they also showed that this comes with a price, namely, with several predictors, a relatively straightforward structure that is manifest through the mean will be harder to detect with SAVE than with SIR. In addition, Zhu and Hastie (2003) found that SAVE can miss detecting important first order differences.

In this paper, we propose a new data-analytical method, called sliced mean variance–covariance regression (SMVCIR) which can readily detect both first and second order differences between the classes. Unlike SAVE, SMVCIR does not necessarily target the central subspace. Instead, by considering variance differences separately from other covariance differences, SMVCIR is able to detect mean and covariance differences when variance differences exist across many of the variables. In particular, we know of no other general procedure, other than SMVCIR, which can readily detect both mean and covariance differences when many variance differences exist. Further unlike SAVE, SMVCIR orders the differences using a QR decomposition with column pivoting of the left unitary matrix of the singular value decomposition (SVD) of the space of first and second order differences. This results in an ordering of the vectors in the super space in terms of relative size. For example, if the standardized differences in variances are dominant over standardized differences in locations or covariances, vectors of standardized variance differences will be ordered first. Using the SVD of the super space, we decide with a scree plot based on relative size of the singular values how many of these ordered vectors of differences to select in the space. A kernel is then constructed based on the selected differences. The eigenvectors from the kernel's spectral decomposition serve as the SMVCIR directions and plots of the data along these directions offer the user visualizations of the data. The ordering of the dimensions based on their relative importance produces fewer possible dimensions and thus guides the user to focus on the most important differences among the groups. The new data-analytic procedure offers great flexibility in that it allows the user to focus on mean differences, variance differences, or covariance differences as well as any combination of these three.

Section 2 contains a motivating example based on generated data which demonstrates the advantages of SMVCIR over both SIR and SAVE when variance differences exist across many of the variables along with other differences. In Section 3 we discuss SMVCIR in detail contrasting it with SIR and SAVE. The new ordering procedure is discussed in Section 4. Section 5 offers a theoretical discussion which shows the relationships among SIR, SMVCIR and SAVE. In particular, we identify the circumstances in which the SMVCIR space is a subset of the SAVE space. We further clarify these relationships using two theoretical (functional) examples. These easily predict the behaviors of these procedures on data and show that each difference in the variance of the predictors across the groups takes up a dimension in SAVE. Thus, variance differences can dominate SAVE while, in contrast, SMVCIR loads all the variance differences between two groups into a single contrast. Loading all the variances into a single direction and ordering dimensions based on their relative importance leads to SMVCIRs increased ability to detect mean and covariance differences between the groups when many variance differences exist. In these two cases, moving the focus away from the central subspace, enables the user to find an initial summary of the data based on a small number of dimensions which highlight mean, variance and covariance differences. In Section 6, we consider two real data sets. The first is the Pen Digit data studied by Zhu and Hastie (2003). The second example is based on a data set from the enology literature which has not yet been discussed in the statistics literature. Finally, Section 7 contains a confirmatory simulation study on the new ordering criteria.

2. Motivating example

We present an example consisting of generated data so that we can compare the procedures SIR, SAVE and SMVCIR when the true model is known. In this example, we generated 100 data points from two multivariate normal distributions each based on 10 variables. In group 1, the means are each 0 and the variance–covariance matrix is the identity matrix. In group 2, the means are each 0 and the variance–covariance matrix has the main diagonal $\{4, 9, 16, 25, 36, 49, 64, 81, 100, 121\}$ and all covariances 0 except that the covariance between the first two variables

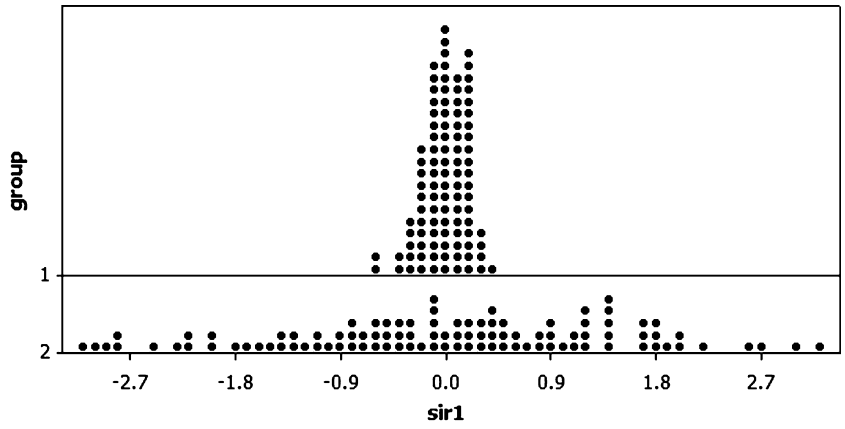

Fig. 1. Dot plot of the SIR dimension against group.

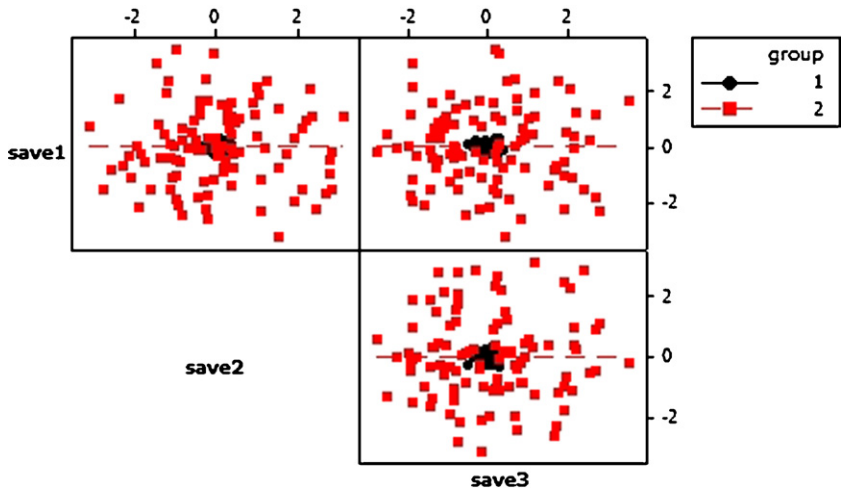

Fig. 2. Plot of the first three SAVE dimensions.

is 5.4. We used the `mvnorm` function of Venables and Ripley (2002) to generate data that has sample means, variances and covariances exactly equal to the population values.

Because $g = 2$, SIR can find at most one dimension, namely one due to a location difference. Fig. 1 shows a dot plot of the SIR dimension against group. Given that there is no location difference between the two groups, Fig. 1 is as expected.

The central subspace for this example is 10-dimensional. As it should, SAVE finds 10 highly significant dimensions each with a permutation test p-value less than 0.001. Dimensions 1–8 of SAVE identify individual variance differences between the groups for variables x_{10}–x_3, respectively (a plot of the first three SAVE dimensions is given in Fig. 2). The covariance difference between groups 1 and 2 is identified by dimensions 9 and 10 (a plot of the last three SAVE dimensions is given in Fig. 3). In practice, though, often only the first few directions are visualized, which in this case would result in missing the difference in covariances.

The new procedure SMVCIR identifies three dimensions which together account for 100% in the scree plot of singular values (see Section 4). The first dimension represents a difference in variance, the second and third dimension represent differences in covariance across the two groups. Fig. 4 shows a plot of the first three SMVCIR dimensions. Hence, unlike SAVE which finds this pertinent information in its last three dimensions (8, 9, and 10), SMVCIR finds it in its first three dimensions.

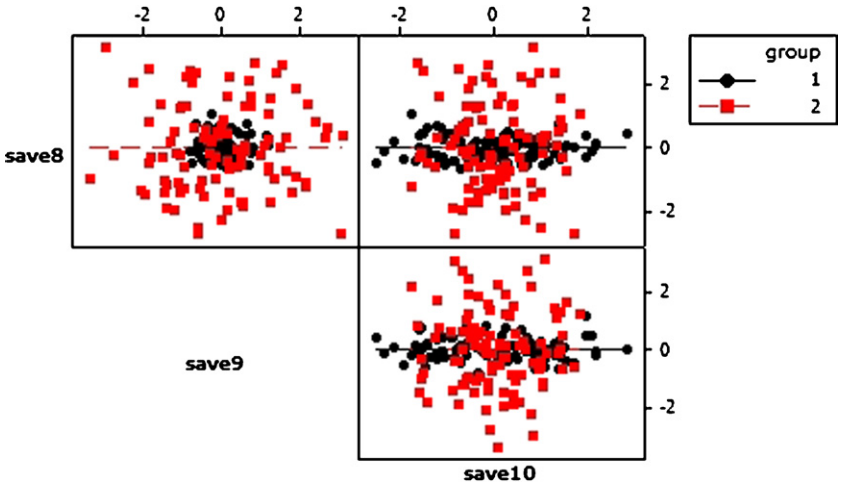

Fig. 3. Plot of the last three SAVE dimensions.

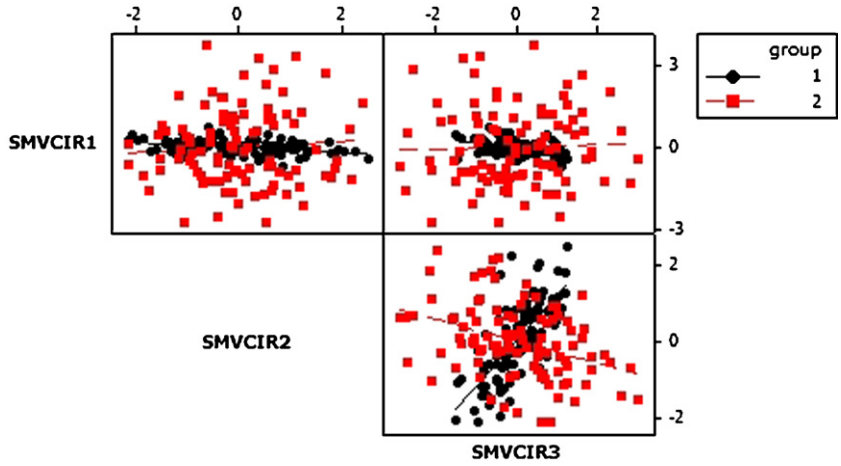

Fig. 4. Plot of the first three SMVCIR dimensions.

Following Cook and Yin (2001), we report standardized coefficients, which use predictors scaled to have standard deviation equal to one, and which are rescaled such that the sum of the squared coefficients add to one. Table 1 gives these standardized coefficients for the first three dimensions of SMVCIR. We see that dimension 1 is a weighted average of variables 2–10. Dimensions 2 and 3, on the other hand, contrast variables 1 and 2. Recalling how the data were generated, variable 2 is highly correlated with variable 1 in group 2. This example illustrates the advantages of moving the focus away from the central subspace in order to find an initial summary of the data based on a small number of dimensions. SMVCIR's three dimensions provide a convenient data-analytic summary, with the first dimension highlighting the variance differences between the groups and the second and third dimensions highlighting the covariance differences. Using simple plots (example, comparison boxplots), one can next identify the variables for which the variances differ across the groups.

In Section 5, we discuss situations as this numerical example but in general settings. We show that the differences among the three procedures, as illustrated in this example, are easily explained.

Table 1
Standardized coefficients of the first three SMVCIR dimensions for the generated data

Variable	SMVCIR_1	SMVCIR_2	SMVCIR_3
1	−0.082	0.865	0.129
2	0.249	−0.501	1.184
3	0.316	0.000	0.003
4	0.330	0.000	0.003
5	0.339	0.000	0.004
6	0.344	0.000	0.004
7	0.347	0.000	0.004
8	0.349	0.000	0.004
9	0.351	0.000	0.004
10	0.352	0.000	0.004

3. SMVCIR

Let $x_1, \ldots, x_k$ be a set of k continuous predictors measured over a set of g groups. For the ith group assume that we have a sample size of n_i of these measurements which constitute the rows of the $n_i \times k$ data matrix X_i. Let $n = \sum_{i=1}^{g} n_i$ denote the total sample size. Let $\overline{x}_i$ denote the $k \times 1$ vector of the sample means for group i and let $\widehat{\Sigma}_i$ denote the $k \times k$ sample covariance matrix. Take as the grand mean vector the usual weighted average of the $\overline{x}_i$'s, which is given by

$$\bar{x} = \sum_{i=1}^{g} \frac{n_i}{n} \overline{x}_i.$$

We define the pooled estimate of the covariance matrix by

$$\widehat{\Sigma}_{\text{p}} = \sum_{i=1}^{g} \frac{n_i - 1}{n - g} \widehat{\Sigma}_i.$$

Finally define the overall $n \times k$ data matrix to be $X = [X_1' X_2' \cdots X_g']'$.

Under the assumption of a common variance–covariance matrix among the groups, classical linear discriminant coordinates finds a visualization (coordinate system) which best separates the data into their groups based on the sample means. Recall (see for example Seber, 1984), that this maximum separation is obtained by maximizing

$$\max_{c \neq 0} \frac{c' \{\sum_{i=1}^{g} (n_i/n)(\overline{x}_i - \overline{x})(\overline{x}_i - \overline{x})'\} c}{c' \widehat{\Sigma}_{\text{p}} c}.$$

The matrix in braces in this expression is, of course, the between sums-of-squares matrix, while the matrix in the denominator is the within sums-of-squares matrix. The maximum is achieved by the normalized eigenvector c_{D1} corresponding to the maximum eigenvalue λ_{D1} of the spectral decomposition of the matrix

$$B_{\text{DISCRIM}} = \widehat{\Sigma}_{\text{p}}^{-1} \sum_{i=1}^{g} \frac{n_i}{n} (\overline{x}_i - \overline{x})(\overline{x}_i - \overline{x})'. \tag{3.1}$$

The vector c_{D1} is called the principal discriminant direction. Denote the k eigenvectors of this spectral decomposition by $\lambda_{\text{D1}} \geqslant \cdots \geqslant \lambda_{\text{D}k} \geqslant 0$ and let $c_{\text{D1}}, \ldots, c_{\text{D}k}$ denote the corresponding orthonormal eigenvalues. The vectors c_{D1} through $c_{\text{D}k}$ are called, respectively, the first through the kth principal discriminant directions. Let $C_{\text{D}} = [c_{\text{D1}} \cdots c_{\text{D}k}]$ and define the $n \times k$ matrix

$$W_{\text{D}} = X C_{\text{D}}. \tag{3.2}$$

The elements of the matrix W_{D} are called the discriminate coordinates. Scatterplots of the columns of W_{D} show the visualization of the data according to best separation of the data in terms of differences in group averages.

Note that the spectral decomposition of the related matrix

$$B_{\text{DISCRIM}} = \widehat{\Sigma}_{\text{p}}^{-1/2} \sum_{i=1}^{g} \frac{n_i}{n} (\overline{x}_i - \overline{x})(\overline{x}_i - \overline{x})' \widehat{\Sigma}_{\text{p}}^{-1/2}. \tag{3.3}$$

results in the same eigenvalues as the spectral decomposition of matrix (3.1). The corresponding orthonormal eigenvectors, though, need to be transformed. That is, if $v_1, \ldots, v_k$ are the eigenvectors of (3.3) then,

$$c_{\text{D}i} = \widehat{\Sigma}_{\text{p}}^{-1/2} v_i, \quad i = 1, \ldots, k.$$

Because matrix (3.3) is symmetric, as opposed to matrix (3.1), it is easier to work with numerically.

Recall that X is the $n \times k$ data matrix. Let $\overline{X}$ be the matrix of column averages; that is, each row of $\overline{X}$ is the grand mean vector $\overline{x}'$. Then more recently, Li (1991) proposed a coordinate representation based on the spectral decomposition of the matrix,

$$B_{\text{SIR}} = \widehat{\Sigma}_x^{-1/2} \sum_{i=1}^{g} \frac{n_i}{n} (\overline{x}_i - \overline{x})(\overline{x}_i - \overline{x})' \widehat{\Sigma}_x^{-1/2},$$

where $\widehat{\Sigma}_x = (X - \overline{X})'(X - \overline{X})/(n-1)$ is the total sums-of-squares matrix. The acronym SIR means sliced inverse regression. Li calls the matrix B_{SIR} the SIR kernel. Let $\lambda_{\text{SIR}1} \geqslant \cdots \geqslant \lambda_{\text{SIR}k} \geqslant 0$ denote the eigenvalues of the matrix B_{SIR} and let $v_{\text{SIR}1}, \ldots, v_{\text{SIR}k}$ denote the corresponding set of orthonormal eigenvectors. As with discriminate coordinates, the eigenvectors of interest are the transformed ones, i.e.,

$$c_{\text{SIR}i} = \widehat{\Sigma}_x^{-1/2} v_{\text{SIR}i}, \quad i = 1, \ldots, k.$$

Define the matrix $C_{\text{SIR}} = [c_{\text{SIR}i} \cdots c_{\text{SIR}k}]$. Then the k coordinate SIR directions are given by the columns of the matrix,

$$W_{\text{SIR}} = X C_{\text{SIR}}.$$

The components of W_{SIR} are called the SIR discriminant coordinates. Other than another standardization, the visualizations of the SIR discriminant coordinates are based on the differences in sample means and are usually quite similar to the visualizations based on (3.2); see Kent (1991) and Chen and Li (2001) for discussion.

The SIR procedure allows us to confine ourselves to *standardized variables*. Denote the standardized variables by $Z = \widehat{\Sigma}_x^{-1/2}[X - \overline{X}]$. Note that the group means and variance–covariance matrices of the standardized variables are given, respectively, by

$$\begin{aligned} \overline{z}_i &= \widehat{\Sigma}_x^{-1/2}(\overline{x}_i - \overline{x}), \\ \widehat{\Sigma}_{z,i} &= \widehat{\Sigma}_x^{-1/2} \widehat{\Sigma}_i \widehat{\Sigma}_x^{-1/2}, \end{aligned} \tag{3.4}$$

for $i = 1, 2, \ldots, g$. Denote the weighted average of sample variance–covariance matrices by

$$\overline{\Sigma}_z = \sum_{i=1}^{g} \frac{n_i}{n} \widehat{\Sigma}_{z,i}. \tag{3.5}$$

Define the $k \times 1$ vectors $\widehat{v}_i$ by, $\widehat{v}_i = \sqrt{n_i/n}\,\overline{z}_i$, $i = 1, \ldots, g$. Under this notation, the SIR kernel is given by $B_{\text{SIR}} = \sum \widehat{v}_i \widehat{v}_i^{\text{T}}$.

In addition to differences between means, we want our visualization procedure to be based on the coordinate system determined from differences between variances and covariances. Subspace notation clarifies these additions and is explicitly used in our ordering procedure. Let

$$S(\widehat{v}) = [\widehat{v}_2 \cdots \widehat{v}_g]. \tag{3.6}$$

Due to the constraint imposed by the sample means, only $g-1$ vectors are needed to span the column space. The space spanned by the columns of $S(\widehat{v})$ is called the space of SIR. We refer to this space as $\Omega(\widehat{v}) = \langle \widehat{v}_2 \cdots \widehat{v}_g \rangle$.

SIRs visualization is based on differences in sample means. We next incorporate differences in sample variances in the same way as means are incorporated. Denote the differences in variances by the vectors $\widehat{\delta}_i = \sqrt{n_i/n}(\widehat{\sigma}_{z,i}^2 - \overline{\sigma}_z^2)$,

$i = 1, \ldots, g$, where $\widehat{\sigma}^2_{z,i}$ is the main diagonal of the matrix $\widehat{\Sigma}_{z,i}$, (3.4) (the vector of sample variances for the ith group based on the standardized variables) and $\overline{\sigma}^2$ is the main diagonal of (3.5) (the weighted average of these vectors of sample variances). We thus expand the space $\Omega(\widehat{v})$, (3.6), to the space,

$$\Omega(\widehat{v}, \widehat{\delta}) = \langle \widehat{v}_2, \ldots, \widehat{v}_g, \widehat{\delta}_2, \ldots, \widehat{\delta}_g \rangle.$$

Means and variances are often the building blocks of statistics. But we also want our visualization procedure to detect differences in covariance relationships from group to group. Let

$$\widehat{\Delta}_i = \sqrt{\frac{n_i}{n}}[\widehat{\Sigma}_{z,i} - \overline{\Sigma}_z], \quad i = 1, \ldots, g.$$

We already have accounted for differences in variances; hence, the matrix we want is $\widehat{\Delta}_i$ minus its main diagonal; that is,

$$\widehat{\Delta}_i^{(0)} = \widehat{\Delta}_i - \mathrm{diag}\{\widehat{\delta}_i\}, \quad i = 1, \ldots, g.$$

Thus, the matrices $\widehat{\Delta}_i^{(0)}$'s contain only information concerning differences in covariances. Our final space is the expanded space,

$$\Omega_{\mathrm{SMVCIR}} = \Omega(\widehat{\Delta}^{(0)}, \widehat{v}, \widehat{\delta}) = \langle \widehat{\Delta}_2^{(0)}, \ldots, \widehat{\Delta}_g^{(0)}, \widehat{\delta}_2, \ldots, \widehat{\delta}_g, \widehat{v}_2, \ldots, \widehat{v}_g \rangle. \tag{3.7}$$

We have called the space SMVCIR, for sliced mean variance–covariance-inverse-regression.

The kernel of the SMVCIR space is

$$B_{\mathrm{SMVCIR}} = S_{\mathrm{SMVCIR}} S'_{\mathrm{SMVCIR}}, \tag{3.8}$$

where $S_{\mathrm{SMVCIR}} = [\widehat{\Delta}_2^{(0)} \cdots \widehat{\Delta}_g^{(0)} \widehat{\delta}_2 \cdots \widehat{\delta}_g \widehat{v}_2 \cdots \widehat{v}_g]$. As with the preceding methods, the coordinate directions are obtained from the spectral decomposition of B_{SMVCIR}. Let $\lambda_{\mathrm{SMVCIR},1} \geqslant \cdots \geqslant \lambda_{\mathrm{SMVCIR},k} \geqslant 0$ be the eigenvalues of this decomposition and $\{v_{\mathrm{SMVCIR},1}, \ldots, v_{\mathrm{SMVCIR},k}\}$ be the corresponding eigenvectors. Let $\{c_{\mathrm{SMVCIR},1}, \ldots, c_{\mathrm{SMVCIR},k}\}$ be the transformed eigenvectors; i.e., $c_{\mathrm{SMVCIR},i} = \widehat{\Sigma}_x^{-1/2} v_{\mathrm{SMVCIR},i}$, $i = 1, \ldots, k$. Define the $k \times k$ matrix C_{SMVCIR} to be $[c_{\mathrm{SMVCIR},1} \cdots c_{\mathrm{SMVCIR},k}]$. The SMVCIR coordinate directions are the columns of the matrix

$$W_{\mathrm{SMVCIR}} = X C_{\mathrm{SMVCIR}}.$$

The visualization procedure SAVE is described in a series of articles, including Cook and Weisberg (1991), Cook and Lee (1999), Cook and Yin (2001). The paper by Cook and Yin (2001) offers a review of SAVE and SIR and it is followed by the comments of several discussants. The SAVE space is given by

$$\Omega_{\mathrm{SAVE}} = \Omega(\widehat{\Delta}, \widehat{v}) = \langle \widehat{\Delta}_2, \ldots, \widehat{\Delta}_g, \widehat{v}_2, \ldots, \widehat{v}_g \rangle.$$

Note that in contrast to SAVE, SMVCIR allows the user to focus on mean differences, variance differences, and covariance differences as well as any combination of these three.

We next turn to ordering the vectors in the SMVCIR subspace based on their relative size or importance. Our visualization coordinate system is then based on vectors drawn from this ordering. This often results in a much smaller number of vectors as seen in the examples.

4. Ordering the columns of the SMVCIR space

The second part of our procedure is based on the ordering of the vectors in the SMVCIR space which we now describe in detail. It is based on the QR decomposition with column pivoting of the left unitary matrix of the SVD of the SMVCIR space $\Omega(\widehat{\Delta}^{(0)}, \widehat{v}, \widehat{\delta})$. Note that the vectors $\widehat{\delta}$ are the diagonal entries of the differences in the variance–covariance matrices which span the SAVE space. Hence, we are using the same scaled units as SAVE. This results in an ordering of the vectors in terms of relative size. For example, if the differences in locations are dominant over differences in variances or covariances, vectors of location differences will be ordered first. Hence, this ordering allows us to visualize the data using a smaller set of the vectors in the SMVCIR space, which still capture much of the SMVCIR space.

4.1. Ordering procedure based on the SVD of the SMVCIR space

Let $h = (g-1)(k+2)$. The SMVCIR space is the column space of the $k \times h$ matrix $S(\widehat{\Delta}^{(0)}, \widehat{\nu}, \widehat{\delta})$. It has rank less than or equal to k. We seek a *working space*, a subspace of the SMVCIR space consisting of r columns of the matrix $S(\widehat{\Delta}^{(0)}, \widehat{\nu}, \widehat{\delta})$. We will discuss the choice of r later, so, for now, assume that it is fixed. Based on r, we seek a permutation matrix Π such that

$$S(\widehat{\Delta}^{(0)}, \widehat{\nu}, \widehat{\delta})\Pi = [B_1 : B_2],$$

where B_1 is $k \times r$ matrix which will serve as our approximation to $S(\widehat{\Delta}^{(0)}, \widehat{\nu}, \widehat{\delta})$. We want a permutation matrix which selects a sufficiently independent set of r columns in B_1 that yield a good approximation to the spanning set.

Consider the SVD of $S(\widehat{\Delta}^{(0)}, \widehat{\nu}, \widehat{\delta})$ given by,

$$S(\widehat{\Delta}^{(0)}, \widehat{\nu}, \widehat{\delta}) = UDV',$$

where U and V are, respectively, $k \times k$ and $h \times h$ orthogonal matrices and D is the $k \times h$ matrix

$$D = [D_1 \quad O],$$

where $D_1 = \text{diag}\{\sigma_1, \ldots, \sigma_k\}$. The diagonal entries of D_1 are the singular values of the SMVCIR space. These are nonnegative and, without loss of generality, we can assume that $\sigma_1 \geqslant \cdots \geqslant \sigma_k \geqslant 0$.

Write the matrix V as

$$V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}, \tag{4.1}$$

and let

$$\Pi' V = \begin{bmatrix} \widetilde{V}_{11} & \widetilde{V}_{12} \\ \widetilde{V}_{21} & \widetilde{V}_{22} \end{bmatrix}, \tag{4.2}$$

where $\widetilde{V}_{11}$ is $r \times r$. We will often denote the jth singular value of a matrix B by $\sigma_j(B)$.

From a result in Golub and Van Loan (1996, Chapter 12) we have the inequality,

$$\frac{\sigma_r}{\|\widetilde{V}_{11}^{-1}\|_2} \leqslant \sigma_r(B_1) \leqslant \sigma_r, \tag{4.3}$$

where $\|\widetilde{V}_{11}^{-1}\|_2$ denotes the 2-norm of $\widetilde{V}_{11}^{-1}$.

Think of r as the working dimension of the spanning matrix $S(\widehat{\Delta}^{(0)}, \widehat{\nu}, \widehat{\delta})$. Recall that the idea was to select the permutation matrix Π so that we have a sufficiently independent set of columns in B_1 which yield a good approximation to the spanning set. One way of measuring this is to use the rth singular values. If the rth singular value of B_1, $\sigma_r(B_1)$, is close to $\sigma_r(S(\widehat{\Delta}^{(0)}, \widehat{\nu}, \widehat{\delta}))$, then this is an indication that B_1 yields a good approximation to the spanning set. Hence, we want the interval $(\sigma_r/\|\widetilde{V}_{11}^{-1}\|_2, \sigma_r)$, based on (4.3), as tight as possible. One way of obtaining this is to select the permutation matrix Π so that the matrix $\|\widetilde{V}_{11}^{-1}\|_2$ is as well conditioned as possible (i.e., its 2-norm is as close to 1 as possible).

A numerically efficient and stable way of obtaining a well-conditioned matrix $\widetilde{V}_{11}$ is to use a QR decomposition with column pivoting on the matrix $[V_{11}' : V_{21}']$. This results in

$$Q'[V_{11}' : V_{21}']\Pi = [R_{11} : R_{12}],$$

where R_{11} is $r \times r$ is an upper triangular matrix of rank r, Q is an $r \times r$ orthogonal matrix, and Π is an $h \times h$ permutation matrix obtained by the column pivoting. Using the notation in expressions (4.1) and (4.2), we have

$$\begin{bmatrix} \widetilde{V}_{11} \\ \widetilde{V}_{21} \end{bmatrix} = \Pi' \begin{bmatrix} V_{11} \\ V_{21} \end{bmatrix} = \begin{bmatrix} R_{11}'Q' \\ R_{12}'Q' \end{bmatrix}.$$

Hence,

$$\widetilde{V}_{11}^{-1} = Q R_{11}'^{-1}.$$

Because the 2-norm is invariant to orthogonal transformations we have,

$$\|\widetilde{V}_{11}^{-1}\|_2 = \|R_{11}'^{-1}\|_2 = \|R_{11}^{-1}\|_2.$$

Thus, the tightness of inequality (4.3) depends on how well conditioned the matrix R_{11} is.

QR decompositions with column pivoting are well known to produce well-conditioned matrices R_{11}. Discussions can be found in sources such as Golub and Van Loan (1996) and Stewart (1973). As described in detail in the Linpack manual (Dongarra et al., 1979, Chapter 9), the QR decomposition is obtained by applying Householder transformations to $[V_{11}' : V_{21}']$ one column at a time. Under column pivoting, the column with the largest norm is brought in first and the succeeding columns are updated and pivoted while the decomposition is being computed. For the computed matrix R this results in the bound,

$$r_{kk}^2 \geqslant \sum_{i=k}^{j} r_{ij}^2 \quad \text{for } j = k, k+1, \ldots, r. \tag{4.4}$$

We consider r_{jj} negligible if $r_{jj} < \varepsilon$. By (4.4), if $r_{jj} < \varepsilon$ then columns j and beyond are negligible also. This results in a well-conditioned R_{11} matrix.

4.2. The ordering procedure in practice

In practice, we initially take r to be k. This gives us an ordering of the columns of the SMVCIR space matrix $S(\widehat{\Delta}^{(0)}, \widehat{v}, \widehat{\delta})$, assuming that its rank is k. Now that we have an ordering of the h vectors, we need to determine how many to bring into the working subspace. This number is our working value r.

We have found a scree plot based on the singular values of the SMVCIR space to be most helpful in choosing r. Let

$$s = \sum_{i=1}^{k} \sigma_i$$

be the sum (average) of the singular values. Because the singular values are nonnegative, consider the relative percentage of their sum s explained by the first j singular values; i.e., the quantity,

$$s_j = 100 \sum_{i=1}^{j} \sigma_i / s.$$

Our scree plot is a plot of s_j versus j for $j = 1, \ldots, k$. In general, it rises steeply then levels off. We have found that where this leveling occurs is a good choice for r. This of course may not happen as in Example 6.2. In such cases, with parsimony in mind, we recommend entering as few vectors in the spanning set as possible which provide sufficient visualization of the data.

Once r has been determined, we take as our working space the $k \times r$ matrix B_1 consisting of the first r-ordered columns of the matrix $S(\widehat{\Delta}^{(0)}, \widehat{v}, \widehat{\delta})$. Our working kernel is then the matrix

$$B_{\mathrm{SMVCIR}}^{*} = B_1 B_1'.$$

As discussed in Section 3, we proceed by obtaining the spectral decomposition of this kernel. This results in the eigenvalues $\lambda_{\mathrm{SMVCIR},1}^{*} \geqslant \cdots \geqslant \lambda_{\mathrm{SMVCIR},k}^{*} \geqslant 0$. Of course, $k - r$ of these eigenvalues are theoretically zero. Denote the corresponding eigenvectors by $\{v_{\mathrm{SMVCIR},1}^{*}, \ldots, v_{\mathrm{SMVCIR},k}^{*}\}$ and the subsequent transformed eigenvectors by $\{c_{\mathrm{SMVCIR},1}^{*}, \ldots, c_{\mathrm{SMVCIR},k}^{*}\}$; i.e., $c_{\mathrm{SMVCIR},i}^{*} = \widehat{\Sigma}_x^{-1/2} v_{\mathrm{SMVCIR},i}^{*}$, $i = 1, \ldots, k$. Define the $k \times k$ matrix C_{SMVCIR}^{*} to be $[c_{\mathrm{SMVCIR},1} \cdots c_{\mathrm{SMVCIR},k}]$. Then our working coordinate directions are the columns of the matrix

$$W_{\mathrm{SMVCIR}}^{*} = X C_{\mathrm{SMVCIR}}^{*}.$$

As with other procedures described in Section 3, the resulting visualization procedure is based on the scatterplots of the columns of W^*_{SMVCIR}. Because r is generally much smaller than h, it is often easy to determine what groups and whether it is differences in variances, means, or covariances which contribute to the separations and patterns in the plots. As in procedures such as principal component analysis or canonical correlation analysis, consideration of the loadings on eigenvectors often shed light on which variables are involved in the separations.

5. Relationships among SIR, SAVE and SMVCIR

Obviously the population space associated with SIR is a subset of the population space associated with SMVCIR. Because the spanning spaces consist of k-dimensional vectors, if the SAVE space has dimension k, then the SMVCIR space is a subset of the SAVE space. Further, as shown in the Appendix, it follows that if the dimension of the SAVE space is less than k and the variance differences are not a linear combination of the covariance differences then the SMVCIR space is a subset of the SAVE space. Other then these cases, though, there is no general directional relationship between the SMVCIR and SAVE spaces.

In all the real examples that we have looked at, SMVCIR produces less dimensions than SAVE and as the theoretical examples below indicate, SMVCIR often identifies many fewer dimensions than SAVE. This enables the user to see mean differences, variance differences and covariance differences in relatively fewer dimensions.

On the other hand, consider the special case of two groups in which there are no differences in the means and no differences in the variances. In this case SAVE and SMVCIR will have the same number of dimensions. Next, consider the special case of two groups in which there are no differences in the means and the differences in variances are equal to the differences in covariances. Then it can be readily shown at the population (functional) level that SAVE is one-dimensional. In this situation SMVCIR identifies a dimension for variances and dimensions for covariances; hence, it is at least two-dimensional.

We now examine two situations at the population (functional) level. They are easily understood, and, more importantly, they do show basic differences among the SIR, SAVE and SMVCIR analyses. They also easily predict the differences of the three analyses for the generated example of Section 2. Derivations of the results in Sections 5.1 and 5.2 can be found in an unpublished technical report available from the authors. We used the matrix formulation of the SIR and SAVE kernels to derive these results. For the SMVCIR kernels, there is no loss in generality in which group is referenced. In expression (3.8), the first group is referenced. For the situations in Sections 5.1 and 5.2, it was easier to reference the last (second for these situations) group.

5.1. Situation 1

Suppose we have two groups, with the mean vectors $\mu_{X|1} = (\mu, 0, \ldots, 0)^{\mathrm{T}}$ and $\mu_{X|2} = 0$ and with the covariance matrices $\Sigma_{X\,|\,1} = I$ and $\Sigma_{X\,|\,2} = \text{diag}\{\sigma_1^2, \ldots, \sigma_k^2\}$, where $\mu \neq 0$ and $\sigma_i^2 > 0$, $i = 1, \ldots, k$. Thus, across the two groups, the first variable differs in location, all variables differ in variance, but there are no covariance differences. Assume that the groups are equilikely. The SIR kernel is then given by

$$B_{\text{SIR}} = \text{diag}\left\{\frac{\mu^2}{2 + 2\sigma_1^2 + \mu^2}, 0, \ldots, 0\right\}.$$

Hence, the analysis based on the SIR kernel involves one eigenvalue with its eigenvector in the direction of the difference in mean (μ). Of course, it does not show the difference in variances.

The SAVE kernel for this situation is

$$B_{\text{SAVE}} = \text{diag}\left\{\frac{(\frac{1}{2})\mu^2}{2 + 2\sigma_1^2 + \mu^2} + \frac{4(\sigma_1^2 - 1)^2}{(2 + 2\sigma_1^2 + \mu^2)^2}, \frac{4(\sigma_2^2 - 1)^2}{(2 + 2\sigma_2^2)^2}, \ldots, \frac{4(\sigma_k^2 - 1)^2}{(2 + 2\sigma_k^2)^2}\right\}.$$

Thus, the SAVE visualization is k-dimensional. It throws a direction for each variance difference $\sigma_i^2 - 1$, $i = 2, \ldots, k$. It also has one direction which is a combination of the difference in the variances and means of the first variable. As shown in the numerical example presented below, it is easy to construct situations where this combination direction is associated with the smallest eigenvalue; i.e., the location difference comes in last.

For the SMVCIR space, the vector of the of diagonal $\Delta_{Z|1}$ is

$$\delta_{Z|1} = \left(\frac{2(\sigma_1^2 - 1)}{2 + 2\sigma_1^2 + \mu^2}, \frac{2(\sigma_2^2 - 1)}{2 + 2\sigma_2^2}, \ldots, \frac{2(\sigma_k^2 - 1)}{2 + 2\sigma_k^2} \right)^{\mathrm{T}}.$$

The matrices $\Delta_{Z|i}^{(0)} = 0, i = 1, 2$. Hence, the SMVCIR space is spanned by the two vectors $(1/\sqrt{2})\delta_{Z|1}$ and $\frac{1}{\sqrt{2}}E(Z|Y=1)$. We chose the kernel which also involves $\frac{1}{\sqrt{2}}E(Z|Y=2)$ and $\frac{1}{\sqrt{2}}\delta_{Z|2}$. Because these are respective negatives of one another, the SMVCIR kernel is

$$B_{\mathrm{SMVCIR}} = \delta_{Z|1}\delta_{Z|1}^{\mathrm{T}} + B_{\mathrm{SIR}}.$$

Thus, the SMVCIR kernel is the sum of two one-dimensional kernels. It follows that the two nonzero eigenvectors of B_{SMVCIR} are linear combinations of the vectors $\delta_{Z\,|\,1}$ and $E(Z\,|\,Y=1)$. Therefore, the SMVCIR shows two directions. Each direction is a linear combination of directions based on the differences of variances and the mean difference.

The following numerical example clarifies the discussion. Set $\mu = 17.5$ and $\Sigma_{X|2} = \mathrm{diag}\{4, 9, 16, 25, 36, 49, 64, 81, 100, 121\}$. The eigenvalues of the spectral decomposition of B_{SAVE} in their natural order are (0.23, 0.41, 0.61, 0.73, 0.80, 0.88, 0.91, 0.92, 0.94). The corresponding eigenvectors form, of course, the standard basis. Thus, in this case, the location difference which is over 7 standard errors from 0 is the last SAVE direction. In contrast, the SMVCIR analysis gives two directions. The first direction is given by the eigenvector $(0.14, 0.24, 0.29, 0.32, 0.33, 0.34, 0.35, 0.36, 0.36, 0.36)^{\mathrm{T}}$ while the second is given by the eigenvector $(-0.99, 0.03, 0.04, 0.04, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05)^{\mathrm{T}}$. The first shows the differences in variances while the second contrasts variable one with the others, showing the mean difference.

5.2. Situation 2

Suppose the mean and covariance structure for the two groups is $\mu_{X|1} = 0$, $\Sigma_{X|1} = I$ and $\mu_{X|2} = 0$, $\Sigma_{X|2} = \begin{bmatrix} \sigma_1^2 & \eta & 0^{\mathrm{T}} \\ \eta & \sigma_2^2 & 0^{\mathrm{T}} \\ 0 & 0 & D \end{bmatrix}$, where $D = \mathrm{diag}\{\sigma_3^2, \ldots, \sigma_k^2\}$. Thus, across the groups, there are no differences in means among the variables, variables one and two differ in covariance, and all the variables differ in variance. Assume that the groups are equilikely. It follows that the SIR kernel is 0, while the SAVE kernel is given by

$$B_{\mathrm{SAVE}} = (\tfrac{1}{4})\{\Sigma_X^{-1/2}[\Sigma_{X|2} - I]\Sigma_X^{-1/2}\}^2.$$

This matrix has full rank k. Further, from the structure of the matrices that make up its factors, B_{SAVE} is of the form

$$B_{\mathrm{SAVE}} = \begin{bmatrix} A_2 & 0 \\ 0 & D_{k-2} \end{bmatrix},$$

where A_2 is a 2×2 matrix whose entries come from the differences in variances and covariances in variables one and two across the groups while D_{k-2} is a diagonal matrix whose entries come from the difference in variances in the last $(k-2)$ variables across the groups. Thus, the spectral decomposition of B_{SAVE} will show k directions with $(k-2)$ of them involved in individual variance differences in the last $(k-2)$ variables. The other two directions contain information from the covariance differences. As the numerical example presented below shows, it is easy to construct situations where SAVE throws the covariance difference directions last.

For the SMVCIR space, let $\delta_{Z|1}$ be the main diagonal of the matrix $\Delta_{Z|1}$. For this situation,

$$\Delta_{Z|1} = \begin{bmatrix} B_2 & 0 \\ 0 & C_{k-2} \end{bmatrix},$$

where C_{k-2} is a diagonal matrix. Hence, $\Delta_{Z|1}^{(0)}$ has only two nonzero columns, (the first two). Label these vectors v_{01} and v_{02}. Thus, SMVCIR space is spanned by these two nonzero columns and the vector $\delta_{Z|1}$. The vector $\delta_{Z|1}$ contains differences in variances information while the two nonzero columns contain differences in covariance information. The SMVCIR kernel is

$$B_{\mathrm{SMVCIR}} = \delta_{Z|1}\delta_{Z|1}^{\mathrm{T}} + v_{01}v_{01}^{\mathrm{T}} + v_{02}v_{02}^{\mathrm{T}}.$$

We next present a small numerical example. Set $k = 10$. Let $\eta = 5.4$ and let $\sigma_i^2 = (i+1)^2$, $i = 1, \ldots, 10$. In this case, the last two SAVE directions involve the first two variables, while the first eight directions concern the last eight variables, one for each variance difference. So covariance differences are in the last two SAVE directions. The spectral decomposition of the SMVCIR kernel shows three nonzero eigenvalues. The first eigenvector is an average over all variables which will highlight the differences in variances. The second eigenvector weighs the first variable heavily while the third eigenvector weighs the second variable heavily. These last two eigenvectors highlight the covariance difference between variables one and two.

The previous two examples illustrate the benefits of loading all the variance differences between 2 groups into a single dimension, and thus moving away from the central subspace. This approach enables the user to detect mean and covariance differences between the groups when variance differences exist between many of the variables. In addition, this approach enables the user to obtain an overall prediction of whether future observations fall into group 1 or group 2 on the basis of increased variance. In many practical situations this is highly desirable (see for example, Cook and Yin, 2001, pp. 198–199).

6. Real data examples

We offer two examples concerning real data sets, one recently analyzed by Zhu and Hastie concerning handwriting discrimination while the second concerns wine vintages. Both examples demonstrate the data-analytic advantages SMVCIR enjoys over SIR and SAVE.

Example 6.1 (*Pen digit training data*). As discussed by Zhu and Hastie (2003), the pen digit database contains samples of handwritten digits $\{1, 2 \ldots, 9\}$ collected from 44 different writers. Each digit is stored as a vector of 16 attributes. The data set is divided into a training data set and a learning data set. We focus on the training data set. Also, we chose to look at just 0's and 9's since the results in Zhu and Hastie (2003, p. 116) suggest that these are the hardest pair to distinguish between. This results in $n = 1499$ data points with 16 predictors and two groups. These data were analyzed by Zhu and Hastie (2003) using several procedures including SIR and SAVE.

Using a cut-off of 64% in the scree plot of singular values, the new procedure SMVCIR identifies six vectors of differences. The first is a difference in variance. Vectors of differences two–five represent differences in covariance, while the sixth vector of differences represents a location difference. Our SMVCIR space is spanned by these six vectors of differences.

We next formed the kernel and obtained its spectral decomposition and, hence, the SMVCIR directions. Fig. 5 shows a plot of the first three SMVCIR directions. Direction one is significantly more variable for digit 9 than digit 0 (Levene's test p-value is less than 0.001). SMVCIR directions two and three are significantly positively correlated for digit 9 and significantly negatively correlated for digit 0. The third direction, shows the location difference. Table 2 displays the standardized coefficients of the first three SMVCIR directions. The first SMVCIR direction is essentially a

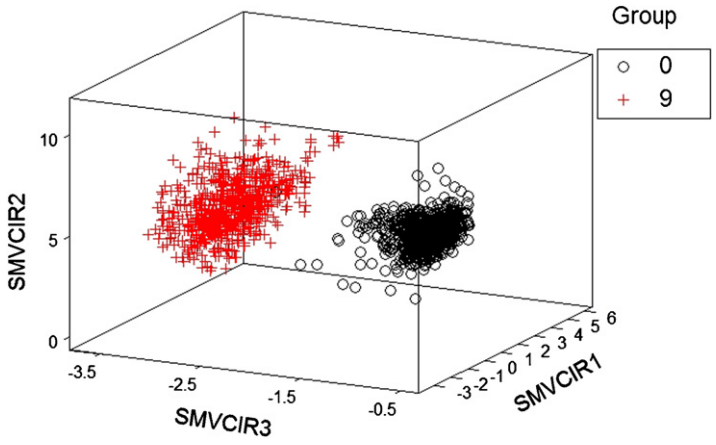

Fig. 5. Plot of the first three SMVCIR dimensions for the pen digit data.

Table 2
Standardized coefficients of the first three SMVCIR dimensions for the pen digit data

Variable	$SMVCIR_1$	$SMVCIR_2$	$SMVCIR_3$
x_1	0.010	−0.128	0.173
x_2	−0.169	0.240	0.043
x_3	−0.367	0.366	0.079
x_4	−0.183	−0.024	0.123
x_5	−0.334	0.183	0.243
x_6	−0.081	−0.242	0.079
x_7	−0.235	0.132	0.065
x_8	−0.406	0.039	0.217
x_9	−0.431	0.408	0.225
x_{10}	0.001	−0.148	−0.197
x_{11}	−0.049	0.078	−0.342
x_{12}	−0.143	−0.040	0.242
x_{13}	−0.125	−0.118	0.618
x_{14}	−0.417	0.595	0.335
x_{15}	−0.253	0.171	0.047
x_{16}	−0.055	−0.295	0.264

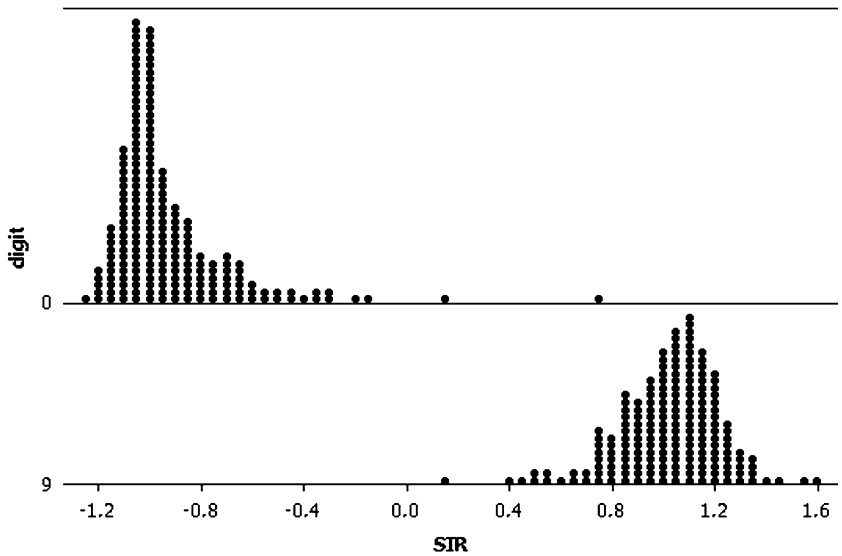

Fig. 6. Dot plot of the SIR dimension for the pen digit data.

weighted average of the predictors with the most variables being variables 8, 9 and 14. The second SMVCIR direction is essentially a contrast between variables 3, 9, 14 and variable 16. Finally, the third SMVCIR direction is essentially a contrast between variables 13, 14 and variable 11. Loading all the variance differences in a single direction allows just three SMVCIR directions to produce, apart from a single point, a separation between the two groups.

Fig. 6 shows the dotplot of the SIR dimension between the two groups. It shows a clear separation in locations between the groups but, of course, it does not show the difference in variations. On the other hand, as shown in Fig. 7, the first five dimensions of SAVE highlight individual variance differences between the groups for some of predictor variables, but a separation in location is not that apparent. See Zhu and Hastie (2003) for further discussion.

Example 6.2 (*Bordeaux harvest data*). The measurement of aspects of berry composition that influence wine flavor and taste is clearly desirable for the optimal production of wine (Francis et al., 1999).

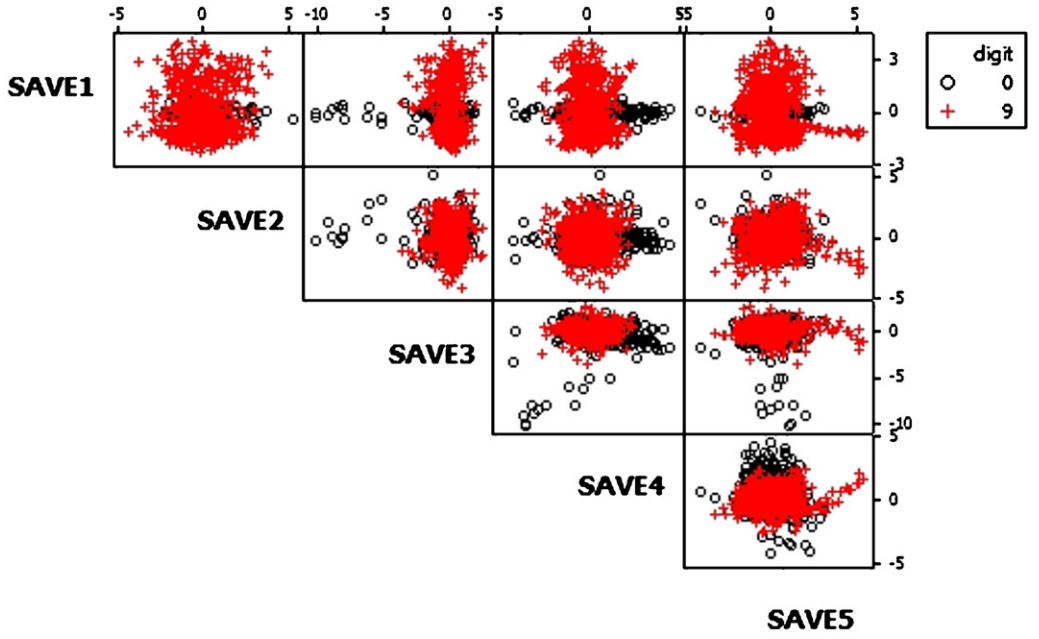

Fig. 7. Plot of the first five SAVE dimensions for the pen digit data.

Ribereau-Gayon et al. (2000, pp. 246–248) present data on various aspects of each vintage in Bordeaux from 1974 to 1996. These and other data were analyzed by Jones and Davis (2000) using multiple linear regression. For each vintage, the data include the phenology, average berry composition, weather and overall vintage quality ratings from 10 to 15 of the top chateaux in Bordeaux (the identities of the chateaux are confidential and known only by the chateaux and the data collectors). The quality of wine from each of the 23 vintages was described as ranging from mediocre (1992) to exceptional (1982, 1985, 1989, 1990, 1995 and 1996). Interest primarily centers on differentiating between exceptional vintages, very good vintages and the rest. Combining good, fairly good and mediocre vintages can be justified by the fact that there is little interest in these vintages—they return much lower prices at auction and are traded much less often than exceptional or very good vintages. Thus, we considered the following three groups: group 1 which denotes the exceptional vintages ($n_1 = 6$); group 2 which denotes the very good vintages ($n_2 = 7$); and group 3 which denotes the good, fairly good or mediocre vintages ($n_3 = 10$).

The following 13 predictor variables were considered:

$x_1 =$ Year,

$x_2 =$ Weight of 100 berries,

$x_3 =$ log(sugar concentration),

$x_4 =$ log(pH),

$x_5 =$ log(total acidity),

$x_6 =$ log(tartaric acid),

$x_7 =$ log(malic acid),

$x_8 =$ Harvest date (measured by day in the year),

$x_9 =$ Number of days from half-bloom to half-veraison,

$x_{10} =$ Number of days from half-veraison to harvest,

x_{11} = Sum of average temperatures in °C from April to September divided by 1000,

x_{12} = Duration of sun exposure in hours from April to September divided by 1000,

x_{13} = Number of days from April to September with temperatures, greater than or equal to 30 °C.

The predictor variables x_1–x_{10} are averages of values recorded across the 10–15 chateau mentioned above. Thus, while the sample size, $n = 23$ is relatively small, this is compensated by the fact that the bulk of the predictors are averages.

Variables x_2–x_7 are based on average values for Cabernet Sauvignon grapes at harvest. The focus on Cabernet Sauvignon grapes can be justified as follows. Virtually all Bordeaux chateaux blend Cabernet Sauvignon with other red grape varieties, such as Merlot with Cabernet Sauvignon often the dominant component of the blend (Parker, 2003, p. 1164). Given that Cabernet Sauvignon grapes ripen after Merlot, it is not surprising that Jones and Davis (2000, p. 249) found that "the wine industry in Bordeaux is more dependent on Cabernet Sauvignon for good vintages than on Merlot". We next provide a brief justification for each of the predictor variables.

A number of authors claim that the quality of modern day Bordeaux wine is much higher than it was even 20 years ago (e.g., Parker, 2003, p. XIII). Reasons given for this include changes to practices in the vineyard, especially the timing of harvesting the grapes and modern wine making techniques (Parker, 2003, p. XIII). Thus, it is important to include year (x_1) as a predictor variable in order to adjust for technology advances.

For red wine, small berries are desirable since they have a greater proportion of skins and seeds to pulp. As such, small red wine grapes have an increased color concentration, flavor compounds and tannins compared to large berries (Krstic et al., 2003, p. 20). We used weight to measure berry size (x_2).

One of the most critical decisions in winemaking is the decision when to harvest the grapes. For example, Peynaud (1984, p. 57) states that "the state of maturity of the grape conditions the quality... of wine" and that "good wine is only achieved with fully ripe grapes." Thus, a great deal of interest has centered on determining those aspects of berry composition which have greatest effect on the quality of the wine produced and monitoring these aspects in order to determine when optimum ripeness occurs. While the sugar content of grapes increases as they ripen, the opposite is true of acidity levels. Thus, attempts have been made to measure the state of maturity of grapes using a sugar/acidity ratio, commonly referred to as the maturation index. In France, this ratio is calculated as sugar concentration over total acidity, or titratable acidity as it is also often called (Ribereau-Gayon et al., 2000, p. 241). Other authors recommend the use of pH instead of total acidity (Van Rooyen et al., 1984). Since higher acidity produces lower pH results, the maturity index is then from sugar × pH. Alternatively, Coombe et al. (1980) recommend an index based on sugar × pH^2 to determine optimum ripeness. More recently, Ribereau-Gayon et al. (2000, p. 241) recommend that these indices should be used with caution. In addition, they recommend variations of each berry constituent (namely, sugar concentration, pH, total acidity, tartaric acid and malic acid) be taken into account separately. We used the log transformed version of these five constituents as predictor variables (that is, x_3–x_7), since interest centers on evaluating the effectiveness of sugar/acidity ratios). In addition, we ran the Box-Cox transformation to multinormality in R (alr3 library in R). It verified that the log transformation was an acceptable transformation to multinormality.

Harvest date is often said to be the best single predictor of vintage quality, with earlier harvest dates generally producing higher quality vintages. For example, Parker (2003, pp. 1171–1172) points out that dating the beginning of the red wine grape harvest for the finest vintages dating back to 1870 almost always occurs before the end of September, while the corresponding dates for notoriously bad vintages are generally into October. Earlier harvest dates mean that the grapes ripened during a warmer and sunnier period, thus benefiting grape quality (Ribereau-Gayon et al., 2000, p. 246). Thus, we used the day in the year on which harvest commenced as a predictor (x_8).

Jones and Davis (2000, p. 253) report that the interval between phenological events is often more important than the actual date of each event, with short intervals being associated with optimum conditions that produce rapid growth. Following Ribereau-Gayon et al. (2000, p. 246), we split the time from half-bloom to harvest into the time before and after half-veraison (x_9 and x_{10}).

Finally, climatic conditions over the growing period from April to September have a major effect on vintage quality. In particular great years are characterized by plentiful amounts of sunshine and heat. Otherwise the grapes never fully ripen (e.g., Parker, 2003, p. 1172). As such we included variables which measure temperature and heat as predictors (x_{11}, x_{12} and x_{13}).

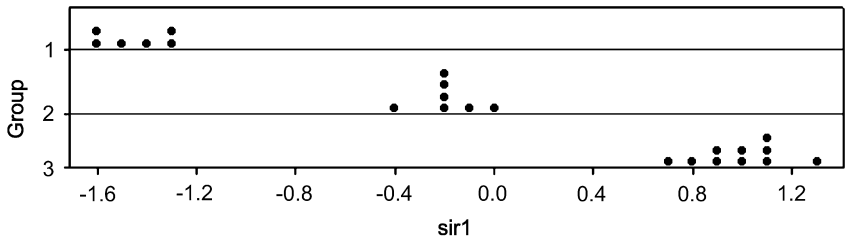

Fig. 8. Dot plot of the SIR direction for the Bordeaux data.

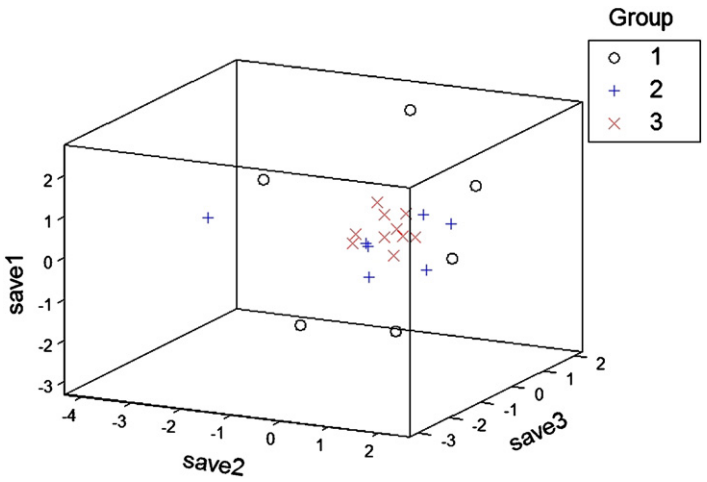

Fig. 9. Plot of the first three SAVE dimensions for the Bordeaux data.

SIR finds one direction which, after removing the three predictor variables with standardized coefficients much less than 0.1 (i.e., LogTartaricAcid, AvTemp and Sun), is statistically significant (permutation p-value $=0.048$). As shown in the dotplot of Fig. 8, this direction of SIR is a location difference between the three groups. Omitting each observation one at a time and recalculating the LDA classification function using the remaining data, and then classifying the omitted observation leads to 21 out of the 23 data points being correctly classified, with one misclassification in group 1 and one in group 2.

SAVE gave no statistically significant dimensions. For example, the permutation test p-values (based on 5000 replications) for the first two SAVE dimensions are 0.714 and 0.845, respectively. A plot of the first three SAVE dimensions is given in Fig. 9. A counter intuitive feature of this plot is that the variability inherent in the SAVE dimensions decreases with group and as such variability is highest in the group corresponding to the highest quality vintages. The location difference evident between the two groups is not shown up by SAVE till dimension 10 (SAVE dimension 10 has correlation with SIR of magnitude equal to 0.94).

For the SMVCIR procedure, the scree plot of singular values rose slowly for this data set. With a parsimonious model in mind, we started with the cut-off value of 35%, which leads to three vectors of differences. The first and third vectors of differences represent covariance differences while the second vector represents a location difference.

These three vectors of differences formed our SMVCIR space. Its kernel and subsequent spectral decomposition were obtained. Fig. 10 shows a plot of the first three SMVCIR directions. The second direction represents a location difference (the correlation between the first SIR direction and this direction is -0.87). Directions one and three show the differences in covariances.

Table 3 contains the raw and standardized coefficients of the first SIR dimension. Recall that the first SIR dimension is such that it is negatively related to quality, that it, low values of it are associated with wines in group 1. The signs of the coefficients in Table 3 are as expected with the following exceptions. Having adjusted for the effects of the other

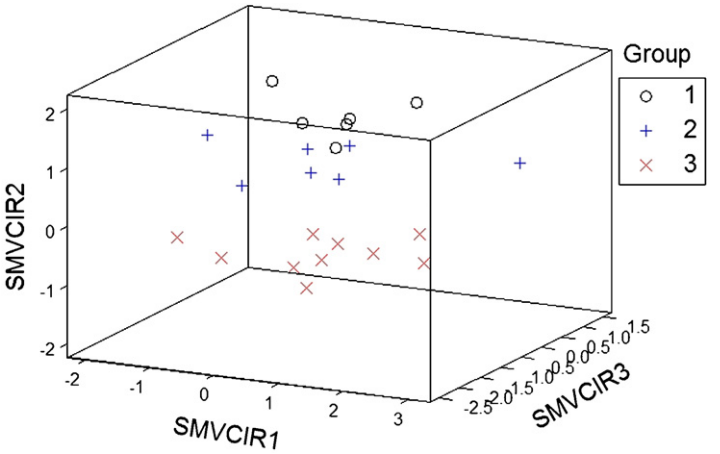

Fig. 10. Plot of the first three SMVCIR directions for the Bordeaux data.

Table 3
Raw and standardized coefficients of the first SIR dimension

SIR predictors	Coefs.	Std. Coefs.
x_1, Year	−0.049	−0.305
x_2, Weight	0.022	0.203
x_3, LogSugar	−5.599	−0.329
x_4, LogpH	−24.560	−0.522
x_5, LogTotalAcidity	−2.108	−0.242
x_6, LogTartaricAcid	−1.223	−0.071
x_7, LogMalicAcid	0.801	0.176
x_8, Harvest	0.060	0.432
x_9, HalfBloom	−0.056	−0.133
x_{10}, HalfVeraison	−0.066	−0.239
x_{11}, AvTemp	−0.106	0.014
x_{12}, Sun	0.218	0.023
x_{13}, HotDays	0.041	0.359

variables, the interval between phenological events was found to be positively related to quality while the number of hot days was found to be negatively related to quality. These results contradict the findings in Jones and Davis (2000).

Table 4 contains the raw and standardized coefficients of the first three SMVCIR dimensions. Recall that dimensions 1 and 3 correspond to differences in covariance, while dimension 2 represents location differences.

Next we compare the first dimension of SIR with the second dimension of SMVCIR. In terms of standardized coefficients the two most important berry constituent variables in the first SIR dimension are x_3, LogSugar and x_4, LogpH. In addition, a best subsets regression of SIR_1 in Minitab identifies these two variables as providing the best two variable approximation to the first SIR dimension in terms of adjusted R^2. Regressing these two variables on SIR_1 and ignoring the intercept term gives

$$\begin{aligned}\mathrm{SIR}_1 &= -7.902\log(\text{sugar}) - 23.963\log(\text{pH})\\ &= -7.902\log(\text{sugar}\times \text{pH}^{23.963/7.902}).\end{aligned}$$

Thus, the relevant term in SIR_1 involving sugar and pH is sugar $\times$ $\text{pH}^{3.0}$. In terms of standardized coefficients of the second SMVCIR dimension the most important acidity measurement is x_5, logTotalAcidity. However, a best subsets regression of SMVCIR_2 in Minitab identifies x_3, LogSugar and x_4, LogpH as providing the best two variable approximation to SMVCIR_2 in terms of adjusted R^2. Proceeding as above we find that the relevant term in SMVCIR_2 involving sugar and pH is sugar $\times$ $\text{pH}^{3.5}$. Fig. 11 shows a dot plot of sugar $\times$ $\text{pH}^{3.0}$ for each of the three groups.

Table 4
Standardized coefficients of the first three SMVCIR dimensions for the Bordeaux harvest data

Variable	$SMVCIR_1$		$SMVCIR_2$		$SMVCIR_3$	
	Coefs.	Std. Coefs.	Coefs.	Std. Coefs.	Coefs.	Std. Coefs.
x_1, Year	−0.022	−0.049	0.013	0.054	0.014	0.028
x_2, Weight	0.009	0.031	−0.007	−0.045	−0.070	−0.203
x_3, LogSugar	−2.995	−0.063	3.079	0.123	13.001	0.240
x_4, LogpH	27.226	0.208	27.783	0.401	−11.426	−0.076
x_5, LogTotalAcidity	−5.199	−0.215	−6.907	−0.540	18.617	0.671
x_6, LogTartaricAcid	−12.837	−0.266	2.048	0.080	−8.771	−0.159
x_7, LogMalicAcid	2.493	0.197	2.306	0.345	−6.353	−0.438
x_8, Harvest	−0.049	−0.128	−0.083	−0.408	0.087	0.195
x_9, HalfBloom	−0.091	−0.078	0.041	0.066	−0.244	−0.181
x_{10}, HalfVeraison	−0.103	−0.135	−0.070	−0.172	0.193	0.220
x_{11}, AvTemp	−15.397	−0.750	−0.902	−0.083	6.426	0.273
x_{12}, Sun	9.814	0.384	0.652	0.048	−2.600	−0.089
x_{13}, HotDays	0.066	0.207	−0.074	−0.438	0.058	0.158

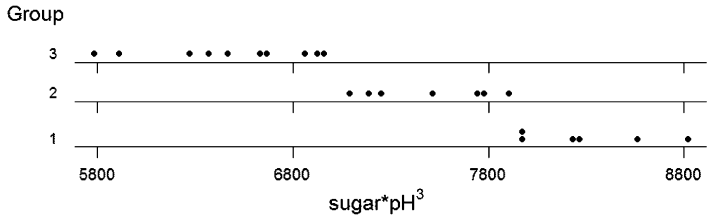

Fig. 11. Dot plot of sugar × $pH^{3.0}$ for the Bordeaux data.

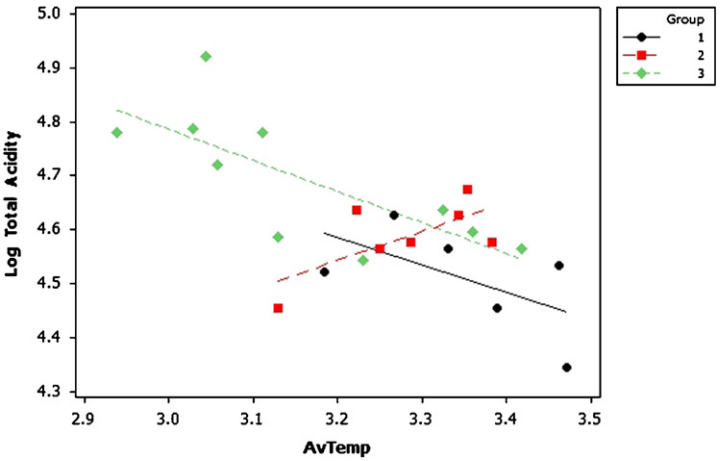

Fig. 12. Plot of Log Total Acidity and AvTemp with regression lines for each group.

(A dot plot of sugar × $pH^{3.5}$ is very similar to that given in Fig. 11.) It is evident from Fig. 11 that the three groups are disjoint in terms of values of this variable. The same is not true for sugar × pH^2 as recommended by Coombe et al. (1980).

Recall that the first and third dimensions of SMVCIR represent covariance differences. It is evident from Table 4 that x_{11}, AvTemp, has the highest standardized coefficient for $SMVCIR_1$ and that x_5, LogTotalAcidity, has the highest standardized coefficient for $SMVCIR_3$.

Fig. 12 contains a plot of x_5, LogTotalAcidity, and x_{11}, AvTemp, with different regression lines fitted for each group. The slopes of the lines are statistically significantly different (p-value = 0.041). The slopes of the lines for groups 1

and 3 are negative, as one would expect, reflecting the fact that total acidity decreases as heat increases. On the other hand, the slope of the regression lines for group 2 is positive. This phenomenon is worthy of further investigation. It may provide an early warning that an otherwise excellent vintage may end up only being very good. A remedy in this circumstance could be to harvest the grapes sooner while the acidity levels are still relatively high.

7. Simulation study on the ordering procedure

In this section, we present the results of a simulation investigation concerning our ordering procedure. We consider two situations. In the first situation the groups differ in location, while in the second situation they differ in variability. It is, thus, clear in each situation which vectors in the spanning set should enter first.

For both situations, we used two groups ($g=2$), four variables ($k=4$), and equal sample sizes ($n_1=n_2=30$). As the error distribution we selected the multivariate normal. Hence, for both situations there are six columns in the spanning sets of the three methods of the form (3.7). The first four columns contain the differences in covariances, column #5 is the vector of differences in variances, and column #6 is the vector of differences in locations.

The SMVCIR procedure used in these simulations was the procedure as discussed in Sections 3 and 4. The observations were prestandardized as discussed in Section 3. A total of 1000 simulations was conducted for each situation.

For situation 1, the mean and variance–covariance structures for the two groups were set at: $E(X_1)=0$, $E(X_2)=(0,0,0,4)'$, $\mathrm{Var}(X_1)=I_4$, and $\mathrm{Var}(X_2)=I_4$. For situation 1, there is only a difference in location for the fourth variable and, hence, the ordering procedure of each method should select column #6 to enter first. Over the 1000 simulations, Column #6 was selected first 930 times by the SMVCIR procedure.

For situation 2, the mean and variance–covariance structures for the two groups were set at: $E(X_1)=0$, $E(X_2)=0$, $\mathrm{Var}(X_1)=I_4$, and $\mathrm{Var}(X_2)=\mathrm{diag}\{1,1,1,25\}$. Hence, the means and covariances for the two groups are the same while the variances differ for the fourth variable. Thus, the ordering procedure of each method should select column #5 to enter first. For this situation, SMVCIR selected Column #5 as the first vector 811 times.

The success rates of SMVCIR are well above the random selection rate of $\frac{1}{6}$ for all distributions, thus demonstrating the high power of our ordering procedure.

8. Conclusion

In this paper we have proposed SMVCIR, a new visualization technique for the classification problem. Unlike SAVE, SMVCIR separates the variance differences among the groups from the covariance differences and thus does not necessarily focus on the central subspace. This enables SMVCIR to detect mean and covariance differences between the groups when variance differences exist between many of the variables. Another key aspect of the new procedure, is an ordering of the dimensions based on their relative importance. This produces fewer possible dimensions and thus guides the user to focus on the most important differences among the groups. The new data-analytic procedure offers great flexibility in that it allows the user to focus on mean differences, variance differences, or covariance differences as well as any combination of these three. In particular, we know of no other general procedure, other than SMVCIR, which can readily detect both mean and covariance differences when many variance differences exist.

Acknowledgment

We would like to thank an anonymous referee whose comments led to an improvement in the paper.

Appendix

Proposition. *If the dimension of the* SAVE *space is less than k and the variance differences are not a linear combination of the covariance differences then* $\Omega_{\mathrm{SMVCIR}} \subset \Omega_{\mathrm{SAVE}}$.

Proof. Let $\mathbf{v}$ be the jth column vector in the covariance difference matrix Δ such that $\mathbf{v}$ has the nonzero difference δ in its jth component. Note that Δ is in the spanning set of SAVE. Then by the hypothesis, if we reduce Δ to column echelon form (or row echelon form since Δ is symmetric), $\mathbf{v}$ becomes the vector $\delta\mathbf{e}_j$, where $\mathbf{e}_j$ is the jth element of the standard basis. Thus, $\delta\mathbf{e}_j$ is in the SAVE space and, hence, so is the jth column of $\Delta^{(0)}$. Since this is true for every nonzero

variance, SAVE contains the vector of variance differences (just sum the $\delta \mathbf{e}_j$'s) and the corresponding columns of the $\Delta^{(0)}$'s. Of course, for the zero variance differences, the corresponding columns of the $\Delta^{(0)}$'s are in SAVE. Therefore, the SAVE space contains the SMVCIR space. □

References

Chen, C.H., Li, K.C., 2001. Generalizations of Fisher's linear discriminant analysis via the approach of sliced inverse regression. J. Korean Statist. Soc. 30, 193–217.

Cook, R.D., Critchley, F., 2000. Identifying regression outliers and mixtures graphically. J. Amer. Statist. Assoc. 95, 781–794.

Cook, R.D., Lee, H., 1999. Dimension reduction in regressions with a binary response. J. Amer. Statist. Assoc. 94, 1187–1200.

Cook, R.D., Weisberg, S., 1991. Discussion of Li, K.C., 1991. Sliced inverse regression for dimension reduction (with discussion). J. Amer. Statist. Assoc. 86, 316–342. J. Amer. Statist. Assoc. 86, 328–332.

Cook, R.D., Yin, X., 2001. Dimension reduction and visualization in discriminant analysis and discussion. Austral. New Zealand J. Statist. 43, 147–199.

Coombe, B.G., Dundon, R.J., Short, A.W.S., 1980. Indices of sugar-acidity as ripeness criteria for winegrapes. J. Sci. Food Agric. 3, 495–502.

Dongarra, J.J., Bunch, J.R., Moler, C.B., Stewart, G.W., 1979. Linpack Users' Guide. SIAM, Philadelphia.

Fisher, R.A., 1936. The use of multiple measurements in taxonomic problems. A. Eug. 7.

Francis, I.L., Iland, P.G., Cynkar, W.U., Kwiatkowski, M., Williams, P.J., Armstrong, H., Botting, D.G., Gawel, R., Ryan, C., 1999. Assessing wine quality with the G–G assay. Proceedings of the 10th Australian Wine Industry Technical Conference, Adelaide, pp. 104–108.

Gnanadesikan, R., 1977. Methods for Statistical Analysis of Multivariate Observations. Wiley, New York.

Golub, G.H., Van Loan, C.F., 1996. Matrix Computations, third ed. John Hopkins University Press, Baltimore.

Jones, G.V., Davis, R.E., 2000. Climate influences on grapevine phenology, grape composition, and wine production and quality for Bordeaux, France. Amer. J. Enol. Vitic. 51, 249–261.

Kent, J.T., 1991. Comment on "sliced inverse regression for dimension reduction". by K.C. Li. J. Amer. Statist. Assoc. 86, 336–337.

Krstic, M., Moulds, G., Panagiotopoulos, B., West, S., 2003. Growing Quality Grapes to Winery Specification. Winetitles, Adelaide.

Li, K.C., 1991. Sliced inverse regression for dimension reduction (with discussion). J. Amer. Statist. Assoc. 86, 316–342.

Parker, R.M., 2003. Bordeaux. fourth ed. Simon & Schuster, New York.

Peynaud, E., 1984. Knowing and Making Wine. Wiley, New York.

Ribereau-Gayon, P., Dubourdieu, D., Doneche, B., Lonvaud, A., 2000. Handbook of Enology, vol. 1. Wiley, Chichester.

Seber, G.A.F., 1984. Multivariate Observations. Wiley, New York.

Stewart, G.W., 1973. Introduction to Matrix Computations. Academic Press, New York.

Van Rooyen, P.C., Ellis, L.P., Du Plessis, C.S., 1984. Interactions between grape maturity and quality for Pinotage and Cabernet Sauvignon wines from four locations. South African J. Enol. Vitic. 5, 29–34.

Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S. Springer, New York.

Zhu, M., Hastie, T.J., 2003. Feature extraction for nonparametric discriminant analysis. J. Comput. Graph. Statist. 12, 101–120.