Colloquium Biometricum 40
2010, 117–133

ANALYSIS OF MULTIVARIATE REPEATED MEASURES DATA

Mirosław Krzyśko[1,2], Wiesław Mądry[3], Stanisław Pluta[4], Michał Skorzybut[1], Waldemar Wołyński[1]

[1]Adam Mickiewicz University
Faculty of Mathematics and Computer Science
Umultowska 87, 61-614 Poznań
[2]President Stanislaw Wojciechowski Higher Vocational State School in Kalisz
Institute of Management
Nowy Świat 4, 62-800 Kalisz
[3] Warsaw University of Life Sciences-SGGW
Department of Experimental Design and Bioinformatics
Nowoursynowska 159, 02-776 Warsaw
[4]Research Institute of Pomology and Floriculture
Fruit Plant Breeding Department
Pomologiczna 18, 96-100 Skierniewice
e-mails: mkrzysko@amu.edu.pl; kdb@sggw.pl; Stanislaw.Pluta@insad.pl; wolynski@amu.edu.pl

Summary

This paper proposes new classifiers under the assumption of multivariate normality for multivariate repeated measures data (doubly multivariate data) with Kronecker product covariance structures. These classifiers are especially useful when the number of observations is not large enough to estimate the covariance matrices, and thus the traditional classifiers fail. Analysis of these data using a MANOVA model is also considered. The quality of these multivariate statistical methods is examined on some real data.

Key words and phrases: classifiers, repeated measures data (doubly multivariate data), Kronecker product covariance structure, a mixed MANOVA model

Classification AMS 2010: 62H30, 62H15, 62H12, 62J10

1. Introduction

The problem we consider in this paper is the following. Given that an object (or person) is known to come from one of K distinct classes, we wish to assign the object to one of these classes on the basis of p characteristics associated with the object, measured at T different time points. Such data are often referred to in the statistical and behavioral science literature as the multivariate repeated measures data or doubly multivariate data. To distinguish the known classes from each other, we associate a unique class label y with each class; the observations are then described as labeled observations. We denote information on the typical object by $\mathbf{x}$, a $(pT \times 1)$ – dimensional column vector obtained by stacking all p characteristics at the first time point, then stacking all p characteristics at the second time point below it and so on.

Next, assume that

$$\mathbf{x} \sim N_{pT}(\boldsymbol{\mu}, \boldsymbol{\Omega})$$

with $pT \times pT$ positive definite covariance matrix $\boldsymbol{\Omega}$.

The optimal Bayes classifier is

$$g(\mathbf{x}) = \arg \max_{1 \le i \le K} \pi_i \, f_i(\mathbf{x})$$

where $\pi_i = P(y = i)$ is the prior probability that $\mathbf{x}$ is a member of a class i, $\pi_1 + \pi_2 + \ldots + \pi_K = 1$, and $f_i(\mathbf{x}) = f_i(\mathbf{x} \mid y = i)$ is the normal probability density function associated with the random vector $\mathbf{x}$ for a class i, $i = 1, 2, \ldots, K$.

When $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$ are unknown, they must by estimated on the basis of the training data set $D_n = \{(\mathbf{x}_i, y_i)\}$, $i = 1, 2, \ldots, n$, observed in the past, that is, a set of n observations is available for which the true categorization is known. The matrix $\boldsymbol{\Omega}$ is positive definite. Its estimate $\hat{\boldsymbol{\Omega}}$ is positive definite with probability one if and only if $n > pT$ (see e.g. Giri (1996), p.93). Hence, estimation of the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$ will require a very large sample, which may not always be feasible. Hence, we assume $\boldsymbol{\Omega}$ to be of the form (Roy and Khattree 2005a, 2005b, 2008):

$$\boldsymbol{\Omega} = \mathbf{V} \otimes \boldsymbol{\Sigma},$$

where $\mathbf{V}$ is a $T \times T$ positive definite covariance matrix and $\boldsymbol{\Sigma}$ is $p \times p$ positive definite covariance matrix. The matrix $\mathbf{V}$ represents the covariance between repeated measures on a given subject and for a given characteristic. Likewise, $\boldsymbol{\Sigma}$

represents the covariance between all characteristics on a given subject and for a given time point. The above covariance structure is subject to an implicit assumption that for all characteristics, the correlation structure between repeated measures remains the same and that covariance between all characteristics does not depend on time and remains constant for all time points.

In this case the estimates of the matrices **V** and **Σ** are positive definite with probability one if and only if $n > \max(p, T)$.

Classification rules for univariate repeated measures data were given by Roy and Khattree (2005a). Classification rules in the case of multivariate repeated measures data under the assumption of multivariate normality for classes and with compound symmetric correlation structure on the matrix **V** were given by Roy and Khattree (2005b). Next, Roy and Khattree (2008) gave the solution of this problem for the case where the correlation matrix **V** has the first order autoregressive (AR(1)) structure. Next, Krzyśko and Skorzybut (2009) gave the solution of this problem in the case when no structures whatsoever are imposed on **V** and **Σ** except that they are positive definite.

Analysis of this data using a one-way MANOVA model is also considered. The problems of interest are to test for the (i) time effect, (ii) group effect, and (iii) the effect of interaction between time and group.

This paper is organized as follows. In Section 2 quadratic and linear classifiers are presented. A mixed effects MANOVA model is considered in Section 3. Section 4 examines the quality of the various multivariate statistical methods on some real data.

2. Quadratic and linear classifiers

Suppose that no structures whatsoever are assumed on **V** and **Σ** except that they are positive definite. In this case the classifier has the form

$$g(\mathbf{x}) = \arg \max_{1 \le i \le K} \ln\left(\pi_i \; f_i(\mathbf{x})\right),$$

where

$$f_i(\mathbf{x}) = (2\pi)^{-\frac{pT}{2}} \left|\mathbf{V}_i\right|^{-\frac{p}{2}} \left|\boldsymbol{\Sigma}_i\right|^{-\frac{T}{2}} \exp\left[-\frac{1}{2}\left(\mathbf{x}-\boldsymbol{\mu}_i\right)'\left(\mathbf{V}_i^{-1} \otimes \boldsymbol{\Sigma}_i^{-1}\right)\left(\mathbf{x}-\boldsymbol{\mu}_i\right)\right],$$

and π_i is the prior probability that an observation $\mathbf{x}$ is from class i.

The parameters $\boldsymbol{\mu}_i$, $\mathbf{V}_i$ and $\boldsymbol{\Sigma}_i$ are unknown and should be estimated relying on K training samples of sizes $n_1, n_2,\ldots, n_K$ from the respective classes.

Let $\mathbf{x}_{\mathrm{ijk}}$ (k = 1, 2,…, T; j = 1, 2,…, n_i; i = 1, 2,…, K) be a $p \times 1$ column vector of measurements on the jth individual in the ith class at the kth time point and

$$\mathbf{x}_{\mathrm{ij}} = \left(\mathbf{x}'_{\mathrm{ij}1}, \mathbf{x}'_{\mathrm{ij}2},\ldots, \mathbf{x}'_{\mathrm{ij}T}\right)'.$$

Then $\mathbf{x}_{\mathrm{ij}}$ is a $pT \times 1$ random observational vector corresponding to the jth individual in the ith class.

We consider a model described as follows:

$$\text{all observations } \mathbf{x}_{\mathrm{ij}} \text{ are independent and } \mathbf{x}_{\mathrm{ij}} \sim \mathrm{N}_{pT}\,(\boldsymbol{\mu}_\mathrm{I}, \mathbf{V}_\mathrm{i} \otimes \boldsymbol{\Sigma}_\mathrm{i}), \tag{2.1}$$

where $\mathbf{V}_\mathrm{i}$ is a $T \times T$ positive definite matrix and $\boldsymbol{\Sigma}_\mathrm{i}$ is a $p \times p$ positive definite matrix, j = 1, 2,…, n_i, $n_\mathrm{i} > \max(p, T)$, i = 1, 2,…, K.

Given a sample of n_i random observations $\mathbf{X}_\mathrm{i} = \left(\mathbf{x}_{i1}, \mathbf{x}_{i2},\ldots, \mathbf{x}_{in_i}\right)$ from $\mathrm{N}_{pT}\,(\boldsymbol{\mu}_\mathrm{I}, \mathbf{V}_\mathrm{i} \otimes \boldsymbol{\Sigma}_\mathrm{i})$, the log likelihood function is given by

$$\begin{aligned}\ln L\left(\boldsymbol{\mu}_i, \mathbf{V}_i^{-1}, \boldsymbol{\Sigma}_i^{-1}; \mathbf{X}_i\right) = {} & -\frac{1}{2}\mathrm{pTn}_i \ln(2\pi) + \frac{1}{2}\mathrm{pn}_i \ln\left|\mathbf{V}_i^{-1}\right| + \frac{1}{2}\mathrm{Tn}_i \ln\left|\boldsymbol{\Sigma}_i^{-1}\right| \\ & -\frac{1}{2}tr\left[\left(\mathbf{V}_i^{-1} \otimes \boldsymbol{\Sigma}_i^{-1}\right)\mathbf{A}_i\right] - \frac{1}{2}\mathrm{n}_i tr\left[\left(\mathbf{V}_i^{-1} \otimes \boldsymbol{\Sigma}_i^{-1}\right)\left(\bar{\mathbf{x}}_i - \boldsymbol{\mu}_i\right)\left(\bar{\mathbf{x}}_i - \boldsymbol{\mu}_i\right)'\right],\end{aligned} \tag{2.2}$$

where

$$\bar{\mathbf{x}}_\mathrm{i} = \frac{1}{n_\mathrm{i}}\sum_{\mathrm{j}=1}^{n_\mathrm{i}} \mathbf{x}_{\mathrm{ij}}$$

and

$$\mathbf{A}_\mathrm{i} = \left(\mathrm{a}_{\mathrm{rs}}^{\mathrm{i}}\right) = \sum_{\mathrm{j}=1}^{n_\mathrm{i}} \left(\mathbf{x}_{\mathrm{ij}} - \bar{\mathbf{x}}_\mathrm{i}\right)\left(\mathbf{x}_{\mathrm{ij}} - \bar{\mathbf{x}}_\mathrm{i}\right)' = \begin{pmatrix} \mathbf{A}_{11}^{\mathrm{i}} & \mathbf{A}_{12}^{\mathrm{i}} & \ldots & \mathbf{A}_{1T}^{\mathrm{i}} \\ \mathbf{A}_{21}^{\mathrm{i}} & \mathbf{A}_{22}^{\mathrm{i}} & \ldots & \mathbf{A}_{2T}^{\mathrm{i}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{T1}^{\mathrm{i}} & \mathbf{A}_{T2}^{\mathrm{i}} & \ldots & \mathbf{A}_{TT}^{\mathrm{i}} \end{pmatrix} \tag{2.3}$$

is the block matrix containing T^2 blocks, where $\mathbf{A}^{i}_{jk} = \left(\mathbf{A}^{i}_{kj}\right)'$.

It can be seen that the maximum likelihood estimate of $\boldsymbol{\mu}_i$ is $\hat{\boldsymbol{\mu}}_i = \bar{\mathbf{x}}_i$. Therefore, substituting $\hat{\boldsymbol{\mu}}_i = \bar{\mathbf{x}}_i$ in (2.2) we obtain

$$\ln L\left(\boldsymbol{\mu}_i, \mathbf{V}_i^{-1}, \boldsymbol{\Sigma}_i^{-1}; \mathbf{X}_i\right) = -\frac{1}{2}\mathrm{pTn}_i \ln(2\pi) + \frac{1}{2}\mathrm{pn}_i \ln\left|\mathbf{V}_i^{-1}\right| + \frac{1}{2}\mathrm{Tn}_i \ln\left|\boldsymbol{\Sigma}_i^{-1}\right| - \frac{1}{2} tr\left[\left(\mathbf{V}_i^{-1} \otimes \boldsymbol{\Sigma}_i^{-1}\right)\mathbf{A}_i\right],\ i = 1, 2, ..., \mathrm{K}.$$

Let

$$\mathbf{A}_i^{*} = \left(a_{rs}^{i*}\right) = \begin{pmatrix} \mathbf{A}_{11}^{i*} & \mathbf{A}_{12}^{i*} & \cdots & \mathbf{A}_{1p}^{i*} \\ \mathbf{A}_{21}^{i*} & \mathbf{A}_{22}^{i*} & \cdots & \mathbf{A}_{2p}^{i*} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{p1}^{i*} & \mathbf{A}_{p2}^{i*} & \cdots & \mathbf{A}_{pp}^{i*} \end{pmatrix}, \qquad (2.4)$$

where $\mathbf{A}_{jk}^{i*} = \left(\mathbf{A}_{kj}^{i*}\right)'$.

The blocks

$$\mathbf{A}_{jk}^{i*} = \left(a_{rs}^{i*jk}\right)$$

are constructed with the elements of the matrix $\mathbf{A}_i = \left(a_{rs}^{i}\right)$:

$$a_{rs}^{i*jk} = a^{i}_{j+(r-1)p,\,k+(s-1)p},$$

r, s = 1, 2,…, T; j, k = 1, 2,…, p; i = 1, 2,…, K.
Let

$$\mathbf{B}_i = \left(\mathrm{tr}\mathbf{A}_{jk}^{i}\boldsymbol{\Sigma}_i^{-1}\right) \qquad (2.5)$$

and

$$\mathbf{C}_i = \left(\mathrm{tr}\mathbf{A}_{jk}^{i*}\mathbf{V}_i^{-1}\right), \qquad (2.6)$$

i = 1, 2,…, K.

Then the following formulae hold:

$$\frac{\partial \mathrm{tr}\left[\left(\mathbf{V}_{\mathrm{i}}^{-1} \otimes \mathbf{\Sigma}_{\mathrm{i}}^{-1}\right)\mathbf{A}_{\mathrm{i}}\right]}{\partial \mathbf{V}_{\mathrm{i}}^{-1}} = 2\mathbf{B}_{\mathrm{i}} - \mathrm{diag}\mathbf{B}_{\mathrm{i}}$$

and

$$\frac{\partial \mathrm{tr}\left[\left(\mathbf{V}_{\mathrm{i}}^{-1} \otimes \mathbf{\Sigma}_{\mathrm{i}}^{-1}\right)\mathbf{A}_{\mathrm{i}}\right]}{\partial \mathbf{\Sigma}_{\mathrm{i}}^{-1}} = 2\mathbf{C}_{\mathrm{i}} - \mathrm{diag}\mathbf{C}_{\mathrm{i}},$$

i = 1, 2,…, K.

Differentiating the $\ln L\left(\boldsymbol{\mu}_i, \mathbf{V}_i^{-1}, \mathbf{\Sigma}_i^{-1}; \mathbf{X}_i\right)$ with respect to $\mathbf{V}_{\mathrm{i}}^{-1}$ and equating it to zero results in the equation

$$\mathbf{V}_{\mathrm{i}} = \frac{1}{pn_{\mathrm{i}}}\mathbf{B}_{\mathrm{i}} \tag{2.7}$$

where $\mathbf{B}_{\mathrm{i}} = \left(\mathrm{tr}\mathbf{A}_{\mathrm{jk}}^{\mathrm{i}}\mathbf{\Sigma}_{\mathrm{i}}^{-1}\right)$, i = 1, 2,…, K.

Differentiating the $\ln L\left(\boldsymbol{\mu}_i, \mathbf{V}_i^{-1}, \mathbf{\Sigma}_i^{-1}; \mathbf{X}_i\right)$ with respect to $\mathbf{\Sigma}_{\mathrm{i}}^{-1}$ and equating it to zero results in the equation

$$\mathbf{\Sigma}_{\mathrm{i}} = \frac{1}{Tn_{\mathrm{i}}}\mathbf{C}_{\mathrm{i}} \tag{2.8}$$

where $\mathbf{C}_{\mathrm{i}} = \left(\mathrm{tr}\mathbf{A}_{\mathrm{jk}}^{\mathrm{i*}}\mathbf{V}_{\mathrm{i}}^{-1}\right)$, i = 1, 2,…, K.

In this case no explicit maximum likelihood estimators are available. The MLEs of $\mathbf{V}_{\mathrm{i}}$ and $\mathbf{\Sigma}_{\mathrm{i}}$ are obtained by solving simultaneously and iteratively the equations (2.7) and (2.8).
The following iterative steps are suggested to obtain the maximum likelihood estimators of $\mathbf{V}_{\mathrm{i}}$ and $\mathbf{\Sigma}_{\mathrm{i}}$, i = 1, 2,…, K.

Algorithm 1

Step 1. Compute $\mathbf{A}_{\mathrm{i}}$ and $\mathbf{A}_{\mathrm{i}}^{*}$ from the equations (2.3) and (2.4), respectively. Get the initial covariance matrix $\mathbf{\Sigma}_{\mathrm{i}}$ of the form

$$\tilde{\boldsymbol{\Sigma}}_i = \mathbf{S}_i = \frac{1}{n_i T} \sum_{j=1}^{n_i} \sum_{k=1}^{T} \left(\mathbf{x}_{ijk} - \overline{\mathbf{x}}_{ik}\right)\left(\mathbf{x}_{ijk} - \overline{\mathbf{x}}_{ik}\right)', \tag{2.9}$$

where

$$\overline{\mathbf{x}}_{ik} = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{x}_{ijk}, \ i = 1,\dots, K, \ k = 1,\dots, T.$$

Step 2. On the basis the initial covariance matrix $\mathbf{S}_i$ and the matrix $\mathbf{A}_i$ compute the matrix $\mathbf{B}_i$ given by (2.5).

Step 3. Compute the matrix $\mathbf{V}_i$ from equation (2.7) using $\mathbf{B}_i$ obtained in Step 2.

Step 4. Compute the matrix $\mathbf{C}_i$ from (2.6) and the matrix $\boldsymbol{\Sigma}_i$ from (2.8).

Step 5. Compute the matrix $\mathbf{B}_i$ given by (2.5) using the matrix $\boldsymbol{\Sigma}_i$ obtained in **Step 4** and next compute the matrix $\mathbf{V}_i$ from (2.7).

Step 6. Compute the matrix $\mathbf{C}_i$ from (2.6) using $\mathbf{V}_i^{(1)}$ obtained in Step 5 and next compute the matrix $\boldsymbol{\Sigma}_i$ from (2.8).

Step 7. Repeat Steps 2 to 6 until convergence is attained. We have selected the following stopping rule. Compute two matrices: (a) a matrix of difference between two successive solutions of (2.7), and (b) a matrix of difference between two successive solutions of (2.8).

Continue the iterations until the maxima of the absolute values of the elements of the matrices
in (a) and (b) are smaller than the pre-specified quantities.

As noted in the literature, see, e.g. Galecki (1994) and Naik and Rao (2001), since

$$(c\mathbf{V}) \otimes (c^{-1}\boldsymbol{\Sigma}) = \mathbf{V} \otimes \boldsymbol{\Sigma},$$

all the parameters of $\mathbf{V}$ and $\boldsymbol{\Sigma}$ are not defined uniquely. Hence, Srivastava et al. (2008) took into consideration the maximum likelihood estimators of $\mathbf{V}$ and $\boldsymbol{\Sigma}$ under the restriction $v_{TT} = 1$, or equivalently under the restriction $\sigma_{pp} = 1$ for $\boldsymbol{\Sigma} = (\sigma_{ij})$ instead $v_{TT} = 1$ making the parameters of $\mathbf{V}$ and $\boldsymbol{\Sigma}$ unique.

For $\mathbf{V}_i = \left(v_{rs}^{(i)}\right)$, we only assume that $v_{TT}^{(i)} = 1$.

Let

$$\mathbf{X}_{ij} = \left(\mathbf{x}_{ij1}, \mathbf{x}_{ij2}, \dots, \mathbf{x}_{ijT}\right),$$

$$\overline{\mathbf{X}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{X}_{ij},$$

$$\mathbf{X}_{\mathrm{ijc}} = \mathbf{X}_{\mathrm{ij}} - \overline{\mathbf{X}}_{\mathrm{i}},$$

and

$$\mathbf{X}_{\mathrm{ijc}} = \left(\mathbf{X}_{\mathrm{ijc}1} : \mathbf{X}_{\mathrm{ijc}T}\right) : \left(p \times (T-1) : p \times 1\right),$$

j = 1,…, n_{i}, i = 1,…, K.

In this case the maximum likelihood estimation equations are of the form (Srivastava et al., 2008):

$$\hat{\boldsymbol{\mu}} = \overline{\mathbf{x}}_{\mathrm{i}} = \mathrm{vec}\left(\overline{\mathbf{X}}_{\mathrm{i}}\right),$$

$$\hat{\mathbf{V}}_{\mathrm{i}} = \frac{1}{n_{\mathrm{i}}\, p}\begin{bmatrix} \sum_{\mathrm{j}=1}^{n_{\mathrm{i}}} \mathbf{X}'_{\mathrm{ijc}1} \hat{\boldsymbol{\Sigma}}_{\mathrm{i}}^{-1} \mathbf{X}_{\mathrm{ijc}1} & \sum_{\mathrm{j}=1}^{n_{\mathrm{i}}} \mathbf{X}'_{\mathrm{ijc}1} \hat{\boldsymbol{\Sigma}}_{\mathrm{i}}^{-1} \mathbf{X}_{\mathrm{ijc}T} \\ \sum_{\mathrm{j}=1}^{n_{\mathrm{i}}} \mathbf{X}'_{\mathrm{ijc}T} \hat{\boldsymbol{\Sigma}}_{\mathrm{i}}^{-1} \mathbf{X}_{\mathrm{ijc}1} & \sum_{\mathrm{j}=1}^{n_{\mathrm{i}}} \mathbf{X}'_{\mathrm{ijc}T} \hat{\boldsymbol{\Sigma}}_{\mathrm{i}}^{-1} \mathbf{X}_{\mathrm{ijc}T} \end{bmatrix} = \frac{1}{n_{\mathrm{i}}\, p} \sum_{\mathrm{j}=1}^{n_{\mathrm{i}}} \mathbf{X}'_{\mathrm{ijc}} \hat{\boldsymbol{\Sigma}}_{\mathrm{i}}^{-1} \mathbf{X}_{\mathrm{ijc}}, \quad (2.10)$$

and

$$\hat{\boldsymbol{\Sigma}}_{\mathrm{i}} = \frac{1}{n_{\mathrm{i}} T} \sum_{\mathrm{j}=1}^{n_{\mathrm{i}}} \mathbf{X}_{\mathrm{ijc}} \hat{\mathbf{V}}_{\mathrm{i}}^{-1} \mathbf{X}'_{\mathrm{ijc}}, \quad (2.11)$$

subject to the condition

$$\sum_{\mathrm{j}=1}^{n_{\mathrm{i}}} \mathbf{X}'_{\mathrm{ijc}T} \hat{\boldsymbol{\Sigma}}_{\mathrm{i}}^{-1} \mathbf{X}_{\mathrm{ijc}T} = n_{\mathrm{i}}\, p, \ \mathrm{i} = 1, 2, \ldots, K. \quad (2.12)$$

In this case no explicit maximum likelihood estimates of $\mathbf{V}_{\mathrm{i}}$ and $\boldsymbol{\Sigma}_{\mathrm{i}}$ are available. The MLEs of $\mathbf{V}_{\mathrm{i}}$ and $\boldsymbol{\Sigma}_{\mathrm{i}}$ are obtained by solving simultaneously and iteratively the equations (2.10) and (2.11) subject to the condition (2.12). This is the so called ”flip-flop” algorithm.

The results given above are summarized in the following theorem:

Theorem 1 (Srivastava et al., 2008) *In the model (2.1) with* $v_{TT} = 1$ *, if* $n_{\mathrm{i}} > \max(p, T)$ *then the maximum likelihood estimation equations given by (2.10) and (2.11) subject to the condition (2.12) will always converge to the unique maximum.*

The following iterative steps are suggested to obtain the maximum likelihood estimates of $\mathbf{V}_i$ and $\mathbf{\Sigma}_i$, i = 1, 2,…, *K*.

Algorithm 2

Step 1. Get the initial covariance matrix $\mathbf{\Sigma}_i$ of the form (2.9), i = 1, 2,…, *K*.

Step 2. On the basis the initial covariance matrix $\mathbf{S}_i$ compute the matrix $\hat{\mathbf{V}}_i$ given by (2.10) and replace all the elements $\hat{v}_{rs}^{(i)}$ by $\hat{v}_{rs}^{(i)} / \hat{v}_{TT}^{(i)}$.

Step 3. Compute the matrix $\hat{\mathbf{\Sigma}}_i$ from the equation (2.11) using the $\hat{\mathbf{V}}_i$ obtained in Step 2.

Step 4. Repeat Steps 2 and 3 until convergence is attained. We have selected the same stopping rule as in Algorithm 1.

So we have two types of estimates of the matrices $\mathbf{V}_i$ and $\mathbf{\Sigma}_i$: without restrictions – as iterative solutions of equations (2.7) and (2.8) and with restrictions – as the iterative solutions of equations (2.10) and (2.11) subject to the condition (2.12).

In practice, if $\hat{\mathbf{V}}_i$ and $\hat{\mathbf{\Sigma}}_i$ are the estimates of the matrices $\mathbf{V}_i$ and $\mathbf{\Sigma}_i$ without restrictions, it is sufficient to divide each element of the matrix $\hat{\mathbf{V}}_i$ by $\hat{v}_{TT}^{(i)}$ and multiply every element in the matrix $\hat{\mathbf{\Sigma}}_i$ by $\hat{v}_{TT}^{(i)}$ to get the estimates of the matrices $\mathbf{V}_i$ and $\mathbf{\Sigma}_i$ with restrictions. Since the classifiers and the test functions considered in this paper do not depend on restrictions (2.12), we will not consider them further.

The form of the obtained classifier is presented in the following theorem.

Theorem 2 (Krzyśko and Skorzybut, 2009) *The classifier based on K training samples of sizes $n_1, n_2, \ldots, n_K$ from the respective classes has the form*

$$\hat{g}_1(\mathbf{x}) = \arg \max_{1 \le i \le K} \hat{\delta}_{i1}(\mathbf{x}),$$

where

$$\hat{\delta}_{i1}(\mathbf{x}) = -\frac{p}{2}\ln\left|\hat{\mathbf{V}}_i\right| - \frac{T}{2}\ln\left|\hat{\mathbf{\Sigma}}_i\right| - \frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}}_i)'\left(\hat{\mathbf{V}}_i^{-1} \otimes \hat{\mathbf{\Sigma}}_i^{-1}\right)(\mathbf{x}-\bar{\mathbf{x}}_i) + \ln\hat{\pi}_i \qquad (2.13)$$

is the quadratic classification function, $\bar{\mathbf{x}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\mathbf{x}_{ij}$, $\hat{\pi}_i = n_i / \sum_{j=1}^{K} n_j$, *and where* $\hat{\mathbf{V}}_i$ *and* $\hat{\mathbf{\Sigma}}_i$ *are obtained by solving simultaneously and iteratively the equations (2.10) and (2.11).*

Suppose now that $\mathbf{V}_1 = \mathbf{V}_2 = \ldots = \mathbf{V}_K = \mathbf{V}$, $\mathbf{\Sigma}_i = \mathbf{\Sigma}_2 = \ldots = \mathbf{\Sigma}_K = \mathbf{\Sigma}$. In this case the maximum likelihood estimation equations are of the form:

$$\hat{\boldsymbol{\mu}} = \overline{\mathbf{x}}_i = \mathrm{vec}\left(\overline{\mathbf{X}}_i\right),$$

$$\hat{\mathbf{V}} = \frac{1}{np}\sum_{i=1}^{K}\sum_{j=1}^{n_i}\mathbf{X}'_{ijc}\hat{\mathbf{\Sigma}}^{-1}\mathbf{X}_{ijc}, \qquad (2.14)$$

$$\hat{\mathbf{\Sigma}} = \frac{1}{nT}\sum_{i=1}^{K}\sum_{j=1}^{n_i}\mathbf{X}_{ijc}\hat{\mathbf{V}}^{-1}\mathbf{X}'_{ijc}, \qquad (2.15)$$

where

$$n_1 + n_2 + \ldots + n_K = n.$$

The starting value of $\hat{\mathbf{\Sigma}}$ can be based on the estimate

$$\mathbf{S} = \frac{1}{nT}\sum_{i=1}^{K}\sum_{j=1}^{n_i}\left(\mathbf{X}_{ij} - \overline{\mathbf{X}}_i\right)\left(\mathbf{X}_{ij} - \overline{\mathbf{X}}_i\right)'.$$

The form of the obtained classifier is presented in the following theorem.

Theorem 3 *The classifier based on K training samples of sizes n_1, n_2,…, n_K from the respective classes has the form*

$$\hat{g}_2(\mathbf{x}) = \arg\max_{1 \le i \le K}\hat{\delta}_{i2}(\mathbf{x}),$$

where

$$\hat{\delta}_{i2}(\mathbf{x}) = \mathbf{x}'\left(\hat{\mathbf{V}}^{-1} \otimes \hat{\mathbf{\Sigma}}^{-1}\right)\overline{\mathbf{x}}_i - \frac{1}{2}\overline{\mathbf{x}}'\left(\hat{\mathbf{V}}^{-1} \otimes \hat{\mathbf{\Sigma}}^{-1}\right)\overline{\mathbf{x}}_i + \ln\hat{\pi}_i \qquad (2.16)$$

is the linear classification function, $\overline{\mathbf{x}}_i$ is given by (2.13), $\hat{\pi}_i = n_i / n$, $n = n_1 + n_2 + \ldots + n_K$, and where $\hat{\mathbf{V}}$ and $\hat{\mathbf{\Sigma}}$ are obtained by solving simultaneously and iteratively the equations (2.14) and (2.15).

3. A mixed effects MANOVA model

Consider a mixed effects MANOVA model (similar to the split-plot design model of the univariate analysis of the usual repeated measures data) with the effects of the subjects within a class being random. Then the MANOVA table (which is similar to the ANOVA table for the split-plot design model) can be given as in Table 1 (Naik and Rao, 2001).

Table 1. MANOVA table for mixed effects model

Source	df	SS and CP	Distribution under H_0
Between Groups			
Groups	$K-1$	$\mathbf{Q}_1$	$W_p(K-1, \boldsymbol{\Sigma})$
Individuals	$n-K$	$\mathbf{Q}_2$	$W_p(n-K, \boldsymbol{\Sigma})$
Within Groups			
Time	$T-1$	$\mathbf{Q}_3$	$W_p(T-1, \boldsymbol{\Sigma})$
Time × Groups	$(T-1)(K-1)$	$\mathbf{Q}_4$	$W_p((T-1)(K-1), \boldsymbol{\Sigma})$
Error	$(T-1)(n-K)$	$\mathbf{Q}_5$	$W_p((T-1)(n-K), \boldsymbol{\Sigma})$
Total	$nT-1$	$\mathbf{X}(\mathbf{I}_{nT}-\frac{1}{nT}\mathbf{J}_{nT})\mathbf{X}'$	

Here **X**, the $nT \times p$ matrix, is defined as

$$\mathbf{X} = (\mathbf{x}_{111},\dots,\mathbf{x}_{11T},\dots,\ \mathbf{x}_{1n_1 1},\dots,\ \mathbf{x}_{1n_1 T},\dots,\ \mathbf{x}_{Kn_K 1},\dots,\ \mathbf{x}_{Kn_K T})'.$$

The matrix quadratic forms $\mathbf{Q}_1 - \mathbf{Q}_1$ are

$$\mathbf{Q}_1 = T\sum_{i=1}^{K} n_i (\bar{\mathbf{x}}_{i..} - \bar{\mathbf{x}}_{...})(\bar{\mathbf{x}}_{i..} - \bar{\mathbf{x}}_{...})' = \mathbf{X}'\mathbf{A}_1\mathbf{X},$$

$$\mathbf{Q}_2 = T\sum_{i=1}^{K}\sum_{j=1}^{n_i} (\bar{\mathbf{x}}_{ij.} - \bar{\mathbf{x}}_{i..})(\bar{\mathbf{x}}_{ij.} - \bar{\mathbf{x}}_{i..})' = \mathbf{X}'\mathbf{A}_2\mathbf{X},$$

$$\mathbf{Q}_3 = n\sum_{k=1}^{T} (\bar{\mathbf{x}}_{..k} - \bar{\mathbf{x}}_{...})(\bar{\mathbf{x}}_{..k} - \bar{\mathbf{x}}_{...})' = \mathbf{X}'\mathbf{A}_3\mathbf{X},$$

$$\mathbf{Q}_4 = \sum_{i=1}^{K} n_i \sum_{k=1}^{T} (\bar{\mathbf{x}}_{i.k} - \bar{\mathbf{x}}_{i..} - \bar{\mathbf{x}}_{..k} + \bar{\mathbf{x}}_{...})(\bar{\mathbf{x}}_{i.k} - \bar{\mathbf{x}}_{i..} - \bar{\mathbf{x}}_{..k} + \bar{\mathbf{x}}_{...})' = \mathbf{X}'\mathbf{A}_4\mathbf{X},$$

$$\mathbf{Q}_5 = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\sum_{k=1}^{T}\left(\overline{\mathbf{x}}_{ijk} - \overline{\mathbf{x}}_{ij.} - \overline{\mathbf{x}}_{i.k} + \overline{\mathbf{x}}_{i..}\right)\left(\overline{\mathbf{x}}_{ijk} - \overline{\mathbf{x}}_{ij.} - \overline{\mathbf{x}}_{i.k} + \overline{\mathbf{x}}_{i..}\right)' = \mathbf{X}'\mathbf{A}_5\mathbf{X}$$

with the appropriate choice of symmetric matrices $\mathbf{A}_1$ – $\mathbf{A}_5$ of order $nT \times nT$ and with the usual notations for the sample average. The matrices $\mathbf{A}_1$ – $\mathbf{A}_5$ can be easily derived (for example, see Geisser and Greenhouse, 1958). The matrix quadratic forms $\mathbf{Q}_1$ – $\mathbf{Q}_5$ are independent of each other and under the appropriate null hypothesis each has a scale multiple of a Wishart distribution with appropriate degrees of freedom. See the work of Khatri (1962), Arnold (1979), Reinsel (1982), and Mathew (1989) in this regard.

Suppose we wish to test H_{01} of no class effect. Then the Wilks' Λ for testing H_{01} is (Naik and Rao, 2001)

$$\Lambda_1 = \frac{|\mathbf{Q}_2|}{|\mathbf{Q}_1 + \mathbf{Q}_2|}. \tag{3.1}$$

We have

$$-\left[n - K - 1 - \frac{1}{2}(p - K)\right]\ln\Lambda_1 \sim \chi^2_{p(K-1)} \text{ approximately,}$$

where $n = n_1 + n_2 + \ldots + n_K$.

For testing H_{02}, that there is no time effect, one can use the Wilks' Λ, which is

$$\Lambda_2 = \frac{|\mathbf{Q}_5|}{|\mathbf{Q}_3 + \mathbf{Q}_5|} \tag{3.2}$$

and the fact that

$$-\left[(\mathrm{n} - \mathrm{K})\mathrm{h} - 1 - \frac{1}{2}(\mathrm{p} + 1 - \mathrm{h})\right]\ln\Lambda_2 \sim \chi^2_{\mathrm{ph}} \text{ approximately,} \tag{3.3}$$

where

$$h = \frac{\left[\mathrm{tr}\left(\mathbf{V} - \frac{1}{T}\mathbf{JV}\right)\right]^2}{\mathrm{tr}\left(\mathbf{V} - \frac{1}{T}\mathbf{JV}\right)^2}.$$

Similarly, for testing H_{03}, that there is no time and group interaction, the Wilks' Λ is

$$\Lambda_3 = \frac{|\mathbf{Q}_5|}{|\mathbf{Q}_4 + \mathbf{Q}_5|} \tag{3.4}$$

and the distribution of the test statistic is

$$-\left[(n-K)h - \frac{1}{2}(p+1-(K-1)h)\right]\ln \Lambda_3 \sim \chi^2_{p(K\text{-}1)h} \text{ approximately.} \tag{3.5}$$

Since in practice **V** is unknown, the degrees of freedom in the χ^2 approximations of (3.3) and (3.5) are unknown. One needs to estimate these so that the distributions in (3.3) and (3.5) can be utilized in applications. For estimating these degrees of freedoms, which are functions of $\mathbf{V} - \frac{1}{T}\mathbf{JV}$, we simply need an estimate of **V**. One can use the maximum likelihood estimate of **V** that is obtained by simultaneously solving the equations (2.14) and (2.15).

4. Example

Plant material. Studies were conducted on 4–6-year-old bushes of the blackcurrant (*Ribes nibrum* L.) being in the full fruit-bearing phase. They were grown in a cultivar trail established in the autumn of 1996 at the Experimental Orchard in Dabrowice (near Skierniewice, central Poland). The field experiment was established in a randomized complete block design in 3 replicates (with 5 plants on every plot). Fourteen genotypes were evaluated, including 8 cultivars with diverse features and originating from different geographical areas of Europe (Pluta 1994, 1996, Pomologia – annex 2003) and 6 breeding clones released at the Institute of Pomology and Floricultures (ISK) in Skierniewice. The list of studied genotypes, their parentage and countries of origin are presented in Table 2. The studies were carried out in 1999–2001.

Table 2. List of blackcurrant genotypes studied in the experiments, their parentage and country of origin

No.	Name of cultivar/ breeding clone	Parentage	Country of Origin
1	Ojebyn	unknown	Sweden
2	Titania	Ałtajskaja D. x (Consort x Kajaanin Musta)	Sweden
3	Ben Lomond	(Consort x Mangus) x (Brodtorp x Janslunda)	Scotland
4	Lentaj	Brodtorp x Minaj Szmyriew	Russia
5	Sjuta Kijewskaja	(Junost x Zoja) x Minaj Szmyriew	Ukraine
6	Czereszniewa	[B-36-16 x (Junost x Zoja)] x [(Minaj Szmyriew x Biełoruskaja Słodkaja)]	Ukraine
7	Czornyj Żemczug	Minaj Szmyriew x Brodtorp	Russia
8	Sanjuta	(Junost x Zoja) x Minaj Szmyriew	Ukraine
9	(PC-1) Gofert	Gołubka x Fert?di-1	Poland
10	PC-3	Biełoruskaja Słodkaja x Titania	Poland
11	PC-8	Smuglianka x Westwick Triumf	Poland
12	PC-9	Ojebyn x Titania	Poland
13	PC-20	Ben Lomond x 7/72	Poland
14	PC-23	Ben Lomond x 7/72	Poland

Measurement and observations. In each year of evaluation, measurements were performed on 210 plants in the experiment (14 genotypes x 15 plants). On every individual plant the following measurements and observations of features were carried out:

1. Number of one-year-old shoots (per bush)
2. Number of strigs on one-year-old bush
3. Number of flowers in raceme
4. Number of fruits in the strig
5. Number of fruits on one-year-old shoots
6. Weight of 100 fruits (g)
7. Fruit yield (kg/bush).

Statistical analysis. A justification for using one-way MANOVA for analysis of data from an experiment carried out in a randomized complete block design (in this experimental design, data are arranged in a two-way classification with one observation in subclasses) is the fact that the effects of blocks in the corresponding ANOVA model were not significant for most traits. To investigate the discriminating power of individual features sequentially, quadratic and linear classifiers were used. First the misclassification errors were found for individual characteristics using the quadratic classifier (2.13). Then the characteristic with the lowest misclassification error was augmented with further characteristics. The pair of characteristics with the lowest misclassification error was augmented with other characteristics, and so on.

It turned out that the optimal subset is a subset of characteristics (7, 6, 2, 5, 3, 4) for which the misclassification error is equal to 4.29%.

Augmenting this subset with characteristic number 1 increases the misclassification error. If we use the linear classifier (2.16), the optimal subset is a subset of characteristics (7, 6, 2, 5, 2). In this case, the misclassification error is equal to 6.67%.

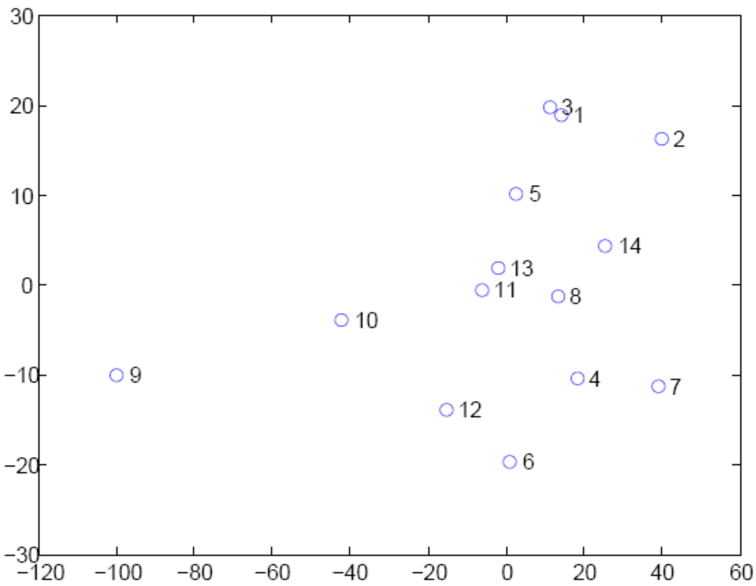

Fig. 1. Plotted values of the first two principal components for the mean of 14 genotypes

Let us now consider the hypothesis H_{01} of the non-differentiation of 14 genotypes of blackcurrants with respect to the 7 observed characteristics. The value of the statistic given in (3.1) is $\Lambda_1 = 0.0091$. Then the test statistic value is equal to 937.5464. Comparing this with $\chi^2_{91}(0.05) = 114.2679$, we clearly reject H_{01}. This means that the tested genotypes differ significantly in terms of at least one characteristic.

Another hypothesis that we want to verify is the hypothesis H_{02}, that there is no time effect. The value of the statistic given in (3.2) is $\Lambda_2 = 0.0952$. Then the test statistic value is equal to 9852.7594. Comparing this with $\chi^2_{13}(0.05) = 22.3620$, we clearly reject H_{02}. This means that during the observation years, the characteristics changed their values (at least one).

The hypothesis H_{03} says that there is no time and genotype interaction. In this case the value of the statistic given in (3.4) is $\Lambda_3 = 0.0395$. Then the test statistic value is equal to 1207.2. Comparing this with $\chi^2_{170}(0.05) = 201.4234$, we clearly reject H_{03}.

Plotted values of the first two principal components given by Deręgowski and Krzyśko (2009) for the mean of 14 genotypes are shown in Figure 1.

References

Arnold S. F. (1979). Linear models with exchangeably distributed errors. *Journal of American Statistical Association* 74, 194–199.

Deręgowski K., Krzyśko M. (2009). Principal component analysis in the case of multivariate repeated measures data. *Biometrical Letters* 46, 163–172.

Galecki A. T. (1994). General class of covariance structures for two or more repeated factors in longitudinal data analysis. *Comun. Statist. – Theory and Methods* 23, 3105–3119.

Geisser S., Greenhouse, S. (1958). An extension of Box's results on the use of the F distribution in multivariate analysis. *Annals of Mathematical Statistics* 29, 885–891.

Giri N. C. (1996). *Multivariate Statistical Analysis*. Marcel Dekker, Inc., New York.

Khatri C. G. (1962). Conditions for Wishartness and independence of second degree polynomials in normal vector. *Annals of Mathematical Statistics* 33, 1002–1007.

Krzyśko M., Skorzybut M. (2009). Discriminant analysis of multivariate repeated measures data with a Kronecker product structured covariance matrices. *Statistical Papers* 50, 817–835.

Mathew T. (1989). MANOVA in the multivariate components of variance model. *Journal of Multivariate Analysis* 29, 30–38.

Naik D. N., Rao S. (2001). Analysis of multivariate repeated measures data with a Kronecker product structured covariance matrix. *J. Appl. Statist*. 28, 91–105.

Pluta S. (1994). *Analiza dialleliczna cech użytkowych porzeczki czarnej*. Praca doktorska, ISK Skierniewice

Pluta S. (1996). *Zeszyty Pomologiczne. Porzeczki i Agrest*. Wyd. ISK, 1–89. *Pomologia odmianoznawstwo roślin sadowniczych – aneks. 2003. Porzeczka czarna. Porzeczka czerwona. Agrest.* Praca zbiorowa pod red. E. Żurawicza, PWRiL, Warszawa 2003, 165–173; 175–180; 182–188.

Reinsel G. (1982). Multivariate repeated measurements or growth curve models with multivariate random-effects covariance structure. *Journal of American Statistical Association* 77, 190–195.

Roy A., Khattree R. (2005a). Discrimination and classification with repeated measures data under different covariance structures. *Communications in Statistics-Simulation and Computation* 34, 167–178.

Roy A., Khattree R. (2005b). On discrimination and classification with multivariate repeated measures data. *Journal of Statistical Planning and Inference* 134, 462–485.

Roy A., Khattree R. (2008). Classification rules for repeated measures data from biomedical research. In: Khattree R., Naik D. N. (eds) *Computational methods in biomedical research*. Chapman and Hall/CRC, 323–370.

Srivastava M. S., von Rosen T., von Rosen D. (2008). Models with a Kronecker product covariance structure: estimation and testing. *Math. Methods Stat*. 17(4), 357–370.

ANALIZA DANYCH WIELOWYMIAROWYCH POCHODZĄCYCH Z POWTARZANYCH POMIARÓW NA TYCH SAMYCH JEDNOSTKACH

Streszczenie

W artykule tym zaproponowano nowe klasyfikatory liniowe i kwadratowe skonstruowane na podstawie danych wielowymiarowych pochodzących z powtarzanych pomiarów (danych podwójnie wielowymiarowych). Rozpatrywany jest również model mieszany wielowymiarowej analizy wariancji. Jakość rozpatrywanych wielowymiarowych metod statystycznych jest weryfikowana na danych pochodzących z wieloletniego doświadczenia z czarną porzeczką.

Słowa kluczowe: klasyfikatory, dane pochodzące z powtarzanych pomiarów (dane podwójnie wielowymiarowe), macierze kowariancji o strukturze iloczynu Kroneckera, model mieszany wielowymiarowej analizy wariancji.

Klasyfikacja AMS 2010: 62H30, 62H15, 62H12, 62J10