



Analysis of multivariate repeated measures data using a MANOVA model and principal components

Mirosław Krzyśko¹, Tadeusz Śmiałowski², Waldemar Wołyński¹

¹Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Umultowska 87, 61-614 Poznań, Poland, e-mail: mkrzyisko@amu.edu.pl, wolyński@amu.edu.pl

²Plant Breeding and Acclimatization Institute, National Research Institute, Radzików, 05-870 Błonie, Poland, e-mail: zshmialo@cyf-kr.edu.pl

SUMMARY

In this paper we consider a set of T repeated measurements on p characteristics on each of n individuals. The n individuals themselves may be divided and randomly assigned to K groups. These data are analyzed using a mixed effect MANOVA model, assuming that the data on an individual have a covariance matrix which is a Kronecker product of two positive definite matrices. Results are illustrated on a data set obtained from experiments with varieties of winter rye.

Key words: multivariate repeated measures data (doubly multivariate data), Kronecker product covariance structure, maximum likelihood estimates, mixed MANOVA model, principal component analysis

1. Introduction

Measurements on a set of p variables (or characteristics) made at several T time points, or under different treatment conditions on the same experimental unit, lead to repeated measures (or longitudinal) data. A typical set of repeated measures data is usually taken on n ($= n_1 + \dots + n_K$) individuals forming K groups over T time points. The problems of interest are to test for (i) the time effect, (ii) the group effect, and (iii) the effect of interaction between time and group.

Analysis of such data is complicated by the existence of correlation among the measurements of different variables as well as correlation among measurements taken at different time points. Several approaches to analyzing these data exist in the literature. In Section 2 we consider a mixed effect MANOVA model with the effects of the subjects within a group being random (Naik and Rao, 2001). We begin with the assumption that the

covariance matrix $\mathbf{\Omega}$ has the structure $\mathbf{V} \otimes \mathbf{\Sigma}$, where \mathbf{V} and $\mathbf{\Sigma}$ are respectively $T \times T$ and $p \times p$ positive definite matrices. The number of unknown parameters of the covariance matrix, in our formulation, is much smaller, $[T(T+1)/2 + p(p+1)/2]$, as opposed to $[pT(pT+1)/2]$ in the case of the general covariance structure.

In Section 3 we derive the maximum likelihood estimators of \mathbf{V} and $\mathbf{\Sigma}$. In Section 4 the principal components are constructed on the basis of the matrix $\hat{\mathbf{V}} \otimes \hat{\mathbf{\Sigma}}$. Next, in Section 5, as an illustration, we apply these results to an example data set.

2. A mixed effects MANOVA model

Let \mathbf{x}_{ijk} , $i = 1, \dots, K$, $j = 1, \dots, n_i$, $k = 1, \dots, T$ be a $p \times 1$ vector of measurements on the j th individual in the i th group at the k th time point, and let $\mathbf{x}_{ij} = (\mathbf{x}'_{ij1}, \dots, \mathbf{x}'_{ijT})'$. Then \mathbf{x}_{ij} is a $pT \times 1$ random observational vector corresponding to the j th individual in the i th group.

Let $\mathbf{x}_{ij} \sim N_{pT}(\boldsymbol{\mu}_i, \mathbf{\Omega})$, for $j = 1, \dots, n_i$, $i = 1, \dots, K$, where $\mathbf{\Omega}$ is a positive definite matrix. Its estimate $\hat{\mathbf{\Omega}}$ is positive definite with probability one if and only if $n > pT$, where $n = n_1 + \dots + n_K$ (see e.g. Giri, 1996, p. 93). Hence estimation of the matrix $\mathbf{\Omega}$ will require a very large sample, which may not always be feasible. We therefore assume $\mathbf{\Omega}$ to be of the form (Roy and Khattree, 2005a, 2005b, 2008):

$$\mathbf{V} \otimes \mathbf{\Sigma},$$

where \mathbf{V} is a $T \times T$ positive definite covariance matrix and $\mathbf{\Sigma}$ is a $p \times p$ positive definite covariance matrix. The matrix \mathbf{V} represents the covariance between repeated measures on a given subject and for a given characteristic. Likewise, $\mathbf{\Sigma}$ 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 \mathbf{V} and $\mathbf{\Sigma}$ are positive definite with probability one if and only if $n > \max(p, T)$.

One method used to analyze multivariate repeated measures data is a mixed effect 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 group being random. Then the MANOVA table 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	$W_p(K - 1, \Sigma)$
Individuals	$n - K$	\mathbf{Q}_2	$W_p(n - K, \Sigma)$
Within Groups			
Time	$T - 1$	\mathbf{Q}_3	$W_p(T - 1, \Sigma)$
Time \times Groups	$(T - 1)(K - 1)$	\mathbf{Q}_4	$W_p((T - 1)(K - 1), \Sigma)$
Error	$(T - 1)(n - K)$	\mathbf{Q}_5	$W_p((T - 1)(n - K), \Sigma)$
Total	$nT - 1$	$\mathbf{X}(\mathbf{I}_{nT} - \frac{1}{nT}\mathbf{J}_{nT})\mathbf{X}'$	

Here \mathbf{X} , the $nT \times p$ matrix, is defined as

$$\mathbf{X} = (\mathbf{x}_{111}, \dots, \mathbf{x}_{11T}, \dots, \mathbf{x}_{1n_11}, \dots, \mathbf{x}_{1n_1T}, \dots, \mathbf{x}_{Kn_K1}, \dots, \mathbf{x}_{Kn_KT})'$$

The matrix quadratic forms $\mathbf{Q}_1, \dots, \mathbf{Q}_5$ are

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

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

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

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

$$\begin{aligned} \mathbf{Q}_5 &= \sum_{i=1}^K \sum_{j=1}^{n_i} \sum_{k=1}^T (\bar{\mathbf{x}}_{ijk} - \bar{\mathbf{x}}_{ij.} - \bar{\mathbf{x}}_{i.k} + \bar{\mathbf{x}}_{i..}) (\bar{\mathbf{x}}_{ijk} - \bar{\mathbf{x}}_{ij.} - \bar{\mathbf{x}}_{i.k} + \bar{\mathbf{x}}_{i..})' = \\ &= \mathbf{X}' \mathbf{A}_5 \mathbf{X} \end{aligned}$$

with the appropriate choice of symmetric matrices $\mathbf{A}_1, \dots, \mathbf{A}_5$ of order $nT \times nT$ and with the usual notations for the sample averages. The matrices $\mathbf{A}_1, \dots, \mathbf{A}_5$ can be easily derived (see for example Geisser and Greenhouse, 1958). The matrix quadratic forms $\mathbf{Q}_1, \dots, \mathbf{Q}_5$ 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 want to test H_{01} , the hypothesis that there is no group effect. Then the Wilks' Λ for testing H_{01} is (Naik and Rao, 2001)

$$\Lambda_1 = \frac{|\mathbf{Q}_2|}{|\mathbf{Q}_1 + \mathbf{Q}_2|}. \quad (1)$$

We have

$$-[n - K - 1 - \frac{1}{2}(p - K)] \ln \Lambda_1 \sim \chi_{p(K-1)}^2 \quad \text{approximately}, \quad (2)$$

where $n = n_1 + \dots + n_K$.

To test H_{02} , the hypothesis that there is no time effect, one can use the following Wilks' Λ :

$$\Lambda_2 = \frac{|\mathbf{Q}_5|}{|\mathbf{Q}_3 + \mathbf{Q}_5|} \quad (3)$$

and the fact that

$$-[(n - K)h - \frac{1}{2}(p + 1 - h)] \ln \Lambda_2 \sim \chi_{ph}^2 \quad \text{approximately}, \quad (4)$$

where

$$h = \frac{[\text{tr}(\mathbf{V} - \frac{1}{T}\mathbf{J}\mathbf{V})]^2}{\text{tr}(\mathbf{V} - \frac{1}{T}\mathbf{J}\mathbf{V})^2}.$$

Similarly, to test H_{03} , the hypothesis that there is no time and group interaction, the Wilks' Λ is

$$\Lambda_3 = \frac{|\mathbf{Q}_5|}{|\mathbf{Q}_4 + \mathbf{Q}_5|} \quad (5)$$

and the distribution of the test statistic is

$$-[(n - K)h - \frac{1}{2}(p + 1 - (K - 1)h)] \ln \Lambda_3 \sim \chi_{p(K-1)h}^2 \quad \text{approximately}. \quad (6)$$

Since, in practice, \mathbf{V} is unknown, the degrees of freedom in the χ^2 approximations of (4) and (6) are unknown. One needs to estimate these so that the distributions in (4) and (6) can be utilized in applications. To estimate these degrees of freedoms, which are functions of $\mathbf{V} - \frac{1}{T}\mathbf{J}\mathbf{V}$, we simply need an estimate of \mathbf{V} .

3. The maximum likelihood estimators

We consider a model described as follows: all observations \mathbf{x}_{ij} are independent and

$$\mathbf{x}_{ij} \sim N_{pT}(\boldsymbol{\mu}_i, \mathbf{V} \otimes \boldsymbol{\Sigma}), \quad (7)$$

where \mathbf{V} is a $T \times T$ positive definite matrix and $\boldsymbol{\Sigma}$ is a $p \times p$ positive definite matrix $i = 1, \dots, K, j = 1, \dots, n_i, n = n_1 + \dots + n_K > \max(p, T)$. In this case the maximum likelihood estimation equations are of the form (Srivastava et al., 2008):

$$\hat{\boldsymbol{\mu}}_i = \text{vec}(\bar{\mathbf{X}}_{i.}) \quad (8)$$

$$\hat{\mathbf{V}} = \frac{1}{np} \sum_{i=1}^K \sum_{j=1}^{n_i} (\mathbf{X}_{ij} - \bar{\mathbf{X}}_{i.})' \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{X}_{ij} - \bar{\mathbf{X}}_{i.}) \quad (9)$$

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{nT} \sum_{i=1}^K \sum_{j=1}^{n_i} (\mathbf{X}_{ij} - \bar{\mathbf{X}}_{i.}) \hat{\mathbf{V}}^{-1} (\mathbf{X}_{ij} - \bar{\mathbf{X}}_{i.})' \quad (10)$$

where \mathbf{X}_{ij} is a $p \times T$ matrix of the form

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

$i = 1, \dots, K, j = 1, \dots, n_i, n = n_1 + \dots + n_K$.

In this case no explicit maximum likelihood estimates of \mathbf{V} and $\boldsymbol{\Sigma}$ are available. The MLEs of \mathbf{V} and $\boldsymbol{\Sigma}$ are obtained by solving simultaneously and iteratively equations (9) and (10). This is the so-called "flip-flop" algorithm. The starting value of $\hat{\boldsymbol{\Sigma}}$ can be based on the estimate

$$\mathbf{S} = \frac{1}{nT} \sum_{i=1}^K \sum_{j=1}^{n_i} (\mathbf{X}_{ij} - \bar{\mathbf{X}}_{i.})(\mathbf{X}_{ij} - \bar{\mathbf{X}}_{i.})'. \quad (11)$$

The following iterative steps are suggested to obtain the maximum likelihood estimates of \mathbf{V} and $\boldsymbol{\Sigma}$.

Algorithm

Step 1 Get the initial covariance matrix $\hat{\boldsymbol{\Sigma}}$ in the form (11).

Step 2 On the basis of the covariance matrix $\hat{\boldsymbol{\Sigma}}$ compute the matrix $\hat{\mathbf{V}}$ given by (9).

Step 3 Compute the matrix $\hat{\boldsymbol{\Sigma}}$ from equation (10) using the $\hat{\mathbf{V}}$ obtained in Step 2.

Step 4 Repeat Steps 2 and 3 until convergence is attained.

We select the following stopping rule. Compute two matrices: (a) the difference between two successive solutions of (9), and (b) the difference between two successive solutions of (10). Continue the iterations until the maxima of the absolute values of the elements of the matrices (a) and (b) are smaller than pre-specified quantities.

4. Principal component analysis

Principal components are constructed on the basis of the matrix $\hat{\mathbf{V}} \otimes \hat{\mathbf{\Sigma}}$ (Deręgowski and Krzyśko, 2009). If $n > \max(p, T)$ then $\hat{\mathbf{V}} \otimes \hat{\mathbf{\Sigma}}$ is a positive definite matrix with probability one, and all of the characteristic roots/eigenvalues of $\hat{\mathbf{V}} \otimes \hat{\mathbf{\Sigma}}$ are real and positive. One of the main reasons for interest in the Kronecker product is the beautifully simple connection between the eigenvalues and the eigenvectors $\hat{\mathbf{V}}$ and $\hat{\mathbf{\Sigma}}$ and $\hat{\mathbf{V}} \otimes \hat{\mathbf{\Sigma}}$ (see e.g. Lancaster and Tismenetsky, 1985, p. 412 or Ortega, 1987, p. 237). If $\alpha_1, \alpha_2, \dots, \alpha_T$ are the eigenvalues of $\hat{\mathbf{V}}$ and $\beta_1, \beta_2, \dots, \beta_p$ are the eigenvalues of $\hat{\mathbf{\Sigma}}$, then eigenvalues of $\hat{\mathbf{V}} \otimes \hat{\mathbf{\Sigma}}$ are the pT numbers $\alpha_r \beta_s$, $r = 1, 2, \dots, T$, $s = 1, 2, \dots, p$.

If $\mathbf{u} = (u_1, u_2, \dots, u_T)'$ is an eigenvector of $\hat{\mathbf{V}}$ corresponding to the eigenvalue α and $\mathbf{w} = (w_1, w_2, \dots, w_p)'$ is an eigenvector of $\hat{\mathbf{\Sigma}}$ corresponding to the eigenvalue β , then an eigenvector $\boldsymbol{\gamma}$ of $\hat{\mathbf{V}} \otimes \hat{\mathbf{\Sigma}}$ associated with $\alpha\beta$ is $\boldsymbol{\gamma} = \mathbf{u} \otimes \mathbf{w} = (u_1 \mathbf{w}', u_2 \mathbf{w}', \dots, u_T \mathbf{w}')'$.

Let $\lambda_1 > \lambda_2 > \dots > \lambda_{pT} > 0$ be the ordered characteristic roots of $\hat{\mathbf{V}} \otimes \hat{\mathbf{\Sigma}}$. Then there exists an orthogonal matrix $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \dots, \boldsymbol{\gamma}_{pT})$, $\boldsymbol{\Gamma} \boldsymbol{\Gamma}' = \mathbf{I}_{pT}$, such that $\boldsymbol{\Gamma}' \hat{\boldsymbol{\Omega}} \boldsymbol{\Gamma} = \mathbf{D}_\lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{pT})$.

Hence if we let

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{pT} \end{bmatrix} = \boldsymbol{\Gamma}' \mathbf{x} = \begin{bmatrix} \boldsymbol{\gamma}'_1 \\ \boldsymbol{\gamma}'_2 \\ \vdots \\ \boldsymbol{\gamma}'_{pT} \end{bmatrix} \mathbf{x} = \begin{bmatrix} \boldsymbol{\gamma}'_1 \mathbf{x} \\ \boldsymbol{\gamma}'_2 \mathbf{x} \\ \vdots \\ \boldsymbol{\gamma}'_{pT} \mathbf{x} \end{bmatrix}$$

then $\text{Cov}(\mathbf{y}) = \mathbf{D}_\lambda$, and the components $y_1 = \boldsymbol{\gamma}'_1 \mathbf{x}$, $y_2 = \boldsymbol{\gamma}'_2 \mathbf{x}$, \dots , $y_{pT} = \boldsymbol{\gamma}'_{pT} \mathbf{x}$ are uncorrelated. The component $y_1 = \boldsymbol{\gamma}'_1 \mathbf{x}$ is called the first principal component, $y_2 = \boldsymbol{\gamma}'_2 \mathbf{x}$ the second principal component, and so on. The variance of y_i is λ_i . Since $\lambda_1 + \lambda_2 + \dots + \lambda_{pT} = \text{tr}(\hat{\mathbf{V}} \otimes \hat{\mathbf{\Sigma}})$, the sum of the variance of the pT principal components is the same as the sum of the variance of the old variables, measured at T different time points. Thus

Table 2. Characteristics of varieties of winter rye

Characteristic	Unit of measurement
Grain yield per plot	g
Winterhardiness	scale (1–9)
Heading	number*
Plant height	cm
1000 grain weight	g
Falling number	s

* number of days from the first of May to heading.

the components with smaller variances can be ignored without significantly affecting the total variance, thereby reducing the number of variables from pT to, say, $k \leq pT$.

Since $\text{tr}(\hat{\mathbf{V}} \otimes \hat{\mathbf{\Sigma}}) - \sum_{j=1}^k \lambda_j = \sum_{j=k+1}^{pT} \lambda_j$, the proportion of the total univariate variance accounted for by k PCs is $\sum_{j=1}^k \lambda_j / \sum_{j=1}^{pT} \lambda_j$, which may be used as a criterion for selecting a subspace of k components from pT .

5. Example

An experiment with populations of different varieties of winter rye was carried out under the project “Analysis of genotypic and environmental variation and genetic determinants of important characteristics in cereals” commissioned by the Polish Ministry of Agriculture and Rural Development (**HOR hn 801-24/13**), and specifically under task **3**: “Analysis of genotypic and environmental variation and genetic determinants of important characteristics in rye.”

Multi-year field and laboratory experiments were carried out in Kraków. A total of 100 caryopses were sown pointwise in 1 m^2 plots, using four repetitions. During vegetation, observations and evaluations were made concerning yield, winterhardiness and susceptibility to disease, measurements were made of plant height, and the time of heading was recorded as a measure of earliness. Grains were harvested from the plots using a HEGE 125 micro-harvester. The harvested grains were weighed at 15% humidity, the weight of 1000 grains was determined, and the falling number was analyzed as a measure of resistance to premature sprouting prior to harvesting and as a measure of the baking quality of the flour. (See Table 2.)

The experiment was conducted on the nine most valuable open-pollinating varieties of winter rye from the register of cultivable varieties, as well as two special populations, SMH-49 and SMH-75, which are used as testers

in genetic studies (see Table 3). The varieties of winter rye are described (with reference to a single-year study) in Ukalski and Śmiałowski (2011).

Table 3. Varieties of winter rye

No.	Name
1.	Dańkowskie Złote
2.	Dańkowskie Diament
3.	Słowiańskie
4.	Bosmo
5.	Agrikolo
6.	Hegro
7.	Kier
8.	SMH-49
9.	SMH-75
10.	Amilo
11.	Rostockie

For each of the six characteristics separately, and for all characteristics simultaneously, the following hypotheses were tested:

1. no variety effect;
2. no year effect;
3. no interaction between varieties and years.

The values of the respective test functions and numbers of degrees of freedom are given in Tables 4-9.

Table 4. Approximate MANOVA (grain yield per plot)

Hypothesis	Wilks' Λ	χ^2	df	p -value
Varieties	0.6801	14.26	10.00	0.1613
Years	0.2325	77.65*	1.65	0.0000
Years \times Varieties	0.7675	16.31	16.48	0.4657

Table 5. Approximate MANOVA (winterhardiness)

Hypothesis	Wilks' Λ	χ^2	df	p -value
Varieties	0.5555	21.75	10.00	0.0164
Years	0.5732	30.60*	1.70	0.0000
Years \times Varieties	0.8661	9.15	17.01	0.9357

The values of the test function and degrees of freedom for the six characteristics jointly are given in Table 10.

Table 6. Approximate MANOVA (heading)

Hypothesis	Wilks' Λ	χ^2	df	p -value
Varieties	0.3526	38.57*	10.00	0.0000
Years	0.0240	212.41*	1.76	0.0000
Years \times Varieties	0.5506	39.30*	17.59	0.0021

Table 7. Approximate MANOVA (plant height)

Hypothesis	Wilks' Λ	χ^2	df	p -value
Varieties	0.3773	36.07*	10.00	0.0000
Years	0.4630	47.34*	1.89	0.0000
Years \times Varieties	0.7219	23.15	18.95	0.2286

Table 8. Approximate MANOVA (1000 grain weight)

Hypothesis	Wilks' Λ	χ^2	df	p -value
Varieties	0.5031	25.42*	10.00	0.0046
Years	0.5567	34.99*	1.84	0.0000
Years \times Varieties	0.7157	23.09	18.43	0.2065

Table 9. Approximate MANOVA (falling number)

Hypothesis	Wilks' Λ	χ^2	df	p -value
Varieties	0.4508	29.48*	10.00	0.0010
Years	0.0589	148.97*	1.63	0.0000
Years \times Varieties	0.4736	45.54*	16.30	0.0001

* significant at $\alpha = 0.01$.**Table 10.** Approximate MANOVA

Hypothesis	Wilks' Λ	χ^2	df	p -value
Varieties	0.0098	159.50*	60.00	0.0000
Years	0.0018	376.28*	11.46	0.0000
Years \times Varieties	0.0784	175.91*	114.64	0.0002

Next, for each characteristic separately and for all characteristics simultaneously, principal components were constructed, and the varieties were presented as points in the system based on the first two principal components. A minimum spanning tree (dendrite) (Florek et al. (1951) and Kruskal (1956)) was constructed on these points (see Figures 1 and 2). The mean and the standard deviation of the lengths of edges in the dendrite were computed. Edges whose length is twice as great as the mean plus two standard deviations are shown with a dotted line. This indicates the division

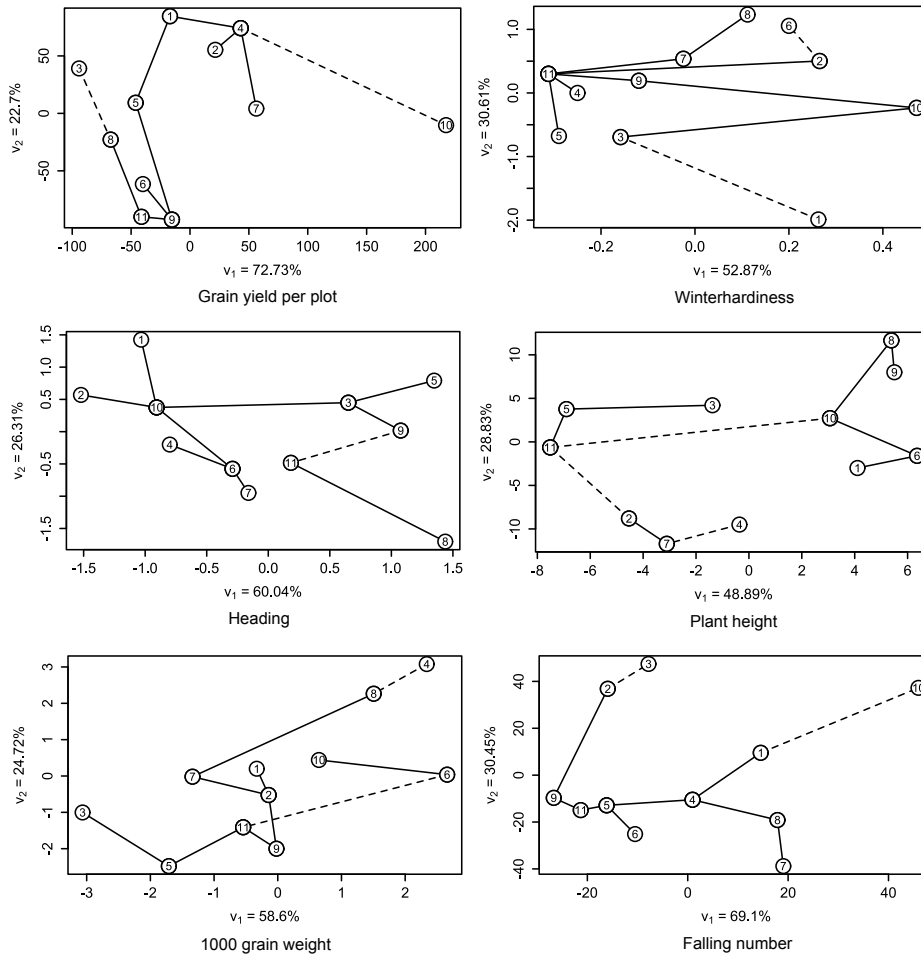


Figure 1. Projection of the 11 selected varieties of winter rye onto the first two principal components, with dendrite (each characteristic separately)

of the varieties into homogeneous groups.

In the case of the six characteristics considered simultaneously, in three consecutive years, the 11 varieties form three homogeneous groups:

1. variety no. 10 (Amilo);
2. variety no. 3 (Słowiańskie);
3. all other varieties.

Słowiańskie is a winter rye variety bred under a special experimental program which aims to increase resistance to fungal diseases of the stalk leaves,

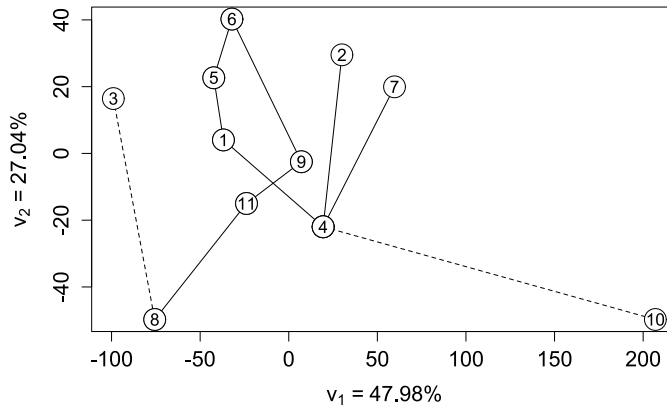


Figure 2. Projection of the 11 selected varieties of winter rye onto the first two principal components, with dendrite (all characteristics simultaneously)

and having a short straw to increase resistance to lodging. This means that in extreme years with significant amounts of rainfall, when the conditions favor the occurrence of diseases and lodging, this variety demonstrates its superiority to other open-pollinating varieties of winter rye.

Amilo is a winter rye variety bred under a special experimental program which aims to increase resistance to premature grain sprouting. It produces grains with superior baking qualities. Amilo is based on source genetic materials originating in Sweden, in particular the Othello variety.

REFERENCES

- Arnold S. F. (1979): Linear models with exchangeably distributed errors. *Journal of American Statistical Association* 74: 194–199.
- Deregowski K., Krzyśko M. (2009): Principal component analysis in the case of multivariate repeated measures data. *Biometrical Letters* 46: 163–172.
- Florek K., Łukaszewicz J., Perkal J., Steinhaus H. and Zubrzycki S. (1951): Sur la liaison et la division des points d'un ensemble fini. *Colloquium Mathematicum* 2: 282–285.
- 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.
- Kruskal J. B. (1956): On the shortest spanning subtree of a graph and the traveling salesman problem. *Proc. Amer. Math. Soc.* 7: 48–50.
- Lancaster P., Tismenetsky M. (1985): *The Theory of Matrices, Second Edition: With Applications.* Academic Press, Orlando.
- 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.
- Ortega J. M. (1987): *Matrix Theory: A Second Course.* Plenum Press, New York.
- 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.
- Ukalski K., Śmiałowski T. (2011): Multivariate analysis of data from preliminary trials with winter rye. *Biuletyn Instytutu Hodowli i Aklimatyzacji Roślin* 260/261: 251–262.