Konstantin Momirović (1932-2004): Methodological Contributions to Multivariate Statistics and Data Analysis

Franjo Prot[1]
Vesna Lužar-Stiffler[2]
Vesna Hljuz Dobrić[3]
Ksenija Bosnar[4]
Zoran Bekić[5]
Marijan Gredelj[6]

[1,4]*Faculty of Kinesiology, University of Zagreb, Croatia*
[1,4]{*pipo, xenia*}*@kif.hr*

[2,3,5]*University Computing Centre - SRCE, University of Zagreb, Croatia*
[2,3,5]{*vesna.luzar-stiffler, vesna.dobric, zoran.bekic*}*@srce.hr*

[6]*Institut for Research and Development of Sustainable Ecosystems, Zagreb, Croatia*
[6]*marijan.gredelj@ires.hr*

Abstract. *Konstantin Momirović (1932-2004) had been an active contributor to both the theoretical and applied developments in multivariate data analysis from its early beginnings in 1960 until his sudden death in 2004. Some of his methodological contributions were published even after 2004. From the point of view of development of his ideas on multivariate data analysis, it is possible to recognize several phases.*

In the first, initialization phase, he worked on methodological topics as an individual assimilating and developing contemporary ideas. In the second phase, he influenced and became the originator and principal figure in the so called "Zagreb Methodological and Multivariate Data Analysis Circle" at the University of Zagreb (Institute of Kinesiology at the Faculty of Physical Education and at the University Computing Centre - SRCE) where completely new ideas emerged. In the third phase he influenced and organized various productive groups of researchers around the Institute for Criminological and Sociological Research, and the Department of Psychology at the Faculty of Philosophy of University of Belgrade. Within these three periods some sub phases are recognized, as well. In each of these phases and sub phases several areas of interest are relatively easy to recognize.

Evolution of some of the more influential ideas, algorithmic solutions and program implementations are demonstrated. Finally, his contributions and influences are divided in the same way as they are grouped and incorporated into the SPSS macro language coded library of seven hundred and sixty macro programs (Prot, F., A. Hosek, K. Bosnar, V. Luzar-Stiffler, V. Hljuz Dobric, Z. Bekić, M. Gredelj (2008)).

This short overview is the core of a more elaborate presentation scheduled for Memorial Session: "Data Mining, Statistics and Biometrics" during the ITI 2008 30th Anniversary Conference at Cavtat.

Keywords. K. Momirović, Multivariate data analysis, Robust Methods, Measurement Theory, SS Statistical System.

1. Introduction

It seems that the easiest way to introduce Konstantin Momirović (1932-2004) to the interested reader is through his personal biography, as follows:

Konstantin Momirović (1982)

Born in 1932 in Tetovo (Macedonia). Graduated from elementary and high school in Belgrade. He received his BS degree in psychology from the Faculty of Philosophy of the University of Zagreb in 1955, and his PhD in psychometrics from the University of Zagreb in 1963. From 1955 to 1959 he was employed as clinical psychologist and head of the Laboratory for Applied Psychology at the Military Hospital in Zagreb. From 1960 to 1966 he acted as head of the Laboratory for Experimental Design and Statistics at the Institute for Developmental Problems of Children and Youths in Zagreb. He was Assistant at the Faculty of Physical Education from 1960 to 1963; Assistant Professor from 1963 to 1966; Associate Professor from 1966 to 1971; Full Professor of Psychology and Quantitative methods from 1971 to 1990 at the same faculty. In the period from 1971 to 1990 he acted as Head of Research and Development Division and as General Manager (1979-1983) of the University Computing Centre - SRCE at the University of Zagreb. From 1991 to 1997 he was Full Professor of Statistics at the Faculty of philosophy of the University of Belgrade, and from 1991 on he acted as Scientific Consultant and Chief Project Manager at the Institute of Criminological and Sociological Research.

His visiting positions include: Professor of Statistics and Computer Programming in graduate and postgraduate studies at the Faculty of Mathematics and Mechanics of Moscow University, Faculty of Applied Mathematics and Cybernetics of Moscow University, Faculty of Medicine of the University of Zagreb, Faculty of Defectology of the University of Zagreb, Faculty of Philosophy of the University of Zagreb, Faculty of Economics of the University of Sarajevo, and faculties of Physical Education at Universities of Belgrade, Ljubljana and Novi Sad. He mentored MA, PhD, MD and BS degree students at Universities of Zagreb, Belgrade, Ljubljana, Sarajevo and Novi Sad.

The main fields of scientific interest of Dr. Momirović are statistics, mathematical psychology and criminology. He published 38 *books and around 500 scientific and professional papers in scientific journals or monographs of statistics, computer science, psychology, sociology, biological anthropology and kinesiology. Dr Momirović is a member of International Statistical Institute, Psychometric Society, European Anthropological Association, International Association of Statistical Computing, Statistical society of Serbia and Anthropological Association of Yugoslavia.*

One of the last project manager positions of Konstantin Momirović was on the national project 14T17 ("Factors of growth of criminal behavior and social deviations in Serbia") at the Institute of Criminological and Sociological Research in Belgrade. He also acted as a scientific consultant to this Institute and formerly professor of statistics at the Faculty of Philosophy of Belgrade University.

This rather short presentation which spans more than four decades hides the most important: How brilliant, outstanding and successful he has been in almost all areas in which he had been involved. Konstantin Momirović (1932-2004), alias Stojan Hadžigalić, among his close friends and colleagues known as "Kosta" has been an exceptionally active and productive contributor to the theoretical and applied developments of multivariate data analysis since the very beginning of his research career in early 1960's, as can be seen from his rich bibliography (Prot, F., A. Hosek, K. Bosnar, V. Luzar-Stiffler, V. Hljuz Dobric, Z. Bekić, M. Gredelj (2008)).

More details on his professional achievements can be found in autobiographic and biographic reports prepared for promotions and rewards, such as in Momirović (1976), Kališnik, Goričar and Ulaga (1979), Bujas, Petz, and Mraković (1983), Momirović (1989, 2004), or in published or unpublished In Memoriams (see e.g., Fajgelj (2004), Macura (2004), Prot (2004), and Vlahović and Kovačević (2005)).

The aim of this paper is to present and illustrate the main areas of his interest and achievements, which made him what he was and what he still represents in the field of multivariate data analysis and statistics, from the point of view of his rich bibliography (Prot et al (2008)).

It is possible to recognize several phases of development of Momirović's ideas on multivariate data analysis and statistics: In the first, initialization phase, from 1957 to 1970 he worked on methodological topics as an individual assimilating and developing contemporary ideas. His contribution to the design of Statistical System (SS) prepared the ground floor for the organized and systematic work; In the second phase, from 1971 to 1990, he became an originator and principal figure of the so called "Zagreb Methodological and Multivariate Data Analysis Circle" at University of Zagreb (Institute of Kinesiology at the Faculty of Physical Education and the University Computing Centre - SRCE) where completely new ideas emerged; In the third phase, from 1991 to 2004, he further extended and developed his research ideas in his new working environment. In each of these phases and sub-phases several areas of interests are rather easily recognized.

One of the first available real evidence of his familiarity with multivariate data analysis is his state-of-the-art (for the time being) and extensive overview of methods for factor analysis, which makes a substantial part of his doctor degree thesis titled "Factor structure of neurotic symptoms" (Momirović,1963, chapters 3 and 4, pages 71 – 195). He concluded his thesis with reaffirmations of modified multi group method (Thurstone, Holtzinger, Burt, Horst, Harman, Momirović) as a mean for structural hypotheses testing in the field of data analysis. Additionally, he fully developed and applied hierarchical algorithm for the modified multigroup method up to the third level of extracted factors (Momirović,1963, chapter 5, pages: 196 – 317).

The most obvious demonstration of his familiarity with interrelated problems and their treatment (by applying various multivariate data analysis methods) can be found in his investigation of validity of psychological tests and measurements (Momirović K., 1966; in Krković, A., K. Momirović, and B. Petz, 1966). Early on, his research interest focused on interrelated application of various methods for factor analysis (including multi group method) as means for estimating construct validity. On the basis of degenerated, simple summation method, algorithms for pseudocanonical correlations analysis, regression analysis and discriminant analyses were developed as methods for establishing pragmatic i.e. predictive and classification validities of measurement instruments. In this phase he examined equivalences between canonical discriminant and Q method of factor analysis, the problem of usefulness of factor scores in canonical discriminant analysis. He presented modified iterative Q method of factor analysis (and the accompanying algorithm) for determination of psychological types to psychologist and to a group of researchers in the emerging field of Kinesiology. Stimulated with the problem of penology treatment evaluation he proposed to evaluate structural changes using the canonical correlation (alienation) model and hierarchical factor solutions where he investigated relations between manifest variables and multidimensional higher-order factor spaces (Momirović, K.; 1969, 1969, 1969). These were research problems which prepared him to open a new stage of systematic research in the field of multivariate methods for transformation and condensation in data analysis (Momirović, K. (1972), "Methods for transformation and condensation of kinesiological data"). On page 303 it was stated that all of the methods and algorithms are going to be coded in FORTRAN IV for IBM series 1130 and 370/165 electronic computers. It had been realized that a collections of statistical and data analysis subroutines (once developed) should be interrelated into a general statistical language.

The research program incorporated in "Methods for transformation and condensation of kinesiological data" resulted in very interesting contributions which considerably influenced algorithm and application development. Full assimilation of Guttman and Harris ideas results with new treatment of problem of initial metrics of variables. The problem of latent structures of manifest variables have been analyzed considering the specific properties of initial metric of variables, and a series of new criteria for determining the number of retained factors were proposed (e.g., PB ("Plum Brandy") criterion (Momirović, K. and J. Štalec, 1971)). The concept of images of variables in Guttman sense has been extended to generalized image transformations of one set of variables to the another set of variables and vice verse (Momirović, K., J. Štalec, E. Zakrajšek,1973). This theoretical contribution anticipated developments in data analysis which will extend and relax already established dominance of classical canonical correlation model as a general and fundamental model for variety of methods for data analysis. Guttman's Image theory and its extensions were also applied to the problems of homogeneity, representativeness and reliability of psychometric measurements. Above mentioned and other related developments were incorporated in real research environment in various areas of applications. Influential (classical) books on multivariate analysis (Anderson, 1958; Harman; 1960, 1967; Horst, 1965; Rao, 1958, 1973; Cattell, 1966, Morrison, 1967; Cooley and Lohnes 1971; Mulaik, 1972; Bock, 1975) have further stimulated deeper insight and research in the area.

On the occasion of this memorial overview only a small number of K. Momirovic's achievements and contributions will be covered in detail, especially some of those which were accomplished while he had been an active researcher at the University Computing Centre - SRCE of the University of Zagreb and which were further developed later on during his extraordinary successful career.

2. Some Influential Methodological Contributions

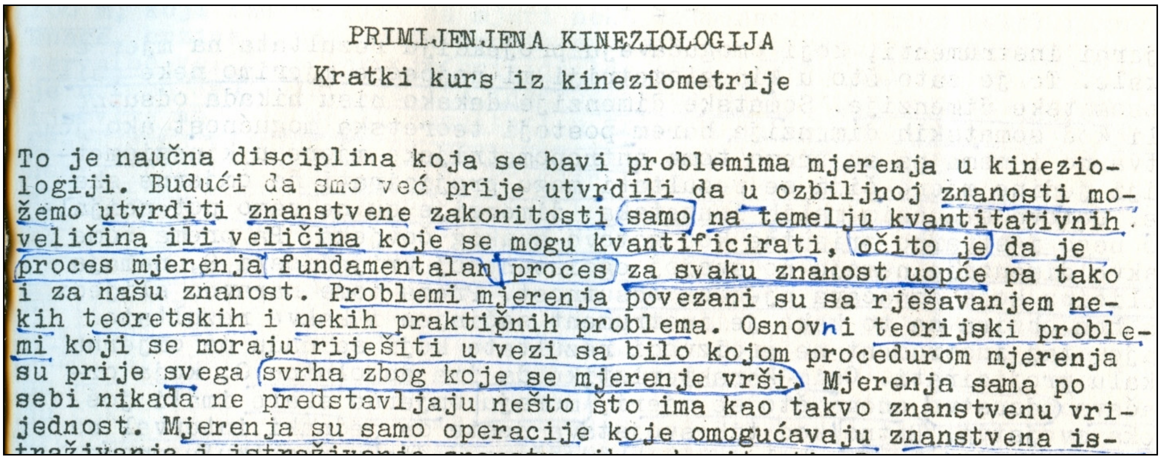
Methodological research and achievements after 1971 could be divided into the following three areas of data analysis:

1. Kinesiometrics (development of new theoretical and applied measurement models in the field of measurement theory)
2. Multivariate data analysis and statistics (new models, methods and algorithms for data analysis)
3. Informatics (a field of computer science related to the development of new software for information systems, data analysis and management)

Proposed names and labels of these sub-fields were introduced and became part of the standard terminology used in the curricula at graduate and postgraduate studies.

2.1. Kinesiometrics (development of new theoretical and applied measurement models in the field of measurement theory)

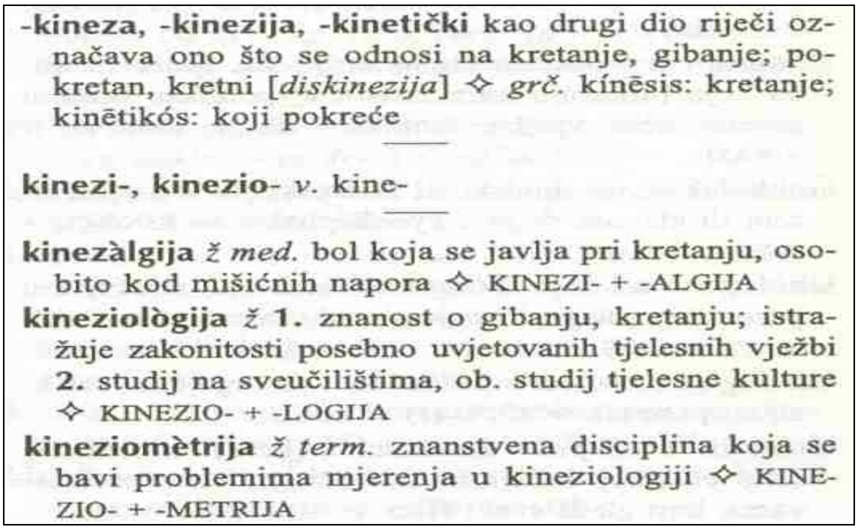
It was K. Momirović, who introduced the term "kinesiometry" in his lecture notes for the course entitled "Short course in kinesiometry"("Kratki kurs iz kineziometrije"; 88 pp) conducted at postgraduate study of Kinesiology during 1971/72, as illustrated in Figure 1. The term had been included in the Anić and Goldstein dictionary (Anić and Goldstein, 1999) at page 675, as illustrated in Figure 2.



PRIMIJENJENA KINEZIOLOGIJA
Kratki kurs iz kineziometrije

To je naučna disciplina koja se bavi problemima mjerenja u kineziologiji. Budući da smo već prije utvrdili da u ozbiljnoj znanosti možemo utvrditi znanstvene zakonitosti samo na temelju kvantitativnih veličina ili veličina koje se mogu kvantificirati, očito je da je proces mjerenja fundamentalan proces za svaku znanost uopće pa tako i za našu znanost. Problemi mjerenja povezani su sa rješavanjem nekih teoretskih i nekih praktičnih problema. Osnovni teorijski problemi koji se moraju riješiti u vezi sa bilo kojom procedurom mjerenja su prije svega svrha zbog koje se mjerenje vrši. Mjerenja sama po sebi nikada ne predstavljaju nešto što ima kao takvo znanstvenu vrijednost. Mjerenja su samo operacije koje omogućavaju znanstvena is-

Figure 1. A part of the first page of student notes for the course on Kinesiometrics ("kineziometrija") at postgraduate study of Kinesiology, years 1971/72
at the College of Physical Education (today Faculty of Kinesiology).



-kineza, -kinezija, -kinetički kao drugi dio riječi označava ono što se odnosi na kretanje, gibanje; pokretan, kretni [*diskinezija*] ✧ *grč.* kínēsis: kretanje; kinētikós: koji pokreće

kinezi-, kinezio- *v.* kine-

kinezàlgija *ž med.* bol koja se javlja pri kretanju, osobito kod mišićnih napora ✧ KINEZI- + -ALGIJA

kineziològija *ž* **1.** znanost o gibanju, kretanju; istražuje zakonitosti posebno uvjetovanih tjelesnih vježbi **2.** studij na sveučilištima, ob. studij tjelesne kulture ✧ KINEZIO- + -LOGIJA

kineziomètrija *ž term.* znanstvena disciplina koja se bavi problemima mjerenja u kineziologiji ✧ KINEZIO- + -METRIJA

Figure 2. The term kinesiometrics "kineziometrija" was introduced to
Anić-Goldstein's Dictionary at page 675 (Anić and Goldstein, 1999)

In measurement theory alternative approaches to classical test theory are constantly being examined. These alternative approaches were applied (by K. Momirović and his associates) in construction and reconstruction of composite measurement instruments (composite tests and questionnaires) (see e.g., Momirović (1966, 1969, 1972, 1974); Zakrajšek, Momirović and Dobrić (1976, 1977); Momirović and Gredelj (1980); Bosnar (1980); Momirović, Gredelj and Dobrić (1981); Momirović, Pavičić and Hošek (1984); Momirović (1988)). A new general model for the estimation of error of measurement, along with the measures of reliability and representativeness were proposed (Momirović, 1974; Momirović and Dobrić, 1976; Zakrajšek, Momirović and Dobrić, 1977; Momirović, Dobrić and Gredelj, 1978; Momirović and Gredelj, 1980; Momirović, Pavičić and Hošek, 1984). Additionally, upper and lower bound of reliability (under the new general model) were derived (Momirović, 1974, 1975; Momirović, Pavičić and Hošek, 1984). That enabled objective definition and estimation of homogeneity independently from the reliability itself. (Momirović, 1974, 1977).

We'll try to illustrate some of these achievements, most of which were developed while K. Momirović has been affiliated to SRCE (Zagreb). Methods for determination of internal metric properties of measurement instruments had been continually in focus of his attention. Based on full assimilation of ideas of Guttman

(1954) and Harris (1963) he generalized the classical test theory, the generalization being that the errors of measurement are permitted to exhibit correlation and nonconstant variability.

Momirović was constantly improving algorithms and implementations of programs for the analysis of metric characteristics of composite measurement instruments ("RTT" programs) from the initial SS program (MAPANAL) through the upgraded versions RTT7 (Statistical System, Momirović, 1980), RTT8 (GENSTAT version, Momirović and Prot 1986), RTT9/10 (SPSS macro version; Momirović and Knežević, 1996), to the most complex and general RTT12G SPSS implementation.

To briefly illustrate some of Momirović's original contributions to the analysis of metric properties of composite measurement instruments, let us define a set E as a set of n entities randomly selected from a homogeneous population P, and let T be a composite measurement instrument composed from m items measured on a continuous or on an ordinal scale. Let **Z** be a matrix of standardized (normalized, if necessary) results which describe set E over the set T. Then $\mathbf{R} = \mathbf{Z}^t\mathbf{Z}n^{-1}$ and $\mathbf{U}^2 = (\text{diag } \mathbf{R}^{-1})^{-1}$ denote the correlation matrix and a diagonal matrix of estimated error variances, respectively. Error variables could be defined as antiimage variables (i.e., $\mathbf{E} = \mathbf{Z}\mathbf{R}^{-1}\mathbf{U}^2$), and true parts as image variables (i.e., $\mathbf{T} = \mathbf{Z}(\mathbf{I} - \mathbf{R}^{-1}\mathbf{U}^2)$. The respective covariance matrices are:

$\mathbf{A} = \mathbf{E}^t\mathbf{E}n^{-1} = \mathbf{U}^2\mathbf{R}^{-1}\mathbf{U}^2$, and

$\mathbf{G} = \mathbf{T}^t\mathbf{T}n^{-1} = \mathbf{R} + \mathbf{U}^2\mathbf{R}^{-1}\mathbf{U}^2 - 2\mathbf{U}^2$.

Then the covariances between observed (**Z**) and true variables (**T**) are $\mathbf{P} = \mathbf{Z}^t\mathbf{T}n^{-1} = \mathbf{R} - \mathbf{U}^2$, and covariances between observed (**Z**) and error variables (**E**) are $\mathbf{Z}^t\mathbf{E}n^{-1} = \mathbf{U}^2$.

Let $\mathbf{M} = \mathbf{Z}\mathbf{U}^{-1}$, represent the observed variables transformed to the "universal" metric, where $\mathbf{H} = \mathbf{M}^t\mathbf{M}n^{-1} = \mathbf{U}^{-1}\mathbf{R}\mathbf{U}^{-1}$ (see e.g. Harris (1962)).

Let **x**, **y**, **v** and **w** denote normalized eigenvectors and let λ^2, δ^2, ω^2 and η^2 denote eigenvalues of the matrices **R**, **G**, **P** and **H**, respectively.

Some examples of Momirović's original contributions regarding representativeness, reliability, homogeneity and convergence of metric indicators are presented in the following paragraphs.

REPRESENTATIVENESS: measures how the measurements are representative regardless of whether the synthetic composite measure is computed or not. Most of these measures were originally proposed by Kaiser (1970). Konstantin Momirovic proposed the following extensions: A measure of absolute lower bound of representativeness ψ_4 (Momirović, Dobrić and Gredelj, 1978), which is identical to absolute lower bound of reliability under Guttman's measurement model $\psi_4 = 1 - \eta^{-4}$, and ψ_5 (a measure proposed by Momirović and Hošek) defined as

$$\psi_5 = ((\mathbf{s}^t(\mathbf{G} * \mathbf{G})\mathbf{s})(\mathbf{s}^t(\mathbf{R} * \mathbf{R})\mathbf{s})^{-1})^{1/2},$$

where **s** is a summation vector of order m.

RELIABILITY: Under the classical test theory model (CTT), the reliability of a test can be defined as $\alpha = \psi^2\sigma^{-2}$, where ψ^2 and σ^2 are variances of true part, and of total score, respectively. Alternatively $\beta = 1 - \varepsilon^2\sigma^{-2}$, where ε^2 is the error variance. Within CTT α and β are equivalent, but under the alternative model (by Momirović referred to as "Guttman's model") α and β are not equivalent. Momirović and his fellow researchers developed a series of reliability measures. Some of them, included in SPSS macro program "RTT12G" are as follows: absolute lower bound for component of standardized items $\mu_1 = 1 - \lambda^{-2}$; Lower bound of β as $\beta_1 = (1 - \lambda^{-2})^2$ and upper bound as $\beta_2 = 1 - \lambda^{-4}$ proposed by Momirović, Dobrić and Gredelj (1977); and $\beta_6 = 1 - (\mathbf{x}^t\mathbf{U}^2\mathbf{x})\ \lambda^{-2}$ proposed by Momirović (1996); Measures of reliability of component of items transformed into image form $\tau = \delta^2\lambda^{-2}$ Momirović (1975) and $\gamma = \omega^2\lambda^{-2}$ Momirović and Knežević (1991). In addition to Guttman-Nicewander $\lambda_6 = 1 - \eta^{-2}$, new measures of reliability (of the first component of standardized items rescaled to Harris universal metric) were proposed, i.e. lower bound $\rho_1 = (1 - \eta^{-2})^2$ as a completely new approach by Momirović and Dobrić (1977), and upper bound $\rho_2 = 1 - \eta^{-4}$ by Zakrajšek, Momirović and Dobrić (1977). A new extension of image transformation, mirror image analysis, and its application to reliability theory was examined by Momirović, Gredelj and Dobrić (1981).

HOMOGENEITY: In addition to the standard measure (i.e., the average of correlations among items), a series of new alternative measures were devised. Measure of relative variance of first principal component of items transformed into image form $h_2 = \delta^2\xi^{-2}$ Momirović (1977); measure defined by the number of components with non negative reliability was proposed by Momirović and Gredelj (1980) as $h_4 = 1 - (k - 1)(m - 1)^{-1}$ where $k = \text{num}(\lambda^2_p > 1)$ and λ^2_p, $p = 1,...,m$ are eigenvalues of matrix **R;** measure $h_5 = 1 - (\theta^2 - \lambda^2)(m - \lambda^2)^{-1}$ (Knežević and Momirović (1995)), where θ^2 is the sum of k eigenvalues of matrix **R,** and another one defined as $h_6 = 1 - (\theta^2 - \lambda^2)(m - 1)^{-1}$; in addition, $h_7 = (\lambda^2 - 1)(\theta^2 - 1)^{-1}$ was proposed by Momirović and Knežević (1995).

CONVERGENCES OF INDICATORS: Momirović proposed one class of convergence measures of indicators which means that proportional conformity of item results to the first principal component. Let

$$\mathbf{G} = \mathbf{R} + \mathbf{U}^2\mathbf{R}^{-1}\mathbf{U}^2 - 2\mathbf{U}^2 \quad \text{and}$$
$$\mathbf{V} = (\text{diag}\mathbf{G}^{-1})^{-1/2}$$

and let δ_1, δ_2 and δ_m be the first, second and mth eigenvalue of the matrix **G**. The following measures of convergences were proposed:

$$\varphi_1 = 1 - \delta_1^{-1};\ \varphi_2 = 1 - \delta_2\,\delta_1^{-1};\ \varphi_3 = \delta_1 m^{-1};$$
$$\varphi_4 = 1 - 2^{-1}(\delta_2 / \delta_m)\delta_1^{-1},\ \varphi_5 = 1 - (\text{trace}\ \mathbf{V})m^{-1}$$

and a measure of convergence defined as relative informativity $\varphi = (1 - \mu_1)^{-1}m^{-1}$ where μ_1 is the absolute lower bound of the first principal component of items.

The new stage of development was achieved using neural network methodology. As an example RTT13HNN (Momirović, 2002) is an emulation of Hopfield neural network (Hopfield, 1982) for the estimation of sampling adequacy, reliability and homogeneity of composite measurement instruments. The basic procedure is the estimation of Guttman anti image and image variables (Guttman, 1953) by a simple modification of Hopfield network, and then the calculation of suitable modifications of Kaiser - Rice measure of sampling adequacy of items, denoted here as ψ_4 (Kaiser and Rice, 1974), Guttman sixth lower bound to reliability of result obtained by simple summation of item scores, usually referred to as λ_6 (Guttman, 1945), Momirović lower bound to reliability of result defined by the first principal component, usually referred to as β_6 (Momirović, 1996) and Momirović second measure of homogeneity h_2 (Momirović, 1977). Construction of Hopfield network for the estimation of anti image and image variables and technique of computation of measures ψ_4, λ_6, β_6 and h_2 can be easily understood from the symbolic code of the program itself (Momirović, 2002).

Taking into account the diversity of initial metrics of variables, i.e. nonstandardized and standardized real, and transformed into Guttman's and Harris' space, different criteria for extractions of latent dimensions, different parsimonious transformations, new explorative and confirmative models were recommended for estimation of validities of measurement instruments (Momirović, 1966, 1970, 1973; Bosnar and Prot, 1981; Prot and Momirović, 1984; Momirović, Erjavec and Radaković, 1988; Bosnar, 1989, Prot and Bosnar, 1989).

2.2. Multivariate Data Analysis and Statistics (new models, methods and algorithms for data analysis)

The field of data analysis, and specially the area of multivariate statistical methods, had been a subject of majority of Momirović's methodological research. At a time, new methods and algorithms were proposed for component analysis (Momirović and Gredelj, 1980; Prot, Viskić-Štalec, Štalec, Bosnar, Momirović and Knap, 1984), taxonomic analysis (Momirović and Zakrajšek, 1973; Momirović, Szirovicza, Gredelj and Dobrić, 1980; Momirović, Hošek, Bosnar and Prot, 1984; Prot, Zenkin, Momirović, Bosnar and Knap, 1984; Prot, Viskić-Štalec, Štalec, Bosnar, Momirović and Knap, 1984; Momirović and Mildner, 1989), multidimensional scaling (Momirović, Bosnar, Štalec and Prot, 1983), canonical correlation (Momirović, Gredelj and Herak, 1980), multivariate regression (Momirović, Szirovicza, Dobrić and Gredelj, 1980) and discriminative analysis (Momirović, Dobrić and Szirovicza, 1979; Momirović, Szirovicza, Dobrić and Gredelj, 1980) analysis of nonnumerical (qualitative) data, which enabled applications of complex procedures in data analysis and hypotheses testing of data with weak metric properties. An algorithm of classical type had been implemented (Zlobec, Varga and Momirović, 1974). More advanced, multivariate approach, based on spectral decomposition of contingency tables was utilized (Bosnar and Pavičić, 1982; Bosnar and Hošek, 1983; Momirović, 1989). New complex integrative algorithms which support interpretation of realistic research data of so called nonquantitative data.

One of the classical problems, the problem of criteria for determination of number of important latent dimensions had been critically examined (Momirović, Kovačević, Ignjatović, Horga, Radovanović, Mejovšek, Štalec and Viskić-Štalec, 1972). Series of new criteria had been proposed and comparatively tested in real standardized metric (Štalec and Momirović, 1971), Guttman's nonstandardized and standardized space (Zakrajšek and Momirović, 1972; Momirović and Štalec, 1973; Momirović, Štalec and Zakrajšek, 1973) and Harris metric as well. Based on these experiences, a new general principle for determination of criterion in all metrics of analyzed variables had been developed (Momirović and Štalec, 1984). Principles of criteria construction for determination of number of important dimensions were generalized to the problem of analyses of relations of sets of variables (Dobrić, Momirović and Gredelj, 1987).

An algorithm for hierarchical component analysis in image space was developed (Dobrić and Momirović, 1984). Concurrent and comparative evaluation studies of component and factor analysis methods were provided (Momirović, Viskić-Štalec, Štalec, Mejovšek, Ignjatović, Radovanović, Horga and Kovačević, 1972; Viskić and Štalec, 1982; Viskić-Štalec 1987; Bosnar and Prot, 1994).

Canonical factor analysis model and component model in Harris' space have been studied and completely new algorithm for pseudocanonical component model has been developed (Bosnar, Prot, Momirović, Lužar and Dobrić, 1982). Estimation of factor score values had been studied as well (Radaković and Momirović, 1987). An algorithm for iterative multigroup method of factor analysis had been improved (Gredelj, Štalec and Momirović, 1983). On the basis of these experiences new different procedures which integrate explorative and confirmative approach in determination of latent spaces of sets of manifest variables were developed (Momirović, 1972; Momirović, Gredelj and Štalec, 1977; Szirovicza, Gredelj and Momirović, 1978; Štalec and Momirović, 1982; Bosnar, Prot, Štalec and Momirović, 1984; Cvitaš and Momirović, 1984; Viskić-Štalec, Štalec and Momirović, 1984; Knezović and Momirović, 1986; Momirović, Erjavec and Radaković, 1988).

Methods for identification of structural similarities or differences of matrices of covariances or crosscovariances were developed for the analyses of structural changes (Dobrić, Karaman and Momirović, 1983; Bosnar and Prot, 1984; Momirović, Prot, Dugić, Bosnar, Erjavec, Gredelj, Kern and Dobrić, 1987; Cvitaš and Momirović, 1985, 1987; Prot, Bosnar, Hošek and Momirović, 1984; Prot, Ivančević and Momirović, 1985; Prot and Bosnar, 1987).

Multivariate analysis of time series had been treated under the spectral decomposition of data matrices of individuals-INDIFF (Momirović and Karaman, 1982; Karaman and Momirović, 1984; Prot, Momirović and Bosnar, 1987), and groups-COLDIFF (Momirović and Karaman, 1982; Pavičić, Karaman and Momirović, 1983; Bosnar, Momirović and Prot, 1987).

Significant results were achieved it the field of taxonomic analysis. The model of polar taxons has been treated from the point of view of different metrics of variables (Momirović, 1978; Momirović, Zakrajšek, Hošek and Stojanović, 1979; Momirović, 1981).

Integration of explorative and confirmative approach in determination of taxonomic dimensions enabled new methods for pattern recognition and multicriteria selection were developed, with or without linear constraints imposed (Prot and Bosnar, 1982; Momirović, Dobrić and Karaman, 1984; Prot, 1985, 1989; Momirović, 1989; Momirović and Mildner, 1989).

Canonical discriminative model had been studied from the formal point of view. Regarding the metrics of initial set of variables, a generalized discriminant procedure (Momirović, Gredelj and Herak, 1981), and a general program for multivariate analysis of variance were developed (Pavičić and Momirović, 1982). It had been found that Mahalanobis transformation of variables offers apropriate numeric and interpretative form for canonical discriminative analysis (Bosnar, Momirović and Prot, 1984).

Systematic application of generalized image transformation of sets of variables (Momirović, Štalec and Zakrajšek, 1973) resulted in definition of some of general measures of associations (Momirović and Dugić, 1986; Gredelj and Momirović, 1988). Symmetric and asymmetric measures of association of sets of variables were defined. Under the concept of generalized image transformation with suitable restrictions, it was possible to develop all of the methods for symmetric and asymmetric relations among sets of variables (Prot, Bosnar and Momirović, 1983, 1983; Bosnar, Prot and Momirović, 1985).

BICANAL method of canonical correlation analysis of images of variables, inspired by Tim and Carlson's (1976) bipartial canonical correlation analysis, has been successfully developed, implemented and applied (Gredelj, Momirović, Dobrić, Herak, Bosnar and Prot, 1982; Hošek, Bosnar, Prot and Momirović, 1984). Bi-partialisation of external sets of variables led to a new algorithms, i.e. bipartial canonical covariance analysis (Momirović, 2001).

Formal properties of relations among sets of variables were examined (Momirović, Štalec, Zakrajšek, 1973; Momirović, Dobrić, 1979; Bosnar, Prot, Momirović, 1985; Bosnar, Prot, 1988).

A class of robust methods for multivariate data analysis were proposed to enable processing of data that do not satisfy assumptions necessary for the application of classical data analysis methods. The advantages of the proposed methods are in their robustness both to outliers and to the artificial associations generated by the few pairs of variables from two different sets. Furthermore, this variables need neither be normally distributed nor their covariance matrices need be regular.

Beside the methods for robust analysis of hypothetical latent dimensions (Štalec and Momirović, 1982; Dobrić. Karaman, Momirović, 1983), and robust analysis of hypothetical taxonomic dimensions (Štalec, Bosnar, Prot, Momirović, 1982), the problem of relations between sets of variables had been successfully treated as well.

Let us present some of these methods, as they were presented by Dobrić (1986).

Let $\mathbf{Z}_1$ (n, m_1) and $\mathbf{Z}_2$ be two centered data matrices, obtained as a description of set E of n entities over two sets V_1 and V_2 of quantitative, elliptically distributed variables. Maximization of covariances between linear composites of variables belonging to the sets V_1 and V_2 is defined, under some constraint, as:

$$\mathbf{Z}_1 \mathbf{x}_p = \mathbf{l}_p \quad | \quad \rho_p = \mathbf{l}_p^t \mathbf{k}_p n^{-1} \rightarrow \max$$
$$\mathbf{Z}_2 \mathbf{y}_p = \mathbf{k}_p \quad | \quad \rho_p > \rho_{p+1}, \quad p=1, \ldots, r-1; \quad r=\min(m_1,m_2)$$
$$\mathbf{x}_p^t \mathbf{x}_q = \mathbf{y}_p^t \mathbf{y}_q = \delta_{pq}$$

and can be reduced to the singular value decomposition of matrix $\mathbf{C}_{12} = \mathbf{Z}_1^t \mathbf{Z}_2 n^{-1}$.

CANONICAL COVARIANCE ANALYSIS (QCR)

In a general case when $m_1 > 1$ and $m_2 > 1$, and symmetrical relations between V_1 and V_2 can be assumed, QCR model is obtained. QCR algorithm seeks such linear composites of $\mathbf{l}_p$ and $\mathbf{k}_p$ of variables from V_1 and V_2 to maximize successively their covariances ρ_p under the constraint of orthonormality of transformation vectors $\{\mathbf{x}_p\}$ and $\{\mathbf{y}_p\}$ respectively. Solution of the problem could be obtained from the characteristic equation

$$\mathbf{C}_{12}\mathbf{C}_{12}^t \mathbf{x}_p = \rho_p^2 \mathbf{x}_p \qquad p=1, \ldots, r$$

with covariances ρ_p equal to $\rho_p = \mathbf{x}_p^t \mathbf{C}_{12} \mathbf{y}_p$, Momirović, Dobrić and Karaman (1983, 1984).

As was found (Lužar, 1985), the test statistic: n $(\mathbf{M}\boldsymbol{\rho})^t (\mathbf{M}\boldsymbol{\psi}\mathbf{M}^t)^{-1} (\mathbf{M}\boldsymbol{\rho})$ is asymptotically χ^2 distributed with r-1 degrees of freedom; here, $\boldsymbol{\rho}$ is a vector of covariances ρ_p , $\mathbf{M}$ is a hypothetical matrix and $\boldsymbol{\psi}$ is a matrix of elements estimated from the elements of crosscovariance matrix of quasicanonical variates $\{\mathbf{l}_p \mathbf{k}_p\}$, p=1, … , r .

Obtained quasicanonical dimensions $\{\mathbf{l}_p\} = \mathbf{L}$ and $\{\mathbf{k}_p\}$ = K are not orthogonal; if $\mathbf{C}_{ii} = \mathbf{Z}_i^t \mathbf{Z}_i n^{-1}$ are covariance matrices of variables from V_i , i=1,2 respectively, then

$$\mathbf{W}_1 = \mathbf{L}^t \mathbf{L}\ 1/n = \mathbf{X}^t \mathbf{C}_1 \mathbf{X} \neq \text{diag}$$

and

$$\mathbf{W}_2 = \mathbf{K}^t \mathbf{K}\ 1/n = \mathbf{Y}^t \mathbf{C}_2 \mathbf{Y} \neq \text{diag} .$$

Denoting by $\mathbf{D}_i^2$ = diag $\mathbf{W}_i$, i =1,2 variances of linear composites $\{\mathbf{l}_p\}$ and $\{\mathbf{k}_p\}$ respectively, the structure matrices $\mathbf{F}_i$, i =1,2 of the first and second set of quasicanonical factors could be obtained as

$$\mathbf{F}_1 = \mathbf{C}_1\mathbf{X}\mathbf{D}_1^{-1}$$

and

$$\mathbf{F}_2 = \mathbf{C}_2\mathbf{X}\mathbf{D}_2^{-1}$$

together with their pattern matrices

$$\mathbf{A}_i = \mathbf{F}_i(\mathbf{D}_i^{-1} \mathbf{W}_i \mathbf{D}_i^{-1})^{-1}$$

Quasicanonical correlations Γ, for diagonal matrix $\boldsymbol{\rho}$, are obtained as

$$\Gamma = \mathbf{D}_1^{-1}\, \rho\, \mathbf{D}_2^{-1}$$

and if η_1 is a canonical correlation, then $\gamma_1 < \eta_1$. This and other relations between the canonical and quasicanonical correlation analysis can be found in Momirović, Bosnar and Prot (1984), Momirović and Dobrić, (1985), and Knežević and Momirović (1996). The influences of initial metric of variables on the results of applied algorithms of quasicanonical analysis of covariances were examined: Harris universal (Dobrić, Momirović and Gredelj, 1985), Guttman's partial image (Momirović, Dobrić and Gredelj, 1985), standardized image (Wolf, Radaković and Momirović, 1988), and between Guttman and Harris metric (Momirović, Dobrić and Gredelj, 1985).

ROBUST REDUNDANCY MODEL

When $m_1 > 1$ and $m_2 > 1$, and when sets V_1 and V_2 have a logical status of predictors and criteria, redundancy analysis is more appropriate then canonical analysis. Robust redundancy model is very simple and is defined as maximization of redundancy

$$\rho^{*2}_{\mathbf{ap}} = \mathbf{a}_p^t\, \mathbf{a}_p\, n^{-1} \rightarrow \max \qquad \mathbf{a}_p = \mathbf{Z}_1^t\, \mathbf{k}_p^*\, n^{-1}$$

and

$$\rho^{*2}_{\mathbf{bp}} = \mathbf{b}_p^t\, \mathbf{b}_p\, n^{-1} \rightarrow \max \qquad \mathbf{b}_p = \mathbf{Z}_2^t\, \mathbf{l}_p^*\, n^{-1}$$

of composites $\mathbf{l}_p^* = \mathbf{Z}_1^t\, \mathbf{x}_p^* n^{-1}$ and $\mathbf{k}_p^* = \mathbf{Z}_2^t\, \mathbf{y}_p^* n^{-1}$ respectively; here $\mathbf{a}_p$ and $\mathbf{b}_p$ are crosscovariance vectors of one set of variables and redundancy variates of another set of variables and redundancy variates of another set of variables, under the constraints of orthonormality of transformation vectors $\mathbf{x}_p^*$ and $\mathbf{y}_p^*$, respectively. It can be easily shown that $\rho^{*2}_{\mathbf{ap}} = \rho^{*2}_{\mathbf{bp}} = \rho^2_p$ and that $\mathbf{x}_p^* = \mathbf{x}_p$ and $\mathbf{y}_p^* = \mathbf{y}_p$, p=1, …, r, so that the maximization of covariances of linear composites is equivalent to the maximization of redundancy of each composite (Prot, Bosnar, and Momirović, 1983).

Under the generalized image transformation with suitable restrictions it was possible to develop different methods for symmetric and asymmetric methods of relations of sets of variables (Prot, Bosnar and Momirović, 1983, 1983; Bosnar, Prot and Momirović, 1985).

SRA MODEL OF REGRESSION ANALYSIS

When $m_1 > 1$ and $m_2 = 1$, SRA model of robust regression analysis is obtained.

Let $\mathbf{Z}_1$ be predictor and $\mathbf{Z}_2 = \mathbf{k}$ the criterion variable. If $\mathbf{Z}_1\, \mathbf{x}^* = \mathbf{k}^*$ is a linear composite of predictors, SRA model is defined as the maximization of covariance ρ^* between $\mathbf{k}$ and $\mathbf{k}^*$

$$\mathbf{k}^t\, \mathbf{k}^*\, n^{-1} = \rho^* = \max,$$

under the constraint:

$$\mathbf{x}^{*t}\, \mathbf{x}^* = \mathbf{1}\ .$$

If we denote $\mathbf{Z}_1^t\, \mathbf{k} = \mathbf{C}_{1k}$, solution of the problem could be obtained from the characteristic equation

$$\mathbf{C}_{1k}\, \mathbf{C}_{1k}^t\, \mathbf{x}^* = \rho^{*2}\, \mathbf{x}^*$$

which makes it obvious that SRA model is a special case of QCR. Analogously as before, SRA multiple correlation $g = \rho^*\, s^{-1}$ is less or equal to the least squares regression (LSR) multiple correlation h , s^2 being the variance of $\mathbf{k}^*$. Furthermore,

- covariance between LSR and SRA estimates of $\mathbf{k}$ is equal to ρ^*
- correlation between LSR and SRA estimates of $\mathbf{k}$ is equal to g / h
- scalar product of LSR and SRA weights vectors is equal to h^2 / ρ^*
- covariances between residuals obtained under LSR and SRA models is equal to $1 - h^2$
- correlation between residuals obtained under LSR and SRA models is equal to the ratio between standard deviation of residuals under LSR and SRA models

All that relations between SRA and LSR regression could be found in Dobrić, Štalec and Momirović (1984). The influences of different metrics and scaling of variables on the results of quasicanonical regression analysis had been also treated (Momirović, Dugić and Gredelj, 1987; Matečić and Momirović, 1988; Momirović, 1988),

SDA MODEL OF ROBUST DISCRIMINANT ANALYSIS

SDA model is defined by the successive maximization of variances

$$\rho_p^{\#2} = \mathbf{k}_p^{\#t}\, \mathbf{k}_p^{\#}\, n^{-1} = \max \qquad p = 1, \dots, r$$

of linear composites

$$\mathbf{k}_p^{\#} = \mathbf{S}\, (\mathbf{S}^t\mathbf{S})^{-1}\, \mathbf{S}^t\, \mathbf{Z}_1\, \mathbf{y}_p^{\#}$$

under the constraints of orthonormality of transformation vectors $\mathbf{y}_p^{\#t}\mathbf{y}_q^{\#} = \delta_{pq}$. Here $\mathbf{S}\, (\mathbf{S}^t\mathbf{S})^{-1}\, \mathbf{S}^t\, \mathbf{Z}_1$ is a matrix of standardized group means of variables from V_1 centered to common zero mean, $\mathbf{S}$ is a selector matrix denoting the membership of each entity to one and only one of the groups and r = min (m_1 , g-1).

The obtained SDA variables $\mathbf{Z}_1\ \mathbf{y}_p^{\#}$ are not orthogonal; if { $\mathbf{y}_p^{\#}$ } = $\mathbf{Y}^{\#}$, covariance matrix $\mathbf{W} = \mathbf{Y}^{\#t}\mathbf{C}_1\mathbf{Y}^{\#} \neq$ diag.

Denoting the variances of SDA variates with $\mathbf{D}^2$ = diag $\mathbf{W}$, the structure $\mathbf{F}$ of SDA factors is equal to $\mathbf{F}=\mathbf{C}_1\mathbf{Y}^{\#}\mathbf{D}^{-1}$ and their pattern matrix $\mathbf{A}$ is equal to $\mathbf{A}=\mathbf{C}_1\mathbf{Y}^{\#}\ (\mathbf{Y}^{\#t}\mathbf{C}_1\mathbf{Y}^{\#})^{-1}\mathbf{D}$.

If we define $\mathbf{Z}_2 = \mathbf{S}\ (\mathbf{S}^t\mathbf{S})^{-1}\mathbf{S}^t\mathbf{Z}_1$, then SDA also maximizes covariances $\mathbf{l}^{*t}\mathbf{k}^*\ n^{-1}$ between linear composites of variables $\mathbf{Z}_1\mathbf{x}_p^* = \mathbf{l}_p^*$ and linear composites $\mathbf{Z}_2\ \mathbf{y}_p^* = \mathbf{k}_p^*$ of these variables projected into the space spanned by vectors from $\mathbf{S}$. These covariances are maximized using the same eigenvectors $\mathbf{x}_p^* = \mathbf{y}_p^* = \mathbf{y}_p^{\#}$ (as in QCR and SRA), and are equal to $\rho_p^{\#2}$. So, SDA is a special case of QCR model too. This fact was utilized to define the test statistics for the parameters of $\rho_p^{\#2}$ (Lužar, 1986). Obtained expression, analogous to the one developed for the QCR model, is again asymptotically χ^2 distributed, with specially defined matrices $\mathbf{M}$ and ψ .

Furthermore, SDA also maximizes redundancy $\mathbf{r}_p^t\mathbf{r}_p$ of SDA factors defined over the same $\mathbf{y}_p^{\#}$ as $_p = \mathbf{Z}_1^t\ \mathbf{S}(\mathbf{S}^t\mathbf{S})^{-1}\mathbf{S}^t\mathbf{Z}_1\ y_p^{\#}$ with their redundancy equal to $\rho_p^{\#4}$, p = 1, … ,r (Štalec and Momirović, 1983; Momirović and Dobrić, 1985).

Furthermore, the properties of that model had been defined in Harris universal space (Milonja, Dobrić and Momirović, 1989), Hotelling space (Gredelj, Dobrić and Momirović, 1989), and Pearson space (Momirović, 1989).

The interrelations of results of canonical covariance analysis, multivariate redundancy analysis and regression analysis suggest a method of spectral analysis of relations of general images of variables named “nuclear analysis” (Hošek and Momirović 2001).

NUCLEAR ANALYSIS

Nucleus of two sets of quantitative variables is defined as the cross covariance matrix of the variables mutually transformed to generalized image form (Hošek and Momirović, 2001). Let $\mathbf{Z}_1$ be a data matrix, in the standard normal form, obtained by the description of a random sample E of n objects on a sample V_1 of m_1 linearly independent quantitative or quantified variables with not necessary logical status of regressors, and let $\mathbf{Z}_2$ be another data matrix, also in the standard normal form, obtained by the description of E on a sample V_2 of m_2 linearly independent quantitative or quantified variables such that $V_1 \cap V_2 = 0$ with not necessary logical status of dependent variables. Suppose, without loss of generalizability, that $m_2 \leq m_1$, and denote by $\mathbf{R}_{11} = \mathbf{Z}_1^t\mathbf{Z}_1$ and $\mathbf{R}_{22} = \mathbf{Z}_2^t\mathbf{Z}_2$ the maximum likelihood estimations of intercorrelation matrices of variables in the sets V_1 and V_2, respectively, and by $\mathbf{R}_{12} = \mathbf{R}_{21}^t = \mathbf{Z}_1^t\mathbf{Z}_2$ the crosscorrelation matrix between variables from V_1 and V_2.

Let $\mathbf{B}_{12} = \mathbf{R}_{11}^{-1}\mathbf{R}_{12}$ be a matrix of the standardized partial regression coefficients obtained by the solution of regression problem $\mathbf{Z}_1\mathbf{B} = \mathbf{Z}_2 - \mathbf{E}_2 \mid$ trace $(\mathbf{E}_2^t\mathbf{E}_2)$ = min, and let $\mathbf{G}_2 = \mathbf{Z}_1\mathbf{B}_{12}$ be the image of the variables from V_2 in the space spanned by the vectors of variables from V_1. In a similar way, let $\mathbf{B}_{21} = \mathbf{R}_{22}^{-1}\mathbf{R}_{21}$ be a matrix of the standardized partial regression coefficients obtained by solving the regression problem $\mathbf{Z}_2\mathbf{B}_{21} = \mathbf{Z}_1 - \mathbf{E}_1 \mid$ trace $(\mathbf{E}_1^t\mathbf{E}_1)$ = min and let $\mathbf{G}_1 = \mathbf{Z}_2\mathbf{B}_{21}$ be the image of the variables from V_1 in the space spanned by the vectors of variables from V_2.

Finally, denote by $\mathbf{A}_{22} = \mathbf{G}_2^t\mathbf{G}_2 = \mathbf{R}_{21}\mathbf{R}_{11}^{-1}\mathbf{R}_{12}$ the covariance matrix of variables from $\mathbf{G}_2$, and by $\mathbf{A}_{11} = \mathbf{G}_1^t\mathbf{G}_1 = \mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}_{21}$ the covariance matrix of variables from $\mathbf{G}_1$. Obviously, because $\mathbf{Z}_2^t\mathbf{G}_2 = \mathbf{A}_{22}$ the matrix $\mathbf{A}_{22}$ is, in the same time, crosscovariance matrix between variables from V_2 and their images in the space spanned by vectors of variables from V_1. For the same reasons, the matrix $\mathbf{A}_{11}$ is, in the same time, crosscovariance matrix between variables from V_1 and their images in the space spanned by vectors of variables from V_2.

It is now natural to define as the nuclear matrix of the relations between variables from V_1 and V_2 the matrix

$$\mathbf{C}_{12} = \mathbf{G}_1^t\mathbf{G}_2 = \mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}_{21}\mathbf{R}_{11}^{-1}\mathbf{R}_{12}$$

that is the matrix of cross covariances between generalized images of variables from V_1 and V_2. The analysis of latent structure of $\mathbf{C}_{12}$ can therefore detect the main factors of $\mathbf{G}_1$ and $\mathbf{G}_2$ responsible for the causes of relations between variables from V_1 and V_2.

Spectral analysis of nuclear matrix gives some information of the latent dimensions responsible for the kernel of relationships between analyzed sets.

Let

$$\mathbf{C}_{12} = \mathbf{Y}_1\mathbf{L}\mathbf{Y}_2^t$$

where $\mathbf{Y}_1^t\mathbf{Y}_1 = \mathbf{I}$, $\mathbf{Y}_2^t\mathbf{Y}_2 = \mathbf{Y}_2\mathbf{Y}_2^t = \mathbf{I}$, $\mathbf{L} = (l_p)$: $l_p > l_{p+1}$ be the basic structure of nuclear matrix $\mathbf{C}_{12}$. Of course, all singular values l_p are not important for the detection of basic dimensions of the kernel of variables from V_1 and V_2. Define the number of important dimensions by Meig criterion, a rule of thumb criterion with good behavior in practice, as

$$k = \text{card}(l_p \geq (\text{trace } \mathbf{L})/m_2),$$

and denote left retained singular vectors as $\mathbf{Y}_1 = (\mathbf{y}_{1p})$, right retained singular vectors as $\mathbf{Y}_2 = (\mathbf{y}_{2p})$, and retained diagonal matrix of singular values as $\mathbf{L} = (l_p)$, p = 1,...,k.

Then the matrices

$$\mathbf{H}_1 = \mathbf{Y}_1\mathbf{L}^{1/2}$$

and

$$\mathbf{H}_2 = \mathbf{Y}_2\mathbf{L}^{1/2}$$

are, respectively, left and right structure matrices of nuclear matrix with the obvious property

$$\mathbf{H}_1\mathbf{H}_2^t = \mathbf{C}_{12} + \mathbf{E}, \text{ trace } (\mathbf{E}^t\mathbf{E}) = \min_k \ \forall \text{ k}.$$

Actually, the proposed method can be considered as the method of canonical covariance analysis (Momirović, Dobrić and Karaman, 1983) of variables from V_1 and V_2 transformed to generalized image form (Momirović, Štalec and Zakrašek, 1973), that is to the solution of the problem

$$\mathbf{G}_1\mathbf{y}_{1p} = \mathbf{u}_p,\ \mathbf{G}_2\mathbf{y}_{2p} = \mathbf{v}_p \ \Big|\ l_p = \mathbf{u}_p^t\mathbf{v}_p = \max,\quad \mathbf{y}_{1p}^t\mathbf{y}_{1q} = \mathbf{y}_{2p}^t\mathbf{y}_{2q} = \delta_{pq}$$

Let h_p and x_p be two unknown Lagrangian multipliers. The solution for $\mathbf{y}_{1p}$, $\mathbf{y}_{2p}$ and l_p can be easily obtained after the differentiation of the function

$$f(\mathbf{y}_{1p}, \mathbf{y}_{2p}, h_p, x_p) = \\ = \mathbf{y}_{1p}^t\mathbf{C}_{12}\mathbf{y}_{2p} - 1/2\ h_p(\mathbf{y}_{1p}^t\mathbf{y}_{1p}-1) - 1/2\ x_p(\mathbf{y}_{2p}^t\mathbf{y}_{2p}-1)$$

with respect to $\mathbf{y}_{1p}$ and $\mathbf{y}_{2p}$ and h_p and x_p, which gives, after some simple algebraic manipulations,

$$\mathbf{y}_{1p}^t\mathbf{C}_{12}\mathbf{y}_{2p} = l_p,$$

that is the singular values decomposition of nuclear matrix. Because spectral analysis of any cross covariance matrix is actually canonical covariance analysis of two sets of variables, transformed to some selected metrics, it was possible to derive identification structures such as pattern, intercorrelations and structure matrices.

Nuclear latent variables for the sets V_1 and V_2 are obviously vectors of the matrices

$$\mathbf{U} = \mathbf{G}_1\mathbf{Y}_1$$

and

$$\mathbf{V} = \mathbf{G}_2\mathbf{Y}_2$$

so that

$$\mathbf{W}_{11} = \mathbf{U}^t\mathbf{U} = \mathbf{Y}_1^t\mathbf{A}_{11}\mathbf{Y}_1$$

and

$$\mathbf{W}_{22} = \mathbf{V}^t\mathbf{V} = \mathbf{Y}_2^t\mathbf{A}_{22}\mathbf{Y}_2$$

are covariance matrices of nuclear variables.

Define $\mathbf{D}_1^2 = \text{diag } \mathbf{W}_{11}$ and $\mathbf{D}_2^2 = \text{diag } \mathbf{W}_{22}$. Then the matrices

$$\mathbf{M}_{11} = \mathbf{D}_1^{-1}\mathbf{W}_{11}\mathbf{D}_1^{-1}$$

and

$$\mathbf{M}_{22} = \mathbf{D}_2^{-1}\mathbf{W}_{22}\mathbf{D}_2^{-1}$$

are correlation matrices of nuclear latent variables, the matrices

$$\mathbf{F}_1 = \mathbf{G}_1^t\mathbf{U}\mathbf{D}_1^{-1} = \mathbf{A}_{11}\mathbf{Y}_1\mathbf{D}_1^{-1}$$

and

$$\mathbf{F}_2 = \mathbf{G}_2^t\mathbf{V}\mathbf{D}_2^{-1} = \mathbf{A}_{22}\mathbf{Y}_2\mathbf{D}_2^{-1}$$

are structural matrices, and

$$\mathbf{P}_1 = \mathbf{F}_1\mathbf{M}_{11}^{-1}$$

and

$$\mathbf{P}_2 = \mathbf{F}_2\mathbf{M}_{22}^{-1}$$

are, respectively, the pattern matrices of standardized nuclear latent variables. It is then possible to define quasicanonical correlations between latent dimensions.

Note that

$$\mathbf{L} = \mathbf{Y}_1^t\mathbf{C}_{12}\mathbf{Y} = \mathbf{U}^t\mathbf{V},$$

the diagonal matrix of singular values, is actually a cross covariance matrix of singular latent variables. Therefore,

$$\mathbf{y} = (y_p) = \mathbf{D}_1^{-1}\mathbf{L}\mathbf{D}_1^{-1}, \quad p = 1,...k$$

is a diagonal matrix of cross correlations between them, i.e., the matrix of quasicanonical correlations between retained nuclear latent variables from V_1 and V_2 with the obvious property $0 \leq y_p \leq 1$.

Because variables from **U** and **V** are asymptotically normally distributed and computed with m_1+m_2 degrees of freedom lost, the asymptotic tests of significance of the hypotheses $H_{0p} = y_p^* = 0$ are obviously

$$f_p = (y^2(1 - y^2))((n - m_1 - m_2)(m_1 + m_2)^{-1}), \quad p = 1,...,k$$

because variables f_p have, asymptotically, under H_{0p} Fisher - Snedecor F distribution with $n_1 = m_1 + m_2$ and $n_2 = n - m_1 - m_2$ degrees of freedom.

In order to clarify the interpretation of obtained results, relationships between robust methods and corresponding classical methods were established and examined. The formal relations between canonical correlation analysis and quasicanonical analysis of covariances were established (Hošek, Bosnar and Prot, 1984; Knežević and Momirović, 1996); quasicanonical analysis of covariances and principal component analysis (Gredelj, Momirović and Dobrić, 1986), least squares regression analysis and robust regression analysis (Dobrić, Štalec and Momirović, 1984). Diagnostic efficacy of robust discriminant analysis had been investigated (Momirović and Dobrić, 1988).

The taxonomic analysis of microsocial structures represented by binary graphs were solved by spectral decomposition of matrix of network of relations (Momirović, Hošek, Bosnar and Prot, 1984) and multidimensional scaling (Petrović and Momirović, 1972).

In Hotteling space (space of left eigenvectors of data matrices) general models were established where classical methods of data analyses could be treated as special cases (Momirović, Gredelj and Herak, 1978; Bosnar, Prot and Momirović, 1985: Momirović and Dobrić, 1988; Gredelj, Dobrić and Momirović, 1989).

Konstantin. Momirović continuously integrated new ideas related to acquisition, analysis and interpretations of results of data analysis procedures. In the field of development and applications of artifical neural networks new algorithms were developed for almost all relevant methods of data analysis. The nonlinear nature of results obtained by these new methods has significantly contributed to further data analysis conteptualisation beyond the classical general linear model based multivariate methods. Interesting results were obtained in taxonomic problems (Momirović, 2003; Popović and Momirović, 2003; Momirović, Hošek, Popović and Boli, 2003; Momirović, 2003; Momirović and Hošek, 2003; Knežević, Momirović, Radovanović and Radović, 2003; Bosnar, Prot, Momirović and Hošek, 2003; Prot, Bosnar, Hošek and Momirović 2003; Popović, and Momirović 2003).

2.3. Informatics/Applied Computer Science (design and development of information systems, designing new software tools for information systems, and multivariate data analysis and management)

The experiences and research problems Momirović has been faced with, and solutions he proposed were related to the multivariate nature of problems, the size of problems and the availability of computer resources (lack or limited access to computer resources) at the time being.

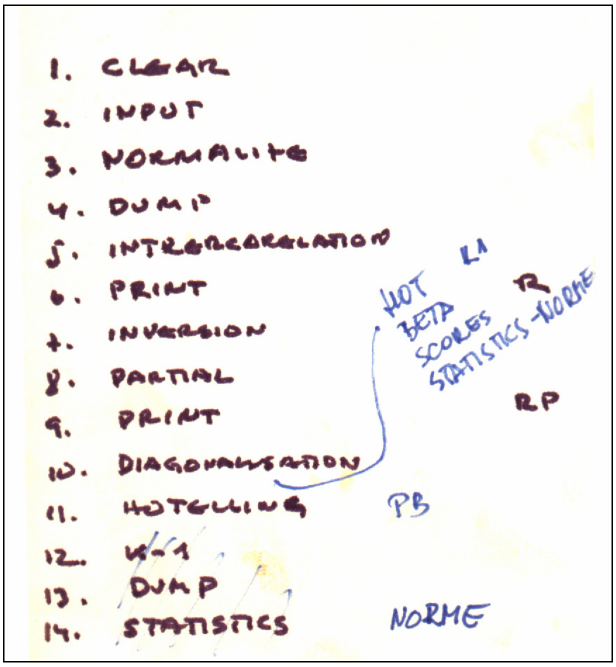
He has been very active in promoting and implementing the ideas of establishment of modern, contemporary, computing centers where ever there has been any chance to initiate them. Computing center of Institute of Kinesiology and the University Computing Centre - SRCE illustrate that (Budin, Jurišić-Kette, Momirović, Peruško, Požar, Simović, Stefanini, Turk, 1974).

The importance of information systems in the fields of education and sports had been anticipated too. Several projects have been proposed: Information system of secondary education (Momirović, Aurer, Maćašović, Gredelj, Gospodnetić, Obelić, Hađina and Stipanović, 1979); Information system of physical culture (Momirović, Gredelj, Štalec, Gospodnetić Milonja, Pavičić, Semenov and Stipanović, 1979); and Information system of top level sport (Ambrožić, Ban, Gospodnetić, Momirović, Pavičić, Pedalo, Semenov and Štalec, 1983).

It has been realized very early on that collections of statistical and data analysis subroutines should be interrelated through general statistical language. So, the first version of SS (statistical system) appeared to be active on the base of close cooperation of Zakrajšek, Štalec and Momirović (1969). Initial versions of SS were developed and coded in FORTRAN IV for IBM 1130 at the Institute for Mathematics and Physics "Jozef Štefan" at the University of Ljubljana. By the year 1971 it was further developed and maintained at the Institute of Kinesiology at the College for Physical Education and at the University of Zagreb for IBM series 1130 and 370/165 of electronic computers (Figures 3 and 4).

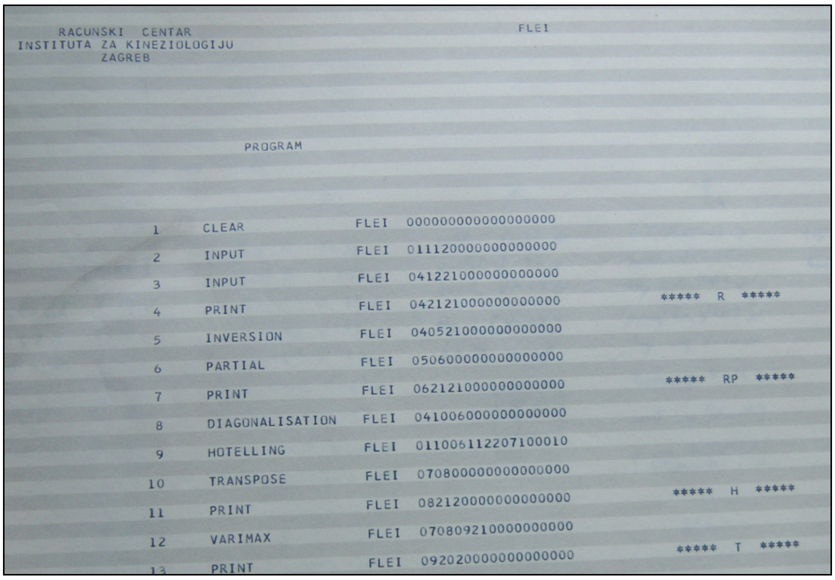
Upgrading the computing facilities in Ljubljana (CDC Cyber) and in Zagreb (UNIVAC

1106) stimulated development of SS at Ljubljana and Zagreb. It went on for some time with cooperation and coordination. After 1974, further maintenance and development went on in Zagreb. SS Statistical system (Zakrajšek, Štalec and Momirović; 1974), the command (linear) programming language became ground floor for multivariate statistical data analysis explorations and education The basic kernel of SS was implemented in FORTRAN V language and systematically improved under the EXEC 8 operational system UNIVAC 1110 and UNIVAC1130 computers at the University Computing Centre - SRCE of the University of Zagreb (Figures 5. and 6.).


1. CLEAR
2. INPUT
3. NORMALIZE
4. DUMP
5. INTERCORRELATION
6. PRINT
7. INVERSION
8. PARTIAL
9. PRINT
10. DIAGONALISATION
11. HOTELLING
12. K-1
13. DUMP
14. STATISTICS

LA
HOT
BETA
SCORES
STATISTICS -NORME
R
RP
PB
NORME

Figure 3. Momirović's handwritten note; Introductory part of SS coded program
(56 executive commands; component analysis in real and image space
with varimax and direct oblimin rotations of initial solutions)
for psychometric analysis of composite tests and computation of composite scores.


RACUNSKI CENTAR
INSTITUTA ZA KINEZIOLOGIJU
ZAGREB

FLEI

PROGRAM

1	CLEAR	FLEI	000000000000000000	
2	INPUT	FLEI	011120000000000000	
3	INPUT	FLEI	041221000000000000	
4	PRINT	FLEI	042121000000000000	***** R *****
5	INVERSION	FLEI	040521000000000000	
6	PARTIAL	FLEI	050600000000000000	
7	PRINT	FLEI	062121000000000000	***** RP *****
8	DIAGONALISATION	FLEI	041006000000000000	
9	HOTELLING	FLEI	011006112207100010	
10	TRANSPOSE	FLEI	070800000000000000	
11	PRINT	FLEI	082120000000000000	***** H *****
12	VARIMAX	FLEI	070809210000000000	
13	PRINT	FLEI	092020000000000000	***** T *****

Figure 4. Introductory part of SS coded program
(58 executive commands; component analysis in real and image space
with varimax and direct oblimin rotations of initial solutions)
program for component analysis (Computing Center of Institute of Kinesiology, Zagreb, 1971)

In the period from 1976 to 1985 new algorithms were developed and numerous computer programs were written and implemented as ordinary scientific and educational tools in Zagreb methodological circle headed by SRCE. Most of these programs were realized in SS (Statistical System) programming environment (Figure 7). New developed algorithms Konstantin Momirović implemented along with his fellow researchers Marijan Gredelj, Vesna Lužar, Vesna Dobrić, Maja Herak, Živan Karaman and Lajos Szirovicza from Universtiy Computing Centre - SRCE and Janez Štalec, Leo Pavičić, Ksenija Bosnar, Franjo Prot, and Nataša Viskić-Štalec from the Faculty of Physical Education.

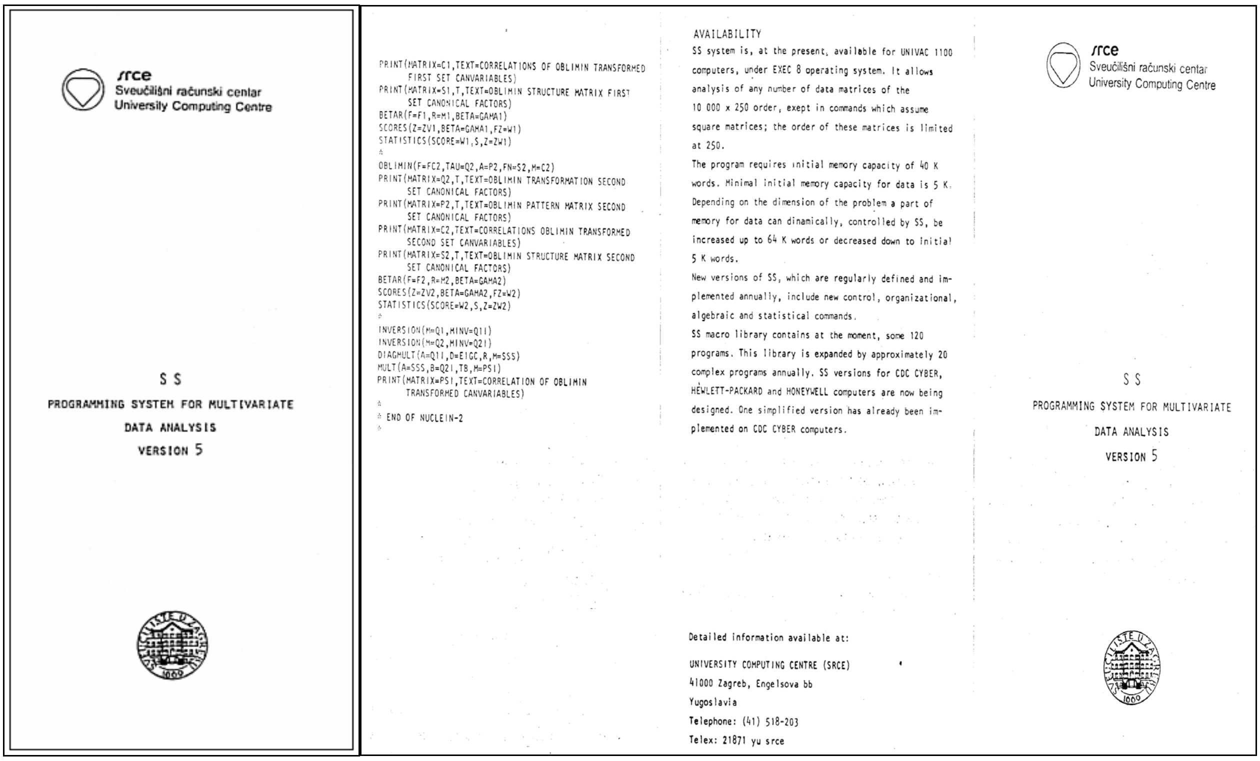
At the time SS has been promoted and accepted as the key tool for multivariate statistical data analysis among potential users, members of the scientific community.

```
*       * * * * *  N U C L E I N  -  2  * * * * *
*
*
OUTPUT(DEVICE=PR,ORIGINALS=2,COPIES=2)
HEADING(TEXT= N U C L E I N  -  2,T)
HEADING(TEXT= ANALYSIS OF CANONICAL CORRELATIONS OF
        LATENT DIMENSIONS)
HEADING(TEXT= A SIMPLIFIED VERSION OF MACROPROGRAM
        NUCLEIN,D)
*
*
*   NUCLEIN-2 ANALYZES THE CANONICAL RELATIONSHIPS OF
* TWO SETS OF VARIABLES REDUCED TO IMPORTANT PRINCIPAL
* COMPONENTS AND TRANSFORMS TO ORTHOBLIQUE POSITION,
* WITH OBLIMIN TRANSFORMATION OF SIGNIFICANT CANONICAL
* FACTORS.
*
*
INPUT(SCORE=S1)
INPUT(SCORE=S2)
CONFORM(IN1=S1,IN2=S2,OUT1=B1,OUT2=B2)
DELETE(MATRIX=S1)
DELETE(MATRIX=S2)
*
STATISTICS(SCORE=B1,S,Z=Z1)
CORRELATION(SCORE=Z1,R=R1)
EIGEN(R=R1,X=X1,EIG=L1)
HOTELLING(X=X1,EIG=L1,F=F1)
ORTHOBLIQUE(F=F1,A=A1,FN=FO1,M=M1,BETA=BET1)
PRINT(MATRIX=A1,T,TEXT=PATERN MATRIX OF FIRST SET)
PRINT(MATRIX=M1,TEXT=CORRELATIONS OF FIRST SET
        DIMENSIONS)
PRINT(MATRIX=FO1,T,TEXT=STRUCTURE MATRIX OF FIRST SET)
SCORES(Z=Z1,BETA=BET1,FZ=V1)
STATISTICS(SCORE=V1,S,Z=ZV1)
*
STATISTICS(SCORE=B2,S,Z=Z2)
CORRELATION(SCORE=Z2,R=R2)
EIGEN(R=R2,X=X2,EIG=L2)
HOTELLING(X=X2,EIG=L2,F=F2,EIGN=EEC)
ORTHOBLIQUE(F=F2,A=A2,FN=FO2,M=M2,BETA=BET2)
PRINT(MATRIX=A2,T,TEXT=PATTERN MATRIX OF SECOND SET)
PRINT(MATRIX=M2,TEXT=CORRELATIONS OF SECOND SET
        DIMENSIONS)
PRINT(MATRIX=FO2,T,TEXT=STRUCTURE MATRIX OF SECOND SET)
SCORES(Z=Z2,BETA=BET2,FZ=V2)
STATISTICS(SCORE=V2,S,Z=ZV2)
*
CROSSCORRELATION(P1=V1,P2=V2,R12=M12)
CANCORR(Z=V1,R1=M1,R2=M2,R12=M12,Q=0.01,F1=FC1,F2=FC2)
PRINT(MATRIX=FC1,T,TEXT=FIRST SET CANONICAL FACTORS)
PRINT(MATRIX=FC2,T,TEXT=SECOND SET CANONICAL FACTORS)
*
OBLIMIN(F=FC1,TAU=Q1,A=P1,FN=S1,M=C1)
PRINT(MATRIX=Q1,T,TEXT=OBLIMIN TRANSFORMATION FIRST SET
        CANONICAL FACTORS)
PRINT(MATRIX=P1,T,TEXT=OBLIMIN PATTERN MATRIX FIRST SET
        CANONICAL FACTORS)
PRINT(MATRIX=C1,TEXT=CORRELATIONS OF OBLIMIN TRANSFORMED
        FIRST SET CANVARIABLES)
PRINT(MATRIX=S1,T,TEXT=OBLIMIN STRUCTURE MATRIX FIRST
        SET CANONICAL FACTORS)
BETAR(F=F1,R=M1,BETA=GAMA1)
SCORES(Z=ZV1,BETA=GAMA1,FZ=W1)
STATISTICS(SCORE=W1,S,Z=ZW1)
*
OBLIMIN(F=FC2,TAU=Q2,A=P2,FN=S2,M=C2)
PRINT(MATRIX=Q2,T,TEXT=OBLIMIN TRANSFORMATION SECOND
        SET CANONICAL FACTORS)
PRINT(MATRIX=P2,T,TEXT=OBLIMIN PATTERN MATRIX SECOND
        SET CANONICAL FACTORS)
PRINT(MATRIX=C2,TEXT=CORRELATIONS OBLIMIN TRANSFORMED
        SECOND SET CANVARIABLES)
PRINT(MATRIX=S2,T,TEXT=OBLIMIN STRUCTURE MATRIX SECOND
        SET CANONICAL FACTORS)
BETAR(F=F2,R=M2,BETA=GAMA2)
SCORES(Z=ZV2,BETA=GAMA2,FZ=W2)
STATISTICS(SCORE=W2,S,Z=ZW2)
*
INVERSION(M=Q1,MINV=Q11)
INVERSION(M=Q2,MINV=Q21)
DIAGMULT(A=Q11,D=EIGC,R,M=SSS)
MULT(A=SSS,B=Q21,TB,M=PS1)
PRINT(MATRIX=PS1,TEXT=CORRELATION OF OBLIMIN
        TRANSFORMED CANVARIABLES)
*
* END OF NUCLEIN-2
*
```

Figure 5. NUCLEIN-2: SS macro coded program for canonical relationships among principal components of two sets of variables (Library SRCE*SS-MACRO of the University Computing Centre - SRCE, Zagreb)

Supported by SRCE this multivariate data analysis program library SRCE*SS-MACRO had been used in various scientific and technical disciplines. Most frequently used macro programs could be classified as: analysis of non-numerical data (8 macro programs); metric (psychometrics, i.e. kinesiometrics) properties of measurements (5 macro programs); explorative and confirmative factor or component analysis (17 macro programs); taxonomic (cluster) analysis, pattern recognition and classification (8 macro programs); multidimensional scaling (11 macro programs); canonical correlation analysis, multivariate regression analysis, redundancy analysis and discriminative analysis (22 macro programs); stochastic process analysis and analysis of changes (7 macro programs).

Multivariate statistical analyzer (MSA) had been designed as a blue print of new version of Statistical system (SS), as a novel generation of specification language for multivariate data analysis and statistics with full set of nested logical and iteration program control structures (Momirović, Aurer, Maćašović, Gredelj, Gospodnetić, Obelić, Hađina and Stipanović, 1979; Momirović, Gredelj, Štalec, Gospodnetić Milonja, Pavičić, Semenov and Stipanović, 1979; Momirović and Štalec, 1983). Limited human resources and financial capacities stopped further development of SS Statistical System.



srce
Sveučilišni računski centar
University Computing Centre

S S
PROGRAMMING SYSTEM FOR MULTIVARIATE
DATA ANALYSIS
VERSION 5

```
PRINT(MATRIX=C1,TEXT=CORRELATIONS OF OBLIMIN TRANSFORMED
      FIRST SET CANVARIABLES)
PRINT(MATRIX=S1,T,TEXT=OBLIMIN STRUCTURE MATRIX FIRST
      SET CANONICAL FACTORS)
BETAR(F=F1,R=M1,BETA=GAMA1)
SCORES(Z=ZV1,BETA=GAMA1,FZ=W1)
STATISTICS(SCORE=W1,S,Z=ZW1)
*
OBLIMIN(F=FC2,TAU=Q2,A=P2,FN=S2,M=C2)
PRINT(MATRIX=Q2,T,TEXT=OBLIMIN TRANSFORMATION SECOND
      SET CANONICAL FACTORS)
PRINT(MATRIX=P2,T,TEXT=OBLIMIN PATTERN MATRIX SECOND
      SET CANONICAL FACTORS)
PRINT(MATRIX=C2,TEXT=CORRELATIONS OBLIMIN TRANSFORMED
      SECOND SET CANVARIABLES)
PRINT(MATRIX=S2,T,TEXT=OBLIMIN STRUCTURE MATRIX SECOND
      SET CANONICAL FACTORS)
BETAR(F=F2,R=M2,BETA=GAMA2)
SCORES(Z=ZV2,BETA=GAMA2,FZ=W2)
STATISTICS(SCORE=W2,S,Z=ZW2)
*
INVERSION(M=Q1,MINV=Q1I)
INVERSION(M=Q2,MINV=Q2I)
DIAGMULT(A=Q1I,D=EIGC,R,M=SSS)
MULT(A=SSS,B=Q2I,TB,M=PSI)
PRINT(MATRIX=PSI,TEXT=CORRELATION OF OBLIMIN
      TRANSFORMED CANVARIABLES)
*
* END OF NUCLEIN-2
*
```

AVAILABILITY

SS system is, at the present, available for UNIVAC 1100 computers, under EXEC 8 operating system. It allows analysis of any number of data matrices of the 10 000 x 250 order, exept in commands which assume square matrices; the order of these matrices is limited at 250.

The program requires initial memory capacity of 40 K words. Minimal initial memory capacity for data is 5 K. Depending on the dimension of the problem a part of memory for data can dinamically, controlled by SS, be increased up to 64 K words or decreased down to initial 5 K words.

New versions of SS, which are regularly defined and implemented annually, include new control, organizational, algebraic and statistical commands.

SS macro library contains at the moment, some 120 programs. This library is expanded by approximately 20 complex programs annually. SS versions for CDC CYBER, HEWLETT-PACKARD and HONEYWELL computers are now being designed. One simplified version has already been implemented on CDC CYBER computers.

Detailed information available at:

UNIVERSITY COMPUTING CENTRE (SRCE)
41000 Zagreb, Engelsova bb
Yugoslavia
Telephone: (41) 518-203
Telex: 21871 yu srce

srce
Sveučilišni računski centar
University Computing Centre

S S
PROGRAMMING SYSTEM FOR MULTIVARIATE
DATA ANALYSIS
VERSION 5

Figure 6. Advertising leaflet for SS: Statistical System Version 5.
from the University Computing Centre - SRCE, Zagreb

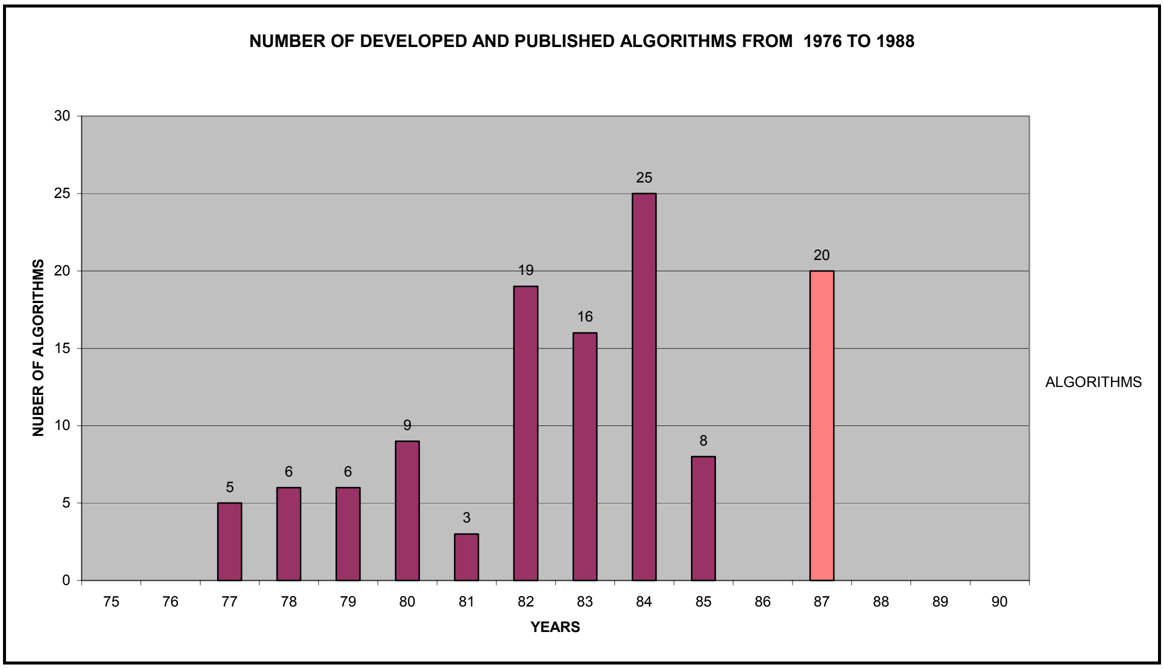


Figure 7. The dynamics of creation, i.e. algorithm development for SRCE*SS-MACRO (1976-1985) and SRCE*GENS-MACRO library (1986-1990)

In the period from 1986. to 1990. development and support of SS macro library was banned and further development was based on power and flexibility of GENSTAT system. Konstantin Momirović led a team of researchers (Marijan Gredelj, Vesna Lužar, Vesna Dobrić and Maja Herak, Ksenija Bosnar, Franjo Prot, Darko Dugić, Zlatko Knezović, Nataša Erjavec, Jovanka Radaković, Vesna Perišić, Josipa Kern) from the University Computing Centre - SRCE, Faculty of Physical Education and School of Public Health "Andrija Štampar" which established new public program library SRCE*GENS-MACRO on UNIVAC 1130 main frame computer (Figure 7.). Most frequently used macro programs from SRCE*GENS-MACRO library could be classified into: analysis of non-numerical data (3 macro programs); metric (psychometrics, i.e. kinesiometrics) properties of measurements (2 macro programs); explorative and confirmative factor or component analysis (4 macro programs); taxonomic (cluster) analysis, pattern recognition, classification and discriminant analysis (5 macro programs); correlation analysis, canonical correlation analysis, multivariate regression analysis, redundancy analysis and discriminative analysis (26 macro programs); stochastic process analysis and analysis of changes (22 macro programs).

In the period from 1996 to 2004 Momirović devoted himself to design and development of a whole new macro library implemented in SPSS macro language as a library of macro programs "IKSI" at the Institute for Criminological and Sociological Research.

Algorithms and macro programs developed from 1996 to 2004 (Figure 4.) could be classified into 21 groups, as shown in Table 1.

In total there are 762 macro programs from which 703 are explicitly referred to as being designed, developed, and coded by Momirović himself (Table 1.). It is important to notice that 390 technical reports are referenced in these almost self documented macros. Technical reports are documenting methodological contributions.

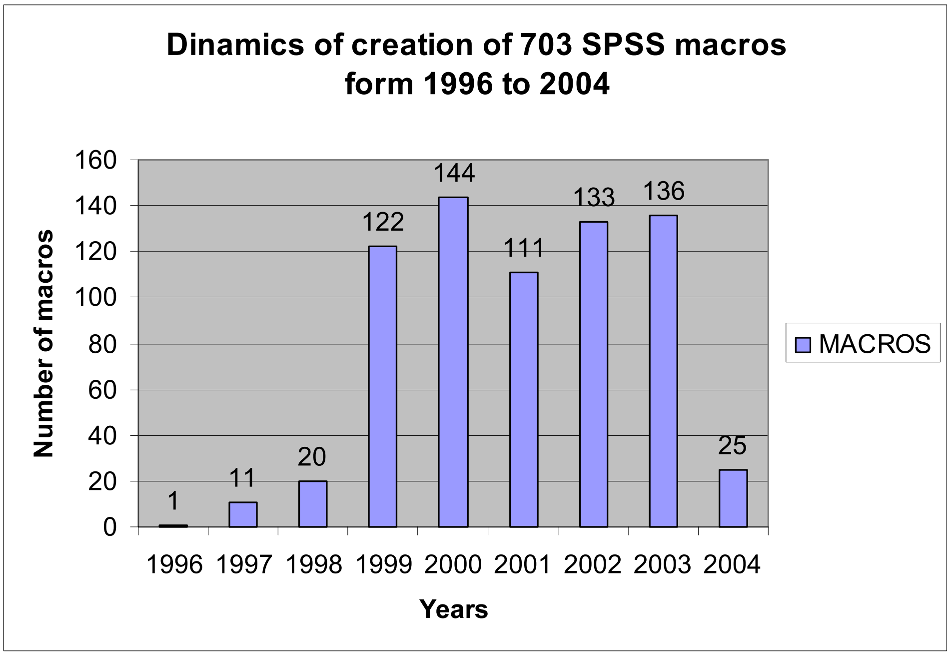


Figure 8. The dynamics of creation, i.e. algorithm development and SPSS macro coded programs (1996 - 2004)

The dynamics of creation of macro programs, i.e. algorithms between 1996 and 2004 demonstrate intensive and continuously high rate productivity for the last six years of his life (Figure 8.).

Phenomenon of Konstantin Momirović's work on this macro library, along with 390 associate technical reports, as a part of his legacy in the field of multivariate statistics and data analysis is worth further study

Table 1. The distribution of SPSS macro programs coded by Momirović, classified in 21 program groups / areas of multivariate data analysis.

PROGRAMS GROUP	No. of macro programs
(01) Analysis of metric characteristics of composite measurement instruments	46
(02) Component analysis	140
(03) Factor analysis	70
(04) Taxonomic analysis	51
(05) Metric multidimensional scaling	7
(06) Latent structure analysis	11
(07) Redundancy analysis	12
(08) Canonical correlation analysis	30
(09) Non-numerical data analysis	29
(10) Canonical covariance analysis	23
(11) Discriminant analysis	30
(12) Multivariate regression analysis	44
(13) Correlation analysis	22
(14) Analysis of qualitative changes	9
(15) Analysis of structural changes or structural differences	18
(16) Pattern recognition	9
(17) Emulation of neural networks	150
(18) Nuclear analysis	21
(19) Network analysis	4
(20) Descriptive and inferential statistical methods	21
(21) Auxiliary mathematical and statistical programs	24
T O T A L	771

3. Concluding Remarks

Konstantin Momirović, alias Stojan Hadžigalić, has been an active contributor to both theoretical and applied developments of multivariate data analysis from the early 1960s till his sudden death in 2004.

Some of his methodological contributions were published even after 2004, and are still published today.

From the point of view of development of his ideas on multivariate data analysis, it is possible to recognize several phases. In the first, initialization phase, he worked on methodological topics as individual assimilating and developing contemporary ideas. His contribution to the design of SS Statistical System for multivariate data analysis prepared the ground floor the organized theoretical and applied team work. In the second phase, he influenced and became the originator and principal figure in the so called "Zagreb Methodological and Multivariate Data Analysis Circle" at the University of Zagreb (Institute of Kinesiology at the Faculty of Physical Education and at the University Computing Centre - SRCE) where completely new ideas emerged. In the third phase he influenced and organized various productive groups of researchers around the Institute for Criminological and Sociological Research, and the Department of Psychology at the Faculty of Philosophy of University of Belgrade

Methodological research and achievements after 1971 could be divided into the following three areas of data analysis:

1. Kinesiometrics (development of new theoretical and applied measurement models in the field of measurement theory)
2. Multivariate data analysis and statistics (new models, methods and algorithms for data analysis)

3. Informatics (a field of computer science related to the development of new software for information systems, data analysis and management)

Evolution of some ideas, algorithmic solutions and programs implementations are evident. The contributions and experiences are finally divided as they are grouped and incorporated in the SPSS macro language coded library of more then seven hundred macro programs. Further research is needed to enable better understanding of the complexities of his ideas implemented in macro programs, especially where artificial neural network approach to multivariate data analysis problems is integrated.

This rather short and illustrative overview is just a preliminary attempt to the deeper insight into phenomenon of Konstantin Momirović's work and his legacy in the field of multivariate statistics and data analysis.

Acknowledgements

This research was partially supported by the Ministry of Science, Education and Sport of the Republic of Croatia, through the following projects:

1. Scientific Research Project No. 034-0342618-2222 "The Development of Algorithms for Multivariate Structural Hypotheses Testing", Principal Investigator: Franjo Prot.
2. Project No. 533-05-07-005 "APIS Reference Center: Data Analysis and Statistics Support for the Academic and Research Community", Project Coordinator: Vesna Hljuz Dobric.
3. Scientific Research Project No. 067-0362980-2484 "Systems for Managing Information, Processes and Knowledge", Principal Investigator: Vesna Bosilj Vuksic.

Photographs used in this paper were reproduced with the permission of:
- Emil Lehpamer, Faculty of Kinesiology (1971)
- Emil Hofman, Faculty of Kinesiology (1982)
- Franjo Prot, Faculty of Kinesiology (1997).

Authors express gratitude to the University Computing Centre - SRCE, organizer of the *ITI 2008 30th Anniversary Conference*, for an opportunity to present the rich scientific legacy of Konstantin Momirović (1932 – 2004) at the Memorial Session: "*Data Mining, Statistics and Biometrics*" dedicated to his memory (Hljuz Dobrić, V. et al. (2008); Prot, F. et al. (2008)).

References [*]

1. Anderson, T.W. (1958): An introduction to multivariate analysis. First edition New York: Willey.
2. Anić, V., I. Goldstein (1999): Foreign Words Dictionary (Rječnik stranih riječi). Zagreb: Novi liber (in Croatian).
3. Bock, D.R.(1975): Multivariate Statistical methods in behavioral research. New York, McGraw-Hill Book Company.
4. Bujas, Z., B. Petz, M. Mraković (1983): Report of 15/04/1983 for the nomination as full professor of "Quantitative Methods" (Professor K. Momirović), Faculty for Physical Education, University of Zagreb (in Croatian).
5. Cattell, R. B. (1966): *Handbook of multivariate* experimental *psychology.* Chicago: McNally.
6. Cooley, W. W.; Lohnes, P. R. (1971): *Multivariate data analysis.* New York: Wiley .
7. Dobrić, V. (1986): On a class of robust methods for multivariate data analysis. In: COMPSTAT 1986. Proceedings in Computational Statistics, Physica-Verlag, Heidelberg: 211-216.
8. Fajgelj, S. (2004): In memoriam Konstantin Momirović 1932-2004. Psihologija (Psychology), 37, 2: 219-225.
9. Guttman, L. (1945): A basis for analysis test - retest reliability. *Psychometrika,* **10**: 255 - 282.
10. Guttman, L. (1953): Image theory for the structure of quantitative variates. *Psychometrika*, **18**: 227-296.
11. Harris, C.W. (1962). Some Gutman-Rao relationships. Psychometrika, 27: 247-263.
12. Harman, H.(1960) Modern factor analysis. Chicago: University of Chicago Press.
13. Harman, H (1967) Modern factor analysis. 2nd edition.Chicago: University of Chicago Press.

[*] Full list of references can be found in: Prot, F., A. Hosek, K. Bosnar, V. Luzar-Stiffler, V. Hljuz Dobric, Z. Bekić, M. Gredelj (2008). Konstantin Momirovic (1932-2004): Biography and Bibliography .

14. Hljuz Dobric, V., F. Prot, V. Luzar-Stiffler, Z. Bekic, K. Bosnar, M. Gredelj (2008):. Konstantin Momirović (1932-2004) Memorial Session: *Data Mining, Statistics and Biometrics*. In: Luzar-Stiffler, V., V. Hljuz-Dobric, Z. Bekic. *Proc. ITI 2008 30th Int. Conf. Information Technology Interfaces*. Zagreb: University Computing Centre – SRCE, pp. 3 (CD version: IEEE Cat.No. CFP08498-CDR; ISBN 978-953-7138-13-4; ISSN 1334-2762).
15. Hopfield, J.J. (1982): Neural networks and physical systems with emergent collective computational abilities. *Proceedings of National Academy of Sciences*, 79:2554-2558
16. Horst, P. (1965): Factor analysis of data matrices. New York: Holt, Rinehart & Winston.
17. Vlahović, P.; P. Kovačević (2005): .... In Memory of Prof. Dr. Konstantin Momirović (1932-2004), (in Serbian)........... Pro Memoria Prof. Dr. Konstantin Momirović: 13. January 1932 - 28. March 2004. Belgrade: Glasnik antropološkog društva Jugoslavije (Revue de la Societa Anthropologique de Yugoslavie), 40:43-46 (in Serbian).
18. Kaiser, H. F.; Caffrey, J. (1965): Alpha factor analysis. *Psychometrika,* **30**: 1-14.
19. Kaiser, H. F. (1970): A second generation Little Jiffy. *Psychometrika*, **35**: 401-415.
20. Kališnik, M., J. Goričar, D. Ulaga (1979): Report of the Nomination Committee on K. Momirović full professorship for the course "Kinesiological Methodology" at the Physical Education College in Ljubljana (in Slovenian).
21. Krković, Anđelko; Momirović, Konstantin; Petz, Boris (1966): Selected chapters from psychometrics and nonparametric statistics. Zagreb: Croatian Society of Psychologists, pp. 120 (In Croatian).
22. Lužar, V. (1985): Inference in quasicanonical correlation analysis. *Proc. 7th International Symposium "Computer at the University"*, Zagreb: University Computing Centre - SRCE, 518: 1-6.
23. Macura, D. (2004): In memoriam Konstantin Momirović (1932-2004), Šport (Sport), 2004: 52-53.
24. Momirović, K. (1963): Factor structure of neurotic symptoms. PhD Thesis, Faculty of Philosophy, University of Zagreb (in Croatian).
25. Momirović, K. (1972): Applied Kynesiology - Short course in Kynesiometry. Lecture notes M.Sc. study in Kynesiology, year 1971/72. Zagreb: College for Physical Education, pp. 88.(in Croatian).
26. Momirović, K. (1976): Biography and Bibliography within the Application for lectureship on the course "Kinesiological Methodology" (Narodni list SRS, 25.12.1976), at the Physical Education College in Ljubljana (in Slovenian)
27. Momirović, K., B. Aurer, P. Maćašović, M. Gredelj, R. Gospodnetić, V. Obelić, Hađina and R. Stipanović, (1979): Information System of Secondary Education in Croatia. Zagreb: University Computing Centre - SRCE (in Croatian).
28. Momirović K., M. Gredelj, J. Štalec, R. Gospodnetić, Z. Milonja, L. Pavičić, I. Semenov i R. Stipanović (1979): Conceptual project of Information System for the Croatian Psysical Education and Sports. Zagreb: Faculty of Physical Education, (12 books, in Croatian)
29. Momirović, K. (1989): Biography for the "Ramiro Bujas Award 1989", Zagreb: Croatian Society of Psychologists (in Croatian).
30. Momirović K., J. Štalec, E. Zakrajšek (1982): A programming language for multivariate data analysis. *COMPSTAT 82, Proceedings in Computational Statistics - Part I, Physica Verlag, Wien*: 87-95.
31. Mulaik, S. A. (1972) Foundations of factor analysis. New York, McGraw-Hill Book Company.
32. Prot, F. (2004): In memoriam prof. dr. Konstantin Momirović (1932-2004). Unpublished manuscript submitted to “Primjenjena psihologija” (Applied Psychology), (in Croatian).
33. Prot, F. and K. Bosnar (1988): An algorithm and program for the analysis of individual changes based on absolute and relative measures of change. *Proc. 10th International Symposium "Computer at the University"*, Zagreb: University Computing Centre - SRCE, 5.4: 1-5.

34. Prot, F., K. Bosnar, M. Mraković (2004): Research methodology in Kinesiology. In Jaklinović, Ž. et al, Editors (2004): Materials for Bibliography, Faculty of Kinesiology 1959 – 2004. Faculty of Kinesiology, Zagreb: 24-37 (in Croatian).

35. Prot, F., A. Hošek, K. Bosnar, (2008): Prof. dr. Konstantin Momirović (1932-2004): Contributions to Biography and Bibliography, pp. 85 (Unpublished manuscript, in Croatian)

36. Prot, F., A. Hosek, K. Bosnar, V. Luzar-Stiffler, V. Hljuz Dobric, Z. Bekić, M. Gredelj (2008): Konstantin Momirovic (1932-2004): Biography and Bibliography In: Luzar-Stiffler, V., V. Hljuz-Dobric, Z. Bekic. *Proc. ITI 2008 30th Int. Conf. Information Technology Interfaces*, Zagreb, University Computing Centre – SRCE, pp.200 (CD version: IEEE Cat.No. CFP08498-CDR; ISBN 978-953-7138-13-4; ISSN 1334-2762).

37. Rao, C. R. (1952): Advanced statistical methods in biometric research. New York: Wiley.

38. Rao, C.R. (1973) Linear statistical inference and its application. New York: Willey.

39. Šipka, P. (2005): Social Studies in Serbia, with and without Momirović: a bibliometric portrait. Psihololologija (Psychology) , 38, 3: 345-360 (in Serbian).

40. Štalec, J., K. Momirović, E. Zakrajšek (1982): Programming in SS: Description of a set of macro programs for multivariate data analysis. *COMPSTAT 82, Proceedings in Computational Statistics - Part II, Physica Veralg, Wien*: 251-252.

41. Zakrajšek, E., J. Štalec, K. Momirović (1974): SS – A System of Programs for Multivariate Data Analysis, *Proc. 1st Int. Sympsium "Computer at the University",* Zagreb: University of Zagreb, University Computing Centre - SRCE, C8:1-16 (in Croatian).