Małgorzata Ćwiklińska-Jurkowska
Department of Theoretical Backgrounds of Biomedical Science
and Medical Informatics, Collegium Medicum in Bydgoszcz,
Nicolaus Copernicus University

EXPLORATORY DATA ANALYSIS FOR THE HEMATOLOGICAL FEATURES. PART I. METHODOLOGY

Abstract: Part I of the work characterizes the data of human blood parameters. It describes examined biomedical data set and used multivariate statistical exploratory methods like PCA, FA, MDS and clusters analysis. To the factor analysis methods belongs the biplots visualization method. Biplots are simply the scatterplots with the superimposition of the variables. The work compares different other alternative multivariate exploratory data analysis procedures from methodological point of view. The described multivariate ordination methods are applied in the Part II of the work "Application".

Key words and phrases: exploratory data analysis, graphical data visualization, principal components, biplots, MDS, cluster analysis

Introduction

In exploratory data analysis (Larose 2005), one usually has not the a priori idea of expected relations between variables. Exploratory analysis allows to discover interesting relationships between variables. It also enables to identify interesting subsets of a data set and to develop initial ideas of possible connections between features and perhaps with the dependent or classifying variable. Graphical exploratory methods in statistics can explore even unknown phenomenon.

Ordination, i.e. geometrical representation of multivariate data as a low dimensional arrangement of points is a required procedure in many applied research problems. A picture is often needed which can provide meaningful interpretation of the data. Hopefully, the picture will supply helpful information about the relationships between the individuals. When one has a priori grouping of the individuals and searches a low dimensional illustration of the data, highlighting differences between them means one can apply e.g. canonical variate analysis or its generalization. Following the application of

the explorative data analysis in a research domain one wants to know what are the important configurations or characteristics that can be seen in this field and how much of the variability is clarified by them.

Very common statistical technique is to plot a scatter diagram showing the pattern of relationships between a set of samples for only two or three original variables. The next step for such scatter diagrams may be looking for trends (e.g. regression lines), clusters, outliers, collinearities or other regularities. The multivariate extensions are needed to represent p-variate set of observations. To display multivariate data in a two dimensional plot, Bartkowiak et. al (1998) applied methods of the grand tour and hierarchical visualization.

In multidimensional statistical problems one wants to reduce a data set containing many variables to a data set containing considerably smaller number of variables, but that still corresponds to a large part of the variability embodied in the original data set. After applying explorative data analysis in a research domain, one wants to know what are the important patterns that appear in this field and how much of the variance is explained by them.

One of the frequent possibilities include, in the first step, the approximation of p-variate space into s-variate space, saving as much information derived by the data as possible. Common methodology is the Principal Component Analysis (PCA). In the principal component analysis p-dimensional scatter of these cases is approximated by their scatter in an r-dimensional sub-space, obtained by orthogonal projection of R_p onto R_s, chosen as this mininimizing the sum of squares of the residuals orthogonal to R_s. The approximation of the bipolot variables is given by the biplot axes.

Useful method of data ordination, including the information of both cases and variables are biplots. Biplots are simply the scatter plots with the superimposition of the variables. Thus, by using this method of data ordination, including the information of both cases and variables, the relationship between the hematological observations and variables may be investigated.

During interventions with extracorporeal circulation, such as hemoperfusion, different blood parameters are changing. The interdependences between different hematological parameters of human blood are interesting. In multivariate statistical problems there is an obvious need for graphical visualization of the data. Simple biplot diagnostic modeling may be applied. Investigation of the structure of the hematological data by studying scatterplot matrices and biplots with compared additional multivariate techniques is the aim of this work. The examined data set includes measurements of blood when hemoperfusion was done with polymer sorbent.

Material

Experiments in vitro with polymeric sorbents and interdependencies between variables describing parameters of human blood are considered. Worse hematological parameters are frequent complications of interventions with extracorporeal circulation, so eleven hematological variables, reported for each of the 36 experiments in vitro, performed in Colleqium Medicum of Nicolaus Copernicus University were examined, where also the feature of the perfusion experiment time was considered. Hemoperfusion of human blood was done with the polymer sorbent.

Table 1

Names and analyzed variables' description

Feature (variable)	Description
ADHEVIN	Level of platelets' adhesion
TIME	TIME trwania perfuzji / Time of perfusion
CZFIVIN	TIME fibrynolizy / Time of fibrinolise
ERYTVIN	Liczba erytrocytów / Number of erythrocytes
FIBGVIN	Fibrinogen concentration
KAOLVIN	Kaolin-kefalin time
KEFAVIN	Kefalin time
LEUKVIN	Leukocytes number
PROTVIN	Prothrombin time
STYPVIN	Stypven-kefalin time
TROMVIN	Thrombocytes number

The description of the examined hematological features is given in Table 1.

"ADHEVIN" – Level of platelets' adhesion; "TIME" – time of perfusion ("CZAS"); "CZFIVIN" – Time of fibrinolise; "ERYTVIN" – Number of erythrocytes; "FIBGVIN" – Fibrinogen concentration; "KAOLVIN" – Kaolin-kefalin time; "KEFAVIN" – Kefalin time; "LEUKVIN" – Leukocytes number; "PROTVIN" – Prothrombin time; "STYPVIN" – Stypven-kefalin time and "TROMVIN" – Thrombocytes number.

Methods

In multivariate analysis looking simultaneously at many variables is difficult. The most advanced and complex methods do not exceed five origi-
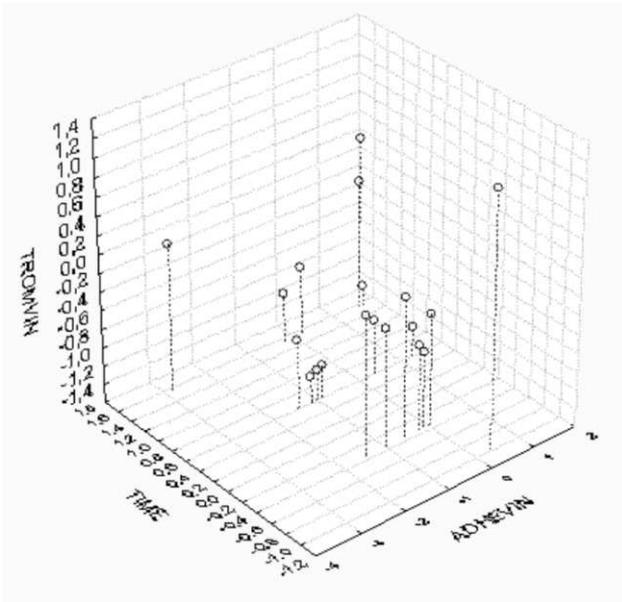
Fig. 1. Scatterplot of three dimensional space based on original variables ADHEVIN, TIME, TROMVIN
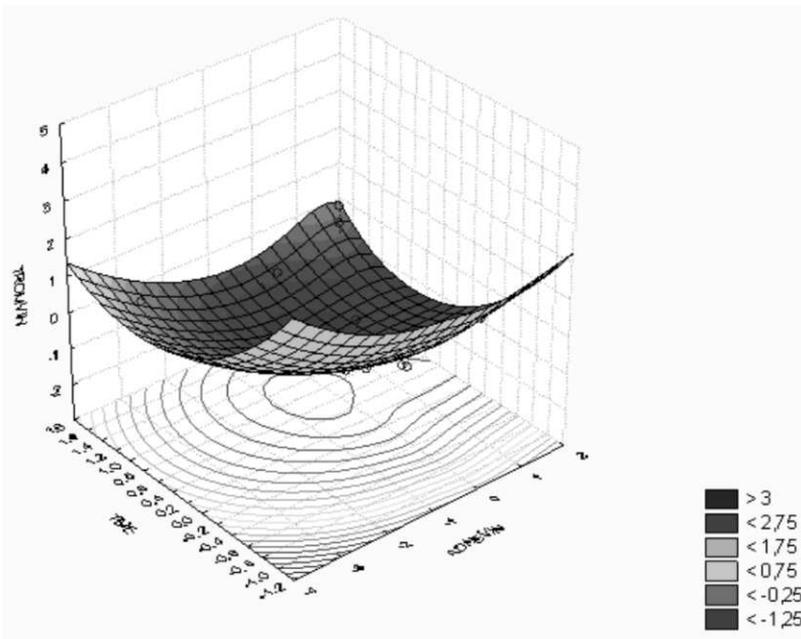

Fig. 2. Scatterplot of three dimensional space based on original variables ADHEVIN, TIME, TROMVIN with overlaid smoothed surface of the relationship between the three variables

nal variables dimensionality. For example, we can look at three-dimensional scatterplot for all three subsets of variables. Fig. 1 presents such three-dimensional scatterplot for applied hematological data set. Fig. 2 contains the same scattterplot with overlaid smooth surface of relationship between the three variables.

A scatterplot matrix permits simultaneous looking at relations between pairs of variables by observing two-dimensional scatterplots presenting the values for fixed pairs of variables. The multivariate extensions (for higher than two or three-dimensional data set) are useful to represent p-variate set of observations. One of the possibilities includes in the first step the approximation of p-variate space into s-variate space, saving as much information derived from the data as possible. Common methodology is the Principal Component Analysis (PCA). In principal, component analysis p-dimensional scatter of these cases is approximated by their scatter in an r-dimensional sub-space, obtained by orthogonal projection of R_p onto R_s, chosen as minimizing the sum of squares of the residuals orthogonal to R_s. Biplots can provide diagnostic aid in analysis.

Biplots are useful procedures in exploratory data analysis and visualization of data sets. The rules of constructing a biplot are given by papers of Gabriel (1971, 1981, 1990) and also can be found in the book written by Krzanowski (1988, 1995).

Principal components are uncorrelated linear combinations of the variables and with variances $\lambda_1, \lambda_2, \ldots, \lambda_p$. Principal components are helpful for reducing the number of variables by finding linear combinations that explain the biggest part of the variability. The coefficients of each principal component are determined by eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$ of covariance matrix Σ. The principal components are sorted by descending (non increasing) order of the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$, which are equal to the variances of the components. Each principal component is a linear combination of the original variables, with coefficients equal to the eigenvalues of the covariance matrix Σ.

The eigenvectors are as a rule taken with unit length. The second component is orthogonal to the first and so on. The coefficients are customarily normalized to 1. The consequence of orthogonality of principal components is a summarizing of variances of all principal components λ_i to overall variance. The usefulness of the principal component is measured by the value of the variance which this component explains.

Using s first principal components (with the highest variances) summarizing most of the variability in the data, one can approximate high-dimensional data in a lower-dimensional linear subspace (Morrison 1976, Mardia

et al. 1979). In this way the dimension can be reduced from p to s. The coordinate system is rotated to make parallel first s coordinates $Y_1, \ldots, Y_s$ with the first s eigenvectors corresponding to eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_s$ of the common sample covariance matrix Σ.

Connections between similar ordination multidimensional methods like canonical variate analysis, principal component analysis and multidimensional metric scaling are recapitulated below in Tab. 2.

Table 2

Relations between methods of multivariate data ordination based on correlations

	Continuous (quantitative) variables	*Mixed quantitative-qualitative variables*	*Qualitative variables*
A priori grouping into k populations	Canonical variate analysis (Rao 1973, Mardia et al. 1979)	Metric scaling for matrix (*pxp*) of inter-population distances (+add a point technique (Krzanowski 1994)	Correspondence analysis (Hill 1974)
No a priori grouping (one population)	Principal Component Analysis	Metric scaling for matrix (*nxn*) of inter-object distances (Mardia et al. 1979)	Correspondence analysis (Hill 1974)

The fundamental tool for obtaining biplots is the PCA. Principal component analysis (PCA, Hotteling 1936) yields usually reduced plots of multivariate observations. A biplot is a graphical representation of information given in nxp matrix X. Biplots commonly use two or three dimensions (Gabriel 1971, Bartkowiak et al. 1996). The term "biplot" is not concerned with the dimension of representation, but means that this is a dual representation, variables and individuals on the same plot. Thus tree-dimensional biplots are also possible, or even higher dimensional, though not clear.

Biplots are constructed on the base of the fact that any cases matrix X (nxp) can be expressed as the product of two matrices G (nxr) and H – (rxp), i.e. $X = GH$, where $r = \text{rank}(X)$. It follows that for any multivariate observation, $x_{ij} = g_i' h_j$ ($i = l, \ldots, n$; $j = 1, \ldots, p$), where g_i'; and h_j are the r-dimensional rows of G and columns of H, respectively. In particular, the g_s' signify the observations and the h's the variables. Thus, in a result biplot a simultaneous representation of data and variables is obtained. For $r = 2$, the biplot can be displayed as a two-dimensional scatterplot of $n + p$ points (n = number of individuals, p = number of variables). Since

$r = \text{rank}(\boldsymbol{X})$, the biplot is only an approximation if $r > 2$. For $r = 3$, the biplot can be displayed as a three-dimensional scatterplot of $n + p$ points. Again, the biplot is only an approximation if $r > 3$.

Biplots give us the possibility of simultaneous graphical representation of both observations (usually marked as points) and variables (usually marked as vectors). These graphs are the scatter plots with the superimposition of the variables. Thus the relationships between the data and variables can be investigated.

Matrices of ranks higher than two cannot be represented exactly by a biplot. However, if a matrix can be acceptably approximated by a rank two matrix, the biplot may allow useful approximate visual inspection of a given matrix. To approximate any rectangular *nxp* matrix of rank r by a *nxp* matrix of lower rank, one may use the singular value decomposition (Gabriel 1971). The goodness of fit of biplot is $\frac{\lambda_1,\ldots,\lambda_s}{\lambda_1\ldots,\lambda_n}$, where s is the number of chosen eigenvectors (the dimensionality of the representation. As we extract consecutive factors, they give an explanation for less and less variability. The decision of when to stop adding consecutive factors mainly depends on when there is only very small rest variability left. The character of this decision may be subjective; on the other hand, various strategies have been developed (Kaiser, 1966; Cattel, 1960). Some authors suggest to take the number of components that ensure obtaining the above value higher than 0.75. Other criterions are Cattel (1966) scree plot principle and Kaiser (1960) criterion (the number of eigenvalues λ_i bigger than 1). The most widely used Kaiser criterion means that if a factor does not extract at least as much as the equivalent of one original variable variability, we drop it.

Biplots have been developed and applied for few decades (Greenacre, 2010; Gower, 2003; Gower at. al 1996, 2010 a,b). An overview of generalizations of the classical linear biplots (based on principal component analysis) is given by Krzanowski (1995). One of the special cases of this generalization is the correspondence analysis (Krzanowski 1995 chap. 12). Biplots may be applied to the principal component analysis (PCA) and the canonical correlation analysis (CCA). Robust biplots have been elaborated by Daigle (2008). In generalized biplots (Gower 1992), both continuous and categorical variables are permitted. The special cases of generalized biplots are classical linear biplots, nonlinear biplots (where the linear axes are replaced by nonlinear trajectories) and, for categorical variables, a new case for multiple correspondence analysis.

Biplots may be also examined in the correspondence analysis (for not measurable data) (Greenacre 1993, Demey et. al. 2003). For classification tasks biplots are also very useful. Interpolative biplot was proposed by Alves

and Oliveira (2003). They showed that while predictive biplots are the best option for interpretation purposes, interpolative biplots are very helpful for classification of new cases (instances) that were not applied for the construction of the principal component or canonical dimension axes. Biplots were generalized also for usage in discriminant analysis (Gardner S. and le Roux 2005). Besides, modern tasks coming from micrarray data sets, biplots are also used, though high dimensionality is met in those problems (Demey et.al 2008). Improved biplot techniques ("better biplots" – Blasius, et. al 2009) are useful when there are many points (perhaps several thousand) and the entire graphical effect of typical biplot can be very confusing. In such situation (Blasius, et. al 2009) propose a number of procedures. For example, the density representation of the points may be applied. Choosing more than one centre of concurrency or the use of colour is also useful. Another possibility is, while respecting the calibrations, moving the axes to new positions more distant from the points, and possibly jointly rotating axes and points. Graphical aspects of biplots in classical and advanced biplot techniques are provided by Gower (2003, 2004).

Biplots may be also performed for higher dimensional representation by only two or three axes, however for graphical interpretation usually two or three dimensional plots are performed (if two or three dimensions represent sufficient variability).

The classical biplot technique makes no provision for missing values on any of the variables to be analyzed. The problem with missing values can be solved by deleting a not complete variable or case or by imputing missing values by mean, median or the value obtained by more elaborated special techniques, e.g. expectation-maximization or propensity score.

Other visualization methods are also applied for comparison purposes. Very close to PCA is the principal factor analysis. The term "factor analysis" includes both principal components and principal factors analysis. In factor analysis, the similarities between objects (e.g., variables) are expressed in the correlation matrix. Factor analysis requires that the underlying data is distributed as multivariate normal, and that the relationships are linear.

PCA and principal factor analysis serve different aims. The principal components analysis is rather a method dimension of dimensionality and principal factors and are often preferred when the goal of the analysis is aimed at the structure recognition. The purpose of PCS is to identify linear orthogonal linear combinations of variables, used for description or for substituting original variables by the smaller number of uncorrelated components. On the other hand, principal factors supply the model of the data and therefore are more complex. In principal components analysis it

is assumed that all variability is supposed be used in the study, whilst in principal factors analysis just the variability common with the other items is employed. Generally, the two methods from factor analysis domain as a rule supply very similar outcomes.

Another applied ordination method is the MDS – multidimensional metric scaling (Tab. 2). The aim of MDS procedure is to rearrange a map from a table of distances between observations points on the plane. Classical multidimensional scaling (Cox & Cox 1994) of a data matrix is also known as the principal coordinates analysis (Gower, 1966). Multidimensional scaling for a set of dissimilarities (or distances) yields returns a set of points such that the distances between the points are approximately equal to the dissimilarities (Digby 1986, Krishnaiah 1977, Mardia et. al 1079). MDS method applies a function minimization procedure that assesses different configurations with the objective of maximizing the goodness-of-fit. Goodness-of-fit statistic (also called "stress measure") based on primal and resulting distances checks how well the distances between objects can be reproduced by the new configuration. The difference with PCA is that in factor analysis, the similarities between objects are expressed in the correlation matrix. The axes from the MDS analysis are arbitrary, and can be rotated in any direction.

MDS identifies important primary dimensions that give the possibility to explain observed similarities or dissimilarities (distances) between the examined objects. With MDS, any kind of similarity or dissimilarity matrices can be considered, among them also the covariance matrix. MDS derives scatterplots of the objects in the different two-dimensional planes, though even three-dimensional MDS plots are possible. The attractiveness of MDS is that we can analyze any kind of distance or similarity matrix. Multidimensional scaling may be also applied for only rank-ordering of distances (or similarities) in the matrix (non-metric multidimensional scaling Mardia et. al 1979). Factor analysis often derive more dimensions than MDS and as a consequence, MDS often provides more clear, interpretable results. In MDS, only the distance is necessary for calculations, in factor analysis the origin data set is needed. Thus, MDS methods are appropriate to a wide range of examinations, because the distance measures can be constructed in many manners.

Cluster analysis

Grouping is often used in many scientific research domains, e.g. in taxonomy. Cluster analysis is a form of unsupervised learning (without the

a priori knowledge of groups). It is used to see if natural groupings are present in the data. A cluster is a group of objects that are similar to one another, and dissimilar to objects in other clusters. In cluster analysis (belonging to the set unsupervised classification procedures), in contrast to supervised classification, there is no target variable for clustering. Thus, cluster analysis can not result in estimate or prediction of the value of the target variable and therefore the cluster procedures' search for the division of the whole objects data is set into homogeneous objects subsets.

Cluster analysis (Hartigan 1975) is an exploratory data analysis technique which classifies (by different unsupervised classification methods) different objects into groups in a manner that the similarity between two clustered is maximal if they belong to the same group and minimal otherwise (Cutsem 1994). The division is conducted into mutually non–overlapping groups. Clustered variables show the similarities between variables. The clustering method applies the dissimilarities (distances) or similarities between objects when creating the clusters. Those measures are used to define the criteria for grouping or separating objects. Different dissimilarity or distance measures may be applied in the cluster analysis (Cutsem 1994), but usually the Euclidean, City-block (Manhattan) distance or Chebychev distance are used.

Cluster analysis is a set of methods of the data exploration. It can be the first step before applying other exploratory analysis procedures. For example, clustering is useful in the selection of best diagnostic features. Variables very close (in the same clusters) may suggest the reduction of features and in the results one may obtain reduced dimensionality. The clustering method has been used lately for gene expression clustering, where very large quantities of genes may exhibit similar behavior. This may reduce the analysis of many genes into important ones.

Hierarchical methods are most popular. If two groups are selected from different partitioning (at different levels) then either these groups are disjoint or one is included in the other. Hierarchical methods may be divided into agglomerative or divisive ones.

The assumptions of linearity and normality, though are often met in multidimensional methods, are not necessary in cluster analysis. However, the representativeness is needed and the collinearity may create unreal picture of clusters.

When the researcher expects that both observations and variables concurrently contribute to the recognition of an important configuration of clusters, the two-way clusters may be applied. In two-way clustering the similarities between different clusters of observations may be induced by

somewhat different subsets of variables. Thus, the resulting arrangement of clusters is by nature not homogeneous. This method offers a powerful exploratory data analysis tool and is often applied in solving bioinformatics problems where gene expression data sets in the so called "heat maps". Genes are clustered in one direction and on perpendicular axe. The studied cases are then grouped, which next can be considered according to the kind of diseases or prediction criterion.

It is worth noting that clustering is often the first step before the applying other, more advanced multidimensional methods from the exploratory data analysis domain, for example in supervised learning. In such case, the first step is clustering, which may be used for the selection or grouping of redundant variables.

Concluding remark

Different ordination methods have different advances and drawbacks and may have different possibilities such as inspecting the data set from different point of views

R E F E R E N C E S

[1]   Alves M. R., Oliveira M. B. Interpolative biplots applied to principal component analysis and canonical correlation analysis. Journal of Chemometrics; Volume 17, Issue 11, pages 594–602, November 2003.

[2]   Bartkowiak A. Liebhart J. Szustalewicz A. 1996. Visualizing the correlation structure by a biplot extended to 3 dimensions. 34th International Center of Biocybernetics. Seminar. Statistics and Clinical Practice. Warsaw, 24–28 June, 1996. pp. 50–52.

[3]   Bartkowiak A. Lisp-Stat. Narzędzie eksploratywnej analizy danych. In Polish. Uniwersytet Wrocławski. 1995.

[4]   Bartkowiak A. Szustalewicz A. 1998. Watchings steps of a grand tour implementations. Machine Graphics & Vision, 7(3), pp. 655–680.

[5]   Blasius, J., Eilers, P. H. C., Gower, J. C. (2009) Better Biplots, CSDA, 53, 3145–3158.

[6]   Cattel R. B. (1966) The scree test for the number of factors. Mutivariate Behavioral Research, 1, 245–276.

[7]   Cuadras C. M, Rao C. R. 1993. Recent advances in biplot methodology. Multivariate analysis: future directions 2, 1993 – North Holland.

[8]   Cox, T. F. and Cox, M. A. A. (1994) Multidimensional Scaling. Chapman and Hall.

[9] Cutsem B. 1994. Classification and Dissimilarity Analysis. Springer-Verlag, New York.

[10] Daigle G. 1992. A robust biplot. Canadian Journal of Statistics. Volume 20, Issue 3, pages 241–255.

[11] Demey, J. R., Vicente-Villardón, J. L., Galindo-Villardón, M. P. and Zambrano, A. Y. (2008). Identifying molecular markers associated with classification of genotypes by External Logistic Biplots.

[12] Digby P. G. N., Gower J. C. Ordination and classification. Le Presses de L'Universite de Montreal 1986.

[13] Gabriel K. R. 1971. The biplot graphics display of matrices with application to principal component analysis. Biometrika 1971 (58), 453–467.

[14] Gabriel K. R. 1981. Biplot Display of Multivariate Matrices for Inspection of Data and Diagnostics, Interpreting Multivariate Data, Ed. V. Barnett London: John Wiley & Sons.

[15] Gardner S., le Roux N.J. Extensions of Biplot Methodology to Discriminant Analysis.. Journal of Classification 22:59–86 (2005).

[16] Gower, J. C. (1966) Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53, 325–328.

[17] Gower J. C. and Harding S. (1988) Non-linear biplots. Biometrika, 73, 445–55.

[18] Gower J.C. (1990). Three-dimensional biplots. Biometrika 1990 77(4):773–785.

[19] Gower J.C. (1992) Generalized biplots. Biometrika 79, 475–493.

[20] Gower J. C. (2003). Unified biplot geometry. Ii: Developments in Applied Statistics. Eds: Ferligoj A, Mrvar A. Metodoloski zvezki, 19, Ljubljana: FDV.

[21] Gower J. C. (2003) Visualisation in Multivariate and Multidimensional Data Analysis. Bulletin of the International Statistical Institute 54, 101–104.

[22] Gower J. C. (2004). The geometry of biplot scaling. Biometrika 91(3): 705–714.

[23] Gower, J. C., van der Velden, M. and Groenen, P. J. F. (2010). Area Biplots. Journal of Computational and Graphical Statistics, 19. 46–61.

[24] Gower, J. C. Hand, D. J. (1996) Biplots. London: Chapman and Hall.

[25] Gower, J. C., Lubbe, S. and LeRoux, N, D. J. (2010) Understanding Biplots. Chichester, John Wiley.

[26] Greenacre, M. (2010). Biplots in Practice. BBVA Foundation, Madrid, Spain.

[27] Greenacre, M. J. 1993. Biplots in correspondence analysis. Journal of Applied Statistics 20: 251–269.

[28] Hartigan J.A. 1975. Clustering Algorithms, Wiley, New York.

[29] Hill, M. O., 1974. Correspondence analysis: a neglected multivariate method. Appl. Statist. 23: 340–354.

[30] Hotelling, H., 1933. Analysis of a Complex of Statistical Variables Into Principal Components, Journal of Educational Psychology, volume 24, pages 417–441 and 498–520.

[31] Kao W. J., Sapatnekar S., Hiltner A., Anderson J. M.: Complement mediated leukocyte adhesion on poly(etherurethane-ureas) under shear stress in vitro. J. Biomed. Mater. Res. 1996, 32 (1): 99–109.

[32] Kaiser H. F. (1960) The application of electronic computers to factor analysis. Educational and Psychological Measurement, 20, 141–151.

[33] Krzanowski W. J., (1988). Principles of Multivariate Analysis, A User's Perspective. Oxford Univ. Press: Clarendon.

[34] Krishnaiah P.R. (ed.) 1977. Multivariate Analysis II. North Holland Vol 2. p. 595.

[35] Krzanowski W. J. (1995). Recent advances in descriptive multivariate analysis. Oxford University Press, New York 1995.

[36] Lane D. A., Bowry S. K. The scientific basis for selection of measures of thrombogenicity. Nephrol. Dial. Transplant. 1994, 9: 18–28.

[37] Larose D. T. 2005. Discovering Knowledge In Data. An Introduction to Data Mining. Wiley and Sons.

[38] Lim F., Yang C. Z., Cooper S. L.: Synthesis, characterization and ex vivo evaluation of polydimethylsiloxane polyurea urethanes. Biomaterials 1994, 15 (6): 408–416.

[39] Mardia K. V., Kent J. T., Bobby J. M. (1979). Multivariate Analysis, Academic Press, London.

[40] Rao C. R. (1973). Linear Statistical Inference and Its Applications, Wiley, New York.