

## INTRODUCTION

The articles we would like to present are the result of the 24<sup>th</sup> Conference on Multivariate Statistical Analysis – MSA 2005 organized by the Chair of Statistical Methods of the University of Łódź. The conference took place on November 7–9, 2005, in Łódź. The conferences on Multivariate Statistical Analysis have been organized every year since 1981. They have focused on the latest achievements in the field of theory of multivariate statistical analysis and its applications. The 24<sup>th</sup> Conference MSA'05 has followed this tradition. Some of the papers contributed to conference are included in this volume.

Czesław Domański (*First statisticians of Łódź*) and Julian Kuciński (*Economic and sociological community of Łódź until the year 1945*) presented interesting studies concerned historical development and popularization of statistic, economy and sociology in Łódź since the turn of XIX and XX century.

The next papers are divided into two thematic groups:

I. Statistical Methods.

II. Applications.

The first part includes papers with methodological aspects of statistical analysis. Some of them are illustrated with examples.

Jacek Białek (*The method of risk in case of stochastic definition of Net Present Value*) considered a stochastic, general definition of Net Present Value. The Net Present Value (*NPV*) rule is a base of modern finance theory. Author proposed the method of measurement of risk in case of using the stochastic definition.

Bronisław Ceranka and Małgorzata Graczyk (*Optimum chemical balance weighing design with correlated errors based on bipartite and ternary design*) studied the estimation problem of individual weights of objects using the chemical balance weighing design under the restriction on the number times in which each object is weighed. It is assumed that the errors have the same variances and they are correlated equally.

Czesław Domański (*Attempt to assess multivariate normality tests*) presented an attempt to assess selected tests from the point of view of their properties as well as possibilities of their applications. The assumption of multivariate

normality is a basis of the classical multivariate statistical methodology and consequences of departures from these assumptions have not been investigated well so far.

Andrzej Dudek (*Internal cluster quality indexes for classification of symbolic data*) described main classification methods used for symbolic data (e.g. data in form of: single quantitative value, categorical value, interval, multivalued variable, multivalued variable with weights) and presented difficulties of measuring clustering quality for symbolic data.

Eugeniusz Gatnar (*Measures of diversity and the classification error in the multiple-model approach*) presented a comparison of the ability of selected diversity measures to predict the accuracy of classifier ensembles.

Tadeusz Gerstenkorn (*Limit property a compound of the generalized negative binomial and beta distributions*) obtained a new limit distribution as a result of compounding of the generalized negative binomial distribution with the generalized beta distribution, interesting also in some special cases.

Tomasz Jurkiewicz (*An influence of distance measures among sample units on efficiency of the modified synthetic estimator: Monte Carlo analysis*) presented Monte Carlo analysis of the efficiency of MES estimator using different distance measures between sample units. Chosen distance measure is one of the crucial factors in using modified synthetic estimator.

Grzegorz Kończak and Maria Czogała (*On multivariate test for stability of the population proportions*) investigated the problem of testing the hypothesis of the stability proportion for multiple attributes. In this paper a test is considered which makes it possible to verify the hypothesis that several proportions fulfill given assumptions simultaneously. The properties of the proposed test are compared with the results achieved in classical approaches.

Grzegorz Kończak and Janusz Wywiół (*On testing linearity of trend function*) considered testing the goodness of fit between a hypothetical trend function and its non-parametric variant. On the basis of one of numeric method some quantiles of the test statistic are evaluated in case when the hypothetical trend is linear one.

Jerzy Korzeniewski (*A proposal of new classification algorithm*) presented a new method of classifying points to a predetermined number of classes. The method is based on the use of the sample/window mean shift technique to obtain a synthetic insight into the data set structure.

Daniel Kosiorowski (*About robust estimators of average shape and variance of shape*) investigated whether in the statistical inference for shapes of economic systems it is useful to modify least squares methods (commonly used Procrustes approach) by applying data depth concept. In the paper theoretical considerations are illustrated with examples of multidimensional financial time series.

Andrzej Mantaj and Wiesław Wagner (*Ordering the spatial units by the non-standard method at various standardization transformations*) used for the purpose of ordering the spatial units the non-standard method in two variants: of the mean and of the median from the modules of transformed sample standardization values. The results of research on ordering were illustrated on the numerical material concerning the communes of the Podkarpackie Province for the year 2001.

Dariusz Parys (*Kruskal-Wallis test in multiple comparisons*) in his paper showed that the Kruskal-Wallis test can be transform to quadratic form among the Mann-Whitney or Kendal  $\tau$  au concordance measures between pairs of treatments. A multiple comparisons procedure based on patterns of transitive ordering among treatments is implement. Author also consider the circularity and non-transitive effects.

Dorota Pekasiewicz (*Sequential probability ratio test for mean based on pseudo-likelihood function*) proposed the application of pseudo-likelihood function instead of likelihood one in the statistic of sequential probability ratio test. Examples of application of the test based on the likelihood function ratio in selected kinds of distributions are presented together with the results of Monte Carlo analysis concerning properties of these tests.

Krystyna Pruska (*Estimation of variance of logistic regression predictors for small areas*) considered the logistic regression predictors for parameter of the Bernoulli distribution and an estimation of variance and MSE for these predictors for small areas. The results of simulation experiments conducted for analysis of properties of estimators of variance and mean square for the predictors are presented.

Joanna Trzęsiok (*The wavelet transform in regression*) presented the wavelet transform in nonparametric regression. The use of wavelets in statistical applications was pioneered by D. Donoho and I. Johnstone. Author discuss their methodology – wavelet shrinkage. The wavelet transform is compared with another nonparametric regression method – splines.

Michał Trzęsiok (*Clustering methods applied to reduce the training sample size in support vector machines*) presented the proposition of reducing the training sample size using the  $K$ -medoids clustering method in the case of very large data sets when it is not possible to train SVMs on the whole training set because of too high computational costs.

Wiesław Wagner (*Remarks on classical means for one and many samples*) presented a uniform optimizing criterion allowing one to determine a class of possible classical means for one and many samples, among them, arithmetic, harmonic and geometric means.

The second part is devoted to the applications of statistical methods.

Andrzej Czajkowski (*Forecasting income distributions of households in Poland on the basis of Markov chains*) presented an application of homogenous

Markov chains in the process of forecasting the income structure of six socio-economic groups of population in Poland for the years 2004, 2006 and 2008. Forecasts are based on results of individual household budgets surveys.

Felicjan Jaguś (*Macroeconomic risk of investment portfolios at the Warsaw Stock Exchange*) in his paper presents the attempt to examine selected macroeconomic risk factors of portfolio investments in Polish capital market as well as to build investment portfolios sensitive to particular risk profiles. The analysis was based on monthly macroeconomic return rates of assets from Warsaw Stock Exchange.

Alina Jędrzejczak (*Comparing income distributions – methods and their application to wage distributions in Poland*) presented the methods useful for ranking income distributions and their application to the analysis of wage distributions in Poland. As a theoretical distribution the Dagum type-I model has been used.

Joanna Kisielińska (*Application of discriminant analysis and neural networks to forecasting the financial standing of farms*) considered a linear discriminant function and neural network that could be applied for financial situation forecasting in Polish farms sector. The construction of discriminant models was based on set of financial indicators and the classification criterion was based on the private farm's income.

Tomasz Kozdraj (*Using artificial neural networks to predict stock prices*) considered the problem of using artificial neural networks to predict stock prices on the example of the Warsaw Stock Exchange. The paper presents the general framework of neural networks, their potential and limitations as well as problems faced by researcher meets while using neural networks in prediction process.

Anna Malarska (*Duration of unemployment in the area of the District Job Centre of Pabianice in the year 2005. Empirical analysis of changes in time by demo-social features*) studied the average duration of unemployment on the local labour market of the area of the District Job Centre of Pabianice. The characteristic is determined by means of individual information about each of the unemployed person and the moments of their registration and deregistration over the period from January 31 to September 30, 2005 with quarterly frequency.

Artur Mikulec (*Evaluation of investing efficiency of Open Pension Funds and their rating by means of the method of cluster analysis*) considered the activity of open pension funds so far. The paper includes rating of open pension funds from the point of view of their efficiency and conducted investment policy, analyzing, at the same time, calculated profitability ratios, rates of return and risk measures.

Andrzej Pawluczuk (*Knowledge indicators for Polish provinces*) used knowledge assessment methodology (KAM 2005) to calculate knowledge index

for Polish provinces, through making modification from macro level to mezo level. All variables are divided into five pillars: variables of economic regime, governance, innovation systems, education, and ICT (information and communication technology). All provinces received own knowledge index, that is the arithmetic average of normalized variables.

Anna Szymańska (*Selected premium estimation methods in automobile liability*) compared two premium estimation methods: the expected value method and the zero utility method. The paper also investigates whether premiums estimated according to the selected methods allow to design an optimal bonus-malus system. The investigation was carried out on real data from an insurance company in Łódź.

Czesław Domański  
Anna Witaszczyk