Use of statistical methods in sociology

Katarina Čobanović[1] - Valentina Sokolovska[2]

Data analysis in human sciences requires the use of statistical methods. Statistical techniques could provide explanations of how and why some statistical methods are used. The aim of the paper is oriented to the possibilities of using statistical methods both in education and in research in the field of sociology. Sociology and other human sciences, such as psychology, demography, economics and others, do require the use of statistical techniques.

This paper is focused on the analysis of data and the use of adequate statistical methods. Some statistical methods are more suitable for quantitative, some others for qualitative data. The paper is oriented to data analysis and the choice of an appropriate method. In sociology, like in other fields of science, it is necessary to deal with some simple and some more complex statistical procedures such as: descriptive statistics, inferential statistics, based on the sample survey, dealing with hypothesis testing, significance tests and analysis of variance, linear and multiple correlation and regression techniques, analysis of variance, linear and multiple correlation and regression techniques, analysis of frequency data using chi-square, nonparametric statistical procedures, multivariate methods etc.

The emphasis in the paper is placed on experience in using statistical methods in education and teaching of statistics and research work in sociology at the University of Novi Sad. The presented examples serve as an illustration of wide application of statistical methods.

Keywords: statistical methods in sociology, education, research work, types of data, analysis of data

[1] Faculty of Agriculture, Department of Agricultural Economics and Rural Sociology, University of Novi Sad, the Republic of Serbia, katcob@polj.ns.ac.yu
[2] Faculty of Philosophy, University of Novi Sad, the Republic of Serbia, valentinas@neobee.net

ACKNOWLEDGEMENT: The paper has been supported by the Ministry of Science of Republic of Serbia (Projects No. 149013D, No. 149007).

1. Introduction

"Statistics is a branch of scientific methodology. It deals with the collection, classification, description and interpretation of data obtained by conducting surveys and experiments. Its essential purpose is to describe and draw inferences about numerical properties of populations" (Ferguson 1966).

The experience in statistics application indicates that a single statistical method can be used in different research areas for dealing with different types of problems. In this respect, we can talk about the universality of a statistical method (Bethea et al.1985; Čobanović et al. 1997). For instance, an experiment conducted in agriculture or in laboratory conditions is an equivalent to a sociological survey (Blalock 1960). Various possibilities which application of a statistical method provides in certain research areas should be considered as an adequate way of dealing with a problem of the research. However, it should be noted that statistics is not a method by which it is possible to solve all the problems in a research. Likewise, statistics is not a mere gathering of data and their presenting (Blalock 1960). Statistical texts cannot be treated as literary texts, since statistical instruments are presented in a very condensed form (Blalock 1960).

When applying a statistical method, it is common to differentiate between quantitative and qualitative features and variables. Nominal and ordinal variables and data are usually considered as qualitative (attributive), while interval variables and ratio variables are considered as quantitative (Ferguson 1966, Krneta 1987). Also, it is common to apply nonparametric statistical methods on nominal and ordinal data, while parametric methods are used for the interval and ratio data (Ferguson 1966).

Statistics and statistical methods have highly significant application in sociology. Functions of statistics are numerous: the methods of descriptive statistics have an important application for describing natural phenomena; inferential statistics is used for inductive reasoning about unknown properties of a larger group using the known indicators of the causes; hypothesis testing most frequently refers to the results of one, two or more causes, on the basis of which it is possible to draw conclusions on the problem of the research, by accepting or refuting an initial hypothesis; regression and correlation analysis, in the most simple case, examines the influence and dependence between two or more variables. If the relationship of a greater number of variables is examined, it is multiple regression and correlation. For a detailed analysis of a multiple regression model it is necessary to introduce partial correlations and explain the relations between variables using partial correlation coefficient; application of chi-square distribution and chi-square test is important in cases with qualitative variables for which it is known or assumed that

are interrelated. Chi-square test is quite a common test based on determining the sum of the quotient of the square of the difference between the observed and expected frequencies and expected frequencies. Chi-square value has multiple applications, for instance, testing equality of distributions, independence test, contingency coefficient C, and so on (Hadživuković 1991, Hinton 2004); nonparametric tests for independent and dependent samples have also important application in research in sociology. These tests primarily refer to attributive variables (Hinton 2004). What should also be mentioned is the importance of nonparametric tests in application of the variance analysis method with a single factor, on the basis of ranked data. The most significant tests are Kruskal-Wallis test i Friedman test (Hadživuković 1991, Hinton 2004). Regarding nonparametric statistics, when examining the correlation and association between variables, it is important to mention Spearman rank correlation coefficient, contingency coefficient C, coefficient fi, which represents an indicator of association and is applied on discrete and dichotomous variables (Ferguson 1966, Hadživukovič 1991).

Application of quantitative methods in sociology by the World War II referred to descriptive statistics and simple methods (Raftery 2000). After the World War II, with the increase of the scope of data, more complex statistical methods started being used in sociology (Raftery 2000). For a great number of more recent statistical methods developed in 20th century it was sociology where they first found their application. **Thus, for example, correspondence analysis, a statistical technique useful in the analysis of categorical data, especially in sociological research, was emphasized in Greenacre's book (Greenacre 2007).**

The aim of this paper is to point to the importance of applying statistics in sociological research and statistics teaching. The paper presents the experience acquired while teaching statistics to sociology students at the University of Novi Sad. Moreover, the paper also presents some experience of applying statistical methods in sociological and demographic research**. Used statistical methods in the paper are results of the previous authors experiences. The authors wanted to describe the use of appropriate statistical methods in different research problems. There were used classical statistical tests in the explanation of Census data results. In the explanation of relationship between different variables, on the sample survey results base, there were used correspondence analysis and chi-square test. The authors wanted to outline the importance of use of statistical methods in social and similar sciences.**

2. Statistical methods in teaching

At the Department of Sociology of the Faculty of Philosophy in Novi Sad, the course in statistics is studied at undergraduate, graduate and doctoral studies. Within

the curriculum for the undergraduate academic studies of sociology, there are two compulsory one-semester courses in statistics on the first year of the studies.

The first course is entitled Descriptive Statistics, and its content is intended to introduce students to the problems of collecting and presenting quantitative data on the sociological issues; descriptive statistical measures; measures of central tendencies, and measures of variability; probability theory and sample method, as well as hypothesis testing. The objective of the course is mastering the basic statistical methodology used for planning quantitative research, for systematizing research results and reaching valid conclusions and decisions.

Within the second compulsory course entitled Statistical Methods, students are introduced to the basic methods of statistical analysis. Special emphasis is placed on the methods of nonparametric statistics: chi-square test, contingency coefficient, Wilcoxon rank test, Kruskal-Wallis, Friedman test, Spearman and Kendall rank correlation coefficient. Furthermore, the course encompasses simple linear regression and correlation, multiple regression and correlation analysis, time series analysis and variance analysis. The objective of the course is mastering statistical methods, including their adequate application and interpretation of obtained results.

During the academic undergraduate studies, students also participate in scientific-research projects of the Department of Sociology. Within the surveys conducted on the territory of the Autonomous Province of Vojvodina, students are engaged primarily to conduct surveys, where they gain fieldwork experience. Afterwards, they store the gathered data in software for statistical analysis and are present when preliminary research results are announced. Thereby, they can have access to a number of phases of research procedure, as well as gain insight into its complexity.

At graduate and doctoral academic studies in sociology students can opt for two optional courses in statistics. The course at graduate studies is entitled Applied Regression Analysis in Sociology, and it is intended to introduce students to the methods of applied regression analysis, as well as to point to the possibilities of its application in sociological analyses. The students also acquire knowledge on determining quantitative dependence of sociological categories, examining the validity of sociological theories hypothesis, as well as anticipating future movements of social phenomena on the basis of the assessed quantitative dependence.

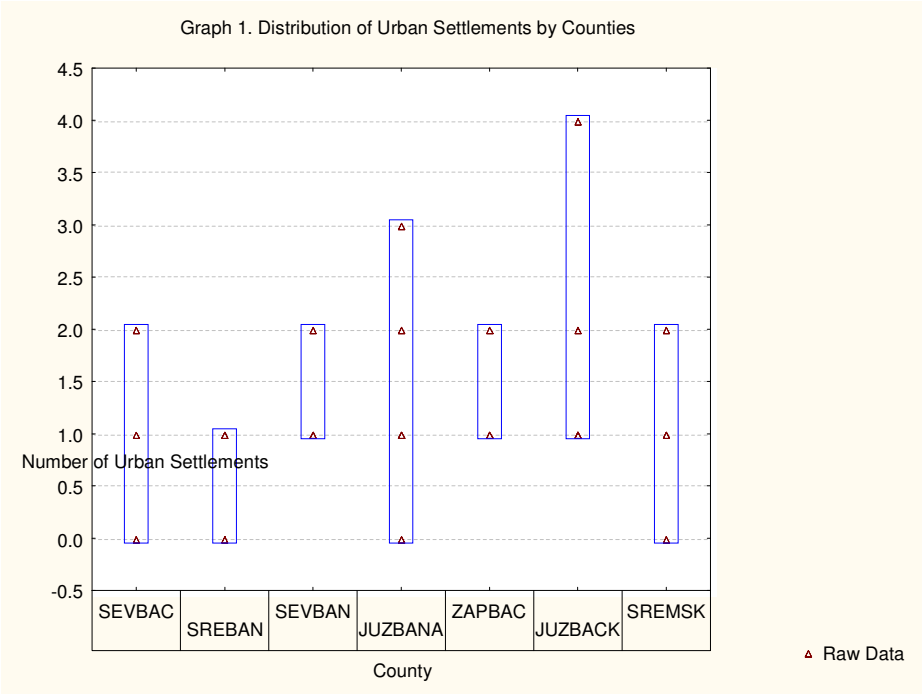
At doctoral studies of sociology, students can study multivariate techniques, also as optional subjects. The course encompasses canonical correlation analysis, multivariate analysis of variance, discrimination analysis, principal components analysis, factor analysis, and grouping and correspondence analysis. Studying and research work includes gathering of adequate empirical material for thesis and its statistical analysis; examining the literature and problem studies from multivariate techniques, using statistical software for data processing, discussions on applied techniques, their scientific scope and limitation in sociological problems analysis.

3. Statistical methods in research

The research on marital status of Vojvodinian population according to the census results from 2002 indicates that the changes of marital status were influenced by socio-political factors, economic factors and a number of cultural factors (Čobanović et al. 2008). Two types of families were analyzed: families with children and families without children at the level of the Autonomous Province, as well as at the level of Vojvodinian counties. Furthermore, the two examined types of families were analyzed also in relation to the types of their settlements: urban settlements and other (rural) settlements. In order to determine the existence of statistically significant differences between the two types of families and types of settlements, t-test was used for 2 independent samples. The number of families according to the type (with and without children) was expressed per capita. The graphical illustration was made using a "box-plot" diagram. In this example, a "box-plot" diagram, as a means of presenting data analysis, confirmed the results of the t-test, i.e. the existence of statistically significant differences between the two types of families and the types of settlements.

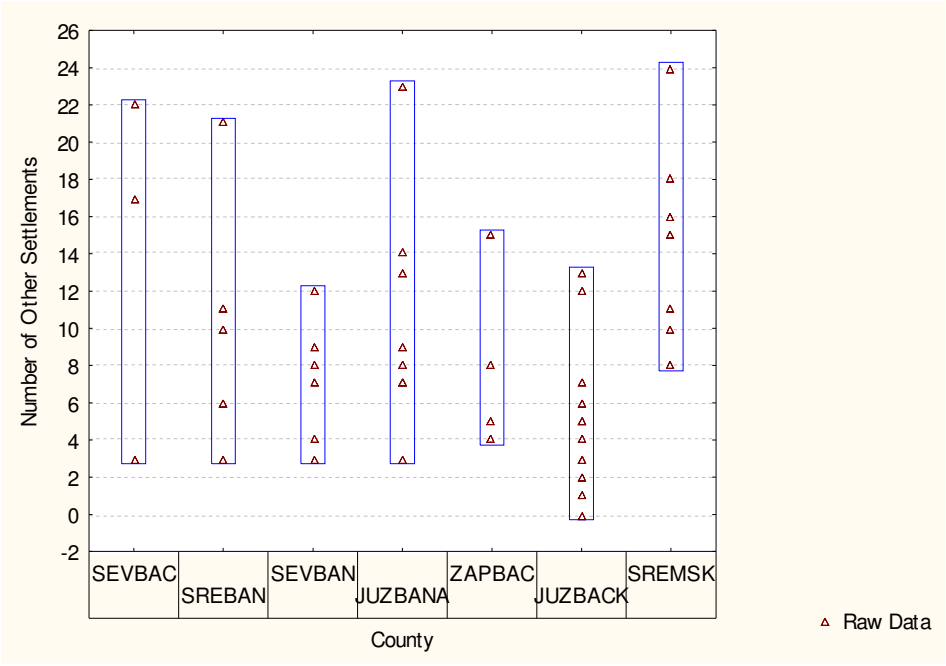
This research, based on the results of the population, households and dwellings census 2002, also included the analysis of the number of municipalities according to the type of settlements in the counties of Vojvodina. The research encompassed the analysis of municipality distribution in relation to the overall number of settlements, to the number of urban settlements and to the number of other settlements in 7 counties in Vojvodina. The municipality distribution (45 municipalities) in counties is graphically presented using a "variability plot" diagram, which in this case proved to be a very suitable way of presenting and analyzing of discontinued variables (the number of settlements) (Graph 1, Graph 2).

Graph 1. Distribution of Urban Settlements by Counties



Source: own creation

Graph 2. Distribution of Other Settlements by Counties



Source: own creation

Regression analysis (for 45 municipalities) in this research, based on the model of multiple linear regression, was used for examining the influence of certain variables on the number of families with children and the number of families without children at the level of urban and rural settlements. The results of regression analysis showed that the number of families with children from urban settlements is significantly determined by the overall number of families (with the positive correlation), the number of inhabitants of rural settlements (with the negative correlation) and the number of families without children of rural settlements (with the negative correlation). The regression model is as follows:

$$\hat{Y}_i = 45.31810 + 0.54817X_1 - 0.44021X_2 - 0.31989X_3$$

$$\left(R_A^2 = 0.999\right) \quad (t = 181.14) \quad (t = 12.011) \quad (t = -2.915)$$

The regression results of the number of families without children in rural settlements on the number of families with children in urban settlements (with the positive correlation), and the overall number of families (with the positive correlation), on the number of families without children in urban settlements (with the positive correlation) and the overall number of inhabitants in urban settlements (with the negative correlation). The regression model is as follows:

$$\hat{Y}_i = 52.79073 + 0.30519X_1 + 0.05356X_2 + 0.64293X_3 - 0.51155X_4$$
$$\left(R_A^2 = 0.984\right) \quad (t = 35.61) \quad (t = 0.26) \quad (t = -4.45) \quad (t = 3.74)$$

In this case, the following explanations can be provided: by the increase of the number of urban inhabitants (X_4) the number of families without children in rural settlements significantly decreases, which can be the result of movement of young population from rural to urban settlements, while in rural settlements there are only old people left (families without children). At the same time, the increase of the number of families without children in urban settlements (X_3) has positive effect on the increase of the families without children in other settlements. This relation can be explained by the assumption that the families without children, in both urban and rural settlements, are the result of the birth and death rate decrease among the population, as well as the process of senilization of Vojvodinian population, which according to a lot of indicators, belongs to the type of old population.

The results of the research, conducted in 2006 by the Department of Sociology of the Faculty of Philosophy in Novi Sad, on the sample of around 1,200 units of observation (respondents), and conducted on the territory of the Autonomous Province of Vojvodina, provided the material for the analysis of socio-economical and cultural aspects of multiculturality (Čobanović et al. 2007).

This research includes 21 selected variables. Some of them are defined as variables expressing grades of socio-economical situation, cultural views, ethnic aspects and the attitudes and opinions about many other features of multiculturalism in Vojvodina. Some variables refer to age, sex, national and confessional affiliation, education, occupation, activity sector, migration characteristics, county, household income, household size and the number of households with members without income. The variables are divided into 2 groups set by the author, keeping in view the assumptions on possible relations that can appear between the variables. In this way, a group of potential "independent" variables and the group of "dependent" variables were formed.

The first group of 6 dependant variables refers to the following phenomena: Evaluation of the family life quality (5 modalities); Evaluation of the most important family moments (7 modalities); Attitude towards joining EU and NATO (6 modalities), Evaluation of the culture preservation method (8 modalities); Attitude of the Serbian nationality members towards members of national minorities (7 modalities); Attitude of the national minority members towards national majority membership (7 modalities).

The second group consisting of 15 independent variables refers to the following phenomena: Sex (2 modalities); Age (6 modalities); National affiliation (11 modalities); Confessional affiliation (11 modalities); Education (6 modalities); Employment status (4 modalities); Occupation (28 modalities); Agency sector (8

modalities); Sojourn (2 modalities); Sojourn Character (2 modalities); County (7 modalities); Previous sojourn vacancy (2 modalities); Household monthly income (20 modalities); Household size (10 modalities); Number of household members without income (10 modalities).

In this research the majority of the analyzed variables are qualitative, given descriptively, with a larger number of modalities. In order to perceive the relations of certain combinations of variables the method of correspondence analysis was applied. Correspondence analysis describes the character of the relations between the occurrences and the structure of attributive (categorical) variables (STATISTICA 7.0, Introductory Overview). The correspondence analysis in this research was based on two-dimensional contingency tables and on determining the overall chi-square value. On the basis of chi-square value the statistical significance of the examined relations, i.e. the combinations of variables, was determined. The results of the correspondence analysis were generated by the statistical programme STATISTICA 7.0. This paper states some interesting results:

- The relation between the level of education and national identification was statistically significant with the chi-square value 132.64^{**}, $(df = 50; p = 0.0000)$;
- The relationship between the national status and employment status was statistically significant $(\chi^2 = 60.473^{**}; df = 30; p = 0.00081)$. It is worth outlining that the categories of employment (employed, unemployed, economically dependants and retired) were very heterogeneous;
- The relationship between the national status and age category and the relationship between the national category and education category were not statistically significant;
- The relation between the aspect of Serbian nationality according to national minorities and employment status was statistically significant $(\chi^2 = 29.645^{**}; df = 8; p = 0.04112)$;
- The contact of national minorities to majority (Serbian) population and the category of sex was not statistically significant $(\chi^2 = 8.0816; df = 6; p = 0.23220)$;
- The relation between the attitude in favour of joining EU and NATO and the age category was not statistically significant $(\chi^2 = 31.20379; df = 25; p = 0.18240)$;
- The relation between the attitude in favour of joining EU and NATO and other groups and the category of education level was statistically significant $(\chi^2 = 50.427^{**}; df = 25; p = 0.00019)$;
- The relation between the decision to join EU and NATO and nationality was statistically significant $(\chi^2 = 117.30^{**}; df = 50; p = 0.00000)$;

– However, the relation between the attitude in favour of joining EU and NATO and type of settlement was not statistically significant;

– The relation between the attitude in favour of joining EU and NATO and employment status was statistically significant $\left(\chi^2 = 26.0610^{*}; df = 15; p = 0.03744\right)$;

– The relation between the category of keeping culture of the nation and the age category was statistically significant, while the relation between the category of keeping culture of the nation and national affiliation was not significant;

– The relation between the category of keeping culture of the nation and the category of confessional affiliation was statistically significant $\left(\chi^2 = 231.80^{**}; df = 70; p = 0.00000\right)$;

– The relation between the category of keeping culture of the nation and the category education level was statistically significant $\left(\chi^2 = 62.001^{**}; df = 35; p = 0.0033\right)$;

– The relations between the categories of perceiving current lifestyle and educational attainment, age groups, employment status and nationality were not statistically significant.

In the doctoral thesis "Acultural Processes of Ethnic groups in Vojvodina" – defended in May 2009 at the Department of Sociology of the Faculty of Philosophy in Novi Sad – correspondence analysis, among others, was used for analyzing the above mentioned survey in 2006. One of the analyzed attitudes is the way the respondents perceive Vojvodina. For the question "How do you personally perceive Vojvodina?" the respondents of Serbian nationality opt mostly for the reply that it is a province of Serbian people and national minorities; the Montenegrins perceive it equally as a province of Serbian people and national minorities, and as a province of its citizens living on its territory; the Croats, Hungarians and Romanians perceive it most as a province of its citizens living on its territory, while the Slovaks see it as a province of different nations and ethnic groups (Graph 3). According to the Tables 1 and 2, which present the results of the correspondence analysis, it can be observed that almost 90% of the responses to this question are assigned to the modality 1, which replaces the reply "as a province of Serbian people".

Table 1. The table of correspondence analysis for How do the respondents perceive Vojvodina regarding the national (ethnic) affiliation of the respondents (survey of 2006)

How do you personally perceive Vojvodina	national (ethnic) affiliation of the respondents						
	Serbian	Montenigrin	Croat	Hungarian	Slovak	Romanian	Active Margin
1	46	2	1	2	1	1	53
2	349	10	4	30	6	5	404
3	135	2	11	56	15	3	222
4	248	10	19	82	13	9	381
5	5	0	0	1	1	0	7
Active Margin	783	24	35	171	36	18	1067

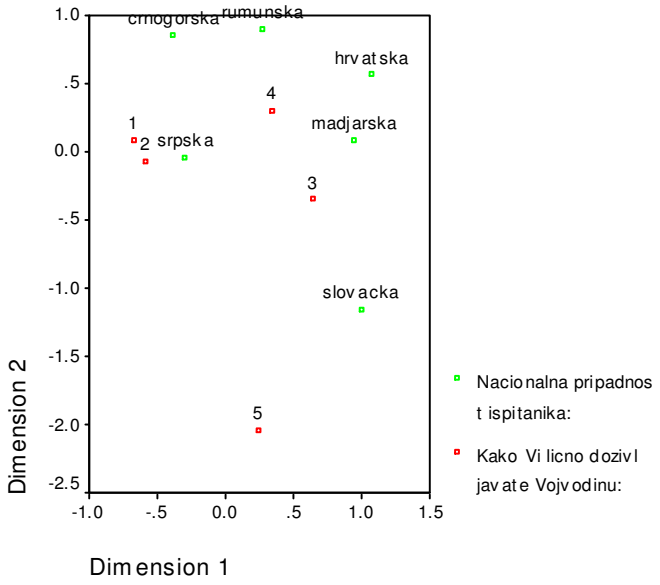
Source: own creation

Table 2. The overview of correspondence analysis for How do the respondents perceive Vojvodina regarding the national (ethnic) affiliation of the respondents (survey of 2006)

Dimen-sion	Singular Value	Inertia	Chi Square	Sig. (20 degrees of freedom)	Proportion of Intertia		Confidence Singular Value	
					Accoun-ted for:	Cumu-lative	Standard Deviation	Corre-lation 2
1	0.280	0.078			0.898	0.898	0.027	0.030
2	0.088	0.008			0.089	0.987	0.036	
3	0.032	0.001			0.012	0.999		
4	0.010	0.000			0.001	1.000		
Total		0.087	93.016	0.000	1.000	1.000		

Source: own creation

Graph 3. How do the respondents perceive Vojvodina regarding the national (ethnic) affiliation of the respondents (survey of 2006)



Source: own creation

The variable "How do you personally perceive Vojvodina?" has the following modalities: 1) As a province of Serbian people. 2) As a province of Serbian people and national minorities, 3) As a province of different peoples and ethnic groups, 4) As a province of its citizens living on its territory, and 5) I do not know.

4. Conclusion

The paper states the importance of statistical methods in sociology. Due to the significance of statistical data analysis it is possible to apply adequate statistical methods, on which the value of acquired results and conclusions is based. The paper presents the experience of applying statistical method in teaching sociology students. There are also some author's results, related to applying statistical methods in the field of sociological and demographic research. Thus, the marital status of Vojvodinian population was examined on the basis of the data from the population, households and dwellings Census 2002 in Serbia. The paper also presents the results of correspondence analysis, referring to the relations of the most significant socio-economic, cultural, ethnic and other aspects of living on the territory of Vojvodina. These results are based on the sample of around 1200 respondents on the territory of

Vojvodina in 2006. In sociology, like in many other disciplines, the use of statistical methods in practise was justified, because of its great importance in explaining the character of data, in explaining the behaviour and relations of investigated phenomena (social phenomena, demographic phenomena and so on) and in making conclusions. Both parametric and nonparametric statistical methods are suitable for use in sociology because of the nature of data. In sociology are common both quantitative and qualitative data.

References

Bethea, R. M.-Duran, B. S.-Boullion, T. L. 1985: *Statistical Methods for Engineers and Scientists*. Marcel Dekker, Inc., New York.

Blalock, H. M. 1960: *Social Statistics.* McGrow Hill Book Company, New York.

Ferguson, G. A.1966: Statistical Analysis in Psychology and Education, 2nd Edition, McGraw-Hill, London.

Greenacre, M. 2007: Correspondence Analysis in Practice, Second edition, Chapman & Hall, Boca **Raton**, FL, USA.

Hadživuković, S. 1991: Statistički metodi s primenom u poljoprivrednim i biološkim istraživanjima, Drugo prošireno izdanje, Poljoprivredni fakultet, Novi Sad.

Hinton, R. R. 2004: Statistics Explained, 2nd Edition, Routledge, New York.

Krneta, M. 1987: Statistika za sociologe, Ekonomika, Ekonomski institut

Raftery, A. E. 2000: Statistics in Sociology, 1950-2000. *Journal of the American Statistical Association*. 95, 450; ABI/INFORM Glober, pp 654-661

Sokolovska, V. 2009: *Acultural Processes of Ethnic groups in Vojvodina*., Doctorial Thesis, Faculty of Phylosophy , University of Novi Sad: (unpublishied material).

STATISTICA 7.0 2007: StatSoft, University Licence, University of Novi Sad

Čobanović K, -Nikolić-Đorić, E.-Mutavdžić, B. 1997: The Role and the Importance of the Application of Statistical Methods in Agricultural Investigations. *Agrieconomica*. Vol. 26. Pp. 457-470.

Čobanović, K.-Sokolovska, V.-Nikolić-Đorić, E. 2007: Statistical Analysis of Multiculturalism Research in Vojvodina, Applied Statistics 2007, Program and Abstracts, Ribno (Bled), Slovenia, p.40

Čobanović, K.-Sokolovska, V.-Nićin, S. 2008: Statistical Analysis of Marital Status of the Population of Vojvodina, Applied Statistics 2008, Program and Abstracts, Ribno (Bled), Slovenia, p. 48