

## EVALUATION OF DISPARITIES IN LIVING STANDARDS OF REGIONS OF THE CZECH REPUBLIC

H. Vostrá Vydrová, Z. Novotná

**Received: January 16, 2012**

### Abstract

VOSTRÁ VYDROVÁ H., NOVOTNÁ Z.: *Evaluation of disparities in living standards of regions of the Czech Republic*. Acta univ. agric. et silvic. Mendel. Brun., 2012, LX, No. 4, pp. 407–414

This paper focuses on regional differences between the regions of the Czech Republic. We will focus on observation of inequalities between indicators of living in different regions of the Czech Republic. The indicators are evaluated at NUTS 3 (regions), using multivariate statistical techniques - factor analysis and cluster analysis. We have identified the twelve indicators of living standards. Base data was reduced using factor analysis on the three emerging factors: 1) basic characteristics, 2) risk groups, 3) environmental variable. Cluster analysis was compiled groups of regions with similar characteristics. Cluster analysis of the breakdown of the county into three clusters based on selected indicators of living standards. They can be described as a group with higher average and lower standard of living. In the first cluster are only two regions (Liberec Region and Karlovy Vary), the third cluster is composed of Prague and the second cluster includes all other regions of the Czech Republic. To verify the evidence of differences between clusters were calculated by multivariate analysis of variance for the various indicators of living standards. An analysis of variance indicates that significant differences between clusters are caused by the standard of living indicators: GDP (regional), the average wage of women, medical equipment, culture entertainment and recreation, higher education, the disabled handicapped and older people. The data were processed in the program STATISTICA 10th.

regions of the Czech Republic, living standards, disparity, multivariate statistical methods, factor analysis, cluster analysis, analysis of variance

Living standards are characterized as the sum of utility values of material, cultural, social and moral, which has a population in a given time and space to meet the necessities of life available and the conditions under which satisfy these needs and create a way of life. To measure living standards can be accessed two different ways. The first one consists in objectively quantify directly consumed goods and services, or financial income and property, leisure, money spent from the budget for public services, education, but also the amount of harmful substances discharged into the air or water, life expectancy, levels of crime, etc. The second way of expressing the standard of living based on the idea that the standard of living you can imagine as a measure of satisfaction of material and immaterial

needs of an individual or household goods and services, rather as a relation between the actual state and between what is perceived as a desirable state or at least satisfactory. The aim of this paper is to examine indicators of living standards and find in them such indicators, which will have significant explanatory power. Or find so many indicators that remain after multivariate statistical methods, with the same predicative ability. There will also be regions of the Czech Republic clustering, the clusters that have similar characteristics. These newly formed clusters of regions of the Czech Republic would have to divide the region according to the quality standard of living of people who live here. Regions of the Czech Republic are uneven in size. Therefore, the interpretation of some between

the regions differences must take into account this fact. Even the capital city of Prague has a very specific position and therefore the interpretation of results obtained in this investigation must be substantially unique. The standard of living in each region records a large extent, the current economy of the region, but also the degree to which unemployment and related labor market is characterized by demand and supply of labor. For this article were elected the following indicators: regional gross domestic product, the average wage in the regions, the number of women per region, the general unemployment rate, infrastructure (number of km of roads in the 1st, 2nd, and 3rd class + highway), arable land (hectares per region), the index of ecological stability, medical equipment (number of hospitals per region), culture, entertainment and recreation, university educated (public and private universities), disabled - handicapped (1st, 2nd and 3rd degree), number of seniors. Data were selected on the basis of the indicators monitored in the agendas of the European Union. For evaluation of this paper were chosen method of factor analysis and cluster analysis. The same procedure is given in the work Žižka *et al.* (Žižka *et al.*, 2009). Navickas and Malakauskaite in their study reported that the most suitable tool for mapping of regional disparities is cluster analysis, which is the main tool for simulating the growth of regional economies (Navickas and Malakauskaite, 2009).

## MATERIALS AND METHODS

### Univariate analysis of selected indicators

First, descriptive statistics were calculated with different input variables. Input variables were selected on the basis of the indicators monitored in the agendas of the European Union to evaluate disparities between regions. These indicators are freely available from the regional database of the Czech Statistical Office and the database of STEP 2010, which is also provided by the Czech Statistical Office. STEP Database 2010 contains more than 500 indicators to evaluate the economic, social, environmental, agricultural situation in the various regions of the Czech Republic. This database is not a public website, but the Czech Statistical Office it provides for a fee. It has been working with 12 indicators. It was working with only a small set of data, since we have only 14 regions of the Czech Republic and to each county, only 12 indicators. The indicators were be standardized of size the region or number of people in the region.

### Methods of multivariate analysis of selected indicators

#### Factor analysis (factor analysis, FA)

This analysis is used to reduce the scope of the original characters from the statistical file based on reasonable assumptions formulated. The analysis is examining the structure and a relationship of the variables to determine whether permits their division into groups, in which the observed

I: Selected indicators for determining the standard of living by region for 2010

GDP	Regional gross domestic product ( million CZK)
Average wage	The average wage of the average gross monthly wage of an employee – an individual for the reporting period – year 2010 ( CZK).
Number of women	The number of women status of the population – women, 31th December reporting year.
General unemployment rate	The ratio of unemployed in the total labor force (in percent), where the numerator and denominator are indicators constructed according to international definitions and recommendations applied in LFS. The indicator is constructed according to Eurostat methodology developed based on recommendations of the International Labour Organisation (ILO).
Infrastructure	Roads 1st, 2nd, 3rd class and motorway (km).
Arable land	The number of hectares for each region.
Environment (coefficient of ecological stability)	The coefficient of ecological stability index number and determines the ratio of the areas of stable and unstable landscape elements in the surveyed area.
Medical equipment	The number of hospitals in the county.
Culture, entertainment, recreation	The number of units of cultural, entertainment and recreation.
University education (public and private universities)	The number of university - educated population by international classification of education (ISCED – 5.6) and the Czech educational system (in thousands).
Disabled – handicapped (1st, 2nd and 3rd degree)	The population of the loss or reduction of work ability (1st, 2nd and 3rd degree).
Number of seniors	The number of pension recipients.

Source: Database of Step 2010 from the Czech Statistical Office

variables from the same groups together more than correlated variables from different groups. With the main aim of understanding the analyzed data, and create new non correlation factors. Factor analysis allows to analyze the interdependencies of variables under consideration and then determine how much importance has watched the character with such properties are joined and how many characters have an impact on their own analysis.

Cluster analysis (cluster analysis, CLU)

The aim of cluster analysis is to find such a grouping of data that each cluster of objects with each other, most similar, but also to other clusters themselves, as many different (Melon et al., 2005) Hierarchical clustering starts with a number of clusters, where each observation is independent noise and ends with one cluster, which in turn contains all the observations. At each step the two closest observations or clusters then combined into a whole new cluster. Among the most commonly used methods include hierarchical clustering method of Ward. To identify observed vectors, which are similar, and their subsequent clustering techniques are used several measures of similarity. The most common functions for the distance between two vectors  $x$  and  $y$  is the Euclidean distance.

To verify the difference between clusters of evidence is appropriate to use methods that reveal these differences. Is seems appropriate to use analysis of variance.

To calculate all the statistical methods used statistical software STATISTICA 10th.

## RESULTS

The following table (Tab. II: Basic characteristics of selected variables) are the basic characteristics of selected indicators. These are the characteristics of the location and variability, which give basic

information about the examined data set in 2010. Calculated on the basis of descriptive characteristics for the 12 selected indicators of evaluation of living in different regions of the Czech Republic, it was stated that: the average regional GDP is 249 034.30 million CZK, the average wage is 22 039.90 CZK, average number of women in the region is 383 140.80, the general unemployment rate is 6.80%, the average number of miles of roads in the region is 3 982.40, arable land accounts for an average of county 214 863.60 hectares, the average coefficient of ecological stability is 1.20%, the average number of hospitals in the region 13.50, the average number of units per culture, entertainment and recreation is 4 304.60, the average number of university – educated population is 88.30 thousand, the average number of handicapped people are 32 225.20 and the average number of seniors 117 531.10 is see Tab. II.

Variability of selected indicators is very high, it is a consequence of the considerable diversity of regions. The City of Prague has a very specific position and the calculated results of individual indicators are significantly higher than the remaining regions CR. The lowest variability was calculated for the average wage, which amounts to 10.77%. The highest values of dispersion of regional GDP (92.50%), units providing culture, entertainment and recreation (61.28%) and number of hectares of arable land per region (65.20%) see Tab. II.

If the eigenvalues of newly created variables (factors) greater than 1, it is possible to consider their further classification (according to Kaiser's rule). The first newly created variable shall be numbered 6.80 and explains 56.68% of variability of the original variables. The second newly created variable shall be numbered 2.28 and explains 19.02%. The third newly created variable shall be numbered 1.28 and explains 10.65%. These three newly created variables

II: Basic characteristics of selected variables for 2010

Variables (indicators)	Average	Median	Minimum	Maximum	Coefficient of variation
GDP	249 034.30	170 481.00	139.52	946 630.00	92.50
Average wage	22 039.90	21 314.00	20 179.00	29 744.00	10.77
Number of women	383 140.80	313 101.50	156 444.00	645 086.00	44.72
General unemployment rate	6.80	6.60	4.50	9.50	23.52
Infrastructure	3 982.40	3 687.50	84.00	9 637.00	56.82
Arable land	214 863.60	194 997.00	14 857.00	551 096.00	65.20
Environment	1.20	1.00	0.30	2.30	45.11
Medicinal device	13.50	10.00	5.00	27.00	54.34
Culture, entertainment, recreation	4 304.60	3 303.00	1 749.00	11 857.00	61.28
University Education	88.30	58.30	20.10	317.40	87.60
Disabled (handicapped)	32 225.20	26 986.00	10 821.00	55 558.00	42.27
Number of seniors	117 531.10	95 808.00	47 189.00	198 350.00	43.36

Source: own processing outputs

III: *Eigen values and percentages of the total scattering factor for newly formed in 2010*

Order	2010		
	Eigenvalues	% of total variance	Cumulative %
1.	6.80	56.68	56.68
2.	2.28	19.02	75.70
3.	1.28	10.65	86.35

Source: own processing outputs

generally explain 86.35% variability of the original variables (see Tab. III).

The newly created factors for 2010:

First factor for 2010 is called the basic characteristics. Chief factor most correlated with regional GDP, average wages, medical facilities, culture, entertainment and recreation and higher education.

Second factor for 2010 is called vulnerable populations. The second factor most correlated with the number of women, the disabled – handicapped and the number of seniors.

Third factor for 2010 is called the environmental variable. The third factor most correlated with the infrastructure, arable land and the environment and also has a strong indirect correlation with the general unemployment rate.

Commented load factor was 0.6 and higher, depending on the newly created variables related to the observed 2010. Factor loadings are presented in Tab. IV.

Variables were first calculated using different methods rotated factor analysis solution. Were also examined based on the value of the Akaike information criteria. Selected solutions were rotated using the Varimax method, which minimizes the number of variables. Or a solution during which all variables between high load with each common factor. About this method of calculation can be said as simplistic method factors. Then we

applied cluster analysis on the data. Specifically used hierarchical cluster analysis and Ward's method for determining the similarity between clusters of Euclidean distance. Graphical output of the cluster analysis vertical hierarchical tree plot (Figure 1: Graph distance connections based on Euclidean distances and 2: Vertical hierarchical tree plot regions of the Czech Republic based on indicators of living standards (Ward method, Euclidean distance). Based on this cluster analysis were assembled group of counties with similar characteristics, which were represented in this article, indicators of living standards. Based on cluster analysis divided the county into three main clusters. The first and largest cluster is a group of counties that connects from an economic point of view: low average wages (compared to other clusters), the low number of university-educated population, very low regional GDP (average over the entire the Czech Republic), low number of seniors, a significantly lower number of hospitals fall the counties in this cluster but also lower the number of handicapped people. This cluster can be named as a cluster with a lower standard of living. These are the following counties: Liberec, Karlovy Vary, Zlin Region, Olomouc Region, Vysočina Region, Pardubice, Hradec Kralove region, Pilsen Region and South Region. The second cluster consists of only four regions: Ústecký Region, Central Region, the South Moravian Region and Moravian-Silesian Region. These counties have significantly higher values than the average for the CR. In this cluster are significantly higher wages than in the first cluster, but also here we find a significantly higher overall unemployment rates, higher regional GDP and a larger number of hospitals added to the region of the cluster. This cluster can be termed as a cluster with an average standard of living. The last cluster is the capital city of Prague. This region is markedly different from all other regions of the Czech Republic. Prague is home to a very high

IV: *Factor loadings newly emerging factors for 2010*

Variables	2010		
	Factor 1	Factor 2	Factor 3
GDP	<b>0.951</b>	0.139	0.136
Average wage	<b>0.979</b>	-0.075	0.055
Number of women	0.407	<b>0.868</b>	0.188
General unemployment rate	0.384	0.261	<b>-0.681</b>
Infrastructure	0.481	0.286	<b>0.601</b>
Arable land	0.338	0.513	<b>0.709</b>
Enviroment	0.408	0.333	<b>0.608</b>
Medicial device	<b>0.672</b>	0.484	0.364
Culture, entertainment, recreation	<b>0.938</b>	0.237	0.187
University education	<b>0.942</b>	0.198	0.046
Disabled - handicapped	-0.338	<b>0.920</b>	0.005
Number of seniors	0.423	<b>0.857</b>	0.154

Source: own processing outputs

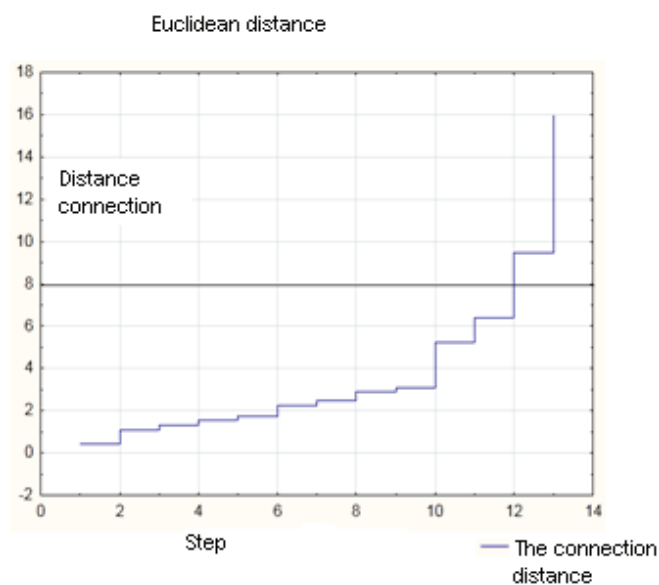
education, above-average wages, a favorable general unemployment rate and a very small number of kilometers of roads and highways and of course the small number of hectares of arable land. Therefore, this cluster can be named as a cluster with a higher standard of living.

Based on the results may be considered to exclude capital (Prague) from the analysis due to significant differences from other regions.

These results were achieved due to the nature of the underlying data. Underlying data (selected indicators) are calculated in absolute terms rather than in shares. Values of the indicators are recalculated according to the size of the region.

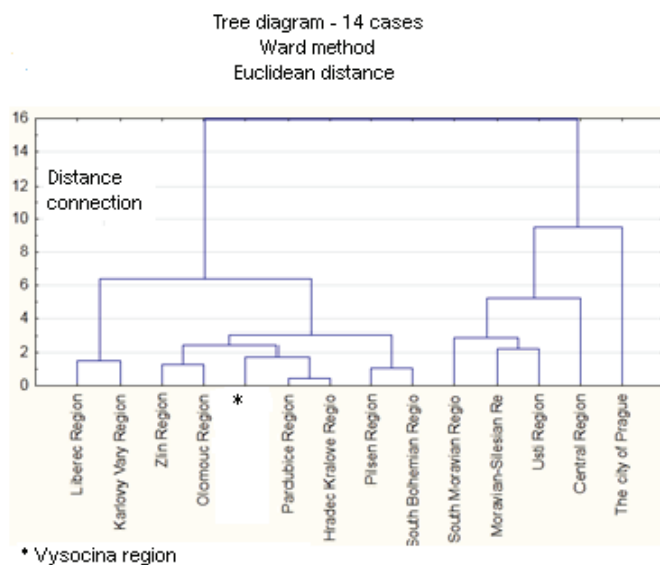
Therefore, the results of cluster analysis show considerably different view of the region than is generally known.

From this analysis of variance indicates that significant differences between clusters are caused by the standard of living indicators: GDP (regional), the average wage of women, medical facilities, culture, entertainment and recreation, university education, the disabled and older people. The conclusions are based on p-values, which were compared with the level of significance ( $\alpha = 0.05$ ). The results are summarized in Tab. V.



1: Graph distance connections based on Euclidean distance

Source: own processing outputs



2: Vertical hierarchical tree plot regions of the Czech Republic based on indicators of living standards (Ward method, Euclidean distance)

Source: own processing outputs





3: Map of the Czech Republic, divided according to the resulting clusters

Source: Own processing of the Czech Republic maps

(Explanation to the map of the Czech Republic: Prague, South Moravia, Pilsen Region, Karlovy Vary Region, Liberec Region, Olomouc Region, Hradec Kralove region, South Bohemia, Pardubice Region, Vysočina Region, Moravian-Silesian Region, Zlín Region, Central Region, Ústí Region).

V: Results of analysis of variance, evidence of differences between clusters

Found significant differences between clusters		Undetected significant differences between clusters	
Variables	p-value	Variables	p-value
HDP	0.000000	General unemployment rate	0.106826
Average wage	0.000001	Infrastructure	0.077840
Number of women	0.000022	Arable land	0.109273
Medical devices	0.000001	Environment (coefficient of ecological stability)	0.073887
Culture, entertainment and recreation	0.000001		
University education	0.000003		
Handicapped	0.000043		
Number of seniors	0.000036		

Source: own processing outputs

## DISCUSSION

In this article we have tried to factor analysis using polished new factors that can reduce the number of indicators that can assess the standard of living. Using factor analysis were compiled three new indicators – basic characteristics, vulnerable groups and environmental variable. Reduce the number of variables from 12 to 3 is beneficial in that our survey. Very similar topic dealt with authors Odehnal and Michálek in 2009. The aim of the paper “Assessment of Competitiveness of European Union Regions Chosen” is a comprehensive evaluation of selected EU regions at NUTS level 2. They have been selected multivariate statistical methods – factor analysis and cluster analysis. A similar procedure to detect differences among the Regions was used in this article. The difference is only in use in the rotated factor analysis solution (varimax normalized). (Odehnal, Michálek, 2009)

Paper “Measuring Living Standards with Proxy Variables,” deals with the measurement standard of

living by proxy variables. According to author Mark R. Montgomery, Michele Gragnolati, Edmund Burke and Kathleen Paredes is a suitable proxy variables used in the examination of living. The authors tested these variables in demographic research in developing countries. “Proxy” variables used in regression models such as regression and correlation analysis, logistic regression or correlation time series. The aim of this paper was comprehensive evaluation indicators of living standards, so they were chosen for analysis of multivariate methods. (Montgomery *et al.*, 2000)

This paper “Measuring Living Standards Using Existing National Sets Data” deals with the general construction index measuring living standards by types of families (Hutten, 1991). This article is another possible way to evaluate the data obtained from the database STEP 2010 for the author of this article.

The paper “The Varimax criterion for analytical rotation in factor analysis” deals with the

determination of a suitable rotation criteria – in particular the use Varimax criterion (Kaiser, 1958). The authors of this article chose rotated solution based on the value Akaike information criteria.

## CONCLUSION

Analyses allowed to assess on the basis of multivariate statistical methods disparity in regions of the Czech Republic. This evaluation is complete, the analysis works with all the monitored variables and their mutual influences. Using multivariate statistical methods suitable for solving our set objectives. Calculated on the basis of descriptive characteristics for the 12 selected indicators of evaluation of living in different regions of the Czech Republic, it was stated that: the average regional GDP is 249 034.30 million CZK, the average wage is 22 039.90 CZK, general unemployment rate is 6.80%, the average coefficient of ecological stability is 1.20%, the average number of university-educated population is 88,30 thousand, the average number of handicapped people are 32 225.20 and the average number of seniors is 117 531.10. Variability of selected indicators is very high, it is a consequence of the considerable diversity of regions. The City of Prague has a very specific position and the calculated results of individual indicators are significantly higher than the remaining regions the Czech Republic. The highest values of dispersion of regional GDP (92.50%) and the number of hectares of arable land per region (65.20%). Furthermore, factor analysis was applied to data from the database STEP 2010 from the Czech Statistical Office. Rated twelve

selected indicators of living standards. Using factor analysis, the original data base reduced to three emerging factors: 1) basic characteristics, 2) risk groups, 3) environmental variable. Subsequently, we used cluster analysis. In this analysis, we chose our discretion based on hierarchical clustering and the method for determining the similarity between clusters of Euclidean distance, which gives the best results in this type of investigation. Based on this cluster analysis were assembled group of counties with similar characteristics, which were represented in this article, indicators of living standards. Cluster analysis of the breakdown of the county into three clusters according to indicators of living standards – a group of counties with low living standards, with an average standard of living and a higher standard of living. To verify the evidence of differences between clusters were calculated by univariate and multivariate analysis of variance for the various indicators of living standards. This analysis shows that significant differences between clusters are caused by the standard of living indicators: GDP (regional), the average wage of women, medical facilities, culture, entertainment and recreation, university education, the disabled – handicapped and number of seniors. These results were achieved due to the nature of the underlying data. Underlying data (selected indicators) are calculated in absolute terms rather than in shares. Values of the indicators are recalculated according to the size of the region. Therefore, the results of cluster analysis show considerably different view of the region than is generally known.

## SUMMARY

This paper focuses on regional differences between the regions of the Czech Republic. The observed inequality in the individual parameters for the standard of living of the population by region has a considerable influence on the overall assessment of a particular region. The indicators are evaluated at NUTS 3 (regions), using multivariate statistical methods - factor analysis and cluster analysis. Solution was chosen for the twelve indicators of the evaluation standard of living. First, the basic characteristics were calculated. Variability of selected indicators seemed very high. It is a consequence of the considerable diversity of regions. The City of Prague has a very specific position and the calculated results of individual indicators are significantly higher than the remaining regions of the Czech Republic. The highest values of dispersion of regional GDP (92.50%). After calculating the basic characteristics were calculated with multivariate statistical methods. First, factor analysis was used. Base data was reduced using factor analysis on the three emerging factors: 1) the economy and leisure time, 2) risk groups, 3) environmental variable. Counties were then examined based on the results of cluster analysis. Cluster analysis was compiled groups of regions with similar characteristics. Cluster analysis of the breakdown of the county into three clusters based on selected indicators of living standards. They can be described as a group with higher average and lower standard of living. To verify the evidence of differences between clusters were calculated by multivariate analysis of variance for the various indicators of living standards. From this analysis of variance indicates that significant differences between clusters are caused by the standard of living indicators: GDP (regional), the average wage of women, medical facilities, culture, entertainment and recreation, higher education, the disabled and older people. These results have been achieved due to the nature of the underlying data recalculation of data by size or region and by population for evaluation of regional disparities is a must. Regions of the Czech Republic are very different in size, population and some indicators of living standards of evaluation.

## Acknowledgement

This article was supported by grant IGA agency. This article is supported by the grant project ID No. 201111170006 - whose theme is the Development trends of selected indicators of living standards in the Czech Republic and grant project No. IGA PEF 20121068 - whose theme is the Analysis of factors influencing the differences between men and women in the labour market.

## REFERENCES

- BLAHUŠ, P., 1985: *Factor analysis and its generalization*. Prague: SNTL – Technical Literature Publishing.
- HEBÁK, P. et al., 2005: *Multivariate Statistical Methods* 3. Prague: Informatorium. ISBN 80-7444-049-4.
- HENDL, J., 2006: *Overview of statistical methods for processing data* 2. ed. Prague: Portal. 468–469 s. ISBN 80-7467-124-9.
- HUTTON, S., 1991: *Measuring Living Standards Using Existing National Data Sets*. Cambridge: University Press. DOI: 10.1017/S0047279400018730.
- KAISER, H. F., 1958: *The varimax criterion for analytic rotation in factor analysis*. Psychometric, Vol. 23, No. 3., 187–200. DOI: 10.1007/BF02889239.
- MELOUN, M., MILITKÝ, J., 2006: *Compendium of statistical data processing. Methods and solved problems*. Prague: Academia. ISBN 80-200-1008-4.
- MELOUN, M., MILITKÝ, J., 2004: *Statistical analysis of experimental data*. Prague: Academia. ISBN 80-200-1254-0.
- MELOUN, M., MILITKÝ, J., HILL, M., 2005: *Computer analysis of multivariate data in the examples*. Prague: Academia. ISBN 80-200-1335-0.
- MONTGOMERY, M., GRAGNOLAT, M., BURKE, K., PAREDES, E., 1999: *Measuring Living Standards with Proxy Variables*. Demography, Vol. 37, No. 2. May 2000. 155–174. Springer. New York.
- NAVICKAS, V., MALAKAUSKAITE, A., 2009: Stimulating Regional Economy Competitiveness through the Formation of Clusters. *Proceedings of the 9th International Conference Liberec Economic Forum* 2009, Technical University of Liberec, pp. 253–259.
- ODEHNAL, J., MICHÁLEK J., 2009: Assessment of Competitiveness of Chosen European Union Regions. *The economic journal*, Vol. 57, No. 2. 113–131. Bratislava. ISSN 0013-3035.
- SVATOŠOVÁ, L., BOHÁČKOVÁ, I., HRABÁNKOVÁ, M., 2005: *Regional Development from Position of Structural Politics*. České Budějovice: JČU. ISBN 80-7040-749-2.
- ŽIŽKA, M., VALENTOVÁ HOVORKOVÁ, V., RYDVALOVÁ, P., 2009: *Statistical comparison of economic situation evaluation in municipalities*. 2009, CULS: Prague, *MATHEMATICAL METHODS IN ECONOMICS*. 366–371, ISBN 978-80-213-1963-9.
- Step 2010 – database provided by the Czech Statistical Office.
- Additional Resources
- The coefficient of ecological stability [online 2011] [cit. 2011-10-15] <[http://storm.fsv.cvut.cz/on\\_line/tok1/stabilita%20vzorce.pdf](http://storm.fsv.cvut.cz/on_line/tok1/stabilita%20vzorce.pdf)>.
- Average wages [online 2011] [cit. 2011-10-07] <[http://www.czso.cz/xp/redakce.nsf/i/prumerne\\_mzdy\\_v\\_cr\\_a\\_krajich\\_v\\_roce\\_2010](http://www.czso.cz/xp/redakce.nsf/i/prumerne_mzdy_v_cr_a_krajich_v_roce_2010)>.

## Address

Ing. Hana Vostrá Vydrová, Ing. Zuzana Novotná, Katedra statistiky, Provozně ekonomická fakulta, Česká zemědělská univerzita v Praze, Kamýcká 129, 165 21 Praha 6 – Suchbátka, Česká republika, e-mail: [vydrova@pef.czu.cz](mailto:vydrova@pef.czu.cz), [novotnaz@pef.czu.cz](mailto:novotnaz@pef.czu.cz)