ACTA UNIVERSITATIS AGRICULTARAE ET SILVICULTARAE MENDELIANAE BRUNENSIS

Volume LX 16 Number 7, 2012

DIMENSIONALITY REDUCTION OF QUALITY OF LIFE INDICATORS

A. Jindrová, J. Poláčková

Received: July 20, 2012

Abstract


Selecting indicators for assessing the quality of life at the regional level is not unambiguous. Currently, there are no precisely defined indicators that would give comprehensive information about the quality of life on a local level. In this paper we focus on the determination (selection) of groups of indicators that can be interpreted, on the basis of studied literature, as factors characterizing the quality of life. Furthermore, on the application of methods to reduce the dimensionality of these indicators, from the source of the database CULS KROK, which provides statistics on the regional and districts level. To reduce the number of indicators and the subsequent creation of derived variables that capture the relationships between selected indicators multivariate statistical analysis methods, especially method of principal components and factor analysis were used. This paper also indicates the methodology grant project “Methodological Approaches to assess Subjective Aspects of the life quality in regions of the Czech Republic”.

quality of life, principal component analysis, dimensionality reduction, Economical indicators, Social and cultural indicators, Environmental indicators, SPSS

Theoretical definition of quality of life can be made from many perspectives. The definition of this complex concept, which goes beyond a single discipline, is not uniform. The quality of life can be defined in terms of psychology, sociology, economics, and also politics. Payne (2005) in his publication “Quality of Life and Health” by the definition of the quality of life according to international companies who see it as a “product of the cooperation of social, health, economic and environmental conditions affecting the development of people”. He also states that in case of a sociological approach to the quality of life the attributes that are emphasized and preferred for the success of social status as a person are property, household goods, education and marital status.

The overall objective of this paper is the establishment and subsequent selection of factors that are suitable for the identification of regional differences at the district level in the research of quality of life and its development.

Measuring of the quality of life is based on many different models and approaches. Accurate, reliable and theoretically satisfactory measurement of quality of life which is agreed by most experts does not exist. This is caused mainly due to the two relatively separated components by which the quality of life is made of. They are objective and subjective (Dvořáková, Dušková, Svobodová, 2006).

A very similar opinion has Vadurová and Mühlpachr (2005), she divides the view of the quality of life on the objective and subjective. The objective quality of life includes materialistic page, social conditions, social status and physical health. Subjective page is then characterized as one's perception of their position in society in the context of its culture and value system.

For evaluating the quality of life it is characteristic to import macro indicators to the regional level. The problem is to choose indicators on which it is not easy to agree on. Issues that often need to be addressed in the selection of indicators at the regional level include the availability of data in time and space or degree of homogeneity of the state. The selection of indicators is influenced by the thematic focus of research itself and the composition of the
research team. Exception is usually the national income, which is undoubtedly related to the quality of life (Charvát, Petr, 2009).

As stated by Potůček (2002) the regional quality of life cannot be measured by the same indicators as the international or national level, because at the regional level there are no data available (the number of indicators is not monitored). Also for the reason that some indicators with good explanatory ability are not important at the regional level and vice versa. The problems of regional development and regional quality of life measurements should therefore be modified with regard to access to available data, the size and character of the evaluated area.

For the application of regional policy, for the measurement of regional disparities, for the level of regional development and the quality of life, are usually evaluating identification and description of regions within the Regional Development Strategy there was developed “Description and comparative analysis of regions of the Czech Republic.” Basic descriptive entity is a descriptor that can be expressed by one or more indicators (identifiers). Individual descriptors are then sorted into 5 groups of relatives that monitor problem areas within the Regional Strategy for Development.

These are:
- The overall characteristics of the region
- The economic potential
- Human potential
- Technical equipment and service areas
- Environment (Kutscherauer, 2004).

When determining the indicator that would be sufficient to inform about the development of the region one must build on the skills of statistical analysis. These analysis need to include indicators that characterize the development of the region and further development of specific indicators for the region (Svatošová, Boháčková, Hrabánková, 2005).

The selection of the optimal subset of key indicators is based on a variety of statistical and mathematical procedures. The multivariate statistical methods, which are among the so-called exploratory methods are used for reducing the dimensionality of variables or for reducing expendable input variables and for avoiding the subsequent creation of new variables that capture the relationships between selected indicators which are used to evaluate the quality of life and its development. These methods have no predefined hypotheses on which basis there would be a decision on their acceptance or rejection. The wide application of multivariate methods show their work in Mederly, Topercer and Nováček (2004), they used these methods in the construction of summary indices used for measurement of quality of life and its development at the global, national and regional level.

The basic source of information for regional socio-economic analysis and definition of statistical indicators are statistics monitored by the Czech Statistical Office, which is the main carrier of the vast extent of the methodology and implementation of statistical services in the Czech Republic. The selection of indicators at the regional level was based on the CULS KROK regionally oriented database that provides statistics on the regional level (NUTS 3) and districts (NUTS 4). The data in this database have an annual periodicity, and are monitored since 1995. Despite the wide range of data in the selected database, one can say that for research at district level (CR 76 districts and the Capital city of Prague), this database is insufficient, especially in terms of lack of economic indicators.

MATERIALS AND METHODS

The basic method for data reduction is an expert selection. It is a method based on subjective opinions. The weights of individual indicators are assigned on the opinion of experts. Hrach and Mihola (2006) state that it can be applied only at a small number of indicators (not more than 10).

Also the correlation coefficients could be used for reducing the number of variables. In case the correlation coefficient has a value greater than 0.9, we can say that a multicollinearity exists in the model. Therefore one of these two variables should be removed from the model.

There are more sophisticated methods of data reduction. The principal component analysis (PCA) and factor analysis (FA) are both variable reduction techniques. These techniques are usually used when variables are highly correlated. The principal component analysis reduces the number of observed variables to a smaller number of principal components accounting for most of the variance of the observed variables. It minimizes the sum of the squared perpendicular distances to the axis of the principal component. By reducing a data set from a group of related variables into a smaller set of components, the PCA achieves parsimony by explaining the maximum amount of common variance using the smallest number of explanatory concepts (more in Field, 2005).

The principal component is a linear combination of weighted observed variables. All linear combinations are related to other variables or to the data structure. Principal components are uncorrelated and orthogonal and explain the maximum amount of variance of the original variables (see Hebáň et al., 2007; Meloun, Militky, 2006, or Rencher, 2002). The first principal component corresponds to the direction of maximum variance; the second principal component corresponds to the direction of maximizing the remaining variance, and so on. Each principal component corresponds to a certain amount of variance of the whole dataset.
The exploratory factor analysis (EFA) is a variable reduction technique, which identifies the number of latent constructs and the underlying factor structure of a set of variables. It decomposes an adjusted correlation matrix. The diagonals were adjusted for the unique factors. The amount of variance explained is the trace (sum of the diagonals) of the decomposed adjusted correlation matrix.

The factor analysis estimates factors, which influence responses on observed variables. The factors account for common variance in a data set. In the FA, observed variables are a linear combination of the underlying factors (estimated factor and a unique factor). Squared multiple correlations are used as communality estimated on the diagonals. Communalinity is the variance in observed variables accounted for by a common factor. Large communalities are strongly influenced by an underlying construct. If communalities are large, close to 1.00, the results of PCA and FA could be similar. For more information see Suhr (2005) or Meloun and Militký (2004).

The rotated PCA methods rotate the PCA eigenvectors, so they point closer to the local clusters of data points. There are several analytical choices of rotation that were proposed in the past. One of them is the varimax method of orthogonal rotation. The varimax rotation criterion maximizes the sum of the variances of the squared coefficients within each eigenvector, and the rotated axes remain orthogonal (Lin, Altman, 2004).

The IBM SPSS Statistics 18 software was used for preparation of a data matrix and for exploratory and multivariate analysis. In PCA, observed variables were standardized, e.g., mean = 0, standard deviation = 1, diagonals of the matrix were equal to 1.

RESULTS AND DISCUSSION

The initial selection and reduction of regional indicators of the quality of life was based on scientific studies “Regional aspects of the quality of life in the Czech Republic” elaborated by Mederly, Tóperčer and Nováček (2004) and “Ways of dealing with regional disparities between the selected regions” by authors Žíveľová, and Jánský (2007). Furthermore, the descriptors from Strategy of Regional Development were used.

The identification of the appropriate indicators for the analysis was affected by a number of variables that are collected by the Czech Statistical Office. District level (NUTS 4) was used as a default level for assessment of regional disparities. The indicators were divided into three main fields: Economical (19 indicators), Social and cultural (42 indicators), and Environmental (16 indicators). The dimensionality reduction of the most important indicators for each of the areas was based on advanced multivariate statistical methods such as factor analysis and its extraction method of principal component. The factor loadings were estimated using the principal component analysis. The principal component analysis including the varimax rotation (orthogonal) was computed in the next step. Rotation is a linear transformation of the solution to make the interpretation easier. With an orthogonal rotation, loadings are equivalent to correlations between the observed variables and the components.

The prior criterion was set to select the number of principal components that will explain the maximal amount of variance. The criteria are that the cumulative variance is at least 75%, and the eigenvalues are greater than one. An eigenvalue greater than one explains at least as much variance as the variance explained by one variable (the ones on the diagonals of the correction matrix).

**Economic variables reduction**

Economic field contained these indicators: density of the rail network; density of the road network; density of other transport areas; unemployment rate; proportion of unemployed under 25 years; average time of registration at the job centre; number of applicants for one job position; number of free job positions; proportion of people receiving unemployment benefits; proportion of people receiving unemployment benefits; proportion of economically active persons in the primary sector; proportion of economically active persons in the secondary sector; proportion of economically active persons in the tertiary sector; average age of applicants; proportion of free job positions for disadvantaged groups; proportion of free job positions for citizens with reduced working ability; proportion of free job positions for graduates and young people; number of people commuting to work longer than 60 minutes.

In principal component analysis, we seek to maximize the variance of a linear combination of the variables. The table 1 below shows the eigenvalues – the proportion of total variance in all the variables, which is accounted for by that factor. The ratio of eigenvalues is the ratio of explanatory importance of the factors with a respect to the variables. The factors with a low eigenvalue (<1.0) were ignored as redundant compared to more important factors. The principal component analysis reduced the 19 observed economic variables into a smaller number of principal components. The first five principal components with eigenvalues greater than one account for 78% of the total variance.

The interpretation of factors was based on the loadings of variables on factors. The first factor is related to the unemployment situation. It largely represents five observed variables, which are logically related. Positive values of loadings have variables: an unemployment rate, a proportion of unemployed under 25, an average time of registration at the job centre, and a number of applicants for one job position. An indicator with high negative loading is a proportion of economically active population employed in the total number of economically active.
The second factor accounting for 18% of the total variance appears to measure the growth of a region. It corresponds to the proportion of various types of communication. The proportion of people receiving unemployment benefits and the proportion of active people in the primary sector reached a high negative value of the loading factor. It means that the higher level of infrastructure is related to a lower proportion of people receiving the unemployment benefits and to a lower proportion of people working in the primary sector.

The third component explaining 11% of the total variance can be called the component of activity in the tertiary sectors. This component is positively correlated with the tertiary sector, and it is related to negative correlation with a proportion of people in the secondary sector.

The fourth component explaining 10% of the total variance is negatively correlated with a variable commuting to work for longer than 60 minutes. Furthermore, this component is positively correlated with a proportion of economically active population and an average age of applicants in the region. The suburban regions, especially around Prague reached low values of the factor scores. High values were reached primarily by cities.

The fifth component explaining 7% of the input information can be called a component of job positions for disadvantages groups. This component correlates with a proportion of free job positions for citizens with reduced working ability and for graduates and young people.

**Social and cultural variables reduction**

Social field contained these indicators: proportion of people living in municipalities with less than 999 inhabitants; proportion of people living in municipalities with 1,000–4,999 inhabitants; proportion of people living in municipalities with 5,000–19,999 inhabitants; proportion of people living in municipalities with 20,000–99,999 inhabitants; proportion of people living in municipalities with more than 100,000 inhabitants; average size of a municipality; birthrate; mortality; natural increase; migration increase; divorce rate; abortion index expressed per 100 births; average age of living persons; proportion of the population younger than 14 years; proportion of the population aged 15–59 years; proportion of the population aged 60–64 years; proportion of the population older than 65 years; old-age index; economic burden index; proportion of sports and recreational areas; proportion of doctors; proportion of dentists; proportion of specialists; proportion of pediatricians; proportion of general practitioners; proportion of medical staff; proportion of pharmacies; average duration of sick leave; average pension; proportion of pension recipients; proportion of recreation cottages; proportion of the individual recreation properties; proportion of the university graduates; proportion of the high school graduates; proportion of the population with a vocational certificate; proportion of the inhabitants with basic education; number of criminal offenses; number of traffic accidents; proportion of apartment households with one car; proportion of apartment households with

---

**I: The Economic variables: total variance explained**

<table>
<thead>
<tr>
<th>Component</th>
<th>Initial Eigenvalues</th>
<th>Rotation Sums of Squared Loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>% of Variance</td>
<td>Cumulative %</td>
</tr>
<tr>
<td>1</td>
<td>6.843</td>
<td>34.217</td>
</tr>
<tr>
<td>2</td>
<td>4.335</td>
<td>21.677</td>
</tr>
<tr>
<td>3</td>
<td>1.974</td>
<td>9.872</td>
</tr>
<tr>
<td>4</td>
<td>1.364</td>
<td>6.820</td>
</tr>
<tr>
<td>5</td>
<td>1.083</td>
<td>5.414</td>
</tr>
</tbody>
</table>

Source: own computation

---

**II: The Social and cultural variables: total variance explained**

<table>
<thead>
<tr>
<th>Component</th>
<th>Initial Eigenvalues</th>
<th>Rotation Sums of Squared Loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>% of Variance</td>
<td>Cumulative %</td>
</tr>
<tr>
<td>1</td>
<td>11.258</td>
<td>26.804</td>
</tr>
<tr>
<td>2</td>
<td>8.606</td>
<td>20.491</td>
</tr>
<tr>
<td>3</td>
<td>5.332</td>
<td>12.696</td>
</tr>
<tr>
<td>4</td>
<td>2.839</td>
<td>6.760</td>
</tr>
<tr>
<td>5</td>
<td>1.932</td>
<td>4.600</td>
</tr>
<tr>
<td>6</td>
<td>1.608</td>
<td>3.829</td>
</tr>
<tr>
<td>7</td>
<td>1.266</td>
<td>3.014</td>
</tr>
<tr>
<td>8</td>
<td>1.095</td>
<td>2.607</td>
</tr>
</tbody>
</table>

Source: own computation
a telephone; proportion of apartment households with a computer; proportion of apartment households with recreation possibilities.

Eight principal components explaining 81% of the total variance were created out of the total 42 input indicators. The first principal component (explaining 21% of input information) appears to measure the intelligence. The component is associated to the proportion of doctors, dentists, and other specialists and medical staff per 1,000 inhabitants. Furthermore, it is related to the proportion of university and high school graduates.

The second component (explaining 19% of total variance) is related to the natural increase. It is positively associated with the natural increase per 1,000 inhabitants and negatively with deaths. This component is also related to the proportion of inhabitants with basic education. Therefore a low natural increase can be expected in the regions with a high proportion of people with low education. It is probably due to the migration of young people to cities.

The third principal component explaining 11% of total variance reflects the economic burden index. The forth component is related to crimes. The fifth component corresponds to the proportion of seniors and to the number of recreation cottages per 1,000 inhabitants. It can be called a component of active ageing. Furthermore, it is possible to reflect components of incapacity of work and recreational areas.

**Environmental variables reduction**

Environmental field contained these indicators: proportion of agricultural land; proportion of non-agricultural land; arable land per capita; proportion of gardens and orchards; proportion of permanent grassland; proportion of forest land; proportion of water surface areas; proportion of green space; proportion of landfill sites; coefficient of ecological stability; proportion of permanently occupied houses with gas; proportion of permanently occupied houses with central rating; proportion of permanently occupied apartments with gas; proportion of permanently occupied apartments with central rating; proportion of permanently occupied apartments connected to sewerage.

Five linear combinations of input variables were created of the total 16 indicators by the rotated principal component analysis. These principal components explain 83% of the total variance. The first principal component explaining 32% of total variance appears to measure the ecological stability. It is positively related to the coefficient of ecological stability, the proportion of non-agricultural land and to the woodland proportion. On the other hand it is negatively associated with the proportion of agricultural and arable land.

The second principal component explaining 16% of total variance is related to the proportion of houses with natural gas and the proportion of gardens and orchards. The third principal component corresponds to the proportion of residential houses with central heating. The fourth factor measures flats connected to the public sewerage system. And the last factor is associated with green vegetation.

**CONCLUSION**

The aim of this paper was to identify the important factors that influence the dynamics of the community development and that have an impact on regional disparities at the district level, especially in terms of the quality of life in particular regions.

The principal component analysis reduced the 19 observed economic variables into five principal components accounting for 78% of the total variance. The most important indicators for the dimensionality reduction were the unemployment indicators. Generally speaking, the unemployment indicators are considered as negative factors affecting the labor market economy as well as the quality of life.

Eight principal components in the social and cultural field were created out of the total 42 input indicators. These components explain 81% of input information. The first one explaining 21% of input information appears to measure the intelligence. The level of education in the region closely corresponds to the number of physicians. Therefore in terms of assessing the quality of life it is possible to consider this component as the important one.

The last analysis reduced the environmental indicators. Five linear combinations of input variables were created out of the total 16 indicators by the rotated principal component analysis. These

<table>
<thead>
<tr>
<th>Component</th>
<th>Initial Eigenvalues</th>
<th>Rotation Sums of Squared Loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>% of Variance</td>
</tr>
<tr>
<td>1</td>
<td>5.742</td>
<td>33.776</td>
</tr>
<tr>
<td>2</td>
<td>4.091</td>
<td>24.066</td>
</tr>
<tr>
<td>3</td>
<td>1.662</td>
<td>9.774</td>
</tr>
<tr>
<td>4</td>
<td>1.524</td>
<td>8.967</td>
</tr>
<tr>
<td>5</td>
<td>1.047</td>
<td>6.161</td>
</tr>
</tbody>
</table>

Source: own computation
components explain 83% of the total variance. The first principal component explains 32% of the total variance and it corresponds to the proportion of non-agricultural and forest land. This component can be called the component of ecological stability.

The authors of this paper are fully aware that the establishment of comprehensive indicators of the quality of life and its development can be based only on numerical quantification of the monitored indicators. Also, a further research is needed to focus on determining indicators based on measurements of the attitudes of citizens in these regions. Public opinion polls are the most common ways of obtaining these kinds of data.

SUMMARY

The paper is focused on a selection of appropriate indicators for evaluating the quality of life at the regional level. The selection of indicators was based on the CULS KROK regionally oriented database. The indicators were monitored by the Czech Statistical Office in 2010. The initial selection and reduction of regional indicators of the quality of life was based on scientific studies. The indicators were divided into three main fields: Economic advancement of the region, Social and cultural spheres, and Environmental sphere.

In the analysis, the selected indicators were transformed into the components by the rotated principal component analysis. The aim of this analysis was to reduce the number of parameters while preserving the information contained in the input variables. Based on the results of the rotated PCA five major factors were found in the Economic field, eight factors in the Social and cultural sphere, and five in the Environmental field.

The final factors could be considered as the comprehensive measure of the quality of life. It can be used to describe the general regional differences and to identify the crucial areas for further analysis. These areas will be examined by the questionnaire survey to analyze the attitudes and opinions of citizens in the field of the quality of their life.

Acknowledgement

The authors gratefully acknowledge the support of the Faculty of Economic and Management, Czech University of Life Sciences, via IGA grant, No. 201111170010, “Methodological approaches to assess subjective aspects of the life quality in the regions of the Czech Republic” and GAČR P403/12/1950, “Design of Experiments for Product Development and Multi-Factor Optimization of Production”.

REFERENCES

SVATOŠOVÁ, L., BOHÁČKOVÁ, I. a HRABÁNKOVÁ, M., 2005: Regionální rozvoj z pozice struk-


Address
Ing. Andrea Jindrová, Ing. Julie Poláčková, Katedra statistiky, Provozně ekonomická fakulta, Česká zemědělská univerzita, Kamýcká 129, 165 21 Praha 6 – Suchdol, Česká republika, e-mail: jindrova@pfe.czu.cz, polackova@pfe.czu.cz