

Methodological starting points of regional development analyses

Metodologická východiska analýz regionálního rozvoje

L. SVATOŠOVÁ

Czech University of Agriculture, Prague, Czech Republic

Abstract: When assessing regional development, it is necessary to assess and analyze many indicators describing the regions as to the natural and geographical conditions, economic situation, human potential, social levels, environment and others. The data analyzed represent a multidimensional statistical population, where many variables are observed and relationships among these variables exist or may exist. The paper deals with the possibilities of the multivariate statistical data analysis application.

Key words: multicriterial assessment, available potential of the region, multivariate statistical methods, principal component analysis, cluster analysis, regression analysis

Abstrakt: Při hodnocení regionálního rozvoje je nutné sledovat a analyzovat řadu ukazatelů, které charakterizují regiony z hlediska přírodních a geografických podmínek, ekonomické situace, lidského potenciálu, sociální úrovně, životního prostředí a dalších. Analyzovaná data tak představují vícerozměrný statistický soubor, ve kterém sledujeme velké množství proměnných, mezi nimiž existují, respektive mohou existovat vzájemné vztahy. Příspěvek diskutuje možnosti využití metod vícerozměrné analýzy dat.

Klíčová slova: vícekriteriální hodnocení, disponibilní potenciál regionu, vícerozměrné statistické metody, analýza hlavních komponent, shluková analýza, regresní analýza

INTRODUCTION

Questions of investigation and evaluation of regional development belong at this time of the Czech Republic accession to the European Union to crucial economic questions. The sense and aim of regional approach is ensuring of the same chance and possibilities of regions development, so that their demographic, natural and economic potential was fully used. The market environment only is not able to ensure a balanced development of the whole state territory. If a steady development of all regions is to be ensured, it is necessary to admit certain measures on regional, state and supranational (EU) levels, which will lead to the reduction or removal of regional differences. To purposefully use the means, which a state or the EU will insert for the development of regions, and to fulfill their purpose, it is necessary to know the situation in region in detail; to found out resources and

reserves; to prove viability and perspectives of the given region. It is necessary to carry out a thorough evaluation of the present development of particular indicators with the aim to create such a system of indicators which would qualify the particular parts of the development potential of regions. This task requires a qualified statistical analysis.

METHODOLOGY

One of the possibilities is to judge selected indicators of regional development separately – to describe independently their development and to forecast future development. This is a good starting point, but for a quick and especially simple evaluation of development potential, it is not very suitable, because an orientation in many indicators is much complicated and a little transparent. Moreover, from analyses of the

particular indicators, there can hardly be presumed the mutual relations and connections. A more suitable form seems to be summary indicators (e.g. in form of coefficient) which include in themselves substantial and characteristic features of the particular parts of regional development. Such indicators will not be indeed suitable for requirements of comprehensibility and simplicity; however, it can be compensated by a thorough and comprehensive methodological description of their construction. Nevertheless, they will intercept the problems in their complexity and interrelations. The indicators then should serve as an information system predicating the rate of regional development and the ability to use means in flowing a region or created in a region. The ability of development has to be continuously analyzed and gained information installed by return in the decision-making process (Svatošová 2003).

The analysis of regional development should contain all indicators the monitoring and evaluation of which is of great importance for determination of the development potential. It regards mainly indicators characterizing:

- Overall description of a region
- Economic situation in a region
- Human potential in a region
- Social level
- Infrastructure
- Environment
- Industry
- Services
- Countryside and agriculture

Each of these above mentioned indicators includes many variables. Statistical analyses stem from a relatively wide database. If such an extensive collection of indicators is worked out, when a starting number of features is considerable and for interpretation not providing an easy survey, some of the methods of multidimensional statistic analyses can be used. A partially reduced collection of variables can be later further analyzed in a substantially easier way. The methods of multidimensional statistic analysis serve to solve several tasks:

- A reduction of excessive number of variables, let us say a compression of information in a smaller number of immeasurable hypothetical magnitudes with as small loss of information as possible, which is contained in the monitored features. In this group of methods, we range the analysis of main components, the factor analysis and the canonical analysis.
- A multidimensional classification – setting of rules for ranging objects from one to several groups on the base of measuring of a certain number of

features, and creation of the description of these ranks. There, a method of discrimination analysis is applied.

- An object typology – an organization, let us say a hierarchical settlement in relatively identical groups, whereas neither their number is known, nor are they defined exactly. At the same time, a sequence of these groups can be set according to the chosen criteria. For this task, it is very suitable to use a cluster analysis.

RESULTS AND DISCUSSION

All above-mentioned methods can be used in the evaluation of the development potential. Each of them gives an important information about the analyzed collection. However, their use and especially the consequent interpretation are relatively demanding, mainly for users – non-statisticians it can seem too complicated. From this point of view, it is suitable to use only one of these methods, which will provide most of the desirable information. Outputs can be then presented in a corresponding form, so that they can be used by users in decision making in the sphere of regional policy.

As an example, the analysis of human resources potential in regions is presented here. The help of the analysis of main components, the cluster analysis, and the multidimensional regression and correlation analyses were utilized for it.

For the analysis of human resources potential in regions, the following variables were selected:

- Population density (number of inhabitants per ha)
- Share of urban inhabitants in the total number of inhabitants in the region
- Share of inhabitants in municipalities up to 2 000 persons from the total number of inhabitants in the region
- Share of women in the total number of inhabitants in the region
- Migration per one thousand of inhabitants of the region (immigrants – emigrants)
- Age index (share of inhabitants over 65 years to the number of inhabitants up to 15 years)
- Share of people with university degree in the total number of inhabitants
- Share of high-school graduates in the total number of inhabitants

Results of the regression and correlation analysis

By the help of the methods of regression and correlation analysis especially, the relations among vari-

ables were investigated and multi-collinearity was determined. It was found out between the following variables:

Population density \Leftrightarrow number of inhabitants in the cadastral area

Migration per one thousand of inhabitants \Leftrightarrow a relative addition of inhabitants

Unemployment rate \Leftrightarrow the share of unemployed up to 25 years of age from the total number of inhabitants

The rate of urban inhabitants \Leftrightarrow the share of inhabitants in municipalities up to 2 000 inhabitants

The second task of the regression analysis – a creation of regression model – supposes the determination of a dependent variable. There are not measured mutual relations among variables, but relations among a dependent variable and independent variables. So, it does not deal with an evaluation of the state of human potential in a region, but with the influence of human factor on the chosen resulting variable. As one of the possibilities, a model describing an influence of human factor on the size of gross domestic product per one inhabitant in the region was investigated. The analysis included all the above mentioned variables and with use of the method of Forward Selection, a relatively quality model was created ($R^2 = 95.8\%$) including 5 variables: the share of urban inhabitants; the share of women; age index; the share of people with university degree; and the share of unemployed up to 25 years.

Results of the cluster analysis

The method of cluster analysis was used in a dual way. In the first stage, it was used as a reduction method – for the selection of relevant variables; in the second stage, as a cluster method which provided information about ranking of particular regions in groups by the level of human resources potential in them.

In the first stage, it regarded clustering of variables. The help of three clusters can describe the whole collection. In the first cluster, there were concentrated variables having an influence on the employment in the region – unemployment rate; the share of unemployed up to 25 years; number of inhabitants per the cadastral area; population density; migration of inhabitants. The second cluster then characterizes predominantly age structure of inhabitants (age index; the share of receivers of old-age pensions); the third cluster then testifies to the structure of inhabitants (share of inhabitants in municipalities up to 2 000

inhabitants; share of people with university degree; with high-school degree; women; share of urban inhabitants).

The second stage – ranking of the particular regions in groups with similar behavior from the viewpoint of human factor – provides too general results. The regions were distinguished to only two clusters which do not enable the desirable detailed comparison.

Results of the analysis of main components

The analysis of main components was realized for the simplified model. On the base of results of the regression and correlation analysis, there were set aside the following variables from the original collection in which multi-collinearity was found out:

- share of urban inhabitants in the total number of inhabitants in the region
- migration per one thousand of inhabitants in the region (immigrants – emigrants)
- number of inhabitants per the cadastral area of the region
- share of unemployed up to 25 years in the total number of inhabitants

Including the left 9 variables in the analysis of main components, the model was explained from 95.25% by the help of five components, and so it is possible to work with only these components which can be characterized in the following way:

The *first component* spends 45.88% of the total variability and correlates in the strongest way with variables:

- Index of age: $r = 0.443$
- Share of inhabitants in municipalities up to 2000 inhabitants: $r = 0.397$
- Share of people with high-school degree: $r = 0.412$
- Unemployment rate: $r = -0.431$

On the base of these correlations, this components can be interpreted as an indicator of the available employment in the region.

The *second component* spends 29% of the total variability and correlates in the strongest way with the following variables:

- Population density: $r = 0.407$
- Share of women in the total number of inhabitants: $r = 0.470$
- Share of people with university degree: $r = 0.464$
- Relative addition of inhabitants: $r = -0.471$

This component can be interpreted as an indicator of potential resources of labour force.

The *third component* spends 10.85% of the total variability and correlates in the strongest way with the following variables:

- Age index: $r = 0.483$
- Share of recipients of old-age pensions: $r = -0.472$
- Relative addition of inhabitants: $r = 0.573$

The third component can be interpreted as an indicator of age structure (age) of inhabitants.

The *fourth component* spends 5.57% from the total dispersion and correlates in the strongest way with the following variables:

- Share of inhabitants in municipalities up to 2 000 inhabitants: $r = 0.491$
- Share of women: $r = 0.409$
- Share of people with high-school degree: $r = -0.425$

This component can be interpreted as an indicator of the structure of inhabitants.

The *fifth component* spends 3.92% of the total variability and correlates in the strongest way with the following variables:

- Share of women: $r = -0.055$
- Share of people with university degree: $r = 0.509$

This component can be interpreted similarly as the fourth component – as a complete indicator of the structure of inhabitants.

Regarding the fact that the fifth component gives a complete information to the fourth components and in addition it spends only a small part of the total model variability (3.97%), it will not be considered further.

So, the total model is explained from 91.33% by four components which represent the most important factors having influence on development of the human potential in regions (Table 1).

By the help of this analysis, the most important factors can be defined which have influence on human potential in the region, and further it is possible to compare particular regions from a viewpoint of a state of this indicator. On the base of weights of components and correlation coefficients of particu-

lar variables, it is possible to create a coefficient of development potential:

Construction of the indicator:

$$I^{LP_i} = \frac{\sum_{i=1}^n x_{ij} w_i}{\sum_{j=1}^m \sum_{i=1}^n x_{ij} w_i} \cdot m$$

where:

i = number of variables in the model

j = number of regions

x_{ij} = value of i -th variable in j -th region

w_i = weight of i -th variable counted from the model of analyses of main component

The coefficient enables comparison of regions, setting of the so-called troubleshooting regions, to which it is necessary to pay attention to, and further it represents a starting point for other analyses of the situation in particular regions.

CONCLUSION

All presented statistic multidimensional methods can be in a corresponding way used for the mentioned analysis. Each of them represents a certain quantification of the given process and provides relevant information. Regarding the set aim and especially user base as the most suitable seems to be the use of one method – the analysis of main component. This method provides the most universal information which quantifies the significance of particular variables – factors of regional development; and further it can serve, as it was implied, for computing of the coefficient, by the help of which it is possible to realize a comparison of regions by the determined criteria. From this point of view, the model brings valuable information for decision making in the regional administration. However, it is necessary to realize that the single coefficient provides only one pieces of information and it is necessary to complete it with a detailed economic evaluation. Its low height does not always necessity inform about the insufficient development potential of so called troubleshooting regions, but it can also regard developed regions which do not have such a strong need of development. The help of the following analyses of relevant indicators can reveal all this.

The mentioned methodology is flexible. In the model of analysis of main components, the form which the construction of the coefficient sets, other variables can be ranked according to on what will be the big-

Table 1. Factors influencing human potential in regions

Component (a factor)	Weight
Employment of inhabitants	0.4588
Potential resources of work powers	0.2903
Age structure of inhabitants	0.1085
Structure of inhabitants	0.0557

gest stress laid in the monitored period. Then the coefficient will give information about the state of use of the expended means.

REFERENCES

Boháčková I., Hrabánková M., Svatošová L. (2003): Role of agriculture in development of agricultural regions. *Agricultural Economics – Czech*, 49: 229–232.

Hebák P., Hustopecký J. (1987): Vícerozměrné statistické metody (Multidimensional statistic methods). SNTL/ALFA, Praha.

Parr Rud O. (2001): *Data Mining*. Computer Press, Praha.

Svatošová L. (2003): Zdroje informací a možnosti analýz rozvojového potenciálu regionů (Sources of information and possibilities of analyses of development potential of regions). Collection of papers from international scientific conference *Agrarian perspectives XII*: 859–865; ISBN 80-213-1056.

Svatošová L. (2003): Analýza klíčových aspektů regionálního rozvoje (Analysis of crucial aspects of regional development). Collection of papers from international conference of departments of statistics and operational research *Kvantitativne metody v ekonomii, metodologické a praktické aspekty výskumu v období vstupu do EÚ*. Račkova dolina, September; ISBN 80-8069-299-8.

Arrived on 10th February 2005

Contact address:

Doc. Ing. Libuše Svatošová, CSc., Česká zemědělská univerzita v Praze, Kamýcká 121, 165 21 Praha 6-Suchdol, Česká republika
e-mail: svatosova@pef.czu.cz
