QUANTITATIVE METHODS IN ECONOMICS
Vol. XIV, No. 2, 2013, pp. 89 – 98

EVALUATION OF VOIVODESHIPS DIVERSIFICATION IN POLAND ACCORDING TO TRANSPORT INFRASTRUCTURE INDICATORS

Aldona Migała–Warchoł, Marek Sobolewski
Department of Quantitative Methods, Rzeszów University of Technology
e-mail: amigala@prz.edu.pl, msobolew@prz.edu.pl

Abstract: The aim of this paper is to determine the level of diversification of Polish voivodeships in terms of the selected indicators of transport infrastructure. The data collected from the Local Data Bank of Central Statistical Office will be used in the paper. By means of some methods of linear ordering and cluster analysis the ordering and classification of the Polish voivodeships will be carried out. The obtained results will allow to make an evaluation of Polish voivodeships in terms of the level of development of transport infrastructure.

Keywords: transport, multivariate statistical analysis, synthetic measures, cluster analysis

INTRODUCTION

The concept of infrastructure, despite the fact that it has been operating in the Polish language for many years, there is still no generally accepted definition, and thus it is not clearly understood. The term *infrastructure* derives from English and it means "the foundation of base i.e. the necessary basis for the economy."

In Polish literature the concept of infrastructure in the most general terms is defined as the basic facilities and institutions necessary for the proper functioning of the economy. These devices include man-made, permanently located, line and point objects for public use, which are the foundation of socio-economic development, in view of their functions to move people and goods (transport), news (communications), electricity (energy) and water (water management) [Rydzkowski 2011].

The purpose of infrastructure is to provide the basic conditions of development of the socio-economic system as a whole and the rest elements of the economy. Accordingly, the transport infrastructure includes man-made, permanently sited basic facilities of roads (linear infrastructure) and transport points (point infrastructure). It affects the economy and society by creating favorable conditions for the movement of people and goods (freight) in the direct and indirect form. From an economic point of view the most important feature of the transport infrastructure is the public nature of its services. There are of course exceptions, i.e. the transport infrastructure services as private goods. Transport infrastructure is formed mainly by three basic groups:

- roads of all modes of transport;
- transport points (airports, ports, etc.);
- ancillary equipment used for the direct control of the roads and transport points [Gołembska 2008].

The aim of this paper is to present a comprehensive analysis of the level of transport infrastructure in Polish voivodeships in 2011[1]. In the first section the criteria for the selection of variables describing the transport infrastructure were described in detail. Pre-selected set of variables was verified statistically, then the synthesis of the information contained therein, by means of taxonomic methods - linear ordering and clustering was done. The obtained results were subjected to the interpretation, and then there were examined the correlations of the level of development of transport infrastructure with the selected macroeconomic indicators.

SUBSTANTIVE SELECTION CRITERIA OF DIAGNOSTIC VARIABLES

The selection of diagnostic variables that describe directly immeasurable social and economic phenomenon, is a point of reference adopted by the researcher. In case of the voivodeships ranking according to the level of development of transport infrastructure an important element is a matter of analyses recipients – either it is an analysis prepared for the individual needs of the residents, or the possibility of development of large manufacturing companies [*Atrakcyjność Inwestycyjna*… 2012].

In the first case, the more important is the functioning of urban and public transport and the quality of local roads, whereas from the point of view of large companies, it is important a network of national roads and the location of the voivodeship with respect to the state border, sea ports or airports. Evaluation of diagnostic variables will be different from the point of view of the citizens who take long trips (and they care about good communication between regions and

[1]It is worth to appeal to a broader context and remind that according to comparative international studies, Poland is at a distant 74. place due to the operation of railway transport, and in the case of road infrastructure is the 134. place (!) [*The Global Competitiveness Report…* 2011].

states) than the use of the transport network within the region. In this paper, an attempt was done to get such diagnostic variables, which would describe the level of development of transport infrastructure in the most comprehensive way, and therefore there were both variables of "global" character (a network of expressways and highways) and "local" one, such as the length of cycle paths or the use of urban transport.

Pre-selection, based on a review of the literature [Atrakcyjność inwestycyjna ... 2012; Wierzbińska, Chudy 2011], led to the creation of a list of 17 diagnostic variables that are listed in the table. The values of some of them are taken directly from the publication of the CSO, some of them are the result of simple calculations of raw data allowing to determine the intensity ratios of certain phenomena. For each variable it was indicated, whether in terms of rankings created, they will be treated as a stimulant or destimulant. The vast majority of the features are stimulants. There is presented the list of diagnostic variables together with the selected descriptive statistics: median, minimum, maximum and coefficient of variation in the first table.

Table 1. List of diagnostic variables together with the selected descriptive statistics

Diagnostic features	**Me**	**min**	**max**	***V***
Traffic at airports [thous. people] ↑	7,6	0,0	88,0	136%
Transportation of cargo at airports [t] ↑	106	0	50 951	227%
Travel time to the nearest sea port [h][a)] ↓	4:59	1:00	8:59	52%
Bridges and overpasses fixed at 100 km of roads ↑	6,6	3,8	20,5	55%
Tunnels and subways, the 1000 km of roads ↑	0,8	0,4	5,6	100%
Hard-surface roads [in km per 100 km^2] ↑	89,8	53,3	179,6	37%
Roads with improved surface [in % of all roads] ↑	91,5	85,3	95,5	3%
Motorways and highways [in km per 100 km^2] ↑	0,67	0,00	1,98	79%
Bicycle paths for 10 thous. km^2 [km] ↑	179,3	83,3	454,1	52%
The urban population served by public transport [%] ↑	75,6	43,9	93,6	19%
Places in urban vehicles per thous. inhabitants ↑	89,8	67,1	147,5	22%
Cars for 1 thous. inhabitants ↑	462,5	409,5	530,8	8%
Lorries for 1 thous. inhabitants ↑	75,2	66,7	108,0	15%
Railway lines per 100 km^2 [km] ↑	6,5	3,8	17,4	46%
Number of passengers of the regional transport per 1 inhabitant ↑	3,1	0,8	20,5	111%
Interregional transport passengers per 1 inhabitant ↑	1,3	0,8	2,4	35%
The share of electrified lines [%] ↑	55,0	29,1	91,1	36%

↑ – stimulant ↓ – destimulant

[a]) Information on road transport was taken from the "map of road connections" available on the website of the General Directorate for National Roads and Motorways (www.gddkia.gov.pl)
Source: own studies

Diagnostic variables concern the existing infrastructure and its actual use in the field of aviation, maritime, road and rail transport. In view of the crucial importance of road transport, this mode of transport is described by the most variables, and the information about the possible use of maritime transport was described by contrast by only one variable which contains information about the average time of travel from the voivodeship to the nearest major sea port (Szczecin or Gdańsk).

It should be noted that some of the features that contain information about the state of the infrastructure - such as rail - can lead to erroneous conclusions. For example, the region with the highest density of railway network is in Śląskie Voivodeship but the share of rail passenger transport is much greater in the Pomorskie and Mazowieckie Voivodeships.

STATISTICAL METHODS

Statistical verification of the diagnostic value of the pre-selected diagnostic variables consisted of two phases:

- analysis of variance (based on classical coefficient of variation);
- correlation analysis (it was used the Potential Information Analysis – PIA).

In the first case, for each diagnostic variable there was determined the coefficient of variation, but for further analysis the characteristics for which the value exceeded 10% were automatically enrolled. When the coefficient of variation did not exceed 10%, it was made another analysis of the substantive value of the given variable, in terms of accuracy of the description of transport for regions and then it was taken the decision whether to exclude or allow further analysis.

PIA method consists in searching the variables which are the most strongly correlated with the others and removing from the analysis those for which the correlation coefficient exceeds a predetermined threshold (a detailed description of these and other alternative methods of reducing the set of diagnostic variables can be found in the literature [Grabiński et al. 1989]). For the study it was adopted a fairly strict level of the correlation coefficient R = 0.80. In the case of variables with a high degree of correlation, the decision on their exclusion from the analysis was preceded by a reassessment of their substantive meaning.

The voivodeships ranking was made by means of unsupervised linear ordering methods, and to standardize the data it was applied the zero unitarisation method [Kukuła 2000]. In order to group the voivodeships according to indicators of the transport development it was used the hierarchical clustering procedure – the Ward's method. Both methods are among the most popular procedures for taxonomic studies used in the socio-economic conditions.

Both while doing the ranking, as well as cluster analysis, there was not used the weighing of diagnostic features. Grouping procedure, using the Ward's method, was based on the feature values that were subjected to classic standardization. The differentiation between objects and clusters, according to the specific of the Ward's method was measured by the square of the Euclidean distance.

To examine the compatibility of the results with other selected indicators of the level of socio-economic development, the correlation analysis methods were used. To avoid distortions associated with the presence of outliers, it was applied nonparametric Spearman rank correlation coefficient. There were also provided the value of test probability *p* which allows to assess whether the tested relationship is not just about "accidental" relationship of these two traits.

The calculations were performed by using *STATISTICA.* There were applied the standard procedures of the program, supplemented by the authors' extensions created in *STATISTICA Visual Basic* language with which it is achieved the complete taxonomic analysis report directly in *WORD.* Minimized in this way the calculation time was used to deepen the interpretation of the results.

VERIFICATION OF DIAGNOSTIC VARIABLES

The values of the variation coefficient *(V)* is given in Table 1. For two characteristics: the share of roads with improved surface and the ratio of passenger cars per 1 thousand inhabitants low coefficient of variation (3 and 8%) constitute the rationale for excluding them from the further analysis.

As it is debatable to what extent the level of motorization (especially when it comes to the number of cars) is associated with the development of transport infrastructure, it was decided to omit this feature in further considerations. The more complex and interesting issue is related with the fraction of roads with improved surfaces in the total length of roads in the region. Although this feature is characterized by low volatility, but it should be noted that this rate of structure can be defined as a fraction of roads with "unimproved" surface, which will change the average level, but not the standard deviation. This reflects some kind of methodological "weakness" of uncritical use of the variation coefficient to reduce the set of diagnostic variables. As for the indicator of the "unimproved" roads the variation coefficient is more than 30%, thus taking into account the substantive importance of this indicator, particularly in terms of accessibility of rural areas, it was decided to leave it for further analysis.

In the second step, for the reduced by one characteristic a set of diagnostic variables there was conducted an analysis of information potential. There were distinguished the characteristics the most correlated with the others (so-called central variables) and it was discussed the exclusion of those which were correlated with them to an extent exceeding a certain threshold value, which is assumed to be $R^* = 0.8$ (so-called satellite variables).

The analysis led to the finding of two central characteristics and two satellite ones that should be excluded from further consideration. The first pair are the rate of passenger air transport (central variable) and the transport of goods at airports (satellite variable). Since these two variables describe the same mode of transport, it seems that without worry to omit some relevant information the satellite variable (transport of goods by air) can be excluded from a further analysis.

We have a different situation in the case of the second pair of variables: density of the rail network per 100 km^2 (central variable) and the ratio of the density of expressways and motorways (satellite variable). As they describe the state of the infrastructure of two different modes of transport, it was decided to leave both of them for further analysis.

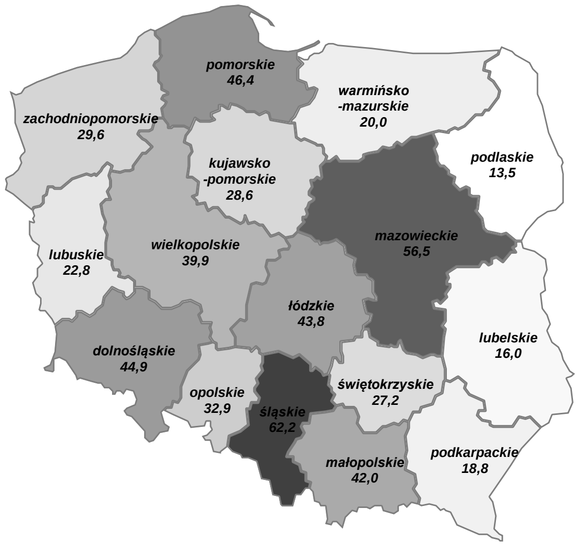
VOIVODESHIPS RANKING ACCORDING TO THE DEVELOPMENT OF TRANSPORT INFRASTRUCTURE

Summing up the considerations discussed in the previous paragraph, from the initial list of variables (Table 1), after the statistical verification there were removed only two features: the motorization indicator and cargo air transport. Thus, the taxonomic analyses will be conducted based on the values of 15 diagnostic variables.

By using the method of linear ordering it was determined the synthetic indicator, which made it possible to prioritize regions because of the level of transport infrastructure (Figure 1).

Śląskie Voivodeship has the highest value of synthetic measure of transport development. The second place in the ranking took Mazowieckie Voivodeship. Another region (Pomorze, Dolny Śląsk) have the value of the synthetic measure much lower than the two leaders indicated above. By far the worst assessment of development and the use of transport infrastructure in the region is characterized by Eastern Poland (Warmia and Mazury, Podkarpackie, Lubelskie and at the very end – Podlaskie Voivodeship).

Figure 1. Ranking of voivodeships by the development of transport infrastructure in 2011



Source: own studies

From a practical point of view, it is interesting how much the level of development of transport is associated with other indicators of socio-economic development. As it results from conducted analysis of correlation (Table 2), a voivodeship with a high degree of development of transport infrastructure has a lower unemployment rate, higher average salary and industrial production index. These correlations are statistically significant and have a relatively high strength. Only the annual growth rate of GDP is correlated to a weaker extent with the level of transport development.

Table 2. Relationship of synthetic measure of transport infrastructure with some selected indicators of socio-economic development

Socio-economic development indicator	**Synthetic measure of transport infrastructure**
Unemployment rate	*R* = -0,70 (*p* = 0,0027**)
Average salary (gross)	*R* = 0,79 (*p* = 0,0003***)
Sold industry production per one inhabitant	*R* = 0,87 (*p* = 0,0000***)
GDP dynamics	*R* = 0,48 (*p* = 0,0596)

R – Spearman's rank correlation coefficient, *p* – test probability value

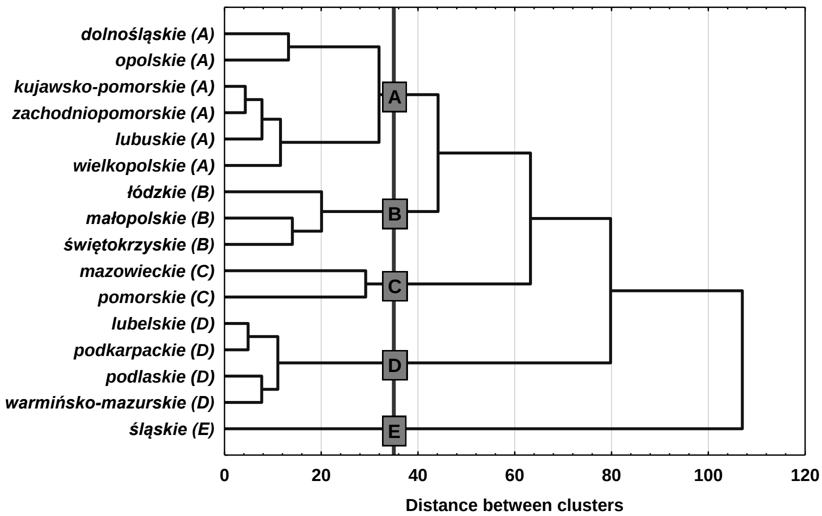
Source: own studies

Of course, the analysis of the relation of transport infrastructure development measure with other indicators should be broadened, taking also into account the dynamics of all the features in the longer term. On the basis of connections to a single, 2011 year, it is impossible to predict the direction of the reason-result relation between the transport level and other indicators of economic development.

VOIVODESHIPS GROUPING ACCORDING TO THE LEVEL OF TRANSPORT INFRASTRUCTURE DEVELOPMENT

As the result of the clustering by the Ward's method the dendrogram was developed which presents the various stages of the agglomeration (Fig. 2). Based on the analysis of the clustering process, taking into account the substantive issues, it was decided to divide the Polish voivodeships into five clusters. The elements of particular groups (designated by the letters A-E) can be read from the dendrogram.

Figure 2. Results of cluster analysis by the Ward's method



Source: own studies

Table 3 shows the group average indicators of all the diagnostic features created for each cluster. The values included in the table are quotients of group average and the average for all the regions. If the ratio is greater than 1, one can say that the group is characterized by relatively preferred values of the feature (contrary to destimuli). The values of the group average lower than 1 indicate a weaker position of the region (or vice versa for destimulants).

Table 3. Relation of group means of individual diagnostic variables to the total mean

Diagnostic variables	Indicators of group mean values				
	A	**B**	**C**	**D**	**E**
People checked in at airports	0,57↓↓	0,94	3,84↑↑	0,16↓↓	1,47↑↑
Time travel to the seaport	0,75↑	1,23↓	0,60↑↑	1,31↓	1,38↓↓
Bridges and overpasses	1,08	1,05	0,72↓	0,77↓	1,81↑↑
Tunnels and subways	0,73↓	0,87↓	1,49↑↑	0,45↓↓	4,24↑↑
Hard-surface roads	0,86↓	1,36↑	0,88↓	0,76↓	1,92↑↑
Unimproved surface roads	1,08	1,30↓	0,82↑	0,78↑	0,86↑
Express roads and highways	1,40↑↑	0,80↓	0,81↓	0,12↓↓	3,08↑↑
Bicycle paths	1,19	0,64↓↓	1,27↑	0,50↓↓	2,38↑↑
Population of cities with public transport	0,85↓	1,16↑	1,13↑	0,97	1,29↑
Places in urban vehicles	1,02	1,11↑	1,28↑	0,80↓	0,84↓
Lorries per 1 thousand inhabitants	0,99	1,09	1,22↑	0,87↓	0,88↓
Density of railways	1,06	0,93	0,83↓	0,67↓	2,50↑↑
Passenger traffic in the region	0,89	0,34↓↓	3,62↑↑	0,43↓↓	0,67↓
Passenger traffic between regions	1,06	1,22↑	1,19↑	0,70↓	0,78↓
The share of electrified lines	0,92	1,45↑↑	1,06	0,64↓	1,44↑↑

Average deviation of the group mean from the total mean:
↑ - favorable
↓ - unfavorable
(the groups which are distinguished by low or high indicators of group means are marked with different shades of gray)

Source: own studies

The table is constructed in such a way as to maximally facilitate the results interpretation. The exact values of group means are also illustrated by the marks.

CONCLUSIONS

In the paper it was made an analysis of the spatial differentiation of Polish voivodeships in respect to quality of transport infrastructure. The leader turned out to be Śląskie Voivodeship, which was slightly ahead of Mazowieckie Voivodeship. The voivodeships with the lowest development level of transport infrastructure were Podlaskie and then Lubelskie and Podkarpackie Voivodeships. The lowest level of transport infrastructure are characterized by far the voivodeships of the Eastern Poland.

A natural extension of the presented in the analyses will be taking into account the time aspect. Based on these same diagnostic variables there will be created the time-space ranking of the level of infrastructure development, which will help to determine the pace of development of the various regions. Particularly valuable is linking the changes in the level of development of transport infrastructure with changes of the selected macroeconomic indicators.

REFERENCES

Investment attractiveness of Polish regions and sub-regions 2012 (2012), the Institute for Market Economy, Gdańsk.

Coyle J. (2002), Zarządzanie logistyczne, Polskie Wydawnictwo Ekonomiczne, Warszawa.

Gołembska E. (2008), Kompendium wiedzy o logistyce, PWN, Warszawa.

Grabiński T. (1992), Metody taksonometrii, AE Kraków.

Grabiński T., Wydymus S., Zeliaś A. (1989), Metody taksonomii numerycznej w modelowaniu zjawisk społeczno-gospodarczych, PWN, Warszawa.

Kukuła K. (2000), Metoda unitaryzacji zerowanej, PWN, Warszawa.

Nowak E. (1990), Metody taksonomiczne w klasyfikacji obiektów społeczno-gospodarczych, Państwowe Wydaw. Ekonomiczne, Warszawa.

Ostasiewicz W. (1998), Statystyczne metody analizy danych, AE Wrocław.

Rydzkowski W. (2011), Usługi logistyczne, Instytut Logistyki i Magazynowania Poznań.

Stanisz A. (2007), Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny, StatSoft, Kraków.

The Global Competitiveness Report 2011-2012 (2011), World Economic Forum, Geneva, Switzerland.

Transport. Results of operations in 2011 (2012), Central Statistical Office, Warsaw.

Wierzbińska M., Chudy-Laskowska K. (2011), Zróżnicowanie województw pod względem infrastruktury transportowej w Polsce – wyniki badań, Zeszyty Naukowe Politechniki Rzeszowskiej, nr 282, Ekonomia i Nauki Humanistyczne, z. 18.