ARGUMENTA
OECONOMICA
CRACOVIENSIA

No 1(20) • 2019
ISSN 1642-168X
e-ISSN 2545-3866
AOC, 2019; 1(20): 71–85
https://doi.org/10.15678/AOC.2019.2005

Grażyna Dehnel
Marek Walesiak
Marek Obrębalski

COMPARATIVE ANALYSIS OF THE ORDERING OF POLISH PROVINCES IN TERMS OF SOCIAL COHESION

Abstract

The article describes an assessment of the social cohesion of Polish provinces. The assessment was based on classical metric and interval-valued data using a hybrid approach combining multidimensional scaling with linear ordering. In the first step, after applying multidimensional scaling, the objects of interest were represented in a two-dimensional space. In the second step, the objects were linearly ordered based on the Euclidean distance from the pattern object. Interval-valued variables characterize the objects of interests more accurately than do metric data. Classic data are of an atomic nature, i.e. an observation of each variable is expressed as a single real number. By contrast, an observation of each interval-valued variable is expressed as an interval. Interval-valued data were derived by aggregating classic metric data on social cohesion at the level of districts to the province level. The article describes a comparative analysis of the results of an assessment of the social cohesion of Polish provinces based on classical metric data and interval-valued data.

Grażyna Dehnel, Poznań University of Economics and Business, Faculty of Informatics and Electronic Economy, Department of Statistics, al. Niepodległości 10, 61-875 Poznań, Poland, e-mail: g.dehnel@ue.poznan.pl, ORCID: https://orcid.org/0000-0002-0072-9681.

Marek Walesiak, Wrocław University of Economics, Department of Econometrics and Computer Science, Nowowiejska 3, 58-500 Jelenia Góra, Poland, e-mail: marek.walesiak@ue.wroc.pl, ORCID: https://orcid.org/0000-0003-0922-2323.

Marek Obrębalski, Wrocław University of Economics, Department of Regional Economy, Nowowiejska 3, 58-500 Jelenia Góra, Poland, e-mail: marek.obrebalski@ue.wroc.pl, ORCID: https://orcid.org/0000-0002-7030-7089.

Keywords: social cohesion, composite indicators, interval-valued data, multidimensional scaling, R software.
JEL Classification: C38, C43, C63.

1. An Overview of Social Cohesion Concepts

Social cohesion is a term which is mainly used in the context of policies conducted by the European Union, the Council of Europe, and OECD. It is frequently invoked in various initiatives and analyses, both in the area of political activities and in research. Assessment of social cohesion is made difficult by the absence of one, unequivocal definition which captures all aspects of this concept. Existing definitions differ in terms of areas of life they focus on, periods they refer to, political ideas they represent, and methods they employ to foster cohesion (*Concerted Development…* 2005, p. 23).

Multi-faceted social cohesion conducted at different levels of territorial organization, including national and regional, is aimed at narrowing spatial development disparities, particularly significant disparities in various social areas. Modern territorial units (e.g. regions) make up a mosaic characterised by varying levels of cohesion. This is clearly confirmed by the results of the 7th cohesion report prepared by the European Commission (*My Region, My Europe, Our Future* 2017). The report refers to three dimensions of cohesion, namely, economic, social and territorial. Each cohesion aspect is defined and measured according to a multi-disciplinary approach.

To track changes in the level of social cohesion across regions, the European Commission also measures social progress. This is defined as "a society's capacity to meet the basic human needs of its citizens, to establish the basis for people and communities to improve and sustain their quality of life and to create the conditions for people to reach their full potential" (*My Region, My Europe, Our Future* 2017, p. 91). In EU practice, one of the indicators used to measure the level of social cohesion is the EU Regional Social Progress Index (EU-SPI). This index comprises three dimensions of social progress (Annoni & Dijkstra 2016, p. 2):

– basic human needs (nutrition and basic medical care, water and sanitation, shelter (housing), personal safety),

– foundations of well-being (access to basic knowledge, access to information and communication, health and wellness, environmental quality),

– opportunity (personal rights, personal freedom and choice, tolerance and inclusion, access to advanced education).

The range of variables taken into consideration in the measurement based on the regional EU-SPI index is a significant example of multi-disciplinary research of social cohesion at the regional level.

The subject literature includes studies describing applications of multivariate statistical methods to measure social cohesion across territorial units at different levels. In the Polish literature, for example, one can mention a study conducted by Balcerzak (2015), which analyzed social cohesion in EU countries based on the development measure proposed by Hellwig (1972). Other studies carried out by Dickes and Valentova (2013), Dickes, Valentova and Borsenberger (2010), and Bottoni (2018) made use of multidimensional scaling, structural equation modelling (SEM), multilevel models, and an aggregate index to measure social cohesion in 47 and 33 European countries, respectively. Rajulton, Ravanera and Beaujot (2007) used the results of factor analysis and standardization to create an aggregate index to measure social cohesion in 49 census metropolitan areas in Canada.

The purpose of this article is to present a comparative analysis of the results of the social cohesion measurement produced for Polish provinces on the basis of classical metric data and interval-valued data. This latter type of data has not been used so far for measuring social cohesion. The assessment of social cohesion in Polish provinces based on these two types of data was conducted by means of a hybrid approach, which combines multidimensional scaling (MDS) and linear ordering to visualize results in a two-dimensional space.

2. Research Methodology

To order objects of interest in terms of their social cohesion, the authors used a two-step procedure proposed by Walesiak (2016), which makes it possible to visualize the results of linear ordering. In the first step, after applying multidimensional scaling, objects of interest are visualized in a two-dimensional space. In the second step, the set of objects is linearly ordered.

The extended research procedure, conducted separately for metric data and interval-valued data, consists of the following steps (cf. Walesiak & Dehnel 2018):

1. Select a complex phenomenon which cannot be measured directly (the level of social cohesion).

2. Select a set of objects and a set of variables (metric and interval--valued), which are substantively related to the complex phenomenon of interest. Add a pattern object (upper pole of development) and an anti--pattern object (lower pole of development) to the set of objects. Identify preference variables (stimulants, destimulants, and nominants) in the set of the variables.

3. Collect data and construct data matrix $\mathbf{X}=[x_{ij}]_{nxm}$ (value of j-th variable for i-th object, $i, k = 1, \ldots, n, j = 1, \ldots, m$) for metric variables or in the form of data table $\mathbf{X}=[x_{ij}]_{nxm}$ $(x_{ij}=[x_{ij}^l, x_{ij}^u])$, where $x_{ij}^l \leq x_{ij}^u$) for interval-valued variables. The pattern object (upper pole) includes the most favourable variable values, whereas the anti-pattern (lower pole) – the least favourable values of the preference variables. In the case of interval-valued variables, coordinates are determined separately for lower and upper limits of the interval.

4. Normalize variable values and arrange the data in the form of a normalized data matrix $\mathbf{Z}=[z_{ij}]_{nxm}$ for metric data or in the form of a normalized data table $\mathbf{Z}=[z_{ij}]_{nxm}$ $(z_{ij}=[z_{ij}^l, z_{ij}^u]$, where $z_{ij}^l \leq z_{ij}^u, z_{ij}$ – normalized observation) for interval-valued variables. The purpose of normalization is to ensure comparability of variables. This is achieved by removing measurement units from the results and standardizing their orders of magnitude. Interval-valued data require a special normalisation treatment. The lower and upper limits of the interval of j-th variable for n objects are combined to form one vector containing $2n$ observations. This approach enables the application of normalization methods suitable for classic metric data. Metric data were normalized using the `data.Normalization` function, and interval-valued data using the `interval_normalization` function from the `clusterSim` package (Walesiak & Dudek 2018a).

5. Select a distance measure for metric data (Manhattan, Euclidean, Chebyshev, Squared Euclidean, GDM1[1] – see, e.g., Everitt *et al.* 2011, pp. 49–50), calculate distances and create a distance matrix $\boldsymbol{\delta}=[\delta_{ik}(\mathbf{Z})]_{nxn}$ $(i, k=1, \ldots, n)$. Select a distance measure for interval-valued data (Ichino-Yaguchi, Euclidean Ichino-Yaguchi, Hausdorff, Euclidean Hausdorff – see Billard & Diday 2006, Ichino & Yaguchi 1994), calculate distances and create a distance matrix $\boldsymbol{\delta} = [\delta_{ik}(\mathbf{Z})]_{nxn}$.

6. Perform multidimensional scaling (MDS): f: $\delta_{ik}(\mathbf{Z}) \rightarrow d_{ik}(\mathbf{V})$ for all pairs (i, k), where f denotes a mapping of distances from m-dimensional

[1] See Jajuga, Walesiak & Bąk (2003).

space $\delta_{ik}(\mathbf{Z})$ into corresponding distances $d_{ik}(\mathbf{V})$ in q-dimensional space ($q < m$). To enable graphical presentation of results, $q = 2$. Distances $d_{ik}(\mathbf{V})$ are unknown. The iterative procedure implemented in the `smacof` algorithm, which makes it possible to find a configuration $\mathbf{V}$ (given q dimensions) and calculate a distance matrix $d_{ik}(\mathbf{V})$, is described in a work by Borg & Groenen (2005, pp. 204–205). To ensure an optimal procedure of multidimensional scaling, we selected methods of normalising variable values, distance measures, and scaling models according to the procedures (for metric and interval-valued data) available in the `mdsOpt` package (Walesiak & Dudek 2018b), which employ the `smacofSym` function from the `smacof` package (Mair *et al.* 2018). More details about the selection of the optimal procedure of multidimensional scaling can be found in Walesiak & Dudek (2017).

7. Finally, after applying multidimensional scaling, a data matrix in 2-dimensional space $\mathbf{V} = [v_{ij}]_{nxq}$ ($q = 2$) is generated. Depending on the location of the pattern and anti-pattern object in the dimensional scaling space $\mathbf{V} = [v_{ij}]_{nx2}$, the coordinate system needs to be rotated by an angle of φ according to the formula:

$$[v'_{ij}]_{nx2} = [v_{ij}]_{nx2} \times D, \tag{1}$$

where: $[v'_{ij}]_{nx2}$ – data matrix in 2-dimensional scaling space after rotating the coordinate system by an angle of φ,

$D = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix}$ – rotation matrix.

The rotation does not change the arrangement of objects relative to one another, but it makes it possible to position the set axis connecting the pattern and anti-pattern, along the identity line, which improves the visualisation of results.

8. Present graphically and interpret the results (of multidimensional scaling – MDS) in a 2-dimensional space. Two points, representing the anti-pattern and pattern, are joined by a straight line to form the so-called set axis in the diagram drawn in the two-dimensional space. Isoquants of development (curves of equal development) are drawn from the pattern point. Objects located between the isoquants represent a similar level of development. The same development level can be achieved by objects located at different points along the same isoquant of development (owing to a different configuration of variable values).

9. Order objects according to the values of aggregate measure d_i based on the Euclidean distance from the pattern object (Hellwig 1981):

$$d_i = 1 - \frac{\sqrt{\sum_{j=1}^{2} (v_{ij} - v_{+j})^2}}{\sqrt{\sum_{j=1}^{2} (v_{+j} - v_{-j})^2}}, \tag{2}$$

where: v_{ij} – j-th coordinate for i-th object in the 2-dimensional MDS space, $v_{+j}(v_{-j})$ – j-th coordinate for the pattern (anti-pattern) object in the 2-dimensional MDS space.

The values of the aggregate measure d_i are included in the interval [0; 1]. The higher the value of d_i, the higher the level of social cohesion of the objects of interest. Target objects are ranked according to the descending values of the aggregated measure (2).

3. Results of the Empirical Study

According to the approach used for the purposes of measuring social cohesion based on the regional EU-SPI index, three dimensions are considered: basic human needs, foundations of well-being, and opportunities. Given this 3-dimensional frame of reference, the social cohesion of Polish provinces was measured using 26 metric variables:

1. Basic human needs (7 variables):

$x1$ – mean monthly wage (in PLN) – stimulant,

$x5$ – total unemployment rate in % – destimulant,

$x9$ – mean useful floor area of a dwelling per inhabitant in m^2 – stimulant,

$x10$ – average number of persons per room – destimulant,

$x11$ – length of the sewerage network in relation to the length of the water supply network in % – stimulant,

$x14$ – number of doctors and dentists per 10,000 of the population – stimulant,

$x25$ – crimes reported (criminal offences, against life and health, and against property) per 10,000 of the population – destimulant.

2. Foundations of well-being (11 variables),

$x12$ – people using water treatment services (% of the total population) – stimulant,

$x13$ – percentage of all dwellings equipped with central heating – stimulant,

$x16$ – children enrolled in day-care centres per 1000 children up to the age of 3 – stimulant,

$x17$ – children enrolled in nursery schools per 1000 children aged 3–5 – stimulant,

$x18$ – students taking obligatory English classes in primary and intermediate schools (% of all students) – stimulant,

$x19$ – number of students per class in secondary schools – destimulant,
$x20$ – members of sports clubs per 1000 of the population – stimulant,
$x21$ – users of public libraries per 1000 of the population – stimulant,
$x22$ – people participating in cultural events (organised by cultural centres and clubs) per 1000 of the population – stimulant,
$x23$ – area of public green space (parks, residential green space) per 10,000 of the population (in ha) – stimulant,
$x24$ – length of district and municipal improved hard surface roads per 10,000 of the population (in km) – stimulant.
3. Opportunities (8 variables):
$x2$ – persons in households (below the income threshold) relying on social assistance per 1000 of the population – destimulant,
$x3$ – age dependency ratio (number of people aged 0–14 and those aged 65 and older per 100 people of working age) – destimulant,
$x4$ – share of women in the labour force in % – nominant (with a nominal value of 50%),
$x6$ – share of young people (up to the age of 25) in the population of registered unemployed in % – destimulant,
$x7$ – share of long-term unemployed (over 12 months) in the population of registered unemployed in % – destimulant,
$x8$ – number of job offers for disabled people per 1000 registered disabled unemployed – stimulant,
$x15$ – places in stationary social welfare facilities per 10,000 of the population – stimulant,
$x26$ – voter turnout in local elections (for municipal authorities and town councils with district rights) in 2014 in % – stimulant.

The statistical data come from the Local Data Bank maintained by the Statistics Poland. The reference year is 2016, except for variable $x26$, which contains data for 2014 (the last local government elections). The $x4$ nominant variable was converted into a stimulant according to the following formula:

$$x_{ij} = -\left| x_{ij}^{N} - \text{nom}_j \right|, \tag{3}$$

where: x_{ij}^{N} – value of j-th nominant observed in i-th object, nom_j – nominal level of j-th variable.

The purpose of the empirical study was to compare alternative measurements of social cohesion in Polish provinces. In the classical approach, the second part of the measurement procedure was applied to a data matrix consisting of 17 objects (16 provinces and the average province) described in terms of 26 metric variables. The second approach was based on interval-valued data, which had to be prepared in two steps. First, we collected classical metric data on social cohesion at the district level (380 districts described in terms of 26 variables), which were then aggregated at the province level to obtain interval-valued data. The lower limit of the interval for each variable in the province was obtained by calculating the first quartile based on district-level data. The upper limit was obtained by calculating the third quartile.

For metric data, the optimal scaling procedure was selected after testing 6 normalisation methods (n1, n2, n3, n5, n5a, n12a – cf. Walesiak & Dudek 2018a), 5 distance measures (Manhattan, Euclidean, Squared Euclidean, Chebyshev, GDM1) and 4 MDS models (ratio transformation, interval transformation, and second and third degree polynomial – Borg & Groenen 2005, p. 202), which amounted to a total of 120 procedures of multidimensional scaling. After applying the `optSmacofSym_mMDS` function from the `mdsOpt` R package, the optimal procedure of multidimensional scaling was selected, which uses the normalisation method n2 (positional standardization), the ratio scaling model, and the Manhattan distance.

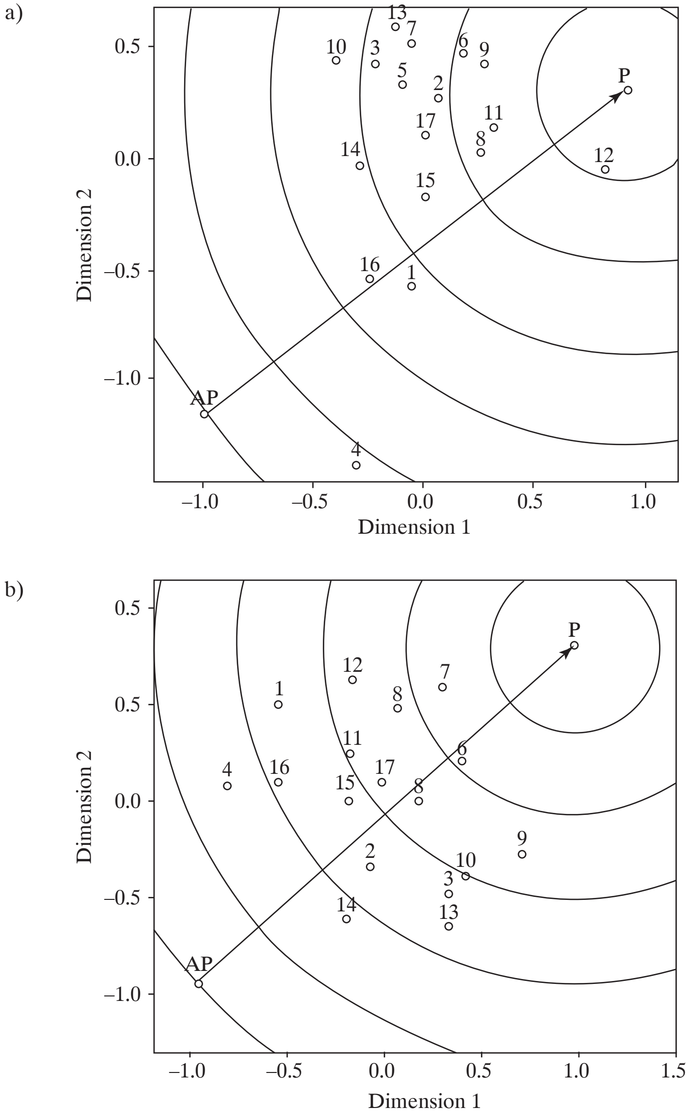
For interval-valued data, the optimal scaling procedure was selected after testing 6 normalisation methods (n1, n2, n3, n5, n5a, n12a), 4 distance measures (Ichino-Yaguchi, Euclidean Ichino-Yaguchi, Hausdorff, Euclidean Hausdorff) and 4 MDS models (ratio transformation, interval transformation, and second and third degree polynomial), which produced a total of 96 procedures of multidimensional scaling. After applying the `optSmacofSym_mMDS` function from the `mdsOpt` R package the optimal procedure of multidimensional scaling was selected, which uses n2 normalisation (positional standardization), the ratio scaling model, and the Euclidean Hausdorff distance.

Figure 1a shows the results of multidimensional scaling of 17 objects (16 provinces and the average province) in terms of the level of social cohesion in 2016 for interval-valued data, whereas Figure 1b shows the corresponding results obtained on the basis of metric data.

In Figures 1a and 1b, the anti-pattern (AP) object and the pattern (P) object are linked by a straight line, known as the set axis. Six isoquants of development were identified by dividing the set axis into 6 equal parts. A longer distance from the isoquant represents a lower level of social cohesion.

a)

b)



1. Dolnośląskie, 2. Kujawsko-Pomorskie, 3. Lubelskie, 4. Lubuskie, 5. Łódzkie, 6. Małopolskie, 7. Mazowieckie, 8. Opolskie, 9. Podkarpackie, 10. Podlaskie, 11. Pomorskie, 12. Śląskie, 13. Świętokrzyskie, 14. Warmińsko-Mazurskie, 15. Wielkopolskie, 16. Zachodniopomorskie, 17. Average province, AP – anti-pattern, P – pattern

Fig. 1. Results of Multidimensional Scaling of 17 Objects in Terms of the Level of Social Cohesion in 2016: a) for interval-valued data, b) for metric data

Source: calculations performed using R software.

Table 1. Ranking of 16 Polish Provinces and the Average Province, Based on Metric and Interval-valued Data, According to the Level of Social Cohesion in 2016 (Values of Measure d_i)

Province	No.	Interval-valued approach		Classical (metric) approach		Δd_i
		d_i	rank	d_i	rank	
Śląskie	12	0.8473	1	0.5534	6	0.2939
Pomorskie	11	0.7420	2	0.5086	8	0.2334
Podkarpackie	9	0.7313	3	0.5755	4	0.1557
Opolskie	8	0.7082	4	0.6251	3	0.0831
Małopolskie	6	0.6856	5	0.6806	2	0.0050
Kujawsko-Pomorskie	2	0.6472	6	0.4042	13	0.2430
Average province	17	0.6150	7	0.5356	7	0.0793
Mazowieckie	7	0.5869	8	0.7269	1	–0.1400
Łódzkie	5	0.5784	9	0.5667	5	0.0117
Wielkopolskie	15	0.5766	10	0.4614	10	0.1152
Świętokrzyskie	13	0.5533	11	0.3934	14	0.1600
Lubelskie	3	0.5252	12	0.4507	11	0.0745
Warmińsko-Mazurskie	14	0.4806	13	0.2983	16	0.1823
Dolnośląskie	1	0.4565	14	0.4051	12	0.0514
Podlaskie	10	0.4524	15	0.4975	9	–0.0452
Zachodniopomorskie	16	0.4018	16	0.3534	15	0.0484
Lubuskie	4	0.1323	17	0.2570	17	–0.1246
Parameters		Value		Value		Difference
Mean		0.5718	×	0.4878	×	0.0840
Standard deviation		0.1591	×	0.1252	×	0.0339
Median		0.5784	×	0.4975	×	0.0809
Median absolute deviation		0.1589	×	0.1370	×	0.0219
Pearson correlation coefficient		0.6780				
Kendall rank correlation coefficient		0.4853				

Source: calculations performed using R software.

Table 1 presents the ranking of 17 objects (16 provinces and the average province) according to the level of social cohesion in 2016 obtained after applying the classical and interval-valued approach. All calculations were performed using the `clusterSim` package (Walesiak & Dudek 2018a) developed for R software (R Core Team 2018).

Based on the results of the study, it was possible to assess the level of social cohesion in Polish provinces. The assessment is based on two measurements: one involving metric data, and the other based on interval-valued data. This approach made it possible to demonstrate how the assessment of social cohesion changes when one moves from mono-parametric measurement (the classical approach) to interval measurement (the interval-valued approach). The consistency of the ranking of provinces according to the values of measure d_i (measured by Kendall rank correlation coefficient $r_{tau} = 0.4853$) is much lower than the correlation of provinces according to the values of measure d_i (measured by the Pearson correlation coefficient $r = 0.6780$). Hence, the ranking of provinces changed to a greater degree than did the values of measure d_i.

Figure 2 shows the spatial distribution of actual differences in measure d_i (Δd_i from Table 1) for Polish provinces between the results for interval-valued data and for metric data.

The proposed modification of the method made it possible to conduct a more in-depth assessment of the multidimensional phenomenon of social cohesion in comparison with the classical approach based on metric data. Provinces were not assessed merely on the basis of the variables' mean values, but also taking into account how the variables varied across districts. After applying the alternative method of measurement, the position of objects changed significantly:

1. The dispersion of objects (provinces) measured by standard deviation and median absolute deviation changed. If the Lubuskie province is excluded from the analysis (outlier No. 4 in Figure 1), the variability in measure d_i in both approaches is at a similar level: $S_{d_i} \approx 0.12$.

2. If the arrangement of provinces is assessed not only in terms of d_i values but also on the basis of their position relative to the set axis, one notices that the dispersion of objects measured using the interval-valued approach is considerably smaller. This is the result of eliminating the impact of extreme values of target variables.

3. The majority of objects (13 provinces and the average province) moved towards the pattern object (see Figure 1). The mean level of social cohesion and the median both increased. Provinces with the highest actual difference in measure d_i include: Śląskie, Pomorskie, and Kujawsko-Pomorskie (the last column in Table 1 and Figure 2). A reverse change – a shift towards the anti-pattern object – was observed in the case of another three provinces: Mazowieckie, Lubuskie, and Podlaskie. This means that the level of social

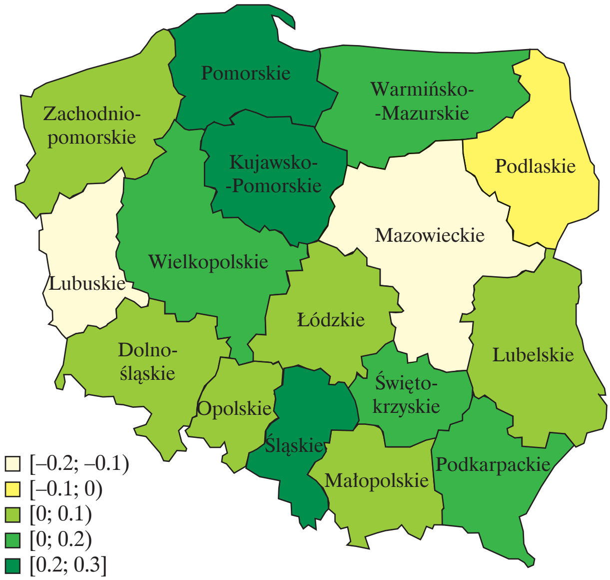


Fig. 2. Spatial Distribution of Actual Differences in Measure d_i (Δd_i) for Provinces of Poland

Source: calculations performed using R software.

cohesion measured using the modified approach is lower compared to the classical approach.

4. The mean, which is the only parameter used in the classical approach, is very sensitive to outliers. In the case of territorial units, the spatial poles of growth behave like extreme observations and can strongly influence the measurement for the entire region to which they belong. Examples of this phenomenon identified in the study involve measurements for the Mazowieckie province (7) strongly affected by the district of Warsaw, and the Lubuskie province (4) being influenced by the district of Zielona Góra. The switch from the classical approach, in which the provinces were assessed exclusively on the basis of the mean values of the target variables, to the interval-valued approach, which accounts for the inter-district variation in these variables, caused these two objects to shift towards the anti-pattern. In the case of Mazowieckie and Lubuskie, this shift, expressed in terms of negative values of the actual difference in measure d_i, was the biggest (see Figure 2).

5. Another change in the position of the objects is their location in relation to the set axis. The measurement based on interval-valued data resulted in 8 provinces shifting above the set axis. Only three provinces remained below the set axis (see Figure 1). This change is mainly due to the fact that the measurement accounted for the asymmetrical distributions of the variables of interest, which were skewed right for the majority of provinces. The provinces located above the set axis are the ones where such right-skewed variables prevailed. It should be added that when evaluating the arrangement of objects on both sides of the set axis, one takes into account not only the direction but also the degree of asymmetry. In the classical approach, such a detailed assessment that accounts for the asymmetry of distributions is not possible.

4. Conclusions

The measurement of the level of social cohesion is a complex task and requires a multi-dimensional approach. In order to rank objects of interest in terms of the level of social cohesion, the authors used a two-step research procedure (multidimensional scaling and linear ordering), which enabled the results to be visualized in a two-dimensional space. By analysing two approaches, it was possible to demonstrate how the assessment of social cohesion changes when one switches from mono-parametric measurement (metric data) to interval measurement (interval-valued data).

The proposed modification made it possible to assess social cohesion in provinces not only on the basis of the variables' mean values, but also by taking into account 50% of districts within each province (the interval between the first and third quartile).

This approach helps to eliminate the influence of outliers on the assessment of social cohesion in Polish provinces.

All the calculations were conducted using scripts written by the authors in R software.

Acknowledgement

The project is financed by the Polish National Science Centre, DEC-2015/17/B/HS4/00905.

Bibliography

Annoni, P. and Dijkstra, L. (2016) *The EU Regional Social Progress Index: Methodological Note*. Brussels: European Commission.

Balcerzak, A. P. (2015) "Wielowymiarowa analiza spójności społecznej w krajach Unii Europejskiej w kontekście strategii Europa 2020" in B. Bartniczak and K. Trzeciak (eds) *Aktualne trendy w zarządzaniu środowiskiem*. Jelenia Góra: Wydawnictwo AD REM.

Billard, L. and Diday, E. (2006) *Symbolic Data Analysis: Conceptual Statistics and Data Mining*. Chichester: John Wiley.

Borg, I. and Groenen, P. J. F. (2005) *Modern Multidimensional Scaling: Theory and Applications*. New York: Springer Science+Business Media.

Bottoni, G. (2018) "A Multilevel Measurement Model of Social Cohesion". *Social Indicators Research* 136 (3): 835–57, https://doi.org/10.1007/s11205-016-1470-7.

Concerted Development of Social Cohesion Indicators: Methodological Guide (2005) Strasbourg: Council of Europe Publishing.

Dickes, P. and Valentova, M. (2013) "Construction, Validation and Application of the Measurement of Social Cohesion in 47 European Countries and Regions". *Social Indicators Research* 113 (3): 827–46, https://doi.org/10.1007/s11205-012-0116-7.

Dickes, P., Valentova, M. and Borsenberger, M. (2010) "Construct Validation and Application of a Common Measure of Social Cohesion in 33 European Countries". *Social Indicators Research* 98 (3): 451–73, https://doi.org/10.1007/s11205-009-9551-5.

Everitt, B.S., Landau, S., Leese, M. and Stahl, D. (2011) *Cluster Analysis*. Chichester: Wiley.

Hellwig, Z. (1972) "Procedure of Evaluating High-level Manpower Data and Typology of Countries by Means of the Taxonomic Method" in Z. Gostkowski (ed.) *Towards a System of Human Resources Indicators for Less Developed Countries: Papers Prepared for UNESCO Research Project*. Wrocław: Ossolineum, The Polish Academy of Sciences Press.

Hellwig, Z. (1981) "Wielowymiarowa analiza porównawcza i jej zastosowanie w badaniach wielocechowych obiektów gospodarczych" in W. Welfe (ed.) *Metody i modele ekonomiczno-matematyczne w doskonaleniu zarządzania gospodarką socjalistyczną*. Warszawa: PWE.

Ichino, M. and Yaguchi, H. (1994) "Generalized Minkowski Metrics for Mixed Feature--type Data Analysis". *IEEE Transactions on Systems, Man, and Cybernetics* 24 (4): 698708, https://doi.org/10.1109/21.286391.

Jajuga, K., Walesiak, M. and Bąk, A. (2003) "On the General Distance Measure" in M. Schwaiger and O. Opitz (eds) *Exploratory Data Analysis in Empirical Research*. Berlin–Heidelberg: Springer-Verlag.

Mair, P., De Leeuw, J., Borg, I. and Groenen, P. J. F. (2018) smacof: Multidimensional Scaling. R package, version 1.10-8, https://CRAN.R-project.org/package=smacof.

My Region, My Europe, Our Future: Seventh Report on Economic, Social and Territorial Cohesion (2017). Brussels: European Commission.

Rajulton, F., Ravanera, Z. R. and Beaujot, R. (2007) "Measuring Social Cohesion: An Experiment Using the Canadian National Survey of Giving, Volunteering, and Participating". *Social Indicators Research* 80 (3): 461–92, https://doi.org/10.1007/s11205-006-0011-1.

R Core Team (2018) *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. Vienna, Austria, https://www.R-project.org.

Walesiak, M. (2016) "Visualization of Linear Ordering Results for Metric Data with the Application of Multidimensional Scaling". *Ekonometria* 2 (52): 9–21, https://doi.org/10.15611/ekt.2016.2.01.

Walesiak, M. and Dudek, A. (2017) "Selecting the Optimal Multidimensional Scaling Procedure for Metric Data with R Environment". *Statistics in Transition: New Series* 18 (3): 521–40, https://doi.org/10.21307/stattrans-2016-084.

Walesiak, M. and Dudek, A. (2018a) clusterSim: Searching for Optimal Clustering Procedure for a Data Set. R package, version 0.47-2, https://CRAN.R-project.org/package=clusterSim.

Walesiak, M. and Dudek, A. (2018b) mdsOpt: Searching for Optimal MDS Procedure for Metric and Interval-valued Symbolic Data. R package, version 0.3-2, https://CRAN.R-project.org/package=mdsOpt.

Walesiak, M. and Dehnel, G. (2018) "Evaluation of Economic Efficiency of Small Manufacturing Enterprises in Districts of Wielkopolska Province Using Interval-valued Symbolic Data and the Hybrid Approach" in M. Papież and S. Śmiech (eds) *The 12th Professor Aleksander Zeliaś International Conference on Modelling and Forecasting of Socio-economic Phenomena. Conference Proceedings*. Cracow: Foundation of the Cracow University of Economics.

Abstract

Analiza porównawcza uporządkowania województw Polski ze względu na spójność społeczną

Ocenę spójności społecznej województw Polski przeprowadzono na podstawie klasycznych danych metrycznych oraz symbolicznych interwałowych z wykorzystaniem podejścia hybrydowego łączącego zastosowanie skalowania wielowymiarowego z porządkowaniem liniowym. W pierwszym kroku w wyniku zastosowania skalowania wielowymiarowego otrzymano wizualizację badanych obiektów w przestrzeni dwuwymiarowej. Następnie przeprowadzono porządkowanie liniowe zbioru obiektów na podstawie odległości Euklidesa od wzorca rozwoju. Zmienne symboliczne interwałowe opisują badane obiekty precyzyjniej niż metryczne dane klasyczne. Dane klasyczne mają charakter atomowy. Obserwacja na każdej zmiennej wyrażona jest w postaci jednej liczby rzeczywistej, z kolei dla zmiennych symbolicznych interwałowych obserwacja na każdej zmiennej ujęta jest w postaci przedziału liczbowego. W celu otrzymania danych symbolicznych interwałowych zastosowano dwustopniowe gromadzenie danych. Najpierw zgromadzono dane klasyczne dotyczące spójności społecznej według powiatów Polski, a następnie poddano je agregacji do poziomu województw, otrzymując dane symboliczne interwałowe. W artykule przeprowadzono analizę porównawczą wyników badania spójności społecznej województw Polski uzyskanych na podstawie klasycznych danych metrycznych oraz danych symbolicznych interwałowych.

Słowa kluczowe: spójność społeczna, miary agregatowe, zmienne symboliczne interwałowe, skalowanie wielowymiarowe, program R.