Non-model methods in the study of regional development – the impact of the aggregation formula on the obtained research results

Rafał Czyżycki
University of Szczecin
ORCID: 0000-0003-4828-1887, e-mail: rafal.czyzycki@wzieu.pl

DOI: 10.14595/CP/01/014

Abstract: The study of the level of socio-economic development is one of the most frequently undertaken scientific considerations in the field of research into regions. Since the very concept of "socio-economic development" is currently not clearly defined and, hence, it is not possible to measure it directly, various attempts are being made to measure it indirectly using methods of multidimensional comparative analysis. The main goal of the article is to assess the impact of the chosen method for obtaining a synthetic measure of development, which is part of the non-model methods of aggregation of diagnostic variables, on the final result of orderingregions in terms of socio-economic development. Meanwhile, the considerations are accompanied by the following research hypothesis: *one of the factors significantly affecting the final result of ordering the regions in terms of socio-economic development is the choice of method used to obtain a synthetic measure of development*. As a result of the research, this hypothesis was confirmed, and proposals aimed at increasing the objectivity of this type of research were also indicated.

Keywords: regional development, non-model methods, synthetic development measure

JEL: C38, O18, R11

Introduction

Issues related to regional development, treated as the development of states, regions, districts, municipalities or otherwise defined relatively homogeneous areas, varying from areas adjacent in terms of specific natural or acquired features [Potoczna, 2006, pp. 86-92], are the subject of many scientific considerations. The analysis of information obtained on the basis of the *Publish or Perish* programme shows that only from the beginning of 2010 to the end of 2017, 120 scientific papers were published whose titles included the term "regional development", 108 papers with "development of regions" in the title and 114 works which used the term "socio-economic development" in their titles. These works, however, have a very diverse cognitive nature, resulting, inter alia, from the fact that the character, dynamics, direction or structure of processes related to regional development are conditioned by a number of factors[1] that influence the phenomena occurring in a region to a different extent [Korenik, 2004, pp 107-113]. The regional development process itself has a multidimensional, extremely heterogeneous character, which means that not only the method of its measurement remains ambiguous [Klóska, 2012, p. 127], but also the grasp of it and description are extremely difficult and lead, as a consequence, to the adoption of various simplifying assumptions [Kosiedowski et al., 2001, p 28]. As a result, research work in the field of regional development focuses in practice either on qualitative methods, or on quantitative methods. Qualitative methods form the basis for describing economic events or determining causal relationships between events and factors affecting these events. In contrast, quantitative methods, using a rich set of mathematical and statistical methods and techniques, enable the quantification of development measures, the measurement of structures and their transformations, as well as the inventory of resources and criteria for their allocation. Quantitative methods also enable the construction of formalised econometric models, including spatial cross-sections (national, regional, micro-regional and zonal) [Kozubek, 1999, pp. 63-71].

The main goal of the article is to assess the impact of the chosen method of obtaining a synthetic measure of development, which is part of the non-model methods of aggregation of diagnostic variables, on the final result of ordering the regions in terms of socio-economic development. It will make it possible to verify the following research hypothesis adopted in the work: *one of the factors significantly affecting the final result of ordering the regions in terms of socio-economic development is the choice of method used to obtain a synthetic measure of development.*

Research methodology

Socio-economic development is a term in the field of multidimensional statistics, directly immeasurable, however describable by a number of diagnostic variables, essentially related to this concept. The ordering of the examined objects from "best" to "worst" is based on the value describing particular objects, obtained from the function aggregating the information contained in the adopted diagnostic variables. The ordering of such a set of objects requires meeting the following assumptions [Walesiak, 1996, p. 125]:

- the set of objects is a non-empty and finite set;

[1] At the subsequent stages of socio-economic development, competition issues also come to the fore, to which particular attention should be devoted, and whose implementation has an impact on the development of the economy. (More in: on Drab-Kurowska, 2013, pp. 501-511).

- there is a primary, synthetic criterion of ordering the elements of this set, which is not subject to direct measurement (e.g. the level of development of the studied region in comparison with other regions);
- a finite set of variables is given, essentially related to the synthetic ordering criterion;
- variables used to describe objects are presented at least on an ordinal scale, meet the demand for uniform preference and are brought to comparability through normalisation;
- the relation that organises the elements of set A is the relation of the majority concerning the numerical values of the synthetic measure of development.

The final research results are determined mainly by the final list of diagnostic variables[2], as well as the selection of the aggregation formula. Due to the existence of many formulas of normalisation of variables, methods of determining weights, or methods of averaging normalised values, a number of different aggregate measures have been described in the scientific literature (used in practice, among others, in the preparation of various types of rankings). However, it should be noted that different aggregation formulas may give different final results, even with respect to the general criterion represented by the same list of diagnostic variables [Czyżycki, 2012, pp. 15-22]. However, the final list of variables included in the study is of key importance, so it should be discussed among experts and recognised as the best representative of the analysed issue. A broad review of literature in this respect was made by R. Klóska [2015, pp. 99-108] who, on the basis of in-depth research, offered 18 indicators, with the help of which it is possible to study regional development in three dimensions:

- from a social perspective: number of infant deaths per 1,000 live births (X_1), at-risk-of-poverty rate ($X_{2\ -\%}$), number of students of tertiary education institutions per 10 000 inhabitants (X_3), registered unemployment rate ($X_{4\ -\%}$), number of road fatalities per 100 000 inhabitants (X_5) and total water consumption for the needs of the national economy and population (hm^3) per 10 000 inhabitants (X_6);
- from an economic perspective: GDP (current prices) per capita (X_7), share of economy entities' financial outlays in the total outlays on research and development activities (X_8 –%), number of newly registered national economy entities in the private sector per 10 000 inhabitants (X_9), number of employees per 1000 inhabitants (X_{10}) and total investments (current prices) per capita in PLN (X_{11});

[2] The set of diagnostic variables substantially related to socio-economic development can be very extensive, taking into account, for example, changes occurring and related to the development of the Internet. M. Czaplewski presents the topic of the impact on the economy of information and communication technologies, including the Internet (Czaplewski, 2011, pp. 20-26).

- from an environmental perspective: percentage of the population using sewage treatment plants (X_{12} -%), forest cover (X_{13} –%), recycling of packaging waste (X_{14} –%), share of devastated and degraded lands requiring reclamation in the total area (X_{15} -%), share of waste (excluding municipal waste) recovered in the amount of waste generated during a year (X_{16} –%), share of the electricity generation from renewable energy sources in the total electricity generation (X_{17} –%) and electricity consumption per 1 million PLN GDP (X_{18} – GWh).

After selecting diagnostic variables, substantially related to the primary criterion according to which the objects (regions) will be ordered, the structure of the synthetic development measure boils down to the following stages:

- unification of the nature of variables subject to aggregation by means of the postulate of uniform preferences of variables, removal of titles from the values of variables, and unification of orders of magnitude in order to bring them to comparability [for more details see: Perkal, 1953, pp. 209-219, Hellwig, 1968, pp. 307-326; Bartosiewicz, 1976, pp. 307-326; Strahl, 1978, pp. 5-7, pp. 205-215; Walesiak, 2014, pp. 363-372];
- weighing standardised diagnostic features, i.e. assigning to individual variables weights defining their significance for the general criterion in comparison with other features;
- selecting an aggregation formula and, based on it, designating the synthetic development measure. There are two types of synthetic variable determination procedures in the literature: model and non-model [Grabiński, 1984, p. 38]. Model methods of aggregation of variables are based on determining distances of individual objects from a certain, defined model object, whereas non-model methods rely on the operation of averaging the values of normalised variables.

In order to bring the proposed diagnostic variables to comparability, the method of zero unitarisation will be used, which in the case of boosters (variables X_2, X_7, X_8, X_9, X_{10}, X_{11}, X_{12}, X_{13}, X_{14}, X_{16} and X_{17}) consists in applying a formula in the form [Kukuła, 2000, p. 226]:

(1)
$$z_{ij} = \frac{x_{ij} - \min_i\{x_{ij}\}}{\max_i\{x_{ij}\} - \min_i\{x_{ij}\}}$$

whereas in the case of inhibitors (X_1, X_3, X_4, X_5, X_6, X_{15}, X_{18}) it is based on the formula:

2)
$$z_{ij} = \frac{\max_i\{x_{ij}\} - x_{ij}}{\max_i\{x_{ij}\} - \min_i\{x_{ij}\}}$$

The normalised variables obtained on the basis of the above formulas are characterised by the adoption of values between <0,1>, where, from the point of view of the general criterion, higher values of normalised variables obtained indicate a higher level of socio-economic development of a given region. Finally, the measure of the development of a given region will be the aggregated value of all normalised diagnostic variables adopted in the study, i.e. the method of standardised sums will be applied, defined as:

$$p_i = \sum_{j=1}^{m} w_j \cdot z_{ij} \tag{3}$$

where: w_j is the weight determining the impact of a given j variable on the adopted general criterion. The determination of individual weights takes place either on the basis of expert opinions, or through the use of specific statistical tools. In the literature on the subject, however, it is recommended that, in the absence of unambiguous indications as to the different meanings and roles of particular features, it is to be silently assumed that all selected diagnostic variables are of the same weight [Kukuła, 2000, p. 64]. This assumption will also be adopted in the article.

Due to the properties of normalised variables, the $p_i$value obtained is normalised in the range <0, m>, where m is the number of diagnostic variables. Using the method of standardised sums in research, a postulate to normalise the final results in the interval <0.1> often appears. An example of such a procedure can be found, among others in J. Dziechciarz [2003, pp. 290-291] and amounts to determining a standard measure of development according to the formula:

$$m_i = \frac{p_i - p_{\_0}}{p_0 - p_{\_0}} \tag{4}$$

where:

$$p_0 = \sum_{j=1}^{m} w_j \cdot z_{0j} \tag{5}$$

$$p_{\_0} = \sum_{j=1}^{m} w_j \cdot z_{\_0j} \tag{6}$$

$$z_{0j} = \max_i z_{ij} \tag{7}$$

$$z_{\_0j} = \min_i z_{ij} \tag{8}$$

The determined m_i measure, irrespective of the previously used method for normalising diagnostic variables, will always take values from the expected interval <0.1>, whereas in the case of the zero unitarisation method proposed in the article, a faster way to obtain a development measure is to calculate the average value of standardised variables, i.e. to determine the value:

(9) $$u_i^1 = \frac{1}{m}\sum_{j=1}^{m} w_j \cdot z_{ij}$$

Among non-model methods, the alternative approaches in the study of the socio-economic development of regions are either the use of an absolute measure of development, defined as [Żmurkow-Poteralska, 2015, p. 187]:

(10) $$u_i^2 = \sum_{j=1}^{m} w_j \cdot z_{ij}^*$$

where: the variable z_{ij}^* is a variable normalised according to the formula:

(11) $$z_{ij}^* = \frac{x_{ij}}{S_j}$$

where: S_j is the standard deviation of the j-th diagnostic variable, or the use of the rank method, which consists in assigning each diagnostic variable an appropriate rank depending on the value of this variable in a given object. In a situation where in two or more objects a given variable assumes the same value, these objects are assigned the same rank, being the arithmetic mean of the subsequent ranks. The measure of development is the arithmetic mean of the ranks assigned to a given object for each diagnostic variable:

(12) $$u_i^3 = \frac{1}{m}\sum_{j=1}^{m} l_{ij}$$

In order to assess the degree of compliance of the received rankings, appropriate correlation coefficients can be used. From the statistical point of view, the values in the selected ranking are the values of the measurable characteristic on the ordinal scale, and this means that statistical measures used to study the interdependence of places in particular rankings, which are often used in this type of research are, among others, Spearman's rank correlation coefficient or the tau-Kendall coefficient (τ). However, because the Spearman coefficient is a derivative of Pearson's linear correlation coefficient and inherits its properties (sensitivity to outliers or lack of normality of distribution of variables) [Kuszewski and Sielska, 2010, p. 156], it is postulated that, instead of the Spearman coefficient, only the tau-Kendall coefficient be used while examining the degree of rankings compliance [Stanisz, 2006, p. 337]. This coefficient takes values from the interval <-1, 1>, where the value 1 indicates full compliance, value 0 indicates the lack of compliance of orderings, while the value -1 indicates their total contradiction. In order to verify the hypothesis about the compliance of the obtained rankings, the test of significance of the tau-Kendall coefficient is used, for which the test statistic defined as:

(13)
$$Z_\tau = \frac{\tau}{\sqrt{\frac{2(2n+5)}{9n(n-1)}}}$$

for n> 10 has an asymptotically normal distribution [Abdi, 2007] (in the case of the study of the socio-economic development of regions in Poland n = 16).

Findings

Using the statistic portal strateg.stat.gov.pl, information on shaping the eighteen diagnostic variables proposed in the article was collected for all regions in Poland. At the same time, it was assumed that socio-economic development would be analysed at the end of 2016 and, in the case of thirteen variables, their values in individual regions, adopted for research, come from that year, while in the case of Gross Domestic Product per capita (X_7), the share of expenditure on R & D. More on the role of R & D: [Budziewicz-Guźlecka, 2014, pp. 9-17] financed from the enterprise sector, in R & D expenditure in total (X_8) and electricity consumption per 1 million GDP – their values from 2015 were taken into account; in the case of recycling of packaging waste (X_{14}) – from 2014, whereas in relation to the share of waste (excluding municipal waste) subjected to recycling in the amount of waste generated during the year (X_{16}), the value taken into consideration was from 2013. For the above five variables, the indicated years were the last for which, at the time of the research, the Central Statistical Office provided information on the value of these variables in individual regions.

Analysing the obtained results characterising the level of socio-economic development of individual regions in Poland in the adopted research period, attention should be paid to the very high correlation between the positions of individual regions resulting from the use of the method of average values of standardised variables (u1) and the rank method (u3) and completely different results obtained when using the absolute development measure (u2) (see table1). In the case of rankings received on the basis of u1 and u3, the maximum difference in the positions occupied concerned the Łódzkie region, for which the rank method indicates the sixth position in terms of socio-economic development in 2016, while the method of average values of normalised variables places this region on the ninth position. The change of two positions in the received rankings can be noticed in the case of the Western Pomeranian and Lublin voivodships, while in the case of nine regions, the positions in the ranking are the same regardless of the method of obtaining the value of the synthetic development measure.

However, the results obtained on the basis of the method of the absolute measure of development suggest a completely different ordering of regions in 2016 in terms of the general criterion adopted. No region occupies the position indicated by the previously discussed methods, what is more, in Masovian voivodship the difference in positions occupied in the rankings amounts to 13 places, and in the case of Warmian-Masurian, Podlasie and Silesian voivodships, differences in rankings ranged from ten to twelve places.

Table 1. The values of the synthetic development measure obtained by the means of the average values of standardised variables method (u11), the absolute measure of development (u2) and the rank method (u3) together with the information on the position of individual regions in the ranking due to the level of socio-economic development in 2016 year

	u_1	position in the ranking	u_2	position in the ranking	u_3	position in the ranking
Lower Silesia	0,4994	7	68,5628	9	144,5	7
Kuyavian-Pomeranin	0,4667	10	62,1353	13	159	10
Lublin	0,3829	12	65,9051	11	192,5	14
Lubusz	0,3895	11	68,7060	8	164,5	11
Łodz	0,4808	9	79,0625	3	144,5	6
Lesser Poland	0,6383	3	88,4474	1	98,5	2
Masovian	0,7796	1	61,5923	14	89,5	1
Opole	0,3566	13	70,2926	6	189	13
Subcarpathian	0,4885	8	62,9894	12	155	9
Podlasie	0,3483	14	79,4964	2	184	12
Pomeranian	0,6517	2	72,9056	5	102,5	3
Silesian	0,5178	5	59,0087	15	134	5
Świętokrzyskie	0,2902	15	58,1770	16	203	15
Warmian-Masurian	0,2793	16	77,5973	4	222	16
Greater Poland	0,6205	4	69,9172	7	113,5	4
Western Pomeranian	0,5044	6	66,1597	10	152	8

Source: own calculations and elaboration.

The large convergence of rankings obtained on the basis of the method of average values of standardised variables and the rank method, as well as the different results obtained in the case of the absolute development measure method, is also indicated by the analysis of the value of the tau-Kendall coefficient (see table 2). On this basis, one can clearly indicate a fairly strong, positive and, most importantly, statistically significant convergence of ordering of the regions examined in terms of the general criterion adopted in the case of applying u1 and u2 measures for this purpose, and the lack of such convergence in the case of the u3 measure.

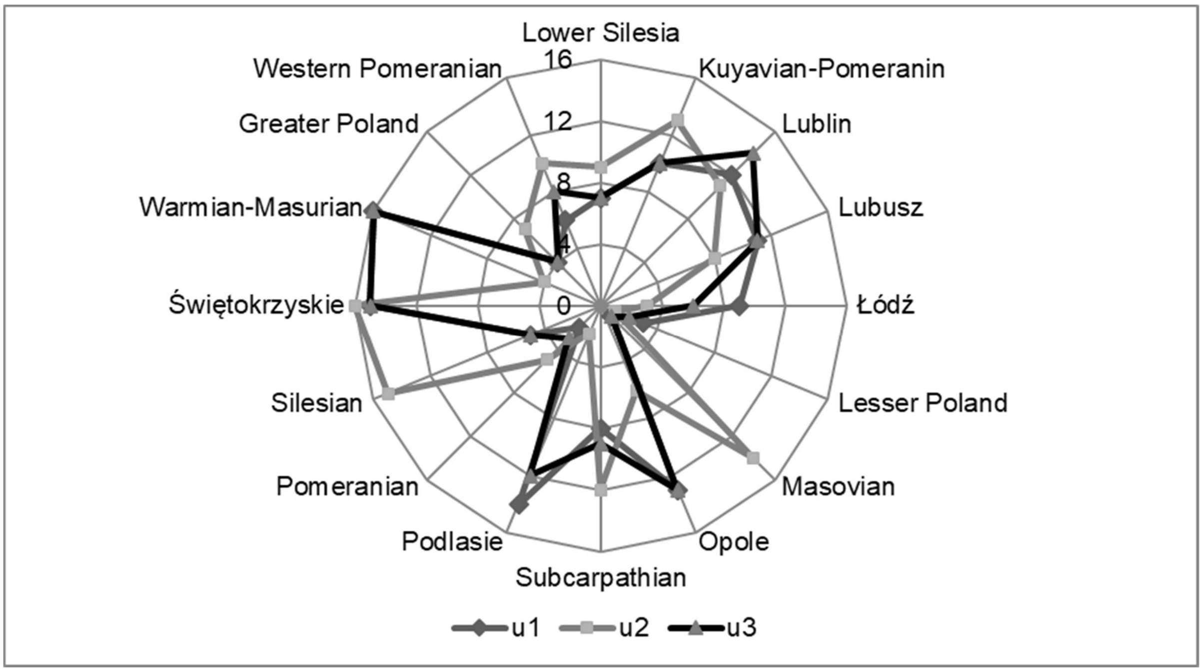
Table 2. Values of the tau-Kendall coefficient examining the compliance of the rankings obtained on the basis of the u1, u2 and u3 measures (values over the main diagonal) together with the ***p-value*** for the test examining the statistical significance of the obtained tau-Kendall coefficient (under the main diagonal)

	u_1	u_2	u_3
u_1	x	0,8667	-0,0500
u_2	2,84E-06	x	0,0833
u_3	0,7871	0,6525	x

Source: own calculations and elaboration.

Also, a graphical comparison of the ordering results of individual regions in 2016, obtained on the basis of the three non-model methods indicated in the article, shows the previously discussed relationships between the obtained results (see Figure 1).

Fig. 1. Positions of Polish regions due to the level of socio-economic development in 2016



Source: own study based on table1.

Summary

On the basis of the conducted studies, it seems justified to draw the following conclusions:

1. Due to the demonstrated dependence between the results of ordering the regions from the adopted method of obtaining a synthetic measure of development, in this type of research, not only the final results of the research should be presented, but also the research methodology should be presented in a fairly detailed way. This will allow, on the one hand, for repeating the conducted analyses and possibly verifying their correctness, and on the other hand, which seems more important, explaining possible differences in the rankings of objects obtained by different researchers, even if they include the same diagnostic variables;
2. In the case of the study of the socio-economic development of regions, the factor “objectifying” the final results may be basing them on the results obtained from several different methods of multidimensional comparative analysis (taking into account the same set of diagnostic variables each time). The final ordering of objects could be based either on the average values of the ranking position of the studied regions from all methods included in the study, or on average values of only those rankings that would be statistically convergent with each other (in this case, such convergence could be determined on the basis of the analysis of the significance of the Kendall-tau coefficient). Considering the latter approach, the convergent approaches in the study presented in the article were approaches based on the method of average values of standardised variables and the rank method. By averaging the results obtained with these methods, one could point out that the best in terms of socio-economic development in 2016 was the Masovian voivodship (in both approaches it took first place), then *ex aequo* Pomeranian and Lesser Poland voivodships (depending on the approach, they occupied second or third place in the rankings), Greater Poland voivodship (which took the fourth position twice) and the Silesian voivodship (fifth position twice).

References

ABDI, H. (2007). *The Kendall rank correlation coefficient*, Encyclopedia of Measurement and Statistics. Sage, Thousand Oaks, CA, pp. 508-510.

BARTOSIEWICZ, S. (1976). *Propozycja metody tworzenia zmiennych syntetycznych*, Prace Naukowe AE we Wrocławiu, 84, pp. 5-7.

BUDZIEWICZ-GUŹLECKA, A. (2012). *Rola działalności naukowo-badawczej w rozwoju gospodarki opartej na wiedzy*, Zeszyty Naukowe Uniwersytetu Szczecińskiego. Ekonomiczne Problemy Usług. T. 1 Nowa gospodarka-spojrzenie interdyscyplinarne, 112, pp. 9-17.

CZAPLEWSKI, M. (2011). *Dobra informacyjne i ich podstawowe cechy ekonomiczne*. Wydawnictwo Naukowe Uniwersytetu Szczecińskiego.

CZYŻYCKI, R. (2012). *Badanie rozwoju społeczno-gospodarczego województw – wpływ metodyki badań na uzyskane wyniki*. Wydawnictwo Wyższej Szkoły Bankowej w Poznaniu, 42, pp. 15-22.

DRAB-KUROWSKA, A. (2013). *Polityka konkurencji na rynku e-commerce*, Zeszyty Naukowe Uniwersytetu Szczecińskiego. Ekonomiczne Problemy Usług. Europejska przestrzeń komunikacji elektronicznej, 104, pp. 501-511.

DZIECHCIARZ, J. (2003). *Ekonometria: metody, przykłady, zadania*. Wydaw. Akademii Ekonomicznej im. Oskara Langego.

GRABIŃSKI, T. (1984). *Wielowymiarowa analiza porównawcza w badaniach dynamiki zjawisk ekonomicznych*, Zeszyty Naukowe, Akademia Ekonomiczna w Krakowie. Seria Specjalna, Monografie, (61).

HELLWIG, Z. (1968). *Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę wykwalifikowanych kadr*, Przegląd statystyczny, 4(1968), pp. 307-326.

KLÓSKA, R. (2012). *Statystyczna analiza poziomu rozwoju społeczno-gospodarczego w Polsce w ujęciu regionalnym. Przestrzeń w nowych realiach gospodarczych*, Zeszyty Naukowe Wyższej Szkoły Bankowej w Poznaniu, 42, pp. 127-136.

KLÓSKA, R. (2015). *Innowacyjność jako determinanta rozwoju regionalnego w Polsce*. Wydawnictwo Naukowe Uniwersytetu Szczecińskiego.

KORENIK, S. (2004). *Rozwój regionalny – nowe tendencje*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu, (1016) Gospodarka – Przestrzeń – Rozwój, pp. 107-113.

KOSIEDOWSKI, W. i in. (2001), *Zarządzanie rozwojem regionalnym i lokalnym: problemy teorii i praktyki*. Towarzystwo Naukowe Organizacji i Kierownictwa Dom Organizatora.

KOZUBEK, Z. (1999). *Etapy i procedury budowy strategii rozwoju regionalnego*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu, (799), pp. 63-71.

KUKUŁA, K. (2000). *Metoda unitaryzacji zerowanej*. Wydaw. Naukowe PWN.

KUSZEWSKI, T., SIELSKA, A. (2010). *Użyteczna sztuka rankingów ekonomiczno-społecznych*, Współczesna Ekonomia, 1(13), pp. 143-162.

PERKAL, J. (1953). *O wskaźnikach antropologicznych*, Przegląd Antropologiczny, 19, pp. 209-219.

POTOCZNA, I. (2006). *Polska polityka regionalna na początku XXI wieku – aspekt teoretyczny*, Biblioteka regionalisty, 1(3), pp. 86-92.

STANISZ, A. (2006). *Przystępny kurs statystyki*, T. 1, Statsoft, Kraków.

STRAHL, D. (1978). *Propozycja konstrukcji miary syntetycznej*, Przegląd Statystyczny, 25(2), pp. 205-215.

WALESIAK, M. (1996). *Metody analizy danych marketingowych*. Wydaw. Naukowe PWN.

WALESIAK, M. (2014). *Przegląd formuł normalizacji wartości zmiennych oraz ich własności w statystycznej analizie wielowymiarowej*, Przegląd Statystyczny. Statistic and Econometrics Committee, Polish Academy of Sciences, 61(4), pp. 363-372.

ŻMURKOW-POTERALSKA, E. (2015). *Regional Innovativeness and Socio-Economic Development*, Folia Oeconomica, 5(307).