STATISTICS IN TRANSITION new series, June 2017
Vol. 18, No. 2, pp. 271–290, DOI: 10.21307/stattrans-2016-070

A MULTIDIMENSIONAL AND DYNAMISED CLASSIFICATION OF POLISH PROVINCES BASED ON SELECTED FEATURES OF HIGHER EDUCATION IN 2002–2013

Wojciech Łukaszonek[1]

ABSTRACT

For close to two decades after the fall of communism in 1989, Polish higher education enjoyed an unprecedented period of development. Favourable political, economic, social and demographic changes led to a fivefold increase in the number of students and the number of higher educational institutions. The dynamic changes and their effects did not occur uniformly, in either space or time. An attempt is made here to identify and analyse the regional differentiation between Polish provinces in terms of features relating to higher education. To investigate the changes in higher education in the period of economic and social transformation, observations were made of fundamental characteristics of higher education in the years 2002–2013. The applied procedure uses new statistical methods applicable to a space of doubly multivariate data. The covariance matrix used to construct principal components is given the structure of a Kronecker product. The results led to the identification of six groups of provinces, including two consisting of a single province – Mazowieckie and Małopolskie provinces – which contain the largest and the highest-ranked[2] higher educational institutions in Poland: the University of Warsaw and Jagiellonian University.

Key words: higher education, doubly multivariate data, cluster analysis, dendrite method, covariance matrix, Kronecker product.

1. Introduction

The systemic changes that took place in Poland in the late 1980s and early 1990s led to transformation of the country in many different areas, including economic, social and cultural ones (Golinowska, ed. 2005). The process of changes in the 1990s also affected education, and its effects made a strong

[1] Faculty of Management, President Wojciechowski Higher Vocational State School, Kalisz, Poland. E-mail: w.lukaszonek@g.pl.

[2] Rankings by *Perspektywy* and *Rzeczpospolita/Gazeta Prawna* (2002–2015); Shanghai Ranking (2010–2015).

impression on the system of higher education. The economic reforms forced changes in the labour market, which led to increased needs for highly qualified staff. One consequence was a rapid growth in the number of people entering higher education. A degree seemed to be a guarantee of well-paid work, the possibility of further development, economic independence and improved social status (Sikorska 1998, Mach 2003). It had previously been an elite attribute, as reflected in the number of graduates in the population. In the centrally planned Polish economy of the 1970s and 1980s, the higher education system was closely controlled by the authorities, and student numbers were centrally regulated (Wnuk-Lipińska 1996, Antonowicz 2012, Kwiek 2014). In 1990 the percentage of the Polish population holding degrees was approximately 6%, this being a result of the policy applied in previous years. The adoption of democratic principles, giving more freedom to citizens, had a strong impact on social behaviours. There was an increase in Poles' educational aspirations, linked to the economic changes that were reflected in the dynamic expansion of the private sector (Ziółkowski 2000, Kwiek 2014). The increased demand for employment was accompanied by demographic changes, manifested in an increase in the population aged 19–24, the time at which higher education is undertaken. In effect, the number of students increased extremely rapidly. Over the years 1990–2005 the total number increased almost fivefold, from approximately 400,000 to almost 2 million (Fig. 1).

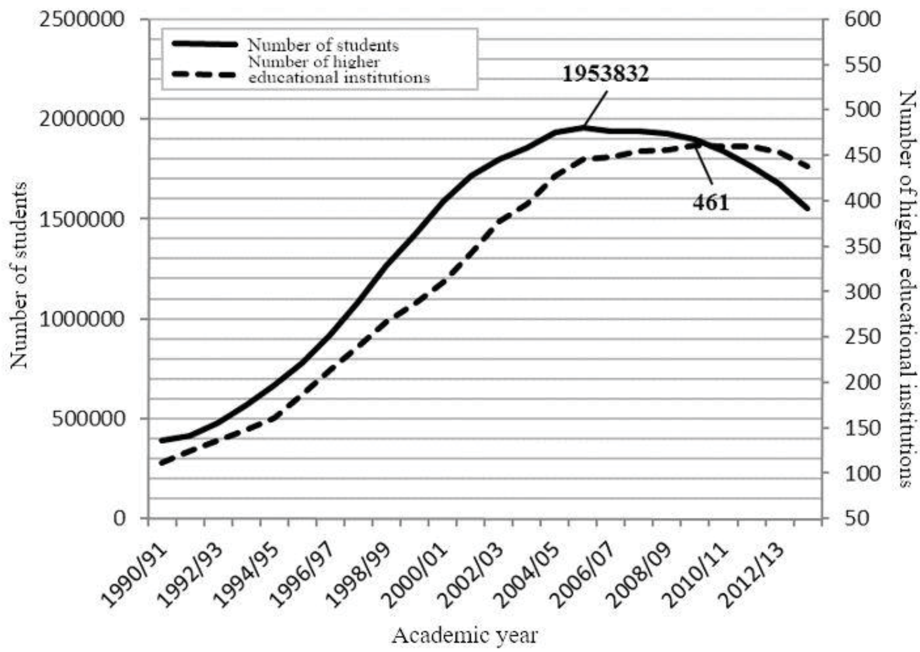

Figure 1. Numbers of students and of higher educational institutions in 1990–2013

Source: based on Higher Education 2014, GUS 2015.

The universities and colleges then existing in Poland were not prepared (particularly in terms of infrastructure) for such a rapid rise in the number of people interested in studying. In 1990 there were 112 higher educational institutions, all of them – apart from the Catholic University of Lublin – being state-run (public) institutions. The free market principles adopted at that time, and the *Act on higher education* (Dz.U. 1990 no. 65 item 385), enabled the foundation of Poland's first private higher educational institutions. The requirements for such an establishment were very liberal and relatively easy to fulfil, which gave an impetus to an unprecedented, and in effect uncontrolled, process of privatisation of higher education (Kwiek 2014). In the 15 years following the fall of communism, a total of 315 non-public (private) higher educational institutions were established in Poland, filling the gap in the market that had arisen due to the increasing public desire to study and the inability of state institutions to meet that need (Misztal 2000, Wasielewski 2013). From 1990 to 2010 the number of higher educational institutions in Poland increased almost fivefold (similarly to the growth in the number of students; Fig. 1).

The educational boom of that period naturally led to an increase in numbers of graduates. The net index of student numbers[3] (among persons aged 19–24) increased from 9.8% in 1990 to 40.8% in 2010, which is in agreement with the growth in the number of students (Fig. 2). Since 2005 the index has remained above 38%, one of the highest values among the countries of the European Union (GUS 2009, 2012, 2015).

[3] On the GUS website, the net index of student numbers is defined as the ratio of the number of persons (in a given age group) in a given level of education at the start of the school year to the total population (at 31 December) in the age group corresponding to that level of education. For example, the net index for the primary school level is calculated by dividing the number of primary school pupils aged 7–12 (the age assigned to that level) at the start of a given school year by the total population aged 7–12 at 31 December of the same year. The result is given as a percentage.

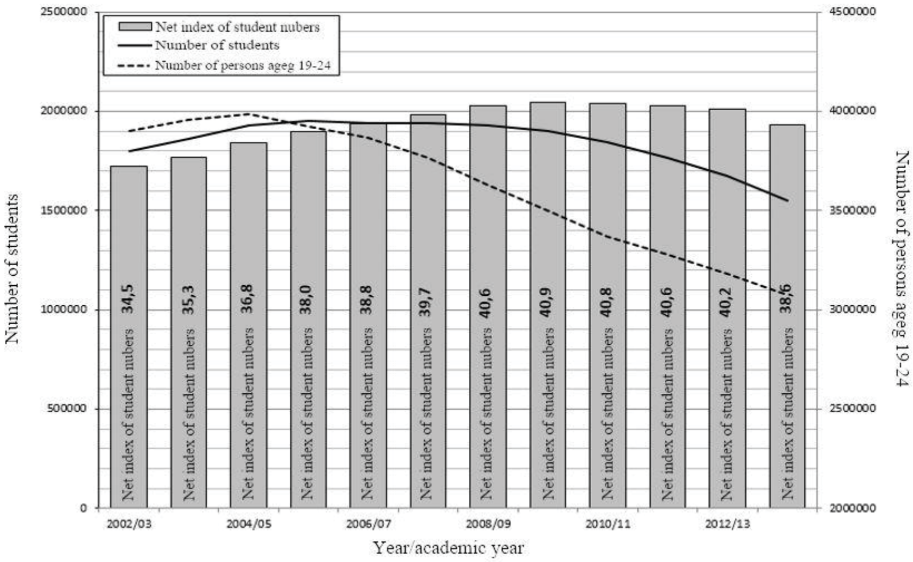

Figure 2. Number of students, number of persons aged 19–24, net index of student numbers, 2002–2013

Source: based on Higher Education 2014, GUS 2015.

The 1990s, which introduced democratic norms into social and political life, and free market values into the economy, brought to light social inequalities (not previously noticed in the post-war period) and the phenomena causing them (including unemployment, educational levels inadequate to the needs of the labour market, physical disability, and other factors). The period saw a growth in the importance of statutory measures aimed at counteracting social exclusion. One such decision was the *Act on vocational colleges* (Dz.U. 1997 no. 96 item 590), whose aims included enabling persons in difficult economic and life situations to undertake higher education. The establishment of national vocational colleges in smaller cities, bringing higher education closer to places distant from large academic centres, enabled the powerful development of local communities. These colleges were intended as a response to the needs of local labour markets, providing teaching oriented towards the professional and practical dimensions of academic subjects. The decision to find such institutions was in line with the international trend towards making higher education more widely available to the general population (Trow 1973).

The speed of changes in the Polish higher education system, whose scale was unprecedented anywhere in Europe, and the consequences of those changes for economic and social development in Poland, provided the motivation for the present work. The aim of the research is to identify and investigate the regional differentiation between Polish provinces in terms of features relating to higher education in the period of economic and social transformation. Observations were made in doubly multivariate way presenting: (i) fundamental characteristics of higher education and (ii) their changes in the years 2002–2013.

The next section of the article will present the features relating to higher education that were selected as diagnostic variables. The third section will describe the research procedure and statistical methods used. Since the study was based on doubly multivariate data, a method of principal component analysis for data of that type was applied. Delimitation of provinces was performed using the dendrite method (Florek et al. 1951, Kruskal 1956, Prim 1957) applied to the obtained principal components. The results of the classification of Polish provinces obtained by the described algorithm are set out in the fourth section. The final conclusions are preceded by an analysis of the identified clusters.

2. Diagnostic variables

The study was carried out using a set of diagnostic variables that are cited in many reports assessing the state of higher education and in analyses of educational systems[4].

The first variable is the number of higher educational institutions per 10,000 population (X1). During the first decade following the fall of communism, approximately 200 non-public higher educational institutions were established in Poland. These had an impact on the availability and diversity of study courses offered within the provinces in which the institutions were located.

The second variable considered is the number of students per 1000 population (X2). The aforementioned rise in student numbers has led to variation between provinces in terms of features relating to higher education over the past decades. Changes in numbers of students have been closely linked to changes in the number of higher educational institutions (Fig. 1). One consequence of the rising number of students is an increase in the number of graduates per 1000 population, which is taken as variable X3 in the model.

The structure and size of teaching and academic staff affect the prestige enjoyed by higher educational institutions. This is reflected in the academic potential of the provinces in which those institutions are situated. The analysis included two values relating to staffing: the number of academic teachers per 10,000 population (X4) and the number of academic teachers with the title of professor per 10,000 population (X5).

Research activity and specialist education are represented by two variables: the number of post-graduate students per 10,000 population (X6) and the number of doctoral students per 10,000 population (X7).

The analysis of the spatial variation between provinces was based on the set of seven diagnostic variables presented above. The data used are taken from the

[4] For example: *Szkolnictwo Wyższe w Polsce 2013*, Ministry of Science and Higher Education; *Szkoły wyższe i ich finanse 2013*, GUS.

Local Data Bank (http://stat.gov.pl/bdl/), the original source being the annual reports of higher educational institutions[5]. Missing values were acquired from the Statistical Yearbooks of Provinces, published by the Central Statistical Office (GUS).

The available absolute figures were divided by the numbers of inhabitants of the relevant provinces. To ensure correctness of the analysis, zero unitarization was applied (Walesiak 2014).

3. Research procedure

The algorithm for spatial delimitation of provinces consisted of three stages: data normalisation, construction of principal components, and cluster analysis.

Data normalisation was performed using the method of zero unitarization (Walesiak 2014). The fact that all of the observed values are stimulants (having positive impact) meant that a single common normalisation formula could be used. The source values of the observed features were transformed according to (1):

$$z_{jp} = \frac{x_{jp} - min\{x_{jp}\}}{max\{x_{jp}\} - min\{x_{jp}\}} \tag{1}$$

where z_{jp} is the normalised value of the pth variable for the jth object, and x_{jp} is the value of the pth variable for the jth object. The method gives normal values of the observed features in the interval <0; 1>, reducing the effect of disproportions in these values on the principal component analysis carried out in the second stage.

The second stage of the procedure involved principal component analysis of doubly multivariate data for a covariance matrix with Kronecker product structure.

Let us assume that we have an n-element sample consisting of objects characterised by p statistical features measured at T different time points. Data of this type are called doubly multivariate. Let $\mathbf{X}_{jk}$ denote the column vector of measurements of p features on the jth object at the kth time point, j=1,2,…,n, k=1,2,…,T. Let $\mathbf{X}_j=(\mathbf{X}_{j1}, \mathbf{X}_{j2},\ldots, \mathbf{X}_{jT})$ be a $p \times T$ matrix, and $\mathbf{x}_j=\mathrm{vec}(\mathbf{X}_j)$ be a pT-dimensional column vector of measurements of p features for the jth object at successive time points k, j=1,2,…,n.

---

[5] The reports are denoted in the GUS databases as follows:
F-01/s: Report on revenue, costs and financial results of higher educational institutions;
S-10: Report on higher education;
S-11: Report on material and social assistance to students and doctoral students;
S-12: Report on academic scholarships, post-graduate and doctoral studies and employment in higher educational institutions and scientific and research institutes.

We assume that $\mathbf{x}_j \sim N_{pT}(\boldsymbol{\mu}, \boldsymbol{\Omega})$, j=1, 2, …, n, where $\boldsymbol{\Omega}$ is a positive definite covariance matrix. Based on the estimator of the covariance matrix $\boldsymbol{\Omega}$ we construct the principal components (Hotelling 1933). The estimator of the matrix $\boldsymbol{\Omega}$ constructed from an n-element sample is positive definite with probability 1 if and only if $n>pT$ (e.g. Giri 1996). This condition implies a need to have a very large sample, which is not always possible. We, therefore, assume that the matrix $\boldsymbol{\Omega}$ has the structure of a Kronecker product (e.g. Gałecki 1994, Naik and Rao 2001, Roy and Khattree 2005, Krzyśko et al. 2011):

$$\boldsymbol{\Omega} = \mathbf{V} \otimes \boldsymbol{\Sigma}, \tag{2}$$

where $\mathbf{V}$ is the positive definite matrix of covariance between time points, with dimension $T\times T$, and $\boldsymbol{\Sigma}$ is the positive definite matrix of covariance between all statistical features, with dimension $p\times p$. When the matrix $\boldsymbol{\Omega}$ has this structure, its estimator is positive definite (with probability 1) if and only if $n>\max(p, T)$, which significantly weakens the condition on the size of the sample.

Bearing in mind that the matrix $\mathbf{V}$ represents variability over time, we may consider three models:

Model 1. We assume that the observations $\mathbf{x}_j$ are independent and that $\mathbf{x}_j \sim N_{pT}(\boldsymbol{\mu}, \mathbf{V} \otimes \boldsymbol{\Sigma})$, where $\mathbf{V}$ is a $T\times T$ positive definite matrix, $\boldsymbol{\Sigma}$ is a $p\times p$ positive definite matrix, and $n>\max(p, T)$. We do not impose any additional restrictions on $\mathbf{V}$.

Model 2. Weadopt the same assumptions as in Model 1, but also assume that the matrix $\mathbf{V}$ is completely symmetric, that is it has the form:

$$\mathbf{V} = -\frac{1}{1-\rho}\left[(1-\rho)\mathbf{I}_T + \rho\mathbf{1}_T\mathbf{1}_T^{\mathrm{T}}\right], \tag{3}$$

where ρ is the coefficient of correlation, and $\mathbf{1}_T$ is a T-dimensional column vector of ones.

Model 3. We adopt the same assumptions as in Model 1, but also assume that the matrix $\mathbf{V}$ has the structure of a first-order autoregression (Krzyśko et al. 2011), that is it has the form:

$$\mathbf{V} = \frac{1}{1-\rho^2}\left(\rho^{|r-s|}\right)_{r,s=1}^{T}, \tag{4}$$

where ρ is the coefficient of correlation.

In all three models the unknown parameters are estimated by the maximum likelihood method, solving appropriate systems of simultaneous equations iteratively until the selected “stop” criterion is attained (Srivastava et. al 2008, Krzyśko and Skorzybut 2009). We construct principal components based on the matrix $\widehat{\mathbf{\Omega}} = \widehat{\mathbf{V}} \otimes \widehat{\mathbf{\Sigma}}$ (Deręgowski and Krzyśko 2009). If $n > \max(p, T)$, then the matrix $\widehat{\mathbf{V}} \otimes \widehat{\mathbf{\Sigma}}$ is positive definite with probability **1**, and so all eigenvalues are real and positive. If $\alpha_1, \alpha_2, .., \alpha_T$ are the eigenvalues of $\widehat{\mathbf{V}}$ and $\beta_1, \beta_2, \dots, \beta_p$ are the eigenvalues of $\widehat{\mathbf{\Sigma}}$, the eigenvalues of $\widehat{\mathbf{V}} \otimes \widehat{\mathbf{\Sigma}}$ are pT numbers of the form $\alpha_r\beta_s$, where $r=1, 2, \dots, T$, $s=1, 2, \dots, p$. Based on the eigenvalues so defined, we construct the principal components of the matrix $\widehat{\mathbf{\Omega}} = \widehat{\mathbf{V}} \otimes \widehat{\mathbf{\Sigma}}$.

The principal components constructed in this way were used in the cluster analysis that formed the last stage of the study procedure. A hierarchical algorithm was used, based on the Wrocław taxonomy (Florek et al. 1951), involving the construction of the shortest dendrite[6] over a set of n objects, based on a selected measure of dissimilarity (Euclidean distance in this case):

$$\rho(\mathrm{x}_u, \mathrm{x}_v) = \left((\mathrm{x}_u - \mathrm{x}_v)'(\mathrm{x}_u - \mathrm{x}_v)\right)^{\frac{1}{2}} = \left(\sum_{i=1}^{p} (\mathrm{x}_{ui} - \mathrm{x}_{vi})^2\right)^{\frac{1}{2}}, \tag{5}$$

where: $u,v=1, 2, \dots, n$, $i=1, 2, \dots, p$.

In the shortest dendrite we determine the mean $\bar{\rho}$ and standard deviation s_ρ of the weights of all edges (distances between objects). The critical value, providing a criterion for the removal of edges from the dendrite, was taken to be the sum of $\bar{\rho}$ and s_ρ. The removal of edges whose weight exceeds the critical value leads to a division of the dendrite, and consequently to the separation of clusters.

4. Classification of provinces

Statistical analysis of the higher education data was performed in several stages (steps), with a different number of provinces considered each time. Each of the stages was based on the dendrite method, where the critical value was taken to be the mean length of an edge of the dendrite plus the standard deviation of the lengths. The method produced a division into six groups of provinces in four steps.

In the **first step**, the principal components were constructed for all 16 provinces, taking account of the three models for the structure of the matrix **V**. The goodness criterion was taken to be the index W, being the ratio of the sum of the variances of the first two principal components to the sum of the variances of all principal components, expressed as a percentage (Table 1). Model 2 was found to preserve the largest proportion of the variation of the data (73.55%).

[6] The shortest dendrite is the tree for which the sum of the weights on the edges is the smallest. The weights are taken as the distance between the tree nodes representing the studied objects. The shortest dendrite was constructed using Kruskal's algorithm (1956).

Table 1. The goodness criterion for the models in the first step

	The goodness criterion		
	Model 1	Model 2	Model 3
W index	63.27	73.55	63.4

Source: own calculations.

Figure 3 shows a projection of the provinces in the plane of the first two principal components, together with a dendrite over the points representing the provinces. The dotted line marks an edge longer than the critical value of 2.1591 (dendrite connections and edge lengths are given in Table 2).

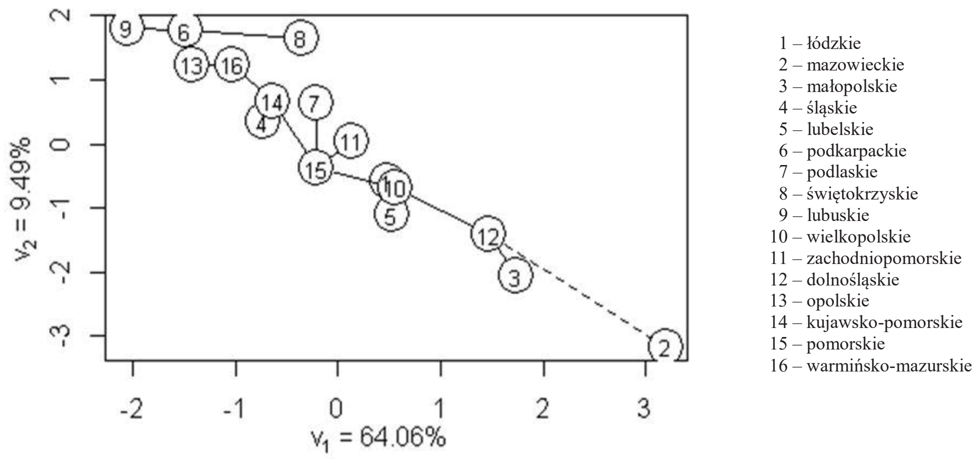

Figure 3. Dendrite over points representing provinces in the plane of the first two principal components, model 2 in the first step

Source: own calculations.

Table 2. Dendrite connections and edge lengths

Pairs of provinces	Edge length
1-10	0.8752
2-12	3.7373
3-12	1.7664
4-14	0.8875
5-10	1.1293
6-8	1.7859
6-9	1.0788
6-13	1.1570

Pairs of provinces	Edge length
7-15	1.4587
10-12	1.4477
10-15	1.2559
11-15	1.3666
13-16	1.2440
14-15	1.5957
14-16	1.3620

Source: own calculations.

The result of the analysis in the **first step** reveals the identification of two clusters, one of which is an isolated (single-element) cluster consisting of Mazowieckie province, further denoted ***Cluster 1***.

The remaining provinces, which make up the second cluster, underwent analysis in the **second step**. The greatest part of the variation, expressed by the index *W*, is preserved by the principal components in the second model (71.16%; Table 3).

Table 3. The goodness criterion for the models in the second step

	The goodness criterion		
	Model 1	Model 2	Model 3
W index	61.65	71.16	61.98

Source: own calculations.

Figure 4 shows a projection of the provinces in the plane of the first two principal components, together with the constructed dendrite. The edges that exceed the critical value (1.5989) are marked by a dotted line.

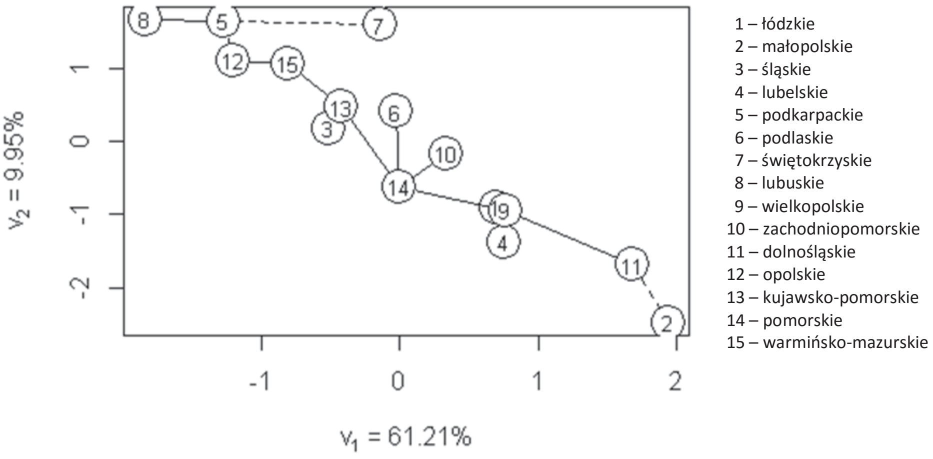

Figure 4. Dendrite over points representing provinces in the plane of the first two principal components, model 2 in the second step

Source: own calculations.

In the second step three clusters were identified, of which two are isolated clusters:

- ***Cluster 2*** consisting of Małopolskie province;
- ***Cluster 6*** consisting of Świętokrzyskie province.

In the **third step**, analysis was applied to the third cluster from the previous stage, consisting of 13 provinces: Śląskie, Podkarpackie, Lubuskie, Opolskie, Kujawsko-Pomorskie, Warmińsko-Mazurskie, Łódzkie, Lubelskie, Podlaskie, Wielkopolskie, Zachodniopomorskie, Dolnośląskie and Pomorskie.

The values of *W* for the three considered cases (Table 4) clearly show that Model 2 is again the most adequate to the data, explaining 67.15% of the variation.

Table 4. The goodness criterion for the models in the third step

	The goodness criterion		
	Model 1	Model 2	Model 3
W index	61.18	67.15	60.37

Source: own calculations.

Figure 5 shows a projection of the 13 provinces in the plane of the first two principal components, together with a dendrite constructed on the points representing them. The edges whose length exceed the critical value (1.462) are marked by dotted lines. This leads to a division into two clusters, consisting of the provinces:

- Łódzkie, Lubelskie, Podlaskie, Wielkopolskie, Zachodniopomorskie, Dolnośląskie, Pomorskie (denoted as ***Cluster 3***);
- Śląskie, Podkarpackie, Lubuskie, Opolskie, Kujawsko-Pomorskie, Warmińsko-Mazurskie (denoted temporarily as *Cluster 4*).

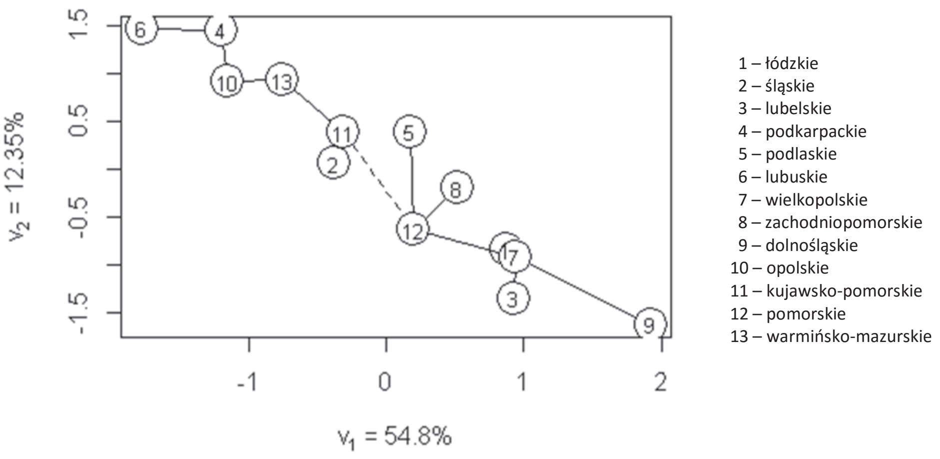

Figure 5. Dendrite over points representing provinces in the plane of the first two principal components, model 2 in the third step

Source: own calculations.

The **fourth step**, the final stage of the analysis, concerned Śląskie, Podkarpackie, Lubuskie, Opolskie, Kujawsko-Pomorskie and Warmińsko-Mazurskie provinces, contained in the temporary *Cluster 4*. The greatest part of the variation (67.75%) is preserved by the first two principal components in Model 2 (Table 5).

Table 5. The goodness criterion for the models in the fourth step

	The goodness criterion		
	Model 1	Model 2	Model 3
W index	55.76	67.75	57.88

Source: own calculations.

The critical edge length value (1.3246) in the dendrite was exceeded by the pairing of Kujawsko-Pomorskie and Warmińsko-Mazurskie provinces. This led to a division of the considered provinces into two clusters:

- ***Cluster 4*** consisting of Śląskie and Kujawsko-Pomorskie;
- ***Cluster 5*** consisting of the remaining provinces: Podkarpackie, Lubuskie, Opolskie and Warmińsko-Mazurskie.

The projection of the six provinces in the plane of the first two principal components, together with the constructed dendrite, is shown in Figure 6.

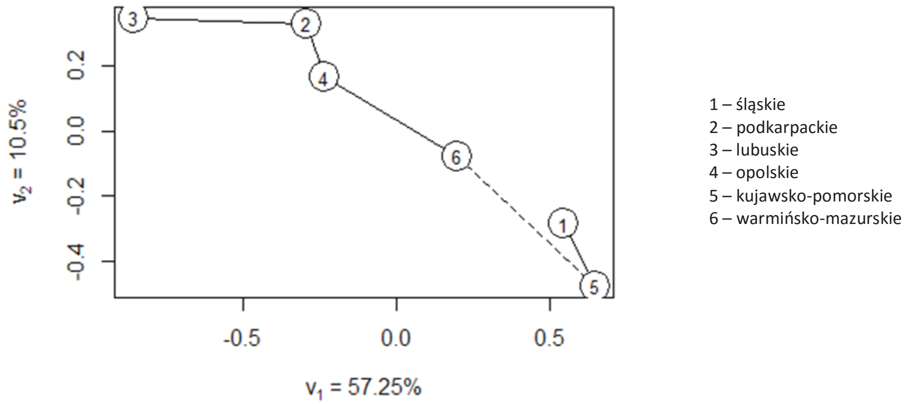

Figure 6. Dendrite over points representing provinces in the plane of the first two principal components, model 2 in the fourth step

Source: own calculations.

The provinces constituting ***Cluster 3*** were also analysed, but no basis was found for any further division of that cluster.

As a result of the four-stage classification process described above, the provinces were divided into a total of six groups (Fig. 7).

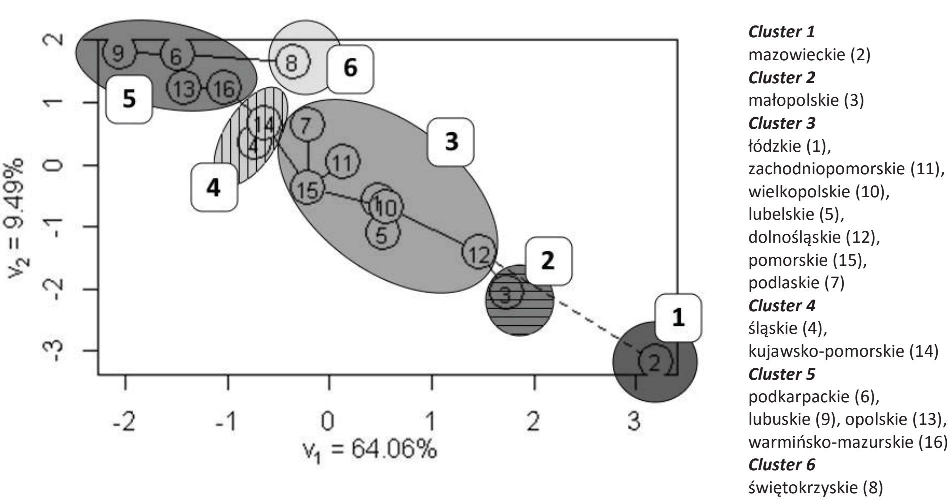

Figure 7. Dendrite for all 16 provinces, with clusters shown

Note: circles contain the numbers of the provinces belonging to the identified groups, and squares contain the numbers used to denote the clusters. In the legend, the numbers assigned to the provinces in the computational procedure are given in brackets.

Source: own calculations.

5. Analysis

In the first two steps of the algorithm, two single-element clusters were identified, consisting of Mazowieckie and Małopolskie provinces (Fig. 8). These regions contain the two largest and most renowned academic centres in Poland: Warsaw and Kraków. In national rankings, the University of Warsaw and Jagiellonian University are the two highest ranked higher educational institutions (*Perspektywy* ranking of higher educational institutions[7], *Polityka* ranking of higher educational institutions). They are also the only Polish institutions to appear on the Shanghai Ranking of the world's 500 leading universities (2015). It should also be noted that Warsaw and Kraków are the largest cities in Poland (in terms of population). As the national capital, Warsaw is also a financial, political and cultural centre. Mention should also be made of other higher educational institutions in these two provinces, which appear in the top ten of the aforementioned ranking: Warsaw University of Technology, the Warsaw School of Economics (SGH), and AGH University of Science and Technology in Kraków. Among non-public institutions offering master's degree courses, the leading ten (in the aforementioned ranking) include six institutions in Warsaw and one in Kraków. The concentration of so many leading institutions in those provinces explains their strong position in the higher education market, and is visible on the dendrite (Fig. 7) in the form of the large distance separating those regions from the remainder.

[7] *Perspektywy* ranking of higher educational institutions 2015.

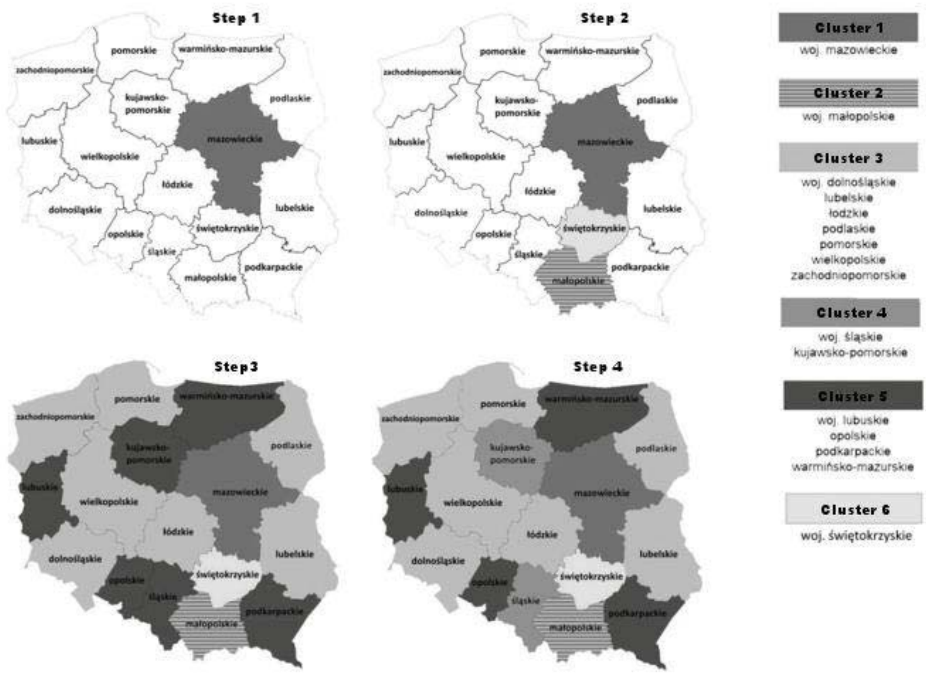

Figure 8. Spatial classification of provinces based on features relating to higher education (2002–2013)

Source: own analysis.

Cluster 3 consists of seven provinces containing higher educational institutions that are well-renowned within Poland and have a long-established tradition[8]. Most of the capitals of provinces in this cluster are among Poland's largest cities: Łódź, Wrocław, Poznań, Gdańsk (together with Gdynia and Sopot), Szczecin. The group also includes Lubelskie and Podlaskie provinces. These two eastern regions owe their membership of this cluster to the presence of higher educational institutions with notable values: the Catholic University of Lublin, Maria Curie-Skłodowska University in Lublin, the University of Białystok,

[8] The top 40 higher educational institutions in the 2015 *Perspektywy* ranking included, in order: Adam Mickiewicz University in Poznań, Wrocław University of Technology, the University of Wrocław, Gdańsk Medical University, Łódź University of Technology, Poznań Medical University, Poznań University of Technology, Wrocław Medical University, the University of Łódź, Gdańsk University of Technology, the University of Gdańsk, Łódź Medical University, Poznań University of Economics, Białystok Medical University, Lublin Medical University, Maria Curie-Skłodowska University in Lublin, the Pomeranian Medical University in Szczecin, Poznań University of Life Sciences, Wrocław University of Environmental and Life Sciences, the Catholic University of Lublin, Lublin University of Technology, and the University of Białystok.

Białystok Medical University, and the theological colleges in Białystok, Łomża and Drohiczyn.

In the course of the delimitation procedure, Cluster 4 was separated from Cluster 5. The dendrite (Fig. 7) shows the closeness of Śląskie and Kujawsko-Pomorskie provinces both to the group of provinces with the smallest potential (Cluster 5) and to the numerous group (Cluster 3), occupying the central part of the diagram.

Cluster 5 contains the Polish provinces with the lowest values of the analysed parameters. The cartogram reveals the peripheral nature of these regions, as well as their relative closeness to regions with higher potential.

The last of the identified clusters, consisting of Świętokrzyskie province, deviates from the axial arrangement of clusters seen on the dendrite. The values of some of the analysed higher education parameters were such as to place this province in the central group (a shift to the right on the horizontal axis of the dendrite) while others indicated that it belonged to the group with the smallest academic potential (a shift upwards on the vertical axis). This dual nature of observed values is well illustrated by the dendrite (Fig. 7).

The above analysis is complemented by a characterisation of the identified clusters in terms of descriptive statistics (Table 6).

Table 6. Descriptive statistics of the diagnostic variables for distinguished (identified) clusters

Variable	Clusters						Overall mean	Coefficient of variation between groups
	1	**2**	**3**	**4**	**5**	**6**		
	Mean value within group							
X1	19.60	9.84	11.60	9.21	6.75	11.20	11.37	38.59%
X2	63.44	60.12	47.33	40.02	34.76	38.59	47.38	25.31%
X3	14.12	12.11	10.84	9.66	8.68	10.80	11.03	17.28%
X4	3.15	3.66	2.69	2.04	1.58	1.43	2.43	36.75%
X5	0.85	0.75	0.59	0.47	0.38	0.39	0.57	34.01%
X6	10.35	4.14	3.37	2.67	2.45	2.86	4.31	70.12%
X7	1.71	1.49	0.87	0.58	0.23	0.08	0.82	80.09%

Meanings of variables: X1 – the number of higher educational institutions per 10,000 population; X2 – the number of students per 1000 population; X3 – the number of graduates per 1000 population; X4 – the number of academic teachers per 1000 population; X5 – the number of academic teachers with the title of professor per 10,000 population; X6 – the number of post-graduate students per 10,000 population; X7 – the number of doctoral students per 10,000 population.

Source: own calculations.

The mean values of the analysed features exhibit variation between the identified clusters. Mazowieckie province has the highest values for six out of the seven features. The single-element Cluster 2, consisting of Małopolskie province, has the highest number of academic teachers per 10,000 population, while in the other categories it lies second only to Mazowieckie province (often coming only slightly behind). The other clusters are separated from the leading two by a significant distance. The values recorded for Świętokrzyskie province clearly reveal its dual nature: the number of higher educational institutions per 10,000 population, the number of graduates per 1000 population and the number of post-graduate students per 10,000 population have values close to those for the high-potential clusters, while the values of number of academic teachers per 1000 population, the number of academic teachers with the title of professor per 10,000 population and the number of doctoral students per 10,000 population would place that province in the weakest group. It should be noted that, in terms of the values of the observed features, Cluster 4 differs from Cluster 3 (with higher potential) to a similar degree as from Cluster 5 (with lower potential), from which it was separated out.

The differentiation of the identified groups is greatest in the case of the variables representing numbers of post-graduate and doctoral students, for which the coefficients of variation are 70% and 80% respectively. The number of graduates per 1000 population, on the other hand, is relatively similar for all clusters, with a coefficient of variation not exceeding 20%.

The large disproportions in the values of variables between the two isolated clusters (Mazowieckie and Małopolskie provinces) and the other groups, the large group of provinces with moderate academic potential containing renowned centres of learning, and the isolated position of Świętokrzyskie province, deviating from the axial arrangement of the other clusters, create a characteristic picture of the spatial variation between Polish provinces based on the selected parameters relating to higher education.

6. Conclusions

The analysis has confirmed the dominance of Mazowieckie and Małopolskie provinces in the Polish higher education market. The higher educational institutions of these regions have been ranked the highest in national rankings for many years, as well as being Poland’s only representatives in important international rankings. The applied delimitation model revealed relations between the provinces in terms of the analysed features. A detailed analysis of the results obtained and consideration of additional parameters relating to economic, demographic and social features would enable a better and more comprehensive presentation of the differences between the regions.

A wider-ranging analysis of the Polish regions, covering features relating to human capital and the quality of life, would appear to be a natural development of the research reported here, and will form a part of the author's future work.

REFERENCES

ANTONOWICZ, D., (2012). External influences and local responses. Changes in Polish higher education 1990–2005, in: KWIEK M., MAASSEN P., (eds.), National Education Reforms in a European Context. Comparative Reflections on Poland and Norway. Frankfurt: Peter Lang, pp. 87–111.

DERĘGOWSKI, K., KRZYŚKO, M., (2009). Principal components analysis in case of multivariate repeated measures data. Biometrical Letters, 46 (2), pp. 163–172.

FLOREK, K., ŁUKASZEWICZ, J., PERKAL, J., STEINHAUS, H., ZUBRZYCKI, S., (1951). Taksonomia wrocławska. Przegląd Antropologiczny, 17, pp. 193–211.

GAŁECKI, A. T., (1994). General class of covariance structures for two or more repeated factors in longitudinal data analysis. Communications in Statistics – Theory and Methods, 23, pp. 3105–3119.

GIRI, N.C., (1996). Multivariate Statistical Analysis. New York: Marcel Dekker, Inc.

GOLINOWSKA, S., ed. (2005). Raport Społeczny Polska 2005. Fundacja im. Friedricha Eberta.

GUS, (2010, 2012, 2015). Rocznik Statystyki Międzynarodowej. Główny Urząd Statystyczny.

HOTELLING, H., (1933). Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology, 24 (6), pp. 417–441.

KRUSKAL, J. B., (1956). On the shortest spanning subtree of a graph and the traveling salesman problem. Proceedings of American Mathematical Society, 7, pp. 48–50.

KRZYŚKO, M., SKORZYBUT, M., (2009). Discriminant analysis of multivariate repeated measures data with a Kronecker product structured covariance matrices. Statistical Papers, 50, pp. 817–835.

KRZYŚKO, M., SKORZYBUT, M., WOŁYŃSKI, W., (2011), Classifiers for doubly multivariate data. Discussiones Mathematicae. Probability and Statistics, 31, pp. 5–27.

KWIEK, M., (2009). The two decades of privatization in Polish higher education. Cost-sharing, equity, and access. Sense Publishers.

KWIEK, M., (2014). Structural changes in the Polish higher education system (1990–2010): A synthetic view. European Journal of Higher Education, 4 (3), pp. 266–280.

MACH, B., (2003). Pokolenie historycznej nadziei i codziennego ryzyka: Społeczne losy osiemnastolatków z roku 1989. Wydawnictwo Instytutu Studiów Politycznych PAN.

MISZTAL, B., ed. (2000). Prywatyzacja szkolnictwa wyższego w Polsce: wyzwania w świetle transformacji systemowej. Kraków: Universitas.

NAIK, D. N., RAO, S., (2001). Analysis of multivariate repeated measures data with a Kronecker product structured covariance matrix. J. App. Statist. 28, pp. 91–105.

PRIM, R. C., (1957). Shortest connection networks and some generalizations. Bell System Technical Journal, 36, pp. 1389–1401.

ROY, A., KHATTREE R., (2005). On implementation of a test for Kronecker product covariance structure for multivariate repeated measures data. Statistical Methodology, 2, pp. 297–306.

SIKORSKA, J., (1998). Konsumpcja: Warunki, zróżnicowania, strategie. Wydawnictwo Instytutu Filozofii i Socjologii PAN.

SRIVASTAVA, M. S., von ROSEN, T., von ROSEN, D., (2008). Models with a Kronecker product covariance structure: estimation and testing. Mathematical Methods of Statistics, 17 (4), pp. 357–370.

TROW, M., (1973). Problems in the transition from elite to mass higher education. Carnegie Commission on Higher Education, Berkeley.

WALESIAK, M., (2014). Przegląd formuł normalizacji wartości zmiennych oraz ich własności w statystycznej analizie wielowymiarowej. Przegląd Statystyczny, 61 (4), pp. 363–372.

WASIELEWSKI, K., (2013). Caught in the trap of mass education – transformations in the Polish higher education after 1989, in: SZAFRANIEC K. and KONSTANTINOVSKIY D. (eds.). Polish and Russian Youth: Education and Work in Changing Society. IS RAS, Moskwa.

WNUK-LIPIŃSKA, E., (1996). Innowacyjność a konserwatyzm: uczelnie polskie w procesie przemian społecznych. Wydawnictwo Instytutu Studiów Politycznych PAN.

ZIÓŁKOWSKI, M., (2000). Przemiany interesów i wartości społeczeństwa polskiego. Teorie, tendencje, interpretacje. Wydawnictwo Fundacji Humaniora, Poznań.