Wiadomości Statystyczne. The Polish Statistician, 2022, vol. 67, 3, 28–45 DOI: 10.5604/01.3001.0015.8065
Statystyka w praktyce / Statistics in practice
Zgłoszony/submitted: 9.08.2021, poprawiony/revised: 21.02.2022, zaakceptowany/accepted: 4.03.2022

Variation of Polish voivodships according to selected indicators referring to people aged 65 and over[1]

Agata Szymańska[a]

Abstract. The aim of the study is the assessment of the spatial variation of voivodships (the largest administrative units in Poland) in terms of selected indicators describing the situation of people aged 65 and over. The study used data from the Local Data Bank of Statistics Poland (Główny Urząd Statystyczny) for the years 2005, 2010, 2015 and 2019. From the obtained set of diagnostic variables, five were used for the final analysis, including those relating to demographics, the pension security system and health infrastructure. The applied empirical method was based on the Euclidean metric as well as cluster analysis with Ward's method.

The performed analyses indicated a variation of voivodships which is reflected by the computed distances and components of the created clusters. The results revealed a spatial variation of voivodships which is consistent with the demographic ageing in Poland. Moreover, the observed distinction of Śląskie Voivodship may have been affected by the fact that variables related to the pension security system were also considered.

Keywords: demographic ageing, people aged 65 and over, cluster analysis, regional variation, Euclidean metric, Ward's method

JEL: J10, J11

Zróżnicowanie województw ze względu na wybrane wskaźniki odnoszące się do osób w wieku 65 lat i więcej

Streszczenie. Celem badania omawianego w artykule jest ocena zróżnicowania przestrzennego województw pod względem wybranych wskaźników opisujących sytuację osób starszych (65 lat i więcej). Badanie dotyczyło lat 2005, 2010, 2015 i 2019; wykorzystano w nim dane publikowane w Banku Danych Lokalnych Głównego Urzędu Statystycznego. Z uzyskanego zestawu zmiennych wybrano do ostatecznych analiz pięć, w tym zmienną demograficzną oraz zmienne związane z zabezpieczeniem emerytalnym i infrastrukturą zdrowotną. Badanie empiryczne przeprowadzono z wykorzystaniem miary odległości euklidesowej oraz analizy skupień metodą Warda.

---

[1] Artykuł został opracowany na podstawie referatu wygłoszonego na konferencji Multivariate Statistical Analysis MSA 2021, która odbyła się w dniach 8–10.11.2021 r. w Łodzi. / The article was based on a paper presented at the Multivariate Statistical Analysis MSA 2021 Conference, held on 8–10 November 2021 in Lodz, Poland.

[a] Uniwersytet Łódzki, Wydział Ekonomiczno-Socjologiczny, Instytut Ekonomii / University of Lodz, Faculty of Economics and Sociology, Institute of Economics, Poland.
ORCID: https://orcid.org/0000-0001-5184-931X. E-mail: agata.szymanska@uni.lodz.pl.

Przeprowadzone analizy wykazały zróżnicowanie województw, co odzwierciedlają zarówno wyliczone odległości, jak i skład wyodrębnionych skupień. Wyniki wskazują na przestrzenne zróżnicowanie województw, które jest zgodne z przestrzennym zróżnicowaniem demograficznego starzenia się ludności w Polsce. Zaobserwowana odrębność woj. śląskiego może ponadto wynikać z uwzględnienia zmiennych dotyczących systemu zabezpieczenia emerytalnego.

Słowa kluczowe: demograficzne starzenie się ludności, osoby starsze, analiza skupień, zróżnicowanie regionalne, odległość euklidesowa, metoda Warda

1. Introduction

In the recent years many countries have experienced various demographic changes, such as the intensification of the ageing of the population, which is an advanced process in many economies, regardless of their development. This process affects different areas of socio-economic life and shapes public policies. Particularly, the phenomenon of demographic ageing can be interpreted as an increase in older people's share in the total population with a simultaneous observable decrease in the share of children (Główny Urząd Statystyczny [GUS], 2014; Holzer, 2003). Demographic ageing is measured by a variable reflecting the percentage of the population aged 65 and over in the total population (e.g. Długosz, 1998; Kowaleski & Majdzińska, 2012). This indicator is widely used by many researchers, organisations and institutions to assess demographic changes and evaluate the possible impact of the ageing process on the socio-economic development of countries, regions or even smaller local communities.

Population ageing affects all countries and regions of the world and influences numerous socio-economic conditions, including the pension system, health care, long-term care and others. Moreover, living conditions may differ between countries, regions or local units. Therefore, this study analyses the spatial variation of voivodships in Poland from the point of view of selected indicators related to the situation of older people using cluster analysis. The set of indicators describes the disparities between voivodships from the point of view of the variables that capture the socio-economic conditions of the well-being of those aged 65 and over.

This paper contributes to the increase of knowledge about the variation of voivodships according to the situation of older people as the conducted research is directly linked to the phenomenon of demographic ageing and a gap in this respect is observed in the literature. The aim of the study was the analysis of the spatial variation of voivodships (the largest administrative units in Poland) from the perspective of selected indicators describing the situation of people aged 65 and over. This was achieved by comparing the changes in the composition of the constructed clusters and analysing the (dis)similarity and the durability of the connections of every single pair of voivodships in the years 2005, 2010, 2015 and 2019.

Cluster analysis allows for the grouping of similar objects into clusters (e.g. Romesburg, 2004; Stanisz, 2007; Zeliaś, 2004) and this approach is consistent with the aim of the study. Moreover, the longest and the shortest distances for each voivodship and the corresponding one in the analysed year (see e.g. Zeliaś, 2004) are presented to distinguish pairs of voivodships with the lowest and highest variation and to compare the potential stability of those links.

2. Demographic ageing – selected facts and background for the study

In the report of 2021 prepared by the Economic Policy Committee's Working Group on Ageing Populations and Sustainability (EPC-AWG; see European Commission, 2021), the forecasted share of the population aged 65 and over in the total population of the European Union for 2070 is approximately 30.3%, denoting an increase of 9.9 p.p. compared to the base year, i.e. 2019. The findings of the report suggest that in the case of the 27 European Union countries the highest indicator for 2070 is foreseen for Poland (34%, an increase of 16 p.p.) and Italy (33.3%, an increase of 10.4 p.p.), whereas the lowest is projected for Sweden (26.3%, an increase of 6.3 p.p.) and Cyprus (27.1%, an increase of 10.9 p.p.). However, it should be noted that in Poland in the base year, the indicator was 17.9%, while the average for the EU-27 was 20.4%, with the highest indicator relating to Italy (23%) and the lowest to Ireland (14.5%).

One aspect strongly connected with the ageing of the population is the level of age-related public expenditure, including pension expenditure which is determined by the number of pensioners and the shape of pension systems. The European Commission (2021) forecasts indicate that the total cost of the ageing-related expenditure as a percentage of GDP (under the AWG reference scenario) will increase by 1.9 p.p., i.e. from 24% in 2019 to 25.9% in 2070. For example, in Poland, an increase of 4 p.p. (from 20.1% to 24.1% over this period) is projected, while the highest increase is forecasted for Slovakia (by 10.8 p.p.). The average spending on gross public pension (as a percentage of GDP) for the EU-27 is predicted to increase by 0.1 p.p. (from 11.6% in 2019 to 11.7% in 2070).

Ageing is not only analysed from the countries' perspective but also in regional dimensions. Considering the European Commission (2021) projections for Poland, the analysis of the older people's population is worth further consideration, especially as Poland's case indicates the spatial diversity of voivodships. Numerous studies have been conducted on the diversity of Polish voivodships analysed from the point of view of the demographic ageing of the population (see e.g. Długosz, 1998; Kowaleski, 2011; Majdzińska, 2017; Podogrodzka, 2016; Roszkowska, 2020) and using different taxonomic approaches.

Table 1 presents the percentage of people aged 65 and over in the total population of Poland and the 16 voivodships in the selected years in the period of 1995–2020, showing that particular voivodships indicated a different development of the index for demographic ageing.

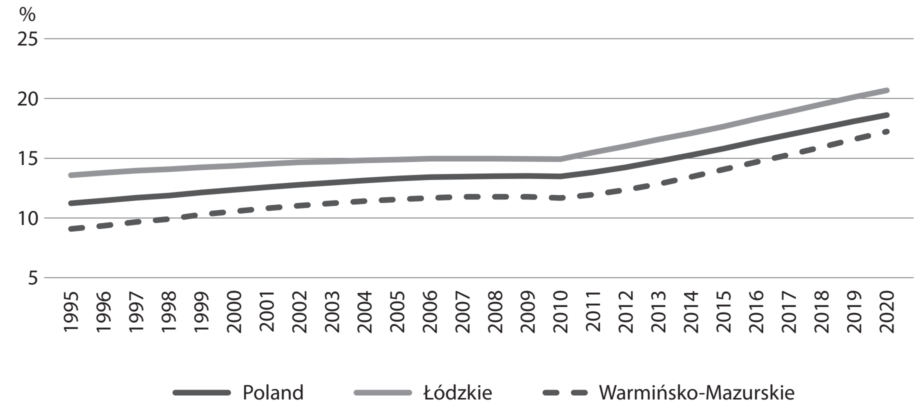
Table 1. Percentage of people aged 65 and over in the total population of Poland by voivodship

Specification	1995	2000	2005	2010	2015	2020	Change between 1995 and 2020 in p.p.
Poland	**11.23**	**12.35**	**13.30**	**13.47**	**15.81**	**18.61**	**7.38**
Dolnośląskie	11.02	12.50	13.47	13.34	16.20	19.65	8.63
Kujawsko-Pomorskie	10.56	11.52	12.32	12.54	15.32	18.31	7.74
Lubelskie	12.83	13.66	14.27	14.38	16.36	19.09	6.26
Lubuskie	9.74	10.86	11.67	11.73	14.77	18.23	8.49
Łódzkie	13.59	14.37	14.88	14.93	17.64	20.67	7.08
Małopolskie	11.24	12.38	13.37	13.58	15.27	17.38	6.15
Mazowieckie	12.95	13.88	14.55	14.43	16.23	18.55	5.60
Opolskie	9.89	11.68	13.66	14.16	16.45	19.35	9.46
Podkarpackie	10.84	11.99	12.91	13.07	14.89	17.35	6.52
Podlaskie	12.49	13.59	14.46	14.60	16.09	18.22	5.73
Pomorskie	9.58	10.82	11.92	12.13	14.68	17.42	7.84
Śląskie	10.08	11.52	13.25	14.22	16.67	19.76	9.68
Świętokrzyskie	12.94	14.03	14.87	14.82	17.17	20.26	7.32
Warmińsko-Mazurskie	9.10	10.55	11.56	11.68	14.05	17.23	8.13
Wielkopolskie	10.79	11.32	11.89	11.85	14.63	17.28	6.48
Zachodniopomorskie	9.50	11.08	12.09	12.24	15.40	19.18	9.68

Source: author's calculations based on data from the Local Data Bank of Statistics Poland.

The highest indicator in Poland, regardless of the analysed year, concerned Łódzkie Voivodship and the lowest – Warmińsko-Mazurskie Voivodship, emphasising the spatial diversity of voivodships in terms of the examined index. As shown in Table 1, the largest change (growth) between 1995 and 2020 occurred in Śląskie and Zachodniopomorskie voivodships (by approximately 9.68 p.p.), with the lowest increase recorded in Mazowieckie (by 5.6 p.p.) and Podlaskie (by 5.73 p.p.). Since the largest and lowest values of the indicator concern Łódzkie and Warmińsko-Mazurskie voivodships, Figure 1 presents the development of the index for these two voivodships from 1995 to 2020 against the background of the Polish average.

Figure 1. Share of people aged 65 and over in the total population of Poland, in Łódzkie Voivodship and in Warmińsko-Mazurskie Voivodship



Source: author's work based on data from the Local Data Bank of Statistics Poland.

The difference in the value of the indicator between Łódzkie and Warmińsko-Mazurskie voivodships ranged from 4.49 p.p. in 1995 to 3.18 p.p. in 2009. However, generally, the disproportion between the voivodship with the lowest and the highest share of people aged 65 and over in the total population was decreasing. It is also emphasised by the scale of the increase of the value of the indicator, which grew by 7.08 p.p. for Łódzkie and even more so (by 8.13 p.p.) for Warmińsko-Mazurskie Voivodship. Generally, an upward trend of the indicator was observed for Poland, suggesting the intensification of the process of demographic ageing.

3. Research method

The analysis of the variation of voivodships was based on an empirical approach using the Euclidean metric as well as cluster analysis with the delimitation of similar objects. The assessment of the obtained results was conducted for three years in five-year intervals, i.e. for 2005, 2010, 2015, as well as for the year 2019. The data were obtained from the Local Data Bank (LDB) of Statistics Poland and 2019 replaced 2020 as some data concerning the latter year were missing.[2] The proposed time span resulted from the availability of data gathered in the LDB.

Cluster analysis is a multivariate statistical method applied in this paper to study the variation of voivodships. The algorithm of cluster analysis is based on the examination of the distance between objects (see Bailey, 1994), and generally, the greater the computed distance, the lower the level of similarity. Thus, an important

[2] Data for this study were accessed in the beginning of July 2021.

element of the algorithm of cluster analysis is to calculate the distance for each pair of objects (e.g. x_i and x_j) to quantify the degree of their (dis)similarity (Everitt et al., 2011). In practice, there is a set of different distance measures, and the most popular choice, i.e. the Euclidean metric (Everitt et al., 2011; Kaufman & Rousseeuw, 2005), is defined by the formula:

$$d_{ij} = \sqrt{\sum_{i=1}^{p} (x_{ik} - x_{jk})^2},$$

where x_{ik} and x_{jk} are the k-th variable value of the p-dimensional observations for individuals i and j, respectively (Everitt et al., 2011).

Ward's method (1963) was adopted to measure the proximity between groups of individuals. In this method, at each step, the pair of clusters for which the increase in total within-cluster variance (defined by weighted squared distance between the cluster centres) achieves a minimum after their merging, is finally merged in the bigger group of clusters. The algorithm begins with particular objects being single and the distance between them is computed according to the Euclidean formula (Kaufman & Rousseeuw, 2005; Romesburg, 2004) and the process stops when the distance between the merged clusters exceeds an arbitrarily established threshold. It means that at each stage at which the objects are combined, from all the possible partitions of the set of objects for which the groups of objects that make it up, the created clusters have the smallest internal variation in terms of the variables describing the objects (Panek & Zwierzchowski, 2013). The advantage of Ward's method is that it is generally used with the (squared) Euclidean distance (Romesburg, 2004; Sarstedt & Mooi, 2014), although it can also be used with any other (dis)similarity measure (Sarstedt & Mooi, 2014). Cluster analysis based on Ward's method is considered effective, despite the fact that it tends to create small-sized clusters (Stanisz, 2007). The dendrogram is the graphical output of the grouping method allowing the analysis of the observed objects and the formation of separate clusters.

In this study, the data source for the set of the chosen explanatory variables was the LDB. Taking into account the potential low variability of the demographic data, the aim was to identify the variables with the longest available period and for each of the 16 voivodships. The availability of data in the LDB influenced the decision to use data covering the period from 2005 to 2019, however, due to the lack of some data, only a narrow set of indicators was suitable for use, and finally the set included what follows:

X_1 – share of people aged 65 and over in the total voivodship population in %;
X_2 – share of people aged 85 and over in the total voivodship population in %;
X_3 – number of people aged 65 and over per 1,000 inhabitants of the voivodship;
X_4 – number of people aged 65 and over in cities per 1,000 of the urban population of the voivodship;
X_5 – number of people aged 65 and over in rural areas per 1,000 of the rural population of the voivodship;
X_6 – deaths of people aged 65 and over per 1,000 inhabitants of the voivodship;
X_7 – deaths of people aged 65 and over per 1,000 inhabitants of the voivodship aged 65 and over;
X_8 – number of pensioners from the non-agricultural insurance system per 1,000 inhabitants of the voivodship;
X_9 – number of pensioners from the non-agricultural insurance system per 1,000 inhabitants of the voivodship aged 65 and over;
X_{10} – number of pensioners from the non-agricultural insurance system per 1,000 beneficiaries of that insurance system;
X_{11} – average pension in relation to the average monthly salary in the voivodship;
X_{12} – average pension in relation to the average monthly salary in the voivodship, calculated within the Polish Classification of Activities (PKD 2007, equivalent to NACE Rev. 2) sections;
X_{13} – number of outpatient clinics per 1,000 inhabitants aged 65 and over located in the voivodship.

However, due to data availability issues, the study period included the years 2005–2019 and each voivodship was expressed by the $X_1 - X_{13}$ set of variables presented for the years 2005, 2010, 2015 and 2019. Selected descriptive statistics are presented in Table A1 in the Appendix. The potential indicators were analysed from the perspective of their informative features, and these variables had to be characterised by a low degree of correlation in order to avoid collinearity. The correlation matrices for all variables are presented in Table A2 in the Appendix. The literature points out that, generally, the correlation between variables should not be strong (Bernstein et al., 1988) and the practice is to use variables with coefficients not exceeding 0.7 (Nowak, 1990). Moreover, the variables must be characterised with an adequate variability in order to ensure the variation of the objects within the examined features. Thus, the final set included indicators X_3, X_8, X_{10}, X_{12} and X_{13}.

The headline indicator, i.e. the index of demographic ageing, was excluded from further analysis due to its high collinearity with the set of other variables. The five variables used in the multivariate analysis were standardised. The used formula took the following form:

$$x'_{ij} = \frac{x_{ij} - \bar{x}_i}{S_i},$$

where: x_i represents the value of the i-th variable for the j-th object (voivodship) in the original data set; $\bar{x}_i$ is the mean of the i-th variable and S_i denotes the standard deviation of the i-th variable.

4. Empirical results

The standardised variables were used to calculate the distance between objects, with distance matrices for the years 2005, 2010, 2015 and 2019, analysed and compared. The computed values allow for the initial analysis of the (dis)similarity of the objects (see e.g. Zeliaś, 2004), showing that in 2005, 2015 and 2019 the largest distance between voivodships concerned the Śląskie and Mazowieckie pairing, whereas in 2010 – Warmińsko-Mazurskie and Mazowieckie voivodships.

Table 2. Voivodships and corresponding voivodships with maximum and minimum distance in the years 2005, 2010, 2015 and 2019 – analysis of the (dis)similarity based on the Euclidean distance

Voivodships	2005		2010		2015		2019	
	min.	max.	min.	max.	min.	max.	min.	max.
DŚL	POM	ŚL	MŁP	W-M	ŚW	W-M	ZPM	W-M
K-P	WLKP	MAZ	POM	MAZ	POM	MAZ	POM	MAZ
LBL	MŁP	ŚL	PKR	ŚL	PKR	ŚL	MŁP	ŚL
LBU	WLKP	MAZ	ZPM	MAZ	MŁP	MAZ	MŁP	MAZ
ŁDZ	OPO	WLKP	OPO	W-M	OPO	W-M	DŚL	WLKP
MŁP	PKR	MAZ	PKR	MAZ	PKR	ŚL	PKR	ŚL
MAZ	DŚL	**ŚL**	DŚL	**W-M**	DŚL	**ŚL**	DŚL	**ŚL**
OPO	ŚW	ŚL	ŁDZ	W-M	ZPM	W-M	ZPM	ŚL
PKR	ZPM	MAZ	MŁP	MAZ	MŁP	ŚL	MŁP	ŚL
PDL	ŚW	ŚL	OPO	ŚL	OPO	ŚL	MŁP	ŚL
POM	K-P	ŚL	PKR	ŚL	PKR	ŚL	MŁP	ŚL
ŚL	ŁDZ	**MAZ**	ŁDZ	MAZ	ZPM	**MAZ**	LBU	**MAZ**
ŚW	PDL	ŚL	MŁP	W-M	DŚL	W-M	ZPM	W-M
W-M	LBU	MAZ	WLKP	**MAZ**	WLKP	MAZ	PKR	MAZ
WLKP	LBU	MAZ	LBU	MAZ	LBU	MAZ	PKR	ŁDZ
ZPM	PKR	MAZ	LBU	MAZ	MŁP	MAZ	OPO	W-M

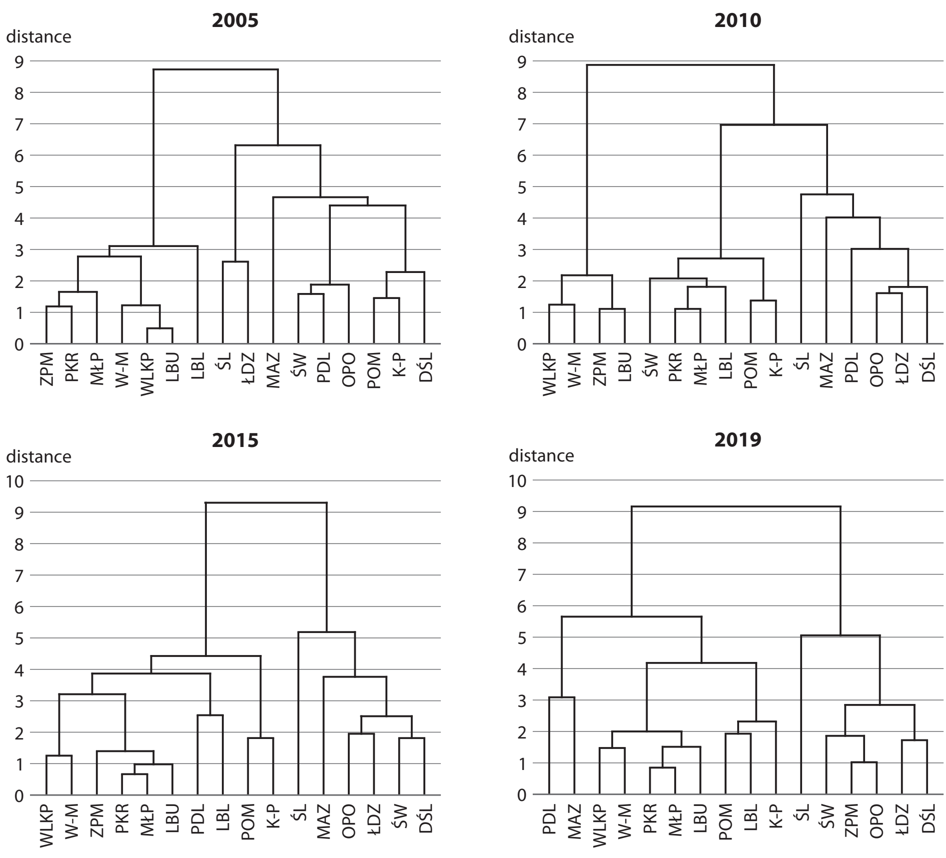
Note. DŚL – Dolnośląskie, K-P – Kujawsko-Pomorskie, LBL – Lubelskie, LBU – Lubuskie, ŁDZ – Łódzkie, MŁP – Małopolskie, MAZ – Mazowieckie, OPO – Opolskie, PKR – Podkarpackie, PDL – Podlaskie, POM – Pomorskie, ŚL – Śląskie, ŚW – Świętokrzyskie, W-M – Warmińsko-Mazurskie, WLKP – Wielkopolskie, ZPM – Zachodniopomorskie. Grey cells – the lowest distance in the analysed year, bolded records – the maximum distance in the analysed year.

Source: author's work based on data from the Local Data Bank of Statistics Poland.

Table 2 presents the results for the minimum and maximum Euclidean distance between a pair of voivodships. The maximum distance between the voivodship from the column titled 'Voivodship' and the corresponding voivodship presented in columns described as 'min.' indicates the lowest similarity, whereas the similarity between the objects is the highest when the distance is the lowest (see columns described as a 'max.').

Generally, most of the computed distances indicate that the maximum distance between particular voivodships concerns their relation to Mazowieckie or Śląskie Voivodship. Moreover, the largest distance in 2005, 2015 and 2019 concerned the pair of Mazowieckie and Śląskie (the maximum distance between these two voivodships in the three years indicates the largest disparity between them, analysed in the context of the set of the applied variables). The lowest distance in 2010, 2015 and 2019 concerned the pair of Małopolskie and Podkarpackie (i.e. suggesting the most similar voivodships).

Figure 2. Dendrograms



Note. Abbreviated names of voivodships as in Table 2.
Source: author's work based on data from the Local Data Bank of Statistics Poland.

Considering all the assumptions made, the selected set of indicators and the five final variables, cluster analysis was employed to group voivodships into clusters of the most similar objects. Figure 2 presents the results of the process of combining individual voivodships into groups, using the cluster analysis algorithm. The same figure also shows the division of the voivodships into two 'large' separate clusters, which were joined together by the last link and proved most different from each other; moreover, their elements changed in each of the analysed years. However, in every year, the 'left' cluster included Wielkopolskie, Lubuskie and Warmińsko-Mazurskie, whereas the 'right' one included Śląskie, Dolnośląskie, Opolskie, Świętokrzyskie and Łódzkie. In general, there are several ways to cut the dendrogram (Panek & Zwierzchowski, 2013). In this study the dendrograms were divided into five clusters in 2005, 2010 and 2019, and four clusters in 2015 (the elements of the clusters for each analysed year are presented in Table 3).

Table 3. Components of the clusters – results of the cluster analysis

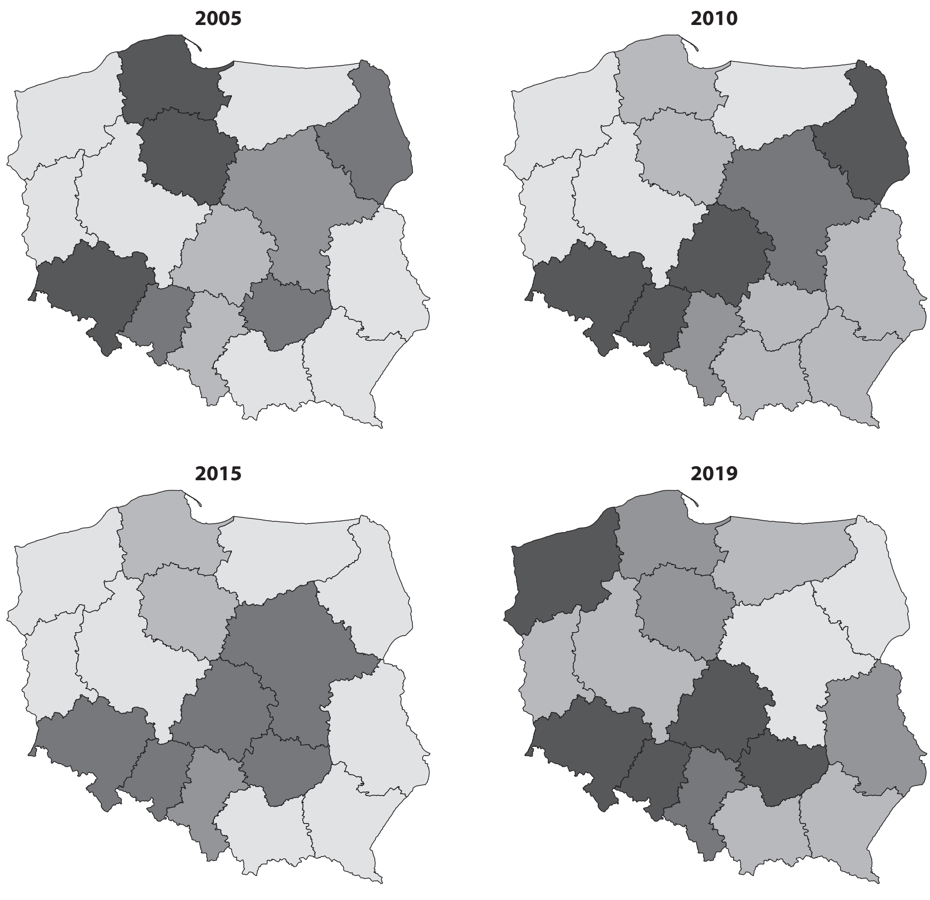
Cluster	2005	2010	2015	2019
1	ZPM, PKR, MŁP, W-M, WLKP, LBU, LBL	WLKP, W-M, ZPM, LBU	WLKP, W-M, ZPM, PKR, MŁP, LBU, PDL, LBL	PDL, MAZ
2	ŚL, ŁDZ	ŚW, PKR, MŁP, LBL, POM, K-P	POM, K-P	WLKP, W-M, PKR, MŁP, LBU
3	MAZ	ŚL	ŚL	POM, LBL, K-P
4	ŚW, PDL, OPO	MAZ	MAZ, OPO, ŁDZ, ŚW, DŚL	ŚL
5	POM, K-P, DŚL	PDL, OPO, ŁDZ, DŚL	.	ŚW, ZPM, OPO, ŁDZ, DŚL

Note. Abbreviated names of voivodships as in Table 2.
Source: author's work based on data from the Local Data Bank of Statistics Poland.

The delimitation of the dendrograms resulted in Mazowieckie Voivodship creating a separate cluster in 2005 and 2010, while a separate cluster was formed by Śląskie Voivodship in 2010, 2015 and 2019. These results were confirmed by the observations resulting from the analysis of the computed distances based on the Euclidean metric (Table 2). In most cases it indicated a maximum distance of individual voivodships from Mazowieckie or Śląskie Voivodship. In each of the analysed years there were clusters whose composition revealed that some elements were repeated: Kujawsko-Pomorskie and Pomorskie voivodships (some elements of cluster 5 in 2005, cluster 2 in 2010 and 2015 and cluster 3 in 2019), Warmińsko-Mazurskie, Wielkopolskie and Lubuskie (included in cluster 1 in 2005, 2010 and 2015, while in 2019 the voivodships were in cluster 2), Małopolskie and Podkarpackie (elements of cluster 1 in 2005 and 2015 and cluster 2 in 2010 and

2019). The maps illustrating the division of voivodships into clusters are presented to visualise the geographical variation of the voivodships (Figure 3).

Figure 3. Classification of voivodships from the point of view of a set of variables



Note. 2015 saw a delimitation into four clusters, while in 2005, 2010 and 2019 into five clusters. For the composition of clusters, see Table 3.
Source: author's work based on Table 3.

The supplementary calculation of the means of the individual variables in the clusters indicates that the lowest values for X_3 (the number of people aged 65 and over per 1,000 inhabitants of the voivodship) were in cluster 1 in 2005 and 2010, and in cluster 2 in 2015 and 2019. The highest mean for X_3 occurred in cluster 3 in 2005, while in cluster 4 in 2010 and 2015 and in cluster 5 in 2019. These results are consistent with the demographic ageing of the voivodships. Generally, voivodships with an indicator for demographic ageing of a low value (e.g. Warmińsko-Mazurskie, Lubuskie, Wielkopolskie) or high value (e.g. Łódzkie, Opolskie, Dolnośląskie) were, except for the year 2005, assigned to the same clusters (see Table 1 and 3 and Figure 1). In the case of X_{12}, the highest mean occurred in cluster 2 in 2005, in cluster 4 in 2019, and in cluster 3 in 2010 and 2015 (in each of the analysed years it was related to the cluster including Śląskie Voivodship, moreover, in 2010, 2015 and 2019 – only Śląskie), while the lowest concerned voivodships in cluster 3 in 2005, cluster 4 in 2010 and 2015, and cluster 1 in 2019 (in each of the analysed years, it was related to the cluster including Mazowieckie). The highest mean for the variable capturing the number of pensioners from the non-agricultural social security system per 1,000 inhabitants of the voivodship is characterised by voivodships included in cluster 2 in 2005, cluster 3 in 2010 and 2015, as well as in cluster 4 in 2019 (in each of the analysed years, it was related to the cluster including Śląskie and only Śląskie in 2010, 2015, and 2019), while the lowest in 2005, 2015 and 2019 for voivodships that created cluster 1 and cluster 2 in 2010.

Moreover, it should be noted that the variation of voivodships in terms of the analysed variables changed in the studied years. However, it is possible to distinguish groups of voivodships that are elements of one cluster in each period.

Considering the analysis of individual variables, in 2010, 2015 and 2019, the ratio of the average pension to the average monthly salary (captured by variable X_{12}) was the highest in Śląskie and the lowest in Mazowieckie Voivodship. The one variable included in the final dataset which indicates health 'infrastructure' was X_{13}, i.e. the average number of outpatient clinics per 1,000 inhabitants aged 65 and over, which in the years 2010, 2015 and 2019 was the highest in Warmińsko-Mazurskie and the lowest in Kujawsko-Pomorskie.

The comparison of the cluster components between 2005 and 2019 shows that at least two clusters included a minimum of two of the same elements. For example, cluster 1 in 2005 (structured by seven elements) and cluster 2 in 2019 (five elements) included Wielkopolskie, Warmińsko-Mazurskie, Podkarpackie, Małopolskie and Lubuskie voivodships; cluster 5 in 2005 (three elements), cluster 2 in 2010 (six elements), cluster 2 in 2015 (two elements) and cluster 3 in 2019 (three elements) included Pomorskie and Kujawsko-Pomorskie; cluster 4 in 2005 (three elements), cluster 4 in 2015 (five elements) and cluster 5 in 2019 (five elements) included Świętokrzyskie and Opolskie.

Figure 3 indicates that clusters of similar voivodships often comprised of spatially remote objects.

The results suggest that the links computed on the basis of Euclidean distances between some voivodships were relatively stable in the analysed years, indicating that the (dis)similarities of these pairs were not as dynamic over time. Moreover, the links between some pairs of voivodships maintained stable in terms of the pairs connected by minimum or maximum distance. For example, in the case of Kujawsko-Pomorskie, Lubuskie, Śląskie and Warmińsko-Mazurskie voivodships, the maximum computed Euclidean distance was with Mazowieckie, whereas Śląskie Voivodship in the case of Lubelskie, Podlaskie and Pomorskie. The lowest and stable distance was created with Podkarpackie Voivodship for Małopolskie Voivodship and Dolnośląskie for Mazowieckie.

Generally, the composition of the formed clusters varied over time, which indicated the changing similarities of the created groups of voivodships. Despite this, some voivodships were included in the same clusters in each year, for example, Podkarpackie and Małopolskie or Pomorskie and Kujawsko-Pomorskie, or Warmińsko-Mazurskie, Wielkopolskie and Lubuskie, as previously mentioned.

It should be noted that one of the limitations of the study involves the exclusion of the headline indicator used in the analysis of demographic ageing, i.e. the share of people aged 65 and over in the total population (due to the high collinearity with other variables). Moreover, the results are sensitive to the set of the included variables, and it may affect the composition of the obtained clusters. Thus, further research should analyse the comparison of a different set of indicators, as well as a different level of spatial data aggregation, for example, gminas or powiats (NTS 5 and NTS 4, respectively).[3]

5. Conclusions

This study attempted to analyse the variation of voivodships in Poland from the point of view of selected variables that capture aspects related to people aged 65 and over and some aspects related to the pension security system in the years 2005, 2010, 2015 and 2019. The cluster analysis of five diagnostic variables identified five clusters for 2005, 2010 and 2019, and four clusters for 2015, showing the variation of voivodships in terms of the analysed variables, especially the distinct position of Śląskie and Mazowieckie voivodships. These two voivodships often showed the highest variation compared to other ones as evidenced by the context of the computed distance and emphasised by the creation of separate clusters. The lowest variation was observed for the pair of Małopolskie and Podkarpackie voivodships, especially in 2010, 2015 and 2019.

[3] NTS – Nomenclature of Territorial Units in Polish Statistics, prepared based on European Nomenclature of Territorial Units of Statistics (NUTS), used in collection, harmonisation, and data release of statistics of regional countries in EU.

Moreover, the obtained clusters emphasise the spatial variation of the voivodships analysed from the point of view of demographic ageing and was confirmed by the composition of the clusters. In many cases, voivodships with the lowest or highest value of the indicator measuring demographic ageing were assigned by the applied algorithm to the same clusters. As a result, the presented outcomes also take into account the spatial variation of demographic ageing, even if the headline indicator for demographic ageing (i.e. the percentage share of people aged 65 and over in the total voivodship population) was excluded from the analysis due to collinearity. However, in the research, the ageing of the population is captured by the variable expressing the number of older people per 1,000 inhabitants of the voivodship. By contrast, in the context of the final set of the chosen determinants, it seems that the distinct position of Śląskie Voivodship was mainly related to the variables that reflect the pension security system.

The study shows the composition of the clustered voivodships analysed from the perspective of the used indicators, presenting the creation of the maximum and minimum distance of every voivodship in 2005, 2010, 2015 and 2019, as well as highlighting the fact that the spatial variation of voivodships is consistent with the demographic ageing observed in Poland. These results could prove informative in policy-oriented discussions and impact the design of socio-economic strategies aimed to improve the situation of older people, especially in the context of the increasing demographic ageing of the Polish society.

Acknowledgements

This research was supported by the National Science Centre in Poland (grant number 2017/27/N/HS4/01878).

References

Bailey, K. D. (1994). *Typologies and taxonomies: An introduction to classification techniques.* Thousand Oaks.

Bernstein, I. H., Garbin, C. P., & Teng, G. K. (1988). *Applied Multivariate Analysis*. Springer-Verlag. https://doi.org/10.1007/978-1-4613-8740.

Długosz, Z. (1998). Próba określenia zmian starości demograficznej Polski w ujęciu przestrzennym. *Wiadomości Statystyczne*, *43*(3), 15–27.

European Commission. (2021). *The 2021 Ageing report. Economic & Budgetary Projections for the EU Member States (2019–2070)*. Publications Office of the European Union.

Everitt, B. S., Landau, S., Morven, L., & Stahl, D. (2011). *Cluster Analysis* (5th edition). John Wiley & Sons.

Główny Urząd Statystyczny. (2014). *Sytuacja demograficzna osób starszych i konsekwencje starzenia się ludności Polski w świetle prognozy na lata 2014–2050*. https://stat.gov.pl/obszary-tematyczne/ludnosc/ludnosc/sytuacja-demograficzna-osob-starszych-i-konsekwencje-starzenia-sie-ludnosci-polski-w-swietle-prognozy-na-lata-2014-2050,18,1.html.

Holzer, J. Z. (2003). *Demografia*. Polskie Wydawnictwo Ekonomiczne.

Kaufman, L., & Rousseeuw, P. J. (2005). *Finding groups in data: An introduction to cluster analysis*. John Wiley & Sons. https://doi.org/10.1002/9780470316801.

Kowaleski, J. T. (Ed.). (2011). *Przestrzenne zróżnicowanie starzenia się ludności Polski. Przyczyny, etapy, następstwa*. Wydawnictwo Uniwersytetu Łódzkiego.

Kowaleski, J. T., & Majdzińska, A. (2012). Miary i skale zaawansowania starości demograficznej. In A. Rossa (Ed.), *Wprowadzenie do gerontometrii* (pp. 7–34). Wydawnictwo Uniwersytetu Łódzkiego. https://doi.org/10.18778/7525-788-5.02.

Local Data Bank of Statistics Poland. https://bdl.stat.gov.pl/BDL/start, access: July, 2021.

Majdzińska, A. (2017). Zróżnicowanie terytorialne starzenia się ludności Polski. *Acta Universitatis Lodziensis. Folia Oeconomica*, *5*(331), 71–90. https://doi.org/10.18778/0208-6018.331.05.

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

Panek, T., & Zwierzchowski, J. (2013). *Statystyczne metody wielowymiarowej analizy porównawczej. Teoria i zastosowania*. Oficyna Wydawnicza SGH.

Podogrodzka, M. (2016). Przestrzenna konwergencja indeksu starości w Polsce. *Acta Universitatis Lodziensis. Folia Oeconomica*, *4*(324), 51–65. https://doi.org/10.18778/0208-6018.324.04.

Romesburg, H. C. (2004). *Cluster analysis for researchers*. Lulu Press.

Roszkowska, E. (2020). Similarity of regions in terms of the structure of the elderly population – proposition of a measure. *Optimum. Economic Studies*, (2), 148–162. http://doi.org/10.15290/oes.2020.02.100.11.

Sarstedt, M., & Mooi, E. (2014). *A Concise Guide to Market Research. The Process, Data, and Methods Using IBM SPSS Statistics* (2nd edition). Springer.

Stanisz, A. (2007). *Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny: vol. 3. Analizy wielowymiarowe*. StatSoft.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*(301), 236–244. https://doi.org/10.2307/2282967.

Zeliaś, A. (Ed.). (2004). *Poziom życia w Polsce i krajach Unii Europejskiej*. Polskie Wydawnictwo Ekonomiczne.

Appendix

Table A1. Selected descriptive statistics

Variable	Mean	Median	Min.	Max.	Variance	Standard deviation
2005						
X_1	13.20	13.31	11.56	14.88	1.40	1.18
X_2	0.91	0.87	0.69	1.14	0.02	0.16
X_3	131.97	133.08	115.62	148.83	139.54	11.81
X_4	128.04	127.63	115.99	144.79	97.51	9.87
X_5	134.82	133.31	95.22	186.15	603.33	24.56
X_6	6.70	6.52	5.64	8.34	0.59	0.77
X_7	50.72	50.11	47.61	56.02	5.22	2.28
X_8	99.39	95.74	75.35	131.79	232.82	15.26
X_9	756.18	761.51	528.08	994.64	12 668.54	112.55
X_{10}	544.02	540.10	472.28	625.99	2786.40	52.79
X_{11}	0.52	0.52	0.39	0.60	≈0.01	0.04
X_{12}	0.55	0.56	0.42	0.64	≈0.01	0.04
X_{13}	2.42	2.44	1.90	2.93	0.09	0.29
2010						
X_1	13.36	13.46	11.68	14.93	1.40	1.19
X_2	1.34	1.27	1.13	1.62	0.03	0.18
X_3	133.56	134.59	116.78	149.31	140.48	11.85
X_4	136.20	136.27	122.44	149.44	89.46	9.46
X_5	127.49	129.18	89.78	176.51	564.21	23.75
X_6	6.83	6.79	5.78	8.44	0.57	0.76
X_7	51.12	50.86	47.61	56.50	5.57	2.36
X_8	125.82	124.44	103.29	162.49	279.03	16.70
X_9	946.01	930.87	718.37	1142.66	15 364.99	123.96
X_{10}	664.43	659.08	611.37	725.26	1340.35	36.61
X_{11}	0.52	0.52	0.41	0.59	≈0.01	0.04
X_{12}	0.55	0.55	0.44	0.63	≈0.01	0.04
X_{13}	3.26	3.17	2.55	4.18	0.24	0.49
2015						
X_1	15.74	15.74	14.05	17.64	1.02	1.01
X_2	1.80	1.76	1.54	2.25	0.05	0.23
X_3	157.39	157.49	140.52	176.44	102.44	10.12
X_4	168.60	170.18	149.44	185.57	103.55	10.18
X_5	139.97	138.54	104.47	178.42	389.67	19.74
X_6	7.53	7.43	6.59	9.07	0.50	0.71
X_7	47.77	47.50	44.88	51.41	2.55	1.60
X_8	127.47	127.05	104.31	164.42	234.80	15.32
X_9	810.15	804.60	648.38	986.42	6710.87	81.92
X_{10}	689.89	685.87	645.80	739.78	811.46	28.49
X_{11}	0.53	0.53	0.43	0.60	0.01	0.04
X_{12}	0.56	0.56	0.45	0.64	0.01	0.04
X_{13}	3.38	3.53	2.56	4.11	0.21	0.45

Table A1. Selected descriptive statistics (cont.)

Variable	Mean	Median	Min.	Max.	Variance	Standard deviation
2019						
X_1	18.09	17.96	16.59	20.12	1.21	1.10
X_2	2.12	2.10	1.75	2.52	0.06	0.24
X_3	80.93	179.64	165.90	201.25	120.62	10.98
X_4	198.20	197.25	174.07	219.74	133.95	11.57
X_5	155.70	153.58	121.24	183.82	297.76	17.26
X_6	8.15	8.18	7.27	9.49	0.39	0.63
X_7	45.03	45.08	42.95	47.17	1.15	1.07
X_8	148.56	144.02	125.99	183.58	234.14	15.30
X_9	820.70	821.42	708.29	955.77	3687.05	60.72
X_{10}	749.10	744.97	713.89	788.03	520.41	22.81
X_{11}	0.47	0.47	0.39	0.54	0.01	0.03
X_{12}	0.50	0.50	0.41	0.57	0.01	0.03
X_{13}	3.14	3.18	2.50	3.69	0.11	0.33

Note. Number of observations – 16.
Source: author's calculations based on data from the Local Data Bank of Statistics Poland.

Table A2. Correlation matrix

Variable	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}
2005													
X_1	1.00	.	.	.	.	.	.	.	.	.	.	.	.
X_2	0.84	1.00	.	.	.	.	.	.	.	.	.	.	.
X_3	1.00	0.84	1.00	.	.	.	.	.	.	.	.	.	.
X_4	0.46	0.36	0.46	1.00	.	.	.	.	.	.	.	.	.
X_5	0.88	0.76	0.88	0.02	1.00	.	.	.	.	.	.	.	.
X_6	0.92	0.88	0.92	0.38	0.83	1.00	.	.	.	.	.	.	.
X_7	0.34	0.53	0.34	0.05	0.36	0.68	1.00	.	.	.	.	.	.
X_8	0.26	−0.01	0.26	0.72	−0.03	0.21	0.01	1.00	.	.	.	.	.
X_9	−0.32	−0.49	−0.32	0.44	−0.54	−0.33	−0.21	0.83	1.00	.	.	.	.
X_{10}	0.58	0.33	0.58	0.49	0.45	0.39	−0.16	0.59	0.24	1.00	.	.	.
X_{11}	−0.36	−0.41	−0.36	−0.35	−0.18	−0.22	0.14	0.04	0.25	−0.43	1.00	.	.
X_{12}	−0.34	−0.41	−0.34	−0.28	−0.20	−0.20	0.13	0.11	0.31	−0.39	0.99	1.00	.
X_{13}	0.01	−0.07	0.01	−0.29	0.24	0.08	0.17	0.09	0.08	−0.13	0.65	0.64	1.00
2010													
X_1	1.00	.	.	.	.	.	.	.	.	.	.	.	.
X_2	0.80	1.00	.	.	.	.	.	.	.	.	.	.	.
X_3	1.00	0.80	1.00	.	.	.	.	.	.	.	.	.	.
X_4	0.62	0.43	0.62	1.00	.	.	.	.	.	.	.	.	.
X_5	0.88	0.74	0.88	0.19	1.00	.	.	.	.	.	.	.	.
X_6	0.91	0.86	0.91	0.57	0.79	1.00	.	.	.	.	.	.	.
X_7	0.23	0.49	0.23	0.14	0.20	0.62	1.00	.	.	.	.	.	.
X_8	0.29	−0.01	0.29	0.67	−0.05	0.29	0.12	1.00	.	.	.	.	.
X_9	−0.38	−0.54	−0.38	0.22	−0.63	−0.32	−0.03	0.77	1.00	.	.	.	.
X_{10}	0.65	0.47	0.65	0.56	0.49	0.53	0.00	0.44	−0.01	1.00	.	.	.
X_{11}	−0.31	−0.58	−0.31	−0.28	−0.19	−0.25	0.00	0.16	0.35	−0.46	1.00	.	.
X_{12}	−0.31	−0.60	−0.31	−0.24	−0.23	−0.26	−0.01	0.20	0.40	−0.43	0.99	1.00	.
X_{13}	−0.41	−0.43	−0.41	−0.59	−0.14	−0.36	−0.06	0.01	0.28	−0.28	0.50	0.50	1.00

Table A2. Correlation matrix (cont.)

Variable	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}
							2015						
X_1	1.00	.	.	.	.	.	.	.	.	.	.	.	.
X_2	0.71	1.00	.	.	.	.	.	.	.	.	.	.	.
X_3	1.00	0.71	1.00	.	.	.	.	.	.	.	.	.	.
X_4	0.67	0.20	0.67	1.00	.	.	.	.	.	.	.	.	.
X_5	0.74	0.81	0.74	0.07	1.00	.	.	.	.	.	.	.	.
X_6	0.97	0.75	0.97	0.60	0.79	1.00	.	.	.	.	.	.	.
X_7	0.77	0.70	0.77	0.37	0.77	0.90	1.00	.	.	.	.	.	.
X_8	0.49	−0.10	0.49	0.64	−0.04	0.40	0.16	1.00	.	.	.	.	.
X_9	−0.06	−0.55	−0.06	0.33	−0.51	−0.14	−0.29	0.84	1.00	.	.	.	.
X_{10}	0.62	0.66	0.62	0.35	0.47	0.55	0.34	0.26	−0.09	1.00	.	.	.
X_{11}	−0.12	−0.53	−0.12	−0.12	−0.05	−0.10	−0.04	0.20	0.30	−0.59	1.00	.	.
X_{12}	−0.10	−0.54	−0.10	−0.07	−0.08	−0.09	−0.05	0.25	0.34	−0.55	0.99	1.00	.
X_{13}	−0.25	−0.12	−0.25	−0.61	0.17	−0.14	0.11	−0.14	−0.01	−0.08	0.27	0.25	1.00
							2019						
X_1	1.00	.	.	.	.	.	.	.	.	.	.	.	.
X_2	0.61	1.00	.	.	.	.	.	.	.	.	.	.	.
X_3	1.00	0.61	1.00	.	.	.	.	.	.	.	.	.	.
X_4	0.77	0.24	0.77	1.00	.	.	.	.	.	.	.	.	.
X_5	0.67	0.86	0.67	0.20	1.00	.	.	.	.	.	.	.	.
X_6	0.97	0.65	0.97	0.69	0.73	1.00	.	.	.	.	.	.	.
X_7	0.55	0.51	0.55	0.21	0.65	0.75	1.00	.	.	.	.	.	.
X_8	0.68	0.00	0.68	0.63	0.10	0.61	0.21	1.00	.	.	.	.	.
X_9	0.11	−0.50	0.11	0.25	−0.42	0.04	−0.16	0.81	1.00	.	.	.	.
X_{10}	0.42	0.70	0.42	0.13	0.42	0.44	0.35	0.06	−0.26	1.00	.	.	.
X_{11}	−0.01	−0.42	−0.01	−0.05	0.02	−0.01	0.01	0.20	0.27	−0.69	1.00	.	.
X_{12}	0.02	−0.42	0.02	−0.01	0.00	0.01	0.00	0.25	0.31	−0.66	0.99	1.00	.
X_{13}	−0.37	0.01	−0.37	−0.65	0.13	−0.24	0.18	−0.23	−0.04	0.07	0.10	0.06	1.00

Source: author’s calculations based on data from the Local Data Bank of Statistics Poland.