ON THE POTENTIAL FOR USING SELECTED PCA-BASED METHODS TO ANALYZE THE CRIME RATE IN POLAND

Małgorzata Misztal

University of Lodz, Lodz, Poland
e-mail: malgorzata.misztal@uni.lodz.pl

ORCID: 0000-0002-8719-2097

Abstract: The aim of the paper is to assess the potential for using some selected PCA-based methods to analyze the spatial diversity of crime in Poland during 2000-2017. Classical principal components analysis (PCA) deals with two-way matrices, usually taking into account objects and variables. In the case of data analyzed in the study, apart from two dimensions (*objects* – voivodships, *variables* – criminal offences), there is also the dimension of time, so the dataset can be seen as data cube: *objects × variables × time*. Therefore, this type of data requires the use of methods handling three-way data structures. In the paper the variability of some selected categories of criminal offences in time (2000--2017) and space (according to voivodships) is analyzed using the between-class and the within-class principal component analysis. The advantage of these methods is, among others, the possibility of the graphical presentation of the results in two-dimensional space with the use of factorial maps.

Keywords: crime, criminal offence, multivariate exploratory data analysis, principal component analysis, factorial maps.

1. Introduction

The Polish Penal Code does not provide a strict definition of a crime, however, it can be defined as "an act or omission of a human being, prohibited by law under the threat of punishment as a crime or a misdemeanor, committed intentionally or unintentionally and socially harmful more than negligibly" [Gardocki 2013, p. 30].

The phenomenon of crime is the subject of research of various scientific disciplines: sociology, psychology, criminology and economics. It also arouses the interest of the public and politicians. As Florczak [2013, p. 7] emphasizes, "ensuring social order and the rule of law is one of the fundamental and universal tasks of all organized societies". In addition, high crime negatively affects the growth of social capital, which in the long term translates significantly into economic growth

[Florczak 2013, p. 7]. The high social costs related to the crime rate make it necessary to conduct research, both in terms of searching for the determinants of crime and the methods of analyzing the phenomenon of crime in a multidimensional perspective.

In Polish literature the most common methods of studying changes in the crime rate over time are the basic methods of trend analysis, statistical indicators (indexes) and simple graphical presentations (line or bar charts). Such analyzes can be found for example in the *Reports on the State of Security in Poland*, available on the police website[1]. More advanced methods of surveying the crime rate in Poland, taking into account potential factors affecting the crime rate and spatial diversity, include: econometric models (see e.g. [Sztaudynger, Sztaudynger 2003; Bieniek, Cichocki, Szczepaniec 2012; Florczak 2013; Kądziołka 2014; 2015a; Lusawa 2016]), spatial data analysis methods (see e.g. [Mordwa 2013; Kądziołka 2016a]), and panel models (e.g. [Kądziołka 2015b]). A proposal to analyze crime in Poland with the use of Weber's median-based taxonomic measure of development was presented in Bąk [2015], Hellwig's taxonomic measure of development for the analysis of public safety in European countries was used by Wierzbicka and Żółtaszek [2015] and the methods of hierarchical clustering were applied by Kądziołka [2015c; 2016b]. An attempt to use redundancy analysis to investigate crime against property in Poland was made in the papers of Misztal [2017; 2018]. The classical principal component analysis was applied by Leżoń [2015] to study the crime rate in EU countries. Classical PCA is also a popular method of studying the level of crime in Nigeria (see e.g. [Olakorede, Adams, Olanrewaju 2017; Faweya, Adeniran, Balogun 2018]) or in Pakistan [Ali, Razzak 2015].

In this article, the analysis of spatial diversity of selected crime categories in Poland in the period 2000-2017 was performed using some PCA-based methods. The advantage of these methods is the possibility of a graphical presentation of the results with the use of factorial maps. It should be noted that classical PCA and PCA-based methods are exploratory data analysis methods. This is "an approach to data analysis that emphasizes the use of informal graphical procedures not based on prior assumptions about the structure of the data or on formal models for the data" [Everrit, Skrondal 2010, p. 157]. The aim of such an analysis is to detect structures and general patterns in relations between different variables, as well as the description and classification of objects characterized by these variables in multidimensional spaces, with minimal use of formal mathematics or statistical methods.

The objective of the paper is to assess the potential for using some selected PCA-based exploratory methods to analyze the crime rate in Poland. In particular, the research is aimed at answering the following questions: (1) Are there any relationships between the individual offence categories? (2) Is it possible to identify clusters of voivodships with regard to the level of the analyzed categories of crimes?

---

[1] For example: https://bip.mswia.gov.pl/bip/raport-o-stanie-bezpie/18405,Raport-o-stanie-bez-pieczenstwa.html (access: 24.03.2019).

(3) Does the crime rate change over time? (4) What are the advantages of using the proposed methods to investigate the crime rate in Poland?

2. Data sources and methods

For the purpose of the study the data made available on the police website[2] were used. Datasets for the voivodships cover the period 2000-2017. Analyzing the three main categories of crime: criminal, economic and traffic offences, it can be easily seen that criminal offences account for the vast majority of all offences committed (Figure 1).
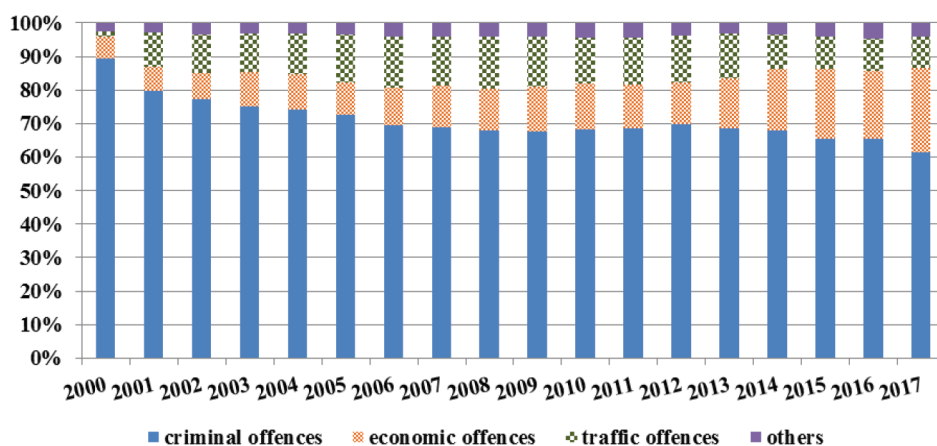

Fig. 1. Structure of crimes recorded in Poland in 2000-2017

Source: own elaboration.

Therefore, further analysis considered only ten selected types of criminal offences: X1 – property damage, X2 – robbery, X3 – burglary, X4 – car theft, X5 – property theft, X6 – participation in violence or assault, X7 – damage to health, X8 – homicide, X9 – rape and X10 – drugs (crimes under The Act On Counteracting Drug Addiction). The choice of these types of crime was dictated mainly by the data availability for individual voivodships. As the voivodships differ in terms of population, all the analyzed crime types were presented in the form of crime intensity per 100 thousand inhabitants.

To assess the differences in the intensity of criminal offences, taking into account the time (18 years: 2000-2017) and space (16 voivodships), an approach using PCA--based methods was proposed.

---

[2] http://www.statystyka.policja.pl (access: 25.03.2019).

Principal component analysis (PCA; [Pearson 1901; Hotelling 1933]) is a method of transforming the original variables into new ones that are uncorrelated and account for decreasing proportions of the variance in the data. These new variables (the principal components) are linear combinations of the original variables. PCA emphasizes correlation structures between variables and provides an ordination of objects that underlies this correlation structure [Thioulouse et al. 2015, p. 85]. The graphical presentation of the analyses with the use of factorial maps (scatterplots and biplots) plays an important role in the interpretation of the PCA results.

Taking into account the period 2000-2017, it is possible to perform 18 separate PCAs, one for every year, or one PCA after concatenating all datasets. However, the analysis of PCA results obtained for the combined datasets can be difficult to interpret and the visualization of the results is often not clear since many labels are superimposed. Star plots with ellipses are then used to facilitate the interpretation of results.

The classical PCA carried out for concatenated datasets mixes both the temporal and the spatial typologies, and the researcher can often decide to separate these two processes [Dufour 2008, p. 7]. Taking the existence of groups of samples in a data table into consideration, Thioulouse et al. [2015, p. 141] suggest using a particular type of analysis, called a between-class analysis, which models the differences between groups by computing the group means and analyses the resulting table. The between-class analysis aims at visually checking the existence of groups and describing the main characteristics of the differences between the groups. In the study, the between-class analysis based on principal component analysis was applied.

Generally speaking, the basis for the between-class analysis is the table of group means. In the presented research, two analyses may be performed: (1) groups correspond to individual years (the values of each variable are averaged across the 16 voivodships) and (2) groups correspond to individual voivodships (the values of each variable are averaged across the 18 years). A more formal description of these methods is presented by Thioulouse et al. [2015].

The between-class analysis may be complemented by the within-class analysis. Within-class analysis operates with the residuals between observed data and the groups means and it aims at looking for structures remaining in the data after removing the differences between groups of samples. The mathematical basis of the within-class PCA can be found in [Thioulouse et al. 2015].

Dufour [2008, pp. 17-18] points out that the within-class and between-class PCA can be regarded as an exploratory generalization of the one-way ANOVA. For each of the considered effects (spatial or temporal) the total inertia (variability) of X (the matrix containing p variables measured on n objects) can be decomposed in two parts: the inertia of X^- (the within model after removing the effect of groups) and the inertia of X^+ (the between model, revealing the effect of groups).

Both methods: the between-class and the within-class PCA were applied in the study to analyze the crime rate in Poland. All the calculations were performed with the use of the R-environment (packages: ade4, adegraphics).

On the potential for using selected PCA-based methods to analyze...

19

3. Results

Figure 2 shows the intensity of criminal offences (in total, per 100 thousand inhabitants) in the period 2000-2017, by voivodships. In addition, the number of criminal offences per 100 thousand inhabitants for Poland is also presented.
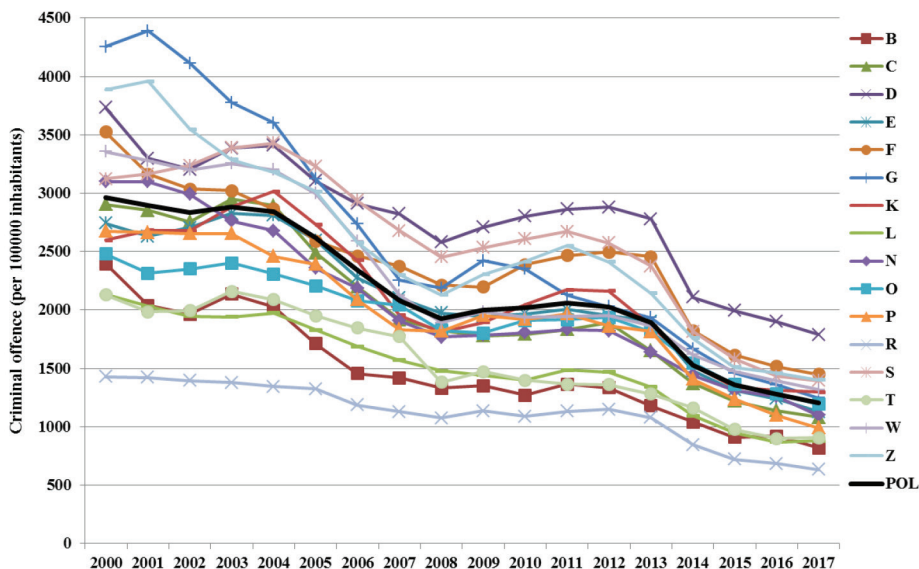
D – Dolnośląskie Voivodship; C – Kujawsko-Pomorskie Voivodship; L – Lubelskie Voivodship; F – Lubuskie Voivodship; E – Łódzkie Voivodship; K – Małopolskie Voivodship; W – Mazowieckie Voivodship; O – Opolskie Voivodship; R – Podkarpackie Voivodship; B – Podlaskie Voivodship; G – Pomorskie Voivodship; S – Śląskie Voivodship; T – Świętokrzyskie Voivodship; N – Warmińsko-Mazurskie Voivodship; P – Wielkopolskie Voivodship; Z – Zachodniopomorskie Voivodship, POL – Poland.

Fig. 2. Changes of the intensity of criminal offences (in total) in period 2000-2017 by voivodships
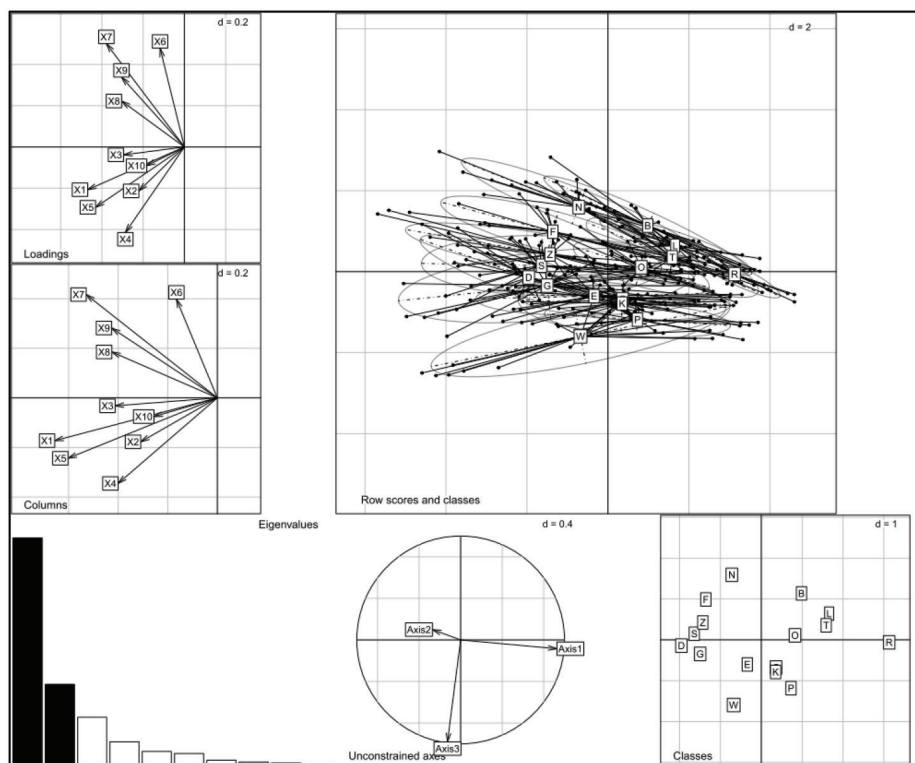
Source: own elaboration.

It can be seen that the changes in the crime rate during 2000-2017 in individual voivodships were rather similar. The number of crimes per 100 thousand inhabitants decreased until 2008, then slowly increased until 2013, and started to decrease again. The voivodships differed in the intensity of criminal offences. The lowest crime intensity was observed in the Podkarpackie Voivodship (R) and the highest in the Pomorskie (G) and Dolnośląskie (D) voivodships. It can also be noted that in the last few years (from 2014) the differences between voivodships have decreased with regard to the level of crime.

To reveal the differences between groups corresponding to individual voivodships and to individual years, two between-class PCAs were performed. Each of them was complemented by the within-class PCA.

3.1. The between-voivodships PCA results

The results of the between-class PCA for groups corresponding to voivodships are presented in Figure 3. The total variability (inertia) in classical PCA (performed on the combined datasets) equals the number of variables (ten in this research). In the presented analysis, the between-class inertia is equal to 3.541, i.e. 35.41% of the total inertia is due to the spatial factor[3].

There are six elementary graphs in Figure 3. The main one (top-right: "Row scores and classes") presents the row scores for the initial data table – the 18 years for each voivodship are grouped with the eighteen-pointed star and an ellipse. Each star and ellipse is labelled with a letter identifying the voivodship, located at the centroid of the star. In addition, the bottom-right graph ("Classes") shows the row scores of the between-class analysis for the 16 voivodships.

Fig. 3. The between-voivodships PCA results

Source: own elaboration (ade4 package).

---

[3] It is also possible to check the significance of the differences between voivodships using the permutation test [Thioulouse et al. 2015, p. 151]. The simulated *p*-value is equal to 0.001, so the differences are highly significant.

The graph labelled "Unconstrained axes" presents the projection of first three axes of the classical PCA into the between-class analysis and it shows the relationships between the classical PCA and the between-class PCA. The first axis in the between--class PCA is the combination of the first and the second axes of the classical PCA. The second axis in the between-class analysis is equivalent to the third axis of the classical PCA.

Eigenvalues are presented with the use of a classical bar chart (in the lower-left graph). The first two eigenvalues are greater than 1.

The last two graphs on the left are labelled "Loadings" and "Columns". They both present the ten variables (the types of crime) used in the analysis and they should be comparable since large differences between these two graphs indicate that the analysis results are not coherent. The "Loadings" graph gives the coefficients of the linear combination that maximize the between-class to total variance ratio and the "Columns" graph shows the scores of the variables (see [Thioulouse et al. 2015, p. 149]). The angles between all vectors on the correlation circle reflect their linear correlations. The direction of the vector corresponds to the direction of the highest variability of a given variable and its length is proportional to the meaning of this variable. Two sets of strongly positively correlated variables can be observed: (1) X6 (participation in violence or assault), X7 (damage to health), X8 (homicide), X9 (rape) and (2) X1 (property damage), X2 (robbery), X3 (burglary), X4 (car theft), X5 (property theft) and X10 (drugs-related crime).

The angles between the vectors representing the set of variables and the principal components (axes) can be used to assess the linear correlation coefficients. The correlation circle shows that the first axis corresponds to the property crime gradient (represented mainly by X1, X5) with high levels of property crime toward the left and absence of property crime on the right. The second axis is strongly positively correlated with X6 and X7 (therefore it corresponds to the crimes against life and health gradient).

The points representing the voivodships can be projected perpendicularly onto the vectors visualizing variables to obtain the approximate ordering of the voivodships in order of increasing intensity of particular crimes (Figure 4[4]).

Analyzing Figure 4 one can observe, among others, that the Podkarpackie (R) Voivodship has the lowest crime rate in all the considered crime categories (on average in the surveyed years). The Mazowieckie Voivodship (W) is the leader due to car theft (X4); at the opposite end there are the Podlaskie (B) and Lubelskie (L) voivodships. The Dolnośląskie (D), Śląskie (S) and Pomorskie (G) voivodships are characterized by a high level of crime against property (X1-X5) and a relatively high

---

[4] The plot presented in Figure 4 is called a "biplot". That term was introduced by Gabriel [1971] and the simplest definition of this term was given by Gower et al. [2015, p. 42]: "A biplot is exactly what it says. It is a plot of two kinds of information displayed together. The 'bi' in biplot refers to the two kinds of information and not to the usual, but not necessary, use of two dimensions". The plot has been scaled up to make it easier to analyse the results.
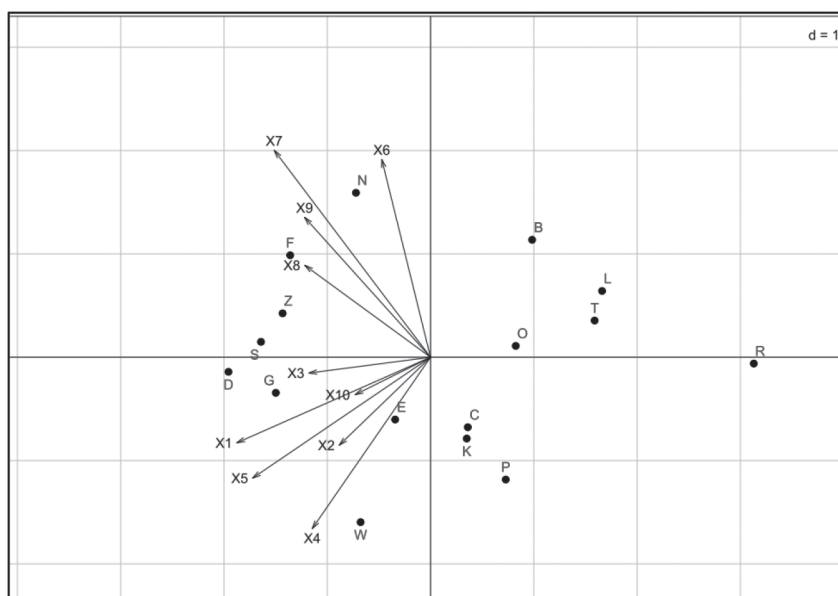
Fig. 4. The between-voivodships PCA biplot

Source: own elaboration (ade4 package).

level of crime against life and health. The Warmińsko-Mazurskie (N) and Lubuskie (F) voivodships are the regions with the highest rates of crime against life and health (X6-X9). The points representing the Łódzkie (E) and Kujawsko-Pomorskie (C) voivodships are located near the coordinate system origin – this indicates an average crime rate in these voivodships in all of the considered crime categories.

To analyze in more detail the similarities and differences between the voivodships and to assess the diversity of crime rates in each voivodship, it is recommended to take a look at the star plots with ellipses. In order to make the analysis easier, the star plots for five selected[5] voivodships are presented in Figure 5.

The 18 years for each voivodship are grouped with the eighteen-pointed star and an ellipse. Each star is labelled with the letter identifying the voivodship, located at the center of gravity of the star.

The ellipse shape makes it possible to assess the diversity of the crime rate within each voivodship. The smallest ellipse is observed for the Podkarpackie Voivodship (R) – this means the smallest variability in the crime rate in that voivodship over the period 2000-2017.

_____

[5] The voivodships were selected due to their location in the two-dimensional between-class PCA space. D and R are located at the opposite ends of the first axis, W and N – at the opposite ends of the second axis and E is near the origin of the coordinate system.
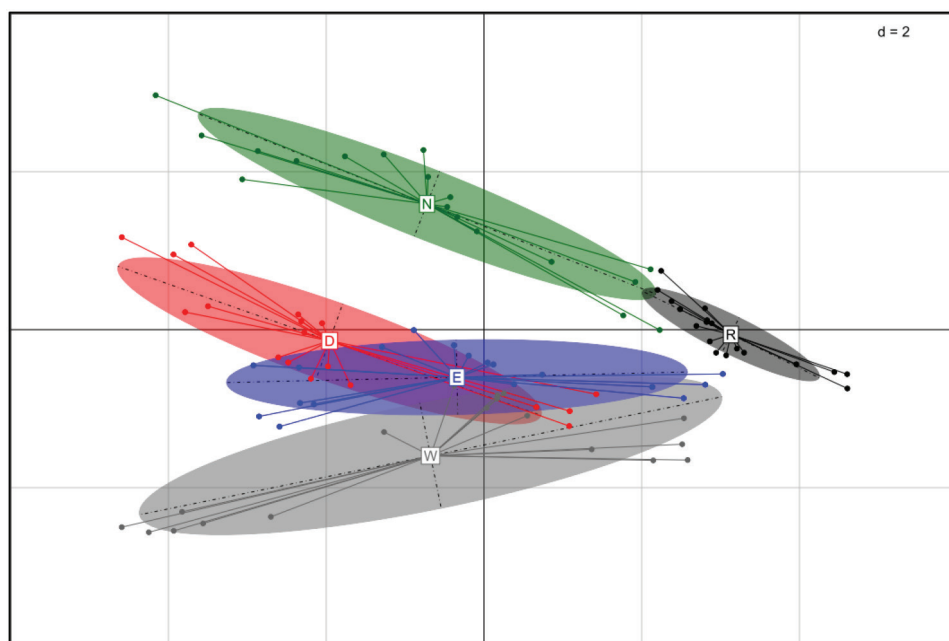
Fig. 5. Star plots with ellipses for five selected voivodships
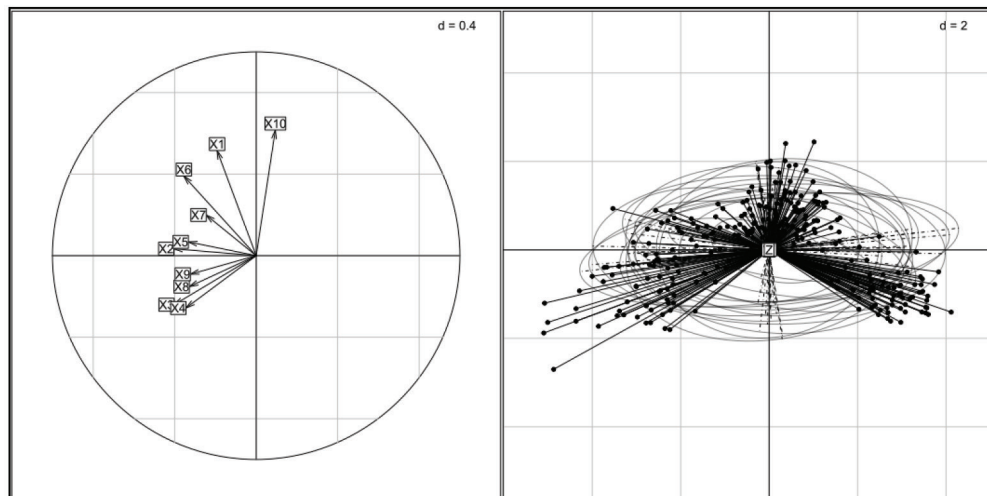
Source: own elaboration (ade4 package).

All the ellipses have elongated horizontal axes but they are differently oriented. The ellipses for the Dolnośląskie (D), Warmińsko-Mazurskie (N) and Podkarpackie (R) voivodships are oriented left-up – this means high variability in the analyzed years due to the intensity of crime against life and health (X6-X9). The ellipse for the Mazowieckie Voivodship is oriented left-down, this means a high variability within the voivodship due to the intensity of some crimes against property: X2 (robbery), X4 (car theft) and X5 (property theft). The ellipse for the Łódzkie Voivodship (E) is located parallel to the first axis of the analysis, representing the property crime gradient. This means a high variability with regard to the level of crime against property, while the rate of crime against life and health is quite stable.

To sum up the results of this analysis, it can be observed that the rays of each star, corresponding to the individual years, can be clustered into three sets, in particular – most to the left – the period 2000-2005, most to the right – the period 2014-2017. That means a decrease in the crime rate in voivodships in the analyzed years.

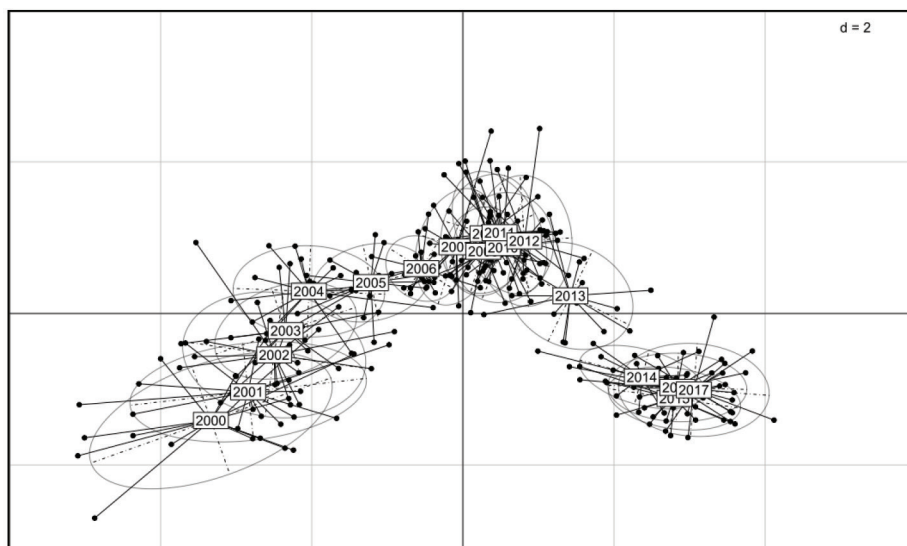### 3.2. The within-voivodships PCA results

The between-voivodships PCA can be complemented by the within-voivodships PCA. As mentioned above, the within-class PCA is based on the residuals between

observed data and the groups' means and it aims at looking for structures remaining in the data after removing differences between groups of samples.

Fig. 6. The within-voivodships PCA results; left panel: correlation circle, right panel: row coordinates grouped by voivodships

Source: own elaboration (ade4 package).

Fig. 7. The within-voivodships PCA results – the annual variation after removing the spatial effect

Source: own elaboration (ade4 package).

The results of the within-voivodships PCA are presented in Figure 6. The spatial effect has been removed – this can be observed in the right panel of Figure 6 (all these row scores are centered by voivodships and all the 16 labels are superimposed at the center of the plot).

The within-voivodships inertia is equal to 6.459, i.e. 64.59% of the total inertia is due to the within PCA.

The annual variation is shown in Figure 7 (the row scores grouped by years) where a strong temporal structure is visible. This means that after removing the differences related to the spatial effect, there is a clear structure in the data related to the temporal effect. Both the crimes against property and against life and health decrease year by year. On the other hand, the number of drug crimes (per 100,000 inhabitants) is steadily increasing (until 2012).

Therefore it can be expected that the temporal effect will be clearly highlighted in the between-years PCA.

3.3. The between-years PCA results

The results of the between-class PCA for groups corresponding to individual years are presented in Figure 8.

In the between-years analysis the values of each crime intensity are averaged across the voivodships. This removes the spatial component in the dataset and may make the temporal structure more apparent. The first two eigenvalues are greater than 1. The between-class inertia is equal to 4.9276, i.e. 49.28% of the total inertia is due to the temporal factor[6].

The first axis of the between-years PCA is equivalent to the first axis of the classical PCA and is correlated with property crimes (X2-X5), homicides (X8) and rapes (X9). The second axis of the between-years PCA represents the second axis of the classical PCA which is correlated with property damage (X1) and drug crimes (X10).

The centroids corresponding to the individual years can be projected perpendicularly onto the vectors visualizing the analyzed crime categories. The analysis of the biplot presented in Figure 9 provides interesting conclusions with regard to the changes in the crime intensity in the analyzed years.

The points representing years are arranged from left to right – opposite to the vectors corresponding to the X1-X9 variables. It can therefore be said that, from one year to another, both the number of crimes against property and against life and health is constantly decreasing. The exception is drug crime (X10) – an increase in this type of crime can be observed from 2000 to 2012, followed by a slight decrease; from 2013 onwards, the intensity of drug crime has remained fairly stable.

---

[6] The simulated *p*-value is equal to 0.001. This means that the temporal effect is highly significant.
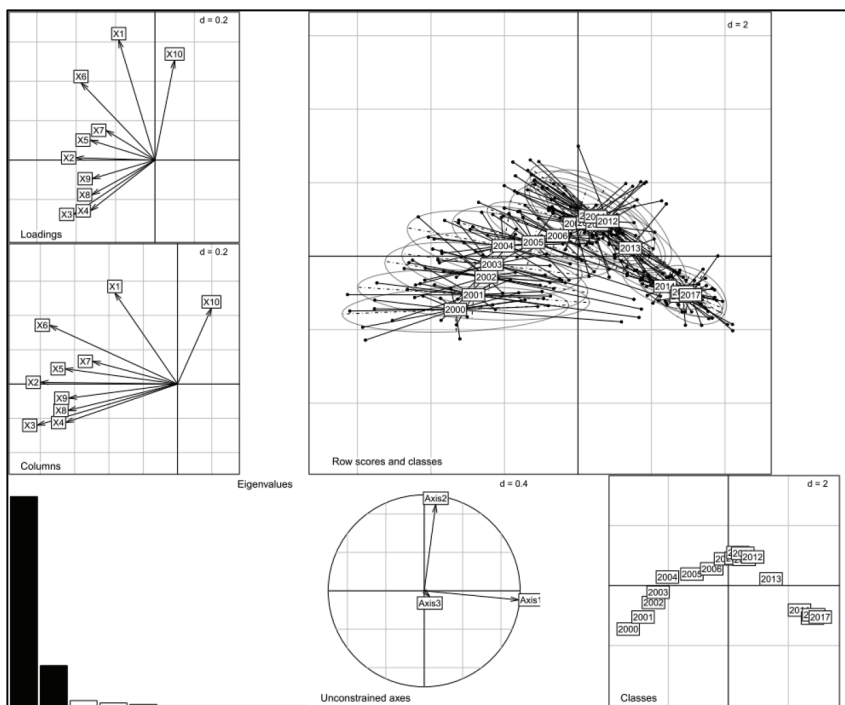
Fig. 8. The between-years PCA results
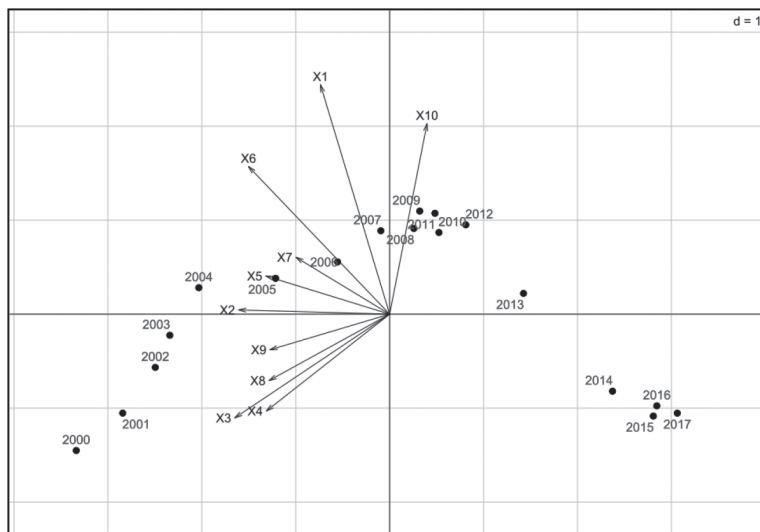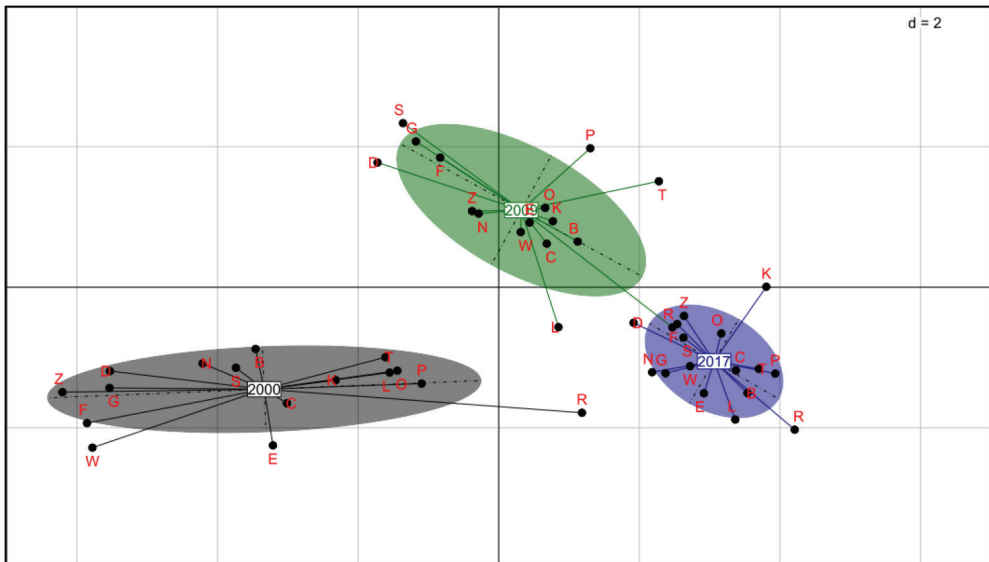
Source: own elaboration (ade4 package).

Fig. 9. The between-years PCA biplot

Source: own elaboration (ade4 package).

In Figure 8 (the row scores and classes graph), the 16 voivodships for each year are grouped with the sixteen-pointed star with an ellipse. The central point of each star is the center of gravity and the points representing the row coordinates for the 16 voivodships are linked to the centroid. The star plots for the three selected[7] years are also shown in Figure 10.

The horizontal arrangement of the long axes of the ellipses along the first between--years PCA axis in period 2000-2005 indicates a large variation between voivodships due to those types of crimes which are most strongly correlated with the horizontal axis (i.e. X2, X3, X4, X9). For later years, a greater diversity of voivodships can be observed due to variables X6, X7. Property damage (X1) and drug crime (X10) have a relatively small impact on the differences between voivodships in the analyzed period of time. Ellipses have been decreasing with time – this means that the differences between the voivodships due to the crime rate start to blur.

When analyzing the arrangement of the rays of each star, corresponding to the individual voivodships, it can be observed, among others, that in every year the Podkarpackie (R) Voivodship is clearly different from the other voivodships, and in particular from the Dolnośląskie (D), Pomorskie (G) and Lubuskie (F) voivodships (Figure 10).

Fig. 10. Star plots with ellipses for 2000, 2009, and 2017

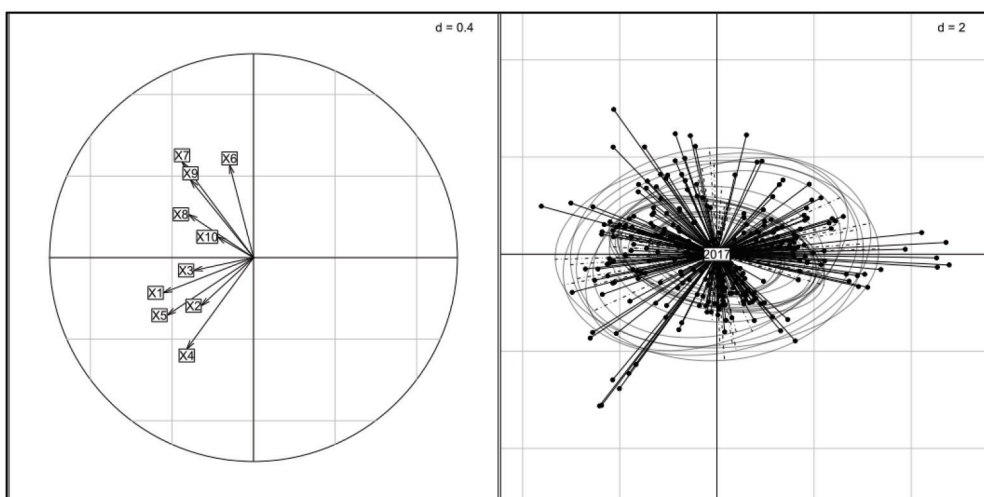Source: own elaboration (ade4 package).

---

[7] Such analyses can be carried out for each year; the first and the last year of the analyzed time interval was selected to highlight the differences between the years and the changes in the crime rate.

It is also possible to identify outlying voivodships in particular years – for example, in 2017 in the Małopolskie Voivodship (K) the highest intensity of drug crime was reported.

3.4. The within-years PCA results

The results of the within-years PCA are presented in Figure 11. The temporal effect has been removed – this can be observed in the right panel of Figure 11 (all these row scores are centered by year and the 18 labels are superimposed at the center of the plot).

The within-years inertia is equal to 5.072, i.e. 50.72% of the total inertia is due to the within PCA.

Fig. 11. The within-years PCA results; left panel: correlation circle, right panel: row coordinates grouped by years

Source: own elaboration (ade4 package).

The spatial variation is shown in Figure 12 (the row scores grouped by voivodships). It can be observed that after removing the differences related to the temporal effect, there is still a structure in the data related to the spatial effect. Moreover, this structure is more pronounced than in the between-voivodships analysis.

The Podkarpackie Voivodship (R) is an outlier – it is characterized by the lowest crime rates in each of the analyzed crime categories. The Dolnośląskie (D), Śląskie (S), Pomorskie (G), Zachodniopomorskie (Z) voivodships are characterized by a high rate of crime against property, at the opposite end there are the Świętokrzyskie

(T), Lubelskie (L) and Podlaskie (B) voivodships. The Warmińsko-Mazurskie (N) and Lubuskie (F) voivodships demonstrate a high crime against life and health rate.
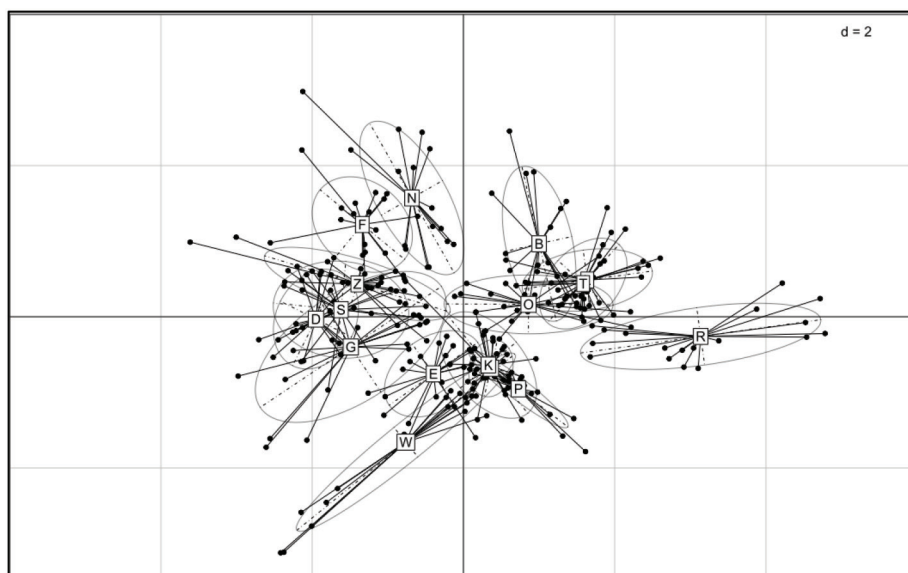
Fig. 12. The within-years PCA results – the voivodships diversity after removing the temporal effect

Source: own elaboration (ade4 package).

4. Conclusions

On the basis of the obtained results and their graphical representation, the variability of the intensity of criminal offences was analyzed taking into account the changes over time (during 2000-2017) and in space (by voivodship). The relationships between the studied crime categories were also determined.

In order to assess the relationships between the individual offence categories, the information provided by the correlation circle should be analyzed. The approximated correlation between two crime categories is equal to the cosine of the angle between the corresponding vectors. Perpendicular vectors indicate the lack of correlation between the variables they represent. An angle less than 90° suggests positive correlation between variables and an angle approaching 180° – strong negative correlation between variables.

The between-voivodships PCA (the data are averaged across 18 years) shows two sets of strongly positively correlated variables: (a) participation in violence or assault, damage to health, homicide and rape – i.e. crimes against life and health, and (b) property damage, robbery, burglary, car theft, property theft – i.e. crimes

against property and drugs-related crime. One can also observe a moderate negative correlation between car theft and participation in violence or assault.

When analyzing the results of the between-years PCA (data are averaged across the 16 voivodships), a negative correlation between the drug crimes' intensity and all the other crime categories, except property damage, should be mentioned. The drug crime rate increases in the subsequent years, while the intensity of other types of crime decreases systematically.

The voivodships are characterized by a fairly high diversity and it is possible to identify clusters of similar voivodships, mainly due to the crime against property rate. The highest property crime rate was observed in the Dolnośląskie, Śląskie, Pomorskie and Mazowieckie voivodships, i.e. in highly urbanized voivodships. Crime against property is typical for urban areas and most of the property crimes are committed in public spaces: parks, shopping centers, railway stations, etc. The Podkarpackie, Świętokrzyskie, Lubelskie and Podlaskie voivodships are at the opposite end, with a low property crime rate. These are voivodships characterized, among others, by a high unemployment rate.

The main structure in the analyzed dataset is the temporal structure. The between-years PCA demonstrated that both the number of crimes against property and against life and health (per 100,000 inhabitants) was steadily decreasing year by year. It should be noted, however, that the noticeable decrease in the intensity of crime against property from 2013 onwards may be the result of the amendment to the law in 2013 changing the method of determining the amount decisive for classifying an act as a crime and not as a misdemeanor (a quarter of the national minimum wage per month). It is also worth noting that in recent years the differences between voivodships have visibly decreased with regard to the analyzed crime categories.

It should be emphasized that all the presented conclusions regarding the spatial diversity of the selected crime types and changes in the level of crime intensity over time were drawn on the basis of the factorial maps (scatterplots and biplots). Factorial maps show the approximate relationships between different points (representing voivodships or years), as well as the relationships between the points and vectors representing the analyzed variables (corresponding to different crime categories). Analysis of factorial maps makes it possible to reveal clusters of points, trends, outliers, patterns in data – all in ways that text and tables do not allow [Gower et al. 2015; Friendly, Sigal 2014]. That is why the PCA-based methods may be recommended as a useful tool for an exploratory and less formal analysis of the data. The next step can be the further examination of data with the use of formal mathematical models. Other advantages of the PCA-based methods are the lack of requirements as to the number of observations compared to the number of variables and the possibility of analyzing numerous variables, including strongly correlated ones.

PCA-based methods (between and within-class analysis) seem to have the potential for their use in the analysis of socio-economic data. So far, the applications

of these methods can only be found in ecological, medical (gene expression) and natural sciences research.

Bibliography

Ali M., Razzak H., 2015, *Principal component analysis of Pakistan crime data*, Sci. Int. (Lahore), 27(5), pp. 4839-4844.

Bąk I., 2015, *Struktura i typologia przestrzenna przestępczości w Polsce*, Ekonometria, 4(50), pp. 43-61.

Bieniek P., Cichocki S., Szczepaniec M., 2012, *Czynniki ekonomiczne a poziom przestępczości – badanie ekonometryczne*, Zeszyty Prawnicze, nr 12(1), pp. 147-172.

Dufour A.B., 2008, *Within PCA and Between PCA*, http://pbil.univ-lyon1.fr/R/pdf/course4.pdf (access: 25.03.2019).

Everitt B.S., Skrondal A., 2010, *The Cambridge Dictionary of Statistics*, Fourth Edition, Cambridge University Press, Cambridge.

Faweya O., Adeniran A.T., Balogun K.O., 2018, *Principal component analysis of the crime rate in Nigeria: a case study of Ekiti and Osun State*, American Journal of Mathematics and Statistics, 8(4), pp. 79-88.

Florczak W., 2013, *Co wywołuje przestępczość i jak ją można ograniczać? Wielowymiarowa analiza makroekonomiczna*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.

Friendly M., Sigal M., 2014, *Some Prehistory of CARME: Visual Language and Visual Thinking*, [in:] Blasius J., Greenacre M. (ed.), *Visualization and Verbalization of Data*, CRC Press, pp. 3-16.

Gabriel K.R., 1971, *The biplot graphical display of matrices with application to principal component analysis*, Biometrika, 58 (3), pp. 453-467.

Gardocki L., 2013, *Pojęcie przestępstwa i podziały przestępstw w polskim prawie karnym*, Annales Universitatis Mariae Curie-Skłodowska Lublin-Polonia, vol. LX(2), Sectio G, pp. 29-40.

Gower J.C., Le Roux N.C., Gardner-Lubbe S., 2015, *Biplots: quantitative data*, WIREs Comput Stat., no. 7, pp. 42-62.

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

Kądziołka K., 2014, *Wpływ wybranych czynników o charakterze społeczno-ekonomicznym na przestępczość przeciwko mieniu w Polsce*, [in:] Szkutnik W. (red.), *Zarządzanie ryzykiem kapitałowym i ubezpieczeniowym oraz społecznymi uwarunkowaniami ryzyka rynku pracy*, Studia Ekonomiczne, nr 181/14, pp. 11-23.

Kądziołka K., 2015a, *Analiza czynników wpływających na przestrzenne zróżnicowanie przestępczości w Polsce na poziomie podregionów*, Współczesna Gospodarka, vol. 6, issue 3, pp. 43-52.

Kądziołka K., 2015b, *Bezrobocie, ubóstwo i przestępczość w Polsce. Analiza zależności na poziomie województw*, Studia Ekonomiczne. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach, nr 242, pp. 71-84.

Kądziołka K., 2015c, *Przestrzenne zróżnicowanie, struktura i dynamika przestępczości w Polsce*, Przestrzeń, Ekonomia, Społeczeństwo, nr 8/II, s. 223-235.

Kądziołka K., 2016a, *Analysis of the crime rate in Poland in spatial and temporal terms*, Central and Eastern European Journal of Management and Economics, vol. 4, no. 1, pp. 81-96.

Kądziołka K., 2016b, *Przestrzenne zróżnicowanie zagrożenia przestępczością w Polsce*, De Securitate et Defensione. O Bezpieczeństwie i Obronności, nr 2(2), pp. 31-43.

Leżoń A., 2015, *Przestępczość w krajach Unii Europejskiej w roku 2012 – rezultaty wielowymiarowej analizy statystycznej*, [in:] Prędki A. (red.), *Wybrane zastosowania narzędzi analitycznych w naukach ekonomicznych*, Mfiles.pl, Kraków, pp. 25-33.

Lusawa R., 2016, *Zróżnicowanie liczby przestępstw stwierdzonych w wybranych powiatach wojewódz-twa mazowieckiego*, Roczniki Naukowe Ekonomii Rolnictwa i Rozwoju Obszarów Wiejskich, t. 103, z. 2, pp. 91-105.

Misztal M., 2017, *On the use of redundancy analysis to study the property crime in Poland*, Acta Universitatis Lodziensis. Folia Oeconomica, 6(332), pp. 99-109.

Misztal M., 2018, *O zastosowaniu analizy redundancji do badania poziomu przestępczości przeciwko mieniu w Polsce w latach 2002-2015*, Taksonomia 31. *Klasyfikacja i analiza danych – teoria i zastosowania*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, nr 508, pp. 157-169.

Mordwa S., 2013, *Zastosowanie GIS w badaniach przestępczości*, Acta Universitatis Lodziensis. Folia Geographica Socio-Economica, 14, pp. 78-92.

Olakorede N.M., Adams S.O., Olanrewaju S.O., 2017, *Principal component analysis of crime data in Gwagwalada Area Command, Abuja from 1995-2015*, American Journal of Theoretical and Applied Statistics, 6(1), pp. 38-43.

Pearson K., 1901, *On lines and planes of closest fit to systems of points in space*, Philosophical Magazine, 6(2), pp. 559-572.

Sztaudynger J.J., Sztaudynger M., 2003, *Ekonometryczne modele przestępczości*, Ruch Prawniczy, Ekonomiczny i Socjologiczny, rok LXV, zeszyt 3, pp. 127-143.

Thioulouse J., Dray S., Dufour A.-B., Siberchicot A., Jombart T., Pavoine S., 2015, *Multivariate Analysis of Ecological Data with ade4*, Springer.

Wierzbicka A., Żółtaszek A., 2015, *Analiza bezpieczeństwa publicznego w krajach europejskich*, Wiadomości Statystyczne, nr 8, pp. 66-80.

O MOŻLIWOŚCI WYKORZYSTANIA WYBRANYCH METOD ANALIZY DANYCH OPARTYCH NA PCA DO BADANIA POZIOMU PRZESTĘPCZOŚCI W POLSCE

Streszczenie: Celem pracy jest ocena możliwości wykorzystania wybranych metod analizy danych, opartych na metodzie głównych składowych (PCA), do badania zróżnicowania przestrzennego wybranych kategorii przestępstw w Polsce w latach 2000-2017. Rozważany w artykule zbiór danych występuje w ujęciu trójwymiarowym w postaci tzw. kostki danych: oprócz dwóch tradycyjnych „wymiarów" (obiekty – województwa, zmienne – przestępstwa kryminalne), pojawia się także „wymiar" czasu. Łączna analiza tego typu danych wymaga wykorzystania innych niż klasyczne metod statystycznych. W pracy dokonano analizy zmienności rozważanych kategorii przestępstw kryminalnych w czasie (w latach 2000-2017) i przestrzeni (według województw), wykorzystując tzw. międzygrupową oraz wewnątrzgrupową analizę głównych składowych. Zaletą tych metod jest m.in. możliwość prezentacji graficznej uzyskanych wyników w przestrzeni dwuwymiarowej z wykorzystaniem map czynnikowych.

Słowa kluczowe: przestępczość, przestępstwa kryminalne, wielowymiarowa eksploracyjna analiza danych, analiza głównych składowych, mapy czynnikowe.