

Scientific Review – Engineering and Environmental Sciences (2018), 27 (3), 310–318
Sci. Rev. Eng. Env. Sci. (2018), 27 (3)
Przegląd Naukowy – Inżynieria i Kształtowanie Środowiska (2018), 27 (3), 310–318
Prz. Nauk. Inż. Kszt. Środ. (2018), 27 (3)
<http://iks.pn.sggw.pl>
DOI 10.22630/PNIKS.2018.27.3.30

Bożena HOŁA, Tomasz NOWOBILSKI

Wydział Budownictwa Lądowego i Wodnego, Politechnika Wroclawska
Faculty of Civil Engineering, Wrocław University of Science and Technology

Klasyfikacja polskich województw pod kątem wybranych wskaźników charakteryzujących budownictwo* **Classification of Polish voivodeships with regards to selected indicators that characterize the construction industry**

Słowa kluczowe: województwo, klasyfikacja, budownictwo, analiza skupień, bezpieczeństwo pracy

Key words: voivodeship, classification, construction industry, cluster analysis, work safety

Wprowadzenie

Na podstawie publikowanych wskaźników rozwoju gospodarczego stwierdza się, że poszczególne województwa Polski są w różnym stopniu rozwinięte gospodarczo (GUS 2009–2017). Nierozłącznym atrybutem każdej działalności gospodarczej jest zjawisko wypadkowości, a jedną z najbardziej wypadkogennych sekcji gospodarki jest branża budowlana (ILO 2014). Analizując dane

statystyczne dotyczące budownictwa, stwierdza się, że wartości wskaźników charakteryzujących budownictwo w poszczególnych województwach są różne (GUS, 2009–2017).

W badaniach naukowych i inżynierskich bardzo często występuje problem identyfikowania obiektów o podobnych cechach. W prowadzeniu takich badań istotne znaczenie ma właściwe zaklasyfikowanie obiektów opisanych wieloma cechami do odpowiednich grup. W takim kontekście przeprowadzono badania, których celem była klasyfikacja województw Polski pod kątem wybranych wskaźników charakteryzujących budownictwo. Do rozwiązania postawionego zadania zastosowano jedną z metod wielowymiarowej analizy sta-

*Artykuł jest wynikiem realizacji przez autorów projektu badawczego 244388 „Model oceny ryzyka wystąpienia katastrof budowlanych, wypadków i zdarzeń niebezpiecznych na stanowiskach pracy z wykorzystaniem rusztowań budowlanych” finansowanego przez NCBiR w ramach Programu Badań Stosowanych na podstawie umowy PBS3/A2/19/2015.

tystycznej – analizę skupień, dostępną w oprogramowaniu Statistica (Stanisz, 2007; Wierzchoń i Kłopotek, 2015).

Proponowana metodyka badań

Przedmiotem badań jest klasyfikacja zbioru województw Polski W pod kątem wybranych wskaźników charakteryzujących budownictwo w aspektach rozwoju gospodarczego i bezpieczeństwa pracy. Każde województwo można opisać wektorem wskaźników I_w :

$$I_w = [I_1, \dots, I_n, \dots, I_N]; w = 1, \dots, W$$

gdzie:

I_n – wskaźnik przyjęty do analizy ($n = 1, 2, 3, \dots, N$).

Zbiór województw W charakteryzuje dwuwymiarowa macierz wskaźników, dana wzorem:

$$I_w = \begin{bmatrix} I_{1,1} & \dots & I_{n,1} & \dots & I_{N,1} \\ I_{1,2} & \dots & I_{n,2} & \dots & I_{N,2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ I_{1,w} & \dots & I_{n,w} & \dots & I_{N,w} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ I_{1,W} & \dots & I_{n,W} & \dots & I_{N,W} \end{bmatrix}$$

gdzie:

w – województwo ($w = 1, 2, 3, \dots, W$),
 n – wskaźnik przyjęty do analizy ($n = 1, 2, 3, \dots, N$).

Przyjęte do obliczeń wartości wskaźników bardzo często różnią się między sobą m.in. jednostką pomiarową lub skalą, co może negatywnie wpłynąć na grupowanie (Stanisz, 2007). Aby temu zaradzić, wszystkie dane liczbowe należy

poddać zabiegowi normalizacji, a wybór właściwej formuły normalizacyjnej zależy od rodzaju danych (Jarocka, 2015). W proponowanej metodyce jako metodę normalizacji przyjęto standaryzację zmiennych według wzoru:

$$P_{n,w} = \frac{I_{n,w} - \bar{I}_n}{\sigma}$$

gdzie:

$P_{n,w}$ – wartości wskaźników po standaryzacji,

$I_{n,w}$ – wartości niestandaryzowanych wskaźników,

\bar{I}_n – wartość średnia w analizowanym zbiorze obiektów,

σ – odchylenie standardowe wartości wskaźnika I_n .

Efektom standaryzacji jest utworzenie zbioru parametrów P_W opisujących analizowany zbiór województw, zapisanych w postaci dwuwymiarowej macierzy, w której każdy wiersz zawiera wartości wszystkich parametrów dotyczących jednego województwa, a każda kolumna zawiera dane dotyczące jednego parametru dla wszystkich województw. Macierz ta dana jest wzorem:

$$P_w = \begin{bmatrix} P_{1,1} & \dots & P_{n,1} & \dots & P_{N,1} \\ P_{1,2} & \dots & P_{n,2} & \dots & P_{N,2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ P_{1,w} & \dots & P_{n,w} & \dots & P_{N,w} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ P_{1,W} & \dots & P_{n,W} & \dots & P_{N,W} \end{bmatrix}$$

Wyliczone parametry stanowią podstawę do grupowania województw z wykorzystaniem analizy skupień. Analiza skupień jest jedną z metod eksploracji

danych, której główną ideą jest pogrupowanie analizowanych obiektów w taki sposób, aby w danej grupie znajdowały się obiekty „podobne” do siebie i zarazem „niepodobne” do obiektów z pozostałych grup (Stanisz, 2007; Wierzchoń i Kłopotek, 2015). Kryterium oceny przynależności obiektu do danej grupy jest miara podobieństwa. W praktycznych rozważaniach, zastosowanie znajduje funkcja odwrotna do miary „podobieństwa”, czyli funkcja „niepodobieństwa” obiektów będąca miarą odległości między nimi. Oznacza to, że jeżeli odległość między obiektem O_a i O_b jest większa niż odległość między obiektami O_a i O_c , czyli:

$$d(O_a, O_b) > d(O_a, O_c); \text{ gdzie } a \neq b \neq c; \\ a, b, c \in \{w\}$$

to obiekt O_a jest bardziej „niepodobny” do obiektu O_b niż do obiektu O_c . W konsekwencji prowadzi to do sytuacji, że obiekty O_a oraz O_c mogą tworzyć skupienie, ponieważ są bardziej do siebie „podobne”. W analizie skupień wykorzystywane są różne miary odległości. Do rozwiązania omawianego zadania zastosowano odległość geometryczną w przestrzeni wielowymiarowej, czyli odległość euklidesową (Stanisz, 2007). Ogólny wzór odległości euklidesowej przyjmuje postać:

$$d(O_a, O_b) = \sqrt{\sum_n (P_{n,a} - P_{n,b})^2}$$

gdzie:

O_a, O_b – badane obiekty, czyli województwa a oraz b , gdzie $a \neq b$, oraz $a, b \in \{w\}$,

$P_{n,a}, P_{n,b}$ – wyznaczone wartości parametrów P_n dla województw a oraz b , gdzie $a \neq b$, oraz $a, b \in \{w\}$.

Do grupowania obiektów proponuje się zastosowanie metody hierarchicznej oraz aglomeracyjnej techniki grupowania. Technika aglomeracyjna, wykorzystywana najczęściej w badaniach (Stanisz, 2007), polega na stopniowym łączeniu obiektów stanowiących odrębne skupienia w nowe skupienia aż do momentu, kiedy wszystkie obiekty utworzą jedno skupienie. Każde połączenie się dwóch skupień nazywane jest krokiem. Istotnym zagadnieniem przy wyznaczaniu właściwej odległości między skupieniami, poza wyborem opisanej wcześniej miary odległości, jest wyznaczenie metody wiązania obiektów. Przeanalizowano różne metody wiązania obiektów, a wśród nich metody, w których odległość jest wyznaczana między konkretnymi miejscami skupień (m.in. danym obiektem lub środkiem ciężkości skupienia), oraz metody, które wykorzystują analizę wariancji – m.in. metoda Warda (Ward, 1963).

Wynikiem hierarchicznej analizy skupień jest wykres w kształcie drzewa, tzw. dendrogram. Obrazuje on, w którym kroku obiekty łączą się ze sobą, nie dając jednak jednoznacznej odpowiedzi co do właściwej liczby skupień. Liczba ta zależy od miejsca odcięcia gałęzi drzewa na wykresie. Dopiero po wybraniu tego miejsca można zidentyfikować skupienia i na ich podstawie dokonać końcowej klasyfikacji analizowanych województw.

Zastosowanie zaproponowanej metodyki na przykładzie klasyfikacji województw Polski

Na podstawie przeglądu literatury przedmiotu (Hoła, 2015; Hoła i Szóstak, 2015; Drozd i Kowalik, 2017; Hoła, Nowobilski, Szer i Szer, 2017; Hoła, Hoła i Szóstak, 2017; Hoła, Sawicki i Szóstak, 2018) oraz danych statystycznych (GUS, 2009–2017) przyjęto następujące wskaźniki do opisu województw: wartość produkcji budowlano-montażowej ($I_{1,r}$), liczbę osób zatrudnionych w budownictwie ($I_{2,r}$), liczbę ludności danego regionu ($I_{3,r}$) oraz liczbę osób poszkodowanych w wypadkach przy pracy ($I_{4,r}$). W obliczeniach uwzględniono dane liczbowe dotyczące lat 2008–2016, uzyskane dla 16 województw Polski, a ich fragment zamieszczono w tabeli 1.

Uzyskane wartości wskaźników poddano zabiegowi standaryzacji. Efektem tego działania było utworzenie parametrów $P_{n,w,r}$ stanowiących zbiór znormalizowanych wskaźników, które

były podstawą klasyfikacji województw. W dalszej części rozdziału zamieszczono sposób wyliczenia przykładowej wartości znormalizowanej dla wskaźnika „wartość produkcji budowlano-montażowej” w 2008 roku. Fragment danych uzyskanych po standaryzacji wszystkich wskaźników przedstawiono w tabeli 2.

Wartość niestandardyzowanego wskaźnika I w 2008 roku dla województwa dolnośląskiego wynosi $I_{1,1,1} = 14\,568,1$. Po wykonaniu obliczeń dla zbioru wszystkich wskaźników opisujących „wartość produkcji budowlano-montażowej” w 2008 roku uzyskano wartość średnią $\bar{I}_{1,1} = 9\,209,4$ oraz odchylenie standardowe: $\sigma = 7\,301,2$. Po podstawieniu do wzoru, otrzymano parametr P_1 (wartość produkcji budowlano-montażowej), którego wartość w 2008 roku dla województwa dolnośląskiego wynosi $P_{1,1,1} = 0,734$:

$$P_{1,1,1} = \frac{I_{1,1,1} - \bar{I}_{1,1}}{\sigma} = \frac{14\,568,1 - 9\,209,4}{7\,301,2} = 0,734$$

TABELA 1. Wartości wybranych wskaźników gospodarczych
TABLE 1. Values of selected economic indicators

v	Wartość produkcji budowlano-montażowej Value of construction and assembly production $I_{1,w,r}$			Liczba osób zatrudnionych w budownictwie Number of people employed in the construction industry $I_{2,w,r}$			Liczba ludności danego regionu Population of a given region $I_{3,w,r}$			Liczba osób poszkodowanych w wypadkach przy pracy Number of people injured in occupational accidents $I_{4,w,r}$		
	2008 $r=1$...	2016 $r=9$	2008 $r=1$...	2016 $r=9$	2008 $r=1$...	2016 $r=9$	2008 $r=1$...	2016 $r=9$
1	14 568,1	...	13 721,2	68 874	...	65 039	2 877 059	...	2 903 710	892	...	413
2	52 88,9	...	6 216,6	43 952	...	42 918	2 067 918	...	2 083 927	607	...	269
...
16	7 325,1	...	9 104	39 288	...	35 483	1 692 957	...	1 708 174	362	...	144

r – rok – year ($r = 1, \dots, 9$), w – województwo – voivodeship ($w = 1, \dots, 16$).

TABELA 2. Wartości parametrów przyjętych do obliczeń
TABLE 2. Values of selected parameters

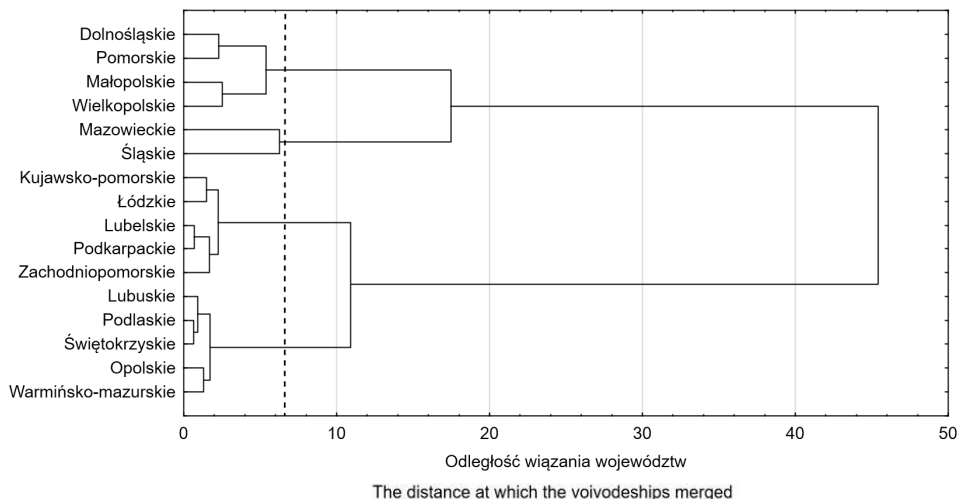
v	Wartość produkcji budowlano-montażowej Value of construction and assembly production $P_{1,w,r}$			Liczba osób zatrudnionych w budownictwie Number of people employed in the construction industry $P_{2,w,r}$			Liczba ludności danego regionu Population of a given region $P_{3,w,r}$			Liczba osób poszkodowanych w wypadkach przy pracy Number of people injured in occupational accidents $P_{4,w,r}$		
	2008 $r = 1$...	2016 $r = 9$	2008 $r = 1$...	2016 $r = 9$	2008 $r = 1$...	2016 $r = 9$	2008 $r = 1$...	2016 $r = 9$
1	0,734	...	0,394	0,481	...	0,291	0,397	...	0,395	0,534	...	0,288
2	-0,537	...	-0,504	-0,264	...	-0,323	-0,254	...	-0,251	-0,110	...	-0,294
...
16	-0,258	...	-0,158	-0,403	...	-0,530	-0,555	...	-0,546	-0,664	...	-0,798

r – rok – year ($r = 1, \dots, 9$), w – województwo – voivodeship ($w = 1, \dots, 16$).

Po przeprowadzeniu obliczeń, z wykorzystaniem oprogramowania Statistica, uzyskano dendrogram przedstawiony na rysunku.

Analizując dendrogram, można zauważyć, że niektóre województwa bar-

dzo szybko tworzą wyraźne skupienia. Zaproponowano odcięcie gałęzi dendrogramu w miejscu zaznaczonym na rysunku linią przerywaną. Przy takim odcięciu otrzymano cztery grupy województw, w których każde województwo charakte-



RYSUNEK. Dendrogram – wykres obrazujący łączenie się poszczególnych województw w kolejnych krokach obliczeniowych (opis w tekście)

FIGURE. Dendrogram – a graph showing the connection of individual voivodeships in the subsequent calculation steps (description in the text)

ryzuje się podobnym poziomem rozwoju budownictwa i bezpieczeństwa pracy. Zidentyfikowane grupy województw zamieszczono w tabeli 3.

nia, w skład których wchodzi od dwóch do pięciu województw. Skupienie I tworzą województwa: dolnośląskie, pomorskie, małopolskie i wielkopolskie, skupienie II

TABELA 3. Otrzymane grupy województw charakteryzujących się podobieństwem tempa rozwoju w budownictwie i poziomem bezpieczeństwa pracy

TABLE 3. The obtained groups of voivodships, which are characterized by a similar speed of development in the construction industry and a similar level of occupational safety

Skupienie Cluster	Województwa Voivodeships	Odległość, przy której nastąpiło połączenie województw The distance at which the voivodeships merged
I	dolnośląskie, pomorskie, małopolskie, wielkopolskie	5,38
II	mazowieckie, śląskie	6,25
III	kujawsko-pomorskie, łódzkie, lubelskie, podkarpackie, zachodniopomorskie	2,26
IV	lubuskie, podlaskie, świętokrzyskie, opolskie, warmińsko-mazurskie	1,73

Analizując otrzymane wyniki, stwierdza się, że duża część województw bardzo szybko łączy się w pary, do których dołączane są kolejne województwa lub skupienia kilku województw. Świadczy to o występowaniu podobieństw między łącznie województwami, które wraz ze wzrostem odległości wiązania słabną. Taka sytuacja wynika z przyjętych do obliczeń założeń, według których w każdym kroku obliczeniowym każdy obiekt jest porównywany ze wszystkimi pozostałymi i łączony z obiektem najbardziej do niego podobnym. Dodatkowo należy stwierdzić, że wraz ze wzrostem odległości wiązania liczba połączeń przy kolejnych krokach obliczeniowych maleje. Świadczy to o istnieniu różnic między poszczególnymi grupami województw, które rosną wraz ze wzrostem odległości wiązania.

Na podstawie analizy otrzymanych wyników wyodrębniono cztery skupie-

tworzą województwa: mazowieckie i śląskie, skupienie III tworzą województwa: kujawsko-pomorskie, łódzkie, lubelskie, podkarpackie, zachodniopomorskie, a skupienie IV tworzą województwa: lubuskie, podlaskie, świętokrzyskie, opolskie, warmińsko-mazurskie.

Zaobserwowano bardzo szybkie tworzenie się wiązań między województwami w skupieniach III i IV. Skupienia te zawierają województwa charakteryzujące się podobnym poziomem wartości produkcji budowlano-montażowej, bezpieczeństwa pracy, liczby osób zatrudnionych w budownictwie oraz liczby osób zamieszkujących województwo. Odległość wiązania w skupieniu III wynosi 2,26, a w skupieniu IV wynosi 1,73, co świadczy o silnym powiązaniu województw wchodzących w skład danego skupienia. Taki stan rzeczy wynika z tego, że wartości wskaźników opisujących

każde województwo w skupieniach III i IV są do siebie zbliżone.

Najbardziej odstającymi od pozostałych są województwa mazowieckie i śląskie. Tworzą one jedno skupienie, odległość, przy której nastąpiło wiązanie między nimi, wynosi 6,25 i jest ponad dwa razy większa niż w skupieniach III i IV. Oznacza to, że powyższa para województw wykazuje najniższy stopień wzajemnego podobieństwa.

Z kolei w skupieniu I zauważa się podobny poziom podobieństwa między parami województw: dolnośląskim i pomorskim oraz małopolskim i wielkopolskim. Odległość wiązania w obu tych parach wynosi odpowiednio 2,31 oraz 2,53, a odległość, przy której nastąpiło wiązanie tych dwóch par, wynosi 5,38. Otrzymane w ten sposób wyniki wykazują na mniejszą siłę powiązania niż występująca w skupieniach III i IV.

Podsumowując, należy stwierdzić, że województwa wchodzące w skład skupień III i IV charakteryzują się podobnym poziomem wartości produkcji budowlano-montażowej, bezpieczeństwa pracy, liczby osób zatrudnionych w budownictwie oraz liczby osób zamieszkujących województwo, między województwami wchodzącymi w skład skupień I i IV różnice w wymienionych wskaźnikach są zaś większe.

Podsumowanie

W artykule zaproponowano własną, uniwersalną metodykę klasyfikacji województw opisanych wybranymi cechami. Uniwersalność metodyki polega na tym, że można ją zastosować także do grupowania innych obiektów gospodar-

czych, np. przedsiębiorstw, a także można uwzględniać w obliczeniach inne wskaźniki charakteryzujące obiekty. Podstawą zaproponowanej metodyki była jedna z metod wielowymiarowej analizy danych statystycznych – analiza skupień. Do obliczeń przyjęto dane statystyczne publikowane przez Główny Urząd Statystyczny, dotyczące wielkości produkcji budowlanej, liczby mieszkańców województwa, liczby osób zatrudnionych w budownictwie oraz liczby osób poszkodowanych w wypadkach przy pracy w budownictwie w latach 2008–2016. Poziom bezpieczeństwa pracy można oceniać przez liczbę osób poszkodowanych w wypadkach przy pracy przypadającą na wartość produkcji budowlano-montażowej lub w odniesieniu do liczby osób zatrudnionych w budownictwie, a także w odniesieniu do liczby mieszkańców danego województwa. W przeprowadzonej klasyfikacji uwzględniono wszystkie wymienione wcześniej wskaźniki, a analiza otrzymanych wyników pozwoliła na zaproponowanie podziału województw na cztery grupy, w których każde województwo charakteryzuje się podobnym poziomem wartości produkcji budowlano-montażowej, bezpieczeństwa pracy, liczby osób zatrudnionych w budownictwie oraz liczby osób zamieszkujących województwo.

Prawidłowa klasyfikacja obiektów opisanych wieloma wskaźnikami może mieć istotne znaczenie dla określenia charakterystycznych cech danej zbiorowości, dokonywania oceny bądź poszukiwania zależności w niej obowiązujących. W prowadzonych przez autorów badaniach naukowych informacje o województwach należących do tego samego skupienia będą podstawą do budowy wieloczynnikowych modeli re-

gresji liniowej służących do predykcji wskaźników opisujących poziom bezpieczeństwa pracy w budownictwie.

Literatura

- Drozd, W., Kowalik, M. (2017). Drzewa decyzyjne w analizie wypadkowości w budownictwie. *Przegląd Naukowy – Inżynieria i Kształtowanie Środowiska*, 26(2), 226-233. DOI: 10.22630/PNIKS.2017.26.2.21
- Główny Urząd Statystyczny – GUS (2009–2017). *Rocznik Statystyczny 2008-2016*. Warszawa: Dział Wydawnictw Statystycznych GUS.
- Hoła, B. (2015). Identification and evaluation of processes in a construction enterprise. *Archives of Civil and Mechanical Engineering*, 15(2), 419-426. DOI: 10.1016/j.acme.2014.11.001
- Hoła, A., Hoła, B., Szóstak, M. (2017). Analysis of the causes and consequences of falls from scaffolding using the Polish construction industry as an example. W *IOP Conference Series: Materials Science and Engineering*, 251, 012050. DOI: 10.1088/1757-899X/251/1/012050
- Hoła, A., Sawicki, M., Szóstak M. (2018). Methodology of Classifying the Causes of Occupational Accidents Involving Construction Scaffolding Using Pareto-Lorenz Analysis. *Applied Sciences*, 8(1), 1-11. DOI: 10.3390/app8010048
- Hoła, B., Nowobilski, T., Szer, I., Szer, J. (2017). Identification of factors affecting the accident rate in the construction industry. *Procedia Engineering*, 11, 35-42. DOI: 10.1016/j.proeng.2017.11.018
- Hoła, B., Szóstak, M. (2015). Analysis of the state of the accident rate in the construction industry in European Union countries. *Archives of Civil Engineering*, 61(4), 19-34. DOI: 10.1515/ace-2015-0033
- International Labour Organization – ILO (2014). *Safety and Health at Work. A Vision for Sustainable Prevention*. Turin: International Training Centre of the ILO. Pobrano z lokalizacji: http://www.ilo.org/wcmsp5/groups/public/@ed_protect/@protrav/@safework/documents/publication/wcms_301214.pdf.

- Jarocka, M. (2015). Wybór formuły normalizacyjnej w analizie porównawczej obiektów wielocechowych. *Economics and Management*, 1, 113-126. DOI: 10.12846/j.em.2015.01.08
- Stanisz, A. (2007). *Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny*. Tom 3: *Modele liniowe i nieliniowe*. Kraków: StatSoft Polska.
- Ward, J.H. Jr. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 58(301), 236–244.
- Wierchoń, S., Kłopotek, M. (2015). *Algorytmy analizy skupień*. Warszawa: WNT.

Streszczenie

Klasyfikacja polskich województw pod kątem wybranych wskaźników charakteryzujących budownictwo. W artykule przedstawiono metodykę klasyfikacji województw Polski pod kątem wybranych wskaźników, takich jak: wartość produkcji budowlano-montażowej, liczba osób zatrudnionych w budownictwie, liczba ludności danego województwa oraz liczba osób poszkodowanych w wypadkach przy pracy. Do rozwiązania postawionego zadania zastosowano analizę skupień, dostępną w oprogramowaniu Statistica.

Summary

Classification of Polish voivodeships with regards to selected indicators that characterize the construction industry. The article presents the methodology of classifying Polish voivodeships with regards to selected indicators, such as the value of construction and assembly production, the number of people employed in the construction industry, the population of a given voivodeship and the number of people injured in occupational accidents. Cluster analysis available in Statistica software was used to solve the problem.

Authors' address:

Tomasz Nowobilski
Politechnika Wroclawska
Wydział Budownictwa Lądowego i Wodnego
Zakład Technologii i Zarządzania
w Budownictwie
ul. Wybrzeże Wyspiańskiego 27
50-370 Wrocław
Poland
e-mail: tomasz.nowobilski@pwr.edu.pl