

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS

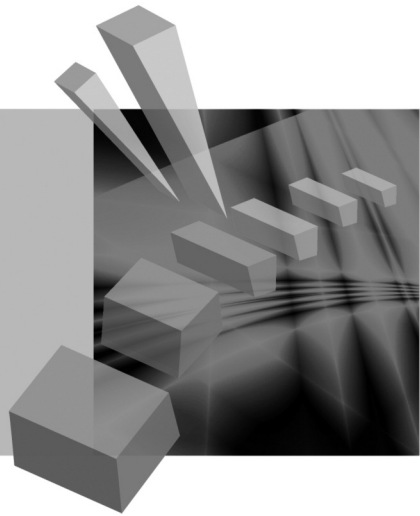
of Wrocław University of Economics

242

Taksonomia 19.

Klasyfikacja i analiza danych

– teoria i zastosowania



Redaktorzy naukowi
Krzysztof Jajuga
Marek Walesiak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2012

Recenzenci: Eugeniusz Gatnar, Elżbieta Gołata, Tadeusz Kufel, Józef Pocięcha,
Miroslaw Szreder, Feliks Wysocki

Redaktor Wydawnictwa: Aleksandra Śliwka

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Małgorzata Czupryńska

Projekt okładki: Beata Dębska

Tytuł sfinansowano ze środków Sekcji Klasyfikacji i Analizy Danych PTS
i Uniwersytetu Ekonomicznego we Wrocławiu

Publikacja jest dostępna na stronie www.ibuk.pl

Streszczenia opublikowanych artykułów są dostępne w międzynarodowej bazie danych
The Central European Journal of Social Sciences and Humanities <http://cejsh.icm.edu.pl>
oraz w The Central and Eastern European Online Library www.ceeol.com,
a także w adnotowanej bibliografii zagadnień ekonomicznych BazEkon [http://kangur.uek.krakow.pl/
bazy_ae/bazekon/nowy/index.php](http://kangur.uek.krakow.pl/bazy_ae/bazekon/nowy/index.php)

Informacje o naborze artykułów i zasadach recenzowania znajdują się
na stronie internetowej Wydawnictwa
www.wydawnictwo.ue.wroc.pl

Kopowanie i powielanie w jakiegokolwiek formie
wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu
Wrocław 2012

ISSN 1899-3192 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM
Nakład: 320 egz.

Spis treści

Wstęp	13
Stanisława Bartosiewicz , Jeszcze raz o skutkach subiektywizmu w analizie wielowymiarowej	17
Andrzej Sokolowski , Q uniwersalna miara odległości	22
Eugeniusz Gatnar , Jakość danych w systemach statystycznych banków centralnych (na przykładzie NBP)	31
Marek Walesiak , Pomiar odległości obiektów opisanych zmiennymi mierzonymi na skali porządkowej – strategię postępowania.....	39
Krzysztof Jajuga, Marek Walesiak , XXV lat konferencji taksonomicznych – fakty i refleksje	47
Józef Pocięcha, Barbara Pawelek , Model SEM w analizie zagrożenia bankructwem przedsiębiorstw w świetle koniunktury gospodarczej – problemy teoretyczne i praktyczne	50
Paweł Lula , Uczące się systemy pozyskiwania informacji z dokumentów tekstowych	58
Ewa Roszkowska , Zastosowanie metody TOPSIS do wspomaganie procesu negocjacji.....	68
Andrzej Młodak , Sąsiedztwo obszarów przestrzennych w ujęciu fizycznym oraz społeczno-ekonomicznym – podejście taksonomiczne	76
Andrzej Bąk , Modele kategorii nieuporządkowanych w badaniach preferencji	86
Jacek Kowalewski , Zintegrowany model optymalizacji badań statystycznych.....	96
Jan Paradysz, Karolina Paradysz , Obszary bezrobocia w Polsce – problem benchmarkowy.....	106
Tomasz Szubert , W co grać, aby jak najmniej przegrać? Próba klasyfikacji systemów gry w zakładach bukmacherskich.....	116
Izabela Szamrej-Baran , Klasyfikacja krajów UE ze względu na ubóstwo energetyczne	126
Sylvia Filas-Przybył, Tomasz Klimanek, Jacek Kowalewski , Analiza dojazdów do pracy za pomocą modelu grawitacji.....	135
Marta Dziechciarz-Duda, Anna Król, Klaudia Przybysz , Minimum egzystencji a czynniki warunkujące skłonność do korzystania z pomocy społecznej. Klasyfikacja gospodarstw domowych	144
Hanna Dudek , Subiektywne skale ekwiwalentności – analiza na podstawie danych o satysfakcji z osiągniętych dochodów	153

Joanicjusz Nazarko, Ewa Chodakowska, Marta Jaročka , Segmentacja szkół wyższych metodą analizy skupień <i>versus</i> konkurencja technologiczna ustalona metodą DEA – studium komparatywne.....	163
Ewa Chodakowska , Wybrane metody klasyfikacji w konstrukcji ratingu szkół.....	173
Bartosz Soliński , Sektor energetyki odnawialnej w krajach Unii Europejskiej – klasyfikacja w świetle strategii zarządzania zmianą.....	182
Krzysztof Szwarz , Klasyfikacja powiatów województwa wielkopolskiego ze względu na sytuację demograficzną.....	192
Elżbieta Gołata, Grażyna Dehnel , Rejestry administracyjne w analizie przedsiębiorczości.....	202
Katarzyna Chudy, Marek Sobolewski, Kinga Stępień , Wykorzystanie metod taksonomicznych w prognozowaniu wskaźników rentowności banków giełdowych w Polsce.....	212
Katarzyna Dębowska , Modelowanie upadłości przedsiębiorstw przy wykorzystaniu metod dyskryminacji i regresji.....	222
Alina Bojan , Wykorzystanie metod wielowymiarowej analizy danych do identyfikacji zmiennych wpływających na atrakcyjność wybranych inwestycji.....	231
Justyna Brzezińska , Analiza logarytmiczno-liniowa w badaniu przyczyn umieralności w krajach UE.....	240
Aneta Rybicka, Bartłomiej Jefmański, Marcin Pelka , Analiza klas ukrytych w badaniach satysfakcji studentów.....	247
Bartłomiej Jefmański , Pomiar opinii respondentów z wykorzystaniem elementów teorii zbiorów rozmytych i środowiska R.....	256
Julita Stańczuk , Porównanie rezultatów wielostanowej klasyfikacji obiektów ekonomicznych z wykorzystaniem analizy dyskryminacyjnej oraz sieci neuronowych.....	265
Jerzy Krawczuk , Skuteczność metod klasyfikacji w prognozowaniu kierunku zmian indeksu giełdowego S&P500.....	275
Anna Czapkiewicz, Beata Basiura , Symulacyjne badanie wpływu zaburzeń na grupowanie szeregów czasowych na podstawie modelu Copula-GARCH.....	283
Radosław Pietrzyk , Ocena efektywności inwestycji funduszy inwestycyjnych z tytułu doboru papierów wartościowych i umiejętności wykorzystania trendów rynkowych.....	291
Aleksandra Witkowska, Marek Witkowski , Zastosowanie metody Panzara-Rosse’a do pomiaru poziomu konkurencji w sektorze banków spółdzielczych.....	306
Marcin Pelka , Podejście wielomodelowe z wykorzystaniem metody <i>boosting</i> w analizie danych symbolicznych.....	315
Justyna Wilk , Analiza porównawcza oprogramowania komputerowego w klasyfikacji danych symbolicznych.....	323

Tomasz Bartłomowicz, Justyna Wilk , Zastosowanie metod analizy danych symbolicznych w przeszukiwaniu dziedzinowych baz danych.....	333
Kamila Migdał-Najman , Propozycja hybrydowej metody grupowania opartej na sieciach samouczących	342
Dorota Rozmus , Porównanie dokładności taksonomii spektralnej oraz zagregowanych algorytmów taksonomicznych opartych na idei metody <i>bagging</i>	352
Krzysztof Najman , Grupowanie dynamiczne z wykorzystaniem samouczących się sieci GNG	361
Małgorzata Misztal , Wpływ wybranych metod uzupełniania brakujących danych na wyniki klasyfikacji obiektów z wykorzystaniem drzew klasyfikacyjnych w przypadku zbiorów danych o niewielkiej liczebności – ocena symulacyjna	370
Mariusz Kubus , Zastosowanie wstępnego uwarunkowania zmiennej objaśnianej do selekcji zmiennych.....	380
Barbara Batóg, Jacek Batóg , Wykorzystanie analizy dyskryminacyjnej do identyfikacji czynników determinujących stopę zwrotu z inwestycji na rynku kapitałowym	387
Katarzyna Wójcik, Janusz Tuchowski , Analiza porównawcza miar podobieństwa tekstów opartych na macierzy częstości i tekstów opartych na wiedzy dziedzinowej	396
Iwona Staniec , Analiza czynnikowa w identyfikacji obszarów determinujących doskonalenie systemów zarządzania w polskich organizacjach	406
Marek Lubicz, Maciej Zięba, Adam Rzechonek, Konrad Pawelczyk, Jerzy Kołodziej, Jerzy Błaszczyk , Analiza porównawcza wybranych technik eksploracji danych do klasyfikacji danych medycznych z brakującymi obserwacjami	416
Iwona Foryś , Wykorzystanie analizy log-liniowej do wyboru czynników determinujących atrakcyjność cenową mieszkań w obrocie wtórnym na przykładzie lokalnego rynku mieszkaniowego.....	426
Ewa Genge , Analiza skupień oparta na mieszankach uciętych rozkładów normalnych.....	436
Jerzy Korzeniewski , Ocena efektywności metody uśredniania zmiennych i metody Ichino selekcji zmiennych w analizie skupień	444
Andrzej Dudek , SMS – propozycja nowego algorytmu analizy skupień	451
Artur Mikulec , Metody oceny wyniku grupowania w analizie skupień.....	460
Małgorzata Machowska-Szewczyk , Algorytm klasyfikacji rozmytej dla obiektów opisanych za pomocą zmiennych symbolicznych oraz rozmytych	469
Artur Zaborski , Analiza PROFIT i jej wykorzystanie w badaniu preferencji	479
Karolina Bartos , Analiza skupień wybranych państw ze względu na strukturę wydatków konsumpcyjnych obywateli – zastosowanie sieci Kohonena	488

Barbara Batóg, Magdalena Mojsiewicz, Katarzyna Wawrzyniak , Klasyfikacja gospodarstw domowych ze względu na bodźce do zawierania umowy o ubezpieczenie z wykorzystaniem modeli zmiennych jakościowych .	496
Izabela Kurzawa , Zastosowanie modelu LA/AIDS do badania elastyczności cenowych popytu konsumpcyjnego w gospodarstwach domowych w relacji miasto–wieś	505
Aleksandra Łuczak, Feliks Wysocki , Metody porządkowania liniowego obiektów opisanych za pomocą cech metrycznych i porządkowych	513
Agnieszka Sompolska-Rzechuła , Porównanie klasycznej i pozycyjnej taksonomicznej analizy zróżnicowania jakości życia w województwie zachodniopomorskim	523
Joanna Banaś, Małgorzata Machowska-Szewczyk , Ocena intensywności wykorzystania skrzynek poczty elektronicznej za pomocą uporządkowanego modelu probitowego	532
Iwona Bąk , Segmentacja gospodarstw domowych emerytów i rencistów pod względem wydatków na rekreację i kulturę	541
Aneta Becker , Zastosowanie metody ANP do porządkowania województw Polski pod względem dynamiki wykorzystania ICT w latach 2008-2010	552
Katarzyna Dębowska , Klasyfikacja sektorów ze względu na ich kondycję finansową przy użyciu metod wielowymiarowej analizy statystycznej	562
Anna Domagała , Propozycja metody doboru zmiennych do modeli DEA (procedura kombinowanego doboru w przód).....	571
Henryk Gierszal, Karina Pawlina, Maria Urbańska , Analiza statystyczna w badaniach zapotrzebowania na usługi teleinformatyczne sieci łączności ruchomej	580
Hanna Gruchociak , Konstrukcja estymatora regresyjnego dla danych o strukturze dwupoziomowej.....	590
Tomasz Klimanek, Marcin Szymkowiak , Zastosowanie estymacji pośredniej uwzględniającej korelację przestrzenną w opisie niektórych charakterystyk rynku pracy	601
Jarosław Lira , Prognozowanie opłacalności produkcji żywca wieprzowego w Polsce	610
Christian Lis , Wykorzystanie metody klasyfikacji w ocenie konkurencyjności portów południowego Bałtyku	619
Beata Bieszk-Stolorz, Iwona Markowicz , Wykorzystanie wielomianowego modelu logitowego do oceny szansy podjęcia pracy przez bezrobotnych .	628
Lucyna Przezbórska-Skobiej, Jarosław Lira , Przestrzeń agroturystyczna Polski i ocena jej atrakcyjności.....	637
Paweł Ulman , Model rozkładu wydatków a funkcje popytu.....	646
Maria Urbańska, Tadeusz Mizera, Henryk Gierszal , Zastosowanie metod analizy statystycznej w badaniach mięczaków	655

Summaries

Stanisława Bartosiewicz , The effects of subjectivism in multivariate analysis revisited.....	21
Andrzej Sokółowski , Q universal distance measure	30
Eugeniusz Gatnar , Data quality in central banks' statistical systems (NBP example)	38
Marek Walesiak , Distance measures for ordinal data – strategies of proceedings.....	46
Krzysztof Jajuga, Marek Walesiak , XXV years of taxonomic conferences – some facts and remarks.....	49
Józef Pocięcha, Barbara Pawelek , General SEM model in researching corporate bankruptcy and business cycles – theoretical and practical problems.....	57
Paweł Lula , Learning-based systems of information extraction from textual resources	67
Ewa Roszkowska , The application of the TOPSIS method to support the negotiation process	75
Andrzej Młodak , Neighborhood of spatial areas in the physical and socio-economic context – a taxonomic approach.....	85
Andrzej Bąk , Models for unordered categories in preference analysis.....	95
Kowalewski Jacek , An integrated model of optimizing statistical surveys	105
Jan Paradysz, Karolina Paradysz , Areas of unemployment in Poland – benchmark problem	115
Tomasz Szubert , How to play to lose the least? Classification of systems in sports bets	125
Izabela Szamrej-Baran , Classification of EU member states in view of fuel poverty	134
Sylvia Filas-Przybył, Tomasz Klimanek, Jacek Kowalewski , An attempt to use the gravity model in the analysis of commuters.....	143
Marta Dziechciarz-Duda, Anna Król, Klaudia Przybysz , Subsistence minimum versus factors influencing tendency to benefit from social care. Classification of households	152
Hanna Dudek , Subjective equivalence scales – analysis based on data about satisfaction with incomes.....	162
Joanicjusz Nazarko, Ewa Chodakowska, Marta Jarocka , Segmentation of universities using cluster analysis versus technological competitors determined by the DEA method – a comparative study	172
Ewa Chodakowska , Selected methods of classification in schools' rating.....	181
Bartosz Soliński , Renewable energy sector in the European Union – classification in the light of change management strategy	191
Krzysztof Szwarc , Classification of Wielkopolska voivodeship due to the demographic situation	201

Elżbieta Gołata, Grażyna Dehnel , Administrative registers in business analysis.....	211
Katarzyna Chudy, Marek Sobolewski, Kinga Stępień , Application of taxonomic methods in forecasting the profitability ratios of listed banks in Poland.....	221
Katarzyna Dębowska , Modeling bankruptcy of firms by using discrimination and regression methods.....	230
Alina Bojan , Identification of variables which influence attractiveness of given investments with the usage of multivariate analysis.....	239
Justyna Brzezińska , Log-linear analysis in the study of mortality in EU.....	246
Aneta Rybicka, Bartłomiej Jefmański, Marcin Pelka , Latent class analysis in student satisfaction surveys.....	254
Bartłomiej Jefmański , The respondent's opinions measurement in the R program with an application of fuzzy sets theory.....	264
Julita Stańczuk , A comparison of the results of multistate classification of economic objects using discriminant analysis and artificial neural networks.....	274
Jerzy Krawczuk , Effectiveness of classification methods in S&P500 stock index direction changes forecasting.....	282
Anna Czapkiewicz, Beata Basiura , The simulation study of the utility of the Copula-GARCH models for clustering financial time series.....	290
Radosław Pietrzyk , Timing and selectivity in mutual funds performance measurement.....	305
Aleksandra Witkowska, Marek Witkowski , Use of the Panzar-Rosse method to assess of the competition level in the cooperative banks sector.....	314
Marcin Pelka , Ensemble learning with the application of <i>boosting</i> in symbolic data analysis.....	322
Justyna Wilk , Comparative study of symbolic data classification software.....	332
Tomasz Bartłomowicz, Justyna Wilk , Application of symbolic data analysis methods for domain database searching.....	341
Kamila Migdał-Najman , A proposal of hybrid clustering method based on self-learning networks.....	351
Dorota Rozmus , Comparison of accuracy of spectral clustering and cluster ensembles stability based on bagging idea.....	360
Krzysztof Najman , A dynamic grouping based on self-learning GNG networks.....	369
Małgorzata Misztal , Influence of data imputation methods on the results of object classification using classification trees in the case of small data sets – simulation assessment.....	379
Mariusz Kubus , The application of pre-conditioning of explanatory variable for feature selection.....	386
Barbara Batóg, Jacek Batóg , Application of discriminant analysis to the identification of factors determining the rate of return on the capital market.....	395

Katarzyna Wójcik, Janusz Tuchowski , Comparative analysis of text documents similarity measures based on frequency matrix and based on domain knowledge.....	405
Iwona Staniec , Factor analysis in the identification of areas that determine the improvement of management systems in Polish organizations.....	415
Marek Lubicz, Maciej Zięba, Adam Rzechonek, Konrad Pawełczyk, Jerzy Kołodziej, Jerzy Błaszczyk , Comparative analysis of selected data mining approaches to the classification of medical data with missing values (covariates).....	425
Iwona Foryś , The log-linear analysis using to select the factors determining the attractiveness of the price of flats on the secondary market on the example of local housing market.....	435
Ewa Genge , Trimming approach to the mixtures of normal distributions.....	443
Jerzy Korzeniewski , Efficiency assessment of Ichino method and mean value method of selecting variables in cluster analysis.....	450
Andrzej Dudek , SMS – proposal of new clustering algorithm.....	459
Artur Mikulec , Evaluation methods for the grouping result in cluster analysis.....	468
Małgorzata Machowska-Szewczyk , Fuzzy clustering algorithm for objects described by symbolic or fuzzy variables.....	478
Artur Zaborski , PROFIT analysis and its using in the research of preferences.....	487
Karolina Bartos , Cluster analysis of selected countries due to the structure of their citizens' consumer expenditures – the use of Kohonen networks.....	495
Barbara Batóg, Magdalena Mojsiewicz, Katarzyna Wawrzyniak , Classification of households according to the impulses of concluding the insurance contract by means of qualitative variable models.....	504
Izabela Kurzawa , The application of LA/AIDS model to examine price elasticities of demand of households in the urban-rural relationship.....	512
Aleksandra Luczak, Feliks Wysocki , Linear ordering methods of objects described by a set of metric and ordinal characteristics.....	522
Agnieszka Sompolska-Rzechuła , The comparison of the classical and positional taxonomic analysis of the quality of life differentiation in Zachodniopomorskie voivodeship.....	531
Joanna Banaś, Małgorzata Machowska-Szewczyk , Evaluation of intensity of mailboxes using with the ordered probit model.....	540
Iwona Bąk , Segmentation of pensioners and annuitants households in terms of expenditures on recreation and culture.....	551
Aneta Becker , Application of ANP method to organize Polish voivodships in terms of dynamics of the use of ICT in 2008-2010.....	561
Katarzyna Dębowska , The classification of sectors' financial situation using the methods of multivariate statistical analysis.....	570

Anna Domagała , Proposal of a new method for variable selection in DEA models (combined forward stepwise selection method).....	579
Henryk Gierszal, Karina Pawlina, Maria Urbańska , Statistical analysis in demand research of ICT services in mobile networks.....	589
Hanna Gruchociak , Construction of regression estimator for two-level data	600
Tomasz Klimanek, Marcin Szymkowiak , Application of spatial models in indirect estimation of some labor market characteristics	609
Jarosław Lira , Forecasting of hog livestock production profitability in Poland	618
Christian Lis , The utilization of taxonomic methods in the appraisal of competitiveness of south Baltic ports	627
Beata Bieszk-Stolorz, Iwona Markowicz , The application of the multinomial logit model in evaluating employment odds for the unemployed job seekers	636
Lucyna Przezbórska-Skobiej, Jarosław Lira , Agritourism space of Poland and its valuation.....	645
Paweł Ulman , Model of expenses distribution and demand functions.....	654
Maria Urbańska, Tadeusz Mizera, Henryk Gierszal , Methods of statistical analysis in research of molluscs	663

Maria Urbańska, Tadeusz Mizera

Uniwersytet Przyrodniczy w Poznaniu

Henryk Gierszal

Uniwersytet im. Adama Mickiewicza w Poznaniu

ZASTOSOWANIE METOD ANALIZY STATYSTYCZNEJ W BADANIACH MIĘCZAKÓW

Streszczenie: W artykule przedstawiono możliwości wykorzystania różnych metod analizy statystycznej do wnioskowania o przyrodzie ożywionej na przykładzie badań mięczaków. Do porównania bogactwa malakofauny na różnych terenach wykorzystano wskaźniki podobieństw, które posłużyły do wykreślenia dendrogramów.

Słowa kluczowe: ślimak, różnorodność, podobieństwo, dyskryminacja.

1. Wstęp

Badania przyrody ożywionej są prowadzone od setek lat. Kiedyś, w dobie odkryć, satysfakcjonującym osiągnięciem było zidentyfikowanie nowego gatunku i możliwość opisanego jego wyglądu. Później przyszedł czas także na obserwacje behawioralne. W dzisiejszych czasach, kiedy rozwój cywilizacji mocno oddziałuje na środowisko naturalne, istnieje potrzeba ilościowego i jakościowego opisanego przyrody ożywionej. Dysponując tymi licznymi danymi, naukowcy starają się określić interakcje między gatunkami, kondycję środowiska i wpływ ludzi. Pomocna w tym staje się analiza statystyczna, która w dobie maszyn liczących staje się łatwo dostępną i szeroko stosowaną.

Przedstawione w tym artykule analizy oparto na obserwacjach empirycznych. Bazując na otrzymanych wynikach, wykorzystano następnie aparat statystyczny w celu wychwycenia pewnych zależności i na tej podstawie wnioskowania o zgrupowaniu ślimaków (*Gastropoda*), tzn. malakofaunie żyjącej na obszarze Biedruska koło Poznania. Istniejący tam poligon wojskowy jest jednym z niewielu miejsc, o którym nie dysponowano informacjami o malakofaunie. Wynika to z faktu, iż od 1904 r. teren ten został przekazany do użytkowania wojsku i automatycznie stał się on niedostępny zarówno dla ludności, jak i naukowców. Dzięki zmianom politycznym, które zaszły w naszym kraju w latach 90., otworzyła się możliwość przebadania obszaru poligonu Biedrusko.

W badaniach do bardziej wnikliwej oceny przyczyn i zachodzących relacji zastosowano narzędzia statystyczne wykorzystujące analizę skupień oraz dyskryminacji. Dla celów badawczych postawiono tezę, że bogactwo ślimaków na obszarze poligonu Biedrusko jest większe niż na otaczających go terenach.

2. Metoda badawcza

Badania malakofauny oparto na klasycznym sposobie absolutnych badań ilościowych fauny lądowej, jakim jest „metoda kwadratów”, która pozwala na przeprowadzenie badań zarówno ilościowych, jak i jakościowych. Ścisłe w odniesieniu do ślimaków metodę tę opracował F. Øekland, który w latach 1927-28 wykonał w lasach Norwegii pierwszą tego typu pracę [Øekland 1929, 1930]. Autor ten zalecał pobranie 50 prób za pomocą ramki (biocenometru) o wymiarach 25×25 cm, co w sumie odpowiada przebadaniu $3\frac{1}{8}$ m². Metoda ta do dzisiaj nie straciła na aktualności. Poddana została jedynie pewnym modyfikacjom dotyczącym przede wszystkim liczby prób w serii oraz wielkości ramki.

Do badań wybrano sześć obiektów znajdujących się na terenie poligonu:

- fitokompleks łągowo-grądowy zabytkowego parku podworskiego w Radojewie (RAD); próbki z 1997 i 1998 r. są oznaczone odpowiednio jako '97 i '98,
- zespół starorzeczy warciańskich położony na północnym-wschodzie od Gołębiowa (CZW),
- kompleks łąkowo-zaroślowo-leśny zwany Sarnią Łąką (SŁ),
- dolina Rowu Północnego (PR),
- rynna Jeziora Glinowieckiego (JG),
- torfowisko przejściowe nad Jeziorem Gogulec koło Złotkowa (GO).

Z dwu pierwszych powierzchni – tj. parku podworskiego w Radojewie oraz zespołu starorzeczy warciańskich – próby pobrano pięciokrotnie w ciągu jednego sezonu badawczego, czyli roku kalendarzowego. Ustalono terminy poszczególnych zbiorów w taki sposób, aby móc przeanalizować wpływ pory roku na liczebność, miejsce występowania czy udział form młodocianych ślimaków. Próby pobierano w następujących terminach:

- | | |
|-------------------------|---------------------|
| • wczesnowiosennym (WW) | 19-20 marca, |
| • późnowiosennym (PW) | 7-8 maja, |
| • letnim (L) | 15-16 lipca, |
| • wczesnojesiennym (WJ) | 5-6 września, |
| • późnojesiennym (PJ) | 24-25 października. |

Do oceny bogactwa malakofauny na obszarach wokół Biedruska zebrane próbki terenowe — zawierające prawie 6000 okazów — zostały poddane obróbce statystycznej wykorzystującej m.in. liczne wskaźniki analityczne i syntetyczne jako wielkości bazowe do wnioskowania o różnorodności i dalszej bardziej zaawansowanej analizie związanej z określeniem np. podobieństwa między terenami pod względem składu gatunkowego i ilościowego ślimaków [Čejka 2011]. Wnioskowanie oparto

także na analizie dyskryminacyjnej pozwalającej dokonać predykcyjnej klasyfikacji przypadków i wyróżnić grupy gatunków ślimaków.

Do porównywania ze sobą badanych terenów w czasie i przestrzeni posłużono się wybranymi wskaźnikami określającymi współwystępowanie gatunków, które określa m.in. podobieństwo między komponentami zoocenozy. W określonych zoocenozach wydzielić można różne grupy gatunków współwystępujących ze sobą. Ponieważ gatunki te są w rozmaity sposób związane z typem środowiska, dlatego istnieje potrzeba uchwycenia takiej charakterystyki cenologicznej, która ułatwiałaby porównanie i klasyfikację zoocenoz. Istnieją dwa rodzaje wskaźników podobieństwa [Morrison 1992]:

- biorące pod uwagę tylko liczebność,
- uwzględniające aktualną i względną liczebność każdego gatunku (odpowiadają one wyznaczaniu korelacji).

Inne opracowanie [Sokal 1963] wyróżnia trzy typy współczynników: asocjacji, korelacji i odległości. Spośród wielu definicji podobieństwa wzięto pod uwagę dwie metryki, które posłużyły ocenie stopnia podobieństwa badanych terenów:

- a) Morisita — metryka należąca do grupy współczynników odległości:

$$C = \frac{2 \sum_i (n_{1i} n_{2i})}{N_1 N_2 \left(\frac{\sum_i n_{1i} (n_{1i} - 1)}{N_1 (N_1 - 1)} + \frac{\sum_i n_{2i} (n_{2i} - 1)}{N_2 (N_2 - 1)} \right)}, \quad (1)$$

gdzie: n_{ji} — liczebność gatunku i w próbce j ,
 N_j — liczebność próbki j , $j \in \{1; 2\}$,

- b) Marczewskiego i Steinhausa [1959] — metryka należąca do grupy współczynników asocjacji:

$$s = \frac{c}{a + b - c}, \quad (2)$$

gdzie: a — liczba gatunków w próbce A,
 b — liczba gatunków w próbce B,
 c — liczba gatunków wspólnie występujących w próbce A i B, czyli $c = \text{card}(A_{\text{gat.}} \cap B_{\text{gat.}})$, a X_{gat} to zbiór gatunków w próbce X.

W pracy [Wolda 1981] podkreślono, że metryka Morisita (1) zasługuje na szczególną uwagę. Ten indeks jest niezależny od wielkości próby [Morisita 1959], a ponadto od jej różnorodności. Jednakże jest on bardzo wrażliwy na zmiany w abudancji bardziej dominujących gatunków. Dlatego też dodatkowo wykonano obliczenia dla metryki Marczewskiego i Steinhausa (2) opisaną m.in. w [Górny, Grüm 1981].

W kolejnym kroku dla obu tych metryk wykreślono dendrogramy reprezentujące graficznie siłę powiązań pomiędzy poszczególnymi terenami. Tereny te różniły się zarówno położeniem na terenie Biedruska, jak i czasem pobierania prób. Do tworzenia dendrogramów posłużono się trzema metodami wiązania rozpatrywanych obiektów (tzn. badanych terenów), takimi jak:

a) łączenie pojedyncze, w którym wykorzystamy „najbliższych sąsiadów” między skupieniami,

b) łączenie pełne (kompletne), które wykorzystuje sąsiadów, którzy są najbardziej od siebie oddaleni,

c) metoda Warda, która do oszacowania odległości między skupieniami wykorzystuje podejście analizy wariancji, biorąc za podstawę podziału bliskie sąsiedztwo obiektów bardzo do siebie podobnych.

Biorąc pod uwagę znane warunki brzegowe, założono, że najbardziej właściwa dla rozpatrywanych badań terenowych technika tworzenia dendrogramu winna maksymalnie zgrupować w:

1) jeden podzbiór, tereny Radojewo'97 i Radojewo'98 (L) spośród obszarów wszystkich badanych latem (Radojewo'97, Radojewo'98 (L), Czaple Warciska (L), Gogulec, Jezioro Glinowieckie, Pełnikowy Rów, Sarnia Łąka) ze względu na przewidywane ich podobieństwo co do położenia geograficznego na terenie Biedruska,

2) dwa niezależne podzbiory, wyniki dwóch całorocznych badań (5 okresów pobrań) zrealizowanych na obszarze Radojewa'98 i Czaplich Warcisk, biorąc pod uwagę wszystkie 15 pobrań zrealizowanych w latach 1997-1998, ponieważ te dwa zestawy pobrań na tych terenach powinny cechować się podobieństwem czasowym (ten sam rok),

3) jedno skupienie, 5 pobrań wykonanych na terenie Radojewa w roku 1998 oraz jedno pobranie na terenie Radojewa w 1997 r.

Dla warunku nr 1 spośród otrzymanych sześciu (3 metody łączenia \times 2 różne metryki) dendrogramów wybrano te, które najwierniej odtworzyły podobieństwo położenia terenów Radojewo'97 i Radojewo'98 (L). Poprawne zgrupowanie tych dwóch obiektów uzyskano dla metryki Marczewskiego i Steinhausa przy łączeniu pełnym oraz przy łączeniu metodą Warda.

Równoległe wykonano obliczenia z uwzględnieniem warunku nr 2, dla którego uzyskano trzy dendrogramy o poprawnym skupieniu wszystkich 5 okresów pobrań dla obszaru Radojewa'98, jak również Czaplich Warcisk. Poprawny wynik uzyskano dla metryki Marczewskiego i Steinhausa i wszystkich trzech rozważanych metod łączenia obiektów.

Dlatego też posłużono się warunkiem nr 3, który spełniły tylko dwa z trzech dendrogramów otrzymanych po uwzględnieniu warunku nr 2. Te dwa dendrogramy uzyskano dla pojedynczego i kompletnego łączenia oraz metryki Marczewskiego i Steinhausa.

Powyższa procedura pozwoliła określić sposób wyznaczenia dendrogramu, który będzie najbardziej właściwy dla całego obszaru Biedruska. Metoda ta wykorzystuje metrykę Marczewskiego i Steinhausa z łączeniem pełnym, gdyż to ona spełnia wszystkie trzy postawione powyżej warunki brzegowe.

Wiele źródeł literaturowych podejmuje szeroką dyskusję nad metodami wyznaczania podobieństwa obiektów w naukach przyrodniczych, wykorzystując w tym celu nawet bardzo zaawansowany aparat matematyczny weryfikowany praktycznie z wykorzystaniem badań terenowych [Horsák, Hájek 2003; Linse i in. 2006]. W naszym przypadku przyjęcie powyższego podejścia do analizy skupień wynika z założonych na wstępie, dość precyzyjnych warunków brzegowych (warunek nr 1 i 2), które pozwalają zawęzić wybór właściwej metryki oraz poprawnej metody grupowania.

Do klasyfikacji terenów wykorzystano także analizę dyskryminacyjną. Dla gatunków występujących w dwóch zespołach A i B wyznacza się wskaźnik syntetyczny Q używany w celu ułatwienia i uproszczenia interpretacji struktury dominacji gatunków. Łączy on dwa wskaźniki C (stałość występowania) i D (dominacja) w jeden wspólny, będący ich średnią geometryczną:

$$Q = \sqrt{C \cdot D}. \quad (3)$$

Stałość występowania C określa obecność danego gatunku w obrębie badanej biocenozy, dominacja D zaś wyraża stosunek liczby wszystkich osobników danego gatunku do liczby wszystkich osobników badanej grupy systematycznej.

Na podstawie wartości Q wykreśla się diagram pokazujący wartość współczynnika dyskryminacji T :

$$T_i = \frac{Q_B - Q_A}{Q_B + Q_A}, \quad (4)$$

gdzie: $1 \leq i \leq$ liczba gatunków,

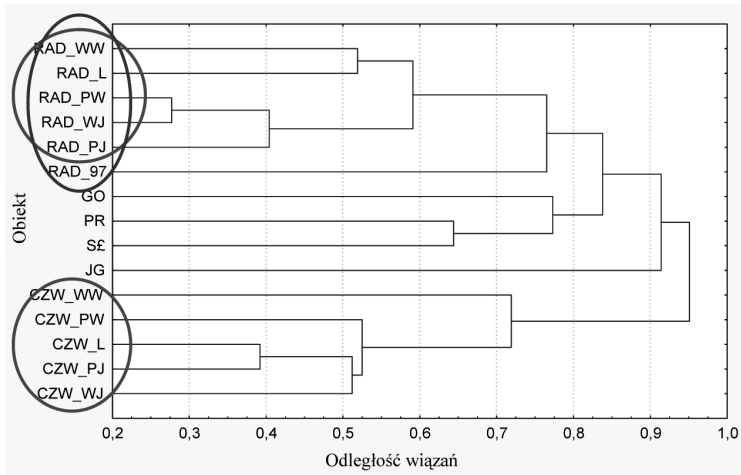
Q_x – wskaźnik Q dla gatunku i w zespole $X \in \{A, B\}$.

W ramach zrealizowanych prac dokonano także badania gleby na analizowanych terenach, na liczebność i bogactwo ślimaków ma bowiem wpływ suma wypadkowa wszystkich czynników środowiskowych (ekologicznych) oddziałujących na dane zgrupowanie. Gleba z jej fizycznymi i chemicznymi właściwościami uważana jest za jeden z ważniejszych czynników wpływających na występowanie i rozmieszczenie ślimaków [Łożek 1962; Dzieczkowski 1969]. Określono wartość pH gleby oraz zawartość pierwiastków (C, N), próchnicy i składników przyswajalnych (K_2O , P_2O_5 , MgO).

3. Wyniki

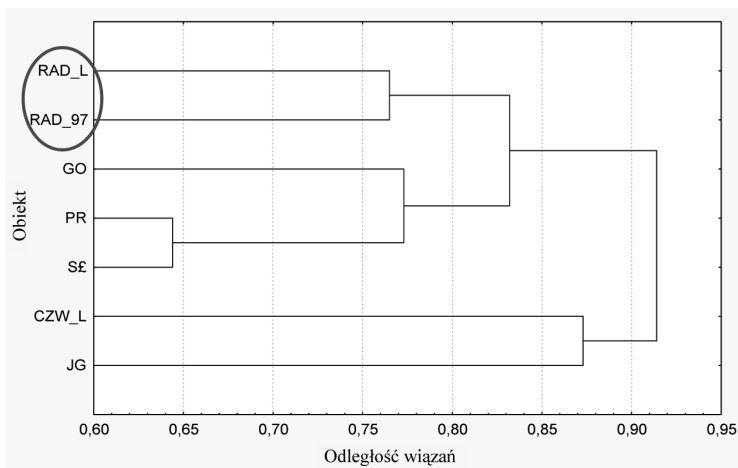
3.1. Analiza podobieństw siedlisk pod względem składu ślimaków

Na rysunku 1 pokazano zgrupowane w zespoły podobne do siebie tereny. Pobrania wykonane w ciągu całego roku (od wiosny do jesieni) na terenie Radojewa '98 i Czaplích Warcisk tworzą dwa podzbiory grupujące próbki z danej lokalizacji przestrzen-



Rys. 1. Dendrogram dla wszystkich 15 analizowanych pobrań

Źródło: opracowanie własne.



Rys. 2. Dendrogram grupujący najbardziej do siebie podobne letnie zgrupowania malakologiczne

Źródło: opracowanie własne.

nej. Dodatkowo w jednym podzbiorze jest obszar parku radojewskiego z dwóch lat: Radojewo'97 i Radojewo'98.

Na rysunku 2 przedstawiono dendrogram dla pobrań letnich. Widać na nim m.in. podobieństwo między tym samym terenem badanym w kolejnych latach: Radojewo'97 i Radojewo'98 (L). Zauważone różnice nie są znaczne i mogą być przejawem naturalnych zmian wewnątrz populacji. W przyrodzie najczęściej spotyka się populacje o cyklicznej dynamice, w których kolejno występują fazy wzrostu i spadku [Krebs 1996]. Obserwowane różnice mogą być właśnie przejawem takiej dynamiki liczebności populacji, jednak aby to stwierdzić, nie można ograniczać się do dwu sezonów, a należałoby przeprowadzić wieloletnie badania ekologiczne.

3.2. Analiza dyskryminacji

Odrębne zestawienie wyników dotyczyło wskaźnika syntetycznego dyskryminacji T w celu scharakteryzowania każdego taksonu w obrębie wyróżnionych dwóch grup pobrań w okresie letnim zidentyfikowanych dzięki dendrogramowi powyżej, tzn.:

- A. Radojewo'97, Radojewo'98 (L), Gogulec, Pełnikowy Rów i Sarnia Łąka,
- B. Jezioro Glinowieckie i Czaple Warciska (L).

Tereny znajdujące się w zespole B charakteryzują się znaczną wilgotnością, której źródłem jest Jezioro Glinowieckie oraz rzeka Warta (dla Czaplich Warcisk). W przypadku zespołu A Gogulec, Pełnikowy Rów oraz Sarnia Łąka są położone w niedalekim sąsiedztwie na obszarze poligonu Biedrusko. Odmienny charakter ma park w Radojewie, lecz jego charakter jest bardziej zbliżony do obszarów, na których nie ma żadnych większych cieków wodnych, dlatego też Radojewo znalazło się w zespole A.

Zespół B charakteryzuje bardzo podobna średnia zawartość azotu w glebie – 0,325% dla Jeziora Glinowieckiego i 0,392% dla Czaplich Warcisk (nawet miara rozrzutu tej wielkości, tzn. odchylenie standardowe, jest bardzo zbliżone). Średnia ilość azotu na pozostałych obszarach zakwalifikowanych do zespołu A jest albo mniejsza, albo większa od wartości uzyskanych dla zespołu B.

Diagram dyskryminacji odwzorowuje przynależność każdego z gatunków do lokalizacji zdefiniowanych przez oba zespoły. Gatunki typowe ($|T| \geq 0,75$) w zespole A to: *Carychium minimum*, *Columella edentula*, *Truncatellina costulata*, *Vallonia pulchella* i *Vitrea crystallina*. Gatunkami charakterystycznymi (tzn. charakteryzującymi się wysoką wartością współczynnika $Q > 10\%$) są dla zespołu B: *Nesovitrea hammonis*, *Nesovitrea petronella* i *Vallonia costata*.

Ponadto zaobserwowano, że średnia wartość stosunku zawartości węgla do azotu w glebie C:N jest najwyższa dla Czaplich Warcisk, Jeziora Glinowieckiego oraz Radojewa. Dwa pierwsze z tych trzech terenów należą do zespołu B, a zatem tę wartość można także wykorzystać, aby móc dyskryminować badane tereny.

4. Podsumowanie

W badaniach ślimaków określono stopień współwystępowania gatunków oraz podobieństw zgrupowań. Dzięki użyciu dendrogramów można było określić różnice i podobieństwa między poszczególnymi zespołami ślimaków dla różnych: (i) sezonów tego samego roku kalendarzowego na danym obszarze badawczym, a także (ii) terenów zbadanych w tym samym okresie. Okazało się, że faktycznie teren poligonu Biedrusko różni się bogactwem malakofauny od sąsiadujących terenów, lecz bogactwo to nie jest aż tak duże jak pobliski park podworski w Radojewie.

Literatura

- Čejka T., *Diversity patterns and freshwater molluscs similarities in small water reservoirs*, „Malacologica Bohemoslovaca” 2011, no 10.
- Dzięczkowski A., *Badania ilościowe ślimaków buczyn południowo-zachodniej Polski*, Praca doktorska wykonana przy Katedrze i Zakładzie Zoologii Ogólnej na UAM w Poznaniu, 1969.
- Górny M., Grüm L. (red.), *Metody stosowane w zoologii gleby*, PWN, Warszawa 1981.
- Horsák M., Hájek M., *Composition and species richness of molluscan communities in relation to vegetation and water chemistry in the western Carpathian spring fens: The poor-rich gradient*, „J. Mollus. Stud.” 2003, no 69 (4).
- Krebs Ch.J., *Ekologia – eksperymentalna analiza rozmieszczenia i liczebności*, Wyd. Nauk. PWN, Warszawa 1996.
- Krebs Ch.J., *Ecology: The Experimental Analysis of Distribution and Abundance*, Benjamin-Cummings Pub. Co, 2008.
- Linse K., Griffiths H.J., Barnes D.K.A., Clarke A., *Biodiversity and biogeography of Antarctic and sub-Antarctic mollusca*, „Deep-Sea Research” 2006, no II 53.
- Ložek V., *Soil Conditions and Their Influence on Terrestrial Gastropoda in Central Europe*, Progress in Soil Zoology, Part 43, Londyn 1962.
- Marczewski E., Steinhaus H., *Odległość systematyczna biotypów*, „Zastosowania Matematyki” 1959, no 4.
- Morisita M., *Measuring of interspecific association and similarity between communities*, „Mem Fac Sci Kyushu Univ, Ser E. Bio.” 1959, no 3.
- Morrison M.L., Marcot B.G., Mannan R.W., *Wildlife-Habitat Relationships: Concepts and Applications*, University of Wisconsin Press, Madison, Wisconsin 1992.
- Øekland F., *Methodik einer quantitativen Untersuchung der Landschneckenfauna*, „Arch. Molluskenk” 1929, no 61.
- Øekland F., *Quantitative Untersuchungen der Landschnecken Fauna Norwegens*, „I. Zeitschr. Morphol. Ökol. Tiere“, Bd. 16(3-4): 748-804, Berlin 1930.
- Sokal R.R., Sneath P.H.A., *Principles of Numerical Taxonomy*, W.H. Freeman and Company, San Francisco 1963.
- Wolda H., *Similarity indices, sample size and diversity*, „Oecologia” 1981, no 50.

METHODS OF STATISTICAL ANALYSIS IN RESEARCH OF MOLLUSCS

Summary: In the paper we present different methods of statistical analysis for deduction about the conditions of the life nature on the example of investigation of invertebrates (Molluscs). To compare the richness of malacofauna among different terrains we applied similarity indices that were used to draw tree plots.

Keywords: snail, diversity, similarity, discrimination.