

SPRAWOZDANIA

KRZYSZTOF JAJUGA, MAREK WALESIAK, FELIKS WYSOCKI

SPRAWOZDANIE Z KONFERENCJI NAUKOWEJ NT. „KLASYFIKACJA I ANALIZA DANYCH – TEORIA I ZASTOSOWANIA”

W dniach 21-23 września 2011 roku w Hotelu Pietrak w Wągrowcu odbyła się XX Jubileuszowa Konferencja Naukowa Sekcji Klasyfikacji i Analizy Danych PTS (XXV Konferencja Taksonomiczna) nt. Klasyfikacja i analiza danych – teoria i zastosowania, organizowana przez Sekcję Klasyfikacji i Analizy Danych Polskiego Towarzystwa Statystycznego, Katedrę Finansów i Rachunkowości Wydziału Ekonomiczno-Społecznego Uniwersytetu Przyrodniczego w Poznaniu oraz Katedrę Statystyki Wydziału Informatyki i Gospodarki Elektronicznej Uniwersytetu Ekonomicznego w Poznaniu. Przewodniczącymi Komitetu Organizacyjnego Konferencji byli prof. dr hab. Feliks Wysocki oraz dr hab. Elżbieta Gołata, prof. nadzw. UEP w Poznaniu, natomiast sekretarzami dr inż. Aleksandra Łuczak i dr Izabela Kurzawa.

Zakres tematyczny konferencji obejmował zagadnienia:

a) teoria (taksonomia, analiza dyskryminacyjna, metody porządkowania liniowego, metody statystycznej analizy wielowymiarowej, metody analizy zmiennych ciągłych, metody analizy zmiennych dyskretnych, metody analizy danych symbolicznych, metody graficzne),

b) zastosowania (analiza danych finansowych, analiza danych marketingowych, analiza danych przestrzennych, inne zastosowania analizy danych – medycyna, psychologia, archeologia, itd., aplikacje komputerowe metod statystycznych).

Zasadniczym celem konferencji SKAD była prezentacja osiągnięć i wymiana doświadczeń z zakresu teoretycznych i aplikacyjnych zagadnień klasyfikacji i analizy danych. Konferencja stanowi coroczne forum służące podsumowaniu obecnego stanu wiedzy, przedstawieniu i promocji dokonań nowatorskich oraz wskazaniu kierunków dalszych prac i badań.

W konferencji wzięło udział 99 osób. Byli to pracownicy oraz doktoranci następujących uczelni: Akademii Górniczo-Hutniczej w Krakowie, Państwowej Wyższej Szkoły Zawodowej im. Prezydenta Stanisława Wojciechowskiego w Kaliszu, Politechniki Białostockiej, Politechniki Łódzkiej, Politechniki Opolskiej, Politechniki Rzeszowskiej, Politechniki Wrocławskiej, Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Uniwersytetu Ekonomicznego w Katowicach, Uniwersytetu Ekonomicznego w Krako-

wie, Uniwersytetu Ekonomicznego w Poznaniu, Uniwersytetu Ekonomicznego we Wrocławiu, Uniwersytetu Gdańskiego, Uniwersytetu im. Adama Mickiewicza w Poznaniu, Uniwersytetu Mikołaja Kopernika w Toruniu, Uniwersytetu Łódzkiego, Uniwersytetu Przyrodniczego w Poznaniu, Uniwersytetu Szczecińskiego, Uniwersytetu w Białymstoku, Wyższej Szkoły Bankowej w Toruniu, Wyższej Szkoły Bankowej we Wrocławiu, Zachodniopomorskiego Uniwersytetu Technologicznego w Szczecinie, a także przedstawiciele NBP, Urzędu Statystycznego w Poznaniu oraz Wydawnictwa C.H. Beck.

W trakcie 3 sesji plenarnych oraz 11 sesji równoległych wygłoszono 64 referaty poświęcone aspektom teoretycznym i aplikacyjnym zagadnień klasyfikacji i analizy danych. Odbyła się również sesja plakatowa, na której zaprezentowano 19 plakatów. Obradom w poszczególnych sesjach konferencji przewodniczyli: prof. dr hab. Józef Pociecha; prof. dr hab. Krzysztof Jajuga; dr hab. Andrzej Sokołowski, prof. nadzw. UEK; prof. dr hab. Iwona Roeske-Słomka; prof. dr hab. Mirosław Krzyśko; dr hab. Janusz Korol, prof. nadzw. US; prof. dr hab. Tadeusz Kufel; dr hab. Barbara Pawełek, prof. nadzw. UEK; dr hab. Andrzej Młodak; dr hab. Jan Paradysz, prof. nadzw. UEP; prof. dr hab. Stanisława Bartosiewicz; prof. dr hab. Eugeniusz Gatnar; prof. dr hab. Jadwiga Suchecka; dr hab. Elżbieta Gołata, prof. nadzw. UEP.

Teksty referatów przygotowane w formie recenzowanych artykułów naukowych stanowią zawartość przygotowywanej do druku publikacji z serii Taksonomia nr 19 (w ramach Prac Naukowych Uniwersytetu Ekonomicznego we Wrocławiu). Zaprezentowano następujące referaty:

Stanisława Bartosiewicz, Jeszcze raz o skutkach subiektywizmu w analizie wielowymiarowej

Referat jest uzupełnieniem artykułu pt. *Opowieść o skutkach subiektywizmu w analizie wielowymiarowej*. Informuje o skutkach subiektywizmu w wyborze metod klasyfikacji obiektów ze względu na zjawisko złożone. Skutkami subiektywizmu są oczywiście zróżnicowane wyniki klasyfikacji. Wnioski z przeprowadzonych badań są następujące: 1) różnice w klasyfikacji obiektów zależą od liczby założonych klas: im więcej klas, tym większe różnice; 2) skutki subiektywizmu najwyraźniej występują przy wyborze metody klasyfikacji (taksonomia wrocławska znacznie mniej różnicuje wyniki, niż metoda środków ciężkości), na drugim miejscu częściej notuje się skutki subiektywizmu przy wyborze rodzaju odległości, na trzecim (ostatnim) miejscu niekiedy, (rzadziej niż przy wyborze rodzaju odległości) pojawiają się skutki subiektywizmu przy wyborze metody normalizacji cech zjawiska złożonego.

Andrzej Sokołowski, Q uniwersalna miara odległości

W zadaniach taksonomicznych coraz częściej mamy do czynienia z danymi zawierającymi zmienne mierzone w różnych skalach. Jedną z możliwych strategii jest wykonanie analizy osobno dla jednorodnej grupy cech, a potem jakieś scalenie wyników. Druga strategia to zastosowanie miary odległości, która ze swej natury umożliwia jednocześnie wykorzystanie cech mierzonych w różnych skalach. Najbardziej dojrzałą

propozycją jest tu miara Marka Walesiaka GDM. W proponowanym referacie przede wszystkim wykraczamy poza tradycyjny podział cech ze względu na skalę pomiaru – na nominalne, porządkowe, przedziałowe i ilorazowe. Wyróżniono 15 rodzajów cech statystycznych. W dalszej kolejności zaproponowano takie przekształcenia tych cech (lub odległości), że składowe sumy w odległości typu Manhattan przyjmują wartości z przedziału $[0,1]$ i są niemianowane. To umożliwia policzenie odległości ogólnej.

Eugeniusz Gatnar, Jakość danych w systemach statystycznych banków centralnych (na przykładzie NBP)

W referacie przedstawiono analizę jakości danych w systemach statystycznych banków centralnych, biorąc jako przykład Narodowy Bank Polski. Dobrym wskaźnikiem jakości tych danych jest pozycja bilansu płatniczego Polski, nazywana „saldo błędów i opuszczeń”. Dane sprawozdawane przez instytucje finansowe oraz Ministerstwo Finansów i GUS nie były dokładne i wskazywały na występowanie wysokiego deficytu na rachunku bieżącym w stosunku do PKB. Podjęte prace przez Departament Statystyki NBP spowodowały rewizję importu w bilansie płatniczym w roku 2011 w wysokości ok. 10 mln PLN.

Marek Walesiak, Pomiar odległości obiektów opisanych zmiennymi mierzonymi na skali porządkowej

W referacie scharakteryzowano trzy strategie postępowania w pomiarze odległości obiektów opisanych zmiennymi mierzonymi na skali porządkowej:

1. Kodowanie kategorii (metody: zastąpienie kategorii rangami, kodowanie liniowe lub nieliniowe), potraktowanie zmiennych porządkowych jako zmiennych mierzonych na skali metrycznej (sztuczne wzmocnienie skali pomiaru zmiennych), a następnie zastosowanie miar odległości właściwych dla danych metrycznych (odległość euklidesowa lub miejska).

2. Kodowanie kategorii (zastąpienie kategorii rangami), a następnie zastosowanie odległości bazujących na rangach (np. odległość Kendalla, odległość Podaniego).

3. Zastosowanie miar odległości wykorzystujących dopuszczalne relacje na skali porządkowej (odległość GDM2).

Przedstawiono odpowiednie formuły odległości dla poszczególnych strategii oraz omówiono ich zalety i wady.

Krzysztof Jajuga, Marek Walesiak, XXV lat konferencji taksonomicznych – fakty i refleksje

W referacie w syntetycznej formie ujęto rys historyczny konferencji taksonomicznych oraz statystykę ogłoszonych i opublikowanych referatów w podziale na poszczególne lata i ośrodki akademickie. Ponadto przedstawiono fakty i refleksje płynące z 25 dotychczasowych konferencji taksonomicznych.

Józef Pociecha, Barbara Pawełek, Model SEM w analizie zagrożenia bankrutwem przedsiębiorstw w zmieniającej się koniunkturze gospodarczej – problemy teoretyczne i praktyczne

W dotychczas stosowanych modelach bankrutwa firm zakłada się, że zmienne występujące w teorii ekonomicznej są bezpośrednio obserwowane. Za źródła losowych odchyłeń przyjmuje się głównie błędy w zachowaniu się obiektów lub błędy w równaniu. W pracy zaproponowano zmianę dotychczasowego ujęcia problemu na podejście alternatywne, w którym za źródło losowej zmienności uznane zostałyby błędy pomiaru. Celem pracy jest przedstawienie propozycji wykorzystania modelu SEM w analizie zagrożenia bankrutwem przedsiębiorstw w zmieniającej się koniunkturze gospodarczej.

Tomasz Górecki, Mirosław Krzyśko, Funkcjonalna analiza składowych głównych

Kiedy obserwujemy dane jako funkcje czasu (np. finansowe szeregi czasowe) nazywamy je danymi funkcjonalnymi. W wielu zastosowaniach statystycznych realizacje ciągłych szeregów czasowych są dostępne jako obserwacje procesu w dyskretnych momentach czasowych. Z tego powodu kluczowym punktem jest konwersja danych dyskretnych na ciągłe funkcje. Konwersja ta może dotyczyć wygładzania. Jedną z procedur wygładzania jest reprezentacja każdej funkcji jako liniowej kombinacji K funkcji bazowych φ_k (Ramsay, J.O., Silverman, B.W., *Functional Data Analysis*, Springer, New York 2006). Analiza składowych głównych (PCA) ma na celu zmniejszenie oryginalnego zestawu zmiennych do mniejszego zbioru ich ortogonalnych liniowych kombinacji. Tak zredukowane dane są następnie wizualizowane (najczęściej na płaszczyźnie) przy zachowaniu maksymalnej zmienności oryginalnych danych. Odpowiednikiem PCA dla danych funkcjonalnych jest funkcjonalna analiza składowych głównych (FPCA).

Paweł Lula, Uczące się systemy pozyskiwania informacji z dokumentów tekstowych

Zasadniczym celem pracy jest prezentacja, klasyfikacja i ocena systemów pozyskiwania informacji z dokumentów tekstowych konstruowanych przy udziale mechanizmów uczących. W początkowej części pracy zdefiniowano pojęcie systemu uczącego się oraz zagadnienie pozyskiwania informacji. Następnie zaprezentowano strukturę oraz sposób funkcjonowania uczącego się systemu pozyskiwania informacji. Kluczowym elementem tego typu rozwiązań jest model zawartości informacyjnej. Jego charakterystyka i rodzaje są zasadniczym tematem kolejnego punktu pracy. Następnie przedstawiono rolę wiedzy zewnętrznej i metod uczenia maszynowego w poszczególnych rozwiązaniach. W kolejnej części referatu zaprezentowano rozważania dotyczące lokalnego lub globalnego charakteru poszczególnych rozwiązań.

Ewa Roszkowska, Zastosowanie metody TOPSIS do wspomaganie procesu negocjacji

Celem pracy jest prezentacja matematycznych podstaw systemu wspomaganie procesu negocjacji z wykorzystaniem procedury TOPSIS. Procedura TOPSIS umożliwia ocenę wartości ofert, ich uporządkowanie od najlepszej do najgorszej, wyznaczenie

ofert alternatywnych, szacowanie i porównanie wartości ustępstw, rozważenie poprawy kompromisu przez poszukiwanie rozwiązań Pareto optymalnych. Istotną zaletą tej metody jest prostota obliczeniowa, łatwość i przejrzystość interpretacji otrzymanych wyników, możliwość uogólnienia na zmienne lingwistyczne, przedziałowe, czy liczby rozmyte.

Andrzej Młodak, Sąsiedztwo obszarów przestrzennych w ujęciu fizycznym oraz społeczno-ekonomicznym – podejście taksonomiczne

W pracy zaprezentowano porównanie koncepcji sąsiedztwa fizycznego oraz społeczno-gospodarczego obszarów przestrzennych. To pierwsze dotyczy ich położenia na mapie administracyjnej oraz ewentualnej wspólnoty granic, zaś drugie – podobieństwa w zakresie złożonego zjawiska społeczno-ekonomicznego. Praca opisuje klasyczną teorię sąsiedztwa i jego macierzy oraz ukazuje jak można zaadoptować ją w przypadku wielowymiarowej analizy danych. Na zakończenie zaproponowano efektywną metodę porównywania obu typów sąsiedztwa wykorzystującą znaną z algebry postać normalną Smitha macierzy całkowitoliczbowych.

Andrzej Bąk, Modele kategorii nieuporządkowanych w badaniach preferencji – zmienne i dane

Wśród mikroekonometrycznych modeli kategorii nieuporządkowanych wyróżnia się najczęściej wielomianowy model logitowy, warunkowy model logitowy i mieszany model logitowy. Podstawę rozróżnienia tych typów modeli stanowi głównie charakter zmiennych objaśniających uwzględnionych w modelu. Rozróżnienie to nie jest jednak jednoznacznie interpretowane. Celem referatu jest wskazanie podstawowych różnic między typami modeli logitowych oraz prezentacja przykładów estymacji różnych typów tych modeli dla różnych typów danych z wykorzystaniem programu R.

Jacek Kowalewski, Zintegrowany model optymalizacji badań statystycznych

Referat stanowi próbę sprecyzowania teoretycznego modelu, który pozwalałby na przeprowadzenie optymalizacji procesu badań statystycznych. Złożoność zagadnienia powoduje, że prezentowany model ma charakter wieloetapowego zagadnienia wielokryterialnego z uwzględnieniem aspektów jakościowych, ekonomicznych oraz aktualności informacji. Część rozważań poświęcono przejściu z kryteriów jednostkowych dla poszczególnych informacji na globalne funkcje celu dla całego systemu badań statystycznych. Przedstawiono także warunki niezbędne do praktycznego zastosowania zaprezentowanego modelu.

Jan Paradysz, Karolina Paradysz, Obszary bezrobocia w Polsce – problem benchmarkowy

W statystyce małych obszarów istnieje wiele estymatorów pośrednich, co do których powstaje problem kryteriów wyboru oraz oceny wyników estymacji. Jedno z bardziej perspektywicznych rozwiązań stwarza statystyka lokalnego rynku pracy. Badanie

aktywności ekonomicznej ludności jest systematyczne, powtarzalne, obejmuje wszystkie gospodarstwa domowe. Można go zintegrować ze statystyką ludności oraz rejestrami administracyjnymi, dzięki czemu dysponujemy względnie stałą bazą zmiennych wspomagających. Szacując wszystkie elementy rynku pracy i sumując realizację odpowiednich estymatorów mogliśmy porównać je z liczbą ludności i ocenić ich dokładność. W wyniku tego porównania trochę lepszym okazał się estymator EB. Estymatory pośrednie charakteryzowały się też mniejszymi standardowymi błędami szacunku.

Tomasz Szubert, W co grać aby jak najmniej przegrać? Próba klasyfikacji systemów gry w zakładach bukmacherskich

W podjętym opracowaniu próbowano udowodnić, że jakakolwiek strategia gry w zakładach bukmacherskich jest nieopłacalna dla graczy, a zyski z tej rozrywki czerpią jedynie firmy organizujące te zakłady. Dla realizacji postawionego zadania wykorzystano informacje o kursach meczów piłkarskich z 6 ostatnich lat w wybranych krajach europejskich. W badaniu posłużono się m.in. analizą korelacji i analizą skupień oraz regresją logistyczną. W toku przeprowadzonych analiz okazało się, że jedynie tzw. progresja z nieograniczoną, maksymalną stawką zakładu jest w stanie przynieść dodatnią stopę zwrotu z gry, ale wymaga ona zaangażowania bardzo dużego kapitału (niekiedy rzędu kilku milionów złotych), co dla wielu graczy jest barierą nie do pokonania, a ponadto w praktyce firmy bukmacherskie stosują limity takich maksymalnych stawek, blokując tym samym szanse osiągnięcia zysku przez graczy.

Izabela Szamrej-Baran, Identyfikacja ubóstwa energetycznego z wykorzystaniem metod taksonomicznych

Głównym celem referatu jest próba identyfikacji i analizy zjawiska ubóstwa energetycznego. Podjęto w nim próbę specyfikacji listy zmiennych opisujących ubóstwo energetyczne. W oparciu o nie skonstruowano ranking krajów UE. Następnie, za pomocą hierarchicznych i niehierarchicznych metod grupowania dokonano ich podziału na grupy krajów podobnych pod względem cech charakteryzujących ubóstwo energetyczne. Przeprowadzono także klasyfikację z wykorzystaniem rozmytej metody *c-średnich*.

Sylwia Filas-Przybył, Tomasz Klimanek, Jacek Kowalewski, Próba wykorzystania modelu grawitacji w analizie dojazdów do pracy

W referacie podjęto próbę wykorzystania modelu grawitacji do opisu ciężenia określonych jednostek podziału przestrzennego kraju do ośrodków miejskich, pełniących istotną rolę w analizie lokalnych i regionalnych rynków pracy. Dotychczas możliwość aplikacji modeli grawitacji w warunkach polskich była ograniczona dostępnością danych, a także ograniczonymi możliwościami technicznymi związanymi z przetwarzaniem informacji przestrzennej. Dopiero prace podjęte w 2008 roku przez Ośrodek Statystyki Miast Urzędu Statystycznego w Poznaniu wypełniły lukę informacyjną w zakresie badań dojazdów do pracy, która istniała w Polsce od 1988 roku.

Marta Dziechciarz-Duda, Anna Król, Klaudia Przybysz, Minimum egzystencji a czynniki warunkujące skłonność do korzystania z pomocy społecznej. Klasyfikacja gospodarstw domowych

Celem podjętych badań była klasyfikacja gospodarstw domowych uprawnionych ze względu na niskie dochody do korzystania z pomocy społecznej oraz próba identyfikacji czynników warunkujących skłonność gospodarstw do korzystania z tej pomocy. Przyjęto założenie, że ocena sytuacji bytowej rodziny nie zależy wyłącznie od dochodu na jednego członka rodziny. Analizowano zatem cechy dotyczące stanu posiadania konkretnych dóbr trwałego użytku oraz subiektywnej oceny sytuacji życiowej gospodarstwa domowego (cechy psychologiczne). Zastosowanie drzew klasyfikacyjnych pozwoliło na identyfikację tych zmiennych, które ze względu na wysoką zdolność dyskryminacyjną różnicowały gospodarstwa korzystające i niekorzystające z pomocy społecznej.

Hanna Dudek, Subiektywne skale ekwiwalentności – analiza na podstawie danych o satysfakcji z osiągniętych dochodów

W referacie przedstawiono wyniki estymacji skal ekwiwalentności na podstawie danych o subiektywnej ocenie własnej sytuacji dochodowej przez gospodarstwa domowe. W tym celu zastosowano metodę częściowo uogólnionych uporządkowanych modeli logitowych. Analizę przeprowadzono na podstawie danych z badań budżetów gospodarstw domowych zrealizowanych przez GUS w 2009 roku. Proponowana metoda jest rozszerzeniem podejścia zaprezentowanego w publikacji [Schwarze J., *Using panel data on income satisfaction to estimate equivalence scale elasticity*, „Review of Income and Wealth” 2003, 49, s. 359-372].

Joanicjusz Nazarko, Ewa Chodakowska, Marta Jarocka, Segmentacja szkół wyższych metodą analizy skupień versus konkurencja technologiczna ustalona metodą DEA — studium komparatywne

W referacie przedstawiono możliwość wykorzystania analizy skupień i idei konkurencji technologicznej w metodzie DEA do segmentacji szkół wyższych. Na przykładzie danych z Rankingu Szkół Wyższych „Perspektyw” i „Rzeczpospolitej”, wykorzystując różne kombinacje zborów zmiennych kryterialnych oraz metod segmentacji, dokonano czterech klasyfikacji uczelni. Następnie określono charakterystyczne cechy analizowanych obiektów należących do wyłonionych jednorodnych grup. Dokonano również oceny zgodności wyników klasyfikacji, wykorzystując miarę Randa oraz sformułowano kilkukryterialne zestawienie porównujące zastosowane w badaniu metody segmentacji.

Ewa Chodakowska, Wybrane metody klasyfikacji w konstrukcji ratingu szkół

W referacie zaprezentowano możliwość wykorzystania analizy skupień (metoda Warda, k-średnich) oraz metod porządkowania liniowego (odległość euklidesową, miarę Hellwiga, miarę GDM Walesiaka) do konstrukcji ratingu szkół. Oceny podobieństw wyników klasyfikacji dokonano za pomocą miar zgodności Randa i Nowaka. Przeprowa-

dzono walidację wyników grupowania, wykorzystując indeks silhouette. Zinterpretowano wyniki otrzymanej klasyfikacji. Celem badania było zweryfikowanie hipotezy o użyteczności wybranych metod klasyfikacji do ratingu efektywności nauczania w szkołach.

Tomasz Górecki, Klasyfikacja szeregów czasowych bazująca na pochodnych

W ostatnich latach popularność szeregów czasowych stale wyrasta. Biorąc pod uwagę szerokie wykorzystanie nowoczesnych technologii informacyjnych, duża liczba szeregów czasowych powstaje w biznesie, medycynie, biologii itp. W konsekwencji nastąpił znaczny wzrost zainteresowania w badaniu tego typu danych, co z kolei spowodowało wysyp wielu prac wprowadzających nowe metody indeksowania, klasyfikacji, grupowania oraz aproksymacji szeregów czasowych. W szczególności, wprowadzono wiele nowych miar odległości pomiędzy szeregami. W pracy zaproponowana została nowa miara odległości bazująca na pochodnych funkcji. W przeciwieństwie do dobrze znanych z literatury podejść, miara ta uwzględnia również kształt szeregu czasowego. W celu kompleksowego porównania, przeprowadzone zostały eksperymenty badające wydajność proponowanej miary na 20 szeregach czasowych z różnych dziedzin. Eksperymenty wykazały, że nasza metoda zapewnia wyższą jakość klasyfikacji.

Bartosz Soliński, Sektor energetyki odnawialnej w krajach UE – klasyfikacja w świetle strategii zarządzania zmianą

Celem referatu jest sklasyfikowanie sektora energetyki odnawialnej w poszczególnych krajach UE, na klasy obserwacji o podobnych właściwościach, dające możliwość wyodrębnienia krajów mających główny wpływ na sukces wykorzystania odnawialnych źródeł energii w świetle strategii zarządzania zmianą. W analizach przeprowadzonych w referacie zostały wykorzystane metody analizy skupień (tzw. drzewa klasyfikacyjne), które pozwoliły na uzyskanie jednorodnych grup krajów będących przedmiotem badania, w ramach których nastąpiło wyodrębnienie ich zasadniczych cech.

Elżbieta Gołata, Grażyna Dehnel, Rejestry administracyjne w analizie przedsiębiorczości

Celem referatu jest ocena możliwości zastosowania metodologii statystyki małych obszarów do szacunku podstawowych charakterystyk ekonomicznych dotyczących małych, średnich i dużych przedsiębiorstw, na podstawie informacji zawartych w źródłach statystycznych i administracyjnych. Dokonano oceny użyteczności informacji zawartych w rejestrach administracyjnych pod kątem ich wykorzystania w estymacji pośredniej typu GREG czy EBLUP. Estymacja ta pozwala na dostarczenie szacunków na niskim poziomie agregacji, na które popyt nieustannie rośnie.

Tadeusz Kufel, Marcin Błazejowski, Paweł Kufel, Modelowanie niefinansowych danych transakcyjnych (tickowych)

W referacie zostały przedstawione płaszczyzny modelowania niefinansowych danych transakcyjnych (tickowych). Niefinansowe dane transakcyjne dotyczące danych

sprzedażowych – danych paragonowych wykorzystywane są w analizie koszykowej. Analiza czasów trwania obsługi klienta, czyli obsługi transakcji, a także przerw w obsłudze klienta pozwala na optymalizację pracy kanałów obsługi klientów. Analiza cen transakcyjnych sprzedaży nieruchomości i pojazdów pozwala na ocenę roli poszczególnych cech nieruchomości (pojazdów) i wyznaczanie cen ofertowych. Normowanie danych transakcyjnych poprzez agregowanie do danych o jednakowych jednostkach czasu (dane minutowe, godzinowe, dobowe, tygodniowe) pozwala na ocenę natężenia zjawiska w czasie, co umożliwia ocenę cykliczności zjawisk.

Katarzyna Chudy, Marek Sobolewski, Kinga Stępień, Wykorzystanie metod taksonomicznych w prognozowaniu wskaźników rentowności banków giełdowych w Polsce

Badania dotyczące rentowności banków wymagają zgromadzenia odpowiednich danych finansowych. Informacje na temat działalności banków są prezentowane w różnych źródłach i różnych okresach sprawozdawczych. Główny problem, jaki występuje na tym etapie to dostęp do ujednoliconych i aktualnych danych. Rozwiązaniem tego problemu może być zastosowanie metod prognostycznych bazujących na danych historycznych. Celem referatu jest skonstruowanie prognozy wartości wskaźników rentowności ROA i ROE dla banków działających na GPW w Warszawie na lata 2010 i 2011. Analizą zostało objętych 14 banków giełdowych. Prognoza została ustalona dla grup banków (metoda Warda), które w 2007 i 2009 roku charakteryzowały się podobnymi wartościami wskaźników ROA i ROE.

Katarzyna Dębkowska, Modelowanie upadłości przedsiębiorstw przy wykorzystaniu metod statystycznej analizy wielowymiarowej

Celem referatu jest ocena sprawności wybranych metod wielowymiarowej analizy statystycznej w prognozowaniu upadłości przedsiębiorstw. Porównano wyniki klasyfikacji trzech metod: drzew klasyfikacyjnych, regresji logitowej oraz analizy dyskryminacyjnej. W ramach badania stworzono bazę polskich przedsiębiorstw reprezentujących różne sektory, wśród których znaleźli się zarówno bankrucci jak i niebankrucci, a proporcja między jednymi a drugimi wyniosła 1:1. Każde przedsiębiorstwo zostało opisane za pomocą zmiennych diagnostycznych w postaci wskaźników finansowych. Dane do analizy zebrano na podstawie informacji zamieszczonych w bazie Emerging Markets Information Service (EMIS).

Alina Bojan, Wykorzystanie metod wielowymiarowej analizy danych do identyfikacji zmiennych wpływających na atrakcyjność wybranych inwestycji

Inwestor ma wiele możliwości na ulokowanie swego kapitału. Są to np. inwestycje w akcje wybranych branż, surowce, waluty czy też obligacje. Wraz ze zmieniającą się sytuacją gospodarczą zmieniać się może również ich atrakcyjność. Dlatego ważne jest umiejętne reagowanie na docierające sygnały i realokacja kapitału w odpowiednim momencie. W referacie dokonano próby identyfikacji zmiennych, które wpływają na to, iż dana grupa inwestycji traci na swej atrakcyjności w porównaniu do pozostałych.

Najpierw za pomocą metod klasyfikacyjnych wyróżnione zostały najbardziej i najmniej atrakcyjne grupy inwestycji z punktu widzenia stopy zwrotu i ryzyka, a następnie wyodrębnione zostały zmienne istotnie wpływające na zmianę segmentu wybranej inwestycji w czasie.

Justyna Brzezińska, Analiza logarytmiczno-liniowa w badaniu przyczyn umieralności w krajach UE

Analiza logarytmiczno-liniowa jest metodą badania niezależności zmiennych niemetrycznych, która pozwala na analizę dowolnej liczby zmiennych przy jednoczesnym uwzględnieniu interakcji zachodzących pomiędzy nimi. Modele, które budowane są hierarchicznie, oceniane są za pomocą statystyki chi-kwadrat, ilorazu wiarygodności oraz kryterium informacyjnego *AIC* oraz *BIC*. W niniejszym referacie zaprezentowane zostanie także kryterium Aitkina. Spośród wszystkich modeli wybrany zostaje model o najmniejszej złożoności, który jest dobrze dopasowany do danych. W programie **R** analiza logarytmiczno-liniowa dostępna jest dzięki funkcji `loglm` w pakiecie *MASS*. W niniejszym referacie analiza ta zaprezentowana zostanie do analizy przyczyn umieralności w krajach UE.

Aneta Rybicka, Bartłomiej Jefmański, Marcin Pełka, Analiza klas ukrytych w badaniach satysfakcji studentów

Badania satysfakcji klientów stanowią jedno z ważniejszych zagadnień poruszanych w ramach badań marketingowych. Prowadzenie tego typu badań oraz analiza ich wyników wymaga od badacza m.in. dużej wiedzy statystycznej bowiem obszar ten dotyczy zarówno zagadnienia zmiennych ukrytych, jak i sposobów analizy zmiennych, których wartości mierzone są na skalach niemetrycznych. Dlatego celem opracowania jest charakterystyka możliwości zastosowania modeli klas ukrytych w badaniach satysfakcji konsumentów, ponieważ spełniają one wszystkie postulaty stawiane analizie danych gromadzonych w ramach badań satysfakcji oraz pozwalają modelować zależności między zmiennymi ukrytymi. Zaprezentowane w referacie badanie empiryczne dotyczące pomiaru i analizy satysfakcji studentów jednego z wydziałów Uniwersytetu Ekonomicznego we Wrocławiu pozwoliło scharakteryzować możliwości aplikacyjne jednego z trzech głównych typów modeli klas ukrytych.

Marta Jarocka, Zastosowanie wybranych metod wielowymiarowej analizy porównawczej w hierarchizacji polskich uczelni

Celem referatu jest przedstawienie możliwości wykorzystania wybranych metod wielowymiarowej analizy porównawczej w konstrukcji rankingu szkół wyższych. Zbiór potencjalnych zmiennych diagnostycznych opracowany przez Kapitułę Rankingu Szkół Wyższych „Perspektyw” i „Rzeczpospolitej” zweryfikowano ze względu na ich wartość informacyjną. Do analizy zdolności dyskryminacyjnej zmiennych zastosowano klasyczny i medianowy współczynnik zmienności, zaś do zbadania stopnia skorelowania zmiennych wykorzystano metodę parametryczną zaproponowaną przez Z. Hellwiga.

Do budowy rankingu uczelni wykorzystano jedną z metod wielowymiarowej analizy porównawczej – metodę wzorcową tzw. miarę Hellwiga.

Alicja Grześkowiak, Agnieszka Stanimir, Badanie efektów kształcenia i warunkujących je czynników

W referacie podjęto próbę oceny wyników egzaminacyjnych osiąganych przez uczniów. Z jednej strony zanalizowano punktację osiąganą przez piętnastolatków w badaniu PISA, z drugiej strony przeprowadzono ocenę krajowych osiągnięć uczniów w trakcie egzaminu gimnazjalnego. Każdy z typów egzaminu ocenia zbliżoną grupę uczniów w podobnych obszarach wiedzy. Zanalizowanie wyników młodych Polaków wraz z wynikami uczniów z innych krajów jest niewątpliwie użyteczne ze względu na ciągłe modyfikacje naszego systemu kształcenia. Jednoczesne wykorzystanie w analizie uzyskanych przez uczniów punktów i innych czynników, które mogą wpływać na poziom wiedzy również jest wskazówką do prowadzenia dalszych zmian systemowych.

Bartłomiej Jefmański, Pomiar opinii respondentów z wykorzystaniem elementów teorii zbiorów rozmytych i środowiska **R**

Opracowanie stanowi propozycję pomiaru opinii respondentów polegającą na połączeniu zastosowania dyskretnych skal szacunkowych oraz liczb rozmytych. Zaprezentowano dwa warianty konstrukcji tzw. pytań pomocniczych w kwestionariuszu ankiety, które umożliwiają przekształcenie wyrażań lingwistycznych w liczby rozmyte (np. o trójkątnych lub trapezoidalnych funkcjach przynależności). Proponowane podejście zastosowano w badaniu opinii doradców zawodowych na temat jakości kształcenia ustawicznego w podregionie wałbrzyskim. Wyniki analizy potwierdziły różnice w interpretacji przez respondentów wartości lingwistycznych stanowiących punkty na zastosowanej w opracowaniu skali szacunkowej. Operacje arytmetyczne na liczbach rozmytych niezbędne do analizy danych ankietowych z zastosowaniem proponowanego podejścia przeprowadzono w programie **R**.

Julita Stańczuk, Porównanie metod statystyki klasycznej i technik sztucznych sieci neuronowych w wielostanowej klasyfikacji obiektów ekonomicznych

Głównym celem referatu jest analiza porównawcza jakości klasyfikacji przedsiębiorstw notowanych na Giełdzie Papierów Wartościowych w Warszawie z wykorzystaniem analizy dyskryminacyjnej oraz sztucznych sieci neuronowych. Wykazano, że sieci neuronowe, a konkretniej perceptron wielowarstwowy, są lepszym klasyfikatorem dla wielostanowej klasyfikacji obiektów ekonomicznych niż analiza dyskryminacyjna.

Daniel Papla, Przykład zastosowania metod analizy wielowymiarowej w badaniu kryzysów finansowych

Kryzysy finansowe są ważnym zjawiskiem dla całej gospodarki, ponieważ podczas kryzysu zwiększają się koszty pośrednictwa oraz koszty kredytów, trudniejszy jest również dostęp do kredytów. Powoduje to ograniczenia działalności sektora realnego, co może prowadzić do kryzysu również i w tym sektorze. Dosyć duża częstość

występowania kryzysów finansowych może prowadzić do wniosku, że sektor finansowy jest szczególnie wrażliwy na różnego rodzaju zaburzenia. Zwłaszcza kryzys z lat 2008-2010 pokazał, jak gospodarka światowa jest wrażliwa na zaburzenia w epoce globalizacji. Aby zbadać uwarunkowania rozprzestrzeniania się kryzysów finansowych wykorzystano w tym referacie metody analizy powiązań między światowymi rynkami kapitałowymi, takie jak wielowymiarowe rozkłady warunkowe, funkcje powiązań czy też dynamiczne modele warunkowej korelacji. Powinny one dać odpowiedź na pytanie, czy podczas kryzysu występuje istotna zmiana w zależnościach między rynkami, co w części tłumaczyłoby tak szybkie rozprzestrzenianie się kryzysu. W pierwszej części pracy omówiono obszerną literaturę dotyczącą kryzysów finansowych oraz metody wykorzystane w dalszej części referatu. Druga część zawiera wyniki badań empirycznych, a w ostatniej zawarta jest interpretacja otrzymanych wyników.

Jerzy Krawczuk, Skuteczność metod klasyfikacji w prognozowaniu kierunku zmian indeksu giełdowego S&P500

Kluczowe w prognozowaniu zachowania rynków finansowych jest określenie kierunku zmiany, czyli określenie czy nastąpi wzrost, czy spadek. Narzędziem, które może zostać wykorzystane do tego celu są używane w eksploracji danych klasyfikatory. Klasyfikator na podstawie zbioru uczącego (danych historycznych), może przydzielić klasę wzrostu bądź spadku. Skuteczność takiej klasyfikacji (prognozy) dla indeksu giełdy amerykańskiej S&P500 została zbadana w niniejszym referacie. Użyto siedmiu popularnych klasyfikatorów, w tym drzew decyzyjnych i klasyfikatora SVM.

Anna Czapkiewicz, Beata Basiura, Symulacyjne badanie wpływu zaburzeń na grupowanie szeregów czasowych w oparciu o modele *Copula*-GARCH

W pracy przedstawiono badanie symulacyjne, dotyczące badania poprawności przyjętej metody grupowania w oparciu o parametr wyznaczony z modelu *Copula*-GARCH. Ponadto zbadano wpływ zaburzeń rozkładów warunkowych procesu GARCH na wynik klasyfikacji. W szczególności zbadano jaki wpływ na wynik klasyfikacji ma nieuwzględnienie istniejącej skośności w modelowaniu szeregów czasowych.

Radosław Pietrzyk, Ocena efektywności inwestycji funduszy inwestycyjnych z tytułu doboru papierów wartościowych i umiejętności wykorzystania trendów rynkowych

Celem referatu jest zbadanie umiejętności zarządzających polskimi funduszami inwestycyjnymi. Efektywność inwestycji zostanie w referacie oceniona pod kątem umiejętności wykorzystania trendów rynkowych i dostosowywania strategii inwestycyjnych do zmieniającej się sytuacji na rynku giełdowym. Drugim czynnikiem oceny efektywności inwestycji jest dodatkowa stopa zwrotu z tytułu doboru papierów wartościowych. Pozwala ona na porównanie portfeli inwestycyjnych między sobą oraz z portfelami zarządzanymi pasywnie.

Aleksandra Witkowska, Marek Witkowski, Zastosowanie metody Panzara-Rosse'a do pomiaru poziomu konkurencji w sektorze banków spółdzielczych

W pracy podjęto próbę pomiaru poziomu konkurencji w sektorze banków spółdzielczych funkcjonujących w Polsce, jako że poziom konkurencji ma istotne znaczenie dla stabilności tych banków, wpływa bowiem na ich dochodowość oraz jakość i dostępność oferowanych produktów. Dążono w niej do uzyskania odpowiedzi na następujące pytania:

1. Jaki był poziom i typ konkurencji w badanym segmencie banków spółdzielczych i jak dalece był on zmienny w czasie,
2. Czy zmiany w poziomie konkurencji są powiązane ze zmianami poziomu koncentracji w przedmiotowym segmencie banków spółdzielczych.

Jako narzędzie badawcze zastosowano jedną z metod WAS – metodę Panzara-Rosse’a.

Marcin Pełka, Podejście wielomodelowe z wykorzystaniem metody *boosting* w analizie danych symbolicznych

Celem referatu jest zaprezentowanie możliwości wykorzystania metody *boosting* w agregacji modeli dla danych symbolicznych z zastosowaniem metody *k*-najbliższych sąsiadów dla danych symbolicznych jako klasyfikatora bazowego. W referacie przedstawiono podstawowe pojęcia z zakresu analizy danych symbolicznych, metody *k*-najbliższych sąsiadów dla danych symbolicznych. Zaprezentowano w nim także możliwość zastosowania podejścia wielomodelowego *boosting* dla danych symbolicznych. W części empirycznej przedstawiono zastosowanie podejścia wielomodelowego dla danych symbolicznych dla przykładowych zbiorów danych.

Justyna Wilk, Analiza porównawcza oprogramowania komputerowego w klasyfikacji danych symbolicznych

Celem referatu jest ocena przydatności dostępnego na rynku oprogramowania statystycznego w klasyfikacji danych symbolicznych. W referacie wyjaśniono podstawowe pojęcia analizy danych symbolicznych. Wskazano metody analizy skupień i procedurę klasyfikacji danych symbolicznych. Scharakteryzowano podstawowe własności pakietów stosowanych w analizie skupień i zakres oprogramowanych metod. Określono ich użyteczność w poszczególnych etapach procedury klasyfikacji danych symbolicznych. Z uwagi na relatywnie duże możliwości aplikacyjne, bliżej scharakteryzowano pakiety R i SODAS.

Tomasz Bartłomowicz, Justyna Wilk, Zastosowanie metod analizy danych symbolicznych w przeszukiwaniu dziedzinowych baz danych

Treścią referatu jest propozycja wykorzystania metodologii analizy danych symbolicznych w filtrowaniu dziedzinowych baz danych. Proponowane rozwiązanie, obok zmiennych klasycznych uwzględnia zmienne symboliczne, które opisują obiekty bez utraty informacji. Ponadto w rozwiązaniu wykorzystano znormalizowaną miarę Ichino-Yaguchiego dla danych symbolicznych. W opinii autorów, połączenie to umożliwi przeszukiwanie baz danych w oparciu o wszystkie możliwe kryteria bez względu na

rodzaj występujących zmiennych. W praktyce oznacza to rozszerzenie możliwości występujących w dostępnych powszechnie mechanizmach filtrowania dziedzinowych baz danych, co zostało zaprezentowane na przykładzie przeszukiwania ofert nieruchomości.

Andrzej Magruk, Analiza skupień w metodyce badawczej foresightu

W referacie przedstawiono wyniki autorskiej – opartej na analizie skupień – klasyfikacji metod badawczych projektów foresight. Autor zidentyfikował bogatą listę metod badawczych możliwych do zastosowania w omawianych projektach. Z uwagi na fakt, że dotychczas wiele metod należy do kilku typów autor pokusił się o dokonanie własnej klasyfikacji, ukazując grupy metod charakteryzujących się podobnymi cechami, odkrywając jednocześnie nieznaną do tej pory strukturę analizowanych danych. Nowatorskie grupowanie może być pomocne, z jednej strony dla początkujących praktyków foresightu, z drugiej natomiast dla ekspertów zajmujących się w sposób profesjonalny metodyką badawczą foresight.

Kamila Migdał Najman, Propozycja hybrydowej metody grupowania opartej na sieciach samouczących

W referacie autorka dokonuje prezentacji hybrydowej metody grupowania opartej na sieciach neuronowych samouczących typu SOM i GNG. Autorka weryfikuje potencjał proponowanej hybrydowej metody grupowania typu: sieć SOM + metoda k-średnich. Proponowane podejście weryfikuje na przykładzie badania preferencji i zachowań komunikacyjnych mieszkańców Gdyni w 2010 roku.

Dorota Rozmus, Porównanie dokładności taksonomii spektralnej oraz zagregowanych algorytmów taksonomicznych opartych na idei metody bagging

Stosując metody taksonomiczne w jakimkolwiek zagadnieniu klasyfikacji ważną kwestią jest zapewnienie wysokiej poprawności wyników grupowania. Stąd też w literaturze wciąż proponowane są nowe rozwiązania, które mają przynieść poprawę dokładności grupowania w stosunku do tradycyjnych metod. Przykładem mogą tu być metody polegające na zastosowaniu podejścia zagregowanego oraz algorytmy spektralne. Głównym celem tego referatu jest porównanie dokładności zagregowanych algorytmów taksonomicznych opartych na idei metody *bagging* oraz spektralnego algorytmu taksonomicznego zaproponowanego przez Ng A. Y., Jordan M. I., Weiss Y., *On spectral clustering: Analysis and an algorithm*, „Advances in Neural Information Processing Systems”, 2001, 849-856.

Krzysztof Najman, Grupowanie dynamiczne w oparciu o samouczące się sieci GNG

Wraz ze stale rozwijającą się techniką informatyczną, dynamicznie zwiększa się także ilość danych zbieranych w różnych systemach komputerowych. Jedną z cech baz danych tworzonych dynamicznie w systemie on-line jest między innymi dynamicznie zmieniająca się struktura grupowa jednostek. Grupowanie jednostek w takiej bazie danych wymaga zastosowania specjalnych metod. W referacie sformułowano wymagania stawiane takiej metodzie, a także zaproponowano użycie w opisywanym celu

samouczącej się sieci neuronowej typu GNG. W referacie w oparciu o badania teoretyczne i eksperyment badawczy weryfikowano hipotezę o przydatności sieci GNG w dynamicznym grupowaniu danych.

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

Drzewa klasyfikacyjne należą do tych algorytmów uczących, które mogą być wykorzystane w sytuacji występowania braków danych w zbiorze danych. W pracy porównano kilka wybranych technik postępowania w sytuacji występowania braków danych. Wykorzystano podejście symulacyjne generując różne proporcje i mechanizmy powstawania braków danych w zbiorach danych pochodzących z repozytorium baz danych na Uniwersytecie Kalifornijskim w Irvine oraz z badań własnych. Celem badań była ocena wpływu wybranych metod imputacji danych na wyniki klasyfikacji obiektów z wykorzystaniem drzew klasyfikacyjnych w przypadku zbiorów danych o niewielkiej liczebności.

Mariusz Kubus, Zastosowanie wstępnego uwarunkowania zmiennej objaśnianej do selekcji zmiennych

Paul D., Bair E., Hastie T., Tibshirani R. [*“Pre-conditioning” for feature selection and regression in high-dimensional problems*, *Annals of Statistics* 36(4), 2008, s. 1595-1618] zaproponowali metodę wstępnego uwarunkowania zmiennej objaśnianej, którą realizuje się w dwóch krokach. W pierwszym przewidywane są wartości zmiennej objaśnianej za pomocą metody głównych składowych z nauczycielem. W drugim budowany jest jeszcze raz model regresji, ale na zbiorze uczącym, w którym rolę zmiennej objaśnianej pełnią jej wartości teoretyczne estymowane w kroku pierwszym. Tu wykorzystuje się LASSO, które pozwala wyeliminować zmienne nieistotne. W referacie zaproponowano modyfikację oryginalnej metody oraz za pomocą symulacji oceniono ich przydatność w zadaniu selekcji zmiennych w przypadku modelu liniowego z interakcjami.

Barbara Batóg, Jacek Batóg, Wykorzystanie analizy dyskryminacyjnej do identyfikacji czynników determinujących stopę zwrotu z inwestycji na rynku kapitałowym

W referacie wykorzystano analizę dyskryminacyjną do wyboru zbioru wskaźników (zmiennych dyskryminacyjnych) determinujących stopę zwrotu z inwestycji na rynku kapitałowym w Polsce. Wśród rozpatrywanych przekrojów badania uwzględniono okres wzrostów oraz okres spadków cen akcji, jak również wybrane sektory gospodarcze. Ocenie poddano dodatkowo trafność uzyskiwanych klasyfikacji w zależności od sposobu podziału badanych obiektów na grupy.

Katarzyna Wójcik, Janusz Tuchowski, Analiza porównawcza miar podobieństwa tekstów bazujących na macierzy częstości z bazującymi na wiedzy dziedzinowej

Zasadniczym celem niniejszej pracy jest próba oceny przydatności znanych z literatury miar podobieństwa tekstów bazujących na macierzy częstości z tymi bazującymi na wiedzy dziedzinowej w postaci ontologii. W kolejnych punktach referatu przedstawione zostały najpierw dokumenty tekstowe, które były porównywane w badaniu, a następnie wybrane miary podobieństwa oparte na macierzy częstości i ich analiza symulacyjna. W dalszej części zaprezentowana została ontologia wykorzystana w badaniach oraz wyniki przeprowadzonej analizy symulacyjnej miar opartych na wiedzy reprezentowanej przez tę ontologię. Na tej podstawie została podjęta próba oceny przydatności tych miar.

Iwona Staniec, Analiza czynnikowa w identyfikacji obszarów determinujących doskonalenie systemów zarządzania w polskich organizacjach

Celem referatu jest przedstawienie empirycznego wykorzystania analizy czynnikowej w badaniu obszarów determinujących doskonalenie systemów zarządzania w polskich organizacjach. Zastosowanie w tym przypadku analizy czynnikowej umożliwia pełniejszą diagnozę współzależności między poszczególnymi elementami systemów zarządzania, a ich istotnością dla doskonalenia oraz zależnością od branży.

Marek Lubicz, Maciej Zięba, Adam Rzechonek, Konrad Pawełczyk, Jerzy Kołodziej, Jerzy Błaszczak, Analiza porównawcza wybranych technik eksploracji danych do klasyfikacji danych medycznych z brakującymi obserwacjami

W praktycznych zadaniach klasyfikacji, na przykład w analizie danych medycznych, występuje stosunkowo często konieczność wnioskowania w oparciu o dane niekompletne. Celem pracy jest porównanie efektywności wybranych podejść do rozwiązywania problemu klasyfikacji z brakującymi obserwacjami dla wybranych klasyfikatorów prostych i złożonych oraz dla różnych metod transformacji cech z brakującymi obserwacjami. W badaniach zastosowano implementacje technik eksploracji danych w środowisku STATISTICA Data Miner oraz systemie uczenia maszynowego WEKA. Jako dane do klasyfikacji wykorzystano bazę danych o pacjentach leczonych operacyjnie z powodu raka płuca we Wrocławskim Ośrodku Torakochirurgii w latach 2000-2011.

Iwona Foryś, Wielowymiarowa analiza wpływu cech mieszkań na ich atrakcyjność cenową w obrocie wtórnym na przykładzie lokalnego rynku mieszkaniowego

Przedmiotem analizy jest obrót na wtórnym rynku mieszkaniowym. Pozytywnie zweryfikowano hipotezę o stałości cech jakościowych wpływających na cenę, niezależnie od sytuacji na rynku mieszkaniowym. W tym celu wykorzystano analizę logliniową. Wykazano, iż rozszerzenie modelu o kolejne zmienne oraz interakcje rzędu wyższego niż drugi nie poprawiało dopasowania modelu. W badanych latach najlepsze dopasowanie modelu przy zmiennej zależnej cena uzyskano dla lokalizacji oraz powierzchni lokalu. Najczęściej nieistotne okazały się zmienne opisujące lokalizację mieszkania w budynku oraz zbywane prawo do lokalu.

Ewa Genge, Analiza skupień oparta na mieszankach uciętych rozkładów normalnych

W referacie przedstawione zostało najnowsze podejście modelowe, polegające na usuwaniu obserwacji oddalonych, w trakcie estymacji parametrów mieszanki. W podejściu tym na macierze kowariancji nakładane są pewne ograniczenia, wyznaczone są funkcje wiarygodności uciętych zmiennych losowych, a parametry takiej mieszanki szacowane są za pomocą zmodyfikowanej wersji algorytmu EM, tj. algorytmu TCLUS [García-Escudero L.A., Gordaliza A., Matrán C., Mayo-Isacar A., *A review of clustering methods*, „Advances of data analysis and classification”, Springer, 2010, 4, s. 89-109]. Przeprowadzone zostały również badania porównawcze z bardziej popularnym, odpornym modelem mieszanek rozkładów normalnych.

Jerzy Korzeniewski, Ocena efektywności metody uśredniania zmiennych i metody Ichino selekcji zmiennych w analizie skupień

Metodę Ichino można stosować do zmiennych mierzonych na dowolnych skalach. Pełka i Wilk zbadali tę metodę, ale badanie ograniczało się do kilku modeli struktur skupień zaś kryterium oceny była stabilność grupowania mierzona indeksem sylwetkowym. Metoda uśredniania zmiennych nieistotnych dla struktury skupień (Fraiman R., Justel A., Svarc M. (2008), Selection of Variables for Cluster Analysis and Classification Rules, JASA, 103) jest nową metodą niezbadaną dotychczas w żadnym obszerniejszym eksperymencie. W referacie przedstawione są wyniki badania symulacyjnego, w którym kryteriami są pamięć, precyzja i indeks Randa. Metody konkurencyjne, z którymi porównywana jest metoda Ichino to; HINoV oraz dwie modyfikacje HINoV (VSKM i VAF z indeksem skupialności). Andrzej Dudek, SMS – propozycja nowego algorytmu analizy skupień

Klasyfikacja spektralna (*spectral clustering* – Ng, A., Jordan, M., Weiss, Y. (2002), *On spectral clustering: analysis and an algorithm*, W: T. Dietterich, S. Becker, Z. Ghahramani (Eds.), *Advances in Neural Information Processing Systems 14*. MIT Press, 849-856) i klasyfikacja za pomocą średniej przesunięcia okna w kierunku wektora średniej (*Mean Shift clustering* – Wang, S., Qiu, W., and Zamar, R. H. (2007), CLUES: *A non-parametric clustering method based on local shrinking*. *Computational Statistics & Data Analysis*, Vol. 52, issue 1, 286-298) to dwa stosunkowo nowe podejścia w analizie skupień, dające, zwłaszcza dla skupień o nietypowych kształtach lepsze rezultaty niż klasyczne metody k-średnich, k-medoidów czy hierarchiczne metody aglomeracyjne. Referat zawiera propozycję algorytmu o roboczej nazwie SMS (*Spectral-Mean Shift*) łączącego cechy obu podejść wyróżniającego się wśród innych algorytmów analizy skupień m.in.: możliwością analizy skupień o nietypowych kształtach; możliwością automatycznego rozpoznawania liczby skupień; lepszą odpornością na zmienne zakłócające.

Artur Mikulec, Metody oceny wyniku grupowania w analizie skupień

W referacie dokonano przeglądu metod oceny wyniku grupowania. Omówiono trzy metody wyboru właściwej liczby skupień dla metod aglomeracyjnych zaproponowane przez Mojenę i Wisharta oraz zaimplementowane w programie *ClustanGraphics 8*. Wspomniane kryteria bazują na relatywnych wartościach różnych poziomów połączeń obiektów na wykresie drzewa – *best cut significance test (upper tail, moving average)* oraz sprawdzaniu losowości podziału obiektów na wykresie drzewa – *tree validation*. Treść referatu zilustrowana została przykładem empirycznym.

Artur Zaborski, Analiza PROFIT i jej wykorzystanie w badaniu preferencji

Celem referatu jest prezentacja analizy PROFIT będącej połączeniem skalowania wielowymiarowego oraz analizy regresji wielorakiej. Dla konfiguracji punktów reprezentujących obiekty otrzymanej za pomocą skalowania wielowymiarowego przeprowadza się analizę regresji wielorakiej, w której zmiennymi objaśniającymi są współrzędne obiektów na mapie percepcyjnej, a zmiennymi zależnymi oceny marek ze względu na poszczególne cechy. Na zakończenie zaprezentowano przykład badania preferencji marek samochodów z wykorzystaniem analizy PROFIT.

Karolina Bartos, Analiza skupień wybranych państw ze względu na strukturę wydatków konsumpcyjnych obywateli – zastosowanie sieci Kohonena

W referacie zaprezentowano wyniki analizy skupień 86 państw ze względu na udział różnego typu wydatków konsumpcyjnych obywateli w ich wydatkach konsumpcyjnych ogółem. W badaniu uwzględniono 12 typów wydatków, takich jak wydatki na: żywność i napoje bezalkoholowe, utrzymanie domu, odzież i obuwie, transport, wypoczynek i rekreację, alkohol i wyroby tytoniowe, telekomunikację, edukację, zdrowie, usługi hotelowe i catering, artykuły gospodarstwa domowego i usługi dla domu, pozostałe dobra i usługi. Do stworzenia grup wykorzystano sieci Kohonena.

Barbara Batóg, Magdalena Mojsiewicz, Katarzyna Wawrzyniak, Klasyfikacja gospodarstw domowych ze względu na bodźce do zawierania umowy o ubezpieczenie z wykorzystaniem modeli zmiennych jakościowych

Celem referatu jest klasyfikacja gospodarstw domowych na rynku ubezpieczeń w Polsce ze względu na bodźce skłaniające gospodarstwa domowe do zawierania umów ubezpieczeniowych na życie, dożycie i pokrycie kosztów leczenia. W badaniu wykorzystano drzewa klasyfikacyjne oraz dwumianowe modele logitowe, w których zmienną objaśnianą był popyt zrealizowany i potencjalny na ubezpieczenia danego rodzaju. Natomiast zmiennymi objaśniającymi były bodźce skłaniające do zawierania umów ubezpieczeniowych takie jak doświadczenia najbliższego otoczenia, strach przed katastrofami, niewydolność społecznego systemu ubezpieczeń i służby zdrowia. Analizę przeprowadzono na próbie 500 gospodarstw domowych.

Izabela Kurzawa, Zastosowanie modelu LA/AIDS do badania elastyczności cenowych popytu konsumpcyjnego w gospodarstwach domowych w relacji miasto-wieś

Celem pracy było zbadanie przydatności liniowej aproksymacji prawie idealnego systemu funkcji popytu LA/AIDS do wyznaczenia elastyczności cenowych popytu dla wybranych grup artykułów żywnościowych w relacji miasto-wieś. Parametry tego modelu oszacowano na podstawie danych z badania indywidualnych budżetów gospodarstw domowych w Polsce w 2006 roku. Na tej podstawie wyznaczono elastyczności cenowe własne i mieszane dla wybranych produktów żywnościowych. Klasa miejscowości zamieszkania gospodarstwa domowego różnicuje elastyczności cenowe.

Aleksandra Łuczak, Feliks Wysocki, Metody porządkowania liniowego obiektów opisanych za pomocą cech metrycznych i porządkowych

W referacie przedstawiono trzy podejścia do konstrukcji syntetycznego miernika rozwoju obiektów opisanych za pomocą cech mierzonych na skalach metrycznych i porządkowej. Podejście pierwsze polega na sprowadzeniu cech metrycznych i porządkowych do określeń lingwistycznych i liczb rozmytych (osłabienie skali cech metrycznych). Podejście drugie wiąże się z zastosowaniem uogólnionej miary odległości GDM zaproponowanej przez Walesiaka (zachowanie skal pomiarowych). Podejście trzecie polega na zamianie wartości cech porządkowych (lingwistycznych) na liczby rozmyte, co prowadzi do wyrażenia wartości tych cech w skali ilorazowej (wzmocnienie skali cech porządkowych). Proponowane podejścia zostały zastosowane do porządkowania liniowego powiatów województwa wielkopolskiego według stanu i jakości infrastruktury technicznej.

Grzegorz Kowalewski, Konstrukcja wskaźników złożonych w testach koniunktury

Wskaźniki złożone w testach koniunktury są zazwyczaj konstruowane w sposób arbitralny. W referacie przedstawiono sposób konstrukcji wskaźników złożonych z wykorzystaniem informacji zawartych w danych. Rozważania przedstawiono na przykładzie badań ankietowych prowadzonych przez GUS w przemyśle przetwórczym.

Agnieszka Sompolska-Rzechuła, Porównanie klasycznej i pozycyjnej taksonomicznej analizy zróżnicowania jakości życia w województwie zachodniopomorskim

Celem referatu było porównanie efektywności dwóch podejść w konstrukcji miernika rozwoju wykorzystujących klasyczny miernik Hellwiga oraz pozycyjny wektor medianowy Webera w badaniu obiektywnej jakości życia w województwie zachodniopomorskim. W analizie wykorzystano wskaźniki z następujących dziedzin: środowisko naturalne, demografia, rynek pracy, infrastruktura komunalna i mieszkania, edukacja, kultura i turystyka, ochrona zdrowia, transport drogowy, dochody i wydatki budżetów jednostek samorządu terytorialnego. Badanie dotyczyło 2008 roku. Podejście klasyczne dało lepsze wyniki pod względem homogeniczności, heterogeniczności i poprawności grupowania.

Krzysztof Szwarc, Klasyfikacja powiatów województwa wielkopolskiego ze względu na sytuację demograficzną

Procesy demograficzne zachodzące w naszym kraju nie przebiegają jednakowo w poszczególnych województwach, powiatach, a nawet gminach. Przemiany, które mają miejsce w Polsce przejawiają się między innymi w ograniczaniu płodności, przesunięciu wieku zawierania małżeństw a także wieku rodzenia pierwszego dziecka, wydłużaniu się przeciętnego dalszego trwania życia. Konsekwencją tych zmian jest starzenie się społeczeństwa. Celem niniejszego referatu jest dokonanie klasyfikacji powiatów badanego województwa ze względu na poziom podstawowych wskaźników, charakteryzujących sytuację demograficzną. Badaniem objęto lata 2002-2009, a źródłem danych była baza demograficzna zamieszczona na stronie internetowej Głównego Urzędu Statystycznego.

Ponadto w trakcie sesji plakatowej zaprezentowano następujące referaty:

Joanna Banaś, Małgorzata Machowska-Szewczyk, Analiza ocen skrzynek poczty elektronicznej za pomocą uporządkowanego modelu probitowego

Celem referatu jest analiza opinii studiujących użytkowników kont pocztowych na temat posiadanych skrzynek poczty. Źródło danych stanowiły kwestionariusze ankiety wypełniane przez studentów wybranych wydziałów szczecińskich uczelni, które zawierały pytania dotyczące zarówno charakterystyki skrzynki poczty elektronicznej, jak i intensywności jej używania. Do oceny obsługi poczty elektronicznej wykorzystano uporządkowany model probitowy, w którym występuje jakościowa zmienna zależna, opisująca kategorie uporządkowane przy założeniu rozkładu normalnego składnika losowego.

Iwona Bąk, Segmentacja gospodarstw domowych emerytów i rencistów pod względem wydatków na rekreację i kulturę

W referacie przedstawiono wyniki badań dotyczące segmentacji gospodarstw domowych emerytów i rencistów pod względem ich przeciętnych miesięcznych wydatków na rekreację i kulturę. Do klasyfikacji gospodarstw domowych wykorzystano drzewo regresyjne. Podstawę informacyjną badań stanowiły nieidentyfikowalne jednostkowe dane o dochodach i wydatkach indywidualnych gospodarstw domowych, pochodzące z badań budżetów gospodarstw domowych przeprowadzonych w 2009 r. przez GUS. Segmentacja została przeprowadzona dla reprezentatywnej próby 1308 gospodarstw domowych emerytów i 327 gospodarstw domowych rencistów.

Aneta Becker, Zastosowanie metody ANP do porządkowania województw Polski pod względem dynamiki wykorzystania ICT w latach 2008-2010

W referacie przedstawiono wyniki uporządkowania województw Polski pod względem wykorzystania technologii informacyjno-telekomunikacyjnych w przedsiębiorstwach (ICT, ang. Information and Communication Technology), w latach 2008-2010. W badaniach wykorzystano metodę wielokryterialnego wspomaganie decyzji ANP (ang. Analytic Network Process). Zastosowany algorytm pozwolił na analizę położenia

województw z uwagi na różne zaawansowanie obszarów w przestrzeni teleinformatycznej.

Katarzyna Dębowska, Ocena sytuacji finansowej sektorów przy użyciu metod wielowymiarowej analizy statystycznej

Przy ocenie sytuacji finansowej przedsiębiorstw pomocne są informacje dotyczące sektorowych wskaźników finansowych. Celem referatu jest dokonanie klasyfikacji sektorów pod względem wskaźników finansowych, wykorzystując metody wielowymiarowej analizy statystycznej: analizę skupień oraz drzewa klasyfikacyjne. Wykorzystane metody pozwoliły na pogrupowanie sektorów ze względu na wskaźniki finansowe. Klasyfikacją objęte zostały 53 sektory ze względu na 14 zmiennych diagnostycznych w postaci wskaźników finansowych. Źródłem informacji do przeprowadzenia klasyfikacji były dane dotyczące wskaźników sektorowych za 2009 r. Otrzymana klasyfikacja może stanowić cenne źródło informacji o sektorach dla przedsiębiorstw lub inwestorów.

Małgorzata Machowska-Szewczyk, Algorytm klasyfikacji rozmytej dla obiektów opisanych za pomocą zmiennych symbolicznych oraz rozmytych

Większość opracowanych metod klasyfikacji umożliwia grupowanie obiektów, opisanych za pomocą zmiennych ustalonego typu. W praktycznych zastosowaniach wiele obiektów może być charakteryzowanych przez różne typy cech. Celem pracy jest prezentacja rozmytego algorytmu klasyfikacji obiektów, które mogą być opisane jednocześnie za pomocą zmiennych numerycznych, symbolicznych lub rozmytych. Algorytm ten został zaproponowany przez Yanga, Hwanga i Chena, którzy zdefiniowali miarę niepodobieństwa między mieszanymi obiektami oraz zmodyfikowali rozmytą metodę c-środków. W referacie przedstawiono także numeryczny przykład zastosowania tej metody do obiektów o cechach mieszanych na podstawie danych rzeczywistych.

Anna Domagała, Propozycja metody doboru zmiennych do modeli DEA (procedura kombinowanego doboru wprzód)

Referat przedstawia propozycję metody doboru zmiennych (nakładów oraz wyników) do modeli DEA (Data Envelopment Analysis). Metoda polega na stopniowym dodawaniu zmiennych do modelu DEA, zgodnie z opracowanym algorytmem. Decyzję o wprowadzeniu danej zmiennej podejmuje się na podstawie dwóch kryteriów – procentowej zmiany średniej efektywności, wywołanej dodaniem takiej zmiennej do układu oraz współczynnika korelacji pomiędzy rezultatami badania bez tej zmiennej, a wynikami badania z jej uwzględnieniem.

Nina Drejerska, Mariola Chrzanowska, Iwona Pomianek, Taksonomiczna analiza porównawcza poziomu rozwoju gmin wiejskich i miejsko-wiejskich województwa mazowieckiego w latach 2002 i 2009

Rozwój społeczno-gospodarczy zachodzący w gminach wiejskich i miejsko-wiejskich w sąsiedztwie dużych aglomeracji powoduje tworzenie się specyficznych obszarów funkcjonalnych. Wyniki porównawczej analizy poziomu rozwoju badanych gmin

metodą k-średnich umożliwiają określenie obszaru funkcjonalnego Warszawy. Rozwój lokalny jednostek administracyjnych jest mierzony głównie wskaźnikami rozwoju społeczno-gospodarczego, stąd do badań wybrano zmienne charakteryzujące aspekty społeczne, gospodarcze i infrastrukturalne oraz uwzględniono zmienną określającą odległość danej gminy od centrum Warszawy. Badaniu poddane zostały wiejskie (229) i miejsko-wiejskie (50) gminy województwa mazowieckiego w latach 2002 i 2009.

Henryk Gierszał, Karina Pawlina, Maria Urbańska, Analiza statystyczna w badaniach zapotrzebowania na usługi teleinformatyczne sieci łączności ruchomej

W referacie zaprezentowano możliwości wykorzystania statystycznych analiz wielu zmiennych w badaniach popytu na usługi telekomunikacyjne na przykładzie planowanej ogólnokrajowej radiowej sieci łączności dyspozytorskiej. Wyniki przedstawiono dla analizy kanonicznej, dyskryminacyjnej, czynnikowej, korespondencji oraz skupień, które pozwoliły zidentyfikować zachowania użytkowników determinujące zapotrzebowanie na tego rodzaju usługi.

Hanna Gruchociak, Konstrukcja estymatora regresyjnego dla danych o strukturze dwupoziomowej

Głównym celem referatu jest przedstawienie przydatności metodologii modelowania dwupoziomowego w szacowaniu wartości zmiennych społeczno-gospodarczych. W pierwszej części opracowania omówiona została idea konstrukcji estymatora dla danych o strukturze dwupoziomowej. W części drugiej przeprowadzone zostało badanie empiryczne, mające na celu zastosowanie opisanego estymatora do szacowania wskaźnika zatrudnienia w przekroju powiatów. Po dokonaniu oszacowań porównano wiarygodności estymatora uwzględniającego dwupoziomową strukturę danych oraz zwykłego estymatora regresyjnego. Przeprowadzona analiza wykazała istotną poprawę jakości oszacowań uzyskanych przy zastosowaniu modelowania dwupoziomowego.

Tomasz Klimanek, Marcin Szymkowiak, Zastosowanie estymacji pośredniej uwzględniającej korelację przestrzenną w opisie niektórych charakterystyk rynku pracy

Referat przedstawia propozycje wykorzystania metod estymacji pośredniej (w tym także tej metody, która uwzględnia korelację przestrzenną) do oszacowania odsetka osób bezrobotnych w populacji osób w wieku 15 lat i więcej w przekroju podregionów w Polsce w I kwartale 2008 roku. Jest to bardziej szczegółowy poziom agregacji przestrzennej niż ten prezentowany w publikacjach Głównego Urzędu Statystycznego opartych na wynikach Badania Aktywności Ekonomicznej Ludności. Drugim celem jest porównanie miar precyzji estymatora bezpośredniego z precyzją estymatora typu EBLUP (*empirical best linear unbiased predictor*) oraz estymatora typu SEBLUP (uwzględniającego korelację przestrzenną).

Wojciech Kuźmiński, Wykorzystanie metod statystycznej analizy wielowymiarowej do oceny realizacji zadań statutowych Agencji Nieruchomości Rolnych

W referacie wykorzystano metody analiz wielowymiarowych do oceny stopnia realizacji zadań stawianych Agencji Nieruchomości Rolnych. Przeprowadzono analizy klasyfikacyjne, celem określenia w których częściach kraju obowiązki wynikające z ustaw realizowane są lepiej, a w których gorzej. Podobnie wykonano grupowanie pod względem podobieństwa regionów.

Jarosław Lira, Prognozowanie opłacalności żywca wieprzowego w Polsce

W pracy analizowano trafność krótkoterminowych prognoz opłacalności produkcji trzody chlewnej na podstawie relacji cenowych żywca wieprzowego w stosunku do cen targowiskowych żyta i jęczmienia oraz cen prosiąt. Do prognozowania targowiskowych cen zbóż zastosowano model wyrównywania wykładniczego Wintersa, a do sporządzenia prognoz targowiskowych cen żywca wieprzowego oraz cen prosiąt wykorzystano zmodyfikowaną metodę klasyczną opartą na modelu multiplikatywnym. Uzyskane prognozy posłużyły do określenia relacji cenowych, które następnie poddano ocenie trafności.

Christian Lis, Wykorzystanie metod taksonomicznych w ocenie konkurencyjności portów południowego Bałtyku w kontekście pogłębienia toru wodnego Szczecin-Świnoujście do głębokości 12,5 m

W badaniu autor wykorzystał metody taksonomiczne do oceny poprawy pozycji konkurencyjnej portu w Szczecinie wśród portów południowego Bałtyku w wyniku podjęcia inwestycji, polegającej na pogłębieniu toru wodnego ze Szczecina do Świnoujścia do głębokości 12,5 m. W referacie wykorzystana została uogólniona miara odległości, dla potrzeb badania nazwana taksonomicznym miernikiem konkurencyjności portów i analiza skupień. Wyniki badań dostarczają istotnych argumentów za koniecznością podjęcia kluczowej inwestycji w ramach Programu SONORA (*South-North Axis*). Program ten ma na celu poprawę infrastruktury transportowej w korytarzu transportowym Bałtyk-Adriatyk.

Beata Bieszk-Stolorz, Iwona Markowicz, Wykorzystanie wielomianowego modelu logitowego do oceny szansy podjęcia pracy przez bezrobotnych

W 2010 roku z Powiatowego Urzędu Pracy w Szczecinie zostało wyrejestrowanych ponad 20 tys. osób bezrobotnych. Przyczyny były bardzo różne. Celem referatu jest analiza szansy na podjęcie pracy oraz ocena prawdopodobieństwa podjęcia pracy o określonym charakterze w zależności od cech osób bezrobotnych: płci, wykształcenia i wieku. Analizę przeprowadzono dwuetapowo. Najpierw, przy wykorzystaniu dwumianowego modelu logitowego, zbadano wpływ cech osób bezrobotnych na szansę wyjścia z bezrobocia poprzez podjęcie pracy. W drugim etapie skupiono się na ocenie prawdopodobieństwa podjęcia określonej formy zatrudnienia. W tym celu zastosowano wielomianowy model logitowy.

Ewa Piotrowska, Statystyczny portret innowacyjności i rozwoju gospodarczego województw Polski

Celem opracowania jest konstrukcja statystycznego portretu innowacyjności i rozwoju gospodarczego województw Polski w oparciu o metody wielowymiarowej analizy statystycznej. Ocenę poziomu innowacyjności regionów przeprowadzono na podstawie odpowiednio dobranych agregatów wskaźników szczegółowych bazujących na metodologii tworzenia syntetycznego indeksu SII. Analizie poddano dane liczbowe z 2008 roku. Na podstawie otrzymanego portretu innowacyjności i rozwoju regionalnego dokonano oceny pozycji województwa podlaskiego w polskiej przestrzeni regionalnej. Otrzymane wyniki mogą zostać wykorzystane w zakresie strategii rozwojowych obejmujących wskazywanie wzorca rozwoju czy niwelowanie dysproporcji rozwojowych.

Lucyna Przezbórska-Skobiej, Jarosław Lira, Przestrzeń agroturystyczna Polski i jej waloryzacja

Celem niniejszego referatu była waloryzacja przestrzeni agroturystycznej Polski na poziomie powiatów ziemskich, z wykorzystaniem cech opisujących tę przestrzeń, uznanych w literaturze jako istotne dla rozwoju agroturystyki. Do przeprowadzenia waloryzacji przestrzeni agroturystycznej wykorzystano zbudowany agregatowy miernik waloryzacji agroturystycznej. Na podstawie wartości miernika wyodrębniono i opisano 7 klas powiatów, o różnej jakości przestrzeni agroturystycznej

Roma Ryś-Jurek, Olga Stefko, Zastosowanie modeli SARIMA do prognozowania cen produktów ogrodnictwa

Celem opracowania było wykazanie przydatności zastosowania modeli SARIMA do prognozowania cen w ogrodnictwie, na przykładzie dwóch produktów, a mianowicie jabłek i kapusty. Posłużono się danymi zbieranymi za pomocą wywiadu bezpośredniego przez serwis informacyjny „Fresh-market.pl”, a dotyczącymi minimalnych dziennych cen sprzedaży jabłek i kapusty w zł/kg, które zebrano między 11 maja 2007, a 24 maja 2011 roku z 10 rynków hurtowych w całej Polsce. Oszacowane modele SARIMA przeszły pozytywnie weryfikację, więc wykorzystano je do celów prognostycznych, sporządzając przykładowe prognozy minimalnych cen sprzedaży jabłek i kapusty na kolejne 30 dni.

Paweł Ulman, Model rozkładu wydatków a funkcje popytu

W pracy podjęto problem wpływu zmian poziomu dochodów na kształtowanie się rozkładu wydatków na dany typ dóbr i usług, a nie jedynie na przeciętny ich poziom. Wykorzystując jako model rozkładu wydatków teoretyczny rozkład Burra typu III (Daguma) oraz uzmienniając jego parametry poprzez odpowiednie funkcje dochodów uzyskano możliwość badania wpływu zmian dochodu na kształtowanie się rozkładu wydatków. Jako funkcje dochodów przyjęto funkcję liniową, potęgową oraz funkcję Törnquista. Odpowiednie modele wydatków oszacowano na podstawie danych z budżetów gospodarstw domowych z 2009 roku przy wykorzystaniu MNW.

Maria Urbańska, Tadeusz Mizera, Henryk Gierszal, Zastosowanie metod analizy statystycznej w badaniach przyrody żywej

W referacie przedstawiono możliwości wykorzystania różnych metod analizy statystycznej do wnioskowania o przyrodzie żywej na przykładzie badań nad dwoma typami królestwa zwierząt obejmujących bezkręgowce (mięczaków) oraz kręgowców (strunowców). W pierwszym przypadku badano gromadę ślimaków (Gastropoda), zaś w drugim gatunek orła bielika *Haliaeetus albicilla*. Do porównania bogactwa malakofauny na różnych terenach wykorzystano wskaźniki podobieństw, które posłużyły do wykreślenia dendrogramów. Z kolei w badaniach nad bielikiem wykorzystano wieloczynnikową analizę wariancji MANOVA oraz regresję wielokrotną w celu określania płci ptaków na podstawie pomiarów biometrycznych.

W pierwszym dniu konferencji odbyło się posiedzenie członków Sekcji Klasyfikacji i Analizy Danych Polskiego Towarzystwa Statystycznego, któremu, po wyborze, przewodniczył prof. dr hab. Feliks Wysocki. Ustalono plan przebiegu zebrania obejmujący punkty:

1. Sprawozdanie z działalności Sekcji Klasyfikacji i Analizy Danych PTS.
2. Zapowiedzi kolejnych konferencji SKAD PTS.
3. Wybór przedstawiciela Sekcji SKAD PTS do IFCS Council.
4. Sprawy różne.

Prof. dr hab. Marek Walesiak otworzył posiedzenie Sekcji SKAD PTS. Przypomniał, że na stronie internetowej Sekcji znajduje się nowy regulamin, a także deklaracja członkowska. Poinformował, że został opublikowany zeszyt z serii „Taksonomia 18”, PN UE we Wrocławiu nr 176, a także podkreślił, aby w przygotowaniu artykułów do kolejnego zeszytu uwzględnić wszystkie wymogi edytorskie zawarte w „Instrukcji dla autorów”. W „Przeglądzie Statystycznym” z. 1/2 z 2011 r. ukaże się sprawozdanie z ubiegłorocznej konferencji SKAD, która odbyła się w Toruniu w dniach 15-17 września.

W następnym punkcie posiedzenia podjęto kwestię kolejnych konferencji SKAD. Mgr inż. Ewa Chodakowska, w imieniu prof. dra hab. Joanicjusza Nazarko, potwierdziła organizację przyszłorocznej konferencji SKAD. Poinformowała, że odbędzie się ona w dniach 12-14 września 2012 r. w miejscowości Borki w hotelu Lipowy Most. Prof. dr hab. Krzysztof Jajuga zasugerował pozostawienie na niezmiennym poziomie opłat konferencyjnych. W związku z tym, że nie padły deklaracje w sprawie organizacji konferencji w 2013 r., prof. dr hab. M. Walesiak zaproponował, że w przypadku braku innych propozycji, podejmie się organizacji konferencji.

W dalszej części zebrania prof. dr hab. Krzysztof Jajuga podsumował, że International Federation of Classification Societies (IFCS) skupia 13 towarzystw klasyfikacyjnych, a SKAD jest trzecim pod względem liczebności. Krótco wyjaśnił organizację i strukturę oraz sposób powoływania prezydenta IFCS. W związku z tym, że w 2011 r. kończy się kadencja prof. UEK dr hab. Andrzeja Sokołowskiego jako członka Rady i jednocześnie przedstawiciela SKAD w IFCS, należało przeprowadzić głosowanie w sprawie wyboru reprezentanta na kolejną kadencję.

Prof. dr hab. Krzysztof Jajuga zaproponował powołanie do Komisji Skrutacyjnej prof. dra hab. Tadeusza Kufła oraz prof. UE dr hab. Elżbietę Gołątę oraz zaproponował

kandydaturę prof. Andrzeja Sokołowskiego. W związku z brakiem innych kandydatur przystąpiono do głosowania. Komisja podliczyła głosy: 45 głosów za, 0 głosów przeciw, 0 osób się wstrzymało. W wyniku głosowania przyjęto kandydaturę prof. Andrzeja Sokołowskiego jako reprezentanta Sekcji SKAD PTS w IFCS Council na kolejną kadencję.

W ostatniej części zebrania prof. Marek Walesiak przypomniał, że w dniach 30 sierpnia – 2 września 2011 r. we Frankfurcie odbyła się konferencja IFCS oraz GfKI. Prof. Andrzej Sokołowski podsumował, że podczas konferencji wygłoszono 175 referatów, z tego znaczna część dotyczyła grupowania szeregów czasowych. Prof. Marek Walesiak podkreślił, że w konferencji udział wzięły osoby z następujących uczelni: UE Wrocław (4 osoby); UWM w Olsztynie (4 osoby); UE Kraków (3 osoby); UE Katowice (2 osoby); Politechnika Opolska (1 osoba); IBS PAN (1 osoba). Poinformował również, że w dniach 21-26 sierpnia w Dublinie odbył się 58th World Statistics Congress of the International Statistical Institute (ISI), w którym udział wziął prof. dr hab. Eugeniusz Gatnar.

Prof. Krzysztof Jajuga zapowiedział, że kolejna konferencja IFCS odbędzie się w Porto w lipcu 2013 r., a w 2015 r., prawdopodobnie w Kolumbii. Zachęcał do udziału w przyszłorocznej konferencji GfKI, która odbędzie się w Hildesheim. Prof. Andrzej Sokołowski podkreślił wagę publikowania artykułów w czasopismach z Listy Filadelfijskiej, w tym w *Journal of Classification*.