

XXXVIII Międzynarodowa Konferencja Naukowa Wielowymiarowa Analiza Statystyczna WAS 2019

The 38th International Scientific Conference on Multivariate Statistical Analysis MSA 2019

W dniach 4–6 listopada 2019 r. odbyła się w Centrum Szkoleniowo-Konferencyjnym Uniwersytetu Łódzkiego (UŁ) XXXVIII Międzynarodowa Konferencja Naukowa Wielowymiarowa Analiza Statystyczna WAS 2019 (The 38th International Scientific Conference on Multivariate Statistical Analysis MSA 2019). Zorganizowały ją: Katedra Metod Statystycznych UŁ, Instytut Statystyki i Demografii UŁ, Łódzki oddział Polskiego Towarzystwa Statystycznego oraz Komitet Statystyki i Ekonometrii Polskiej Akademii Nauk (PAN). Komitetowi Naukowemu przewodniczył prof. dr hab. Czesław Domański; funkcję sekretarzy naukowych konferencji pełniły dr hab. Aleksandra Baszczyńska i dr hab. Katarzyna Bolonek-Lasoń. Finansowego wsparcia przy organizacji wydarzenia udzieliły Ministerstwo Nauki i Szkolnictwa Wyższego (MNiSW)¹, PAN oraz StatSoft Polska.

Głównymi celami konferencji były prezentacja najnowszych osiągnięć z zakresu wielowymiarowej analizy statystycznej oraz wymiana doświadczeń będących wynikiem jej stosowania. Poruszone zostały następujące zagadnienia: rozkłady wielowymiarowe, testy statystyczne, metody nieparametryczne, analiza czynnikowa, analiza skupień, analiza dyskryminacyjna, analiza wariancji i regresji, metody bayesowskie, analizy Monte Carlo, data mining, procedury odporne, analiza danych cenzurowanych, rozpoznawanie obrazów, analizy stochastyczne oraz zastosowanie metod statystycznych w naukach ekonomicznych, marketingu, finansach, ubezpieczeniach, rynku kapitałowym i zarządzaniu ryzykiem.

W konferencji wzięły udział 72 osoby z ośrodków akademickich w Gdańsku, Katowicach, Krakowie, Łodzi, Poznaniu, Rzeszowie, Szczecinie, Warszawie i we Wrocławiu, przedstawiciele Głównego Urzędu Statystycznego (GUS) i urzędów statystycznych w Łodzi, Poznaniu i Rzeszowie, a także goście z Włoch i Niemiec. Łącznie odbyło się 15 sesji plenarnych i panelowych, na których wygłoszono 42 referaty. Konferencję zainaugurowały wystąpienia Czesława Domańskiego, rektora UŁ prof. dr hab. Antoniego Różalskiego oraz prodziekana Wydziału Ekonomiczno-Socjologicznego UŁ dr. hab. Michała Przybylińskiego, prof. UŁ.

¹ Organizacja konferencji Multivariate Statistical Analysis 2019 została sfinansowana w ramach umowy 712/P-DUN/202019 ze środków MNiSW przeznaczonych na działalność upowszechniającą naukę.

W pierwszej sesji plenarnej, której przewodniczył Czesław Domański, przedstawiono referaty dotyczące aktualnych problemów statystyki wielowymiarowej. Prof. dr hab. inż. Jacek Wesołowski (Politechnika Warszawska, GUS) wykazał w wystąpieniu *Optymalna alokacja próbki w schematach warstwowych – metody algebry liniowej i algorytmy*, że powszechnie stosowany rekurencyjny algorytm Neymana zapewnia optymalność alokacji w schemacie warstwowym z ograniczeniami na liczebność próbki w warstwach. Następnie prof. dr hab. Mirosław Krzyśko (Uniwersytet im. Adama Mickiewicza w Poznaniu) przedstawił w referacie *Jądrowe współrzędne dyskryminacyjne w przypadku geograficznie ważonych danych czasowo-przestrzennych z wyborem zmiennych* rozszerzenie analizy jądrowych współrzędnych dyskryminacyjnych na dane czasowo-przestrzenne oraz ważone dane czasowo-przestrzenne wraz z propozycją procedury wyboru zmiennych.

Drugą sesję, pod przewodnictwem Mirosława Krzyśki, poświęcono historii polskiej statystyki. Czesław Domański przybliżył sylwetkę Jakuba Kazimierza Haura, a dr hab. Jerzy T. Kowaleski, prof. UŁ – Marcina Kromera.

Trzecia sesja plenarna miała charakter wspomnieniowy. Przywołano postaci wybitnych polskich statystyków, którzy odeszli w 2017 i 2018 r. Mirosław Krzyśko wspominał Krystynę Katulską, dr Ewa Wycinka (Uniwersytet Gdański, UG) – Mirosława Krzysztofiaka, prof. dr hab. Janusz Wywiół (Uniwersytet Ekonomiczny w Katowicach, UE w Katowicach) – Józefa Kolonkę, zaś dr hab. Stanisław Wanat, prof. UEK (Uniwersytet Ekonomiczny w Krakowie) – Stanisława Wydymusa i Michała Majora.

Następnie rozpoczęły się sesje panelowe. Sesji IV A, prowadzonej w języku angielskim, przewodniczył prof. dr hab. Marek Walesiak (Uniwersytet Ekonomiczny we Wrocławiu, UEW). Dr hab. inż. Jacek Białek, prof. UŁ (GUS), omówił w referacie *Chain drift problem in the CPI measurement based on scanner data* wyniki badań symulacyjnych, które wskazują na sytuacje na rynku generujące największe obciążenie. Występuje ono wtedy, gdy indeks cen różni się od jedności w momencie powrotu cen produktów do wartości bazowej. Dr hab. Tomasz Żądło, prof. UE w Katowicach, w referacie *On generalization of Quatember's bootstrap* zaprezentował uogólnienie algorytmu Quatembera, pozwalającego na bezpośrednie ponowne próbkowanie z oryginalnej próby, oraz przedstawił wyniki badania jego własności wraz z porównaniem z aktualnymi konkurencyjnymi metodami. Referat *Asymptotic properties of duration-based VaR backtests* wygłoszony przez dr Martę Małecką (UŁ) dotyczył własności granicznych testu geometrycznego i geometrycznego testu VaR, zbudowanych na zasadzie ilorazu wiarygodności, jak również własności testu Giniego. Autorka wyprowadziła odpowiednie rozkłady asymptotyczne i określiła użyteczność otrzymanych wyników dla prób o realistycznych rozmiarach.

Podczas sesji IV B, której przewodniczyła dr hab. Iwona Markowicz, prof. US (Uniwersytet Szczeciński), wygłoszono cztery referaty. W prezentacji *Alokacja próby w problemie estymacji wskaźnika struktury w populacjach skończonych* Dominik Siemradzki i prof. dr hab. Wojciech Zieliński (Szkoła Główna Gospodarstwa Wiejskiego w Warszawie, SGGW) przeprowadzili porównanie dokładności szacowania w zależności od wyboru alokacji próby. Przedstawili także wyniki badań, w których zestawili swoją propozycję ze znaną alokacją Neymana oraz alokacją proporcjonalną. Ewa Wycinka i dr hab. Beata Jackowska, prof. UG, w wystąpieniu *Estymacja rozkładu czasu istnienia przedsiębiorstw z uwzględnieniem rodzaju zdarzenia kończącego działalność* omówiły rezultaty badania zastosowania estymatorów uwzględniających rodzaj zdarzenia kończącego działalność w modelowaniu rozkładów czasu istnienia przedsiębiorstw. W badaniu porównano własności estymatorów dla zdarzeń konkurujących: naiwnego estymatora Kaplana-Meiera, estymatora Aalena-Johansena oraz estymatora uwzględniającego informacyjność cenzurowania IPCW (Inverse Probability Censoring Weighted). Następnie dr Stanisław Jaworski (SGGW) wygłosił *Kilka uwag o estymacji poziomu bezrobocia w Polsce*, opisując estymację poziomu bezrobocia za pomocą modelu strukturalnych szeregów czasowych. Ostatnim referatem w tej sesji było wystąpienie Janusza L. Wywiśla i Grzegorza Sitka (UE w Katowicach) *Wariancja wartości własnych macierzy*. Autorzy rozważali problem dokładności estymacji wariancji maksymalnych wartości własnych macierzy kowariancji, omówili konstrukcję tego estymatora i zaproponowali metodę aproksymacji wariancji estymatora korelacji kanonicznej.

Sesję V A (w języku angielskim), pod przewodnictwem dr. hab. Wojciecha Gamrota, prof. UE w Katowicach, otworzył referat *On EBLUP under some linear mixed model with correlated random effects*, którego autorka Małgorzata Krzciuk (UE w Katowicach) przedstawiła wnioski z badań symulacyjnych wpływu występowania zależności między efektami losowymi na własności najlepszego liniowego nieobciążonego predyktora empirycznego dla liniowego modelu mieszanego z dwoma skorelowanymi efektami losowymi. Następnie dr Maciej Beręsewicz (Uniwersytet Ekonomiczny w Poznaniu, UEP, Urząd Statystyczny w Poznaniu) wygłosił przygotowany z Katarzyną Zadrogą (UEP) referat *Estimation of the number of illegally residing foreigners in Poland in 2017 and 2018 using Bayesian non-linear mixed count regression models*. Jego przedmiotem było oszacowanie liczby cudzoziemców nielegalnie przebywających w Polsce w latach 2017–2018, w którym zastosowano bayesowski nieliniowy model mieszany dla danych dyskretnych, wykorzystujący wyłącznie dane zagregowane raportowane przez straż graniczną i policję oraz rejestr PESEL.

Prowadzonej równolegle sesji V B przewodniczył Wojciech Zieliński. W sesji tej dr Wioletta Grzenda (Szkoła Główna Handlowa w Warszawie, SGH) wygłosiła re-

ferat *Bayesowskie wielomianowe modele logitowe dla kategorii nieuporządkowanych w badaniu sytuacji osób młodych na rynku pracy w Polsce*, w którym przedstawiła wyniki badań sytuacji respondentów na rynku pracy z zastosowaniem dwumianowego modelu logitowego. Prelegentka wykazała, że kontynuacja nauki często wynika z problemów ze znalezieniem pracy, a łączenie pracy zawodowej z nauką nie należy do preferowanych form aktywności zawodowej osób młodych. Referat *Triady czy tetrazy? Porównanie niepełnych metod pomiaru podobieństwa preferencji*, przedstawiony przez dr. hab. Artura Zaborskiego (UEW) dotyczył zestawienia dwóch niepełnych metod pomiaru podobieństwa preferencji, tj. metody triad oraz metody tetrad. Zostały one porównane ze względu na ich pracochłonność oraz zdolność odwzorowania znanej struktury obiektów, także w przypadku kiedy nie jest spełniony warunek co do równej liczby par obiektów w podgrupach prezentowanych respondentom. Prezentacja *Uogólniona ekonometria entropii krzyżowej a sprzeczne transgraniczne (duże) źródła danych. Aktualizacja rachunków narodowych* autorstwa dr. Seconda Bwanakare (Wyższa Szkoła Informatyki i Zarządzania w Rzeszowie) oraz dr. Marka Cierpiął-Wolana (Uniwersytet Rzeszowski, Urząd Statystyczny w Rzeszowie) dotyczyła skutecznej metody łączenia danych z różnych źródeł oraz porównania wyników z techniką stosowaną dotychczas w oficjalnych statystykach. Autorzy podzielili się wynikami badań, w których zastosowali związaną z prawem potęgowym metodę dywergencji informacji Kullbacka-Leiblera – znaną z uogólnienia entropii Shannona – do rozwiązania nieliniowych, źle uwarunkowanych problemów odwrotnych za pomocą teorii bayesowskiej.

Drugiego dnia konferencji odbyły się dwie sesje plenarne i dwie równoległe sesje panelowe. Podczas pierwszej sesji plenarnej (w języku angielskim), poprowadzonej przez dr. hab. Alinę Jędrzejczak, prof. UŁ, wygłoszono wykłady gościnne. Dr. hab. Francesca Greselin (University of Milan-Bicocca) przedstawiła prezentację *Advances in learning from contaminated datasets*, przygotowaną wspólnie z Andreą Cappozzo (University of Milan-Bicocca) i prof. dr. hab. Thomasem Brendanem Murphym (University College Dublin). Dotyczyła ona modyfikacji procedury analizy dyskryminacyjnej, stosowanej w przypadku gdy dane rzeczywiste niektórych jednostek w zestawie uczenia się mogą być zawodne (zjawisko definiowane jako szum etykiety) lub część obserwacji może odbiegać od głównej struktury danych (co jest określane jako wartości odstające). Wykład prof. dr. hab. Hansa-Joachima Mittaga (University of Hagen) *A new virtual library containing interactive learning objects for statistics education* zawierał opis projektu służącego opracowaniu interaktywnych obiektów edukacyjnych do celów edukacji statystycznej. Autor zapoznał publiczność z wynikami projektu, opisał dwie biblioteki wirtualne oraz przedstawił dotychczasowe doświadczenia z różnymi scenariuszami uczenia się mieszanego, a także perspektywy

współpracy międzynarodowej. Prof. dr hab. Józef Dziechciarz i dr Marta Dziechciarz-Duda (UEW) w wystąpieniu *Selected aspects of households' well-being measurement* dokonali dogłębnej analizy problemów metodologicznych pomiaru dobrostanu gospodarstw domowych i przedstawili propozycje dotyczące tego zagadnienia.

Drugą sesję, której przewodniczył prof. dr hab. Grzegorz Kończak (UE w Katowicach), otworzyło wystąpienie Iwony Markowicz *Rozbieżności w handlu wewnątrz-wspólnotowym: przypadek Polski*, opracowane wspólnie z dr. Pawłem Baranem (US). Prelegentka przedstawiła wyniki analizy rozbieżności danych w handlu Polski w relacjach: Polska – kraj UE (relacje dwustronne) i Polska – kraje UE (relacja zregulowana kraj – kraje). Dr hab. Grażyna Dehnel, prof. UEP, i Marek Walesiak (UEW) w wystąpieniu *Ocena spójności społecznej województw Polski na podstawie danych klasycznych oraz symbolicznych interwałowych* przedstawili ocenę spójności społecznej województw w 2018 r., opracowaną z wykorzystaniem podejścia hybrydowego, łączącego zastosowanie skalowania wielowymiarowego z porządkowaniem liniowym. Dr hab. Beata Bieszk-Stolorz, prof. US, wygłosiła referat *Wybrane modele zdarzeń wielokrotnych w ocenie ryzyka powtórnej rejestracji w urzędzie pracy*, który zawierał wyniki analizy ryzyka kolejnej rejestracji w urzędzie pracy w zależności od wybranych cech osób bezrobotnych: płci, wieku, wykształcenia oraz stażu pracy. W badaniu porównano wyniki otrzymane dla modelu danych zliczanych Andersona-Gila, modeli warunkowych Prentince'a-Williamsa-Petersona: czasu całkowitego i luki czasowej oraz modelu Wei, Lin i Weissfelda. Sesję zamknął referat Wojciecha Gamrota (UE w Katowicach) *Skala Likerta i współczynnik regresji*. Autor omówił skutki postępowania polegającego na stosowaniu metod wnioskowania statystycznego przeznaczonych dla zmiennych ciągłych, a w szczególności metod opartych na założeniu wielowymiarowej normalności w przypadku modelu regresji liniowej, gdy skala Likerta stosowana jest w statystycznych badaniach kwestionariuszowych.

Sesji III A przewodniczył Tomasz Żądło; w jej trakcie wygłoszono trzy referaty. Dr Piotr Sulewski (Akademia Pomorska w Słupsku) w wystąpieniu *Rozpoznawanie rozkładów zamiast testowania zgodności* przedstawił koncepcję rozpoznawania rozkładów, zgodnie z którą stosowana jest reguła k -najbliższych sąsiadów – jako metoda konkurencyjna wobec przeprowadzania klasycznych testów zgodności wykorzystujących miarę rozbieżności. Referat Dominiki Polko-Zajac (UE w Katowicach) *O permutacyjnych testach porównywania populacji wielowymiarowych* dotyczył permutacyjnej, jednoczesnej procedury identyfikacji różnic występujących między wektorami wartości przeciętnych oraz macierzami wariancji – kowariancji w dwóch badanych populacjach. Autorka przedstawiła również wyniki badania symulacyjnego w celu określenia rozmiaru i mocy tych testów. Następnie Krzysztof Szymoniak-Książek (UE w Katowicach) w prezentacji *Własności nieparametrycznych testów izotropii* omówił wyniki

badania symulacyjnego polegającego na wygenerowaniu ciągów realizacji pola losowego o zadanym rozkładzie teoretycznym, dla których testowano hipotezę zerową stanowiącą o izotropii. Autor wyznaczył empiryczne prawdopodobieństwa odrzuceń i porównał je z zakładanym z góry poziomem istotności.

Sesję III B, której przewodniczyła Grażyna Dehnel, otworzył referat dr Katarzyny Budny (UEK) *Wielowymiarowa nierówność Czebyszewa – oszacowania dla prawdopodobieństwa przyjmowania przez wektor losowy wartości z kuli euklidesowej*. Autorka zaprezentowała wybrane wielowymiarowe uogólnienia klasycznej nierówności Czebyszewa oraz oszacowania dla prawdopodobieństwa przyjmowania przez wektor losowy wartości z kuli euklidesowej, wyrażone za pomocą momentów wektora losowego opartych na definicji potęgi wektora. Wystąpienie dr. hab. Michała Bernardelliego (SGH) *Identyfikacja punktów zwrotnych szeregów czasowych z rynku kryptowalut* dotyczyło wykorzystania ścieżek Viterbiego do analizy jednowymiarowych szeregów cen z rynku kryptowalut za pomocą ukrytych modeli Markowa. Adam Juszcak (UŁ, Narodowy Bank Polski) w referacie *Zastosowanie danych scrapowanych w pomiarze inflacji* rozpatrzył pozytywne i negatywne aspekty wykorzystania web scrapingu przy obliczaniu wskaźnika dóbr i usług konsumpcyjnych (CPI).

Sesji IV A, prowadzonej w języku angielskim, przewodniczyła prof. dr hab. Grażyna Trzpiot (UE w Katowicach). Autorka pierwszego referatu dr hab. Aneta Ptak-Chmielewska (SGH) w prezentacji *Application of multidimensional classification to prediction of SME* dokonała porównania efektywności analizy dyskryminacyjnej z dyskryminacją wielowymiarową, taką jak metoda wektorów wspierających (*support vector machines*), z wykorzystaniem próby przedsiębiorstw MŚP. Dr hab. Jerzy Korzeniewski, prof. UŁ, w wystąpieniu *Determining semantic relatedness of concepts – modifications proposals* przedstawił metody badania podobieństwa semantycznego pojęć wraz z własną propozycją modyfikacji metody Leacocka i Chodorowa stosowanej dla pojęć, których wzajemne powiązania są bardzo słabe. Następnie dr Elżbieta Roszko-Wójtowicz (UŁ) i dr hab. Maria M. Grzelak, prof. UŁ, wygłosiły referat *Innovation activities and competitiveness of manufacturing divisions in Poland in the years 2009–2017*. Autorki przeprowadziły analizę pomiaru i oceny wpływu działalności innowacyjnej na konkurencyjność działów przetwórstwa przemysłowego, wykorzystując w tym celu modele panelowe, zarówno statystyczne z opóźnieniami, jak i dynamiczne, co pozwoliło na wnikliwe prześledzenie zmian w czasie w odniesieniu do rozpatrywanych zmiennych. Ostatnim głosem w sesji było wystąpienie dr. Łukasza Wawrowskiego (UEP) *Impact of dependent variable transformation on poverty rate estimates in poviats* dotyczące badania ubóstwa z wykorzystaniem danych o niskim stopniu agregacji przestrzennej. Prelegent zapoznał uczestników z wynikami estymacji stopy ubóstwa na poziomie powiatów z wykorzystaniem

danych z Europejskiego Badania Dochodów i Warunków Życia i Narodowego Spisu Powszechnego Ludności i Mieszkań oraz metod estymacji pośredniej bazujących na liniowych modelach mieszanych i symulacjach Monte Carlo.

W sesji IV B, której przewodniczył dr hab. Andrzej Dudek, prof. UEW, wygłoszono cztery referaty. Dr Agnieszka Stanimir (UEW) w wystąpieniu *Metody wielowymiarowej analizy statystycznej w analizie pytań wielokrotnego wyboru* przedstawiła wyniki zastosowania metod wielowymiarowej analizy statystycznej w sytuacjach, gdy wybór kategorii nie ma charakteru następstw oraz gdy z punktu widzenia badacza istotne jest poznanie sekwencji dokonywanych wyborów kategorii w ramach jednej zmiennej. Referat Łukasza Ziarki (UŁ) *O możliwości zastosowania analizy asocjacji do określenia wzorca zachowania się wykonawców w przetargach publicznych* dotyczył możliwości zastosowania analizy asocjacji (analizy koszykowej) do identyfikacji nielegalnych porozumień zawieranych pomiędzy wykonawcami ubiegającymi się o udzielenie zamówienia publicznego. Prelegent pokusił się również o próbę oceny zaproponowanego podejścia. Następnie dr Anna Denkowska i Stanisław Wanat (UEK) w prezentacji *Wzajemne powiązania a ryzyko systemowe w europejskim sektorze ubezpieczeniowym. Nowe wyniki bazujące na wykorzystaniu dynamicznych minimalnych drzew rozpinających* przedstawili wyniki badań wzajemnych powiązań europejskich instytucji ubezpieczeniowych oraz ich wkład w ryzyko systemowe związane z wykorzystaniem sieci korelacyjnych. Autorki ostatniego referatu *Zastosowanie modelu Zengi do opisu rozkładu dochodów ludności Polski dla grup społeczno-ekonomicznych* Alina Jędrzejczak i dr Kamila Trzcińska (UŁ) wykazały, że model Zengi bardzo dobrze opisuje rozkład dochodów ludności Polski dla grup społeczno-ekonomicznych.

Trzeciego dnia konferencji odbyły się dwie sesje plenarne. Pierwszej przewodniczył prof. dr hab. Andrzej Bąk (UEW). Dwa referaty wygłosili dr hab. Małgorzata Graczyk i prof. dr hab. Bronisław Ceranka (Uniwersytet Przyrodniczy w Poznaniu). Wystąpienie *Uwagi o wysoce D-efektywnych sprzężynowych układach wagowych* autorzy poświęcili związanym z nowymi metodami konstrukcji sprzężynowym układom wagowym o wysokiej D-efektywności w klasach, w których nie istnieje układ D-optymalny. W drugim wystąpieniu *Nowe wyniki dotyczące metod konstrukcji D-optymalnych chemicznych układów wagowych* omówili doświadczenie, w którym wyznaczone są nieznanne miary p obiektów przy użyciu n operacji pomiarowych, zgodnie z modelem chemicznego układu wagowego, oraz podali warunki optymalności i serie parametrów układów D-optymalnych. Grażyna Trzpiot poświęciła swój referat *Seniorzy w miastach a miasta przyjazne seniorom – analiza dla wybranych miast Polski* zmianom i tempu procesu starzenia się wybranych miast w Polsce z wykorzystaniem modelu opisowego oraz odpornego podejścia taksonomicznego.

Grzegorz Kończak w wystąpieniu *Wielowymiarowe permutacyjne rozszerzenie testu McNemara* zreferował propozycję k -wymiarowego ($k > 2$) testu dla różnic pomiędzy prawdopodobieństwami sparowanych wektorów losowych o rozkładach dwupunktowych; propozycja ta stanowi wielowymiarowe rozszerzenie testu McNemara.

W finalnej sesji, pod przewodnictwem Czesława Domańskiego, Andrzej Bąk wystąpił z referatem *Metody imputacji brakujących danych z wykorzystaniem programu R na przykładzie Banku Danych Lokalnych*. Przedstawiał w nim wyniki badań związanych z uzupełnieniem brakujących danych, wykorzystujących metody zaproponowane w literaturze przedmiotu, jak również zestawy programu R. Wystąpienie Czesława Domańskiego *Uwagi o testach normalności opartych na charakterystykach procesów stochastycznych* zawierało rozważania o testach normalności, w szczególności tych, w których wykorzystywane są różnorodne charakterystyki procesów stochastycznych.

Czesław Domański, po zakończeniu i podsumowaniu wydarzenia, poinformował że kolejna konferencja Wielowymiarowa Analiza Statystyczna WAS odbędzie się w dniach 16–18 listopada 2020 r.

Aleksandra Baszczyńska (Uniwersytet Łódzki)