

Konferencja naukowa Metodologia Badań Statystycznych MET2019

Między 3 a 5 lipca br. odbyła się w Warszawie konferencja naukowa Metodologia Badań Statystycznych MET2019, zorganizowana przez Główny Urząd Statystyczny (GUS) we współpracy z Polskim Towarzystwem Statystycznym (PTS). Konferencja zgromadziła blisko 250 osób z kilku krajów.

W roli keynote speakerów wystąpiło dwóch wybitnych statystyków – prof. Graham Kalton z USA (Westat) i prof. Danny Pfeffermann z Izraela. Kalton zdobył międzynarodowe uznanie dzięki dorobkowi na polu metodologii badań sondażowych. Jest jednym z twórców Joint Program in Survey Methodology prowadzonego na University of Maryland, członkiem m.in. American Statistical Association, American Association for the Advancement of Science, International Statistical Institute oraz National Associate of the National Academy of Sciences (NAS). Pfeffermann specjalizuje się m.in. we wnioskowaniu analitycznym ze złożonych badań reprezentacyjnych, korektach sezonowych i szacowaniu trendów oraz w statystyce małych obszarów. Jest profesorem statystyki na Uniwersytecie w Southampton w Wielkiej Brytanii oraz emerytowanym profesorem Uniwersytetu Hebrajskiego w Jerozolimie, a także członkiem American Statistical Association i International Statistical Institute. Stoi na czele izraelskiego Centralnego Biura Statystycznego (Central Bureau of Statistics, CBS).

Program konferencji ukierunkowany był na problemy metodologiczne badań statystycznych w kontekście zmieniających się warunków funkcjonowania statystyki oficjalnej. Wystąpienia dotyczyły zarówno innowacyjnych przedsięwzięć statystyki publicznej, jak i sposobów komunikacji i udostępniania wyników badań, cechujących się wysoką jakością i wiarygodnością. Wydarzenie ma zainicjować cykl spotkań ekspertów statystyki publicznej z przedstawicielami środowisk naukowych, praktykami zarządzania i biznesu oraz użytkownikami danych statystycznych.

Tematyka konferencji obejmowała zagadnienia takie, jak:

- statystyka matematyczna, metoda reprezentacyjna i statystyka małych obszarów, statystyka ludności, społeczna, gospodarcza oraz statystyka regionalna;
- analiza i klasyfikacja danych, big data i dane statystyczne, statystyka polska na arenie międzynarodowej oraz miary i wskaźniki w statystyce publicznej;
- historia polskiej statystyki, metody statystyczne w badaniach historycznych, komunikacja statystyki i edukacja statystyczna.

Konferencję otworzył prezes GUS dr Dominik Rozkrut, który ogłosił Manifest 5 „O” na rzecz otwartości i transparentności w statystyce publicznej. Zaprezentował także nowy Portal Naukowy GUS, na którym zamieszczane są czasopisma i monografie naukowe wydawane przez Urząd. Powstanie również portal dotyczący programu European Master in Official Statistics (EMOS) jako платфор-

ma komunikacji, promocji i wymiany informacji o wydarzeniach związanych z EMOS. Natomiast na stronie dekompozycje.stat.gov.pl będzie można projektować i monitorować strategie rozwoju regionalnego, a także analizować wzrost i różnice w poziomie rozwoju gospodarczego województw i makroregionów.

W pierwszym dniu konferencji odbyło się dziewięć sesji, podczas których wygłoszono 31 referatów. Wystąpienie prof. dr hab. Grażyny Trzpiot i prof. dr hab. inż. Jacka Szołtyśka dotyczyło miast przyjaznych seniorom. Autorzy przyjęli za punkt wyjścia definicję WHO, wskazującą cztery główne kryteria, według których można ocenić, czy dane miasto jest przyjazne osobom starszym, są to: zasady, usługi, otoczenie i struktury, które wspierają proces aktywnego starzenia się. W ramach badania stworzono model opisowy, a następnie dokonano oceny wybranych miast Polski, wykorzystując w tym celu podejście taksonomiczne.

Prof. dr hab. Czesław Domański i dr hab. Alina Jędrzejczak poruszyli temat problemów etycznych w procesie generowania danych w internecie. Zwrócili uwagę na towarzyszące badaniom statystycznym przekonanie, że dostęp do wiarygodnych informacji niesie społeczne korzyści. Statystycy powinni brać pod uwagę prawdopodobieństwo konsekwencji zbierania i upowszechniania danych, eliminując możliwe do przewidzenia przypadki ich nieprawidłowej interpretacji lub niewłaściwego wykorzystania.

Zagadnienie cząstkowych indeksów polaryzacyjnych omówili dr Jan Zwierchowski i prof. dr hab. Tomasz Panek. W okresie objętym przeprowadzonym przez nich badaniem gospodarstwa domowe znajdujące się w środkowej klasie dochodowej przesunęły się w kierunku niższej lub wyższej klasy dochodowej, natomiast gospodarstwa domowe znajdujące się w klasach niższej i wyższej przemieszczały się w kierunku mediany rozkładu dochodów. Co więcej, konwergencja obserwowana w niższych i wyższych klasach dochodowych była silniejsza niż polaryzacja w środkowej klasie dochodowej. Przepływy gospodarstw domowych do środkowej klasy dochodowej były wyższe niż wypływy z tej klasy.

Referat *Wkład polskiej myśli statystycznej do statystyki światowej* wygłosili Czesław Domański i prof. dr hab. Włodzimierz Okrasa. Przedstawili dorobek wybranych statystyków z Polski, których wkład w rozwój statystyki wart jest uwidocznienia i popularyzacji.

Referat dr. Marka Cierpień-Wolana i Sebastiana Wójcika dotyczył budowy innowacyjnego systemu integracji danych w dziedzinie turystyki. Pojawienie się źródeł danych takich jak big data wymaga nowego podejścia do tworzenia wiarygodnych informacji statystycznych. Ma to związek z rozproszonym charakterem tych źródeł, a także z zamiarem wykorzystania ich na dużą skalę w oficjalnych statystykach. To podejście, często określane terminem *trusted smart statistics*, będzie w najbliższych latach stanowiło wyzwanie dla oficjalnej statystyki. Autorzy podkreślili, że innowacyjne metody zbierania i łączenia danych pozwalają na zwiększenie poziomu ich dezagregacji, poprawę jakości operatów oraz

opracowywanie szybkich szacunków w dziedzinie turystyki, co odegra kluczową rolę w rozwoju spójnego systemu informacji turystycznej.

W drugim dniu konferencji odbyło się 12 sesji ukierunkowanych na zagadnienia związane m.in. z tematyką statystyki gospodarczej, big data i danych statystycznych oraz statystyki ludności. Wygłoszono 47 referatów. Dr Katarzyna Brzozowska-Rup i dr Mariola Chrzanowska omówiły badania zjawiska szarej strefy. Zaprezentowały koncepcję oraz zastosowanie modeli strukturalnych zmiennych ukrytych (Structural Equation Models, SEM). Mogą one służyć do bezpośredniego pomiaru aktywności gospodarczych prowadzonych bez pełnej wiedzy państwa, co w konsekwencji skłania do wykorzystania modeli ze zmiennymi nieobserwowalnymi i niemierzalnymi. Prelegentki podkreśliły istnienie niewykorzystanego potencjału eksplikacyjnego modeli równań strukturalnych, w związku z czym należałoby podjąć próbę specyfikacji i rozszerzenia tych modeli w kontekście badań szarej strefy.

Wybrane problemy reprezentacyjnych badań powtarzalnych przedstawiły dr hab. Barbara Kowalczyk i dr Dorota Juszcak. Dokonały one przeglądu metod estymacji na podstawie danych otrzymanych z badań rotacyjnych. Następnie zaprezentowały wyniki teoretyczne oraz omówiły wyniki badań symulacyjnych.

Zagadnienia dotyczące wykorzystania bootstrapowanych modeli ekonometrycznych dla danych indywidualnych w prognozowaniu dotyczącym nowych lokalizacji w zbiorach big data poruszyła dr hab. Katarzyna Kopczewska. Typowe przestrzenne modele ekonometryczne, takie jak SDM, SDEM czy SAC, szacowane dla danych punktowych geo-lokalizowanych cechują się ograniczoną możliwością prognozowania w przypadku nowych punktów w przestrzeni spoza próby. Przedstawiona nowatorska metodologia kalibruje zarówno przestrzeń, jak i model przy użyciu bootstrapu i teselacji. Bootstrap umożliwia kalibrację modelu ekonometrycznego bez potrzeby szacowania na całym zestawie danych. To rozwiązanie przetestowano na podstawie lokalizacji firm z rejestru REGON.

Dr hab. Mirosława Kaczmarek podjęła temat wykluczenia pokolenia Baby Boomers (BB) z rynku e-commerce. Prelegentka skupiła się na określeniu pozycji Polski na tle krajów Unii Europejskiej (UE) pod względem korzystania z handlu i usług dostępnych online przez osoby z pokolenia BB; zwróciła uwagę na płeć i wykształcenie jako czynniki determinujące wykluczenie z rynku e-commerce. Polska znajduje się w grupie 13 krajów, które charakteryzują się niskimi wskaźnikami korzystania z e-handlu, a na wykluczenie z rynku e-commerce osób z generacji BB istotnie wpływa właśnie poziom ich wykształcenia.

Trzeciego dnia konferencji odbyło się sześć sesji dotyczących m.in. statystyki regionalnej, statystyki gospodarczej i statystyki matematycznej, podczas których wygłoszono 26 referatów. Wystąpienie dr Katarzyny Cheby i dr hab. Iwony Bąk *Europejski Ranking Innowacyjności i jego regionalny wymiar – analiza taksonomiczna* dotyczyło problemu znacznego wewnętrznego zróżnicowania poziomu

innowacyjności krajów i regionów UE. Zauważono, że uśrednianie wyników uzyskiwanych w ramach poszczególnych badanych obszarów i przedstawianie ich w postaci jednego miernika w publikowanych przez Komisję Europejską raportach European Innovation Scoreboard i Regional Innovation Scoreboard może prowadzić do zbyt dużych uogólnień i opacznej klasyfikacji regionów do grup typologicznych.

Dominika Rogalińska i dr Marek Pieniążek wygłosili referat *Aplikacyjne aspekty kształtowania strumieni danych statystycznych w państwie na przykładzie badań przestrzennych*. Jak stwierdzili, obecny system przepływu danych charakteryzuje wysoki poziom złożoności, a dane gromadzone są według różnych metod i w różnych formach zapisu, a to ogranicza ich użyteczność w kompleksowym monitorowaniu. W związku z tym warto rozważyć wdrożenie rozwiązań stosowanych m.in. w krajach skandynawskich, w których statystyka publiczna pełni funkcję kluczowego węzła wymiany danych, dzięki czemu nakłady na realizację obowiązku sprawozdawczego zmniejszyły się, a jakość i spójność danych wzrosła. Przykładem działań podjętych w tym kierunku przez GUS są prace związane z budową Systemu Monitorowania Usług Publicznych (SMUP), który ma być pierwszym kompleksowym rozwiązaniem wspierającym doskonalenie świadczenia usług dla każdego obywatela w jednostkach samorządu terytorialnego. Prelegenci podkreślili, że usprawnienie systemu przekazywania danych jest szczególnie istotne na poziomie regionalnym, ponieważ dostęp do aktualnych i spójnych danych oraz możliwość ich przetwarzania stały się kluczowym czynnikiem rozwoju.

Dr Bohdan Wyżnikiewicz w referacie *Cena prawdy w statystyce* przybliżył zagadnienia związane z przypadkami podważania prawdziwości podawanych informacji statystycznych w polskiej statystyce publicznej. Zaproponował także schemat postępowania w odpowiedzi na krytykę danych statystycznych ze strony mediów i niektórych ekspertów.

Na dzień przed rozpoczęciem konferencji odbyły się warsztaty towarzyszące, które poprowadził Graham Kalton. Ich program obejmował omówienie metod próbkowania populacji stosowanych w ankietach, od prostego próbkowania losowego, poprzez systematyczne próbkowanie, grupowanie i próbkowanie wieloetapowe, po projekty paneli czy stosowanie powtarzających się ankiet. Kalton wiele uwagi poświęcił kwestiom związanym z praktycznym prowadzeniem badań sondażowych, w tym problemom wynikającym z braku odpowiedzi, niedokładności pomiarów lub pojawienia się duplikatów. Szczególnie owocna była dyskusja na temat prowadzenia badań statystycznych w krajach rozwijających się (np. afrykańskich), w których występują trudności związane z ankietowaniem zarówno lokalnej ludności, jak i miejscowej administracji. Jednym ze sposobów pokonania tego rodzaju problemów może być zastosowanie siatki selekcji, zaproponowanej po raz pierwszy przez Lesliego Kisha. Inne stosowane kroki to np. metoda Troidahl-Cartera czy metoda zaproponowana przez Louisa Rizzo. Kalton

omówił również metodę ACE (Address Coverage Enhancement), stworzoną w celu uzupełnienia metody ABS, oraz najnowsze trendy w badaniach sondażowych, obejmujące głównie metody web scrapingu stron takich jak Facebook czy Twitter.

Daniel Koprowicz (Urząd Statystyczny w Rzeszowie)