

Jednym z kluczowych zagadnień w statystyce publicznej stanowiących istotne wyzwanie metodologiczne jest estymacja wielkości populacji. Informacja o liczbie osób mieszkających w danym kraju lub regionie, czy też stanowiących populację określoną w inny sposób aniżeli przestrzennie, ma bowiem istotny wpływ na wiele aspektów życia, takich jak podział funduszy europejskich, środków na ochronę zdrowia czy kalkulację PKB. Jednocześnie strategia i wizja rozwoju spisów powszechnych formułowana przez Eurostat kładzie główny nacisk na ideę odejścia od tradycyjnego spisu organizowanego w odstępach dziesięcioletnich na rzecz kroczącego spisu wirtualnego bazującego na źródłach administracyjnych uzupełnionych o badania reprezentacyjne. *Powyższe działania są wynikiem funkcjonowania w statystyce publicznej nowego paradygmatu, w którym kładzie się nacisk na ponowne wykorzystanie wszystkich dostępnych źródeł danych zamiast tworzenia nowych.*

Wykorzystanie „niestatystycznych” źródeł danych prowadzi jednak do powstania problemów natury metodologicznej. Jako przykład można podać zrealizowaną przez GUS pracę eksperymentalną *Populacja cudzoziemców w Polsce w czasie COVID-19*, w której w oparciu o dziewięć rejestrów administracyjnych oszacowano, że w dniu w dniu 31.12.2019 w Polsce przebywało 2,1 mln. obcokrajowców. Można zadać pytanie, jak bliska prawdy jest ta wartość? Jaki wpływ na ten szacunek mają błędy nielosowe i w jaki sposób możemy go ewentualnie skorygować? Ważna jest również kwestia precyzji uzyskanych szacunków i sposobu jej wyznaczania. Z ekonomicznego punktu widzenia kluczowa jest bowiem wiarygodność szacunków wskaźników, takich jak stopa bezrobocia czy PKP, służy bowiem ona do oceny tego, w jakim stopniu poszerzają one wiedzę z zakresu sytuacji gospodarczej. Z tej perspektywy istotne jest zatem określenie, na ile informacje zawarte w rejestrach administracyjnych są wystarczające jako źródło wiedzy. *Uzyskanie odpowiedzi na wyżej postawione pytania z jednej strony wypełni lukę metodologiczną, z drugiej zaś strony służyć będzie wypracowaniu metody estymacji charakterystyk populacji cudzoziemców w Polsce, która pozwoli poszerzyć wiedzę na temat tej populacji.*

Głównym celem projektu jest opracowanie metod estymacji wielkości i charakterystyk populacji cudzoziemców w Polsce w oparciu o dostępne źródła danych. Ze względu na ubogość informacji w rejestrach administracyjnych planuje się wykorzystanie alternatywnych źródeł danych, takich jak informacje pozyskane ze smartfonów. Wykorzystanie wielu zbiorów ma pozwolić na uzyskanie szacunków na wzór spisu, jednak bez konieczności przeprowadzania badania w tradycyjnej formie. Wiąże to jednak ze sobą wyzwania metodologiczne wynikające z łączenia wielu źródeł danych oraz ich niestatystycznej natury. W związku z powyższym zespół badawczy kładzie duży nacisk na uwzględnienie w procesie estymacji oraz ocenie jakości uzyskanych szacunków wpływu integracji danych oraz błędów nielosowych (np. błędów pokrycia, błędów łączenia, nieignorowalnych braków danych), którymi obarczone są administracyjne oraz alternatywne źródła danych.

W projekcie wykorzystane będą dane administracyjnych dostępne w ramach statystyki publicznej, w tym zakresie planowana jest ścisła współpraca z GUS. Ponadto wykorzystane zostaną zanonimizowane dane jednostkowe z rejestrów policji, zagregowane dane ze Straży Granicznej i Państwowej Inspekcji Pracy oraz dane pochodzące z badań reprezentacyjnych realizowanych przez Narodowy Bank Polski. Projekt obejmował będzie również zastosowanie danych big data (ze smartfonów) o cechach behawioralnych i społeczno-demograficznych cudzoziemców.

Nowatorski charakter projektu polega na: 1) zaproponowaniu nowych estymatorów oraz metodę oceny ich jakości uwzględniających wpływ błędów nielosowych 2) opracowaniu metod estymacji liczby nielegalnie pracujących cudzoziemców w Polsce na bazie danych pochodzących ze smartfonów; 3) opracowaniu metod wnioskowania statystycznego na bazie zbiorów nielosowych, biorących pod uwagę takie zjawiska jak auto-selekcja czy nieignorowalne braki danych.

Szacunki w oparciu o wypracowane metody będą stanowiły punkt wyjścia do przeprowadzania złożonych analiz statystycznych w ujęciu krajowym oraz regionalnym, które przyczynią się do otrzymania wiarygodnych informacji oraz prognoz dotyczących roli cudzoziemców w polskiej gospodarce.

Należy także podkreślić, że wypracowane metody mają uniwersalny charakter i nie będą ograniczać się tylko i wyłącznie do populacji cudzoziemców, ale będą mogły również zostać wykorzystane do generowania statystyk w przypadku innych trudnych do zbadania populacji.

Na potrzeby projektu, polski zespół badawczy nawiązał współpracę w obszarze problematyki cudzoziemców z naukowcami z *University of Southampton, Utrecht University, Istat i Sapienza University of Rome*. Tak utworzony międzynarodowy zespół ekspertów, mających odpowiednie kompetencje i doświadczenie, będzie stanowił gwarant realizacji założonych w projekcie celów.