



Szacowanie liczby ludności przez służby statystyki publicznej na potrzeby samorządów lokalnych

Marcin Szymkowiak
Maciej Beręsewicz

Uniwersytet Ekonomiczny w Poznaniu
Urząd Statystyczny w Poznaniu

Mieszkaniec, czyli kto? – zmierzyć niemierzalne
Warszawa, 24-25.03.2025 r.



Plan prezentacji

Wprowadzenie

Metodyka szacunku wielkości populacji bez spisów

Nowe dane w szacunkach wielkości populacji

Podsumowanie

Wprowadzenie

Znaczenie dokładnych danych ludnościowych

- ▶ **Podstawa planowania strategicznego** w samorządach lokalnych.
- ▶ **Fundament podziału środków budżetowych** i dotacji.
- ▶ **Narzędzie do projektowania** infrastruktury i usług publicznych.

Wykorzystanie szacunków ludności w zarządzaniu lokalnym

Obszar	Zastosowanie
Planowanie przestrzenne	Projektowanie infrastruktury adekwatnej do potrzeb
Budżetowanie	Alokacja środków do jednostek pomocniczych
Edukacja	Planowanie sieci szkół i przedszkoli
Ochrona zdrowia	Określanie potrzeb w zakresie POZ
Transport publiczny	Optymalizacja siatki połączeń

Konsekwencje niedokładnych szacunków

- ▶ **Finansowe:** Utrata dochodów z subwencji i dotacji.
- ▶ **Infrastrukturalne:** Niedostosowanie usług publicznych do rzeczywistych potrzeb.
- ▶ **Strategiczne:** Błędne decyzje rozwojowe oparte na nieprawidłowych danych.
- ▶ **Społeczne:** Niezaspokojenie potrzeb faktycznych mieszkańców.

Tradycyjne metody szacowania liczby ludności

Metody bazujące na rejestrach

- ▶ Ewidencja ludności (PESEL).
- ▶ Rejestry meldunkowe.
- ▶ Dane z urzędów stanu cywilnego.

Metody bazujące na spisach

- ▶ Narodowy Spis Powszechny Ludności i Mieszkań (co 10 lat).
- ▶ Mikrospisy i badania reprezentacyjne.

Nowoczesne metody estymacji

Metody statystyczne

- ▶ Metody capture-recapture dedykowane populacjom trudnym do zbadania.
- ▶ Poszukiwanie śladów życia w rejestrach.
- ▶ Metody statystyki małych obszarów.

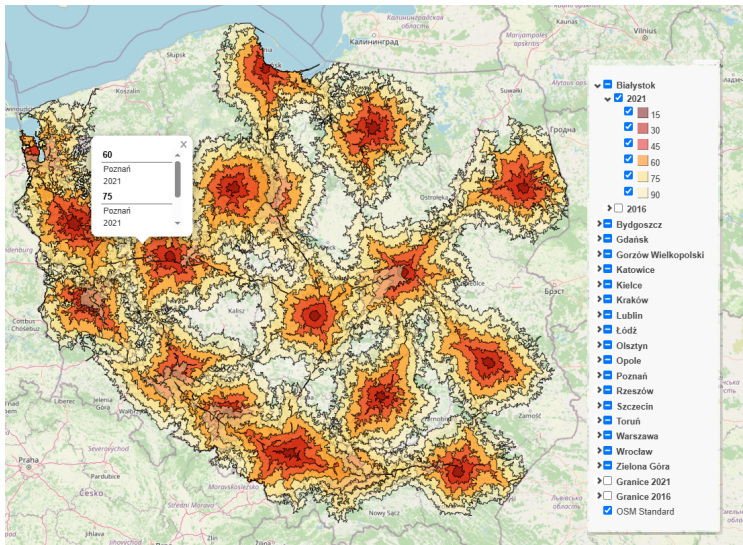
Wykorzystanie alternatywnych źródeł danych

- ▶ Dane z sieci komórkowych.
- ▶ Dane z mediów społecznościowych.
- ▶ Dane z systemów energetycznych i wodociągowych.

Dylematy związane z szacowaniem liczby ludności

- ▶ Różne definicje populacji: populacja rezydentów, faktyczna, rzeczywista itd.
- ▶ Konieczność badania populacji komplementarnych (dziennych, nocnych, sezonowych) - ang. supplementary population bases.
- ▶ Konieczność badania populacji trudnych do zbadania - ang. hard-to-reach population.
- ▶ Różnorodność źródeł statystycznych (spisy, rejestry administracyjne, badania reprezentacyjne) oraz innych danych (od operatorów telefonii komórkowych, zużycie energii, wody, wywóz śmieci, zdjęcia satelitarne).
- ▶ Kwestia oceny jakości niestandardowych źródeł danych.
- ▶ Faktyczne miejsce zamieszkania, a *fractional counting*.
- ▶ Kwestie mobilności ludności (dojazdy do pracy i szkół).
- ▶ Probabilistyczna integracja rejestrów administracyjnych.

Dostępność czasowo-przestrzenno miast



Dostępność czasowo-przetrzenna miast

- ▶ **Analiza populacji w zasięgu czasowym** - szacowanie liczby mieszkańców znajdujących się w poszczególnych strefach czasowych (np. do 30 min, 30-60 min, 60-90 min od stolicy województwa).
- ▶ **Prognozowanie migracji wewnętrznych** - obszary o dobrej dostępności czasowej do ośrodków wojewódzkich mogą przyciągać nowych mieszkańców.
- ▶ **Planowanie infrastruktury** - identyfikacja obszarów o dużej gęstości zaludnienia i słabej dostępności czasowej, które wymagają inwestycji transportowych.
- ▶ **Potencjał rynku pracy** - analiza liczby potencjalnych pracowników w zasięgu dojazdów do głównych ośrodków gospodarczych.

Metodyka szacunku wielkości populacji bez spisów

Pomiar wielkości populacji – ujęcie nowoczesne

▶ **Spisy wirtualne oparte o rejestry administracyjne**

- ▶ niski koszt w porównaniu do spisów powszechnych,
- ▶ wymagające dostępu do rejestrów i współpracy gestorów,
- ▶ wymagające określonych wysokich kompetencji za relatywnie niskie wynagrodzenie,
- ▶ ale możliwe do realizacji w krótkim czasie i dostarczające aktualnych informacji.

▶ **Spisy wirtualne oparte o Big Data**

- ▶ niski koszt w porównaniu do spisów powszechnych,
- ▶ dane zbierane pasywnie przez systemy informacyjne,
- ▶ znaczące błędy pokrycia i pomiaru,
- ▶ umożliwiające publikacje wyników w czasie bliskim do rzeczywistego.

Wyzwanie: wyznaczenie populacji nieobserwowanej w tych źródłach.

Estymacja wielkości populacji – najprostsze podejście

Tab.: Przykład integracji dwóch źródeł danych

Źródło A	Źródło B		
	Tak	Nie	
Tak	n_{11}	n_{10}	n_{1+}
Nie	n_{01}	$n_{00} = ?$	$?$
	n_{+1}	$?$	N

W takim przypadku, najprostsze podejście umożliwiające oszacowanie wielkości populacji, możemy zastosować estymator Lincolna-Petersen'a (por. *capture-recapture*, *multiple system estimation*)

$$\hat{N} = \frac{n_{1+} n_{+1}}{n_{11}}.$$

- ▶ Probabilistyczne łączenie rekordów – łączenie rejestrów bez kluczy (np. PESEL) wprowadza dodatkowe błędy.
- ▶ Błędy nadreprezentacji i ślady życia – osoby przebywające poza Polską.
- ▶ Nieaktualne rejestry administracyjne w zakresie miejsc pobytu – problem z ustaleniem miejsca zamieszkania – możliwe rozwiązanie: *fractional counting*.

Może w takim warto wykorzystać źródła *big data*?

Nowe dane w szacunkach wielkości populacji

Dane ze smartfonów: obywatele Ukrainy

Tab.: Liczba obywateli Ukrainy (18+) w źródłach danych (na stan 2021.12.30)

Źródło	Liczba osób
Selectivv (2021Q4)	1,262,765

Dane ze smartfonów: obywatele Ukrainy

Tab.: Liczba obywateli Ukrainy (18+) w źródłach danych (na stan 2021.12.30)

Źródło	Liczba osób
Selectivv (2021Q4)	1,262,765
Rejestr PESEL	1,226,816

Dane ze smartfonów: obywatele Ukrainy

Tab.: Liczba obywateli Ukrainy (18+) w źródłach danych (na stan 2021.12.30)

Źródło	Liczba osób
Selectivv (2021Q4)	1,262,765
Rejestr PESEL	1,226,816
Rejestr NFZ	1,266,265

Dane ze smartfonów: obywatele Ukrainy

Tab.: Liczba obywateli Ukrainy (18+) w źródłach danych (na stan 2021.12.30)

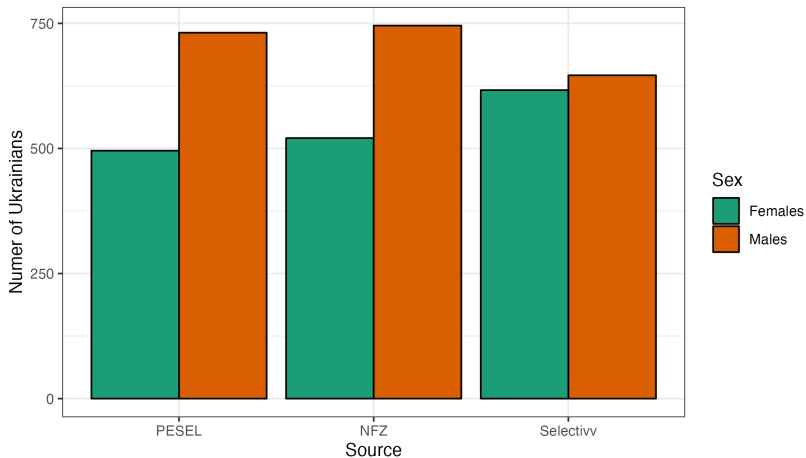
Źródło	Liczba osób
Selectivv (2021Q4)	1,262,765
Rejestr PESEL	1,226,816
Rejestr NFZ	1,266,265
Rejestr ZUS	703,008

Dane ze smartfonów: obywatele Ukrainy

Tab.: Liczba obywateli Ukrainy (18+) w źródłach danych (na stan 2021.12.30)

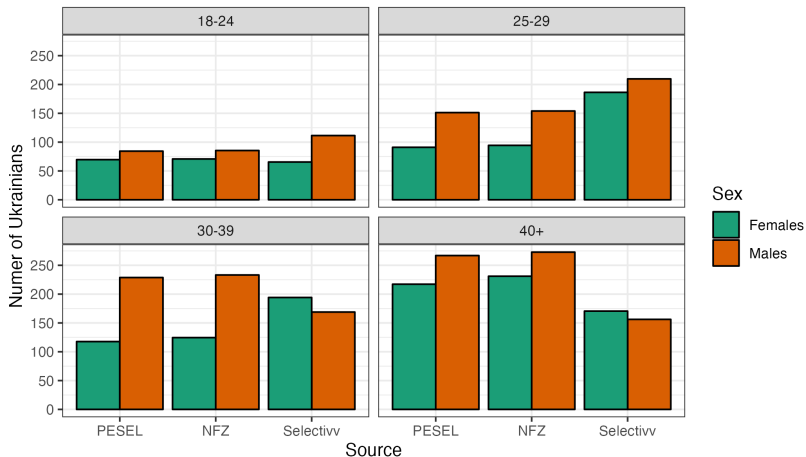
Źródło	Liczba osób
Selectivv (2021Q4)	1,262,765
Rejestr PESEL	1,226,816
Rejestr NFZ	1,266,265
Rejestr ZUS	703,008
Rejestr UDSC	272,927

Dane ze smartfonów: obywatele Ukrainy – wg płci



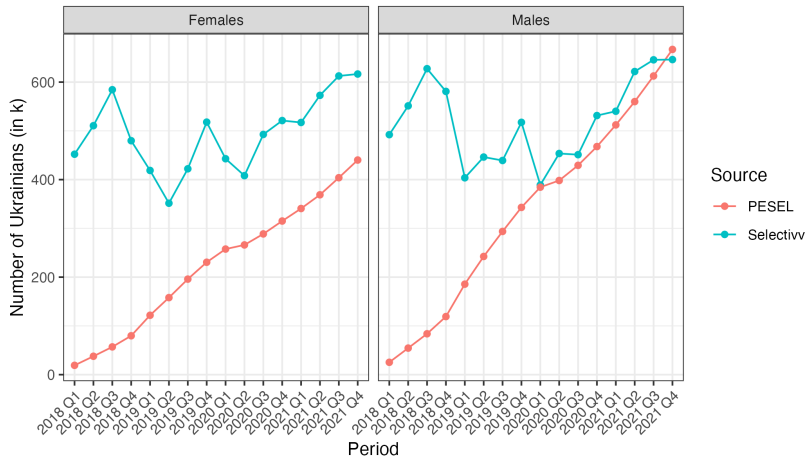
Rys.: Liczba obywateli Ukrainy (18+) wg płci w źródłach danych (na stan 2021.12.30)

Dane ze smartfonów: obywatele Ukrainy – wg wieku



Rys.: Liczba obywateli Ukrainy (18+) wg płci i wieku w źródłach danych (na stan 2021.12.30)

Dane ze smartfonów: obywatele Ukrainy – wg czasu



Rys.: Liczba obywateli Ukrainy (18+) wg okresu w źródłach danych

Porównanie populacji cudzoziemców (18+) Selectiv z NSP 2021

Ludność ogółem wg NSP 2021 (tys.)

Grupy Krajów	Spis	Selectiv
Europa (bez UE)	1,032.2	1,171.2
Azja	133.5	436.0
UE	130.3	630.4
Inne	36.0	61.7
Ogółem	1,459.2	2,919.3

Imigranci przebywający czasowo wg NSP 2021 (tys.)

Długość pobytu	Spis	Selectiv
30 dni – 3 miesiące	519.5	925.4
3 miesiące – 12 miesięcy	604.9	986.2
ponad 12 miesięcy	309.3	1,007.6
Ogółem	1,433.8	2,919.3

Wiek 20-29 (tys.)

Grupy Krajów	Spis	Selectiv
Europa (bez UE)	327.7	554.0
Azja	36.9	200.9
UE	18.5	292.3
Inne	9.7	23.1
Ogółem	392.8	1,070.3

Imigranci przebywający czasowo wg NSP 2021 (m. Poznań, tys.)

Długość pobytu	Spis	Selectiv
30 dni – 3 miesiące	16.2	24.6
3 miesiące – 12 miesięcy	13.7	28.6
ponad 12 miesięcy	9.9	28.3
Ogółem	39.7	81.6



Próba walidacyjna – wyniki

Tab.: Dokładność danych Selectivv dla zmiennych społeczno-demograficznych

Zmienna	Kategoria	Błąd klasyfikacji	Wielkość próby
Kraj	Białoruś	4.0	101
	Polska	3.2	247
	Ukraina	6.5	153
Płeć	Kobiety	12.7	221
	Mężczyźni	10.7	280
Wiek	18-24	11.9	236
	25-29	15.2	151
	30-39	7.8	64
	40+	4.0	50
Długość pobytu	3m	38.6	44
	3m-12m	21.4	112
	12m+	2.1	97

Podsumowanie

Podsumowanie

- ▶ **Wyzwania:** estymacja wielkości populacji bez spisów powszechnych, błędy klasyfikacji, opracowanie metodyki śladów życia, długości pobytu i miejsca zamieszkania, populacja dzienna i nocna.
- ▶ **Dyskusja:** zmiana sposobu liczenia populacji – *fractional counting* – uwzględniamy niepewność i nie liczymy "głów" (ang. *head count*), a prawdopodobieństwo przebywania (np. 80% Poznań, 20% Radojewo).
- ▶ **Ilu nas jest:** przejście od "zliczania" do "szacowania" (z odpowiednim błędem szacunku). Czy jesteśmy przygotowani na określanie wielkości populacji nie jako jedna liczba ale przedział?



Dziękujemy za uwagę

<https://poznan.stat.gov.pl/osrodki/osrodek-statystyki-miast-1234/>

<https://poznan.stat.gov.pl/osrodki/osrodek-metodologii-badan-ludnosciowych/>