



## Raport końcowy

Praca badawcza

# Cudzoziemcy na krajowym rynku pracy w ujęciu regionalnym

## *Foreigners on the national labour market – regional approach*

realizowana w ramach projektu „Statystyka dla polityki spójności. Wsparcie systemu monitorowania polityki spójności w perspektywie finansowej 2014-2020 oraz programowania i monitorowania polityki spójności po 2020 roku”



Warszawa 2018

### Centrum Badań i Edukacji Statystycznej GUS

Jachranka 81 k/Warszawy, 05-140 Serock  
Jachranka@stat.gov.pl  
cbies.stat.gov.pl



Rzeczpospolita  
Polska

Unia Europejska  
Fundusz Spójności



**Jednostka opracowująca raport:**

Centrum Badań i Edukacji Statystycznej GUS

**Kierownik projektu:**

Dorota Szattys

**Zespół badawczy:**

Michał Adamski, Maciej Beręsewicz, Mariusz Chmielewski, Piotr Filip, Grzegorz Gudaszewski, Daniel Godlewski, Tomasz Józefowski, Paweł Kaczorowski, Zofia Kostrzewa, Jacek Kowalewski, Arleta Olbrot-Brzezińska, Artur Owczarkowski, Joanna Stańczak, Karina Stelmach, Marcin Szymkowiak, Anna Wysocka.

## Przedmowa

Istnieje duże zapotrzebowanie na określenie rzeczywistych rozmiarów wielkości zasobu cudzoziemców przebywających w Polsce czasowo. (z uwzględnieniem nierejestrowanej/nieudokumentowanej imigracji), w szczególności w obszarze imigracji pracowniczych. Dane dotyczące zasobów imigracyjnych są ważnym elementem prowadzenia polityki spójności, zwłaszcza, jeśli weźmie się pod uwagę fakt, iż imigracja cudzoziemców nie jest zjawiskiem dotyczącym w równym stopniu obszaru całego kraju. Cudzoziemcy tworzą zazwyczaj „klastry” w sensie rozmieszczenia geograficznego, z tego względu zagadnienie imigracji powinno być analizowane w kontekście regionalnych rynków pracy. Istotnym powodem realizacji badania była potrzeba spełnienia oczekiwań licznych regionalnych i lokalnych odbiorców informacji, zgłaszających od wielu lat zapotrzebowanie na informacje o rynku pracy, uwzględniające określony poziom szczegółowości oraz zdefiniowane agregacje. Obecnie nie ma miarodajnego i bezpośredniego źródła danych, które dostarczałoby wiarygodnych informacji w tym zakresie.

Celem niniejszej pracy badawczej było opracowanie metody szacunku zasobów cudzoziemców przebywających w Polsce czasowo, ze szczególnym uwzględnieniem cudzoziemców pracujących na terytorium Polski, w oparciu o administracyjne i statystyczne źródła danych oraz przy wykorzystaniu metod bazujących na technikach wielokrotnych połowów (*capture-recapture*). Praca badawcza stanowi innowacyjne w tym zakresie podejście do opracowywania wyników badań prowadzonych przez statystykę publiczną. Rezultaty zaprezentowane w niniejszym raporcie należy traktować jako eksperymentalne i niebędące oficjalnymi statystykami publikowanymi przez Główny Urząd Statystyczny.

Przekazując do rąk Państwa niniejsze opracowanie zwracamy się z prośbą o uwagi i sugestie, które będą stanowić cenną wskazówkę w dalszych pracach nad rozwojem i doskonaleniem zastosowanej metody.

## Preface

There is a substantial demand for determination of the real size of foreigners' stock staying temporarily in Poland (including unregistered/undocumented immigration), in particular in the field of labour immigration.

Data on the stock of immigration are an important element to conduct of cohesion policy, especially when it comes to the fact that immigration of foreigners is not a phenomenon affecting the entire territory of the country equally. Foreign citizens usually form "clusters" in the sense of geographical distribution, therefore the issue of immigration should be analyzed in the context of regional labour markets. An important reason for the study was the need to meet the expectations of numerous regional and local users of information, reporting for many years the demand for information about the labour market, including a specific level of detail and defined aggregations. Currently, there is no reliable and direct source of data that would provide reliable information in this area.

The aim of this research work was to develop a method for estimating the stock of foreigners staying temporarily in Poland, with particular emphasis on foreigners working on the territory of Poland based on administrative and statistical data sources and using capture-recapture methods . Therefore, study work is an innovative approach to the development of research results carried out by official statistics. Results presented in the report should be considered as experimental and thus not considered as published by Statistics Poland.

Handing over the following report, we will be grateful for any comments and suggestions that will be a valuable source to further development and improvement of the method.

## Spis treści Contents

<b>Przedmowa</b> .....	<b>3/4</b>
<i>Preface</i>	
<b>Objaśnienia skrótów</b> .....	<b>9</b>
<i>Abbreviations</i>	
<b>Wstęp</b> .....	<b>10/12</b>
<i>Introduction</i>	
<b>Synteza</b> .....	<b>14/17</b>
<i>Executive summary</i>	
<b>1. Wyniki badania</b> .....	<b>20</b>
<i>Executive summary</i>	
1.1. Liczebność cudzoziemców.....	20
<i>Number of foreigners</i>	
1.2. Rozmieszczenie cudzoziemców według województw.....	20
<i>Territorial distribution of foreigners by voivodships</i>	
1.3. Charakterystyka cudzoziemców.....	21
<i>Characteristics of foreigners</i>	
1.3.1. Płeć i wiek .....	21
<i>Sex and age</i>	
1.3.2. Kraj obywatelstwa.....	24
<i>Country of citizenship</i>	
1.3.3. Status na rynku pracy .....	25
<i>Labour market status</i>	
1.3.4. Sekcje PKD.....	26
<i>PKD sections</i>	
1.3.5. Rozmieszczenie terytorialne cudzoziemców pracujących w Polsce .....	27
<i>Territorial distribution of foreigners working in Poland</i>	
1.3.6. Okres przebywania w Polsce .....	28
<i>Period of residence in Poland</i>	
1.3.7. Poziom wykształcenia .....	28
<i>Level of education</i>	
1.3.8. Charakter miejsca zamieszkania .....	29
<i>Type of the place of residence</i>	
1.4. Cudzoziemcy w podregionach .....	29
<i>Foreigners in subregions</i>	
<b>2. Metodologia badania</b> .....	<b>31</b>
<i>Survey methodology</i>	
2.1. Zakres podmiotowy badania .....	31
<i>The subject of study</i>	
2.2. Zakres przedmiotowy badania.....	31
<i>The scope of study</i>	
2.3. Zakres terytorialny.....	32
<i>Territorial scope</i>	
2.4. Przyjęte definicje i listy kodowe.....	32
<i>Definitions and code lists</i>	

2.5.	Przepisy prawne regulujące kwestię pobytu i zatrudnienia cudzoziemców w Polsce .....	33
	<i>Polish legal acts regulating the stay and employment of foreigners in Poland</i>	
2.6.	Charakterystyka potencjalnych źródeł o cudzoziemcach .....	33
	<i>Characteristics of the potential sources of foreigners</i>	
2.7.	Przygotowanie zbiorów danych na potrzeby badania .....	35
	<i>Data preparation and processing</i>	
2.7.1.	Dobór podmiotowy i przedmiotowy .....	35
	<i>Selection of data sources</i>	
2.7.2.	Wyliczanie cech pochodnych na podstawie przekształceń surowych danych .....	36
	<i>Key variables identified in the data</i>	
2.7.3.	Redukcja nadmiarowych danych .....	36
	<i>Data cleaning and deduplication</i>	
2.7.4.	Wyodrębnianie podstawowych jednostek/podmiotów badania – operacje agregowania i restrukturyzacji zbiorów.....	36
	<i>Unit identification and delimitation</i>	
2.7.5.	Łączenie (parowanie) rekordów z różnych zbiorów .....	37
	<i>Deterministic record linkage of register data</i>	
2.8.	Szacunek liczby cudzoziemców w Polsce .....	37
	<i>Estimation of number of foreigners in Poland</i>	
2.8.1.	Szczegółowa informacja o założeniach przyjętych w pracy badawczej .....	37
	<i>Assumptions of the proposed approach</i>	
2.8.2.	Opis wykorzystanych źródeł danych .....	39
	<i>Description of selected data sources</i>	
2.8.3.	Opis zmiennych pochodzących z analizowanych rejestrów.....	40
	<i>Description of variables used in the models</i>	
2.9.	Opis procedury estymacji wielkości populacji cudzoziemców w Polsce oraz szczegółowych przekrojach w 2015 i 2016 r. ....	43
	<i>Description of the capture-recapture method to estimate number of foreigners in Poland in 2015 and 2016</i>	
2.9.1.	Estymacja wariancji estymatora wielkości populacji .....	43
	<i>Estimation of variance of number of foreigners in Poland</i>	
2.9.2.	Wybór modelu .....	44
	<i>Model selection</i>	
2.9.3.	Opis wybranego modelu log-liniowego.....	46
	<i>Description of the selected log-linear model</i>	
2.10.	Otrzymane szacunki i ich interpretacja .....	48
	<i>Estimates and interpretation of the results</i>	
2.10.1.	Podstawowe informacje dotyczące interpretacji szacunków .....	48
	<i>Basic information on the estimates interpretation</i>	
2.10.2.	Szacunki wielkości populacji wraz z ich precyzją .....	49
	<i>Estimates of population size and its precision</i>	
2.10.3.	Informacja o szacunkach dla podregionów .....	49
	<i>Estimates for subregions</i>	
2.11.	Ocena spełnienia kryteriów jakości danych .....	51
	<i>Evaluation of data quality criteria</i>	
<b>3.</b>	<b>Podsumowanie i rekomendacje .....</b>	<b>53</b>
	<i>Conclusions and recommendations</i>	

<b>Bibliografia</b> .....	<b>55</b>
---------------------------	-----------

*Bibliography*

<b>Spis załączników</b> .....	<b>56</b>
-------------------------------	-----------

*List of annexes*

## Spis tablic

### List of tables

Cudzoziemcy w wieku 18 lat i więcej przebywający w Polsce w latach 2015 i 2016 według województw i grup wieku .....	23
Zestawienie liczby cudzoziemców w wieku 18+ według PESEL, UdSC i ZUS zgodnie ze stanem na 31.12.2015 r. ....	39
Zestawienie liczby cudzoziemców w wieku 18+ według PESEL, UdSC i KEP zgodnie ze stanem na 31.12.2015 r. ....	39
Zestawienie liczby cudzoziemców w wieku 18+ według PESEL, UdSC i ZUS zgodnie ze stanem na 31.12.2016 r. ....	39
Zestawienie liczby cudzoziemców w wieku 18+ według PESEL, UdSC i KEP zgodnie ze stanem na 31.12.2016 r. ....	40
Wybrane miary jakości modeli log-liniowych według kombinacji źródeł oraz roku .....	45
Porównanie modelu 2s według kombinacji źródeł danych według miar Pseudo-R <sup>2</sup> .....	46
Postać ostatecznego modelu log-liniowego (notacja nawiasowa) wraz z oszacowaną wielkością populacji cudzoziemców w Polsce oraz 95% przedziałem ufności .....	46
Zestawienie zmiennych oraz zmienności przez nie wyjaśnionej dla 2015 roku .....	47
Zestawienie zmiennych oraz zmienności przez nie wyjaśnionej dla 2016 roku .....	47
Szacunek wielkości populacji cudzoziemców w Polsce .....	49
Szacunek wielkości populacji cudzoziemców w Polsce według kraju obywatelstwa .....	49
Szacunek wielkości populacji cudzoziemców w Polsce według kraju obywatelstwa na podstawie modelu uwzględniającego podregiony .....	50
Wybrane podregiony charakteryzujące się nieakceptowalną precyzją przed imputacją danych na podstawie modelu uwzględniającego podregiony .....	50

## Spis wykresów

### List of charts

Cudzoziemcy w wieku 18 lat i więcej przebywający w Polsce w latach 2015 i 2016 według płci .....	21
Cudzoziemcy w wieku 18 lat i więcej przebywający w Polsce w latach 2015 i 2016 według województw i płci .....	22
Cudzoziemcy w wieku 18 lat i więcej przebywający w Polsce w latach 2015 i 2016 według wieku .....	23
Cudzoziemcy w wieku 18 lat i więcej przebywający w Polsce w latach 2015 i 2016 według kraju obywatelstwa (w %) .....	24
Cudzoziemcy w wieku 18 lat i więcej przebywający w Polsce w latach 2015 i 2016 według województw i kraju obywatelstwa (w %) .....	25
Cudzoziemcy w wieku 18 lat i więcej przebywający w Polsce w latach 2015 i 2016 według statusu na rynku pracy .....	26
Cudzoziemcy w wieku 18 lat i więcej przebywający w Polsce w latach 2015 i 2016 według sekcji PKD .....	26
Cudzoziemcy w wieku 18 lat i więcej przebywający w Polsce w latach 2015 i 2016 według okresu przebywania w Polsce .....	28
Cudzoziemcy w wieku 18 lat i więcej przebywający w Polsce w latach 2015 i 2016 według poziomu wykształcenia .....	29

Cudzoziemcy w wieku 18 lat i więcej przebywający w Polsce w latach 2015 i 2016 według charakteru miejsca zamieszkania (miasto/wieś) .....	29
Rozkład cudzoziemców według płci, lat oraz kombinacji rozważanych źródeł.....	41
Rozkład cudzoziemców według wieku, lat oraz kombinacji rozważanych źródeł.....	41
Rozkład cudzoziemców według kraju obywatelstwa, lat oraz kombinacji rozważanych źródeł.....	42
Rozkład cudzoziemców według województw (sklasyfikowanych jako ustalony i nieustalony), lat oraz kombinacji rozważanych źródeł .....	42

## Spis kartogramów

### List of cartograms

Cudzoziemcy w wieku 18 lat i więcej przebywający w Polsce w latach 2015 i 2016 według województw .....	20
Pracujący cudzoziemcy w wieku 18 lat i więcej przebywający w Polsce w latach 2015 i 2016 według województw .....	27
Cudzoziemcy w wieku 18 lat i więcej przebywający w Polsce w latach 2015 i 2016 według podregionów (NUTS 2016) .....	30



## Objaśnienia skrótów

### Abbreviations

Skrót <i>Abbreviation</i>	Znaczenie <i>Meaning</i>
BAEL	Badanie aktywności ekonomicznej ludności <i>Labour force survey</i>
GUS	Główny Urząd Statystyczny <i>Statistics Poland</i>
EFTA	Europejskie Stowarzyszenie Wolnego Handlu <i>European Free Trade Association</i>
KEP	Krajowa Ewidencja Podatników <i>National Taxpayer Register</i>
MiIR	Ministerstwo Inwestycji i Rozwoju <i>Ministry of Investment and Economic Development</i>
MSWiA	Ministerstwo Spraw Wewnętrznych i Administracji <i>Ministry of Interior and Administration</i>
MRPiPS	Ministerstwo Rodziny Pracy i Polityki Społecznej <i>Ministry of Labour, Family and Social Affairs</i>
NSP	Narodowy spis powszechny ludności i mieszkań <i>National Population and Housing Census</i>
NUTS	Klasyfikacja Jednostek Terytorialnych do Celów Statystycznych <i>Nomenclature of territorial units for statistics</i>
OBM	Operacyjna Baza Mikrodanych <i>Operational Microdata Database</i>
OECD	Organizacja Współpracy Gospodarczej i Rozwoju <i>Organisation for Economic Co-operation and Development</i>
PBSSP	Program badań statystycznych statystyki publicznej <i>Programme of statistical surveys of public statistics</i>
PESEL	Powszechny elektroniczny system ewidencyjny ludności <i>Common Electronic System of Population Register</i>
PKD	Polska Klasyfikacja Działalności <i>Statistical classification of economic activities</i>
RDS	<i>Respondent-Driven Sampling</i>
UE	Unia Europejska <i>European Union</i>
UdSC	Urząd do Spraw Cudzoziemców <i>Office for Foreigners</i>
ZUS	Zakład Ubezpieczeń Społecznych <i>Social Insurance Fund</i>

## Wstęp

Niniejszy dokument jest raportem stanowiącym podsumowanie pracy badawczej pt. „Cudzoziemcy na krajowym rynku pracy w ujęciu regionalnym” realizowanej w ramach projektu „Wsparcie systemu monitorowania polityki spójności w perspektywie finansowej 2014-2020 oraz programowania i monitorowania polityki spójności po 2020 roku” współfinansowanego przez Unię Europejską ze środków Programu Operacyjnego Pomoc Techniczna 2014-2020.

Coraz częściej zarówno na szczeblu rządowym, jak i samorządowym oraz lokalnym podnoszona jest kwestia konieczności dysponowania rzeczywistą skalą zasobu cudzoziemców przebywających w Polsce czasowo i podejmujących pracę. Szczególnie istotne dla realizowania polityk ludnościowych migracyjnych i gospodarczych są informacje o cechach demograficzno-społecznych i ekonomicznych cudzoziemców. Kolejna ważna kwestia dotyczy skali imigracji nierejestrowanej, tj. pozostającej poza ewidencją. Obecnie nie ma miarodajnego i bezpośredniego źródła danych, które dostarczałoby wiarygodnych informacji w tym zakresie. Zwrócić należy uwagę na fakt, iż imigracja cudzoziemców nie jest zjawiskiem dotyczącym obszaru całego kraju w równym stopniu i podlega zróżnicowaniu przestrzennemu, zwłaszcza w kontekście regionalnych rynków pracy.

Opracowane w projekcie badawczym szacunki wielkości zasobów imigracyjnych według cech demograficzno-społecznych i ekonomicznych cudzoziemców w ujęciu regionalnym mogą stanowić dobrą podstawę do wyprowadzenia (wtórnie) różnego rodzaju wskaźników dla wyodrębnionych jednostek terytorialnych, takich jak np. bilans migracyjny netto czy wskaźnik aktywności zawodowej cudzoziemców. Te ostatnie zaś mogą być wykorzystywane, przez władze samorządowe, m.in. do monitorowania wielkości zatrudnienia, wysokości stopy bezrobocia czy popytu na pracę cudzoziemców o wysokich kwalifikacjach oraz oceny wpływu powierzania pracy cudzoziemcom na wysokość płac pracowników rodzimych. Dodatkowo opracowane w projekcie szacunki mogą być wykorzystywane do monitorowania grup narażonych na wykluczenie zawodowe i społeczne poprzez zapobieganie substytucji rodzimych zasobów pracy przez cudzoziemców. Wreszcie dane te będą mogły stanowić podstawę do prowadzenia analiz statystycznych na temat sytuacji społeczno-gospodarczej poszczególnych regionów kraju oraz prognoz ich rozwoju.

Posiadanie informacji na temat liczby cudzoziemców przebywających czasowo, uwzględniając w tym nierejestrowanych imigrantów, stanowi dla służb statystyki publicznej w Polsce istotne wyzwanie metodologiczne. Wynika to z faktu, że zbiorowość taka stanowi populację trudną do zbadania przy wykorzystaniu tradycyjnych metod statystycznych (ang. *hard to reach population*). Populacja taka charakteryzuje się bowiem brakiem dostępnego operatu losowania oraz trudnością w identyfikacji jednostek badania. O ile rozpoznanie problemów występujących w populacjach trudnych do zbadania jest możliwe na gruncie badań statystycznych (można zastosować przykładowo dobór jednostek do badania w oparciu o metodę kuli śnieżnej czy metodę RDS<sup>1</sup>), o tyle proces estymacji wielkości takiej zbiorowości jest z metodologicznego punktu widzenia poważnym wyzwaniem badawczym. W literaturze przedmiotu istnieją jednak odpowiednie metody statystyczne, które umożliwiają estymację wielkości populacji trudnych do zbadania. Zaliczyć tutaj można przykładowo rozwiązania bazujące na technikach *capture-recapture*<sup>2</sup>, analizie klas ukrytych czy modelowaniu z uwzględnieniem tzw. modeli mieszanych z efektami losowymi. Skuteczne wykorzystanie tych technik w praktyce w dużej mierze zależy od dostępności danych statystycznych i jest uwarunkowane koniecznością spełnienia odpowiednich założeń leżących u podstaw poszczególnych metod.

---

<sup>1</sup> Metoda kuli śnieżnej, respondent driven sampling (RDS) to niekodowe techniki doboru respondentów do badania próby, wykorzystujące sieci powiązań osób należących do badanej populacji lub osób z nią powiązanych oraz schemat geograficzny (location sampling; time-location sampling). Wszystkie te metody opierają się na założeniu, że członkowie badanej populacji, osoby blisko z nimi związane lub uczęszczające do tych samych miejsc, mają większe możliwości wskazania i dotarcia do założonej liczby osób należących do danej populacji niż sami badacze.

<sup>2</sup> W polskiej literaturze trudno znaleźć adekwatne tłumaczenie pojęcia *capture-recapture*. Można próbować je tłumaczyć jako metodę wielokrotnego połowu, ponieważ wywodzi się z badań biologicznych (m.in. dla określania populacji ryb). Jednakże w raporcie pozostawiono pojęcie w jego oryginalnym brzmieniu.

W pracy badawczej podjęto próbę oszacowania wielkości zasobów cudzoziemców przebywających w Polsce w końcu 2015 i 2016 roku na poziomie kraju, województw i podregionów. Opracowano również podstawową charakterystykę cudzoziemców uwzględniającą m.in. cechy demograficzno-społeczne, obywatelstwo, status na rynku pracy. W tym celu posłużono się dwoma modelami różniącymi się wykorzystanymi zmiennymi – jednym dla oszacowania liczby cudzoziemców na poziomie podregionów oraz drugim dla oszacowania wszystkich pozostałych zmiennych.

Należy podkreślić, że wyniki otrzymane z obydwu modeli są nieporównywalne, tj. wyniki szacunku z modelu pierwszego nie będą tożsame z wynikami z modelu drugiego. Głównym powodem wyboru takiego podejścia był niski poziom agregacji przestrzennej, który jest trudny do osiągnięcia na podstawie wykorzystanych danych. Podobne podejście stosuje się w przypadku estymacji dla małych domen, gdzie na wysokich poziomach agregacji stosuje się estymatory bezpośrednie, a na niskich – podejście wspierane lub oparte na modelu.

W pierwszym etapie pracy badawczej dokonano analizy aktów prawnych regulujących pobyt i zatrudnienie cudzoziemców na terytorium RP, zinventaryzowano źródła danych statystycznych i pozastatystycznych, w tym rejestry i bazy danych administracyjnych oraz oceniono ich przydatność dla realizacji projektu. Zaprojektowano także wstępny zestaw wskaźników, możliwych do oszacowania na poziomie NUTS 3 wg klasyfikacji NUTS 2013 i 2016. Zaproponowano także metody szacunku (typu *capture-recapture* oraz bezpośredniego wyznaczania populacji) w celu ich przetestowania podczas naliczania danych w drugim etapie pracy badawczej.

W ramach drugiego etapu prac określono finalną metodę szacunku, testowo naliczono zestaw wskaźników zaprojektowanych w pierwszym etapie. Przeprowadzono następnie weryfikację poszczególnych wskaźników i dokonano wyboru ostatecznego zestawu, w którym uwzględniono również niezbędne modyfikacje w stosunku do pierwotnej listy wskaźników. Efektem przeprowadzonych prac jest niniejszy raport końcowy. Wszystkie prezentowane dane dotyczą 2015 i 2016 r.

Autorzy sugerują aby uzyskane wyniki rozważać w kontekście przedziałów ufności niż oszacowań punktowych.

Strukturę raportu podporządkowano realizacji celu głównego, celów szczegółowych oraz empirycznej weryfikacji zastosowanej metody szacunku. Na raport składają się: przedmowa, wstęp, synteza, trzy rozdziały, bibliografia oraz aneks.

Cele pracy, podsumowanie wyników pracy badawczej oraz informacje na temat możliwości ich wykorzystania w statystyce publicznej zostały zaprezentowane w rozdziale zatytułowanym „Synteza”. Kluczowym elementem przeprowadzonej pracy badawczej była przestrzenna analiza danych, którą przedstawiono w rozdziale analitycznym pn. „Wyniki badań” dla ostatecznie wybranych wskaźników. Opis metody szacunku oraz ocenę jakości danych zawarto w rozdziale poświęconym metodologii. W zakończeniu sformułowano rekomendacje uzyskane w toku prowadzonych prac oraz przedstawiono perspektywy dalszych badań.

Mając na względzie dużą złożoność problemu, innowacyjny charakter pracy oraz liczne trudności napotkane w trakcie jej realizacji (m.in. dotyczące niejednorodności oraz ograniczonego dostępu do źródeł danych) uzyskane wyniki wskazują, iż prace powinny być kontynuowane a metoda badawcza doskonalona.

## Introduction

This document is a report summarizing the research work entitled “Foreigners on the national labour market – regional approach” conducted under the project “Support monitoring system of cohesion policy in the financial perspective 2014–2020 and programming and monitoring cohesion policy after 2020” co-financed by the European Union from the Operational Program Technical Assistance 2014–2020.

Increasingly at both the governmental and local government level, the issue of the necessity to dispose of the real scale of the stock of foreigners staying temporarily in Poland and undertaking work is raised. The information on the demographic and social features of the foreigners is particularly important for the implementation of population and migration policies. Another important issue concerns the scale of undeclared immigration, i.e. remaining outside the register. Currently, there is no reliable and direct source of data that would provide reliable information in this area. Attention should be paid to the fact that immigration of foreigners is not a phenomenon affecting the entire area of the country to an equal extent and is subject to spatial differentiation, especially in the context of regional labour markets.

Having adequate statistical data on immigration stock would be an important contribution to conducting a proper migration policy related to the socio-economic aspects in relation to foreigners. Such data would also play a significant role in the process of building a long-term migration strategy of Poland, one of the objectives of which is to counteract the phenomenon of illegal immigration.

The estimates of the size of immigration stock according to the socio-demographic and economic characteristics of foreigners in regional terms developed in the research project may constitute a good basis for deriving (in secondary way) various types of indicators for separate territorial units, such as the net migration or the professional activity rate of foreigners. The latter may be used by local authorities, among others to monitor the employment size, the level of unemployment or the demand for work of highly qualified foreigners and the assessment of the impact of entrusting work to foreigners on the amount of wages and salaries of national workers. Additionally, the estimates developed in the project may be used to monitor groups exposed to professional and social exclusion by preventing substitution of domestic labour resources by foreigners. Finally, these data will be a basis for conducting statistical analyzes on the socio-economic situation of individual regions of the country and projections of their development.

Getting information on the number of foreigners staying temporarily, including undeclared immigrants, is therefore a significant methodological challenge for public statistics services in Poland. This is due to the fact that such a population is a population difficult to estimate using traditional statistical methods (hard to reach population). This population is characterized by the lack of available sampling frame and the difficulty in determining the units of the study. While recognition of problems occurring in difficult to study populations is possible on the basis of statistical surveys (for example, the selection of units based on snowball or respondent driven sampling), the estimation of the size of such a population is a serious research challenge from a methodological point of view. There are, however, adequate statistical methods in the literature on the subject, which allow the estimation of the size of the population difficult to study. Included here are, for example, solutions based on capture-recapture methods, analysis of latent classes or modeling including so-called mixed models with random effects. The effective use of these techniques in practice largely depends on the availability of statistical data and is conditioned by the need to meet the appropriate assumptions underlying the individual methods.

In the research work an attempt was made to estimate the amount of stock of foreigners staying in Poland at the end of 2015 and 2016 at the level of the country, voivodships and subregions. The basic characteristics of foreigners were also considered, including demographic and social variables, citizenship, status in the labour market. To achieve this goal two models with different covariates were applied – first for estimates on subregion level and second for estimation for all remaining variables.

It should be emphasized that results obtained from the first model are not coherent with the second model. The main reason for this approach is the level of spatial aggregation that is hard to achieve with available data. The approach is similar to small area estimation where estimates on a country level are based on direct estimators while model-assisted or model-based approach are used for the small

domains.

In the first stage of the research work, legal acts regulating the stay and employment of foreigners on the territory of the Republic of Poland were analyzed, sources of statistical and non-statistical data, including registers and administrative databases, were inventoried with the assessment of their usefulness for project implementation. A preliminary set of indicators, which can be estimated at the NUTS 3 level according to the NUTS 2013 and 2016 classification, was also designed. Methods of estimation (capture-recapture and direct population determination) were also proposed in order to test them while calculating data in the second stage of research work.

As part of the second stage of work, the final method of estimation was determined, and a set of indicators designed in the first stage was tested. The verification of individual indicators was then carried out and the final set was selected, which also included the necessary modifications in relation to the original list of indicators. The result of this work is this final report. All presented data relate to 2015 and 2016.

The authors suggest that the results should be considered in the context of confidence intervals rather than point estimates.

The structure of the report was subordinated to the implementation of the main objective, detailed objectives and empirical verification of the applied method of estimation. The report consists of: preface, introduction, executive summary, three chapters, bibliography and list of annex.

The objectives of the work, a summary of the results of the research work and information on the possibilities of their use in official statistics are presented in the chapter entitled Executive summary. The key element of the research work was the spatial data analysis, which was presented in the analytical chapter for the finally selected indicators. The description of the estimation method and the assessment of the quality of the data are included in the chapter on methodology. In conclusion, the proposals and recommendations obtained in the course of the conducted work were formulated and the prospects for further research were presented.

Taking into consideration the complexity of the problem, the innovative nature of the work and the many difficulties encountered during its implementation, among others regarding the heterogeneity of data sources, the results indicate that the work should be continued and the research method should be improved.

## Synteza

W pracy badawczej założono opracowanie metody szacunku zasobu cudzoziemców przebywających w Polsce czasowo według ich cech demograficznych (płci, ekonomicznych grup wieku), kraju poprzedniego zamieszkania, kraju obywatelstwa, statusu na rynku pracy oraz poziomu wykształcenia – dla Polski ogółem, województw i podregionów (NUTS 2016). Zasób cudzoziemców miał zostać oszacowany według czasu pobytu w Polsce cudzoziemców, tzn. z wyodrębnieniem pobytów krótkookresowych (od 3 do 12 miesięcy) oraz długookresowych (12 miesięcy i więcej). Zamierzano również rozpoznać możliwość prezentowania danych w podziale miasto/wieś oraz uwzględnienie sekcji PKD dla cudzoziemców pracujących. W przypadku cech demograficznych (płci, ekonomicznych grup wieku), kraju obywatelstwa dla Polski ogółem, województw i podregionów zastosowano modele log-liniowe. Pozostałe przekroje, tj. status na rynku pracy, poziom wykształcenia, czas pobytu, podział miasto/wieś, sekcje PKD, zostały uzyskane na podstawie rozszacowania wyników uzyskanych z modeli log-liniowych, zakładając odpowiednie struktury.

Podejmowano starania, aby powyższe założenia zostały zrealizowane w całości, jednakże niektóre z nich – ze względu na zakres danych w wykorzystywanych źródłach – musiały zostać w pewnym stopniu zmodyfikowane:

- Wyniki pracy badawczej opracowano nie dla całej zbiorowości cudzoziemców, a jedynie dla cudzoziemców w wieku 18 lat i więcej. Wynikało to z faktu, iż wykorzystywane zbiory ZUS i KEP praktycznie nie zawierały danych o osobach w wieku 0-17 lat. Uznano jednocześnie, że populacja w wieku 18 lat i więcej będzie odpowiednia dla charakterystyki pracujących cudzoziemców;
- Wyniki dotyczą ogółu cudzoziemców przebywających w Polsce bez wyraźnego wyodrębnienia cudzoziemców przebywających czasowo. W większości rejestrów (poza rejestrem PESEL) nie ma informacji o charakterze zamieszkania cudzoziemców w Polsce (czy cudzoziemiec mieszka stałe, czy przebywa czasowo) – stąd nie było możliwe wyodrębnienie populacji cudzoziemców przebywających czasowo i scharakteryzowanie wyłącznie tej populacji. Podano jednakże liczbę osób zameldowanych w Polsce na pobyt stały ustaloną na podstawie rejestru PESEL, zatem liczba cudzoziemców przebywających czasowo może zostać oszacowana. Charakterystyka osób dotyczy jednak ogółu cudzoziemców przebywających w Polsce w wieku 18 lat i więcej. Mając na uwadze fakt, że liczba cudzoziemców mieszkających w Polsce na stałe jest stosunkowo mała, można uznać, że prezentowane wyniki dotyczące cech demograficzno-społecznych dobrze opisują cudzoziemców przebywających w Polsce czasowo.
- Wyniki nie są prezentowane według kraju poprzedniego zamieszkania, jednakże w zdecydowanej większości kraj ten pokrywa się z krajem obywatelstwa, dla którego dane zostały opracowane; rezygnacja ze zmiennej „kraj poprzedniego zamieszkania” wynikała z braku dostępności tej informacji w analizowanych rejestrach;
- Kraje EFTA włączono do kategorii inne;
- Ze względu na niezadowalającą precyzję uzyskanych oszacowań zdecydowano o prezentowaniu większości danych dotyczących zasobu cudzoziemców w szczegółowych podziałach jedynie dla Polski ogółem. Dla województw zaprezentowano dane w podziale na płeć, grupy wieku i kraje obywatelstwa cudzoziemców. Dla pozostałych zmiennych trudno było uzyskać miarodajną strukturę, pozwalającą zaprezentować je na niższym poziomie terytorialnym. Dla podregionów i regionów warszawskiego stołecznego i mazowieckiego regionalnego – ze względów szerzej przedstawionych poniżej – zostały zaprezentowane jedynie dane ogólne. Dane o cudzoziemcach pracujących zostały opracowane dla Polski ogółem i dla województw;
- W przypadku szacunków dla podregionów zastosowano inny model niż dla pozostałych zmiennych, co powoduje, że wyniki są nieporównywalne. W związku z małą liczbą reprezentantów oraz niskim poziomem agregacji przestrzennej niektóre podregiony charakteryzują się bardzo niską precyzją. Z tego względu do otrzymanych wyników należy podchodzić z dużą ostrożnością. Autorzy nie rekomendują ich wykorzystywania. Ponieważ precyzja danych ogólnych dla

podregionów i regionów warszawskiego stołecznego i mazowieckiego regionalnego nie jest w pełni zadawalająca, prezentowanie danych dla tych jednostek według poszczególnych cech cudzoziemców (czyli danych o wiele bardziej szczegółowych) byłoby niecelowe. Ze względu na niską precyzję dane te nie byłyby miarodajne, zatem nie zdecydowano się na ich zaprezentowanie. Ta decyzja została podyktowana również tym, iż pokazanie tych danych o niedostatecznej jakości mogłoby umożliwić ich rozpowszechnienie oraz wykorzystywanie – bez właściwej interpretacji.

Według szacunków liczba cudzoziemców w wieku 18 lat i więcej przebywających w Polsce w końcu 2015 r. wynosiła 507,7 tys. – mieściła się w przedziale 369,1 tys. do 724,4 tys. (95% poziom ufności). Liczba ta – oprócz cudzoziemców zameldowanych na pobyt czasowy – obejmowała również cudzoziemców zameldowanych na pobyt stały (dla porównania takich osób według rejestru PESEL było 39,1 tys.). Wśród cudzoziemców przeważali mężczyźni (stanowili ponad 67% osób w wieku 18 lat i więcej), zdecydowaną większość stanowiły osoby w wieku produkcyjnym (ponad 96% osób). Najwięcej cudzoziemców zamieszkiwało w Warszawie (20,5% ogółu). Na drugim miejscu plasował się Kraków (prawie 8%), a na kolejnych – Wrocław i podregion Warszawski Zachodni (po prawie 6%). Odsetek pracujących cudzoziemców wynosił prawie 67%. Najwięcej cudzoziemców przebywało w Polsce przez okres od 3 do 12 miesięcy. Najliczniejszą grupę stanowili obywatele Ukrainy (283,7 tys. – 95% przedział ufności: od 203,9 tys. do 415,7 tys.).

Analogiczne dane oszacowano dla 2016 r. Liczbę cudzoziemców w wieku 18 lat i więcej przebywających w Polsce w końcu 2016 r. oszacowano na 743,7 tys. Liczba ta mieściła się w przedziale 600,8–943,1 tys. (95% poziom ufności). Liczba ta – oprócz cudzoziemców zameldowanych na pobyt czasowy – obejmowała również cudzoziemców zameldowanych na pobyt stały (dla porównania takich osób wg rejestru PESEL było 41,4 tys.). Podobnie jak w 2015 r. odnotowano przewagę mężczyzn (66% cudzoziemców w wieku 18 lat i więcej) i zdecydowaną przewagę osób w wieku produkcyjnym (prawie 97%). W Warszawie mieszkało najwięcej cudzoziemców (ponad 22%), w Krakowie 7%, a w podregionie Warszawskim Zachodnim ponad 6%. Odsetek cudzoziemców pracujących wynosił prawie 64%. Podobnie jak w 2015 r. dominowali cudzoziemcy przebywający od 3 do 12 miesięcy. Liczba obywateli Ukrainy wzrosła stosunku do roku poprzedniego i wynosiła 455,0 tys. (95% poziom ufności: 361,5 tys. – 584,7 tys.). W 2016 r. odnotowano wyraźny wzrost liczby cudzoziemców w stosunku do roku poprzedniego, wzrosła liczba obywateli Ukrainy, Białorusi, Rosji, Wietnamu i innych krajów spoza UE – liczba obywateli UE nieznacznie zmniejszyła się. Struktura wieku i płci pozostawała na podobnym poziomie.

Praca badawcza stanowi materiał kompleksowy i innowacyjny ze względu na zastosowaną metodę opracowania szacunku oraz wykorzystane źródła danych. Należy jednak zaznaczyć, że w trakcie realizacji projektu napotkano poważne trudności. Najważniejszą ich przyczyną był fakt, że metoda szacunku oparta była na źródłach danych administracyjnych, pozyskiwanych w ramach PBSSP dla innych badań, zatem o zakresie informacyjnych zdefiniowanym przez określoną jednostkę/departament realizującą własne badanie. Jako przykłady można wskazać następujące zbiory: rejestry ZUS pozyskiwane na potrzeby realizacji badań z zakresu rynku pracy nie zawierały wszystkich ubezpieczonych cudzoziemców. Bardzo cenne zbiory dotyczące zezwoleń na pracę i oświadczeń pracodawców o zamiarze powierzenia pracy cudzoziemcowi nie zawierały cech identyfikacyjnych, w związku z czym nie można było ich połączyć deterministycznie z innymi zbiorami. Podobny problem wystąpił w przypadku zbiorów Urzędu do Spraw Cudzoziemców pochodzących z systemu „Pobyt”. Zbiór PESEL z kolei nie obejmował cudzoziemców przebywających czasowo i zameldowanych w gminach, którzy nie posiadali numeru PESEL. Co więcej, wszystkie wykorzystane rejestry były aktualne na dzień 31.12.2016, co mogło wpłynąć na wyniki uzyskane na dzień 31.12.2015. Tym samym wtórne wykorzystanie rejestrów i zawartych w nich zmiennych miało istotny wpływ zarówno na sam wybór źródeł danych oraz metodę, jak i w konsekwencji na konstrukcję wskaźników.

Wnioski i rekomendacje sformułowane na zakończenie pracy badawczej dotyczą przede wszystkim konieczności doskonalenia metody szacunku w kolejnych latach. Jednakże warunkiem koniecznym jest pozyskanie danych z rejestrów i baz danych w zakresie, które literalnie będą odpowiadały założonym celom badawczym oraz aktualnych – na jednoznacznie określony dzień. Zatem niezbędne będzie

rozszerzenie zakresu danych w obecnie pozyskiwanych przez statystykę publiczną rejestrach oraz pozyskanie nowych źródeł pozastatystycznych (np. rejestry policji w zakresie cudzoziemców podejrzanych o popełnienie przestępstw na terytorium RP).

Rezultaty badania w postaci tablic wynikowych zamieszczono na stronie internetowej Głównego Urzędu Statystycznego w części *Statystyka dla polityki spójności*. Mogą one być wykorzystane wtórnie na potrzeby innych obszarów badawczych, np. rynku pracy, rachunków narodowych, mieszkalnictwa i budownictwa. Wyniki badania będą też mogły być wykorzystane przez jednostki administracji państwowej (w tym rządowej) odpowiedzialne za realizację polityki spójności, jednostki samorządu terytorialnego, instytucje zarządzające, pośredniczące i wdrażające Programy Operacyjne, środowisko naukowe, związki zawodowe, sektor prywatny, a także instytucje prowadzące badania i analizy statystyczne, w tym organizacje międzynarodowe, takie jak agendy Unii Europejskiej i Organizacja Współpracy Gospodarczej i Rozwoju (OECD). Wyniki pracy badawczej stanowią też rozszerzenie informacji dla badania znajdującego się w programie badań statystycznych statystyki publicznej pn. *Zasoby migracyjne*.



## Executive summary

The research assumes the development of a method for estimating the stock of foreigners staying temporarily in Poland according to their demographic characteristics (sex, economic age groups), country of previous residence, country of citizenship, status on the labour market and education level – for Poland in general, voivodships and sub-regions (NUTS 2016). The stock of foreigners was to be estimated according to the length of stay of foreigners in Poland, i.e. the identification of short-term stays (from 3 to 12 months) and long-term stays (12 months and more). It was also intended to recognize the possibility of presenting data in urban / rural division and including the PKD sections for working foreigners.

In case of demographic characteristics (sex, economic age groups), country of citizenship for Poland in general, voivodships and subregions log-linear models were applied. Remaining variables: status on the labour market, education level, the length of stay, urban / rural division and the PKD sections were obtained by assuming known structures from external data sources.

Efforts have been made to realize these assumptions completely, however, some of them – due to the scope of data in the sources used – had to be modified to a certain extent:

- The results of the research work were developed not for the entire population of foreigners, but only for foreigners aged 18 and over. This was due to the fact that the used collections of ZUS and KEP practically did not contain data on persons aged 0-17. At the same time, it was acknowledged that a population aged 18 and over would be suitable for the characteristics of working foreigners;
- The results apply to all foreigners staying in Poland without the explicit separation of foreigners staying temporarily. In most registers (except for the PESEL register) there is no information about the residence of foreigners in Poland (whether the foreigner resides permanently or temporarily) – hence it was not possible to identify a population of foreigners staying temporarily and characterize only this population. However, the number of persons registered in Poland for permanent residence determined on the basis of the PESEL register is provided, so that the number of foreigners staying temporarily can be estimated. However, the characteristics of people apply to all foreigners staying in Poland aged 18 and over. Bearing in mind the fact that the number of foreigners residing in Poland permanently is relatively small, it can be concluded that the presented results on demographic and social variables well describe foreigners staying temporarily in Poland.
- The results are not presented according to the country of previous residence, however, in the vast majority of cases, this country coincides with the country of citizenship for which the data were prepared;
- The EFTA was included into “other” category;
- Due to the unsatisfactory precision of the obtained estimates, it was decided to present the majority of data on the stock of foreigners in detailed breakdowns only for Poland as a whole. For voivodships, data broke down by sex, age groups and countries of citizenship of foreigners were presented. For other variables, it was difficult to obtain a reliable distributions, which would allow to use them for breakdowns at lower territorial levels. For sub-regions and regions warszawski stołeczny and mazowiecki regionalny – for reasons discussed below – only general data are presented. Data on working foreigners were prepared for Poland as a whole and for voivodships;
- In the case of estimates for sub-regions, a different model was applied than for the other variables, which means that the results are not comparable. Due to the small number of individuals and the low level of spatial aggregation, data for some subregions and regions warszawski stołeczny and mazowiecki regionalny are characterized by very low precision. Therefore, the obtained results should be approached with great caution. The authors do not recommend using them. Since the precision of general data for subregions and regions warszawski stołeczny and mazowiecki regionalny is not entirely satisfactory, presenting data for these areas according to individual characteristics of foreigners (meaning much more detailed

data) would be pointless. Due to their low precision, these data would not be reliable, so the authors decided not to present them. This decision was motivated by the fact that presenting the data of insufficient quality would lead to their dissemination and use – without proper interpretation.

According to estimates, the number of foreigners aged 18 and more staying in Poland at the end of 2015 was 507.7 thous. – it was in the range of 369.1 thous. up to 724.4 thous. (95% confidence interval). This number – apart from foreigners registered for temporary stay – also included foreigners registered for permanent residence (for comparison of such persons according to the PESEL register, there were 39.1 thous.). Men predominated among foreigners (they constituted over 67% of people aged 18 and more), The vast majority were people of working age (over 96% of people). The majority of foreigners lived in Warsaw (20.5% of the total). Kraków came second (almost 8%), followed by Wrocław and the West Warsaw sub-region (almost 6% each). The percentage of working foreigners was almost 67%. Most foreigners stayed in Poland for a period of 3 to 12 months. The largest group were citizens of Ukraine (283.7 thous. – 95% confidence interval: from 203.9 thous. to 415.7 thous.).

Analogical data was estimated for 2016. The number of foreigners aged 18 and more staying in Poland at the end of 2016 was estimated at 743.7 thous. The number was in the range of 600.8-943,1 thous. (95% confidence interval). This number – apart from foreigners registered for temporary stay – also included foreigners registered for permanent residence (for comparison of such persons according to the PESEL register there were 41.4 thous.). As in 2015, the majority of men (66% of foreigners aged 18 and more) and a definite advantage of people of working age (almost 97%) were noted. Most of the foreigners lived in Warsaw (over 22%), in Kraków 7%, and in the Western Warsaw subregion over 6%. The percentage of working foreigners was almost 64%. As in 2015, foreigners staying for 3 to 12 months dominated. The number of Ukrainian citizens increased in relation to the previous year and amounted to 455.0 thous. (95% confidence interval: 361.5 thous. – 584.7 thous.) In 2016 there was a clear increase in the number of foreigners compared to the previous year, the number of citizens of Ukraine, Belarus, Russia, Vietnam and other countries outside the EU increased – the number EU citizens slightly decreased. The age and sex structure remained at a similar level

The research work is a comprehensive and innovative material due to the method used to develop the estimate and the data sources used. However, it should be noted that serious difficulties have been encountered during the realisation of the project. The most important reason was the fact that the method of estimation was based on administrative data sources acquired under *the Programme of statistical surveys of public statistics* for other surveys, thus on the information scope defined by a specific unit / department performing its own research. As examples, the following collections can be identified: Social Insurance Fund registers obtained for the purposes of conducting labour market survey did not include all insured foreigners. Very valuable collections concerning work permits and employer's declarations of intention to entrust work to a foreigner did not contain identification variables, and therefore they could not be combined deterministically with other collections. A similar problem occurred in the case of collections of the Office for Foreigners coming from the POBYT system. The PESEL collection, in turn, did not include foreigners staying temporarily and registered in gminas who did not have a PESEL number. Moreover, data sources were as of 31.12.2016, which could have influenced the results obtained on 31.12.2015. Thus, the secondary use of registers and the variables contained in them had a significant impact both on the choice of data sources and the method as well as on the construction of indicators.

Conclusions and recommendations formulated at the end of the research work mainly concern the need to improve the method of estimation in the following years. However, a prerequisite is to obtain data from registers and databases to the extent that they strictly meet the assumed research goals. Therefore, it will be necessary to extend the scope of data in the registers currently acquired by public statistics and to acquire new non-statistical sources (for example registers of Polish Police referring to foreigners who have committed a crime on the territory of the Republic of Poland).

The results of the research in the form of result tables are available on the website of the Statistics Poland in the Statistics for Cohesion Policy section. They can be used secondarily for the needs of other research areas, such as the labour market, national accounts, housing and construction.



**Rzeczpospolita  
Polska**

**Unia Europejska**  
Fundusz Spójności



The results of the study will also be able to be used by state administration (including government) units responsible for the implementation of cohesion policy, local government units, managing institutions, intermediaries and implementing Operational Programs, scientific environment, trade unions, private sector, as well as research and development institutions. statistical analyzes, including international organizations, such as the European Union's agendas and the Organisation for Economic Co-operation and Development (OECD).

## 1. Wyniki badania *Survey results*

Poniżej zaprezentowano wyniki szacunku liczby cudzoziemców w Polsce. Cudzoziemcy zostali scharakteryzowani według płci, wieku, kraju obywatelstwa, okresu przebywania w Polsce, statusu na rynku pracy. Przedstawiono rozmieszczenie terytorialne populacji cudzoziemców, a dla cudzoziemców pracujących wyszacowano sekcje PKD. Wyniki te w postaci tabelarycznej zostały zamieszczone w załączniku 1 do niniejszego raportu.

Prezentowane wyniki są eksperymentalne, ponieważ uzyskano je na podstawie modelu log liniowego. Należy je traktować z ostrożnością, ponieważ przyjęto szereg założeń, które zostały opisane w dalszej części raportu. Wyniki należy traktować jako przybliżające prawdziwą wielkość populacji cudzoziemców. Sugeruje się, aby wyniki interpretować w kontekście przedziałów ufności, a nie oszacowań punktowych.

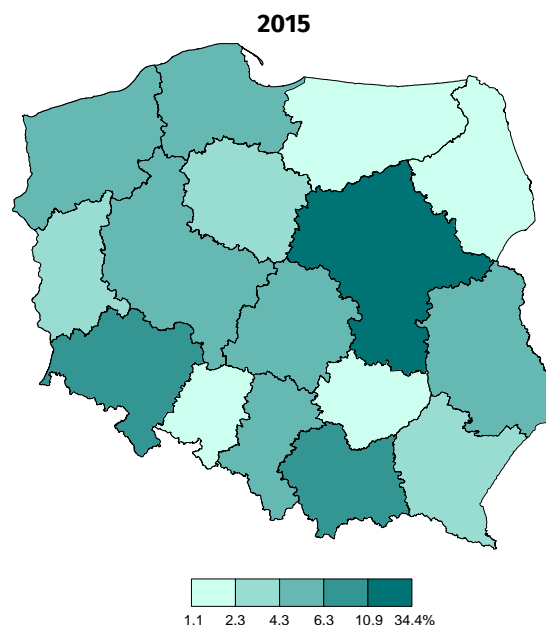
### 1.1. Liczebność cudzoziemców

Według oszacowania punktowego zbiorowość cudzoziemców w wieku 18 lat i więcej przebywających w Polsce w końcu 2015 r. liczyła 507,7 tys., a w końcu 2016 r. – 743,7 tys. osób. W kategoriach estymacji przedziałowej, przy założeniu 95% poziomu ufności, liczbę pełnoletnich cudzoziemców w Polsce na koniec 2015 r. pokrywa przedział od 369,1 tys. do 724,4 tys., zaś na koniec 2016 r. od 600,8 tys. do 943,1 tys.

### 1.2. Rozmieszczenie cudzoziemców według województw

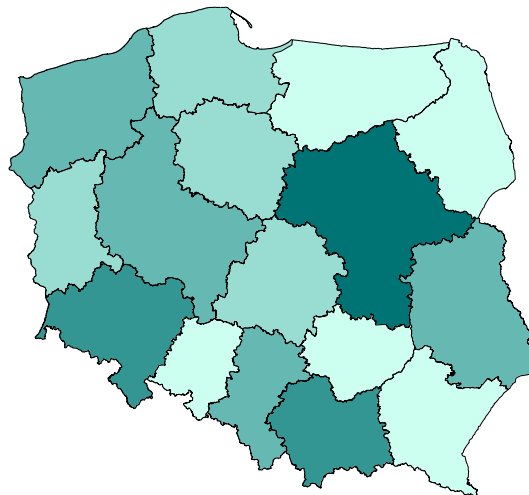
Rozmieszczenie cudzoziemców na terenie kraju nie jest równomierne. Zdecydowanie najwięcej przebywa ich i pracuje na terenie województwa mazowieckiego. W 2015 r. przebywało w nim prawie 33%, a w 2016 r. ponad 34% ogółu cudzoziemców. Co dziesiąty cudzoziemiec mieszkał na terenie województwa małopolskiego, a podobny odsetek odnotowano na terenie województwa dolnośląskiego. Czwarte miejsce w rankingu województw zajmuje woj. wielkopolskie (w 2015 r. – 6% a w 2016 r. – 6,7%). Najmniej cudzoziemców przebywa na terenie woj. świętokrzyskiego (odpowiednio: 1,1 oraz 1,2%). Rzadko wybierane jest również woj. warmińsko-mazurskie (1,3 i 1,4%) i podlaskie (1,7 i 1,6%).

**Kartogram 1. Cudzoziemcy w wieku 18 lat i więcej przebywający w Polsce w latach 2015 i 2016 według województw**





2016



Źródło: opracowanie własne

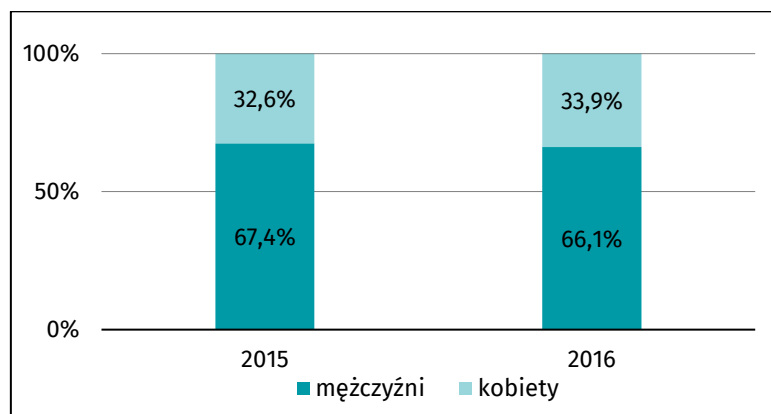
W badanych latach zaobserwowano bardzo niewielkie zmiany w rankingu województw ustalonym według odsetka cudzoziemców przebywających na terenie danego województwa z ogólnej liczby cudzoziemców, w 2016 r. w każdym województwie zwiększyła się jednak znacząco liczba cudzoziemców w stosunku do roku poprzedniego.

### 1.3. Charakterystyka cudzoziemców

#### 1.3.1. Płeć i wiek

Wśród cudzoziemców – zarówno w 2015 jak i 2016 r. dominowali mężczyźni. Oszacowano ich liczebność odpowiednio na 342,2 tys. (95% przedziału ufności: 245,9 tys. – 488,7 tys.) i 491,7 tys. (95% przedziału ufności: 393,6 tys. – 627,8 tys.). Mężczyźni stanowili zatem odpowiednio 67,4 i 66,1% ogółu szacowanej populacji.

**Wykres 1.1. Cudzoziemcy w wieku 18 lat i więcej przebywający w Polsce w latach 2015 i 2016 według płci**



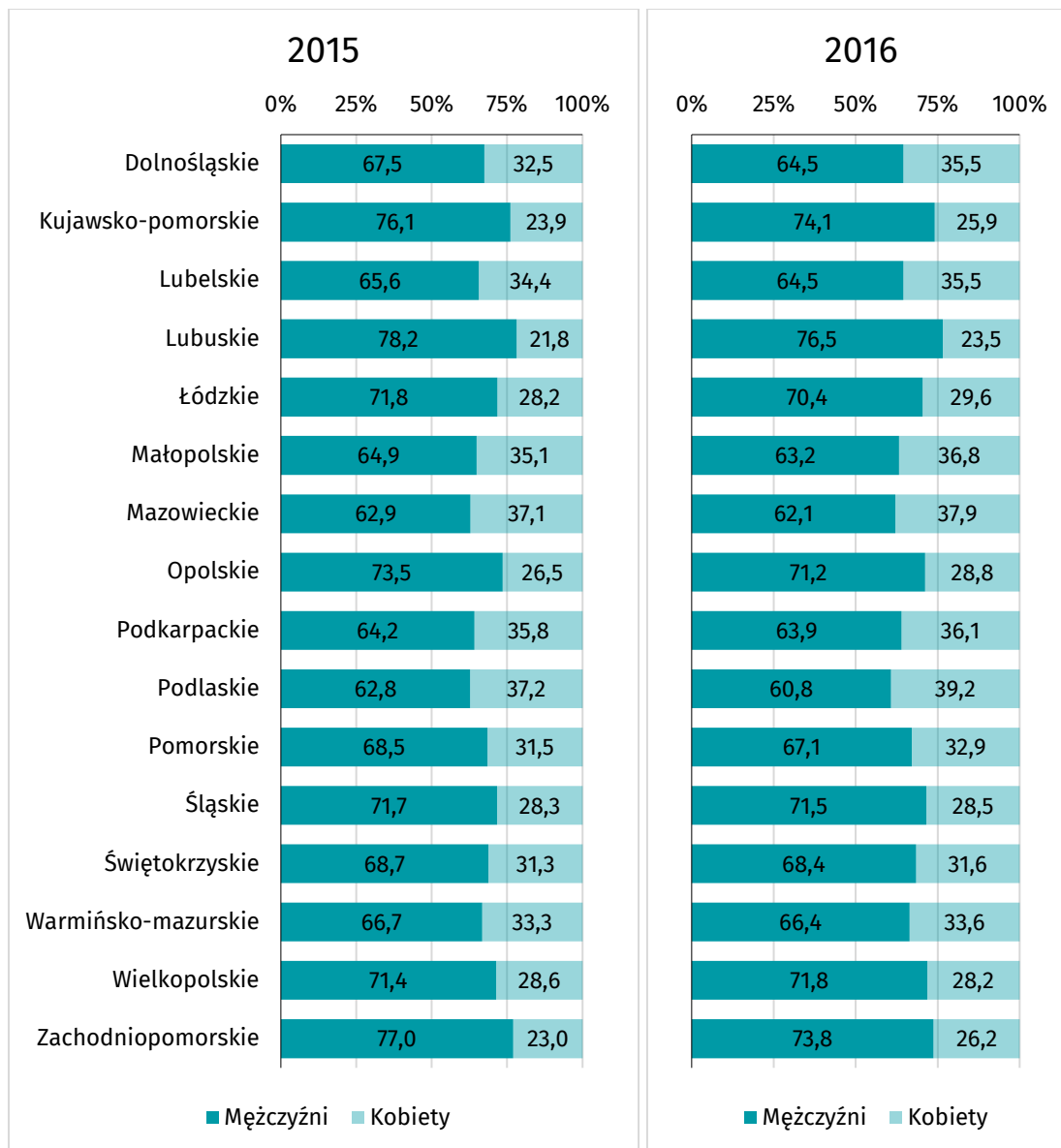
Źródło: opracowanie własne

Mężczyźni przeważali we wszystkich województwach, chociaż wielkość ich odsetka była zróżnicowana i w 2016 r. wahała się od 60,8% do 76,5% ( w 2015 r. od 62,8% do 78,2%).

Największa przewaga mężczyzn występowała w woj. lubuskim, najmniejsza w woj. podlaskim. W województwach wschodniej i południowo-wschodniej Polski (podlaskie, podkarpackie, małopolskie, lubelskie, warmińsko-mazurskie) zaobserwowano większy udział kobiet niż w województwach Polski zachodniej (lubuskie, zachodniopomorskie). W woj. mazowieckim, w którym przebywa najwięcej

cudzoziemców i są oni najbardziej zróżnicowani pod względem obywatelstwa, odsetek mężczyzn wyniósł w 2016 r. 62,1% (w 2015 r. utrzymywał się na podobnym poziomie – wynosił prawie 63%).

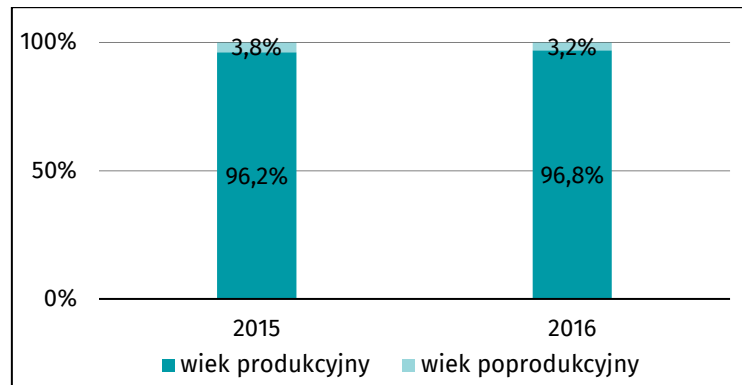
**Wykres 1.2. Cudzoziemcy w wieku 18 lat i więcej przebywający w Polsce w latach 2015 i 2016 według województw i płci**



Źródło: opracowanie własne

Zdecydowanie przeważają osoby w wieku produkcyjnym, ponieważ do Polski imigrują ludzie stosunkowo młodzi w celach zarobkowych lub podjęcia studiów. Ludność w wieku produkcyjnym oszacowano dla 2015 r. na 488,4 tys. (95% przedziału ufności: 354,5 tys. – 699,0 tys.), dla 2016 r. – na 720,2 tys. (95% przedziału ufności: 580,3 tys. – 915,1 tys.), co w 2015 r. stanowiło 96,2%, a w 2016 r. – 96,8% całej badanej populacji pełnoletnich cudzoziemców. W ciągu 2016 r. nastąpił zatem znaczący wzrost liczby cudzoziemców w wieku produkcyjnym, zwiększając tym samym zasoby pracy w Polsce.

**Wykres 1.3. Cudzoziemcy w wieku 18 lat i więcej przebywający w Polsce w latach 2015 i 2016 według wieku**



Źródło: opracowanie własne

We wszystkich województwach wśród cudzoziemców w wieku 18 lat i więcej odsetek osób w wieku poprodukcyjnym był niewielki – w 2016 r. wahał się od 2,2% (w woj. małopolskim) do 5,7% (w woj. świętokrzyskim). W 2015 r. w każdym z województw odsetek osób w wieku poprodukcyjnym był nieco wyższy (wahał się od 2,9% do 7,5%).

**Tabela 1. Cudzoziemcy w wieku 18 lat i więcej przebywający w Polsce w latach 2015 i 2016 według województw i grup wieku (w %)**

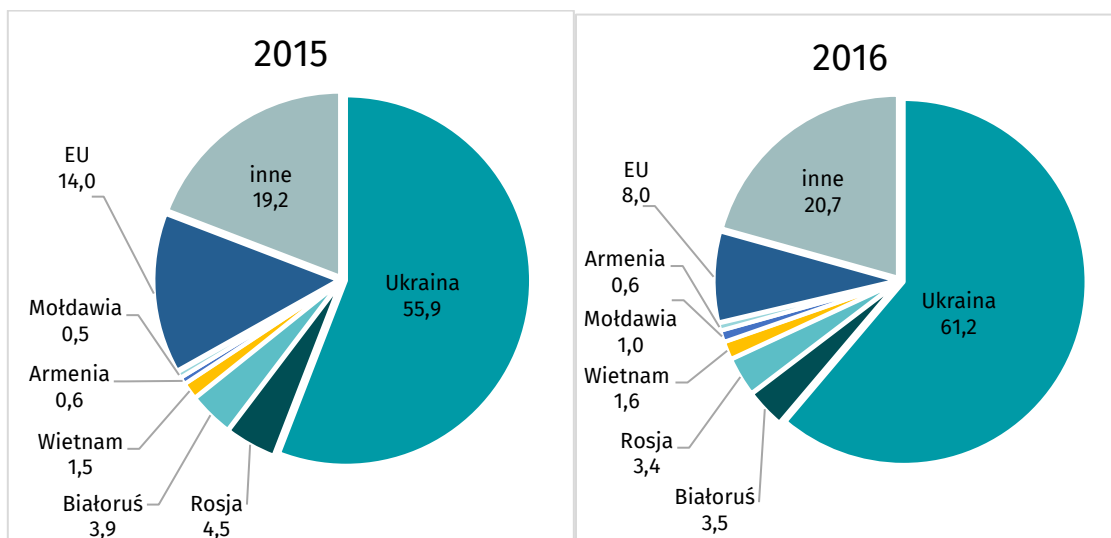
Województwa	2015		2016	
	produkcyjny	poprodukcyjny	produkcyjny	poprodukcyjny
Dolnośląskie	96,6	3,4	97,4	2,6
Kujawsko-pomorskie	95,6	4,4	96,5	3,5
Lubelskie	96,8	3,2	97,4	2,6
Lubuskie	96,2	3,8	97,0	3,0
Łódzkie	96,8	3,2	96,9	3,1
Małopolskie	97,1	2,9	97,8	2,2
Mazowieckie	96,6	3,4	97,0	3,0
Opolskie	96,5	3,5	96,8	3,2
Podkarpackie	94,7	5,3	95,3	4,7
Podlaskie	94,2	5,8	95,0	5,0
Pomorskie	95,2	4,8	95,7	4,3
Śląskie	95,9	4,1	96,7	3,3
Świętokrzyskie	92,5	7,5	94,3	5,7
Warmińsko-mazurskie	94,1	5,9	95,2	4,8
Wielkopolskie	95,2	4,8	96,7	3,3
Zachodniopomorskie	95,3	4,7	95,9	4,1

Źródło: opracowanie własne

### 1.3.2. Kraj obywatelstwa

Wśród cudzoziemców przebywających w Polsce zdecydowanie przeważają obywatele krajów trzecich<sup>3</sup>. Wynika to z faktu, że polski rynek pracy jest atrakcyjny dla cudzoziemców z naszej wschodniej granicy, z jednej strony z powodu bliskości geograficznej, sieci migracyjnych, które pozwalają zminimalizować koszty pobytu przynajmniej w pierwszych tygodniach, zdecydowanie wyższych zarobków niż w krajach rodzimych, z drugiej – liberalizacji zasad dostępu obywateli do polskiego rynku pracy. Uregulowania prawne wprowadzające uproszczoną procedurę zezwalają na podejmowanie pracy (oświadczenia pracodawców o powierzeniu pracy cudzoziemcowi) przez obywateli sześciu krajów trzecich: Armenii, Białorusi, Gruzji, Mołdawii, Rosji i Ukrainy. Spośród krajów trzecich to właśnie obywatele Ukrainy stanowią największą zbiorowość. Szacuje się, że w 2015 r. przebywało 283,7 tys. (95% przedział ufności: 203,9 tys. – 415,7 tys.), a w 2016 r. 455,0 tys. (95% przedział ufności: 361,5 tys. – 584,7 tys.) obywateli tego kraju.

**Wykres 1.4. Cudzoziemcy w wieku 18 lat i więcej przebywający w Polsce w latach 2015 i 2016 według kraju obywatelstwa (w %)**



Źródło: opracowanie własne

Zasób cudzoziemców przebywających w poszczególnych województwach jest zróżnicowany pod względem kraju obywatelstwa. Wprawdzie we wszystkich województwach – poza woj. podlaskim – przeważają obywatele Ukrainy, ale ich odsetek jest różny. Największy jest w woj. opolskim (w 2015 r. prawie 79%, w 2016 r. ponad 83%), a następnie w woj. podkarpackim (odpowiednio: ponad 72% i prawie 77%) oraz w woj. lubelskim (71,5% i 72,5%). W woj. podlaskim odsetek obywateli Ukrainy jest zdecydowanie mniejszy (w 2015 r. 27%, w 2016 r. prawie 32%), ponieważ w tym województwie szczególnie widoczni są – zapewne z powodu sąsiedztwa – obywatele Białorusi (w 2015 r. ponad 28%, w 2016 r. ponad 31%), a także Rosjanie.

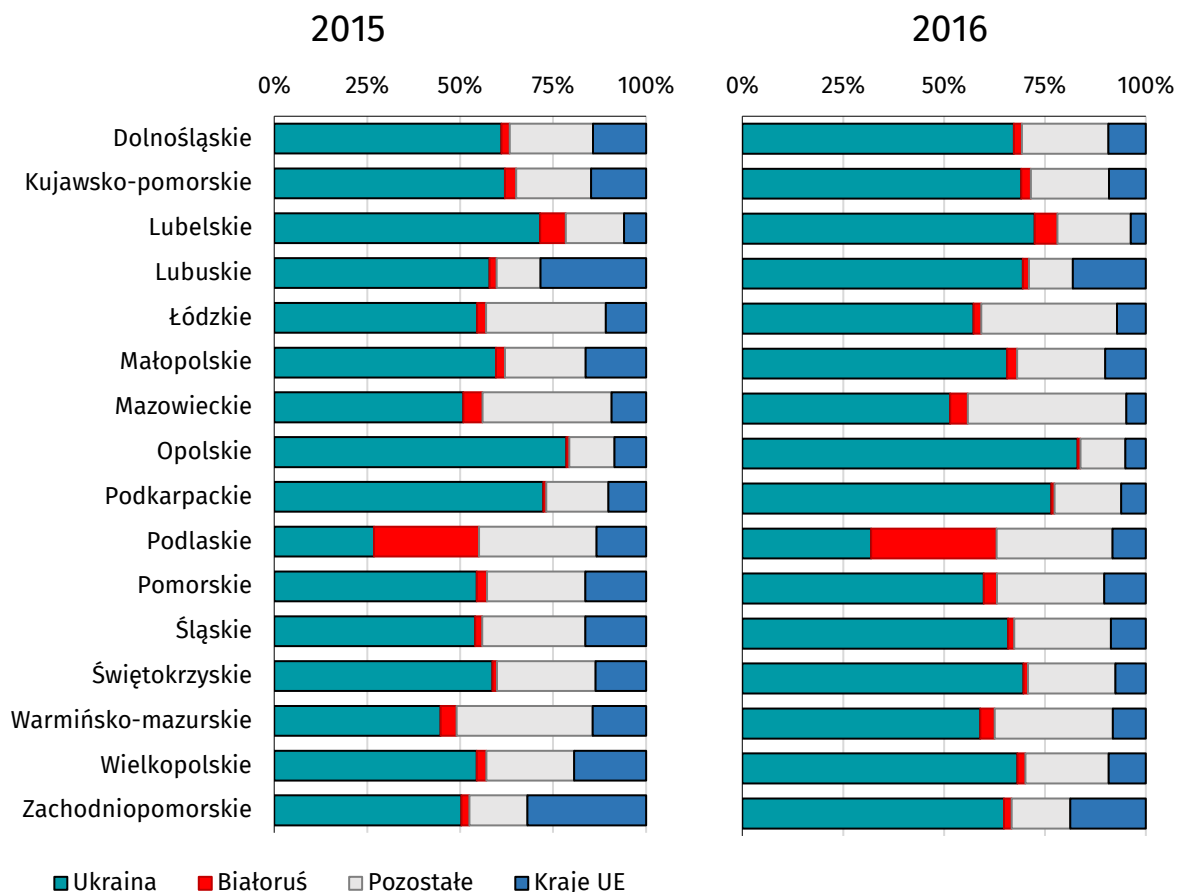
Najbardziej zróżnicowana grupa cudzoziemców przebywa na terenie woj. mazowieckiego. Tu największy jest odsetek cudzoziemców z obywatelstwem innym niż te, które zostały wyodrębnione w niniejszej pracy badawczej. Podczas, gdy dla Polski obywatelstwa inne stanowiły w 2016 r. prawie 21% (patrz wykres 1.4), to w woj. mazowieckim odsetek ten był o 10 punktów proc. większy. W porównaniu z innym województwami największy jest też odsetek Wietnamczyków (w 2015 r. 3,5%, w 2016 r. 3,8% ogółu cudzoziemców przebywających w tym województwie).

<sup>3</sup> Według rozporządzenia (WE) nr 862/2007 PE i Rady z dn. 11 lipca 2007 r. ws. statystyk wspólnoty z zakresu migracji i ochrony międzynarodowej „obywatel kraju trzeciego” oznacza każdą osobę, która nie jest obywatelem Unii Europejskiej w rozumieniu art. 17 ust. 1 Traktatu, w tym bezpaństwowców.



W strukturze obywatelstw, cudzoziemcy z krajów UE najbardziej widoczni są w woj. zachodniopomorskim i lubuskim, czyli województwach sąsiadujących z Niemcami (w 2016 r. stanowili powyżej 18% ogółu cudzoziemców zamieszkujących na tych terenach). Najwięcej jednak – w liczbach bezwzględnych – obywateli UE przebywa na terenie woj. mazowieckiego. Należy zaznaczyć, że ze względu na znaczną przewagę wielkości zasobu cudzoziemców przebywających w tym województwie – w porównaniu z ogólną liczbą cudzoziemców w innych województwach – również obywatele innych badanych krajów są na tym terenie najliczniejsi.

**Wykres 1.5. Cudzoziemcy w wieku 18 lat i więcej przebywający w Polsce w latach 2015 i 2016 według województw i kraju obywatelstwa (w %)**



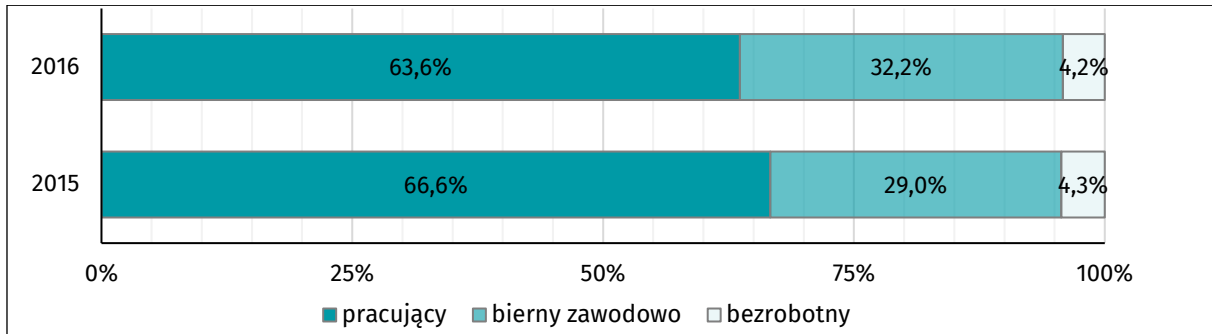
Źródło: opracowanie własne

### 1.3.3. Status na rynku pracy

Większość cudzoziemców przebywających w Polsce pracuje. W 2015 r. pracowało 2/3 ogółu, w 2016 r. – odsetek pracujących był niewiele mniejszy. Mimo podobnych udziałów procentowych, liczba pracujących w 2016 r. była znacznie wyższa niż w roku poprzednim. Szacuje się, że w 2015 r. pracowało w Polsce przeszło 338 tys., a w 2016 r. ponad 473 tys. cudzoziemców.

Ustalenie statusu na rynku pracy było dużym wyzwaniem, ponieważ w rejestrach przyjętych do oszacowania populacji pozarejestrowej nie ma informacji pozwalającej rozszacować populację cudzoziemców według tej zmiennej. W efekcie przyjęto strukturę według BAEL, która jest bardziej odpowiednia dla cudzoziemców przebywających przez okres co najmniej 12 miesięcy, w tym dla cudzoziemców mieszkających w Polsce na stałe. Wśród tej grupy jest więcej osób biernych zawodowo, część z nich jest w wieku poprodukcyjnym.

**Wykres 1.6. Cudzoziemcy w wieku 18 lat i więcej przebywający w Polsce w latach 2015 i 2016 według statusu na rynku pracy**



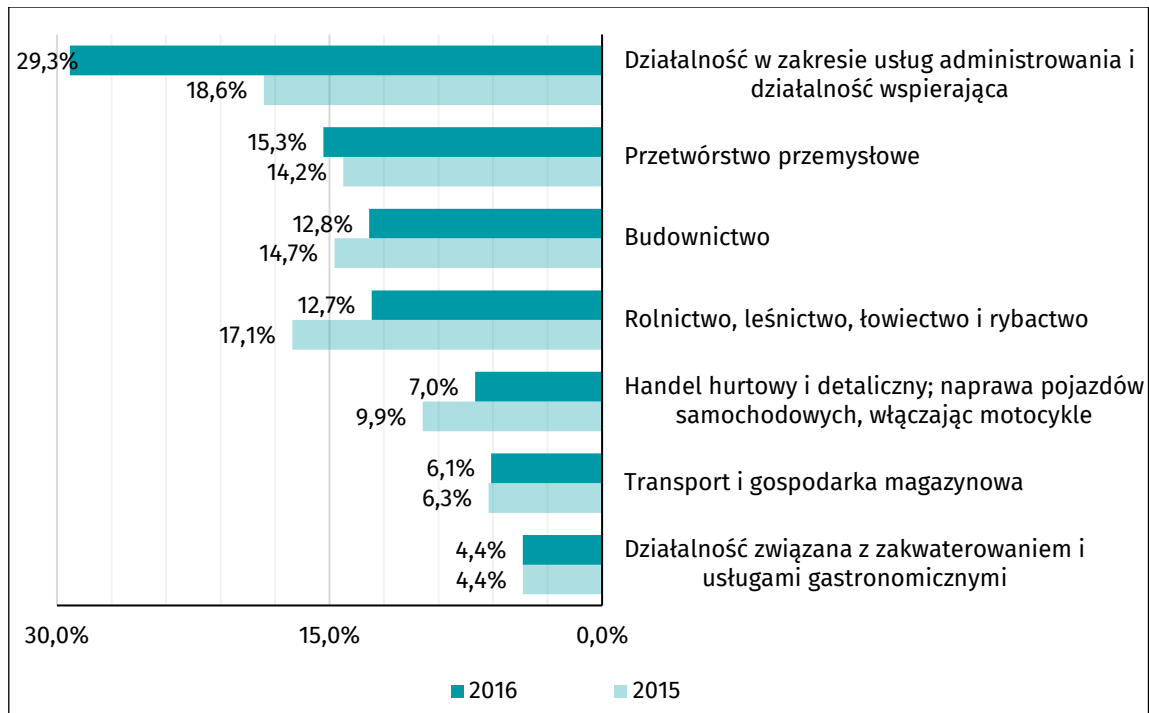
Źródło: opracowanie własne

Wśród cudzoziemców przebywających w Polsce do 12 miesięcy odsetek pracujących jest prawdopodobnie większy, ponieważ imigracja o charakterze krótkookresowym ma w ostatnich latach charakter głównie zarobkowy. Bardziej miarodajna struktura statusu na rynku pracy zostanie ustalona w trakcie doskonalenia metody szacowania populacji cudzoziemców.

#### 1.3.4. Sekcje PKD

Struktura sekcji PKD została ustalona na podstawie informacji zawartych w rejestrach MRPiS dotyczących zezwoleń na pracę cudzoziemców w Polsce oraz oświadczeń pracodawców o powierzenie pracy cudzoziemcowi. Z informacji tych wynika znacząca rola sekcji „Działalność w zakresie usług administrowania i działalność wspierająca”. Do sekcji tej należą agencje pracy tymczasowej, które często pośredniczą w zatrudnieniu cudzoziemców. Oznacza to, że większość cudzoziemców, którym przypisano tę sekcję nie pracuje w niej, ale u innych pracodawców – w innych sekcjach.

**Wykres 1.7. Cudzoziemcy w wieku 18 lat i więcej przebywający w Polsce w latach 2015 i 2016 według sekcji PKD**



Źródło: opracowanie własne

Spośród innych sekcji w 2015 r. cudzoziemcy najczęściej pracowali w sekcji „Rolnictwo, leśnictwo, łowiectwo i rybactwo”, ale w 2016 r. sekcja ta przesunęła się na dalsze miejsce. Wyżej w rankingu znalazło się „Przetwórstwo przemysłowe” i „Budownictwo”. Należy jednak podkreślić, że te trzy sekcje są najpopularniejsze wśród cudzoziemców.

Ponadto, cudzoziemcy pracują w sekcjach: „Handel hurtowy i detaliczny; naprawa pojazdów samochodowych, włączając motocykle”, „Transport i gospodarka magazynowa” i „Działalność związana z zakwaterowaniem i usługami gastronomicznymi”.

### 1.3.5. Rozmieszczenie terytorialne cudzoziemców pracujących w Polsce

Najwięcej cudzoziemców pracowało – zarówno w 2015, jak i 2016 r. – w woj. mazowieckim, jednakże odsetek cudzoziemców pracujących w tym województwie w ogólnej liczbie cudzoziemców pracujących w Polsce zmniejszył się w ciągu 2016 r. o prawie 6 punktów proc. Oznacza to, że rynki pracy innych województw coraz częściej potrzebują i stają się atrakcyjne dla cudzoziemców.

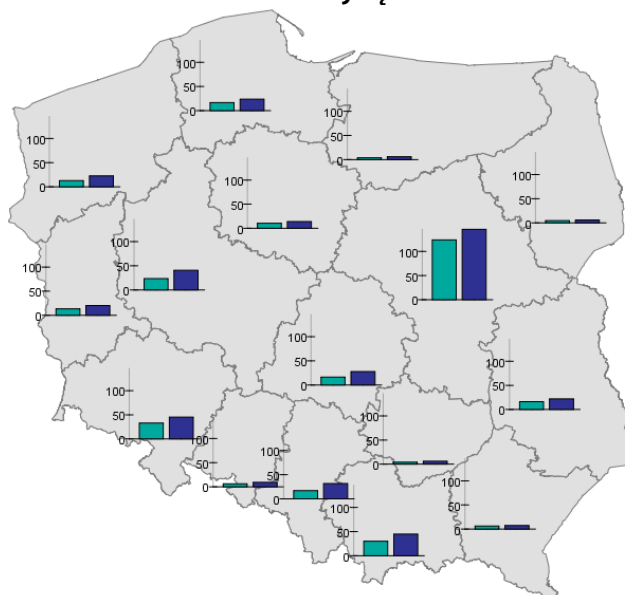
W 2016 r. liczba cudzoziemców pracujących w każdym województwie – podobnie jak ogólna liczba cudzoziemców – wzrosła w stosunku do 2015 r. Największy wzrost (prawie o 86%) odnotowano w woj. śląskim. Znaczna dynamika wzrostu wystąpiła również w woj. zachodniopomorskim, wielkopolskim i łódzkim. Najmniejszy procentowy wzrost odnotowano w woj. mazowieckim (17,5%), mimo iż wzrost liczby cudzoziemców, co do bezwzględnej wartości, był największy – na rynku pracy woj. mazowieckiego przybyło prawie 22 tys. cudzoziemców. W woj. śląskim, gdzie procentowy wzrost był największy, pracowało o około 15 tys. cudzoziemców więcej niż w roku poprzednim.

Rozszacowanie ogólnej liczby pracujących cudzoziemców na poszczególne województwa dokonano na podstawie struktury, która powstała jako wypadkowa z uśrednienia dwóch rozkładów według województw, tj. tego, jaki uzyskano w ramach oszacowania zasobu cudzoziemców oraz tego, jaki ustalono w zbiorze połączonych danych o zezwoleniach na pracę cudzoziemców w Polsce i oświadczeniach pracodawców o powierzenie pracy cudzoziemcom, sporządzonym na podstawie rejestrów MRPiPS.

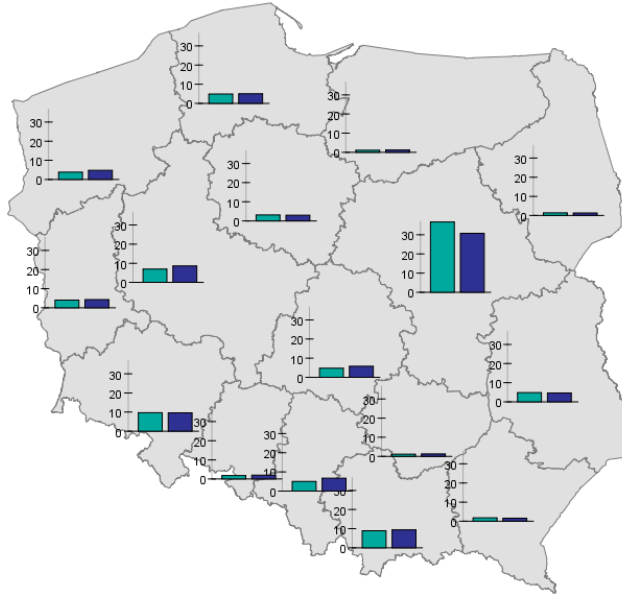
Ponieważ – jak zaznaczono wcześniej – obecnie nie ma miarodajnego i bezpośredniego źródła danych, które dostarczałoby wiarygodnych informacji o cudzoziemcach pracujących w Polsce, liczba ta została wyszacowana. Wyszacowane również zostały liczby pracujących na terenie poszczególnych województw – stąd przytoczone powyżej dane należy traktować jako przybliżone.

**Kartogram 2. Cudzoziemcy w wieku 18 lat i więcej pracujący w Polsce w latach 2015 i 2016 według województw**

**A. W tysiącach**



### B. W % ogółu cudzoziemców pracujących w Polsce



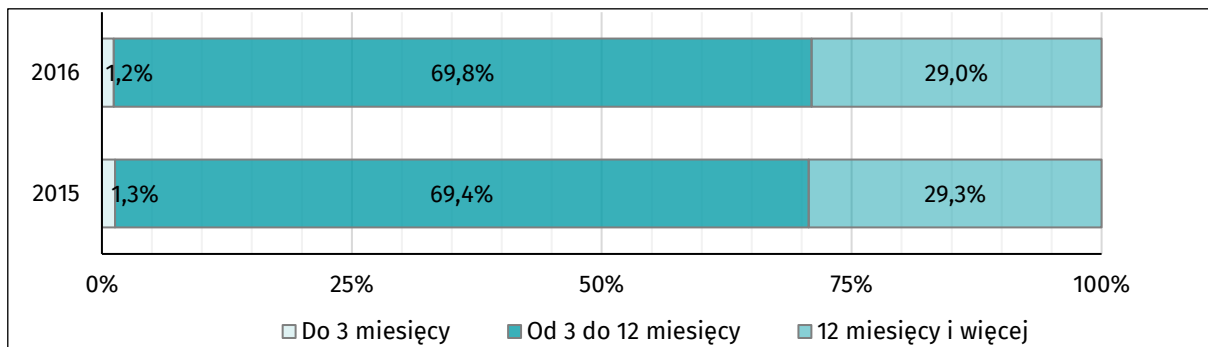
Źródło: opracowanie własne

#### 1.3.6. Okres przebywania w Polsce

Zdecydowana większość (ponad 71%) cudzoziemców przebywa w Polsce krócej niż rok. Najliczniejsza jest grupa, która przebywa od 3 do 12 miesięcy (ok. 70%).

Szacuje się, że w 2015 r. prawie 143 tys. cudzoziemców, a w 2016 r. prawie 209 tys. przebywało w Polsce co najmniej rok.

**Wykres 1.8. Cudzoziemcy w wieku 18 lat i więcej przebywający w Polsce w latach 2015 i 2016 według okresu przebywania w Polsce**



Źródło: opracowanie własne

Powyższa struktura została ustalona na podstawie informacji MRPIPS dotyczącej zezwoleń na pracę cudzoziemców i oświadczeń pracodawców o powierzeniu pracy cudzoziemcowi, zatem stosunkowo dobrze obrazuje długość przebywania w Polsce cudzoziemców pracujących.

#### 1.3.7. Poziom wykształcenia

Rejestry nie dostarczają informacji na temat wykształcenia cudzoziemców. Do rozszacowania badanej populacji cudzoziemców według poziomów wykształcenia wykorzystano odpowiednią strukturę uzyskaną na podstawie wyników NSP 2011. Założono, że struktura wykształcenia nie uległa w ostatnich latach znaczącym zmianom.

Wśród cudzoziemców dominują osoby z wykształceniem średnim (ponad 43%). Znaczącą grupę stanowią również osoby posiadające wykształcenie wyższe (ponad 36%). Tylko co piąta osoba ma wykształcenie niższe niż średnie. Cudzoziemcy są zatem stosunkowo dobrze wykształceni i stanowią cenny kapitał z punktu widzenia rynku pracy w Polsce.

Szacuje się, że w 2016 r. przebywało w Polsce prawie 590 tys. cudzoziemców z wykształceniem co najmniej średnim, w tym prawie 270 tys. z wykształceniem wyższym.

**Wykres 1.9. Cudzoziemcy w wieku 18 lat i więcej przebywający w Polsce w latach 2015 i 2016 według poziomu wykształcenia**

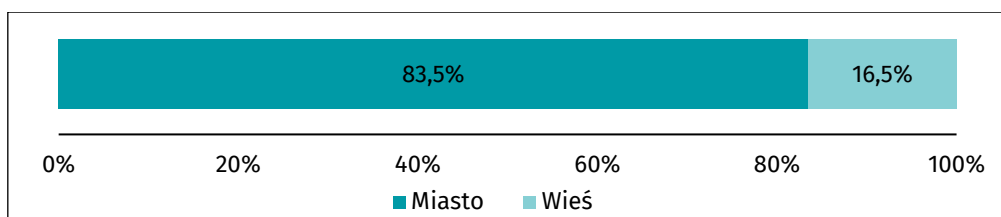


Źródło: opracowanie własne

### 1.3.8. Charakter miejsca zamieszkania

Zdecydowana większość cudzoziemców mieszka w miastach. Struktura miejsca zamieszkania została ustalona na podstawie wyników z NSP 2011. W trakcie doskonalenia metody szacunku struktura ta zostanie uaktualniona i ustalona bardziej miarodajnie.

**Wykres 1.10. Cudzoziemcy w wieku 18 lat i więcej przebywający w Polsce w latach 2015 i 2016 według charakteru miejsca zamieszkania (miasto/wieś)**



Źródło: opracowanie własne

## 1.4. Cudzoziemcy w podregionach

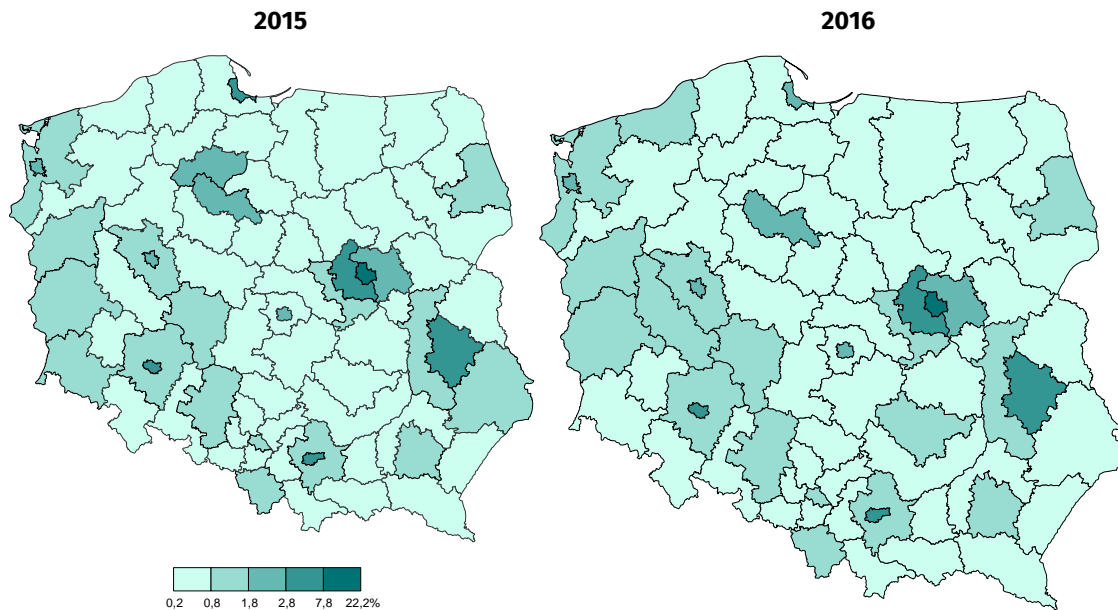
Należy podkreślić, że do oszacowania wielkości populacji cudzoziemców na poziomie podregionów wykorzystano inny model niż ten dla pozostałych zmiennych, o czym szerzej w dalszej części raportu oraz załączniku 3. W rezultacie, uzyskane wyniki dla podregionów i pozostałych cech nie są w pełni zgodne – w tym również na poziomie Polski ogółem – co można zaobserwować zestawiając odpowiednie dane zaprezentowane w Załączniku 1 – „Wyniki pracy badawczej”.

Niezależnie od powyższego, wypada zauważyć, że miasto Warszawa jest najczęściej wybieranym przez przybyszów z zagranicy podregionem, na który w 2015 r. przypadało niespełna 20,5%, a w 2016 – przeszło 22% z ogólnej liczby cudzoziemców w wieku 18 lat i więcej. Wysoko plasują się również podregiony

otaczające Warszawę – warszawski zachodni (odpowiednio: 5,7 i 6,4%) i warszawski wschodni (2,6% i 2,7%). Cudzoziemcy najczęściej przebywają w miastach. Poza Warszawą znacząca liczba cudzoziemców przebywa w Krakowie (7,7 i 7,0%), we Wrocławiu (5,7 i 5,8%), w podregionie trójmiejskim (3,2 i 2,6%), Szczecinie (2,5 i 2,4%), Łodzi (2,1 i 2,2%) i Poznaniu (2,2 i 2,1%). Dość wyraźnie odznacza się na tle kraju również podregion lubelski (3,0 i 3,3%).

Niewielu cudzoziemców przebywa na terenie podregionów: wrocławskiego, chojnickiego, suwalskiego, krośnieńskiego i grudziądzkiego. W każdym z tych podregionów przebywa po ok. 0,2% ogółu cudzoziemców.

**Kartogram 3. Cudzoziemcy w wieku 18 lat i więcej przebywający w Polsce w latach 2015 i 2016 według podregionów (NUTS 2016)**



Źródło: opracowanie własne

## 2. Metodologia badania

### *Survey methodology*

Badanie realizowano w dwóch etapach. W pierwszym etapie dokonano analizy aktów prawnych regulujących pobyt i zatrudnienie cudzoziemców na terytorium RP, zinwentaryzowano źródła danych i oceniono ich przydatność dla realizacji projektu, dokonano przeglądu literatury z zakresu estymacji charakterystyk populacji trudnych do zbadania oraz statystycznej integracji danych, zaproponowano metodę szacowania typu *capture-recapture* – bazującą na modelach log-liniowych oraz bezpośredniego wyznaczania populacji w celu ich przetestowania w drugim etapie prac, a także określono wstępną listę wskaźników możliwych do oszacowania na poziomie podregionów (wg NUTS 2013 i 2016).

W ramach drugiego etapu prac określono finalną metodę szacunku, testowo naliczono zestaw wskaźników zaprojektowanych w pierwszym etapie. Przeprowadzono następnie weryfikację poszczególnych wskaźników i dokonano wyboru ostatecznego zestawu, w którym uwzględniono również niezbędne modyfikacje w stosunku do pierwotnej listy wskaźników. Skontrolowano poprawność i kompletność pozyskanych danych, dokonano analizy danych i przedstawiono jej wyniki oraz przygotowano raport końcowy.

Poniżej przedstawiono zakres czasowy, podmiotowy, przedmiotowy i terytorialny pracy badawczej, definicje i klasyfikacje oraz listy kodowe:

### 2.1. Zakres podmiotowy badania

Zakresem podmiotowym badania jest zasób cudzoziemców przebywających w Polsce czasowo według stanu na koniec 2015 i 2016 roku.

Ustalenia zakresu podmiotowego badania (kluczowego dla wyznaczania charakteryzowanej zbiorowości) w ramach analizowanych i łączonych zbiorów danych, dokonywano na podstawie wartości odpowiednich cech, a także wiedzy i doświadczenia ekspertów zespołu badawczego wykonujących prace oparte na źródłach administracyjnych.

### 2.2. Zakres przedmiotowy badania

Zakres przedmiotowy badania przygotowany został na podstawie cech dostępnych w źródłach administracyjnych i badaniach statystycznych.

W zakresie przedmiotowym badania założono uwzględnienie następujących cech cudzoziemców:

- płeć,
- ekonomiczne grupy wieku (przedprodukcyjny, produkcyjny, poprodukcyjny),
- poziom wykształcenia (wyższe, średnie, zawodowe, podstawowe, bez wykształcenia),
- kraj poprzedniego zamieszkania (podstawowe grupowania: kraje UE, EFTA, kraje trzecie – znaczące kraje imigracji (w tym: Ukraina, Białoruś, Mołdawia, Rosja, Armenia, Wietnam),
- kraj obywatelstwa: kraje UE, EFTA, kraje trzecie – znaczące kraje imigracji (w tym: Ukraina, Białoruś, Mołdawia, Rosja, Armenia, Wietnam),
- status na runku pracy (pracujący, bezrobotny, bierny zawodowo),
- okres pobytu czasowego.

Dodatkowo planowano rozpoznanie możliwości opracowania metodologii i naliczenia wskaźników opisanych:

- w podziale na miasto, wieś,
- z uwzględnieniem sekcji PKD dla pracujących,
- z uwzględnieniem cudzoziemców przebywających w Polsce do 3 miesięcy.

## 2.3. Zakres terytorialny

Objęmuje: kraj, województwo, podregion (miasto i wieś)

## 2.4. Przyjęte definicje i listy kodowe

Na potrzeby realizacji projektu przyjęto definicję, która determinuje określenie badanej zbiorowości:

**Cudzoziemiec** – osoba nieposiadająca obywatelstwa polskiego lub bezpieczeństwa (podstawa prawna – ustawa z dnia 12 grudnia 2013 r. o cudzoziemcach, Dz.U. z 2016 r. poz. 1990 z późn. zm.).

**Przyjęto następujące listy kodowe, porządkujące charakterystykę cudzoziemców:**

### Kryterium czasu przebywania w Polsce:

- Do 3 miesięcy,
- Od 3-12 miesięcy,
- 12 miesięcy i dłużej.

### Ekonomiczne grupy wieku:

- Przedprodukcyjny (0-17 lat),
- Produkcyjny (kobiety 18-59 lat, mężczyźni 18-64 lata),
- Poprodukcyjny (kobiety 60 lat i więcej, mężczyźni 65 lat i więcej).

### Poziom wykształcenia:

- Wyższe,
- Średnie,
- Zawodowe,
- Podstawowe,
- Bez wykształcenia.

### Kraj obywatelstwa:

- Kraje UE (28 krajów) – ogółem,
- EFTA (ogółem),
- Kraje trzecie (ogółem) – w tym: Ukraina, Białoruś, Gruzja, Mołdawia, Rosja, Armenia, Wietnam.

**Kraje EFTA:** Islandia, Lichtenstein, Norwegia, Szwajcaria

### Status na rynku pracy:

- Pracujący,
- Bezrobotny,
- Bierny zawodowo.

### Sekcje PKD:

- Rolnictwo, leśnictwo, łowiectwo i rybactwo;
- Górnictwo i wydobywanie;
- Przetwórstwo przemysłowe;
- Wytwarzanie i zaopatrywanie w energię elektryczną, gaz, parę wodną, gorącą wodę i powietrze do układów klimatyzacyjnych;
- Dostawa wody; gospodarowanie ściekami i odpadami oraz działalność związana z rekultywacją;
- Budownictwo;
- Handel hurtowy i detaliczny; naprawa pojazdów samochodowych, włączając motocykle;
- Transport i gospodarka magazynowa;
- Działalność związana z zakwaterowaniem i usługami gastronomicznymi;
- Informacja i komunikacja;
- Działalność finansowa i ubezpieczeniowa;
- Działalność związana z obsługą rynku nieruchomości;
- Działalność profesjonalna, naukowa i techniczna;



- Działalność w zakresie usług administrowania i działalność wspierająca;
- Administracja publiczna i obrona narodowa; obowiązkowe zabezpieczenia społeczne;
- Edukacja;
- Opieka zdrowotna i pomoc społeczna;
- Działalność związana z kulturą, rozrywką i rekreacją;
- Pozostała działalność usługowa;
- Gospodarstwa domowe zatrudniające pracowników, gospodarstwa domowe produkujące wyroby i świadczące usługi na własne potrzeby;
- Organizacje i zespoły eksterytorialne.

## 2.5. Przepisy prawne regulujące kwestię pobytu i zatrudnienia cudzoziemców w Polsce

Analiza dostępnych w rejestrach i systemach informacyjnych danych o cudzoziemcach byłaby niemożliwa bez znajomości przepisów prawnych regulujących pobyt i zatrudnienie cudzoziemców na terytorium RP.

Poniżej przedstawiona została lista aktów prawnych w tym obszarze:

- ustawa z dnia 12 grudnia 2013 r. o cudzoziemcach (Dz.U. z 2016 r. poz. 1990 z późn. zm.);
- ustawa z dnia 20 kwietnia 2004 r. o promocji zatrudnienia i instytucjach rynku pracy (Dz.U. z 2017 r. poz. 1065 z późn. zm.);
- ustawa z dnia 13 czerwca 2003 r. o udzielaniu cudzoziemcom ochrony na terytorium Rzeczypospolitej polskiej (Dz.U. z 2016 r. poz. 1836 z późn. zm.);
- ustawa z dnia 14 lipca 2006 r. o wjeździe na terytorium Rzeczypospolitej Polskiej, pobytku oraz wyjeździe z tego terytorium obywateli państw członkowskich Unii Europejskiej i członków ich rodzin (Dz.U. z 2017 r. poz. 900);
- rozporządzenie Ministra Pracy i Polityki Społecznej z dnia 1 kwietnia 2015 r. w sprawie wydawania zezwolenia na pracę cudzoziemca (Dz.U. poz. 543);
- rozporządzenie Ministra Pracy i Polityki Społecznej z dnia 29 stycznia 2009 r. w sprawie określenia przypadków, w których zezwolenie na pracę cudzoziemca jest wydawane bez względu na szczegółowe warunki wydawania zezwoleń na pracę cudzoziemców (tj. Dz.U. z 2015 r. poz. 97);
- rozporządzenie Ministra Pracy i Polityki Społecznej z dnia 21 kwietnia 2015 r. w sprawie przypadków, w których powierzenie wykonywania pracy cudzoziemcowi na terytorium Rzeczypospolitej Polskiej jest dopuszczalne bez konieczności uzyskania zezwolenia na pracę (Dz.U. poz. 588);
- ustawa z dnia 15 czerwca 2012 r. o skutkach powierzania wykonywania pracy cudzoziemcom przebywającym wbrew przepisom na terytorium Polski (Dz.U. poz. 769).

## 2.6. Charakterystyka potencjalnych źródeł o cudzoziemcach

Cele pracy badawczej realizowane były z wykorzystaniem informacji pochodzących z zasobów informacyjnych statystyki publicznej (PBSSP – Program badań statystycznych statystyki publicznej – 2015, 2016), w szczególności dane administracyjne i statystyczne gromadzone w ramach badań: „Zasoby migracyjne w Polsce”, „Cudzoziemcy w Polsce. Legalizacja pobytu cudzoziemców na terytorium RP”, „Operat do Badań Społecznych”, „Charakterystyka demograficzno-społeczna i ekonomiczna gospodarstw domowych i rodzin”, „Badanie Aktywności Ekonomicznej Ludności”.

W toku prac przeanalizowano zasoby informacyjne Głównego Urzędu Statystycznego oraz zasoby innych jednostek administracji publicznej. Wytypowane do wykorzystania w pracy badawczej zbiory danych musiały spełniać następujące warunki dotyczące zawartych w nich informacji:

- dostępność,
- odpowiednio wysoka jakość,
- pokrycie badanej populacji,
- aktualność przypadająca na 2015 r. i 2016 r. (tj. najbardziej aktualne zbiory w momencie rozpoczęcia prac), istotna z punktu widzenia sporządzenia, po zakończeniu prac, rekomendacji i oceny możliwości wdrożenia wyników badania do produkcji statystycznej.

Ocenę jakości przeprowadzono dla zbiorów administracyjnych. Sformułowana została w tym celu następująca lista wskaźników jakościowych:

1. Pokrycie podmiotowe (liczba osób z obywatelstwem innym niż polskie oraz nieustalone obywatelstwo),
2. Pokrycie przedmiotowe (kompletność cech: stopień wypełnienia pól dla cechy  $x$ = liczba rekordów, w których istnieje wartość cechy/liczba rekordów, w których wartość powinna się znajdować\*100%, w tym dla wypełnień słownikowych),
3. Zgodność czasowa – aktualność: stan na dzień ...,
4. Zgodność z przyjętymi definicjami i klasyfikacjami (Zgodność z definicją cudzoziemca: (0-2), gdzie 0 – oznacza brak zgodności w stosowaniu definicji cudzoziemca, 2 – zgodna),
5. Stosowanie TERYT (0-1-2, gdzie 2 – oznacza stosowanie identyfikatorów jednostek podziału administracyjnego, identyfikatorów miejscowości, identyfikatorów adresowych ulic, nieruchomości, budynków i mieszkań, 1 – wyłącznie identyfikatory jednostek podziału administracyjnego, 0 – brak),
6. Zgodność ze słownikami krajów (0-2, gdzie 0 – oznacza brak zgodności, 2 – zgodna),
7. Stosowanie w zbiorach unikatowego identyfikatora rekordu (osób) nr PESEL (0-2, gdzie 0 oznacza brak stosowania, 2 – jest stosowany).

Przy wyborze źródeł administracyjnych brano także pod uwagę potencjalną metodę szacunku i założenia metodologiczne; jakie w niej przyjęto, tj. takie same we wszystkich źródłach definicje populacji cudzoziemców, niezależność źródeł danych (w sensie prawnym), zamkniętą populację, zapisanie każdej jednostki w każdym ze źródeł (przy czym prawdopodobieństwo nie musi być jednakowe), zidentyfikowanie i połączenie bez błędów każdej jednostki za pomocą nr PESEL.

Do oszacowania zasobu cudzoziemców na rynku pracy w Polsce w 2015 i 2016 r. wykorzystano metodę *capture-recapture* – bazującą na modelu log-liniowym. Zastosowano dwa modele – jeden dla podregionów, drugi dla pozostałych zmiennych. Modele log-liniowe wykorzystane w niniejszej pracy badawczej zostały opisane w Załączniku 2.

Po dokonaniu analizy i oceny źródeł danych jako główne źródła administracyjne wykorzystywane w modelu ostatecznie wskazano zasoby informacyjne:

- Urzędu do Spraw Cudzoziemców system „Pobyt” – zbiór rejestrów, ewidencji i wykazu w sprawach cudzoziemców,
- Ministerstwa Rodziny, Pracy i Polityki Społecznej – w zakresie zezwoleń na pracę i oświadczeń pracodawców o powierzeniu pracy cudzoziemcom,
- Ministerstwa Cyfryzacji – rejestr PESEL – w zakresie cudzoziemców zameldowanych na pobyt stały lub czasowy,
- Zakładu Ubezpieczeń Społecznych – Centralny Rejestr Ubezpieczonych w zakresie ubezpieczonych cudzoziemców oraz członków ich rodzin,
- Centralny Rejestr Podmiotów – Krajowa Ewidencja Podatników (CRP KEP) w zakresie danych o osobach fizycznych – cudzoziemcach.

Natomiast spośród badań statystycznych wskazano zbiory NSP 2011 (Narodowy Spis Powszechny Ludności i Mieszkań) dotyczące cudzoziemców oraz BAEL (Badanie Aktywności Ekonomicznej Ludności). Dane te zostały wykorzystane w projekcie badawczym jako źródło uzupełniające w zakresie cech

niedostępnych w zbiorach jednostkowych, tj. do opracowania charakterystyki społeczno-ekonomicznej cudzoziemców.

Przeprowadzona analiza zbiorów administracyjnych wykazała, że:

1. każdy ze zbiorów zawiera pełny zakres podmiotowy cudzoziemców w zakresie zarejestrowanych faktów, których dotyczy dany zbiór (właściwych dla danego zbioru). Z punktu widzenia potrzeby szacowania pełnej zbiorowości cudzoziemców, analizowane zbiory dostarczają informacji o pewnych podzbiorowościach cudzoziemców przebywających w Polsce. Podzbiorowości te stanowiły podstawę do oszacowania populacji cudzoziemców przebywających w Polsce.

Nie jest możliwe ustalenie pokrycia podmiotowego każdego ze zbiorów w stosunku do całej populacji cudzoziemców (tj. stosunku liczby cudzoziemców w zbiorze do ogólnej populacji cudzoziemców), ponieważ na tym etapie nie jest znana jej wielkość.

2. zakres zmiennych w zbiorach pozwala na częściową charakterystykę cudzoziemców:
  - zbiory zawierają informację o płci i wieku cudzoziemców,
  - zbiory (w większości) zawierają symbol gminy zamieszkania pozwalający na określenie rozmieszczenia terytorialnego szacowanej populacji cudzoziemców,
  - w większości analizowanych zbiorów administracyjnych brak jest informacji o kraju poprzedniego zamieszkania, poziomie wykształcenia, statusie na rynku pracy i sekcji PKD,
  - w części zbiorów niepełna informacja lub brak o okresie przebywania cudzoziemców w Polsce.
3. zbiory zostały przygotowane tak, by mogły dostarczyć informacji na koniec 2015 i 2016 r., zatem zapewniona jest zgodność czasowa.
4. definicja cudzoziemca w ocenianych zbiorach jest zgodna z definicją przyjętą w projekcie:
  - zbiory zawierają identyfikatory jednostek podziału administracyjnego zgodne z TERYT, jednakże w niektórych zbiorach informacja ta jest niepełna.
  - zbiory zawierają informacje o obywatelstwie cudzoziemców – kody krajów są zgodne ze słownikiem krajów.
5. numer PESEL jest rejestrowany w większości zbiorów, z punktu widzenia realizacji projektu są to zbiory referencyjne, w zbiorach, w których brak jest identyfikatora PESEL (w zakresie zezwoleń na pracę i oświadczeń pracodawców o powierzeniu pracy cudzoziemcom) wykorzystane zostały inne cechy m.in. kraj obywatelstwa, kod TERYT, inne specyficzne dla danego zbioru

## 2.7. Przygotowanie zbiorów danych na potrzeby badania

Wykorzystywane w omawianym projekcie badawczym wejściowe zbiory z badań statystycznych oraz rejestrów i systemów informatycznych zostały poddane przetwarzaniu, umożliwiającemu porównywanie, łączenie i analizę danych z różnych źródeł oraz oszacowanie wyników. W tym obszarze prac można wyróżnić grupy działań, wzajemnie się przenikających i uzupełniających, które zostały opisane poniżej.

### 2.7.1. Dobór podmiotowy i przedmiotowy

W tej fazie prac – na podstawie wstępnej analizy ich zawartości zbiorów wejściowych oraz stosownie do przyjętego zakresu przedmiotowego badania i przesłanek metodologicznych – dokonano selekcji potencjalnie przydatnych zmiennych ze zbiorów, jak np. wybór niezbędnych spośród blisko 650 zmiennych w zbiorze osób z NSP 2011 r. Z kolei, stosownie do zakresu podmiotowego badania, zastosowano doборы rekordów danych, tzn. dotyczących cudzoziemców (np. w przypadku zbiorów z badań obejmujących szersze kategorie ludności) i odpowiednich do założonych w badaniu momentów obserwacji (31 XII 2015 i 2016 r.) pod względem okresu przebywania w Polsce (w przypadku zbiorów rejestrowych odnotowujących fakty i daty dotyczące pobytu).

## 2.7.2. Wyliczanie cech pochodnych na podstawie przekształceń surowych danych

W ramach tej grupy działań wykonano szereg wyliczeń i przekształceń surowych danych, mających przede wszystkim na celu:

- **utworzenie (wyprowadzenie) cech potrzebnych do opisu badanej populacji**, czyli tego typu operacje, jak np. wyodrębnienie z numeru PESEL daty urodzenia i płci, obliczanie wieku na podstawie daty urodzenia, a następnie, na podstawie wieku i płci, wyprowadzenie cechy ekonomiczne grupy wieku lub np. wyliczanie okresu pobytu cudzoziemca na podstawie dat zarejestrowanych w dokumentach;
- **zapewnienie zgodności definicyjnej i zakresowej cech pochodzących z różnych źródeł** – np. dostosowanie różnorodnych konwencji zapisów kraju obywatelstwa do ujednoliconego słownika kodów krajów czy przekształcanie zapisów symboli działów PKD (na różnych poziomach szczegółowości) na kody reprezentujące sekcje PKD.

## 2.7.3. Redukcja nadmiarowych danych

**Deduplikacja w obrębie pojedynczych zbiorów danych**, polegająca na wykrywaniu i usuwaniu zwykłych (ewidentnych) dubli – powielonych rekordów danych, czyli takich, które pomimo różnych technicznych (bazodanowych) identyfikatorów rekordu, zawierały dokładne powtórzenie wszystkich wartości/danych.

**Niwelowanie redundancji podmiotowej danych w kilku zbiorach jednego rejestru.** W wielu wypadkach w różnych (odrębnych) zbiorach udostępnianych przez tego samego gestora i należących do jednego systemu informatycznego występują – zarówno merytorycznie zasadne (np. ze względu na różne konteksty prawne pobytu), jak i nieuzasadnione powtórzenia rekordów danych o tych samych osobach. Zaistniała zatem potrzeba wyboru najbardziej adekwatnych, z punktu widzenia założeń badania, danych z określonego rejestru. Poprzez łączenie (parowanie) poszczególnych zbiorów w ramach danego rejestru i wykrywanie rekordów dotyczących tych samych podmiotów (osób), a następnie – w oparciu o przesłanki merytoryczne i utworzone na ich podstawie hierarchie adekwatności – dokonano wyboru najbardziej odpowiedniego rekordu reprezentującego danego cudzoziemca w rejestrze. W konsekwencji, w odniesieniu do określonego rejestru powstawał – w zależności od potrzeb – jeden zbiór zawierający dane dotyczące unikalnych jednostek lub kilka zbiorów, ale podmiotowo rozłącznych.

## 2.7.4. Wyodrębnianie podstawowych jednostek/podmiotów badania – operacje agregowania i restrukturyzacji zbiorów

W wielu zbiorach rejestrowych dedykowanych cudzoziemcom podstawowe jednostki danych (rekordy) nie odnoszą się bezpośrednio do pojedynczych osób, lecz do różnego rodzaju faktów dotyczących osób. Tym samym, w tego typu zbiorach w naturalny (i poprawny) sposób jednej osobie może odpowiadać kilka rekordów danych – odpowiednio do liczby odnotowanych faktów dotyczących osoby, jak np. wielokrotnie wydanych i zarejestrowanych osobnych zezwoleń na pracę lub innych tego rodzaju dokumentów. Stąd niezbędne były przekształcenia i transformacje zbiorów wejściowych, w wyniku których otrzymywano rekordy danych odnoszące się do osób. W szczególności były to działania oparte na:

- grupowaniu (agregowaniu) rekordów danych, w ramach którego utworzono rekordy dla osób oraz wyprowadzono za pomocą operacji i funkcji agregujących przewidziane w badaniu cechy charakteryzujące cudzoziemców (np. wyliczenia na podstawie zapisów w kilku dokumentach długość okresu przebywania osoby w Polsce) lub cechy pomocnicze, jak np. liczba dokumentów obowiązujących w referencyjnym momencie badania czy liczba wariantów cechy zapisanych w dokumentach (np. liczba różnych województw podawanych przez jedną osobę w różnych oświadczeniach);

- restrukturyzacji (transpozycji) danych, w wyniku których pewne różnorodne wartości dotyczące jednej osoby zarejestrowane w kilku rekordach (różne warianty cechy) zapisywano w kolumnach jednego rekordu odnoszącego się do osoby, jak np. różne rodzaje podawanych obywatelstw.

## 2.7.5. Łączenie (parowanie) rekordów z różnych zbiorów

Jedną z podstawowych operacji zastosowanych na zbiorach w ramach pracy badawczej było łączenie (parowanie) rekordów z różnych zbiorów, mające na celu porównywanie i weryfikację danych, wyliczanie określonych w założeniach badania cech (charakterystyk) cudzoziemców, a także integrację danych i ostatecznie – oszacowanie wartości wyników badania. W ramach realizacji projektu zastosowano łączenie danych z różnych zbiorów, zarówno na poziomie równorzędnych jednostek (parowanie rekordów osobowych z dwóch źródeł), jak i łączenie danych na różnych poziomach agregacji (dołączanie do jednostkowego zbioru osób danych dotyczących całych grup/kategorii osób lub jednostek terytorialnych).

W zależności od sytuacji (dostępności odpowiednich kolumn danych), stosowane było parowanie za pomocą pojedynczych kolumn – unikalnych kluczy, czyli różnego typu identyfikatorów rekordów/osób czy kodów terytorialnych, ale także na podstawie kilku cech charakteryzujących jednostki (osoby) zbiorów danych, jak np. data urodzenia, płeć, obywatelstwo itp. Operacje łączenia przeprowadzane były zarówno w obrębie zbiorów pochodzących z jednego rejestru – zazwyczaj na podstawie przygotowanego przez gestora sztucznego identyfikatora rekordów/osób – jak i kojarzenia zbiorów z różnych rejestrów czy badań, w tym wypadku – na ogół za pomocą uniwersalnego identyfikatora (numer PESEL) lub na podstawie kombinacji wartości kilku cech.

## 2.8. Szacunek liczby cudzoziemców w Polsce

### 2.8.1. Szczegółowa informacja o założeniach przyjętych w pracy badawczej

W związku z wykorzystaniem w niniejszym opracowaniu metody *capture-recapture* bazującej na modelach log-liniowych<sup>4</sup>, do oszacowania liczby cudzoziemców poza dostępnymi statystycznymi źródłami danych, należy w pierwszej kolejności określić przyjęte założenia metodologiczne. Poniżej przedstawiono listę kluczowych założeń, których spełnienie jest niezbędne z punktu widzenia przyjętych rozwiązań modelowych. Wskazano również działania, które miały na celu ich spełnienie. Do głównych założeń należy zaliczyć:

- **definicje populacji we wszystkich rozważanych źródłach są takie same** – w projekcie określono populację cudzoziemców jako osoby posiadające obywatelstwo inne niż polskie w wieku 18 lat i więcej, które przebywały w Polsce w końcu 2015 i 2016 roku;
- **źródła danych są niezależne** – w przypadku źródeł administracyjnych systemy powinny być niezależne (na przykład w sensie prawnym), aby możliwe było zastosowanie metody *capture-recapture* wykorzystującej modele log-liniowe. Ostatecznie na potrzeby prezentowanej pracy badawczej wykorzystano kombinację trzech źródeł danych, które umożliwiają spełnienie tego założenia (tj. PESEL, UdSC i ZUS oraz PESEL, UdSC i KEP)<sup>5</sup>. Głównym uzasadnieniem wyboru tych źródeł danych było ich bieżące wykorzystywanie w statystyce publicznej na potrzeby innych badań (nie wymagało to pozyskania danych spoza PBSSP) oraz pokrycie tej samej populacji;
- **populacja jest zamknięta** – zakłada się, że w badanym okresie wielkość populacji jest stała. Ponadto należy podkreślić, że wszystkie rejestry były aktualne na ten sam dzień, tj. 31.12.2016 r.

<sup>4</sup> Szczegółowy opis modeli log-liniowych wykorzystywanych w kontekście populacji trudnych do zbadania zamieszczony został Załączniku 3 do niniejszej pracy badawczej.

<sup>5</sup> W dalszej części raportu autorzy posługują się zapisem PESEL, UdSC i KEP/ZUS co oznacza, że raz wykorzystano zbiór KEP, a drugim razem ZUS.

oraz na potrzeby analiz, zgodnie z zapisami w studium wykonalności, wybrano odpowiednie okresy dla 2015 i 2016 r.

- **brak błędów nadreprezentacji i duplikatów** – zakłada się, że źródła są pozbawione błędów nadreprezentacji, tj. źródła zawierają wyłącznie jednostki z badanej populacji oraz zostały zdeduplikowane. Podstawowym źródłem był zintegrowany zbiór danych powstały na podstawie łączenia kilku rejestrów administracyjnych i zdeduplikowany, który został stworzony na podstawie kilku rejestrów administracyjnych i zdeduplikowany. Rejestr ten zawiera również zmienną dotyczącą jakości danego rekordu, który jest przybliżeniem błędu nadreprezentacji. Dodatkowo, przyjęto przy tym założenie, że cudzoziemcy będący w rejestrach przebywają na terenie Polski niezależnie od tego czy mają ustalone miejsce pobytu. Jest to kluczowe zwłaszcza w przypadku systemu „Pobyt”, którego gestorem jest UdSC.
- **każda jednostka ma szansę znalezienia się w każdym z rozpatrywanych źródeł danych (przy czym prawdopodobieństwo nie musi być jednakowe)** – aby spełnić to założenie w procesie modelowania wykorzystano dwa modele zawierające następujące zmienne: 1) kraj obywatelstwa, 2) płeć, 3) wiek (2 grupy) i 4) województwo (16 oraz nieustalone) oraz : 1) kraj obywatelstwa, 2) płeć, 3) wiek (2 grupy) i 4) podregion (73 oraz nieustalone). Wybór podyktowany był zapisami projektu;
- **każdą jednostkę będzie można zidentyfikować i połączyć między źródłami bez błędów** – w tym celu zastosowano jedynie integrację danych za pomocą identyfikatora PESEL (łączenie deterministyczne).

Należy zaznaczyć, że do oszacowania wielkości populacji na poziomie podregionów wykorzystano inny model niż ten dla pozostałych zmiennych. Oznacza to, że wyniki uzyskane między modelami nie są porównywalne. W obydwu przypadkach zastosowano tę samą procedurę estymacji oraz doboru modeli, zaprezentowaną poniżej. Szczegółowy opis dotyczy modelu zbudowanego z wykorzystaniem zmiennej województwo. Rezultaty dla podregionów przedstawiono w oddzielnym podrozdziale.

W pierwszym kroku prac postanowiono wykorzystać dwie kombinacje zbiorów danych: PESEL, UdSC i ZUS oraz PESEL, UdSC i KEP aby zobaczyć, które z kombinacji będą właściwą podstawą do dalszych szacunków. Wyboru dokonano poddając analizie miary jakości wykorzystywane do oceny uogólnionych modeli liniowych, które opisane zostały szczegółowo w załączniku poświęconemu teorii estymacji log-liniowej.

Głównym celem projektu było oszacowanie liczby cudzoziemców w Polsce w 2015 oraz 2016 roku na regionalnych rynkach pracy. W związku z faktem, że otrzymane dane były aktualne na 31.12.2016 r. podjęto następujące kroki przy wyodrębnianiu populacji na wyżej wymienione lata:

- dla 2015 roku:
  - Na podstawie PESEL, UdSC i ZUS/KEP wybrano tylko osoby urodzone przed 31.12.2015 r.,
  - Na podstawie UdSC wybrano tylko te osoby, które miały decyzję umożliwiającą pobyt w Polsce wydaną między 01.01.2015 r. oraz 31.12.2015 r.
- dla 2016 roku:
  - Na podstawie PESEL, UdSC i ZUS/KEP wybrano tylko osoby urodzone przed 31.12.2016 r.,
  - Na podstawie UdSC wybrano tylko tych, którzy mieli decyzję umożliwiającą pobyt w Polsce wydaną między 01.01.2016 r. oraz 31.12.2016 r.

W przypadku rejestru UdSC nie wyłączono z analizy cudzoziemców, którym upłynęła data ważności dokumentu wydanego przez Urząd do Spraw Cudzoziemców w ciągu 2016 r., tj. przed 31.12.2016 r. W 2015 roku wszyscy cudzoziemcy w zbiorze UdSC mieli aktualne dokumenty.

Dla rekordów występujących w PESEL, ZUS i KEP wyodrębniono te, dla których określono kod jakości 1 oznaczający „*sytuacja referencyjna (istnienie osoby potwierdzone)*”, 3 wskazujący „*osoby w wieku 90+*” oraz kod 6 oznaczający „*osobę zidentyfikowaną tylko w jednym rejestrze*”, który był wyznaczony przed dołączeniem rejestru UdSC.

W przypadku następujących przekrojów: 1) ogółem, 2) kraj obywatelstwa, 3) wiek, 4) płeć, 5) województwo i 6) podregion dokonano szacunków w oparciu o model log-liniowy, a precyzję szacunków

określono na podstawie metody *bootstrap* wyznaczając 95% przedziały ufności. Wykorzystano w tym celu kod języka R stanowiący załącznik do artykułu Petera i inn. (2012). Kod ten na potrzeby prac analitycznych w pracy badawczej został odpowiednio zmodyfikowany. Natomiast w przypadku pozostałych przekrojów dokonano rozszacowania wielkości populacji uwzględniając powyższe przekroje oraz struktury z innych źródeł danych. W tym przypadku prezentowane są jedynie oszacowania punktowe.

## 2.8.2. Opis wykorzystanych źródeł danych

W Tabelach 2-5 przedstawiono zestawienie liczby cudzoziemców według występowania w trzech źródłach według badanych lat. Wartość „TAK” oznacza, że cudzoziemiec został zidentyfikowany, a wartość „NIE” oznacza, że nie został zidentyfikowany w danym źródle. W przypadku kombinacji PESEL, UdSC, ZUS w 2015 roku wykorzystano informacje o ponad 68,5 tys., a w 2016 roku dla blisko 154 tys. cudzoziemców. W przypadku kombinacji PESEL, UdSC i KEP w 2015 roku wykorzystano dane o blisko 69 tys., a w 2016 o 154,5 tys. cudzoziemców.

W odniesieniu do 2015 roku jedynie 9 871 cudzoziemców było zidentyfikowanych jednocześnie w PESEL, UdSC i ZUS oraz 11 870 w przypadku PESEL, UdSC i KEP. W 2016 roku te liczby przedstawiały się następująco: 18 951 występowało jednocześnie w PESEL, UdSC i ZUS oraz 21 544 występowało jednocześnie w PESEL, UdSC i KEP. W tabelach pojawia się również wartość “--”, która oznacza nieznaną liczbę cudzoziemców będących poza wymienionymi rejestrami.

**Tabela 2. Zestawienie liczby cudzoziemców w wieku 18+ według PESEL, UdSC i ZUS zgodnie ze stanem na 31.12.2015 r.**

PESEL	ZUS	UdSC		SUMA
		NIE	TAK	
NIE	NIE	--	30 090	30 090
	TAK	3 821	5 583	9 404
TAK	NIE	7 042	7 476	14 518
	TAK	4 620	9 871	14 491
	<b>SUMA</b>	<b>15 483</b>	<b>53 020</b>	<b>68 503</b>

Źródło: opracowanie własne.

**Tabela 3. Zestawienie liczby cudzoziemców w wieku 18+ według PESEL, UdSC i KEP zgodnie ze stanem na 31.12.2015 r.**

PESEL	KEP	UdSC		SUMA
		NIE	TAK	
NIE	NIE	--	28 230	28 230
	TAK	4 011	7 443	11 454
TAK	NIE	5 731	5 477	11 208
	TAK	5 931	11 870	17 801
	<b>SUMA</b>	<b>15 673</b>	<b>53 020</b>	<b>68 693</b>

Źródło: opracowanie własne.

**Tabela 4. Zestawienie liczby cudzoziemców w wieku 18+ według PESEL, UdSC i ZUS zgodnie ze stanem na 31.12.2016 r.**

PESEL	ZUS	UdSC		SUMA
		NIE	TAK	
NIE	NIE	--	92 106	92 106
	TAK	3 821	11 224	15 045
TAK	NIE	7 115	16 549	23 664
	TAK	4 641	18 951	23 592
	<b>SUMA</b>	<b>15 577</b>	<b>138 830</b>	<b>154 407</b>

Źródło: opracowanie własne.

**Tabela 5. Zestawienie liczby cudzoziemców w wieku 18+ według PESEL, UdSC i KEP zgodnie ze stanem na 31.12.2016 r.**

PESEL	KEP	UdSC		SUMA
		NIE	TAK	
NIE	NIE	--	89 695	<b>89 695</b>
	TAK	4 012	13 635	<b>17 647</b>
TAK	NIE	5 825	13 956	<b>19 781</b>
	TAK	5 931	21 544	<b>27 475</b>
	<b>SUMA</b>	<b>15 768</b>	<b>138 830</b>	<b>154 598</b>

Źródło: opracowanie własne.

Głównym celem niniejszej pracy badawczej jest oszacowanie liczby cudzoziemców będących poza tymi rejestrami, tj. nieznannej wartości liczbowej na przecięciu pól: PESEL = NIE, UdSC = NIE i ZUS = NIE lub KEP = NIE. Na potrzeby estymacji tej liczebności wykorzystano wspomniane już modele log-liniowe, których teorię w odniesieniu do populacji trudnych do zbadania przedstawiono w Załączniku 3 pracy badawczej.

### 2.8.3. Opis zmiennych pochodzących z analizowanych rejestrów

Zgodnie z literaturą<sup>6</sup> poświęconą szacowaniu wielkości nieznannej populacji założono, że prawdopodobieństwo pokrycia przez określone źródła danych nie jest jednakowe. Dlatego na potrzeby procesu modelowania wykorzystano następujące zmienne:

- Płeć:
  - 1 = Mężczyzna
  - 2 = Kobieta
- Wiek:
  - Produkcyjny – 18-59 dla kobiet, 18-64 dla mężczyzn
  - Poprodukcyjny – 60+ dla kobiet, 65+ dla mężczyzn
- Kraj obywatelstwa:
  - UE
  - Armenia
  - Mołdawia
  - Białoruś
  - Rosja
  - Ukraina
  - Wietnam
  - Inne
- Województwo
  - 16 województw kodowanych 1,...16
  - BD – brak danych jeżeli nie zostało ustalone miejsce pobytu
- Podregion:
  - 73 podregiony kodowane 1,...,73
  - BD – brak danych jeżeli nie zostało ustalone miejsce pobytu

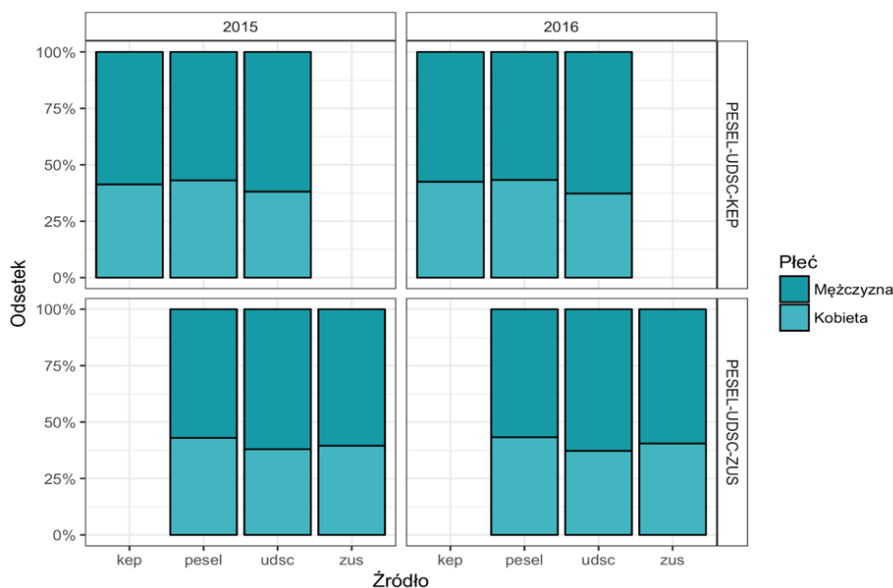
Na kolejnych wykresach przedstawiono rozkłady analizowanych zmiennych w poszczególnych rejestrach. Uwzględniono zatem następujące zmienne: płeć, wiek, kraj obywatelstwa i informację czy lokalizacja (podregion zamieszkania) danej osoby była ustalona czy nie.

Największe różnice w rozkładach zmiennych zaobserwowano w przypadku rejestru UdSC, w którym odsetek mężczyzn był wyższy niż w pozostałych rejestrach. Ta relacja jest stała w czasie ponieważ różnice są zauważalne zarówno dla 2015, jak i 2016 roku.

<sup>6</sup> Wykaz najważniejszych pozycji literatury przedmiotu dotyczących populacji trudnych do zbadania przedstawiono w załączniku 2, jak również w niniejszym raporcie w rozdziale Bibliografia.



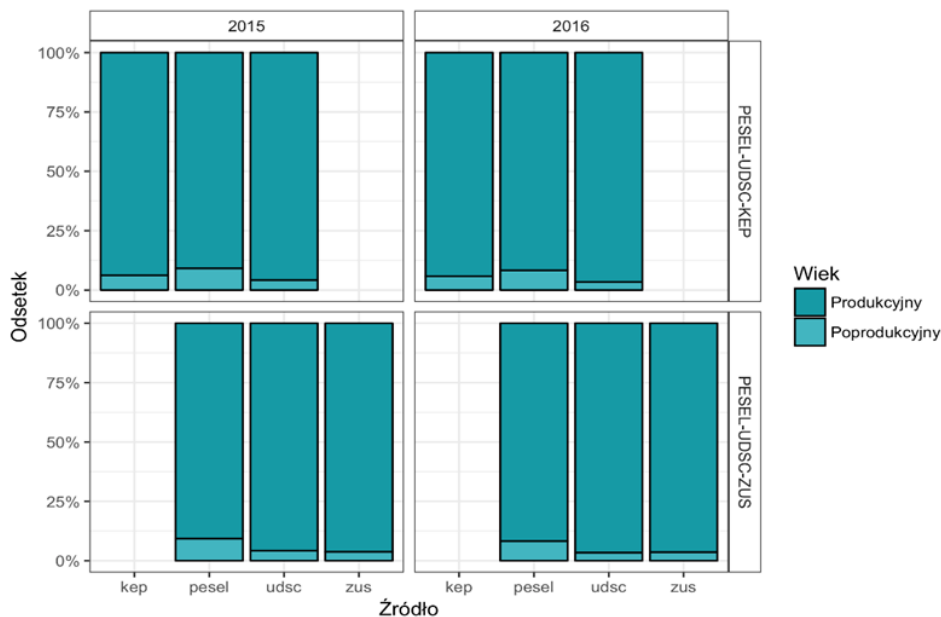
**Wykres 2.1. Rozkład cudzoziemców według płci, lat oraz kombinacji rozważanych źródeł**



Źródło: Opracowanie własne.

W przypadku zmiennej wiek, największe różnice zauważyć można porównując rejestr PESEL do pozostałych źródeł danych. Odsetek osób w wieku poprodukcyjnym jest znacznie wyższy niż w innych rejestrach. Wydaje się, że sytuacja ta ma swoje uzasadnienie w tym, że rejestr PESEL obejmuje osoby, które długo przebywają w danym kraju, a w ZUS, KEP i UdSC znajdują się osoby, które przyjeżdżają do pracy w Polsce. Bez względu jednak na rejestr można zaobserwować znaczną przewagę cudzoziemców w wieku produkcyjnym w porównaniu z liczbą cudzoziemców w wieku poprodukcyjnym. Ma to oczywiście związek z faktem, że cudzoziemcy ci przybywają do Polski w celach zarobkowych.

**Wykres 2.2. Rozkład cudzoziemców według wieku, lat oraz kombinacji rozważanych źródeł**



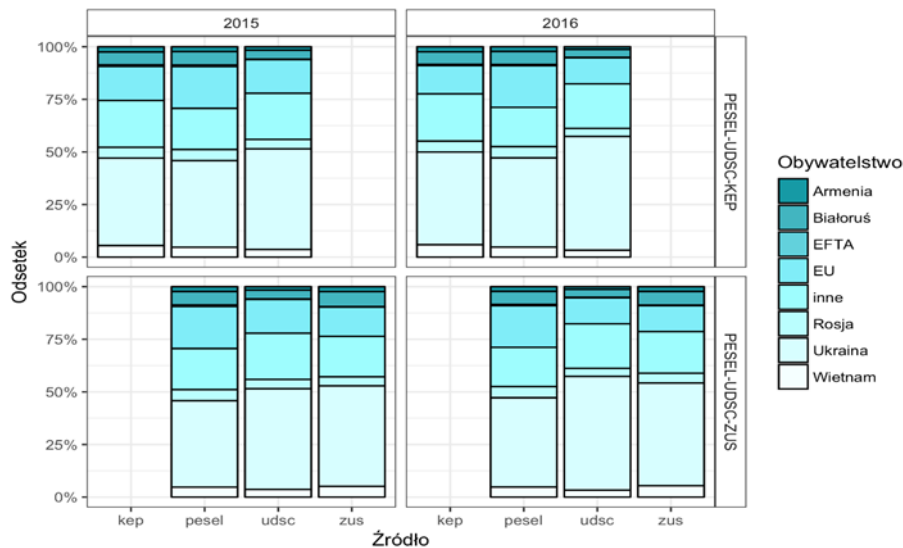
Źródło: Opracowanie własne.

Na wykresie 2.3 przedstawiono z kolei rozkład liczby cudzoziemców według trzech rozważanych źródeł danych oraz ich kraju obywatelstwa. Można zauważyć, że bez względu na źródło przeważają

obywatele Ukrainy. Stosunkowo znaczny jest również odsetek obywateli z krajów Unii Europejskiej oraz innych krajów.

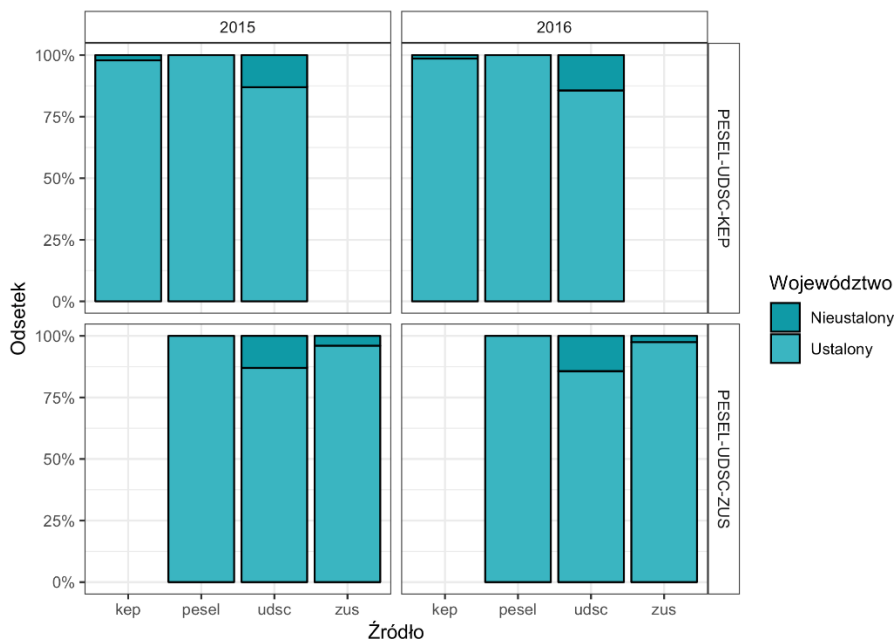
Na wykresie 2.4 przedstawiony został z kolei rozkład cudzoziemców według podregionów. Bez względu na źródło można sklasyfikować podregion (miejsce pobytu cudzoziemca) jako ustalony. Dotyczy to przede wszystkim rejestru PESEL, a w dużej mierze również KEP oraz ZUS. Pewne odstępstwo stanowi system „Pobyt” UdSC, ze znacznie większym odsetkiem cudzoziemców, w porównaniu do trzech wspomnianych wyżej rejestrów, w którym podregion przebywania cudzoziemca został zaklasyfikowany jako nieustalony.

**Wykres 2.3. Rozkład cudzoziemców według kraju obywatelstwa, lat oraz kombinacji rozważanych źródeł**



Źródło: Opracowanie własne.

**Wykres 2.4. Rozkład cudzoziemców według województw (sklasyfikowanych jako ustalony i nieustalony), lat oraz kombinacji rozważanych źródeł**



Źródło: Opracowanie własne.

## 2.9. Opis procedury estymacji wielkości populacji cudzoziemców w Polsce oraz w szczegółowych przekrojach w 2015 i 2016 r.

Na potrzeby estymacji liczby cudzoziemców w Polsce w 2015 i 2016 r. w założonych przekrojach odnoszących się do płci, grup wieku, województw oraz kraju obywatelstwa wykorzystano modele log-liniowe. W przypadku podregionów zastosowano model, w którym wykorzystano zamiast zmiennej województwo zmienną podregion. W tych przekrojach, oprócz podania oszacowanej wartości punktowej liczby cudzoziemców, podano również informację o 95% przedziałach ufności. W odniesieniu do sekcji PKD oraz okresu przebywania cudzoziemców w Polsce (przyjęto ostatecznie dwie kategorie: do 12 miesięcy i 12 miesięcy i więcej) dokonano rozszacowania oszacowanej liczby cudzoziemców w oparciu o rozkłady uzyskane z MRPiPS na podstawie oświadczeń i zezwoleń dla cudzoziemców. W przypadku liczby cudzoziemców w przekrojach zdefiniowanych przez ich wykształcenie, status na rynku pracy (pracujący, bezrobotni i bierni zawodowo) czy miejsce zamieszkania (miasto/wieś) dokonano rozszacowania uzyskanej liczby cudzoziemców zgodnie z rozkładami analizowanych cech z NSP 2011 (wykształcenie i miejsce zamieszkania) i BAEL (status na rynku pracy) w odniesieniu do grupy cudzoziemców. Należy jednak podkreślić, że w tym przypadku (jak i w odniesieniu do sekcji PKD i czasu przebywania) przedstawiono jedynie wartości punktowe liczby cudzoziemców, a niemożliwe było podanie odpowiednich przedziałów ufności. Co więcej, uzyskane wyniki dla liczby cudzoziemców w latach 2015 i 2016, zwłaszcza w odniesieniu do ich wykształcenia czy miejsca zamieszkania należy interpretować ze szczególną ostrożnością. Wyniki te mogą być bowiem obciążone. Przyjmuje się tutaj dość silne założenie, że rozkłady tych cech dla cudzoziemców z NSP 2011 nie zmieniły się w stosunku do roku 2015 i 2016. Ze względu jednak na brak danych odnoszących się do tych zmiennych w trzech wziętych pod uwagę rejestrach był to jedyny możliwy sposób przedstawienia takich struktur.

Jak to zostało podkreślone, na potrzeby estymacji liczby cudzoziemców w Polsce w latach 2015 i 2016, w wyodrębnionych przekrojach, wykorzystano modele log-liniowe. Ich teoretyczne podstawy zostały sformułowane w Załączniku 3, który stanowi integralną część prezentowanego raportu badawczego. Na potrzeby budowy odpowiedniego modelu log-liniowego wykorzystano dane pochodzące z trzech wspomnianych wcześniej rejestrów a w charakterze zmiennych dodatkowych wzięto pod uwagę płeć, kraj obywatelstwa, podregion oraz grupy wieku<sup>7</sup>. W raporcie opisany syntetycznie został jedynie sposób oszacowania wariancji estymatora wielkości populacji, którą wykorzystano do zbudowania odpowiednich przedziałów ufności.

### 2.9.1. Estymacja wariancji estymatora wielkości populacji

Do oszacowania wariancji estymatora wielkości populacji wykorzystano procedurę zaproponowaną przez Petera i inn. (2012), która składała się z następujących kroków:

1. Dokonanie oszacowania wielkości populacji we wszystkich przekrojach na podstawie ostatecznego modelu.
2. Wyliczenie wielkości populacji  $\hat{N} = \hat{N}_1 + \dots + \hat{N}_P$ , gdzie  $p = 1, \dots, P$  to odpowiedni przekrój wyznaczony przez kombinację trzech źródeł, kraju obywatelstwa, płci, grupy wieku oraz podregionu.
3. Wygenerowanie z rozkładu wielomianowego wektora o długości  $p \times 1$  odpowiadającego z populacji o wielkości  $\hat{N}$  z prawdopodobieństwami  $\hat{\pi}_p = (\hat{N}_1/\hat{N}, \dots, \hat{N}_P/\hat{N})^T$ . W wyniku otrzymano wektor pseudo-wielkości populacji we wszystkich  $P$  przekrojach, które oznaczono przez  $\mathbf{N}^* = (N_1^*, \dots, N_P^*)^T$ .
4. Dołączenie wektora z punktu 3 do pełnego zbioru danych z punktu 1.

<sup>7</sup> Sposób wykorzystania modeli log-liniowych dla dwóch i trzech źródeł danych z uwzględnieniem dodatkowych zmiennych, w kontekście szacowania populacji trudnych do zbadania, został również szczegółowo przedstawiony w Załączniku 3.

5. Oszacowanie modelu log-liniowego jedynie dla danych obserwowanych w rejestrach administracyjnych.
6. Dokonanie oszacowania dla przekrojów nieobserwowanych w rejestrze na podstawie modelu z punktu 5.
7. Powtórzenie kroku 3-6  $B$  razy, gdzie  $B=500$ .

Następnie na podstawie otrzymanych oszacowań  $\hat{N}^B$  wyznaczono wartość oczekiwaną, wariancję, względny błąd standardowy oraz 95% przedział ufności metodą percytylową na podstawie następujących wzorów:

- wartość oczekiwana:

$$\hat{N} = \frac{\sum_{b=1}^B \hat{N}^B}{B},$$

- wariancja oszacowań:

$$\hat{V}(\hat{N}) = \frac{\sum_{b=1}^B (\hat{N}^B - \hat{N})^2}{B-1},$$

- względny błąd standardowy:

$$\hat{CV}(\hat{N}) = \frac{\sqrt{\hat{V}(\hat{N})}}{\hat{N}} \times 100\%,$$

- 95% przedział ufności wyznaczony metodą percytylową:

$$\left[ \hat{N}_{2,5\%}, \hat{N}_{97,5\%} \right]$$

Szacunków dokonano dla wszystkich przekrojów: ogółem, kraj pochodzenia, wiek, płeć, województwo oraz podregion.

## 2.9.2. Wybór modelu

Na potrzeby wyboru końcowego modelu log-liniowego, który wykorzystano w procesie estymacji liczby cudzoziemców w Polsce w odpowiednich przekrojach, w pierwszej kolejności dokonano zakodowania zmiennych, zgodnie z symbolicznym zapisem (notacja nawiasowa) charakterystycznym dla modeli log-liniowych<sup>8</sup>. Poniższy opis dotyczy wyników uzyskanych dla modelu wykorzystującego województwa. Przyjęto zatem następujące oznaczenia:

- PESEL = P
- UdSC = U
- ZUS = Z
- KEP = K
- Płeć = S (od angielskiego słowa sex)<sup>9</sup>
- Wiek = A (od angielskiego słowa age)
- Kraj obywatelstwa = C (od angielskiego słowa citizenship)
- Województwo = V (od angielskiego słowa voivodeship)

Procedurę modelowania przeprowadzono oddzielnie dla lat 2015 i 2016 oraz dla kombinacji źródeł. Oznacza to, że ostatecznie przeprowadzono cztery niezależne procedury szacunku wielkości populacji.

<sup>8</sup> Szerzej o notacji – w Załączniku 3 poświęconemu teorii modeli log-liniowych.

<sup>9</sup> Użyto angielskich skrótów, aby wyraźnie dokonać rozróżnienia symboli podczas budowania modeli log-liniowych. W ten sposób uniknięto użycia tej samej litery np. przypadku rejestru PESEL i płci, co byłoby źródłem niejasności.

W przypadku każdego ze źródeł danych rozważono następujące kombinacje zmiennych (według układów trzech rozpatrywanych źródeł) zapisane zgodnie z notacją nawiasową modeli log-liniowych:

- Efekty główne:
  - [P][U][Z][S][A][C][V],
  - [P][U][K][S][A][C][V].
- Efekty główne wraz z interakcjami pierwszego rzędu<sup>10</sup>
  - [P][U][Z][S][A][C][V][PU][PZ][PS]...
  - [P][U][K][S][A][C][V][PU][PK][PS]...

Nie rozważano modelu z trzema interakcjami, ponieważ liczba możliwych interakcji byłaby bardzo duża, co skomplikowałoby model i mogłoby doprowadzić do jego przeuczenia przy relatywnie niewielkiej liczbie obserwacji. Następnie dla modeli z interakcjami pierwszego stopnia zastosowano procedurę krokową wykorzystującą informacyjne kryterium AIC w celu wyboru najlepszego modelu<sup>11</sup>. Procedurę wykorzystano zgodnie z literaturą statystyczną w zakresie *capture-recapture*.

Tabela 6 zawiera zestawienie wybranych miar jakości dla zastosowanych modeli log-liniowych według kombinacji źródeł oraz roku. W przypadku pierwszej kombinacji źródeł (PESEL, UdSC i KEP) dla obydwu lat (2015, 2016) model z interakcjami oraz doбором krokowym okazał się najlepszy. Potwierdzają to kryteria informacyjne AIC oraz BIC<sup>12</sup>, które są niższe dla modelu 2s<sup>13</sup> w porównaniu do modelu 2. Ten wynik został zaznaczony na czerwono w Tabeli 5.

**Tabela 6. Wybrane miary jakości modeli log-liniowych według kombinacji źródeł oraz roku**

Kombinacja	Rok	Model	Dewiancja modelu pustego	df modelu pustego	G <sup>2</sup>	AIC	BIC	Dewiancja	df resztowy
PESEL UdSC KEP	2015	1	216 638	2 316	-27 547	55 153	55 320	46 360	2 288
		2	216 638	2 316	-9 707	19 944	21 467	10 679	2 052
		2s	216 638	2 316	-9 707	19 944	21 467	10 679	2 052
	2016	1	656 534	2 517	-55 973	112 003	112 172	101 828	2 489
		2	656 534	2 517	-13 320	27 171	28 716	16 523	2 253
		2s	656 534	2 517	-13 321	27 169	28 709	16 524	2 254
PESEL UdSC ZUS	2015	1	216 672	2 218	-24 972	50 003	50 168	41 580	2 190
		2	216 672	2 218	-8 909	18 349	19 860	9 453	1 954
		2s	216 672	2 218	-8 909	18 349	19 860	9 453	1 954
	2016	1	651 403	2 399	-50 619	101 295	101 463	9 1543	2 371
		2	651 403	2 399	-12 260	25 051	26 583	14 827	2 135
		2s	651 403	2 399	-12 260	25 049	26 575	14 827	2 136

Źródło: opracowanie własne. Wyjaśnienia: 1 = model wyłącznie z efektami głównymi, 2 = model z efektami głównymi i interakcjami pierwszego rzędu, 2s = model 2 z zastosowaną procedurą krokową (s pochodzi od *step*, które odnosi się do pojęcia regresji krokowej, ang. *stepwise selection*; *stepwise regression*). Wyjaśnienia: df = stopnie swobody, df modelu pustego – stopnie swobody modelu jedynie z wyrazem wolnym (inaczej modelu pustego), df resztowy – różnica między liczbą stopni swobody modelu pustego, a modelu w danym wierszu.

W przypadku modelu opartego na kombinacji PESEL, UdSC i ZUS model 2 i 2s okazał się identyczny w 2015 roku (co potwierdzają kryteria informacyjne), a w przypadku 2016 roku model 2s był nieznacznie lepszy od modelu 2, ponieważ zarówno kryteria informacyjne AIC, jak i BIC, są niższe.

Wyniki zaprezentowane w Tabeli 5 nie pozwalają na bezpośrednie porównanie modeli budowanych na różnych kombinacjach źródeł danych, ponieważ liczba obserwacji jest różna (por. liczbę

<sup>10</sup> Ze względu na bardzo złożoną postać modeli, w notacji nawiasowej zapisano tylko ich początkowe człony. Pominięto zatem wszystkie możliwe interakcje, które wzięte zostały pod uwagę, podając tylko kilka pierwszych ich wariantów.

<sup>11</sup> Postać kryterium informacyjnego Akaike (AIC) przedstawiono w załączniku 3 pracy badawczej.

<sup>12</sup> Podobnie jak AIC również kryterium informacyjne BIC zostało opisane w załączniku 3 niniejszej pracy badawczej.

<sup>13</sup> objaśnienia modeli znajdują się w legendzie pod Tabelą 5.

stopni swobody dla modelu pustego, kolumna – df modelu pustego). W przypadku kombinacji PESEL, UdSC, KEP liczba rekordów wynosiła odpowiednio 2 316 dla 2015 i 2 517 dla 2016 roku, a dla PESEL, UdSC i ZUS 2 218 i 2 399 dla 2015 i 2016 roku. Dlatego do porównania tych modeli wykorzystano miary Pseudo-R<sup>2</sup>, które informują o jakości dopasowania modelu i są analogią do miar R<sup>2</sup> wykorzystywanych w modelach liniowych<sup>14</sup>.

Tabela 7 zawiera porównanie modelu 2s według kombinacji źródeł danych na podstawie miar Pseudo-R<sup>2</sup>. Wszystkie miary sugerują, że model z wykorzystaniem ZUS wyjaśnia większą część wariancji niż model z KEP. Obydwa modele wyjaśniają ponad 93% zmienności dla 2015 roku i ponad 96% zmienności liczby cudzoziemców w 2016 roku.

**Tabela 7. Porównanie modelu 2s według kombinacji źródeł danych według miar Pseudo-R<sup>2</sup>**

Rok	Miara	PESEL, UdSC, KEP	PESEL, UdSC, ZUS
2015	Współczynnik R <sup>2</sup> Effrona <sup>15</sup>	94,82	95,39
	Skorygowany współczynnik R <sup>2</sup> (skor. wsp. McFaddena)	95,86	96,47
2016	Współczynnik R <sup>2</sup> Effrona	97,4	97,64
	Skorygowany współczynnik R <sup>2</sup> (skor. wsp. McFaddena)	98,43	98,62

Źródło: opracowanie własne.

Na podstawie danych zawartych w Tabeli 6 podjęto ostatecznie decyzję o wykorzystaniu oszacowań z modelu log-liniowego wykorzystującego PESEL, UdSC i ZUS jako źródła danych w estymacji liczby cudzoziemców w Polsce w 2015 i 2016 roku.

### 2.9.3. Opis wybranego modelu log-liniowego

Tabela 8 przedstawia postać ostatecznego modelu wraz z oszacowaną wielkością populacji cudzoziemców w Polsce w latach 2015 i 2016 oraz 95% przedziałem ufności. Model dla 2015 r. nieznacznie różni się od modelu dla 2016 r., ponieważ posiada dwa dodatkowe elementy – interakcję między płcią a wiekiem (oznaczone w tabeli jako [SA]) oraz między płcią a zbiorem UdSC (oznaczone w tabeli jako [US]).

Wynik modelowania sugeruje, że prawdopodobieństwa pokrycia przez badane źródła danych jest stałe, co powoduje, że zastosowany model jest stabilny w czasie. Przemawia to na korzyść zastosowanego w pracy badawczej podejścia.

**Tabela 8. Postać ostatecznego modelu log-liniowego (notacja nawiasowa) wraz z oszacowaną wielkością populacji cudzoziemców w Polsce oraz 95% przedziałem ufności**

Rok	Postać modelu	$\hat{N}$	95% przedział ufności
2015	[P][Z][U][V][S][A][C] [PZ][PU][PV][PS][PA][PC][ZU][ZV] [ZA][ZC][UV][UC][VS][VA][VC] [SA][AC][UA][SC][US] [ZS]	507 693	(369 135, 724 407)
2016	[P][Z][U][V][S][A][C] [PZ][PU][PV][PS][PA][PC][ZU][ZV] [ZA][ZC][UV][UC][VS][VA][VC] [SA][AC][UA][US][SC]	743 665	(600 796, 943 124)

Źródło: Opracowanie własne.

<sup>14</sup> Miarę Pseudo-R<sup>2</sup> również opisano w załączniku 3 pracy badawczej.

<sup>15</sup> Współczynnik ten nie został wyszczególniony w załączniku 3 gdyż jest on analogiczny do klasycznego współczynnika R<sup>2</sup> znanego z regresji liniowej. Informacje na jego temat ( oraz wielu innych miar dopasowania R<sup>2</sup>) można znaleźć pod adresem [http://www.glmj.org/archives/articles/Smith\\_v39n2.pdf](http://www.glmj.org/archives/articles/Smith_v39n2.pdf).

W tabeli 9 oraz 10 przedstawiono wyniki procedury analizy wariancji typu II, która wskazuje jaką część wariancji wyjaśniły poszczególne czynniki. Anova typu drugiego ma przewagę nad anową typu I, ponieważ kolejność zmiennych nie odgrywa roli w momencie wyliczenia wariancji wyjaśnionej przez poszczególne zmienne/interakcje. Kolumna LR informuje o wartości statystyki ilorazu wiarygodności, która informuje o tym jaką część wariancji wyjaśnił określony czynnik (niezależnie od poziomów, które przyjmuje).

W obydwu modelach zmienną wyjaśniającą największą część wariancji były kraj obywatelstwa (C), województwo (V), wiek (A), rejestr ZUS (Z) oraz interakcja między województwem, a obywatelstwem (RC). Wyniki odzwierciedlają różnice zaprezentowane na wcześniejszych wykresach porównujących źródła danych ponieważ informacja o braku lokalizacji znacząco różniła rejestr PESEL od UdSC czy ZUS. Podobna sytuacja dotyczy wieku. W rejestrze PESEL obserwowaliśmy znacząco więcej osób w wieku poprodukcyjnym w porównaniu do pozostałych źródeł.

**Tabela 9. Zestawienie zmiennych oraz zmienności przez nie wyjaśnionej dla 2015 roku**

Zmienna	LR	df	Zmienna	LR	df
C	69,08	7	UC	1,19	7
V	59,93	16	ZC	1,15	7
A	54,51	1	VS	0,72	16
VC	8,58	112	PC	0,68	7
Z	7,10	1	ZU	0,57	1
PZ	6,66	1	SA	0,51	1
S	3,72	1	PA	0,40	1
ZV	2,12	16	PS	0,29	1
P	2,00	1	ZA	0,28	1
PV	1,79	15	VA	0,19	16
UV	1,59	16	UA	0,15	1
U	1,46	1	ZS	0,01	1
SC	1,34	7	US	0,01	1
AC	1,19	7	PU	0,00	1

Źródło: Opracowanie własne.

**Tabela 10. Zestawienie zmiennych oraz zmienności przez nie wyjaśnionej dla 2016 roku**

Zmienna	LR	df	Zmienna	LR	df
C	200,10	7	PC	2,25	7
A	146,38	1	VS	1,53	16
V	145,60	16	ZC	1,36	7
Z	41,40	1	UC	1,22	7
VC	20,98	112	SA	1,07	1
PZ	20,03	1	PA	0,83	1
P	17,74	1	PS	0,77	1
U	15,74	1	ZU	0,42	1
S	9,80	1	ZA	0,40	1
ZV	4,92	16	UA	0,39	1
PV	3,44	15	VA	0,35	16
SC	2,81	7	US	0,02	1
AC	2,53	7	PU	0,01	1
UV	2,50	16			

Źródło: Opracowanie własne.

## 2.10. Otrzymane szacunki i ich interpretacja

### 2.10.1. Podstawowe informacje dotyczące interpretacji szacunków

Wyniki zostały podzielone na dwie części. W pierwszej przedstawiono szacunek wielkości populacji wraz z przedziałami ufności oraz precyzją według następujących poziomów<sup>16</sup>:

- Ogółem,
- Kraj obywatelstwa,
- Płeć,
- Wiek,
- Województwo,
- Podregion.

W tym przypadku wyniki oparte są na modelu i zmiennych, które bezpośrednio były wykorzystane do oszacowania wielkości populacji cudzoziemców. Prezentowane wyniki są nieobciążone, a przedziały ufności odzwierciedlają prawdziwą niepewność przy założeniu 1) poprawności modelu, 2) braku błędów pokrycia w poszczególnych źródłach oraz przy spełnieniu założeń prezentowanych na początku raportu. Niemniej jednak w związku z obiektywnymi przesłankami wyniki należy traktować jako przybliżające prawdziwą wielkość populacji cudzoziemców. **Dlatego rekomenduje się, aby interpretować wyniki w kontekście przedziałów ufności, a nie oszacowań punktowych.**

Dodatkowa uwaga dotyczy wielkości populacji według podregionów i województw. W zmiennej województw (i podregion) pojawiała się kategoria „nieustalony”, która oznaczała brak możliwości wskazania gdzie dana osoba przebywa (głównie dla UdSC i ZUS). W związku z wymogiem podania informacji o liczbie cudzoziemców w poszczególnych województwach zastosowano imputację danych metodą hotdeck, w której grupy zostały utworzone według kraju obywatelstwa, płci oraz wieku, a kategoria „nieustalony” została zaimputowana wykorzystując losowanie cudzoziemców z rozkładu wielomianowego zgodnie z prawdopodobieństwami wynikającymi ze struktury cudzoziemców, dla których była informacja. Podejście to zakłada, że struktura osób, które mają status nieustalony, jest podobna do struktury tych, które posiadają określone miejsce przebywania (założenie to może nie być prawdziwe aczkolwiek jest trudne w weryfikacji, jednak w badaniu takie założenie zostało przyjęte).

Dlatego w przypadku zmiennej podregiony i województwa wyniki zaprezentowano w dwóch ujęciach – uwzględniając kategorię „nieustalony” oraz z wartościami zaimputowanymi. W przypadku drugiego ujęcia należy podchodzić z dużą ostrożnością do wyników ponieważ przedziały ufności nie uwzględniają niepewności wynikającej z imputacji danych.

W drugiej części przedstawiono pozostałe zmienne (status na rynku pracy, miejsce zamieszkania (miasto/wieś), wykształcenie, czas przebywania, sekcja PKD) uzyskane w wyniku rozszacowania uzyskanej wielkości populacji cudzoziemców według rozkładów znanych z innych źródeł i zakładając, że struktury z innych zbiorów (NSP 2011, BAEL, zezwoleń i oświadczeń z MRPiPS) odzwierciedlają struktury populacji cudzoziemców. W związku z powyższym do wyników tych należy podchodzić, jak to zostało wcześniej wspomniane z dużą ostrożnością. Przyjmuje się bowiem silne założenie, że rozkłady cudzoziemców według analizowanych zmiennych w dodatkowych źródłach danych odzwierciedlają rzeczywiste struktury w populacji cudzoziemców w Polsce i są niezmiennie w czasie (przykładowo zakłada się, że rozkład wykształcenia cudzoziemców z NSP 2011 odpowiada rozkładowi cudzoziemców według wykształcenia w latach 2015 i 2016 tj. na czas dokonywania szacunków).

Ponieważ dokonano rozszacowania wielkości populacji zgodnie ze strukturami z innych źródeł nie ma możliwości, w tym przypadku, dostarczenia rzetelnej miary jakości tych wyników. W związku z tym wyniki zaprezentowane zostały tylko w postaci oszacowań punktowych. Należy do nich podchodzić z bardzo dużą ostrożnością w związku z dość silnym założeniem dotyczącym niezmienności struktur.

<sup>16</sup> W tej części opracowania zamieszczono jedynie tabele dla wybranych zmiennych. Wyniki zostały zamieszczone w załączniku 1 „Wyniki pracy badawczej”.



## 2.10.2 Szacunki wielkości populacji wraz z ich precyzją

Na potrzeby estymacji wariancji estymatora zastosowano opisaną wcześniej metodę bootstrap. Niestety, w nielicznych przypadkach szacunki otrzymane z tej procedury były znacząco wyższe od pozostałych (np. wielkość populacji 7 mln). Ten problem dotyczył grup o najmniejszej liczebności, obywateli Armenii i Wietnamu. Podjęto decyzję o usunięciu problematycznych próbek bootstrapowych, co spowodowało ich zmniejszenie z 1000 na 946 w 2015 roku i 1000 na 991 w 2016 roku. W tabelach 11 i 12 przedstawiono wybrane wyniki estymacji uzyskane z wykorzystaniem modeli log-liniowych (wraz z precyzją – w %). Pełen wykaz tablic (na przykład dla podregionów) zawarto w Załączniku 1 „Wyniki pracy badawczej”.

**Tabela 11. Szacunek wielkości populacji cudzoziemców w Polsce**

Rok	$\hat{N}$	95% przedział ufności		Precyzja
		Dolna granica	Górna granica	
2015	507 693	369 135	724 407	17,64
2016	743 665	600 796	943 124	11,70

Źródło: Opracowanie własne.

**Tabela 12. Szacunek wielkości populacji cudzoziemców w Polsce według kraju obywatelstwa**

Rok	Kraj	$\hat{N}$	95% przedział ufności		Precyzja
			Dolna granica	Górna granica	
2015	Armenia	3 168	2 263	4 505	18,33
2016	Armenia	4 773	3 897	6 032	11,35
2015	Białoruś	19 868	14 429	27 951	17,38
2016	Białoruś	25 813	20 832	32 569	11,81
2015	Mołdawia	2 693	1 613	4 227	25,59
2016	Mołdawia	7 580	5 355	10 617	17,99
2015	Rosja	22 611	16 040	32 237	18,62
2016	Rosja	25 534	20 685	32 344	12,07
2015	Ukraina	283 714	203 946	415 732	18,55
2016	Ukraina	454 974	361 512	584 696	12,27
2015	Wietnam	7 408	5 554	9 942	15,45
2016	Wietnam	11 728	10 008	14 170	9,10
2015	kraje EU	70 901	53 579	97 126	15,63
2016	kraje EU	59 571	50 914	71 169	8,77
2015	inne	97 329	70 037	138 339	17,86
2016	inne	153 692	124 170	196 140	12,06

Źródło: Opracowanie własne.

## 2.10.3 Informacja o szacunkach dla podregionów

W pracy rozważano model, który zamiast zmiennej województwo uwzględnił zmienną podregion (73 podregiony oraz nieustalony). Niestety, model ten charakteryzował się niestabilnością dla 2015 roku objawiającą się zmniejszoną precyzją szacunków już na poziomie całego kraju. Poniżej rozkład precyzji szacunków liczby cudzoziemców ogółem oraz wybranych podregionów dla 2015 i 2016 roku. Na czerwono zaznaczono kraje obywatelstwa, dla których precyzja szacunku była nieakceptowalnie wysoka.

**Tabela 13. Szacunek wielkości populacji cudzoziemców w Polsce według kraju obywatelstwa na podstawie modelu uwzględniającego podregiony**

Rok	Kraj	$\hat{N}$	95% przedział ufności		Precyzja
			Dolna granica	Górna granica	
2015	Armenia	7264	3135	16847	48,97
2016	Armenia	6379	4971	8232	12,84
2015	Białoruś	20683	14476	32604	23,78
2016	Białoruś	26393	21132	32895	11,66
2015	Mołdawia	4211	2564	7028	26,4
2016	Mołdawia	9159	6801	12388	16,07
2015	Rosja	23548	15910	37051	25,05
2016	Rosja	27506	21795	34174	12,16
2015	Ukraina	279357	189916	439565	24,95
2016	Ukraina	427320	345488	528071	11,71
2015	Wietnam	12239	6210	23860	40,45
2016	Wietnam	12780	10745	15228	9,51
2015	EU	67587	49929	99839	20,63
2016	EU	57622	50117	67741	8,19
2015	inne	99345	68603	153658	24,27
2016	inne	153290	123111	192099	11,74

Źródło: Opracowanie własne.

**Tabela 14. Wybrane podregiony charakteryzujące się nieakceptowalną precyzją przed imputacją danych na podstawie modelu uwzględniającego podregiony**

Rok	Podregion	$\hat{N}$	95% przedział ufności		Precyzja
			Dolna granica	Górna granica	
2015	ŚWIECKI	7251	124	24238	103,57
2016	ŚWIECKI	381	271	539	16,97
2015	CHOJNICKI	353	150	902	62,42
2016	ŚWIECKI	313	250	394	11,77
2015	INOWROCŁAWSKI	471	230	1212	58,88
2016	ŚWIECKI	497	362	760	21,82

Źródło: Opracowanie własne.

W związku z tym, z racji niestabilności wyników dla roku 2015, postanowiono zrezygnować z tego modelu i skupić się na modelu wykorzystującym województwa zamiast podregionów. Należy zaznaczyć, że podjęto próbę wykorzystania modelu z 72 podregionami (i poziomem nieustalonym) jednak podobnie, jak w przypadku 73 podregionów, wynik dla 2015 roku był niestabilny.

## 2.11. Ocena spełnienia kryteriów jakości danych

Po przeprowadzeniu badania dokonano analizy i oceny jakości uzyskanych oszacowań pod względem:

- ich przydatności dla potencjalnych odbiorców,
- dokładności
- dostępności i przejrzystości,
- terminowości i punktualności,
- porównywalności
- bezpieczeństwa danych

Celem analizy **przydatności** danych jest ich ocena pod kątem wymagań odbiorców. Bezpośrednim odbiorcą przygotowanych danych będą podmioty zainteresowane monitorowaniem polityki spójności w obszarach dotyczących wspierania zatrudnienia i mobilności oraz włączenia społecznego i zwalczania ubóstwa oraz w szczególności: Ministerstwo Rodziny, Pracy i Polityki Społecznej (MRPiPS), Ministerstwo Inwestycji i Rozwoju (MIiR), Ministerstwo Spraw Wewnętrznych i Administracji (MSWiA) a także instytucje gromadzące i przetwarzające dane o cudzoziemcach jak Urząd ds. Cudzoziemców (UdSC), Zakład Ubezpieczeń Społecznych (ZUS), ponadto jednostki samorządu terytorialnego, instytucje zarządzające, pośredniczące i wdrażające Programy Operacyjne, środowisko naukowe, związki zawodowe, sektor prywatny

Dane zawarte w tablicach wynikowych (patrz Załącznik 1) mogą być wykorzystywane w analizach m.in. przy kreowaniu polityk ludnościowych i migracyjnych, w tym działań związanych z integracją cudzoziemców. Wyniki pracy badawczej będą mogły być również uwzględnione w pracach Międzyresortowych Zespołów: ds. Migracji (kierowany przez MSWiA) oraz ds. Społeczno-gospodarczych aspektów polityki migracyjnej (kierowany przez MIiR). Ponadto wyniki mogą być cennym źródłem dla opracowań naukowych oraz prac dyplomowych powiązanych z tematyką migracji.

Wyniki pracy badawczej stanowią też rozszerzenie informacji dla badania znajdującego się w programie badań statystycznych statystyki publicznej: *Zasoby migracyjne w Polsce*.

**Jednak jak zaznaczono, wyniki te należy traktować jako eksperymentalne i niebędące oficjalnymi statystykami.** Prace badawczo-rozwojowe mające na celu doskonalenie metody będą kontynuowane. Planuje się wdrożenie metody po 2022 r. po weryfikacji wynikami otrzymanymi ze spisu ludności i mieszkań 2021.

Ocena **dokładności danych** została szczegółowo zawarta w podrozdziale: 2.10 *Otrzymane szacunki i ich interpretacja*. Uzyskane wyniki dla populacji ogółem, kraju obywatelstwa, płci, wieku, województwa, podregionu oparte są na modelu i zmiennych, które bezpośrednio były wykorzystane do oszacowania badanej populacji odzwierciedlają prawdziwą niepewność przy założeniu 1) poprawności modelu, 2) braku błędów pokrycia w poszczególnych źródłach oraz przy spełnieniu założeń przyjętych w pracy badawczej. Interpretacja wyników powinna być prowadzona w kontekście przedziałów ufności, a nie oszacowań punktowych.

Z kolei w przypadku zmiennej „podregion i województwo” wyniki zaprezentowano w dwóch ujęciach – uwzględniając kategorię „nieustalony” oraz z wartościami zainputowanymi. W przypadku drugiego ujęcia wyniki należy interpretować z ostrożnością z uwagi na fakt, iż przedziały ufności nie uwzględniają niepewności wynikającej z imputacji danych.

Wielkość badanej populacji wg pozostałych zmiennych (status na rynku pracy, miejsce zamieszkania miasto/wieś, poziom wykształcenia, czas przebywania, sekcje PKD) została rozszacowana zgodnie ze strukturami następujących źródeł: NSP-2011, BAEL, zezwolenia i oświadczenia MRPiPS. Jednak w tym przypadku nie ma możliwości dostarczenia rzetelnej miary jakości tych wyników. Zatem wyniki zostały zaprezentowane wyłącznie w postaci oszacowań punktowych. Należy do nich podchodzić z bardzo dużą ostrożnością w związku z dość silnym założeniem dotyczącym niezmienności struktur.

Ocena **dostępności i przejrzystości danych** dotyczyła łatwości dostępu do danych dla różnych podmiotów.

- Pozyskane dane będą udostępnione w formie: niniejszego raportu końcowego z analizą danych wynikowych w formie opisowej i graficznej,
- tablic wynikowych w formacie xls.,
- syntetycznego opisu w języku nietechnicznym,
- prezentacji wyników pracy badawczej.

W raporcie uwzględniono opis metody szacowania zasobu cudzoziemców na krajowym rynku, a także załączono dodatkowe opracowania, tj. przegląd najważniejszych metod statystycznych w kontekście estymacji populacji trudnych do zbadania zawierający syntetyczny opis poszczególnych technik oraz wskazanie praktycznego ich zastosowania w wybranych krajach, w tym również w odniesieniu do populacji nieudokumentowanych/nierejestrowanych cudzoziemców (załącznik 2) oraz opis modeli log-liniowych zastosowanych w pracy badawczej (załącznik 3).

Raport końcowy udostępniony został na stronie internetowej Głównego Urzędu Statystycznego w obszarze *Statystyka dla polityki spójności*.

**Ocena terminowości i punktualności.** Praca badawcza realizowana była w 2017 i 2018 roku. Czas pomiędzy zakończeniem badania a datą udostępnienia wyników wynosi 3 miesiące. Wszystkie prace przewidziane w harmonogramie zostały wykonane zgodnie z założonymi terminami. Przeprowadzona ocena punktualności i terminowości badania była pozytywna. Przeprowadzone badanie dotyczyło lat 2015 i 2016. Zapewniło to ich **porównywalność** w czasie oraz w przestrzeni geograficznej, tj. pomiędzy województwami.

Ocena aspektów związanych z **bezpieczeństwem danych**. Wszystkie działania na zbiorach danych odbywały się w bezpiecznym środowisku informatycznym Operacyjnej Bazy Mikrodanych (OBM) oraz ze stosowaniem środków technicznych i organizacyjnych mających na celu zabezpieczenie tych danych przed ich udostępnieniem osobom nieupoważnionym, a także dopuszczaniem do tych danych wyłącznie osób posiadających odpowiednie upoważnienie i prowadzeniem ewidencji osób upoważnionych do przetwarzania danych.

### 3. Podsumowanie i rekomendacje

#### Conclusions and recommendations

Wybrana do oszacowania zasobu cudzoziemców na krajowym rynku pracy metoda *capture-recapture* bazująca na modelach log-liniowych jak dotąd nie była stosowana w badaniach statystycznych GUS. Jedynymi doświadczeniami, z których można było skorzystać są empiryczne badania zrealizowane przez holenderskich i norweskich badaczy.

W pracy badawczej oszacowano wielkość zasobu cudzoziemców przebywających w Polsce w końcu 2015 r. i 2016 r. w podziale na województwa. Biorąc pod uwagę zmienne planowane do opracowania, uzyskano wyniki zgodnie założeniami, tj. opracowano wszystkie planowane cechy z listy obowiązkowej, ponadto oszacowano sekcję PKD dla pracujących cudzoziemców, okres przebywania do 3 miesięcy oraz charakter miejsca przebywania (miasto, wieś). Wszystkie wyniki zostały zaprezentowane na poziomie ogólnopolskim, a część na poziomie województw (dane według płci, grup wieku, kraju obywatelstwa przedstawiono w układzie wojewódzkim).

Wyszacowano również wielkość populacji cudzoziemców pracujących w Polsce i zaprezentowano dane również w układzie wojewódzkim. Przedstawiono zatem informacje o zróżnicowaniu terytorialnym zjawiska zatrudniania cudzoziemców w Polsce. Niestety, dostępne źródła danych w znacznym stopniu uniemożliwiły otrzymanie szczegółowych wyników na poziomie podregionów i regionów warszawskiego stołecznego i mazowieckiego regionalnego z akceptowalną precyzją. Dla podregionów i wyżej wymienionych regionów zaprezentowano statystyki dotyczące zasobu cudzoziemców.

Konieczne będzie prowadzenie dalszych prac rozwojowo-badawczych, doskonalących metodę badawczą. Podstawowym warunkiem jest rozszerzenie zakresu zbiorów administracyjnych – pozyskiwanych obecnie przez statystykę publiczną w ramach PBSSP – m.in. o cechy identyfikacyjne, tak aby zakres ten odpowiadał literalnie celom badawczym. Niezbędne jest również podjęcie prac nad rozpoznaniem innych źródeł, które ze względu na swój charakter i zakres mogą być niezwykle cennym źródłem do weryfikacji charakterystyki cudzoziemców, np. rejestr policji dot. cudzoziemców podejrzanych o popełnienie przestępstw, zbiory Państwowej Inspekcji Pracy w zakresie kontroli legalności zatrudniania cudzoziemców, Komendy Głównej Straży Granicznej w zakresie legalności pobytów lub zbiory Ministerstwa Spraw Zagranicznych w zakresie wiz. W tym celu konieczne będzie nawiązanie bądź zintensyfikowanie współpracy z gestorami poszczególnych rejestrów i baz danych.

W pracach nad doskonaleniem metody badawczej niezbędne będzie przetestowanie metody bezpośredniego wyznaczania populacji pracujących cudzoziemców w ramach zintegrowanego zbioru danych, powstałego na podstawie łączenia (zarówno deterministycznego jak i probabilistycznego) wszystkich dostępnych a zarazem istotnych źródeł danych administracyjnych oraz zbiorów z badań statystycznych. Warto również w przyszłości rozważyć inne metody szacowania populacji trudnych do zbadania, m.in. wykorzystany w Norwegii na potrzeby estymacji cudzoziemców hierarchiczny model mieszany Poissona. Jego zastosowanie wymagać będzie jednak wspomnianego rejestru policji w zakresie cudzoziemców podejrzanych o popełnienie przestępstw.

Będą podejmowane także próby uzyskania bardziej miarodajnych oszacowań dla podregionów i regionów warszawskiego stołecznego i mazowieckiego regionalnego oraz rozkładów cech: okres przebywania, status na rynku pracy, PKD oraz charakter miejsca zamieszkania wraz z uwzględnieniem podziału według województw. Oszacowania zostaną uzupełnione o cechę „grupy zawodów”. Ponadto podjęte zostaną działania nad wyodrębnieniem z oszacowanej populacji tej jej części, która pozostaje poza zbiorami rejestrowymi.

Doskonalenie metody badawczej będzie realizowane w trybie ciągłym aż do spisu ludności i mieszkań 2021, którego wyniki pozwolą na ocenę jakości uzyskanych oszacowań. Planuje się, że po 2022 roku metoda szacowania zasobu cudzoziemców na krajowym rynku pracy według pożądaných cech demograficzno-społecznych i ekonomicznych zostanie wdrożona do ciągłej produkcji statystycznej. Niemniej jednak wychodząc naprzeciw zapotrzebowaniu zgłaszanemu przez użytkowników danych planuje się udostępnianie rezultatów oszacowań (jako wyniku badania eksperymentalnego) uzyskanych w ramach



Rzeczpospolita  
Polska

Unia Europejska  
Fundusz Spójności



testowania wskazanych powyżej metod, tj. metod wykorzystujących łączenie deterministyczne i probabilistyczne oraz analizy wrażliwości na złamania założeń metody *capture-recapture*, jak i stosowanych modeli.

## Bibliografia Bibliography

- Brzezińska J. (2015), *Analiza logarytmiczno-liniowa. Teoria i zastosowania z wykorzystaniem programu R*, Wydawnictwo C.H. Beck, Warszawa.
- Coumans A.M. Cruyff M., Van der Heijden., Wolf J., Schmeets H. (2017), *Estimating homelessness in the Netherlands using a capture-recapture approach*, *Social Indicators Research*, 130(1): 189-212.
- Gerritse S.Ch. (2016), *An application of population size estimation to official statistics: Sensitivity of model assumptions and the effect of implied coverage*, PhD thesis, Utrecht University.
- Goodman L.A. (1964), *Simple Methods for Analyzing Three-factor Interaction in Contingency Tables*, *Journal of the American Statistical Association*, 59, 319-352.
- Goodman L.A. (1968), *The Analysis of Cross-Classified Data: Independence, Quasi-independence, and Interactions in Contingency Tables with or without Missing Values*, *Journal of the American Statistical Association*, 63, 1091-1131.
- Goodman L.A. (1969), *On Partitioning  $\chi^2$  and Detecting Partial Association in Three-way Contingency Tables*, *Journal of the Royal Statistical Society. Series B (Methodological)*, 486-498.
- Peter G.M., Van Der Heijden., Whittaker J., Cruyff M., Bakker B., Van der Vliet R. (2012), *People Born in the Middle East but Residing in the Netherlands: Invariant Population Size Estimates and the Role of Active and Passive Covariates*, *The Annals of Applied Statistics*, Vol. 6, No. 3, 831-852.
- Stokes M.E., Davis C.S., Koch G.G. (2012), *Categorical Data Analysis Using SAS, Third Edition*, SAS Institute Inc., Cary, NC, USA.
- Zwane E., van der Heijden P.G.M. (2005), *Population estimation using the multiple system estimator in the presence of continuous covariates*, *Statistical Modelling*, 5(1): 39-52.



Fundusze  
Europejskie  
Pomoc Techniczna



Rzeczpospolita  
Polska

Unia Europejska  
Fundusz Spójności



## Spis załączników *List of annexes*

1. Załącznik 1 Wyniki pracy badawczej
2. Załącznik 2 Przegląd literatury w zakresie szacowania wielkości populacji trudnych do zbadania
3. Załącznik 3 Opis modeli log-liniowych wykorzystanych w pracy badawczej