

GŁÓWNY URZĄD STATYSTYCZNY
STATISTICS POLAND

Raport końcowy
Final report

Prace eksperymentalne
Experimental studies

Stratyfikacja dochodowa mieszkańców miast

Income stratification of city dwellers



Warszawa, 2020
Warsaw, 2020

GŁÓWNY URZĄD STATYSTYCZNY
STATISTICS POLAND

Raport końcowy
Final report

Prace eksperymentalne
Experimental studies

Stratyfikacja dochodowa mieszkańców miast

Income stratification of city dwellers



Warszawa, 2020
Warsaw, 2020

Jednostka opracowująca raport

The unit preparing the report

Centrum Badań i Edukacji Statystycznej GUS
Centre for Statistical Research and Education

Kierownik projektu

Project manager

Sylwia Filas-Przybył

Zespół autorski

Editorial team

Krzysztof Basarbowicz, Sylwia Filas-Przybył, Maciej Kaźmierczak, Tomasz Klimanek, Jacek Kowalewski, Kazimierz Kruszka, Krystyna Siwiak, Dorota Stachowiak, Małgorzata Stawikowska, Anna Stępnia

Tłumaczenie

Translation

Grzegorz Grygiel

ISSN

ISBN

Raport dostępny na stronie

Report available on

<http://stat.gov.pl/>

Przy publikowaniu danych GUS prosimy o podanie źródła

When publishing Statistics Poland data — please indicate the source

Druk publikacji współfinansowany ze środków Unii Europejskiej w ramach projektu „Statystyka dla polityki spójności 2019-2020. Wsparcie systemu monitorowania i ewaluacji polityki spójności w perspektywie finansowej 2014-2020 oraz monitorowania i ewaluacji polityki spójności po 2020 r.”

Publication print co-financed by the European Union within the project „Statistics for Cohesion Policy 2019-2020. Support for the Monitoring System of Cohesion Policy in Financial Perspective 2014-2020 as well as Monitoring and Evaluation of Cohesion Policy after 2020

Przedmowa

Niniejszy raport „Stratyfikacja dochodowa mieszkańców miast” jest propozycją nowego podejścia metodologicznego w badaniach dochodów ludności na poziomie lokalnym. Wypełnia ono częściowo lukę informacyjną istniejącą w statystyce publicznej w tym zakresie.

Opracowanie to jest jedną z prac eksperymentalnych prowadzonych w GUS w ramach projektu „Statystyka dla polityki spójności 2019-2020. Wsparcie systemu monitorowania i ewaluacji polityki spójności w perspektywie finansowej 2014-2020 oraz monitorowania i ewaluacji polityki spójności po 2020 r.” Praca ta jest zgodna z nowoczesnym paradygmatem prowadzenia badań statystycznych, w myśl którego należy rezygnować z obciążenia respondentów koniecznością wypełnienia kolejnych kwestionariuszy badawczych, na rzecz jak najszerszego wykorzystania informacji zawartych w rejestrach administracyjnych.

Jedną z podstawowych zmiennych opisujących poziom życia ludności są jej dochody. Dochody osobiste stanowią o możliwości zaspokojenia różnych potrzeb ludności. W zasobach statystyki publicznej, zwłaszcza w odniesieniu do miast i obszarów wewnątrzmijskich, brakowało informacji dotyczących charakterystyk opisujących poziom i zróżnicowanie dochodów ludności. Dotychczasowe próby podejmowania tej tematyki ograniczały się do prowadzenia indywidualnych badań ankietowych na wybranych obszarach.

Publikacja zawiera ogólną charakterystykę metodyki przeprowadzonego badania „Stratyfikacja dochodowa mieszkańców miast” oraz niektóre jego wyniki. Pokazano je w formie licznych kartogramów i tablic z danymi, a całość wzbogacono komentarzem analitycznym opisującym wybrane aspekty badanego przedmiotu.

Przekazując do rąk Państwa niniejsze opracowanie, mam nadzieję, że zawarty w niej materiał będzie przydatny wszystkim, których interesuje prezentowana tematyka. Ze względu na eksperymentalny charakter badania, zachęcam do podzielenia się uwagami po lekturze tego raportu.

Preface

The following report on Income stratification of city dwellers is a proposal of a new methodological approach to the study of personal incomes at the local level. It partially bridges the existing information gap in official statistics on this topic.

The study is an example of experimental research carried out by Statistics Poland as part of the project “Statistics for the Cohesion Policy 2019-2020. Support for the system of monitoring and evaluation of cohesion policy in the financial perspective 2014-2020 and for the monitoring and evaluation of cohesion policy after 2020”. The study is in line with the modern paradigm of statistical data collection, which stipulates that instead of burdening respondents with the obligation of completing multiple questionnaires, national statistical institutions should make the widest possible use of information contained in administrative registers.

One of the basic variables describing the population’s standard of living is income. Personal income earned by individuals enables them to meet their various needs. So far, official statistics, especially concerning cities and inner-city areas, have not included information about the characteristics describing the level of and variation in personal incomes. Previous attempts to bridge this gap have been limited to conducting individual surveys in selected areas.

The report contains a general description of the methodology applied to obtain Income stratification of city dwellers and some of the results, which are shown in the form of numerous choropleth maps and tables and are accompanied by analytical commentary explaining selected aspects of the subject matter.

I hope that the insights provided by the report will be useful to all those interested in the topic. Given the experimental nature of the study, all readers are encouraged to share their comments concerning its content.

Spis treści

Contents

Str.
Page

Przedmowa	3
Preface	4
Spis tablic	7
List of tables.	7
Spis map	9
List of maps	9
Spis rycin	10
List of figures	10
Objaśnienia skrótów i symboli	11
List of symbols, acronyms and abbreviation.	11
Wstęp	12
Introduction	15
Synteza	18
Executive summary	19
I. Metodologia badania	20
I. Research methodology.	20
1. Zakres podmiotowy i czasowy badania	20
1. The scope and reference period of the study.	20
2. Zakres terytorialny badania	20
2. The territorial coverage of the study	20
3. Źródła danych	20
3. Data sources	20
3.1. Dane Ministerstwa Finansów	20
3.1. Data of the Ministry of Finance	20
3.2. Dane Głównego Urzędu Statystycznego	21
3.2. Data of Statistics Poland.	21
3.3. Dane przestrzenne	22
3.3. Spatial data	22
3.4. Środowisko programistyczne	23
3.4. Programming environment	23
4. Przegląd literatury pod kątem mierników charakteryzujących sytuację dochodową mieszkańców miast oraz metod klasyfikacji i analiz przestrzennych	23
4. A review of literature on measures of income of the urban population and methods of classification and spatial analysis	23
5. Wypracowanie metod postępowania z danymi atrybutowymi i przestrzennymi.	27
5. Development of methods of handling attribute and spatial data	27

5.1. Przygotowanie danych atrybutowych	27
5.1. Preparation of attribute data	27
5.2. Przygotowanie danych przestrzennych	30
5.2. Preparation of spatial data	30
6. Detekcja i sposoby postępowania z obserwacjami odstającymi.	31
6. Methods of outlier detection and treatment.	31
II. Analiza uzyskanych wyników.	48
II. Analysis of results	48
III. Ocena spełnienia kryteriów jakości otrzymanych wyników.	85
III. Evaluation of the degree to which the results meet the quality criteria	85
IV. Uwagi końcowe	89
IV. Concluding remarks.	89
Bibliografia	91
References	91
Załączniki	94
Appendices	94

Spis tablic

List of tables

	Str. Page
Tablica 1. Struktura Zintegrowanej Bazy Populacji	21
Tablica 2. Miary poziomu	24
Tablica 3. Miary zróżnicowania i nierówności dochodowej	25
Tablica 4. Mierniki autokorelacji przestrzennej.	26
Tablica 5. Zmienne ze zbiorów Ministerstwa Finansów istotne z punktu widzenia badania .	27
Tablica 6. Zmienne w Bazie Przychodów Mieszkańców Miast	28
Tablica 7. Zmienne w Bazie Dochodów Mieszkańców Miast	29
Tablica 8. Zmienne w Bazie Dochodów Mieszkańców Miast Wojewódzkich przypisanych do siatki kwadratów.	29
Tablica 9. Liczba kwadratów o boku 500 metrów oraz 1000 m według jednostek przestrzennych	30
Tablica 10. Wykaz zastosowanych podejść i ujęć w analizie rozkładu przychodów/dochodów.	34
Tablica 11. Wpływ zastosowanych podejść na wybrane charakterystyki poziomu i zróżnicowania przychodów podatników („top-down”).	36
Tablica 12. Wpływ zastosowanych podejść na wybrane charakterystyki poziomu i zróżnicowania dochodów pracowników najemnych („top-down”)	37
Tablica 13. Wpływ zastosowanych podejść na wybrane charakterystyki poziomu i zróżnicowania przychodów podatników („bottom-up”)	37
Tablica 14. Wpływ zastosowanych podejść na wybrane charakterystyki poziomu i zróżnicowania dochodów pracowników najemnych („bottom-up”).	38
Tablica 15. Wskaźniki zróżnicowania przychodów mieszkańców miast według grup wielkościowych miast w 2018 r.	51
Tablica 16. Wartość środkowa przychodów (w złotych) według województw i grup wielkościowych miast w 2018 r.	54
Tablica 17. Dochody według grup decylowych w 2018 r.	56
Tablica 18. Wskaźniki zróżnicowania dochodów mieszkańców miast według grup wielkościowych miast	59
Tablica 19. Podatnicy i dochody w najwyższym przedziale dochodowym w 2018 r. według województw	65
Tablica 20. Wybrane miary nierówności dochodów w 2018 r. według województw	66
Tablica 21. Mediana dochodów mieszkańców miast	69
Tablica 22. Wskaźniki zróżnicowania dochodów mieszkańców miast.	70
Tablica 23. Wskaźniki zróżnicowania dochodów mieszkańców miast	72
Tablica 24. Wskaźniki zróżnicowania dochodów mieszkańców miast	72

Tablica 25. Liczba deklaracji podatkowych pozyskanych do badania – przychody	85
Tablica 26. Liczba podatników – przychody	86
Tablica 27. Liczba złożonych formularzy przez podatników – przychody	86
Tablica 28. Liczba deklaracji podatkowych pozyskanych do badania – dochody	87
Tablica 29. Liczba podatników – dochody	87

Spis map

List of maps

	Str. Page
Mapa 1. Mediana przychodów (w złotych) ludności w 2018 r. według gmin w kraju.	49
Mapa 2. Mediana przychodów (w złotych) według województw i grup wielkościowych miast w 2018 r.	53
Mapa 3. Mediana dochodów (w złotych) ludności w 2018 r. według gmin w kraju	55
Mapa 4. Mediana dochodów (w złotych) według województw i grup wielkościowych miast w 2018 r.	64
Mapa 5. Współczynnik Giniego w miastach ogółem oraz grupach wielkościowych miast w 2018 r. według województw	67
Mapa 6. Mediana przychodów (w złotych) ludności w 2018 r. według miast	68
Mapa 7. Mediana dochodów (w złotych) ludności w 2018 r. według miast	69
Mapa 8. Mediana dochodów – Poznań	76
Mapa 9. Relacja decyla dziewiątego do decyla pierwszego (P90/P10) – Poznań	77
Mapa 10. Wskaźnik zróżnicowania kwintylowego (S80/S20) – Poznań	78
Mapa 11. Współczynnik Giniego – Poznań	79
Mapa 12. Relacja decyla dziewiątego do decyla pierwszego (P90/P10) – LISA – Poznań. . .	81
Mapa 13. Wskaźnik zróżnicowania kwintylowego (S80/S20) – LISA – Poznań	82
Mapa 14. Współczynnik Giniego – LISA – Poznań	83
Mapa 15. Obszary o wysokim stopniu zróżnicowania dochodów – synteza – Poznań	84

Spis rycin

List of figures

	Str. Page
Rycina 1. Poszukiwanie optymalnej wartości parametru lambda do transformacji Box'a-Cox'a.	34
Rycina 2. Rozkład przychodów podatników zamieszkujących miasta po transformacji Box'a-Cox'a.	35
Rycina 3. Wpływ zastosowanych podejść na medianę przychodów podatników w miastach („bottom-up”)	39
Rycina 4. Wpływ zastosowanych podejść na medianę przychodów podatników w miastach („bottom-up”)	40
Rycina 5. Wpływ zastosowanych podejść na współczynnik Giniego przychodów podatników w miastach („bottom-up”)41
Rycina 6. Wpływ zastosowanych podejść na współczynnik Giniego przychodów podatników w miastach („bottom-up”) cd.	42
Rycina 7. Wpływ zastosowanych podejść na współczynnik Giniego przychodów podatników w miastach („bottom-up”) dok.	43
Rycina 8. Wpływ zastosowanych podejść na współczynnik Giniego dochodów pracowników najemnych w miastach („bottom-up”).	44
Rycina 9. Wpływ zastosowanych podejść na współczynnik Giniego dochodów pracowników najemnych w miastach („bottom-up”) cd.	45
Rycina 10. Wpływ zastosowanych podejść na współczynnik Giniego dochodów pracowników najemnych w miastach („bottom-up”) dok.	46
Rycina 11. Mediana przychodów w złotych według grup wielkościowych miast w 2018 r.51
Rycina 12. Kwartyły dochodów (w złotych) według grup wieku i płci w 2018 r.57
Rycina 13. Mediana dochodów (w złotych) według grup wielkościowych miast w 2018 r.	58
Rycina 14. Krzywa Lorenza – współczynnik Giniego dla miast dużych i średnich	60
Rycina 15. Krzywa Lorenza – współczynnik Giniego dla miast dużych i małych61
Rycina 16. Krzywa Lorenza – współczynnik Giniego dla miast średnich i małych	62
Rycina 17. Krzywa Lorenza – współczynnik Giniego dla Nowych Skalmierzyc i Podkowy Leśnej.71
Rycina 18. Krzywa Lorenza – współczynnik Giniego dla Poznania i Warszawy74
Rycina 19. Krzywa Lorenza – współczynnik Giniego dla Poznania i Białegostoku75

Objaśnienia skrótów i symboli

List of symbols, acronyms and abbreviation

Skrót <i>Abbreviation</i>	Znaczenie <i>Meaning</i>
EUROSTAT	Urząd Statystyczny Unii Europejskiej <i>Statistical Office of the European Union</i>
PESEL	Powszechny Elektroniczny System Ewidencji Ludności <i>Universal Electronic System for Registration of the Population</i>
PIT	podatek dochodowy od osób fizycznych – podatek bezpośredni obejmujący dochody uzyskiwane przez osoby fizyczne. „PIT” to również nazwa druków urzędowych udostępnianych w Polsce przez Ministerstwo Finansów, na których podatnicy składają deklaracje rozliczeniowe dotyczące podatku od dochodów osobistych <i>Personal Income Tax – a direct tax levied on incomes earned by natural persons. „PIT” is also the acronym used to refer to official tax forms published by the Ministry of Finance, which are used for filing a personal income tax return</i>
PRG	państwowy rejestr granic i powierzchni jednostek podziałów terytorialnych kraju <i>the register of borders and areas for units of the country's territorial division</i>
SHP	format plików grafiki wektorowej, stosowany dla danych geoprzestrzennych używanych w Systemach Informacji Geograficznej <i>a geospatial vector data format for geographic information system (GIS)</i>
SIMC	system identyfikatorów i nazw miejscowości <i>a system of identifiers and names of localities</i>
TERYT	krajowy rejestr urzędowy podziału terytorialnego kraju <i>the Territorial Division Register</i>
ULIC	Centralny Katalog Ulic <i>the Central Catalogue of Street Names</i>
ZBP	Zintegrowana Baza Populacji <i>the Integrated Population Database</i>

Wstęp

Niniejszy raport jest podsumowaniem pracy badawczej pt. „Stratyfikacja dochodowa mieszkańców miast”, zrealizowanej w ramach projektu „Statystyka dla polityki spójności 2019–2020. Wsparcie systemu monitorowania i ewaluacji polityki spójności w perspektywie finansowej 2014–2020 oraz monitorowania i ewaluacji polityki spójności po 2020 r.” Badanie współfinansowane było ze środków Programu Operacyjnego Pomoc Techniczna 2014-2020, w ramach celu tematycznego „Wspieranie zatrudnienia i mobilności pracowników” (8. cel tematyczny) oraz „Wspieranie włączenia społecznego i walka z ubóstwem” (9. cel tematyczny).

Celem pracy badawczej było dokonanie stratyfikacji dochodowej mieszkańców miast w Polsce na podstawie danych pochodzących ze źródeł administracyjnych. W kontekście tak określonego celu baza źródłowa pozyskana z Ministerstwa Finansów umożliwiła określenie struktury (rozkładu) dochodów tylko tych podatników mieszkających w miastach Polski w 2018 roku, którzy byli pracownikami najemnymi. Dla całej zbiorowości podatników analizę można było przeprowadzić tylko na podstawie informacji o uzyskanych przez nich przychodach. Nie jest to rozwiązanie w pełni satysfakcjonujące, ale daje jednak pewne przybliżenie do sytuacji pożądanej w badaniu i może stanowić punkt wyjścia do dalszych poszukiwań.

Główny Urząd Statystyczny opracowuje i publikuje dane o dochodach (zwłaszcza w odniesieniu do gospodarstw domowych) pochodzące z badań reprezentacyjnych. Stąd najczęściej dostępne informacje dotyczą poziomu wojewódzkiego lub wyższych poziomów przestrzennych, które zdaniem zainteresowanych tematem badaczy nie mogą być wykorzystane do bardziej precyzyjnych analiz przestrzenno-ekonomicznych. Brak danych dla niskich poziomów przestrzennych jest szczególnie dotkliwy właśnie dlatego, że zróżnicowanie dochodów osobistych mieszkańców uwidacznia się przede wszystkim na niższych poziomach agregacji przestrzennej. Informacje dotyczące strony dochodowej mieszkańców są niezbędne w kształtowaniu polityki regionalnej i społecznej, ale przede wszystkim stanowią miernik poziomu rozwoju społeczno-gospodarczego danego obszaru.

Ranga problemu, którym są nierówności przestrzenne w dochodach ludności na poziomie lokalnym, przy jednoczesnym braku odpowiednich informacji, skłoniła statystykę publiczną do podjęcia wysiłku zmierzającego do wypracowania metodologii badania w oparciu o przegląd literatury pod kątem mierników charakteryzujących sytuację dochodową oraz metod klasyfikacji i analiz przestrzennych, a także dostarczenia informacji na temat poziomu i zróżnicowania rozkładu dochodów mieszkańców miast w Polsce. W badaniu wykorzystano dane pochodzące z rejestrów administracyjnych, dzięki czemu nie zwiększa ono obciążenia respondentów statystyki publicznej.

Nierówności społeczne i ekonomiczne, w tym dochodowe mają istotny wpływ na rozwój regionalny i lokalny. Z tego też powodu ww. zagadnienia znalazły odzwierciedlenie w szeregu światowych, europejskich oraz krajowych politykach i programach. Zgodnie z Agendą na rzecz zrównoważonego rozwoju 2030 przyjętą we wrześniu 2015 r. przez kraje należące do Organizacji Narodów Zjednoczonych „współczesny wysiłek modernizacyjny powinien koncentrować się na wyeliminowaniu ubóstwa we wszystkich jego przejawach, przy równoczesnej realizacji szeregu celów gospodarczych, społecznych i środowiskowych”. Jednym z podstawowych celów Agendy jest „uczynienie miast i osiedli ludzkich bezpiecznymi, stabilnymi, zrównoważonymi oraz sprzyjającymi włączeniu społecznemu”.

Znaczenie zrównoważonego rozwoju społeczno-gospodarczego, konieczność przeciwdziałania dysproporcjom i wykluczeniu społecznemu oraz potrzeba wspierania terytoriów i społeczności dotkniętych marginalizacją zostały podkreślone w trzech kluczowych, krajowych dokumentach strategicznych, na których opiera się polityka miejska w Polsce: „Krajowej Polityce Miejskiej 2023”, „Strategii na rzecz Odpowiedzialnego Rozwoju do roku 2020 (z perspektywą do 2030 r.)” oraz „Krajowej Strategii Rozwoju Regionalnego 2030”.

Celem strategicznym przyjętej w październiku 2015 r. „Krajowej Polityce Miejskiej 2023” jest wzmocnienie zdolności miast i miejskich obszarów funkcjonalnych do tworzenia zrównoważonego rozwoju, miejsc pracy i poprawy jakości życia mieszkańców. W dokumencie zwrócono uwagę na fakt, że „dysproporcje społeczne wewnątrz miast są często większe niż pomiędzy regionami”.

„Strategia na rzecz Odpowiedzialnego Rozwoju do roku 2020 (z perspektywą do 2030 r.)” przyjęta w lutym 2017 r. wskazuje, że „zrównoważony rozwój społeczny i regionalny to harmonijny rozwój całego kraju, wrażliwy na terytorialną różnorodność i jej atuty, a jednocześnie dbający o zapewnienie całemu społeczeństwu wysokiej jakości życia. Efektywność działań rozwojowych jest uwarunkowana adekwatnym dostosowaniem ich zakresu i mechanizmu wdrażania do specyfiki poszczególnych obszarów: indywidualnych potencjałów, deficytów i wzajemnych zależności. Zakładane efekty działań rozwojowych podejmowanych w tym obszarze to przede wszystkim nowe miejsca pracy dobrej jakości, przekładające się na wyższe dochody mieszkańców i większą bazę dochodową samorządów”.

W przyjętej we wrześniu 2019 r. „Krajowej Strategii Rozwoju Regionalnego 2030” podkreślono, że „wdrażaniu działań rozwojowych na obszarach zagrożonych trwałą marginalizacją i w miastach średnich tracących funkcje społeczno-gospodarcze nie sprzyja utrzymujące się ubóstwo, braki w wyposażeniu podstawowej infrastruktury oraz niska dostępność transportowa i usługowa tych obszarów”. Dla polityki regionalnej oznacza to „konieczność koncentracji wsparcia na przeciwdziałaniu nadmiernym dysproporcjom rozwojowym, zarówno między regionami, jak i w samych regionach. Z drugiej strony, obok działań o charakterze wyrównawczym, interwencje w ramach polityki regionalnej odnosząc się będą do wszystkich terytoriów uwzględniając ich pozycję rozwojową, w tym w szczególności jakość kapitału ludzkiego”.

Mając na uwadze zapisy zawarte w ww. dokumentach strategicznych, konieczne jest dostarczenie twardych danych statystycznych, pozwalających na monitorowanie zmian i stopnia osiągnięcia celów.

Zaspokojenie zgłaszanych potrzeb informacyjnych w zakresie dochodów ludności na poziomie lokalnym stało się inspiracją do podjęcia pracy badawczej „Stratyfikacja dochodowa mieszkańców miast”, gdzie istotnym komponentem jest wykorzystanie danych zawartych w źródłach administracyjnych. Wyniki pracy badawczej przedstawiono w niniejszym raporcie.

Zasadnicze części raportu - część metodologiczna i analityczna – zostały poprzedzone rozdziałem zatytułowanym Synteza. Wypunktowano w niej najistotniejsze informacje dotyczące przychodów podatników/dochodów pracowników najemnych zamieszkujących miasta w 2018 r.

W rozdziale poświęconym metodologii badania opisano szczegółowo wszystkie etapy prac poczynawszy od omówienia źródeł danych po opis detekcji i sposobów postępowania z obserwacjami odstającymi.

Część analityczna zawarta w rozdziale Analiza uzyskanych wyników prezentuje zarówno sytuację odnoszącą się do przychodów wszystkich podatników oraz dochodów pracowników najemnych zamieszkujących miasta przedstawioną w różnych aspektach przestrzennych i typologicznych.

W Uwagach końcowych znalazły się m.in. spostrzeżenia i sugestie dotyczące perspektywy dalszych prac z wykorzystaniem źródeł administracyjnych w kontekście kolejnych programów badań statystycznych statystyki publicznej oraz konieczności stałej merytorycznej współpracy służb statystyki publicznej z gestorami danych administracyjnych, w tym z Ministerstwem Finansów.

Efektom przeprowadzonych prac jest niniejszy raport końcowy, a jego integralną część stanowią załączniki do raportu:

- Struktury pozyskanych zbiorów danych;
- Schemat przygotowania baz danych;
- Kartogramy dla miast wojewódzkich;
- Baza danych zawierająca informacje dla:

- miast imiennie, zarówno dla przychodów podatników (mediana), jak i dochodów pracowników najemnych (mediana, relacja decyla dziewiątego do decyla pierwszego P90/P10, wskaźnik zróżnicowania kwintylowego S80/S20, współczynnik Giniego),
- miast wojewódzkich w układzie siatki kwadratów dla dochodów pracowników najemnych (mediana w przedziałach kwartylowych, relacja decyla dziewiątego do decyla pierwszego P90/P10, wskaźnik zróżnicowania kwintylowego S80/S20, współczynnik Giniego).

W pracy badawczej wykorzystano dane zawarte w rejestrach podatkowych Ministerstwa Finansów za rok 2018. Pozyskane dane dotyczące przychodów obejmowały dziesięć rodzajów deklaracji podatkowych PIT – formularze PIT-28, PIT-28S, PIT-36, PIT-36S, PIT-36L, PIT-36LS, PIT-37, PIT-38, PIT-39 oraz PIT-40A/11A. Natomiast dane dotyczące dochodów pracowników najemnych (tj. otrzymujących wynagrodzenie z tytułu stosunku pracy, służbowego, spółdzielczego i z pracy nakładczej, a także zasiłki pieniężne z ubezpieczenia społecznego wypłacane przez zakład pracy) pochodziły z formularzy PIT-11. Zarówno dochody jak i przychody podawane są jako sumy roczne.

Introduction

The following report summarises the results of the study entitled “Income stratification of city dwellers” conducted as part of the project “Statistics for the Cohesion Policy 2019-2020. Support for the system of monitoring and evaluation of cohesion policy in the financial perspective 2014-2020 and for the monitoring and evaluation of cohesion policy after 2020”. The study was co-financed under the Operational Programme ‘Technical Assistance 2014-2020’ to boost employment and employee mobility (8th thematic objective) and to foster social inclusion and the fight against poverty (9th thematic objective).

The purpose of the study was to stratify Poland’s city dwellers in terms of income using data from administrative sources. Based on information included in a database obtained from the Ministry of Finance, it was possible to identify the structure (distribution) of taxable personal incomes earned in 2018 only by city dwellers who were in paid employment. As regards the entire population of taxpayers, the analysis could only be conducted based on information about taxable revenues (including deductible expenses). While not being the most satisfactory option, it does provide an approximation of what the study aims to investigate and can be regarded as the starting point for further research.

Because Statistics Poland processes and publishes income data (especially for households) from sample surveys, the resulting statistics are usually available only at the level of province or even higher levels of spatial aggregation, which are of little use to researchers interested in conducting more detailed socio-economic analyses. The lack of data for lower levels of spatial aggregation is particularly aggravating precisely because the variation in personal incomes becomes evident mainly at these lower levels. Moreover, information about incomes earned by the population is indispensable for conducting regional and social policy, but, above all, it is used as a measure of socio-economic development of a given area.

The relevance of spatial inequalities in personal incomes at the local level, combined with the lack of information in this respect, have prompted Statistics Poland to develop a methodological approach based on the review of the literature on measures of income and methods of classification and spatial analysis and to provide statistics about the level and variation in the distribution of incomes earned by inhabitants of Polish towns and cities. Because the study is based on data from administrative registers, it does not increase the respondent burden.

Social and economic inequalities, including income inequalities, have a considerable impact on the regional and local development. For this reason, these issues have been addressed in numerous global, European and national policies and programmes. According to the 2030 Agenda for Sustainable Development, launched in September 2015 during the United Nations Summit in New York, “contemporary modernization efforts should focus on eliminating poverty in all its manifestations, while achieving a number of commercial, social, and environmental goals”. One of the basic goals of the Agenda is to “make cities and human settlements inclusive, safe, resilient and sustainable”.

The importance of sustainable socio-economic development, the need to counteract social disproportions and exclusion and to support marginalised territories and communities have been emphasised in three key strategic documents, which outline the framework of Poland’s urban policy: the National Urban Policy 2023, the Strategy for Responsible Development and the National Strategy for Regional Development 2030.

The strategic objective of the National Urban Policy, adopted in October 2015, is to strengthen the capacity of cities and urban functional areas to develop sustainably, to create jobs and improve the quality of life for their inhabitants. The document highlights the fact that “social disproportions within cities are often bigger than between regions”.

The Strategy for Responsible Development, adopted in February 2017, indicates that “a socially and territorially sustainable development entails a harmonious development of the whole country, recognising the territorial diversity and its advantages, as well as trying to ensure a high quality of life for the whole society. The effectiveness of development activities depends on how adequately their scope and implementation mechanisms are adjusted to characteristics of particular areas: their individual potential, deficits and mutual dependencies. The expected outcomes of development activities in these areas include, above all, new high-quality jobs, which generate higher income for their inhabitants and higher tax revenues for units of local government”.

The National Strategy for Regional Development, adopted in September 2019, states that “the implementation of development activities in areas at risk of permanent social exclusion and in medium-sized cities losing their socio-economic functions is inhibited by continuing poverty, deficiencies regarding the basic infrastructure and poor availability of public transport and other services in these areas”. From the perspective of regional policy, this means “the need to concentrate support efforts on combating excessive disproportions in development, between regions and within them. On the other hand, in addition to compensatory actions, interventions undertaken as part of regional policy will apply to all territories, taking into account their stage of development, particularly their quality of human capital”.

In view of the observations made in the strategic documents mentioned above, it is necessary to provide reliable statistical data that can be used to monitor changes and the degree of goal achievement.

The goal of providing information about personal incomes at local level was the motivation behind the study entitled *Income stratification of city dwellers*, with the crucial requirement that the goal was to be achieved using data from administrative sources. The results of this study are presented in the following report.

The two main parts of the report – the methodological and analytical part – are preceded by an Executive Summary, which highlights the key findings concerning gross incomes of taxpayers/net incomes earned by employed city dwellers in 2018.

The chapter devoted to the methodological aspects of the study provides a detailed description of all stages, starting from the data sources and ending with ways of detecting and handling outliers.

The analytical part, included in the chapter entitled *Analysis of the results*, presents findings concerning taxable revenues of all taxpayers and taxable incomes of employees living in towns and cities, taking into various spatial and typological classifications. The chapter called *Concluding remarks* provides, among other things, comments about and suggestions for future research based on administrative registers in the context of subsequent programme of statistical surveys and the need for permanent cooperation between agencies of official statistics and holders of administrative data, including the Ministry of Finance.

The following final report is the main outcome of the study, including four appendices:

- Descriptions of the structure of administrative data files provided by their holders;
- Descriptions of the preparation of databases;
- Choropleth maps for provincial capital cities;
- A database containing information for cities in alphabetical order, in terms of (median) taxable revenue and taxable income (median, Gini coefficient, income quintile ratio S80/S20, P90/P10 ratio).

The study is based on data from tax registers for 2018, maintained by the Ministry of Finance. The data on taxable revenues (including deductible expenses) come from information provided on 10 kinds of PIT tax forms: PIT-28, PIT-28S, PIT-36, PIT-36S, PIT-36L, PIT-36LS, PIT-37, PIT-38, PIT-39 and PIT-40A/11A. Data about taxable incomes (excluding deductible expenses) of people in paid employment (i.e. receiving remuneration in the context of an employment, public service or cooperative relationship or for home-based work, as well as social insurance benefits paid by employers) come from PIT-11 forms. Both taxable revenues and taxable incomes were given as annual totals.

Synteza

- Zamierzenia badawcze zostały w dużej części osiągnięte, chociaż wyniki odnoszące się do dochodów dotyczą wyłącznie pracowników najemnych mieszkających w miastach. Zakres danych, pozyskiwanych dotychczas przez statystykę publiczną w ramach programu badań statystycznych statystyki publicznej podyktował przeprowadzenie analizy dla całej zbiorowości podatników na podstawie informacji na temat uzyskanych przez nich przychodów.
- Chociaż przeprowadzono szczegółowe badanie wpływu podejść związanych z detekcją i sposobami postępowania z obserwacjami odstającymi, uwzględniono szereg przesłanek skutkujących oparciem stratyfikacji dochodowej mieszkańców wyłącznie o dane surowe.
- Mediana przychodów mieszkańców miast osiągała najwyższe wartości głównie w miastach województwa śląskiego i mazowieckiego. W odniesieniu do dochodów najwyższy poziom mediany osiągały mieszkańcy miast województwa mazowieckiego i dolnośląskiego. Na południu kraju można zauważyć wyraźny wzorec przestrzenny wysokich wartości mediany dla miast leżących wzdłuż autostrady A4.
- Zauważalna jest nadreprezentacja miast województwa mazowieckiego wśród jednostek osadniczych o najwyższych wartościach mediany dochodu. Aż cztery z pięciu miast o największych wartościach mediany leżą w województwie mazowieckim.
- Najwyższym poziomem dochodów mierzonym medianą charakteryzują się mieszkańcy miast dużych, a najniższym – mieszkańcy miast małych. Dla populacji podatników zamieszkujących w miastach średnich mediana wyniosła 36691,97 zł, była więc niższa o 5311,73 zł od analogicznej wartości odnotowanej dla miast dużych i o 2208,23 zł wyższa od mediany dla miast małych.
- Dochody podatników w miastach charakteryzują się dużym zróżnicowaniem i nie są równomiernie rozłożone pomiędzy wszystkie jednostki populacji. Łączne dochody podatników znajdujących się relatywnie w najgorszej sytuacji dochodowej (1. grupa decylowa) stanowiły w 2018 r. zaledwie 1% dochodów, podczas gdy suma dochodów podatników z najwyższej grupy decylowej stanowiła 30,6% ogółu dochodów uzyskanych przez podatników.
- Duże zróżnicowanie dochodów mieszkańców miast i stosunkowo silna ich koncentracja znajdują swoje odzwierciedlenie także w wartości wskaźnika zróżnicowania kwintylowego i we współczynniku Giniego, które osiągnęły wartości odpowiednio: 14,3 i 0,43.
- Najsilniejszą koncentrację dochodów można zauważyć u mieszkańców miast dużych, wskaźniki zróżnicowania kwintylowego i współczynnik Giniego wyniosły odpowiednio: 15,42 i 0,45, nieco mniejsze wartości tych mierników, ale wciąż stosunkowo duże, charakteryzowały mieszkańców miast średnich – 13,07 i 0,41. Mieszkańcy miast małych wśród wszystkich mieszkańców miast charakteryzowali się najślabszą koncentracją dochodów (12,67 i 0,40).
- Mediana dochodów uzyskanych w 2018 r. przez podatników zamieszkujących w miastach tracących funkcje społeczno-gospodarcze wyniosła 35099,27 zł i była o 4501,72 zł niższa od analogicznej miary wyliczonej dla pozostałych miast w kraju.
- Mieszkańcy miast tracących funkcje społeczno-gospodarcze charakteryzowali się mniejszym zróżnicowaniem dochodów i słabszą koncentracją dochodów w stosunku do mieszkańców pozostałych miast w kraju.
- Zastosowanie siatki kilometrowej pozwoliło na detekcję i analizę wewnątrzmijskich rozkładów przestrzennych wybranych charakterystyk poziomu i zróżnicowania dochodów mieszkańców miast wojewódzkich. Dla każdego z nich udało się zidentyfikować obszary o wysokim stopniu zróżnicowania dochodów pracowników najemnych.

Executive summary

- The research objectives have largely been achieved, although the results to do with personal incomes concern only employees living in cities. Given the scope of data, currently obtained in the course of implementing the programme of statistical surveys, the analysis was conducted for the entire population of taxpayers based on information about their incomes.
- Although a thorough analysis of the impact of various methods of outlier detection and treatment was conducted, a number of considerations were identified in favour of performing income stratification exclusively on the basis of raw data.
- The median personal income of city dwellers was found to be the highest mainly in cities of ŚLĄSKIE and MAZOWIECKIE provinces. The highest income levels were recorded for city inhabitants of MAZOWIECKIE and DOLNOŚLĄSKIE. In the south of the country, there is an evident spatial pattern of high values of the median income for cities located along the A4 motorway.
- There is a clear overrepresentation of cities from MAZOWIECKIE in the group of localities with the highest values of the median income. 4 out of 5 cities with the highest values of the median income are located in MAZOWIECKIE.
- The highest level of personal incomes in terms of the median can be observed for inhabitants of large cities, while the lowest – for inhabitants of small towns. For the population of taxpayers living in medium-sized towns, the median equals PLN 36,691.97, which is PLN 5,311.73 less than the corresponding value for large cities and PLN 5,311.73 more than the median for small towns.
- Personal incomes of city dwellers are characterised by considerable variation and are not equally distributed among all individuals. Combined incomes of taxpayers who are relatively worst off (1st decile) accounted for 1% of all incomes in 2018, while the combined incomes of those in the highest decile accounted for 30.6% of all incomes earned.
- The high degree of variation in incomes of city dwellers and their relatively strong concentration are also reflected by the income quintile share ratio and the Gini coefficient, which were equal to 14.3 and 0.43 respectively.
- The highest degree of income concentration can be observed for inhabitants of large cities, with values of the income quintile ratio and the Gini coefficient amounting to 15.42 and 0.45 respectively; in the case of inhabitants of medium-sized towns, the corresponding values are somewhat smaller but still relatively high: 13.07 and 0.41. The degree of income concentration was the lowest for inhabitants of small towns: 12.67 and 0.40.
- The median income earned in 2018 by taxpayers living in cities losing their socio-economic functions was equal to PLN 35,099.27 and was PLN 4501.72 less than the corresponding value calculated for other cities.
- Incomes of inhabitants of cities losing their socio-economic functions were less variable and concentrated compared to those of other cities.
- The use of 1 km x 1km grid made it possible to detect and analyse innercity spatial distributions of selected characteristics of the level and variation in incomes of people living in provincial capital cities. For each of these cities, it was possible to identify areas with a high degree of variation in incomes of employees.

I. Metodologia badania

I. Research methodology

1. Zakres podmiotowy i czasowy badania

1. *The scope and reference period of the study*

Badaniem objęto podatników rozliczających się na formularzach PIT-28, PIT-28S, PIT-36S, PIT-36LS, PIT-36, PIT-37, PIT-38, PIT-39, PIT-40A/11A oraz PIT-11 za rok 2018.

2. Zakres terytorialny badania

2. *The territorial coverage of the study*

Wszystkie miasta w Polsce, w tym 18 miast wojewódzkich, dla których dodatkowo dane zaprezentowano w układzie wewnętrznym na siatce kwadratów.

3. Źródła danych

3. *Data sources*

3.1. Dane Ministerstwa Finansów

3.1. *Data of the Ministry of Finance*

Źródłem danych atrybutowych wykorzystanych w badaniu są informacje zawarte w rejestrach podatkowych. Zgodnie z zapisami w Programie Badań Statystycznych Statystyki Publicznej 2018 są one przekazywane przez ich gestora na potrzeby statystyki publicznej.

W trakcie prowadzonych prac wykorzystano informacje przekazywane na potrzeby następujących badań statystycznych:

- 1.04.01 Stowarzyszenia, fundacje, samorząd gospodarczy i zawodowy oraz społeczne jednostki wyznaniowe,
- 1.04.06 Partie polityczne,
- 1.04.08 Partnerzy dialogu społecznego – organizacje pracodawców i związki zawodowe,
- 1.24.01 Wynagrodzenia w gospodarce narodowej,
- 1.70.02 Statystyczny system informacyjny o miastach.

Zestawy danych z systemów informacyjnych Ministerstwa Finansów zawierają dane pochodzące z następujących formularzy podatkowych:

- PIT-11 - Informacja o przychodach z innych źródeł oraz o dochodach i pobranych zaliczkach na podatek dochodowy,
- PIT-28/PIT-28S - Zeznanie o wysokości uzyskanego przychodu, wysokości dokonanych odliczeń i należnego ryczałtu od przychodów ewidencjonowanych,
- PIT-36/PIT-36S - Zeznanie o wysokości osiągniętego dochodu (poniesionej straty),
- PIT-36L/PIT-36LS - Zeznanie o wysokości osiągniętego dochodu (poniesionej straty),
- PIT-37 - Zeznanie o wysokości osiągniętego dochodu (poniesionej straty),
- PIT-38 - Zeznanie o wysokości osiągniętego dochodu (poniesionej straty),
- PIT-39 - Zeznanie o wysokości osiągniętego dochodu (poniesionej straty),

- PIT-40A/11A – Roczne obliczenie podatku przez organ rentowy, informacja o dochodach uzyskanych od organu rentowego.

Wszystkie dane pozyskane z formularzy podatkowych dotyczą roku 2018.

Szczegółowa struktura poszczególnych zbiorów Ministerstwa Finansów wykorzystanych na potrzeby badania przedstawiona została w załączniku 1.

Wtórne wykorzystanie danych pozyskanych na potrzeby innych badań statystyki publicznej spowodowało, że analizowaną w pracy badawczej zmienną są przychody ludności. W przypadku wprowadzenia badania na stałe do programu badań statystycznych statystyki publicznej nie przewiduje się konieczności wprowadzenia zmian w przygotowanej metodologii badania.

3.2. Dane Głównego Urzędu Statystycznego

3.2. *Data of Statistics Poland*

Zintegrowana Baza Populacji to zbiór danych stworzony przez Departament Badań Demograficznych oraz Departament Systemów Teleinformatycznych, Geostatystyki i Spisów Głównego Urzędu Statystycznego. Dane te należy traktować jako eksperymentalne i niebędące oficjalnymi statystykami publikowanymi przez Główny Urząd Statystyczny. Baza powstała w oparciu o zintegrowane źródła administracyjne pochodzące z Ministerstwa Cyfryzacji, Ministerstwa Finansów, Zakładu Ubezpieczeń Społecznych, Kasy Rolniczego Ubezpieczenia Społecznego, Narodowego Funduszu Zdrowia czy Agencji Restrukturyzacji i Modernizacji Rolnictwa.

Dane pozyskane ze Zintegrowanej Bazy Populacji dotyczą roku 2018 i zawierają m.in. informacje o miejscu zamieszkania ludności na terenie Polski, w tym dane adresowe (nazwa ulicy, numer budynku) oraz współrzędne geograficzne budynków będących miejscem zamieszkania. Wspomniana Zintegrowana Baza Populacji ma następującą strukturę:

Tablica 1. Struktura Zintegrowanej Bazy Populacji

Zmienna	Opis
PESEL	numer identyfikacyjny PESEL
PLEC	płeć (1 – mężczyzna, 2 – kobieta)
WIEK	wiek (w 2018 r.)
NAZWA_GMINY	nazwa gminy zamieszkania lub dzielnicy w przypadku Warszawy, Krakowa, Łodzi, Wrocławia i Poznania
KOD_GMINY	kod TERYT gminy zamieszkania lub dzielnicy w przypadku Warszawy, Krakowa, Łodzi, Wrocławia i Poznania
NAZWA_MIEJSCOWOSCI	nazwa miejscowości zamieszkania lub dzielnicy w przypadku Warszawy, Krakowa, Łodzi, Wrocławia i Poznania
KOD_MIEJSCOWOSCI	kod SIMC miejscowości zamieszkania lub dzielnicy w przypadku Warszawy, Krakowa, Łodzi, Wrocławia i Poznania
NAZWA_ULICY	nazwa ulicy (osiedla, alei, ...) zamieszkania
PRZEDROSTEK_ULICY	przedrostek ulicy (osiedla, alei, ...) zamieszkania
KOD_ULICY	kod ulicy (osiedla, alei, ...) zamieszkania
NUMER_BUDYNKU	numer budynku zamieszkania

Zmienna	Opis
NUMER_LOKALU	numer lokalu zamieszkania
WSP_X	współrzędna X miejsca zamieszkania w układzie 92 (długość)
WSP_Y	współrzędna Y miejsca zamieszkania w układzie 92 (szerokość)

Źródło: Opracowanie własne.

W projekcie wykorzystano również dane systemu TERYT, które posłużyły do weryfikacji poprawności i kompletności danych przestrzennych oraz danych atrybutowych.

3.3. Dane przestrzenne

3.3. Spatial data

W ramach realizowanego projektu wykorzystano następujące dane przestrzenne:

- mapę numeryczną podziału administracyjnego Polski w układzie gmin (z wyróżnieniem części miejskiej oraz części wiejskiej w jednostkach o statusie gminy miejsko-wiejskiej), stanowiącą zasób państwowego rejestru granic i powierzchni jednostek podziałów terytorialnych kraju (PRG),
- punkty adresowe stanowiące zasób państwowego rejestru granic i powierzchni jednostek podziałów terytorialnych kraju (PRG),
- kilometrową siatkę kwadratów przygotowaną przez Europejskie Forum ds. Geostatystyki EFGS¹ i opublikowaną na stronie: http://www.efgs.info/data/eurogrid/Grid_ETRS89_LAEA_PL_1K.zip/view.

Siatka kwadratów została wykonana w odwzorowaniu azymutalnym równopowierzchniowym Lamberta (ETRS 1989 LAEA), dzięki czemu wszystkie przecięcia siatki mają pełne współrzędne, co umożliwia agregację danych do pola kwadratu bez konieczności stosowania zaawansowanych narzędzi do analizy przestrzennej, a jedynie przy wykorzystaniu podstawowych funkcji dostępnych w każdym oprogramowaniu bazodanowym.

Mapy numeryczne podziału administracyjnego Polski zostały przygotowane w Państwowym Układzie Współrzędnych Geodezyjnych 1992 (ETRS89 / Poland CS92, kod EPSG: 2180). Mapy te zawierały podział administracyjny Polski w układzie:

- gminy (z wyróżnieniem części miejskiej oraz części wiejskiej w jednostkach o statusie gminy miejsko-wiejskiej),
- dzielnice, delegatury (w przypadku miast mających dzielnice lub delegatury),
- powiaty,
- województwa.

Pozyskane mapy podziału administracyjnego Polski aktualne są na dzień 1 stycznia 2018 r.

Dane numeryczne pochodzące z Państwowego Rejestru Granic zawierały identyfikatory oraz nazwy jednostek podziału terytorialnego zgodne z systemem TERYT, co umożliwiło połączenie danych przestrzennych z utworzonymi danymi atrybutowymi.

Dane adresowe (punkty adresowe) pozyskano z Państwowego Rejestru Granic.

Wszystkie dane przestrzenne zostały przygotowane w postaci plików SHP.

¹ Europejskie Forum Geografii i Statystyki (European Forum for Geography and Statistics).

3.4. Środowisko programistyczne

3.4. *Programming environment*

Dane atrybutowe wygenerowano w systemie SAS Enterprise Guide w wersji 4.2. Jednocześnie program ten wykorzystywany był do prac analitycznych i obliczeniowych.

Prace związane z łączeniem danych statystycznych z informacją przestrzenną przeprowadzono z wykorzystaniem również programu R w wersji 3.4.0 oraz programów ArcGIS for Desktop Basic (ArcView) w wersji 10.4.1. i QGIS w wersji 3.10.0-A Coruña.

Do prezentowania danych przestrzennych wykorzystano QGIS w wersji 3.10.0-A Coruña oraz ArcGIS for Desktop Basic (ArcView) w wersji 10.4.1.

4. **Przegląd literatury pod kątem mierników charakteryzujących sytuację dochodową mieszkańców miast oraz metod klasyfikacji i analiz przestrzennych**

4. *A review of literature on measures of income of the urban population and methods of classification and spatial analysis*

W oparciu o literaturę przedmiotu, w początkowej fazie prac, rozważano szeroką listę mierników charakteryzujących rozkład dochodów (patrz Tab. 2., Tab. 3., Tab. 4.). Każdy z mierników poddano analizie pod kątem: przydatności w ramach niniejszego badania, złożoności obliczeniowej, skorelowania z innymi charakterystykami rozkładu dochodów w celu uniknięcia redundancji informacji, jaką dany miernik niesie w ramach całej listy, odporności miernika na występowanie obserwacji odstających, a także ugruntowanej pozycji miernika w literaturze przedmiotu.

Celem analizy struktury jest charakterystyka rozkładu dochodów w całej badanej populacji za pomocą wybranych parametrów (mierników) tego rozkładu, najczęściej opisujących poziom i różnicowanie dochodów.

Tablica 2. Miary poziomu

Nazwa miernika	Wzór	Przykłady aplikacji w literaturze
Suma (wartość globalna)	$T = \sum_{i=1}^N x_i,$ <p>gdzie: $i = 1, 2, \dots, N,$ N – oznacza liczbę jednostek w populacji, x_i – oznacza dochody/przychody i-tej jednostki w populacji.</p>	
Średnia arytmetyczna	$m = \frac{\sum_{i=1}^N x_i}{N},$ <p>gdzie: $i = 1, 2, \dots, N$ N – oznacza liczbę jednostek w populacji, x_i – oznacza dochody/przychody i-tej jednostki w populacji.</p>	(Canberra Group, Handbook on Household Income Statistics, 2011), (Canberra Group, Expert Group on Household Income Statistics, 2001), (Kerm, 2007); (Moser i Schnetzer, 2014); (Neri, Gagliardi, Ciampalini, Verma i Betti, 2009); (Vecchi, 2018)
Udział (frakcja)	$w_k = \frac{\sum_{i \in k} x_i}{\sum_{i=1}^N x_i},$ <p>gdzie: $i = 1, 2, \dots, N,$ $k = 1, 2, \dots, N_k,$ N_k – oznacza liczbę jednostek w podzbiorowości k wyodrębnionej w całej populacji, N – oznacza liczbę jednostek w populacji x_i – oznacza dochody/przychody i-tej, jednostki w populacji.</p>	
Minimum	$x_{min} = \min_{1 \leq i \leq n} x_i.$ <p>gdzie: x_i – oznacza dochody/przychody i-tej jednostki w populacji.</p>	(Neri, Gagliardi, Ciampalini, Verma i Betti, 2009)
Maksimum	$x_{max} = \max_{1 \leq i \leq n} x_i,$ <p>gdzie: x_i – oznacza dochody/przychody i-tej jednostki w populacji.</p>	(Neri, Gagliardi, Ciampalini, Verma i Betti, 2009)
Kwantyl (kwartył, kwintyl, decyl, percentyl) rozkładu rzędu p	$Q(p) = \inf\{x \in \mathbb{R}: p \leq F(x)\},$ <p>gdzie: \inf – kres dolny zbioru, x – zmienna oznaczająca dochody/przychody, p – rząd kwantyla rozkładu, $F(x_i) = P(x \leq x_i)$ jest dystrybuantą rozkładu zmiennej x.</p>	(Canberra Group, Handbook on Household Income Statistics, 2011), (Canberra Group, Expert Group on Household Income Statistics, 2001), (Kerm, 2007); (Neri, Gagliardi, Ciampalini, Verma i Betti, 2009); (Vecchi, 2018)

Źródło: Opracowanie własne.

Tablica 3. Miary zróżnicowania i nierówności dochodowej

Nazwa miernika	Wzór	Przykłady aplikacji w literaturze
Rozstęp	$R = x_{max} - x_{min}$	(Cowell, 2011)
Rozstęp międzykwartylowy	$IQR = Q(0,75) - Q(0,25)$, gdzie: $Q(0,75)$ – oznacza kwantyl rzędu 0,75, czyli trzeci kwantyl rozkładu, $Q(0,25)$ – oznacza kwantyl rzędu 0,25, czyli pierwszy kwantyl rozkładu.	(Filzmoser, Gussenbauer i Templ, 2016)
Wskaźnik zróżnicowania kwantylowego (kwintylowego, decylnego, percentylowego)	$W_{Q_{p_1}, Q_{p_2}} = \frac{Q(p_1)}{Q(p_2)}$, gdzie: $Q(p_1)$ – oznacza kwantyl rzędu p_1 , $Q(p_2)$ – oznacza kwantyl rzędu p_2	(Canberra Group, Handbook on Household Income Statistics, 2011) – P90/P10, P80/P20, P80/P50, P50/P20; (Kerm, 2007) – P90/P10, P82/P20; (Martín-Román, Ayala, Vicente i Juan, 2017) – P90/P10, P90/P50, P10/P50, P75/P25; (Moser i Schnetzer, 2014) – P90/P10, P90/P50, P50/P10; (OECD, OECD Framework for Statistics on the Distribution of Household Income, Consumption and Wealth, 2013); (Vecchi, 2018) – P90/P10, P75/P25, P90/P50
Relacja do mediany	$r_{me} = \frac{Q(p)}{Q(0,50)}$, gdzie: $Q(p)$ – oznacza kwantyl rzędu p , $Q(0,50)$ – oznacza kwantyl rzędu 0,50, czyli medianę rozkładu.	(Canberra Group, Handbook on Household Income Statistics, 2011); (Neri, Gagliardi, Ciampalini, Verma i Betti, 2009) – mean/P50, P80/P50, P90/P50, P95/P50, P99/P50 (P90/P50, P91/P50, P92/P50, P93/P50, P94/P50, P95/P50, P96/P50, P97/P50, P98/P50, P99/P50, P100/P50);
Relacja do średniej arytmetycznej	$r_m = \frac{m_{Q(p)}}{m}$, gdzie: $m_{Q(p)}$ – oznacza średnią arytmetyczną w grupie kwantylowej rzędu p , m – oznacza średnią arytmetyczną w całej populacji.	(Bruil i Koymans, 2014); (Fesseau i Mattonetti, 2013)
Współczynnik Giniego	$G = \left(\frac{1}{2N^2m}\right) \sum_{i=1}^N \sum_{j=1}^N x_i - x_j $, gdzie: $i = 1, 2, \dots, N$, $j = 1, 2, \dots, N$, N – oznacza liczbę jednostek w populacji, m – oznacza średnią arytmetyczną w całej populacji, x_i – oznacza dochody/przychody i -tej jednostki w populacji, x_j – oznacza dochody/przychody j -tej jednostki w populacji.	(Canberra Group, Handbook on Household Income Statistics, 2011), (Canberra Group, Expert Group on Household Income Statistics, 2001), (Kerm, 2007); (Martín-Román, Ayala, Vicente i Juan, 2017); (Mastronardi i Cavallo, 2020); (Moser i Schnetzer, 2014); (OECD, OECD Framework for Statistics on the Distribution of Household Income, Consumption and Wealth, 2013); (Osier, 2009); (Panek, 2014); (Panzer i Postiglione, 2020); (Szulc, 2018); (Vecchi, 2018); (Cowell, 2011)
Wskaźnik zróżnicowania kwintylowego	$S80/S20 = \frac{\sum_{x_i \geq Q(0,8)} x_i}{\sum_{x_i \leq Q(0,2)} x_i}$, gdzie: $Q(0,8)$ – oznacza kwantyl rzędu 0,8, $Q(0,2)$ – oznacza kwantyl rzędu 0,2, x_i – oznacza dochody/przychody i -tej jednostki w populacji.	(GUS, 2019), (Bruil i Koymans, 2014), (EUROSTAT, 2003), (Fesseau i Mattonetti, 2013), (Kerm, 2007) – S80/S20, S90/S10; (Osier, 2009); (Szulc, 2018)

Źródło: Opracowanie własne.

Tablica 4. Mierniki autokorelacji przestrzennej

Nazwa miernika	Wzór	Przykłady aplikacji w literaturze
Statystyka Getisa-Orda	$G_i(d) = \frac{\sum_j w_{ij}(d)x_j - W_i\bar{x}(i)}{s(i) \left\{ \frac{[(n-1)S_{1i} - W_i^2]}{(n-2)} \right\}^{\frac{1}{2}}}, i \neq j$ <p>gdzie: x – analizowana zmienna, x_i – wartość analizowanej zmiennej dla i-tego regionu, n – oznacza liczbę obserwacji, i, j – regiony, w_{ij} – macierz przestrzenna wag, która zawiera informację na temat sąsiedztwa regionów i oraz j, $W_i = \sum_{i \neq j} w_{ij}(d)$ – suma wag, $S_{1i} = \sum_j w_{ij}^2$, \bar{x} – średnia analizowanej zmiennej. $s(i)$ – odchylenie standardowe analizowanej zmiennej.</p>	(Moser i Schnetzer, 2014)
Statystyka lokalna Morana	$I_i = \frac{x_i - \bar{x}}{S_i^2} \sum_{j=1, j \neq i}^n w_{ij}(x_j - \bar{x})$ <p>gdzie: x – analizowana zmienna, x_i – wartość analizowanej zmiennej dla i-tego regionu, n – oznacza liczbę regionów, i, j – regiony, w_{ij} – macierz wag przypisanych sąsiadom i-tego regionu, \bar{x} – średnia analizowanej zmiennej, $S_i^2 = \frac{\sum_{j=1, i \neq j}^n (x_j - \bar{x})^2}{n-1} - \bar{x}^2$.</p>	(Mastronardi i Cavallo, 2020) (Basarbowicz i inni, 2015)

Źródło: Opracowanie własne.

Zastosowane w badaniu stratyfikacji dochodowej mieszkańców miast metody klasyfikacji miały na celu wyodrębnienie bardziej jednolitych podzbiorowości podlegających analizie. W odniesieniu do mieszkańców miast zastosowano przede wszystkim grupowanie typologiczne według płci, wieku, źródeł przychodów podatkowych oraz grupowanie wariacyjne oparte na przedziałach kwantylowych (kwartyłowych, kwintylowych, decylowych, percentylowych). W przypadku gdy przedmiotem analizy były miasta w Polsce, wykorzystano przede wszystkim grupowanie typologiczne (w odniesieniu do charakterystyk ważnych ze względu na prowadzoną politykę rozwoju kraju, grup wielkościowych miast, podziału administracyjnego). W analizie zróżnicowania sytuacji dochodowej mieszkańców miast wojewódzkich, gdzie obiektem analizy jest siatka kwadratów, klasyfikacji przestrzennych efektów aglomeracyjnych dokonano z wykorzystaniem klastrów o niskich lub wysokich wartościach charakterystyk dochodowych.

5. Wypracowanie metod postępowania z danymi atrybutowymi i przestrzennymi

5. Development of methods of handling attribute and spatial data

5.1. Przygotowanie danych atrybutowych

5.1. Preparation of attribute data

Mając na celu ustalenie listy zmiennych charakteryzujących sytuację dochodową podatników, dokonano analizy danych pochodzących z systemów informacyjnych Ministerstwa Finansów, zarówno w zakresie przychodów, jak i dochodów.

Pozyskane z Ministerstwa Finansów dane dotyczące przychodów były podzielone na dwa zbiory – jeden przedstawiający informacje dotyczące przychodu podatnika, a drugi zawierający jego dane adresowe (struktury zbiorów zostały przedstawione w załączniku 1. Identyfikatory terytorialne (identyfikatory gminy) zostały ujęte w podziale tylko na gminy miejskie, wiejskie oraz miejsko-wiejskie bez rozróżnienia na część miejską i część wiejską.

W przypadku gmin miejsko-wiejskich w procesie przygotowania danych do analiz wyróżniono część miejską oraz część wiejską gminy, a dla miast: Warszawa, Kraków, Łódź, Wrocław i Poznań, wyróżniono odpowiednio dzielnice lub delegatury.

Przygotowanie danych atrybutowych pozyskanych z Ministerstwa Finansów, a przetworzonych przez Główny Urząd Statystyczny, rozpoczęto od połączenia odpowiednich zbiorów dotyczących przychodu podatnika oraz zawierających jego dane adresowe. Dla poszczególnych formularzy PIT wykorzystano zmienną identyfikującą *tech_id_PIT*. W kolejnym kroku wykonano deduplikację danych, usuwając z nich powtarzające się wiersze, pozostawiając pojedyncze wpisy dla każdego podatnika. Z kolei po uzupełnieniu brakujących kodów TERYT usunięto rekordy, dla których zmienna *adr_kod_gmn_PIT*, przechowująca informację o kodzie TERYT gminy zamieszkania podatnika miała puste wartości.

W kolejnym kroku dokonano kontroli logicznej otrzymanych zbiorów.

W przygotowanych w ten sposób bazach wybrano poniżej zaprezentowane zmienne.

Tablica 5. Zmienne ze zbiorów Ministerstwa Finansów istotne z punktu widzenia badania

Zmienna	Opis
<i>KOD_FORM</i>	kod formularza rozliczenia podatkowego
<i>TECH_ID</i>	techniczny identyfikator podatnika
<i>PLEC</i>	płeć podatnika
<i>ROK_URODZENIA</i>	rok urodzenia podatnika
<i>PRZYCHOD</i>	przychód
<i>KOD_TERYT</i>	kod TERYT gminy zamieszkania podatnika
<i>MIEJSCOWOSC</i>	nazwa miejscowości zamieszkania podatnika

Zmienna	Opis
KOD_POCZTOWY	kod pocztowy miejsca zamieszkania podatnika
ULICA	nazwa ulicy zamieszkania podatnika
NR_DOMU	numer budynku zamieszkania podatnika

Źródło: Opracowanie własne.

Po zunifikowaniu struktury zbiorów zostały one połączone w jedną bazę zawierającą dane o przychodach podatników rozliczanych na różnych formularzach PIT. W celu uzyskania bazy z informacjami o łącznym przychodzie każdego z podatników wybrano unikatową zmienną *TECH_ID* wraz z danymi adresowymi podatnika, a następnie zsumowano występujące w bazie przychody dla każdego z identyfikatorów *TECH_ID*. Ponadto dodano kolumnę informującą o liczbie złożonych formularzy PIT przez danego podatnika. Otrzymana w ten sposób baza, zwana dalej *Bazą Przychodów Mieszkańców Miast (PMM)*, charakteryzuje się następującą strukturą:

Tablica 6. Zmienne w Bazie Przychodów Mieszkańców Miast

Zmienna	Opis
TECH_ID	techniczny identyfikator podatnika
PLEC	płeć podatnika
ROK_URODZENIA	rok urodzenia podatnika
KOD_TERYT	kod TERYT gminy zamieszkania podatnika
MIEJSCOWOSC	nazwa miejscowości zamieszkania podatnika
KOD_POCZTOWY	kod pocztowy miejsca zamieszkania podatnika
ULICA	nazwa ulicy zamieszkania podatnika
NR_DOMU	numer budynku zamieszkania podatnika
PRZYCHOD – zmienna wyliczona	łączny przychód rozliczany na wszystkich złożonych przez podatnika formularzach PIT
LICZBA_FORM – zmienna wyliczona	liczba złożonych przez podatnika formularzy PIT

Źródło: Opracowanie własne.

Pozyskany z Ministerstwa Finansów zbiór danych dotyczący dochodów zawierał informacje pochodzące z rozliczanych formularzy PIT-11.

Ze zbioru MF PIT-11 wybrano te rekordy, które odnosiły się do osób uzyskujących dochód z tytułu *wynagrodzenia ze stosunku: pracy, służbowego, spółdzielczego i z pracy nakładczej, a także zasiłki pieniężne z ubezpieczenia społecznego wypłacone przez zakład pracy, o którym mowa w art. 31 ustawy, oraz płatników, o których mowa w art. 42e ust. 1 ustawy*. Po odfiltrowaniu wpisów niepozwalających na zidentyfikowanie podatnika (puste wartości w kolumnie *podat_PESEL_PIT*), otrzymano bazę rozliczeń dla pracowników najemnych.

Następnie bazę tą połączono ze Zintegrowaną Bazą Populacji, celem uzupełnienia niepełnych informacji adresowych. W przypadku braku danych oraz dla gmin miejsko-wiejskich wykorzystano odpowiednią zmienną ZBP. Po uzupełnieniu informacji wybrano tych spośród podatników, którzy w 2018 roku mieszkali w miastach.

W kolejnym kroku wybrano unikatowy zbiór podatników, przypisując każdemu z nich wartość uzyskanego dochodu z tytułu *wynagrodzenia ze stosunku: pracy, służbowego, spółdzielczego i z pracy nakładczej, a także zasiłki pieniężne z ubezpieczenia społecznego wypłacone przez zakład pracy, o którym mowa w art. 31 ustawy, oraz płatników, o których mowa w art. 42e ust. 1 ustawy.* – przepisana w przypadku jednego stosunku pracy lub zsumowaną w przypadku występowania kilku wpisów.

W ten sposób na podstawie zmiennych zawartych w bazie PIT-11 oraz w Zintegrowanej Bazie Populacji, utworzono *Bazę Dochodów Mieszkańców Miast (DMM)* charakteryzującą się następującą strukturą:

Tablica 7. Zmienne w Bazie Dochodów Mieszkańców Miast

Zmienna	Opis
NR_ID	numer identyfikacyjny podatnika
ROK_UR – zmienna wyliczona	rok urodzenia podatnika
PLEC	płeć (1 – mężczyzna, 2 – kobieta)
KOD_TERYT	kod TERYT gminy podatnika
MIEJSCOWOSC	nazwa miejscowości zamieszkania podatnika
KOD_POCZTOWY	kod pocztowy miejsca zamieszkania podatnika
ULICA	nazwa ulicy zamieszkania podatnika
NR_DOMU	numer budynku zamieszkania podatnika
DOCHOD – kolumna wyliczona	łączy dochód z wynagrodzenia ze stosunku: pracy, służbowego, spółdzielczego z pracy nakładczej, a także zasiłki pieniężne z ubezpieczenia społecznego wypłacone przez zakład pracy, o którym mowa w art. 31 ustawy, oraz płatników, o których mowa w art. 42e ust. 1 ustawy

Źródło: Opracowanie własne.

Ponadto, dla podatników - mieszkańców miast wojewódzkich zaczerpnięto z ZBP oraz Punktów Adresowych PRG informacje o współrzędnych geograficznych miejsca ich zamieszkania, celem umożliwienia zagregowania ich do siatki kwadratów wewnątrzmijskich. W następnym kroku każdemu z podatników przypisano numer identyfikacyjny kwadratu, w którym znajdowało się jego miejsce zamieszkania. Pozyskana baza, zwana dalej *Bazą Dochodów Mieszkańców Miast Wojewódzkich przypisanych do siatki kwadratów (DMG)*, charakteryzowała się następującą strukturą:

Tablica 8. Zmienne w Bazie Dochodów Mieszkańców Miast Wojewódzkich przypisanych do siatki kwadratów

Zmienna	Opis
NR_ID	numer identyfikacyjny podatnika
ROK_UR – zmienna wyliczona	rok urodzenia podatnika
PLEC	płeć (1 – mężczyzna, 2 – kobieta)
KOD_TERYT	kod TERYT gminy podatnika
MIEJSCOWOSC	nazwa miejscowości zamieszkania podatnika
KOD_POCZTOWY	kod pocztowy miejsca zamieszkania podatnika
ULICA	nazwa ulicy zamieszkania podatnika

Zmienna	Opis
NR_DOMU	numer budynku zamieszkania podatnika
DOCHOD – kolumna wyliczona	łączny dochód z wynagrodzenia ze stosunku: pracy, służbowego, spółdzielczego i z pracy nakładczej, a także zasiłki pieniężne z ubezpieczenia społecznego wypłacone przez zakład pracy, o którym mowa w art. 31 ustawy, oraz płatników, o których mowa w art. 42e ust. 1 ustawy
ID_GRID	numer identyfikacyjny oczka siatki kwadratów, w którym znajdowało się miejsce zamieszkania podatnika

Źródło: Opracowanie własne.

5.2. Przygotowanie danych przestrzennych

5.2. Preparation of spatial data

Dane przestrzenne przygotowano w dwóch ujęciach. Pierwsze polegało na sprawdzeniu kompletności i poprawności danych atrybutowych zawartych w zbiorach Państwowego Rejestru Granic (PRG). Chodziło o ustalenie, czy zbiory PRG zawierają wszystkie gminy, powiaty i województwa ujęte w rejestrze TERYT oraz czy symbole poszczególnych jednostek są zgodne z symbolami w rejestrze TERYT.

Drugie ujęcie prowadziło do powstania zbioru zawierającego siatkę kwadratów obejmującą wyłącznie obszary miast wojewódzkich, w ich granicach administracyjnych. Pierwszym krokiem było stworzenie warstw obiektów powierzchniowych z granicami administracyjnymi miast wojewódzkich. Następnie w oparciu o grupę pozyskanych danych przestrzennych (Grid_ETRS89_LAEA_PL_1K) dla obszaru całego kraju utworzono siatkę kwadratów o boku 500 metrów, która na kolejnym etapie prac została ograniczona do obszaru leżącego w obrębie granic miast wojewódzkich. W celu usprawnienia dalszych prac związanych z analizą danych przestrzennych postanowiono ograniczyć wielkość zbioru zawierającego siatkę kwadratów, podlegającego dalszej analizie, poprzez utworzenie 18 zbiorów z siatką kwadratów o boku 500 metrów oraz 18 zbiorów z siatką kwadratów o boku 1000 metrów - osobno dla każdego miasta wojewódzkiego. Liczebność kwadratów w poszczególnych zbiorach przedstawia zamieszczona poniżej Tablica 9.

Tablica 9. Liczba kwadratów o boku 500 metrów oraz 1000 m według jednostek przestrzennych

Miasto wojewódzkie	Liczba kwadratów o boku 500 metrów	Liczba kwadratów o boku 1000 metrów
Białystok	472	133
Bydgoszcz	822	231
Gdańsk	1210	339
Gorzów Wielkopolski	405	116
Katowice	768	218
Kielce	517	146
Kraków	1446	396
Lublin	680	190
Łódź	1285	349
Olsztyn	418	121

Miasto wojewódzkie	Liczba kwadratów o boku 500 metrów	Liczba kwadratów o boku 1000 metrów
Opole	470	209
Poznań	1183	324
Rzeszów	552	169
Szczecin	1378	379
Toruń	547	159
Warszawa	2231	601
Wrocław	1299	350
Zielona Góra	1246	340

Źródło: Opracowanie własne.

Ze względu na konieczność zachowania tajemnicy statystycznej przeanalizowano liczbę osób oraz budynków znajdujących się w poszczególnych jednostkach siatki kwadratów. Tę operację wykonano zarówno dla siatki kwadratów o boku 500 metrów jak i o boku 1000 metrów. W celu określenia badanej populacji postanowiono uwzględnić tylko kwadraty o boku 1000 metrów, w których liczba ludności wyniosła nie mniej niż 10 osób i liczba budynków wyniosła nie mniej niż 3.

Kolejnym etapem prac było dołączenie do danych atrybutowych współrzędnych X i Y budynku zamieszkania. Wszystkie prace poprzedzone zostały dokonaniem konwersji danych geograficznych (zawartych w Zintegrowanej Bazie Populacji i PRG) z układu odniesienia 1992 do układu współrzędnych zgodnego z układem siatki kwadratów (ETRS 1989 LAEA). Tak przygotowane dane w postaci plików SHP zostały zaimportowane do środowiska SAS. Samo przyłączenie współrzędnych X i Y nastąpiło w dwóch etapach. W pierwszym etapie przyłączono współrzędne geograficzne zawarte w Zintegrowanej Bazie Populacji. W drugim etapie, do danych atrybutowych, którym nie udało się przypisać współrzędnych geograficznych, przyłączono dane zawarte w punktach adresowych pochodzących z PRG. Następnie przy pomocy pakietu ArcGIS (w środowisku SAS każdy z punktów uzyskał identyfikator kwadratu, w obrębie którego został zlokalizowany). Następnie dane pochodzące z Ministerstwa Finansów zostały zagregowane według identyfikatora jednostki przestrzennej (według TERYT-u miasta) oraz identyfikatora kwadratu. Tak przygotowana baza danych stanowiła punkt wyjścia do dalszych analiz przestrzennych oraz wizualizacji.

6. Detekcja i sposoby postępowania z obserwacjami odstającymi

6. *Methods of outlier detection and treatment*

Problem tzw. obserwacji odstających (outliers) w przypadku analizowania dochodów/przychodów jest szczególnym wyzwaniem. Fakt występowania w zbiorze danych wartości wyraźnie różniących się od pozostałych może bowiem wynikać nie tylko z błędów pomiarowych, przekłamań, ale także z samej natury zjawiska, jakim są dochody ludności i z ich nierównomiernego rozkładu w populacji (Ostasiewicz, 1998). Co więcej, w przypadku analizy dochodów ludności - zakładając, że dane zostały zebrane bez jakichkolwiek błędów - także należy przeprowadzić analizę obserwacji odstających. Jak wskazuje P. Kerm (Kerm, 2007), niektóre ze wskaźników sytuacji dochodowej (np. klasyczne miary tendencji centralnej) nie są odporne na bardzo niskie bądź bardzo wysokie wartości. Podobna sytuacja dotyczy także pomiaru nierówności dochodowych czy ubóstwa. Stąd, zwłaszcza w przypadku prowadzenia analiz porównawczych, należy wykryć obserwacje odstające i je odpowiednio skorygować, aby zneutralizować wpływ nietypowych jednostek w populacji na wnioski formułowane na podstawie mierników poziomu i zróżnicowania dochodów ludności.

W literaturze istnieje szereg metod detekcji obserwacji odstających i ich klasyfikacji. Filzmoser i in. (Filzmoser, Gussenbauer i Templ, 2016) oraz Denderski (Denderski, 2019) rozróżniają przed wszystkim metody adekwatne w przypadku analizy zbiorowości jednowymiarowych (jedna zmienna) i wielowymiarowych (wiele zmiennych). Należy przy tym podkreślić, że w przypadku rozkładów niesymetrycznych jeszcze przed przystąpieniem do wykrywania obserwacji odstających należy dokonać takiej transformacji danych, aby po tym zabiegu ich rozkład był zbliżony do normalnego.

Opis zastosowanych podejść.

Na wstępnym etapie prac rozpatrywano następujące podejścia w detekcji i sposobach postępowania z obserwacjami odstającymi w analizie rozkładu przychodów podatników/dochodów pracowników najemnych zamieszkujących miasta:

1. dane surowe, bez żadnej ingerencji;
2. dane ucięte (TRIM), gdzie odcięciu podlegało po 0,5% skrajnych obserwacji z obydwu końców rozkładu przychodów/dochodów;
3. winsoryzację (WINS), gdzie zastosowano prostą imputację polegającą na zastąpieniu oryginalnych 0,5% skrajnych obserwacji z obydwu końców rozkładu przychodów/dochodów, wartościami kwantyli rzędu odpowiednio 0,005 dla dolnego końca rozkładu przychodów/dochodów oraz 0,995 dla górnego końca rozkładu przychodów/dochodów;
4. transformację Box'a-Cox'a (BOXCOX_IQR) i „odcięcie” obserwacji skrajnych z obydwu końców rozkładu przychodów/dochodów wyznaczonych z wykorzystaniem rozstępu międzykwartylowego;
5. transformację Box'a-Cox'a (BOXCOX_MAD) i „odcięcie” obserwacji skrajnych z obydwu końców rozkładu przychodów/dochodów wyznaczonych z wykorzystaniem medianowego odchylenia przeciętnego;
6. transformację Box'a-Cox'a (BOXCOX_IQR2) i zastąpienie obserwacji skrajnych z obydwu końców rozkładu przychodów/dochodów, wyznaczonych z wykorzystaniem rozstępu międzykwartylowego, wartością wyznaczoną na podstawie wzoru (2) dla obserwacji skrajnych z dolnego końca rozkładu przychodów/dochodów oraz na podstawie wzoru (3) dla obserwacji skrajnych z górnego końca rozkładu przychodów/dochodów;
7. transformację Box'a-Cox'a (BOXCOX_MAD2) i zastąpienie obserwacji skrajnych z obydwu końców rozkładu przychodów/dochodów, wyznaczonych z wykorzystaniem medianowego odchylenia przeciętnego, wartością wyznaczoną na podstawie wzoru (6) dla obserwacji skrajnych z dolnego końca rozkładu przychodów/dochodów oraz na podstawie wzoru (7) dla obserwacji skrajnych z górnego końca rozkładu przychodów/dochodów.

Należy podkreślić ponadto, że wszystkie wymienione powyżej podejścia, zakładające pewną ingerencję w skrajne wartości rozkładu dochodów/przychodów, analizowane były w ramach dwóch odmiennych ujęć. Pierwsze z nich, nazwane roboczo „*top-down*”, polegało na zastosowaniu podejść 2. – 7. na całym zbiorze podatników/pracowników najemnych, podczas gdy drugie ujęcie, nazwane roboczo „*bottom-up*”, polegało na zastosowaniu podejść 2. – 7. odrębnie dla każdego z 930 miast, a następnie złożenie populacji podatników/pracowników najemnych zamieszkujących miasta w Polsce poprzez połączenie subpopulacji dla wszystkich 930 miast.

W zastosowanych wyżej podejściach miały zastosowanie następujące formuły:

Rozstęp międzykwartylowy:

$$y_l = med(x) - c * S_{IQR}, \quad (2)$$

$$y_h = med(x) + c * S_{IQR}, \quad (3)$$

gdzie:

$x = (x_1, x_2, \dots, x_n)$ – oznacza uporządkowany niemalejąco zbiór, x_i – oznacza i -ty element uporządkowanego niemalejąco zbioru,

$med(x)$ – oznacza medianę uporządkowanego niemalejąco zbioru,

y_l – oznacza punkt odcięcia w dolnym ogonie rozkładu,

y_h – oznacza punkt odcięcia w górnym ogonie rozkładu,

$$S_{IQR} = \frac{IQR}{1,35}, \quad (4)$$

$$IQR = Q(0,75) - Q(0,25), \quad (5)$$

c – oznacza stałą, która najczęściej przyjmuje wartości 3 (Filzmoser, Gussenbauer i Templ, 2016), a nawet 5 (Dupriez, 2007).

Medianowe odchylenie przeciętne:

$$y_l = med(x) - c * S_{MAD}, \quad (6)$$

$$y_h = med(x) + c * S_{MAD}, \quad (7)$$

gdzie:

$x = (x_1, x_2, \dots, x_n)$ – oznacza uporządkowany niemalejąco zbiór, x_i – oznacza i -ty element uporządkowanego niemalejąco zbioru,

$med(x)$ – oznacza medianę uporządkowanego niemalejąco zbioru, y_l – oznacza punkt odcięcia w dolnym ogonie rozkładu,

y_h – oznacza punkt odcięcia w górnym ogonie rozkładu,

$$S_{MAD} = \frac{MAD}{0,675}, \quad (8)$$

$$MAD = med|x - med(x)|, \quad (9)$$

c – oznacza stałą, która najczęściej przyjmuje wartości 3 (Filzmoser, Gussenbauer i Templ, 2016), bądź 3,5 (Denderski, 2019).

W przypadku podejść 4. – 7. można wyróżnić następujące etapy prac:

Etap 1. Transformacja Box'a-Cox'a danych oryginalnych w celu doprowadzenia do rozkładu zbliżonego do normalnego;

Etap 2. Detekcja outlierów z wykorzystaniem rozstępu międzykwartylowego/ medianowego odchylenia przeciętne (w zależności od oceny uzyskanych wyników);

Etap 3. „Odcięcie” wartości odstających bądź zaimputowanie dla nich wielkości wynikających z formuł (2) i (3), bądź (6) i (7) ustalonych na etapie 2;

Etap 4. Transformacja wsteczna Box'a-Cox'a.

Schematycznie zakres wszystkich zastosowanych podejść w detekcji i sposobach postępowania z obserwacjami odstającymi w analizie rozkładu przychodów podatników/dochodów pracowników najemnych zamieszkujących miasta przedstawia Tablica 10. Należy zaznaczyć, że ujęcia „top-down” oraz „bottom-up” przy zastosowaniu danych surowych są tożsame, tzn. odnoszą się do dokładnie tej samej populacji (brak jakiegokolwiek ingerencji w oryginalne dane).

Tablica 10. Wykaz zastosowanych podejść i ujęć w analizie rozkładu przychodów/dochodów

Zastosowane podejście	Przychody podatników		Dochody pracowników najemnych	
	„top-down”	„bottom-up”	„top-down”	„bottom-up”
dane surowe	x		x	
TRIM	x	x	x	x
WINS	x	x	x	x
BOXCOX_IQR	x	x	x	x
BOXCOX_MAD	x	x	x	x
BOXCOX_IQR2	x	x	x	x
BOXCOX_MAD2	x	x	x	x

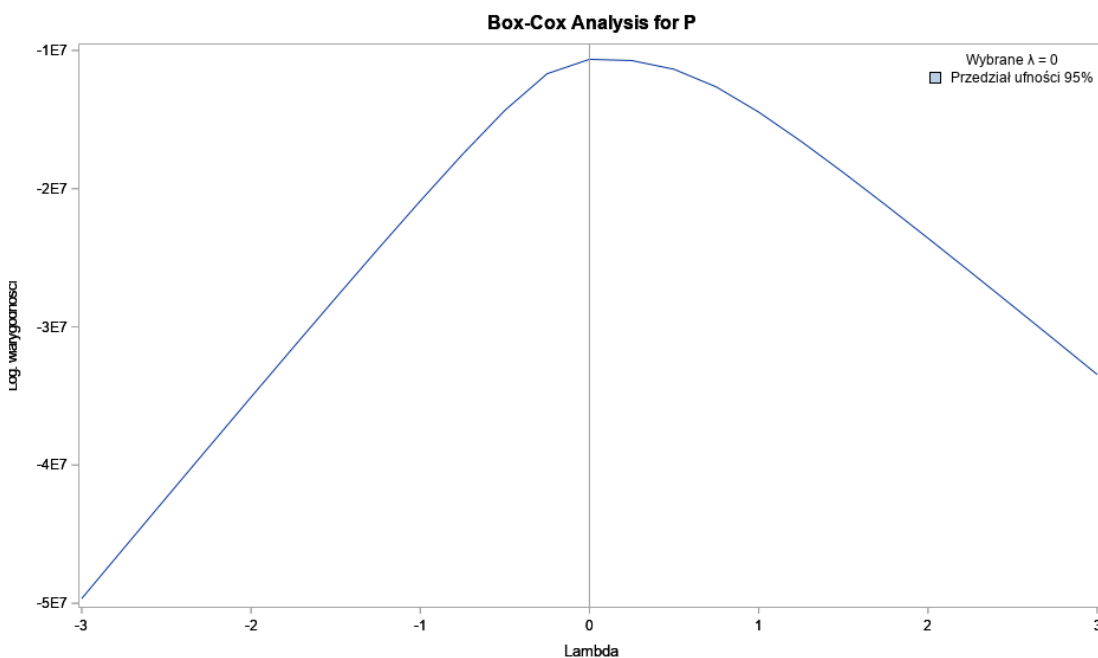
Źródło: Opracowanie własne.

Dodatkowo poniżej zestawione zostały podstawowe informacje na temat zastosowanej transformacji Box'a-Cox'a oraz podejść opartych na rozstępie międzykwartylowym i medianowym odchyleniu przeciętnym.

A. Wyznaczenie parametru lambda (λ) w transformacji Box'a-Cox'a.

W celu znalezienia optymalnej wartości parametru lambda (λ), od której zależy postać zastosowanej transformacji, zgodnie z wzorem (1), zastosowano procedurę TRANSREG oprogramowania SAS (por. <https://support.sas.com/documentation/onlinedoc/stat/141/transreg.pdf>). Takie postępowanie zastosowano zarówno dla przychodów podatników, jak i dochodów pracowników najemnych.

Rycina 1. Poszukiwanie optymalnej wartości parametru lambda do transformacji Box'a-Cox'a



Źródło: obliczenia własne – PROC TRANSREG (SAS).

Do transformacji Box'a-Cox'a przychodów podatników/dochodów pracowników najemnych dla obydwu ujęć „top-down” oraz „bottom-up”² przyjęto parametr $\lambda = 0$, stąd zastosowane przekształcenie we wszystkich analizowanych podejściach ma postać $z_i = \ln x_i$.

B. Wyznaczenie stałych c w określaniu obserwacji skrajnych dla podejść: BOXCOX_IQR, BOXCOX_MAD.

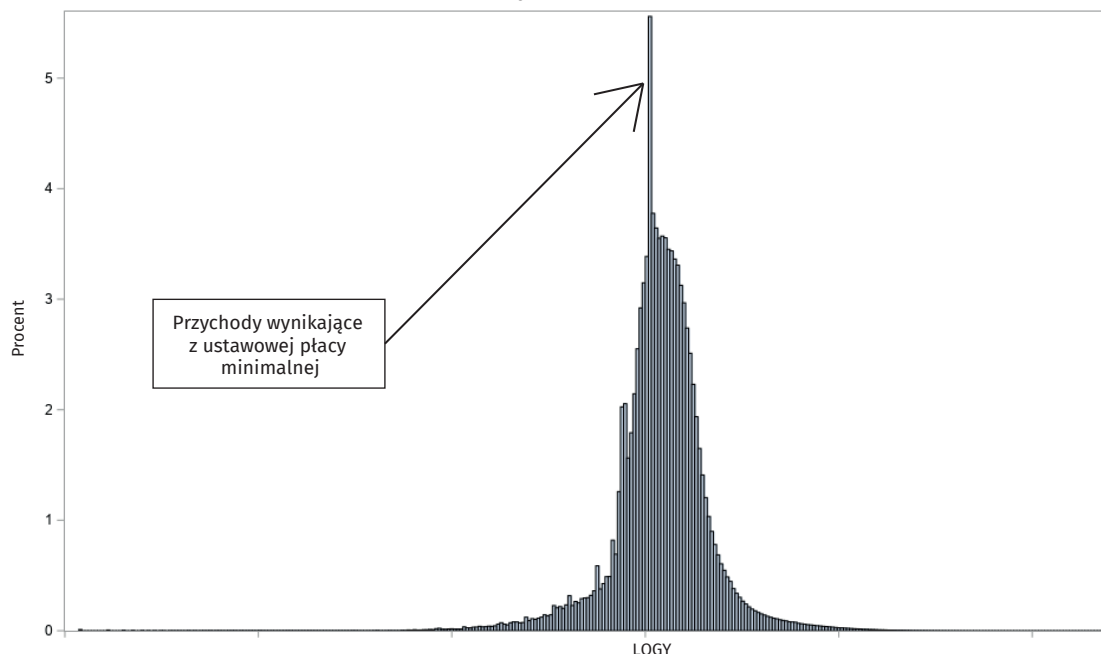
Zastosowanie rozstępu międzykwartylowego oraz medianowego odchylenia przeciętnego w określaniu obserwacji skrajnych z obydwu końców rozkładu przychodów/dochodów, zakłada przyjęcie pewnych stałych. Zgodnie z sugestiami wynikającymi z przeglądu literatury w zakresie analizy rozkładów dochodów, a także po wstępnej analizie kilku zestawów stałych c, przyjęto ostatecznie następujące wartości:

- dla rozstępu międzykwartylowego $c = 5$ (por. Dupriez, 2007);
- dla medianowego odchylenia przeciętnego $c = 3,5$ (por. Denderski, 2019).

C. Wyznaczenie stałych c w określaniu obserwacji skrajnych dla podejść: BOXCOX_IQR2, BOXCOX_MAD2.

W podejściach BOXCOX_IQR2 oraz BOXCOX_MAD2, po transformacji logarytmicznej przychodów/dochodów, wykorzystano charakterystyki rozkładu powyżej mediany do modelowania lewej strony rozkładu, tzn. poniżej mediany. Zabieg ten miał zniwelować ewentualny wpływ tzw. strukturalnych wartości (np. ustawowych minimalnych przychodów/dochodów i ich pochodnych) na wyznaczenie obserwacji skrajnych w rozkładzie. Schematycznie przedstawia to poniższa rycina.

Rycina 2. Rozkład przychodów podatników zamieszkujących miasta po transformacji Box'a-Cox'a



Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

² Ze względu na czasochłonność procedury TRANSREG dla dużych zbiorów danych, a takimi są zbiory Ministerstwa Finansów, i ograniczenia czasowe związane z realizacją niniejszej pracy, nie prowadzono odrębnych analiz dla każdego z 930 miast w zakresie wyszukiwania optymalnej wartości parametru lambda. Zastosowano natomiast uproszczone założenie dla ujęcia „bottom-up”, przyjmując dla każdego z 930 miast transformację z wykorzystaniem parametru $\lambda = 0$, czyli takiego, który uzyskano dla podejścia „top-down”.

W celu określenia obserwacji skrajnych z obydwu końców rozkładu przychodów/dochodów, przyjęto następujące wartości:

- dla rozstępu międzykwartylowego $c = 5$ (por. Dupriez, 2007);
- dla medianowego odchylenia przeciętnego $c = 3,5$ (por. Denderski, 2019).

Następnie jednostkom o oryginalnych wartościach zlogarytmowanych przychodów/dochodów mniejszych niż wyznaczony punkt odcięcia w dolnym ogonie rozkładu, przypisano wartości zlogarytmowanych przychodów/dochodów równe wartości wyznaczonego punktu odcięcia w dolnym ogonie rozkładu. Analogicznie postąpiono z wartościami zlogarytmowanych przychodów/dochodów większych niż wyznaczony punkt odcięcia w górnym ogonie rozkładu.

Charakterystyka liczbowa i graficzna wybranych wyników.

Zastosowane podejścia mają różnoraki wpływ na poszczególne mierniki poziomu i zróżnicowania dochodów. Wpływ ten można oceniać zarówno w odniesieniu do rozkładu opartego na pojedynczych podatnikach/pracownikach najemnych (por. Tab. 11. – Tab. 14.), ale także w odniesieniu do rozkładu gdzie jednostkami są miasta (por. Ryc. 3. – Ryc. 6.).

Wpływ zastosowanych podejść na medianę rozkładu przychodów podatników w ujęciu top-down jest niewielki – mediana przychodów waha się w granicach od 33204,50 zł do 33458,15 zł, przy medianie dla danych surowych na poziomie 33204,48 zł. Wyraźnie większe oddziaływanie mają zastosowane podejścia na charakterystyki zróżnicowania przychodów podatników, przy czym największy wpływ widać w odniesieniu do wskaźnika zróżnicowania kwintylowego i współczynnika Giniego. Współczynnik zróżnicowania kwintylowego dla podejścia Box-Cox_MAD stanowił zaledwie 28%, a współczynnik Giniego 64% poziomu odpowiednich mierników obliczonych dla danych surowych. Najmniejszym wpływem charakteryzowało się podejście WINS oraz Box-Cox_IQR2, wyróżnione w Tablicy 11.

Tablica 11. Wpływ zastosowanych podejść na wybrane charakterystyki poziomu i zróżnicowania przychodów podatników („top-down”)

Podejście	Mediana	IQR	P90/P10	P90/P50	P50/P10	S80/S20	Wsp. Giniego
Dane surowe	33204,48	40576,51	10,39	3,29	3,16	45,43	0,75
TRIM	33204,48	40051,76	9,62	3,20	3,01	20,44	0,58
WINS	33204,48	40576,51	10,39	3,29	3,16	26,62	0,63
Box-Cox_IQR	33338,64	40054,48	9,27	3,19	2,90	19,89	0,58
Box-Cox_MAD	33458,14	38863,26	8,21	3,01	2,73	12,82	0,48
Box-Cox_IQR2	33204,50	41774,86	10,81	3,29	3,29	27,58	0,65
Box-Cox_MAD2	33204,50	41774,86	10,81	3,29	3,29	20,19	0,57

Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Wpływ zastosowanych podejść na medianę rozkładu dochodów pracowników najemnych w ujęciu top-down jest niewielki. Znacznie większe oddziaływanie mają zastosowane podejścia na charakterystyki zróżnicowania przychodów podatników, przy czym największy wpływ widać w odniesieniu do relacji decyla dziewiątego do decyla pierwszego oraz wskaźnika zróżnicowania kwintylowego. Pierwsza z powyżej wymienionych charakterystyk zróżnicowania dochodów w podejściu Box-Cox_MAD2 stanowiła zaledwie 42%, a druga stanowiła 49% poziomu odpowiednich mierników obliczonego dla danych surowych. Najmniejszym wpływem w tym przypadku charakteryzowało się podejście oparte na winsoryzacji, wyróżnione w Tablicy 12.

Tablica 12. Wpływ zastosowanych podejść na wybrane charakterystyki poziomu i zróżnicowania dochodów pracowników najemnych („top-down”)

Podejście	Mediana	IQR	P90/P10	P90/P50	P50/P10	S80/S20	Wsp. Giniego
Dane surowe	38577,83	37663,50	12,79	2,32	5,51	14,30	0,43
TRIM	38577,44	37290,16	11,81	2,29	5,17	12,53	0,41
WINS	38577,83	37663,50	12,79	2,32	5,51	13,74	0,42
Box-Cox_IQR	39398,08	37930,70	9,85	2,29	4,30	11,66	0,42
Box-Cox_MAD	40600,41	38702,69	7,77	2,25	3,46	8,72	0,39
Box-Cox_IQR2	38577,85	36735,25	5,39	2,32	2,32	7,09	0,38
Box-Cox_MAD2	38577,85	36735,25	5,39	2,32	2,32	6,97	0,38

Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Zestawiając ze sobą ujęcia *top-down* oraz *bottom-up* (odpowiednio Tab. 11. z Tab. 13. oraz Tab. 12. z Tab. 14.), nie dostrzegamy znaczącego ich wpływu na charakterystyki poziomu i zróżnicowania przychodów podatników/dochodów pracowników najemnych w ujęciu całego rozkładu. Taki wpływ powinien być dostrzegalny na poziomie analiz prowadzonych dla poszczególnych miast, przy założeniu, że obserwacje odstające nie będą rozkładać się równomiernie na obydwu ogonach rozkładów dla wszystkich 930 miast jednocześnie.

Podobnie jak w przypadku ujęcia *bottom-up*, w ujęciu *top-down* dla przychodów najmniejszy wpływ zastosowanych podejść na charakterystyki rozkładu przychodów mają podejście oparte na winsoryzacji (WINS) oraz podejście Box-Cox_IQR2 (por. Tab. 13.). W przypadku dochodów pracowników najemnych taką własnością charakteryzuje się podejście oparte na winsoryzacji (por. Tab. 14.)

Tablica 13. Wpływ zastosowanych podejść na wybrane charakterystyki poziomu i zróżnicowania przychodów podatników („bottom-up”)

Podejście	Mediana	IQR	P90/P10	P90/P50	P50/P10	S80/S20	Wsp. Giniego
Dane surowe	33204,48	40576,51	10,39	3,29	3,16	45,43	0,75
TRIM	33203,76	40054,92	9,62	3,20	3,01	20,75	0,58
WINS	33204,48	40576,51	10,39	3,29	3,16	26,96	0,64
Box-Cox_IQR	33351,36	39998,10	9,25	3,18	2,90	20,04	0,58
Box-Cox_MAD	33513,36	38836,42	8,19	3,00	2,73	13,03	0,49
Box-Cox_IQR2	33300,00	41457,69	10,42	3,28	3,18	26,60	0,65
Box-Cox_MAD2	33300,00	41457,69	10,42	3,28	3,18	19,63	0,57

Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Tablica 14. Wpływ zastosowanych podejść na wybrane charakterystyki poziomu i zróżnicowania dochodów pracowników najemnych („bottom-up”)

Podejście	Mediana	IQR	P90/P10	P90/P50	P50/P10	S80/S20	Wsp. Giniego
Dane surowe	38577,83	37663,50	12,79	2,32	5,51	14,30	0,43
TRIM	38576,56	37294,72	11,83	2,29	5,17	12,73	0,41
WINS	38577,83	37663,50	12,79	2,32	5,51	13,87	0,42
Box-Cox_IQR	39381,77	37929,88	9,91	2,29	4,32	11,76	0,42
Box-Cox_MAD	40617,86	38750,37	7,79	2,25	3,46	8,88	0,39
Box-Cox_IQR2	38953,31	36646,22	5,36	2,30	2,33	6,98	0,38
Box-Cox_MAD2	38953,31	36646,22	5,36	2,30	2,33	6,90	0,38

Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

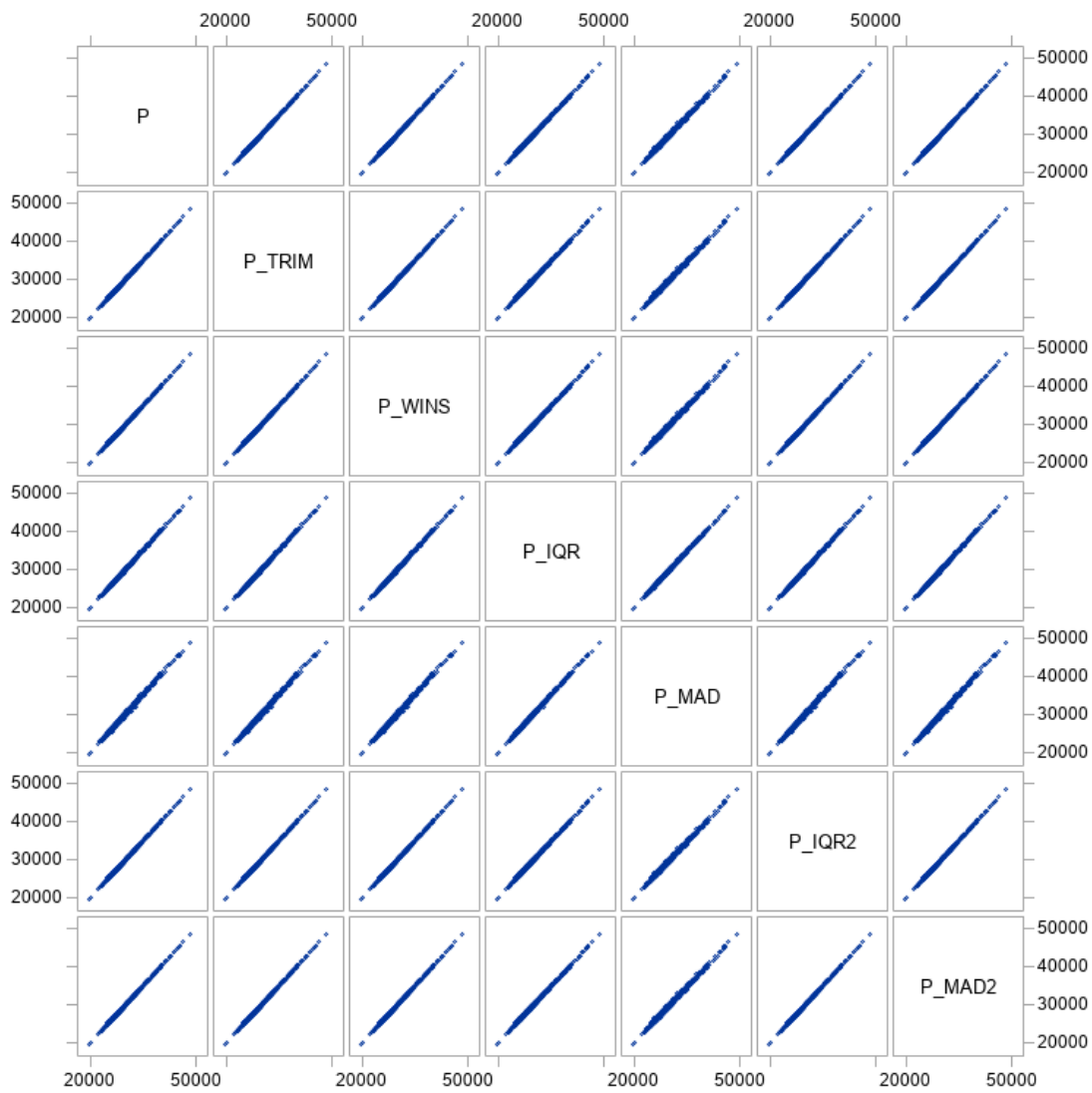
Charakterystyczną własnością mediany, która znajduje potwierdzenie na poniżej zamieszczonych rycinach (por. Ryc. 3. i Ryc. 4.) jest jej odporność na wartości odstające. Wszystkie zastosowane podejścia w detekcji i sposobach postępowania z obserwacjami odstającymi w analizie rozkładu przychodów podatników/dochodów pracowników najemnych zamieszkujących miasta dotyczą bowiem obserwacji znajdujących się na obydwu końcach rozkładów. Jak można łatwo zauważyć na Ryc. 3. zarówno mediany przychodów podatników wyliczone dla 930 miast w oparciu o dane surowe, jak i w każdym z zastosowanych podejść są ze sobą zgodne, tworząc smugę punktów leżących na prostej o równaniu $y = x$. Podobną sytuację możemy zaobserwować na Ryc. 4. w stosunku do dochodów pracowników najemnych.

Uzupełnieniem tablic 11-14 są ryciny 3-10, które prezentują wpływ zastosowanych podejść na wybrane charakterystyki poziomu i zróżnicowania przychodów podatników oraz dochodów w przypadku pracowników najemnych. Konstrukcja rycin jest przemyślana w ten sposób, aby odbiorca mógł porównać każdą parę zastosowanych podejść między sobą za pomocą wykresów zwanych trellisami³. W przypadku mediany przychodów i dochodów w ujęciu bottom-up (Ryc. 3 i Ryc. 4) wszystkie podejścia udało się zestawić w ramach jednego wykresu typu trellis, gdzie na przekątnej znajdują się nazwy⁴ zastosowanych podejść, a poza przekątnymi znajdują się wykresy rozrzutu. Na wykresach rozrzutu punktami są mediany przychodów/dochodów dla miast o współrzędnych, którymi są wartościami przyjmowane w ramach dwóch porównywanych podejść. W przypadku współczynnika Giniego (Ryc. 5- Ryc. 10), należało zastosować drobną modyfikację polegającą na rozbiściu jednego wykresu typu trellis na trzy prezentujące wpływ zastosowanych podejść w porównaniu do wskaźników obliczanych na podstawie danych surowych.

³ Szerzej na temat wykresów zwanych trellisami jako narzędzi kompleksowej prezentacji graficznej zjawisk traktuje Suhecki (2010).

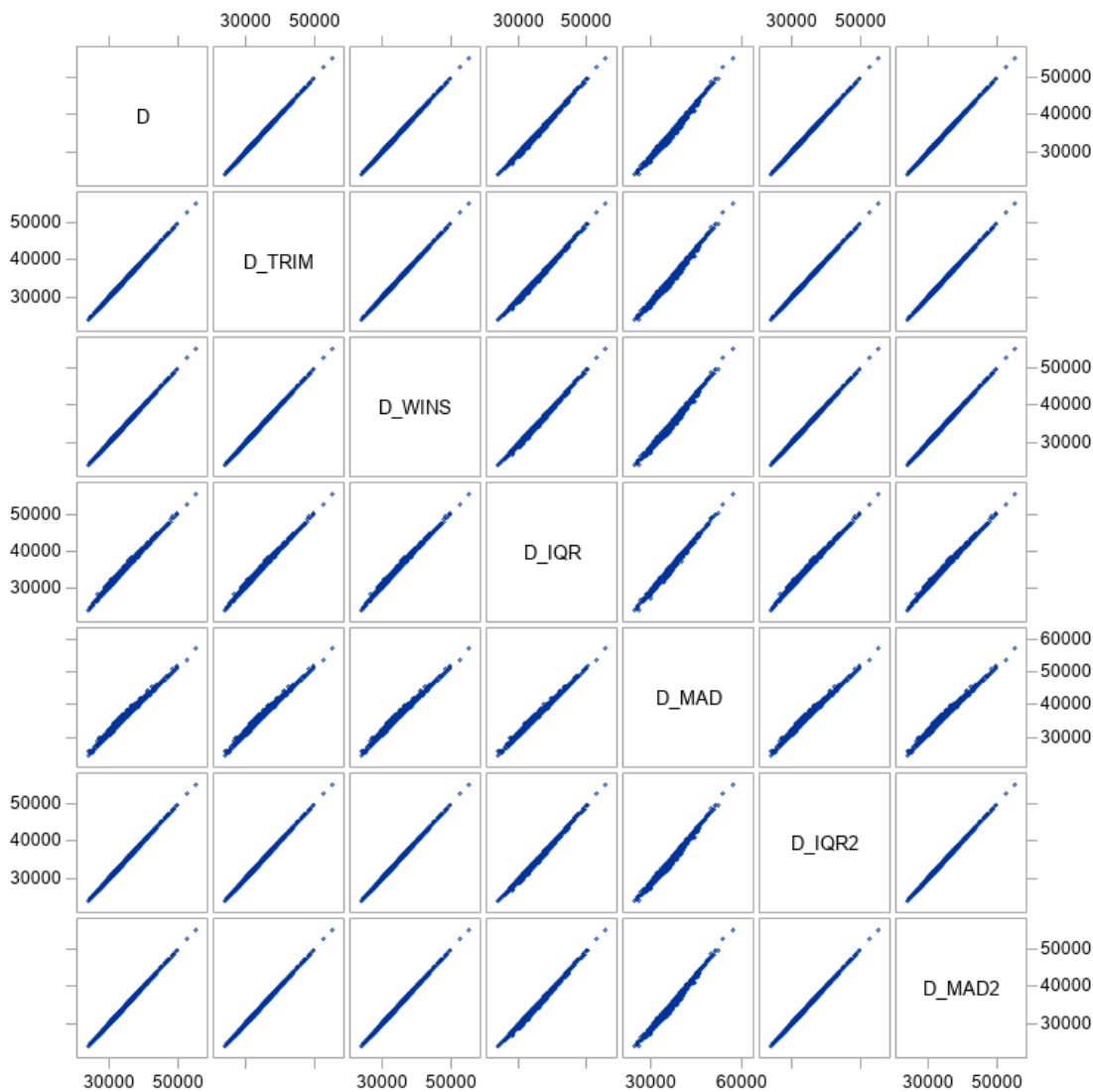
⁴ Nazwy podejść zostały zasymbolizowane następująco: dane surowe dla przychodów – P, dane surowe dla dochodów – D, dane ucięte dla przychodów – P_TRIM, dane ucięte dla dochodów – D_TRIM, itd.

Rycina 3. Wpływ zastosowanych podejść na medianę przychodów podatników w miastach („bottom-up”)



Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

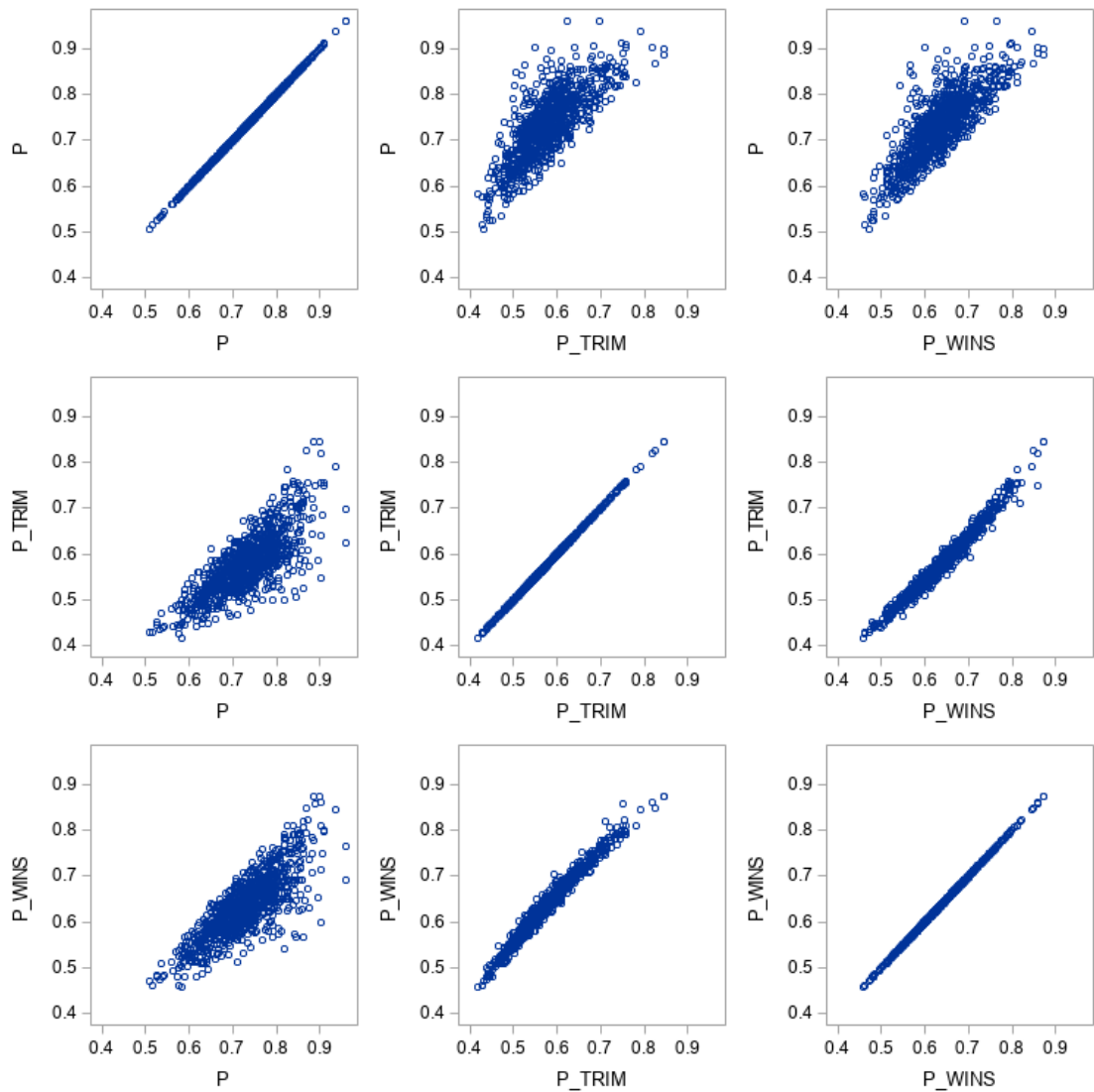
Rycina 4. Wpływ zastosowanych podejść na medianę przychodów podatników w miastach („bottom-up”)



Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

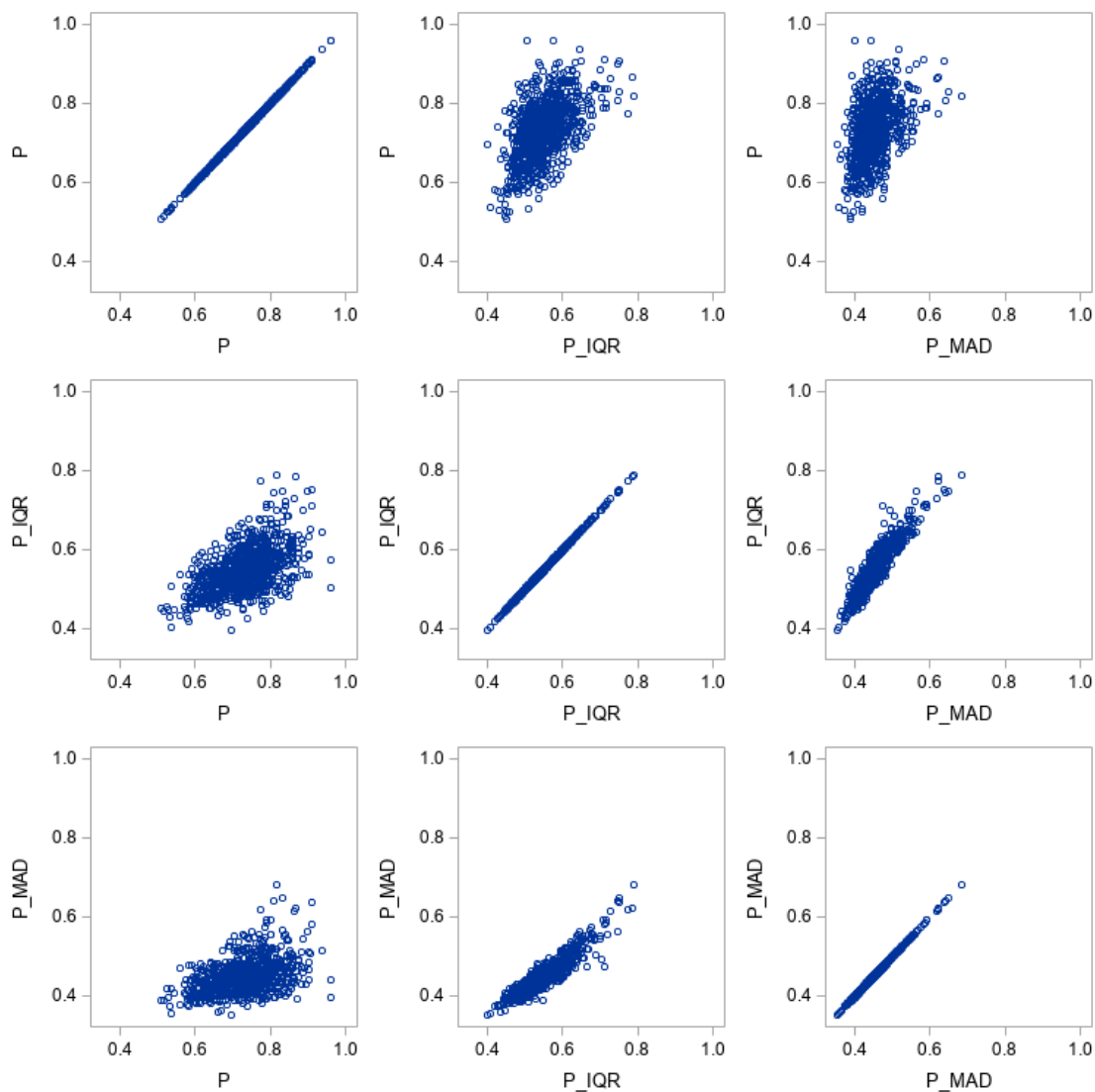
Zupełnie odmienna sytuacja występuje w odniesieniu do mierników zróżnicowania. Ze względu na ograniczenia związane z niniejszym raportem przedstawione zostaną jedynie wykresy rozrzutu dla współczynnika Giniego w ujęciu *bottom-up* dla przychodów podatników (Ryc. 5.) oraz dochodów pracowników najemnych (Ryc. 6.). Widać wyraźnie, że zastosowane podejścia znacząco, w stosunku do danych oryginalnych, zmniejszają wartości współczynnika Giniego dla praktycznie każdego z 930 miast. Ma to miejsce zarówno w odniesieniu do przychodów podatników, jak i dochodów pracowników najemnych. Tutaj modyfikacja obserwacji znajdujących się na obydwu końcach rozkładów ma bezpośredni wpływ na kształtowanie się mierników zróżnicowania. Wydaje się, że stosunkowo najmniejszy wpływ na kształtowanie się współczynnika Giniego w stosunku do danych oryginalnych w przypadku przychodów podatników mają podejścia oparte na winsoryzacji (WINS) oraz Box-Cox_IQR2. Natomiast w przypadku dochodów pracowników najemnych są to podejścia oparte na winsoryzacji i Box-Cox_IQR.

Rycina 5. Wpływ zastosowanych podejść na współczynnik Giniego przychodów podatków w miastach („bottom-up”)



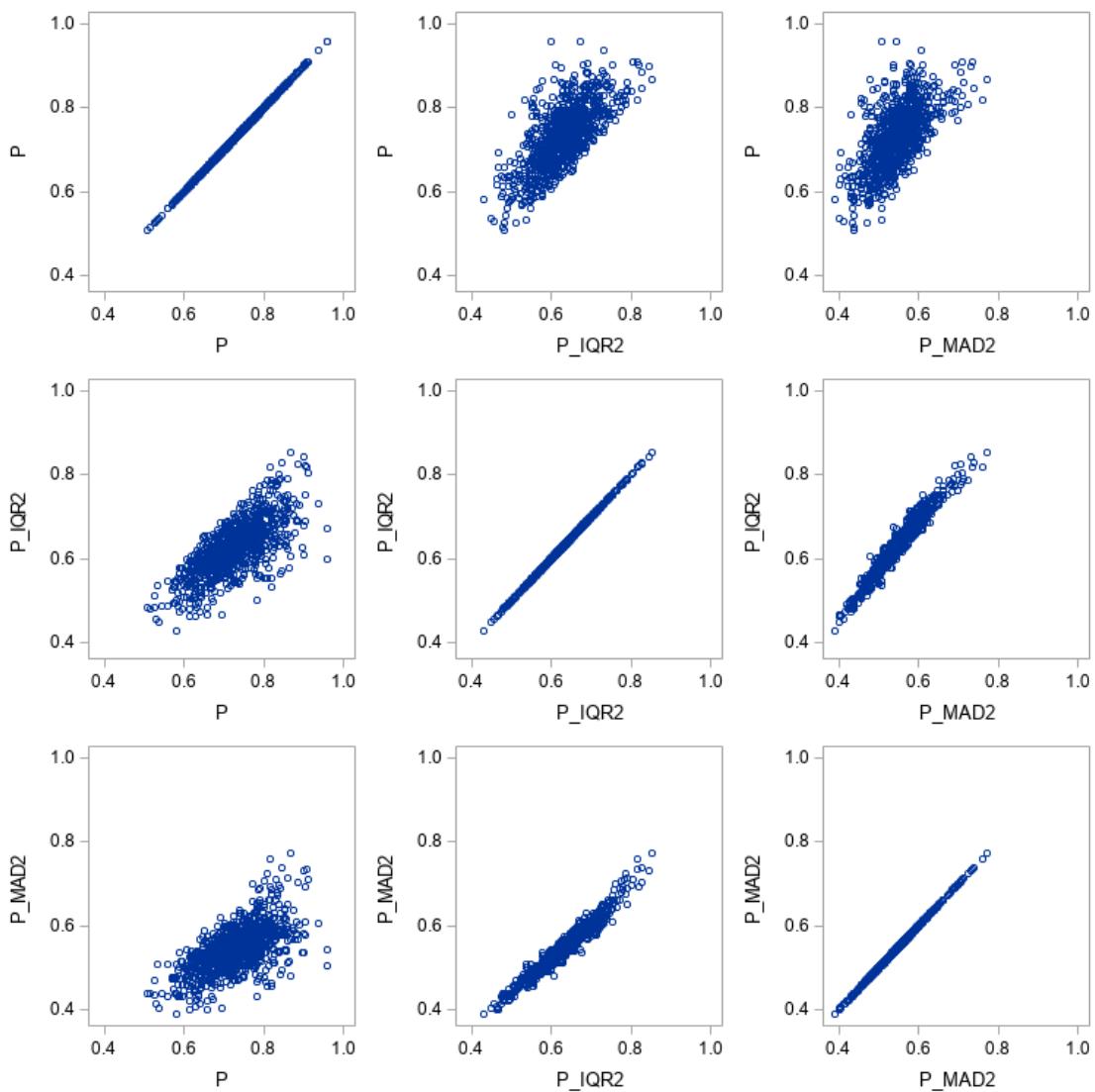
Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Rycina 6. Wpływ zastosowanych podejść na współczynnik Giniego przychodów podatków w miastach („bottom-up”) cd.



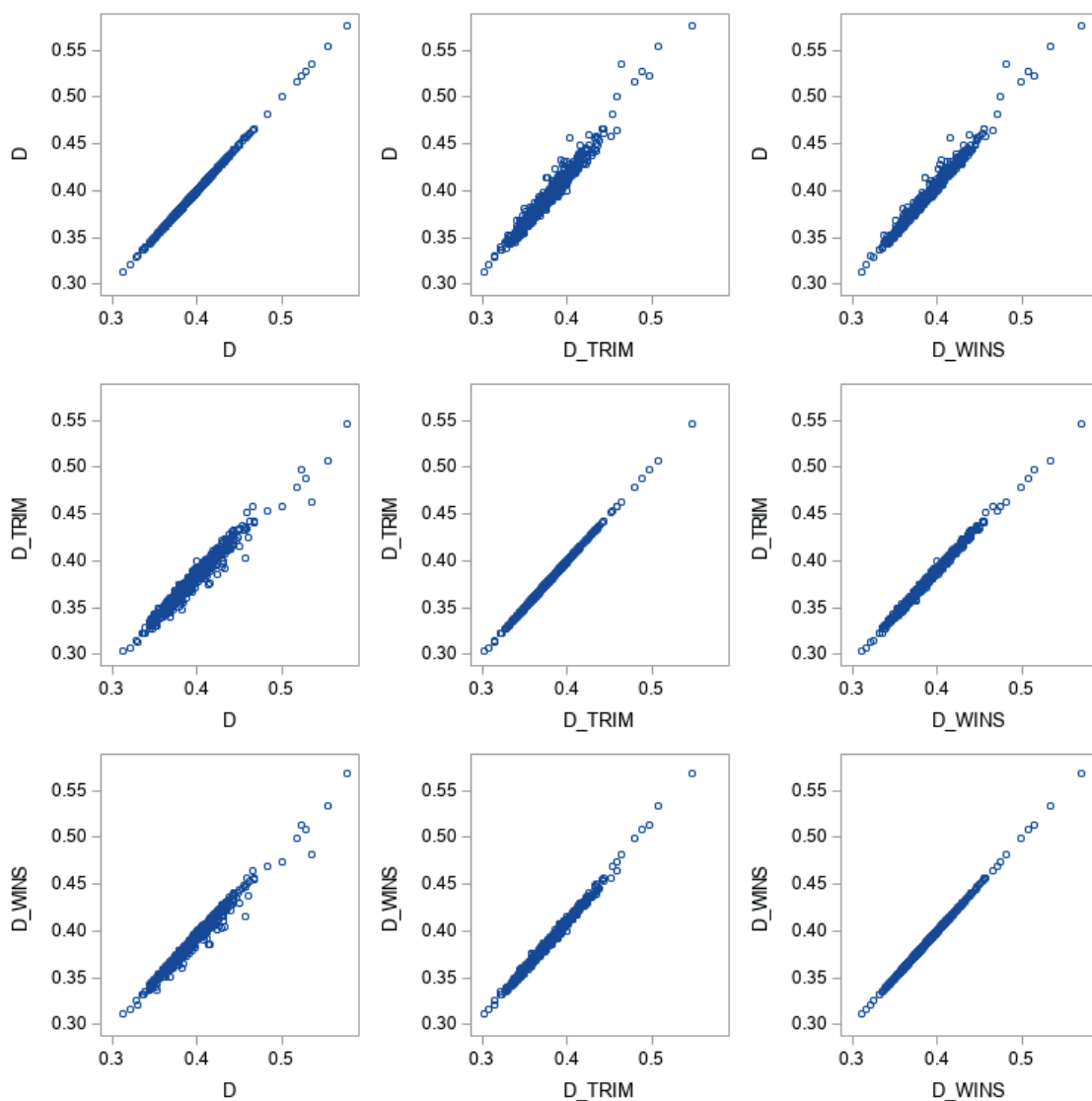
Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Rycina 7. Wpływ zastosowanych podejść na współczynnik Ginię przychodów podatków w miastach („bottom-up”) dok.



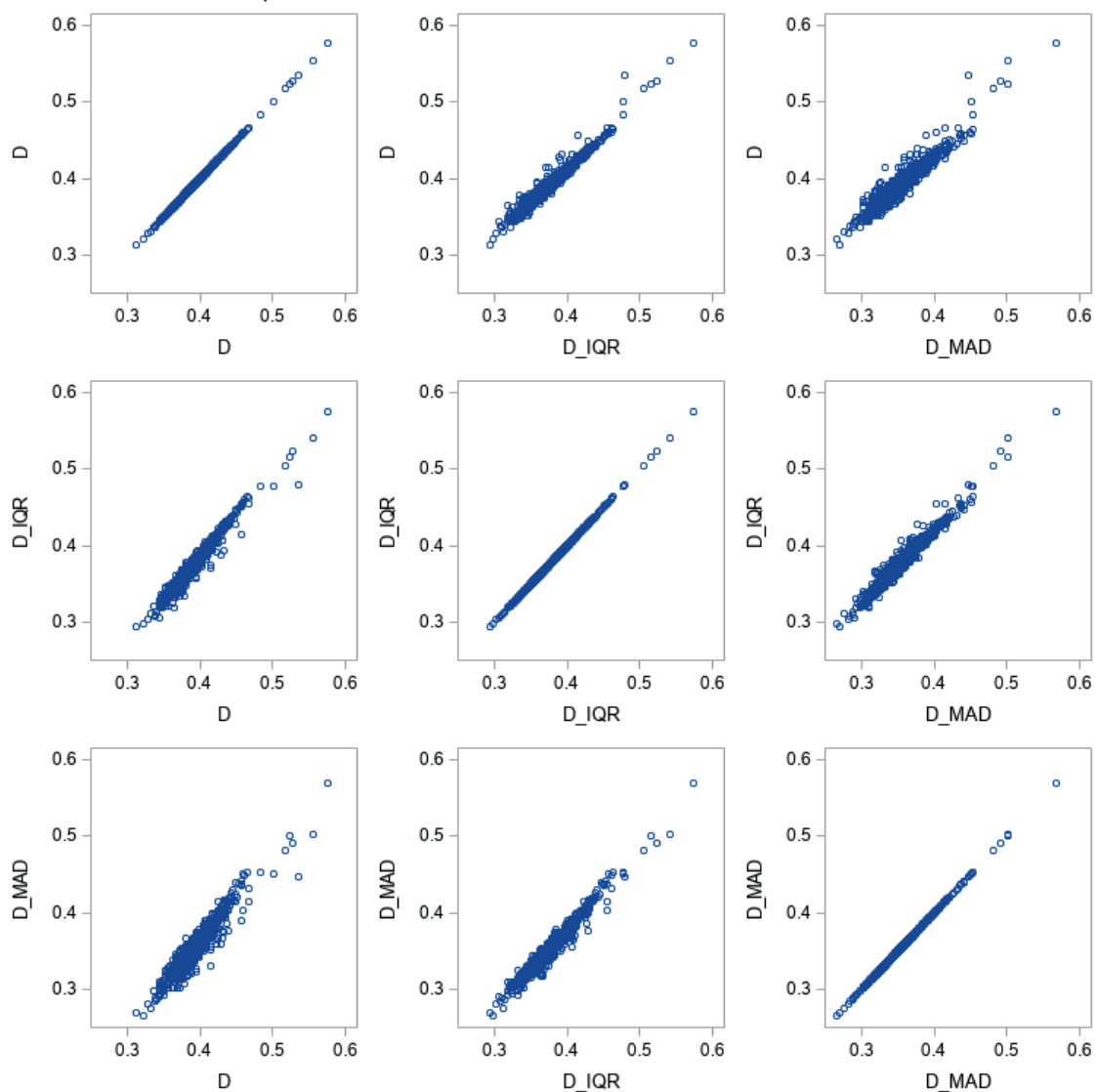
Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Rycina 8. Wpływ zastosowanych podejść na współczynnik Giniego dochodów pracowników najemnych w miastach („bottom-up”).



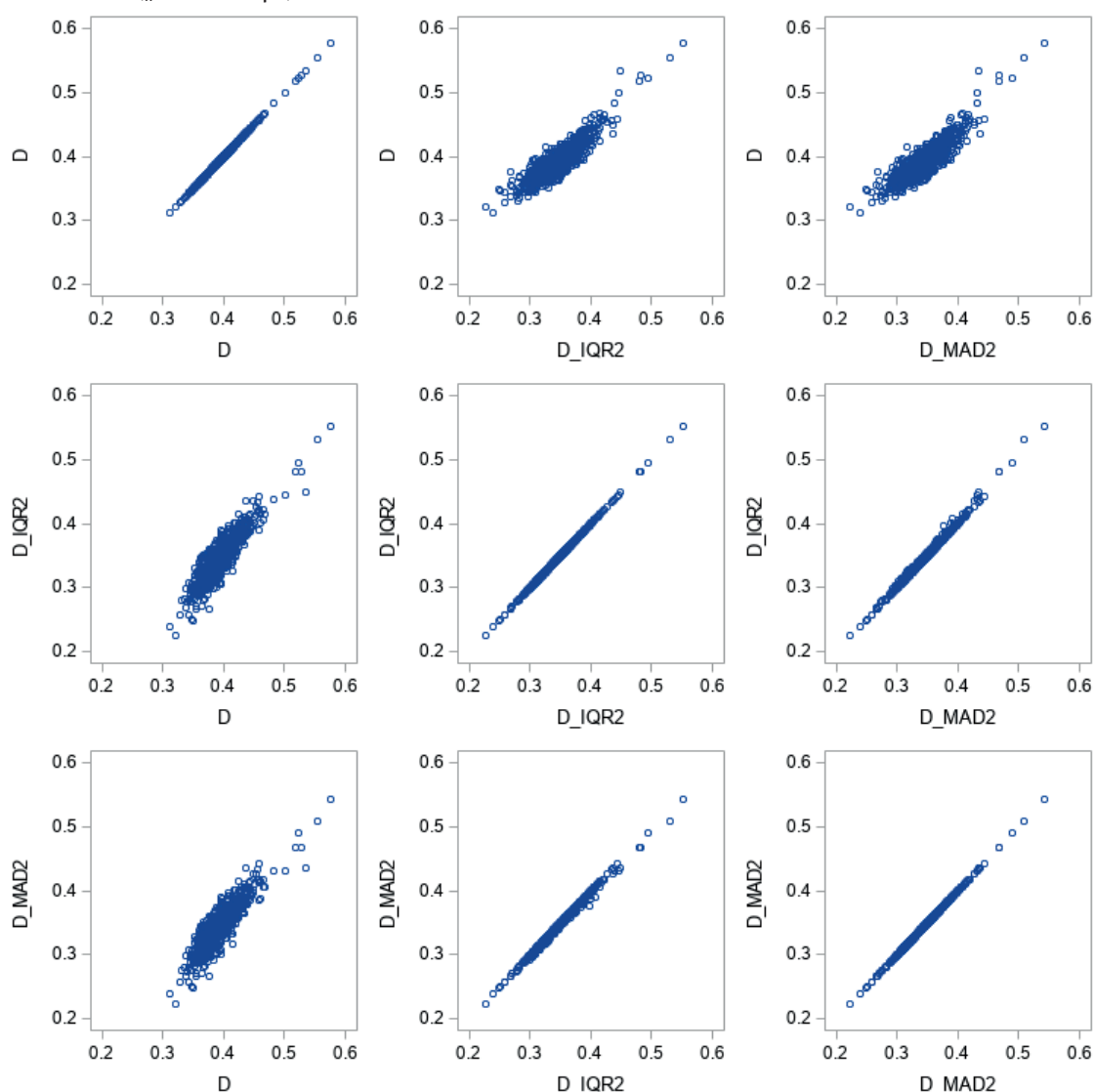
Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Rycina 9. Wpływ zastosowanych podejść na współczynnik Giniego dochodów pracowników najemnych w miastach („bottom-up”) cd.



Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Rycina 10. Wpływ zastosowanych podejść na współczynnik Giniego dochodów pracowników najemnych w miastach („bottom-up”) dok.



Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Wnioski z analizy poświęconej detekcji i sposobom postępowania z obserwacjami odstającymi

W analizie rozkładu przychodów podatników/dochodów pracowników najemnych zamieszkujących miasta autorzy niniejszej pracy przeprowadzili szczegółowe badanie wpływu zastosowanych podejść w detekcji i sposobach postępowania z obserwacjami odstającymi. Na potrzeby niniejszej pracy zdecydowano jednak o przeprowadzeniu stratyfikacji dochodowej mieszkańców miast wyłącznie w oparciu o dane surowe. Wynika to z następujących przesłanek:

1) Cel pracy

Kluczowym aspektem stratyfikacji dochodowej mieszkańców miast jest analiza zróżnicowania przychodów/dochodów. Przeprowadzone zabiegi analityczne wskazują, że wszystkie z zastosowanych podejść w detekcji i sposobach postępowania z obserwacjami odstającymi mają istotny wpływ na najczęściej stosowane mierniki zróżnicowania przychodów/dochodów, tzn. relację decyla dzie-

wiątego do pierwszego P90/P10, wskaźnik zróżnicowania kwintylowego S80/S20 oraz współczynnik Giniego. To zniekształcenie pomiaru zróżnicowania oraz kolejna przesłanka (patrz następny akapit) stanowią kluczowe argumenty przemawiające za oparciem analiz jedynie na danych surowych.

2) Brak dodatkowych informacji o jakości danych o przychodach/dochodach w zbiorach Ministerstwa Finansów.

Autorzy niniejszego raportu oparli analizę detekcji i sposobów postępowania z obserwacjami odstającymi w rozkładach przychodów podatników/dochodów pracowników najemnych zamieszkujących miasta na zbiorach udostępnionych przez Ministerstwo Finansów. Należy podkreślić, że na tym etapie prac nie dysponowali żadnym dodatkowymi informacjami na temat jakości danych w przekazanych zbiorach. Wydaje się zatem, że oparcie analizy tylko na statystycznych metodach wykrywania obserwacji odstających bez uwzględnienia aspektów merytorycznych, stanowi zbyt wątplą przesłankę do „odrzućcia” metod imputacji dla nawet niewielkiej części informacji w bazach uzyskanych od Ministerstwa Finansów. Należy mieć jednak nadzieję, że niniejsza praca będzie stanowiła pretekst do podjęcia ściślejszej współpracy z gestorem, dzięki której powstaną przesłanki oparte na metodach statystycznych i wiedzy merytorycznej, pozwalające odpowiednio radzić sobie z „problematycznymi” danymi.

3) Jak to robi gestor danych - Ministerstwo Finansów?

Dodatkową przesłanką, która wzmacnia dwie poprzednie, jest publikacja *Wybrane aspekty systemu podatkowo-składkowego na podstawie danych PIT i ZUS 2016*, wydana przez Ministerstwo Finansów w ramach cyklu *Opracowania i analizy*, w której autorzy stosują w zasadzie jedynie dwa podejścia, tzn. z uwzględnieniem zerowych przychodów/dochodów bądź bez ich uwzględnienia. Należy podkreślić, że żadnych innych metod wykrywania obserwacji odstających i radzenia sobie z nimi analitycy Ministerstwa Finansów nie stosowali.

II. Analiza uzyskanych wyników

II. Analysis of results

Znacznym utrudnieniem badań przestrzenno-ekonomicznych jest brak dostępności do informacji statystycznych lub brak odpowiedniej agregacji przestrzennej. Z perspektywy polityki społecznej czy regionalnej ważną dziedziną są dochody ludności.

W zasobach statystyki publicznej znajdują się dane dotyczące dochodów (głównie w odniesieniu do gospodarstw domowych) pochodzące z badań reprezentacyjnych. Stąd dostępne dane dotyczą poziomu wojewódzkiego lub wyższych poziomów przestrzennych. Zauważalny jest brak informacji opisujących poziom i źródła dochodów na niższych poziomach agregacji, np. w odniesieniu do miast i obszarów wewnątrzmijskich. Brak danych dla niskich poziomów agregacji jest istotny ze względu na fakt, że zróżnicowanie dochodów osobistych jest widoczne głównie na niższych poziomach przestrzennych.

Jak zaznaczono na wstępie, badanie charakteryzowane w niniejszym raporcie dotyczy przychodów wszystkich podatników, jak i dochodów wybranej grupy podatników, tzn. osób zatrudnionych na podstawie umowy o pracę (osoby otrzymujące należności ze stosunku: pracy, służbowego, spółdzielczego i z pracy nakładczej, a także zasiłki pieniężne z ubezpieczenia społecznego wypłacone przez zakład pracy). Zarówno sytuacja odnosząca się do przychodów, jak i dochodów podatników została przedstawiona w różnych aspektach:

- przestrzennym - kraj, województwa, w części tabelarycznej stanowiącej załącznik do Raportu także dla miast imiennie, a dla miast wojewódzkich również w układzie siatki kwadratów;
- typologicznym - miasta duże, średnie i małe, miasta tracące funkcje społeczno-gospodarcze, miasta wojewódzkie.

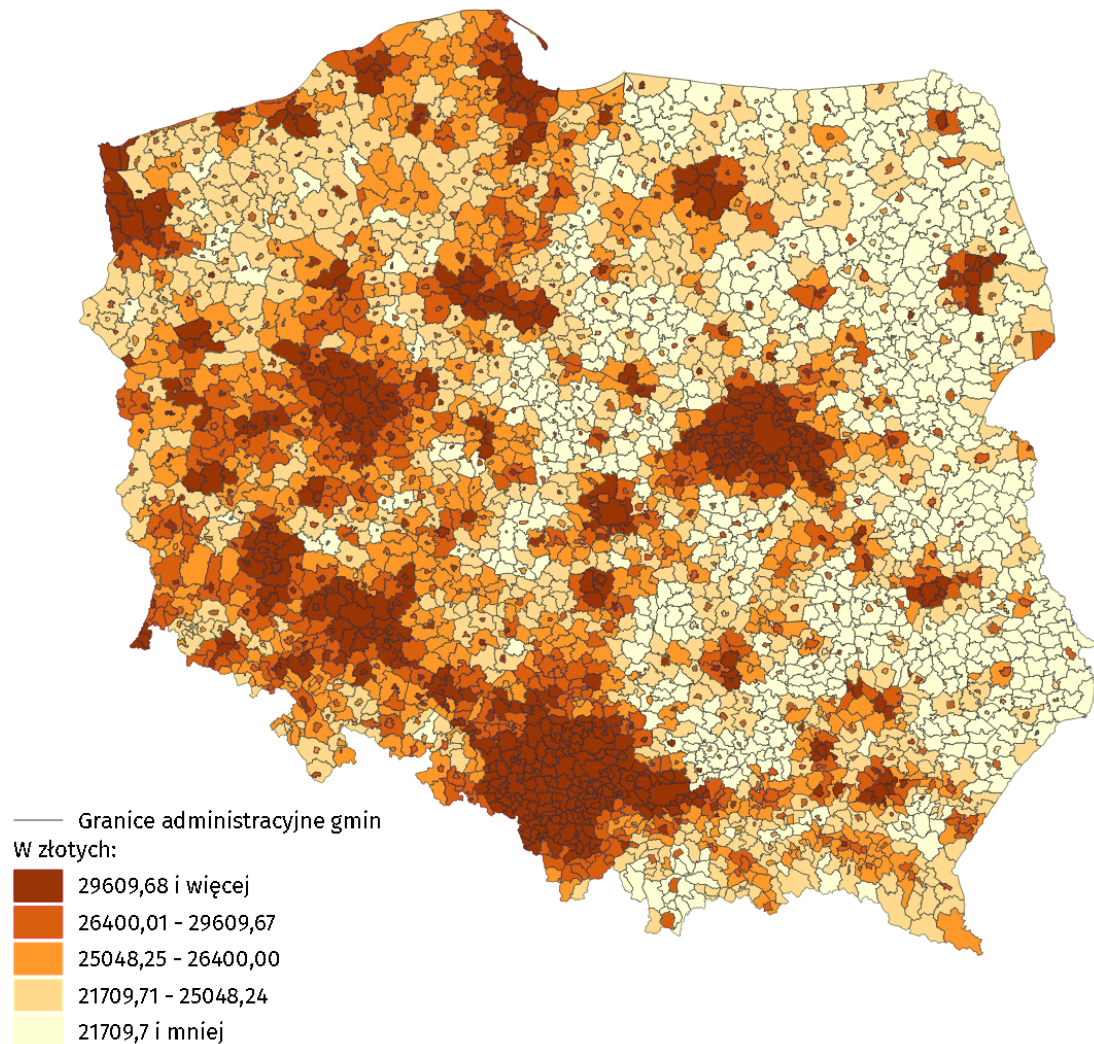
Do scharakteryzowania "sytuacji dochodowej" wykorzystano miary położenia oraz zróżnicowania, których szczegółowy opis znajduje się w części metodologicznej raportu.

Przychody podatników mieszkających w miastach

Celem tej części opracowania jest zaprezentowanie przychodów podatników, dla których zostały złożone do urzędów skarbowych zeznania podatkowe na stosownych formularzach PIT (wykaz tych formularzy przedstawiono w części dotyczącej metodologii badania).

W zbiorowości wszystkich gmin kraju wielkość przychodów, reprezentowaną przez ich wartość środkową, ilustruje niżej zamieszczony kartogram (Mapa 1.).

Mapa 1. Mediana przychodów (w złotych) ludności w 2018 r. według gmin w kraju



Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Powyższy kartogram ma jedynie charakter poglądowy służący zobrazowaniu rozkładu przychodów w przestrzeni całego kraju i jednocześnie ma podkreślić pozycję miast, które są przedmiotem dalszej analizy.

W zbiorowości podatników zamieszkujących w miastach wartość środkowa przychodów (mediana) wyniosła 33204,48 złotych.

Miarą pozwalającą określić, jak bardzo różnią się od siebie przychody podatników uzyskujących bardzo małe i bardzo duże przychody jest relacja decyla dziewiątego i decyla pierwszego (P90/P10). Im większe są nierówności w przychodach podatników, tym większa jest wartość tego wskaźnika. Dla ogółu miast w kraju w 2018 r. wskaźnik ten uzyskał wartość 10,39. Oznacza to, że osoba na granicy 10% populacji o najwyższych przychodach osiągnęła w 2018 r. przychody będące 10-krotnością przychodów uzyskanych przez osobę znajdującą się na granicy 10% populacji o najniższych przychodach.

Miarą zróżnicowania przychodów jest także wskaźnik zróżnicowania kwintylowego (S80/S20) i współczynnik Giniego. Stosunek S80/S20 jest relacją sumy przychodów wszystkich podatników znajdujących się w grupie 20% osób o najwyższych przychodach i sumy przychodów wszystkich podatników znajdujących się w grupie 20% osób o najniższych przychodach. W 2018 r. w miastach ogółem w kraju wskaźnik ten wyniósł 45,43, natomiast współczynnik Giniego wyniósł 0,75, co wskazuje na duże zróżnicowanie przychodów tej subpopulacji podatników.

Wyniki pracy badawczej pozwoliły na scharakteryzowanie nierówności w przychodach według płci i wieku podatników.

W subpopulacji mężczyzn mediana przychodów uzyskanych w 2018 r. wyniosła 40590,94 zł i była wyższa o 12436,68 zł od wartości dla kohorty kobiet oraz o 7386,46 zł od wartości dla kraju. O większych nierównościach przychodów wśród mężczyzn świadczy także rozstęp międzykwartylowy (IQR), który dla tej subpopulacji wynosił 47871,15 zł (dla kobiet odpowiednio - 32668,47 zł).

W kontekście przychodów podatników ważna jest informacja o ich kształtowaniu się w aspekcie ekonomicznych grup wieku. Najwyższa wartość mediany wystąpiła w kohorcie podatników będących w wieku produkcyjnym niemobilnym i wyniosła 46020,00 zł. Była ona o 38,6% wyższa od analogicznej miary dla ogółu podatników w miastach. Niższą wartość mediany odnotowano w grupie osób będących w wieku produkcyjnym mobilnym 38388,30 zł. W pozostałych dwóch grupach wieku ekonomicznego, tj. w wieku poprodukcyjnym i przedprodukcyjnym, wyniosła odpowiednio: 25392,30 zł i 1622,47 zł.

Również największe zróżnicowanie badanej cechy - mierzone rozstępem ćwiartkowym (IQR) - wystąpiło w grupie osób będących w wieku produkcyjnym niemobilnym i wyniosło 51565,10 zł. Mniejszy rozstęp ćwiartkowy wystąpił wśród podatników w wieku produkcyjnym mobilnym i poprodukcyjnym (odpowiednio: 48964,73 zł i 18797,02 zł).

Relacja górnej wartości dziewiątego i pierwszego decyla rozkładu przychodów (wskaźnik P90/P10) osiągnęła największą wartość w grupie podatników będących w wieku produkcyjnym mobilnym (23,39). Wskazuje ona na znacznie większe zróżnicowanie przychodów w tej grupie w porównaniu z pozostałymi grupami wiekowymi podatników.

Wśród podatników w wieku produkcyjnym mobilnym relacja S80/S20 osiągnęła wartość 70,22, a w subpopulacji osób będących w wieku produkcyjnym niemobilnym - 63,16.

Na znaczne zróżnicowanie w rozkładzie przychodów między poszczególnymi subpopulacjami według ekonomicznych grup wieku wskazuje również współczynnik Giniego (od 0,62 do 0,82).

Analiza rozkładu przychodów w populacji podatników została poszerzona o aspekt typologiczny i przestrzenny miast.

- **Aspekt typologiczny**

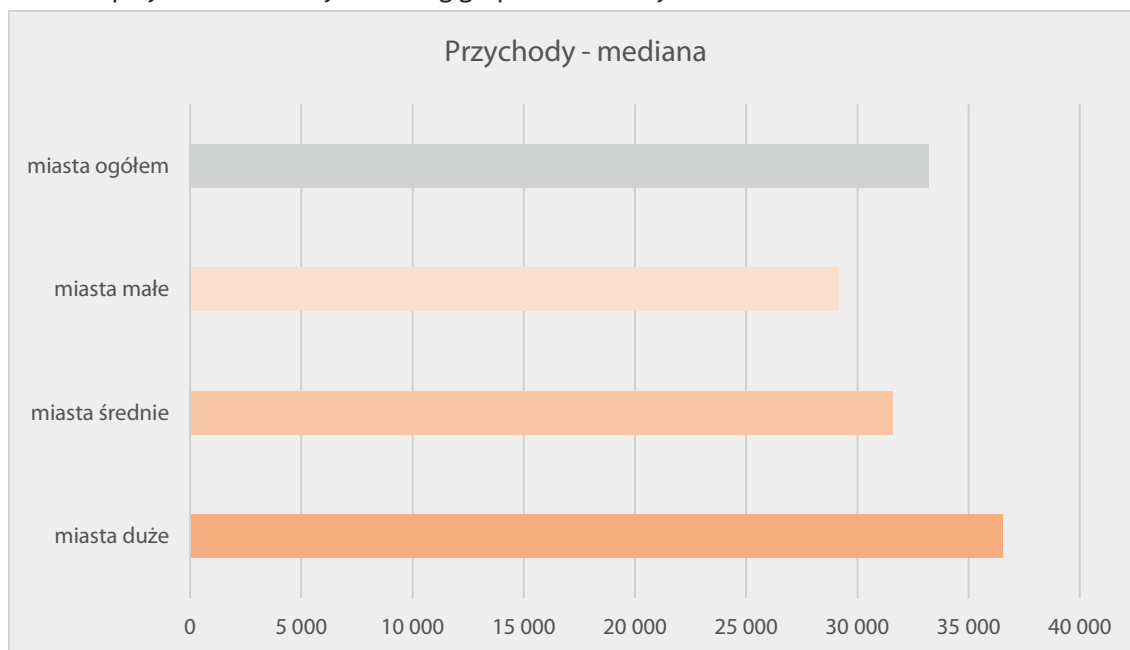
Wykorzystując typologię miast opartą wyłącznie na kryterium ludnościowym, poniżej przedstawiono wyniki analizy nierówności w przychodach podatników w podziale na trzy kategorie miast: małe, średnie i duże.

W całym 2018 r. w miastach ogółem w kraju połowa podatników objętych badaniem uzyskała przychody nie wyższe niż 33204,48 zł. Najbogatszymi pod względem wartości mediany byli podatnicy mieszkający w miastach dużych, w których mediana przychodów wyniosła 36544,22 zł. Z kolei najuboższymi byli mieszkańcy miast małych (29142,24 zł). Dla populacji podatników zamieszkujących w miastach średnich mediana była niższa o 4985,43 zł od analogicznej wartości odnotowanej dla miast dużych i o 2416,55 zł wyższa od mediany dla miast małych.

Wartość rozstępu ćwiartkowego (IQR) wskazuje, że największe zróżnicowanie przychodów miało miejsce w grupie miast dużych (45916,26 zł), a najmniejsze w miastach małych (34556,85 zł).

Identyczne tendencje uwidacznia analiza wskaźnika P90/P10, który w rozkładzie przychodów był najwyższy w grupie miast dużych i wyniósł 11,49 (w miastach: średnich - 9,34, a małych - 9,07).

Rycina 11. Mediana przychodów w złotych według grup wielkościowych miast w 2018 r.



Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Tablica 15. Wskaźniki zróżnicowania przychodów mieszkańców miast według grup wielkościowych miast w 2018 r.

Miasta	P90/P10	S80/S20	Wsp. Giniego
Ogółem	10,39	45,43	0,75
Duże	11,49	47,22	0,75
Średnie	9,34	41,98	0,74
Małe	9,07	44,48	0,76

Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Zróżnicowanie przychodów mierzone wskaźnikiem zróżnicowania kwintylowego (S80/S20) osiągnęło najwyższą wartość w populacji miast dużych i wyniosło - 47,22. Stosunek ten w pozostałych dwóch grupach wielkościowych miast wyniósł odpowiednio w miastach średnich - 41,98, a małych - 44,48.

Współczynnik Giniego we wszystkich grupach wielkościowych miast kształtował się na zbliżonym poziomie i zamykał się w przedziale od 0,74 w miastach średnich do 0,76 w miastach małych.

W zbiorowości miast dużych funkcjonują miasta będące stolicami województw, które rzutują na wielkości przychodów tej kategorii miast ogółem.

Mediana przychodów uzyskanych w 2018 r. przez podatników zamieszkujących w tych miastach wyniosła 37843,76 zł. Była ona wyższa o 1299,54 zł od wartości dla całej zbiorowości miast dużych i o 4639,28 zł od wartości dla ogółu miast w kraju.

Miasta wojewódzkie charakteryzują się także największym zróżnicowaniem przychodów opisywanym przez wskaźnik zróżnicowania kwintylowego (S80/S20), który wyniósł 48,85.

Współczynnik Giniego odnotowany dla miast wojewódzkich (0,75) był na tym samym poziomie jak w grupie miast dużych i w całej zbiorowości miast.

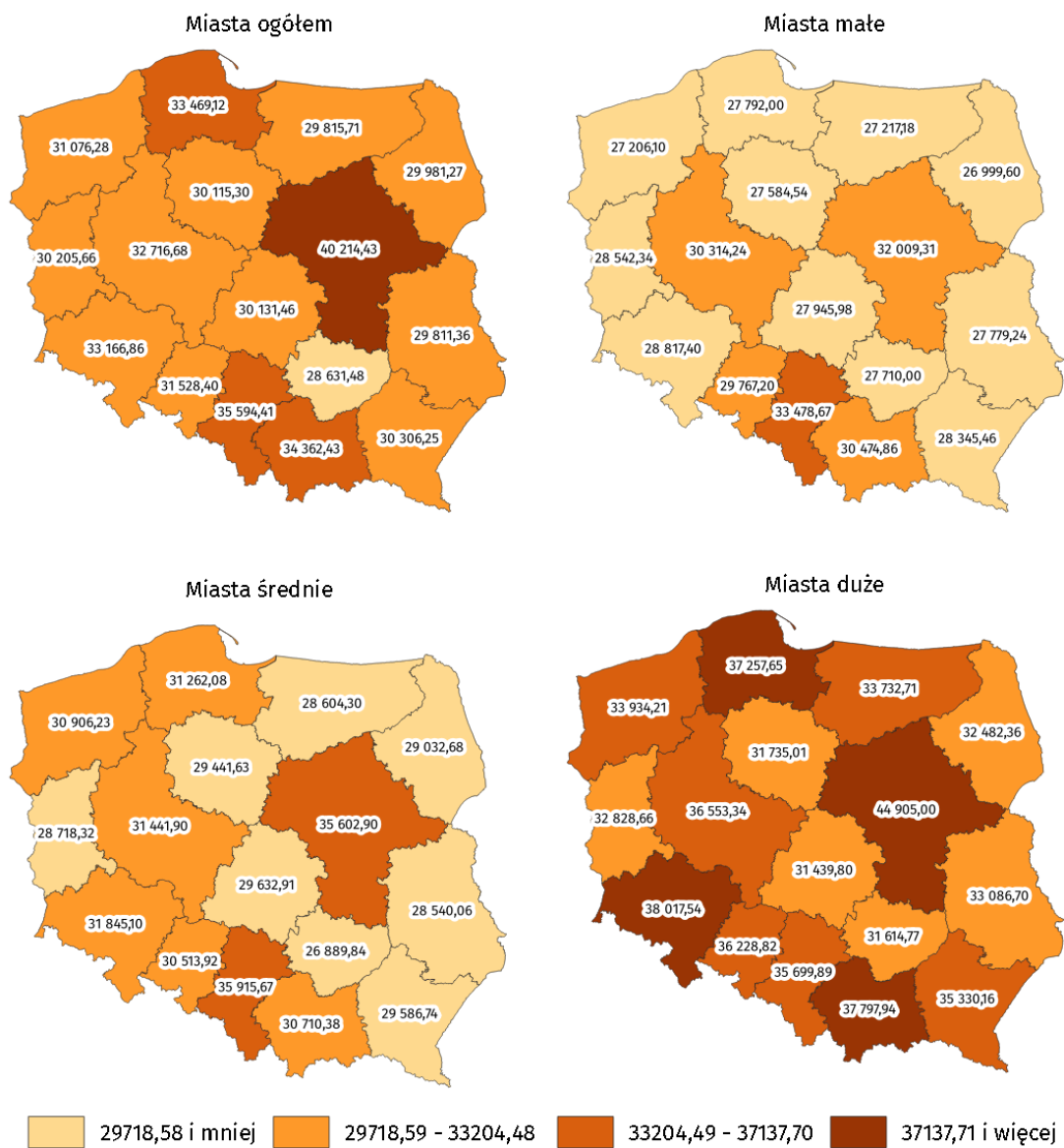
- **Aspekt przestrzenny**

Przestrzenne zróżnicowanie przychodów podatników w przekroju województw uwzględniające również trzy grupy wielkościowe miast zaprezentowano poniżej za pomocą mediany.

W 2018 r. w zbiorowości miast poszczególnych województw najbogatsi (pod względem wartości mediany) mieszkańcy zamieszkiwali w województwie mazowieckim (40214,43 zł), a najbiedniejsi w województwie świętokrzyskim (28631,48 zł). Wartość mediany powyżej 40 tys. złotych wystąpiła jedynie w miastach województwa mazowieckiego, natomiast w miastach czterech województw mediana przychodów nie przekroczyła 30 tys. złotych.

W zbiorowości województw rozpiętość wartości mediany była największa w grupie miast dużych (13465,20 zł), a najmniejsza wśród miast małych (6479,07 zł). Skrajne wartości tej miary w miastach małych zamykają się w przedziale od 26999,60 zł (woj. podlaskie) do 33478,67 zł (woj. śląskie). W grupie miast średnich skrajne wartości mediany wystąpiły dla miast województwa świętokrzyskiego (26889,84 zł) i śląskiego (35915,67 zł), a w grupie miast dużych w łódzkim - 31439,80 zł i w mazowieckim - 44905,00 zł.

Mapa 2. Mediana przychodów (w złotych) według województw i grup wielkościowych miast w 2018 r.



Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Tablica 16. Wartość środkowa przychodów (w złotych) według województw i grup wielkościowych miast w 2018 r.

Województwo	Miasta ogółem	Grupy wielkościowe miast		
		małe	średnie	duże
Dolnośląskie	33166,86	28817,40	31845,10	38017,54
Kujawsko-pomorskie	30115,30	27584,54	29441,63	31735,01
Lubelskie	29811,36	27779,24	28540,06	33086,70
Lubuskie	30205,66	28542,34	28718,32	32828,66
Łódzkie	30131,46	27945,98	29632,91	31439,80
Małopolskie	34362,43	30474,86	30710,38	37797,95
Mazowieckie	40214,43	32009,31	35602,90	44905,00
Opolskie	31528,40	29767,20	30513,92	36228,82
Podkarpackie	30306,25	28345,46	29586,74	35330,16
Podlaskie	29981,27	26999,60	29032,68	32482,36
Pomorskie	33469,12	27792,00	31262,08	37257,66
Śląskie	35594,41	33478,67	35915,67	35699,89
Świętokrzyskie	28631,48	27710,00	26889,84	31614,77
Warmińsko-mazurskie	29815,71	27217,18	28604,30	33732,72
Wielkopolskie	32716,68	30314,24	31441,90	36553,34
Zachodniopomorskie	31076,28	27206,10	30906,23	33934,21

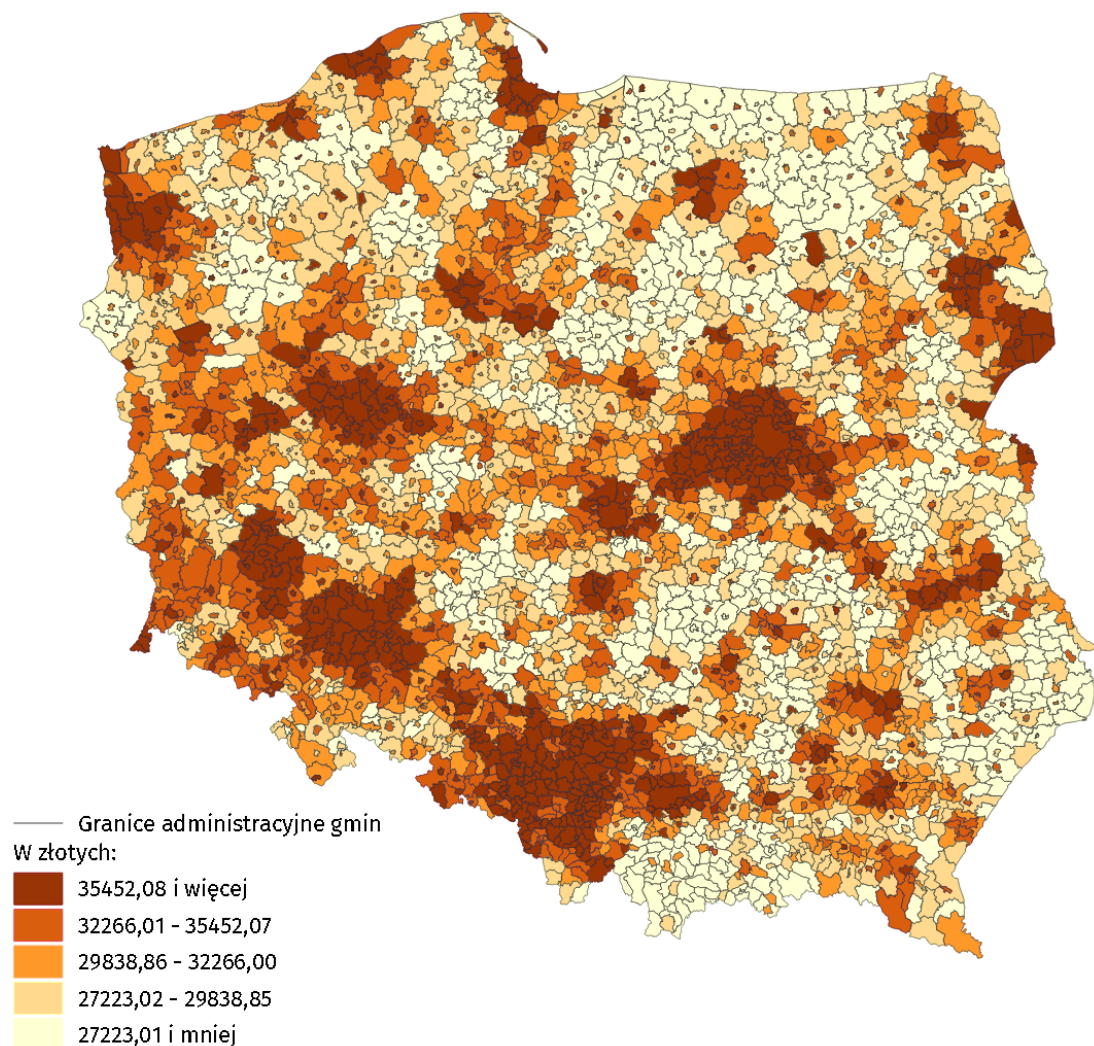
Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Dochody pracowników najemnych mieszkających w miastach

Niżej przedstawione wyniki analizy dotyczą podatników - mieszkańców miast - dla których w 2018 r. źródłem dochodów były "należności ze stosunków: pracy, służbowego, spółdzielczego i z pracy nakładczej, a także zasiłki pieniężne z ubezpieczenia społecznego wypłacane przez zakład pracy". Dla zwiększenia czytelności, w dalszej części opracowania będą nazywane osobami "zatrudnionymi na podstawie umowy o pracę". Dla tej grupy podatników analizie poddano ich dochody (przychód minus koszty uzyskania przychodu).

W zbiorowości wszystkich gmin kraju wielkość dochodów, reprezentowana przez ich wartość środkową, pokazuje niżej zamieszczony kartogram (Mapa 3.).

Mapa 3. Mediana dochodów (w złotych) ludności w 2018 r. według gmin w kraju



Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Podobnie jak w przypadku przychodów, kartogram ten ma jedynie charakter ilustracyjny służący zobrazowaniu rozkładu dochodów w przestrzeni całego kraju i jednocześnie pokazujący pozycję miast, które są przedmiotem dalszej analizy.

Dochód podatników (ponad 8 milionów osób) uzyskany z wyżej opisanego źródła przychodu, jak informuje Ministerstwo Finansów⁵, stanowił około 60% dochodów ogółem uzyskanych w 2018 r.

W zbiorowości podatników zatrudnionych na podstawie umowy o pracę wartość środkowa dochodów (mediana) wynosiła 38 577,83 złotych.

Dochody podatników znajdujących się relatywnie w najgorszej sytuacji dochodowej (1 grupa decylowa) stanowią w 2018 r. zaledwie 0,64% dochodów, podczas gdy suma dochodów podatników kwalifikujących się do ostatniej grupy decylowej (relatywnie najbogatszych) stanowi 30,60% ogółu dochodów uzyskanych przez podatników.

⁵ Informacja dotycząca rozliczenia podatku dochodowego od osób fizycznych za 2018 r.

Tablica 17. Dochody według grup decylowych w 2018 r.

Grupa decylowa	Udział w liczbie podatników	Udział w dochodach ogółem	Wysokość dochodów
1	10%	0,64%	poniżej 7005,39
2	10%	2,62%	7005,39 - 18626,82
3	10%	4,72%	18626,82 - 24275,83
4	10%	5,77%	24275,83 - 31204,72
5	10%	7,27%	31204,72 - 38577,83
6	10%	8,89%	38577,83 - 46691,60
7	10%	10,67%	46691,60 - 55760,58
8	10%	12,80%	55760,58 - 67487,24
9	10%	16,04%	67487,24 - 89564,01
10	10%	30,60%	powyżej 89564,01

Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Inną miarą zróżnicowania pozwalającą określić, jak bardzo różnią się od siebie dochody podatników uzyskujących bardzo małe i bardzo duże dochody, jest stosunek P90 i P10. Im większa jest nierówność dochodowa, tym większa jest wartość tego wskaźnika.

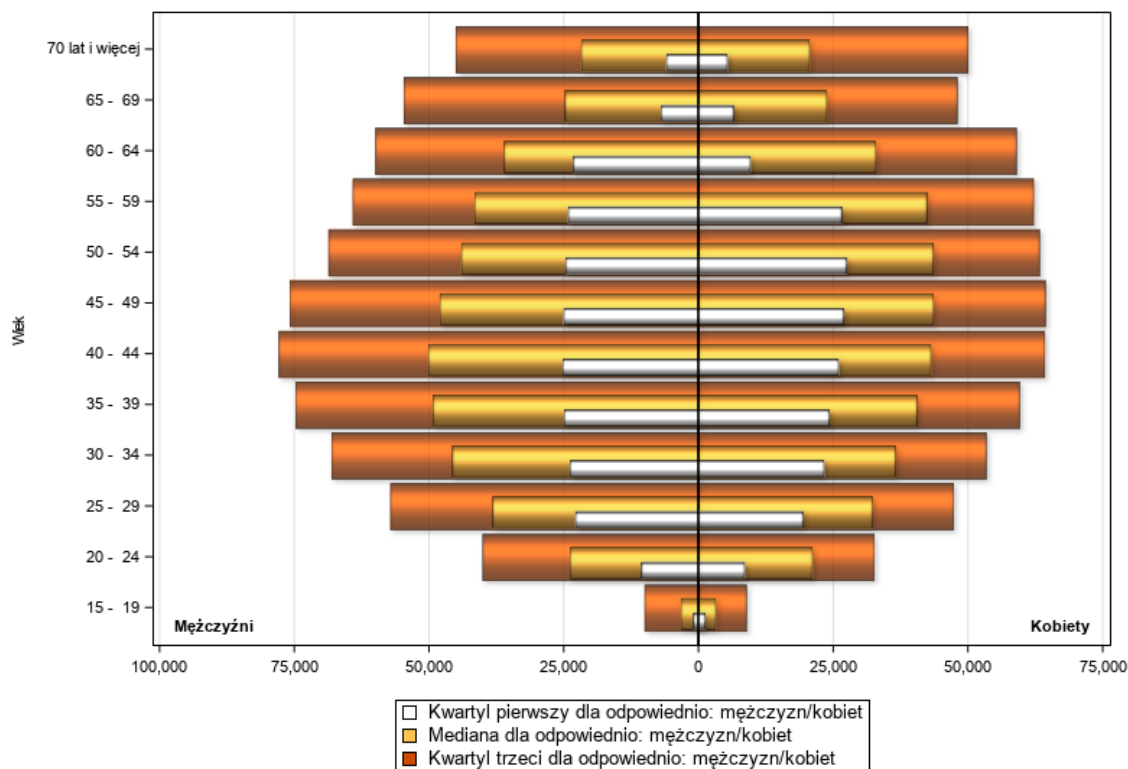
Dla ogółu miast w kraju w 2018 r. wskaźnik ten uzyskał wartość 12,79. Oznacza to, że osoba na granicy 10% populacji o najwyższych dochodach osiągnęła w 2018 r. dochody będące niemal 13-krotnością dochodów uzyskanych przez osobę znajdującą się na granicy 10% populacji o najniższych dochodach.

Miarą zróżnicowania dochodów są także wskaźnik S80/S20 i współczynnik Giniego. Wskaźniki te znajdują się na liście miar stosowanych przez Unię Europejską do monitorowania poziomu życia. Relacja S80/S20 jest stosunkiem sumy dochodów uzyskanych przez 20% osób o najwyższym poziomie dochodów do sumy dochodów uzyskanych przez 20% osób o najniższym poziomie dochodów. W 2018 r. w miastach ogółem w kraju wskaźnik ten wyniósł 14,30. Natomiast współczynnik Giniego (najczęściej wykorzystywana miara rozkładu nierówności o wartościach z przedziału od 0 do 1) - wyniósł 0,43.

Wyniki badania pozwoliły na ustalenie nierówności dochodowych osób zatrudnionych na podstawie umowy o pracę w podziale według płci i wieku podatników.

W subpopulacji mężczyzn mediana przychodów uzyskanych w 2018 r. wyniosła 40590,94 zł i była wyższa o 12436,68 zł od wartości dla kohorty kobiet oraz o 7386,46 zł od wartości dla kraju. O większych nierównościach przychodów między mężczyznami a kobietami w ujęciu bezwzględny świadczy także większy rozstęp międzykwartyłowy (IQR), który dla subpopulacji mężczyzn wyniósł 47871,15 zł, a dla kobiet 32668,47 zł. Należy przy tym pamiętać, że zróżnicowanie przychodów w tym ujęciu dotyczy połowy obszaru zmienności rozkładu, czyli obszaru między kwartylem pierwszym a trzecim, a nie całego rozkładu.

Rycina 12. Kwartyle dochodów (w złotych) według grup wieku i płci w 2018 r.



Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

W kontekście dochodów podatników ważna jest informacja o ich kształtowaniu się w aspekcie ekonomicznych grup wieku⁶.

Najwyższa wartość mediany wystąpiła w grupie podatników będących w wieku produkcyjnym niemobilnym i wyniosła 43158,40 zł. Była ona o 11,9% wyższa od analogicznej wartości dla ogółu podatników w miastach. Znacznie niższą wartość mediany odnotowano wśród osób będących w wieku produkcyjnym mobilnym - 37015,82 zł. W pozostałych dwóch ekonomicznych grupach wieku tj. poprodukcyjnym i przedprodukcyjnym, wyniosła odpowiednio: 30947,50 zł i 301,37 zł. Największe zróżnicowanie badanej cechy - mierzone rozstępem ćwiartkowym - wystąpiło w grupie osób będących w wieku poprodukcyjnym i wyniosło 47653,83 zł. Mniejszy rozstęp ćwiartkowy wystąpił wśród podatników w wieku produkcyjnym: niemobilnym i mobilnym (odpowiednio: 40275,88 zł i 36680,70 zł).

Relacja decyla dziewiątego i pierwszego rozkładu dochodów (wskaźnik P90/P10) osiągnęła najwyższą wartość w grupie podatników będących w wieku poprodukcyjnym (38,71), co wskazuje na znaczne zróżnicowanie dochodów tej grupy wiekowej.

O stosunku sumy dochodów uzyskanych przez 20% osób o najwyższym poziomie dochodów do sumy dochodów uzyskanych przez 20% osób o najniższym poziomie dochodów informuje relacja S80/S20. Wskaźniki dla poszczególnych grup podatników informują, że ich największa wartość wystąpiła wśród podatników w wieku poprodukcyjnym, gdzie suma dochodów najbogatszych jest po-

⁶ Wiek przedprodukcyjny - wiek, w którym ludność nie osiągnęła jeszcze zdolności do pracy, tj. grupa wieku 0 - 17 lat.

Wiek produkcyjny - wiek zdolności do pracy, tj. dla mężczyzn grupa wieku 18-64 lata, dla kobiet - 18-59 lat.

Wiek produkcyjny mobilny - grupa wieku produkcyjnego obejmująca ludność w wieku 18-44 lata.

Wiek produkcyjny niemobilny - grupa wieku produkcyjnego obejmująca ludność w wieku: mężczyźni - 45-64 lata, kobiety - 45-59 lat.

Wiek poprodukcyjny - wiek, w którym osoby zazwyczaj kończą pracę zawodową, tj. dla mężczyzn - 65 lat i więcej, dla kobiet - 60 lat i więcej.

nad czterdziestokrotnie (40,89) wyższa od sumy dochodów uzyskanych w 2018 r. przez relatywnie najuboższych. Najniższa wartość wskaźnika (9,25) dotyczy dochodów uzyskanych przez zatrudnionych w wieku produkcyjnym niemobilnym.

Współczynnik Giniego wskazuje, że wśród osób w wieku produkcyjnym niemobilnym wystąpiła najbardziej równomierna dystrybucja dochodów, a wskaźnik osiągnął wartość 0,40. Nieco wyższy wskaźnik (0,43) odnotowano dla dochodów osób będących w wieku produkcyjnym mobilnym. W pozostałych dwóch grupach (wiek przedprodukcyjny i poprodukcyjny) współczynnik Giniego nieznacznie przekroczył 0,5.

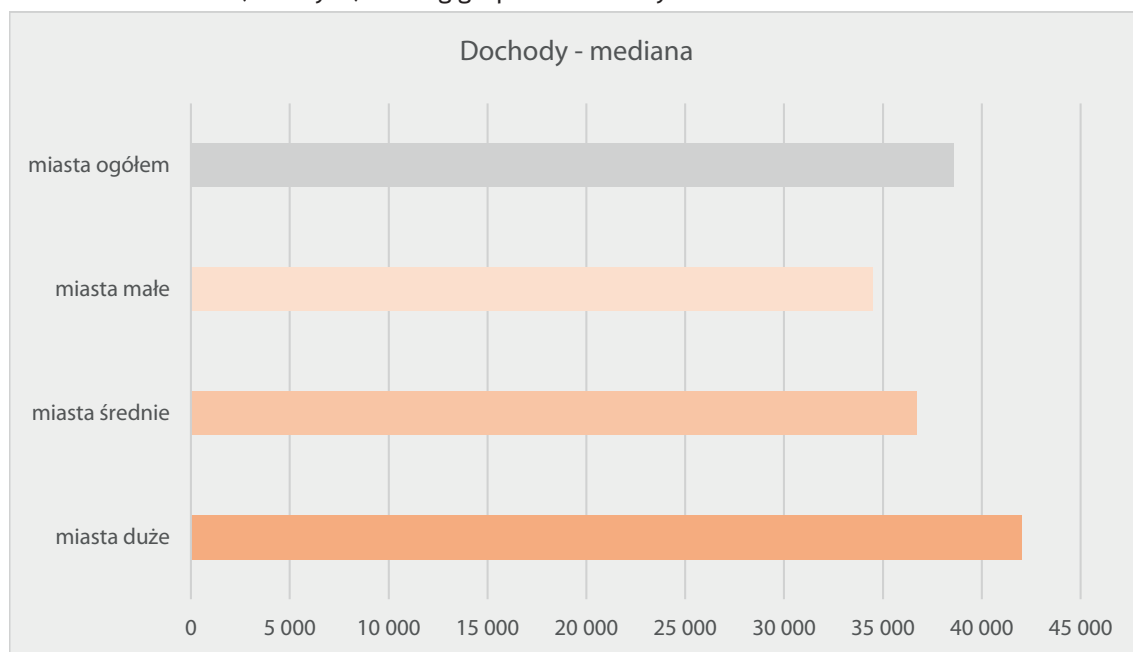
Analiza rozkładów dochodów w populacji podatników została poszerzona o aspekt typologiczny i przestrzenny miast.

• Aspekt typologiczny

Wśród mieszkańców miast obserwowana jest zależność między wysokością dochodów, a wielkością miasta. Wykorzystując typologię miast opartą wyłącznie na kryterium ludnościowym, poniżej została przedstawiona analiza nierówności dochodowych w podziale na trzy kategorie miast: małe, średnie i duże.

W całym 2018 r. w miastach ogółem w kraju połowa podatników objętych badaniem uzyskała dochody nie wyższe niż 38577,83 zł. Najbogatszymi, pod względem wielkości mediany, byli podatnicy mieszkający w miastach dużych, w których mediana dochodów wyniosła 42003,70 zł. Z kolei „najuboższymi” byli mieszkańcy miast małych (34483,74 zł). Dla populacji podatników zamieszkujących w miastach średnich mediana wyniosła 36691,97 zł, była więc niższa o 5311,73 zł od analogicznej wartości odnotowanej dla miast dużych i o 2208,23 zł wyższa od mediany dla miast małych.

Rycina 13. Mediana dochodów (w złotych) według grup wielkościowych miast w 2018 r.



Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Wartość rozstępu ćwiartkowego wskazuje, że największe zróżnicowanie dochodów ma miejsce w grupie miast dużych, a najmniejsze w miastach małych.

Identyczne tendencje uwidacznia analiza wskaźnika P90/P10. Relacja decyla dziewiątego do decyla pierwszego w rozkładzie dochodów była najwyższa w grupie miast dużych i wyniosła 13,35 (w miastach: średnich - 12,24, małych - 12,15).

Tablica 18. Wskaźniki zróżnicowania dochodów mieszkańców miast według grup wielkościowych miast

Miasta	P90/P10	S80/S20
Ogółem	12,79	14,30
Duże	13,35	15,42
Średnie	12,24	13,07
Małe	12,15	12,69

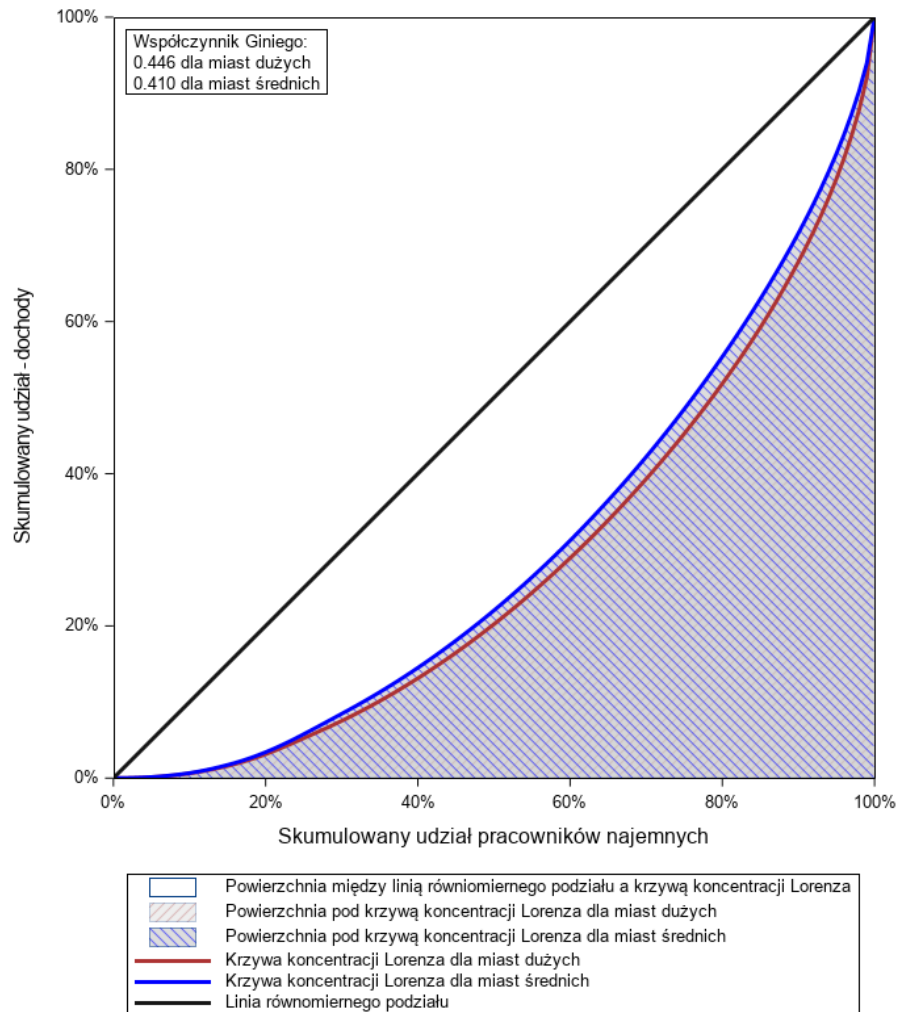
Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Kolejne ważne miary zróżnicowania dochodów określane przez wskaźnik S80/S20 i współczynnik Giniego osiągnęły również najwyższe wartości w populacji miast dużych i wynoszą odpowiednio: 15,42 i 0,45.

Stosunek sumy dochodów w najwyższej grupie kwintylowej do sumy dochodów w najniższej grupie kwintylowej w pozostałych dwóch grupach miast wyniósł w miastach: średnich - 13,07, a małych - 12,69.

Natomiast współczynnik Giniego w tych dwóch grupach miast kształtował się na zbliżonym poziomie i wyniósł w miastach: średnich - 0,41, a małych - 0,40.

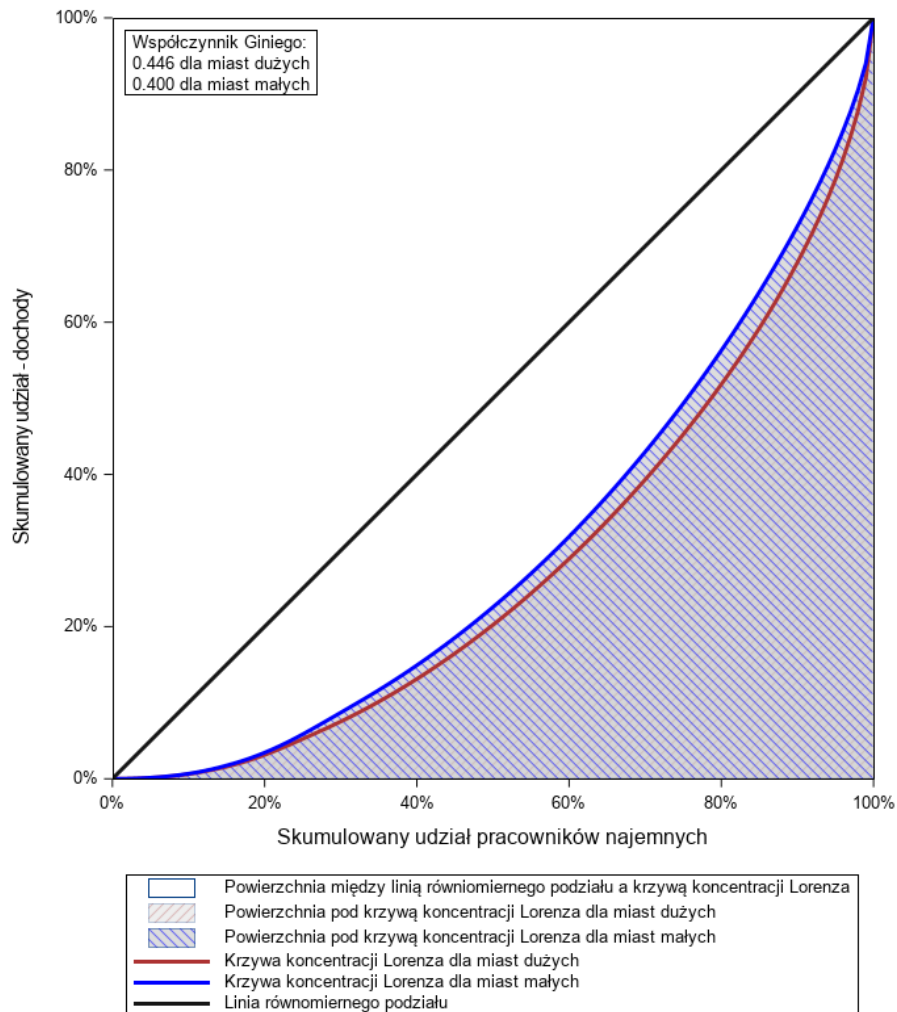
Rycina 14. Krzywa Lorentza – współczynnik Giniego dla miast dużych i średnich



Na podstawie danych zagregowanych - grup percentylowych

Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

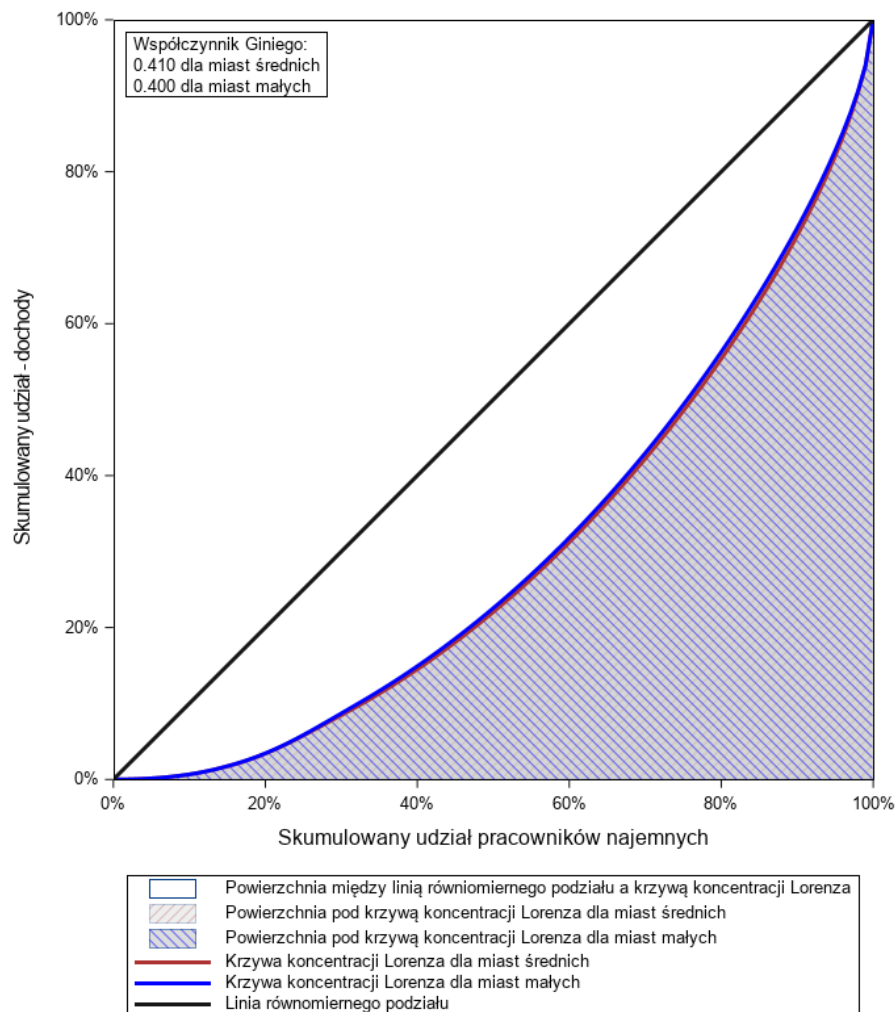
Rycina 15. Krzywa Lorenza – współczynnik Giniego dla miast dużych i małych



Na podstawie danych zagregowanych - grup percentylowych

Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Rycina 16. Krzywa Lorenza – współczynnik Giniego dla miast średnich i małych



Na podstawie danych zagregowanych - grup percytylowych

Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Analizując zbiorowość miast dużych, trudno nie wspomnieć o miastach będących stolicami województw, które jako najbardziej rozwinięte rynki pracy oraz centra usług (np. biznesowych czy administracyjnych) wpływają na sytuację dochodową mieszkańców tej kategorii miast.

Mediana dochodów uzyskanych w 2018 r. w grupie podatników zamieszkujących w tych miastach wyniosła 43838,68 zł i była wyższa o 1834,98 zł od analogicznej miary dla całej zbiorowości miast dużych i o 5260,85 zł od wartości mediany dla ogółu miast w kraju.

Rozstęp ćwiartkowy (IQR) wskazuje na większe zróżnicowanie dochodów w stolicach województw (45531,03 zł) w odniesieniu zarówno do ogółu miast dużych, jak i miast w kraju.

Miasta wojewódzkie charakteryzują się także największym zróżnicowaniem dochodów co obrazuje wskaźnik zróżnicowania kwintylowego (S80/S20), który wyniósł 15,64 (odpowiednio miasta duże - 15,42, a miasta ogółem w kraju - 14,30).

Współczynnik Giniego (0,45) wskazuje, że zróżnicowanie dochodów w miastach wojewódzkich było podobne do notowanego w całej grupie miast dużych, ale większe w odniesieniu do całej zbiorowości miast w kraju (0,43).

Dokumenty strategiczne takie jak np. „Strategia na rzecz Odpowiedzialnego Rozwoju do roku 2020 (z perspektywą do 2030 r.)”, czy „Krajowa Strategia Rozwoju Regionalnego 2030” zwracają uwagę na obszary o szczególnej koncentracji problemów rozwojowych. W tej grupie obszarów znajdują się miasta tracące swoje funkcje społeczno-gospodarcze. Miasta te cechuje regres w ich rozwoju powodowany różnymi przyczynami, co niewątpliwie wpływa na pogarszanie się warunków życia. Stąd monitorowanie sytuacji dochodowej mieszkańców tej grupy miast wydaje się niezwykle istotne.

Mediana dochodów uzyskanych w 2018 r. przez podatników zamieszkujących w tej grupie miast wyniosła 35099,27 zł i była o 4501,72 zł niższa od analogicznej miary wyliczonej dla pozostałych miast w kraju. Natomiast rozstęp kwartyłowy w miastach tracących funkcje społeczno-gospodarcze (32560,08 zł) jest mniejszy niż w pozostałych miastach kraju, co wskazuje na mniejsze zróżnicowanie dochodów.

Również relacja decyla dziewiątego do pierwszego w rozkładzie dochodów osiąga mniejszą wartość od analogicznej miary wyznaczonej dla pozostałych miast.

Pozostałe miary zróżnicowania i koncentracji dochodów mieszkańców miast tracących funkcje społeczno-gospodarcze również cechują się mniejszymi wartościami. Wskaźnik zróżnicowania kwintylowego S80/S20 był na poziomie 12,63, a współczynnik Giniego - 0,40 (dla miast pozostałych odpowiednio: 14,61 i 0,44).

- **Aspekt przestrzenny**

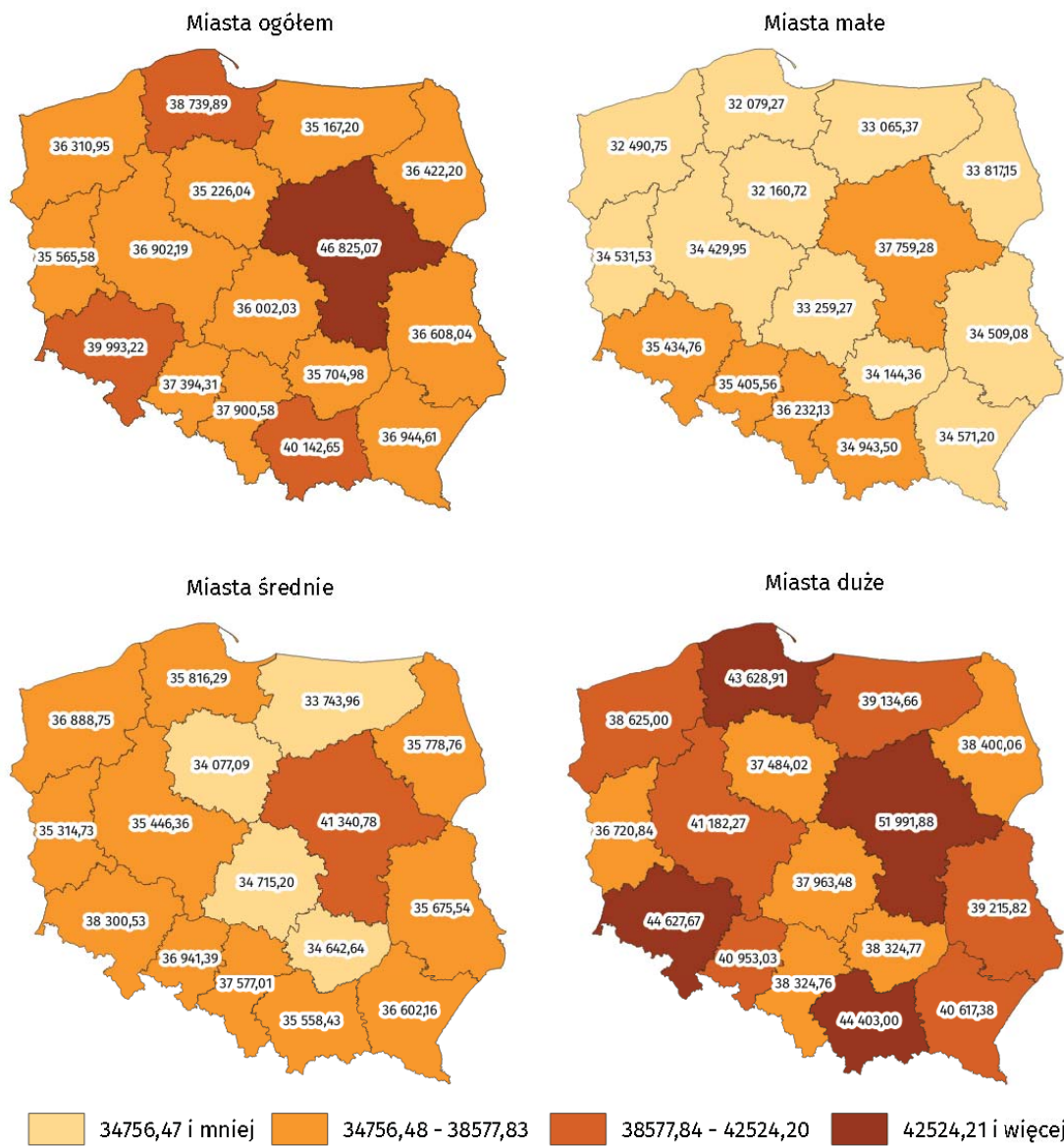
Obserwacja i analiza zróżnicowania dochodów w aspekcie przestrzennym (regionalnym) stanowi ważny element polityki każdego państwa. Przestrzenne zróżnicowanie dochodów podatników w przekroju województw uwzględniające również trzy grupy wielkościowe miast zaprezentowano poniżej za pomocą wcześniej wspomnianych miar.

W 2018 r. najbogatsi mieszkańcy miast (pod względem wartości mediany) zamieszkiwali w województwie mazowieckim (46825,07 zł), a najbiedniejsi w województwie warmińsko-mazurskim (35167,20 zł). Mediana powyżej 40 tys. złotych wystąpiła jeszcze tylko wśród mieszkańców miast województwa małopolskiego.

W zbiorowości województw rozpiętość mediany była największa w grupie miast dużych (15271,04 zł), a najmniejsza wśród miast małych (5680,01 zł). We wszystkich województwach mediana dochodów mieszkańców miast małych nie przekroczyła 40 tys. zł. Skrajne wartości tej miary zamykają się w przedziale od 32079,27 zł (woj. pomorskie) do 37759,28 zł (woj. mazowieckie). W grupie miast średnich skrajne wartości mediany wystąpiły w miastach województw: warmińsko-mazurskiego (33743,96 zł) i mazowieckiego (41,340,78 zł), a dużych w województwach: lubuskim - 36720,84 zł i mazowieckim - 51991,88 zł. W miastach dużych wartości mediany powyżej 40 tys. zł odnotowano również w miastach województw: dolnośląskiego, małopolskiego, opolskiego, podkarpackiego, pomorskiego i wielkopolskiego.

Terytorialne zróżnicowanie mediany dochodów w 2018 r. przedstawiają poniżej zaprezentowane kartogramy.

Mapa 4. Mediana dochodów (w złotych) według województw i grup wielkościowych miast w 2018 r.



Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Wykorzystując grupy decylowe dla kraju ogółem, porównano udział podatników i ich dochodów dla miast ogółem według województw. Wyniki tego porównania dla grupy dziesiątej (powyżej 89564,01 zł) przedstawia poniżej zaprezentowana tablica.

Tablica 19. Podatnicy i dochody w najwyższym przedziale dochodowym w 2018 r. według województw

Województwo	Udział	
	podatników	dochodów
	w %	
Dolnośląskie	11,54	33,24
Kujawsko-pomorskie	5,72	18,62
Lubelskie	7,19	22,03
Lubuskie	5,91	18,94
Łódzkie	7,87	25,08
Małopolskie	11,65	33,76
Mazowieckie	18,27	48,41
Opolskie	6,78	20,87
Podkarpackie	6,61	20,63
Podlaskie	5,62	17,68
Pomorskie	10,17	30,55
Śląskie	9,10	26,73
Świętokrzyskie	6,60	21,45
Warmińsko-mazurskie	5,34	17,09
Wielkopolskie	7,82	24,89
Zachodniopomorskie	6,57	20,13

Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Jak wynika z Tab. 19., 18% podatników znajdujących się w najwyższej grupie dochodowej w województwie mazowieckim generuje ponad 48% dochodów uzyskanych przez wszystkich podatników (zatrudnionych na podstawie umowy o pracę) zamieszkałych w miastach tego województwa. W województwach: małopolskim i dolnośląskim dochody niespełna 12% podatników generują odpowiednio: 33,76% i 33,24% dochodów ogółem.

Pozycyjna miara nierówności, którą jest relacja decyla dziewiątego do pierwszego w rozkładzie dochodów w poszczególnych województwach, zamyka się w przedziale od 8,80 (woj. podlaskie) do 14,53 (woj. dolnośląskie), co wskazuje na największe nierówności w dochodach podatników mieszkających w województwie dolnośląskim. Wysoką wartość tego parametru odnotowano także w województwie mazowieckim - 14,13.

Kwintylowe zróżnicowanie dochodów mieszkańców miast według województw mierzone wskaźnikiem S80/S20 zamyka się w przedziale od 9,75 (woj. podlaskie) do 17,09 (woj. mazowieckie). Wysokie wartości tej relacji wystąpiły również w województwach: dolnośląskim, pomorskim i małopolskim (odpowiednio 15,35; 14,84 i 14,36).

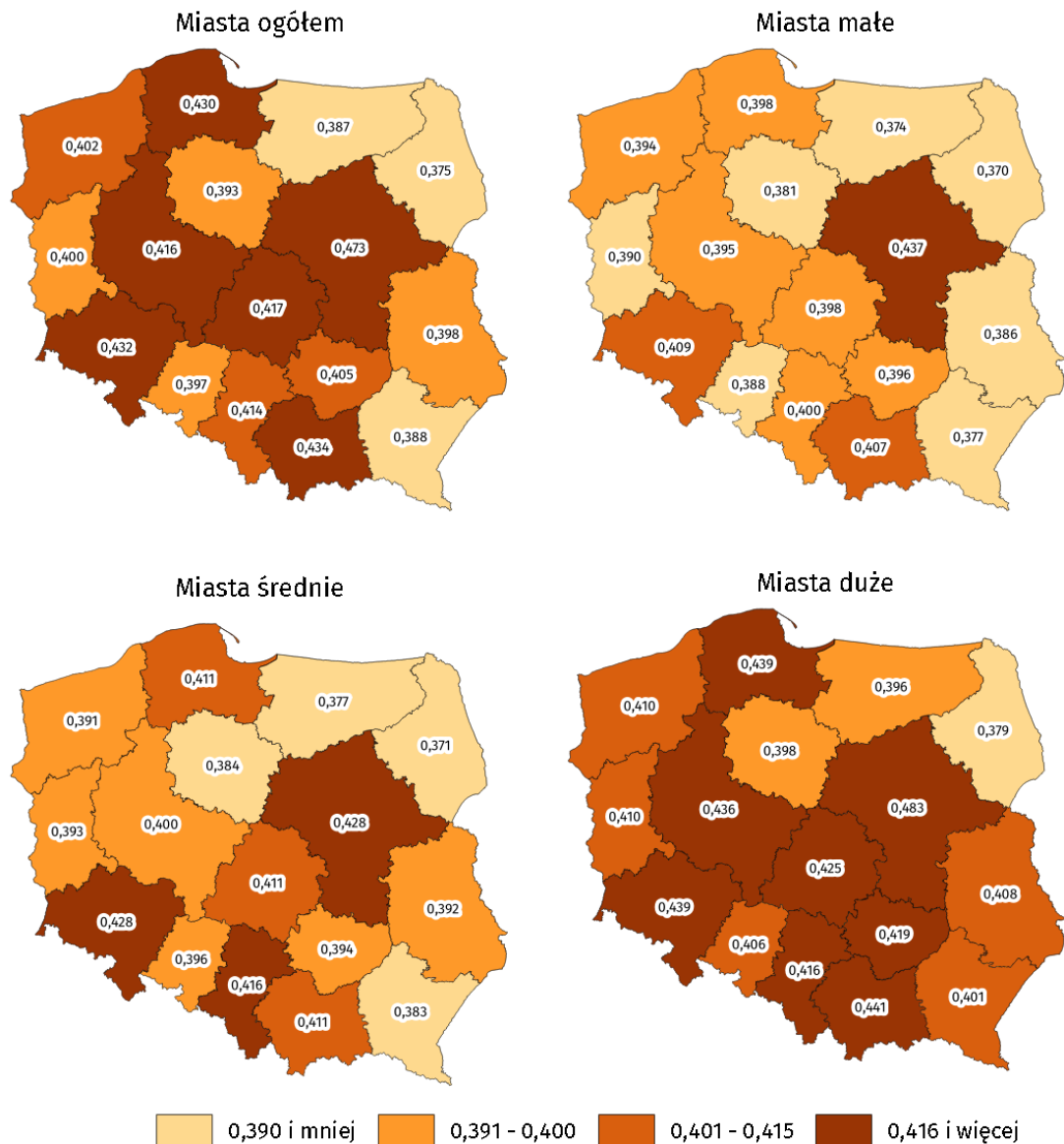
Tablica 20. Wybrane miary nierówności dochodów w 2018 r. według województw

Województwo	P90/P10	S80/S20
Dolnośląskie	14,53	15,35
Kujawsko-pomorskie	11,62	12,09
Lubelskie	11,17	11,95
Lubuskie	12,96	13,36
Łódzkie	12,50	13,52
Małopolskie	12,73	14,36
Mazowieckie	14,13	17,09
Opolskie	12,32	12,72
Podkarpackie	9,45	10,46
Podlaskie	8,80	9,75
Pomorskie	13,59	14,84
Śląskie	12,73	13,57
Świętokrzyskie	11,83	12,71
Warmińsko-mazurskie	11,06	11,53
Wielkopolskie	12,81	13,76
Zachodniopomorskie	13,77	13,95

Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Zróżnicowanie dochodów mierzone współczynnikiem Giniego w przekroju województw zamyka się w przedziale od 0,38 w woj. podlaskim do 0,47 w woj. mazowieckim. Przestrzenne zróżnicowanie współczynnika Giniego obrazuje Mapa 5.

Mapa 5. Współczynnik Giniego w miastach ogółem oraz grupach wielkościowych miast w 2018 r. według województw



Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

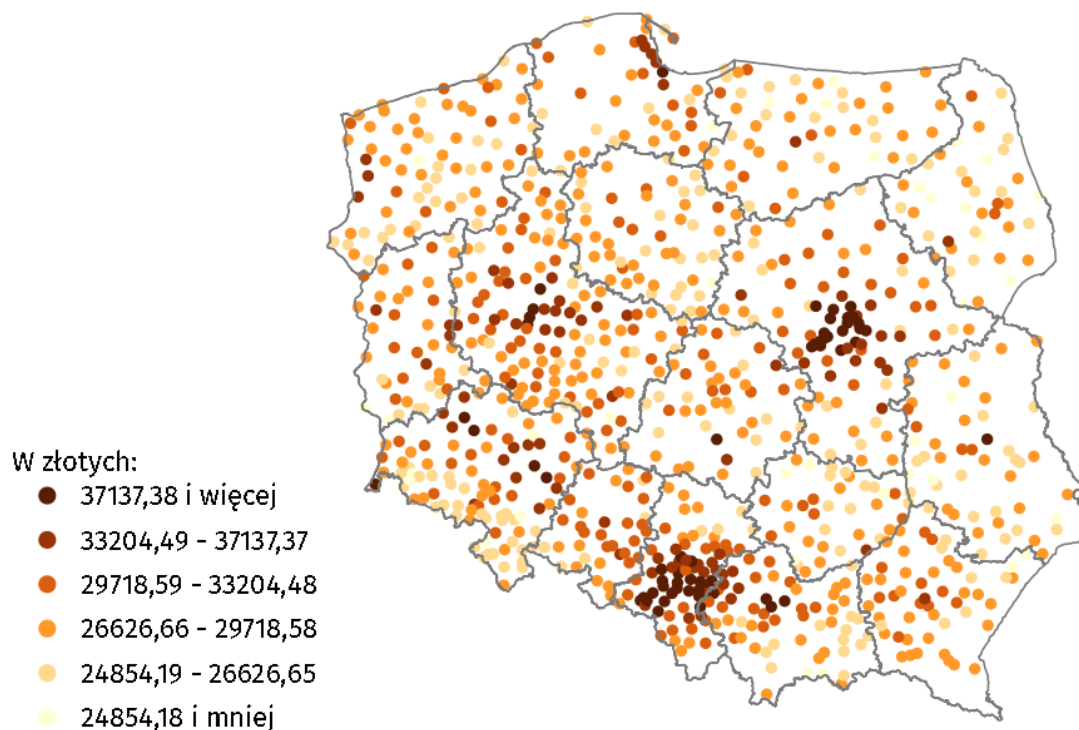
W przekroju województw największe zróżnicowanie tej miary wystąpiło w zbiorowości miast dużych (od 0,38 – woj. podlaskie do 0,48 – woj. mazowieckie).

Przychody i dochody mieszkańców poszczególnych miast

Ustalenia wynikające z kwerendy literatury przedmiotu wskazują na wagę i rosnące znaczenie problemu, którym są nierówności przestrzenne w dochodach osobistych na poziomie lokalnym. W kolejnej części analizy podjęto próbę charakterystyki poziomu i zróżnicowania przychodów podatników/dochodów pracowników najemnych w ujęciu poszczególnych miast.

Mediana przychodów mieszkańców miast osiągnęła wartości najwyższe głównie w miastach województwa śląskiego (22 miasta) i mazowieckiego (20 miast).

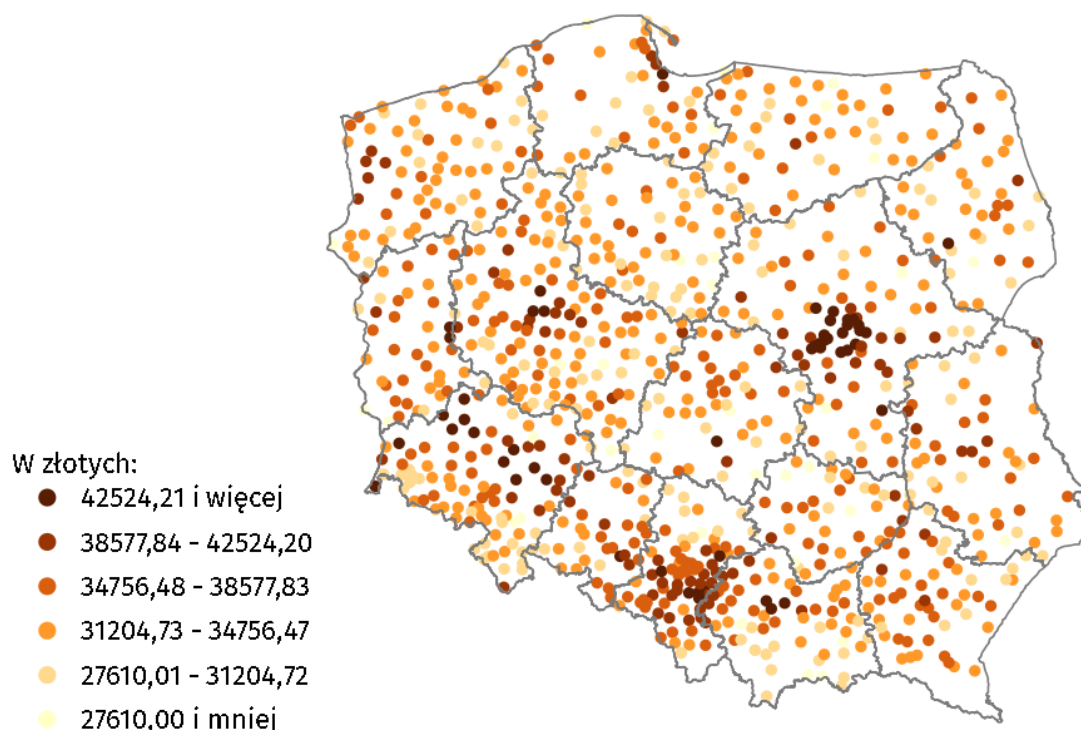
Mapa 6. Mediana przychodów (w złotych) ludności w 2018 r. według miast



Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

W przypadku dochodów mieszkańców miast zauważa się nieco odmienną koncentrację miast o najwyższych wartościach mediany. Skupiają się one głównie w województwie mazowieckim (21 miast) oraz dolnośląskim (12 miast). Ponadto wysokie wartości tej miary wystąpiły również wśród mieszkańców miast położonych wzdłuż autostrady na południu kraju.

Mapa 7. Mediana dochodów (w złotych) ludności w 2018 r. według miast



Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Największą wartość mediany dochodów mieli mieszkańcy Warszawy, a najmniejszą mieszkańcy Cedyni leżącej w województwie zachodniopomorskim. Na uwagę zasługuje fakt, że aż cztery z pięciu miast o największych wartościach mediany leżą w województwie mazowieckim.

Tablica 21. Mediana dochodów mieszkańców miast

Wskaźnik	Miasta o wartościach wskaźnika					
	największych			najmniejszych		
	nazwa	wartość	województwo	nazwa	wartość	województwo
Mediana (zł)	Warszawa	55 110,27	mazowieckie	Cedynia	23 865,00	zachodniopomorskie
	Podkowa Leśna	52 525,22	mazowieckie	Łęknica	23 865,25	lubuskie
	Bogatynia	49 721,99	dolnośląskie	Dzierżgoń	24 231,42	pomorskie
	Piaseczno	49 672,55	mazowieckie	Krynica Morska	24 334,34	pomorskie
	Ożarów Mazowiecki	49 655,92	mazowieckie	Przedbórz	24 527,00	łódzkie

Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Wskaźnik zróżnicowania dochodów P90/P10 zamykał się w przedziale od 26,37 do 6,67. Najwyższą wartość omawianej relacji w całej zbiorowości miast odnotowano w mieście Biały Bór, a najniższą w mieście Białzowa. Dla Białego Boru oznacza to, że dziewiąty decyl rozkładu dochodów był ponad 26 razy większy od decyla pierwszego w tym rozkładzie. W przypadku miasta Białzowa relacja pomiędzy tymi dwoma poziomami dochodów wynosiła niecałe 7 razy.

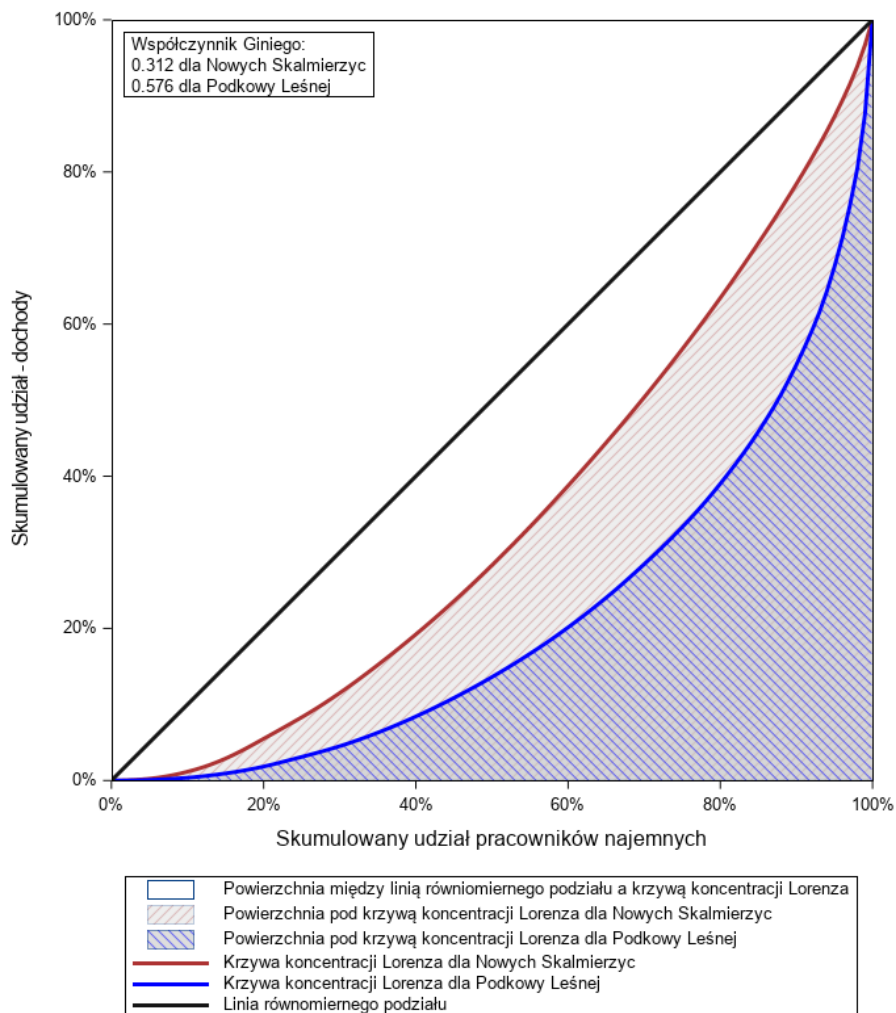
Tablica 22. Wskaźniki zróżnicowania dochodów mieszkańców miast

Wskaźnik	Miasta o wartościach wskaźnika					
	największych			najmniejszych		
	nazwa	wartość	województwo	nazwa	wartość	województwo
P90/P10	Biały Bór	26,37	zachodnio-pomorskie	Białzowa	6,21	podkarpackie
	Podkowa Leśna	25,41	mazowieckie	Piława	6,23	mazowieckie
	Dzierżgoń	24,48	pomorskie	Nowe Skalmierzyce	6,36	wielkopolskie
	Paczków	23,75	opolskie	Choroszcz	6,55	podlaskie
	Krynica Morska	23,55	pomorskie	Czyżew	6,67	podlaskie
S80/S20	Podkowa Leśna	33,09	mazowieckie	Nowe Skalmierzyce	6,92	wielkopolskie
	Krynica Morska	25,56	pomorskie	Białzowa	6,96	podkarpackie
	Konstancin-Jeziorna	25,33	mazowieckie	Górzno	7,19	kujawsko-pomorskie
	Puszczykowo	24,15	wielkopolskie	Piława	7,26	mazowieckie
	Zakopane	24,14	małopolskie	Hajnówka	7,68	podlaskie
Wsp. Giniego	Podkowa Leśna	0,58	mazowieckie	Nowe Skalmierzyce	0,31	wielkopolskie
	Konstancin-Jeziorna	0,55	mazowieckie	Zbąszynek	0,32	lubuskie
	Puszczykowo	0,53	wielkopolskie	Lubawa	0,33	warmińsko-mazurskie
	Łomianki	0,53	mazowieckie	Górzno	0,33	kujawsko-pomorskie
	Józefów	0,52	mazowieckie	Hajnówka	0,34	podlaskie

Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Kolejna miara zróżnicowania dochodów mieszkańców miast, jaką jest wskaźnik zróżnicowania kwintylowego S80/S20, osiągnęła największą wartość w mieście Podkowa Leśna (33,09), a najmniejszą w mieście Nowe Skalmierzyce (6,92). Dla Podkowy Leśnej wartość ta oznacza, że łączne dochody 20% mieszkańców o najwyższych dochodach były ponad 33 razy większe od łącznych dochodów 20% mieszkańców o najniższych dochodach. W przypadku Nowych Skalmierzyc relacja ta wyniosła niecałe 7 razy. Również dla dochodów mieszkańców tych miast odnotowano skrajne wartości współczynnika Giniego, co obrazuje niżej przedstawiona krzywa Lorenza. Podobnie jak w przypadku mediany, aż cztery z pięciu miast o największych wartościach tego miernika leżą w województwie mazowieckim.

Rycina 17. Krzywa Lorenza – współczynnik Giniego dla Nowych Skalmierzyc i Podkowy Leśnej



Na podstawie grup percentylowych

Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Z przeprowadzonej analizy wynika zasadniczy wniosek, że w Polsce występuje duże zróżnicowanie zarówno przychodów, jak i dochodów mieszkańców poszczególnych jednostek osadniczych (miast).

• **Dochody mieszkańców miast wojewódzkich**

Wprowadzeniem do analizy zróżnicowania wewnątrzmięjskiego na przykładzie miasta Poznania jest informacja o nierównościach dochodowych w poszczególnych miastach wojewódzkich w postaci swobodnego rankingu i krótkiego komentarza analitycznego.

Tablica 23. Wskaźniki zróżnicowania dochodów mieszkańców miast

Miasto (w kolejności pod względem liczby ludno- ści)	Mediana		P90/P10		S80/S20		Wsp. Giniego	
	Wartość w złotych	poz.	wartość	poz.	wartość	poz.	wartość	poz.
Warszawa	55 110,27	1.	15,16	1.	18,55	1.	0,48	1.
Kraków	45 506,27	3.	12,96	7.	14,86	5.	0,44	2.
Łódź	37 963,48	15.	13,12	5.	14,33	7.	0,43	7.
Wrocław	46 434,38	2.	13,99	4.	15,40	3.	0,44	4.
Poznań	42 621,00	5.	14,09	3.	15,66	2.	0,44	3.
Gdańsk	44 241,75	4.	13,02	6.	14,64	6.	0,44	5.
Szczecin	39 051,52	11.	14,59	2.	14,99	4.	0,41	10.
Bydgoszcz	38 349,21	13.	10,98	14.	11,62	15.	0,39	16.
Lublin	39 215,82	10.	10,94	15.	12,14	14.	0,41	11.
Białystok	38 400,06	12.	8,35	18.	9,49	18.	0,38	18.
Katowice	41 027,36	7.	12,35	10.	14,12	8.	0,43	6.
Toruń	37 384,74	17.	11,49	13.	12,61	12.	0,41	12.
Kielce	38 324,77	14.	12,03	11.	13,33	10.	0,42	8.
Rzeszów	40 617,38	9.	9,68	16.	10,91	16.	0,40	14.
Olsztyn	41 158,81	6.	9,26	17.	10,18	17.	0,38	17.
Zielona Góra	37 891,61	16.	12,87	8.	13,81	9.	0,42	9.
Opole	40 953,03	8.	11,52	12.	12,53	13.	0,41	13.
Gorzów Wlkp.	35 500,30	18.	12,80	9.	13,22	11.	0,40	15.

Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Wyniki pracy badawczej wskazują na bardzo znaczącą wagę miast wojewódzkich w rozkładzie dochodów mieszkańców wszystkich miast. Zarówno mediana jak i wszystkie miary nierówności rozkładu dochodów ich mieszkańców są wyższe od analogicznych mierników dla wszystkich miast.

Tablica 24. Wskaźniki zróżnicowania dochodów mieszkańców miast

Miasta	Mediana	P90/P10	S80/S20	Wsp. Giniego
Wojewódzkie	43838,68	13,50	15,64	0,45
Ogółem	38577,83	12,79	14,30	0,43

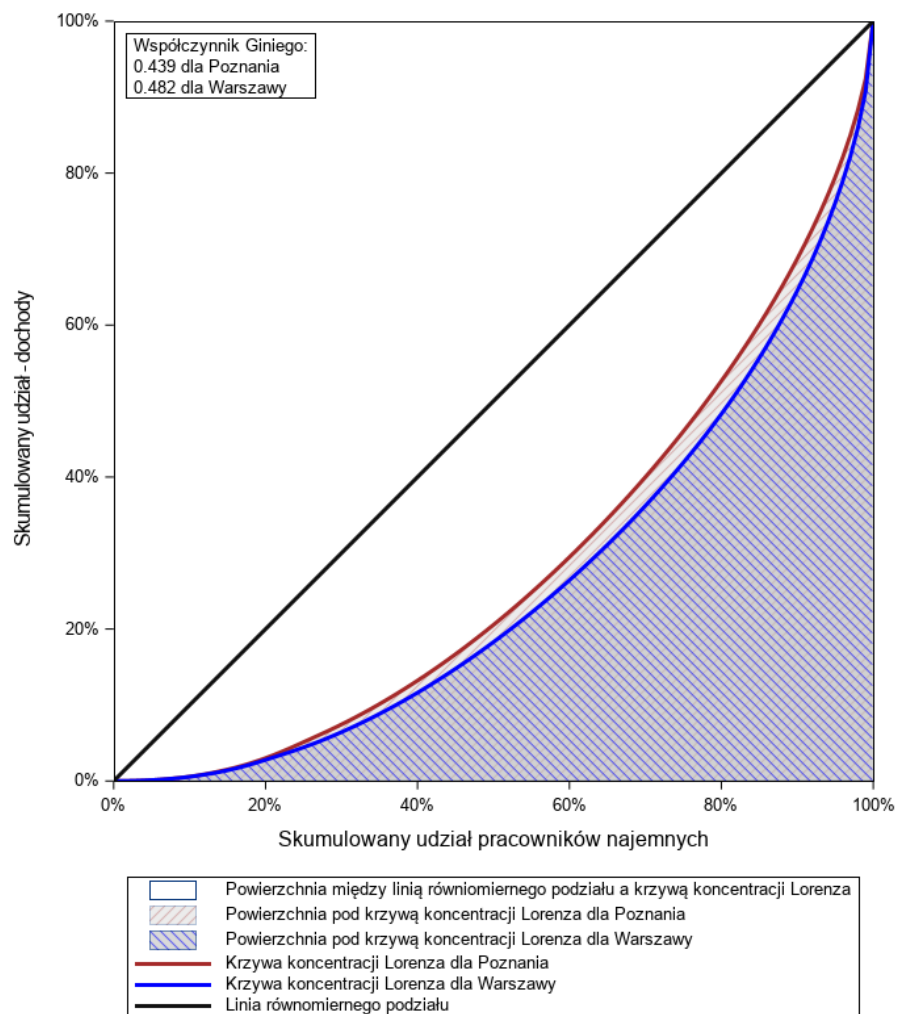
Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Wśród miast wojewódzkich najmniejsze wartości mediany dochodów (poniżej 38 tys.) mieli mieszkańcy Gorzowa Wielkopolskiego, Torunia, Zielonej Góry oraz Łodzi. Natomiast największą medianą dochodów charakteryzowali się mieszkańcy Warszawy. Poznań w tym rankingu uplasował się na 5. miejscu. Mediana dochodów mieszkańców zaledwie czterech miast wojewódzkich, tj. Warszawy, Wrocławia, Krakowa oraz Gdańska była większa od mediany dochodów mieszkańców wszystkich miast wojewódzkich (43838,68 zł).

Analizując zróżnicowanie dochodów mieszkańców miast wojewódzkich za pomocą trzech miar: P90/P10, S80/S20 oraz współczynnika Giniego, zauważa się, że najbardziej zróżnicowane dochody mieli mieszkańcy stolicy. Wartości każdej z tych trzech miar były najwyższe spośród wszystkich miast wojewódzkich. Najmniej natomiast były zróżnicowane dochody mieszkańców Białegostoku, gdzie każda z trzech analizowanych miar osiągnęła wartości najniższe. Warto również podkreślić, że trzy miasta: Warszawa, Olsztyn i Białystok uplasowały się na tych samych pozycjach (odpowiednio 1., 17. i 18.) dla wszystkich trzech analizowanych wskaźników zróżnicowania rozkładu dochodów ich mieszkańców.

W poszczególnych miastach wojewódzkich współczynnik Giniego oscylował w granicach od 0,38 do 0,48, co obrazują niżej przedstawione krzywe Lorenza. Wartość współczynnika Giniego dla Poznania była niższa tylko od analogicznej wartości dla Warszawy i Krakowa.

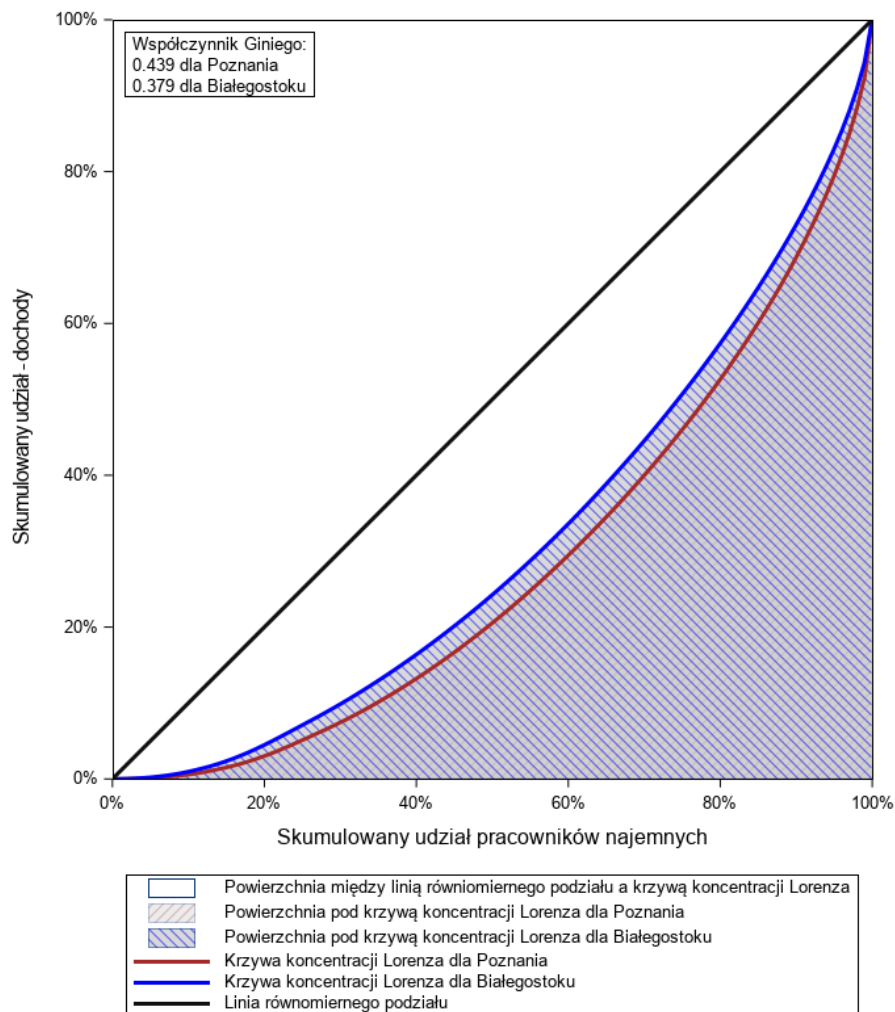
Rycina 18. Krzywa Lorenza – współczynnik Giniego dla Poznania i Warszawy



Na podstawie grup percentylowych

Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Rycina 19. Krzywa Lorenza – współczynnik Giniego dla Poznania i Białegostoku



Na podstawie grup percentylowych

Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

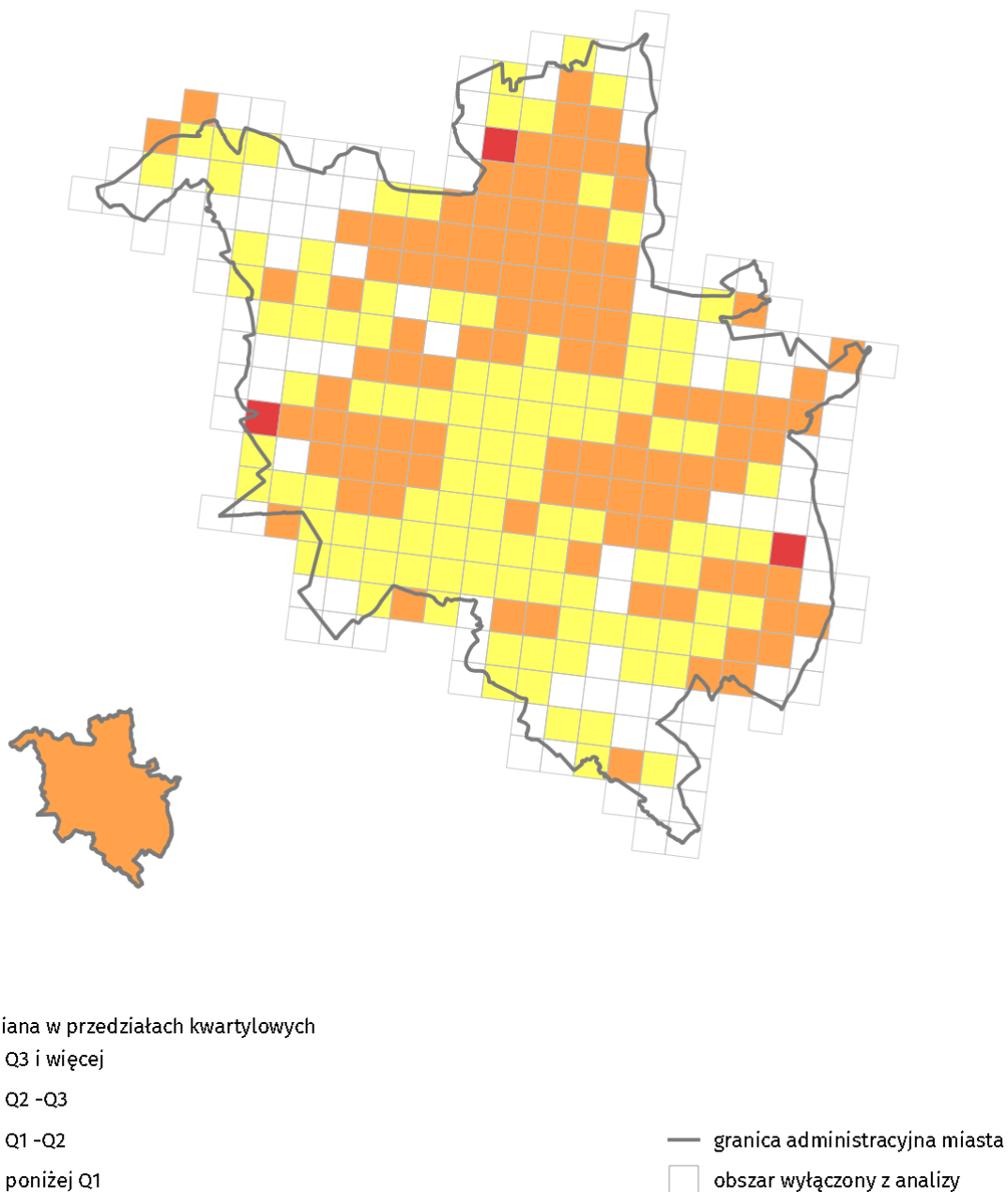
- **Wewnętrzniejskie zróżnicowanie dochodów mieszkańców miasta Poznania**

W niniejszej pracy, wychodząc naprzeciw zapotrzebowaniu odbiorców dokonano próby przedstawienia wybranych mierników charakteryzujących poziom oraz nierówności dochodowe w miastach wojewódzkich w układzie siatki kwadratów o boku 1000 metrów. Taka prezentacja danych dla miast choć w części spełni oczekiwania zgłaszane przez użytkowników i będzie pomocna w prowadzeniu działań w ramach różnych programów lokalnych.

Dla każdego z miast wojewódzkich wykonano i załączono do Raportu po 8 kartogramów obrazujących wewnętrzniejskie zróżnicowanie dochodów mieszkańców danego miasta. W Raporcie przedstawione zostały wyniki analizy na przykładzie miasta Poznania.

Do pomiaru poziomu i zróżnicowania dochodów ludności wykorzystuje się wiele miar. W pracy badawczej statystycznego opisu obszarów wewnątrzmijskich dokonano na podstawie czterech miar: mediany dochodu, relacji decyla dziewiątego do pierwszego (P90/P10), wskaźnika zróżnicowania kwintylowego (S80/S20) oraz współczynnika Giniego.

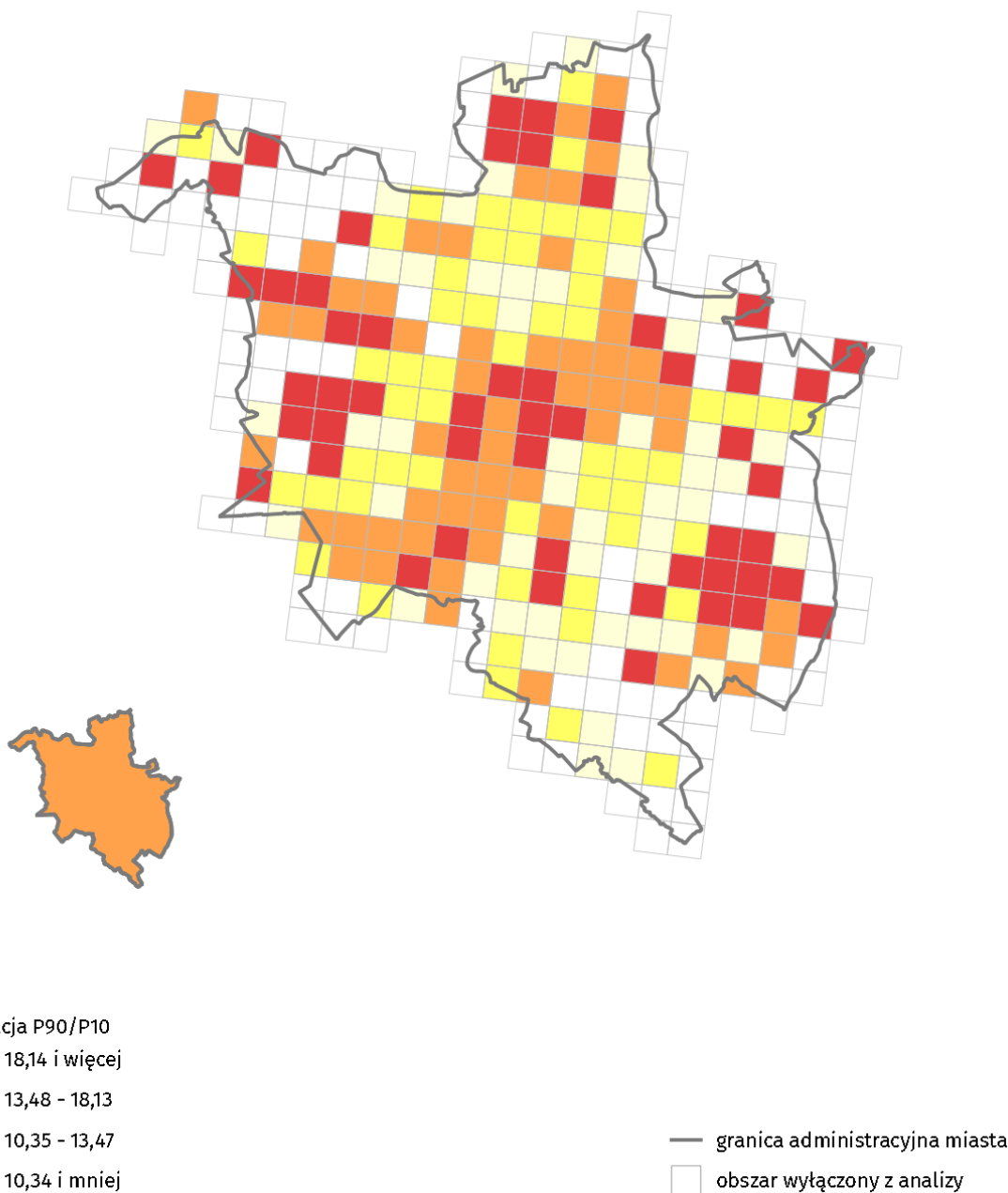
Mapa 8. Mediana dochodów – Poznań



Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

W północnej części Poznania zauważa się największe skupisko kwadratów o wartościach mediany powyżej drugiego kwartyla (Q2) dla całego miasta. Natomiast najwięcej kwadratów o wartościach mediany poniżej wspomnianego kwartyla znajduje się w centralnej i południowej części miasta. Jedynie w trzech oczkach o boku 1 km mediana dochodów mieszkańców jest wyższa niż trzeci kwartyl (Q3) rozkładu dochodów mieszkańców Poznania. Równocześnie w żadnym z kwadratów nie zauważa się mediany niższej niż pierwszy kwartyl (Q1) rozkładu dochodów dla całego miasta.

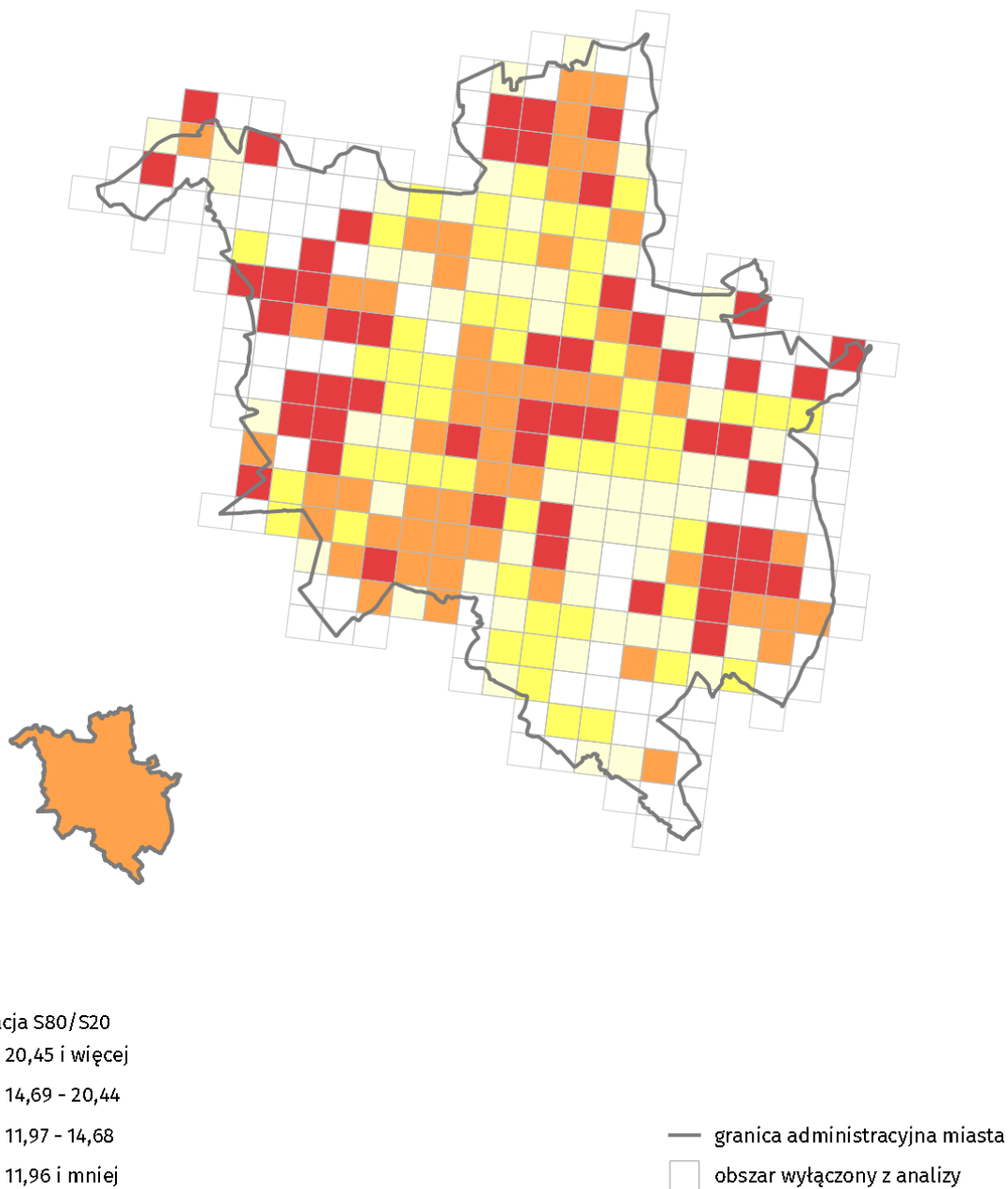
Mapa 9. Relacja decyla dziewiątego do decyla pierwszego (P90/P10) – Poznań



Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

W przestrzeni miasta Poznania zauważa się pasmo kwadratów z południowego zachodu po północny wschód o stosunkowo wysokich wartościach wskaźnika P90/P10. Po obu stronach tego pasma występują kwadraty o niższych wartościach wskaźnika. Natomiast najbardziej zróżnicowane dochody mają podatnicy mieszkający na obrzeżach Poznania (na północy, zachodzie i południowym wschodzie miasta).

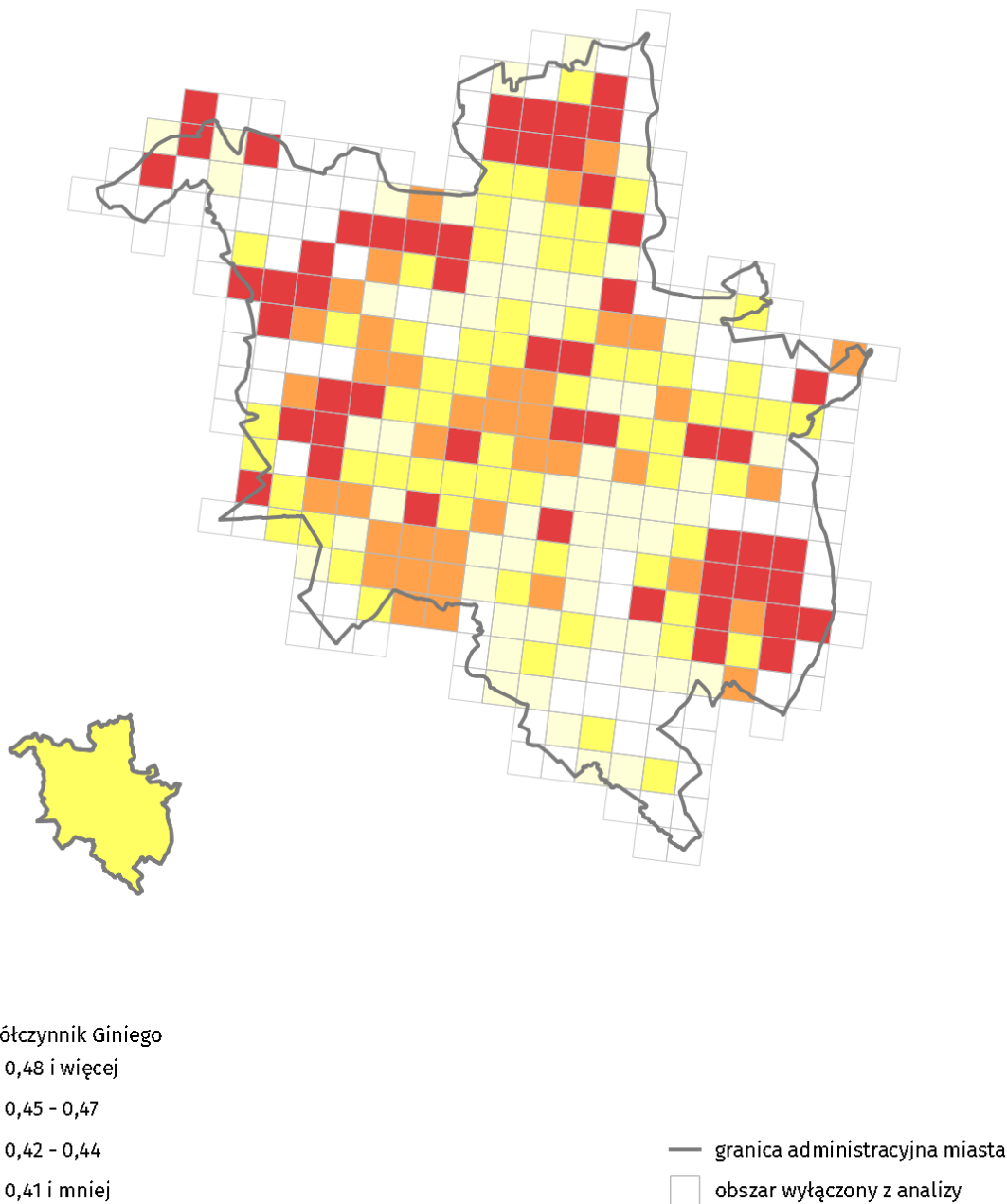
Mapa 10. Wskaźnik zróżnicowania kwintylogowego (S80/S20) – Poznań



Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Na powyższym kartogramie wewnątrzmięskie zróżnicowanie dochodów zaprezentowano za pomocą wskaźnika S80/S20. Stosunek ten wskazuje na podobne tendencje w przestrzennych nierównościach rozkładu dochodów mieszkańców Poznania jak relacja P90/P10.

Mapa 11. Współczynnik Giniego – Poznań



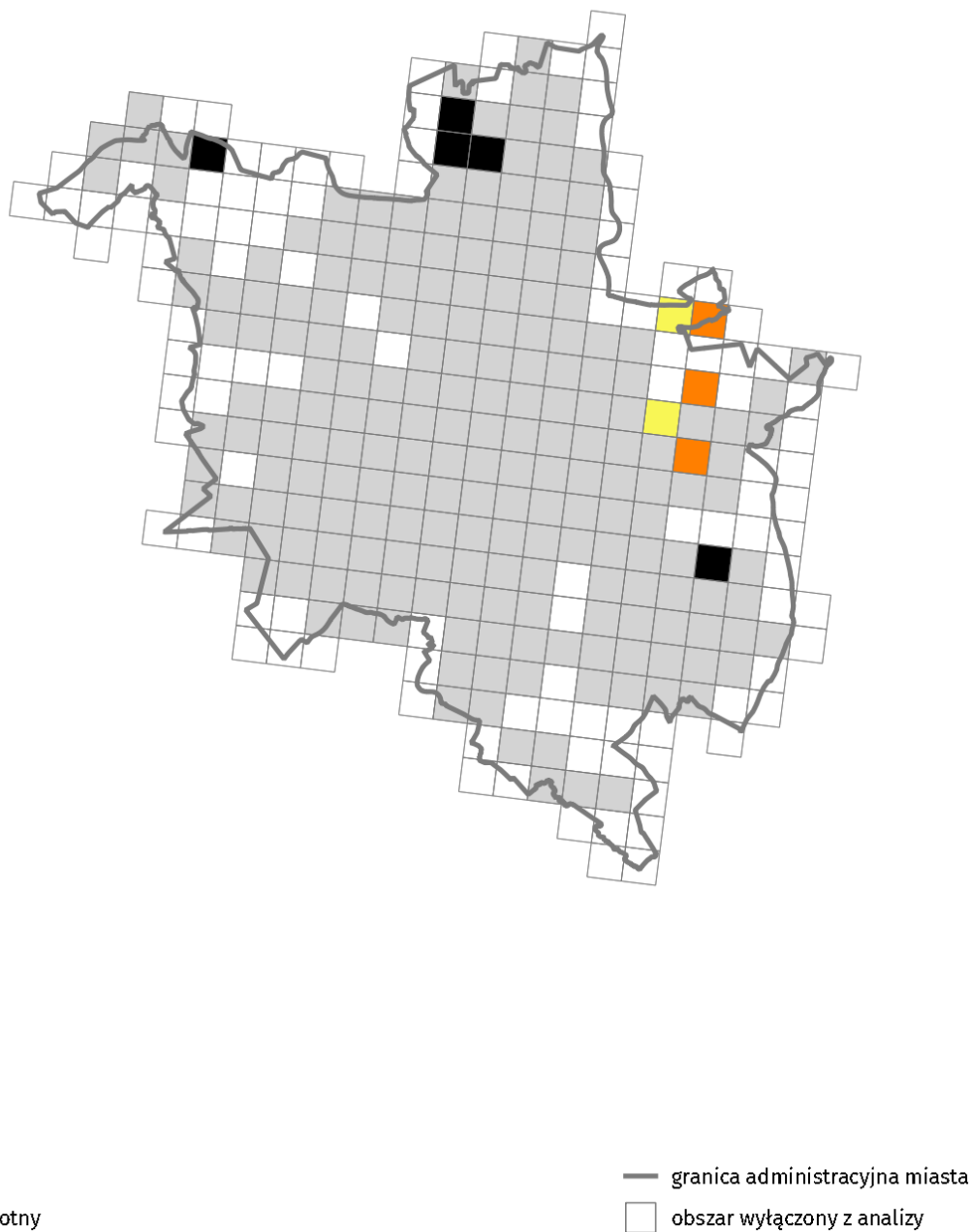
Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Najbardziej zróżnicowane pod względem dochodów mieszkańców są tereny zlokalizowane w północnej, północno-zachodniej i południowo-wschodniej części miasta. Nieco mniejszymi wartościami wskaźnika charakteryzują się kwadraty położone w południowo-zachodniej i centralnej części Poznania.

Do analizy koncentracji nierówności w rozkładzie dochodów mieszkańców Poznania wykorzystano statystykę lokalną Morana I^7 opracowaną dla trzech wspomnianych wyżej miar (oprócz mediany). Dzięki niej możliwa była identyfikacja efektów przestrzennych w postaci klastrow o niskich lub wysokich wartościach badanej zmiennej. Otrzymano dwa typy klastrow: High-High (HH) jednostek o wysokich wartościach badanej cechy, otoczonych przez jednostki o również wysokich wartościach, Low-Low (LL) jednostek o niskich wartościach otoczonych przez podobnych sąsiadów. Wyznaczono również dwa typy obserwacji odstających: Low-High (LH) i High-Low (HL). Końcowym wynikiem analizy są kartogramy klastrow i obserwacji odstających – Mapy 12, 13 i 14.

⁷ Suhecki (2010)

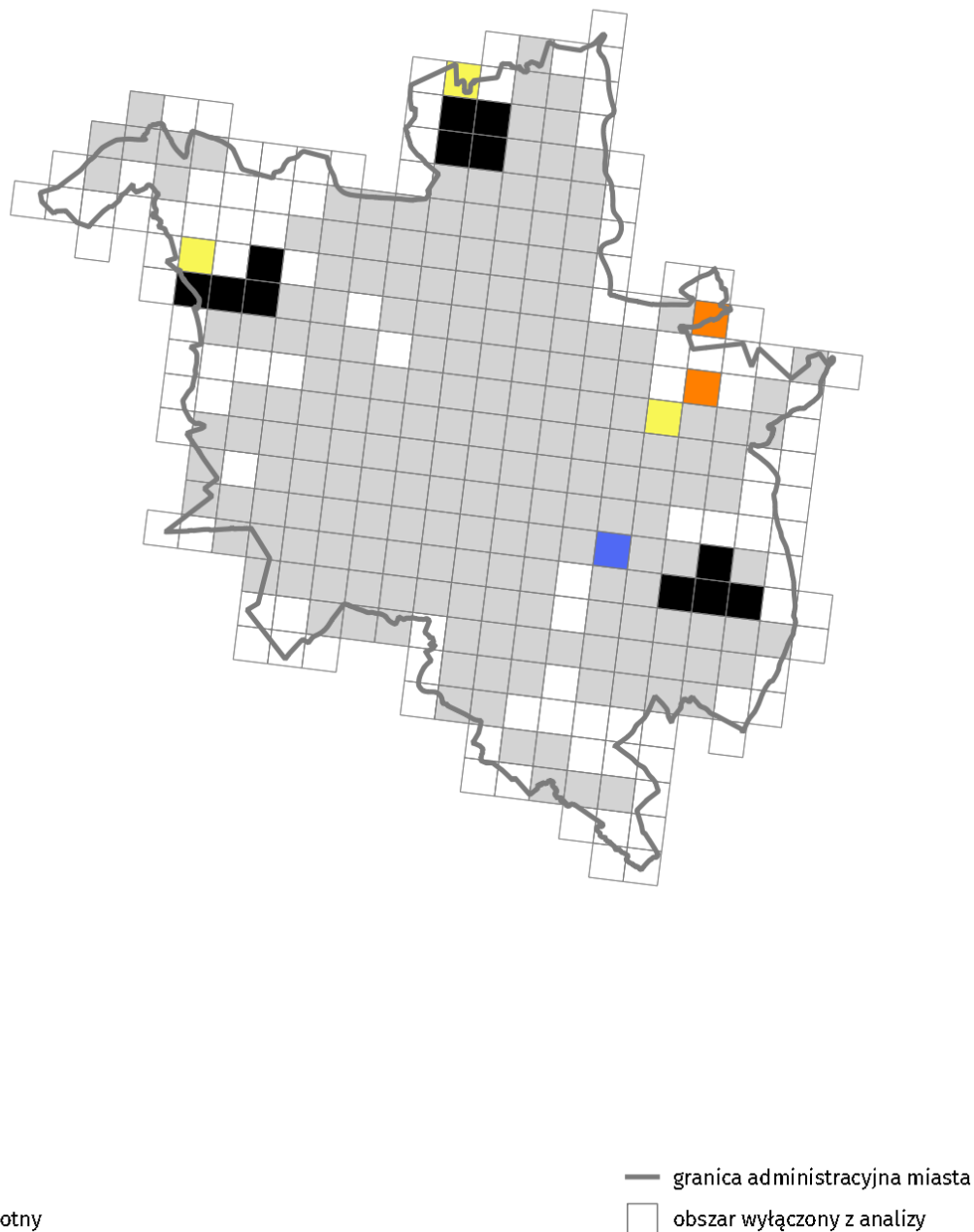
Mapa 12. Relacja decyla dziewiątego do decyla pierwszego (P90/P10) – LISA⁸ – Poznań



Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

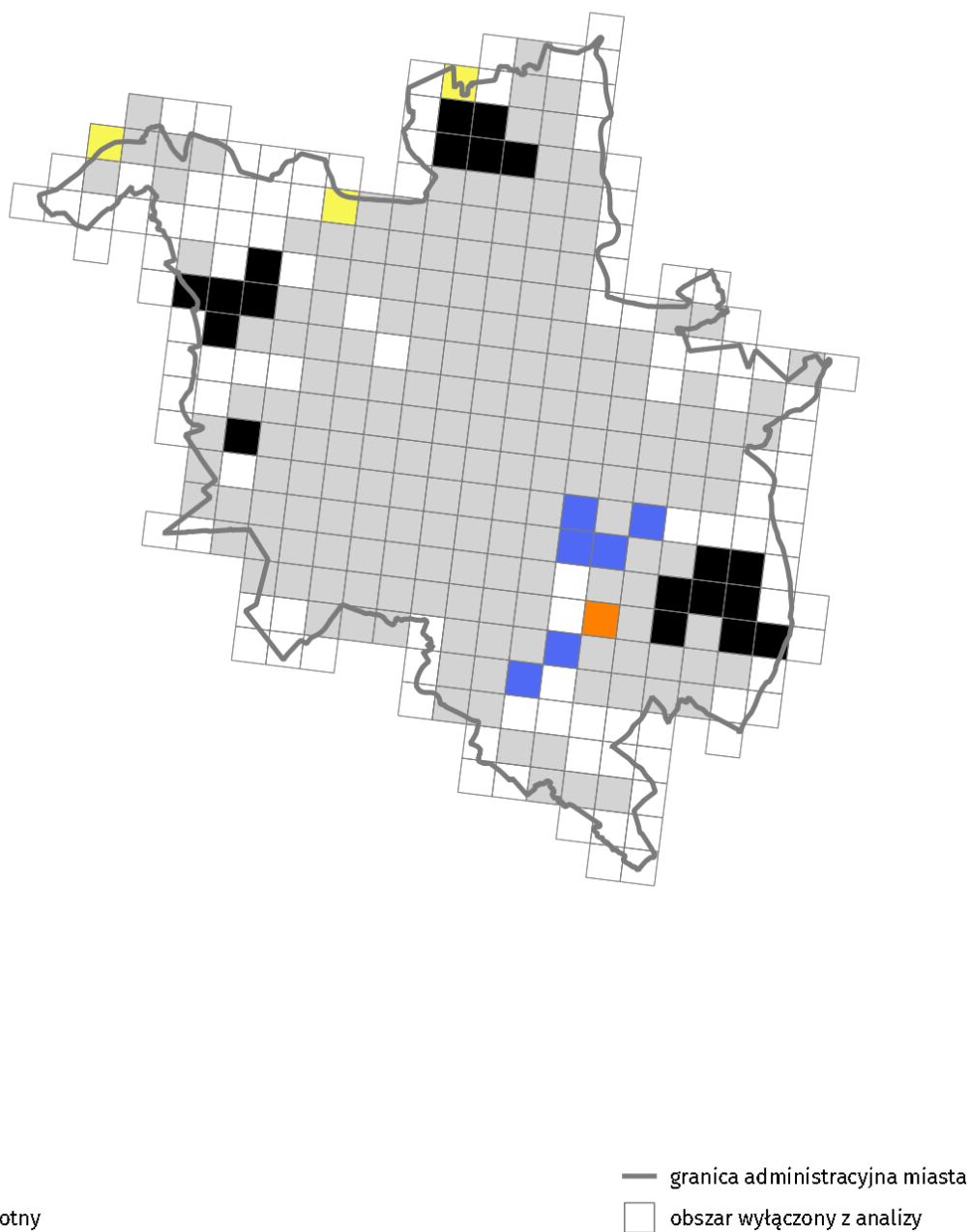
⁸ Suhecki (2010)

Mapa 13. Wskaźnik zróżnicowania kwintylnego (S80/S20) – LISA – Poznań



Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

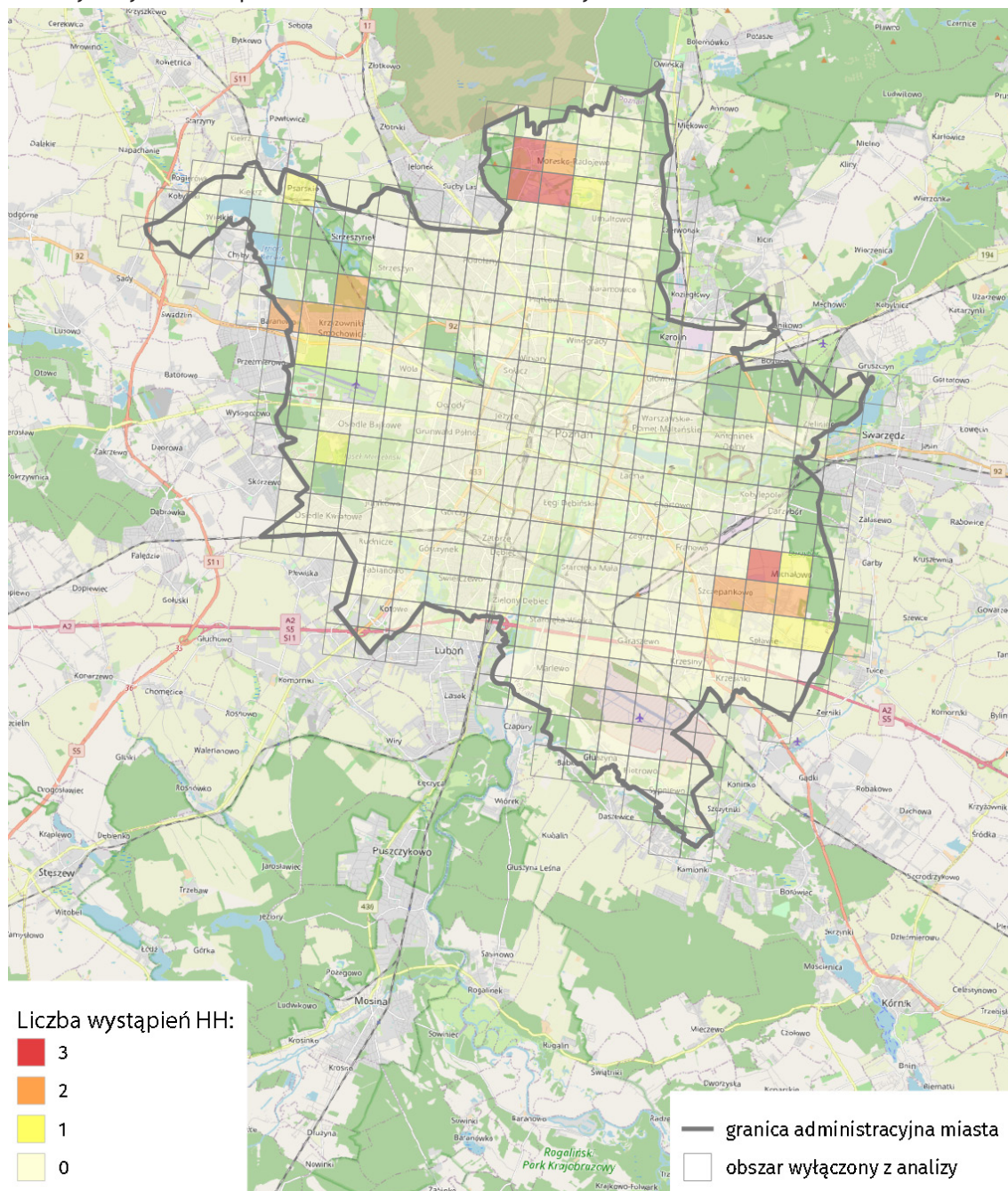
Mapa 14. Współczynnik Giniego – LISA – Poznań



Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Nakładając na siebie trzy warstwy utworzone dla wyżej zobrazowanych miar ustalono częstotliwość wystąpienia w tej samej jednostce przestrzennej wysokich wartości (HH) (oprócz mediany).

Mapa 15. Obszary o wysokim stopniu zróżnicowania dochodów – synteza – Poznań



Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Wyniki syntezy, jak obrazuje kartogram, wskazują na wyraźnie rysujące się w przestrzeni Poznania trzy klastry o najwyższym zróżnicowaniu dochodów uzyskanych z tytułu umowy o pracę. Skupiska te zlokalizowane są w północnej, północno-zachodniej oraz południowo-wschodniej części miasta.

III. Ocena spełnienia kryteriów jakości otrzymanych wyników

III. Evaluation of the degree to which the results meet the quality criteria

Po przeprowadzeniu badania dokonano oceny spełnienia kryteriów jakości uzyskanych wyników pod względem:

- przydatności danych,
- dokładności danych,
- terminowości i punktualności danych,
- porównywalności danych,
- spójności i zastępowalności danych,
- dostępności i przejrzystości danych.

Przydatność danych

W zasobach statystyki publicznej, zwłaszcza w odniesieniu do miast i obszarów wewnątrzmijskich, brakuje informacji dotyczących charakterystyk opisujących poziom i źródła dochodów ludności. Zaspokojenie zgłaszanych potrzeb informacyjnych w zakresie dochodów ludności na poziomie lokalnym stało się inspiracją do podjęcia pracy badawczej „Stratyfikacja dochodowa mieszkańców miast,” z wykorzystaniem danych zawartych w źródłach administracyjnych.

Odbiorcami wyników badania będą instytucje zajmujące się monitorowaniem polityki spójności, jednostki administracji rządowej i samorządowej, środowisko naukowe i inni.

Dokładność danych

Dane otrzymane z Ministerstwa Finansów za rok 2018 dotyczące przychodów obejmowały dziesięć deklaracji podatkowych PIT – formularze PIT-28, PIT-28S, PIT-36, PIT-36S, PIT-36L, PIT-36LS, PIT-37, PIT-38, PIT-39 oraz PIT-40A/11A⁹. Ogółem otrzymano blisko 28,4 miliona deklaracji spośród których wyłączono z badania 5,1% z nich ze względu na brak informacji o wartości przychodu. Szczegółowe zestawienie według formularzy podatkowych przedstawia Tablica 25.

Tablica 25. Liczba deklaracji podatkowych pozyskanych do badania – przychody

Formularz	Liczba złożonych deklaracji	W tym deklaracje z brakiem informacji o przychodzie w %
PIT-28	1 260 749	7,2
PIT-28S	22	36,4
PIT-36	2 789 162	11,1
PIT-36S	43	4,7
PIT-36L	629 327	3,4
PIT-36LS	70	12,9
PIT-37	17 667 257	5,6
PIT-38	277 172	12,5

⁹ Przedstawione statystyki mogą nieznacznie różnić się od publikacji Ministerstwa Finansów „Informacja dotycząca rozliczenia podatku dochodowego od osób fizycznych za 2018” z powodu innej daty pobrania danych oraz niewielkich różnic definicyjnych.

Formularz	Liczba złożonych deklaracji	W tym deklaracje z brakiem informacji o przychodzie w %
PIT-39	106 503	0,8
PIT-40A/11A	5 640 730	0,0
Łącznie	28 371 035	5,1

Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Liczba podatników, na podstawie których badane były przychody musiała zostać ograniczona ze względu na brak możliwości przypisania im kodu TERYT ich miejsca zamieszkania oraz wspomniane wcześniej braki wartości przychodu. Stanowili oni 5,5% wszystkich podatników. Szczegółowe zestawienie liczby podatników składających formularze przedstawia Tablica 26.

Tablica 26. Liczba podatników – przychody

Formularz	Liczba złożonych deklaracji	Podatnicy uwzględnieni w analizie w % wszystkich
PIT-28	1 260 728	91,7
PIT-28S	22	63,6
PIT-36	2 789 122	88,4
PIT-36S	43	95,3
PIT-36L	629 305	96,3
PIT-36LS	70	87,1
PIT-37	17 666 828	93,9
PIT-38	277 146	86,9
PIT-39	106 483	96,4
PIT-40A/11A	5 640 726	99,1
Łącznie	26 746 099	94,5

Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Podatnicy uwzględnieni w analizie – to podatnicy, którym udało się przyporządkować kod TERYT miejsca zamieszkania oraz którzy mieli podaną wartość przychodów.

Jeśli podatnik złożył co najmniej dwa formularze podatkowe różnego typu, to wartości przychodów z tych formularzy były zsumowane dla każdego z takich podatników. 94,5% podatników składających zeznania podatkowe za rok 2018 złożyło zeznanie na jednym rodzaju formularzu, 5,2% na dwóch. Szczegółowe zestawienie liczby złożonych deklaracji podatkowych przedstawiono w Tablicy 27.

Tablica 27. Liczba złożonych formularzy przez podatników – przychody

Liczba złożonych formularzy	Liczba podatników
1	23 895 968
2	1 304 006

Liczba złożonych formularzy	Liczba podatników
3	78 564
4	4 321
5	106
Ogółem	25 282 965

Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Dane dotyczące dochodów pracowników najemnych (tj. otrzymujących wynagrodzenie z tytułu *stosunku pracy, służbowego, spółdzielczego...*) za 2018 rok pochodziły z formularza PIT-11¹⁰. Pozyskano blisko 15,3 miliona deklaracji dla pracowników najemnych, spośród których z badania wyłączono zaledwie 0,2% ze względu na brak informacji o wartości dochodu – Tablica 28.

Tablica 28. Liczba deklaracji podatkowych pozyskanych do badania – dochody

Formularz	Liczba złożonych deklaracji	W tym deklaracje z brakiem informacji o dochodzie w %
PIT-11	15 270 252	0,2

Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

PIT-11 - dotyczy jedynie podatników otrzymujących wynagrodzenie z tytułu stosunku pracy, służbowego, spółdzielczego... (pierwszy wiersz PIT-11) oraz zidentyfikowanych w bazie wejściowej numerem PESEL.

Podobnie jak w przypadku przychodów, liczba podatników, których dochody zostały poddane badaniu, została ograniczona o te osoby, którym nie można było przypisać kodu TERYT ich miejsca zamieszkania oraz te, które nie miały podanej wartości dochodu. Stanowili oni jednak zaledwie 0,3% wszystkich podatników – Tablica 29.

Tablica 29. Liczba podatników – dochody

Formularz	Liczba podatników	Podatnicy uwzględnieni w analizie w % wszystkich
PIT-11	13 125 709	99,7

Źródło: Opracowanie własne na podstawie danych z Ministerstwa Finansów.

Podatnicy uwzględnieni w analizie – to podatnicy, którym udało się przyporządkować kod TERYT miejsca zamieszkania oraz którzy mieli podaną wartość dochodów.

Terminowość i punktualność danych

Dane dotyczą roku podatkowego 2018. Dane ostateczne zostały przekazane w lipcu 2020 roku. Praca została zrealizowana w grudniu 2020 roku.

Porównywalność danych

Praca badawcza ma charakter badania pilotażowego, w ramach którego została wypracowana metoda umożliwiająca jego kontynuację w kolejnych latach.

¹⁰ Przedstawione statystyki mogą nieznacznie różnić się od publikacji Ministerstwa Finansów „Informacja dotycząca rozliczenia podatku dochodowego od osób fizycznych za 2018” z powodu innej daty pobrania danych oraz niewielkich różnic definicyjnych.

Spójność i zastępowalność danych

Badanie „Stratyfikacja dochodowa mieszkańców miast” zalicza się do grona prac eksperymentalnych. Jego wyników nie da się porównać z danymi o dochodach publikowanymi przez Główny Urząd Statystyczny choćby z dwóch powodów. Z jednej strony Główny Urząd Statystyczny opracowuje i publikuje dane o dochodach w odniesieniu do gospodarstw domowych – z drugiej ma to związek z jego innowacyjnością polegającą na wykorzystaniu danych jednostkowych pochodzących z Ministerstwa Finansów. Główny Urząd Statystyczny opracowuje i publikuje dane o dochodach pochodzące z badań reprezentacyjnych. Stąd najczęściej dostępne informacje dotyczą poziomu wojewódzkiego lub wyższych poziomów przestrzennych, które zdaniem zainteresowanych tematem badaczy nie mogą być wykorzystane do bardziej precyzyjnych analiz przestrzenno-ekonomicznych.

Dostępność i przejrzystość danych

Na całość wyników pracy badawczej składają się:

- Raport, w którym zostały zaprezentowane wybrane dane na temat przychodów i dochodów podatników zamieszkujących miasta zagregowane do poziomu kraju, województw, miast małych, miast średnich, miast dużych (w tym miast wojewódzkich), miast tracących funkcje społeczno-gospodarcze a także miast imiennie;
- Załączniki do raportu:
 - Struktury pozyskanych zbiorów danych;
 - Schemat przygotowania baz danych;
 - Kartogramy dla miast wojewódzkich;
 - Baza danych zawierająca informacje dla miast imiennie, zarówno dla przychodów podatników (mediana), jak i dochodów pracowników najemnych (mediana, relacja decyla dziewiątego do decyla pierwszego P90/P10, wskaźnik zróżnicowania kwintylogowego S80/S20, współczynnik Giniego);
 - Baza danych zawierająca informacje dla miast wojewódzkich w układzie siatki kwadratów dla dochodów pracowników najemnych (mediana w przedziałach kwartylogowych, relacja decyla dziewiątego do decyla pierwszego P90/P10, wskaźnik zróżnicowania kwintylogowego S80/S20, współczynnik Giniego).

Wyniki pracy badawczej zostały udostępnione na stronie internetowej Głównego Urzędu Statystycznego w obszarze Statystyki eksperymentalne.

Zakres danych, możliwych do pozyskania na potrzeby tego badania zdeterminował przeprowadzenie analizy dla całej zbiorowości podatników wyłącznie na podstawie informacji o uzyskanych przez nich przychodów. Jedynie dla pracowników najemnych mieszkających w miastach istniała możliwość zaprezentowania ich dochodów.

IV. Uwagi końcowe

IV. Concluding remarks

Najważniejszym źródłem danych atrybutowych, tj. danych o przychodach/dochodach ludności, wykorzystanych w badaniu są informacje zawarte w rejestrach podatkowych. W trakcie prowadzonych prac zidentyfikowano pięć badań zapisanych w Programie Badań Statystycznych Statystyki Publicznej 2018 na potrzeby, których przekazywane są dane Ministerstwa Finansów: 1.04.01, 1.04.06, 1.04.08, 1.24.01, 1.70.02.

Okazuje się, że zakres danych pozyskiwanych w ramach ww. badań determinuje sposób ujęcia sytuacji finansowej ludności. Koncentrują się one bowiem na kategorii przychodów, nie uwzględniając elementów kosztów uzyskania przychodów bądź wymiaru zysku/straty, dookreślających dochód. W odniesieniu jedynie do pracowników najemnych istnieje możliwość przedstawienia ich dochodów.

Oczywistym wnioskiem dla statystyki publicznej w kontekście kolejnych programów badań statystycznych statystyki publicznej jest **przemyślenie jednego, spójnego zakresu danych pozyskiwanych od jednego gestora na potrzeby różnych badań.**

Doświadczenia zebrane podczas pracy badawczej „Stratyfikacja dochodowa mieszkańców miast” wskazują na potrzebę udoskonalenia metodyki badania i rozszerzenie jego zakresu o zmienną „dochody” nie tylko dla pracowników najemnych. Konieczna w tym celu wydaje się współpraca służb statystyki publicznej i Ministerstwa Finansów.

Ponadto należy również rozważyć możliwość rozszerzenia zakresu terytorialnego badania. Wykorzystane w badaniu źródło administracyjne daje szansę uzyskania danych o dochodach mieszkańców miast, lecz również wszystkich gmin w kraju.

Ważnym źródłem danych była także Zintegrowana Baza Populacji - zbiór stworzony przez Departament Badań Demograficznych oraz Departament Systemów Teleinformatycznych, Geostatystyki i Spisów Głównego Urzędu Statystycznego dzięki, której możliwe było uzupełnienie, zaktualizowanie oraz zgeokodowanie danych adresowych zawartych w rejestrach podatkowych.

Nie można na podstawie danych Ministerstwa Finansów ustalić dochodów gospodarstw domowych. Jednak **powiązanie ich z informacjami w planowanym module „Rodziny” Zintegrowanej Bazy Populacji umożliwiłoby charakterystykę sytuacji dochodowej nie tylko podatników, lecz także rodzin wyodrębnionych w wyniku procesu parentyzacji.**

Należy mieć nadzieję, że **dobra jakościowo Zintegrowana Baza Populacji** stanie się zbiorem referencyjnym dla statystyki publicznej w zakresie wielu badań demograficznych i społecznych.

Mając na uwadze rosnącą wagę opisu tego, co dzieje się wewnątrz miast i obszarów je otaczających, **niezwykle istotne jest wypracowanie odpowiedniego instrumentarium składającego się z metod i zasad prezentowania informacji na siatce kwadratów zgodnie z warunkami tajemnicy statystycznej wynikającymi z ustawy.**

Środowisko przetwarzania danych przestrzennych składa się z oprogramowania dostępnego w środowisku OBM, tzn. R-Studio i SAS oraz oprogramowania do wizualizacji wyników analiz w postaci oprogramowania QGIS oraz ArcGIS. Taki dualizm wynika z charakteru przetwarzanych danych – dane jednostkowe przetwarza się zdalnie w bezpiecznym środowisku OBM, natomiast dane zagregowane, podlegające wizualizacji przetwarza się lokalnie na stacji roboczej.

Przedstawione w niniejszym raporcie wyniki postępowania badawczego dają określony obraz rozwarstwienia dochodowego mieszkańców miast. Wymaga on jeszcze korekt i uzupełnień, ale – zdaniem autorów – jest znaczącym krokiem na drodze do rozpoznania umożliwiającego monitorowanie

i kształtowanie sytuacji bytowej w miastach. Rozwijając rozpoczęte badania, należy ustalać nie tylko liczbową charakterystykę tej sytuacji, ale też szukać odpowiedzi o przyczyny i konsekwencje opisywanych stanów.

Bibliografia

References

- [1]. Alfons, A. i Templ, M. (2013). Estimation of Social Exclusion Indicators from Complex Surveys: The R Package laeken. *JSS Journal of Statistical Software*, 54(15).
- [2]. Alfons, A., Holzer, J., Templ, M. i Haider, A. (2020). *Estimation of Indicators on Social Exclusion and Poverty*.
- [3]. Atkinson, A. i Bourguignon, F. (2000). *Handbook of Income Distribution 1*.
- [4]. Atkinson, A. i Bourguignon, F. (2015). *Handbook of Income Distribution 2*.
- [5]. Basarbowicz, K., Dąbrowski, A., Filas-Przybył, S., Kaźmierczak, M., Klimanek, T., Kowalewski, J., Kruszka, K., Pawlikowski, D., Stachowiak, D., Stawikowska, M. i inni (2015). *Identyfikacja obszarów specjalnych wewnątrz miast wojewódzkich oraz na ich obszarach funkcjonalnych uwzględniających sytuację demograficzną i ekonomiczną ich mieszkańców na podstawie analiz przestrzennych z wykorzystaniem Geographic Information System (GIS)*. Centrum Badań i Edukacji Statystycznej GUS, Warszawa.
- [6]. Bruil, A. i Koymans, M. (2014). *Measuring Inequalities in the Dutch Household Sector*. Statistics Netherlands.
- [7]. Canberra Group. (2001). *Expert Group on Household Income Statistics*. The Canberra Group.
- [8]. Canberra Group. (2011). *Handbook on Household Income Statistics* (wyd. 2nd).
- [9]. Chakravarty, S. (2009). *Inequality, Polarization and Poverty: Advances in Distributional Analysis*.
- [10]. Chrostek, P., Klejdysz, J., Kormiluk, D., Skawiński, M. (2019). *Wybrane aspekty systemu podatkowo-składkowego na podstawie danych PIT i ZUS 2016*, MF Opracowania i Analizy
- [11]. Cowell, F. i Flachaire, E. (2007, 12). Income distribution and inequality measurement: The problem of extreme values. *Journal of Econometrics*, 141(2), 1044-1072
- [12]. Cowell, F. i Victoria-Feser, M.-P. (1996). Poverty measurement with contaminated data: A robust approach. *European Economic Review*, 40, 1761-1771.
- [13]. Cowell, F. i Victoria-Feser, M.-P. (1996, 1). Robustness Properties of Inequality Measures. *Econometrica*, 64(1), 77.
- [14]. de Waal, T., Pannekoek, J. i Scholtus, S. (2011). *Handbook of Statistical Data Editing and Imputation*. Wiley.
- [15]. Denderski, P. (2019). *Metody identyfikacji obserwacji odstających*.
- [16]. Dupriez, O. (2007). *Building a household consumption database for the calculation of poverty ppps. Technical report*. Washington: World Bank.
- [17]. EUROSTAT. (2003). *'Laeken' indicators - detailed calculation methodology*.
- [18]. Fesseau, M. i Mattonetti, L. (2013). *Distributional measures across household groups in a national accounts framework*. EUROSTAT&OECD.
- [19]. Filzmoser, P., Gussenbauer, J. i Templ, M. (2016). *Detecting outliers in household consumption survey data*. Vienna University of Technology, Vienna.

- [20]. G. Brys, M. H. (2003). A robust measure of skewness. *Journal of Computational and Graphical Statistics*, strony 996-1017.
- [21]. Główny Urząd Statystyczny. (2019). *Budżety gospodarstw domowych w 2018*.
- [22]. Główny Urząd Statystyczny. (2019). *Dochody i warunki życia ludności Polski - raport z badania EU-SILC 2018*.
- [23]. GUS. (2019). *Budżety gospodarstw domowych w 2018 r.* Warszawa: Główny Urząd Statystyczny.
- [24]. GUS. (2019). *Dochody i warunki życia ludności Polski – raport z badania EU-SILC 2018*. Warszawa: Główny Urząd Statystyczny.
- [25]. Hubert, E. V. (2008). An adjusted boxplot for skewed distributions. *Computational Statistics and Data Analysis*, strony 5186-5201.
- [26]. *Informacja dotycząca rozliczenia podatku dochodowego od osób fizycznych za 2018 rok (2019)*, Ministerstwo Finansów, Departament Podatków Dochodowych, Warszawa
- [27]. Kerm, P. (2007). *Extreme incomes and the estimation of poverty and inequality indicators from EU-SILC**. Luxembourg.
- [28]. Martín-Román, Ayala, J., Vicente, L. i Juan. (2017). *Regional inequality in decentralized countries: a multi-country analysis using LIS*.
- [29]. Mastronardi, L. i Cavallo, A. (2020). The spatial dimension of income inequality: An analysis at municipal level. *Sustainability (Switzerland)*, 12(4).
- [30]. Moser, M. i Schnetzer, M. (2014). *The Geography of Average Income and Inequality: Spatial Evidence from Austria*.
- [31]. Neri, L., Gagliardi, F., Ciampalini, G., Verma, V. i Betti, G. (2009). *Outliers at upper end of income distribution (EU-SILC 2007)*.
- [32]. OECD. (2013). *OECD Framework for Statistics on the Distribution of Household Income, Consumption and Wealth*.
- [33]. OECD. (2013). *OECD Guidelines for Micro Statistics on Household Wealth*. OECD.
- [34]. Osier, G. (2009). Variance Estimation for Complex Indicators of Poverty and Inequality Using Linearization Techniques. *Survey Research Methods*, 3(3), 167-195.
- [35]. Ostasiewicz, W. (1998). *Statystyczne metody analizy danych*. Wrocław: Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu.
- [36]. Panek, T. (2014). *Miary nierówności*. SGH, Warszawa.
- [37]. Panzera, D. i Postiglione, P. (2020). Measuring the Spatial Dimension of Regional Inequality: An Approach Based on the Gini Correlation Measure. *Social Indicators Research*, 148(2), 379-394.
- [38]. Rousseeuw, P. i Croux, C. (1993). Alternatives to the Median Absolute Deviation. *Journal of the American Statistical Association*, 88(424), 1273-1283.
- [39]. Sączewska-Piotrowska, A. (2018). *Income Stratification in Poland*. *Studia Ekonomiczne / Uniwersytet Ekonomiczny w Katowicach*, nr 373, 42-53.
- [40]. Szulc, A. (2018). *Pieniężne mierniki zamożności*. SGH, Warszawa.

- [41]. Śliwicki, D. i inni (2018), *Opracowanie metodologii i oszacowanie liczby pracujących w gospodarce narodowej według miejsca zamieszkania i głównego miejsca pracy na poziomie NTS 4, stopy bezrobocia rejestrowanego na poziomie NTS 5 oraz miar wynagrodzeń brutto na poziomie NTS 4*, Centrum Badań i Edukacji Statystycznej GUS, Warszawa
- [42]. Suchecki B. (red.), *Ekonometria przestrzenna: metody i modele analizy danych przestrzennych*, Wydawnictwo C.H. Beck, Warszawa 2010.
- [43]. UN. (2015). *Inequality Measurement*.
- [44]. Vecchi, G. (2018). *Robust measures of income and wealth inequality*.

Załączniki

Appendices

Załącznik 1. Struktury pozyskanych zbiorów danych

Szczegółową strukturę poszczególnych zbiorów Ministerstwa Finansów wykorzystanych na potrzeby badania „Stratyfikacja dochodowa mieszkańców miast” przedstawiają poniżej zamieszczone zestawienia tabelaryczne.

Struktura zbiorów Ministerstwa Finansów dotyczących PIT-28 i PIT-28S

Zmienna	Opis
<i>rok_podat_PIT</i>	rok rozliczenia podatkowego
<i>formularz_PIT</i>	kod formularza rozliczenia podatkowego
<i>tech_id_PIT</i>	techniczny identyfikator podatnika
<i>podat_plec_PIT</i>	płeć podatnika
<i>podat_rok_ur_PIT</i>	rok urodzenia podatnika
<i>adr_kod_gmn_PIT</i>	kod TERYT gminy
<i>przych_art17_PIT</i>	przychód określony przez organ podatkowy na podstawie art. 17 ustawy
<i>podat_przych_PIT</i>	przychód
<i>przych_najem_dzierz_PIT</i>	przychód z najmu, podnajmu lub dzierżawy (1 – TAK, 0 – NIE)
<i>przych_dzial_naz_ryczalt_PIT</i>	przychód z działalności prowadzonej na własne nazwisko opodatkowany ryczałtem (1 – TAK, 0 – NIE)
<i>przych_dzial_sp_ryczalt_PIT</i>	przychody z działalności prowadzonej w formie spółki opodatkowany ryczałtem (1 – TAK, 0 – NIE)
<i>kwota_1proc_opp_PIT</i>	wnioskowana kwota 1% do przekazania na organizację pożytku publicznego
<i>KRS_PIT</i>	numer KRS organizacji pożytku publicznego, na którą został przekazany 1% podatku należnego
<i>darowizny_opp_PIT</i>	darowizny - przekazane org. na dział. pożytku publicznego - art. 26 ust. 1 pkt 9 lit. a ustawy
<i>darowizny_kult_relig_PIT</i>	darowizny - przekazane org. na dział. pożytku publicznego - art. 26 ust. 1 pkt 9 lit. a ustawy
<i>darowizny_kult_relig_PIT</i>	darowizny - przekazane na cele kultu religijnego - art. 26 ust. 1 pkt 9 lit. b ustawy
<i>darowizny_krew_PIT</i>	darowizny - przekazane na krwiodawstwa - art. 26 ust. 1 pkt 9 lit. c ustawy
<i>darowizny_inne_PIT</i>	darowizny wynikające z odrębnych ustaw

Źródło: Opracowanie własne.

Struktura zbiorów Ministerstwa Finansów dotyczących PIT-36, PIT-37 i PIT-40A/11A

Zmienna	Opis
<i>rok_podat_PIT</i>	rok rozliczenia podatkowego
<i>formularz_PIT</i>	kod formularza rozliczenia podatkowego
<i>tech_id_PIT</i>	techniczny identyfikator podatnika
<i>podat_plec_PIT</i>	płeć podatnika
<i>podat_rok_ur_PIT</i>	rok urodzenia podatnika
<i>adr_kod_gmn_PIT</i>	kod TERYT gminy
<i>id_roz_wpolnie_PIT</i>	identyfikator wspólnego rozliczenia
<i>wsk_dochod_emer_rent_PIT</i>	uzyskanie dochodów z emerytury/renty (1 – TAK, 0 – NIE)
<i>wsk_dochod_dzial_gosp_PIT</i>	uzyskanie dochodów z działalności gospodarczej (1 – TAK, 0 – NIE)
<i>podat_przych_PIT</i>	przychód
<i>kwota_1proc_opp_PIT</i>	wnioskowana kwota 1% do przekazania na organizację pożytku publicznego
<i>KRS_PIT</i>	numer KRS organizacji pożytku publicznego, na którą został przekazany 1% podatku należnego
<i>darowizny_opp_PIT</i>	darowizny - przekazane org. na dział. pożytku publicznego - art. 26 ust. 1 pkt 9 lit. a ustawy
<i>darowizny_kult_relig_PIT</i>	darowizny - przekazane na cele kultu religijnego - art. 26 ust. 1 pkt 9 lit. b ustawy
<i>darowizny_krew_PIT</i>	darowizny - przekazane na krwiodawstwa - art. 26 ust. 1 pkt 9 lit. c ustawy
<i>darowizny_inne_PIT</i>	darowizny wynikające z odrębnych ustaw

Źródło: Opracowanie własne.

Struktura zbiorów Ministerstwa Finansów dotyczących PIT-36L, PIT-38 i PIT-39

Zmienna	Opis
<i>rok_podat_PIT</i>	rok rozliczenia podatkowego
<i>formularz_PIT</i>	kod formularza rozliczenia podatkowego
<i>tech_id_PIT</i>	techniczny identyfikator podatnika
<i>podat_plec_PIT</i>	płeć podatnika
<i>podat_rok_ur_PIT</i>	rok urodzenia podatnika
<i>adr_kod_gmn_PIT</i>	kod TERYT gminy
<i>podat_przych_PIT</i>	przychód
<i>kwota_1proc_opp_PIT</i>	wnioskowana kwota 1% do przekazania na organizację pożytku publicznego
<i>KRS_PIT</i>	numer KRS organizacji pożytku publicznego, na którą został przekazany 1% podatku należnego

Źródło: Opracowanie własne.

Struktura zbiorów Ministerstwa Finansów dotyczących PIT-36S i PIT-36LS

Zmienna	Opis
<i>rok_podat_PIT</i>	rok rozliczenia podatkowego
<i>formularz_PIT</i>	kod formularza rozliczenia podatkowego
<i>tech_id_PIT</i>	techniczny identyfikator podatnika
<i>podat_plec_PIT</i>	płeć podatnika
<i>podat_rok_ur_PIT</i>	rok urodzenia podatnika
<i>adr_kod_gmn_PIT</i>	kod TERYT gminy
<i>wsk_dochod_emer_rent_PIT</i>	uzyskanie dochodów z emerytury/renty (1 – TAK, 0 – NIE)
<i>wsk_dochod_dzial_gosp_PIT</i>	uzyskanie dochodów z działalności gospodarczej (1 – TAK, 0 – NIE)
<i>podat_przych_PIT</i>	przychód
<i>kwota_1proc_opp_PIT</i>	wnioskowana kwota 1% do przekazania na organizację pożytku publicznego
<i>KRS_PIT</i>	numer KRS organizacji pożytku publicznego, na którą został przekazany 1% podatku należnego
<i>darowizny_opp_PIT</i>	darowizny - przekazane org. na dział. pożytku publicznego - art. 26 ust. 1 pkt 9 lit. a ustawy
<i>darowizny_kult_relig_PIT</i>	darowizny - przekazane na cele kultu religijnego - art. 26 ust. 1 pkt 9 lit. b ustawy
<i>darowizny_krew_PIT</i>	darowizny - przekazane na krwiodawstwa - art. 26 ust. 1 pkt 9 lit. c ustawy
<i>darowizny_inne_PIT</i>	darowizny wynikające z odrębnych ustaw

Źródło: Opracowanie własne.

Struktura zbiorów Ministerstwa Finansów dotyczących danych adresowych w PIT-28, PIT-28S, PIT-36, PIT-36S, PIT-36L, PIT-36LS, PIT-37, PIT-38, PIT-39, PIT-40A/11A

Zmienna	Opis
<i>tech_id_PIT</i>	techniczny identyfikator podatnika
<i>adr_kod_gmn_PIT</i>	kod TERYT gminy zamieszkania podatnika
<i>adr_naz_msc_PIT</i>	nazwa miejscowości zamieszkania podatnika
<i>adr_kod_pocz_PIT</i>	kod pocztowy miejsca zamieszkania podatnika
<i>adr_naz_ulica_PIT</i>	nazwa ulicy zamieszkania podatnika
<i>adr_naz_ulica_PIT</i>	przedrostek ulica zamieszkania podatnika
<i>adr_nr_bud_PIT</i>	numer budynku zamieszkania podatnika

Źródło: Opracowanie własne.

Struktura zbioru dotyczącego PIT-11

Zmienna	Opis
<i>rok_podat_PIT</i>	rok rozliczenia podatkowego
<i>id_dok_PIT</i>	id dokumentu
<i>platnik_NIP_PIT</i>	numer NIP płatnika
<i>podat_adr_kod_kraj_PIT</i>	trzycyfrowy kod kraju
<i>podat_adr_kod_gmn_PIT</i>	kod TERYT gminy zamieszkania podatnika
<i>podat_PESEL_PIT</i>	numer identyfikacyjny PESEL
<i>podat_plec_PIT</i>	płeć (1 – mężczyzna, 2 – kobieta)
<i>suma_przych_WP_PIT</i>	przychód wynagrodzenia ze stosunku: pracy, służbowego, spółdzielczego i z pracy nakładczej, a także zasiłki pieniężne z ubezpieczenia społecznego wypłacone przez zakład pracy, o którym mowa w art. 31 ustawy, oraz płatników, o których mowa w art. 42e ust. 1 ustawy
<i>suma_koszt_WP_PIT</i>	koszty uzyskania przychodu ze stosunku pracy, służbowego, spółdzielczego i z pracy nakładczej, a także zasiłki pieniężne z ubezpieczenia społecznego wypłacone przez zakład pracy, o którym mowa w art. 31 ustawy, oraz płatników, o których mowa w art. 42e ust. 1 ustawy
<i>suma_dochod_WP_PIT</i>	dochód z wynagrodzenia ze stosunku: pracy, służbowego, spółdzielczego i z pracy nakładczej, a także zasiłki pieniężne z ubezpieczenia społecznego wypłacone przez zakład pracy, o którym mowa w art. 31 ustawy, oraz płatników, o których mowa w art. 42e ust. 1 ustawy
<i>dochod_zwol_WP_PIT</i>	dochód z wynagrodzenia ze stosunku: pracy, służbowego, spółdzielczego i z pracy nakładczej, a także zasiłki pieniężne z ubezpieczenia społecznego wypłacone przez zakład pracy, o którym mowa w art. 31 ustawy, oraz płatników, o których mowa w art. 42e ust. 1 ustawy zwolniony od podatku
<i>suma_zaliczki_WP_PIT</i>	zaliczka pobrana przez płatnika z wynagrodzenia ze stosunku pracy, służbowego, spółdzielczego i z pracy nakładczej, a także zasiłki pieniężne z ubezpieczenia społecznego wypłacone przez zakład pracy, o którym mowa w art. 31 ustawy, oraz płatników, o których mowa w art. 42e ust. 1 ustawy
<i>suma_wyplat_dni_SP_PIT</i>	przychód z należności z tytułu członkostwa w rolniczej spółdzielni produkcyjnej lub innej spółdzielni zajmującej się produkcją rolną oraz zasiłki pieniężne z ubezpieczenia społecznego
<i>dochod_wyplat_dni_SP_PIT</i>	dochód z należności z tytułu członkostwa w rolniczej spółdzielni produkcyjnej lub innej spółdzielni zajmującej się produkcją rolną oraz zasiłki pieniężne z ubezpieczenia społecznego
<i>suma_zaliczki_dni_SP_PIT</i>	należne zaliczki z tytułu członkostwa w rolniczej spółdzielni produkcyjnej lub innej spółdzielni zajmującej się produkcją rolną oraz zasiłki pieniężne z ubezpieczenia społecznego
<i>suma_przych_EZ_PIT</i>	przychód z emerytury i renty z zagranicy
<i>suma_dochod_EZ_PIT</i>	dochód z emerytury i renty z zagranicy
<i>suma_zaliczki_EZ_PIT</i>	zaliczka pobrana przez płatnika z emerytury i renty z zagranicy
<i>suma_przych_AR_PIT</i>	przychód z tytułu należności za pracę przypadające tymczasowo aresztowanym lub skazanym
<i>suma_dochod_AR_PIT</i>	dochód z tytułu należności za pracę przypadające tymczasowo aresztowanym lub skazanym

Zmienna	Opis
<i>suma_zaliczki_AR_PIT</i>	zaliczka pobrana przez płatnika z tytułu należności za pracę przypadające tymczasowo aresztowanym lub skazanym
<i>suma_wypłat_FN_PIT</i>	suma wypłat z tytułu świadczeń wypłacanych z Funduszu Pracy i Funduszu Gwarantowanych Świadczeń Pracowniczych
<i>suma_dochod_FN_PIT</i>	dochód z tytułu świadczeń wypłacanych z Funduszu Pracy i Funduszu Gwarantowanych Świadczeń Pracowniczych
<i>suma_zaliczki_FN_PIT</i>	zaliczka pobrana przez płatnika z tytułu świadczeń wypłacanych z Funduszu Pracy i Funduszu Gwarantowanych Świadczeń Pracowniczych
<i>suma_przych_DG_PIT</i>	przychód z działalności wykonywanej osobiście - art.13 pkt.2,4 i 6 (z wyjątkiem czynności związanych z pełnieniem obowiązków społecznych lub obywatelskich) i pkt 8
<i>suma_koszt_DG_PIT</i>	koszty uzyskania przychodu z działalności wykonywanej osobiście - art.13 pkt.2,4 i 6 (z wyjątkiem czynności związanych z pełnieniem obowiązków społecznych lub obywatelskich) i pkt 8
<i>suma_dochod_DG_PIT</i>	dochody z działalności wykonywanej osobiście - art.13 pkt.2,4 i 6 (z wyjątkiem czynności związanych z pełnieniem obowiązków społecznych lub obywatelskich) i pkt 8
<i>suma_zaliczki_DG_PIT</i>	zaliczka pobrana przez płatnika z działalności wykonywanej osobiście - art.13 pkt.2,4 i 6 (z wyjątkiem czynności związanych z pełnieniem obowiązków społecznych lub obywatelskich) i pkt 8
<i>suma_przych_OS_PIT</i>	przychód z tytułu czynności związanych z pełnieniem obowiązków społecznych lub obywatelskich (art. 13 ust 5 i 6)
<i>suma_koszt_OS_PIT</i>	koszty uzyskania przychodu z tytułu czynności związanych z pełnieniem obowiązków społecznych lub obywatelskich (art.13 ust 5 i 6)
<i>suma_dochod_OS_PIT</i>	dochód z tytułu czynności związanych z pełnieniem obowiązków społecznych lub obywatelskich (art.13 ust 5 i 6)
<i>suma_zaliczki_OS_PIT</i>	zaliczka pobrana przez płatnika z tytułu czynności związanych z pełnieniem obowiązków społecznych lub obywatelskich (art. 13 ust 5 i 6)
<i>suma_przych_PM_PIT</i>	przychód z tytułu z praw autorskich i innych praw - art.18 ustawy
<i>suma_koszt_PM_PIT</i>	koszty uzyskania przychodu z tytułu z praw autorskich i innych praw - art.18 ustawy
<i>suma_dochod_PM_PIT</i>	dochód z tytułu z praw autorskich i innych praw - art.18 ustawy
<i>suma_zaliczki_PM_PIT</i>	zaliczka pobrana przez płatnika z tytułu z praw autorskich i innych praw - art.18 ustawy
<i>suma_przych_UA_PIT</i>	przychód z tytułu należności wynikających z umowy aktywizacyjnej- art.35a ust.1 ustawy
<i>suma_koszt_UA_PIT</i>	koszty uzyskania przychodu z tytułu należności wynikających z umowy aktywizacyjnej - art.35a ust.1 ustawy
<i>suma_dochod_UA_PIT</i>	dochód z tytułu należności wynikających z umowy aktywizacyjnej - art.35a ust.1 ustawy
<i>suma_zaliczki_UA_PIT</i>	zaliczka pobrana przez płatnika z tytułu należności wynikających z umowy aktywizacyjnej - art.35a ust.1 ustawy
<i>suma_przych_IZ_PIT</i>	przychód z innych źródeł (wg deklaracji płatnika)

Zmienna	Opis
<i>suma_koszt_IZ_PIT</i>	koszty uzyskania przychodu z innych źródeł (wg deklaracji płatnika)
<i>suma_dochod_IZ_PIT</i>	dochód z innych źródeł (wg deklaracji płatnika)
<i>suma_dochod_zwol_IZ_PIT</i>	dochód z innych źródeł zwolniony od podatku (wg deklaracji płatnika)
<i>suma_zaliczki_WP_PIT</i>	zaliczka pobrana przez płatnika z innych źródeł (wg deklaracji płatnika)
<i>suma_odl_ubez_spol_PIT</i>	składki na ubezpieczenia społeczne, o których mowa w przepisach ustawy, podlegające odliczeniu od dochodu
<i>suma_odl_ubez_spol_art26_PIT</i>	zagraniczne składki na ubezpieczenia społeczne podlegające odliczeniu od dochodu, o których mowa w art. 26 ust. 1 pkt 2a ustawy
<i>suma_odl_ubez_zdr_PIT</i>	składki na ubezpieczenie zdrowotne, o których mowa w przepisach ustawy, podlegające odliczeniu od podatku
<i>suma_odl_ubez_zdr_art27b_PIT</i>	zagraniczne składki na ubezpieczenie zdrowotne podlegające odliczeniu od podatku, o których mowa w art. 27b ust. 1 pkt 2 ustawy
<i>suma_przych_pomoc_zgr_PIT</i>	przychody pochodzące ze środków bezzwrotnej pomocy zagranicznej, o których mowa w art. 21 ust. 1 pkt 46 ustawy

Źródło: Opracowanie własne.

Załącznik 2. Schemat przygotowania baz danych

Szczegółowy wykaz procedur przygotowujących zbiory MF PIT-11 i MF PIT na potrzeby badania „Stratyfikacja dochodowa mieszkańców miast” opisano poniżej.

Wykaz procedur przygotowujących zbiór wyników ze zbioru MF PIT-11:

Zbiór MF PIT-11 to zbiór pozyskany z Ministerstwa Finansów. Każdy rekord zbioru zawiera informacje z jednego złożonego formularza PIT-11.

Informacje dotyczące rozliczeń formularzy pracowników najemnych (tj. pracowników, którzy otrzymują wynagrodzenie ze stosunku pracy, służbowego, spółdzielczego i z pracy nakładczej, a także zasiłki pieniężne z ubezpieczenia społecznego wypłacone przez zakład pracy, o którym mowa w art. 31 ustawy, oraz płatników, o których mowa w art. 42e ust. 1 ustawy) został zawarte w rekordach o niepustych kolumnach *suma_przych_WP_PIT*, *suma_koszt_WP_PIT* i *suma_dochod_WP_PIT*.

Opis prac przeprowadzonych na zbiorze MF PIT-11 w celu zidentyfikowania pracowników najemnych, ich miejsca zamieszkania oraz ich dochodu:

1. Wybór ze zbioru MF PIT-11 rekordów dotyczących rozliczeń pracowników najemnych.

Zgodnie z zawartą w Załączniku 1. strukturą zbioru MF PIT-11 zawiera on informacje dotyczące wszystkich rodzajów źródeł dochodów objętych formularzem PIT-11. Ponieważ w raporcie rozpatrywany był jedynie przypadek pracowników najemnych, pozostałe rekordy (dla innych źródeł dochodów) należało pominąć w dalszym przetwarzaniu bazy. Wybrano zatem jedynie rekordy o niepustych wartościach w kolumnach *suma_przych_WP_PIT*, *suma_koszt_WP_PIT* i *suma_dochod_WP_PIT*.

Ponadto, zbiór MF PIT-11 zawierał rekordy z brakiem danych w kolumnie *podat_PESEL_PIT*. Dla planowanych obliczeń i statystyk istotna była identyfikacja wpisów należących do tego samego podatnika, konieczne było pominięcie wierszy z brakiem danych w kolumnie *podat_PESEL_PIT*.

2. Przekształcenie zbioru MF PIT-11 ograniczonego do rozliczeń pracowników najemnych na bazę podatników-pracowników najemnych.

Na tym etapie przygotowania, baza zawierała informacje o rozliczeniach PIT-11, a kolumną identyfikującą wiersze był identyfikator rozliczenia/dokumentu. W celu utworzenia bazy podatników (pracowników najemnych) zsumowano zadeklarowane dochody dla każdego z nich.

3. Uzupelnienie braków kodów TERYT gminy adresu zamieszkania na podstawie informacji ze zbioru ZBP.

Na podstawie numerów identyfikacyjnych PESEL przyłączono do bazy operacyjnej BO2 kolumnę zawierającą kod TERYT z bazy ZBP.

4. Aktualizacja kodów TERYT gminy adresu zamieszkania.

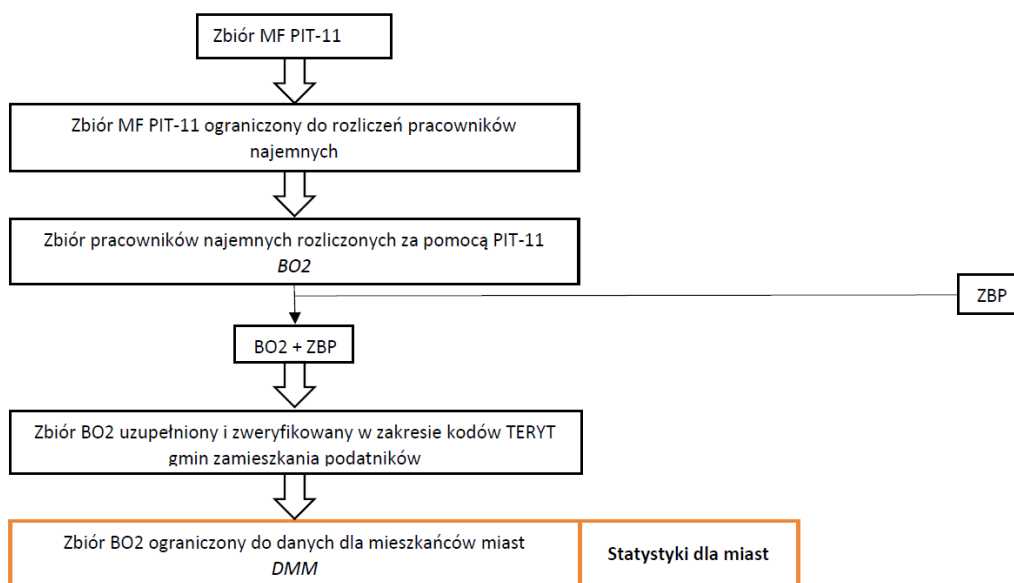
- jeżeli kod TERYT gminy podany w zbiorze MF PIT-11 przyjmował wartość niezgodną z rejestrem TERYT ze stanem na 31.12.2018 roku, to wartość tę zastąpiono odpowiednią wartością zgodną z tym rejestrem.
- Ponadto dokonano imputacji kodów TERYT gminy zamieszkania w opisanych poniżej sytuacjach, do wykonania tej operacji wykorzystując jedynie te rekordy bazy ZBP, dla których podany został kod TERYT gminy zamieszkania.
 - W przypadku gmin miejsko-wiejskich konieczne było rozbieżność gminy na część miejską i wiejską. Wówczas, w przypadku zgodności pierwszych sześciu cyfr kodu TERYT gminy zamieszkania podatnika w BO2 z pierwszymi sześcioma cyframi kodu TERYT gminy zamieszkania podatnika z ZBP zastąpiono kod TERYT BO2 kodem TERYT zaczerpniętym z ZBP. W przeciwnym przypadku, TERYT gminy zamieszkania podatnika został sklasyfikowany jako nieustalony, a wiersz dotyczący podatnika pominięty w dalszym przetwarzaniu bazy.
 - Jeżeli kod TERYT gminy nie został podany w zbiorze MF PIT-11, dokonano imputacji danych zaczerpniętych z bazy ZBP (po wcześniejszym zastąpieniu kodów TERYT dla delegatur kodem TERYT gminy nadrzędnej dla tej delegatury – dotyczy to pięciu największych miast: Warszawy, Krakowa, Łodzi, Wrocławia i Poznania). Jeżeli podatnikowi nie udało się przypisać kodu TERYT gminy zamieszkania z bazy ZBP, kod ten uznano jako nieustalony, a wiersz dotyczący podatnika pominięto w dalszym przetwarzaniu bazy.

5. Wybór z bazy rekordów dotyczących pracowników najemnych zamieszkujących w miastach.

Populację pracowników najemnych zamieszkujących w miastach otrzymano wybierając według kodu TERYT podatników zamieszkujących w gminach miejskich (ostatnia cyfra kodu TERYT: 1) oraz częściach miejskich gmin miejsko-wiejskich (ostatnia cyfra kodu TERYT: 4).

Otrzymano w ten sposób bazę DMM.

Schemat 1. Procedury przygotowujące bazę DMM



Źródło: Opracowanie własne.

Kolejny etap obejmował przekształcenie bazy DMM i przygotowanie zbioru danych potrzebnych w celu przeprowadzenia analiz wewnątrzmijskich dla miast wojewódzkich. W tej części szczególnie istotne było odniesienie przestrzenne miejsca zamieszkania podatnika wewnątrz miasta wojewódzkiego.

- Wybór z bazy rekordów dotyczących pracowników najemnych zamieszkujących w miastach wojewódzkich.

Dokonano ograniczenia bazy DMM do mieszkańców miast wojewódzkich według kodu TERYT gminy zamieszkania podatnika (baza DMMW).

- Uzupełnienie współrzędnych adresu zamieszkania na podstawie informacji ze zbioru ZBP oraz PRG_PA.

- W pierwszym kroku, na podstawie numerów identyfikacyjnych PESEL oraz kodów TERYT gminy zamieszkania przyłączono do bazy DMMW kolumny zawierające współrzędne budynku zamieszkania podatnika z bazy ZBP.
- Następnie, dla rekordów, którym nie udało się przyłączyć współrzędny z bazy ZBP, dołączono na podstawie adresu zamieszkania podatnika (kod TERYT gminy, miejscowość, ulica, numer domu) współrzędne z bazy PRG_PA.

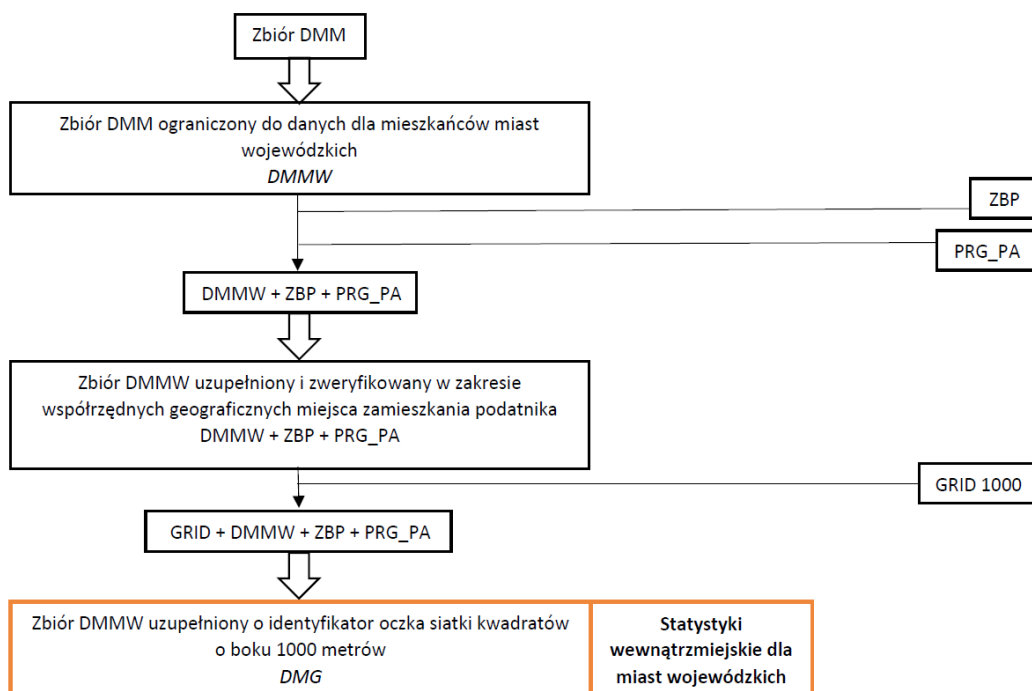
Jeżeli nie udało się przypisać podatnikowi współrzędnych ani z ZBP, ani z PRG_PA, to podatnik taki nie był uwzględniany w analizie wewnątrzmijskiej.

- Przypisanie adresów zamieszkania podatników do siatki kwadratów o boku 1000 metrów (GRID 1000).

Dla każdego rekordu w bazie odpowiadającemu jednemu podatnikowi przypisano jednoznacznie numer identyfikacyjny kwadratu, w którym znajdowało się jego miejsce zamieszkania.

Otrzymano w ten sposób bazę DMG.

Schemat 2. Procedury przygotowujące bazę DMG



Źródło: Opracowanie własne.

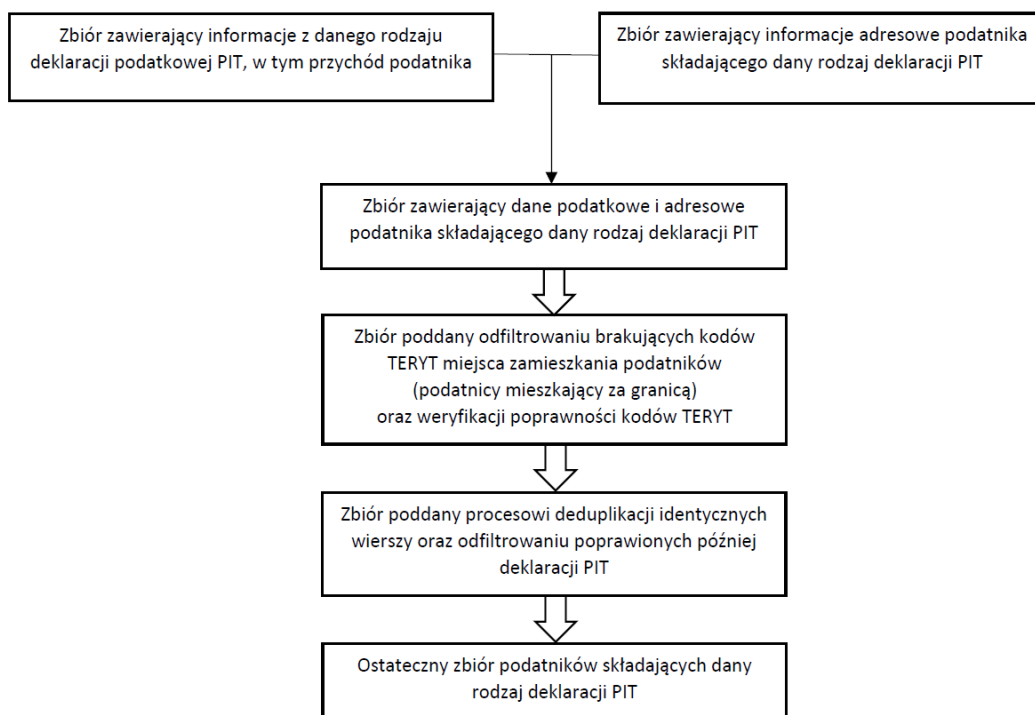
Wykaz procedur przygotowujących zbiór wyników ze zbiorów MF PIT:

Określenie zbioru MF PIT oznacza zbiory pozyskane z Ministerstwa Finansów dotyczące przychodów podatników rozliczających się na deklaracjach PIT-28, PIT-28S, PIT-36, PIT-36S, PIT-36L, PIT-36LS, PIT-37, PIT-38, PIT-39 i tych spośród podatników, którzy zostali rozliczeni wyłącznie przez ZUS na formularzu PIT-40A/11A. Dla każdej z tych dziesięciu deklaracji pozyskano dwa zbiory. Pierwszy z nich zawierał informacje o przychodach podatników, natomiast drugi – informacje na temat miejsca zamieszkania danego podatnika. Identyfikatorami dla obu rodzaju zbiorów były zmienne *tech_id*, które pozwalały na jednoznaczne ich połączenie.

Opis prac przeprowadzonych na zbiorach MF PIT w celu zidentyfikowania miejsca zamieszkania podatników oraz ich przychodu:

Dla każdego z analizowanych dziesięciu rodzajów formularzy PIT dokonano połączenia danych o przychodach z danymi adresowymi wykorzystując w tym celu zmienną *tech_id*, a następnie dokonano deduplikacji zbioru i weryfikacji w zakresie poprawności kodów TERYT oraz usunięto rekordy z brakiem danych dla kodu TERYT oraz wartości przychodu.

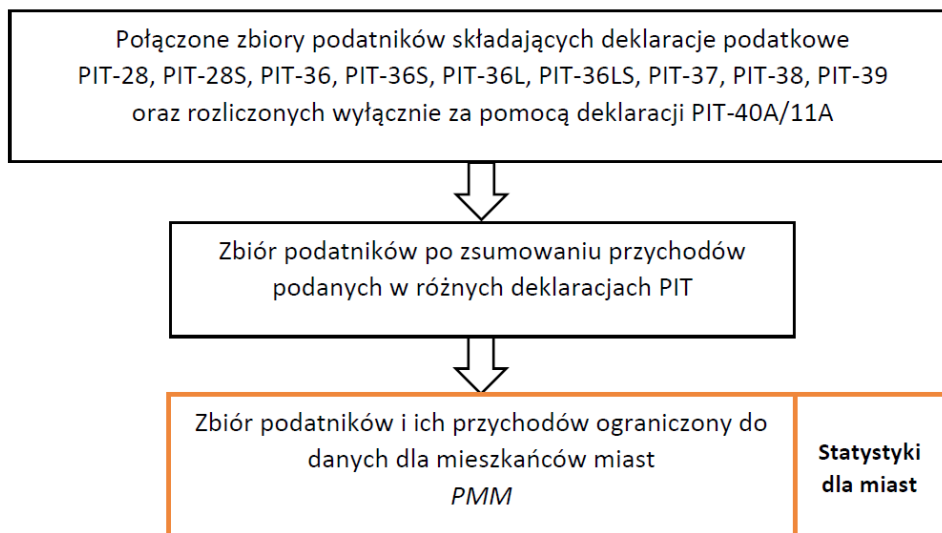
Schemat 3. Procedury przygotowujące bazy PIT, wykonane dla każdego rodzaju deklaracji podatkowej PIT



Źródło: Opracowanie własne.

Po wyborze zmiennych istotnych z punktu widzenia badania, przygotowane w ten sposób dziesięć zbiorów połączono w jedną bazę. Dalsze etapy przygotowania opisuje umieszczony poniżej schemat:

Schemat 4. Procedury przygotowujące bazę PMM



Źródło: Opracowanie własne.

Załącznik 3. Mapy obrazujące wewnętrzne zróżnicowanie sytuacji dochodowej mieszkańców miast wojewódzkich w 2018 r.

Kartogramy zostały udostępnione na stronie internetowej Głównego Urzędu Statystycznego w obszarze Statystyki eksperymentalne.

Załącznik 4. Baza danych dla miast w zakresie wybranych cech (2018 r.)

Baza danych została udostępniona na stronie internetowej Głównego Urzędu Statystycznego w obszarze Statystyki eksperymentalne. Zawiera ona informacje dla:

- miast imiennie, zarówno w zakresie przychodów podatników (mediana), jak i dochodów pracowników najemnych (mediana, relacja decyla dziewiątego do decyla pierwszego P90/P10, wskaźnik zróżnicowania kwintylowego S80/S20, współczynnik Giniego),
- miast wojewódzkich w układzie siatki kwadratów w zakresie dochodów pracowników najemnych (mediana w przedziałach kwartylowych, relacja decyla dziewiątego do decyla pierwszego P90/P10, wskaźnik zróżnicowania kwintylowego S80/S20, współczynnik Giniego).