

Analiza porównawcza rozkładów dochodów w regionie centralnym i wschodnim

Projekt przeprowadzony w ramach przedmiotu:
"Rozkłady dochodów i dobrobytu"

Studia stacjonarne II stopnia na kierunku
Analityka Gospodarcza semestr II 2021/2022

Autorzy:

Radosław Czapp

Emil Filipowicz



Spis treści

Wstęp	3
1. Materiał statystyczny i metody	5
2. Nieparametryczny i parametryczny opis rozkładów	20
3. Nierówności ekonomiczne w rozkładach	24
4. Wyniki analizy ubóstwa ekonomicznego	28
5. Teoretyczne modele rozkładów	30
Podsumowanie i wnioski końcowe	39

Spis tabel i rysunków

Tabela 1 Statystyka parametryczna badanych regionów.	23
Rysunek 1 Histogram rozkładu dochodów ekwiwalentnych w badanych regionach.	20
Rysunek 2 Wykres estymacji funkcji gęstości metodą kernel dla badanych regionów.	21
Rysunek 3 Wykres estymacji funkcji gęstości dla badanych regionów oraz ogółu.	22
Rysunek 4 Wykres krzywej Lorenza dla badanych regionów.	24
Rysunek 5 Wyniki estymacji współczynników Giniego dla badanych regionów.	25
Rysunek 6 Wyniki estymacji współczynników uogólnionej entropii dla badanych regionów.	26
Rysunek 7 Wyniki estymacji współczynników Atkinsona dla badanych regionów.	27
Rysunek 8 Wyniki estymacji miary FGT_0 (HCR) dla badanych regionów z linią ubóstwa na poziomie 500 zł.	28
Rysunek 9 Wyniki estymacji miary FGT_0 (HCR) dla badanych regionów z linią ubóstwa na poziomie 1282,09 zł.	29
Rysunek 10 Wykres estymacji (Q-Q) rozkładu logarytmiczno-normlanego w badanych regionach.	30
Rysunek 11 Wykres estymacji (P-P) rozkładu logarytmiczno-normlanego w badanych regionach.	31
Rysunek 12 Wyniki estymacji uogólnionego rozkładu beta drugiego rodzaju dla regionu centralnego.	32
Rysunek 13 Wyniki estymacji uogólnionego rozkładu beta drugiego rodzaju dla regionu wschodniego.	33
Rysunek 14 Wykres estymacji uogólnionego rozkładu beta drugiego rodzaju dla regionu centralnego.	34
Rysunek 15 Wykres estymacji uogólnionego rozkładu beta drugiego rodzaju dla regionu wschodniego.	34
Rysunek 16 Wyniki estymacji rozkładu Daguma dla regionu centralnego.	35
Rysunek 17 Wyniki estymacji rozkładu Daguma dla regionu wschodniego.	36
Rysunek 18 Wyniki estymacji rozkładu Singha-Maddali dla regionu centralnego.	37
Rysunek 19 Wyniki estymacji rozkładu Singha-Maddali dla regionu wschodniego.	38

Wstęp

W pracy przeprowadzono analizę dwóch rozkładów: dochodów w regionie centralnym (rozkład 1) i dochodów w regionie wschodnim (rozkład 2) w Polsce w 2016 roku.

Celem projektu jest poszukiwanie odpowiedzi na następujące pytania badawcze:

Czy i jak różnią się analizowane rozkłady pod względem:

1. Poziomu dochodów
2. Poziomu nierówności ekonomicznych
3. Poziomu ubóstwa ekonomicznego
4. Formy teoretycznej

Przeprowadzona analiza umożliwia postawienie następujących hipotez:

1. Rozważane rozkłady różnią się pod względem miar położenia, zmienności, asymetrii i koncentracji.
2. Nierówności ekonomiczne w rozkładzie 1 są większe niż w rozkładzie 2.
3. Rozkład 1 wykazuje większą skalę i głębię ubóstwa niż rozkład 2.
4. Rozważane rozkłady charakteryzują się jednakową postacią teoretyczną, ale o różnych parametrach.

Niniejsza praca składa się z pięciu rozdziałów. Pierwszy zawiera ogólną charakterystykę danych statystycznych wykorzystanych w projekcie oraz informuje o metodach opisu nieparametrycznego i parametrycznego rozkładów dochodów. Przedstawione zostały również metody pomiaru nierówności i ubóstwa, a także teoretyczne modele rozkładów dochodów. W drugim rozdziale zaprezentowano nieparametryczny i parametryczny opis rozkładów dochodów, w którym przedstawiono wyniki estymacji funkcji gęstości i parametrów rozkładów. W trzecim rozdziale przedstawiono wyniki pomiaru nierówności ekonomicznych badanych rozkładów dochodów oraz testowania różnic pomiędzy miarami nierówności. Czwarty rozdział zawiera wyniki przeprowadzonej analizy ubóstwa ekonomicznego w regionie centralnym i wschodnim. Ostatni - piąty rozdział prezentuje wyniki dopasowania rozkładu log-normalnego i uogólnionego rozkładu beta drugiego rodzaju.

Motywy podjęcia badań

Głównym motywem podjęcia badań jest analiza zróżnicowania dochodów, sformułowanie wniosków dotyczących wielkości dochodów, ich nierównomierności, poziomu ubóstwa czy bogactwa w wybranych regionach (centralnym i wschodnim). W dotychczas przeprowadzonych badaniach, zróżnicowanie dochodów w wybranych regionów nie było podejmowane lub było podejmowane w szerszym zakresie.

Pierwszym przykładem może być praca dr Beaty Kasprzyk i dr Małgorzaty Leszczyńskiej na temat: "Dochody i wydatki jako determinanty dobrobytu ekonomicznego gospodarstw domowych w Polsce - ujęcie regionalne"¹, w której okres badawczy przypadł na lata 2000-2015 oraz zostało zaprezentowane ogólne ujęcie wszystkich regionów Polski.

Drugim przykładem jest "Zróżnicowanie rozkładów dochodów gospodarstw domowych w rolników w Polsce według regionów"² autorstwa Aliny Jędrzejczak i Doroty Pekasiewicz, gdzie zaprezentowano dane na temat dochodów gospodarstw rolniczych, wzięto pod uwagę wszystkie regiony Polski oraz użyto danych za rok 2015.

Omówienie danych statystycznych i metod badań

W projekcie wykorzystano dane o dochodach gospodarstw domowych w regionach centralnym i wschodnim w Polsce w 2016 roku pochodzące z bazy danych Głównego Urzędu Statystycznego.

Do analizy rozkładów dochodów użyto nieparametrycznych oraz parametrycznych miar statystycznych. W celu pomiaru nierówności ekonomicznych skorzystano z indeksów: Giniego, uogólnionej entropii oraz Atkinsona. Pomiar ubóstwa został przeanalizowany przy wykorzystaniu miar FGT (Fostera-Greera-Thorbecka).

¹ B. Kasprzyk, M. Leszczyńska - "Dochody i wydatki jako determinanty dobrobytu ekonomicznego gospodarstw domowych w Polsce - ujęcie regionalne", Uniwersytet Rzeszowski

²A. Jędrzejczak, D. Pekasiewicz - "Zróżnicowanie rozkładów dochodów gospodarstw domowych w rolników w Polsce według regionów", Zagadnienia Ekonomiki Rolnej

1. Materiał statystyczny i metody

1.1. Ogólna charakterystyka danych statystycznych wykorzystanych w projekcie

Autorzy oświadczają, iż udostępnione nieidentyfikowalne mikrodane o dochodach gospodarstw domowych wykorzystano wyłącznie dla realizacji naukowych celów niniejszego projektu.

Dochód gospodarstwa domowego (inaczej dochód rozporządzalny) stanowi sumę bieżących dochodów danego gospodarstwa domowego pochodzących ze wszystkich źródeł pomniejszoną o zaliczki na podatek dochodowy, inne podatki oraz składki ubezpieczeniowe i zdrowotne. Do podstawowych składników dochodu rozporządzalnego gospodarstwa domowego należą, m.in.:

- Wynagrodzenie za pracę
- Przychody z działalności gospodarczej (prowadzonej na własny rachunek lub z działalności rolniczej)
- Przychody z majątku trwałego
- Transfery z budżetu państwa

Analizie poddane zostały dwa regiony Polski: centralny oraz wschodni. Na region centralny składają się województwa mazowieckie i łódzkie, natomiast region wschodni obejmuje województwa podlaskie, lubelskie, świętokrzyskie i podkarpackie. Ogół danych znajdujących się w projekcie stanowią takie zmienne jak:

- **DDP** – dochód rozporządzalny gospodarstwa domowego
- **LOS** – liczba osób w gospodarstwie domowym
- **MN** – mnożnik GUS
- **LIS** – jednostka ekwiwalentna (pierwiastek z liczby osób w gospodarstwie domowym)
- **DDP_LIS** – dochód ekwiwalentny
- **MN_LOS** – mnożnik GUS pomnożony przez liczbę osób w gospodarstwie domowym

1.2. Metody opisu nieparametrycznego i parametrycznego rozkładów dochodów

Opis nieparametryczny

Metody opisu nieparametrycznego należą do odpornych metod statystycznych, ponieważ są mało wrażliwe na wartości odstające. Powszechnie używane są przy badaniu zmiennych na skali porządkowej (np. jako ocena punktowa filmu od 1 do 5 lub kolejność preferencji klienta).

Metody parametryczne podczas analizy danych zawierających obserwacje odstające, dają błędne wyniki, gdyż obserwacje odstające są z definicji przykładem odstępstwa od założonego rozkładu. Metody nieparametryczne są mniej wrażliwe na takie błędy, ze względu na mniejszą liczbę założeń. W pewnych przypadkach, nawet jeśli użycie metod parametrycznych jest uzasadnione, metody nieparametryczne są prostsze w zastosowaniu.

W przypadku, gdy rozkład jest bardzo zbliżony do założonego, metody parametryczne dają lepsze rezultaty od nieparametrycznych (generują mniejszy błąd i bardziej istotne statystycznie wyniki).

Histogram

Jest najpopularniejszą metodą prezentowania rozkładu empirycznego (otrzymanego na podstawie danych empirycznych) danej cechy. Histogram to inaczej szereg przedziałowy, składający się z przedziałów oraz liczby obserwacji, które się w nich znajdują. Ponadto, służy do analizy zbiorowości i pozwala określić, przedział najliczniejszy (dominanta), symetryczność/asymetryczność, przedziały odstające.

Estymacja funkcji gęstości metodą kernel

Estymator nieparametryczny, służący do określania gęstości rozkładu zmiennej losowej na podstawie uzyskanej próby (wartości, jakie badana zmienna przyjęła podczas pomiarów). Niech x_1, x_2, \dots, x_n będzie próbą losową z rozkładu zmiennej losowej o nieznannej gęstości f . Estymator jądrowy gęstości, wyznaczony na podstawie tej próby, definiujemy wzorem:

$$\hat{f}(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - x_i}{h_n}\right)$$

$n \in N$ - liczność próby

(h_n) – współczynnik wygładzania, spełniający warunki: $h_n \rightarrow 0$ oraz $nh_n \rightarrow \infty$

$K: R \rightarrow [0, \infty)$ – mierzalna funkcja nazywana jądrem, spełniająca: $\int_R K(x)dx = 1$, gdzie K jest symetryczna względem zera oraz posiada w $x = 0$ słabe maksimum globalne.

Te własności zapewniają, między innymi, że dla każdej realizacji próby, estymator jądrowy jest gęstością pewnego rozkładu prawdopodobieństwa. W przypadku pojedynczej obserwacji x_i , funkcja K , przesunięta o wektor x_i oraz wygładzona parametrem h_n , przedstawia szacowany rozkład zmiennej losowej X_i przy założeniu, że przyjęła ona wartość x_i . W przypadku n -elementowej realizacji zmiennej losowej X , rozkład jest szacowany przez sumę obserwacji pojedynczych. Wówczas współczynnik $\frac{1}{nh_n}$ standaryzuje otrzymaną funkcję, aby zachodził warunek: $\int_R \hat{f}(x)dx = 1$ (podstawowe żądanie gęstości rozkładu prawdopodobieństwa).

Opis parametryczny

Metody opisu parametrycznego służą do badania i opisu zbiorowości statystycznych i umożliwiają szybki przegląd danych oraz ich analizę. Dzięki zastosowaniu tych metod, można zapoznać się z pewnymi prawidłowościami w zakresie zjawisk masowych, np. problemy ekonomiczne, demograficzne. Można również korzystać z testowania hipotez, które polega na szacowaniu, czy pewne założenia będą prawdziwe dla badanego zbioru danych. Podczas analizy danych, wynik testowania hipotez jest statystycznie istotny, gdy wyniki nie mogły się zdarzyć losowo. Do metod opisu parametrycznego można zaliczyć:

- **Miary położenia**, służące do określenia położenia zmiennych, stosowane w stosunku do zbiorowości jednorodnej:
 - **Średnia** – liczona jako zwykła średnia arytmetyczna.
 - **Moda** – zwana również dominantą (z uwagi na swą specyfikę), będąca wartością pojawiającą się najczęściej.
 - **Mediana** – zwana również kwantylem rzędu $\frac{1}{2}$, interpretowana jako taka wartość zmiennej losowej, dla której połowa populacji przyjmuje wartości mniejsze bądź

równe, a druga połowa wartości większe bądź równe. Często bywa nazywana wartością środkową.

- **Kwantyle innych rzędów** – Kwantyl rzędu p to taka wartość, że $p * 100\%$ populacji przyjmuje wartości niższe lub równe temu kwantylowi, natomiast $(1-p) * 100\%$ populacji wartości wyższe lub równe.
- **Miary zmienności**, pozwalające na ustalenie zróżnicowania (zmienności) badanej zbiorowości pod względem danej cechy:
 - **Wariancja** – stopień rozrzutu wartości zmiennej losowej wokół wartości oczekiwanej.
 - **Odchylenie standardowe** – określa przeciętne odchylenie wartości zmiennej losowej od jej wartości oczekiwanej - im jest większe, tym większe jest zróżnicowanie w badanej populacji.
 - **Współczynnik zmienności** – stosowany w sytuacji, kiedy porównujemy zróżnicowanie tej samej zmiennej losowej w dwóch lub więcej populacjach.
 - **Rozstęp** – różnica pomiędzy największą, a najmniejszą wartością badanej zmiennej w próbie.
- **Miary asymetrii**, określające czy większa część populacji plasuje się powyżej, czy poniżej przeciętnego poziomu badanej cechy. Asymetrię rozkładu można obliczyć za pomocą współczynnika asymetrii (skośności). Dodatnia wartość tego współczynnika oznacza asymetrię prawostronną (większa część populacji przyjmuje wartości poniżej przeciętnej), natomiast ujemna - lewostronną (większa część populacji przyjmuje wartości powyżej przeciętnej).
- **Miary spłaszczenia**, nazywane miarami koncentracji, określają skoncentrowanie rozkładu wokół wartości oczekiwanej:
 - **Kurtoza** – informuje na ile obserwacje są skoncentrowane wokół średniej.
 - **Eksces** – określa, czy koncentracja wartości zmiennej badanej wokół średniej w danym rozkładzie jest większa czy mniejsza w porównaniu do zbiorowości o rozkładzie normalnym.

1.3. Metody pomiaru nierówności ekonomicznych

Nierówność jest znacznie szerszym pojęciem aniżeli ubóstwo, ponieważ definiowana jest w odniesieniu do badanej populacji - nie koncentruje się wyłącznie na jednostkach ubogich. Rozkład dochodów, w którym wszystkie gospodarstwa domowe dysponują takim samym dochodem nazywany jest rozkładem egalitarnym, czyli pozbawionym nierówności. Podczas pomiaru nierówności badany jest stopień odchylenia rozkładu dochodów od rozkładu egalitarnego. Do własności miar nierówności należą następujące aksjomaty:

- **Aksjomat o monotoniczności** – zmniejszenie dochodu ubogiego gospodarstwa domowego, zwiększa wartość indeksu.
- **Aksjomat o transferze** – transfer dochodu z ubogiego gospodarstwa domowego do uboższego powoduje spadek wartości indeksu i odwrotnie.
- **Aksjomat o stałości skali** – wartość indeksu nie ulega zmianie, gdy takim samym proporcjonalnym zmianom ulegają dochody gospodarstw domowych oraz granica ubóstwa.
- **Aksjomat o koncentracji na ubóstwie** – wartość indeksu nie zmienia się wraz ze wzrostem dochodów nie ubogich gospodarstw domowych.
- **Aksjomat o wrażliwości transferu** – wpływ transferu dochodów z ubogiego gospodarstwa domowego do uboższego na wzrost wartości indeksu. Przy stałej wysokości transferu, wpływ jest większy im wyższy jest poziom dochodów gospodarstwa, z którego dokonano transferu.
- **Aksjomat o dekomponowalności** – wartość indeksu badanej populacji powinna być dekomponowalna ze względu na podpopulacje (ogólny wskaźnik ubóstwa powinien być możliwy do obliczenia jako średnia ważona ze wskaźników dla podpopulacji).
- **Aksjomat o zgodności w podpopulacjach** – wartość indeksu dla całej badanej populacji ulegnie zmniejszeniu, jeśli wartość indeksu o identycznej formule zmaleje dla jednej z podpopulacji przy braku zmian jego wartości w pozostałych podpopulacjach.
- **Aksjomat o symetryczności** – wymiana dochodów pomiędzy dowolną parą gospodarstw domowych nie powinna powodować zmian wartości indeksu.

- **Aksjomat o łatwości interpretacyjnej** – indeks powinien posiadać jasną i klarowną interpretację ekonomiczną.
- **Aksjomat o unormowaniu** – istnieje dolna granica wartości współczynnika równa zero, którą współczynnik przyjmuje w przypadku rozkładu egalitarnego.
- **Aksjomat o stałości replikacji** – wartość współczynnika nie ulega zmianie przy dowolnej liczbie replikacji populacji badanej.

Funkcja Lorenza

Argumentami funkcji Lorenza są skumulowane odsetki gospodarstw domowych uporządkowane według dochodów niemalejących, a wartościami – odsetki ich dochodów. Przyjmujemy, że dochód ekwiwalentny gospodarstwa domowego jest dodatnią zmienną losową ciągłą Y^e , o funkcji gęstości $f(y^e)$ oraz dystrybuancie $F(y^e) = P(Y^e \leq y^e)$. Dystrybuanta $F(y^e)$ reprezentująca skumulowaną frakcję gospodarstw domowych w badanej populacji, których dochód ekwiwalentny jest nie większy niż y^e , ma postać:

$$F(y^e) = \int_0^{y^e} f(y^e) dy^e$$

Przyjmując, że μ jest wartością przeciętną zmiennej Y^e , można zdefiniować funkcję $F(y^e)$ wskazującą jaki udział w sumie dochodów całej zbiorowości gospodarstw domowych mają łączne dochody ekwiwalentne tych gospodarstw, których dochody ekwiwalentne nie przekraczają poziomu y^e :

$$F(y^e) = \frac{1}{\mu} \int_0^{y^e} y^e f(y^e) dy^e$$

Funkcja Lorenza rośnie monotonicznie wraz ze wzrostem dochodów ekwiwalentnych oraz jest funkcją wypukłą. Kumulacja frakcji gospodarstw domowych oraz kumulacja ich dochodów ekwiwalentnych prowadzi do liniowej funkcji Lorenza, co oznacza, że p procent gospodarstw domowych w badanej populacji uzyskuje dokładnie p procent łącznego dochodu ekwiwalentnego populacji. Funkcja ta obrazuje fakt występowania lub braku występowania nierówności dochodowych oraz odzwierciedla stopień nierówności dochodowych. Im bardziej funkcja Lorenza odchyła się od rozkładu egalitarnego, tym badany rozkład dochodów jest bardziej nierównomierny.

Indeks Giniego

Nazywany wskaźnikiem nierówności społecznej, stosowny jest do liczbowego wyrażania nierównomiernego rozkładu dóbr, zwłaszcza dochodu. Skala współczynnika wyrażona jest w procentach i zawarta w przedziale od 0 (wszyscy ludzie uzyskują takie same dochody) do 100 (jedno gospodarstwo koncentruje cały dochód). Indeks Giniego można zdefiniować następująco:

$$G = \frac{\Delta}{2\mu}$$

$$\Delta = \frac{1}{n^2} \sum_{i=1}^n \sum_{i'=1}^n |y_i^e - y_{i'}^e|$$

n – liczba gospodarstw domowych

$y_i^e, y_{i'}^e$ – wartości średnich dochodów ekwiwalentnych

Podwojona wartość współczynnika Giniego określa jaką frakcję przeciętnych dochodów ekwiwalentnych μ stanowi przeciętna różnica bezwzględna pomiędzy dochodami ekwiwalentnymi pary losowo wybranych gospodarstw domowych.

Indeks uogólnionej entropii

Wывodząca się z teorii informacji (miara redundancji danych interpretowana jako nielosowość lub kompresja danych) mierzy nierówności dochodów w populacji. Ogólną miarę opartą na pojęciu entropii można zdefiniować:

$$GE(\alpha) = \begin{cases} \frac{1}{N\alpha(\alpha-1)} \sum_{i=1}^N \left[\left(\frac{y_i}{\bar{y}} \right)^\alpha - 1 \right], & \alpha \neq 0, 1, \\ \frac{1}{N} \sum_{i=1}^N \frac{y_i}{\bar{y}} \ln \frac{y_i}{\bar{y}}, & \alpha = 1, \\ -\frac{1}{N} \sum_{i=1}^N \ln \frac{y_i}{\bar{y}}, & \alpha = 0. \end{cases}$$

Wartość miary: $GE \in [0, +\infty)$, w przypadku rozkładu egalitarnego przyjmuje wartość 0. Wzrost nierównomierności rozkładu dochodów powoduje wzrost wartości tej miary. Parametr α może być dowolną liczbą rzeczywistą reprezentującą wagę nadawaną różnicy pomiędzy

dochodami w różnych częściach rozkładu dochodów. Dla mniejszych wartości α miara GE jest bardziej wrażliwa na zmiany dochodów w dolnej części rozkładu. Dla wyższych wartości wzrasta wrażliwość miary GE w górnej części rozkładu.

Indeks Atkinsona

Na podstawie tego wskaźnik można ustalić, który koniec przedziału rozkładu dochodów miał najistotniejszy wpływ na zaobserwowane nierówności. Wartość tego współczynnika określa o jaki odsetek aktualna suma dochodów ekwiwalentnych badanej populacji gospodarstw domowych mogłaby być mniejsza, gdyby ich rozkład był równomierny, aby osiągnęły ten sam poziom zamożności jaki posiadają przy aktualnych dochodach ekwiwalentnych.

Współczynnik Atkinsona możemy zdefiniować:

$$A = 1 - \left\{ \frac{1}{n} \sum_{i=1}^n \left(\frac{y_i^e}{\bar{y}^e} \right)^{1-\epsilon} \right\}^{\frac{1}{1-\epsilon}}$$

ϵ – współczynnik awersji do nierówności

Współczynnik awersji do nierówności ϵ reprezentuje wagę nadawaną transferom dochodów pomiędzy różnymi częściami rozkładu dochodów. Wraz ze wzrostem tego współczynnika przywiązujemy coraz większą wagę do transferów dochodów do dolnej części ich rozkładu. W sytuacji ekstremalnej, gdy $\epsilon \rightarrow \infty$ bierzemy pod uwagę wyłącznie transfery do bardzo nielicznej grupy gospodarstw domowych o najniższych dochodach. Jeżeli $\epsilon = 0$ wówczas wszystkim transferom dochodów nadawana jest jednakowa waga.

1.4. Metody pomiaru ubóstwa ekonomicznego

Wybór linii ubóstwa

Wśród najczęściej wykorzystywanych linii ubóstwa rozróżnia się te, ustalane metodami zobiektywizowanymi oraz subiektywnymi (na podstawie osądów badanych osób czy rodzin).

Owe zobiektywizowanie polega na tym, że przyjmowane (w konstrukcji tych mierników) założenia o charakterze wartościującym tylko po części mają uzasadnienie i umocowanie naukowe, po części zaś noszą charakter subiektywnych wyborów (dokonywanych przez badaczy i polityków społecznych). Mierniki określane jako obiektywne można podzielić na te, które wyznaczają bezwzględne granice ubóstwa oraz te, które mają charakter względny.

W konstruowaniu subiektywnych metod wyznaczania linii ubóstwa za podstawę przyjmowane są opinie badanych osób (respondentów badań reprezentacyjnych) w kwestii wysokości dochodu niezbędnego do zaspokojenia minimum potrzeb egzystencjalnych i społecznych badanej jednostki.

Linie ubóstwa ustalane metodami zobiektywizowanymi

W rzeczywistości większość potrzeb zaspokajanych jest w obrębie gospodarstwa domowego, które stanowi podstawowy podmiot, dla którego budowane są koszyki towarów i usług oraz warunków mieszkaniowych. W konstrukcji koszyka – zarówno minimum egzystencji, jak i minimum socjalnego – określane są potrzeby wszystkich osób w gospodarstwie domowym, zróżnicowane ze względu na ich wiek i poziom aktywności fizycznej, a także potrzeby gospodarstwa domowego bezpośrednio niezależne od jego liczebności. Przy takim rozwiązaniu nie stosuje się skal ekwiwalentności. Skale ekwiwalentności wynikające z porównania wartości koszyków minimum socjalnego opracowanych dla gospodarstw domowych o różnych cechach demograficznych i społecznych, żyjących w warunkach połowy lat 90. niewiele różnią się od stosowanych w krajach OECD; natomiast skale ekwiwalentności koszyków minimum egzystencji – różnią się znacznie.

W pierwszej połowie lat 80. dostosowano koszyk do sytuacji narastającego kryzysu gospodarczego i pogarszających się warunków życia, nazywając go kryzysowym. Od połowy lat 80. stosowane w konstrukcji tego koszyka normy w zasadzie nie ulegały zmianie. Natomiast dokonywano okresowej waloryzacji minimum socjalnego na podstawie informacji statystycznych

o zmianach cen towarów i usług konsumpcyjnych. Koszyk minimum egzystencji opracowany został po raz pierwszy dla warunków rynkowych 1994 roku, a jego wycena i waloryzacja są kontynuowane systematycznie. Od 1994 r. zmieniono także koszyk minimum socjalnego, dostosowując go do warunków rynkowych i zasad funkcjonowania gospodarstw domowych w nowym systemie gospodarki rynkowej. Wśród metod pomiaru ubóstwa ekonomicznego wyznaczanych metodami zobiektywizowanymi wyróżnić można:

- a) **Minimum egzystencji** – zakres i poziom zaspokajanych potrzeb według standardu minimum egzystencji wyznacza granicę, poniżej której występuje biologiczne zagrożenie życia oraz psychofizycznego rozwoju człowieka. Jeżeli bieżący dochód gospodarstwa domowego nie może zapewnić konsumpcji wyznaczonej tym standardem, to wówczas zagrożona zostaje trwałość życia, przez utratę możliwości zachowania zdrowia i zdolności do pracy.
- b) **Minimum socjalne** – Koszyk towarów i usług według standardu minimum socjalnego jest tak skonstruowany, aby na każdym etapie rozwoju człowieka zapewniał mu warunki bytowania, które umożliwiają reprodukcję jego sił życiowych, posiadanie i wychowanie potomstwa oraz utrzymywanie więzi społecznych w procesie nauki, pracy i wypoczynku.
- c) **Linie według standardu europejskiego** – W badaniach prowadzonych przez Eurostat przyjmuje się trzy linie ubóstwa. Stanowią one 40%, 50% i 60% przeciętnych wydatków na jednostkę ekwiwalentną przy zastosowaniu skali ekwiwalentności OECD. W Polsce w badaniach prowadzonych przez GUS stosuje się min. linię ubóstwa stanowiącą 50% przeciętnych wydatków gospodarstw domowych, w przeliczeniu na jednostkę ekwiwalentną wg skali OECD.
- d) **Linia ubóstwa jako kryterium świadczeń pomocy społecznej w Polsce** – W ustawie obowiązującej do końca 1996 r. funkcję linii ubóstwa pełniła najniższa emerytura (o zasiłek z pomocy społecznej mogła się ubiegać osoba, której dochód nie przekroczył tej granicy). Później granica ta została określona kwotowo na osobę samotną i/lub rodzinę (przy zastosowaniu skali ekwiwalentności OECD) i podlega corocznej waloryzacji. Jednakże obok tego kryterium dochodowego (znacznie niższego od minimum socjalnego a zbliżonego do minimum egzystencji) obowiązują także inne, charakteryzujące społeczne położenie świadczeniobiorcy.

Linie ubóstwa ustalane metodami subiektywnymi

Podstawą do ustalania linii ubóstwa o charakterze subiektywnym są wyniki badań ankietowych polegających na gromadzeniu opinii na temat wysokości dochodu zabezpieczającego minimum potrzeb rodziny bądź zjawiska biedy w skali społecznej, wraz z oceną własnej sytuacji materialnej. Aktualnie w Polsce badania tego typu są prowadzone przez CBOS, Instytut Filozofii i Socjologii PAN, a w ostatnich latach także przez GUS i SGH.

Przy ustalaniu subiektywnej linii ubóstwa istotne znaczenie mają zarówno pytania zadawane respondentom, jak i metody przetwarzania uzyskanych odpowiedzi. Przykładowo w metodzie lejdejskiej istotną rolę spełnia tzw. dochodowa funkcja dobrobytu (wyrażająca zależność osiąganą przez gospodarstwo domowe satysfakcji z konsumpcji (dobrobytu) od wysokości dochodu tego gospodarstwa). Parametry tej funkcji są szacowane na podstawie wyników badań ankietowych gospodarstw domowych na temat ich oceny różnych poziomów dochodu. Ustalana arbitralnie pewna niska wartość dochodowej funkcji dobrobytu daje z kolei podstawę do wyznaczenia wysokości dochodu stanowiącego subiektywną linię ubóstwa.

Miary FGT (Fostera-Greera-Thorbecka)

Miary FGT to zestaw wskaźników pozwalających określić poziom ubóstwa w obrębie danej grupy społecznej lub w określonym obszarze geograficznym. Zostały opracowane przez Jamesa Fostera, Joela Greera oraz Erika Thorbecke'a w 1984 roku. Wzór ogólny można zdefiniować następująco:

$$FGT_{\alpha} = \frac{1}{N} \sum_{i=1}^H \left(\frac{z - y_i}{z} \right)^{\alpha}$$

z – granica ubóstwa

N – wielkość całej populacji

H – liczba biednych osób (uzyskujących dochody poniżej poziomu z)

y – dochód jednostki

Jeśli współczynnik α jest niski, wówczas wskaźnik FGT nadaje wszystkim gospodarstwom domowym o dochodach poniżej poziomu z taką samą wagę. Im wyższa wartość tego współczynnika, tym większy nacisk kładziony jest na najbiedniejsze gospodarstwa domowe. Ogólnie im wyższa jest wartość indeksu FGT, tym większe występuje ubóstwo w badanej populacji.

W przypadku, gdy $\alpha = 0$, wówczas mamy do czynienia ze wskaźnikiem HCR (Head count ratio), który sprowadza się do podzielenia liczby ludzi żyjących poniżej linii ubóstwa przez całkowitą liczebność populacji.

$$FGT_0 = \frac{H}{N}$$

Dla $\alpha = 1$, wskaźnik zostaje zredukowany do wskaźnika luki ubóstwa, będącego miarą natężenia ubóstwa, który szacuje głębokość ubóstwa, biorąc pod uwagę jak daleko, badane jednostki znajdują się od granicy ubóstwa.

$$FGT_1 = \frac{1}{N} \sum_{i=1}^H \left(\frac{z - y_i}{z} \right)$$

Podczas, gdy $\alpha = 2$, podobnie jak w powyższym przypadku, większa waga przywiązywana jest do gospodarstw domowych znajdujących się daleko od granicy ubóstwa, lecz w sposób bardziej proporcjonalny. Przykładowo, gospodarstwo domowe o dochodach dwukrotnie niższych o pozostałych będzie miało ponad dwukrotnie większą wagę w stopie ubóstwa (zakładając, że dochody tych gospodarstw są poniżej granicy ubóstwa). Wówczas wzór przyjmuje następującą postać:

$$FGT_2 = \frac{1}{N} \sum_{i=1}^H \left(\frac{z - y_i}{z} \right)^2$$

Możliwe są również inne dekompozycje miar FGT. Do najczęściej wykorzystywanych należy Indeks Sen – zaproponowany w 1976 roku, oparty na różnych typach wskaźników FGT. Definiowany jest w celu uwzględnienia trzech czynników: zapadalności, intensywności oraz nierówności. Zatem wskaźnik ten nadaje większą lub mniejszą wagę jednostkom w zależności od ich rangi w rozkładzie dochodów.

$S = FGT_0 \cdot G + FGT_1 \cdot (1 - G)$, gdzie G jest indeksem Giniego stosowanym dla populacji znajdujących się poniżej linii ubóstwa.

Kolejnym wskaźnik jest **Indeks Wattsa**, w którym zastosowano inny system ważenia:

$$W = \frac{1}{N} \sum_{i=1}^N \ln \frac{z}{y_i}$$

Umożliwia on obliczenie średniego czasu wyjścia z ubóstwa, definiowane następująco:

$$T = \frac{W}{g}$$

T – czas wyjścia z ubóstwa

W – wskaźnik Wattsa

g – oczekiwana stopa wzrostu PKB

Jednak jedną z głównych wad tego wskaźnika jest użycie logarytmu we wzorze: oznacza to, że nie ma on zastosowania w momencie, gdy dochód jest ujemny lub zerowy.

1.5. Teoretyczne modele rozkładów dochodów

Dobry rozkład teoretyczny powinien dostosowywać się do zmian w kształcie rozkładów empirycznych poprzez zmiany wielkości parametrów, wtedy będzie dobrze opisywał rozkłady w różnych przekrojach i okresach. Innym problemem jest możliwość odzwierciedlenia dochodów zerowych i ujemnych.

Większość znanych rozkładów teoretycznych może być stosowana jedynie dla dochodów dodatnich – wprowadzenie dochodów ujemnych i zerowych wymaga dodatkowych zabiegów. Do wyjątków należy rozkład Daguma typu II. Nie mniej istotnym zagadnieniem jest dopasowywanie rozkładu teoretycznego do sytuacji, w których minimalny dochód jest większy od zera. Powinno być to możliwe przez zmiany w parametrach rozkładu bez konieczności „ucinania go” (przykładem może być rozkład Daguma typu III lub trójparametrowy rozkład logarytmiczno-normalny).

Za jedną z cech teoretycznych rozkładu prawdopodobieństwa, która może być przydatna w analizie płac i dochodów, uznaje się możliwość przejścia w rozkład zeromodalny. W praktyce zdarzają się sytuacje, kiedy rozkłady dochodów nie osiągają w żadnym punkcie maksimum. Może to mieć miejsce, gdy rozważamy rozkłady dochodów krajów rozwijających się o dużej liczbie ludności, mających znaczny udział zatrudnionych w rolnictwie lub w przypadku krajów z niewielkim odsetkiem osób bardzo bogatych i dużym odsetkiem bardzo biednych. Warto również dodać, że rozkłady zamożności są zawsze zeromodalne. Do najczęściej stosowanych teoretycznych modeli rozkładów dochodów należą:

- a) Rozkład Pareto** – charakteryzuje się ciągłym rozkładem prawdopodobieństwa, spełniając przy tym warunek potęgowego prawa skalowania. Oryginalnie używany był do przedstawienia rozmieszczenia dóbr w społeczeństwie, głosi, że większa część bogactwa w danej populacji posiadana jest przez niewielki procent członków w niej występujących. Rozkład Pareto wyrażany jest także jako zasada Pareto, mówiąca o tym, że 20% danej populacji posiada 80% bogactwa w niej zawartej.
- b) Rozkład logarytmiczno-normalny** – jest ciągłym rozkładem prawdopodobieństwa dodatniej zmiennej losowej, której logarytm cechuje się rozkładem normalnym. Znajduję zastosowanie w statystyce, ze względu na fakt, że wiele zmiennych występujących naturalnie w zastosowaniach jest nie ujemnych. W tym przypadku funkcja gęstości ma następującą postać:

$$f_{L-N} = \frac{1}{\sigma x \sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right)$$

μ – parametr położenia, σ – parametr kształtu

c) Rozkład Daguma – powstał na podstawie obserwacji empirycznych i jest jednym z najczęściej stosowanych modeli do opisu rozkładów płac i dochodów. Rozkład ten może być zarówno jednomodalny, jak i zeromodalny. W przypadku wysokich wartości dochodów rozkład Daguma jest zbliżony z rozkładem Pareto, dzięki czemu posiada niewielką liczbę skończonych momentów, co sprawia, że charakteryzuje się wysoką zgodnością z danymi empirycznymi.

d) Rozkład Singha-Maddali – funkcja gęstości rozkładu Singha-Maddali ma następującą postać:

$$f_{S-M}(x) = \alpha \beta^{-1} k \left(\frac{x}{\beta}\right)^{\alpha-1} \left(1 + \left(\frac{x}{\beta}\right)^{\alpha}\right)^{-k-1} \text{ dla } x > 0, \text{ gdzie } \alpha, \beta, k - \text{parametry rozkładu}$$

Wartość średnia i odchylenie standardowe dane są wzorami:

$$\mu_{S-M} = \beta \Gamma\left(\frac{1+\alpha}{\alpha}\right) \Gamma\left(\frac{k\alpha-1}{\alpha}\right) \Gamma^{-1}(k)$$

$$\sigma_{S-M} = \frac{\beta}{\Gamma(k)} \sqrt{\Gamma\left(\frac{\alpha+2}{\alpha}\right) \Gamma(k) \Gamma\left(\frac{k\alpha-2}{\alpha}\right) - \Gamma^2\left(\frac{\alpha+1}{\alpha}\right) \Gamma^2\left(\frac{k\alpha-1}{\alpha}\right)}$$

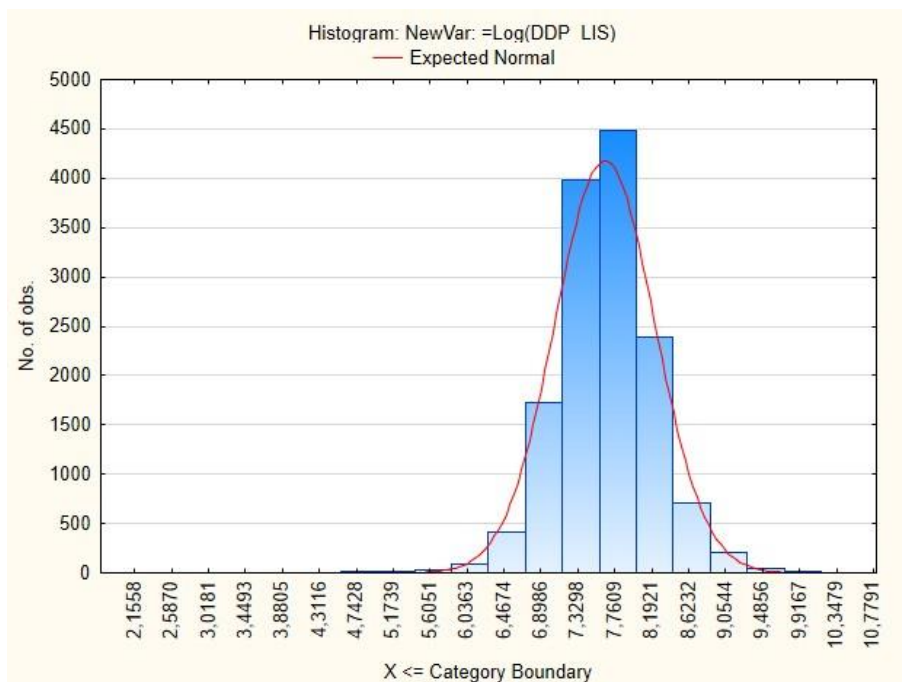
Gdzie Γ – funkcja gamma

e) Rozkład Fiska – log-logistyczny, ciągły rozkład prawdopodobieństwa. Służy do badania czasu trwania określonego zdarzenia, którego intensywność najpierw wzrasta, w późniejszych okresach maleje. W ekonomii służy do modelowania nierówności dochodów.

f) Rozkład Zengi – skonstruowany w 2010 roku teoretyczny model opisu rozkładów za pomocą trzyparametrowej funkcji gęstości, dla nieujemnych wartości dochodów. Powstał w oparciu o rozkład Pareto. Funkcja gęstości rozkładu Zengi przyjmuje różne kształty, bardziej różnorodne niż tradycyjne trójparametrowe modele rozkładów dochodów, co pozwala na dobre dopasowanie także w przypadku niskich grup dochodowych.

2. Nieparametryczny i parametryczny opis rozkładów

Histogram rozkładu dochodów ekwiwalentnych w badanych regionach

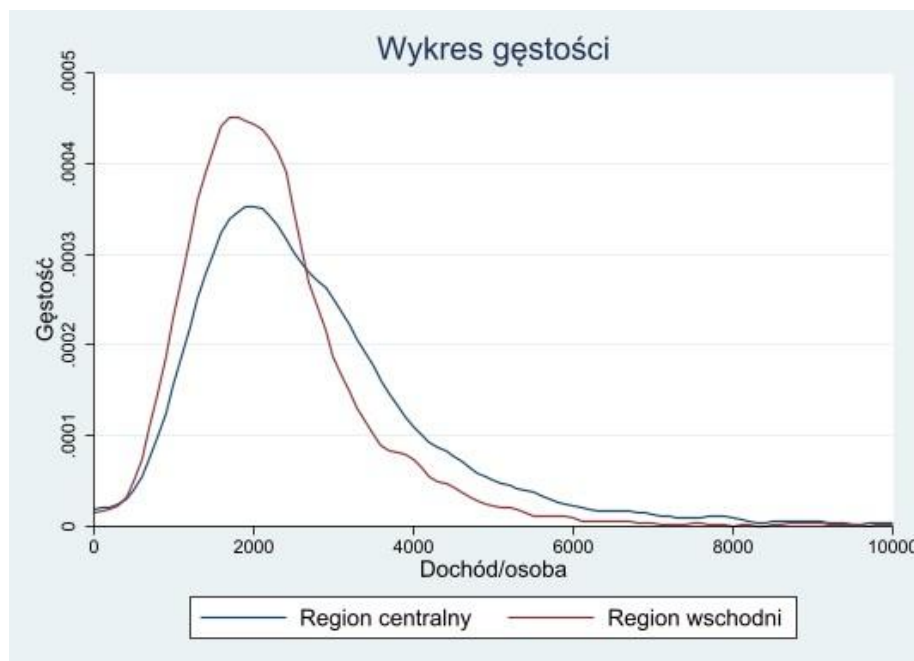


Rysunek 1 Histogram rozkładu dochodów ekwiwalentnych w badanych regionach.

Histogram wykonany na potrzeby analizy prezentuje rozkłady dochodów w regionie centralnym oraz wschodnim. Na rzędnych przedstawiono liczbę gospodarstw domowych, natomiast na osi odciętych zlogarytmowaną wartość dochodów, aby zminimalizować wpływ wartości odstających na wynik analizy. Rozkład ten jest zbliżony do rozkładu normalnego, co informuje, że większość dochodów gospodarstw domowych skupiona jest wokół najczęściej występujących wartości danego przedziału. Wynika z tego, że w badanej populacji niespełna 4500 gospodarstw domowych zadeklarowało wysokość dochodów na przeciętnym poziomie – co odpowiada dochodom w okolicach 2400 zł.

Estymacja funkcji gęstości metoda kernel

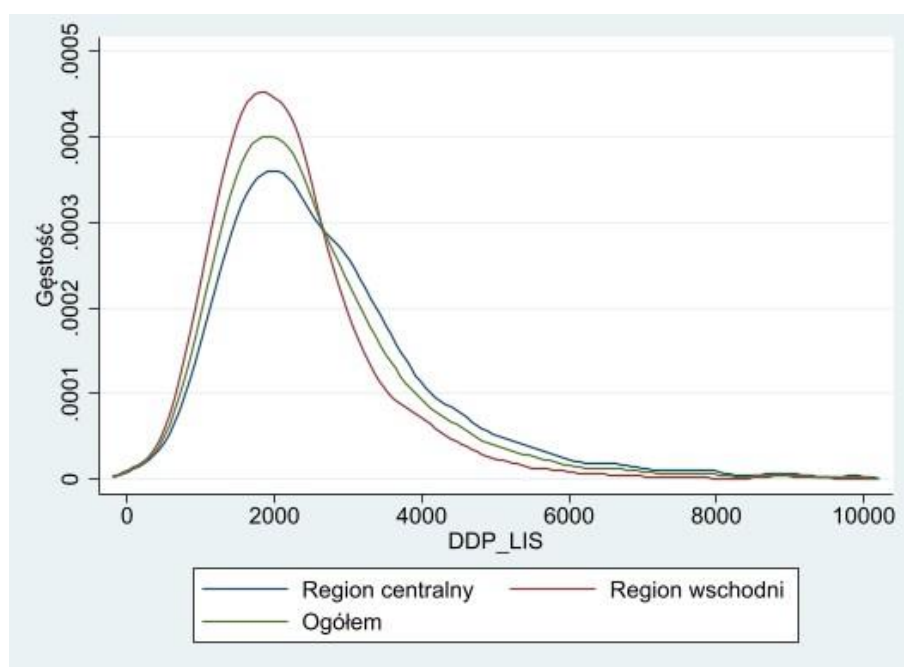
Po przeprowadzeniu estymacji, wykorzystując funkcję gęstości metodą kernel, otrzymano zaprezentowany poniżej rozkład gęstości oddzielnie dla obu badanych regionów. Warto zauważyć, że wartość parametru szerokości pasma dla opracowanych estymacji została przyjęta na poziomie 200.



Rysunek 2 Wykres estymacji funkcji gęstości metodą kernel dla badanych regionów.

Z powyższego wykresu można wywnioskować, że w przypadku obu regionów mamy do czynienia z rozkładem asymetrycznym, prawostronnym, dodatnim. Z rozkładu gęstości wynika, że w regionie wschodnim dochody ekwiwalentne koncentrują się głównie na poziomie ok. 2000 zł. Zauważalne jest również, że liczba osób posiadająca dochody ekwiwalentne w przedziale 4000-6000 zł drastycznie maleje, a powyżej 6000 zł liczba osób jest bliska zeru. Świadczy to o tym, że w regionie wschodnim mniejsza liczba gospodarstw domowych osiąga wyższe dochody.

Natomiast, w regionie centralnym można zauważyć podobną tendencję koncentracji dochodu ekwiwalentnego na poziomie 2000 zł, aczkolwiek analizując prawą stronę rozkładu, większa liczba gospodarstw domowych charakteryzuje się wyższym poziomem dochodu ekwiwalentnego niż w regionie wschodnim.



Rysunek 3 Wykres estymacji funkcji gęstości dla badanych regionów oraz ogółu.

Dodatkowo, w celu porównania, autorzy uzupełnili wykres gęstości o rozkład dochodu ekwiwalentnego dla całej badanej populacji, co ilustruje powyższy wykres.

Metody opisu parametrycznego

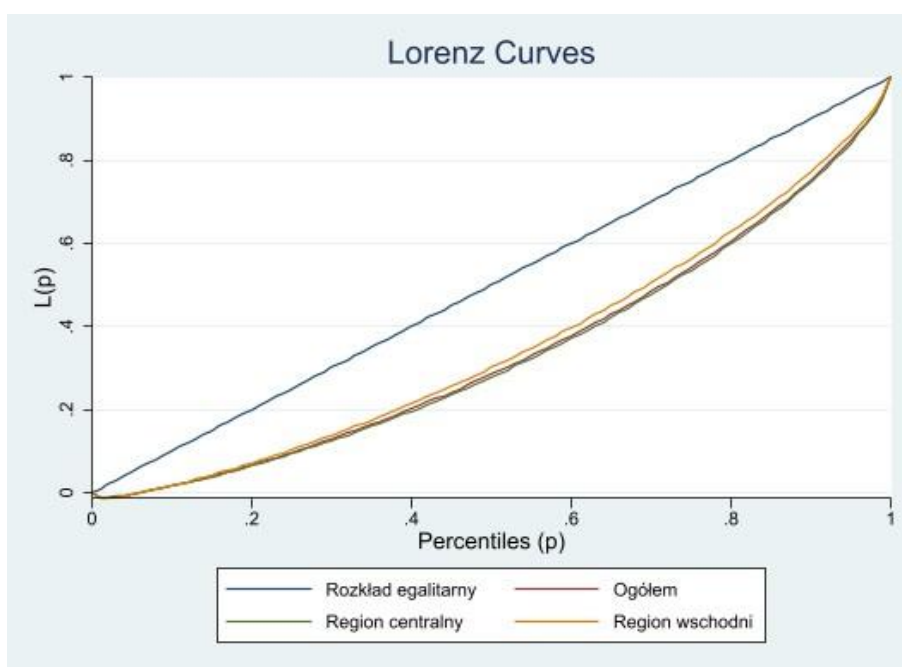
Parametr	Region centralny	Region wschodni	Ogółem
Średnia	2654.320206	2099.220632	2410.943797
Mediana	2267.670013	1893.397114	2082.866152
Dominanta	2500	2500	2500
Odchylenie standardowe	2134.779713	1528.002171	1912.720716
Wariancja próbki	4557284.423	2334790.635	3658500.538
Kurtoza	66.41451478	454.0357321	140.7649332
Skośność	0.833835843	-8.903036459	-1.209118295
Minimum	-46478.1	-60155.46	-60155.46
Maksimum	32308.76	21743.17997	32308.76
Suma	21367277.66	13193601.67	34560879.33
Licznik	8050	6285	14335
Poziom ufności [95%]	46.64107261	37.78357773	31.31392286

Tabela 1 Statystyka parametryczna badanych regionów.

3. Nierówności ekonomiczne w rozkładach

Funkcja Lorentza

Krzywa Lorentza dla badanych obszarów ilustruje, że w regionie wschodnim rozkład dochodów ekwiwalentnych jest bardziej równomierny niż w regionie centralnym, ponieważ znajduje się bliżej rozkładu egalitarnego, co wskazuje, że w tym regionie występują mniejsze rozbieżności w uzyskiwanych dochodach ekwiwalentnych w gospodarstwach domowych. Rozkład dochodów ekwiwalentnych w gospodarstwach domowych w regionie centralnym wykazuje, że w tym regionie występują większe rozbieżności w dochodach.



Rysunek 4 Wykres krzywej Lorentza dla badanych regionów.

Ze względu na fakt, że w niektórych punktach krzywe rozkładów wzajemnie się nakładają – w tym przypadku lepiej byłoby zastosować indeks Giniego, który znacznie bardziej szczegółowo zaprezentuje różnice pomiędzy rozporządzalnymi dochodami ekwiwalentnymi w gospodarstwach domowych.

Indeks Giniego

W celu analizy porównawczej warto zauważyć, iż w 2016 roku wartość indeksu Giniego dla Polski wynosiła 0,312³. Dla regionu wschodniego wartość indeksu wyniosła 0,299574, co wskazuje, że rozbieżności w dochodach ekwiwalentnych w tym regionie są znacznie mniejsze niż w przypadku całego kraju. Potwierdza to również wnioski wysunięte podczas analizy krzywej Lorenza, mówiące o występowaniu mniejszych rozbieżności w uzyskiwanych dochodów ekwiwalentnych w gospodarstwach domowych w regionie wschodnim.

W regionie centralnym wartość indeksu Giniego wynosiła 0,334504, co świadczy o większych rozbieżnościach w dochodach ekwiwalentnych w tym regionie w porównaniu do regionu wschodniego oraz całego kraju.

Index	:	Gini index				
Household size	:	los				
Group variable	:	reg				

	Group		Estimate	STE	LB	UB

	1		0.334504	0.005497	0.323729	0.345279
	3		0.299574	0.009773	0.280419	0.318730

Population			0.325348	0.005180	0.315195	0.335502

Rysunek 5 Wyniki estymacji współczynników Giniego dla badanych regionów.

³ Na podstawie danych Banku Światowego dostępnych pod linkiem: [Gini index \(Poland\)](#)

Indeks uogólnionej entropii

Na potrzeby analizy, przyjęto wartość parametru teta na poziomie 1, co de facto oznacza korzystanie z indeksu Theila, który służy do pomiaru nierówności społecznych – w tym również dochodów – jako maksymalna miara entropii. Z przeprowadzonej analizy wynika, że rozkład dochodów ekwiwalentnych w gospodarstwach domowych w regionie wschodnim jest bardziej zbliżony do rozkładu egalitarnego niż ma to miejsce w przypadku regionu centralnego Polski.

Index	:	Entropy index
Parameter theta	:	1
Household size	:	lis
Group variable	:	reg

Group	Estimate	STE	LB	UB
1	0.179521	0.005768	0.168215	0.190827
3	0.134237	0.005482	0.123492	0.144982
Population	0.168049	0.004286	0.159649	0.176450

Rysunek 6 Wyniki estymacji współczynników uogólnionej entropii dla badanych regionów.

Indeks Atkinsona

W przeprowadzonej analizie wartość współczynnika awersji do nierówności przyjęto na poziomie 2,9 – co reprezentuje wagę nadawaną transferom dochodów. Na podstawie tego wskaźnika można ustalić, że w przypadku regionu wschodniego (gdybyśmy mieli do czynienia z rozkładem równomiernym) gospodarstwa domowe osiągnęłyby taki sam poziom życia i zamożności przy 25% aktualnej sumy dochodów. Natomiast w przypadku regionu centralnego odsetek ten wynosi niespełna 27%.

Dodatkowo wartości uzyskane z estymacji indeksu Atkinsona ukazują, że badane regiony Polski charakteryzują się znacznymi rozbieżnościami w uzyskanych dochodach ekwiwalentnych w gospodarstwach domowych.

Index	:	Atkinson index			
Parameter epsilon	:	2.9			
Household size	:	lis			
Group variable	:	reg			
<hr/>					
Group		Estimate	STE	LB	UB
<hr/>					
1		0.730511	0.049464	0.633556	0.827466
3		0.750703	0.073993	0.605667	0.895740
<hr/>					
Population		0.747068	0.048325	0.652345	0.841791
<hr/>					

Rysunek 7 Wyniki estymacji współczynników Atkinsona dla badanych regionów.

4. Wyniki analizy ubóstwa ekonomicznego

Z opisanych w pierwszym rozdziale metod pomiaru ubóstwa ekonomicznego do badań określonego w projekcie problemu autorzy wybrali metodę miar FGT. Na potrzeby analizy oraz w celu obliczenia wskaźnika HCR (Head count ratio), wartość współczynnika α przyjęto na poziomie równym 0.

W pierwszym przypadku linia ubóstwa została wytyczona na poziomie 500 zł dochodu ekwiwalentnego, co mniej więcej odpowiada poziomowi minimum egzystencji. W regionie wschodnim można zauważyć, że 2,27% gospodarstw domowych egzystuje poniżej wyznaczonej linii ubóstwa. Natomiast, w regionie centralnym odsetek ten oscyluje wokół wartości 2,43%, co świadczy, że w obrębie województw mazowieckiego i łódzkiego większa liczba gospodarstw domowych boryka z problemami ubóstwa.

Poverty index	:	FGT index
Household size	:	lis
Group variable	:	reg
Parameter alpha	:	0.00

Group	Estimate	STE	LB	UB	Pov. line
1	0.024281	0.001789	0.020775	0.027788	500.00
3	0.022660	0.001974	0.018791	0.026530	500.00
Population	0.023556	0.001326	0.020958	0.026155	500.00

Rysunek 8 Wyniki estymacji miary FGT_0 (HCR) dla badanych regionów z linią ubóstwa na poziomie 500 zł.

W drugim przypadku wartość linii ubóstwa odpowiada 60% mediany, stanowiąc równowartość 1282,09 zł dochodu ekwiwalentnego. Tak podniesiony próg ubóstwa spowodował znaczne zwiększenie udziału gospodarstw domowych, znajdujących się w gorszej sytuacji materialnej – poniżej linii ubóstwa. W regionie wschodnim odsetek gospodarstw domowych egzystujących poniżej wyznaczonej linii ubóstwa wyniósł 18,91%, a w regionie centralnym – 14,06%.

Poverty index	:	FGT index				
Household size	:	lis				
Group variable	:	reg				
Parameter alpha	:	0.00				
Group		Estimate	STE	LB	UB	Pov. line
1		0.140660	0.003818	0.133176	0.148145	1282.09
3		0.189124	0.004885	0.179550	0.198699	1282.09
Population		0.162339	0.002953	0.156550	0.168127	1282.09

Rysunek 9 Wyniki estymacji miary FGT₀ (HCR) dla badanych regionów z linią ubóstwa na poziomie 1282,09 zł.

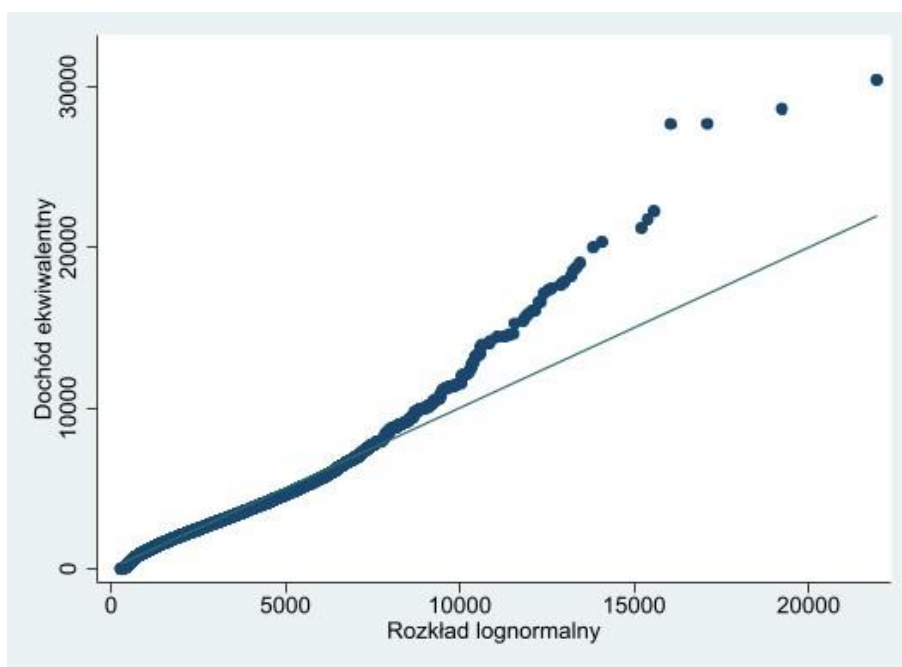
Z przeprowadzonej analizy porównawczej wynika, że według indeksu HCR, w regionie wschodnim mniejsza część gospodarstw domowych uzyskuje dochody ekwiwalentne poniżej wyznaczonej linii ubóstwa na poziomie 500 zł, natomiast w regionie centralnym, odsetek ten jest minimalnie większy, co informuje, że większa liczba gospodarstw domowych w regionie centralnym deklaruje dochody poniżej poziomu linii ubóstwa.

Sytuacja zmienia się w przypadku podniesienia linii ubóstwa do poziomu 1282,09 zł dochodu ekwiwalentnego. W tym przypadku, w regionie centralnym odsetek gospodarstw domowych egzystujących poniżej progu ubóstwa jest znacznie mniejszy niż w regionie wschodnim, co informuje, że w przedziale 500-1282,09 zł różnica pomiędzy badanymi regionami jest jeszcze większa.

5. Teoretyczne modele rozkładów

Wyniki dopasowania rozkładu logarytmiczno-normalnego

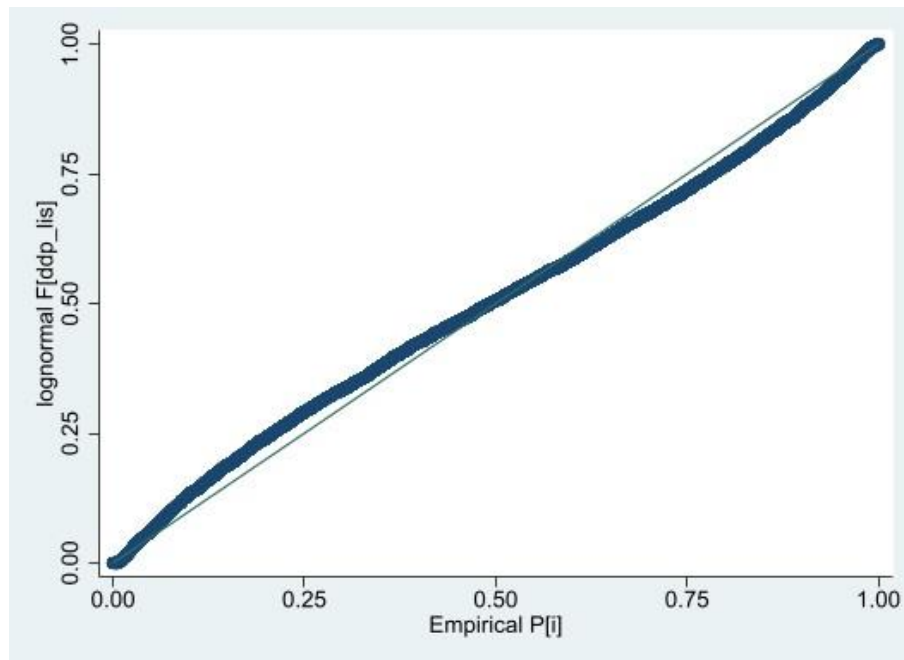
Zaprezentowany poniżej wykres przedstawia wyniki przeprowadzenia testu kwantyl-kwantyl (Q-Q), który ukazuje prawidłowość rozkładu reszt względem rozkładu normalnego.



Rysunek 10 Wykres estymacji (Q-Q) rozkładu logarytmiczno-normalnego w badanych regionach.

W powyższym przypadku zachodzi sytuacja, w której wartości reszt odbiegają od linii wyznaczającej rozkład normalny, co oznacza, że przy wyższych wartościach dochodu ekwiwalentnego mamy do czynienia z nieprawidłowym dopasowaniem względem wartości teoretycznych rozkładu.

Na poniższym wykresie zilustrowano wyniki przeprowadzenia analizy testu empirycznego prawdopodobieństw (P-P) służącego do oceny stopnia zgodności danych i przedstawiającego skumulowane funkcje rozkładu względem siebie.



Rysunek 11 Wykres estymacji (P-P) rozkładu logarytmiczno-normalnego w badanych regionach.

W powyższej sytuacji rozkład dochodu ekwiwalentnego dla obu badanych regionów nie pokrywa się idealnie z wartościami teoretycznymi, co informuje, że dane służące do opisu rozkładu dochodów nie są odpowiednio dopasowane oraz rozkład nie jest idealnie zgodny z wartościami teoretycznymi.

Wyniki dopasowania uogólnionego rozkładu beta drugiego rodzaju

Jest ciągłym rozkładem prawdopodobieństwa, opisywanym za pomocą następujących parametrów, w tym:

- a, b – parametry skali
- p – prawdopodobieństwo zaistnienia danego zdarzenia
- q – równe 1-p

Dzięki zastosowaniu estymacji uogólnionego rozkładu beta drugiego rodzaju otrzymujemy także statystyki opisowe dotyczące rozkładu, są to m.in. dominanta, średnia, odchylenie standardowe, wariancja czy współczynnik zmienności.

W przypadku regionu centralnego, możemy zaobserwować, że 1% najuboższych gospodarstw domowych osiąga dochód ekwiwalentny w wysokości 521,29 zł lub mniej, natomiast najbogatsza część regionu deklaruje dochód na poziomie 10 100 zł lub więcej.

ML fit of GB2 distribution				Number of obs	=	7948
Log likelihood = -68561.908				Wald chi2(0)	=	.
				Prob > chi2	=	.
DDP_LIS		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
a	_cons	4.128606	.2754183	14.99	0.000	3.588796 4.668416
b	_cons	2565.094	43.07665	59.55	0.000	2480.666 2649.523
p	_cons	.6671557	.0583593	11.43	0.000	.5527736 .7815378
q	_cons	.7559691	.0702196	10.77	0.000	.6183412 .893597
Quantiles		Cumulative shares of total DDP_LIS (Lorenz ordinates)				
1%	521.29353	0.00134				
5%	937.82044	0.01204				
10%	1.21e+03	0.03107				
20%	1.59e+03	0.08058				
25%	1.74e+03	0.10980				
30%	1.89e+03	0.14162				
40%	2.16e+03	0.21265	Mode	2.09e+03		
50%	2.45e+03	0.29342	Mean	2.85e+03		
60%	2.76e+03	0.38462	Std. Dev.	2.06e+03		
70%	3.15e+03	0.48803				
75%	3.40e+03	0.54538	Variance	4.24e+06		
80%	3.70e+03	0.60753	Half CV^2	0.26084		
90%	4.74e+03	0.75304				
95%	5.97e+03	0.84534	p90/p10	3.90487		
99%	1.01e+04	0.94811	p75/p25	1.95014		

Rysunek 12 Wyniki estymacji uogólnionego rozkładu beta drugiego rodzaju dla regionu centralnego.

W przypadku regionu wschodniego deklarowane dochody ekwiwalentne są na niższym poziomie niż ma to miejsce w przypadku regionu centralnego. Najuboższa część gospodarstw domowych osiąga dochód w wysokości 462,62 zł, natomiast najbogatszy odsetek deklaruje dochody ekwiwalentne w wysokości 7 060 zł lub więcej.

ML fit of GB2 distribution				Number of obs	=	6223
Log likelihood = -51829.061				Wald chi2(0)	=	.
				Prob > chi2	=	.

	DDP_LIS	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]

a						
	_cons	5.945429	.5172255	11.49	0.000	4.931685 6.959172

b						
	_cons	2211.601	31.68272	69.80	0.000	2149.504 2273.698

p						
	_cons	.4601856	.0487796	9.43	0.000	.3645794 .5557918

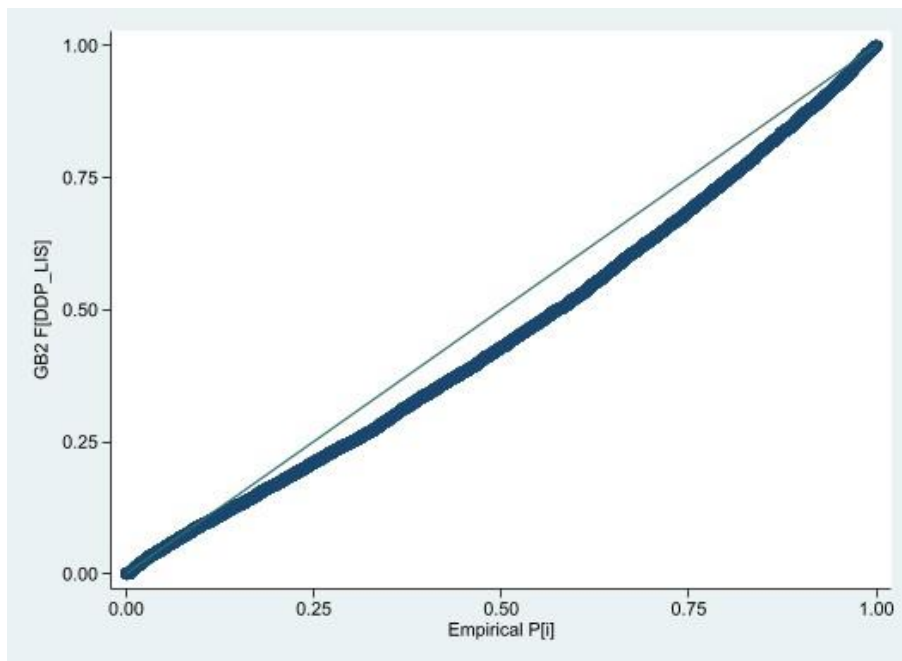
q						
	_cons	.5851815	.0660349	8.86	0.000	.4557555 .7146075

Quantiles		Cumulative shares of total DDP_LIS (Lorenz ordinates)				

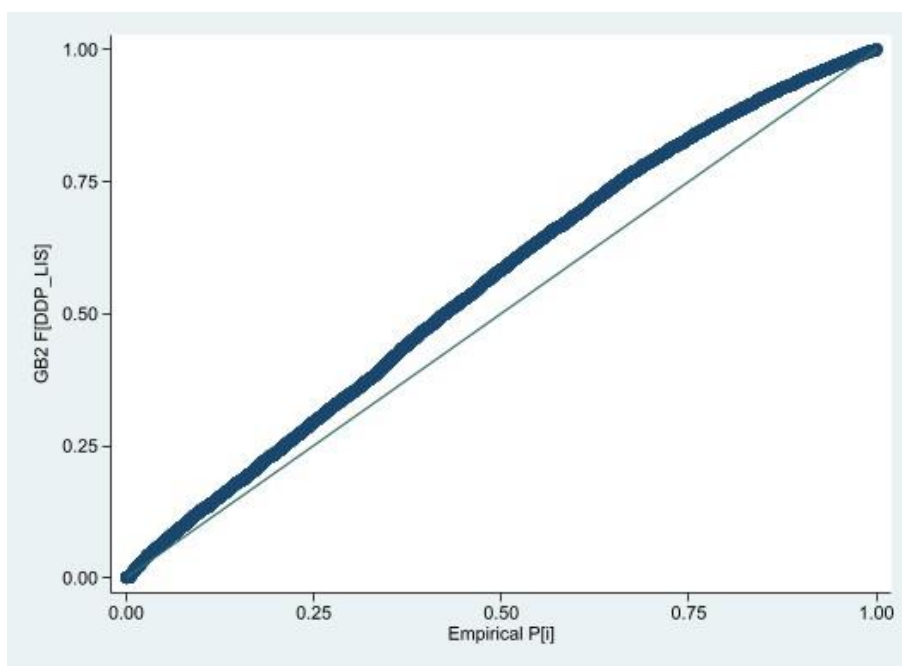
1%	462.62445	0.00148				
5%	833.39899	0.01333				
10%	1.08e+03	0.03437				
20%	1.39e+03	0.08877				
25%	1.52e+03	0.12061				
30%	1.63e+03	0.15507				
40%	1.85e+03	0.23118	Mode	1.89e+03		
50%	2.06e+03	0.31644	Mean	2.29e+03		
60%	2.28e+03	0.41114	Std. Dev.	1.36e+03		
70%	2.55e+03	0.51655				
75%	2.72e+03	0.57411	Variance	1.86e+06		
80%	2.93e+03	0.63575	Half CV^2	0.17789		
90%	3.62e+03	0.77686				
95%	4.44e+03	0.86367	p90/p10	3.37050		
99%	7.06e+03	0.95668	p75/p25	1.79142		

Rysunek 13 Wyniki estymacji uogólnionego rozkładu beta drugiego rodzaju dla regionu wschodniego.

Zauważalne jest zróżnicowanie w osiąganych dochodach w badanych regionach Polski w poszczególnych grupach. Region centralny wydaje się być bardziej zamożnym od regionu wschodniego.



Rysunek 14 Wykres estymacji uogólnionego rozkładu beta drugiego rodzaju dla regionu centralnego.



Rysunek 15 Wykres estymacji uogólnionego rozkładu beta drugiego rodzaju dla regionu wschodniego.

Z powyższych wykresów P-P wynika, że uogólniony rozkład beta drugiego rodzaju nie zapewnia wystarczająco dobrego dopasowania rozkładu dochodów w badanych regionach, ponieważ widoczne rozkłady odbiegają od wartości teoretycznych.

Rozkład Daguma

W przypadku, gdy w uogólnionym rozkładzie beta drugiego rodzaju, parametr q przyjmuje wartość bliską 1, można skorzystać z rozkładu dystrybucji dochodu Daguma. Wyniki z estymacji dla regionu centralnego przedstawiają się następująco:

ML fit of Dagum distribution				Number of obs	=	7948
Log likelihood = -68566.319				Wald chi2(0)	=	.
				Prob > chi2	=	.

	DDP_LIS	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]

a						
	_cons	3.422449	.0577895	59.22	0.000	3.309184 3.535715

b						
	_cons	2625.987	44.8781	58.51	0.000	2538.028 2713.947

p						
	_cons	.8452412	.0310469	27.22	0.000	.7843904 .906092

Quantiles		Cumulative shares of total DDP_LIS (Lorenz ordinates)				

1%	535.13954	0.00140				
5%	940.29543	0.01228				
10%	1.21e+03	0.03140				
20%	1.58e+03	0.08093				
25%	1.73e+03	0.11015				
30%	1.88e+03	0.14201				
40%	2.16e+03	0.21323	Mode	2.05e+03		
50%	2.45e+03	0.29446	Mean	2.83e+03		
60%	2.77e+03	0.38646	Std. Dev.	1.91e+03		
70%	3.17e+03	0.49104				
75%	3.42e+03	0.54911	Variance	3.64e+06		
80%	3.73e+03	0.61204	Half CV^2	0.22668		
90%	4.74e+03	0.75900	Gini coeff.	0.30468		
95%	5.90e+03	0.85142	p90/p10	3.92024		
99%	9.57e+03	0.95219	p75/p25	1.97388		

Rysunek 16 Wyniki estymacji rozkładu Daguma dla regionu centralnego.

W przypadku tego regionu, możemy zaobserwować, że 1% najuboższych gospodarstw domowych osiąga dochód ekwiwalentny w wysokości 535,14 zł lub mniej, natomiast najbogatsza część regionu deklaruje dochód na poziomie 9 570 zł lub więcej.

ML fit of Dagum distribution					Number of obs	=	6223
					Wald chi2(0)	=	.
Log likelihood = -51840.844					Prob > chi2	=	.

	DDP_LIS		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]

a							
	_cons		4.092724	.0793802	51.56	0.000	3.937142 4.248307

b							
	_cons		2308.492	35.09755	65.77	0.000	2239.702 2377.282

p							
	_cons		.7223691	.0279884	25.81	0.000	.6675129 .7772254

Quantiles				Cumulative shares of total DDP_LIS (Lorenz ordinates)			

1%	486.43416		0.00160				
5%	841.30931		0.01382				
10%	1.07e+03		0.03506				
20%	1.38e+03		0.08939				
25%	1.50e+03		0.12112				
30%	1.62e+03		0.15548				
40%	1.84e+03		0.23157	Mode	1.83e+03		
50%	2.05e+03		0.31724	Mean	2.27e+03		
60%	2.29e+03		0.41290	Std. Dev.	1.24e+03		
70%	2.58e+03		0.51993				
75%	2.75e+03		0.57853	Variance	1.54e+06		
80%	2.96e+03		0.64132	Half CV^2	0.14950		
90%	3.63e+03		0.78460	Gini coeff.	0.26800		
95%	4.37e+03		0.87158	p90/p10	3.39018		
99%	6.55e+03		0.96175	p75/p25	1.83092		

Rysunek 17 Wyniki estymacji rozkładu Daguma dla regionu wschodniego.

W przypadku regionu wschodniego deklarowane dochody ekwiwalentne są na niższym poziomie niż ma to miejsce w przypadku regionu centralnego. Najuboższa część gospodarstw domowych osiąga dochód w wysokości 482,43 zł, natomiast najbogatszy odsetek deklaruje dochody ekwiwalentne w wysokości 6 550 zł lub więcej.

Z estymacji za pomocą rozkładu Daguma zauważalne jest, że grupy najuboższe osiągają minimalnie większe dochody niż w przypadku uogólnionego rozkładu beta drugiego rodzaju, natomiast w grupach najbogatszych odnotowano spadek wysokości dochodów względem poprzedniego modelu.

Ponownie, zauważalne jest zróżnicowanie w osiągniętych dochodach w badanych regionach Polski w poszczególnych grupach. Region centralny wydaje się być bardziej zamożnym od regionu wschodniego.

Rozkład Singha-Maddali

W przypadku, gdy w uogólnionym rozkładzie beta drugiego rodzaju, parametr p przyjmuje wartość bliską 1, można skorzystać z rozkładu dystrybucji dochodu Singha-Maddali. Wyniki z estymacji dla regionu centralnego przedstawiają się następująco:

```
ML fit of Singh-Maddala distribution                                Number of obs   =    7954
                                                                    Model chi2(0)    =      .
                                                                    Prob > chi2      =      .

Log Likelihood = -68702.5861570
```

	DDP_LIS	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
p1						
	_cons	.6909236	.0241808	28.57	0.000	.64353 .7383172
p2						
	_cons	7.883842	.0215149	366.44	0.000	7.841673 7.92601
p3						
	_cons	.1818709	.0417357	4.36	0.000	.1000704 .2636714

```
a = 1+exp(p1) = 2.99556; std. err. = 0.04025; z = 62.07859
b = 1+exp(p2) = 2.66e+03; std. err. = 57.10159; z = 46.49694
q = exp(p3) = 1.19946; std. err. = 0.05006; z = 23.96029
```

Singh-Maddala model estimates for distribution of DDP_LIS

Percentiles Cumulative shares of total DDP_LIS (Lorenz ordinates)				
1%	538.74770	0.00142		
5%	933.68096	0.01226		
10%	1.20e+03	0.03120		
20%	1.56e+03	0.08021		
25%	1.72e+03	0.10915		
30%	1.86e+03	0.14074		
40%	2.15e+03	0.21153		
50%	2.45e+03	0.29252	Mean	2.83e+03
60%	2.78e+03	0.38453	Std. Dev.	1.88e+03
70%	3.19e+03	0.48947		
75%	3.44e+03	0.54787	Variance	3.53e+06
80%	3.76e+03	0.61124	Half CV^2	0.21992
90%	4.78e+03	0.75944	Gini coeff.	0.30649
95%	5.94e+03	0.85256	p90/p10	3.99515
99%	9.50e+03	0.95340	p75/p25	2.00454

Rysunek 18 Wyniki estymacji rozkładu Singha-Maddali dla regionu centralnego.

W przypadku tego regionu, możemy zaobserwować, że 1% najuboższych gospodarstw domowych osiąga dochód ekwiwalentny w wysokości 538,75 zł lub mniej, natomiast najbogatsza część regionu deklaruje dochód na poziomie 9 500 zł lub więcej.

```

ML fit of Singh-Maddala distribution
Log Likelihood = -51873.8469212

Number of obs   = 6224
Model chi2(0)   = .
Prob > chi2     = .

-----
      DDP_LIS |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
p1      _cons |   .8403763   .0251629    33.40   0.000    .7910578   .8896947
p2      _cons |   7.726625   .0217216   355.71   0.000    7.684051   7.769198
p3      _cons |   .2546702   .0470435     5.41   0.000    .1624667   .3468738
-----
a = 1+exp(p1) =   3.31724; std. err. =   0.05831; z = 56.89113
b = 1+exp(p2) =  2.27e+03; std. err. =  49.26312; z = 46.05746
q = exp(p3) =   1.29004; std. err. =   0.06069; z = 21.25693

Singh-Maddala model estimates for distribution of DDP_LIS
-----
Percentiles Cumulative shares of
              total DDP_LIS (Lorenz ordinates)

1%  525.69874      0.00177
5%  863.44042      0.01446
10% 1.08e+03      0.03594
20% 1.37e+03      0.09009
25% 1.49e+03      0.12154
30% 1.61e+03      0.15556
40% 1.83e+03      0.23085
50% 2.05e+03      0.31575      Mean      2.28e+03
60% 2.29e+03      0.41082      Std. Dev. 1.25e+03
70% 2.59e+03      0.51756
75% 2.77e+03      0.57619      Variance 1.55e+06
80% 2.98e+03      0.63917      Half CV^2 0.14950
90% 3.68e+03      0.78338      Gini coeff. 0.26962
95% 4.43e+03      0.87119      p90/p10    3.40571
99% 6.60e+03      0.96206      p75/p25    1.85179

```

Rysunek 19 Wyniki estymacji rozkładu Singha-Maddali dla regionu wschodniego.

W przypadku regionu wschodniego deklarowane dochody ekwiwalentne są na niższym poziomie niż ma to miejsce w przypadku regionu centralnego. Najuboższa część gospodarstw domowych osiąga dochód w wysokości 525,70 zł, natomiast najbogatszy odsetek deklaruje dochody ekwiwalentne w wysokości 6 600 zł lub więcej.

Z estymacji za pomocą rozkładu Singha-Maddali zauważalne jest, że grupy najuboższe ponownie osiągają minimalnie większe dochody niż w przypadku uogólnionego rozkładu beta drugiego rodzaju oraz rozkładu Daguma, natomiast w grupach najbogatszych również odnotowano spadek wysokości dochodów względem poprzednich modeli.

Ponownie, zauważalne jest zróżnicowanie w osiąganych dochodach w badanych regionach Polski w poszczególnych grupach. Region centralny wydaje się być bardziej zamożnym od regionu wschodniego, choć po estymacji tym modelem, rozbieżności między badanymi regionami uległy zmniejszeniu.

Podsumowanie i wnioski końcowe

Przeprowadzona analiza dostarczyła niezwykle cennych informacji dotyczących rozkładów dochodów w regionach centralnym oraz wschodnim Polski w 2016 roku.

W pierwszym rozdziale przedstawiono ogólną charakterystykę danych statystycznych wykorzystywanych w projekcie oraz niezbędną wiedzę teoretyczną potrzebną do przeanalizowania badanych rozkładów dochodów. Opis dotyczył m.in. metod parametrycznych oraz nieparametrycznych, takich jak histogram, estymacja funkcją gęstości metodą kernel. Przedstawiono również metody pomiaru nierówności ekonomicznych za pomocą indeksu Giniego, uogólnionej entropii, a także indeksu Atkinsona. Zaprezentowano metody pomiaru ubóstwa ekonomicznego oraz teoretyczne modele rozkładów dochodów.

W drugim rozdziale zawarto wyniki estymacji funkcji gęstości i parametrów rozkładów. Histogram prezentował rozkłady dochodów w regionach. Rozkład ten okazał się być zbliżonym do rozkładu normalnego, co wykazało, że większość dochodów gospodarstw domowych skupiona była wokół najczęściej występujących wartości danego przedziału. Po przeprowadzeniu estymacji, wykorzystując funkcję gęstości metodą kernel można było wywnioskować, że w przypadku obu regionów mieliśmy do czynienia z rozkładem asymetrycznym, prawostronnym, dodatnim. W przypadku regionu wschodniego, rozkład był bardziej skoncentrowany wokół przeciętnego dochodu, natomiast w regionie centralnym rozbieżności w dochodach były większe.

Rozdział trzeci zawierał wyniki pomiaru nierówności ekonomicznych badanych rozkładów i testowania różnic pomiędzy miarami nierówności. Krzywa Lorenza, indeks Giniego, uogólnionej entropii oraz Atkinsona ukazały te same wyniki dla badanych obszarów, informując, że w regionie wschodnim rozkład dochodów ekwiwalentnych okazał się być bardziej równomierny niż w regionie centralnym.

W czwartym rozdziale zaprezentowano wyniki analizy ubóstwa ekonomicznego za pomocą miary FGT_0 (indeksu HCR). Zostały wyznaczone odpowiednie linie ubóstwa, które poinformowały oraz scharakteryzowały występujące zależności pomiędzy badanymi regionami.

Piąty rozdział poświęcony był przedstawieniu wyników dopasowania rozkładu logarytmiczno-normalnego oraz uogólnionego rozkładu beta drugiego rodzaju, a także jego specyficznym odmianom (rozkładowi Daguma oraz Singha-Maddali). Zauważalne było zróżnicowanie w osiągniętych dochodach w badanych regionach Polski w poszczególnych grupach

oraz odnotowano niewystarczające dopasowanie rozkładów związanych z dochodami ekwiwalentnymi do wartości teoretycznych modeli.

Głównymi celem projektu było podstawienie hipotez wymienionych we wstępie do projektu. Po przeprowadzonej analizie i zgromadzeniu niezbędnych informacji oraz wiedzy teoretycznej można potwierdzić, że rozważane rozkłady różnią się pod względem miar położenia, zmienności, asymetrii i koncentracji oraz, że nierówności ekonomiczne w rozkładzie regionu centralnego są większe niż w rozkładzie regionu wschodniego. Natomiast odrzucić można stwierdzenie, jakoby rozkład regionu centralnego wykazywał większą skalę i głębię ubóstwa niż rozkład regionu wschodniego oraz, że rozważane rozkłady charakteryzują się jednakową postacią teoretyczną, ale o różnych parametrach.

Niniejszy projekt odpowiedział na postawione na wstępie pytania badawcze oraz wykazał, że region centralny charakteryzuje się większą rozbieżnością dystrybucji dochodów w gospodarstwach domowych jednocześnie będąc bogatszym regionem Polski. Natomiast region wschodni, choć uboższy okazał się bardziej równomiernie rozłożonym pod względem deklarowanych dochodów. Uzyskane wyniki potwierdzają ogólne przeświadczenie, jakoby region wschodni był biedniejszym regionem Polski względem regionu centralnego.

■ ■ ■