

**PODZIAŁ POLSKI NA OBSZARY O ZBLIŻONYCH
WARUNKACH ROZWOJU GOSPODARCZEGO I
SPOŁECZNEGO ZA POMOCĄ ANALIZY
DYSKRYMINACYJNEJ**

Jadwiga Zaród

*Zachodniopomorski Uniwersytet Technologiczny w
Szczecinie*

Katedra Zastosowań Matematyki w Ekonomii

Cel pracy

Celem pracy jest podział Polski na obszary o zbliżonych warunkach rozwoju gospodarczego i społecznego, uszeregowanie ich (od najlepiej do naj słabiej rozwiniętych) oraz wyłonienie kompleksów pretendujących do miana metropolii.

Material badawczy

Badaniami objęto 66 podregionów Polski. Każdy podregion opisano za pomocą zmiennych zwanych diagnostycznymi:

- x_1 – liczba ludności,
- x_2 – gęstość zaludnienia
- x_3 – procent ludności w wieku produkcyjnym,
- x_4 – stopa bezrobocia,
- x_5 – przeciętne miesięczne wynagrodzenie brutto (zł),
- x_6 – przeciętna powierzchnia użytkowa 1 mieszkania (m^2),
- x_7 - przeciętna powierzchnia użytkowa mieszkania na 1 osobę (m^2),
- x_8 – procent ludności korzystającej z instalacji wodociągowej,
- x_9 – procent ludności korzystającej z instalacji kanalizacyjnej,
- x_{10} – procent ludności korzystającej z instalacji gazowej,
- x_{11} – wskaźnik skolaryzacji brutto dla liceów ogólnokształcących (%),
- x_{12} – wskaźnik skolaryzacji brutto dla szkół policealnych (%),
- x_{13} – liczba uczniów przypadających na 1 komputer z dostępem do Internetu w gimnazjach,
- x_{14} – liczba uczniów przypadających na 1 komputer z dostępem do Internetu w liceach,
- x_{15} – liczba ludności na 1 zakład opieki zdrowotnej,
- x_{16} – liczba czytelników bibliotek na 1000 ludności,
- x_{17} – liczba ludności na 1 miejsce w kinach,
- x_{18} – procent ludności obsługiwanej przez oczyszczalnie ścieków,
- x_{19} – liczba podmiotów gospodarki narodowej wpisanych do rejestru REGON.

Material badawczy

Na podstawie wartości zmiennych diagnostycznych dokonano wstępnego podziału podregionów na 7 obszarów (grup). Liczbę grup k ustalono na podstawie wzoru:

$$k = 1 + 3,322 \log n$$

gdzie $n = 66$ – liczba podregionów.

Dane wejściowe poddano standaryzacji, aby uniezależnić wyniki analiz od jednostek pomiaru poszczególnych zmiennych za pomocą wzoru:

$$Z_{ij} = \frac{x_{ij} - \bar{x}}{s}$$

gdzie:

x_{ij} - wartość j -tej zmiennej dla i -tego podregionu,

\bar{x} - wartość średnia danej zmiennej,

s – odchylenie standardowe.

Zmienna x_4 jest destymulantą, poprzez przemnożenie jej wartości przez minus jeden dokonano przekształcenia w stymulantę.

Material badawczy

Analiza dyskryminacyjna zakłada, że dane (ujęte w postaci zmiennych) reprezentują próbę z wielowymiarowego rozkładu normalnego. Po wykonaniu histogramów rozkładów liczebności oraz testów χ^2 i Kołmogorowa-Smirnowa, służących do oceny normalności rozkładu zredukowano liczbę zmiennych do: $x_2, x_3, x_4, x_6, x_7, x_9, x_{10}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}, x_{18}, x_{19}$. Dalsze ograniczenie listy zmiennych spowodowało badanie ich mocy dyskryminacyjnej za pomocą testu Lambda Wilksa i testu F.

Materiał badawczy

Tabela 1. Wyniki testu Lambda Wilksa i testu F

Zmienne	Lambda Wilksa	F	p	Tolerancja	1-Tolerancja (R kwadrat)
x2	0,0017	30,0748	0,0000	0,4803	0,5197
x3	0,0005	4,5137	0,0009	0,7027	0,2973
x4	0,0009	13,5048	0,0000	0,4736	0,5264
x6	0,0008	10,0522	0,0000	0,4375	0,5625
x7	0,0007	6,5911	0,0000	0,5103	0,4893
x10	0,0012	18,0542	0,0000	0,5936	0,4064
x13	0,0006	5,7866	0,0001	0,8173	0,1827
x14	0,0006	5,1299	0,0003	0,8316	0,1683
x19	0,0005	3,5442	0,0052	0,4823	0,5177

Metody badawcze

Analiza dyskryminacyjna, główna metoda badawcza, stanowi zespół metod dyskryminacyjnych i klasyfikacyjnych.

Metody dyskryminacyjne podziela Polskę na obszary jak najbardziej do siebie podobne ze względu na opisujące je cechy.

Metody klasyfikacyjne ustala, do którego z utworzonych obszarów należy przyporządkować dany podregion, wykorzystując w tym celu te zmienne, które miały największą moc dyskryminacyjną

Metody dyskryminacyjne

Do podziału wykorzystano funkcje dyskryminacyjne o postaci:

$$F = A^T Z$$

gdzie:

F – macierz zmiennych dyskryminacyjnych

A^T – transponowana macierz współczynników zmiennych dyskryminacyjnych

Z – wystandaryzowana macierz obserwacji

Maksymalna liczba obliczonych funkcji jest równa liczbie grup minus jeden (czyli 6).

Funkcja dyskryminacyjna o postaci:

$$F = 1,0136x_2 - 0,0913x_3 + 0,5019x_4 - 0,4324x_6 + 0,241x_7 + 0,2868x_{10} + 0,4361x_{13} + 0,3684x_{14} - 0,7295x_{19}$$

wyjaśnia 72,95% wariacji międzygrupowej, ma najniższą wartość testu Lambda Wilksa (0,0004) i dlatego będzie stanowiła podstawę dalszych analiz.

Metody dyskryminacyjne

Na podstawie funkcji dyskryminacyjnej można obliczyć wartości kanoniczne dla wszystkich podregionów i przeciętne dla każdego obszaru.

Tabela 2. Wartości średnie funkcji dyskryminacyjnej obszarów

Obszary	Średnie wartości kanoniczne
1	12,7112
2	2,0356
3	-0,6347
4	-1,9783
5	-2,2033
6	-2,5722
7	-3,5035

Metody klasyfikacyjne

Do klasyfikacji obiektów wykorzystano funkcje klasyfikacyjne o postaci:

$$K_r = c_r + c_{r1}x_1 + c_{r2}x_2 + \dots + c_{rj}x_j \quad r = 1, 2, \dots, k, \quad j = 1, 2, \dots, m$$

gdzie:

K_r – wartość r -tej zmiennej klasyfikacyjnej (dla r -tego obszaru)

c_{r0} – stała dla r -tego obszaru

c_{rj} – współczynniki zmiennych o istotnej mocy dyskryminacyjnej

x_j – wartości obserwowane (standaryzowane) dla j -tej zmiennej.

Funkcje o takiej postaci zostały oszacowane dla 7 obszarów. Dla każdego podregionu obliczono wartości wszystkich funkcji klasyfikacyjnych. Dany podregion przyporządkowano do obszaru, dla którego ma on największą wartość klasyfikacyjną.

Metody klasyfikacyjne

Tabela 3. Klasyfikacja wstępna podregionów

Obszar	Poprawność kwalifikacji (%)	Liczba podregionów w poszczególnych obszarach						
		1	2	3	4	5	6	7
1	100,00	8						
2	87,50		7	1				
3	60,00		1	6		2	1	
4	45,45		1		5	3	1	1
5	40,00			2	1	4	1	2
6	70,00				1	1	7	1
7	66,67						3	6
średnio	65,15	8	9	9	7	10	13	10

wiersze – klasyfikacja wstępna,

kolumny – podział na podstawie funkcji klasyfikacyjnych

Metody klasyfikacyjne

Tabela 5. Wyniki klasyfikacji końcowej

Obszary	Poprawność kwalifikacji (%)	Liczba podregionów w poszczególnych obszarach						
		1	2	3	4	5	6	7
1	100,00	8						
2	100,00		8					
3	100,00			9				
4	100,00				3			
5	100,00					7		
6	100,00						19	
7	100,00							12
średnio	100,00	8	8	9	3	7	19	12

Metody klasyfikacyjne

Tabela 4. Współczynniki funkcji klasyfikacyjnych

grupy zmienne	1	2	3	4	5	6	7
x2	47,5333	-1,8067	-7,8812	-5,0519	-2,3941	-9,3674	-7,0822
x3	-1,2695	0,3008	-0,9034	3,0636	-2,1689	0,1318	1,6138
x4	10,9205	6,8368	1,2482	-0,5058	-3,8266	-0,0852	-10,2807
x6	-8,2732	-5,4338	-1,1144	6,3984	8,7328	-0,0159	3,3077
x7	3,5598	1,177	2,8806	-2,3557	-4,569	-1,3185	0,0235
x10	6,4938	0,3178	1,5649	-2,5656	7,3734	-4,6412	-2,026
x13	8,886	3,7178	0,7762	-0,4711	-5,797	-2,2508	-1,9325
x14	4,891	2,8126	-0,5442	0,4726	0,3953	-1,3188	-2,9882
x19	-11,2246	-0,6017	1,0305	2,1593	0,5339	2,1438	2,8657
stała	83,6793	-8,5103	-5,6274	-11,4088	-16,2546	-5,9872	-11,9682

Metody klasyfikacyjne

Dodatkowo dla każdego klasyfikowanego podregionu obliczono jego odległości Mahalanobisa od środka ciężkości grupy według wzoru:

$$d_{ir} = \left[\sum_{j=1}^m \sum_{l=1}^m (x_{ij} - x_{rj})(x_{il} - x_{rl}) s_{jl} \right]^{\frac{1}{2}} \quad i = 1, \dots, n \quad (r = 1, \dots, k)$$

gdzie: x_{ij} – wartość j -tej zmiennej dla i -tego podregionu

x_{rj} – środek ciężkości dla j -tej zmiennej w r -tym obszarze

x_{il} – wartości l -tej zmiennej dla i -tego podregionu

x_{rl} – środek ciężkości dla l -tej zmiennej w r -tym obszarze

s_{jl} - jl -ty element macierzy odwrotnej do macierzy kowariancji.

Dany podregion zaklasyfikowano do tego obszaru, dla którego jego odległość od środka ciężkości jest najmniejsza.

Klasyfikacja podregionów do odpowiednich obszarów

Obszar	Jednostki terytorialne	Wartość funkcji dyskryminacyjnej	Kwadrat odległości Mahalanobisa
1	miasto Warszawa	14,8546	44,6771
	miasto Łódź	14,4015	28,2521
	miasto Kraków	14,0408	6,3572
	podregion katowicki	13,1490	8,5008
	miasto Poznań	12,9241	7,7455
	miasto Wrocław	12,4466	3,0116
	podregion trójmiejski	11,1607	8,1908
	miasto Szczecin	8,7225	22,2309

Klasyfikacja podregionów do odpowiednich obszarów

Obszar	Jednostki terytorialne	Wartość funkcji dyskryminacyjnej	Kwadrat odległości Mahalanobisa
2	podregion tyski	3,5160	112,3671
	podregion gliwicki	3,4766	108,3994
	podregion bytomski	2,9340	132,1094
	podregion rybnicki	2,4293	121,2634
	podregion byd.-toruński	1,8412	158,4313
	podregion sosnowiecki	1,6301	142,7497
	podregion leg.-głogowski	0,3423	168,2386
	podregion częstochowski	0,1151	179,2295

Klasyfikacja podregionów do odpowiednich obszarów

Obszar	Jednostki terytorialne	Wartość funkcji dyskryminacyjnej	Kwadrat odległości Mahalanobisa
3	podregion bielski	0,9007	161,6359
	podregion krakowski	-0,4680	199,2843
	podregion białostocki	-0,5482	189,4791
	podregion war. zachodni	-0,5759	216,8261
	podregion oświęcimski	-0,8967	193,5659
	podregion lubelski	-0,8977	198,2236
	podregion war. wschodni	-1,0238	205,3626
	podregion łódzki	-1,0601	206,2136
	podregion poznański	-1,1426	218,4594

Klasyfikacja podregionów do odpowiednich obszarów

Obszar	Jednostki terytorialne	Wartość funkcji dyskryminacyjnej	Kwadrat odległości Mahalanobisa
4	podregion leszczyński	-1,5600	222,3660
	podregion gdański	-1,9190	230,0851
	podregion kaliski	-2,4558	275,1436

Klasyfikacja podregionów do odpowiednich obszarów

Obszar	Jednostki terytorialne	Wartość funkcji dyskryminacyjnej	Kwadrat odległości Mahalanobisa
5	podregion rzeszowski	-1,3272	221,4506
	podregion tarnowski	-1,4865	229,4837
	podregion tarnobrzeski	-2,0483	240,3925
	podregion krośnieński	-2,1567	266,9303
	podregion wrocławski	-2,2475	246,1074
	podregion nowosądecki	-3,2022	266,4871
	podregion przemyski	-4,3383	277,3135

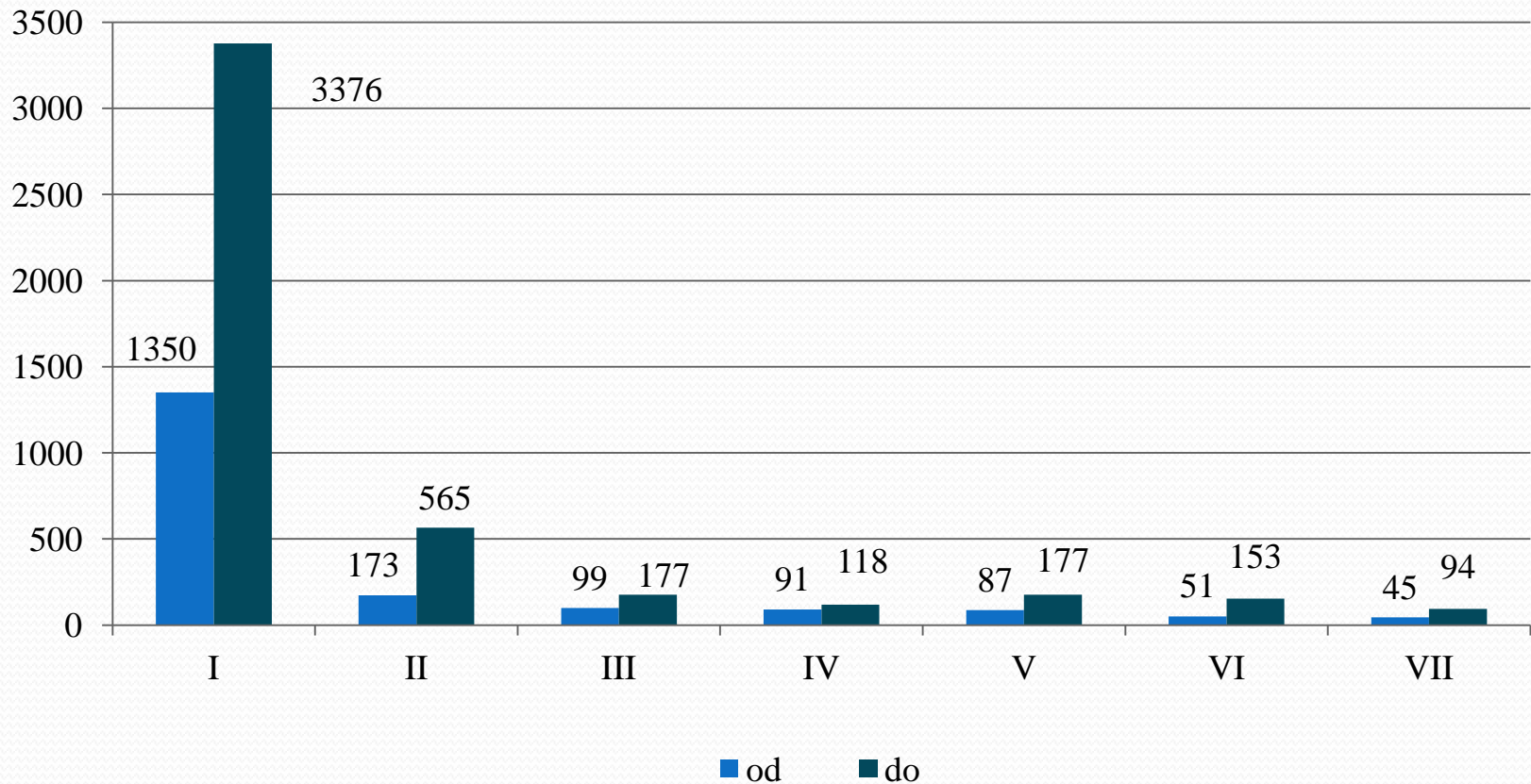
Klasyfikacja podregionów do odpowiednich obszarów

Obszar	Jednostki terytorialne	Wartość funkcji dyskryminacyjnej	Kwadrat odległości Mahalanobisa
6	podregion gorzowski	-1,3975	211,3141
	podregion piotrkowski	-1,3981	208,5928
	podregion skierniewicki	-1,6473	229,7509
	podregion starogardzki	-2,0259	224,3660
	podregion łomżyński	-2,0875	240,2193
	podregion kielecki	-2,2661	230,8468
	podregion puławski	-2,4468	244,7464
	podregion piski	-2,6172	240,4931
	podregion opolski	-2,6542	248,2406
	podregion elbląski	-2,7476	245,8443
	podregion ciechociński-płocki	-2,7489	242,7799
	podregion koniński	-2,7766	252,9043
	podregion suwalski	-2,8101	255,0524
	podregion chełmiński-zamojski	-2,9682	252,7898
	podregion ostrołęcko-siedlecki	-3,1522	259,2272
	podregion sandomier.-jędrzejowski	-3,2165	265,1809
podregion bialski	-3,4976	273,1995	
podregion sieradzki	-3,6419	286,9468	

Klasyfikacja podregionów do odpowiednich obszarów

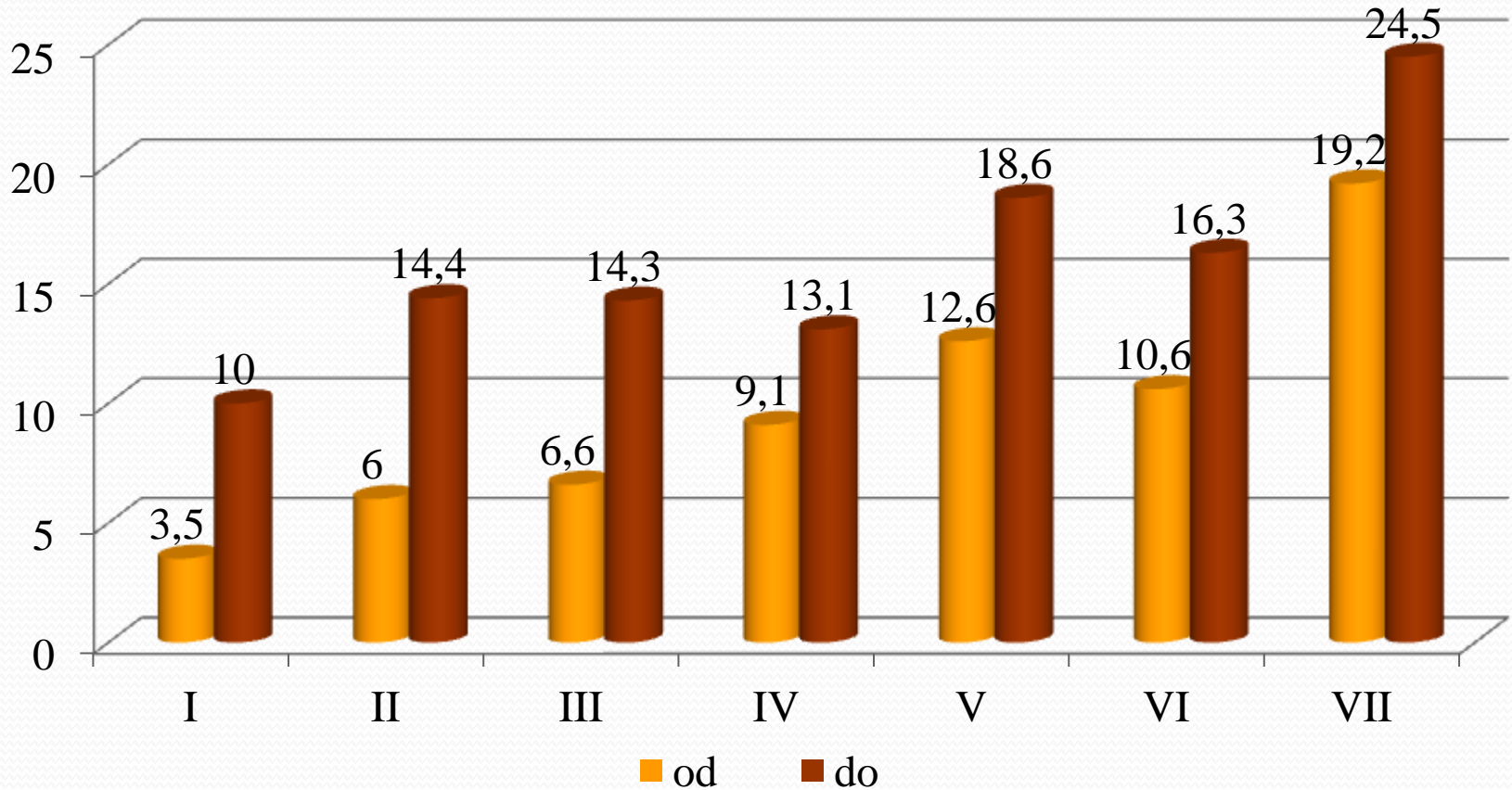
Obszar	Jednostki terytorialne	Wartość funkcji dyskryminacyjnej	Kwadrat odległości Mahalanobisa
7	podregion wałbrzyski	-2,4446	241,2780
	podregion olsztyński	-2,6406	242,4755
	podregion jeleniogórski	-2,7127	244,2529
	podregion zielonogórski	-2,7724	244,5128
	podregion radomski	-3,3256	270,3143
	podregion nyski	-3,5115	268,1817
	podregion ełcki	-3,5322	276,2134
	podregion słupski	-3,6289	271,6938
	podregion szczeciński	-3,6826	285,4871
	podregion koszaliński	-3,8574	285,5106
	podregion grudziądzki	-4,1286	288,2251
	podregion stargardzki	-4,2868	300,3961
	podregion włocławski	-4,2882	295,29,08

Charakterystyka obszarów



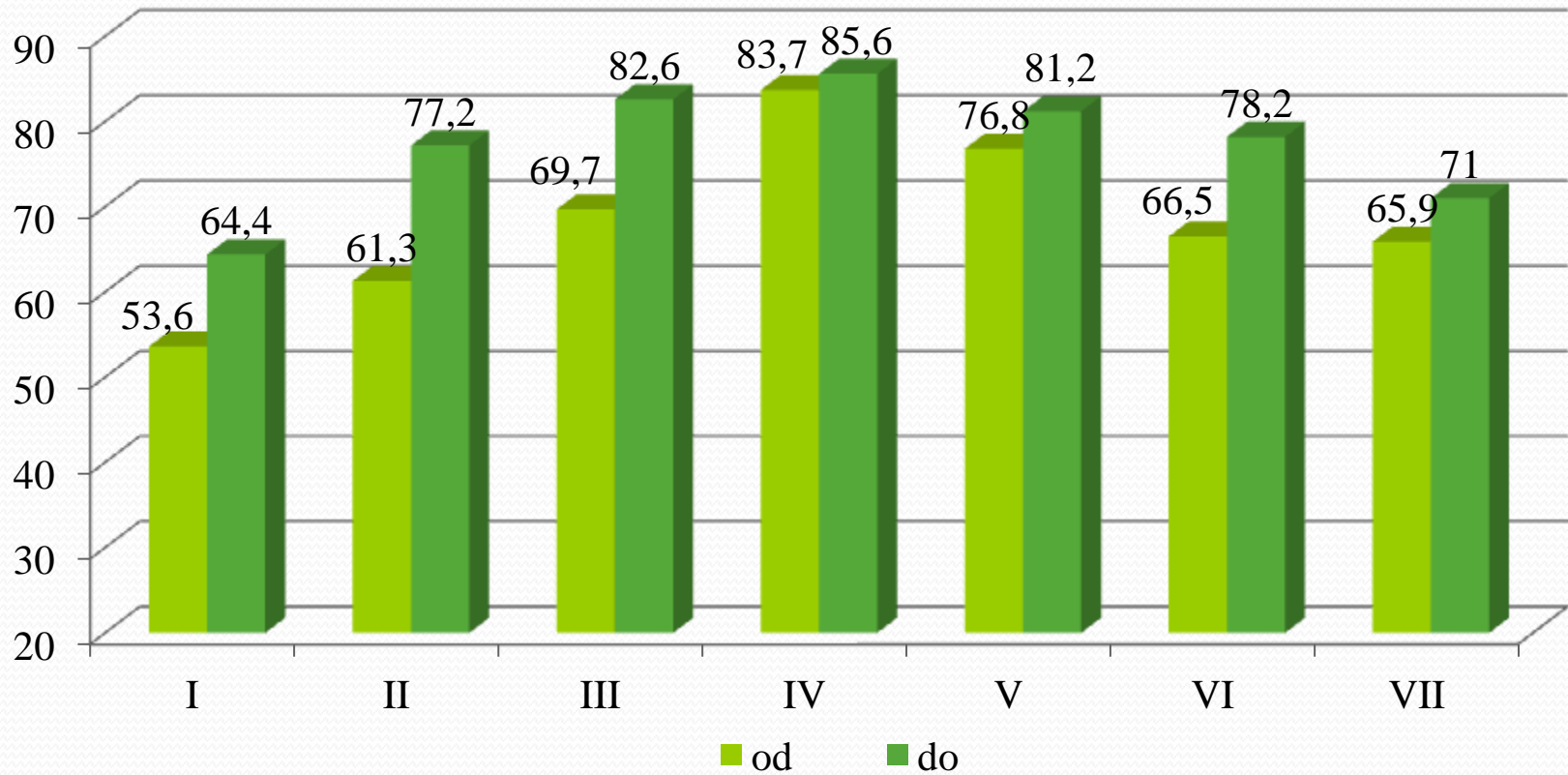
Wykres 1. Gęstość zaludnienia (osób /km²)

Charakterystyka obszarów



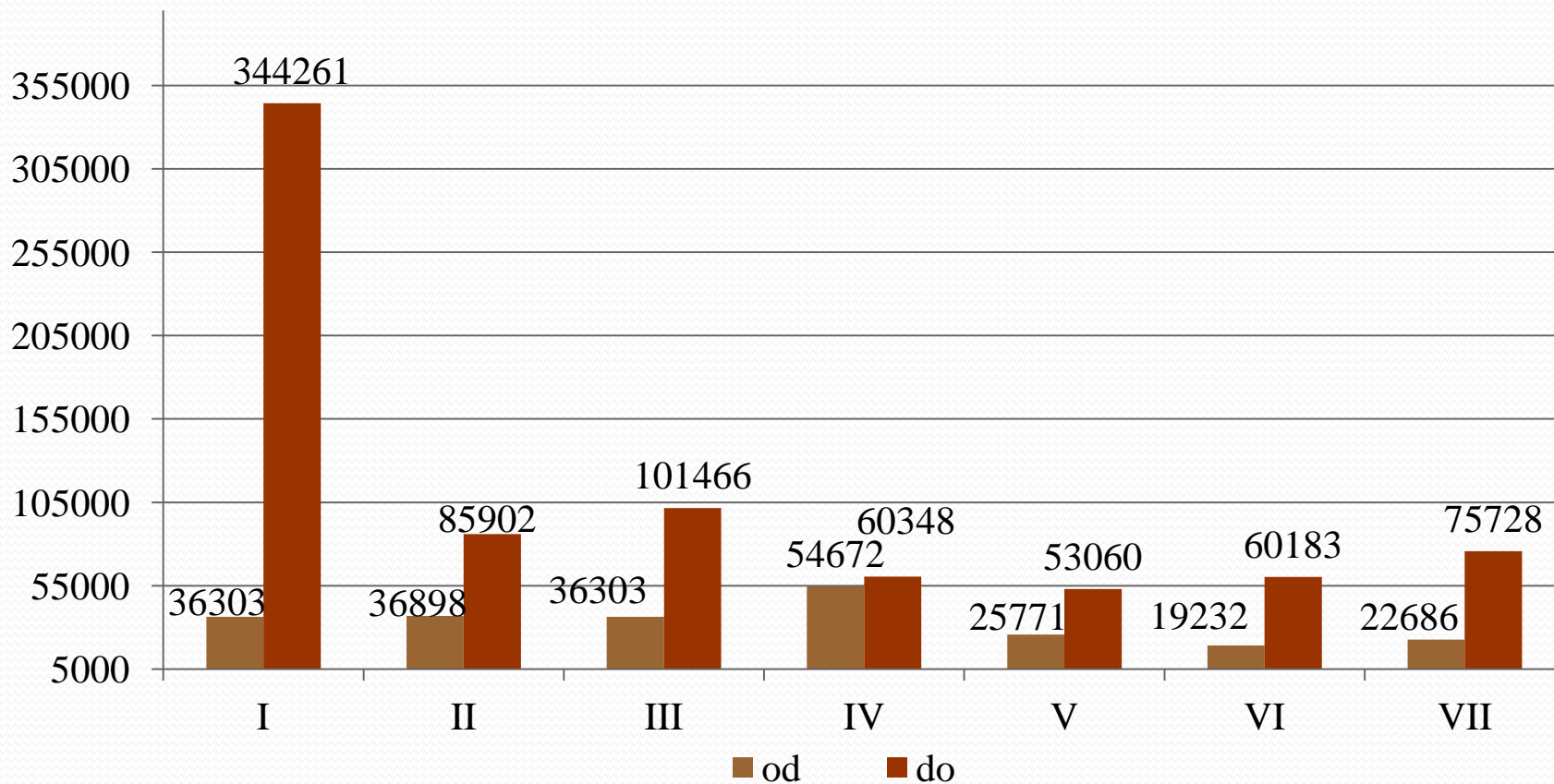
Wykres 2. Stopa bezrobocia (%)

Charakterystyka obszarów



Wykres 3. Powierzchnia mieszkania (m²)

Charakterystyka obszarów



Wykres 4. Liczba podmiotów gospodarczych

Wnioski

1. Funkcje dyskryminacyjne pozwalają podzielić Polskę na obszary jak najbardziej do siebie podobne ze względu na wybrane zmienne.
Trafność podziału potwierdzają niskie współczynniki testu Lambda Wilksa (0,0004)
2. Wyznaczone wartości funkcji dyskryminacyjnej wskazują podregiony pretendujące do miana obszarów metropolitalnych i uszeregowują wyłonione kompleksy od najlepiej do najslabiej rozwiniętych.
3. Funkcje klasyfikacyjne pozwalają przydzielić poszczególne podregiony do wyznaczonych obszarów.

Wnioski

4. Trafność klasyfikacji potwierdzają obliczone odległości Mahalanobisa.
5. Analiza dyskryminacyjna może być wykorzystywana jako narzędzie wspomagające podział podregionów na obszary o różnych możliwościach rozwoju.



Dziękuję