



## WPŁYW AKTYWNOŚCI MIESZKAŃCÓW NA POZIOM ŻYCIA W GMINACH WOJ. MAŁOPOLSKIEGO

*Mariusz Łapczyński, Akademia Ekonomiczna w Krakowie,  
Katedra Analizy Rynku i Badań Marketingowych*

### Aktywność mieszkańców gmin woj. małopolskiego

Celem niniejszego opracowania jest próba zbadania zależności między aktywnością mieszkańców i władz gminnych a poziomem życia w tych gminach<sup>1</sup>. Aktywność ta traktowana jest tutaj w trzech wymiarach: 1) jako aktywność ekonomiczna mieszkańców, 2) jako aktywność ekonomiczna władz, 3) oraz jako aktywność obywatelska mieszkańców.

**Pierwszy wymiar** ma charakter ekonomiczny i jest wyrażony czterema wskaźnikami:

- ◆ liczbą podmiotów gospodarczych (PGN<sup>2</sup>) na 1000 mieszkańców,
- ◆ udziałem prywatnych podmiotów gospodarczych w całkowitej liczbie PGN,
- ◆ współczynnikiem aktywności zawodowej,
- ◆ oraz wskaźnikiem zatrudnienia.

Podmiotami gospodarki narodowej są osoby prawne, jednostki organizacyjne nie mające osobowości prawnej oraz osoby fizyczne prowadzące działalność gospodarczą<sup>3</sup>. Nie są nimi natomiast indywidualne gospodarstwa rolne i jednostki lokalne (zakłady, oddziały, filie). Wykorzystując wskaźnik liczby podmiotów na 1000 mieszkańców, można zaobserwować dość duże zróżnicowanie jego wartości. Najmniej przedsiębiorczą gminą jest Radgoszcz z powiatu dąbrowskiego (30,05 podmiotów/1000 mieszkańców), a najbardziej przedsiębiorczą – Zakopane z powiatu tatrzańskiego (175,78 podmiotów/1000 mieszkańców).

Kolejny wskaźnik to udział prywatnych podmiotów gospodarczych w całkowitej liczbie PGN. Zróżnicowanie tego zjawiska w poszczególnych gminach województwa jest nieco mniejsze: od 85,45% w Mędrzechowie (powiat dąbrowski) do 98,73% w Zielonkach (powiat krakowski). Inny ze wskaźników aktywności ekonomicznej mieszkańców gmin to

---

<sup>1</sup> Badania zostały przeprowadzone wiosną 2005 r. przez Centrum Analiz Samorządowych Małopolskiego Instytutu Samorządu Terytorialnego i Administracji (MISTiA); szczegółowe wyniki badań są dostępne na stronie internetowej [www.mistia.org.pl](http://www.mistia.org.pl).

<sup>2</sup> PGN = Podmioty Gospodarki Narodowej.

<sup>3</sup> „Wybrane dane o powiatach i gminach województwa małopolskiego w 2002 r.”, uwagi metodyczne, s. 50.



współczynnik aktywności zawodowej (w %)<sup>4</sup>. Dostęp do danych wyłącznie na poziomie powiatów spowodował, że zdecydowano się przyjąć średnią wartość dla powiatu do wszystkich gmin tego powiatu. Trzeba tu zaznaczyć, że wskaźnik ten informuje niejako o potencjale gmin (powiatów), gdyż wskazuje na odsetek osób zdolnych do pracy. Gminami o najwyższych wartościach tego wskaźnika są gminy powiatu proszowickiego (60,70%), zaś gminami o najniższych wartościach są gminy powiatu chrzanowskiego (48,80%). Następnym przejawem aktywności ekonomicznej ludności jest wskaźnik zatrudnienia (w %)<sup>5</sup>. W tym przypadku również zdecydowano się przyjąć średnie wartości na poziomie powiatu dla wszystkich gmin tego powiatu. Podobnie jak poprzednio, gminami o najwyższych wartościach tego wskaźnika są gminy powiatu proszowickiego (54,30%), natomiast gminami o najniższych wartościach – gminy powiatu chrzanowskiego (37,90%).

**Drugi wymiar** aktywności mieszkańców gmin ma także charakter ekonomiczny, chociaż dotyczy władz gminnych, a w szczególności sumy pozyskanych przez nie środków w latach 1999-2004. W niniejszym opracowaniu uwzględniono nie tylko środki budżetowe (w tym te z Kontraktów Regionalnych), ale również środki pomocowe (m.in. PAOW, TEMPUS, PHARE), środki z Wojewódzkiego Funduszu Ochrony Środowiska i Gospodarki Wodnej (WFOŚiGW), środki z Urzędu Kultury Fizycznej i Sportu (UKFiS), środki z ZDW, MZMiUW oraz tzw. środki na projekty miękkie (kultura, polityka społeczna, sport)<sup>6</sup>.

Na środki pomocowe pozyskane przez gminy w latach 1999-2004 składały się m.in. środki z programów: INRED, STRUDER II, RAPID, TEMPUS i SOCRATES, TURIN III, INICJATYWA II, PAOW, ODBUDOWA I i II, PHARE CBC i PHARE SSG. Łączna suma środków pomocowych wynosiła 299 945 111,35 zł. Tylko 6 gmin nie otrzymało żadnej pomocy, a 11 uzyskało środki przekraczające 6,2 mln zł. Kwota środków z WFOŚiGW pozyskanych w latach 1999-2004 wynosiła 593 027 973 zł. Były to wydatki związane m.in. z budową kanalizacji sanitarnej, oczyszczalni ścieków, kotłowni itp. 13 gmin ze 182 nie otrzymało żadnych pieniędzy na ten cel, a 84 otrzymało więcej niż 2 mln zł. W latach 1999-2004 gminy województwa małopolskiego pozyskały 85 083 000 zł z Urzędu Kultury Fizycznej i Sportu (UKFiS). Kwota ta została przede wszystkim przeznaczona na budowę, rozbudowę i modernizację sal sportowych i pływalni. Stosunkowo duża liczba gmin (60) nie otrzymała żadnych środków na ten cel.

<sup>4</sup> Udział aktywnych zawodowo (pracujących i bezrobotnych) w ogólnej liczbie aktywnych zawodowo i biernych zawodowo; terminy „aktywny” i „bierny” zawodowo są stosowane przez GUS, EUROSTAT i Międzynarodową Organizację Pracy; **aktywnym zawodowo** jest osoba powyżej 15 roku życia pracująca lub bezrobotna – można go w skrócie nazwać „siłą roboczą”; **biernym zawodowo** jest osoba powyżej 15 roku życia, która nie została sklasyfikowana jako pracująca lub bezrobotna, czyli ta, która w badanym tygodniu 1) nie pracowała, nie miała pracy i jej nie poszukiwała, 2) nie pracowała, poszukiwała pracy, ale nie była gotowa do jej podjęcia w ciągu 2 tygodni następujących po tygodniu badanym; 3) nie pracuje i nie poszukuje pracy, gdyż ma załatwioną i oczkuje na jej rozpoczęcie; wśród biernych zawodowo znajdują się osoby zniechęcone, które nie poszukują pracy, gdyż są przekonane, że jej nie znajdą [źródło: GUS].

<sup>5</sup> Wskaźnik zatrudnienia to udział aktywnych zawodowo pracujących w ogólnej liczbie aktywnych (pracujących i bezrobotnych) i biernych zawodowo.

<sup>6</sup> Dane o środkach pozyskanych przez gminy pochodzą z Urzędu Marszałkowskiego woj. małopolskiego.



Suma środków przeznaczonych na inwestycje drogowe i mostowe oraz remonty kapitalne dróg i mostów (ZDF + IS)<sup>7</sup> wynosiła w latach 1999-2004 – 216 978 693,74 zł. Aż 46% gmin województwa nie otrzymało na ten cel żadnych środków. Łączna kwota środków uzyskanych z Małopolskiego Zarządu Melioracji i Urządzeń Wodnych w Krakowie (MZMiUW) wynosiła w latach 1999-2004: 300 171 910,58 zł. Pieniądze zostały wykorzystane głównie na regulację rzek, potoków, usuwanie szkód powodziowych, budowę stopni wodnych, wzmacnianie wałów przeciwpowodziowych, zabudowę wyrw itp. Spośród 182 małopolskich gmin środki otrzymało 110.

Za „projekty miękkie” uznaje się środki przeznaczone na kulturę, sport i politykę społeczną. Wydatki na kulturę związane są m.in. z odbudową i konserwacją zabytków, finansowaniem imprez kulturalnych (festiwali, koncertów, konkursów) oraz wspieraniem działalności towarzystw i organizacji prowadzących działalność kulturalną. Wydatki na sport to głównie finansowanie imprez sportowych oraz wspieranie zrzeszeń i klubów sportowych, natomiast wydatki na politykę społeczną to m.in. wspieranie związków, stowarzyszeń, fundacji i towarzystw, których działalność związana jest z pomocą osobom niepełnosprawnym, pochodzącym z rodzin patologicznych, uzależnionym itp. Suma środków na „projekty miękkie” w latach 1999-2004 wynosiła 15 939 943 zł, przy czym 63 gminy nie otrzymały żadnych pieniędzy na ten cel.

Łączna suma środków (ze wszystkich źródeł) pozyskanych przez gminy woj. małopolskiego wynosiła w latach 1999-2004: 1 851 389 108,96 zł. Dane z tabeli 1 pokazują, że aż 40% gmin uzyskało w tym czasie kwotę mniejszą niż 5 mln 176 tys. zł.

Tabela. 1. Wszystkie środki pozyskane przez gminy woj. Małopolskiego w latach 1999-2004.

SUMA ŚRODKÓW	Liczba gmin	Skumulowana liczba gmin	Procent	Skumulowany procent
0 zł	1	1	1%	1%
(0,01 zł - 5 176 148,62 zł)	72	73	40%	40%
(5 176 148,63 zł - 15 528 445,86 zł)	84	157	46%	86%
(15 528 445,87 zł - 25 880 743,10 zł)	17	174	9%	96%
(25 880 743,11 zł - 36 233 040,34 zł)	3	177	2%	97%
> 36 233 040,34 zł	5	182	3%	100%

Źródło: opracowanie własne na podstawie danych z Urzędu Marszałkowskiego

**Trzeci wymiar** aktywności mieszkańców gmin woj. małopolskiego związany jest z ich obywatelskimi postawami, czego wskaźnikiem może być frekwencja w kolejnych wyborach w latach 2000-2004. Analizie poddano frekwencję z pięciu kolejnych wyborów: 1) prezydenckich z 2000 r., 2) parlamentarnych z 2001 r., 3) samorządowych z 2002 r., 4) z referendum unijnego z 2003 r. oraz 5) wyborów do Parlamentu Europejskiego w 2004 r.

<sup>7</sup> W tym środki na zadania realizowane w ramach Kontraktu Wojewódzkiego.



W 2000 r. odbyły się wybory Prezydenta RP. Frekwencja w całym woj. małopolskim wynosiła ok. 62 %, choć zróżnicowanie między poszczególnymi gminami było stosunkowo duże: od 46,19% w Raclawicach do 69,45% w Chełmku. W 2001 r. odbyły się wybory do Parlamentu RP. Spośród 2,42 mln uprawnionych do głosowania w woj. małopolskim do urn udało się 1,18 mln, co daje frekwencję na poziomie około 49%. Najwyższą frekwencję odnotowano w Wietrzychowicach (59%), zaś najniższą w Igołomii-Wawrzeńcycach (34%). W czasie kolejnych wyborów – samorządowych w 2002 r. – liczba uprawnionych do głosowania w woj. małopolskim wynosiła ok. 2,45 mln osób, a udział aktywnych (tych, którzy oddali głos) wynosił w skali województwa ok. 46%. Podobnie jak w latach poprzednich zróżnicowanie między gminami było dość duże. Najwyższą frekwencję zaobserwowano w Tymbarku (66,88%), zaś najniższą w Liskach (33,65%)

W 2003 r. odbyło się referendum unijne. Frekwencja w woj. małopolskim wynosiła około 60%, co oznaczało, że spośród 2,48 mln uprawnionych do głosowania, do urn udało się 1,48 mln osób. Różnica między najniższą frekwencją w regionie 37,39% (Igołomia-Wawrzeńcycy) a najwyższą frekwencją w regionie 68,52% (Oświęcim miasto) przekraczała 30 pkt. procentowych. W dniu 1 maja 2004 r. Polska została członkiem Unii Europejskiej, a wkrótce po tym wydarzeniu odbyły się wybory do Parlamentu Wspólnoty. Frekwencja w całym województwie małopolskim była względnie niska (najniższa ze wszystkich dotychczasowych od 2000 r.) i wynosiła ok. 23%. Gminą o najwyższym udziale głosujących był Kraków (30,62%), a gminą o najniższym udziale – ponownie Igołomia-Wawrzeńcycy (8,47%).

## Poziom życia w gminach woj. małopolskiego – analiza metodą *k*-średnich

Poziom życia w gminach będą tutaj opisywać następujące wskaźniki:

- ◆ ekonomiczne (przeciętne miesięczne wynagrodzenie w zł, dochody własne gminy na 1 osobę w zł);
- ◆ mieszkaniowe – wyrażone przeciętną powierzchnią użytkową mieszkania na 1 osobę w m<sup>2</sup>;
- ◆ „konsumpcyjne” (zużycie gazu z sieci w ciągu roku na 1 odbiorcę w m<sup>3</sup>, zużycie energii elektrycznej w ciągu roku na 1 odbiorcę w kW-h);
- ◆ związane z infrastrukturą techniczną (długość sieci wodociągowej w km, długość sieci kanalizacyjnej w km);
- ◆ dotyczące migracji ludności (przyjęto, że gminy o dodatnim saldzie migracji stałych można traktować jako bardziej atrakcyjne, jeśli chodzi o wybór miejsca zamieszkania, a tym samym jako gminy o wyższym poziomie życia).

Niestety nie udało się dotrzeć do innych bardzo istotnych wskaźników poziomu życia, jakimi są bez wątpienia: PKB *per capita* wg gmin czy wyposażenie gospodarstw domowych w dobra trwałe użytku. Pewnym utrudnieniem jest też fakt, że niektóre dane dotyczyły



powiatów (np. średnie miesięczne wynagrodzenie), co oznacza, że przyjęcie tych wartości dla poszczególnych gmin tych powiatów jest obarczone błędem. Niemniej jednak są to informacje na tyle istotne, że większą stratą byłoby całkowite wykluczenie ich z analizy.

Jednym z ekonomicznych wskaźników poziomu życia w regionie jest przeciętne miesięczne wynagrodzenie. Najwyższy poziom dochodów występuje w Krakowie (3 300,64 zł), zaś najniższy w gminach powiatu ziemskiego tarnowskiego (1 573,07 zł). Drugi wskaźnik – dochody własne gminy w przeliczeniu na 1 mieszkańca – charakteryzuje się bardzo dużym zróżnicowaniem: od nieco ponad 200 zł w Słopnicach (powiat limanowski) do ponad 2000 zł w Krakowie. Ważnym wyznacznikiem poziomu życia są zasoby mieszkaniowe danej gminy, a w szczególności przeciętna powierzchnia użytkowa mieszkania przypadająca na 1 osobę. Wskaźnik dla województwa wynosi 20,4 m<sup>2</sup>, podczas gdy w „najgorszej” pod tym względem gminie – Słopnicach (powiat limanowski) wynosi 14,4 m<sup>2</sup>, a w najlepszej – Zielonkach (powiat krakowski) 27,8 m<sup>2</sup>

Przez wskaźniki „konsumpcyjne” poziomu życia rozumiana jest wielkość zużycia gazu i energii elektrycznej w poszczególnych gminach. Wprawdzie intuicyjnie lepszym wskaźnikiem jest wyposażenie gospodarstw domowych w dobra trwałego użytku, np. samochody, zmywarki do naczyń itp., jednak dostęp do takich informacji okazał się niemożliwy. Wskaźniki te mają pewne wady. Po pierwsze są to informacje w skali powiatu, które przełożono na poszczególne gminy z tych powiatów, a po drugie nie w każdym przypadku informują o wysokim poziomie życia. Przykładem jest powiat tatrzański charakteryzujący się największym zużyciem energii, co wynika jednak z liczby odwiedzających ten region turystów. Pomimo tych niedogodności zdecydowano się włączyć te charakterystyki do analizy, ze względu na dużą wartość poznawczą. Największym zużyciem energii elektrycznej charakteryzują się gminy powiatu tatrzańskiego (3 446 kW-h na 1 osobę w roku), zaś najmniejszym – gminy powiatu dąbrowskiego (1 403 kW-h na 1 osobę w roku). Największym zużyciem gazu z sieci charakteryzują się gminy powiatu tatrzańskiego (2 934 m<sup>3</sup> na 1 osobę w roku), natomiast najmniejszym – gminy powiatu olkuskiego (515 m<sup>3</sup> na 1 osobę w roku).

„Infrastrukturalnymi” wskaźnikami poziomu życia są: długość sieci wodociągowej i długość sieci kanalizacyjnej w poszczególnych gminach. Długość sieci wodociągowej w woj. małopolskim jest bardzo zróżnicowana. Aż dziewięć gmin nie ma wodociągów w ogóle, a tylko w sześciu długość sieci przekracza 240 km. Gminami o najdłuższej sieci wodociągowej są: Kraków (1 094,9 km) i Wieliczka (318,0 km), zaś gminami o najkrótszej<sup>8</sup> sieci są: Budzów z powiatu nowosądeckiego (1,5 km) oraz Nawojowa z powiatu suskiego (0,6 km). Długość sieci kanalizacyjnej w poszczególnych gminach woj. małopolskiego jest także bardzo zróżnicowana. 35 gmin wcale nie posiada kanalizacji, a aż 100 gmin posiada sieć nie dłuższą niż 30-kilometrową. Gminami o najdłuższej sieci kanalizacyjnej są: Kraków (995,0 km) i Tarnów miasto (213,1 km), natomiast gminami o najkrótszej sieci są: Wietrzychowice z powiatu tarnowskiego (0,5 km) i Tokarnia z powiatu myślenickiego (0,3 km).

<sup>8</sup> Wyłączając gminy nie posiadające sieci wodociągowej.

Innym wskaźnikiem poziomu życia jest saldo migracji stałej. Można bowiem wnioskować, że gminy, gdzie jest ono wysokie, są pod jakimś względem atrakcyjne (chodzi o m.in. łatwość zdobycia pracy, wyższy poziom wynagrodzeń itp.), skoro przyciągają mieszkańców z innych okolic. I odwrotnie, jeśli gminy charakteryzują się wysokim ujemnym saldem migracji, tzn. że jakość życia jest tam niższa. Wyjątkiem są duże miasta z ujemnym saldem i ich przedmieścia z dodatnim, ale ogólnie rzecz ujmując (w skali całego województwa), analiza tego zjawiska pozwala wskazać regiony, gdzie żyje się „lepiej”. Gminami o najwyższym dodatnim saldzie migracji są: Kraków (+1 839 osób) i Wieliczka (+388 osób), natomiast gminami o najwyższym ujemnym saldzie migracji są: Gorlice miasto (-222 osoby) i Tarnów miasto (-552 osoby).

Za pomocą metody *k*-średnich pogrupowano 178 gmin woj. małopolskiego w 4 jednorodne skupienia<sup>9</sup>. Zmiennymi włączonymi do analizy były: 1) przeciętne miesięczne wynagrodzenie w zł (2002 r.), 2) dochody własne gminy na 1 mieszkańca w zł (2003 r.), 3) długość sieci kanalizacyjnej w km (2002 r.), 4) długość sieci wodociągowej w km (2002 r.), 5) powierzchnia użytkowa mieszkania na 1 osobę w m<sup>2</sup> (2002 r.) i 6) saldo migracji stałych (2002 r.)<sup>10</sup>.

Przed analizą wystandaryzowano zmienne i sprawdzono, czy korelacja między nimi nie jest zbyt wysoka. Analiza macierzy korelacji (rys. 1 – z lewej strony) pozwala stwierdzić, że współczynniki korelacji są względnie niskie (najwyższy istotny statystycznie przyjmuje wartość 0,42). Wyboru 4 skupisk dokonano w oparciu o analizę stosunku międzygrupowej sumy kwadratów odchyłeń do wewnątrzgrupowej sumy kwadratów odchyłeń dla wariantów od 2 do 9 skupisk (przykład analizy wariancji dla 4 skupień znajduje się na rys. 1 z prawej strony).

Zmienna	Korelacje					
	doch-S	migr-S	wod-S	kan-S	pensje-S	mieszk-S
doch-S	1,00	-0,02	0,24	0,40	0,32	0,20
migr-S	-0,02	1,00	0,29	0,05	0,19	0,32
wod-S	0,24	0,29	1,00	0,42	0,32	0,34
kan-S	0,40	0,05	0,42	1,00	0,07	-0,01
pensje-S	0,32	0,19	0,32	0,07	1,00	0,28
mieszk-S	0,20	0,32	0,34	-0,01	0,28	1,00

Zmienna	Analiza wariancji					
	Między SS	df	Wewn. SS	df	F	Istotn. p
doch-S	65,5071	3	111,4929	174	34,07765	0,000000
migr-S	102,1445	3	74,8555	174	79,14426	0,000000
wod-S	76,7402	3	100,2598	174	44,39395	0,000000
kan-S	104,0062	3	72,9938	174	82,64214	0,000000
pensje-S	82,2802	3	94,7198	174	50,38289	0,000000
mieszk-S	54,4417	3	122,5583	174	25,76423	0,000000

Rys. 1. Macierz korelacji między zmiennymi – wskaźnikami poziomu życia (z lewej strony) oraz analiza wariancji dla 4 skupisk (z prawej strony) – źródło: opracowanie własne

Ze względu na ograniczoną objętość pracy nie wymieniono zawartości poszczególnych skupień, a podano jedynie ich wartości średnie oraz wartości skrajne (minimalne i maksymalne).

<sup>9</sup> Z analizy wyłączono miasta na prawach powiatów, gdyż ze względu na skrajnie wysokie wartości niektórych zmiennych uznano je za przypadki odstające.

<sup>10</sup> Z analizy wyłączono zmienne związane ze zużyciem energii elektrycznej i gazu, gdyż skrajnie wysokie wartości w powiecie tatrzańskim powodowały, że powiat ten każdorazowo stanowił odrębne skupienie; jak wspomniano wcześniej, nie oznacza to wysokiego poziomu życia w tym regionie, gdyż konsumpcja energii wynika z liczby odwiedzających ten powiat turystów.

Tabela 2. Średnie wartości zmiennych w poszczególnych skupieniach.

skupienia	średnie wartości zmiennych w każdym skupieniu					
	przeciętne miesięczne wynagrodzenie w zł	dochody własne gminy na 1 mieszkańca w zł	długość sieci wodociągowej w km	długość sieci kanalizacyjnej w km	powierzchnia użytkowa mieszkania na 1 osobę w m <sup>2</sup>	saldo migracji stałych
skupienie 1 (104 gminy)	<u>średnia</u> <b>1728,4</b> <u>min</u> 1573,1 <u>max</u> 2139,6	<u>średnia</u> <b>433,68</b> <u>min</u> 206,30 <u>max</u> 1031,00	<u>średnia</u> <b>39,03</b> <u>min</u> 0,00 <u>max</u> 169,70	<u>średnia</u> <b>13,90</b> <u>min</u> 0,00 <u>max</u> 75,00	<u>średnia</u> <b>19,93</b> <u>min</u> 14,40 <u>max</u> 25,80	<u>średnia</u> <b>+5</b> <u>min</u> -96 <u>max</u> +142
skupienie 2 (52 gminy)	<u>średnia</u> <b>1967,2</b> <u>min</u> 1674,6 <u>max</u> 2139,6	<u>średnia</u> <b>601,79</b> <u>min</u> 280,70 <u>max</u> 1289,30	<u>średnia</u> <b>116,40</b> <u>min</u> 5,30 <u>max</u> 269,10	<u>średnia</u> <b>14,37</b> <u>min</u> 0,00 <u>max</u> 55,30	<u>średnia</u> <b>22,69</b> <u>min</u> 19,30 <u>max</u> 26,60	<u>średnia</u> <b>+19</b> <u>min</u> -44 <u>max</u> +165
skupienie 3 (16 gmin)	<u>średnia</u> <b>1803,1</b> <u>min</u> 1573,1 <u>max</u> 2069,1	<u>średnia</u> <b>893,57</b> <u>min</u> 286,90 <u>max</u> 1454,60	<u>średnia</u> <b>116,31</b> <u>min</u> 20,60 <u>max</u> 245,10	<u>średnia</u> <b>77,05</b> <u>min</u> 21,50 <u>max</u> 138,40	<u>średnia</u> <b>20,18</b> <u>min</u> 18,40 <u>max</u> 22,70	<u>średnia</u> <b>-67</b> <u>min</u> -222 <u>max</u> +55
skupienie 4 (6 gmin)	<u>średnia</u> <b>1986,1</b> <u>min</u> 1573,1 <u>max</u> 2139,6	<u>średnia</u> <b>834,17</b> <u>min</u> 577,40 <u>max</u> 1216,70	<u>średnia</u> <b>202,92</b> <u>min</u> 122,20 <u>max</u> 318,00	<u>średnia</u> <b>74,15</b> <u>min</u> 47,80 <u>max</u> 113,60	<u>średnia</u> <b>23,55</b> <u>min</u> 19,90 <u>max</u> 27,80	<u>średnia</u> <b>+256</b> <u>min</u> +126 <u>max</u> +388

Źródło: opracowanie własne.

## Aktywność mieszkańców a poziom życia w gminach woj. małopolskiego – analiza z wykorzystaniem drzew klasyfikacyjnych CART

Pierwsza próba zbadania zależności między aktywnością mieszkańców gmin a poziomem życia w tychże gminach opiera się na prostych współczynnikach korelacji Pearsona między wszystkimi wskaźnikami aktywności a wszystkimi wskaźnikami poziomu życia. Z badań wyłączono miasta na prawach powiatów: Kraków, Nowy Sącz i Tarnów, których zdecydowanie najwyższe wartości poszczególnych wskaźników spowodowały, że gminy te uznano za obserwacje nietypowe. Niestety analiza nie przyniosła zaskakujących wyników. Spośród związków istotnych statystycznie tylko kilka można uznać za średnio silne.



Dotyczą one:

- ◆ liczby podmiotów gospodarczych na 1000 mieszkańców i dochodów własnych gminy na 1 mieszkańca:  $r=0,58$  – średnio silny związek (im wyższy wskaźnik liczby podmiotów, tym wyższe dochody własne gminy);
- ◆ frekwencji w referendum unijnym i dochodów własnych gminy na 1 mieszkańca:  $r=0,48$  – średni związek (im zamożniejsza gmina, tym wyższa frekwencja w referendum unijnym<sup>11</sup>);
- ◆ wskaźnika zatrudnienia i powierzchni użytkowej mieszkania na 1 osobę:  $r=0,45$  – średni związek (im wyższy wskaźnik zatrudnienia, tym większa powierzchnia mieszkania na osobę);
- ◆ liczby podmiotów gospodarczych na 1000 mieszkańców i długości sieci kanalizacyjnej:  $r=0,42$  – średni związek (im dłuższa sieć kanalizacyjna, tym więcej podmiotów gosp. na 1000 mieszkańców<sup>12</sup>);
- ◆ frekwencji w wyborach samorządowych w 2002 r. i długości sieci kanalizacyjnej:  $r=-0,41$  – średni ujemny związek (im wyższa frekwencja w wyborach, tym mniej rozbudowana sieć kanalizacyjna i odwrotnie: im niższa frekwencja, tym większa długość sieci kanalizacyjnej<sup>13</sup>).

Jak widać, interpretacja współczynnika korelacji liniowej Pearsona, mimo względnej prostoty obliczeniowej, niekoniecznie jest oczywista. Związki między zmiennymi mogą być jednokierunkowe (A jest przyczyną B, B zależy od A), dwukierunkowe/wzajemne (A jest przyczyną B i jednocześnie B jest przyczyną A) albo pozorne (B jest pozornie zależne od A). Wnioskowanie musi być oparte na dobrej znajomości badanego zjawiska.

Innym sposobem zbadania zależności między wskaźnikami aktywności gmin a poziomem życia w tych gminach jest zastosowanie drzew klasyfikacyjnych. Z wcześniejszych analiz wynika, że gminy woj. małopolskiego można podzielić na 4 względnie jednorodne skupiska. Celem tej analizy będzie sprawdzenie, które zmienne decydują o przynależności danej gminy do danego skupiska. Na rys. 2 (poniżej) znajduje się model drzewa klasyfikacyjnego dla zmiennej zależnej „przynależność do skupiska”.

Za regułę podziału wybrano indeks Giniego, kryterium stopu – minimalną liczebność węzła końcowego (5 przypadków), sposób szacowania jakości modelu – 10-krotną walidację krzyżową. Prawdopodobieństwa *a priori* pojawienia się klas ustalono na równe, co jest równoznaczne z tym, że „szansa” na pojawienia się każdego skupiska jest

---

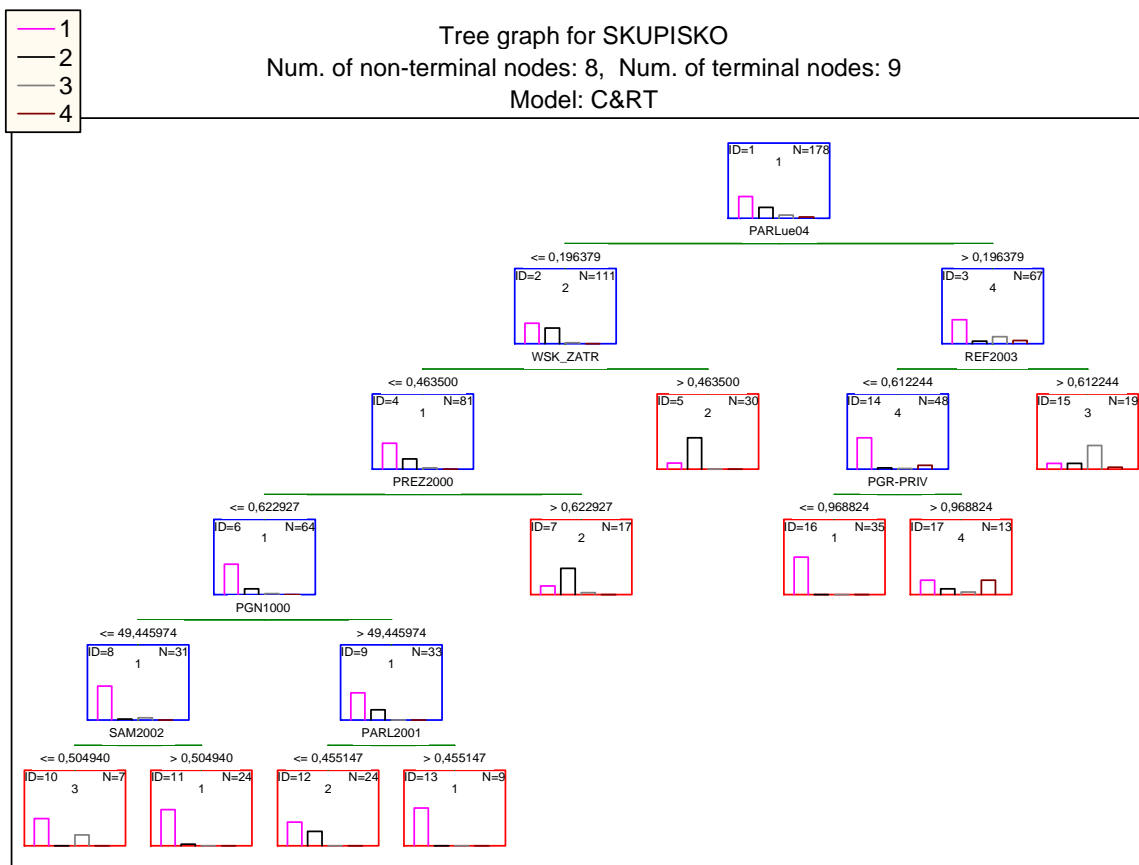
<sup>11</sup> Współczynnik korelacji informuje jedynie o sile związku między zmiennymi, nie wyjaśnia natomiast, co jest przyczyną, a co skutkiem.

<sup>12</sup> Tu z kolei można przypuszczać, że rozbudowana infrastruktura przyciąga przedsiębiorców, choć nie można wykluczyć, że to podatki płacone przez przedsiębiorców zwiększają możliwości inwestycyjne gminy, która dba o rozwój infrastruktury; jest to przykład tzw. zależności wzajemnej, gdzie A jest przyczyną B i jednocześnie A jest skutkiem B.

<sup>13</sup> Można tu przypuszczać, że w gminach, gdzie sieć kanalizacyjna jest słabo rozbudowana (lub nie ma jej wcale), mieszkańcy uczestniczą czynnie w wyborach władz gminnych w nadziei na poprawę tego stanu rzeczy.



jednakowa (0,25). Najłatwiejszym sposobem oceny dopasowania modelu do danych jest analiza macierzy błędnych klasyfikacji (rys. 3).



Rys. 2. Model drzewa klasyfikacyjnego (źródło: opracowanie własne)

Classification matrix						
Response: SKUPISKO						
Model: C&RT						
	Observed	Predicted 1	Predicted 2	Predicted 3	Predicted 4	Row Total
Number	1	67	24	8	5	104
Column Percentage		98.53%	33.80%	30.77%	38.46%	
Row Percentage		64.42%	23.08%	7.69%	4.81%	
Total Percentage		37.64%	13.48%	4.49%	2.81%	58.43%
Number	2	1	46	3	2	52
Column Percentage		1.47%	64.79%	11.54%	15.38%	
Row Percentage		1.92%	88.46%	5.77%	3.85%	
Total Percentage		0.56%	25.64%	1.69%	1.12%	29.21%
Number	3		1	14	1	16
Column Percentage		0.00%	1.41%	53.85%	7.69%	
Row Percentage		0.00%	6.25%	87.50%	6.25%	
Total Percentage		0.00%	0.56%	7.87%	0.56%	8.99%
Number	4			1	5	6
Column Percentage		0.00%	0.00%	3.85%	38.46%	
Row Percentage		0.00%	0.00%	16.67%	83.33%	
Total Percentage		0.00%	0.00%	0.56%	2.81%	3.37%
Count	All Groups	68	71	26	13	178
Total Percent		38.20%	39.89%	14.61%	7.30%	

Rys. 3. Macierz błędnych klasyfikacji (źródło: opracowanie własne)

Co istotne, w przypadku modeli *data mining*, o jakości modelu nie świadczy wyłącznie stosunek błędnie sklasyfikowanych przypadków do wszystkich przypadków (tu błąd predykcji wynosi 25,8%), ale trafność predykcji poszczególnych klas. Powyższy model „najlepiej” opisuje skupienie 2 (błąd równy 11,54%), skupienie 3 (błąd wynosi 12,5%), skupienie 4 (16,67%).

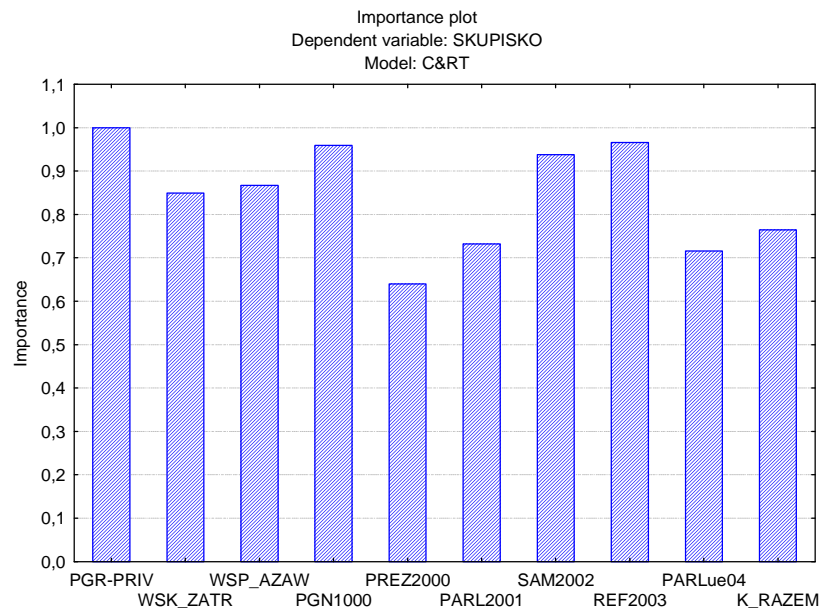
Na rys. 4 zamieszczono ranking ważności predyktorów. Każda zmienna niezależna – wskaźnik aktywności mieszkańców gmin – została porangowana na skali od 0 do 100. Im wyższa pozycja w tym rankingu (większa liczba punktów), tym większy wpływ danego wskaźnika aktywności mieszkańców na przynależność gminy do skupienia. Najważniejszymi wskaźnikami są:

- ◆ odsetek prywatnych podmiotów gospodarczych,
- ◆ frekwencja w referendum unijnym w 2003 r.,
- ◆ liczba podmiotów gospodarczych na 1000 mieszkańców,

zaś najmniej ważnymi:

- ◆ frekwencja w wyborach parlamentarnych w 2001 r.,
- ◆ frekwencja w wyborach do Parlamentu UE w 2004 r.,
- ◆ frekwencja w wyborach prezydenckich w 2000 r.

Precictor importance	
Response: SKUPISKO	
Model: C&RT	
Variable Rank	Importance
PGR-PRIV	100 1,000000
WSK_ZATR	85 0,849197
WSP_AZAW	87 0,867023
PGN1000	96 0,959321
PREZ2000	64 0,639882
PARL2001	73 0,732000
SAM2002	94 0,937929
REF2003	97 0,966128
PARLue04	72 0,715907
K_RAZEM	76 0,764701



Rys. 4. Ranking ważności predyktorów (źródło: opracowanie własne)

Każdy model drzewa można zamienić na zestaw zdań warunkowych typu „jeżeli..., to...”. Liczba liści jest zawsze równa liczbie reguł, co przy rozbudowanym modelu sprawia, że ich wygenerowanie bywa dosyć kłopotliwe. Dlatego warto najpierw sprawdzić, które węzły są najbardziej interesujące z punktu widzenia problemu badawczego (rys. 5 z lewej strony), a następnie opisać je, bądź to odczytując prosto z modelu, bądź korzystając z generatora reguł w języku SQL (rys. 5. z prawej strony).



Result of terminal nodes					
Response: SKUPISKO					
Model: C&RT					
	Class 1	Class 2	Class 3	Class 4	Gain
10	5	0	2	0	7,00000
11	23	1	0	0	24,00000
12	15	9	0	0	24,00000
13	9	0	0	0	9,00000
7	4	12	1	0	17,00000
5	5	25	0	0	30,00000
16	35	0	0	0	35,00000
17	5	2	1	5	13,00000
15	3	3	12	1	19,00000

```

/* Selecting cases related to Node 16 */
SELECT * FROM <TABLE>
WHERE ( (PARLue04 > 1.96378560425794e-001)
And (REF2003 <= 6.12244386682518e-001)
And ("PGR-PRIV" <= 9.68824459210257e-001)
);

/* Assigning values related to Node 16 */
UPDATE <TABLE>
SET NODEID=16, PREDVAL = '1', PROBVAL = 1.00000000000000e+000
WHERE ( (PARLue04 > 1.96378560425794e-001)
And (REF2003 <= 6.12244386682518e-001)
And ("PGR-PRIV" <= 9.68824459210257e-001)
);

```

Rys. 5. Zawartość węzłów końcowych i przykładowa reguła w języku SQL  
(źródło: opracowanie własne)

## Literatura

1. Grabiński T., Wydymus S., Zeliaś A., Metody taksonomii numerycznej w modelowaniu zjawisk społeczno-gospodarczych, PWN, Warszawa 1989.
2. *Województwo Małopolskie 2003*, Urząd Marszałkowski Województwa Małopolskiego, Kraków 2004.
3. Wybrane dane o powiatach i gminach województwa małopolskiego w 2002 r., Urząd Statystyczny w Krakowie, Kraków 2003.
4. Breiman L. i in., *Classification and Regression Trees*. Chapman and Hall, New York 1993.