

STATYSTYKA W PRAKTYCE

Dorota ROZMUS
Joanna TRZĘSIOK

Główne składowe rozwoju inteligentnego Polski¹

Streszczenie. *W systemie STRATEG, utworzonym przez GUS na potrzeby monitorowania polityki spójności, zgromadzono ponad 300 wskaźników do pomiaru rozwoju inteligentnego. Z uwagi na dużą liczbę zmiennych zasadne wydaje się zbadanie możliwości konstrukcji wskaźników syntetycznych jak najlepiej reprezentujących zmienne pierwotne. Celem badania, które przeprowadzono na podstawie danych z baz STRATEG i BDL za 2015 r., jest wskazanie — dzięki zastosowaniu analizy czynnikowej — głównych składowych rozwoju inteligentnego w Polsce, informujących o najważniejszych obszarach charakteryzowanych przez wskaźniki rozwoju inteligentnego.*

W efekcie uzyskano siedem głównych składowych, łącznie wyjaśniających ponad 94% wariacji pierwotnego zbioru wskaźników, dla których zaproponowano merytoryczną interpretację.

Słowa kluczowe: polityka spójności, rozwój inteligentny, metoda głównych składowych.

JEL: C38

Głównym celem polityki regionalnej Unii Europejskiej (UE) jest wyrównywanie poziomu gospodarczego poszczególnych regionów poprzez dofinansowywanie

¹ Artykuł opracowano na podstawie referatu wygłoszonego na konferencji *Rozwój gospodarczy i przestrzenny Polski a realizacja polityki spójności*, która odbyła się 25 i 26 maja 2017 r. w Katowicach.

rozwoju oraz pomoc słabiej rozwiniętym krajom. Takie zintegrowane działanie, noszące nazwę polityki spójności UE, jest prowadzone w czterech kierunkach:

- rozwoju zrównoważonego, który ma na celu wspieranie gospodarki konkurencyjnej, korzystającej z zasobów naturalnych w sposób racjonalny i przyjazny dla środowiska;
- rozwoju inteligentnego, czyli dążenia do postępu opartego na wiedzy i innowacjach;
- włączania społecznego, stawiającego sobie za cel wspieranie gospodarki opartej na wysokim poziomie zatrudnienia, zapewniającej spójność społeczną i gospodarczą;
- wzmacniania administracji, którego celem jest zwiększanie skuteczności działań administracji publicznej.

Artykuł poświęcono zagadnieniu rozwoju inteligentnego. Program ten koncentruje się na przedsięwzięciach służących budowie gospodarki opartej na wiedzy, konkurencyjnym sektorze nauki oraz efektywnych instytucjach otoczenia biznesu, które umożliwiają inteligentny rozwój gospodarczy kraju. Jego cele to:

- wsparcie podmiotów gospodarczych w działaniach badawczych oraz na rzecz innowacji;
- podniesienie jakości badań naukowych i prac rozwojowych oraz zwiększenie stopnia ich umiędzynarodowienia;
- zwiększenie stopnia przystosowania jednostek do warunków gospodarki rynkowej.

Na potrzeby programowania i monitorowania polityki spójności GUS stworzył system STRATEG², w którym zgromadzono wskaźniki wykorzystywane do oceny realizacji strategii obowiązującej w Polsce oraz UE. W zakresie rozwoju inteligentnego zaproponowano ponad 300 wskaźników, zebranych w pięć grup tematycznych. Z uwagi na dużą liczbę tych mierników zbadano możliwość konstrukcji zmiennych kompleksowych, które reprezentują zmienne pierwotne. Celem omawianego w artykule badania jest wskazanie głównych składowych rozwoju inteligentnego Polski, które informują o najważniejszych dziedzinach programu.

REDUKCJA LICZBY ZMIENNYCH

Analiza czynnikowa stanowi zespół metod i procedur statystycznych pozwalających na badanie wzajemnych relacji w zbiorze zmiennych oraz wykrywanie ukrytych uwarunkowań, które wyjaśniają występowanie tych zmiennych. Umożliwia sprowadzenie dużej liczby badanych zmiennych do mniej licznych zbioru wzajemnie niezależnych (nieskorelowanych) czynników. Wyodrębnione czynniki zachowują znaczną część informacji zawartych w zbiorze pierwotnym.

² <http://strateg.stat.gov.pl>.

Analiza głównych składowych (jedna z metod analizy czynnikowej) jest metodą transformacji zmiennych pierwotnych we wzajemnie nieortogonalne nowe zmienne, tzw. główne składowe, które są liniową kombinacją pierwotnych zmiennych. Dokładna analiza głównych składowych umożliwia wskazanie tych zmiennych początkowych, które mają duży wpływ na ich postać. Taka główna składowa jest wówczas reprezentantem tej grupy.

Metoda głównych składowych jest powszechnie znana, dlatego nie będzie tu dokładnie omówiona³ poza przedstawieniem algorytmu metody, który można zapisać następująco:

1. Dobór zmiennych, które zostaną wykorzystane w analizie.
2. Wybór kryterium określającego liczbę głównych składowych.
3. Wyodrębnienie głównych składowych.
4. Rotacja składowych w celu uzyskania łatwiejszej interpretacji.
5. Identyfikacja i interpretacja otrzymanych głównych składowych.

PRZEBIEG ANALIZY ORAZ WYNIKI

Przedmiotem badania, jak już wspomniano, był zestaw wskaźników opisujących rozwój inteligentny Polski. W systemie STRATEG zebrano ich ponad 300 i pogrupowano w pięć kategorii:

- jakość edukacji, umiejętności i uczenia się przez całe życie;
- jakość i dostępność technologii informacyjno-komunikacyjnych (ICT);
- podnoszenie konkurencyjności małych i średnich przedsiębiorstw, sektorów rolnego oraz rybołówstwa i akwakultury;
- wspieranie badań naukowych, rozwoju technologicznego i innowacji;
- zrównoważony transport i wysoka przepustowość kluczowych sieci infrastrukturalnych.

W pierwotnym zbiorze danych zgromadzono 214 wskaźników opisujących 16 województw Polski w 2015 r. Dane pochodziły z systemu STRATEG oraz Banku Danych Lokalnych GUS⁴. Tak duży zbiór pozwala na wieloaspektowe i złożone monitorowanie rozwoju inteligentnego Polski. Jednak wykorzystanie go w pierwotnej postaci do dalszej analizy, np. taksonomicznej, jest praktycznie niemożliwe, ponieważ zbyt duża liczba zmiennych przesądza o znikomej wartości poznawczej zbudowanego na nich modelu. Na ogół redukuje się zatem liczbę zmiennych, co można przeprowadzić na dwa sposoby:

- merytorycznie, najlepiej z pomocą eksperta, dobierając tylko część wskaźników do analizy;
- z wykorzystaniem analitycznych metod doboru zmiennych.

³ Szczegółowy opis metody można znaleźć m.in. w pracach: Kim i Mueller (1978a, 1978b); Walesiak i Gatnar (2009); Gatnar i Walesiak (2004).

⁴ <https://bdl.stat.gov.pl/bdl/start>.

W artykule zastosowano drugie podejście, a do redukcji posłużono się analizą głównych składowych.

Dobór zmiennych

W pierwszym etapie analizy głównych składowych określono, które z 214 pierwotnych zmiennych należy wprowadzić do modelu. W tym celu obliczono wartości miary Kaisera-Mayera-Olkina (*KMO*) (Zakrzewska, 1994):

$$KMO = \frac{\sum_j \sum_{h \neq j} r_{jh}^2}{\sum_j \sum_{h \neq j} r_{jh}^2 + \sum_j \sum_{h \neq j} \hat{r}_{jh}^2} \quad (1)$$

gdzie:

r_{jh} — współczynnik korelacji między zmiennymi o numerach j oraz h , dany wzorem

$$r_{jh} = \frac{cov_{jh}}{s_j s_h} \quad (2)$$

cov_{jh} — kowariancja między zmiennymi o numerach j oraz h ,

s_j, s_h — odchylenia standardowe odpowiednio dla zmiennych j oraz h ,

\hat{r}_{jh} — współczynnik korelacji cząstkowej między tymi zmiennymi

$$\hat{r}_{jh} = -\frac{R_{jh}}{\sqrt{R_{jj}R_{hh}}} \quad (3)$$

R_{jh}, R_{jj} oraz R_{hh} — dopełnienia algebraiczne odpowiednich elementów macierzy $\mathbf{R} = [r_{hj}]$.

KMO pozwala na sprawdzenie, w jakim stopniu zmienne są skorelowane. Niskie wartości tej miary wskazują na słabe związki między zmiennymi i tym samym niewielkie prawdopodobieństwo, że pozwolą one na utworzenie silnych i łatwych w interpretacji składowych (Walesiak i Gatnar, 2009). Często przyjmowaną w analizach wartością graniczną miary *KMO* jest 0,7. Wartości niższe od 0,7 wskazują na potrzebę usunięcia części zmiennych.

W celu eliminacji pojedynczych zmiennych oblicza się, w podobny sposób, miarę adekwatności doboru każdej indywidualnej zmiennej (MSA_h) (Zakrzewska, 1994):

$$MSA_h = \frac{\sum_{h \neq j} r_{jh}^2}{\sum_{h \neq j} r_{jh}^2 + \sum_{h \neq j} \hat{r}_{jh}^2} \quad (4)$$

Usuwa się te zmienne, które mają najniższe wartości MSA_h .

Wartość KMO obliczona dla całego zbioru danych wyniosła 0,682 (czyli nie przekraczała progu 0,7), co — razem ze zbyt dużą liczbą zmiennych — wskazywało na konieczność usunięcia części z nich. Eliminacji zmiennych dokonano na podstawie dwóch kryteriów:

- formalnego — polegającego na wykluczeniu wszystkich zmiennych, dla których $MSA_h < 0,6$;
- merytorycznego — na podstawie którego usuwano wskaźniki informujące o tym samym zjawisku⁵.

Ostatecznie w zbiorze pozostawiono 55 zmiennych⁶ opisujących rozwój inteligentny Polski.

Wybór kryterium określającego liczbę głównych składowych

W kolejnym kroku określono liczbę głównych składowych wyodrębnionych do dalszego badania. Najczęściej w tym celu wykorzystuje się kryteria: wartości własnej, wyjaśnionej wariancji, osypiska lub heurystyczne⁷. W omawianym badaniu zastosowano kryterium wartości własnej (np. Kim i Mueller, 1978b; Walesiak i Gatnar, 2009), które polega na pozostawieniu tych składowych, dla których spełniony jest warunek:

$$\lambda_i > 1 \quad (5)$$

gdzie λ_i jest wartością własną obliczoną dla i -tej składowej⁸.

Zastosowane kryterium opiera się na założeniu, że wartość własna reprezentuje ilość zmienności w oryginalnych zmiennych, która jest związana z daną główną składową. Oznacza to, że w modelu powinny pozostać tylko te składowe, które mają wartość własną większą od 1, gdyż te o wartości własnej mniejszej od 1 nie są bardziej przydatne niż pojedyncze zmienne, które po standaryzacji mają wariancję równą 1. Otrzymana główna składowa, jako kombinacja liniowa zmiennych pierwotnych, powinna wyjaśniać więcej zmienności niż pojedyncza zmienna.

⁵ Przykładowo, jeśli wśród zmiennych, dla których $MSA_h \geq 0,6$., znajdują się: *nakłady na działalność B+R* (badania i rozwój) w mln zł oraz *nakłady na B+R na 1 mieszkańca* w zł, to do analizy wprowadzamy tylko jedną zmienną, tę z wyższą wartością miary MSA_h . W tym przypadku była to zmienna *nakłady na B+R na 1 mieszkańca* w zł, dla której $MSA_h = 0,83$.

⁶ Zmienne te przedstawiono w tabl. 2.

⁷ Te kryteria szczegółowo opisują m.in. Zakrzewska (1994) oraz Walesiak i Gatnar (2009).

⁸ Obliczanie wartości własnych λ_i jest istotnym elementem analizy czynnikowej. Więcej informacji na ten temat można znaleźć m.in. w pracy Gatnar i Walesiak (2004).

Okazało się, że dla siedmiu głównych składowych obliczone wartości własne są większe od 1, zatem te właśnie składowe wyodrębniono, aby je poddać dalszej analizie.

Wyodrębnienie głównych składowych

Siedem głównych składowych łącznie wyjaśniało aż 94,19% wariacji wszystkich zmiennych, z czego pierwsza składowa — ponad 38%. Obliczone dla każdej składowej wartości własne, odsetek wariacji wyjaśnionej przez poszczególne składowe oraz skumulowany odsetek wariacji po wyodrębnieniu siedmiu głównych składowych przedstawiono w tabl. 1. Pogrubioną czcionką zaznaczono te składowe, których wartość własna jest większa od 1 i które starano się zinterpretować.

TABL. 1. WYNIKI ANALIZY GŁÓWNYCH SKŁADOWYCH

Składowe	Wartości po wyodrębnieniu głównych składowych		
	wartości własne	odsetek wyjaśnionej wariacji	odsetek skumulowany wariacji
1	38,426	69,865	69,865
2	4,686	8,519	78,384
3	2,668	4,850	83,234
4	1,819	3,306	86,541
5	1,622	2,949	89,489
6	1,360	2,473	91,962
7	1,223	2,224	94,186
8	0,876	1,594	95,779
9	0,610	1,110	96,889
10	0,451	0,821	97,710
11	0,408	0,741	98,451
12	0,283	0,514	98,965
13	0,253	0,461	99,425
14	0,187	0,341	99,766
15	0,129	0,234	100,000
...
55	0,000	0,000	100,000

Źródło: obliczenia własne z wykorzystaniem programu SPSS.

Rotacja składowych

Obliczono również ładunki czynnikowe, czyli współczynniki dla głównych składowych (a_{ji}), które odgrywają istotną rolę w ich interpretacji, ponieważ można je traktować jak współczynniki korelacji pomiędzy zestandaryzowaną zmienną pierwotną Z_j a główną składową S_i (Gatnar i Walesiak, 2004):

$$a_{ji} = r_{Z_j S_i} \quad (6)$$

Klarowną interpretację składowych umożliwia sytuacja, gdy każda zmienna ma wysoki ładunek czynnikowy dla jednej tylko głównej składowej i jednocześnie

dla każdej składowej można wskazać zmienne, które mają z nią wysokie ładunki a_{ji} . W praktyce oznacza to, że poszukiwane są takie zmienne Z_j , dla których $|a_{ji}| \rightarrow 1$ dla tylko jednej składowej S_i , zaś dla pozostałych wartość ta jest jak najbliższa 0. W przypadku gdy interpretacja wyodrębnionych składowych jest trudna, zaleca się zastosowanie rotacji, która ma na celu takie obrócenie układu osi, by uzyskana w jej efekcie macierz ładunków czynnikowych pozwalała na łatwiejsze ich objaśnienie.

Siedem głównych składowych uzyskanych w trakcie analizy okazało się trudnych do interpretacji. Wynikało to przede wszystkim z dużej liczby zmiennych i znacznej ich redukcji (z 55 do 7). Zastosowano więc rotację układu odniesienia, przy zachowaniu wcześniej wyznaczonej liczby głównych składowych oraz zasobów zmienności każdej składowej. Wykorzystano zarówno rotacje ortogonalne, jak i ukośne⁹. Na podstawie analizy otrzymanych macierzy ładunków czynnikowych stwierdzono, że najłatwiejsze do interpretacji są wyniki uzyskane za pomocą ukośnej rotacji *oblimin* (tabl. 2).

TABL. 2. MACIERZ ŁADUNKÓW CZYNNIKOWYCH PO ROTACJI OBLIMIN

Zmienne	Główne składowe						
	1	2	3	4	5	6	7
<i>Liczba jednostek prowadzących działalność B+R na 100 tys. ludności</i>	0,880	0,357	-0,103	-0,084	-0,145	-0,205	-0,044
<i>Udział kapitału zagranicznego ulokowanego w podmiotach w województwie w całkowitej wartości tego kapitału w Polsce ...</i>	0,864	-0,110	-0,043	0,039	-0,002	0,142	-0,045
<i>Wartość kapitału zagranicznego w podmiotach z udziałem kapitału zagranicznego w mln zł</i>	0,864	-0,110	-0,043	0,039	-0,002	0,142	-0,046
<i>Nakłady sektora przedsiębiorstw na działalność B+R w mln zł</i>	0,833	0,047	0,034	-0,068	-0,149	0,038	-0,088
<i>Liczba przedsiębiorstw z kapitałem zagranicznym</i>	0,796	-0,154	-0,048	-0,025	-0,020	0,179	-0,061
<i>Zatrudnieni w B+R w sektorze przedsiębiorstw</i>	0,789	0,108	-0,012	-0,085	-0,191	0,060	-0,061
<i>Ruch pasażerów w portach lotniczych — przyjazdy i wyjazdy w osobach</i>	0,770	-0,062	0,119	-0,106	-0,128	0,085	-0,161
<i>Słuchacze studiów podyplomowych na 10 tys. ludności w wieku 25—64 lat</i>	0,743	-0,162	0,028	0,081	-0,091	-0,045	-0,255
<i>Nakłady na działalność B+R na 1 mieszkańca</i>	0,722	0,044	0,097	0,008	-0,408	-0,069	-0,070
<i>Wartość brutto środków trwałych w przedsiębiorstwach na 1 mieszkańca w tys. zł</i>	0,720	-0,320	-0,144	-0,271	0,177	0,168	0,011
<i>Udział podmiotów gospodarki narodowej z udziałem kapitału zagranicznego w ogólnej liczbie podmiotów wpisanych do rejestru REGON</i>	0,699	-0,373	-0,069	-0,270	-0,030	0,112	0,084
<i>Nakłady na działalność innowacyjną w przedsiębiorstwach w tys. zł</i>	0,647	-0,031	-0,118	-0,147	-0,085	0,217	-0,090
<i>Widzowie i słuchacze w teatrach i instytucjach muzycznych</i>	0,638	-0,083	0,024	-0,113	-0,013	0,228	-0,278
<i>Stopa bezrobocia wśród absolwentów szkół wyższych według BAEL w osobach</i>	0,627	0,292	-0,039	0,208	-0,273	0,273	0,001

⁹ Różne metody rotacji opisano szczegółowo m.in. w pracy Gatnar i Walesiak (2004).

TABL. 2. MACIERZ ŁADUNKÓW CZYNNIKOWYCH PO ROTACJI OBLIMIN (cd.)

Zmienne	Główne składowe						
	1	2	3	4	5	6	7
Odsetek osób w wieku 15 lat i więcej z wyższym wykształceniem	0,571	-0,134	0,228	0,013	-0,420	0,073	-0,104
Długość linii krajowych regularnej komunikacji autobusowej w km	0,564	-0,360	-0,159	0,380	0,050	0,034	-0,267
Przedsiębiorstwa prowadzące e-sprzedaż poprzez stronę internetową	0,547	0,066	-0,206	-0,042	-0,123	0,313	-0,115
Liczba ofiar śmiertelnych w wypadkach drogowych	0,514	-0,033	-0,226	0,144	-0,175	0,237	-0,197
Przewozy pasażerów transportem samochodowym zarobkowym	0,506	0,006	-0,332	-0,037	0,044	-0,181	-0,436
Udział przedsiębiorstw z dostępem do Internetu o prędkości przynajmniej 100 Mb/s	0,497	0,003	0,119	-0,001	-0,127	0,158	-0,400
Przewozy pasażerów środkami komunikacji miejskiej w mln osób	0,470	-0,080	-0,067	-0,052	-0,215	0,237	-0,260
Udział podmiotów gospodarki narodowej o liczbie pracujących do 9 osób w ogólnej liczbie podmiotów wpisanych do rejestru REGON	0,415	-0,034	-0,223	-0,111	-0,216	0,263	-0,190
Liczba przedsiębiorstw niefinansowych o liczbie pracujących do 9 osób (mikroprzedsiębiorstwa)	0,374	0,000	-0,254	-0,091	-0,248	0,271	-0,188
Produkcja budowlano-montażowa (zrealizowana na terenie województwa) ogółem w tys. zł	0,362	0,003	-0,173	-0,150	-0,305	0,233	-0,202
Liczba nowo zarejestrowanych podmiotów gospodarki narodowej w sektorze prywatnym	0,322	-0,050	-0,250	-0,138	-0,304	0,206	-0,219
Współczynnik skolaryzacji brutto (szkoły zawodowe i ogólnozawodowe)	0,044	0,932	0,058	-0,021	0,180	-0,072	0,063
Odsetek uczniów szkół średnich zawodowych razem wśród uczniów szkół ponadgimnazjalnych ogółem	-0,030	0,755	0,245	-0,104	0,112	-0,045	0,333
Drogi publiczne ogółem o nawierzchni twardej ulepszonej na 100 km ² w km	-0,084	0,697	-0,178	-0,123	-0,207	0,327	-0,196
Liczba podmiotów gospodarki narodowej w rejestrze REGON na 1000 mieszkańców w wieku produkcyjnym	0,264	-0,511	-0,135	-0,480	-0,176	-0,015	-0,070
Łączna długość dróg ekspresowych i autostrad w km	-0,225	-0,077	-0,876	-0,224	-0,103	0,119	-0,037
Przewóz ładunków transportem samochodowym (z zagranicy)	0,189	0,026	-0,533	-0,193	-0,012	0,361	-0,143
Długość dróg krajowych o twardej nawierzchni w km	0,377	-0,291	-0,415	0,248	-0,181	0,223	-0,073
Drogi publiczne o twardej nawierzchni na 10 tys. mieszkańców w km	0,036	-0,042	0,230	0,698	0,028	-0,131	0,290
Odsetek gospodarstw domowych wyposażonych w komputer osobisty z dostępem do Internetu	0,397	0,102	-0,176	-0,642	-0,193	-0,017	0,064
Łączna długość dróg wojewódzkich i powiatowych o twardej nawierzchni w km	0,290	-0,253	-0,335	0,466	-0,428	0,156	-0,002
Uczestnicy studiów doktoranckich na 10 tys. ludności	-0,063	-0,184	0,109	-0,060	-0,903	-0,011	-0,163
Relacja liczby studentów do liczby ludności w wieku 19—24 lata w osobach	-0,002	-0,157	-0,016	-0,109	-0,830	-0,085	-0,250
Udział zatrudnionych w B+R w pracujących ogółem	0,187	0,064	-0,068	-0,113	-0,817	0,009	0,057
Liczba zgłoszeń wynalazków krajowych w Urzędzie Patentowym RP na 1 mln mieszkańców	0,087	-0,103	-0,222	-0,224	-0,632	0,184	0,007
Linie kolejowe ogółem (eksploatowane) na 10 tys. mieszkańców w km	-0,104	-0,433	0,045	-0,215	0,599	-0,082	0,255

TABL. 2. MACIERZ ŁADUNKÓW CZYNNIKOWYCH PO ROTACJI OBLIMIN (dok.)

Zmienne	Główne składowe						
	1	2	3	4	5	6	7
Wydatki z budżetów powiatów na oświatę i wychowanie w zł	0,248	0,117	-0,458	0,028	-0,598	-0,057	-0,018
Imprezy organizowane przez domy, ośrodki kultury, kluby i świetlice	0,016	0,193	-0,183	-0,211	-0,558	0,346	-0,069
Absolwenci studiów na poziomie magisterskim ogółem	0,341	0,082	-0,130	-0,034	-0,526	0,156	-0,122
Liczba studentów szkół wyższych	0,352	0,020	-0,131	-0,054	-0,504	0,133	-0,149
Liczba pracowników naukowych pełnozatrudnionych w województwie z tytułem naukowym profesora	0,389	-0,044	-0,183	0,007	-0,450	0,134	-0,145
Liczba udzielonych patentów na wynalazki krajowe przez Urząd Patentowy RP	0,295	-0,022	-0,162	-0,027	-0,373	0,347	-0,122
Udział patentów udzielonych w województwie w liczbie udzielonych patentów na wynalazki krajowe przez Urząd Patentowy RP ogółem	0,300	-0,023	-0,162	-0,027	-0,371	0,347	-0,120
Liczba zbiorów elektronicznych w bibliotekach publicznych	0,064	-0,068	-0,128	0,009	0,175	0,922	-0,010
Udział zgłoszeń wynalazków w województwie w liczbie zgłoszeń wynalazków krajowych w Urzędzie Patentowym RP ogółem	0,034	0,205	-0,138	-0,136	0,040	0,679	-0,317
Odsetek osób korzystających z Internetu w kontaktach z administracją publiczną ...	0,175	-0,190	0,168	-0,139	-0,429	0,643	0,293
Udział przedsiębiorstw innowacyjnych w ogólnej liczbie przedsiębiorstw przemysłowych	-0,096	0,291	0,523	-0,163	-0,201	0,527	-0,211
Liczba udzielonych praw ochronnych na krajowe wzory użytkowe przez Urząd Patentowy RP	0,200	0,316	-0,194	-0,070	-0,119	0,402	-0,358
Liczba przedsiębiorstw niefinansowych o liczbie pracujących do 49 osób (małe przedsiębiorstwa)	0,236	0,187	-0,304	-0,132	-0,204	0,373	-0,216
Przedsiębiorstwa (powyżej 9 pracujących) posiadające stronę internetową w %	-0,017	-0,079	0,038	-0,079	-0,040	0,008	-0,946
Osoby dorosłe uczestniczące w kształceniu lub szkoleniu (w wieku 25—64 lata)	0,386	-0,280	0,368	-0,154	-0,276	-0,040	-0,467

Źródło: jak przy tabl. 1.

Identyfikacja i interpretacja głównych składowych

Otrzymane wyniki — pomimo zastosowania rotacji — nie pozwalają na jednoznaczny interpretację wszystkich głównych składowych. W niektórych przypadkach trudno o nadanie wyznaczonym składowym definitywnej i klarownej nazwy, gdyż opisywane są one przez zmienne przyporządkowane do różnych kategorii rozwoju inteligentnego. Ponadto część zmiennych nie bierze udziału w interpretacji głównych składowych, ponieważ obliczone dla nich ładunki czynnikowe a_{ji} wskazują na istotną korelację z więcej niż jedną składową.

Mimo trudności zaproponowano nazwy dla wszystkich otrzymanych głównych składowych. Propozycje te, wraz ze zmiennymi, które najlepiej charakteryzują składowe, przedstawiono w tabl. 3.

TABL. 3. GŁÓWNE SKŁADOWE OTRZYMANE W WYNIKU ANALIZY

Zmienne opisujące główne składowe	Ładunki czynnikowe
1 — nakłady na badania i rozwój oraz kapitał zagraniczny	
Liczba jednostek prowadzących działalność B+R na 100 tys. ludności	0,880
Udział kapitału zagranicznego ulokowanego w podmiotach w województwie w całkowitej wartości tego kapitału w Polsce	0,864
Wartość kapitału zagranicznego w podmiotach z udziałem kapitału zagranicznego w mln zł	0,864
Nakłady sektora przedsiębiorstw na działalność B+R w mln zł	0,833
Liczba przedsiębiorstw z kapitałem zagranicznym	0,796
Zatrudnieni w B+R w sektorze przedsiębiorstw	0,789
Ruch pasażerów w portach lotniczych — przyjazdy i wyjazdy w osobach	0,770
Słuchacze studiów podyplomowych na 10 tys. ludności w wieku 25—64 lata	0,743
Nakłady na działalność B+R na 1 mieszkańca	0,722
Wartość brutto środków trwałych w przedsiębiorstwach na 1 mieszkańca w tys. zł	0,720
Udział podmiotów gospodarki narodowej z udziałem kapitału zagranicznego w ogólnej liczbie podmiotów wpisanych do rejestru REGON	0,699
Nakłady na działalność innowacyjną w przedsiębiorstwach w tys. zł	0,647
Widzowie i słuchacze w teatrach i instytucjach muzycznych	0,638
Stopa bezrobocia wśród absolwentów szkół wyższych według BAEL w osobach	0,627
2 — skolaryzacja	
Współczynnik skolaryzacji brutto (szkoły zawodowe i ogólnozawodowe)	0,932
Odsetek uczniów szkół średnich zawodowych razem wśród uczniów szkół ponadgimnazjalnych ogółem	0,755
Drogi publiczne ogółem o nawierzchni twardej ulepszonej na 100 km ² w km	0,697
3 — transport	
Łączna długość dróg ekspresowych i autostrad w km	-0,876
Przewóz ładunków transportem samochodowym (z zagranicy)	-0,533
Długość dróg krajowych o twardej nawierzchni w km	-0,415
4 — postęp	
Drogi publiczne o twardej nawierzchni na 10 tys. mieszkańców w km	0,698
Odsetek gospodarstw domowych wyposażonych w komputer osobisty z dostępem do Internetu	-0,642
5 — szkolnictwo wyższe i innowacje	
Uczestnicy studiów doktoranckich na 10 tys. ludności	-0,903
Relacja liczby studentów do liczby ludności w wieku 19—24 lata w osobach	-0,830
Udział zatrudnionych w B+R w pracujących ogółem	-0,817
Liczba zgłoszeń wynalazków krajowych w Urzędzie Patentowym RP na 1 mln mieszkańców	-0,632
6 — technologie informacyjno-komunikacyjne	
Liczba zbiorów elektronicznych w bibliotekach publicznych	0,922
Udział zgłoszeń wynalazków w województwie w liczbie zgłoszeń wynalazków krajowych w Urzędzie Patentowym RP ogółem	0,679
Odsetek osób korzystających z Internetu w kontaktach z administracją publiczną	0,643
7 — innowacje w przedsiębiorstwach	
Przedsiębiorstwa (powyżej 9 pracujących) posiadające własną stronę internetową w %	-0,946

Podsumowanie

Przeprowadzona analiza pozwoliła na wyodrębnienie siedmiu głównych składowych, jednak ze względu na bardzo dużą liczbę zmiennych pierwotnych niełatwo było jednoznacznie zinterpretować uzyskane wyniki, choćby dlatego, że nie wszystkie zmienne można było przypisać tylko do jednej składowej. Problematyczne okazało się również znalezienie odpowiednich nazw dla głównych składowych. Analizując wyniki przedstawione w tabl. 3, można zauważyć, że pierwsza składowa jest skorelowana głównie ze zmiennymi charakteryzującymi województwa pod względem działalności badawczo-rozwojowej, innowacyjnej oraz zaangażowania kapitału zagranicznego, choć obejmuje również informacje o liczbie widzów i słuchaczy w teatrach i instytucjach muzycznych oraz o stopie bezrobocia wśród absolwentów szkół wyższych. Wydaje się, że zmienne te nie są powiązane, jednak po głębszym zastanowieniu można dojść do wniosku, że wysokie nakłady na badania i rozwój generują zapotrzebowanie na odpowiednio wykształconych pracowników, w tym absolwentów szkół wyższych. Z kolei wyższy odsetek osób z wykształceniem wyższym w danym województwie rodzi zapotrzebowanie na odpowiednio sprofilowaną dla nich ofertę kulturalną. Ostatecznie (mając świadomość niedoskonałości) otrzymanym głównym składowym nadano nazwy: *nakłady na badania i rozwój oraz kapitał zagraniczny, skolaryzacja, transport, postęp, szkolnictwo wyższe i innowacje, technologie informacyjno-komunikacyjne oraz innowacje w przedsiębiorstwach*.

Uzyskane główne składowe w niewielkim stopniu są zgodne z podziałem zaproponowanym w systemie STRATEG. Zmienne wchodzące w skład większości składowych (oprócz transportu i innowacji w przedsiębiorstwach) pochodzą z co najmniej dwóch kategorii tego systemu. Najbardziej zróżnicowana pod tym względem jest pierwsza składowa, która obejmuje zmienne przypisane pierwotnie przez STRATEG aż do czterech różnych grup.

Pomimo problemów z interpretacją i nazewnictwem głównych składowych zaletą przedstawionego podejścia jest istotna redukcja zbioru danych. Nowe zmienne — główne składowe — można wykorzystać w dalszej analizie, np. taksonomicznej.

dr Dorota Rozmus, dr Joanna Trzęsiok — *Uniwersytet Ekonomiczny w Katowicach*

LITERATURA

- Gatnar, E., Walesiak, M. (red.). (2004). *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*. Wydawnictwo Akademii Ekonomicznej we Wrocławiu.
- Kim, J. O., Mueller, C. W. (1978a). *Factor Analysis. Statistical Methods and Practical Issues*. Beverly Hills: Sage.
- Kim, J. O., Mueller, C. W. (1978b). *Introduction to Factor Analysis. What it is and How to do it*. Beverly Hills: Sage.

Walesiak, M., Gatnar, E. (red.). (2009). *Statystyczna analiza danych z wykorzystaniem programu R*. Warszawa: Wydawnictwo Naukowe PWN.

Zakrzewska, M. (1994). *Analiza czynnikowa w budowaniu i sprawdzaniu modeli psychologicznych*. Poznań: Wydawnictwo Uniwersytetu Adama Mickiewicza.

Summary. *The STRATEG system created by Statistics Poland for monitoring cohesion policy, contains more than 300 indicators for measuring smart development. Due to a large number of variables, it seems appropriate to examine the possibility of constructing synthetic indicators that best represent the primary variables. The aim of this study, which was conducted on the basis of data for 2015 from the STRATEG and BDL databases, is to identify, owing to the use of factor analysis, the main components of smart development in Poland, informing about the most important areas characterised by the indicators of smart development.*

As a result 7 main components were obtained, which together account for more than 94% of variance from the original set of indicators and for which a substantive interpretation was proposed.

Keywords: cohesion policy, smart development, principal component analysis.