EkoMiasto#Gospodarka

Zrównoważony, inteligentny i partycypacyjny rozwój miasta

pod redakcją
Aleksandry Nowakowskiej

WYDAWNICTWO UNIWERSYTETU ŁÓDZKIEGO

FSS

norway grants

http://dx.doi.org/10.18778/7969-968-1.08

Elżbieta Antczak*

MODELOWANIE ROZWOJU GOSPODARCZEGO MIASTA

*Dr, Uniwersytet Łódzki, Wydział Ekonomiczno-Socjologiczny, Katedra Ekonometrii Przestrzennej, e-mail: wiszniewska@uni.lodz.pl

8.1. Metody ilościowe w ocenie rozwoju gospodarczego miasta

Miasta są miejscem rozwoju gospodarczego. Są ośrodkami koncentracji kapitału, innowacji, umiejętności i wiedzy. Z drugiej strony obszary miejskie tworzą także przestrzeń wielu dysproporcji oraz wzajemnych relacji społecznych, ekologicznych i ekonomicznych. Zwiększające się nierówności dochodowe (zob. rozdział *Zrozumieć fenomen miasta*), polaryzacja i segregacja społeczna (zob. *EkoMiasto#Społeczeństwo*, rozdział *Jakość życia w mieście*) i rozlewanie się miast (zob. rozdział *Ekonoma miasta*) prowadzą do obniżania się poziomu jakości życia w mieście. Jednym z możliwych rozwiązań problemów miasta jest zrównoważone podejście do ich rozwoju. Dzisiaj powiązania zewnętrzne miast, uwarunkowania lokalne, zależności przestrzenne stanowią o rozwoju jednostki nie tylko w skali lokalnej, ale i regionalnej. Granice administracyjne miast nie odzwierciedlają już warunków ich rozwoju, gdyż rozwój jednego obszaru miejskiego często determinuje poziom rozwoju innych jednostek w przestrzeni.

Dla zapewnienia zrównoważonego rozwoju obszarów miejskich niezbędne są diagnoza oraz analiza determinujących go wyzwań i czynników. Do tego celu przydatne stają się metody z zakresu statystyki opisowej (miary położenia, asymetrii, podobieństwa struktur) oraz miary współzależności (współczynniki korelacji). Z kolei syntetyczny obraz rozwoju miasta w odniesieniu do innych jednostek można uzyskać z zastosowaniem metod porządkowania liniowego (metody wzorca rozwoju). Kwantyfikacją mechanizmu powiązań między zjawiskami czy też innymi obszarami miejskimi (zależności przestrzennych) zajmuje się ekonometria (klasyczna i przestrzenna). Ostatecznie w celu określenia możliwości wykorzystania potencjału miast, szczególnie w ramach wdrażania strategii zrównoważonego rozwoju, wyznacza się prognozy krótko- i długookresowe.

8.2. Statystyczny obraz rozwoju miasta

Miary przeciętne (klasyczne, pozycyjne)

Najpowszechniejszymi miarami świadczącymi o podobieństwie jednostek są miary przeciętne (np. średnia arytmetyczna, mediana i dominanta). Średnia arytmetyczna jest ilorazem sumy wariantów obserwowanej cechy i liczebności analizowanej zbiorowości [Sobczyk, 2010, s. 49] (wzór opisujący tę miarę przedstawiono w: *EkoMiasto#Środowisko*, rozdział *Ocena i prognozowanie zmian środowiskowych*). Ciekawym wariantem opisanej wyżej średniej jest tzw. średnia arytmetyczna ważona. Stosuje się ją wówczas, gdy pewnym pomiarom chcemy nadać większe znaczenie [Stanisz, 2006, s. 116]:

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i \cdot w_i}{\sum_{i=1}^{k} w_i} \tag{8.1}$$

gdzie: $w_i > 0$ to wagi w postaci np. liczebności jednostek odpowiadającej poszczególnym wartościom zmiennej lub arbitralnie nadane inne wartości.

Zebrano dane dotyczące liczby nowo zarejestrowanych podmiotów na 10 tys. ludności w wieku produkcyjnym dla roku 2009 i 2013 oraz wybranych dwunastu miast w Polsce. Na podstawie informacji okazało się, że średnia liczba podmiotów cechujących zbiorowość w roku 2009 wynosiła 162 podmioty, a w 2013 roku 180 podmiotów (o 18 podmiotów więcej). Dodatkowo dla roku 2013 wyznaczono średnią ważoną, wzór (8.1), gdzie wagami (w_i) była liczba miast charakteryzująca się daną wielkością cechy.

Liczba nowo zarejestrowanych podmiotów na 10 tys. ludności w wieku produkcyjnym w 2013 roku

Wartość cechy (liczba podmiotów, x_i)	Liczba miast (w_i)	$x_1 \times w_1$
90	1	90
91	1	91
142	2	284
187	1	187
188	3	564
189	2	378
262	1	262
307	1	307
Suma	12	2163

Wartość średniej ważonej wynosi $\bar{x} = \frac{2163}{12} = 180$ podmiotów w 2013. Tu okazało się, że uzyskane wyniki opisujące średnią wielkość badanego zjawiska są identyczne. Jednakże w praktyce nadanie innych wag poszczególnym wariantom cechy może zmienić wartość średniej arytmetycznej.

Źródło: opracowanie własne na podstawie danych GUS.

Wartość średniej arytmetycznej (klasycznej, ważonej) jest wynikową wielkości wszystkich obserwacji, co oznacza, że duży wpływ na nią mają tzw. obserwacje ekstremalne, skrajne. Jest to zarówno wada, jak i zaleta tej miary [Stanisz, 2006, s. 116]. W przypadku występowania w badanym szeregu wyraźnie dużych i małych wartości większą wartość poznawczą ma mediana (kwartyl drugi, wartość środkowa, Me). Mediana jest wartością jednostki położnej w zbiorowości, tak że dzieli szereg obserwacji na dwie równe części. Medianę dla szeregu szczegółowego (pojęcie szeregu szczegółowego wyjaśniono w części *EkoMiasto#Środowisko*, rozdział *Ocena i prognozowanie zmian środowiskowych*) o nieparzystej liczbie obserwacji uzyskuje się, wskazując element środkowy:

$$Me = x_{\frac{N+1}{2}} \tag{8.2}$$

gdzie:

x – wartość cechy,

N – liczebność próby.

Natomiast dla szeregu o parzystej liczbie wyrazów medianę stanowi średnia arytmetyczna dwóch sąsiadujących ze sobą pomiarów środkowych [Kukuła, 2003, s. 14]:

$$Me = \frac{\left(x_{\frac{N}{2}} + x_{\frac{N}{2}+1} \right)}{2}$$

Wskazanie mediany

Na podstawie wykorzystanych wcześniej informacji dotyczących liczby nowo zarejestrowanych podmiotów wskażemy medianę nowo zarejestrowanych podmiotów na 10 tys. ludności w wieku produkcyjnym w 2013 roku i porównamy ją z wartością klasycznej średniej arytmetycznej. Dane uporządkowano w sposób rosnący.

Liczba nowo zarejestrowanych podmiotów na 10 tys. ludności w wieku produkcyjnym w 2013 roku i w wybranych miastach

Miasto	2013
Jastrzębie-Zdrój	90
Ruda Śląska	91
Gliwice	142
Sosnowiec	142
Łódź	187
Leszno	188
Legnica	188
Bielsko-Biała	188
Koszalin	189
Olsztyn	189
Poznań	262
Warszawa	307

Liczba obserwacji szeregu jest parzysta (12 analizowanych miast). Zatem wartość mediany wyznaczymy ze wzoru (8.3): $x_{\frac{N}{2}} = x_{\frac{12}{2}} = 188_6$ *i* $x_{\frac{N}{2}+1} = x_{\frac{12}{2}+1} = x_{\frac{13}{2}} = 188_7$ gdzie wartość mediany wynosi: $Me = \frac{188 + 188}{2} = 188$ nowo zarejestrowanych podmiotów na 10 tys. ludności w wieku produkcyjnym. Oznacza to, że w 2013 roku połowa analizowanych miast cechowała się liczbą badanych podmiotów nie większą bądź równą 188, i druga część obszarów miejskich posiadała dokładnie 188 podmiotów bądź mniej. Ze względu na występujące w szeregu obserwacje skrajne wartość mediany jest różna od wartości średniej arytmetycznej ($\bar{x} = 180$).

Źródło: opracowanie własne na podstawie danych GUS.

W grupie miar pozycyjnych, do których należy również opisana wyżej mediana, znajduje się też wartość najczęstsza, czyli dominanta (*Do*, modalna, *Mo*). Wyznaczanie wartości mediany i dominanty dla szeregów rozdzielczych przedziałowych obarczone jest pewnymi warunkami [więcej w: Sobczyk, 2010, s. 55–56]. Modalna jest wartością cechy statystycznej, która w badanym szeregu występuje najczęściej i której odpowiada największa liczebność [Suchecka, 2002, s. 28]. Dominanta ma szczególne zastosowanie przy zmiennych nominalnych (z których nie można wyznaczyć średniej ani mediany). Zmienna nominalna to taka zmienna, której wartości wyrażane są liczbowo. Niemniej jednak liczby te są umowne, zatem nie można wykonywać na nich działań arytmetycznych ani ich porównywać.

Wskazanie wartości najczęściej występującej (dominanty)

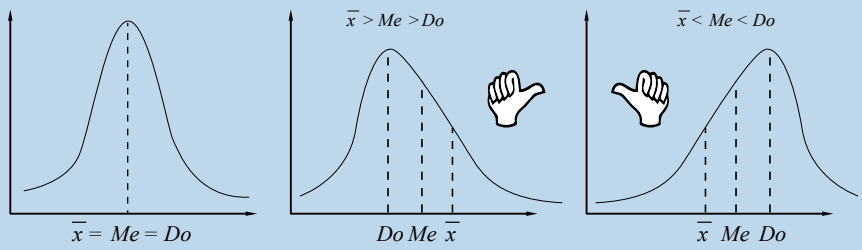
Korzystając z danych dotyczących nowo zarejestrowanych podmiotów, można stwierdzić, że w 2013 roku w badanym szeregu obserwacji występuje wartość najczęstsza. Dominującą liczbą nowo zarejestrowanych podmiotów na 10 tys. ludności w wieku produkcyjnym wśród badanych miast jest 188.

Źródło: obliczenia własne na podstawie danych GUS.

Miary asymetrii

Z punktu widzenia prowadzonej analizy statystycznej istotne są również informacje na temat zróżnicowania zbiorowości i różnic między jednostkami. Warto wówczas przeprowadzić badanie asymetrii (skośności) rozkładu. Asymetrię rozkładu (kierunek asymetrii) można określić, porównując wartości trzech miar: średniej arytmetycznej, mediany i dominanty [Sobczyk, 2007, s. 57]. Rozkład zbiorowości pod względem badanej cechy jest symetryczny, gdy wszystkie wymienione wyżej miary są identyczne. To oznacza, że przeważająca liczba jednostek tworzących zbiorowość ma wartości cechy równe przeciętnemu poziomowi zjawiska (zbiorowość nie jest zróżnicowana). Gdy spełniona jest nierówność: $\overline{x} > Me > Do$, to rozkład wykazuje asymetrię prawostronną, co oznacza, że przeważająca liczba jednostek ma wartości niższe od przeciętnego poziomu zjawiska (zbiorowość jest zróżnicowana pod względem cechy). Natomiast jeżeli spełniona jest nierówność: $\overline{x} < Me < Do$, wówczas rozkład cechuje się asymetrią lewostronną. Zatem przeważająca liczba jednostek ma wartości wyższe od przeciętnego poziomu zjawiska, czyli zbiorowość jest zróżnicowana pod względem cechy (rys. 8.1).

Rysunek 8.1. Położenie miar tendencji centralnej w rozkładach: symetrycznym i asymetrycznych



Źródło: opracowanie własne na podstawie Sobczyk, 2007, s. 58.

Zarówno o kierunku, jak i sile asymetrii mówi współczynnik asymetrii. Jest to miara umowna i niemianowana, co umożliwia porównanie ze sobą asymetrii różnych rozkładów [Sobczyk, 2007, s. 58]. Współczynnik ten można wyznaczyć ze wzoru:

$$As = \frac{\bar{x} - Do}{s} \tag{8.4}$$

gdzie:

$\bar{x}$ – średnia arytmetyczna,

Do – dominanta,

s – odchylenie standardowe (sposób obliczania i interpretacji miary odchylenia standardowego zob. *EkoMiasto#Środowisko*, rozdział *Ocena i prognozowanie zmian środowiskowych*).

Współczynnik asymetrii równy zeru wskazuje na symetrię rozkładu zmiennej, wartość dodatnia oznacza asymetrię prawostronną, a wartość ujemna – asymetrię lewostronną ($-1 \leq As \leq 1$). W celu interpretacji otrzymanego wyniku przyjmuje się następujące przedziały oceny siły skośności (wartości bezwzględnych) [Sobczyk, 2010, s. 70]: |0,0–0,2| – bardzo słaba, |0,3–0,4| – słaba, |0,5–0,6| – umiarkowana, |0,7–0,8| – silna, |0,9–1,0| – bardzo silna.

Ocena asymetrii rozkładu

Na podstawie informacji dotyczących liczby nowo zarejestrowanych podmiotów na 10 tys. ludności w wieku produkcyjnym w wybranych miastach w roku 2013 i miar asymetrii, wzór (8.4), dokonamy analizy zróżnicowania zbiorowości pod względem badanej cechy.

Wartości miar asymetrii cechujące liczbę nowo zarejestrowanych podmiotów na 10 tys. ludności w wieku produkcyjnym w wybranych miastach w 2013 roku

Miary	Wyniki obliczeń
Kierunek asymetrii	$\bar{x} < Me < Do$, $180 < 188 = 188$, zatem: $\bar{x} < Do$
Siła asymetrii (współczynnik asymetrii)	$As = \frac{\bar{x} - Do}{s} = \frac{180 - 188}{62} = -0{,}13$

Na podstawie wartości obliczonych miar można wnioskować o bardzo słabej asymetrii lewostronnej rozkładu analizowanych podmiotów w miastach Polski w 2013 roku. Większość ośrodków miejskich cechuje się liczbą podmiotów większą od przeciętnego poziomu zjawiska. Bardzo słaba siła asymetrii (współczynnik asymetrii, As = -0,13) świadczy zatem o niewielkim zróżnicowaniu zbiorowości miast pod względem analizowanej cechy

Źródło: obliczenia własne na podstawie GUS.

Wskaźnik struktury

Istotnym etapem analizy statystycznej jest badanie podobieństwa różnych zbiorowości ze względu na jedną cechę. W tym celu stosuje się tzw. wskaźnik podobieństwa struktur [Suchecka, 2002, s. 17]:

$$\omega_p = \sum_{i=1}^{k} min(\omega_{1i}, \omega_{2i}) \tag{8.5}$$

gdzie: $\omega_i = \frac{n_i}{n}$ – wskaźnik struktury będący stosunkiem liczby jednostek o danej wartości cechy (n_i) do łącznej liczebności próby (n). Suma wskaźników struktur równa się jedności (bądź 100%), a wartości wskaźnika spełniają nierówność: $0 \leq \omega_p \leq 1$; ω_{1i}, ω_{2i} to wskaźniki struktury jednej i drugiej zbiorowości (przy czym zbiorowości może być więcej niż dwie). Przyjmuje się następującą interpretację wartości wskaźnika podobieństwa struktur [Sobczyk, 2010, s. 81]: 0–0,2 – podobieństwo bardzo niskie, 0,3–0,4 – podobieństwo niskie, 0,5–0,6 – podobieństwo umiarkowane, 0,7–0,8 – podobieństwo duże, 0,9–1,0 – podobieństwo bardzo duże (jeżeli wartość wskaźnika wynosi jeden, świadczy to o identyczności struktur, gdy zero – o kompletnej odmienności).

Obliczanie wskaźników struktury i współczynnika podobieństwa

W celu weryfikacji skali podobieństwa miast pod względem nowo zarejestrowanych podmiotów na 10 tys. ludności w wieku produkcyjnym w roku 2009 i 2013 wykorzystamy dane zawarte w tabeli. Zamieszczono w niej również obliczenia pomocnicze służące zilustrowaniu sposobu wyznaczania wskaźnika podobieństwa struktur, wzór (8.5).

Rozkład liczby nowo zarejestrowanych podmiotów na 10 tys. ludności w wieku produkcyjnym w wybranych miastach w dwóch latach

Miasto	Liczba podmiotów		Wskaźniki struktury		min (ω_{1i}, ω_{2i}) (w %)
	2009	2013	2009 (ω_{1i}) (w %)	2013 (ω_{2i}) (w %)	
Jastrzębie-Zdrój	109	90	6	4	4
Ruda Śląska	101	91	5	4	4
Gliwice	147	142	8	7	7
Sosnowiec	126	142	6	7	6
Łódź	167	187	9	9	9
Leszno	180	188	9	9	9
Legnica	176	188	9	9	9
Bielsko-Biała	133	188	7	9	7
Koszalin	211	189	11	9	9
Olsztyn	176	189	9	9	9
Poznań	216	262	11	12	11
Warszawa	207	307	11	14	11
Suma	1949	2163	100	100	$\omega_p = 93$

Z otrzymanych obliczeń wynika, że struktura miast pod względem nowo zarejestrowanych podmiotów w roku 2009 i 2013 jest bardzo podobna (wartość współczynnika podobieństwa struktur wynosi 93%). Zatem można się również spodziewać niewielkich zmian w poziomie zjawiska na przełomie badanych okresów.

Źródło: opracowanie własne na podstawie danych GUS.

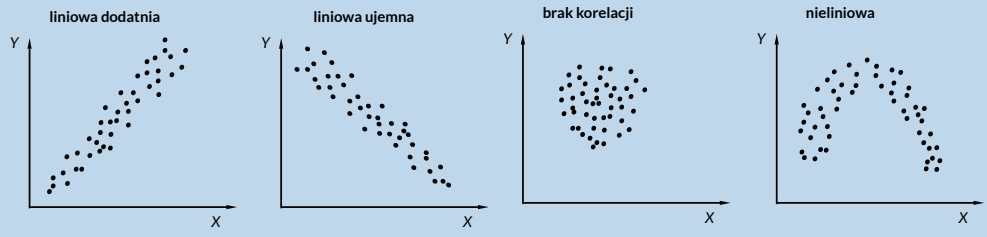
8.3. Analizy powiązań

W rzeczywistości dana zbiorowość opisywana jest nie przez jedną, ale wiele cech (mierzalnych i niemierzalnych). Co więcej, w praktyce czynniki charakteryzujące jednostki wzajemnie się warunkują [Sobczyk, 2007, s. 220]. Oznacza to, że przypuszczalnie istnieje między nimi korelacja. Zatem taki związek korelacyjny (zależność statystyczną, korelacyjną) charakteryzuje się tym, że konkretnym wartościom jednej zmiennej (cechom, zjawiskom) odpowiadają ściśle określone wartości średnie drugiej zmiennej [Kukuła, 2003, s. 149]. W ten sposób można ustalić, jak zmieni się poziom zmiennej, której wielkości chcemy objaśnić (zależnej, objaśnianej, *Y*) w zależności od zmiany wartości zmiennej objaśniającej, która objaśnia kształtowanie się *Y*. Dział statystyki służący określeniu, czy między badanymi zmiennymi zachodzą zależności, jaka jest ich siła, jaki jest ich kształt i kierunek nazywa się teorią korelacji (współzależności, powiązań). Badanie korelacji ma sens wówczas, gdy między zmiennymi istnieje logiczny, merytorycznie uzasadniony związek przyczynowo-skutkowy. Zależność między cechami może przybrać charakter liniowy bądź nieliniowy, dodatni, ujemny lub można również mówić o braku powiązań.

Wykresy korelacyjne

W celu zilustrowania związków korelacyjnych (wstępnego ich wykrycia) rysuje się tzw. wykresy korelacyjne. W prostokątnym układzie współrzędnych na osi odciętych zaznaczamy wartości jednej zmiennej, a na osi rzędnych wartości drugiej zmiennej. Punkty, które odpowiadają poszczególnym wariantom cechy, tworzą opisany wykres [Stanisz, 2006, s. 290] (rys. 8.2).

Rysunek 8.2. Wykresy korelacyjne – rodzaje korelacji pomiędzy cechami



Źródło: opracowanie własne na podstawie Stanisz, 2006, s. 291.

Dodatnia korelacja występuje wówczas, gdy wzrost wartości jednej zmiennej przyczynia się do wzrostu średnich wartości drugiej zmiennej. Korelacja ujemna ma miejsce, gdy wzrostowi wartości jednej zmiennej towarzyszy spadek średnich wartości drugiej zmiennej. Korelacja to powiązanie nie tylko dwóch cech. Zależność może dotyczyć jednej zmiennej objaśnianej i wielu objaśniających, wielu zmiennych zależnych i jednej niezależnej oraz wielu zmiennych objaśnianych i wielu objaśniających [Stanisz, 2006, s. 289]. Zatem współczynnik korelacji jest miernikiem kierunku i siły zależności między cechami. Po uprzednim stwierdzeniu o istotnej korelacji między cechami można przeprowadzić analizę regresji (analiza regresji zostanie omówiona w podrozdziale *Związki przyczynowo-skutkowe*).

W rezultacie otrzymamy odpowiedź na pytanie, jakiej średniej zmiany Y (zmiennej objaśnianej) można oczekiwać pod wpływem jednostkowej zmiany zmiennej niezależnej (X).

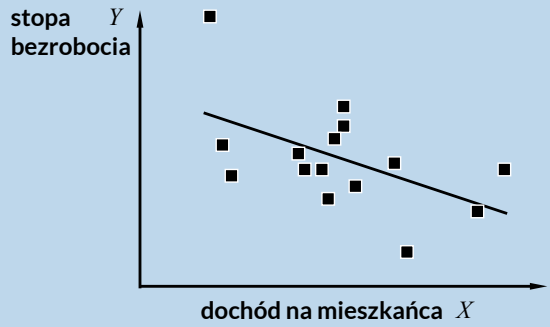
Analiza powiązań (korelacji) między zmiennymi

Zebrano dane dotyczące dochodów budżetów miast na prawach powiatu w zł na mieszkańca oraz stopy bezrobocia rejestrowanego w % dla roku 2013 i dla piętnastu polskich miast. Zakładając, że pomiędzy wielkością dochodów a stopą bezrobocia istnieje merytorycznie uzasadniony związek przyczynowo-skutkowy, dokonano analizy powiązań między zmiennymi.

Wielkość stopy bezrobocia i dochodów budżetów miast na prawach powiatu w wybranych miastach Polski i 2013 roku

Miasto	Stopa bezrobocia w %, y_i	Dochód w zł na mieszkańca, x_i
Warszawa	5	7114
Kraków	6	4922
Bielsko -Biała	6	4296
Tarnów	11	4758
Nowy Sącz	11	4559
Siedlce	12	4580
Skierniewice	12	4034
Łódź	12	4997
Piotrków Trybunalski	13	5215
Płock	13	5558
Częstochowa	14	4180
Ostrołęka	16	4565
Piekary Śląskie	17	3091
Bytom	21	3898
Radom	23	4505

Rysunek 8.2. Wykresy korelacyjne – rodzaje korelacji pomiędzy cechami



Opierając się na wykresach korelacyjnych zamieszczonych na rysunku 8.2, możemy stwierdzić, że odczyt informacji ze skonstruowanego wykresu korelacyjnego dla dwóch analizowanych zmiennych wskazuje na korelację ujemną. Zatem wzrost poziomu dochodów w zł na mieszkańca przyczynia się do spadku średniej wielkości stopy bezrobocia w roku 2013 w analizowanych miastach Polski.

Źródło: opracowanie własne na podstawie danych GUS.

Współczynnik korelacji cech mierzalnych

Do pomiaru siły, ale i kierunku związku liniowego dwóch cech mierzalnych stosuje się współczynnik korelacji liniowej Pearsona (r_{xy} lub r_{yx}):

$$r_{xy} = r_{yx} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_j - \bar{y})}{s(x)\cdot s(y)} \tag{8.6}$$

gdzie:

x_i, y_i – wartości cech i-tej i j-tej obserwacji,
x, y – średnie wartości zmiennych,
n – liczebność próby,
$s(x), s(y)$ – odchylenia standardowe.

Współczynnik korelacji liniowej przyjmuje wartości z przedziału [-1;+1]. Znak miernika informuje o kierunku zależności, a wartość bezwzględna o sile powiązań. W celu uproszczenia interpretacji siły zależności proponuje się następującą skalę [Kukuła, 2003, s. 150]: r_{xy} = 1 lub r_{xy} = -1 to korelacja doskonała, r_{xy} = 0 oznacza brak korelacji (bądź korelację nieliniową –wstępnie wynikającą z wykresu korelacyjnego), gdy $|r_{xy}| < 0{,}2 \rightarrow$ zależność niewyraźna, praktycznie brak zależności, $0{,}2 \le |r_{xy}| \le 0{,}4 \rightarrow$ zależność liniowa wyraźna, lecz niewielka, $0{,}4 \le |r_{xy}| \le 0{,}7 \rightarrow$ zależność liniowa wyraźna, $0{,}7 \le |r_{xy}| \le 0{,}9 \rightarrow$ zależność liniowa znacząca, $|r_{xy}| \ge 0{,}9 \rightarrow$ zależność liniowa silna.

Obliczanie siły i kierunku korelacji

Na podstawie zebranych danych dotyczących wielkości dochodów miast na prawach powiatu w zł na mieszkańca (Y) a stopą bezrobocia rejestrowanego w % (X), korzystając ze wzoru (8.6), obliczymy wartość współczynnika korelacji liniowej Pearsona wskazującą na siłę i kierunek zależności między tymi cechami. Po podstawieniu wielkości obliczeń pomocniczych do wzoru (8.6) wartość współczynnika korelacji wynosi:

$$r_{xy} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_j - \bar{y})}{s(x)\cdot s(y)} = \frac{\frac{1}{15}\cdot(-33800)}{893\cdot 5} = -0{,}50$$

Znak współczynnika świadczy o korelacji ujemnej pomiędzy analizowanymi zmiennymi, co stanowi potwierdzenie zależności z wykresu korelacyjnego, natomiast wartość bezwzględna miary mówi o wyraźnej zależności łączącej obie cechy.

Źródło: obliczenia własne.

Współczynnik korelacji cech niemierzalnych

Do ustalenia korelacji pomiędzy cechami, z których obie albo co najmniej jedna mają charakter jakościowy, wykorzystuje się tablice wielodzielne. Zamieszcza się w nich liczebności poszczególnych wariantów, a nie ich wartości [Sobczyk, 2007, s. 245]. Najprostszą tablicą wielodzielną jest tablica asocjacji (dwie kolumny na dwa wiersze) (tab. 8.1).

Tabela 8.1. Tablica asocjacji

	y_1	y_2	Razem
x_1 x_2	a c	b d	$a+b$ $c+d$
Razem	$a+c$	$b+d$	N

Źródło: opracowanie własne na podstawie Sobczyk, 2010, s. 120.

Symbole a, b, c, d oznaczają liczbę jednostek posiadających warianty cechy x_1, x_2, y_1 i y_2. W celu sprawdzenia istnienia związku pomiędzy cechami należy obliczyć jeden z wielu proponowanych współczynników kontyngencji [Sobczyk, 2010, s. 121], np. współczynnik Yule'a–Kendala:

$$Q = \frac{ad - bc}{ad + bc} \tag{8.7}$$

Współczynnik ten przyjmuje wartości od -1 do +1, znak wskazuje na kierunek korelacji, a wartość bezwzględna na siłę związku.

Korzystając z zebranych danych dotyczących dochodów miast na prawach powiatu w zł na mieszkańca i stopy bezrobocia rejestrowanego w %, dokonano pewnych obliczeń w celu przekształcenia tych zmiennych mierzalnych w zmienne wyrażone na skali nominalnej (jakościowe). Uprzednio wyznaczono średnie wielkości zmiennych cechujące analizowane miasta ($\bar{x} = 13\%$ oraz $\bar{y} = 4685$ zł) i ustalano następujące warianty cechy: x_1 – wielkość dochodu powyżej średniego poziomu zmiennej, x_2 – wielkość dochodu poniżej średniego poziomu z próby, y_1 – wielkość stopy bezrobocia powyżej średniego poziomu zmiennej, y_2 – wielkość stopy bezrobocia poniżej średniego poziomu z próby. Następnie wariantom przypisano odpowiednie liczebności (liczbę miast spełniających dane kryterium). Wyniki zamieszczono w tabeli asocjacji w celu zbadania korelacji między zmiennymi, korzystając ze wzoru na współczynnik korelacji Yule'a–Kendala, wzór (8.7).

Liczebności miast w zależności od kryterium dotyczącego średniej wielkości stopy bezrobocia i dochodu własnego miast na prawach powiatu w roku 2013

		Stopa bezrobocia, Y		
		Powyżej średniej	Poniżej średniej	
Dochód na mieszkańca, X		y_1	y_2	Razem
Powyżej średniej	x_1	4	4	8
Poniżej średniej	x_2	5	2	7
Razem		9	6	15

Korzystając z danych zawartych w tabeli, dokonano odpowiednich obliczeń, otrzymując wynik: $Q = \frac{ad - bc}{ad + bc} = \frac{8 - 20}{8 + 20} = \frac{-12}{28} = -0{,}43$. Wartość współczynnika Q wskazuje na korelację ujemną pomiędzy dochodem na mieszkańca a stopą bezrobocia (czyli wraz ze wzrostem średniego poziomu dochodu na mieszkańca w badanych miastach spada średnia wielkość stopy bezrobocia). Wartość bezwzględna miernika świadczy o wyraźnej zależności liniowej pomiędzy cechami (różnica w wartości współczynnika korelacji Q i Pearsona jest rezultatem utraty pewnej ilości informacji w wyniku dokonanej agregacji danych).

Źródło: opracowanie własne na podstawie danych GUS.

8.4. Wzorzec i antywzorzec rozwoju gospodarczego miasta

Z praktycznego punktu widzenia analizy rozwoju gospodarczego miast kwestią istotną staje się nie tylko badanie struktury oraz wzajemnych powiązań zjawisk warunkujących ten rozwój, ale także uzyskanie syntetycznego obrazu rozwoju wszystkich jednostek. Co więcej, opracowywanie koncepcji rozwoju gospodarczego miasta wymaga wiedzy o tym, czy dany ośrodek miejski wykazuje silniejszy rozwój od innej jednostki, czy zachodzi relacja odwrotna bądź równość w poziomach roz-

woju. Kompleksowe ujęcie rozwoju gospodarczego pozwala diagnozować dziedziny oraz jednostki, w których konieczne jest zintensyfikowanie działań strategicznych. Wskazanie wzorców i antywzorców rozwoju jest możliwe poprzez budowę taksonomicznych mierników rozwoju [Hellwig, 1968; Młodak, 2006, s. 119]. Na podstawie wiedzy merytorycznej (np. teorii ekonomicznej) sporządzana jest lista potencjalnych zmiennych opisujących analizowane zjawisko. W kolejnym etapie kandydatki dzieli się na stymulanty (zmienne, których wzrost wartości oznacza rozwój zjawiska), destymulanty (wzrost wartości nie sprzyja wzrostowi zjawiska) i nominanty (charakteryzują się pewnym optymalnym poziomem, od którego jakiekolwiek odchylenia traktowane są jako niekorzystnie wpływające na określone zjawisko) [Szanduła, 2014, s. 147]. Następnie liczbę zmiennych redukuje się za pomocą metod formalnych polegających na weryfikacji ich statystycznych właściwości (wartości współczynnika zmienności Ve oraz korelacji). Zmienna zostaje włączona do analizy dla $Ve > 10\%$ oraz niskiej wartości współczynnika korelacji.

Dobór zmiennych diagnostycznych

W celu zobrazowania rozwoju gospodarczego wybranych miast Polski w 2013 roku zebrano zmienne diagnostyczne: nakłady inwestycyjne w przedsiębiorstwach w zł na 1 mieszkańca w wieku produkcyjnym (*NI*), nowo zarejestrowane podmioty na 10 tys. ludności w wieku produkcyjnym (*NZP*), zmieszane odpady zebrane w ciągu roku w kg na mieszkańca (*ZO*), drogi powiatowe o twardej nawierzchni w km na 100 km^2 (*DP*), stopa bezrobocia rejestrowanego w % (*SB*), przeciętne miesięczne wynagrodzenia brutto w zł (*MW*). Następnie dokonano podziału cech ze względu na ich charakter. Do grupy stymulant zaliczono: *NI*, *NZP*, *DP*, *MW*, do zbioru destymulant: *ZO* i *SB*. Po weryfikacji statystycznych właściwości zmiennych ze zbioru usunięto: nakłady inwestycyjne w przedsiębiorstwach w zł na 1 mieszkańca w wieku produkcyjnym (stwierdzono wysoką korelację tej cechy ze stopą bezrobocia i przeciętnymi miesięcznymi wynagrodzeniami, współczynnik korelacji liniowej Pearsona przekroczył wartość bezwzględną równą 0,70. Wartości współczynników zmienności cech przekraczały 10%, gdyż mieściły się w przedziale [17%;72%]).

Źródło: opracowanie własne.

Zmienne diagnostyczne mają zazwyczaj różne miana, różny zakres zmienności, co uniemożliwia ich bezpośrednie porównanie. Zatem sprowadzamy je do porównywalnych form, stosując różne metody normalizacji, unitaryzacji czy standaryzacji [Młodak, 2006, s. 39–40].

Powszechnie w celu doprowadzenia zmiennych do wspólnego miana stosuje się wzór:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_i} \tag{8.8}$$

gdzie:

i – ta jednostka terytorialna (obiekt), dla $i = 1, 2, \ldots, n$,
j – ta zmienna (wartość cechy), dla: $j = 1, 2, \ldots, m$,
$\bar{x}_j$ – średnia arytmetyczna,
s_j – odchylenie standardowe.

Zmienne determinujące rozwój gospodarczy miast poddano standaryzacji (uzyskano w ten sposób cechy porównywalne, pozbawione jednostek).

Wartości cech diagnostycznych (przed standaryzacją i po standaryzacji) rozwoju gospodarczego wybranych miast dla 2013 roku

Miasto	*NZP*	*ZO*	*DP*	*SB*	*MW*	*NZP*	*ZO*	*DP*	*SB*	*MW*
	z_{ij}									
Łódź	5823	262	253	12	3711	-0,32	0,10	-0,97	-0,06	0,03
Piotrków Trybunalski	4844	174	270	13	2769	-0,50	-1,91	-0,71	0,03	-1,40
Skierniewice	1936	311	362	12	3082	-1,03	1,23	0,67	-0,10	-0,93
Płock	12737	300	213	13	4799	0,93	0,97	-1,57	0,07	1,69
Ostrołęka	19953	198	374	16	3913	2,24	-1,37	0,85	0,64	0,34
Siedlce	4054	284	408	12	3533	-0,65	0,61	1,35	-0,16	-0,24
Radom	4132	202	240	23	3523	-0,63	-1,29	-1,16	1,92	-0,26
Warszawa	17246	307	371	5	5226	1,75	1,14	0,80	-1,48	2,34
Kraków	7893	270	293	6	3998	0,05	0,28	-0,37	-1,29	0,47
Nowy Sącz	5919	250	372	11	3098	-0,31	-0,16	0,82	-0,37	-0,90
Tarnów	6441	259	324	11	3448	-0,21	0,04	0,10	-0,41	-0,37
Bielsko-Biała	9879	228	416	6	3781	0,41	-0,68	1,49	-1,20	0,14
Bytom	2398	256	252	21	3464	-0,95	-0,04	-0,99	1,60	-0,35
Piekary Śląskie	3310	303	298	17	3329	-0,78	1,05	-0,30	0,82	-0,55
x_j	7612	257	317	13	3691	$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$				
s_j	5508	43	66	5	656					

Źródło: opracowanie własne na podstawie danych GUS.

Zgodnie z założeniem Hellwiga konstrukcję miernika najpowszechniej opiera się na odległości euklidesowej obiektów od quasi (tymczasowego) wzorca rozwojowego, tj. $\phi_j = \phi_1, \phi_2, \ldots, \phi_m$:

$$d_i = \sqrt{\sum_{j=1}^{m} (z_{ij} - \phi_j)^2} \tag{8.9}$$

dla $i = 1, 2, \ldots, n$, gdzie: $\phi_j = \max_{i=1,2,\ldots,n} z_{ij}$, $j = 1, 2, \ldots, m$ dla cech o własnościach stymulant oraz $\phi_j = \min_{i=1,2,\ldots,n} z_{ij}$ dla cech o własnościach destymulant. Następnie docelowe wartości miar rozwoju wyznacza się na podstawie następującego wzoru:

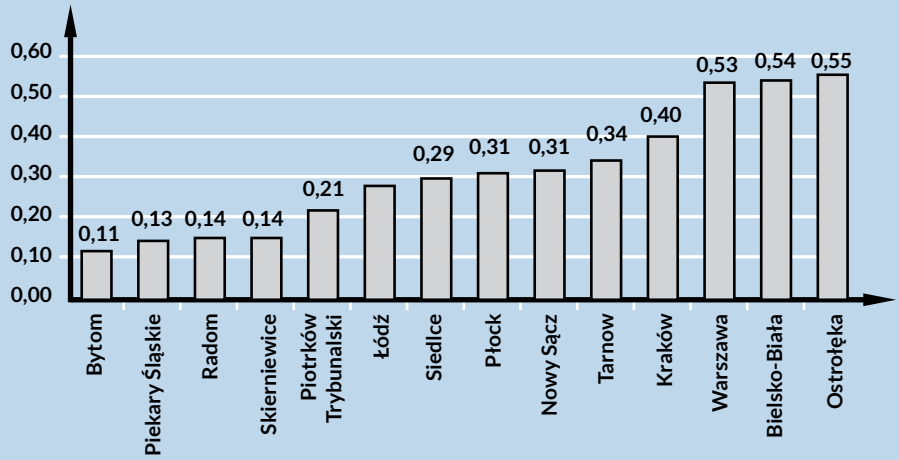
$$\mu_i = 1 - \frac{d_i}{d_-} \tag{8.10}$$

dla $i = 1, 2, \ldots, n$, gdzie: d_- – krytyczna (graniczna) odległość danej jednostki od wzorca $d_- = \bar{d} + 2s_d$ przy czym: $\bar{d}$ – średnia arytmetyczna wektora $\mathbf{d} = (d_1, d_2, \ldots, d_n)$, sd – odchylenie standardowe wektora $\mathbf{d}$ [Antczak, 2013, s. 37–53].

Na podstawie wartości taksonomicznych mierników rozwoju (zazwyczaj mieszczących się w przedziale [0;1]) porządkujemy obiekty tak, że im wyższy poziom badanego zjawiska w danym obiekcie, tym wyższa wartość miary rozwoju tej jednostki. W ten sposób wyznaczamy obiekt o najwyższym poziomie analizowanego zjawiska (z najwyższą wartością miary), który można uznać za empiryczny wzorzec rozwoju, jednostkę o niższym poziomie zjawiska (najniższej wartości miernika) – empiryczny antywzorzec rozwoju oraz obiekty przestrzenne znajdujące się pomiędzy wzorcem a antywzorcem rozwoju.

Syntetyczne miary rozwoju gospodarczego

Na podstawie danych opisujących rozwój gospodarczy miast i obliczeń według wzorów (8.9) i (8.10) wyznaczono wartości syntetycznych mierników rozwoju gospodarczego miast dla roku 2013.



Z wykresu można odczytać, że wzorcem rozwoju gospodarczego (pod względem zebranych determinant) w 2013 roku było miasto Ostrołęka (0,55), tuż za nią: Bielsko-Biała (0,54) i Warszawa (0,53). Z kolei antywzorcem rozwoju okazał się Bytom (0,11), a ostatnie pozycje w rankingu miast zajęły Piekary Śląskie (0,13), Radom (0,14) i Skierniewice (0,14).

Źródło: opracowanie własne.

8.5. Związki przyczynowo-skutkowe

Związek przyczynowo-skutkowy zmiennych, gdzie jedna jest przyczyną (zjawiskiem objaśniającym), a druga skutkiem (zjawiskiem objaśnianym) może być wyrażony w postaci funkcji łączącej rozpatrywane zjawiska (regresji). Dzięki temu możliwe jest przewidywanie kształtowania się wartości jednej zmiennej przy znanych wartościach drugiej zmiennej (regresja dwóch zmiennych) lub, wprowadzając do funckji większą liczbę zmiennych objaśniających (niezależnych) możemy mówić o regresji wielorakiej [Sobczyk, 2010, s. 128]. Zatem funkcja regresji jest narzędziem analitycznym (zapisanym w formie wzoru) służącym do badania przyczynowo-skutkowych powiązań między zmiennymi w taki sposób, że średnim wartościom zmiennej objaśnianej (zależnej, skutku) przypisane są konkretne wartości zmiennej objaśniającej (niezależnej, będącej regresorem, przyczyną) [Stanisz, 2006, s. 299].

Model ekonometryczny

Najprostszą funkcją regresji jest jednorównaniowy liniowy model ekonometryczny z jedną zmienną objaśniającą:

$$y_i = \alpha_0 + \alpha_1 \cdot x_i + \varepsilon_i \tag{8.11}$$

gdzie:

y_i – zmienna objaśniana,

x_i – zmienna objaśniająca,

i-ta obserwacja dla $i = 1, 2, \ldots, n$,

$n > 2$ – liczebność próby,

ε_i – składnik losowy (główne przyczyny występowania składnika losowego to: brak ważnych zmiennych objaśniających w modelu, postać analityczna funkcji może być nieadekwatna do rzeczywistych powiązań między zmiennymi, nieprzewidywalność zachowań podmiotów występujących na rynku, błędy obarczające dane statystyczne [Borkowski, Dudek Szczęsny, 2004, s. 18]),

α_0, α_1 – parametry strukturalne modelu.

Potwierdzenie zależności pomiędzy potencjalnymi zmiennymi

Zebrano dane dotyczące dochodów miast na prawach powiatu w zł na mieszkańca (x_1) oraz ilości zebranych zmieszanych odpadów komunalnych w kg na mieszkańca (y) w latach 2005–2013 w Warszawie. Wzór (8.11) wyraża wówczas liniową zależność ilości odpadów komunalnych od dochodów (współczynnik korelacji liniowej Pearsona wskazuje na ujemną zależność pomiędzy zmiennymi r_{xy} = - 0,74, a wartość bezwzględna na znaczącą siłę).

Źródło: opracowanie własne.

Weryfikacja modelu powinna obejmować weryfikację merytoryczną (merytoryczny sens znaków oszacowanych ocen parametrów) oraz statystyczną (jakość modelu).

Parametry strukturalne (estymacja, istotność statystyczna, interpretacja)

Wartości nieznanych parametrów modelu ekonometrycznego α_0, α_1 szacuje się (estymuje) klasyczną metodą najmniejszych kwadratów (KMNK). Idea tej metody

polega na wyznaczeniu takich oszacowań parametrów strukturalnych (α_0, α_1) będących ocenami (a_0, a_1), że suma kwadratów reszt modelu (różnica pomiędzy wartością empiryczną zmiennej objaśnianej (y_i) a jej wartością teoretyczną $\hat{y}_i$) $\sum_{i=1}^{n} e_i^2$) osiąga minimum (pojęcie reszt modelu zostało wyjaśnione w części *EkoMiasto#Środowisko*, rozdział *Ocena i prognozowanie zmian środowiskowych*; dokładany opis KMNK zob. np. w [Borkowski, Dudek, Szczęsny, 2004, s. 25–27]). Wartości teoretyczne oblicza się z funkcji (8.12), w wyniku czego otrzymujemy wartości ocen parametrów strukturalnych (estymatory) i wynik zapisujemy w postaci równania:

$$\hat{y}_i = a_0 + a_1 \cdot x_i \tag{8.12}$$

gdzie: $a_0 = \bar{y} - a_1\bar{x}$, $a_1 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$ [Kukuła, 2003, s. 157]. Ocena parametru stojącego przy zmiennej objaśniającej podlega interpretacji, ale uprzednio musimy ocenić istotność statystyczną wpływu zmiennej stojącej przy tym parametrze na zmienną objaśnianą. Do tego celu powszechnie stosuje się test *t*-Studenta [Kufel, 2007, s. 55]. Hipoteza zerowa dla tego testu ma postać: H_0: $a_j = 0$ (ocena parametru jest statystycznie nieistotna, równa zeru), przy hipotezie alternatywnej H_1: $a_j \neq 0$ (istotność statystyczna oceny parametru), a wartość statystyki wyznacza się ze wzoru: $t_j = a_j/S(a_j)$, gdzie: $S(a_j)$ – standardowy błąd szacunku parametru. Jeżeli hipoteza zerowa jest prawdziwa (z punktu widzenia istotności oceny parametru jest to sytuacja niekorzystna), to bezwzględna wartość statystyki nie powinna przekraczać wartości krytycznej (odczytanej z tablic *t*-Studenta): $|tj| < t^*$. Wówczas mówimy o braku istotnego wpływu zmiennej objaśniającej na zmienną objaśnianą. W sytuacji gdy $|tj| \geq t^*$, to hipotezę zerową odrzucamy, czyli ocena parametru statystycznie istotnie różni się od zera. Zatem zmienna niezależna ma istotny wpływ na kształtowanie się poziomu zmiennej zależnej.

Szacowanie ocen parametrów modelu ekonometrycznego

Na podstawie danych dotyczących wielkości dochodów w zł na mieszkańca (x_1) i ilości zebranych odpadów komunalnych w kilogramach na mieszkańca (y) w Warszawie w latach 2005–2013 oszacowano parametry funkcji zależności pomiędzy zmiennymi klasyczną metodą najmniejszych kwadratów i otrzymano równanie: $\hat{y} = 872 - 0{,}08 \cdot x_i$. Następnie dokonano weryfikacji istotności statystycznej ocen parametrów i okazało się, że wartość statystyki testu *t*-Studenta dla oszacowanych ocen parametrów ($t_{a0} = 7{,}1$, $t_{a1} = 3{,}8$) przewyższa wartość krytyczną $t^* = 1{,}9$). Zatem zmienna objaśniająca ma istotny wpływ na kształtowanie się wielkości zmiennej zależnej, tak że w latach 2005–2013 wzrost dochodów na mieszkańca o 1 zł w Warszawie powoduje spadek ilości zebranych odpadów średnio o 0,08 kg na mieszkańca. Może być to tłumaczone efektywnością działań podejmowanych w zakresie gospodarki odpadami (zbiórka selektywna, wyższy stopień świadomości ekologicznej ludności, promowanie idei ekorozwoju).

Źródło: opracowanie własne na podstawie danych GUS.

Po stwierdzeniu istotności statystycznej ocen parametrów w oszacowanym modelu (8.12) a_1 informuje, o ile wzrósł (gdy $a_1 > 0$) lub zmalał (gdy $a_1 < 0$) średni poziom zmiennej zależnej y pod wpływem zwiększania się zmiennej x o jednostkę, przy

innych czynnikach stałych. Ocena parametru a_0 nie podlega interpretacji, jednakże w celu zachowania walorów praktycznej przydatności funkcji regresji ocena ta musi być istotna statystycznie.

Ocena jakości funkcji regresji – parametry stochastyczne

Parametry stochastyczne służą do oceny zgodności modelu z danymi empirycznymi. Miary określające stopień efektywności oszacowań ocen parametrów z ich rzeczywistymi wartościami obliczane są na podstawie reszt (zapis opisanych parametrów w formie wzorów można znaleźć np. w: [Jóźwiak, Podgórski, 2012, s. 367–370]). Współczynnikiem, który określa dopasowanie funkcji do danych empirycznych i wyjaśnia procent kształtowania się zmienności Y, jest współczynnik determinacji (R^2). Jego wartość mieści się w przedziale [0;1]. Im wyższa wartość współczynnika, tym dopasowanie modelu wyższe. Interpretacje podaje się w procentach. O odchyleniu (różnicy) pomiędzy przeciętnymi empirycznymi wartościami zmiennej objaśnianej a wartościami teoretycznymi tej zmiennej (wyznaczonymi z modelu) mówi odchylenie standardowe reszt (standardowy błąd estymacji) Se. Błąd wyrażony jest w jednostkach zmiennej zależnej i z tego względu o jakości estymacji bardziej obrazowo informuje tzw. współczynnik zmienności resztowej (losowej) V. Współczynnik wyrażony jest w procentach i stanowi udział błędu standardowego estymacji w średniej wartości zmiennej zależnej (zakłada się, że wartość tej miary nie powinna przekraczać 10–15%). Ostatnią miarą świadczącą o jakości modelu jest współczynnik zbieżności φ^2. Określa, jaka część zmienności zmiennej zależnej nie została wyjaśniona przez model (jest to wpływ czynników losowych, nie uwzględnionych). Miara wyrażona jest w procentach. Im wartość φ^2 jest niższa, tym wyższa jest wartość współczynnika determinacji i tym lepiej dany model opisuje kształtowanie się badanego zjawiska.

Ocena jakości funkcji regresji

Na podstawie wyników estymacji funkcji regresji dokonano oceny jakości oszacowanego modelu na podstawie wartości czterech parametrów stochastycznych: $R^2 = 0{,}69$, $Se = 52{,}4$, $V = 13\%$, $\varphi^2 = 0{,}31$. Na podstawie powyższych wyników można powiedzieć, że model w 69% wyjaśnia kształtowanie się ilości zebranych zmieszanych odpadów, a 31% jest to wpływ czynników losowych. Zatem wynik nie jest zadowalający, ale również nie dyskwalifikuje modelu, tym bardziej że udział błędu estymacji w średniej wartości zmiennej zależnej wynosi dopuszczalne 13%, gdyż wartości empiryczne zmiennej objaśnianej odchylają się od wartości teoretycznych wyznaczonych z modelu o 52,4 kg na mieszkańca.

Źródło: opracowanie własne.

Własności reszt modelu

Badanie własności reszt modelu prowadzi się w celu weryfikacji założeń KMNK [Borkowski i in., 2004, s. 77]. Analiza własności odchyleń losowych obejmuje m.in. badanie losowości (adekwatności postaci funkcyjnej modelu), normalności (związane z analizą istotności ocen parametrów), autokorelacji czasowej (braku powiązań składników losowych w czasie) i homoskedastyczności (jednorodności). Większość z pakietów statystycznych podaje wartości statystyk oraz wartości poziomu pseudoistotności (p) odpowiednich testów służących weryfikacji własności reszt. Na przy-

kład normalność rozkładu odchyleń losowych weryfikuje się następującym zestawem hipotez: H_0 – składnik losowy modelu ma rozkład normalny, H_1 – brak normalności rozkładu składnika losowego. Na podstawie wartości poziomu p oraz statystyk takich testów, jak: Jarque'a–Bera, Shapiro–Wilka czy Kołomorgowa–Lilieforsa dokonuje się trafnego wyboru odpowiedniej hipotezy. Model ekonometryczny powinien cechować się normalnością rozkładu reszt, homoskedastycznością oraz w przypadku modeli trendu (bądź ze zmienną czasową, zob. *EkoMiasto#Środowisko*, rozdział *Ocena i prognozowanie zmian środowiskowych*) brakiem autokorelacji czasowej.

Ocena rozkładu reszt modelu

Oszacowania funkcji regresji dotyczącej wielkości dochodów w zł na mieszkańca (x_1) i ilości zebranych odpadów komunalnych w kilogramach na mieszkańca (y) wskazują na poprawność merytoryczną uzyskanych wyników, a jakość modelu nie budzi większych wątpliwości. Potwierdzają to również testy dotyczące własności reszt modelu. Rozkład reszt modelu jest normalny bądź zbliżony do normalnego, odchylenia losowe są homoskedastyczne i nie cechują się autokorelacją czasową. Możemy powiedzieć, że uzyskane wyniki są wiarygodne, a narzędzie badawcze efektywne. Zatem model może być wykorzystany np. do konstrukcji prognozy.

Źródło: opracowanie własne.

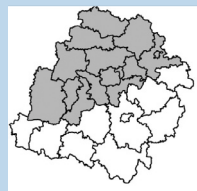
8.6. Zależności przestrzenne w mieście i między miastami

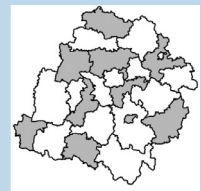
Zależności przestrzenne

Regiony, gminy, miasta nie są izolowane w przestrzeni i mogą podlegać wpływom innych jednostek w zależności od odległości między nimi [Tobler, 1970]. Wynika z tego, że rozwój miasta związany jest również z jego położeniem przestrzennym. Oznacza to, że na poziom rozwoju danego obszaru miejskiego może mieć wpływ poziom zjawiska w jednostkach z jego otoczenia (wpływy sąsiedztwa, rola odległości, zależności przestrzenne). Co więcej, może się zdarzyć, że obszary sąsiadujące są podobne pod względem badanej zmiennej (autokorelacja dodatnia, skupianie się podobnych wartości w przestrzeni, występowanie klastrów). Jest to proces warunkowany dyfuzją [Kopczewska, 2006, s. 15]. W przypadku zróżnicowania sąsiednich obszarów pod względem badanej cechy, bardziej niż wynikałoby to z rozkładu losowego, mówi się o występowaniu autokorelacji ujemnej. W przestrzeni powstaje tzw. efekt szachownicy (rys. 8.3). Z kolei gdy wartości w danym obszarze nie zależą od wartości zaobserwowanych w jednostkach sąsiednich, wówczas mamy do czynienia z brakiem autokorelacji przestrzennej (wskazuje ona na przestrzenną losowość).

Rysunek 8.3. Typy autokorelacji przestrzennej

Autokorelacja dodatnia





Autokorelacja ujemna

Źródło: opracowanie własne.

Ujęcie zależności przestrzennych

Narzędzia wykorzystywane w analizach przestrzennych służą testowaniu zależności przestrzennych, tj. wykryciu procesów autokorelacji przestrzennej zachodzących między jednostkami danej przestrzeni, ocenie jej zasięgu, siły i struktury.

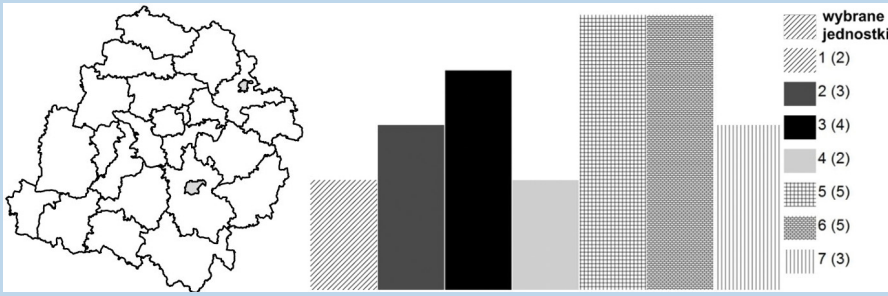
Kluczowym elementem analiz przestrzennych jest macierz wag przestrzennych (sposób konstrukcji macierzy zob. *EkoMiasto#Środowisko*, rozdział *Ocena i prognozowanie zmian środowiskowych*) – począwszy od macierzy sąsiedztwa, która obrazuje przestrzenne powiązania (bliskość obserwacji, sąsiedztwo, liczbę wspólnych granic), przez macierz dystansu uwzględniającą jednostki pozostające w pewnej odległości od siebie (geograficznej, ekonomicznej, kosztowej, społecznej), po macierz reprezentującą siłę potencjalnych zależności między ośrodkami (macierz kierunkowa, asymetryczna, wybranych najbliższych sąsiadów). Macierz wag przestrzennych najczęściej oznaczana jest symbolem **W**. Ze względu na ogromną liczbę potencjalnych powiązań odgórnie przyjmuje się jedną z możliwych macierzy **W**. Jednym ze sposobów wizualizacji powiązań występujących w zbudowanej macierzy jest histogram powiązań.

Macierz sąsiedztwa pierwszego rzędu

Wykres histogramu przedstawia liczbę powiązań między powiatami województwa łódzkiego występujących w macierzy **W**: sąsiedztwa pierwszego rzędu.

Powiaty posiadające 1 sąsiada

Liczba sąsiadów pierwszego rzędu i przypisana im liczba powiatów



Histogram odczytujemy w ten sposób, że po prawej stronie w legendzie znajdują się opisy liczby sąsiadów pierwszego rzędu, a w nawiasach przyporządkowana im liczba powiatów. Na przykład w województwie łódzkim są dwa powiaty, które mają po jednym powiecie sąsiednim rzędu pierwszego (zaznaczone na mapie). Z kolei, z wykresu można jeszcze wnioskować, że najwięcej powiatów (pięć) posiada po pięciu i sześciu sąsiadów pierwszego rzędu.

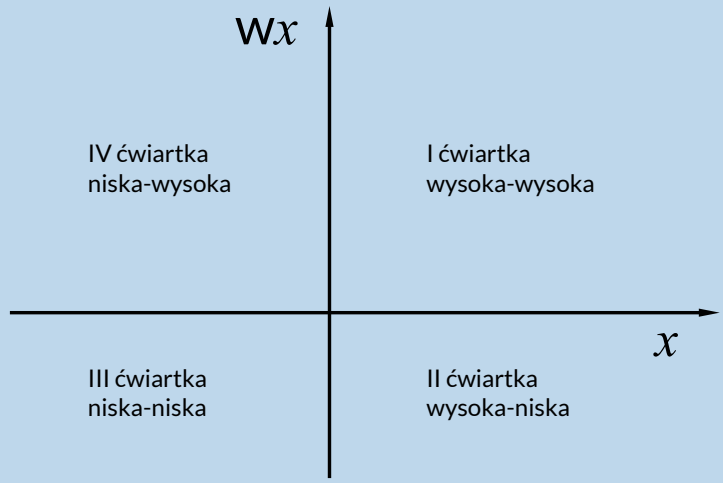
Źródło: opracowanie własne w OpenGeoDa.

Pomiar autokorelacji

Testowanie siły, kierunku i zasięgu autokorelacji przestrzennej prowadzi się za pomocą globalnych i lokalnych statystyk Morana [Moran, 1950] (opis statystyk zob. *EkoMiasto#Środowisko*, rozdział *Ocena i prognozowanie zmian środowiskowych*). Miary globalne (w postaci jednej liczby) są wskaźnikiem globalnej autokorelacji prze-

strzennej lub podobieństwa obiektów (uśrednionej, dla całego zbioru obiektów). Z kolei statystyki lokalne wyznacza się dla każdego regionu osobno. W ten sposób uzyskamy odpowiedź na pytanie o podobieństwo lub odmienność danego regionu względem jego otoczenia [Kopczewska, 2006, s. 69]. Ciekawym sposobem weryfikacji występowania zależności przestrzennych (poza wartościami obliczanymi statystyk Morana) jest tzw. moranowski wykres rozproszenia [Anselin, 1988] (rys. 8.4).

Rysunek 8.4. Schemat moranowskiego wykresu rozproszenia



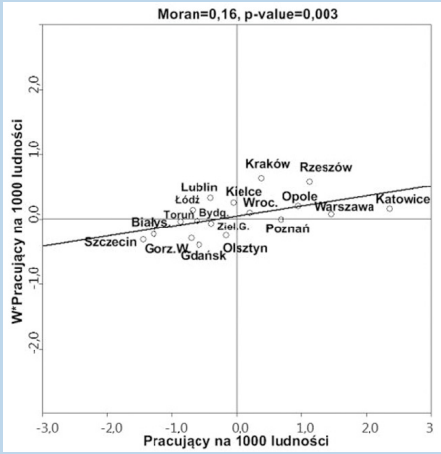
Źródło: opracowanie własne na podstawie Suchecki, 2010, s. 127.

Linia biegnąca przez I i III ćwiartkę na moranowskim wykresie rozproszenia wskazuje na występowanie autokorelacji przestrzennej dodatniej. Lokalnie oznacza to podobieństwo badanych regionów w określonym sąsiedztwie lub odległości. Jednostki są rozłożone w przestrzeni tak, że tworzą tzw. klastry wartości podobnych: niskie obok niskich, wysokie obok wysokich (wskazują na to również obliczone wartości statystyk lokalnych, LISA). I tak wzrost wartości w danym obiekcie może powodować wzrost poziomu cechy w regionie będącym jego sąsiadem, uznanym za sąsiada w macierzy **W** (wysoka-wysoka) lub średni spadek (niska-niska). Z kolei obiekty układające się na linii przebiegającej przez II i IV ćwiartkę świadczą o zróżnicowaniu badanych obiektów (autokorelacji ujemnej). Lokalnie oznacza to występowanie obiektów o niskich wartościach zmiennej obok jednostek o wysokim poziomie cechy obok siebie, czyli zdecydowanie innych wartości. Zatem wzrost wartości zmiennej powoduje średni wzrost bądź spadek poziomu zjawiska w regionie uznanym za sąsiada (zgodnie z macierzą **W**).

W przestrzeni występują również tzw. obserwacje odstające, które mogą przyjmować znacząco różne wartości badanej cechy względem otaczających obiektów (nie prowadzi się ich interpretacji, ale mają one strategiczne i wymowne znaczenie w badaniu). Wartości statystyk globalnych i lokalnych nabierają mocy interpretacyjnej (można je interpretować jako stopień podobieństwa/zróżnicowania) wówczas, gdy wykazują istotność statystyczną. W praktyce oznacza to przyjęcie hipotezy o istotności statystyki (H_0) na podstawie pseudopoziomu istotności. Wymaga to przeprowadzenia testu randomizacji [Kopczewska, 2006, s. 90].

Zebrano dane dotyczące liczby pracujących na 1000 ludności w 2013 roku cechujące ośrodki miejskie będące siedzibami województw. Wykorzystując macierz wag przestrzennych zbudowaną według ustalonego promienia odległości (do 340 km), wyznaczono wartość globalnej statystyki Morana I i oceniono jej istotność. Wstępne wyniki zaprezentowano na wykresie rozproszenia. Tym samym uzyskano odpowiedź na pytanie: Czy liczba pracujących w miastach w 2013 roku jest determinowana zależnościami przestrzennymi (sąsiedztwem innych miast)? Inaczej mówiąc, czy na poziom zjawiska w danym mieście ma wpływ wielkość emisji tego związku w miastach oddalonych o odległości do 340 km. Jaki jest kierunek tych zależności?

Moranowski wykres rozproszenia dla liczby pracujących na tys. ludności w 2013 roku w 18 miastach

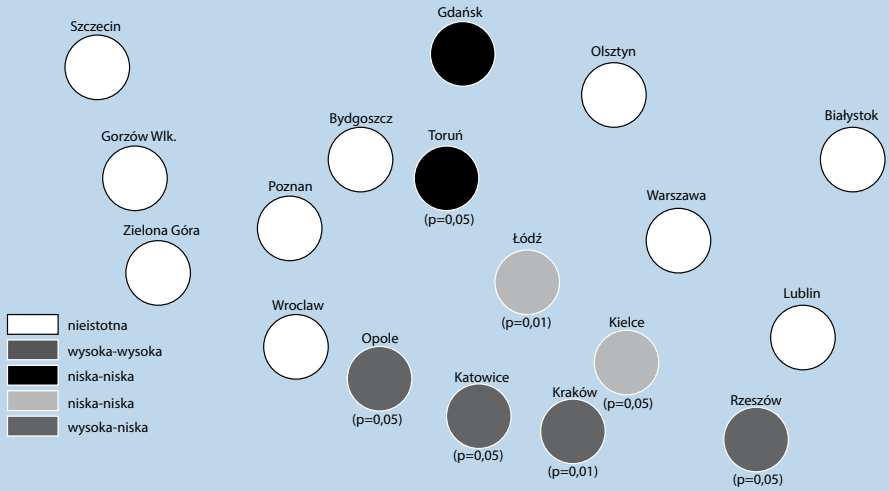


Wartość statystyki Morana I odzwierciedlająca zależności przestrzenne między miastami pod względem cechy wynosi $I = 0{,}16$, dla poziomu pseudoistotności $p = 0{,}003$. Zatem na przyjętym poziomie istotności $\alpha = 0{,}05$ wyznaczona statystyka globalnych zależności przestrzennych jest dodatnia i istotna statystycznie: $p < \alpha$. Odrzucamy hipotezę zerową na korzyść alternatywnej o występowaniu autokorelacji przestrzennej. Znak statystyki wskazuje na grupowanie się miast o podobnych wielkościach pracujących obok siebie (wysokich obok wysokich, niskich obok niskich). Z moranowskiego wykresu rozproszenia można również odczytać konkretne miasta, które cechują się wysokim poziomem zatrudnienia i w obrębie 340 km otoczone są miastami z równie wysoką liczbą pracujących (ćwiartka I). Są to np.: Warszawa, Katowice, Kraków, Rzeszów, Wrocław. Z kolei miasta znajdujące się w ćwiartce IV wykresu są to ośrodki, w których liczba pracujących jest na niskim poziomie i są otoczone miastami o równie niskim poziomie (według macierzy **W**). Należą do nich np.: Szczecin, Gorzów Wielkopolski, Gdańsk, Zielona Góra, Białystok. Na przykład Lublin, Łódź, Kielce należą do obserwacji odstających. Stanowią grupę miast o niskim poziome zmiennej i są otoczone ośrodkami o wysokim poziomie cechy. Nie wszystkie z wyżej wymienionych miast odgrywają istotną rolę w kształtowaniu się zależności przestrzennych.

Źródło: opracowanie własne w OpenGeoDa.

Jak już wspomniano, dokładniejszy obraz tych interakcji lokalnych oraz ich istotność można uzyskać, wyznaczając lokalne indeksy autokorelacji przestrzennej (LISA).

Lokalne zależności poszczególnych miast determinowane liczbą pracujących na tys. ludności w 2013 roku



Na rysunku można dostrzec klaster niskich wartości cechy (*cold spots*). Tworzą go Gdańsk i Toruń. Zatem wzrost liczby pracujących na 1000 ludności w tych miastach w 2013 roku powoduje spadek bądź brak zmian w poziomie cechy w miastach będących ich sąsiadami (zdefiniowanych w macierzy **W**). Natomiast miastami, w których wzrost poziomu zjawiska przyczynia się do wzrostu liczby pracujących w miastach będących sąsiadami (w odległości do 340 km od środka geograficznego badanej jednostki), są: Kraków, Rzeszów, Opole, Katowice (*hot spots*). Można również dostrzec dwie obserwacje odstające: Łódź i Kielce, gdzie poziom cechy jest niski, ale otaczają go miasta (w obrębie do 340 km) o wyższej liczbie pracujących. Mapa obrazuje również poziom istotności statystycznej indeksów LISA, czyli siłę interakcji przestrzennych zachodzących między miastami (wartości poziomu pseudoistotności podano w nawiasach). Najwyższa siła zależności (poziom $p = 0{,}01$) cechuje Łódź i Kraków, a niższa siła zależności pozostałe wyróżnione miasta ($p = 0{,}05$).

Źródło: opracowanie własne w programie OpenGeoDa.

8.7. Prognozowanie rozwoju miast

Prognozowanie jest procesem polegającym na racjonalnym, naukowym przewidywaniu przyszłości (przyszłych wartości zjawisk cechujących regiony, rozwój miast, jednostek, obiektów). Z pojęciem prognozowania wiąże się pojęcie prognozy. Jest to empirycznie weryfikowalny i akceptowalny osąd, formułowany w oparciu o dorobek nauk ekonomicznych, społecznych, przyrodniczych z użyciem adekwatnych metod [Cieślak, 2004, s. 16]. Jest wiele narzędzi prognostycznych również służących przewidywaniu rozwoju jednostek (w tym miast), począwszy od prognoz na podstawie szeregów czasowych, poprzez prognozowanie analogowe, heurystyczne [Radzikowska, 2004], po przewidywanie przyszłości na podstawie modeli przyczy-

nowo-skutkowych [Cieślak, 2004, s. 104–130]. Ze względu na złożoność procesów determinujących rozwój obszarów miejskich szczegółowo omówione zostaną pojęcie i etapy prognozy z wykorzystaniem wspomnianego modelu ekonometrycznego w podrozdziale *Związki przyczynowo-skutkowe*. Model ekonometryczny jako narzędzie przedstawiające relacje między badanym zjawiskiem a jego determinantami musi spełniać założenia, by służył konstrukcji prognozy. Zatem:

- model powinien być zbudowany na podstawie teorii ekonomii;
- w modelu występuje istotność statystyczna zmiennej/zmiennych objaśniającej/objaśniających;
- model jest modelem dobrym jakościowo (co zostaje określone poprzez parametry stochastyczne);
- składnik losowy ma rozkład normalny i jest stabilny w przekroju (homoskedastyczność) oraz czasie (brak autokorelacji);
- znane są przyszłe wartości zmiennej/zmiennych objaśniającej/objaśniających (przewidywane np. w oparciu o ekstrapolację trendu, więcej w: *EkoMiasto#Środowisko*, rozdział *Ocena i prognozowanie zmian środowiskowych*);
- wartości zmiennej objaśnianej można ekstrapolować w przyszłość.

Model spełniający powyższe założenia można wykorzystać w procesie prognozowania krótkookresowego (do sześciu okresów w przód). Prognozę wyznacza się wówczas na podstawie tzw. prognozy nieobciążonej, czyli podstawiając do modelu prognozowane wartości zmiennej/zmiennych objaśnianej/objaśnianych. Ocena dopuszczalności, akceptowalności, wiarygodności prognozy polega na weryfikacji wielkości błędów prognoz (*ex-post* i *ex-ante*). Do błędów *ex-post*, które określają trefność prognoz wygasłych (zrealizowanych), można zaliczyć powszechnie stosowany średni względny błąd prognozy (MAPE):

$$MAPE = \frac{1}{T-n}\sum_{t=n+1}^{T}\left|\frac{y_t - y_t^p}{y_t}\right| \qquad (8.13)$$

gdzie:

T – okres, na który konstruowana jest prognoza,

n – liczba obserwacji szeregu czasowego użyta do wyznaczenia prognozy,

y_t^p – rzeczywista wartość zmiennej prognozowanej,

y_t – wyprognozowana wartość zmiennej na czas $t > n$ [Cieślak, 2004, s. 49–50].

Błędy *ex-ante* omówiono w *EkoMiasto#Środowisko*, rozdział *Ocena i prognozowanie zmian środowiskowych*. Dopuszczalną wartość błędu określa się arbitralnie (może być to przedział od 5 do 10%).

Ekstrapolacja trendu i ocena wyników prognozowania

Na podstawie modelu opisującego kształtowanie się ilości zebranych w ciągu roku odpadów komunalnych w zależności od poziomu dochodu na mieszkańca w Warszawie w latach 2005–2013 ($\hat{y} = 872 - 0{,}08 \cdot x_i$) wyznaczono wielkości prognoz punktowych zmiennej zależnej na trzy kolejne lata: 2014–2016. Wiedząc, że wartości zmiennej objaśniającej ustalono na drodze ekstrapolacji trendu, a wielkości parametrów stochastycznych oraz własności reszt modelu $\hat{y} = 872 - 0{,}08 \cdot x_i$ upoważniają do wyznaczenia prognoz, do równania podstawiamy kolejne wartości zmiennej niezależnej. Ustalamy dopuszczalność prognoz na poziomie 10% wartości błędów *ex-post* (wzór 8.13) i *ex-ante*. W tabeli zawarto prognozy zmiennej niezależnej, wartości prognoz zmiennej zależnej na deklarowane lata oraz wielkości błędów.

Wyniki prognozowania na podstawie modelu ekonometrycznego

Rok prognozy	Zmienna x_i	Warszawa: $\hat{y} = 872 - 0{,}08 \cdot x_i$	V_T	V_T (w %)
2014	7650	$y^*_{2014} = 872 - 0{,}08 \cdot 7650 = 293$	22	7
2015	7974	$y^*_{2015} = 872 - 0{,}08 \cdot 7974 = 269$	24	9
2016	8297	$y^*_{2016} = 872 - 0{,}08 \cdot 8297 = 244$	25	10
MAPE	8,9%			

Na podstawie wartości prognoz można stwierdzić, że ilość zbieranych w roku odpadów komunalnych w Warszawie w 2016 roku wyniesie 244 kg na mieszkańca. Wielkość prognozowanej ilości odpadów wskazuje na spadek zmiennej w 2016 roku w odniesieniu do roku 2005 o 50%. Na podstawie obliczonych względnych błędów prognoz można stwierdzić, że otrzymana prognoza może być uznana za dopuszczalną. Błędy *ex-ante* prognozowanych wielkości zjawiska wahają się od 7 do 10%. Przykładowa interpretacja błędu bezwzględnego dla prognozowanego zjawiska na 2016 rok jest następująca: prognozując ilość zbieranych w ciągu roku odpadów komunalnych w Warszawie w 2016 roku, można się pomylić średnio o ±25 kg na mieszkańca. Wielkość błędu prognoz wygasłych MAPE kształtująca się na poziomie 8,8% potwierdza dopuszczalność i wiarygodność wyników. A zatem różnica między wartościami prognoz a rzeczywistym poziomem zmiennej zależnej wynosi średniorocznie ok. 9%.

Źródło: opracowanie własne.

Bibliografia

Anselin L. (1988), *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, Dordrecht.

Antczak E. (2013), *Przestrzenny taksonomiczny miernik rozwoju*, „Wiadomości Statystyczne", nr 7.

Borkowski B., Dudek H., Szczęsny W. (2003), *Ekonometria. Wybrane zagadnienia*, PWN, Warszawa.

Cieślak M. (red.) (2004), *Prognozowanie gospodarcze. Metody i zastosowania*, Wydawnictwo Naukowe PWN, Warszawa.

Hellwig Z. (1968), *Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę wykwalifikowanych kadr*, „Przegląd Statystyczny", nr 4.
Jóźwiak J., Podgórski J. (2012), *Statystyka od podstaw*, wyd. 7, PWE, Warszawa.
Kopczewska K. (2006), *Ekonometria i statystka przestrzenna z wykorzystaniem programu R CRAN*, CeDeWu.PL, Warszawa.
Kufel T. (2007), *Ekonometria. Rozwiązywanie problemów z wykorzystaniem programu GRETL*, PWN, Warszawa.
Kukuła K. (2003), *Elementy statystyki w zadaniach*, „Wydawnictwo Naukowe PWN, Warszawa.
Młodak A. (2006), *Analiza taksonomiczna w statystyce regionalnej*, Difin, Warszawa.
Moran P.A.P. (1950), *Notes on Continuous Stochastic Phenomena*, „Biometrika", nr 37(1).
Radzikowska B. (2004), *Metody prognozowania, zbiór zadań*, Wydawnictwo Akademii Ekonomicznej im. Oskara Lengego we Wrocławiu, Wrocław.
Sobczyk M., (2007), *Statystyka*, PWN, Warszawa.Sobczyk M. (2010), Statystyka opisowa, wyd. 1, C.H. Beck, Warszawa.
Stanisz A. (2006), *Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny*, t. 1–3, wyd. 1, StatSoft, Kraków.
Suchecka J. (2004), *Metody statystyczne: zarys teorii i zadania*, Wydział Zarządzania Politechniki Częstochowskiej, wyd. 2, Częstochowa.
Suchecki B. (2010), *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*, C.H. Beck, Warszawa.
Szanduła J. (2014), *Uwagi do unitaryzacji zmiennych w referencyjnym systemie granicznym*, „Przegląd Statystyczny", R. LXI, z. 2.
Tobler W. (1970), *A computer model simulating urban growth in the Detroit region*, „Economic Geography", nr 46(2).
Zeliaś A. (2000), *Metody statystyczne*, PWE, Warszawa.

WYDAWNICTWO
UNIWERSYTETU
ŁÓDZKIEGO

www.wydawnictwo.uni.lodz.pl
e-mail: ksiegarnia@uni.lodz.pl
tel. (42) 665 58 63, faks (42) 665 58 62

ISBN 978-83-7969-968-1

9 788379 699681

Projekt „EkoMiasto. Kształcenie na rzecz zrównoważonego i partycypacyjnego rozwoju miasta" finansowany ze środków funduszy norweskich oraz środków krajowych.

KATEDRA GOSPODARKI
REGIONALNEJ I ŚRODOWISKA

FSS

norway grants