

Iwona Markowicz, Beata Stolorz

Uniwersytet Szczeciński

KLASYFIKACJA BEZROBOTNYCH ZE WZGLĘDU NA WARTOŚCI ILORAZU SZANS PODJĘCIA ZATRUDNIENIA PRZY ZASTOSOWANIU MODELU LOGITOWEGO

Streszczenie: Celem artykułu jest klasyfikacja bezrobotnych do grup o podobnym ilorazie szans podjęcia zatrudnienia, której podstawą był model logitowy. Autorki przeprowadziły grupowanie według cech bezrobotnych, odmienne od stosowanego przez urzędy pracy. Proponowane w opracowaniach statystycznych grupy według wieku czy wykształcenia mogą być niejednorodne ze względu na przyczyny wyrejestrowania z urzędu. Zbadano bezrobotnych wyrejestrowanych z Powiatowego Urzędu Pracy w Szczecinie w 2008 r. ($n = 19\ 768$). W artykule zastosowano rzadko stosowane kodowanie zmiennej objaśniającej $-1; 1$, które umożliwia porównanie szans analizowanych podgrup z szansą średnią dla grupy.

1. Wstęp

Celem artykułu jest klasyfikacja bezrobotnych do grup o podobnym ilorazie szans podjęcia zatrudnienia. Autorki podjęły próbę przeprowadzenia grupowania według cech bezrobotnych, odmiennego od stosowanego przez urzędy pracy. Proponowane w opracowaniach statystycznych grupy według wieku czy wykształcenia mogą być niejednorodne ze względu na przyczyny wyrejestrowania z urzędu.

Proponowane podejście do klasyfikacji oparte jest na szansie przejścia ze stanu bezrobocia do zatrudnienia. Punktem wyjścia jest logistyczny model regresji. W artykule wykorzystany zostanie model z dychotomiczną zmienną objaśnianą (wyrejestrowanie bezrobotnego z powodu podjęcia pracy lub innego powodu) oraz wieloma zmiennymi objaśniającymi (cechy bezrobotnych mające wpływ na możliwość uzyskania zatrudnienia: płeć, wiek, wykształcenie¹). Dzięki transformacji logitowej [$\text{logit}(p)$] dokonuje się przekształcenia prawdopodobieństwa p z przedziału $[0, 1]$ na przedział $(-\infty; +\infty)$. W celu porównania dwóch grup o różnych wariantach analizowanej zmiennej objaśniającej wyznaczany jest iloraz szans, który wskazuje na różnice między grupami. Ilorazy szans mogą być zatem wykorzystane

¹ Autorki w pracy [Markowicz, Stolorz 2008] wskazały, że cechy te są determinantami wychodzenia z bezrobocia.

do oceny podobieństwa grup bezrobotnych o różnych wariantach cechy determinującej możliwość znalezienia pracy.

Omówione podejście klasyfikacyjne zostanie zaprezentowane na przykładzie danych dotyczących bezrobotnych wyrejestrowanych z Powiatowego Urzędu Pracy w Szczecinie w 2008 r. ($n = 19768$).

W artykule, poza nowym podejściem do grupowania, zastosowane zostanie rzadko stosowane kodowanie zmiennej objaśniającej $-1; 1$, które umożliwia porównanie szans analizowanych podgrup z szansą średnią dla grupy.

2. Dane statystyczne wykorzystane w badaniu

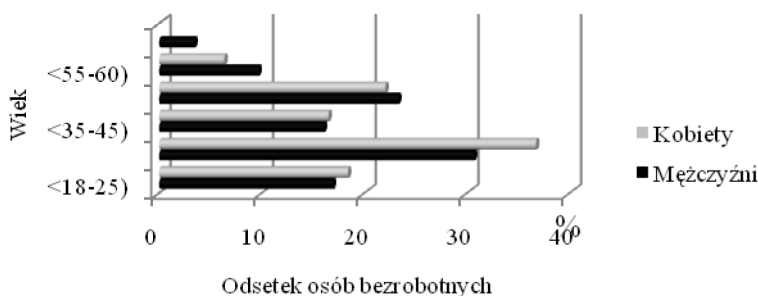
W badaniu wykorzystano indywidualne dane osób bezrobotnych wyrejestrowanych w Szczecinie w 2008 r. Dzięki współpracy z Powiatowym Urzędem Pracy w Szczecinie pozyskano informacje o 19 768 osobach dotyczące dat zarejestrowania i wyrejestrowania z urzędu, płci, wieku i wykształcenia. Przyczyny wyrejestrowania były różne, ale podzielono je na dwie grupy – podjęcie pracy (zdarzenie sprzyjające) lub niepodjęcie pracy. Strukturę badanych osób według tych grup i płci zawarto w tab. 1.

Zarówno wśród kobiet, jak i wśród mężczyzn przeważali bezrobotni w wieku od 25 do 35 lat oraz dużą grupę stanowiły osoby między 45 a 55 rokiem życia (rys. 1).

Tabela 1. Struktura osób bezrobotnych wyrejestrowanych z PUP w Szczecinie w 2008 r. według powodu wyrejestrowania i płci

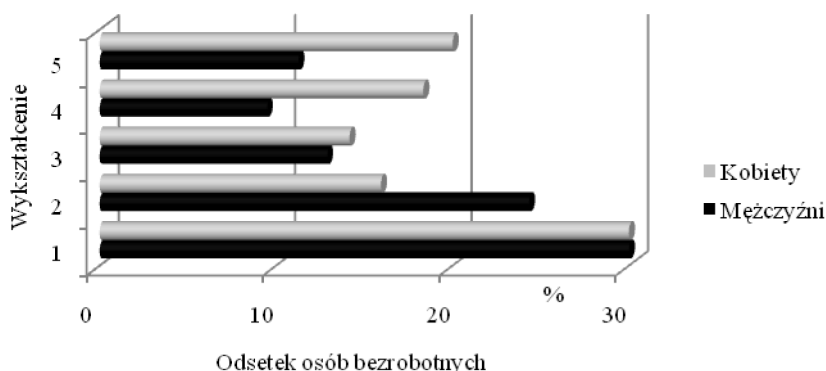
Powód wyrejestrowania	Płeć		Razem
	kobiety	mężczyźni	
Podjęcie pracy	3 244	2282	5 526
Niepodjęcie pracy	6 991	7251	14 242
Razem	10 235	9533	19 768

Źródło: opracowanie własne.



Rys. 1. Struktura kobiet i mężczyzn wyrejestrowanych z PUP w Szczecinie w 2008 r. według grup wieku (w %)

Źródło: opracowanie własne.



Rys. 2. Struktura kobiet i mężczyzn wyrejestrowanych z PUP w Szczecinie w 2008 r. według grup wykształcenia (w %)

Źródło: opracowanie własne.

Wśród badanych przeważają bezrobotni z wykształceniem co najwyżej gimnazjalnym. Druga najlicniejsza grupa wśród mężczyzn ma wykształcenie zasadnicze zawodowe, a wśród kobiet wyższe (rys. 2). Grupy wykształcenia przyjęto za PUP: 1 – brak lub niepełne podstawowe, podstawowe, gimnazjalne, 2 – zasadnicze zawodowe, 3 – średnie ogólnokształcące, 4 – średnie zawodowe 4-letnie, średnie zawodowe, pomaturalne/policealne, 5 – wyższe (w tym licencjat).

3. Model logitowy jako podstawa klasyfikacji

Zaproponowany w pracy sposób klasyfikacji oparty jest na modelu logitowym²:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \alpha_0 + \sum_{i=1}^k \alpha_i x_i, \quad (1)$$

gdzie p określa prawdopodobieństwo sukcesu (lub porażki), wyrażenie $\frac{p}{1-p}$

oznacza szansę (lub ryzyko) wystąpienia określonego zdarzenia.

Im większa wartość współczynnika regresji α_i , tym większy wpływ zmiennej objaśniającej na prawdopodobieństwo sukcesu (lub porażki). Jednak do interpretacji wykorzystuje się najczęściej iloraz szans (lub ryzyka), czyli e^{α_i} .

Iloraz szans (OR – *Odds Ratio*) przy zmianie jednej ze zmiennych objaśniających x_i o jednostkę (pozostałe zmienne objaśniające pozostają bez zmian) określony jest wzorem:

² Szerzej o transformacji logitowej i jej własnościach w pracach [Gatnar, Walesiak 2004, s. 158-160; Ostasiewicz 1999, s. 407-410].

$$\begin{aligned}
 OR &= \frac{P(x_{j+1})}{1 - P(x_{j+1})} \div \frac{P(x_j)}{1 - P(x_j)} = \\
 &= \frac{e^{\alpha_0} \cdot e^{\alpha_i x_{j+1}}}{e^{\alpha_0} \cdot e^{\alpha_i x_j}} = e^{\alpha_i(x_{j+1} - x_j)} = e^{\alpha_i \cdot 1} = e^{\alpha_i}.
 \end{aligned}
 \tag{2}$$

Model logitowy, w którym przyjmuje się rozkład logistyczny, jest jednym z modeli wykorzystywanych do analizy zero-jedynkowej zmiennej objaśnianej czy tzw. modeli prawdopodobieństwa. Model ten jest pozbawiony wad klasycznego modelu regresji liniowej, szacowanego klasyczną metodą najmniejszych kwadratów. Wady te to możliwość przyjęcia wartości prawdopodobieństwa spoza przedziału [0; 1] oraz heteroskedastyczność składnika losowego [Jajuga 1990]. Autor ten podkreśla też wadę modeli prawdopodobieństwa, która jest istotna z punktu widzenia praktycznych zastosowań. Do uzyskania zarówno efektywnych ocen parametrów modelu, jak i prawdopodobieństw teoretycznych niezbędna jest duża liczba obserwacji dla każdego stanu wektora zmiennych objaśniających. Warunek ten w badaniu prezentowanym w niniejszym artykule został spełniony.

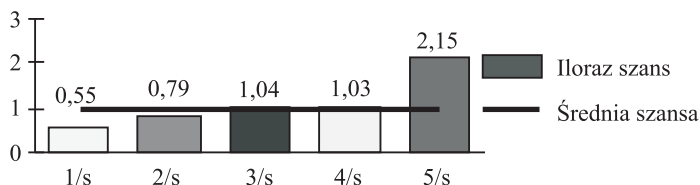
Najczęściej spotykanym w literaturze sposobem kodowania dychotomicznych zmiennych objaśniających jest kodowanie 0;1. W ten sposób możliwe jest wyznaczenie szans badanych grup w stosunku do jednej wybranej grupy. Możliwe jest również inne podejście³. Jest to kodowanie -1,1, które umożliwia obliczenie szans wszystkich grup w stosunku do średniej szansy tych grup. W pracy autorki zastosowały to drugie podejście. Umożliwiło to wyznaczenie i porównanie szans na podjęcie pracy wyrejestrowanych bezrobotnych poszczególnych grup wieku, wykształcenia i płci w stosunku do średniej szansy.

4. Klasyfikacja bezrobotnych według wykształcenia i wieku

Informacje o bezrobociu podawane w statystyce publicznej są to informacje zbiorcze przedstawiane według ujednocionej klasyfikacji. Ten sposób prezentacji wyników umożliwia porównania w czasie i przestrzeni. W przypadku badania konkretnej zbiorowości bezrobotnych i dysponując danymi indywidualnymi, można dokonać innego podziału na podgrupy. Autorki zaproponowały klasyfikację opartą na wartościach ilorazów szans. W celu wskazania różnic wewnątrzgrupowych, czyli między bezrobotnymi o danym wykształceniu zaliczonymi do jednej grupy, przedstawiono ilorazy szans obliczone dla podziału stosowanego przez PUP (rys. 3) i dla poszczególnych wariantów wykształcenia (rys. 4). Poszczególne warianty wykształcenia oznaczono następująco: 1 – brak lub niepełne podstawowe, 2 – podstawowe, 3 – gimnazjalne, 4 – zasadnicze zawodowe, 5 – średnie ogólnokształcące,

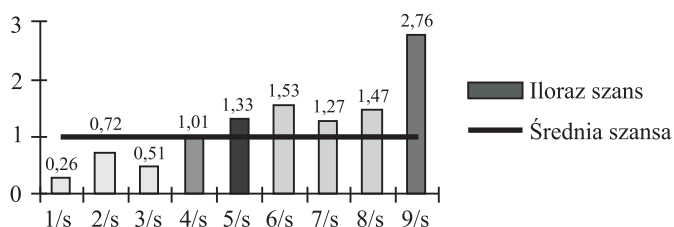
³ Obie procedury przedstawione zostały w pracy [Hosmer, Lemeshow 2000].

6 – średnie zawodowe 4-letnie, 7 – średnie zawodowe, 8 – pomaturalne/policealne, 9 – wyższe (w tym licencjat).



Rys. 3. Iloraz szans podjęcia pracy przez bezrobotnych według grup wykształcenia (podział na grupy stosowany przez PUP)

Źródło: opracowanie własne.



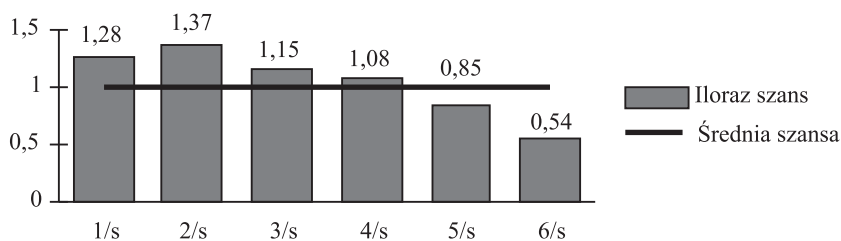
Rys. 4. Iloraz szans podjęcia pracy przez bezrobotnych według poszczególnych wariantów wykształcenia

Źródło: opracowanie własne.

Porównując oba wykresy, należy podkreślić różnice wewnątrz grup 1 i 4 (według klasyfikacji PUP). W grupie pierwszej, obejmującej wykształcenie co najwyżej gimnazjalne, największe szanse na podjęcie pracy mieli bezrobotni z wykształceniem podstawowym. Natomiast wśród osób z grupy czwartej największe szanse miały osoby z wykształceniem średnim zawodowym czteroletnim. Oczywiście w obu przypadkach bezrobotni z wyższym wykształceniem mieli największe szanse na zatrudnienie. Ze względu na różnice w ilorazach szans wyznaczonych dla poszczególnych wariantów wykształcenia zrezygnowano z tworzenia podgrup.

Na rysunku 5 przedstawiono ilorazy szans podjęcia pracy przez bezrobotnych według sześciu grup wieku na podstawie podziału stosowanego przez PUP. Oznaczono je w następujący sposób: 1-<18,25), 2-<25,35), 3-<35,45), 4-<45,55), 5-<55,60), 6-<60,65).

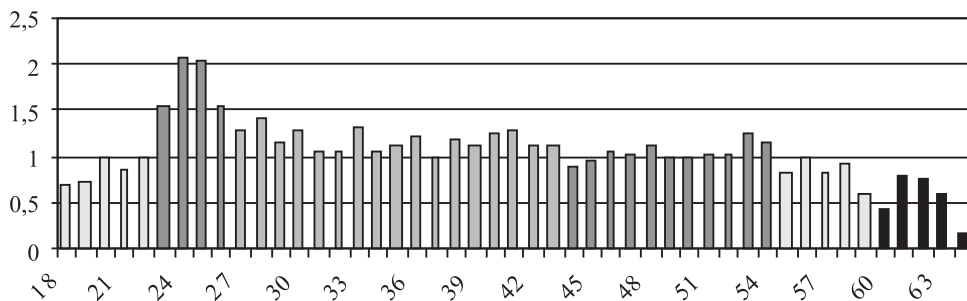
W celu zbadania różnic istniejących wewnątrz tych grup wyznaczono ilorazy szans na zatrudnienie dla bezrobotnych pogrupowanych według ukończonych pełnych lat życia (rys. 6). Największym prawdopodobieństwem wyrejestrowania z powodu podjęcia pracy charakteryzowały się osoby bezrobotne w wieku 24 i 25 lat. Było ono dwukrotnie większe niż średnia liczona dla wszystkich grup.



Rys. 5. Iloraz szans podjęcia pracy przez bezrobotnych według grup wieku (podział na grupy stosowany przez PUP)

Źródło: opracowanie własne.

W pierwszym podziale osoby te znalazły się w różnych grupach wieku. Podjęto decyzję o klasyfikacji bezrobotnych do grup o podobnych ilorazach szans⁴. Utworzono również sześć grup (na rys. 6 oznaczone są one tym samym odcieniem szarości) i ponumerowano następująco: 1-<18,23), 2-<23,27), 3-<27,44), 4-<44,55), 5-<55,60), 6 – <60,65). Największe zmiany dotyczą pierwszych trzech grup wieku. Druga z nich w obu przypadkach obejmuje osoby o największym prawdopodobieństwie zatrudnienia. W nowym podziale grupa druga obejmuje jedynie osoby w wieku: 23, 24, 25 i 26 lat. Dzięki takiemu podziałowi uwidoczniła się ich wyraźna przewaga. Trzy grupy właściwie nie uległy zmianie. Ostatnią pozostawiono bez zmian ze względu na wiek emerytalny kobiet.

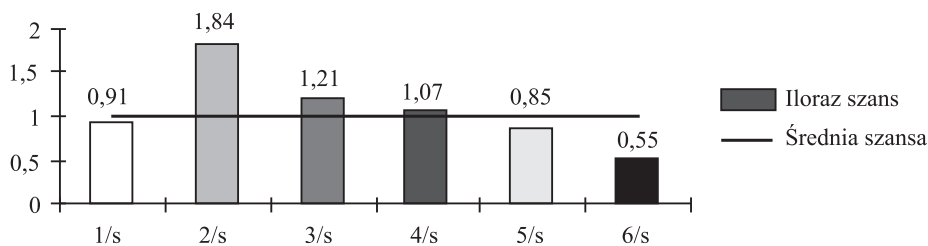


Rys. 6. Iloraz szans podjęcia pracy przez bezrobotnych według ukończonych pełnych lat życia

Źródło: opracowanie własne.

Porównując oba podziały bezrobotnych według grup wieku (rys. 5 – podział PUP, rys. 7 – proponowany podział), należy zauważyć, że drugi z podziałów umożliwia wyraźniejsze wskazanie wpływu wieku na możliwości wyjścia z bezro-

⁴ Wstępnego podziału dokonano wizualnie na podstawie rys. 6, a następnie sprawdzano, czy zmiany granic nowych przedziałów wpływają na zmniejszenie odchylenia standardowego podgrupy. Ponieważ nie wpływały, pozostano przy tym podziale.



Rys. 7. Iloraz szans podjęcia pracy przez bezrobotnych według nowych grup wieku

Źródło: opracowanie własne.

bocia. Zmiana granicy pierwszej grupy wieku prowadzi do wniosku, że osoby najmłodsze (do 23 lat) mają niższe szanse na podjęcie pracy zarówno w stosunku do średniej wszystkich grup, jak i w stosunku do osób w wieku średnim. Nie można było tego zauważyć w przypadku pierwszego podziału.

5. Podsumowanie

W artykule podjęto próbę zastosowania innej niż w statystyce publicznej klasyfikacji osób bezrobotnych na podstawie ilorazu szans wyznaczonego na podstawie modelu logitowego. Powszechnie stosowane grupowanie ze względu na wykształcenie i wiek jest bardzo przydatne ze względu na możliwość porównań w czasie i przestrzeni. Proponowana klasyfikacja może być uzupełnieniem w prowadzeniu analiz dotyczących konkretnej zbiorowości. Istnieje wówczas możliwość wyodrębnienia grup o największej lub najmniejszej szansie na wyjście z bezrobocia i podjęcie pracy. Ten rodzaj analizy może być wykorzystany w przypadku badania lokalnego rynku pracy, a wyniki mogą posłużyć do kształtowania regionalnej polityki społecznej.

Literatura

- Gatnar E., Walesiak M. (red.), *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*, AE, Wrocław 2004.
- Hosmer D.W., Lemeshow S., *Applied Logistic Regression*, John Wiley & Sons, Inc., 2000.
- Jajuga K., *Modele z dyskretną zmienną objaśnianą*, [w:] *Estymacja modeli ekonometrycznych*, red. S. Bartosiewicz, PWE, Warszawa 1990.
- Markowicz I., Stolorz B., *Analiza determinant czasu poszukiwania pracy na rynku lokalnym – na przykładzie Szczecina*, Studia Regionalne i Lokalne nr 4 (34), Wydawnictwo Naukowe Scholar, Uniwersytet Warszawski, Warszawa 2008.
- Ostasiewicz W. (red.), *Statystyczne metody analizy danych*, AE, Wrocław 1999.

THE LOGIT MODEL CLASSIFICATION OF THE UNEMPLOYED ACCORDING TO THEIR EMPLOYMENT ODDS

Summary: The aim of the article is to classify the unemployed to groups with similar employment odds ratio. The basis of the classification is the logit model. The authors formed the groups according to the unemployed person's profile that was different from the one applied by labour offices. The groups classified according to age or qualifications as proposed in statistical studies can be non-uniform as far as the reasons to sign off from the job centre are concerned. The authors studied a group of job seekers who had signed off from the Local Labour Office in Szczecin in 2008 ($n = 19\ 768$). In the article, a rarely used independent variable encoding $-1, 1$ was applied which makes it possible to compare the analyzed sub-groups with the mean chance for the whole group.