

# PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

# RESEARCH PAPERS

of Wrocław University of Economics

Nr 427

**Taksonomia 27**

**Klasyfikacja i analiza danych –  
teoria i zastosowania**



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu  
Wrocław 2016

Redaktor Wydawnictwa: Agnieszka Flasińska

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Beata Mazur

Projekt okładki: Beata Dębska

Tytuł dofinansowany ze środków Narodowego Banku Polskiego  
oraz ze środków Sekcji Klasyfikacji i Analizy Danych PTS

Informacje o naborze artykułów i zasadach recenzowania  
znajdują się na stronach internetowych  
[www.pracnaukowe.ue.wroc.pl](http://www.pracnaukowe.ue.wroc.pl)  
[www.wydawnictwo.ue.wroc.pl](http://www.wydawnictwo.ue.wroc.pl)

Publikacja udostępniona na licencji Creative Commons  
Uznanie autorstwa-Użycie niekomercyjne-Bez utworów zależnych 3.0 Polska  
(CC BY-NC-ND 3.0 PL)



© Copyright by Uniwersytet Ekonomiczny we Wrocławiu  
Wrocław 2016

**ISSN 1899-3192** (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)  
**e-ISSN 2392-0041**  
**ISSN 1505-9332** (Taksonomia)

Wersja pierwotna: publikacja drukowana

Zamówienia na opublikowane prace należy składać na adres:  
Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu  
ul. Komandorska 118/120, 53-345 Wrocław  
tel./fax 71 36 80 602; e-mail:[econbook@ue.wroc.pl](mailto:econbook@ue.wroc.pl)  
[www.ksiegarnia.ue.wroc.pl](http://www.ksiegarnia.ue.wroc.pl)

Druk i oprawa: TOTEM

## Spis treści

<b>Wstęp</b> .....	9
<b>Beata Bal-Domańska:</b> Propozycja procedury oceny zrównoważonego rozwoju w układzie <i>presja – stan – reakcja</i> w ujęciu przestrzennym / Proposal of the assessment of poviats sustainable development in the pressure – state – response system in spatial terms.....	11
<b>Tomasz Bartłomowicz:</b> Pomiar preferencji konsumentów z wykorzystaniem metody <i>Analytic Hierarchy Process</i> / Analytic Hierarchy Process as a method of measurement of consumers' preferences.....	20
<b>Maciej Beręsewicz, Marcin Szymkowiak:</b> Analiza skupień wybranych lokalnych rynków nieruchomości w Polsce z wykorzystaniem internetowych źródeł danych / Cluster analysis of selected local real estate markets in Poland based on Internet data sources.....	30
<b>Beata Bieszk-Stolorz:</b> Wybrane modele przeciętnego efektu oddziaływania w analizie procesu wychodzenia z bezrobocia / Chosen average treatment effect models in the analysis of unemployment exit process.....	40
<b>Justyna Brzezińska:</b> Modele IRT i modele Rascha w badaniach testowych / IRT and Rasch models in test measurement.....	49
<b>Mariola Chrzanowska, Nina Drejerska:</b> Geograficznie ważona regresja jako narzędzie analizy poziomu rozwoju społeczno-gospodarczego na przykładzie regionów Unii Europejskiej / Geographically weighted regression as a tool of analysis of socio-economic development level of regions in the European Union.....	58
<b>Sabina Denkowska:</b> Zastosowanie analizy wrażliwości do oceny wpływu nieobserwowanej zmiennej w <i>Propensity Score Matching</i> / The application of sensitivity analysis in assessing the impact of an unobserved confounder in Propensity Score Matching.....	66
<b>Adam Depta:</b> Zastosowanie analizy czynnikowej do wyodrębnienia aspektów zdrowia wpływających na jakość życia osób jękających się / The application of factor analysis to the identification of the health aspects affecting the quality of life of stuttering people.....	76
<b>Mariusz Doszyń, Sebastian Gnat:</b> Taksonomiczno-ekonometryczna procedura wyceny nieruchomości dla różnych miar porządkowania / Taxonomic and econometric method of real estate valuation for various classification measures.....	84

<b>Marta Dziechciarz-Duda, Anna Król:</b> Segmentacja konsumentów smartfonów na podstawie preferencji wyrażonych / Segmentation of smartphones' consumers on the basis of stated preferences .....	94
<b>Ewa Genge:</b> Zmienne towarzyszące w ukrytym modelu Markowa – analiza oszczędności polskich gospodarstw domowych / Latent Markov model with covariates – Polish households' saving behaviour .....	103
<b>Joanna Górna, Karolina Górna:</b> Modelowanie wzrostu gospodarczego z wykorzystaniem narzędzi ekonometrii przestrzennej / Economic growth modelling with the application of spatial econometrics tools .....	112
<b>Alicja Grześkowiak:</b> Wielowymiarowa analiza kompetencji zawodowych według grup wieku ludności / Multivariate analysis of professional competencies with respect to the age groups of the population .....	122
<b>Agnieszka Kozera, Feliks Wysocki:</b> Problem ustalania współrzędnych obiektów modelowych w metodach porządkowania liniowego obiektów / The problem of determining the coordinates of model objects in object linear ordering methods .....	131
<b>Mariusz Kubus:</b> Lokalna ocena mocy dyskryminacyjnej zmiennych / Local evaluation of a discrimination power of the variables.....	143
<b>Paweł Lula, Katarzyna Wójcik, Janusz Tuchowski:</b> Analiza wydzźwięku polskojęzycznych opinii konsumenckich ukierunkowanych na cechy produktu / Feature-based sentiment analysis of opinions in Polish.....	153
<b>Aleksandra Łuczak, Agnieszka Kozera, Feliks Wysocki:</b> Ocena sytuacji finansowej jednostek samorządu terytorialnego z wykorzystaniem rozmytych metod klasyfikacji i programu R / Assessment of financial condition of local government units with the use of fuzzy classification methods and program R .....	165
<b>Dorota Rozmus:</b> Badanie stabilności taksonomicznej czynnikowej metody odległości probabilistycznej / Stability of the factor probability distance clustering method .....	176
<b>Adam Sagan, Aneta Rybicka, Justyna Brzezińska:</b> <i>Conjoint analysis</i> oparta na modelach IRT w zagadnieniu optymalizacji produktów bankowych / An IRT-approach for conjoint analysis for banking products preferences.....	184
<b>Michał Stachura:</b> O szacowaniu centrum populacji określonego obszaru na przykładzie Polski / On estimating centre of population of a given territory. Poland's case .....	195
<b>Michał Stachura, Barbara Wodecka:</b> Wybrane aspekty i zastosowania modeli zdarzeń ekstremalnych / Selected facets and application of models of extremal events .....	205
<b>Iwona Staniec, Jan Żółtowski:</b> Wykorzystanie analizy log-liniowej do wyboru czynników determinujących współpracę w przedsiębiorczości	

---

technologicznej / Use of log-linear analysis for the selection determinants of cooperation in technological entrepreneurship.....	215
<b>Marcin Szymkowiak, Wojciech Roszka:</b> Potencjał gospodarczy gmin aglomeracji poznańskiej w ujęciu taksonomicznym / The economic potential of municipalities of the Poznań agglomeration in the light of taxonomy analysis.....	224
<b>Lucyna Wojcieszka:</b> Zastosowanie modeli klas ukrytych w badaniu opinii respondentów na temat roli państwa w gospodarce / Implementation of latent class models in the respondents' survey on the role of the country in economy.....	234

## **Wstęp**

W dniach 14–16 września 2015 r. w Hotelu Novotel Gdańsk Marina w Gdańsku odbyła się XXIV Konferencja Naukowa Sekcji Klasyfikacji i Analizy Danych PTS (XXIX Konferencja Taksonomiczna) „Klasyfikacja i analiza danych – teoria i zastosowania”, zorganizowana przez Sekcję Klasyfikacji i Analizy Danych Polskiego Towarzystwa Statystycznego oraz Katedrę Statystyki Wydziału Zarządzania Uniwersytetu Gdańskiego.

W trakcie dwóch sesji plenarnych oraz 13 sesji równoległych wygłoszono 58 referatów poświęconych aspektom teoretycznym i aplikacyjnym zagadnienia klasyfikacji i analizy danych. Odbyła się również sesja plakatowa, na której zaprezentowano 14 plakatów.

Teksty 24 recenzowanych artykułów naukowych stanowią zawartość prezentowanej publikacji z serii Taksonomia nr 27. Teksty 25 recenzowanych artykułów naukowych znajdują się w Taksonomii nr 26.

*Krzysztof Jajuga, Marek Walesiak*

**Sabina Denkowska**

Uniwersytet Ekonomiczny w Krakowie  
e-mail: Sabina.Denkowska@uek.krakow.pl

---

**ZASTOSOWANIE ANALIZY WRAZLIWOSCI  
DO OCENY WPLYWU NIEOBSERWOWANEJ  
ZMIENNEJ W *PROPENSITY SCORE MATCHING***

---

**THE APPLICATION OF SENSITIVITY ANALYSIS  
IN ASSESSING THE IMPACT OF AN UNOBSERVED  
CONFOUNDER IN *PROPENSITY SCORE MATCHING***

---

DOI: 10.15611/pn.2016.427.07

**Streszczenie:** Coraz częściej do przeprowadzania ewaluacji projektów i programów współfinansowanych ze środków unijnych zalecana jest metoda *Propensity Score Matching* (PSM). Metoda ta umożliwia redukcję obciążenia selekcyjnego w szacowaniu efektu oddziaływania na jednostki poddane oddziaływaniu ATT (Average Treatment Effect on Treated). W badaniach eksperymentalnych randomizacja powoduje zbalansowanie wszystkich zmiennych, również nieobserwowanych. W badaniach obserwacyjnych tak nie jest, a oszacowany efekt może być spowodowany niekoniecznie przez oddziaływanie, może być wynikiem braku zbalansowania nieuwzględnionej w badaniu zmiennej, wpływającej zarówno na zmienną wynikową, jak i na selekcję do grupy poddanej oddziaływaniu. Metoda analizy wrażliwości zaproponowana przez P.R. Rosenbauma umożliwia ocenę odporności oszacowanego efektu oddziaływania na występowanie takiej potencjalnej nieobserwowanej zmiennej. W artykule metodę granic Rosenbauma zastosowano do zbadania odporności oszacowanego za pomocą PSM efektu netto staży zorganizowanych przez jeden z małopolskich PUP w 2013 r.

**Słowa kluczowe:** *Propensity Score Matching*, analiza wrażliwości, granice Rosenbauma.

**Summary:** Propensity Score Matching (PSM) is a method recommended to evaluate projects and programmes co-financed by the European Union. It allows for the reduction of the selection bias while estimating the average treatment effect on treated (ATT). In experimental studies randomisation ensures balance of covariates between the treated and control groups, including unobserved ones. In observational studies an estimated effect can be caused not necessarily by the treatment, but may result from the lack of balance of a variable not accounted for in the study, which influences both the outcome and the selection process. Sensitivity analysis proposed by P.R. Rosenbaum allows for the assessment of the robustness of the treatment effect on the occurrence of such a potential unobserved confounder. The article presents the application of Rosenbaum's bounds method to analyse (with the use of PSM) the robustness of the net effect of internships organised in 2013 by a District Employment Office in Małopolska.

**Keywords:** Propensity Score Matching, sensitivity analysis, Rosenbaum's bounds.

## 1. Wstęp

Metoda *Propensity Score Matching* (PSM) jest jedną z metod kontrfaktycznych coraz częściej zalecanych [European Commission 2014, s. 6, 7] do przeprowadzania ewaluacji projektów i programów współfinansowanych ze środków unijnych. Metoda ta umożliwia redukcję obciążenia selekcyjnego występującego w szacowaniu przeciętnego efektu oddziaływania na jednostki poddane oddziaływaniu (ATT). W badaniach społeczno-ekonomicznych oddziaływaniem może być np. pewnego rodzaju interwencja polegająca na dawaniu środków na organizację programów, projektów, organizowaniu szkoleń dla pewnych grup społecznych, a beneficjentami mogą być osoby, gospodarstwa domowe bądź instytucje. Efekt interwencji mierzony jest za pomocą zmiennej wyjściowej, którą mogą być np. zarobki lub zatrudnienie.

Metody oparte na dopasowywaniu polegają na łączeniu jednostek z grupy poddanej interwencji z jednostkami z grupy kontrolnej, wyselekcjonowanej z puli kontrolnej jednostek niepoddanych oddziaływaniu na podstawie obserwowanych charakterystyk  $X$ . W PSM dopasowywanie odbywa się na podstawie oszacowanego w oparciu o obserwowane charakterystyki  $X$ , *propensity score*.

Szacowane, za pomocą metod opartych na dopasowywaniu jednostek, przeciętne efekty oddziaływań są nieobciążone, pod warunkiem że wszystkie zmienne wpływające zarówno na wynik zmiennej wynikowej  $Y$ , jak i na włączenie do grupy poddanej oddziaływaniu, zostały uwzględnione w badaniu. Główny zarzut krytyków pod adresem badań opartych na łączeniu jednostek, w tym również metody PSM, dotyczy nieuwzględnienia w badaniu ważnej zmiennej, która wpływa zarówno na wynik zmiennej wynikowej, jak i na selekcję do grupy poddawanej oddziaływaniu. Nie zawsze jest to zarzut krytyczny względem etapu projektowania badania i zbierania danych. Zmienna nieobserwowana  $U$  może być niemierzalna lub trudna do zmierzenia. A ponieważ przy dopasowaniu grup nie uwzględniono tej zmiennej, więc jednostki z obu grup mogą nie być porównywalne w sensie zmiennej nieobserwowanej  $U$ . Zatem oszacowany efekt może być spowodowany niekoniecznie przez oddziaływanie, może być wynikiem braku zbalansowania<sup>1</sup> nieuwzględnionej w badaniu zmiennej  $U$ , która wpływa zarówno na wynik, jak i na selekcję do grupy poddanej oddziaływaniu<sup>2</sup>. Dlatego P.R. Rosenbaum [2005] zaleca, by badania obserwacyjne oparte na dopasowywaniu jednostek standardowo uzupełniać analizą wrażliwości, która umożliwi ocenę odporności oszacowanego efektu na występowanie takiej potencjalnej zmiennej.

<sup>1</sup> Zbalansowanie zmiennych oznacza podobieństwo rozkładów rozumiane jako brak różnic systematycznych w rozkładach zmiennych.

<sup>2</sup> Problem ten nie występuje w badaniach eksperymentalnych opartych na randomizacji, która balansuje wszystkie zmienne obserwowane i zmienne nieobserwowane, tak więc różnice występujące w wartościach zmiennej wyjściowej dla grupy eksperymentalnej i kontrolnej są wynikiem oddziaływania na jednostki w grupie eksperymentalnej.



Metoda analizy wrażliwości zaproponowana przez P.R. Rosenbauma [2002] umożliwia ocenę odporności oszacowanego efektu oddziaływania na występowanie potencjalnej, nieobserwowanej zmiennej wpływającej zarówno na wynik, jak i na selekcję do grupy poddanej oddziaływaniu. W artykule metodę granic Rosenbauma zastosowano do zbadania odporności oszacowanego za pomocą PSM efektu netto staży zorganizowanych przez jeden z małopolskich PUP w 2013 r. [Denkowska 2015].

## 2. Metoda *Propensity Score Matching*

Niech  $X$  oznacza wektor obserwowanych charakterystyk,  $D$  zaś – oddziaływanie ( $D \in \{0, 1\}$ ), przy czym  $D = 1$  oznacza, że jednostka została poddana oddziaływaniu, a  $D = 0$  oznacza, że nie została poddana oddziaływaniu. Dla każdej  $i$ -tej jednostki z  $N$ -elementowej populacji możliwy jest tylko jeden z dwóch wyników zmiennej wyjściowej  $Y$ :

$$Y_i = D \cdot Y_i^1 + (1 - D) \cdot Y_i^0 = \begin{cases} Y_i^0, & \text{gdym } D = 0 \\ Y_i^1, & \text{gdym } D = 1 \end{cases} \quad (1)$$

W badaniach ewaluacyjnych zazwyczaj celem jest estymacja przeciętnego efektu oddziaływania ATE (*Average Treatment Effect*):

$$\tau_{\text{ATE}} = E[Y^1 - Y^0] \quad (2)$$

lub przeciętnego efektu oddziaływania na jednostki poddane oddziaływaniu ATT (*Average Treatment Effect on Treated*):

$$\tau_{\text{ATT}} = E[(Y^1 - Y^0) | D = 1]. \quad (3)$$

Przeciętny efekt oddziaływania na jednostki poddane oddziaływaniu można przedstawić jako różnicę:

$$\tau_{\text{ATT}} = (E[Y^1 | D = 1] - E[Y^0 | D = 0]) - (E[Y^0 | D = 1] - E[Y^0 | D = 0]), \quad (4)$$

w której odjemnik to tzw. obciążenie selekcyjne wynikające z braku zbalansowania zmiennych obserwowanych oraz zmiennych nieobserwowanych, a także z powodu niepokrywających się obszarów określoności w grupie poddanej oddziaływaniu i w puli kontrolnej.

Metoda *Propensity Score Matching* umożliwia redukcję obciążenia selekcyjnego w szacowaniu przeciętnego efektu oddziaływania na jednostki poddane interwencji ATT. Polega ona na dopasowywaniu do grupy poddanej interwencji takiej grupy kontrolnej wyselekcjonowanej z puli kontrolnej osób niepoddanych oddziaływaniu, że rozkłady obserwowanych charakterystyk wektora  $X$  w obu grupach

będą zbalansowane. W PSM do dopasowywania jednostek wykorzystuje się funkcję balansującą (zob. [Rosenbaum, Rubin 1983]) zmienne obserwowane  $X$ , *propensity score*  $p$ :

$$p(\mathbf{x}) := P(D=1|X=\mathbf{x}) = E(D|X=\mathbf{x}). \quad (5)$$

Kluczowym założeniem<sup>3</sup> metody PSM jest założenie warunkowej niezależności (*Conditional Independence Assumption*), które oznacza, że cały proces selekcji musi być oparty jedynie na obserwowanych charakterystykach wektora  $X$ , jak również wszystkie zmienne wpływające na udział w programie oraz na wynik zmiennej wyjściowej  $Y$  są obserwowane przez badacza [Caliendo, Kopeinig 2008]. Założenie to w praktyce badawczej jest łatwe do podważenia, a tym samym łatwo podważyć wyniki otrzymane za pomocą metody PSM.

### 3. Analiza wrażliwości oparta na metodzie granic Rosenbauma

P.R. Rosenbaum [2002] zaproponował metodę badania wrażliwości, którą można stosować w przypadku metod opartych na dopasowywaniu jednostek na podstawie obserwowanych charakterystyk  $X$ . W metodach tych, zakłada się, że dowolne dwie jednostki  $l$  oraz  $k$  o tych samych charakterystykach  $X(\mathbf{x}_k = \mathbf{x}_l)$  mają to samo prawdopodobieństwo poddania oddziaływaniu<sup>4</sup>  $\pi_k = \pi_l$ . P.R. Rosenbaum zaproponował, by analizę wrażliwości przeprowadzać, analizując, przy jakiej różnicy pomiędzy prawdopodobieństwami poddania oddziaływaniu – występującej z powodu nieobserwowanej zmiennej  $U$  mającej wpływ zarówno na zmienną wynikową, jak i na selekcję – zależność pomiędzy oddziaływaniem  $D$ , a wynikami zmiennej wyjściowej  $Y$  okazałaby się nieistotna statystycznie. Zamiast jednak analizować prawdopodobieństwa poddania oddziaływaniu, zaproponował, by analizę wrażliwości prowadzić w oparciu o szanse poddania oddziaływaniu. W metodzie zaproponowanej przez P.R. Rosenbauma [2002] badamy, jaka jest maksymalna wartość parametru  $\Gamma$  (wynikająca z występowania zmiennej nieobserwowanej  $U$ ) spełniającego nierówności:

$$\frac{1}{\Gamma} \leq \frac{\frac{\pi_k}{1-\pi_k}}{\frac{\pi_l}{1-\pi_l}} \leq \Gamma, \quad (6)$$

przy której oszacowany efekt interwencji pozostanie istotny statystycznie.

P.R. Rosenbaum [2002] proponuje spojrzeć na  $\Gamma$  przez pryzmat regresji logistycznej. Jeżeli istnieje taka nieobserwowana zmienna<sup>5</sup>  $U$  wpływająca na prawdo-

<sup>3</sup> To i pozostałe założenia metody PSM, takie jak założenie wspólnego obszaru określoności (*common support*) oraz założenie SUTVA (*Stable Unit Treatment Assumption*) można znaleźć np. u P.R. Rosenbauma i D.B. Rubina [1983], P. Strawińskiego [2008], S. Denkowskiej [2015].

<sup>4</sup> Gdy nie występuje ukryte obciążenie.

<sup>5</sup> Przybierająca wartości z przedziału  $(0;1)$ .

podobieństwo poddania oddziaływaniu, to szansę poddania oddziaływaniu dla jednostki  $k$  o obserwowanych charakterystykach  $\mathbf{x}_k$  oraz wartości  $u_k$  nieobserwowanej zmiennej  $U$  można zapisać następująco:

$$\frac{\pi_k}{1 - \pi_k} = \exp(\kappa(\mathbf{x}_k) + \gamma u_k), \quad \text{gdzie } 0 \leq u_k \leq 1. \quad (7)$$

Zatem dla dwóch jednostek  $l$  o tych samych prawdopodobieństwach poddania oddziaływaniu ( $\pi_k = \pi_l$ ) oszacowanych na podstawie obserwowanych charakterystyk  $X$ , otrzymujemy, że iloraz szans poddania oddziaływaniu dla tych jednostek zależy od  $\gamma$  oraz różnicy ( $u_k - u_l$ ):

$$\frac{\frac{\pi_k}{1 - \pi_k}}{\frac{\pi_l}{1 - \pi_l}} = \frac{\exp(\kappa(\mathbf{x}_k) + \gamma u_k)}{\exp(\kappa(\mathbf{x}_l) + \gamma u_l)} = \exp(\gamma(u_k - u_l)). \quad (8)$$

P.R. Rosenbaum [2002] pokazał, że z powodu nieobserwowanej zmiennej zachodzi następujący związek dla ilorazu szans poddania oddziaływaniu jednostek sparowanych na podstawie obserwowanych charakterystyk:

$$\frac{1}{\exp(\gamma)} \leq \frac{\frac{\pi_k}{1 - \pi_k}}{\frac{\pi_l}{1 - \pi_l}} \leq \exp(\gamma). \quad (9)$$

Ze wzorów (6) i (9) otrzymujemy, że  $\Gamma = \exp(\gamma)$ , gdzie  $\gamma$  jest parametrem regresji występującym przy zmiennej nieobserwowanej  $U$ . Zatem znając  $\Gamma$ , możemy stwierdzić, że dwa obiekty (jednostki), które przed poddaniem oddziaływaniu wydają się takie same pod względem obserwowanych charakterystyk  $X$  (lub oszacowanego *propensity score*) mogą różnić się pod względem nieobserwowanej zmiennej  $U$  wpływającej zarówno na zmienną wynikową  $Y$ , jak i na selekcję, w ten sposób, że jeden obiekt może mieć  $\Gamma$ -krotnie większą szansę trafienia do grupy poddawanej oddziaływaniu niż drugi obiekt [Rosenbaum 2005].

W badaniach eksperymentalnych opartych na randomizacji, w których nie występuje ukryte obciążenie, wartość  $\Gamma$  wynosi 1. W analizie wrażliwości, parametr  $\Gamma$  traktujemy jako miarę stopnia odejścia od badań wolnych od ukrytego obciążenia [Rosenbaum 2002]. Stwierdzoną zależność (między oddziaływaniem a zmienną wynikową) można uznać za odporną na występowanie nieobserwowanej zmiennej (wpływającej zarówno na selekcję, jak i na zmienną wynikową), gdy dopiero bardzo duże wartości parametru  $\Gamma$  wpłynęłyby na zmianę wniosków z badań (zależność przestałaby być istotna statystycznie). Jeśli wartości  $\Gamma$  bliskie 1 są w stanie zmienić wnioski otrzymane przy założeniu, że nie występuje ukryte obciążenie, to taką zależność określamy jako wrażliwą na występowanie takiego nieobserwowanego czynnika.

Analiza wrażliwości na występowanie nieobserwowanej zmiennej  $U$  zaproponowana przez P.R. Rosenbauma [2002] oparta jest na różnych testach randomiza-

cyjnych, które zależą od typu zmiennej wynikowej  $Y$  (zob. np.: [Liu, Kuramoto, Stuart 2013; Keele 2010]). W przypadku binarnej zmiennej wynikowej  $Y$  analiza wrażliwości oparta jest na teście<sup>6</sup> McNemary. Test McNemary jest wykorzystywany do sprawdzenia, czy oddziaływanie ma istotny wpływ na wynik zmiennej wyjściowej  $Y$ . Informacje dotyczące sparowanych jednostek zestawiane są w tablicy kontyngencji o wymiarach  $2 \times 2$ . Dla jednostek sparowanych na podstawie *propensity score* szanse dostania się do grupy poddawanej interwencji teoretycznie są takie same. W analizie wrażliwości Rosenbauma analizujemy, jaki iloraz szans poddania oddziaływaniu dla jednostek sparowanych, występujący z powodu nieobserwowanej zmiennej  $U$ , wpłynąłby na zmianę wniosków z badań.

Niech  $T$  oznacza liczbę wszystkich par, w których wyniki zmiennej wyjściowej  $Y$  różniły się między sobą i niech  $a$  oznacza liczbę par, w których jednostka poddana interwencji miała pozytywny wynik zmiennej wyjściowej, a niepoddana – negatywny. Dla różnych zadanych wartości  $\Gamma$  wyznaczane są prawdopodobieństwa  $p^- = \frac{1}{1+\Gamma}$  oraz  $p^+ = \frac{\Gamma}{1+\Gamma}$ , które następnie wykorzystywane są do wyznaczenia dolnej oraz górnej granicy dla prawdopodobieństwa testowego  $p$ . Obie granice dla prawdopodobieństw testowych wyznaczane są analogicznie jak prawdopodobieństwo testowe w teście dwumianowym:

$$p_{\text{dolne}} = \sum_{i=a}^T \binom{T}{i} (p^-)^i (1-p^-)^{T-i} \quad \text{oraz} \quad p_{\text{górne}} = \sum_{i=a}^T \binom{T}{i} (p^+)^i (1-p^+)^{T-i}. \quad (10)$$

Obliczenia są powtarzane dla różnych wartości  $\Gamma$ , a celem jest znalezienie największej wartości  $\Gamma$ , dla której wyliczone prawdopodobieństwo<sup>7</sup>  $p_{\text{górne}}$  nie przekroczy 0,05. Im większe wartości  $\Gamma$  są potrzebne, by prawdopodobieństwo  $p_{\text{górne}}$  przekroczyło wartość 0,05, tym silniejsza zależność musiałaby występować pomiędzy zmienną nieobserwowaną  $U$ , a dostaniem się do grupy poddanej oddziaływaniu, żeby zmienić wyniki analizy, czyli spowodować brak ich statystycznej istotności.

#### 4. Analiza wrażliwości – przykład empiryczny

Analiza wrażliwości została zastosowana do badań autorki [Denkowska 2015] nad efektem netto staży dla bezrobotnych zorganizowanych przez Powiatowy Urząd Pracy w Tarnowie w 2013 r. W badaniach została wykorzystana metoda *Propensity Score Matching*. Badania prowadzone były na podstawie danych z systemu ewidencji Syriusz. Zmienną wyjściową było zatrudnienie po 3 miesiącach od zakończenia stażu.

<sup>6</sup> W przypadku zmiennej wynikowej typu ciągłego analiza wrażliwości oparta jest na teście znaków Wilcoxa oraz punktowej estymacji Hodgesa-Lehmanna dla testu znaków [Keele 2010].

<sup>7</sup> Dolna granica dla prawdopodobieństwa testowego  $p$ ,  $p_{\text{dolne}}$  jest zawsze mniejsza od zaobserwowanego prawdopodobieństwa testowego  $p$  i tym samym jest mniej ważna.

W stażach organizowanych przez PUP w Tarnowie w 2013 r. wzięły udział 1623 osoby. Pulę kontrolną stanowiło 19217 osób nieobjętych aktywizacją w 2013 r. Zastosowana w badaniu Metoda Najbliższego Sąsiada ze zwracaniem (1 : 1, *caliper* = 0,25) dała bardzo dobre zbalansowanie wszystkich 37 zmiennych, interakcji oraz zmiennej (*wiek*)<sup>2</sup> (np. standaryzowane różnice średnich nie przekraczały na moduł 2%) uwzględnionych w modelu regresji logistycznej (zob. [Denkowska 2015]). Zastosowanie metody z limitem spowodowało usunięcie dwóch osób z grupy stażystów. Oszacowany efekt netto staży wyniósł 7,895% ( $p = 9,1492e-08$ ).

Mimo starań, by w jak najpełniejszym stopniu wykorzystać możliwości systemu ewidencji Syriusz, i tak pojawiły się wątpliwości czy obserwowana zależność pomiędzy uczestnictwem w stażach, a zatrudnieniem nie jest *de facto* spowodowana przez nieuwzględnioną w badaniu zmienną. Na przykład takie predyspozycje, jak przedsiębiorczość czy komunikatywność niewątpliwie mają silny wpływ na zatrudnienie, nasuwa się więc pytanie, jak silny musiałby być wpływ takiego nieobserwowanego czynnika na proces selekcji do grupy stażystów, by wnioski z badań na temat wpływu staży na zatrudnienie przestały być istotne statystycznie?

Analizę wrażliwości metodą granic Rosenbauma można przeprowadzić tylko w przypadku metod łączenia bez zwracania. Dlatego ponownie podjęto kroki mające na celu znalezienie zarówno modelu, jak i metody łączenia bez zwracania, dających jak najlepsze zbalansowanie zmiennych. Ostatecznie okazało się, że najlepsze zbalansowanie otrzymano w przypadku MNS bez zwracania (1 : 1, *caliper* = 0,25). Tym razem zbalansowanie 37 zmiennych, interakcji i zmiennej (*wiek*)<sup>2</sup> okazało się nieco gorsze, jednak nadal można je było uznać za dopuszczalne (zob. [Caliendo, Kopeinig 2008; Denkowska 2015]) (wszystkie standaryzowane różnice po dopasowaniu zmniejszyły się i na moduł nie przekraczały 3,5174%, a testy *t*-Studenta nie wykazały istotnych różnic między średnimi). Ocena dopasowania uzupełniono o analizę wykresów kwantyl-kwantyl oraz pudełkowych. Podobieństwo rozkładów dla zmiennych typu ciągłego oraz dyskretnego „potwierdził” bootstrapowy test Smirnova-Kołmogorowa<sup>8</sup>. Zastosowana w badaniu metoda z limitem spowodowała usunięcie 3 stażystów, dla których nie było wystarczająco dobrych dopasowań w puli kontrolnej. Tym razem oszacowany<sup>9</sup> efekt netto staży wyniósł 9,3827% z błędem standardowym wynoszącym 1,7738% ( $p = 1,2259e-07$ ).

W celu przeprowadzenia analizy wrażliwości metodą granic Rosenbauma wyniki dotyczące 1620 par zebrano w tablicy kontyngencji (tab. 1). Liczba wszystkich par, w których wyniki zmiennej wyjściowej *Y* różniły się między sobą wyniosła 840 ( $T = 496 + 344$ ), a par w których tylko stażyści mieli zatrudnienie 496 (*a*).

---

<sup>8</sup> Bootstrapowa wersja testu Smirnova-Kołmogorowa umożliwia badanie podobieństwa rozkładów w przypadku zarówno rozkładów zmiennych losowych ciągłych, jak i dyskretnych [Abadie 2002; Sekhon 2011].

<sup>9</sup> Badanie przeprowadzono w pakiecie *matching* w R.

**Tabela 1.** Tablica kontyngencji dla sparowanych jednostek

		Stażyści		Suma
		Zatrudnienie	Brak	
Grupa kontrolna	Zatrudnienie	<b>455</b>	<b>344</b>	799
	Brak	<b>496</b>	<b>325</b>	821
Suma		951	669	1620

Źródło: opracowanie własne w pakiecie *matching* w R.

W kolejnym kroku dla różnych, zadanych wartości  $\Gamma$  wyznaczone były prawdopodobieństwa  $p^-$  oraz  $p^+$ , wykorzystywane następnie do wyznaczenia dolnej oraz górnej granicy dla prawdopodobieństwa testowego  $p$  według wzorów (10). W tabeli 2 przedstawiono wyniki obliczeń dla wybranych wartości  $\Gamma$ . Największa wartość<sup>10</sup> parametru  $\Gamma$ , dla której prawdopodobieństwo  $p_{\text{górne}}$  było mniejsze od 0,05, wynosiła 1,28.

**Tabela 2.** Granice dla prawdopodobieństwa testowego  $p$ , w przypadku wybranych wartości  $\Gamma$ 

Gamma	Prawdopodobieństwa	
	$p_{\text{dolne}}$	$p_{\text{górne}}$
1,00	0,0000	0,00000
1,15	0,0000	0,00067
1,25	0,0000	0,02240
<b>1,28</b>	<b>0,0000</b>	<b>0,04784</b>
<b>1,29</b>	<b>0,0000</b>	<b>0,06003</b>
1,30	0,0000	0,07440

Źródło: obliczenia własne.

Podsumowując otrzymane wyniki, możemy stwierdzić, że oszacowany efekt netto staży przestałby być istotny statystycznie<sup>11</sup> w przypadku istnienia niewzględnionej w badaniu zmiennej silnie wpływającej na znalezienie pracy, która 1,29-krotnie zwiększałaby szansę dostania się do grupy stażystów. W przypadku, gdyby ta nieobserwowana zmienna zwiększała szansę dostania się na staż najwyżej 1,28-krotnie, obserwowana zależność pomiędzy uczestnictwem w stażu, a znalezieniem pracy nadal pozostawałaby istotna statystycznie. Zatem dla dwóch sparowanych jednostek, które wydają się takie same z punktu widzenia oszacowanego na podstawie obserwowanych zmiennych *propensity score*, jedna z nich może mieć 1,28-krotnie większą szansę poddania oddziaływaniu niż druga z powodu zmiennej

<sup>10</sup> W badaniach społecznych wartość parametru  $\Gamma$  zazwyczaj wyznaczana jest z dokładnością najwyżej do dwóch miejsc po przecinku.

<sup>11</sup> Na poziomie istotności 0,05.

nieobserwowanej  $U$  [Rosenbaum 2005]. Wartość  $\Gamma = 1,28$  w badaniach społecznych dotyczących rynku pracy może być interpretowana<sup>12</sup> jako najwyższej umiarkowana odporność na występowanie takiej ukrytej zmiennej  $U$ .

Otrzymane wyniki (w tab. 2) potwierdziła analiza przeprowadzona z wykorzystaniem pakietu *rbound* w R [Keele 2010]. W pakiecie tym metoda granic Rosenbauma jest oprogramowana dla binarnej, porządkowej oraz ciągłej zmiennej wynikowej dla wariantu łączenia<sup>13</sup>  $1 : k$ .

## 5. Podsumowanie

W badaniach obserwacyjnych nie możemy mieć pewności, że uwzględniliśmy wszystkie zmienne wpływające na wynik końcowy oraz na proces selekcji. Dlatego P.R. Rosenbaum [2005] zaleca, by badania oparte na dopasowywaniu jednostek prowadzić dwuetapowo, stosując najpierw podejście klasyczne oparte na założeniu warunkowej niezależności, a następnie w kolejnym kroku przeprowadzić ocenę wrażliwości na występowanie nieobserwowanego czynnika. Ocena wrażliwości pozwala określić, jak dużą wagę należy przykładać do otrzymanych wyników badań (oszacowanych efektów oddziaływań). Wysoka wartość gamma dodaje pewności, że obserwowana zależność ma charakter przyczynowy.

Analiza wrażliwości daje decydentom pełniejszą wiedzę o efektach interwencji i może być pomocna w podejmowaniu decyzji.

## Literatura

- Abadie A., 2002, *Bootstrap tests for distributional treatment effect in instrumental variable models*, Journal of the American Statistical Association, vol. 97 (457), s. 284–292.
- Caliendo M., Kopeinig S., 2008, *Some practical guidance for the implementation of Propensity Score Matching*, Journal of Economic Surveys, vol. 22, no. 1, s. 31–72.
- cran.r-project.org/web/packages/rbounds/rbounds.pdf (wersja z 2.07.2014).
- Denkowska S., 2015, *Wybrane metody oceny jakości dopasowania w Propensity Score Matching*, [w:] K. Jajuga, M. Walesiak (red.), *Taksonomia 24. Klasyfikacja i analiza danych – teoria i zastosowania*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, nr 384, s. 60–74.
- European Commission, 2014, *The Programming Period 2014–2020. Guidance Document on Monitoring and Evaluation. European Regional Development Fund and Cohesion Fund. Concepts and Recommendations*.
- Keele L., 2010, *An overview of rebounds: An R Package for Rosenbaum bounds sensitivity analysis with matched data*, [www.personal.psu.edu/ljk20/rbounds%20vignette.pdf](http://www.personal.psu.edu/ljk20/rbounds%20vignette.pdf) (15.05.15).
- Liu W., Kuramoto S.K., Stuart E.A., 2013, *An introduction to sensitivity analysis for unobserved confounding in non-experimental prevention research*, Prevention Science, vol. 14, no. 6, s. 570–580.

<sup>12</sup> W badaniach społecznych wartości  $\Gamma$  są zazwyczaj z przedziału od 1 do 2 [Keele 2010].

<sup>13</sup> [cran.r-project.org/web/packages/rbounds/rbounds.pdf](http://cran.r-project.org/web/packages/rbounds/rbounds.pdf) (wersja z 2.07.2014).

- Rosenbaum P.R., 2002, *Observational Studies*, Springer, New York.
- Rosenbaum P.R., 2005, *Observational study*, [w:] B.S. Everitt, D.C. Howell (red.), *Encyclopedia of Statistics in Behavioral Science*, vol. 3, John Wiley & Sons, Hoboken, NJ.
- Rosenbaum P.R., Rubin D.B., 1983, *The central role of propensity score in observational studies for casual effects*, *Biometrika*, vol. 70, no. 1, s. 41–55.
- Sekhon J.S., 2011, *Multivariate and propensity score matching software with automated balance optimization: The matching package for R*, *Journal of Statistical Software*, vol. 42, no. 7, s. 1–52.
- Strawiński P., 2008, *Quasi-eksperymentalne metody ewaluacji*, [w:] A. Haber (red.), *Środowisko i warsztat ewaluacji*, PARP, Warszawa, s. 193–220.