

MAŁGORZATA RÓSZKIEWICZ¹

IDENTYFIKACJA DETERMINANT BRAKU ODPOWIEDZI W BADANIU POLSKICH GOSPODARSTW DOMOWYCH

1. WSTĘP

Występowanie braków odpowiedzi w badaniach empirycznych i będące tego konsekwencją obciążenia wyników są wyznacznikiem jakości danych i prowadzonych na ich podstawie analiz. Wyniki wielu badań terenowych wskazują na nasilanie się występowania tego zjawiska, a dodatkowo wraz z tą tendencją, badacze odnotowują zmianę jego uwarunkowań, co prowadzi do zmian natury obciążeń z tego tytułu (por. Stinchcombe i inni, 1981; Martin, Matheson, 1999; De Heer, 1999; Steeh i inni, 2001; Stoop, 2005; Groves, 2006). Szacowanie obciążeń spowodowanych brakiem odpowiedzi wymaga ustalenia, czy są one konsekwencją braku kontaktu, czy też odmowy udziału w badaniu, a także posiadania wiedzy o mechanizmach ich powstawania. Rozpoznanie przyczyn braków odpowiedzi jest szczególnie ważne w badaniach porównawczych, gdyż ten sam poziom wskaźnika odpowiedzi (ang. *response-rate*), może być efektem diametralnie odmiennej konfiguracji dostępności respondentów i ich gotowości do współpracy, co będzie prowadzić do błędów interpretacyjnych.

W sytuacjach gdy wyniki pilotażu sugerują możliwość wystąpienia wysokiego poziomu wskaźnika braków (ang. *non-response rate*), rozważa się różne strategie redukcji wynikających stąd błędów nielosowych. Strategie te dotyczą zarówno modyfikacji metod doboru próby, jak i metod dokonywania pomiarów, a także modyfikacji metod analizy danych. W pierwszej kolejności rozważa się strategie gwarantujące osiągnięcie wymaganej liczebności próby. Wśród nich, z racji charakterystycznych działań, można wyróżnić strategie ekstensywne i intensywne. Do strategii ekstensywnych można zaliczyć działania polegające na zwielokrotnieniu rozmiaru próby losowanej lub przygotowaniu transz tzw. prób rezerwowych. Działania te nie wymagają rozpoznania przyczyn wysokiej wartości wskaźnika braków odpowiedzi, a jego poziom agreguje wówczas wszystkie potencjalne przyczyny niezrealizowania wywiadu, a więc agreguje zarówno błąd operatu, błąd braku reakcji i błąd braku realizacji oraz zagubienie i odrzucenie. Strategie te wiążą się na ogół z wysokimi kosztami realizacji badania, gdyż wymagają wylosowania większej liczby jednostek wyboru, zaangażowania i przeszkolenia

¹ Szkoła Główna Handlowa w Warszawie, Instytut Statystyki i Demografii, ul. Madalińskiego 6/8, 02-513 Warszawa, Polska, e-mail: mroszki@sgh.waw.pl.

większej grupy ankieterów, a także wydłużają proces realizacji badania terenowego. Ograniczony budżet oraz założony harmonogram prac skłaniają do stosowania strategii intensywnych. Strategie te obejmują z kolei działania, które dążą do poprawy wskaźnika odpowiedzi w założonym rozmiarze próby. Ich powodzenie zależy jednak nie tylko od działań stymulujących uczestnictwo w badaniu lecz również od postaw wobec procesu badawczego wybranych do badania jednostek. W strategiach tych, by była możliwa poprawa poziomu wskaźnika odpowiedzi, konieczne jest zrozumienie natury braków i ich determinant.

Ramy konceptualne dla wyjaśnienia uczestnictwa w badaniach terenowych opierają się na takich teoriach jak teoria wymiany społecznej (Goyder, 1987; Dillman, 2000), koncepcja społeczeństwa obywatelskiego (Brehm, 1993) oraz teoria społecznej integracji (Goyder, 1987). Z kolei teorią, która koncentruje się na interakcjach między charakterystykami jednostek w próbie i przyjętą metodologią badania jest teoria dźwigni-wyróżnika (ang. *leverage-salience theory*) (Groves i inni, 2000). Teorie te wśród czynników dostępności i gotowość współpracy jednostek próby wyróżniają zarówno czynniki leżące po stronie badanych jednostek, a które obejmują deskryptory ich ulokowania w strukturze społecznej oraz postawy wobec samego procesu badawczego, jak i czynniki leżące po stronie realizatorów badania, czyli cechy osób realizujących badanie oraz przyjętą metodologię. Ujęta w tych teoriach sekwencja zmian uwarunkowanych przyczynowo pozwala, zgodnie z definicją zaproponowaną przez Sztompkę, problem braków odpowiedzi traktować jako proces społeczny (Sztompka, 1999). Tak też jest ujmowany w literaturze tematu (Lin, Schaeffer, 1995; Groves, 2006). Z kolei sugerowany w tych teoriach zestaw uwarunkowań braku odpowiedzi pozwala sądzić, że struktura tego procesu jest wielopoziomowa, gdzie na pierwszym poziomie plasują się jego uwarunkowania związane z cechami jednostek populacji i uwarunkowania związane z warunkami realizacji badania w terenie, na drugim, uwarunkowania leżące po stronie realizatorów badanie obsługujących określone wiązki jednostek, zaś na trzecim, cechy samego procesu badawczego, który przebiega w określonej rzeczywistości i dotyczy wybranej tematyki.

Dane dotyczące zaangażowania respondentów w proces badawczy w przypadku polskich doświadczeń są rzadko, jeśli w ogóle, publikowane. Na ogół badacze ograniczają się do podania wskaźnika odpowiedzi nie charakteryzując struktury i uwarunkowań jego poziomu. W literaturze polskiej głównie są podejmowane wątki koncentrujące się na metodach redukcji błędów nielosowych (Paradysz, Szymkowiak, 2007; Piasecki, 2014). W badaniach gospodarstw domowych realizowanych przez Główny Urząd Statystyczny, w latach 2000–2012 wskaźnik odpowiedzi wahał się, w zależności od tematyki badania, od 40% do 70% (Warno, Żyra, 2013). Skala występowania tego zjawiska pozwala uznać problem rozpoznania uwarunkowań braków odpowiedzi w polskich badaniach terenowych za ważki problem badawczy. Celem artykułu jest próba rozpoznania uwarunkowań braku odpowiedzi w badaniu typu *face-to-face* zrealizowanym wśród polskich gospodarstw domowych w 2013 r. w ramach projektu Uwarunkowania

Decyzji Edukacyjnych (UDE)². W projekcie tym populację badaną stanowiły gospodarstwa domowe i ich członkowie, w których funkcjonują osoby w wieku 16–65 lat. Odrębnie przeanalizowano uwarunkowania braku kontaktu oraz odmowy udziału w tym badaniu.

W niniejszej analizie ograniczono się do rozważenia jedynie dwóch typów uwarunkowań spośród wyróżnionych wyżej, tj. cech badanych gospodarstw domowych oraz warunków realizacji badania w terenie. Efekt uwarunkowań związanych z kompetencjami ankieterów obsługujących w tym projekcie poszczególne wiązki jednostek badania, został potwierdzony przez Rószkiewicz we wcześniejszych analizach (Rószkiewicz, 2014). Efekt ten ujawnił się jako koincydencja generowania braku kontaktu i odmowy udziału w badaniu, co oznacza, że ankieterzy mniej efektywni w nawiązywaniu kontaktu byli również mniej skuteczni w nakłanianiu jednostek do współpracy. Dysponując danymi z jednego projektu badawczego nie było możliwe rozważenie uwarunkowań związanych z cechami samego procesu badawczego, który przebiegał w określonej rzeczywistości i należy być świadomym ograniczeń z tego powodu. Wybór projektu badawczego był podyktowany dostępnością danych i należy traktować go jako przykładowy dla rozważanego zagadnienia.

W diagnozowaniu determinant braku odpowiedzi odnoszących się do warunków realizacji badania w terenie wykorzystano pełną historię kontaktów z wszystkimi wylosowanym gospodarstwami domowymi. Wśród tych 122 831 wybranych gospodarstwach domowych wystąpiły jednostki, z którymi w ogóle nie nawiązano kontaktu, oraz które definitywnie odmówiły udziału w badaniu. W drugiej kolejności podjęto próbę bardziej szczegółowego rozpoznania uwarunkowań braku odpowiedzi, wśród których rozważono cechy deskryptywne jednostek badanej populacji. W tej analizie odrębnie rozważono uwarunkowania braku kontaktu oraz odmowy udziału w badaniu. W tym celu wykorzystano bazę danych, obejmującą tylko zbadane jednostki, czyli 34 753 gospodarstwa domowe, w której obok badanych cech odnotowano również liczbę wizyt ankietera w celu nawiązania kontaktu oraz w celu nakłonienia wylosowanego gospodarstwa do udziału w badaniu. Sposób konstrukcji próby (plan losowania i schemat losowania) oraz zastosowany system wag gwarantował możliwość szacowania wartości wiodących cech z dostateczną precyzją na poziomie NUTS1 oraz NUTS2. Wagi zostały tak skonstruowane, aby uzyskać zgodność struktur demograficznych gospodarstw domowych i osób z danymi pochodzącymi z zewnętrznych źródeł (Narodowy Spis Powszechny z 2011 roku).

² Badanie zostało prowadzone w ramach projektu systemowego pod nazwą „Badanie jakości i efektywności edukacji oraz instytucjonalizacja zaplecza badawczego” współfinansowanego ze środków Europejskiego Funduszu Społecznego w ramach Programu Operacyjnego Kapitał Ludzki. Projekt koordynowany przez Instytut Badań Edukacyjnych, którego metodologię zaproponował zespół badaczy ze Szkoły Głównej Handlowej w Warszawie. Pełne informacje o projekcie są ulokowane na stronie: <http://eduentuzjasci.pl/ude>.

2. PODEJŚCIE BADAWCZE

W diagnozowaniu uwarunkowań braku odpowiedzi leżących po stronie jednostek populacji podstawową trudnością jest niemożność dysponowania danymi o cechach tych, z którymi w ogóle nie można było nawiązać kontaktu, oraz którzy definitywnie odmówili udziału w badaniu. W przełamaniu tej trudności wykorzystuje się pogląd dominujący w literaturze tematu głoszący, że nie ma empirycznych dowodów na występowanie grup jednostek permanentnie nieuczestniczących w badaniach (ang. *hard-core*), które rezygnują zawsze i kontestują każdy proces badawczy. Uznaje się, że jeśli grupy takie istnieją, to są bardzo nieliczne (Goyder, 1987). Według Schnell (1997) i Goyder (1987) najtrudniejsi respondenci staliby się nieuczestniczącymi w badaniu, gdyby nie podejmowano dodatkowych wysiłków skłaniających ich do współpracy. Na tej podstawie, za Lin, Schaeffer (1995), jednostki, które zbadano tylko i wyłącznie dzięki podjęciu większych wysiłków można utożsamiać z tymi, które w badaniu nie uczestniczyły z powodu definitywnego braku możliwości nawiązania kontaktu lub definitywnej odmowy. W niniejszej analizie za miarę trudności w pozyskaniu respondenta przyjęto liczbę wizyt złożonych przez ankieterów zanim przeprowadzono wywiad, a za tzw. „trudnych” respondentów uznano tych, którzy wymagali więcej niż jednej wizyty, czy to w celu nawiązania kontaktu, czy też w celu nakłonienia do współpracy. Różnice między deskryptorami grupy „łatwych” respondentów oraz respondentów „trudnych”, z którymi utożsamiono brak odpowiedzi, pozwoliły na wnioskowanie zarówno o determinantach braku dostępności jak i braku gotowości do współpracy. W analizie czynników leżących po stronie jednostek populacji wykorzystano model regresji logistycznej, odrębnie dla braku kontaktu oraz braku współpracy.

W celi weryfikacji założenia o analogii między jednostkami, z którymi w ogóle nie nawiązano współpracy mimo kolejnych prób, i tzw. jednostkami „trudnymi”, z którymi tę współpracę nawiązano dopiero w wyniku wielokrotnych wizyt, dokonano porównania grupy jednostek badania, z którymi nawiązano kontakt po co najmniej dwukrotnych próbach z grupą jednostek, których dotyczył definitywny brak odpowiedzi. Podstawą tej oceny była pełna historia kontaktów, w której zarejestrowano wszystkie próby podjęcia współpracy z wylosowanymi jednostkami, niezależnie od ich wyniku. Zakres porównań dotyczył charakterystyk odnoszących się do warunków realizacji badania w terenie, gdyż tylko takie cechy wylosowanych jednostek były możliwe do odnotowania przy definitywnym braku odpowiedzi (całkowity brak kontaktu lub definitywna odmowa udziału w badaniu). Analiza porównawcza miała odpowiedzieć na pytanie, czy warunki realizacji badania w terenie determinują w inny sposób definitywny brak odpowiedzi oraz uzyskanie odpowiedzi dopiero po wielokrotnych próbach, co podważałoby tezę o zakładanej analogii. W weryfikacji tej tezy przeprowadzono prostą analizę porównawczą rozkładów wyróżnionych cech w obu podpopulacjach oraz wykorzystano model hybrydowy łącząc procedurę drzewa klasyfikacyjnego z wielomianowym modelem logitowym, w celu uwzględnienia nie tylko efektów głównych rozważanych warunków realizacji badania w terenie ale również interakcji między nimi.

3. ZAKRES REAKCJI ORAZ REALIZACJI BADANIA W RAMACH PROJEKTU UDE

W zrealizowanym w 2013 r. badaniu terenowym w ramach projektu UDE nawiązano kontakt z 75,6% gospodarstw domowych spośród 122 831 wylosowanych do próby, z których 40% wzięło udział w badaniu. By osiągnąć wymaganą liczebność próby na poziomie 34 753 gospodarstw domowych konieczne było wykorzystanie prób rezerwowych. Jedynie 27,53% jednostek badania, wśród których efektywnie zrealizowano wywiady pochodziło z próby podstawowej, 22,81% pochodziło z pierwszej rezerwy, 17,94% z drugiej rezerwy, zaś pozostałe 31,73% zbadanych jednostek pochodziło z trzeciej lub kolejnych prób rezerwowych. Najczęściej nie podawano przyczyny odmowy udziału w badaniu.

Wskaźnik reakcji, kształtujący się na dość wysokim poziomie, wykazywał znaczne zróżnicowanie regionalne. Najłatwiej respondenci byli osiągalni w województwie zachodniopomorskim oraz wielkopolskim, gdzie niespełna 16% kontaktów skończyło się definitywnym niepowodzeniem. Najtrudniej zaś, w województwach dolnośląskim, podlaskim i opolskim, gdzie częstość braku kontaktu była ponad dwukrotnie wyższa. Trudność w nawiązaniu kontaktu przekładała się na konieczność podejmowania wielokrotnych prób. Jednakże skuteczność w nawiązywaniu kontaktu przy kolejnych wizytach była niższa niż w pierwszym i drugim podejściu. Jak ilustrują to dane zestawione w tabeli 1, w ramach pierwszej wizyty udawało się nawiązać kontakt w 84% przypadków (wylosowanych adresów), natomiast w ramach dalszych wizyt kontakt nawiązywano już tylko w około 59% przypadków.

Tabela 1.

Wskaźnik reakcji w badaniu UDE w 2013 r. z uwzględnieniem liczby wizyt ankieratorów

Reakcja	Liczba wizyt					Ogółem
	1	2	3	4	5	
Brak kontaktu	16,0%	23,1%	61,3%	49,8%	48,5%	24,4%
Kontakt nawiązany	84,0%	76,9%	38,7%	50,2%	51,5%	75,6%
Ogółem	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%

Źródło: opracowanie własne na podstawie danych UDE.

Nawiązanie kontaktu nie gwarantowało jeszcze zrealizowania wywiadów. Jak ilustrują to dane zestawione w tabeli 2, przy pierwszej wizycie udało się zrealizować wywiady jedynie wśród 37,4% gospodarstw należących do populacji badanej. Kolejne wizyty po nawiązaniu kontaktu dawały tylko trochę lepsze rezultaty.

Podobnie jak wskaźnik reakcji również wskaźnik realizacji, ujawniający zakres odmawiania współpracy, wśród większości jednostek, z którymi nawiązano kontakt wykazywał znaczne zróżnicowanie regionalne. Najtrudniej respondenci podejmowali współpracę w województwie warmińsko-mazurskim oraz dolnośląskim, gdzie nie-

spełna 33% nawiązanych kontaktów skończyło się podjęciem współpracy. Najłatwiej zaś w województwach podlaskim i świętokrzyskim, przy czym tam liczba zrealizowanych wywiadów wśród nawiązanych kontaktów ledwie przekroczyła połowę.

Tabela 2.

Wskaźnik realizacji wśród nawiązanych kontaktów w badaniu UDE w 2013 r.
z uwzględnieniem liczby wizyt ankieterów

Realizacja	Liczba wizyt					Ogółem
	1	2	3	4	5	
Odmowa	62,6%	52,8%	54,8%	52,1%	38,0%	59,9%
Wywiad zrealizowany	37,4%	47,2%	45,2%	47,9%	62,0%	40,1%
Ogółem	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%

Źródło: opracowanie własne na podstawie danych UDE.

Zróżnicowanie regionalne wskaźnika braku reakcji wskazuje na występowanie zróżnicowania reguł dostępności respondentów. Z kolei zróżnicowanie regionalne wskaźnika braku realizacji sugeruje występowanie zróżnicowania reguł uczestnictwa w badaniu.

4. ZAKRES ANALOGII MIĘDZY JEDNOSTKAMI GENERUJĄCYMI BRAK ODPOWIEDZI I TZW. JEDNOSTKAMI „TRUDNYMI”

W celu weryfikacji poprawności założonej analogii w pierwszej kolejności dokonano porównania rozkładów cech opisujących warunki realizacji badania w obu grupach jednostek, tj. w grupie jednostek „trudnych”, z którymi współpracę nawiązano dopiero w wyniku wielokrotnych wizyt, i grupą jednostek, których dotyczył definitywny brak odpowiedzi. Wśród cech odnoszących się do warunków realizacji procesu badawczego w terenie wyróżniono następujące charakterystyki: makroregion, województwo, charakter miejsca zamieszkania, dzień tygodnia i porę dnia ostatniego kontaktu, kiedy podejmowano próbę realizacji badania. Odnotowano bardzo wysokie podobieństwo tych struktur, a w każdym przypadku współczynnik V-Cramera kształtował się poniżej 0,08, potwierdzając brak zależności między cechami opisującymi warunki realizacji badania i przynależnością do jednej z dwóch porównywanych grup jednostek. W drugiej kolejności podjęto próbę sprawdzenia podobieństwa wielowymiarowego. Przyjęto, że uzasadnieniem tezy o analogii między rozważanymi tu grupami jednostek będą charakteryzujące się wysokim podobieństwem parametry strukturalne modelu opisującego reakcję jednostek w procesie badawczym uwzględniającego odrębnie respondentów „trudnych” jak i generujących braki odpowiedzi. Podejścia takie jest możliwe z wykorzystaniem wielomianowego modelu regresji logistycznej.

nej. W modelowaniu prawdopodobieństwa reakcji jednostki w procesie badawczym wykorzystano podejście hybrydowe, łącząc wielomianowy model logitowy z modelem drzewa klasyfikacyjnego, poprzez uznanie przynależność do podklas wyodrębnionych w procedurze drzewa klasyfikacyjnego za dodatkowy predyktor w modelu logitowym. Za takim rozwiązaniem przemawiał fakt, że model drzewa klasyfikacyjnego pozwala wyodrębnić podklasy badanych jednostek ze względu na istotne interakcje między predyktorami braku odpowiedzi, jednakże nie daje możliwości oszacowania prawdopodobieństwa braku odpowiedzi dla każdej z rozważanych konfiguracji predyktorów, czyli *de facto* dla każdego przypadku empirycznego, co z kolei jest możliwe w modelu regresji logistycznej. Ponadto w modelu regresji logistycznej, w przeciwieństwie do modelu drzewa klasyfikacyjnego, istnieje możliwość oceny siły wpływu poszczególnych predyktorów na prawdopodobieństwo braku odpowiedzi. Natomiast żadne z tych podejść oddzielnie nie daje możliwości oceny wpływu efektów interakcji między predyktorami na to prawdopodobieństwo. Możliwości takie daje natomiast zastosowanie podejścia hybrydowego.

W celu dokonania hybrydyzacji, w pierwszym etapie analizy zastosowano procedurę drzewa klasyfikacyjnego, ograniczając ją do trzech rekurencji³, które wyznaczyły 18 podklas jednostek badania istotnie różniących się wskaźnikiem braku odpowiedzi. Jakość uzyskanego rozwiązania wyrażał współczynniki poprawnej klasyfikacji gospodarstw domowych na podstawie uzyskanego modelu ze względu na ich faktyczne zachowanie w zrealizowanym badaniu, który wyniósł 57,76%, przy czym poprawność klasyfikacji samych braków odpowiedzi opisywana tzw. współczynnikiem czułości, wyniosła 58,9%. Charakterystykę wyznaczonych 18 podklas w modelu drzewa klasyfikacyjnego zestawia tabela 3.

Tabela 3.

Charakterystyki podklasy wyodrębnionych w modelu CHAID

Podklasa	Charakterystyki podklasy według kategorii predyktorów	Rozmiar	Oszacowane prawdopodobieństwo braku odpowiedzi
Podklasa nr 1	miasto 10 tys.–19,9 tys.	7,1%	0,793
Podklasa nr 2	miasto do 10 tys.	4,7%	0,759
Podklasa nr 3	mazowieckie, dolnośląskie, miasta 500 tys.–999,9 tys.	7,1%	0,906
Podklasa nr 4	mazowieckie, lubelskie, podlaskie, pomorskie, 20 tys.–99,9 tys.	4,5%	0,856

³ Takie przycięcie rozwiązania było podyktowane potrzebą uzyskania w miarę czytelnego rozwiązania w modelu hybrydowym, w którym wyodrębnione podklasy pełniły funkcję kategorii predyktora opisującego interakcje między wyodrębnionymi cechami warunków realizacji badania w terenie. Zbyt duża liczba kategorii prowadziłaby do nadmiernie rozbudowanego rozwiązania, trudnego w przejrzystej interpretacji.

Tabela 3. (cd.)

Podklasa	Charakterystyki podklasy według kategorii predyktorów	Rozmiar	Oszacowane prawdopodobieństwo braku odpowiedzi
Podklasa nr 5	małopolskie, śląskie, dolnośląskie, 20 tys.–99,9 tys.	5,7%	0,826
Podklasa nr 6	kujawsko-pomorskie, opolskie, lubuskie, wielkopolskie, 20 tys.–99,9 tys.	3,8%	0,739
Podklasa nr 7	gmina wiejska, rano (do godz. 12.00)	5,8%	0,814
Podklasa nr 8	mazowieckie, lubelskie, podlaskie, świętokrzyskie, pomorskie, miasto 200 tys.–499,9 tys.	8,6%	0,895
Podklasa nr 9	śląskie, kujawsko-pomorskie, zachodniopomorskie, 200 tys.–499,9 tys.	3,6%	0,764
Podklasa nr 10	mazowieckie, śląskie, dolnośląskie, miasto 100 tys.–199,9 tys.	4,1%	0,885
Podklasa nr 11	małopolskie, warmińsko-mazurskie, opolskie, zachodniopomorskie, podkarpackie, miasto 100 tys.–199,9 tys.	6,5%	0,843
Podklasa nr 12	kujawsko-pomorskie, lubuskie, wielkopolskie, miasto 100 tys.–199,9 tys.	4,3%	0,773
Podklasa nr 13	Warszawa, miasto 500 tys.–999,9 tys., małopolskie, łódzkie, wielkopolskie, rano (do 12.00) i popołudniu (12.00–17.00)	3,9%	0,818
Podklasa nr 14	Warszawa, miasto 500 tys.–999,9 tys., małopolskie, łódzkie, wielkopolskie, wieczorem (po 17.00)	3,3%	0,846
Podklasa nr 15	miasto 20 tys.–49,9 tys., warmińsko-mazurskie, świętokrzyskie, łódzkie, zachodnio-pomorskie, wtorek, środa, czwartek	3,7%	0,789
Podklasa nr 16	miasto 20 tys.–49,9 tys., warmińsko-mazurskie, świętokrzyskie, łódzkie, zachodnio-pomorskie, piątek, sobota, niedziela, poniedziałek	4,0%	0,835
Podklasa nr 17	gmina wiejska, mazowieckie, lubelskie, podlaskie, małopolskie, łódzkie, kujawsko-pomorskie, świętokrzyskie, lubuskie, podkarpackie, popołudniu (12.00–17.00) i wieczorem (po 17.00)	12,3%	0,686
Podklasa nr 18	gmina wiejska, warmińsko-mazurskie, opolskie, łódzkie, pomorskie, zachodniopomorskie, wielkopolskie, dolnośląskie, popołudniu (12.00–17.00) i wieczorem (po 17.00)	7,0%	0,768

Źródło: opracowanie własne.

Przy ograniczeniu rozwiązania do trzech rekurencji podklasy istotnie różniące się wskaźnikiem braku odpowiedzi objęły koincydencję takich cech spośród rozważanych warunków realizacji badania jak: województwo, klasa miejsca zamieszkania, dzień tygodnia oraz pora dnia ostatniego kontaktu. Tylko dwie podklasy odnosiły się do kategorii, które opisywały efekty główne, a dotyczyły dwóch kategorii miejsca zamieszkania. Podklasy te, z racji uwzględnienia ich jako odrębne kategorie również w zmiennej opisującej klasę miejsca zamieszkania, zostały w dalszej analizie pominięte. W pozostałych przypadkach podklasy dotyczyły interakcji drugiego stopnia, obejmujących województwo i klasę miejsca zamieszkania, a w jednym przypadku – klasę miejsca zamieszkania i porę dnia, oraz interakcji trzeciego stopnia odnoszących się do województwa, klasy miejsca zamieszkania oraz dnia tygodnia lub pory dnia ostatniego kontaktu.

W drugim etapie oszacowano wielomianowy model regresji logistycznej dla trzech możliwych reakcji jednostki w procesie badawczym, tj. odpowiedzi w pierwszym kontakcie ($s = 1$), odpowiedzi po wielokrotnych próbach kontaktu ($s = 2$) oraz definitywnym braku odpowiedzi ($s = 3$), w którym obok predyktorów odnoszących się do wymienionych cech, opisujących warunki realizacji badania w terenie, uwzględniono również interakcje między nimi, wyodrębnione w procedurze drzewa klasyfikacyjnego, jako typy podklas badanych jednostek. Jako kategorię odniesienia przyjęto odpowiedź respondenta w pierwszym kontakcie ($s = 1$). Na tej podstawie formuła wielopoziomowego modelu regresji logistycznej wielomianowej przyjęła postać:

$$\log\left(\frac{\pi_{ij}^s}{\pi_{ij}^1}\right) = \boldsymbol{\beta}^{(s)T} \mathbf{x}_{ij}^{(s)} \quad (1)$$

gdzie:

$\pi_{ij}^{(s)}$ – prawdopodobieństwo reakcji jednostki, czyli $\pi_{ij}^{(s)} = P(y_{ij} = s)$ dla $s = 1, 2, 3$,
 $\mathbf{x}_{ij}^{(s)}$ – wektor predyktorów odnoszących się do cech warunków badania w terenie,
 $\boldsymbol{\beta}^{(s)}$ – wektor współczynników modelu.

Wyniki oszacowanego modelu hybrydowego zestawiają druga i trzecia kolumna tabeli 4. Tabela ta zestawia oceny parametrów modelu odnoszących się do wybranych cech realizacji badania w terenie, które pokazują ich wpływ na możliwość podjęcia współpracy dopiero po wielokrotnych próbach oraz na możliwość definitywnego braku odpowiedzi, w porównaniu do możliwości podjęcia współpracy w pierwszym kontakcie uwzględniając kategorię referencyjną, przy kontrolowanym wpływie pozostałych zmiennych. Dla poszczególnych predyktorów przyjęto następujące kategorie referencyjne: dla makroregionu był to makroregion małopolski, dla województw było to województwo zachodniopomorskie, dla klasy miejsca zamieszkania było to miasto Warszawa, dla dnia tygodnia była to niedziela i dla pory dnia był to wieczór (po godzinie 17.00). Dla uwzględnionych w modelu efektów interakcji między predyktorami kategorię odniesienia stanowiła ostatnia podklasa wymieniona w tabeli 3.

Tabela 4.

Oceny współczynników regresji logistycznej w modelach hybrydowych, wielomianowym i binarnym, dla braku odpowiedzi na podstawie wyników projektu UDE w 2013 r.

Predyktor/kategoria odniesienia	Wielomianowy model regresji logistycznej		Model regresji binarnej
	Pozyskanie odpowiedzi w kolejnych próbach względem próby pierwszej	Definitywny brak odpowiedzi względem odpowiedzi w pierwszej próbie	Brak odpowiedzi względem odpowiedzi w pierwszej próbie
Stała	-2,297	0,603	0,824
Makroregion/ małopolski			
centralny	0,012	0,852**	0,0392**
wielkopolski	-0,096	-0,492**	-0,566**
śląski	0,110	0,161**	0,018
zachodni	-0,648*	-0,398*	-0,422*
pomorski	0,305	0,343*	-0,135
północno-wschodni	-0,366	0,711**	0,423**
wschodni	-0,012	0,870**	0,673**
Województwo/ zachodniopomorskie			
dolnośląskie	1,313**	1,355**	1,071**
kujawsko-pomorskie	0,233	0,954**	0,899**
lubelskie	1,020**	-0,508**	-0,820
lubuskie	0,871**	0,771**	0,731**
łódzkie	0,701**	-0,169	-0,165
małopolskie	0,953**	0,597**	0,483**
mazowieckie	0,630**	-0,181	-0,524
opolskie	0,424	0,440*	0,382*
podkarpackie	0,792**	0,544**	0,473**
podlaskie	1,618**	-0,247*	-0,510**
pomorskie	0,341**	-0,207**	0,282**
śląskie	0,749*	0,485**	0,476*
świętokrzyskie	1,029**	0,269	-0,367*
warmińsko-mazurskie	0,730**	0,079	0,222*

Predyktor/kategoria odniesienia	Wielomianowy model regresji logistycznej		Model regresji binarnej
	Pozyskanie odpowiedzi w kolejnych próbach względem próby pierwszej	Definitywny brak odpowiedzi względem odpowiedzi w pierwszej próbie	Brak odpowiedzi względem odpowiedzi w pierwszej próbie
wielkopolskie	0,765**	0,953**	0,975**
Klasa miejsca zamieszkania/Warszawa			
gmina wiejska	0,909**	-0,011	-0,023
miasto do 10 tys. mieszkańców	0,832**	-0,019	0,038
miasto 10.000–19.999 mieszkańców	0,773**	0,221*	-0,007
miasto 20.000–49.999 mieszkańców	0,664**	0,056*	-0,126**
miasto 50.000–99.999 mieszkańców	0,792**	0,207*	0,045
miasto 100.000–199.999 mieszkańców	0,981**	0,019	0,228
miasto 200.000–499.999 mieszkańców	0,696**	-0,032	-0,253
miasto 500.000–999.999 mieszkańców	1,123**	0,416**	1,063**
Dzień tygodnia ostatniego kontaktu/niedziela			
poniedziałek	-0,122*	-0,191**	-0,086**
wtorek	-0,188**	-0,310**	-0,122**
środa	-0,067	-0,503**	-0,324**
czwartek	-0,232**	-0,324**	-0,08*
piątek	-0,229**	-0,291**	0,043
sobota	-0,254**	-0,019	0,083
Pora dnia ostatniego kontaktu/wieczorem (po 17.00)			
rano (do 12.00)	-0,135**	0,02	0,179**
popołudniu (12.00–17.00)	-0,27**	-0,09**	0,058**
Podklasy interakcji/podklasa nr 18			
podklasa nr 3	1,110**	0,645**	0,751**
podklasa nr 4	0,390**	0,395**	0,642**
podklasa nr 5	0,151	-0,013	0,285**
podklasa nr 6	0,081	-0,225**	-0,093
podklasa nr 7	0,002	0,179**	0,133**

Tabela 4. (cd.)

Predyktor/kategoria odniesienia	Wielomianowy model regresji logistycznej		Model regresji binarnej
	Pozyskanie odpowiedzi w kolejnych próbach względem próby pierwszej	Definitywny brak odpowiedzi względem odpowiedzi w pierwszej próbie	Brak odpowiedzi względem odpowiedzi w pierwszej próbie
podklasa nr 8	0,635**	0,852**	0,932**
podklasa nr 9	0,001	0,002	0,014
podklasa nr 10	0,253*	0,604**	0,723**
podklasa nr 11	0,027	0,241**	0,441**
podklasa nr 12	0,002	0,0001	0,090
podklasa nr 13	-0,242**	-0,143*	-0,839**
podklasa nr 14	0,001	0,0001	-0,538**
podklasa nr 15	-0,146	-0,069	0,252**
podklasa nr 16	0,002	0,005	0,395**
podklasa nr 17	-0,457**	-0,401**	-0,427**

* – istotność na poziomie 0,05, ** – istotność na poziomie 0,01.

Źródło: opracowanie własne na podstawie danych UDE.

Jak wskazują zestawione w tabeli 4 wartości, w analizie wielowymiarowej uzyskano jedynie częściową zbieżność wyników oszacowań odnoszących się do wyodrębnionych reakcji jednostek w procesie badawczym. W większości, ale nie we wszystkich przypadkach, wystąpiła zgodność kierunku oddziaływań poszczególnych kategorii predyktorów zarówno na ryzyko odpowiedzi dopiero po wielokrotnych próbach kontaktu jak i ryzyko definitywnego braku odpowiedzi. Zgodność kierunku oddziaływań oznacza, że kategorie te są wspólnymi stymulantami (wartości dodatnie oszacowanych współczynników regresji) lub destymulantami (wartości ujemne dla oszacowanych współczynników regresji) dla występowania tzw. „trudnych” respondentów oraz występowania definitywnego braku odpowiedzi. Do wspólnych stymulant po stronie warunków realizacji badania należą: lokalizacja w województwie dolnośląskim, lubuskim, małopolskim, podkarpackim, śląskim i wielkopolskim, miasta o liczbie ludności od 10 tys. do 19,9 tys. mieszkańców, a także od 20 tys. do 49,9 tys. mieszkańców, od 50 tys. do 99,9 tys. mieszkańców oraz powyżej 500 tys. mieszkańców, oraz interakcje drugiego stopnia warunków badania definiujące podklasy nr 3, 4, 8 i 10. Do wspólnych destymulant należą: makroregion zachodni, realizacja badania w poniedziałek, wtorek, czwartek i piątek, popołudniu (między 12.00 a 17.00) oraz interakcje trzeciego stopnia warunków badania definiujące podklasy nr 13 i 17.

Zarówno wysoka zgodność jednowymiarowych rozkładów poszczególnych predyktorów w obu analizowanych grupach jednostek badania oraz znaczna zbieżność wyników analizy wielowymiarowej, odnotowane dla gospodarstw domowych objętych projektem UDE, dała podstawę do uznania tezę o analogii między tzw. „trudnymi” respondentami, z którymi współpracę nawiązano dopiero w wyniku wielokrotnych wizyt i jednostkami, których dotyczył definitywny brak odpowiedzi. Na bazie tego założenia finalnie oszacowano współczynniki modelu regresji logistycznej dla prawdopodobieństwa braku odpowiedzi w pierwszym kontakcie ze względu na wyróżnione warunki badania. Wyniki zestawia kolumna czwarta tabeli 4. Różnice w wartościach oszacowanych współczynników dla tego uproszczonego modelu ryzyka braku, tzn. bazującego na założonej analogii, z wartościami oszacowanymi dla modelu ryzyka definitywnego braku odpowiedzi, pokazuje zakres niedokładności wynikający z tego założenia. Zbieżność rozwiązania, pod względem kierunku oddziaływania poszczególnych kategorii predyktorów na prawdopodobieństwo braku odpowiedzi, dotyczy wszystkich poza jednym makroregionem (pomorski), prawie wszystkich województwami (poza województwem pomorskim i świętokrzyskim), miast o liczbie mieszkańców powyżej 500 tys., kolejnych czterech dni tygodnia (od poniedziałku do czwartku) oraz realizacji badania popołudniu (od 12.00 do 17.00), a także dwunastu podklas określających interakcje drugiego i trzeciego stopnia rozważanych warunków badania.

Przyjęcie tezy o występującej analogii między tzw. „trudnymi” respondentami i jednostkami, które generują brak odpowiedzi dało możliwość identyfikacji tych cech gospodarstw domowych, które determinują zarówno ryzyko braku kontaktu jaki i ryzyko braku odpowiedzi.

5. IDENTYFIKACJA WPLYWU CECH JEDNOSTEK BADANIA NA RYZYKO BRAKU KONTAKTU ORAZ RYZYKO ODMOWY UDZIAŁU W BADANIU

W identyfikacji cech badanych gospodarstw na ryzyko braku kontaktu oraz ryzyko braku odpowiedzi uwzględniono następujące charakterystyki głowy i samego gospodarstwa domowego: płeć, wiek, stan cywilny, poziom wykształcenia status na rynku pracy, źródło utrzymania, klasa miejsca zamieszkania oraz wielkość gospodarstwa domowego. Wyniki oszacowanych modeli regresji logistycznej, odrębnie dla braku kontaktu oraz odmowy udziału w badaniu, zestawiono w tabeli 5. W oszacowanych modelach przyjęto następujące kategorie referencyjne: dla płci – kobieta, dla stanu cywilnego – po rozwodzie lub w separacji, dla wykształcenia – brak lub nieukończone wykształcenie podstawowe, dla statusu na rynku pracy – bierność zawodowa, dla klasy miejsca zamieszkiwanie – gmina wiejska oraz dla źródła utrzymania – źródło niezarobkowe inne niż kapitał lub nieruchomości.

Tabela 5.

Oceny współczynników regresji logistycznej dla predyktorów prawdopodobieństwa braku kontaktu oraz prawdopodobieństwa odmowy udziału w badaniu UDE w 2013 r., których chociaż jedna kategoria okazała się istotna statystycznie

Predyktor/kategoria odniesienia	Odmowa		Brak kontaktu	
	Współczynnik	Istotność	Współczynnik	Istotność
Źródło utrzymania/niezarobkowe				
praca najemna	-0,071	0,131	0,370	0,006
praca na własny rachunek	-0,013	0,846	-0,836	0,000
rolnictwo	0,152	0,042	-1,083	0,002
kapitał	2,398	0,000	-17,379	0,999
nieruchomość	-20,425	0,998	0,130	0,839
emerytura	0,011	0,836	0,205	0,182
renta	-0,331	0,000	-0,374	0,101
Płeć/kobieta	0,079	0,006	-0,271	0,000
Wiek	0,003	0,010		
Stan cywilny/rozwidziony, rozwiedziona, w separacji				
wolny	0,010	0,828		
związek	-0,158	0,010		
wdowieństwo	-0,221	0,000		
Wykształcenie/brak lub niepełne podstawowe				
podstawowe/gimnazjalne	0,105	0,352		
zasadnicze zawodowe	0,239	0,025		
średnie	0,263	0,013		
pomaturalne/policealne	0,367	0,002		
licencjackie/inżynierskie	0,184	0,117		
magisterskie	0,359	0,001		
stopień naukowy	-0,578	0,034		
Aktywność zawodowa /bierny				
pracujący	0,175	0,000		
bezrobotny	-0,283	0,000		

Predyktor/kategoria odniesienia	Odmowa		Brak kontaktu	
	Współczynnik	Istotność	Współczynnik	Istotność
Klasa miejsca zamieszkania/gmina wiejska				
miasto do 10 tys.	-0,236	0,004	-0,233	0,426
miasto 10 tys.–19,9 tys.	0,309	0,000	0,048	0,809
miasto 20 tys.–49,9 tys.	0,373	0,000	1,252	0,000
miasto 50 tys.–99,9 tys.	0,299	0,000	1,521	0,000
miasto 100 tys.–199,9 tys.	0,430	0,000	1,307	0,000
miasto 200 tys.–499,9 tys.	1,109	0,000	1,450	0,000
miasto 500 tys.–999,9 tys.	0,965	0,000	0,157	0,613
Warszawa	0,952	0,000	2,707	0,000
Liczba osób w gospodarstwie domowym	0,084	0,000	0,167	0,000
Stała	-2,211	0,000	-4,816	0,000

Źródło: opracowanie własne na podstawie danych UDE.

Wyniki oszacowań wskazują, że kombinacja cech składających się na kategorie referencyjne ogranicza ryzyko braku kontaktu. Dla tej grupy ryzyko braku kontaktu było niższe średnio o 47% względem pozostałych gospodarstw. Ponadto odnotowano, że ryzyko braku kontaktu wzrasta wraz z wiekiem głowy gospodarstwa domowego, a w przypadku gdy głową gospodarstwa domowego jest mężczyzna brak kontaktu staje się bardziej prawdopodobny przy kontrolowanym wpływie pozostałych cech. W porównaniu z grupą referencyjną, funkcjonowanie w związkach oraz wdowieństwo zmniejsza ryzyko braku kontaktu, podobnie posiadanie stopni naukowych. Dla wszystkich pozostałych poziomów wykształcenia, ryzyko braku kontaktu jest wyższe w porównaniu z grupą referencyjną. Najwyższym ryzykiem braku kontaktu względem grupy referencyjnej charakteryzują się gospodarstwa, których głowa legitymuje się wykształceniem pomaturalnym/policealnym lub ma ukończone studia magisterskie. W porównaniu z grupą referencyjną, status pracującego zwiększa ryzyko braku kontaktu, natomiast status bezrobotnego ryzyko to zmniejsza. W porównaniu z grupą referencyjną, im większy ośrodek miejski tym ryzyko braku kontaktu jest wyższe. Najłatwiej nawiązać kontakt w małych miejscowościach, a najtrudniej w dużych miastach. Z kolei jeśli źródłem utrzymania był kapitał, rolnictwo lub emerytura to ryzyko braku kontaktu było wyższe porównaniu z grupą referencyjną, zaś niższe w przypadku pracy lub renty. Stymulantą ryzyka braku kontaktu okazała się również wielkość gospodarstwa domowego – im większe gospodarstwo domowe tym ryzyko to było wyższe. W przypadku uwarunkowań odmowy udziału w badaniu spośród

rozważanych cech głowy i gospodarstwa domowego istotnie statystycznie okazały się jedynie płeć, miejsce zamieszkania, główne źródło utrzymania oraz wielkość gospodarstwa domowego. Wyniki oszacowań wskazują, że kombinacja cech składających się na kategorie referencyjne sprzyja realizacji wywiadu. W takiej grupie gospodarstw ryzyko odmowy przy nawiązaniu kontaktu jest dwukrotnie mniejsze względem pozostałych gospodarstw domowych. Ponadto odnotowano, że w przypadku, gdy głową gospodarstwa domowego jest kobieta, odmowa udziału w badaniu staje się bardziej prawdopodobna. W porównaniu z grupą referencyjną, im większy ośrodek miejski tym ryzyko odmowy jest wyższe. Jeśli już kontakt zostaje nawiązany, to najłatwiej przeprowadzić wywiad na terenach wiejskich, najtrudniej zaś w mniejszych, ale nie najmniejszych miejscowościach. W porównaniu z grupą referencyjną, jeśli źródłem utrzymania jest praca na własny rachunek lub rolnictwo, tym ryzyko odmowy jest niższe, zaś wyższe w przypadku pracy najemnej. Jedynie te trzy źródła utrzymania istotnie zmieniały ryzyko odmowy względem grupy referencyjnej. Również w przypadku ryzyka odmowy, wielkość gospodarstwa domowego okazała się stymulantą.

6. DYSKUSJA

Uzyskane wyniki bazują na założeniu o analogii między jednostkami, z którymi w ogóle nie nawiązano współpracy w kolejnych próbach i jednostkami, z którymi tę współpracę nawiązano dopiero w wyniku wielokrotnych wizyt. Pozwala to nie tylko rozszerzyć wiedzę o pochodzeniu braków odpowiedzi, ale również identyfikować segmenty respondentów wymagające większych wysiłków w zakresie perswazyjności i budowania pozytywnego nastawienia do badania. Wykazano, że istnieje związek między cechami społeczno-ekonomicznymi badanych gospodarstw i ich dostępnością oraz ich gotowością do współpracy. Powyższe wyniki wskazują bowiem, że obie składowe braku odpowiedzi, czyli brak kontaktu oraz odmowa, mają swoje uwarunkowania. Brak kontaktu jest powiązany z cechami gospodarstw domowych określających ich status społeczno-ekonomiczny. Należą do nich płeć i wiek, stan cywilny, poziom wykształcenia status na rynku pracy głowy gospodarstwa domowego oraz miejsce zamieszkania, wielkość gospodarstwa domowego, a także główne źródło utrzymania. Z kolei odmowa udziału w badaniu jest powiązana z cechami gospodarstw domowych określających ich środowisko. Należą do nich miejsce zamieszkania, wielkość gospodarstwa domowego i główne źródło utrzymania. Ponadto, jeśli brak dostępności i brak gotowości współpracy posiadają wspólne istotne predyktory, to ich poszczególne kategorie wywołują odmienne efekty.

Analiza czynników leżących po stronie jednostek populacji, które determinują brak kontaktu oraz brak współpracy, wykazała zatem, że reguły uczestnictwa w badaniu różnicują się w podpopulacjach badanych gospodarstw domowych wyznaczonych przez konfigurację tych czynników. Natomiast odmiennosć uwarunkowań dostępności i gotowości współpracy dowodzi, iż stanowią one odrębne wymiary, charakteryzujące

odrębne własności badanych jednostek. Mają one charakter ukryty, a obserwowane są jedynie ich efekty w postaci konkretnych zachowań respondentów przy nawiązaniu kontaktu oraz przy próbie podejmowania współpracy. W tym sensie mogą być rozważane jako zmienne opisujące wymiary w przestrzeni postaw, na których lokują się respondenci w zależności od ich natężenia. Celem przeprowadzonej analizy było rozpoznanie uwarunkowań każdej z nich, a wyniki potwierdziły ich odrębność.

Ujawniono również efekty związane z warunkami realizacji badania w terenie. Model hybrydowy pozwolił ujawnić efekty interakcji między lokalizacją gospodarstwa domowego a sposobem prowadzenia badania w terenie na kształtowanie prawdopodobieństwa braku odpowiedzi, co wskazuje, że różnicowanie zasad realizacji badań w terenie może poprawić ich efektywności.

LITERATURA

- Brehm J., (1993), *The Phantom Respondents: Opinion Surveys and Political Representation*, Ann Arbor, University of Michigan Press.
- De Heer W., (1999), International Response Trends, Results of International Survey, *Journal of Official Statistics*, 15 (2), 129–142.
- Dillman D. A., (2000), *Mail and Internet Surveys: The Tailored Design Method*, 2nd edition, New York, Wiley.
- Goyder J., (1987), Surveys on Surveys: Limitation and Potentialities, *Public Opinion Quarterly*, 50, 27–41.
- Groves R. M., (2006), Nonresponse Rates on Nonresponse Bias In Household Surveys, *Public Opinion Quarterly*, 70, 646–675.
- Groves R. M., Couper M. P., (1998), *Noresponse in Household Interview Surveys*, New York, Wiley.
- Groves R. M., Singer E., Corning A., (2000), Leverage-Saliency Theory of Survey Participation. Description and Illustration, *Public Opinion Quarterly*, 64, 299–308.
- Lin I-F., Schaeffer N. C., (1995), Using Survey Participants to Estimate the Impact of Nonparticipation, *Public Opinion Quarterly*, 59, 236–258.
- Lynn P., (2002), PEDAISOIS: Methodology for Collecting Data about Survey Non-respondents, *Working Papers of the Institute for Social and Economic Research*, paper 2002-05, Colchester: University of Essex, www.iser.essex.ac.uk/pubs/workpaps/isr/2002-05.pdf
- Martin J., Matheson J., (1999), Responses to Declining Response Rate on Government Surveys, *Survey Methodology Bulletin*, 45, 33–37.
- Paradysz J., Szymkowiak M., (2007), Imputacja i kalibracja jako remedium na braki odpowiedzi w badaniu budżetów gospodarstw domowych, *Taksonomia*, 14, *Klasyfikacja i analiza danych – teoria i zastosowania*, 74–81.
- Piasecki T., (2014), Metody imputacji w badaniach gospodarstw domowych, *Wiadomości Statystyczne*, 9, 1–20.
- Rószkiewicz M., (2015), Milcząca większość. Uwarunkowania poziomu wskaźnika braku odpowiedzi w środowisku gospodarstw domowych w Polsce w 2013 r. Próba diagnozy, *PN Taksonomia*, 24, *Klasyfikacja i analiza danych – teoria i zastosowania*.
- Schnell R., (1997), Nonresponse in Bevölkerungsumfragen, Opladen (Leske + Budrich).
- Steeh Ch., Kirgies N., Cannon B., De Witt J., (2001), Are They Really as Bad as They Seem? Nonresponse Rate at the End of Twentieth Century, *Journal of Official Statistics*, 17 (2), 227–247.
- Stinchcombe A. L., Jones C., Sheatsley P., (1981), Nonresponse Bias for Attitude Questions, *Public Opinion Quarterly*, 45, 359–375.
- Stoop I. A. L., (2005), *The Hunt for the Last Respondent. Nonresponse in Sample Surveys*, Social and Cultural Planning Office of the Netherlands, The Hague.

- Sztompka P., (1999), Stawanie się społeczeństwa: pomiędzy strukturą a zmianą, w: Kurczewska J., (red.), *Zmiana społeczna*, 39–54.
- Żyra M., Warno K., (2013), Braki odpowiedzi w badaniach gospodarstw domowych w Polsce, http://stat.gov.pl/cps/rde/xbcr/lodz/ASSETS_Konferencja_MRS_2013_Zyra_M_Warno_K_Braki_odpowiedzi_w_badaniach_gospodarstw_domowych_w_Polsce.pdf

IDENTYFIKACJA DETERMINANT BRAKU ODPOWIEDZI W BADANIU POLSKICH GOSPODARSTW DOMOWYCH

Streszczenie

W artykule podjęto próbę identyfikacji czynników mających wpływ na poziom wskaźnika braku odpowiedzi w badaniu polskich gospodarstw domowych zrealizowanym w 2013 r. w ramach projektu Uwarunkowania Decyzji Edukacyjnych. Podstawą analizy były wyniki historii kontaktów z wylosowanymi jednostkami oraz wyniki zrealizowanych wywiadów. Odrębnie przeanalizowano uwarunkowania braku kontaktu oraz odmowy odpowiedzi. W identyfikacji czynników związanych z warunkami realizacji badania istotnie wpływających na prawdopodobieństwo braku odpowiedzi wykorzystano model hybrydowy łączący wielomianowy model regresji logistycznej oraz procedurę drzewa klasyfikacyjnego. Model regresji logistycznej wykorzystano również do identyfikacji cech gospodarstw domowych wpływających na brak odpowiedzi. Wyniki wskazały, że brak kontaktu i brak odpowiedzi stanowią odrębne wymiary, na których lokują się jednostki wybrane do badania, a reguły uczestnictwa w badaniu różnicują się względem cech społeczno-demograficznych. Skuteczna organizacja procesu badawczego powinna uwzględniać zarówno zróżnicowanie regionalne w dostępności i gotowości współpracy jak również w preferencjach respondentów względem sposobu nawiązywania kontaktów.

Słowa kluczowe: wskaźnik braku odpowiedzi, błąd braku reakcji, błąd braku realizacji, model regresji logistycznej, procedura drzewa klasyfikacyjnego, model hybrydowy

AN ATTEMPT TO DIAGNOSE DETERMINANTS OF NON-RESPONSE RATE IN POLISH HOUSEHOLDS SURVEYS

Abstract

Article focuses on the identification of factors affecting the non-response in Polish household surveys. The analyse uses data from the survey realized on a random sample of Polish households in 2013 in project Determinant of Educational Decisions. Logistic regression model and classification tree procedure and hybridization of this approach was used to identify factor affecting probability of non-responses. Noncontact and noncooperation in the study was considered separately. Results confirmed that noncontact and noncooperation are two entirely different processes and rules for participation in the study significantly differentiate into subpopulations of Polish households varied by socio-economic features. Efficient organization of the research process should take into account both regional differences in the availability and willingness of cooperation as well as the respondents' preferences in regarding the way in making contacts.

Keywords: no-response rate, noncontact, noncooperation, logistic regression, classification tree procedure, hybrid model