

Małgorzata Rószkiewicz

Szkoła Główna Handlowa w Warszawie
e-mail: mroszki@sgh.waw.pl

WNIOSKOWANIE O WPLYWIE PROCESU BADAWCZEGO NA POZIOM WSKAŹNIKA ODPOWIEDZI

AN ATTEMPT TO IDENTIFY THE IMPACT OF RESEARCH PROCESS ON THE LEVEL OF RESPONSE-RATE

DOI: 10.15611/pn.2017.468.20

JEL Classification: C 83

Streszczenie: W artykule podjęto próbę identyfikacji wpływu procesu badawczego na gotowość współpracy z ankierem wśród polskich gospodarstw domowych. Podstawą analizy były wyniki historii kontaktów z jednostkami wylosowanymi do badania ankietowego, zrealizowanego w 2013 r. na losowej próbie ponad 34 tys. gospodarstw domowych. Za efekt procesu badawczego przyjęto liczbę odmów odpowiedzi na stawiane w kwestionariuszu pytania. Za zmienną zależną przyjęto liczbę prób nakłaniania do współpracy podjętą przez ankiera, zanim zrealizowano wywiad. Wykorzystano podejście wielopoziomowe, w którym jako zmienną niezależną drugiego poziomu uwzględniono status społeczno-ekonomiczny respondenta. Wyniki wskazały, że wprawdzie wyższy status społeczno-ekonomiczny ograniczał gotowość do współpracy z ankierem, to jednak łagodził jej wrażliwość na negatywny wpływ procesu badawczego.

Słowa kluczowe: wskaźnik braku odpowiedzi, błąd braku realizacji, efekt procesu badawczego, podejście wielopoziomowe, wielowymiarowa analiza korespondencji.

Summary: The article presents an attempt to identify the effect of the research process on the level of the response-rate in the study of Polish households conducted in 2013. The number of refusals to answer questions posed in the questionnaire was assumed as the effect of the research process and the number of attempts to urge the cooperation undertaken by the interviewer before the interview was carried out was assumed as the dependent variable. Multilevel approach has been applied, in which the socio-economic status of the respondent was taken as the independent variable on second level. The results indicated that while the higher socio-economic status enforced additional number of attempts to cooperate by the interviewer, it softened its sensitivity to the effect of the research process.

Keywords: nonresponse-rate, noncooperation-rate, effect of research process, multilevel approach, multivariate correspondence analysis.

1. Wstęp

Teorie wyjaśniające uczestnictwo w badaniach terenowych wyróżniają wśród czynników dostępności i gotowości współpracy jednostek próby 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. Założone w tych teoriach uwarunkowania przy czynowe pozwalają problem braków odpowiedzi traktować jako proces społeczny [Lin, Schaeffer 1995; Groves 2006]. Sugerowany zaś 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 badania obsługujących określone wiązki jednostek, zaś na trzecim cechy samego procesu badawczego, który przebiega w określony sposób i dotyczy wybranej tematyki. Zrealizowane do tej pory analizy dla zbiorowości polskich gospodarstw domowych pozwoliły rozpoznać zakres oddziaływania cech deskryptywnych jednostek badania na wskaźnik odpowiedzi oraz cech leżących po stronie organizacji pomiaru w terenie, z oddziaływaniem ankierów włącznie. Uzyskane wyniki wskazywały, że czynniki pierwszego poziomu różnicujące udział w badaniu są związane ze statusem społeczno-ekonomicznym jednostek próby [Rószkiewicz 2015, 2016]. Potwierdzono również występowanie efektu ankiera obsługującego określone wiązki jednostek badania. Obecnie poddano weryfikacji hipotezę o interakcji między uwarunkowaniami pierwszego i trzeciego poziomu i na tej podstawie podjęto próbę oceny siły efektu cech procesu badawczego, związanego z tematyką badania, w kształtowaniu gotowości współpracy z ankierem, z uwzględnieniem zróżnicowania tego efektu względem statusu społeczno-ekonomicznego badanych jednostek.

2. Dane i metoda

Podstawą analizy były wyniki historii kontaktów z jednostkami wylosowanymi do badania ankietowego zrealizowanego przez Instytut Badań Edukacyjnych w 2013 r. w ramach projektu Uwarunkowania Decyzji Edukacyjnych na losowej próbie ponad 34 tys. gospodarstw domowych.

Zmienna objaśniana w rozważanym modelu zależności, odnosząca się do skłonności respondenta do współpracy, została zoperacjonalizowana jako liczba wizyt ankiera niezbędna, by zrealizować wywiad. Zmienne niezależne zaś dotyczyły statusu społeczno-ekonomicznego respondenta oraz efektu tematyki badania. Przy czym, ponieważ wpływ statusu społeczno-ekonomicznego na wskaźnik odpowiedzi został już wykazany we wcześniejszych opracowaniach [Rószkiewicz 2015, 2016], obecnie skoncentrowano się na efekcie interakcji statusu społeczno-ekonomiczne-

go z efektem tematyki badania, zakładając, że reakcja respondentów na tematykę badania różnicuje się względem ich statusu społeczno-ekonomicznego. Założono zatem relacje dwupoziomowe, gdzie status społeczno-ekonomiczny jest moderatorem zależności gotowości współpracy respondenta względem odbioru tematyki badania. Nie dysponując wynikami innych badań ankietowych o odmiennej tematyce, co uniemożliwiło zastosowanie wprost analizy porównawczej, za efekt tematyki badania przyjęto liczbę odmów odpowiedzi na pytania zadawane według kwestionariusza ankietowego, co wyznaczało zakres kompletności wypełnionej ankiety. Założono zatem, że respondenci bardziej skłonni do współpracy ze względu na tematykę badania w ogóle lub bardzo rzadko odmawiają odpowiedzi na poszczególne pytania ankiety, zaś respondenci mniej skłonni do współpracy ze względu na tematykę badania częściej omijają pytania, odmawiając odpowiedzi. W tym sensie podjęto próbę jedynie pośredniej oceny efektu cech procesu badawczego związanego z tematyką badania w kształtowaniu gotowości współpracy z ankierem. Takie postępowanie było podyktowane również tym, że w przypadku definitywnej odmowy udziału w badaniu nie udało się ustalić przyczyn, co uniemożliwiło wyodrębnienie grupy, która w takim przypadku kierowałaby się tematyką badania.

Za wskaźnik statusu społeczno-ekonomicznego przyjęto jednowymiarową zmienną¹ agregującą poziom dochodu ekwiwalentnego na 1 osobę w gospodarstwie, poziom wykształcenia i pozycję społeczno-zawodową głowy gospodarstwa domowego. Zmienna skonstruowana w procedurze wielowymiarowej analizy korelacji charakteryzowała się dostateczną jakością i posiadała dobre własności w diagnozowaniu tego statusu². Posłużenie się zmiennymi ilościowymi stworzyło możliwość modelowania zależności gotowości współpracy z ankierem względem obu rozważanych czynników według formuły dwupoziomowego modelu regresji.

Model wielopoziomowy jest uzasadniony, jeśli dane mają strukturę hierarchiczną w tym sensie, że wartości w ramach występujących klas/grup jednostek są mniej zróżnicowane niż pomiędzy nimi, co podważa założenie o niezależności obserwacji. Zakres występujących wówczas zależności może być wyrażony współczynnikiem korelacji międzyklasowej (międzygrupowej) [Goldstein 1999; Kret, de Leeuw 1998; Snijders, Bokser 1994]. W modelu regresji wielopoziomowej, podobnie jak w modelu regresji prostej, przyjmuje się założenia o normalności i liniowości.

W modelu dwupoziomym od grupy/klasy zależy zarówno poziom zmiennej zależnej, jak i wpływ zmiennej niezależnej na zmienną zależną. Wyjaśnienie tego

¹ Przyjęto zasady konstrukcji zmiennej skalującej status społeczno-ekonomiczny według podejścia zaproponowanego przez Górniaka [Górniak 2000, s. 79-93].

² Współczynnik inercji przyjął wartość 36,1%. Odnotowano wyższe wartości średnie zmiennej skalującej status społeczno-ekonomiczny w grupach gospodarstw domowych o relatywnie wyższym średnim miesięcznym dochodzie na osobę oraz o wyższym poziomie wykształcenia głowy gospodarstwa domowego, a także w takich grupach gospodarstw, których głowa pełni funkcje kierownicze, jest samodzielnym specjalistą lub utrzymuje się z pracy na rachunek własny. W każdym przypadku różnice między średnimi, testowane w procedurze analizy wariancji, okazały się istotne dla $p < 0,001$.

wpływu odbywa się za pomocą zmiennej niezależnej obserwowanej jedynie na poziomie grup/klas. Rozważane zależności można zapisać w postaci układu równań, w którym pierwsze równanie jest modelem zależności zachodzących na pierwszym poziomie, zaś pozostałe dwa określają zależności na poziomie drugim. W rozważanym tu modelu zdefiniowano po jednej zmiennej niezależnej na każdym poziomie, tzn. założono, że zmienną niezależną pierwszego poziomu jest liczba odmów odpowiedzi na pytania zawarte w kwestionariuszu ankietowym przez respondenta, zaś zmienną niezależną drugiego poziomu jest jego status społeczno-ekonomiczny, co definiuje następujący układ równań:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij}, \quad (1)$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + u_{0j}, \quad (2)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + u_{1j}, \quad (3)$$

gdzie: β_{0j} , β_{1j} – parametry (odpowiednio wyraz wolny oraz współczynnik regresji) modelu pierwszego poziomu regresji liniowej, uzależniający zmienną Y od zmiennej pierwszego poziomu X ; e_{ij} – składnik losowy modelu pierwszego poziomu, o którym zakłada się, że jego wartość średnia wynosi zero i ma stałą wariancję w każdej klasie; γ_{00} , γ_{01} – parametry (odpowiednio wyraz wolny oraz współczynnik regresji) modelu regresji liniowej wyrazu wolnego z modelu pierwszego poziomu względem zmiennej drugiego poziomu Z ; γ_{10} , γ_{11} – parametry (odpowiednio wyraz wolny oraz współczynnik regresji) modelu regresji liniowej współczynnika regresji z modelu pierwszego poziomu względem zmiennej drugiego poziomu Z ; u_{0j} , u_{1j} – składniki losowe równań modelu drugiego poziomu.

Powyższy model można również przedstawić w postaci jednego równania:

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}X_{ij}Z_j + u_{1j}X_{ij} + u_{0j} + e_{ij}. \quad (4)$$

Zakłada się, że składniki losowe u_{0j} oraz u_{1j} posiadają rozkłady o wartości średniej równej zero i wariacjach równych odpowiednio σ_{u0}^2 i σ_{u1}^2 oraz kowariancję $\sigma_{u0,u1}$ różną od zera. Są one niezależne względem składnika losowego e_{ij} . Składnik modelu $(\gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}X_{ij}Z_j)$ wyraża warunki (parametry) ustalone i jest określany mianem ustalonej lub zdeterminowanej części modelu. Składnik pozostały $(u_{1j}X_{ij} + u_{0j} + e_{ij})$ wyraża warunki zakłóceń losowych i jest określany mianem losowej lub stochastycznej części modelu. Składnik $(X_{ij}Z_j)$ odnosi się do warunku interakcji, która ujawnia się zmiennym współczynnikiem regresji zmiennej Y względem zmiennej X ze względu na warunki określone przez zmienną Z . Jest to interakcja międzypoziomowa wyrażana równaniem prostej regresji jednej zmiennej.

Najpopularniejszą metodą estymacji parametrów modelu wielopoziomowego jest metoda największej wiarygodności. Estymatory uzyskane tą metodą są asymptotycznie efektywne. Dla dużych prób pozwalają utrzymać założenia o normalności. W przypadku pełnej metody największej wiarygodności (FML) współczynniki re-

gresji traktuje się jako ustalone (stałe), lecz o nieznanych wartościach, zaś składniki wariancji są przedmiotem estymacji. W przypadku ograniczonej metody największej wiarygodności (RML) oszacowanie składników wariancji ma miejsce po usunięciu efektów stałych z modelu [Hox 2002; Snijders, Bokser 1994].

3. Wyniki

W pierwszej kolejności poddano weryfikacji założenie, iż oddziaływanie tematyki badania oraz statusu społeczno-ekonomicznego respondentów ma charakter hierarchiczny w tym sensie, że struktura zależności skłonności respondentów do współpracy względem tematyki badania jest dwupoziomowa i ulega modyfikacji względem statusu społeczno-ekonomicznego jednostki.

Wartość współczynnika korelacji międzyklasowej, przy założeniu, że grupa/klasa wyznacza status społeczno-ekonomiczny jednostek, oszacowana w procedurze tzw. modelu zerowego regresji wielopoziomowej na poziomie 0,24198, potwierdziła założenie o hierarchicznej strukturze danych. Uzasadnieniem wykorzystania modelu dwupoziomowego, uwzględniającego efekty oddziaływania statusu społeczno-ekonomicznego na gotowość współpracy respondenta z ankieterem, były wartości wskaźników dopasowania modelu do danych, wyznaczonych jako kryteria informacyjne (AIC, AICC, CAIC i BIC, a także podstawowa dla ich wartości wielkość $(-2 \times \logarytm \text{ funkcji wiarygodności})$), które dla modelu jednopozomowego kształtowały się na znacznie wyższym poziomie. Różnice w dopasowaniu na korzyść modelu dwupoziomowego wskazują na występowanie zmienności międzygrupowej zmiennej zależnej. Wartości kryteriów informacyjnych wykorzystanych w ocenie jakości dopasowania modelu do danych dla poszczególnych modeli dwupoziomowych zbudowanych przy odpowiednich założeniach odnoszących się do ich parametrów wraz z ocenami tych parametrów zestawia tabela 1.

Tabela 1. Wartości kryteriów informacyjnych wykorzystanych w ocenie jakości dopasowania modelu do danych poszczególnych modeli dwupoziomowych zbudowanych przy odpowiednich założeniach odnoszących się do ich parametrów oraz oceny tych parametrów uzyskane na podstawie wyników próby losowej

Wyszczególnienie	Model jedynie z wyrazem wolnym, tzw. model zerowy	Model regresji jednopozomowy	Model jedynie z losowym wyrazem wolnym	Model z losowym wyrazem wolnym i losowym współczynnikiem regresji bez interakcji	Model z losowym wyrazem wolnym i losowym współczynnikiem regresji z efektem interakcji
1	2	3	4	5	6
Kryterium informacyjne					
$-2 \logarytm$ wiarygodności	35 823,451	39 049,652	35 782,790	35 011,845	34 989,377

1	2	3	4	5	6
Kryterium informacyjne Akaike (AIC)	35 829,451	39 055,652	35 790,790	35 021,845	35 001,377
Kryterium Hurvicha i Tsaia (AICC)	35 829,452	39 055,652	35 790,791	35 021,847	35 001,380
Kryterium Bozdogana (CAIC)	35 857,218	39 083,418	35 827,812	35 068,123	35 056,910
Kryterium bayesowskie Schwarza (BIC)	35 854,218	39 080,418	35 823,812	35 063,123	35 050,910
Oceny parametrów i ich błędy szacunku					
Wyraz wolny modelu	1,198281*	1,219244*	1,195111*	1,197884*	1,199490*
Błąd standardowy	0,009354	0,002851	0,009377	0,009339	0,009140
Współczynnik regresji: odmowy odpowiedzi		0,063695*	0,063418*	0,056498*	0,063871*
Błąd standardowy		0,009422	0,009941	0,025577	0,025488
Ocena parametrów wariancji: składnik drugiego poziomu dla stałej modelu	0,062151*		0,062344*	0,058134*	0,053003*
Błąd standardowy	0,003955		0,003962	0,003805	0,003731
Ocena parametrów wariancji: składnik drugiego poziomu dla współczynnika regresji				0,155837*	0,200250*
Błąd standardowy				0,017496	0,025929
Interakcja					-0,044886*
Błąd standardowy					0,009535
Reszta	0,194693*	0,231108*	0,194392*	0,186290*	0,186258*
Błąd standardowy	0,001663	0,001938	0,001660	0,001598	0,001597

* Parametr istotny przy poziomie $< 0,01$.

Źródło: opracowanie własne.

W każdym z oszacowanych modeli wyraz wolny wyraża średnią liczbę prób podejmowania współpracy przez ankietera. Różnice między wartościami wyrazu wolnego oraz współczynnika regresji dla modelu jednopozomowego oraz modeli zakładających relacje międzypozomowe wynikają z przyjętych w tych modelach założeń o hierarchicznej strukturze danych.

4. Wnioski

Najwyższe dopasowanie do danych empirycznych i istotne parametry uzyskano w formule modelu dwupozomowego z losowym wyrazem wolnym i losowym współczynnikiem regresji, przy założeniu skorelowania współczynnika regresji z wyrazem

wolnym. Oznacza to, że status społeczno-ekonomiczny głowy gospodarstwa domowego miał nie tylko wpływ na średni poziom gotowości podjęcia współpracy z ankieterem, ale również na jej zmienność. Wyniki potwierdziły występowanie istotnie stymulującego efektu pierwszego oraz drugiego poziomu, jak i efekt interakcji międzypoziomowej, który okazał się istotnie ujemny. Ponieważ efekty losowe opisywane są przez zmienne, co do których przyjmuje się, że ich wartość średnia wynosi 0, a ich odchylenia standardowe opisywane są przez przeciętne odchylenia parametrów modelu w zależności od przynależności do klasy definiowanej przez status społeczno-ekonomiczny, to uzyskane wyniki wskazują, że wzrost statusu społeczno-ekonomicznego o jedno odchylenie standardowe powoduje wzrost średniej liczby prób podjęcia współpracy średnio o 0,230224, czyli ponad trzyipółkrotnie względem średniego poziomu opisanego wyrazem wolnym, oraz wzrost efektu projektu badawczego średnio o 0,447493, czyli o 37,3% względem średniego efektu opisanego współczynnikiem regresji. Natomiast ujemna istotna kowariancja odnosząca się do zależności wpływów statusu społeczno-ekonomicznego i efektów procesu badawczego na gotowość współpracy respondenta z ankieterem wskazuje, że prawidłowości te są łagodzone interakcją relacji międzypoziomowych. Wprawdzie wyższy status społeczno-ekonomiczny zwiększa konieczną liczbę prób podejmowania współpracy, to obniża jej wrażliwość na negatywny efekt procesu badawczego. Wyniki wskazują zatem, że im wyższy był status społeczno-ekonomiczny respondentów, tym słabsza była stymulacja liczby prób podejmowania współpracy przez efekt procesu badawczego.

Literatura

- Goldstein H., 1999, *Multilevel Statistical Models*, Institute of Education, London.
- Górnjak J., 2000, *My i nasze pieniądze*, AUREUS, Kraków.
- Groves R.M., 2006, *Nonresponse rates on nonresponse bias in household surveys*, Public Opinion Quarterly, vol. 70, s. 646-675.
- Hox J., 2002, *Multilevel Analysis. Technique and Applications*, Lawrance Erlbaum Associates, Publishers, Mahwah-London.
- Kret I., de Leeuw J., 1998, *Introduction Multilevel Modeling*, SAGE Publication Ltd., London-Thousand Oaks-New Delhi.
- Lin I.-F., Schaeffer N.C., 1995, *Using survey participants to estimate the impact of nonparticipation*, Public Opinion Quarterly, vol. 59, s. 236-258.
- Rószkiewicz M., 2015, *Milcząca większość. Uwarunkowania poziomu wskaźnika braku odpowiedzi w środowisku gospodarstw domowych w Polsce w 2013 r. Próba diagnozy*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, nr 385, Taksonomia 24, s. 219-228.
- Rószkiewicz M., 2016, *Czynniki różnicujące efektywność pracy ankietera w wywiadach face-to-face w środowisku polskich gospodarstw domowych*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, nr 426, Taksonomia 26, s. 166-173.
- Snijders T.A.B., Bokser R.J., 1994, *Modeled variance in two-level models*, Sociological Methods & Research, no. 22, s. 342-363.